

JOINT INSTITUTE FOR NUCLEAR RESEARCH
LABORATORY OF INFORMATION TECHNOLOGY

DISTRIBUTED COMPUTING AND GRID-
TECHNOLOGIES IN SCIENCE
AND EDUCATION

Book of abstract of the 8th International Conference

Dubna, 10 – 14 September, 2018

РАСПРЕДЕЛЕННЫЕ ВЫЧИСЛЕНИЯ
И ГРИД – ТЕХНОЛОГИИ В НАУКЕ
И ОБРАЗОВАНИИ

Тезисы докладов 8 - й Международной конференции

Дубна, 10 – 14 сентября, 2018

УДК[004.7+004.9](063)
ББК[32.988-5я431+32.97я431]
D63

Organized by the Joint Institute for Nuclear Research,
Laboratory of Information Technologies
under the Sponsorship of the
IBS Platformix, Jet Infosystems, SoftLine, NIAGARA,
Dell, Intel, Huawei, RSC Group,
Cisco, IBM, Schneider Electric, Nvidia

Distributed Computing and Grid-Technologies in Science and Education: Book of Abstr. Of the 8th Intern. Conf. (Dubna, 10–14 September, 2018).–Dubna: JINR, 2018.- 174 p.
ISBN -5-9530-

The book contains abstracts of the reports submitted to the 8th International Conference “Distributed Computing and Grid-Technologies in Science and Education” (GRID 2018), which is held every two years by the JINR Laboratory of information technologies.

The main purpose of the Conference is to discuss the current Grid operation and the future role of the distributed and cloud computing, HPC, BigData etc. in Russia and worldwide. The Conference provides a platform for discussing fresh results and for establishing contacts for closer cooperation in future.

The submitted abstracts are described the development of the global grid-infrastructures and such thriving technologies of intensive data processing as distributed, heterogeneous, cloud computing, BigData, etc. Full texts of the reports approved by the GRID2018 Advisory Committee will be published separately.

Распределенные вычисления и Грид-технологии в науке и образовании: Тез. докл. 8-й междунар. конф. (Дубна, 10 – 14 сентября, 2018 г.).- Дубна: ОИЯИ, 2018. – 174 с.
ISBN -5-9530-

В сборник включены аннотации докладов, представленных на 8-ю международную конференцию «Распределенные вычисления и Грид-технологии в науке и образовании» (GRID 2018), которая проводится раз в два года Лабораторией информационных технологий ОИЯИ.

Основной целью конференции является обсуждение текущего состояния и будущего развития распределенных вычислений, грид - и облачных технологий, высокопроизводительных вычислений, технологий BigData в России и во всем мире.

Представленные аннотации описываются развитие глобальной Грид-инфраструктур и бурно развивающиеся технологии интенсивной обработки данных, такие как распределенные, гетерогенные вычисления, облачные вычисления, BigData т.д. Полные тексты докладов, отобранных программным комитетом, будут опубликованы отдельно.

ISBN -5-9530-

УДК[004.7+004.9](063)
ББК[32.988-5я431+32.97я431]
©Joint Institute for Nuclear research, 2018

Advisory committee:

Afanasiev A.P.	IITP RAS, Moscow
Andreeva J.	IT CERN, Switzerland
Avetisyan A.I.	Institute for System Programming of the RAS, Moscow
Bird I.	IT CERN, Switzerland
Bogdanov A.V.	SPbSU, St.Petersburg
Brun R.	PH CERN, Switzerland
Buzatu F.D.	IFA, Magurele-Bucharest, Romania
Chetverushkin B.N.	Keldysh Institute of Applied Mathematics, Moscow
Dubnicka S.	IoP SAS, Bratislava, Slovakia
Fuhrmann P.	DESY, Hamburg, Germany
Heiss A.	KIT, Karlsruhe, Germany
Ilyin V.A.	NRC "Kurchatov Institute", Moscow
Kaushik De	University of Texas at Arlington, USA
Klimentov A.	BNL, USA
Kostov L.	Bulgarian Nuclear Regulatory Agency, Republic of Bulgaria
Kryukov A.P.	SINP Moscow State University
Lakhno V.D.	IMPB Russian Academy of Sciences
Legre Y.	CNRS National Grid & Cloud Institute, France
Levai P.	Wigner RCP, Hungary
Li W.	IHEP CAS, China
Lin S.C.	Academia Sinica, Taipei, Taiwan
Lippert Th.	Jülich Supercomputing Centre, Germany
Lokajicek M.	PI AS CzR, Czech Republic
Mamedov N.T.	Institute of Physics of Azerbaijan National Academy of Sciences, Azerbaijan
Maron G.	CNAF INFN, Bologna, Italy
Matveev V.A.	JINR
Livny M.	Condor project, USA
Prangishvili A.	GTU, Tbilisi, Georgia
Sahakyan V.G.	IIAP NAS Armenia
Sithole H.	Centre for High Performance Computing, South Africa
Smirnova O.G.	NDGF/University of Lund, Sweden
Sokolov I.A.	Federal Research Center "Informatics and Management", Moscow
Tsaregorodtsev A.	CPPM-IN2P3-CNRS, France
Valentey S.D.	PRUoE, Moscow
Velikhov V.E.	NRC "Kurchatov Institute", Moscow
Voevodin V.V.	SRCC Moscow State University
Volkov D.V.	Open System Publication, Moscow
Zinovjev G.	BITP, Kiev, Ukraine
Zolotarev V.	CC SPbU, St.Petersburg

Програмный комитет:

Аветисян А.И.	Институт системного программирования РАН, Москва, Россия
Андреева Ю.	ЦЕРН
Афанасьев А.П.	ИЦМС ИПС РАН, Москва, Россия
Берд Я.	ЦЕРН
Богданов А.В.	ИВВиБД, Россия
Брун Р.	ЦЕРН
Бузату Ф.Д.	IFA, Magurele-Bucharest, Румыния
Валентей С.Д.	РЭУ им. Г.В. Плеханова, Москва
Велихов В.Е.	НИЦ «Курчатовский институт», Москва, Россия
Воеводин В.В.	НИВЦ МГУ, Москва, Россия
Волков Д.В.	Издательство «Открытые системы», Москва, Россия
Дубничка С.	Институт физики САН, Братислава, Словакия
Зиновьев Г.	ИТФ, Киев, Украина
Золотарев В.	КЦ СПбСУ, Санкт-Петербург
Ильин В.А.	НИИЯФ МГУ / НИЦ «Курчатовский институт», Москва
Каушик Де.	Университет Техаса в Арлингтоне, США
Клементов А.	БНЛ, США
Костов Л.	Болгарское агентство по ядерному регулированию, Болгария
Крюков А.П.	НИИЯФ МГУ, Москва, Россия
Лакно В.Д.	ИМПБ РАН, Россия
Легре Я.	CNRS National Grid & Cloud Institute, Франция
Левай П.	ИЦФ им. Ю.П. Вигнера, Венгрия
Ливный М.	Condor project, США
Ли В.	CAS IHEP, Китай
Лин С.К.	Академия Синика, Тайбэй, Тайвань
Липперт Т.	Суперкомпьютерный центр, Юлих, Германия
Локайчик М.	Институт физики, Чешская республика
Мамедов Н.Т.	Институт физики НАНА, Азербайджан
Марон Г.	CNAF INFN, Болонья, Италия
Матвеев В.А.	ОИЯИ
Прангишвили А.	ГТУ, Тбилиси, Грузия
Саакян В.Г.	ИПИА, Армения
Сизоле Х.	Centre for High Performance Computing, Южная Африка
Смирнова О.Г.	NDGF/Университет Лунда, Швеция
Соколов И.А.	ФИЦ ИУ РАН, Москва
Фурман П.	DESY, Гамбург, Германия
Хейс А.	KIT, Карлсруе, Германия
Царегородцев А.	SRPM-IN2P3-CNRS, Франция
Четверушкин Б.Н.	ИПМ им М.В. Келдыша, Москва, Россия

Organizing Committee (JINR):

Korenkov V.V. - Chairman
Strizh T.A. - Vice-Chairman
Podgainy D.V. - Scientific Secretary
Adam S.

Aristarkhova M.V.

Belov O.V.

Grafov A.N.

Katraseva T.I.

Korotchik O.M.

Kutovsky N.A.

Lukyanov S.O.

Mazhitova E.

Nechaevskiy A.V.

Pelevanyuk I.S.

Rumyantseva O.Yu.

Stankus D.B.

Streltsova O.I.

Torosyan Sh.G.

Voytishin N.N.

Vorontsov A.S.

Zaikina A.G.

Zaikina T.N.

Zrelov P.V.

Zuev M.I.

Организационный комитет (ОИЯИ):

Кореньков В.В. – председатель

Стриж Т.А. – зам. председателя

Подгайный Д.В. – ученый секретарь

Адам С.

Аристархова М.В.

Белов О.В.

Войтишин Н.Н.

Воронцов А.С.

Графов А.Н.

Заикина А.Г.

Заикина Т.Н.

Зрелов П.В.

Зуев М.И.

Катрасева Т.И.

Коротчик О.М.

Кутовский Н.А.

Лукьянов С.О.

Мажитова Е.

Нечаевский А.В.

Пелеванюк И.С.

Румянцева О.Ю.

Станкус Д.Б.

Стрельцова О.И.

Торосян Ш.Г.

CONTENTS

PLENARY REPORTS

Supercomputer "GOVORUN" — new prospects for heterogeneous computations at JINR.....	21
<i>Gheorghe Adam, Vladimir Korenkov, Dmitry Podgainy, Oksana Streltsova</i>	
CRIC: the information system for LHC Distributed Computing.....	22
<i>Alexey Anisenkov</i>	
The CMS Tier1 at JINR: five years of operations.....	23
<i>Andrey Baginyan, Anton Balandin, Sergey Belov, Andrey Dolbilov, Alexey Golunov, Natalia Gromova, Ivan Kadochnikov, Ivan Kashunin, Vladimir Korenkov, Valery Mitsyn, Igor Pelevanyuk, Sergey Shmatov, Tatiana Strizh, Vladimir Trofimov, Nikolay Voytishin, Victor Zhiltsov</i>	
The ATLAS EventIndex and its evolution based on Apache Kudu storage	24
<i>Dario Barberis</i>	
Big data as the future of information technology	25
<i>Alexander Bogdanov, Alexander Degtyarev, Vladimir Korkhov, Thurein Kyaw</i>	
THE designing of cloud infrastructure consisting of geographically distributed data centers	26
<i>Petr Fedchenkov, Nikita Samokhin, Sergey Khoruzhnikov, Oleg Lazo, Andrey Shevel</i>	
Cloud-based Computing for LHAASO experiment at IHEP	27
<i>Qiulan Huang, Weidong Li, Yaodong Cheng, Haibo Li, Tao Cui, Xiaowei Jiang</i>	
PIK Computing Centre	28
<i>Andrey Kiryanov</i>	
Real-time event reconstruction and analysis in the CBM experiment at FAIR using HPC	29
<i>Ivan Kisel (for the CBM collaboration)</i>	
BigPanDA Experience on Titan for the ATLAS Experiment at the LHC	30
<i>Alexei Klimentov, Kaushik De, Danila Oleynik, Ruslan Mashinistov, Jack Wells</i>	
JINR Multifunctional Information and Computing Complex: Status and Perspectives	31
<i>Vladimir Korenkov, Andrey Dolbilov, Valery Mitsyn, Tatiana Strizh, Nikolay Kutovskiy, Dmitry Podgainy, Oksana Streltsova, Vladimir Trofimov</i>	
Multicomponent cluster management system for the computing center at IHEP	32
<i>Anna Kotliar, Ekaterina Popova, Victoria Ezhova, Viktor Kotliar</i>	
Electronic, Dynamical and Thermodynamic Properties of DNA.....	33
<i>Victor Lakhno</i>	
File Transfer Service at Exabyte scale.....	34
<i>Andrea Manzi</i>	

Large scale simulations with parallel annealing algorithm	35
<i>Lev Shchur</i>	
Tier-1 centre at NRC "Kurchatov institute" between LHC Run2 and Run3 ...	36
<i>Igor Tkachenko, Eugene Ryabinkin, Alexander Rogovski, Ilya Lyalin</i>	
DIRAC services for scientific communities	37
<i>Andrei Tsaregorodtsev</i>	
Deep machine learning and pattern/face recognition based on quantum neural networks and quantum genetic algorithm.....	38
<i>Sergey Ulyanov, Andrey Reshetnikov, Nikita Ryabov</i>	
Advanced global network services to support research excellence.....	39
<i>Rudolf Vohnout, Vincenzo Capone</i>	
Организация доступа к экспериментальным данным установки ИТЭР в режиме удаленной пультовой	40
<i>Екатерина Миронова, Игорь Семенов, Олег Семенов, Сергей Портоне</i>	

TECHNOLOGIES, ARCHITECTURES, MODELS OF DISTRIBUTED COMPUTING SYSTEMS

ALICE DCS preparation for Run 3.....	43
<i>Alexander Kurepin, Andre Augustinus, Peter Chochula, Ombretta Pinazza, Mateusz Lechman, Peter Matthew Bond, Kevin Cifuentes Salas, John Larry Lang</i>	
The distributed grid site of Institute of Physics	44
<i>Alexandr Mikula, Martin Adam, Jiří Chudoba, Dagmar Adamová, Jana Uhlířová, Petr Horák, Petr Vokáč</i>	
INP BSU grid site.....	44
<i>Vladimir Mossolov, Dmitry Yermak, Vitaly Yermolchik</i>	
Properties of The Parallel Discrete Event Simulation Algorithms on Small- World Communication Networks	45
<i>Liliia Ziganurova, Lev Shchur</i>	
Грид и облачная инфраструктура дата-центра Института Физики НАН Азербайджана	45
<i>Алексей Бондяков</i>	

OPERATION, MONITORING, OPTIMIZATION IN DISTRIBUTED COMPUTING SYSTEMS

The BigPanDA monitoring system architecture	49
<i>Aleksandr Alekseev, Alexei Klimentov, Siarhei Padolski, Tatiana Korchuganova, Torre Wenaus (on behalf of the ATLAS Collaboration)</i>	
The BigPanDA self-monitoring alarm system for ATLAS	49
<i>Aleksandr Alekseev, Siarhei Padolski, Tatiana Korchuganova</i>	

Improving Networking Performance of a Linux Node	50
<i>Alexander Bogdanov, Nabil Ahmed, Vladimir Gaiduchok</i>	
Application of unified monitoring system in LHAASO	50
<i>Haibo Li, Qingbao Hu, Qiulan Huang, Wei Zheng, Xiaowei Jiang</i>	
BigData tools for the monitoring of the ATLAS EventIndex	51
<i>Andrei Kazymov, Evgeny Alexandrov, Fedor Prokoshin</i>	
Performance measurements for the WLCG Cost Model.....	51
<i>Victoria Matskovskaya</i>	
Search for Anomalies in the Computational Jobs of the ATLAS Experiment with the Application of Visual Analytics	52
<i>Alexei Klimentov, Igal Milman, Maria Grigorieva, Mikhail Titov, Tatiana Korchuganova, T. Galkin, Victor Pilyugin</i>	
Mechanisms for ensuring the integrity of information in distributed computing systems in the long-term period of time	53
<i>A.S. Minzov, A.Y. Nevskii, O.R. Baronov</i>	
Development of JINR Tier-1 service monitoring system	53
<i>Igor Pelevanyuk, Ivan Kadochnikov, Tatiana Strizh, Valery Mitsyn, Vladimir Korenkov</i>	
Trigger information data flow for the ATLAS EventIndex	54
<i>Fedor Prokoshin, Mikhail Mineev</i>	
Исследование особенностей Интернет-трафика в магистральном канале	54
<i>Иванов Виктор, Зрелов Петр, Крюков Юрий, Татаринов Иван, Иванов Валерий</i>	
Разработка перспективной системы сбора данных на основе TRB-3	55
<i>Александр Курепин, Андрей Кондратьев, Константин Сказыткин, Николай Курепин</i>	

MIDDLEWARE AND SERVICES FOR PRODUCTION- QUALITY INFRASTRUCTURE

Creating tools to assist in development of CMS software.....	59
<i>George Adamov</i>	
New methods of minimizing the errors in the software	59
<i>Elizaveta Dorenskaya</i>	
Current status of software development for the MPD and BM@N experiments	60
<i>Konstantin Gertsenberger</i>	
The ATLAS Production System Predictive Analytics service: an approach for intelligent task analysis	60
<i>Alexei Klimentov, Dmitry Golubkov, Mikhail Borodin, Mikhail Titov</i>	
Event building from free streaming data at the CBM	61
<i>Ivan Korolko, Mikhail Prokudin</i>	

Experience with ITEP-FRRC HPC facility.....	62
<i>Ivan Korolko, Mikhail Prokudin, Victor Kolosov</i>	
DDS – The Dynamic Deployment System	62
<i>Anar Manafov, Andrey Lebedev</i>	
The concept of proactive protection in a distributed computing system.....	63
<i>Anatoly Minzov, Pavel Osipov</i>	
COMPASS Production System: Processing on HPC	63
<i>Artem Petrosyan</i>	
Participation of Russian institutions in the processing and storage of ALICE data	64
<i>Andrey Zarochentsev</i>	

SCIENTIFIC, INDUSTRY AND BUSINESS APPLICATIONS IN DISTRIBUTED COMPUTING SYSTEMS

Usage of the distributed computing system in the recovery of the spectral density of sea waves	67
<i>Ilya Busko</i>	
Selection of rational composition of IT-services of information system with the purpose of increase of efficiency of transport logistics companies functioning.....	68
<i>Alexander Degtyarev, Gennady Dik, Alexander Dik</i>	
Virtual testbed for naval hydrodynamic problems	69
<i>Alexander Degtyarev, Yury Pylnev, Anatoly Eibozhenk, Alexander Bogdanov, Vladimir Korkhov, Ivan Gankevich</i>	
.NET Core technology in scientific tasks	69
<i>Victor Dorokhin</i>	
Different Approaches for Elastic Imaging using Multiprocessor Computing Systems	70
<i>Vasily Golubev, Alena Favorskaya</i>	
Cache-friendly memory traversal to improve performance of grid-characteristic method	71
<i>Andrey Ivanov, Nikolay Khokhlov</i>	
Scalable semantic virtual machine framework for language-agnostic static analysis.....	71
<i>Maxim Menshchikov</i>	
Enabling Biology, Chemistry and Other Sciences on Titan through BigPanDA	72
<i>Alexei Klimentov, Kaushik De, Danila Oleynik, Ruslan Mashinistov, Pavlo Svirin, Jack Wells, Sergey Panitkin</i>	

Molecular dynamic simulation of water vapor interaction with various types of pores using hybrid computing structures	73
<i>Eduard Nikonov, Vladimir Korenkov, Maria Popovičová</i>	
Using extended reality technologies in distributed computer systems	74
<i>Nadezhda Vozdvizhenskaya</i>	
Разработка децентрализованной платежной системы на основе технологии blockchain с учетом специфики мобильных платформ	74
<i>Андрей Илюхин, Эдуард Никонов</i>	
Имитационная модель БРЛК с синтезированной апертурой антенны в сети распределенных вычислений MarGrid	75
<i>Владимир Безродный, Анатолий Леухин, Андрей Воронин</i>	
NRV web knowledge base: scientific and educational applications.....	77
<i>Aleksandr Alekseev, Alexander Karpov, Andrey Denikin, Mikhail Naumenko, Viacheslav Samarin, Vladimir Rachkov, Vyacheslav Saiko</i>	
Essential aspects of IT training technology for processing, storage and intellectual analysis of the big data using the virtual computer lab.....	77
<i>Evgenia Cheremisina, Mikhail Belov, Nadezhda Tokareva, Yury Kryukov</i>	

CLOUD COMPUTING, VIRTUALIZATION

About some block chain problems	81
<i>Alexander Bogdanov, Alexander Degtyarev, Magdalyne Kamande, Oleg Iakushkin, Vladimir Korkhov</i>	
Experiments with JupyterHub at the Saint Petersburg State University	81
<i>Andrey Erokhin, Andrey Zarochentsev</i>	
Cloud Meta-Scheduler for Dynamic VM Reallocation	82
<i>Nikita Balashov</i>	
Approaches to the automated deployment of the cloud infrastructure of geographically distributed data centers.....	82
<i>Petr Fedchenkov, Nikita Samokhin, Sergey Khoruzhnikov, Oleg Lazo, Andrey Shevel</i>	
Kubernetes testbed cluster for the Lightweight Sites project	83
<i>Iuliia Gavrilenko</i>	
Design and implementation of a service for performing HPC computations in cloud environment.....	83
<i>Ruslan Kuchumov, Vladimir Korkhov</i>	
New features of the JINR cloud.....	84
<i>Nikolay Kutovskiy, Nikita Balashov, Alexandr Baranov, Yelena Mazhitova, Roman Semenov</i>	

Improving the load of supercomputers based on job migration using container virtualization	84
<i>Stanislav Polyakov, Minh-Duc Nguyen, Elena Petrova</i>	
Creation of cloud infrastructure of INP'S Astana branch - private establishment "NULITS" and its integration with the distributed JINR cloud infrastructure	85
<i>Boris Potapchuk, Mikhail Mazhitov, Sergey Belov, Yelena Mazhitova</i>	
Clouds of JINR, University of Sofia and INRNE - current state of the project.....	85
<i>Hristov Svetoslav, Kouzma Kouzmov, Nikita Balashov, Nikolay Kutovskiy, Radoslava Hristova, Vladimir Dimitrov, Vladimir Korenkov</i>	
The service for parallel applications based on the JINR Cloud and HybriLIT resources.....	86
<i>Elena Zemlyanaya, Pavel Goncharov, Andrey Nechaevskiy, Ruslan Kuchumov, Oksana Streltsova, Nikolay Kutovskiy, Maksim Bashashin, Nikita Balashov, Gennady Ososkov, Ivan Sokolov</i>	

DESKTOP GRID TECHNOLOGIES AND VOLUNTEER COMPUTING

Integration of the BOINC system and additional software packages.....	89
<i>Nikolay Khrapov</i>	
The activity of Russian Chapter of International Desktop Grid Federation	89
<i>Ilya Kurochkin</i>	
Modeling of task scheduling in desktop grid systems at the initial stage of development.....	90
<i>Ilya Kurochkin, Gerk Evgeny</i>	
Supporting Efficient Execution of Many-Task Applications with Everest	91
<i>Oleg Sukhoroslov, Sergey Smirnov, Sergey Volkov</i>	
The game character of collaboration in volunteer computing community	92
<i>Victor Tishchenko</i>	
Orthogonality-based classification of diagonal Latin squares of order 10.....	94
<i>Eduard Vatutin, Vitaly Titov, Oleg Zaikin, Stepan Kochemazov, Maxim Manzyuk, Natalia Nikitina</i>	
BOINC-based comparison of the geoacoustic inversion algorithms efficiency	96
<i>Oleg Zaikin, Pavel Petrov, Ilya Kurochkin</i>	
Реализация вычислений с динамическими зависимостями задач в среде десктоп грид с использованием Everest и Templet Web	96
<i>Востокин Сергей, Сухорослов Олег, Бобылева Ирина, Стефан Попов</i>	

HIGH PERFORMANCE COMPUTING, CPU ARCHITECTURES, GPU, FPGA

A software package for studying the system of long Josephson junctions on hybrid computing architectures	101
<i>Maksim Bashashin, Ksenia Erofeeva, Oksana Streltsova, Elena Zemlyanaya, Maksim Zuev</i>	
Network Infrastructure of heterogeneous platform "HybriLIT"	101
<i>Dmitry Belyakov, Yuriy Butenko, Andrey Dolbilov, Mikhail Matveev, Vladislav Moiseev, Martin Vala</i>	
On porting of applications to new heterogeneous systems.....	102
<i>Alexander Bogdanov, Vladimir Mareev, Nikita Storublevtzev, Denis Manyashin</i>	
Algorithms for the calculation of nonlinear processes on hybrid architecture clusters.....	103
<i>Alexander Bogdanov, Vladimir Mareev, Nikita Storublevtzev</i>	
DistrIbUted virtual cluster management system	103
<i>Amissi Cubahiro, Vladimir Gaiduchok, Magdalyne Kamande, Vladimir Korkhov, Alexander Degtyarev</i>	
The problem of symbolic-numeric computation of the eigenvalues and eigenfunctions of the leaky modes in a regular homogeneous open waveguide	104
<i>Andrey Drevitskiy, Dmitriy Divakov</i>	
Real-time visualization of ship and wavy surface motions based on GPGPU computations	105
<i>Anton Gavrikov, Andrei Ivashchenko, Ivan Gankevich, Nataliia Kulabukhova, Alexander Bogdanov, Alexander Degtyarev, Vladimir Rukovchuk</i>	
The Usage of HPC Systems for Simulation of Dynamic Earthquake Process	105
<i>Yulia Golubeva, Vasily Golubev</i>	
OpenFOAM wave modelling optimization with heterogeneous systems application porting.....	106
<i>Nikita Nizovtsov, Alexander Degtyarev</i>	
Accelerating real-time ship motion simulations using general purpose GPU computations	106
<i>Ivan Petriakov, Ivan Gankevich, Alexander Degtyarev, Vladimir Korkhov</i>	
GPGPU implementation of Schrödinger's Smoke for Unity3D	107
<i>Oleg Iakushkin, Anastasia Iashnikova, Olga Sedova</i>	
RSC BasIS Platform - micro-agent platform to manage computer clusters.....	107
<i>Pavel Lavrenko, Mikhail Malkov</i>	
Storage-on-demand: storage and compute united. RSC Tornado hyper-converged solution for data processing	108
<i>Pavel Lavrenko</i>	

Ways to improve the productivity of fire simulation tools on modern equipment.....	108
<i>Victor Smirnov</i>	
Information-software environment for the GOVORUN supercomputer.....	109
<i>Shushanik Torosyan, Oksana Streltsova, Dmitry Podgainy, Martin Vala, Margarit Kirakosya</i>	
Merging multidimensional histograms via hypercube algorithm.....	109
<i>Martin Vala, Andrey Bulatov, Yurii Butenko</i>	
SALSA - Scalable Adaptive Large Structures Analysis.....	110
<i>Martin Vala, Yurii Butenko, Branislav Beke, Fedor Matej</i>	
HybriLIT monitoring system.....	110
<i>Martin Vala, Ivan Kashunin, Yurii Butenko, Dmitry Belyakov</i>	
Govorun supercomputer engineering infrastructure. Monitoring system of engineering infrastructure.....	111
<i>Alexey Vorontsov, Andrey Dolbilov, Dmitry Podgainy, Oksana Streltsova, Vladimir Korenkov, Tatiana Strizh</i>	
Библиотеки и пакеты прикладных программ, доступные пользователям ЭВМ ОИЯИ.....	111
<i>Людмила Попкова, Александр Сапожников, Татьяна Сапожникова</i>	
О согласовании вычислительного эксперимента при интерактивном моделировании гидромеханики корабля в штормовом море.....	112
<i>Василий Храмушин, Александр Дегтярев, Александр Богданов, Иван Ганкевич</i>	
Контекстная графическая среда пространственной визуализации результатов вычислительных экспериментов в механике сплошных сред	114
<i>Василий Храмушин</i>	

CONSOLIDATION AND INTEGRATION OF DISTRIBUTED RESOURCES

Current status of data center for cosmic rays based on KCDC	119
<i>Dmitriy Kostunin, Victoria Tokareva</i>	
The interoperability problem during the implementation of the FOURTH PARADIGM	119
<i>Alexander Oleynikov, Andrey Kamenshchikov</i>	
DIRAC at JINR - purpose, experience, future.....	120
<i>Igor Pelevanyuk, Andrei Tsaregorodtsev</i>	
MODERN E - INFRASTRUCTURE FOR SCIENCE AND EDUCATION IN MOLDOVA BASED ON THE RENAM-GEANT PLATFORM	121
<i>Grigore Secrieru, Peter Bogatencov, Nicolai Iliuha, Nichita Degteariov</i>	
Discrete and Global Optimization in Everest Distributed Environment by Loosely Coupled Branch-and-Bound Solvers	121
<i>Vladimir Voloshinov, Sergey Smirnov</i>	

DATABASES, DISTRIBUTED STORAGE SYSTEMS, DATALAKES

Geometry Database for the CBM experiment and its first application to experiments in the NICA project	125
<i>Elena Akishina, Evgeny Alexandrov, Igor Alexandrov, Irina Filozova, Volker Friese, Victor Ivanov, Oleg Rogachevskiy, Konstantin Gertsenberger</i>	
A new approach to the development of provenance metadata management systems for large scientific experiments	125
<i>Andrey Demichev, Alexander Kryukov</i>	
Problems of date and time data types in relational model of data	126
<i>Vladimir Dimitrov</i>	
Data Knowledge Base for the ATLAS collaboration.....	126
<i>Marina Golosova, Maria Grigorieva</i>	
WLCG data lake prototype for HL-LHC	127
<i>Ivan Kadochnikov, Ian Bird, Gavin McCance, Jaroslava Schovancova, Maria Girone, Simone Campana, Xavier Espinal Currul</i>	
NRC "KI" participation in DataLake project	128
<i>Andrey Kiryanov, Andrey Zarochentsev</i>	
Integrating LEAF to data management workflow in LHAASO	128
<i>Haibo Li, Yaodong Cheng, Qi XU, Qiulan Huang</i>	
A distributed data warehouse system for astroparticle physics.....	129
<i>Minh Duc Nguyen</i>	
Evolution of Data Storage and Handling at Italian WLCG Data Center	129
<i>Vladimir Sapunenko, Luca Dell'Agnello, Gaetano Maron, Daniele Cesini</i>	
Новый вероятностно-статистический подход расчета информационных потерь в распределенных системах хранения и обработки данных физических экспериментов.....	130
<i>Андрей Нечаевский, Геннадий Ососков, Дарья Пряхина, Владимир Трофимов</i>	

BIG DATA ANALYTICS, MACHINE LEARNING

Semantic information management: the approach to semantic assets development lifecycle.....	135
<i>Yury Akatkin, Elena Yasinovskaya, Michael Bich</i>	
Checking foreign counterparty companies using Big Data	135
<i>Lazar Badalov, Sergey Belov, Ivan Kadochnikov</i>	
Labour market monitoring system	136
<i>Sergey Belov, Irina Filozova, Ivan Kadochnikov, Vladimir Korenkov, Roman Semenov, Petr Zrellov, Pavel Smelov</i>	
Text segmentation on photorealistic images.....	136
<i>Valery Grishkin, Aleksaner Ebral, Nikolai Stepenko, Jean Sene</i>	

Deep Learning Methodology for Prediction of Long-Term Dynamics of Financial Derivatives	137
<i>Alexander Bogdanov, Vladimir Rukovchuk, Kirill Lysov</i>	
Cluster analysis of scientific payload to execute it efficiently in distributed computing environment	137
<i>Maksim Gubin, Mikhail Titov, Maria Grigorieva</i>	
Russian-Language speech recognition system based on deepspeech	138
<i>Oleg Iakushkin, George Fedoseev, A. Shaleva, Olga Sedova, Alexander Degtyarev</i>	
Texture generation for archaeological reconstructions	139
<i>Oleg Iakushkin, Anna Fatkina, Dmitry Selivanov, Alexander Degtyarev</i>	
Machine learning for natural language processing tasks	139
<i>Aleksey Kulnevich, Vladislav Radishevskiy</i>	
Building corpora of transcribed speech from open access sources	140
<i>Oleg Iakushkin, G. Fedoseev, Anna Shaleva</i>	
Optimization of Neural networks training with Vector-Free heuristic on Apache Spark	140
<i>Kamil Khamitov, Nina Popova</i>	
Particle identification in ground-based gamma-ray astronomy using convolutional neural networks	141
<i>Evgeny Postnikov, Alexander Kryukov, Stanislav Polyakov, Dmitry Shipilov, Dmitriy Zhurov</i>	
Botnet in PyPy to speed up the work of the Earley parser	142
<i>Vladislav Radishevskiy, Aleksey Kulnevich, Nikolai Stepenko, Jean Sene</i>	
Improving the efficiency of smart grids of energy consumption with application of systems of artificial intellect	143
<i>Eugene Shchetinin, Mikchail Berezhkov</i>	
Using TensorFlow to solve the problems of financial forecasting for high-frequency trading	143
<i>Alexey Stankus</i>	
Comparison of explicit and not explicit mathematical methods of financial forecasting	144
<i>Alexey Stankus</i>	
Agent Technology Situational Express Analysis in Assessment of Technological Development Level of the BRICS Countries	144
<i>Evgeny Tretyakov, Diana Koshlan, Vladimir Korenkov, Boris Onykij, Alexey Artamonov</i>	
Combining satellite imagery and machine learning to predict atmospheric heavy metal contamination	145
<i>Alexander Uzhinskiy, Gennady Ososkov, Pavel Goncharov, Marina Frontsyeva</i>	
Architecture and basic principles of the multifunctional platform for plant disease detection	145
<i>Alexander Uzhinskiy, Pavel Goncharov, Gennady Ososkov, Andrey Nechaevskiy</i>	

Comparison of different convolution neural network architectures for the solution of the problem of emotion recognition by facial expression.....	146
<i>Anton Vorontsov, Alexey Averkin</i>	
Time Series and Data Analysis Based on Hybrid models of Deep Neural Networks and Neuro-Fuzzy Networks.....	147
<i>Sergey Yarushev, Alexey Averkin</i>	
Многомерный анализ данных о продажах на основе технологии OLAP	147
<i>Сауле Сарсимбаева, Владимир Димитров</i>	
Сверточная нейронная сеть в системе стереозрения мобильного робота.....	148
<i>Сергей Ульянов, Кирилл Кошелев</i>	
О методах и технологиях интеллектуального энергосбережения в коммерческих зданиях.....	149
<i>Евгений Щетинин, Евгения Попова</i>	

BIOINFORMATICS

Direct Simulation of the Charge Transfer along Oligonucleotides at T=300K	153
<i>Nadezhda Fialko, Victor Lakhno</i>	
Data consolidation and analysis system for brain research.....	154
<i>Vladimir Korkhov, Alexander Bogdanov, Vladislav Volosnikov, Andrey Vorontsov, Natalia Zalutskaya, Kirill Gribkov, Nikolay Neznanov, Natalia Ananyeva, Alexande Degtyarev</i>	
Development of software for face retrieval Systems modeling.....	155
<i>Nadezhda Shchegoleva, Varvara Petrova</i>	
Хаотическая динамика мгновенного сердечного ритма и его фазовое пространство.....	155
<i>Виктор Цветков, Алексей Кудинов, Александр Иванов, Илья Цветков, Сергей Михеев</i>	
Визуализация квантового фазового пространства мгновенного сердечного ритма	156
<i>Илья Цветков, Виктор Цветков, Сергей Михеев</i>	

POSTERS

Numerical solution of diffraction problem on the joint of two open three-layered waveguides	159
<i>Veniamin Chupritskiy, Dmitriy Divakov</i>	
An Image Verification Framework Development	159
<i>Ahmed Elaraby, Andrey Nechaevskiy</i>	

Efficiency measurement system for the computing cluster at IHEP	160
<i>Victoria Ezhova, Viktor Kotliar</i>	
Combined Explicit-Implicit Taylor Series Methods.....	160
<i>Ivan Hristov, Zafar Tukhliev, Stefka Dimova, Radoslava Hristova, Igor Puzynin, Taisiya Puzynina, Nikolay Shegunov, Zarif Sharipov</i>	
Data gathering and wrangling for the monitoring of the Russian labour market.....	161
<i>Javad Javadzade</i>	
Application of Hubzero platform for the educational process in astroparticle physics.....	162
<i>Yuliya Kazarina, Oleg Fedorov, Andrey Mikhailov, Igor Bychkov, Andreas Haungs, Alexander Kryukov, Dmitriy Kostunin, Alexey Shigarov, Dmitriy Shipilov, Daria Chernykh</i>	
Event-Driven Automation and chat-ops on IHEP computing cluster	162
<i>Anna Kotliar, Viktor Kotliar</i>	
Using binary file format description languages for verifying raw data in astroparticle physics experiments	163
<i>Alexander Kryukov, Elena Korosteleva, Dmitriy Kostunin, Igor Bychkov, Alexe Khmel'nov, Andrey Mikhailov, Oleg Fedorov, Alexey Shigarov</i>	
Pseudo-random number generator based on neural network.....	164
<i>Dmitry Kulyabov, Anna Korolkova, Migran Gevorkyan, Anastasiya Demidova</i>	
Methods & tools of the RSC BasIS distributed micro agent platform for managing compute, network and storage resources to efficiently process data	165
<i>Mikhail Malkov</i>	
Modern hyper-converged platform for computational- and I/O- heavy environments	165
<i>Mikhail Malkov</i>	
Modernization of web service for the data center simulation program	166
<i>Dmitry Marov, Daria Priakhina</i>	
Parallel calculations of ground states of 6,7,9,11Li nuclei by Feynman's continual integrals method.....	166
<i>Mikhail Naumenko, Viacheslav Samarin</i>	
Analysis of the features of the optimal logical structure of distributed databases	167
<i>Elena Nurmatova, Victor Gusev, Viktor Kotliar</i>	
Using multivariate quantile function for solving bioinformatics problems	167
<i>Sergey Poluyan, Nikolay Ershov</i>	
A way of anomaly detection in engineering equipment characteristics of Symmetra at IHEP IT center.....	168
<i>Ekaterina Popova, Viktor Kotliar</i>	

Sensitivity Analysis in a problem of ReaxFF molecular-dynamic force field optimization.....	168
<i>Konstantin Shefov, Stepanova Margarita</i>	
Optimisation of TensorFlow applications on the workstation Intel® Xeon® Platinum.....	169
<i>Svetlana Shikota, Lev Shchur, Alexander Russkov</i>	
Scalability of the Parallel Strongin Algorithm in the Problem of Optimizing a Molecular-Dynamic Force Field	169
<i>Margarita Stepanova, Konstantin Shefov</i>	
Convolutional neural networks for self-driving cars on GPU.....	170
<i>Boris Tyulkin, Nataliia Kulabukhova</i>	
Possible application areas of machine learning techniques at MPD/NICA experiment and their implementation prospects in distributed computing environment.....	171
<i>Dmitry Zinchenko, Alexander Zinchenko, Eduard Nikonov</i>	
Применение эволюционных и роевых алгоритмов оптимизации для решения модельной задачи предсказания структуры белка	171
<i>Максим Быстров, Николай Ершов</i>	
Анализ параллельной структуры популяционных алгоритмов оптимизации	172
<i>Николай Ершов, Сергей Полуян</i>	
Применение сети Хопфилда для автоматизированной подборки КПЭ....	173
<i>Денис Кравченко, Эдуард Никонов</i>	
Когнитивно-интеллектуальная система адаптации и обучения детей-аутистов	173
<i>Алла Мамаева, Андрей Шевченко</i>	
Когнитивно интеллектуальная система диагностики, адаптации и обучения детей-аутистов. Модуль обработки данных	174
<i>Андрей Шевченко, Алла Мамаева</i>	

PLENARY REPORTS

SUPERCOMPUTER "GOVORUN" — NEW PROSPECTS FOR HETEROGENEOUS COMPUTATIONS AT JINR

**Gheorghe Adam, Vladimir Korenkov, Dmitry Podgainy^a,
Oksana Streltsova**

¹Joint Institute for Nuclear Research

E-mail: ^apodgainy@jinr.ru

The report provides description of the “HybriLIT” heterogeneous platform that is a component of the Multipurpose information and computing complex (MICC) of JINR. HybriLIT includes “GOVORUN” supercomputer and education and testing polygon; its platform is based on the latest computation architectures (processors; co-processors; graphical accelerators), and also modern software such as Intel Cluster Studio, CUDA, Matlab, etc.; thus, allowing to carry out extra-massive computations and reach sufficient acceleration. “GOVORUN” supercomputer meets all the requirements for modern HPC systems:

1. Possibilities for dynamic expansion of the cluster by means of adding new computation nodes;
2. Possibilities for synchronous updates of the software on computation nodes;
3. Swift installation and maintainability of nodes on the cluster after failures and reloads.

For efficient use of computation resources of the supercomputer, information environment has been developed. It includes services that provide possibilities for interaction with users, development of applications, notifications on the upcoming events, and organization of tutorials on parallel programming technologies.

Also, the report provides information on the current use of the “GOVORUN” supercomputer resources for the tasks being solved at JINR.

CRIC: THE INFORMATION SYSTEM FOR LHC DISTRIBUTED COMPUTING

Alexey Anisenkov

Budker Institute of Nuclear Physics

E-mail: alexey.anisenkov@cern.ch

The Worldwide LHC Computing Grid infrastructure links about 200 participating computing centers affiliated with several partner projects. It is built by integrating heterogeneous compute and storage resources in diverse data centers all over the world and provides CPU and storage capacity to the LHC experiments to perform data processing and physics analysis at petabytes scale data operations. Moreover the experiments extend the capability of WLCG distributed environment by actively connecting opportunistic Cloud platforms, HPC and volunteer resources. In order to be effectively being used by the experiments, these distributed resources should be well described, which implies easy service discovery and detailed description of service configuration. CRIC represents the evolution of ATLAS Grid Information System (AGIS) into the common experiment independent high-level information framework which has been evolved in order to serve not just ATLAS Collaboration needs for the description of distributed environment but any other virtual organization relying on large scale distributed infrastructure as well as the WLCG on the global scope. CRIC collects information from various information providers, complements it with experiment-specific configuration required for computing operations, performs data validation and provides coherent view and topology description to the LHC VOs for service discovery and usage configuration. In this contribution we describe the design and overall architecture of the system, recent developments and most important aspects of the CRIC framework components implementation and features like flexible definition of information models, built-in collectors, user interfaces, advanced fine-granular authentication/authorization and others.

THE CMS TIER1 AT JINR: FIVE YEARS OF OPERATIONS

**Andrey Baginyan, Anton Balandin, Sergey Belov, Andrey Dolbilov,
Alexey Golunov, Natalia Gromova, Ivan Kadochnikov,
Ivan Kashunin, Vladimir Korenkov, Valery Mitsyn,
Igor Pelevanyuk, Sergey Shmatov, Tatiana Strizh^a,
Vladimir Trofimov, Nikolay Voytishin, Victor Zhiltsov**

Joint Institute for Nuclear Research

E-mail: ^astrizh@jinr.ru

This talk summarizes five years of operational experience of the WLCG Tier1 computer centre at the Laboratory of Information Technologies of Joint Institute for Nuclear Research, which serves the CMS experiment at LHC. In early 2013 the Tier1 prototype was deployed and its trial operation began. March, 2015 is the date of finalization of this complex and commissioning a full-scale Tier1 centre for CMS at JINR. Since its inception it was continuously adapted to the new requirements, introducing new hardware and technologies as they became available. The resources provided by the centre to the CMS experiment have increased significantly and it is on top reliability levels as compared to other Tier1 centers processing data for CMS. A special Tier1 network centre was developed to provide scalability of its network infrastructure. Additional work has been done in the recent years in hardware and services monitoring. Future modernization and increase of the Tier1 performance will provide possibilities of efficient and fast processing and reliable storage of the CMS data to cope with high luminosity and high energy of the collisions of LHC run 3.

THE ATLAS EVENTINDEX AND ITS EVOLUTION BASED ON APACHE KUDU STORAGE

Dario Barberis

University and INFN Genova

E-mail: dario.barberis@ge.infn.it

The ATLAS experiment produced hundreds of petabytes of data and expects to have one order of magnitude more in the future. This data are spread among hundreds of computing Grid sites around the world. The EventIndex catalogues the basic elements of these data - real and simulated events. It provides the means to select and access event data in the ATLAS distributed storage system, and provides support for completeness and consistency checks and data overlap studies. The EventIndex employs various data handling technologies like Hadoop and Oracle databases, and is integrated with other elements of the ATLAS distributed computing infrastructure, including systems for data, metadata, and production management (AMI, Rucio and PANDA). The project is in operation since the start of LHC Run 2 in 2015, and is in permanent development in order to fit the analysis and production demands and follow technology evolutions. The main data store in Hadoop, based on MapFiles and HBase, can work for the rest of Run 2 but new solutions are explored for the future. Kudu offers an interesting environment, with a mixture of BigData and relational database features, which looked promising at the design level and is now used to build a prototype to measure the scaling capabilities as a function of data input rates, total data volumes and data query and retrieval rates. An extension of the EventIndex functionalities to support the concept of Virtual Datasets produced additional requirements that are tested on the same Kudu prototype, in order to estimate the system performance and response times for different internal data organisations. This talk reports on the current system performance and on the first measurements of the new prototype based on Kudu.

BIG DATA AS THE FUTURE OF INFORMATION TECHNOLOGY

**Alexander Bogdanov^a, Alexander Degtyarev , Vladimir Korkhov,
Thurein Kyaw**

St.Petersburg State University

E-mail: ^abogdanov@csa.ru

Currently, the problem of "Big Data" is one of the most, if not the most urgent in computer science. Its solution implies the possibility of processing uncorrelated and heterogeneous data of large volume, the implementation of their integration from distributed sources by consolidation or federalization methods and ensuring the security of access and storage of these data. Only the creation of technology that provides processing and storage of dissimilar, uncorrelated data of large volume can be considered a breakthrough result corresponding to the world level. To effectively address these issues, a new definition of this concept is proposed, namely, "Big Data" is characterized by the situation when the conditions for implementing the CAP theorem are relevant. The CAP theorem is a heuristic statement that in any realization of distributed computations, it is impossible to provide the following three properties: Consistency, Availability and Partition Tolerance. Thus, depending on which of the properties cannot be implemented, we are dealing with different types of "Big Data". And this, in turn, means that a standard approach based on the MapReduce concept has a limited scope of applicability. Various possibilities for implementing data processing in different cases are discussed, and a conclusion is made about the need to create an ecosystem of "Big Data". The work will review the world market of Big Data technologies, and also describe the state of work on this problem in various countries. At the end of the article, we will talk about the opportunities that the solution of the problem opens for various fields of science and business.

THE DESIGNING OF CLOUD INFRASTRUCTURE CONSISTING OF GEOGRAPHICALLY DISTRIBUTED DATA CENTERS

**Petr Fedchenkov¹, Nikita Samokhin¹, Sergey Khoruzhnikov¹,
Oleg Lazo¹, Andrey Shevel^{1,2,a}**

¹*ITMO University*

²*NRC "Kurchatov Institute" PNPI*

E-mail: ^ashevel_aye@pnpi.nrcki.ru

University ITMO (ifmo.ru) is designing the cloud of geographically distributed data centers under centralized administration to control the distributed virtual storage, virtual data links, virtual machines, and data center infrastructure management. Resulted cloud has to be tolerant to hardware and software failures of any type. The integrated set of programs is developed to implement mentioned goals. Each program of the set is relatively independent program agent in form of VM or container which can run on different hardware servers. Any agent might send the request for specific service to another agent with developed protocol. The cloud system of distributed data centers assumes well known functionality. The creation, management, and provision of services with defined SLA: virtual machines, long-term data storage, data links with ability to encrypt the transferred data, and so on. In presented approach most of above functions are implemented in form of program agents. The installation of the system in a number of data centers is implemented with a range of automated deployment steps. Many FOSS components like Openstack, CEPH, SALT, Grafana/Kibana, Zabbix, etc were used as toolkits in this design. The developed cloud is now under heavy testing/modifications.

CLOUD-BASED COMPUTING FOR LHAASO EXPERIMENT AT IHEP

**Qiulan Huang^a, Weidong Li, Yaodong Cheng, Haibo Li, Tao Cui,
Xiaowei Jiang**

Institute of High Energy Physics, Chinese Academy of Sciences

E-mail: ^ahuangql@ihep.ac.cn

Mass data processing and analysis contribute much to the development and discoveries of a new generation of High Energy Physics. The LHAASO(Large High Altitude Air Shower Observatory) experiment of IHEP located in Daocheng, Sichuan province (altitude 4410 m), is expected the most sensitive project to study the problems in Galactic cosmic ray physics, and requires massive storage and computing power, which is urgent to explore new solutions based on cloud computing models to integrate distributed heterogeneous resources. However, it faces with high operation and maintenance costs, system instability and other issues. To address these issues, we introduce cloud computing technology for LHAASO in order to make sure the system availability and stability, as well as simplify system deployment and significantly reduce the maintenance cost. Particularly, we discuss the cloud-based computing architecture to federate distributed resources across regions for LHAASO experiment, including distributed resource management, job scheduling, distributed monitoring and automated deployment. Also container orchestration is introduced to make use of the load balancing and fault tolerance to improve system availability. The prototype is based on Openstack and HTCondor to achieve a unified resource management and scheduling job across regions transparently, located in Beijing, Chengdu and Daocheng, some commercial cloud also included like Alibaba Cloud. We also report a dynamic resource provisioning approach to achieve the resource expansion on demand and the efficient job scheduling, so as to improve the overall resource utilization. Considering serving data access from remote site, we design a remote storage system named LEAF to provide the unified data view, data cache as well as high performance of data access over WLAN. This presentation also will discuss the cloud-based open platform for LHAASO. The feature of the platform makes it possible to do data analysis by the web browser as data, software and computing resources are available in cloud. Finally, a proposal of integrating the HPC facility into the current computing system to faster LHAASO reconstruction jobs will be discussed.

PIK COMPUTING CENTRE

Andrey Kiryanov

NRC "Kurchatov Institute" PNPI

E-mail: globus@pnpi.nw.ru

In the framework of the PIK nuclear reactor reconstruction project, a new PIK Computing Centre was commissioned in 2017, the main task of which is storage and processing of PIK experiments data. The Centre's capacity is also used by other scientific groups at PNPI for solving problems in different areas of science such as computational biology and condensed matter physics. It also becomes an integral part of computing capacities of NRC "Kurchatov Institute". The PIK Computing Centre has a heterogeneous structure and consists of several types of computing nodes suitable for a wide range of tasks and two independent data storage systems, all of which are interconnected with a fast InfiniBand network. The engineering infrastructure provides redundant main power and two independent UPS installations for computing equipment and for cooling system.

REAL-TIME EVENT RECONSTRUCTION AND ANALYSIS IN THE CBM EXPERIMENT AT FAIR USING HPC

Ivan Kisel (for the CBM collaboration)

Goethe University Frankfurt

E-mail: i.kisel@compeng.uni-frankfurt.de

CBM is a future heavy-ion experiment at FAIR, Darmstadt. It is characterised by up to 10 MHz collision rates, a large amount of produced particles, non-homogeneous magnetic fields and a very complex detector system. Event reconstruction is the most complicated and time consuming task of the data analysis in the CBM experiment with up to one thousand particles per central collision. An additional complication is a continuous data stream represented in form of time slices. All of the above mentioned makes it necessary to develop fast and efficient reconstruction algorithms and to optimise them for running on heterogeneous many-core HPC clusters. The First-Level Event Selection (FLES) package is intended to reconstruct online the full event topology including trajectories (tracks) of charged particles and short-lived particles without low-level hardware triggers. The FLES package consists of several modules: Cellular Automaton (CA) track finder, Kalman Filter (KF) based track fitter, KF based short-lived particle finder and physics selection. The FLES package is platform and operating system independent. It is portable to different many-core CPU architectures. The package is vectorised using SIMD instructions and parallelised between CPU cores. All algorithms are optimised with respect to the memory usage and the speed. The CA track finder takes 10 ms per minimum bias event on a single CPU core. The KF track fitter estimates parameters of about 400 particles in a μs on a GPU card. Decays of short-lived particles in more than 100 channels are searched and analysed in about 2 ms. The whole FLES package shows a strong scalability many-core CPU systems achieving the reconstruction speed of 1700 events/s on a 80-cores server and $2.2 \cdot 10^5$ events/s on an HPC cluster with 100 nodes. The developed FLES algorithms can be of interest for other HEP experiments as well.

BIGPANDA EXPERIENCE ON TITAN FOR THE ATLAS EXPERIMENT AT THE LHC

**Alexei Klimentov^{1,a}, Kaushik De², Danila Oleynik³,
Ruslan Mashinistov⁴, Jack Wells⁵**

¹*Brookhaven National Lab*

²*University of Texas at Arlington*

³*Joint Institute for Nuclear Research*

⁴*NRC "Kurchatov Institute"*

⁵*Oak Ridge National Laboratory*

E-mail: ^aalexei.klimentov@cern.ch

The PanDA software is used for workload management on distributed grid resources by the ATLAS experiment at the LHC. An effort was launched to extend PanDA, called BigPanDA, to access HPC resources, funded by the US Department of Energy (DOE-ASCR). Through this successful effort, ATLAS today uses over 25 million hours monthly on the Titan supercomputer at Oak Ridge National Laboratory. Many challenges were met and overcome in using HPCs for ATLAS simulations. ATLAS uses two different operational modes at Titan. The traditional mode uses allocations - which require software innovations to fit the low latency requirements of experimental science. New techniques were implemented to shape large jobs using allocations on a leadership class machine. In the second mode, high priority work is constantly sent to Titan to backfill high priority leadership class jobs. This has resulted in impressive gains in overall utilization of Titan, while benefiting the physics objectives of ATLAS. For both modes, BigPanDA has integrated traditional grid computing with HPC architecture. This talk will summarize the innovations to successfully use Titan for LHC physics goals

JINR MULTIFUNCTIONAL INFORMATION AND COMPUTING COMPLEX: STATUS AND PERSPECTIVES

**Vladimir Korenkov^a, Andrey Dolbilov, Valery Mitsyn,
Tatiana Strizh^b, Nikolay Kutovskiy, Dmitry Podgainy,
Oksana Streltsova, Vladimir Trofimov**

Joint Institute for Nuclear Research

E-mail: ^akorenkov@jinr.ru, ^bstrizh@jinr.ru

JINR possesses a complex informational-computational infrastructure. The uninterrupted functioning of all its elements at the right level is mandatory for the fulfillment of the JINR scientific research programmes. The support of this infrastructure fully functional is a major task of the Laboratory of Information Technologies. Follows from the noticeable diversity of the scientific targets defined by the JINR research the Multifunctional Information and Computing Complex (MICC) was developed as distributed computing infrastructure fulfilling all the needs. The MICC should meet the requirements for a modern highly performant scientific computing complex: multi-functionality, high performance, task adapted data storage system, high reliability and availability, information security, scalability, customized software environment for different existing user groups, high performance telecommunications and modern local network. Dedicated MICC components are: the CMS Tier1 grid site; JINR Tier2 grid site providing support to the virtual organizations (VOs) concerning the JINR participation in the LHC experiments (ATLAS, ALICE, CMS, LHCb), other VOs within large-scale international collaborations with the JINR groups and, traditionally, the sequential computing tasks of non-grid JINR; cloud computing structure aimed at expanding the range of services provided to the users and at creating an integrated cloud environment of the JINR Member States; high performance heterogeneous computing platform HybriLIT, the main part of which is supercomputer “Govorun”. A brief status overview of each component is presented. Particular attention is given to the development of distributed computations performed in collaboration with CERN, BNL, FNAL, FAIR, China, and JINR Member States. We present our plans to further develop MICC as a center for scientific computing within the multidisciplinary research environment of JINR and JINR Member States, and particularly for megascience projects, such as NICA.

MULTICOMPONENT CLUSTER MANAGEMENT SYSTEM FOR THE COMPUTING CENTER AT IHEP

Anna Kotliar, Ekaterina Popova, Victoria Ezhova, Viktor Kotliar^a

NRC "Kurchatov Institute" IHEP

E-mail: ^avictor.kotlyar@ihep.ru

Cluster management system is a core part of any computing infrastructure. Such system includes components for allocating and controlling over resources for different computing tasks, components for configuration management and software distribution on the computing hardware, components for monitoring and management software for the whole distributed infrastructure. The main goals of such system are to create autonomic computing system with functional areas such as self-configuration, self-healing, self-optimization and self-protection or to help to reduce the overall cost and complexity of IT management by simplifying the tasks of installing, configuring, operating, and maintaining clusters. In the presented work current implementation of the multicomponent cluster management system for the IHEP computing center will be shown. For the moment this system consists of event-driven management system, configuration management system, monitoring and accounting system and a chat-ops technology which is used for the administration tasks.

ELECTRONIC, DYNAMICAL AND THERMODYNAMIC PROPERTIES OF DNA

Victor Lakhno

*Institute of Mathematical Problems of Biology RAS,
Keldysh Institute of Applied Mathematics of RAS*

E-mail: lak@impb.ru

The idea to use DNA molecule as a base element for nanobioelectronics is discussed. It could be considered as some molecular wire where a typical charge transfer/transport pattern can physically be viewed as a polaron and/or soliton which mobility can be very low. A computer experiment demonstrates that mobile breather excited near one of the ends of DNA can trap the polarons. The formed quasiparticle can move along the molecule for a long distance and does not require the electric field. The dynamics of charge migration was modeled to calculate temperature dependencies of its thermodynamic equilibrium values such as energy, electronic heat capacity and reaction constants for different nucleotide sequences. The mechanism of charge transfer for a long distance due to polaron melting is considered. Special attention is given to: dynamical behavior of electrons in regular polynucleotide chains, dynamics of polaron states formation in Peyrard – Bishop- Dauxois chain, polaron motion in an electric field, the role of dispersion, Bloch oscillations and breather states.

The work was supported by RSF project 16-11-10163.

FILE TRANSFER SERVICE AT EXABYTE SCALE

Andrea Manzi

CERN, the European Organization for Nuclear Research

E-mail: andrea.manzi@cern.ch

The File Transfer Service (FTS) is an open source solution for data transfers developed at CERN. It has been used for almost 4 years for the CERN LHC experiments data distribution in the WLCG infrastructure and during this period the usage has been extended to non-CERN and non-HEP users, reaching in 2017 almost an Exabyte of transferred data volume.

The talk will focus on the service architecture and main features, like the transfer optimizer, the multi protocol support, cloud extensions and monitoring. The ongoing and future activities around FTS are also going to be presented, like the integration with OpenID Connect and CDMI.

LARGE SCALE SIMULATIONS WITH PARALLEL ANNEALING ALGORITHM

Lev Shchur

Landau Institute for Theoretical Physics, Science Center in Chernogolovka

E-mail: levshchur@gmail.com

Population annealing algorithm designed for the simulations of the statistical mechanics systems with rugged free energy landscape. We report on the realization of the algorithm for the use on the hybrid computing architecture combining CPU and GPGPU. Algorithm is fully scalable. We report application of the developed realization to several interesting problems. Algorithm can be applied to any system of statistical mechanics, described by partition function.

Work is supported by the grant of Russian Science Foundation 14-21-00158.

TIER-1 CENTRE AT NRC “KURCHATOV INSTITUTE” BETWEEN LHC RUN2 AND RUN3

**Igor Tkachenko^a, Eygene Ryabinkin, Alexander Rogovskiy,
Ilya Lyalin**

National Research Center "Kurchatov Institute"

E-mail: ^atia@grid.kiae.ru

The issues of development and modernization of the Tier-1 center at the National Research Center "Kurchatov Institute" are considered in accordance with the changing requirements of experiments at the Large Hadron Collider. Increasing requirements for computing resources, driven by increase in simulations, led to an increase in their volumes, which in turn required the development of automation of the processes of managing the installation and configuration of the software. To improve the quality of the software management system and minimize the risk of errors during the nodes automatic configuration a code review system was implemented. A new system for in-demand provision of additional computing resources from supercomputing cluster was developed and implemented.

DIRAC SERVICES FOR SCIENTIFIC COMMUNITIES

Andrei Tsaregorodtsev

Center for Particle Physics of Marseilles IN2P3 CNRS

E-mail: atsareg@in2p3.fr

The software framework developed by the DIRAC Project provides all the necessary components for building distributed computing systems for large-scale scientific collaborations. It supports both workload and data management tasks, therefore providing an integral solution for computing models of various scientific communities. DIRAC is used by several large High Energy Physics experiments, most notably by the LHCb Collaboration at CERN. The DIRAC services provided by several grid infrastructures are giving access to the framework to researchers from other scientific domains with applications largely differing in their scale and properties. The DIRAC development plans are strongly influenced by these new communities aiming to satisfy their specific needs. In this contribution we will present recent DIRAC evolution for enhancing services provided by grid infrastructure projects, in particular those provided by the EOSC-Hub project for the users of the European Grid Infrastructure.

DEEP MACHINE LEARNING AND PATTERN/FACE RECOGNITION BASED ON QUANTUM NEURAL NETWORKS AND QUANTUM GENETIC ALGORITHM

Sergey Ulyanov^a, Andrey Reshetnikov, Nikita Ryabov

Dubna State University

E-mail: ^aulyanovsv@mail.ru

In report a new approach for deep machine learning and pattern recognition based on quantum neural network and quantum genetic algorithm is described. The structure of quantum fuzzy neural network is considered. Examples of pattern recognition is described. The method of global optimization in control problems is considered on example of quantum genetic algorithm. The structure on quantum genetic algorithm is introduced. Information technology of intelligent control system design based on quantum soft computing is presented. Example of quantum genetic algorithm application for control of nonlinear “carte-pole” system is described. Application of modified Grover quantum search algorithm in unstructured big database is discussed. Quantum soft computing optimizer of knowledge bases is presented. This report discusses the development of robust intelligent control systems. Special attention is paid to the algorithm of quantum fuzzy inference, in particular to the stage of determining the type and form of quantum correlation. Automating the choice of the type of quantum correlation can be done with the help of a quantum genetic algorithm whose analysis and choice are considered in this report.

ADVANCED GLOBAL NETWORK SERVICES TO SUPPORT RESEARCH EXCELLENCE

Rudolf Vohnout^{1,2,a}, Vincenzo Capone¹

¹*GÉANT*
²*CESNET*

E-mail: ^arudolf.vohnout@cesnet.cz

With the focus on cloud services and the benefits of using the exclusive contracts with major cloud providers made by the effort of the community, GÉANT provides unique access to the cloud resources for all its members. Based on CERN experience and data generated by LHC, especially with CMS experiment, working with physics community has been always challenging for GÉANT so as the motivation to provide best services possible to accelerate top research on European and thanks to cooperation with overseas partners also at global scale. As there is a strong cooperation between physical research teams in Geneva and Dubna, GÉANT is investigating ways to optimize the connection between the research institutions and labs. This paper (and hopefully presentation) will focus on possible solutions of interconnection of the cooperating research teams in JINR and CERN in order to support their research activities. The situation in Russia was so far more complicated than in other countries due to the existence of more National Research and Education Networks (aka NRENs) providing services to different customers and they were not so keen to cooperate in order to be able to deliver optimal supportive environment to the community it deserves.

ОРГАНИЗАЦИЯ ДОСТУПА К ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ УСТАНОВКИ ИТЭР В РЕЖИМЕ УДАЛЕННОЙ ПУЛЬТОВОЙ

**Екатерина Миронова, Игорь Семенов^а, Олег Семенов,
Сергей Портоне**

Проект ИТЭР

E-mail: ^аi.semenov@iterrf.ru

В настоящее время в России, на базе Проектного Центра ИТЭР, создан прототип Центра удаленного участия в экспериментах на крупных физических установках - Remote Participation Center. Цель – создание единой исследовательской среды научно-исследовательских центров, лабораторий и университетов, участвующих в исследованиях по управляемому термоядерному синтезу. Основной задачей Центра также является отработка систем управления диагностических комплексов, поставляемых Российской Федерацией в Международный проект ИТЭР. Проект ИТЭР (ITER – International Thermonuclear Experimental Reactor) в настоящее время является одним из наиболее сложных международных научно-технических мега проектов. Собственно, установка сооружается во Франции в Центре атомных исследований Кадараш. В проекте участвуют семь стран: Россия, Объединённая Европа, Китай, Индия, Корея, США, Япония. Стоимость проекта около 20 млрд. долларов (Доля России ~2 млрд. долларов). Завершение строительства (первая плазма) в 2026 году. В основу проекта заложен принцип создания высокотемпературной плазмы (150 млн. градусов) на основе установки ТОКАМАК. Предполагается, что ИТЭР будет производить порядка 10-15 Пбайт экспериментальной информации в год. В докладе представлен обзор работ по организации удаленного доступа к экспериментальным данным и дистанционному управлению диагностическим оборудованием на современных термоядерных установках, а также рассматриваются вопросы организации передачи больших потоков данных в условиях ограниченной пропускной способности линий передачи данных. Доклад представляет интерес для физиков и инженеров, работающих на крупных физических установках в области информационных технологий.

Работа выполнена по Контракту с Государственной Корпорацией РОСАТОМ №Н.4а.241.9Б.17.1001.

**TECHNOLOGIES,
ARCHITECTURES, MODELS
OF DISTRIBUTED COMPUTING
SYSTEMS**

ALICE DCS PREPARATION FOR RUN 3

**Alexander Kurepin^a, Andre Augustinus, Peter Chochula,
Ombretta Pinazza, Mateusz Lechman, Peter Matthew Bond,
Kevin Cifuentes Salas, John Larry Lang**

CERN, the European Organization for Nuclear Research

E-mail: ^aalexander.kurepin@cern.ch

The ALICE experiment is heavy ion collision detector at the CERN LHC. Its goal to study extreme phase of matter – called quark-gluon plasma. It is collaboration of 41 countries and more than 1800 scientists. A large number of complex subsystems requires supervision and control system. ALICE Control Coordination (ACC) is the functional unit mandated to coordinate the execution of the Detector control system (DCS).

In 2020, the ALICE experiment at CERN will start collecting data with upgraded detector. The ALICE upgrade addresses the challenge of reading out and inspecting the Pb-Pb collisions at rates of 50 kHz, sampling the pp and p-Pb at up to 200 kHz. ALICE O2 project merges online and offline into one large system with ~8400 optical links, data rate 1.1 TB/s, data storage ~60PB/year. From DCS O2 requires continuous data flow with ~100 000 conditions parameters for event reconstruction. Data has to be injected into each 20ms data frame.

DCS-O2 interface consists of electronics and software modules for configuring CRU controllers and provide continuous dataflow to O2 system.

In this talk, we will describe the architecture and functionality of the ADAPOS mechanism. We will discuss the requirements and results obtained during the test campaign. We will also provide a description of a new front-end access mechanism allowing for detector control in parallel to the data acquisition.

A new mechanism, called ADAPOS has been developed. Its role is to collect condition parameters from the distributed SCADA system and provide this data to O2 - A new combined online-offline facility.

THE DISTRIBUTED GRID SITE OF INSTITUTE OF PHYSICS

**Alexandr Mikula^{1,2,a}, Martin Adam¹, Jiří Chudoba¹,
Dagmar Adamová³, Jana Uhlířová¹, Petr Horák¹, Petr Vokáč⁴**

¹*Institute of Physics of the Czech Academy of Sciences*

²*CESNET*

³*Nuclear Physics Institute of the Czech Academy of Sciences*

⁴*Czech Technical University in Prague*

E-mail: ^amikula@fzu.cz

The Computing Center of the Institute of Physics (IOP) of the Czech Academy of Sciences serves a broad spectrum of users with various computing needs. The Computing Center hosts a WLCG Tier-2 for ATLAS and ALICE experiments. There are also other supported experiments from astroparticle physics, namely Cherenkov Telescope Array and Pierre Auger Observatory. Center also supports OSG stack for the NOvA and DUNE experiments. Computing resources are also utilized by local users from IOP through HTCondor batch system. Hosted storage capacity is divided between grid services (DPM and XrootD) and locally accessible NFS storage. Computing resources are distributed among several locations in the Czech Republic. This contribution will describe mentioned topics in more detail. It will also give insight in our experience with different classes of hardware and with different approaches of administration and monitoring of all services.

INP BSU GRID SITE

Vladimir Mossolov, Dmitry Yermak, Vitaly Yermolchik^a

Institute for Nuclear Problems, Belarusian State University

E-mail: ^ayermolchik@hep.by

The main goal of INP BSU grid site is to provide access for scientists and students from our institute to computational power of WLCG and make contribution to data processing. We are involved in the two greatest modern experiments in particle physics – CMS and ATLAS at the Large Hadron Collider at CERN. INP BSU grid site is the only certified and production EGI grid site in Belarus. Also we integrated our cloud resources with JINR cloud.

An overview of INP BSU computational facilities usage and development is presented.

PROPERTIES OF THE PARALLEL DISCRETE EVENT SIMULATION ALGORITHMS ON SMALL-WORLD COMMUNICATION NETWORKS

Liliia Ziganurova^{1,3,a}, Lev Shchur^{2,3}

¹*Scientific Center in Chernogolovka,*

²*Landau Institute for Theoretical Physics*

³*National Research University Higher School of Economics*

E-mail: ^aziganurova@gmail.com

We discuss synchronization aspects in the method of large-scale simulation, known as parallel discrete event simulation (PDES). We build models of the evolution of simulation time profile in two PDES algorithms, in conservative algorithm and in optimistic one. The models capture the essential properties of the algorithms, namely, the scalability and the degree of desynchronization. We investigate the models on small-world communication networks (SW), which constructed as regular lattices with addition of small fraction of long-range communication links. SW networks are characterized by the small length of average shortest path and by the large value of clustering coefficient. We show that synchronization is better, when processing elements are arranged in SW topology, rather than in regular lattices. In PDES algorithms on SW network the desynchronization remains constant in the limit of infinite number of processing elements, and the same time the average utilization remains positive. We also find, that the degree of clustering in networks has no influence on the synchronization between processing elements, and the synchronization is mainly affected by the length of average shortest path. We present the results of our simulations and compare them with the case-study simulations.

The work is done under the grant RSCF 14-21-00158.

ГРИД И ОБЛАЧНАЯ ИНФРАСТРУКТУРА ДАТА-ЦЕНТРА ИНСТИТУТА ФИЗИКИ НАН АЗЕРБАЙДЖАНА

Алексей Бондяков

Объединенный институт ядерных исследований

E-mail: aleksey@jinr.ru

Основными направлениями развития дата-центра Института физики НАН Азербайджана являются грид и облачные технологии. Пользуясь тем что грид-сегмент дата-центра интегрирован в инфраструктуру EGI/WLCG в качестве грид-сайта уровня Tier3, пользователи дата-центра получают возможность участвовать в таких международных проектах как ATLAS (CERN). Сотрудничество с международными научными центрами ОИЯИ, ЦЕРН в области информационных технологий способствует эффективному развитию дата центра, а также помогает решать задачи пользователей в таких научных областях, как физика высоких энергий, физика твердого тела и т. д.

**OPERATION, MONITORING,
OPTIMIZATION IN DISTRIBUTED
COMPUTING SYSTEMS**

THE BIGPANDA MONITORING SYSTEM ARCHITECTURE

**Aleksandr Alekseev¹, Alexei Klimentov², Siarhei Padolski²,
Tatiana Korchuganova^{1,a}, Torre Wenaus²**
(on behalf of the ATLAS Collaboration)

¹*National Research Tomsk Polytechnic University*

²*Brookhaven National Lab*

E-mail: ^atatianakorchuganova@tpu.ru

Currently-running large-scale scientific projects involve unprecedented amounts of data and computing power. For example, the ATLAS experiment at the Large Hadron Collider (LHC) has collected 140 PB of data over the course of Run 1 and this value increases at rate of ~800 MB/s during the ongoing Run 2 and recently has reached 350 PB. Processing and analysis of such amounts of data demands development of complex operational workflow and payload systems along with building top edge computing facilities. In the ATLAS experiment a key element of the workflow management is the Production and Distributed Analysis system (PanDA). It consists of several core components and one of them is the monitoring. The latter is responsible for providing a comprehensive and coherent view of the tasks and jobs executed by the system, from high level summaries to detailed drill-down job diagnostics. The BigPanDA monitoring has been in production since the middle of 2014 and it continuously evolves to satisfy increasing demands in functionality and growing payload scales. Today it effectively keeps track of more than 2 million jobs per day distributed over 170 computing centers worldwide in the largest instance of the BigPanDA monitoring: the ATLAS experiment. In this paper we describe the monitoring architecture and its principal features.

THE BIGPANDA SELF-MONITORING ALARM SYSTEM FOR ATLAS

Aleksandr Alekseev^{1,a}, Siarhei Padolski², Tatiana Korchuganova¹

¹*National Research Tomsk Polytechnic University*

²*Brookhaven National Lab*

E-mail: ^afrt@tpu.ru

The BigPanDA monitoring system is a Web application created to deliver the real-time analytics, covering many aspects of the ATLAS experiment distributed computing. The system serves about 35000 requests daily and provides critical information used as input for various decisions: from distribution of the payload among available resources to issue tracking related to any of 350k jobs running simultaneously. It evolves intensively; in particular, in 2017, the system received 933 commits, delivering new features and expanding the scope of the presented data. The experience of operating BigPanDA in 24/7 mode led to development of a multilevel self-monitoring alarm system. This ELK-stack based solution covers all critical components of the BigPanda: from user authentication to management of the number of connections to the DB backend. The developed solution provides an intelligent error analysis, delivering to the operators only those notifications that need human intervention. We describe the architecture, principal features, and operation experience of self-monitoring, as well as its adaptation possibilities.

IMPROVING NETWORKING PERFORMANCE OF A LINUX NODE

Alexander Bogdanov¹, Nabil Ahmed², Vladimir Gaiduchok^{2,a}

¹*St Petersburg State University*

²*Saint Petersburg Electrotechnical University "LETI"*

E-mail: ^agvladimiru@gmail.com

Linux networking performance is excellent. Linux networking stack is perfect, it is very effective. There are many options that could be configured for different cases. This article is devoted to questions related to networking stack implementation and configuration. Questions related to effective configuration will be discussed. Possible problems and solutions will be shown. Different options for configuration will be discussed. Questions will be discussed on the example of a computational cluster as well as on the example of a generic node. Effective usage of existing solutions is very important in both cases. Different aspects of configuration of the networking stack in general, configuration of TCP/IP stack, configuration of networking interfaces, real as well as virtual, are discussed in this article. In the end several recommendations are given.

APPLICATION OF UNIFIED MONITORING SYSTEM IN LHAASO

Haibo Li, Qingbao Hu^a, Qiulan Huang, Wei Zheng, Xiaowei Jiang

Institute of High Energy of Physics, Chinese Academy of Science

E-mail: ^ahuqb@ihep.ac.cn

LHAASO The on line machinecomputer room of LHAASO experiment located at has high altitude and poor natural environment. As t, and there is no permanent resident maintenance manpowerpersonnel, so it needs to deploy an automatic operation and maintenance system for the remote management. According to the characteristics of the LHAASO cluster management, we have designed a distributed monitoring framework, to support the site monitoring and management. In this framework, the monitoring data is collected in real-time at remote site, then the data is compressed and transferred the data back to IHEP. The servers at IHEP and used to analyze the data and display the running status via web page. This monitoring system monitors can monitoring and displays the machine performance of both physical machine and virtual machine, cluster's service status, job status and equipment energy consumption information in real-time. The system; detects abnormal equipment and givesfor real-time alarm; It creates and destroys the virtual machines based on physical machine states to maintain adequate computing capacity; provide accurate cause of failure for temporary maintenance staff at LHAASO personnel.

BIGDATA TOOLS FOR THE MONITORING OF THE ATLAS EVENTINDEX

Andrei Kazymov^{1,a}, Evgeny Alexandrov^{1,b}, Fedor Prokoshin²

¹*Joint Institute for Nuclear Research*

²*Universidad Tecnica Federico Santa Maria*

E-mail: ^akazymai@jinr.ru, ^baleksand@jinr.ru

The ATLAS EventIndex collects event information from data both at CERN and Grid sites. It uses the Hadoop system to store the results, and web services to access them. Its successful operation depends on a number of different components, that have to be monitored constantly to ensure continuous operation of the system. Each component has completely different sets of parameters and states and requires a special approach. A scheduler runs monitoring tasks, which gather information by various methods: querying databases, web sites and storage systems, parsing logs and using CERN host monitoring services. Information is then fed to Grafana dashboards via InfluxDB. Using this platform allowed much faster performance and flexibility compared to the previously used Kibana system.

PERFORMANCE MEASUREMENTS FOR THE WLCG COST MODEL

Victoria Matskovskaya

Plekhanov Russian University of Economics

E-mail: viktoriyams@mail.ru

High energy physics community needs metrics that allow to characterize the resource usage of the experiments workloads detailed enough so that the impact of changes in the infrastructure or the workload implementations can be quantified with a precision high enough to guide design decisions towards improved efficiencies. This model has to express the resource utilization of the workloads in terms of fundamental capabilities that computing systems provide, such as storage, memory, network, computational operations, latency, bandwidths etc. To allow sites and user communities use this model to improve also their cost efficiency, an approach to map these capabilities to local costs is highly desirable. This can't be achieved at a global level, since the conditions at different grid sites are too different, but the model should be constructed in such a way that this mapping on a local level can be done easily, following given examples. Decisions on the evolution of workloads, workflows and infrastructures impact the quantity and quality of human resources required to build and operate the system. It is important that a cost and performance model at the system level takes these adequately into account to allow to optimize the global infrastructure cost into a constrained budget. In this report there are presented methods and results of grid sites benchmarking with typical HEP tasks. Comparative analysis and correlation studies of these results against the data from accounting portals (Rebus, etc.) are discussed.

SEARCH FOR ANOMALIES IN THE COMPUTATIONAL JOBS OF THE ATLAS EXPERIMENT WITH THE APPLICATION OF VISUAL ANALYTICS

**Alexei Klimentov¹, Igal Milman², Maria Grigorieva^{3,a},
Mikhail Titov³, Tatiana Korchuganova⁴, T. Galkin²,
Victor Pilyugin²**

¹Brookhaven National Lab

²National Research Nuclear University “MEPHI”

³National Research Center “Kurchatov Institute”

⁴National Research Tomsk Polytechnic University

E-mail: ^amaria.grigorieva@cern.ch

ATLAS is the largest experiment at the LHC. It generates vast volumes of scientific data accompanied with auxiliary metadata. These metadata represent all stages of data processing and Monte-Carlo simulation, as well as characteristics of computing environment, such as software versions and infrastructure parameters, detector geometry and calibration values. The systems responsible for data and workflow management and metadata archiving in ATLAS are called Rucio, ProdSys2, PanDA and AMI. Terabytes of metadata were accumulated over the many years of systems functioning. These metadata can help physicists carrying out studies to evaluate in advance the duration of their analysis jobs. As all these jobs are executed in a heterogeneous distributed and dynamically changing infrastructure, their duration may vary across computing centers and depends on many factors, like memory per core, system software version and flavour, volumes of input datasets and so on. Ensuring the uniformity in jobs execution requires searching for anomalies (for example, jobs with too long execution time) and analyzing the reasons of such behavior to predict and avoid the recurrence in future. The analysis should be implemented on the basis of all historical jobs metadata that are too large to be processed and analyzed by standard means. Detailed analysis of the archive can benefit from application of visual analytics methods providing more easy way of navigation within the multiple internal data correlations. Presented research is the starting point in this direction. The slice of ATLAS jobs archive was analyzed visually, demonstrating the most and the less efficient computing sites. Then, the efficient sites will be compared to inefficient to find out parameters affecting jobs execution time or indicating possible time delays. Further work will concentrate on the increasing of the amount of analyzed jobs and the development of the interactive 3-dimensional visual models, facilitating the interpretation of analysis results.

MECHANISMS FOR ENSURING THE INTEGRITY OF INFORMATION IN DISTRIBUTED COMPUTING SYSTEMS IN THE LONG-TERM PERIOD OF TIME

A.S. Minzov^a, A.Y. Nevskii, O.R. Baronov

National Research University "MEI"

E-mail: ^aminzovas@mpei.ru

The article discusses issues of ensuring the integrity of information over a long period of time. This task was not raised earlier. However, experience shows that in the long periods of time in electronic archives there can be an uncontrolled change in information and even its disappearance. Attacks on the integrity of electronic archives can be targeted. This requires the creation of information technology to ensure the integrity of archives. The work is devoted to the mechanism of the integrity of information in the electronic archive by creating a distributed managed trusted environment. This allows you to track the processes, data, user actions and make decisions about the choice of the owners of the archive, restore the archive with a partial loss of information in it and meet attacks on the integrity of the archive.

DEVELOPMENT OF JINR TIER-1 SERVICE MONITORING SYSTEM

**Igor Pelevanyuk^a, Ivan Kadochnikov, Tatiana Strizh,
Valery Mitsyn, Vladimir Korenkov**

Joint Institute for Nuclear Research

E-mail: ^agavelock@gmail.com

Tier-1 center for CMS in JINR has been successfully operating since 2015. Monitoring is an important aspect of ensuring its performance. Hardware monitoring of the Tier-1 center had been introduced at construction time and was constantly upgraded with the center. The scientific community makes use of the resources through Grid services that depend on more low-level services. A dedicated monitoring system has been developed to keep an eye on the state of all services related to Tier-1 operations. The main object of the monitoring system is to collect data from different sources, process it and provide a comprehensive overview on a web page. The mechanism was implemented to allow determining status by analyzing collected data. The notion of event was introduced to allow reactions on ongoing changes of all services. The whole system consists of core libraries and monitoring modules. A monitoring module may unite functionality related to data collection, analysis, visualization, possible statuses and events, and reactions on events. This allows building flexible monitoring modules which together form a Tier-1 service monitoring system.

TRIGGER INFORMATION DATA FLOW FOR THE ATLAS EVENTINDEX

Fedor Prokoshin¹, Mikhail Mineev^{2,a}

¹*Universidad Tecnica Federico Santa Maria*

²*Joint Institute for Nuclear Research*

E-mail: ^amineev@jinr.ru

The trigger information is an important part of the ATLAS event data. In the EventIndex trigger information is collected for various use cases, including event selection and overlap counting. Decoding the trigger information from the event records, stored as a bit mask, requires additional input from the conditions metadata database, as trigger configurations evolve with time. It depends on the run number for the real data and from the simulation settings for Monte-Carlo data. We describe trigger information handling in the EventIndex and the interfaces used to access it.

ИССЛЕДОВАНИЕ ОСОБЕННОСТЕЙ ИНТЕРНЕТ- ТРАФИКА В МАГИСТРАЛЬНОМ КАНАЛЕ

**Иванов Виктор¹, Зрелов Петр¹, Крюков Юрий²,
Татаринов Иван^{3a}, Иванов Валерий¹**

¹*Объединенный институт ядерных исследований*

²*Государственный университет «Дубна»*

³*Kaspersky Lab*

E-mail: ^aivan-tatarinov@yandex.ru

В работах [1, 2] анализировались статистические особенности потоков информации на входном шлюзе локальной сети среднего размера (250-300 компьютеров) и было показано, что при агрегировании измерений трафика формируется (начиная с некоторого порогового значения окна агрегирования: в нашем случае 1 с) статистическое распределение величины потока, которое не меняет своей формы при дальнейшем росте окна агрегирования (вплоть до 10 с). Было показано, что указанное распределение с высокой точностью аппроксимируется логнормальным распределением. Авторами работы [3] также проводилось агрегирование измерений трафика и был сделан вывод о том, что “гистограмма байтовой интенсивности соответствует логнормальному распределению”. Однако никакого обоснования этого утверждения не было приведено. Более того, в указанной гистограмме наблюдается пик в области малых интенсивностей, не согласующийся с логнормальным законом. В работе [4] исследовались статистические характеристики Интернет-трафика в магистральном канале при трех значениях времени агрегирования: 1 мс, 10 мс и 100 мс. Как отмечали авторы данной работы, полученные ими графики плотностей распределения вероятностей величины интенсивности трафика не удалось аппроксимировать каким-либо из известных распределений. В настоящем исследовании на основе анализа измерений Интернет-трафика в магистральном канале,

взятых с того же сайта [5], что и в работе [4], показано, что при агрегировании измерений трафика формируются статистические распределения, которые, в зависимости от периода наблюдения, с высокой точностью аппроксимируются логнормальным, либо двумя логнормальным распределениями.

Список литературы:

[1] I. Antoninou, V.V. Ivanov, Valery V. Ivanov, and P.V. Zrelov: On the Log-Normal Distribution of Network Traffic, *Physica D* 167 (2002) 72-85.

[2] I. Antoninou, V.V. Ivanov, Valery V. Ivanov, and P.V. Zrelov: Statistical Model of Network Traffic, “Физика элементарных частиц и атомного ядра” (ЭЧАЯ), 2004, Т.35, Вып.4 (984-1019).

[3] Ю.А. Крюков, Д.В. Чернягин: Исследование самоподобия трафика высокоскоростного канала передачи пакетных данных// III Международная научная конференция “Современные проблемы информатизации в системах моделирования, программирования и телекоммуникациях”. Электронный ресурс. Труды конференции. Москва, 2009. 8 стр. URL: <http://econf.rae.ru/article/4819>.

[4] Д.В. Симаков, А.А. Кучин: Анализ статистических характеристик Интернет-трафика в магистральном канале, //Т-Comm: Телекоммуникации и транспорт. 2015, Том 9, №5, С. 31-35.

[5] MAWI Working Group Traffic Archive. URL: <http://mawi.wide.ad.jp/mawi/>

РАЗРАБОТКА ПЕРСПЕКТИВНОЙ СИСТЕМЫ СБОРА ДАННЫХ НА ОСНОВЕ TRB-3

**Александр Курепин¹, Андрей Кондратьев^{2,а},
Константин Сказыткин³, Николай Курепин³**

¹*ЦЕРН, Европейская организация по ядерным исследованиям*

²*Объединенный институт ядерных исследований*

³*Институт ядерных исследований РАН*

E-mail: ^аkondratyev@jinr.ru

В связи с увеличением объема информации получаемой в ходе эксперимента ALICE на Большом Адронном Коллайдере, повышаются требования к системам сбора данных с детекторов, например, увеличение пропускной способности. Одним из возможных методов решения данной проблемы является использование TRB-3 платформы. Решение представляет собой глубокую модернизацию существующей модели сбора данных.

**MIDDLEWARE AND SERVICES
FOR PRODUCTION-QUALITY
INFRASTRUCTURE**

CREATING TOOLS TO ASSIST IN DEVELOPMENT OF CMS SOFTWARE

George Adamov

Joint Institute for Nuclear Research

E-mail: adamov@jinr.ru

Software packages, created for the modern physics experiments, present a sets intertwined code structures, written by different people and bundled together to perform range of different reconstruction and analysis tasks on the experimental data. Sometimes due to complicated nature of such frameworks a new set of tools is required to simplify their further development. In this work we investigate an example of such tool, created for the CMS experiment to analyse the structure of it's software components.

NEW METHODS OF MINIMIZING THE ERRORS IN THE SOFTWARE

Elizaveta Dorenskaya

NRC "Kurchatov Institute" ITEP

E-mail: dorenskaya@itep.ru

One of the qualitative ways to minimize software errors is to check the code by as many users as possible (the “many eyes” principle). We propose a new approach which goes well with this technique. This is the minimization of human participation in writing program codes. The implementation of this approach has be made in such a way that the machine itself creates the program code according to the description provided by the user. Due to the re-assignment of writing the program code to the machine the process of its generation is simplified simultaneously and the number of program errors is reduced. The latter happens due to the reduction of the human factor influence. By simplifying the writing of the program code, the number of people capable of generating it increases and the period of training in programming and the time spent on writing a separate program are reduced. Our methods do not completely eliminate software errors, because they can be both in the user’s own description and in the interpreter. But, nevertheless, it is very important to maximally minimize the number of software errors, because already now the software determines many aspects of our life and the number of its applications is increasing. Even now, for example, such important areas of our life as health and finance may depend from the quality of software and the number of errors in it.

There were written several programs which implement some ways to minimize errors in the “Perl” language. The author's certificates were received. Each of these programs includes a method to minimize errors. In total, these programs are three:

* Dialog method. It is implemented so that the program itself asks questions to the user and selects the desired code according to his answers. This method has a significant drawback is

to describe and predict all the options of the dialogue is impossible, because their number tends to infinity.

* Method for determining the appropriate condition by the specified values. It is implemented with the help of “Gnuplot” charting program. It consists in the fact the program chooses which chart to build based on the given data.

* A method for interpreting a condition using special words. It consists in the fact that a special simplified language or a set of words is invented, with the help of which more complex conditions are written. This simplifies the process of writing code.

CURRENT STATUS OF SOFTWARE DEVELOPMENT FOR THE MPD AND BM@N EXPERIMENTS

Konstantin Gertsenberger

Joint Institute for Nuclear Research

E-mail: gertsen@jinr.ru

The software for detector simulation, reconstruction and analysis of physics data is an essential part of each high-energy physics experiment. A new generation of the experiments for the relativistic nuclear physics is expected to be started up in the nearest years at the Nuclotron-based Ion Collider facility (NICA) being under construction at the Joint Institute for Nuclear Research in Dubna: the fixed target experiment BM@N (Baryonic Matter at Nuclotron), whose technical runs were started in 2015, and the future MPD (Multi-Purpose Detector) experiment on ion collisions, which will operate at the storage rings of the NICA facility. The event data model of the experiments is shown. The status of the software frameworks MpdRoot and BmnRoot developed for the MPD and BM@N is considered. For these tasks many additional systems, such as raw data converter, monitoring systems, event display, databases, parallelization tools and others have been developed.

THE ATLAS PRODUCTION SYSTEM PREDICTIVE ANALYTICS SERVICE: AN APPROACH FOR INTELLIGENT TASK ANALYSIS

**Alexei Klimentov¹, Dmitry Golubkov², Mikhail Borodin³,
Mikhail Titov^{4,a}**

¹Brookhaven National Lab

²Institute for High Energy Physics

³The University of Iowa

⁴NRC "Kurchatov Institute"

E-mail: ^amikhail.titov@cern.ch

The second generation of the Production System (ProdSys2) of the ATLAS experiment (LHC, CERN), in conjunction with the workload management system - PanDA (Production and Distributed Analysis), represents a complex set of computing components that are responsible for defining, organizing, scheduling, starting and executing payloads in a distributed

computing infrastructure. ProdSys2/PanDA are responsible for all stages of (re)processing, analysis and modeling of raw and derived data, as well as simulation of physical processes and functioning of the detector using Monte Carlo methods. The prototype of the ProdSys2 Predictive Analytics (P2PA) is an essential part of the growing analytical service for the ProdSys2 and will play the key role in the ATLAS computing. P2PA uses such tools as Time-To-Complete (TTC) estimation towards units of the processing (i.e., tasks, chains and groups of tasks) to control the processing state and rate, and to be able to highlight abnormal operations and executions (e.g., to discover stalled processes). It uses methods and techniques of machine learning to obtain corresponding predictive models and metrics that are aimed to characterize the current system's state and its changes over the close period of time.

EVENT BUILDING FROM FREE STREAMING DATA AT THE CBM

Ivan Korolko, Mikhail Prokudin^a

NRC "Kurchatov Institute" ITEP

E-mail: ^amikhail.prokudin@cern.ch

The CBM will be the first experiment employing a new data streaming technique. All data collected from the CBM detector will be transported to computer farm. Physical objects (such as tracks and vertexes) will be reconstructed at the real time and interesting events will be stored for further detailed analysis. A unit of data in this approach is a timeslice – all data collected from the detector in a given period of time. Each timeslice contains data from many heavy ions collisions and may be treated independently from other timeslices at different nodes of computing farm. Data produced by particles originating from individual heavy ions collision (event) should be used for physical analysis rather than free streaming data. Event building can be performed at different data levels. The simplest event building technique works at the level of individual activations of readout electronics channels (digis). Each digi contains information about activation time, readout channel number, etc. Event building can be divided on two steps. The first one is an event finding and the second one is an event composition, when data corresponding to found event collected from several subdetectors. Event finding is performed using data only from a given CBM subdetector. This subdetector should be fast, has good time resolution and low noise levels. In general, event is found if the number of digis in a given time window exceeds a given threshold which depends on colliding system and interaction rate. Currently STS and BFTC considered as a data sources for event finding. General event composition method for different subdetectors has been developed and tested. The digi from a given subdetector is attributed to the event if its time after correction lies in an acceptance time window. The acceptance time window should be extended according to time resolution of the subdetector. r/c should be subtracted from the digi time for the correction, where r is a distance between triggered readout channel and c is a speed of light. This event composition method works for all subdetectors except calorimeters.

EXPERIENCE WITH ITEP-FRRC HPC FACILITY

Ivan Korolko^a, Mikhail Prokudin, Victor Kolosov

NRC "Kurchatov Institute" ITEP

E-mail: ^aivan.korolko@cern.ch

ITEP-FRRC HPC facility was built in 2011-2014 as a common project of State Atomic Energy Corporation ROSATOM and Helmholtz Association of German Research Centers. It utilizes the concept of “green computing” which was invented by GSI/FAIR scientists. Facility is used for FAIR related studies by various groups from ITEP and other Russian physics centers. After 7 years of successful HPC facility operation we want to summarize the experience we got running the hardware and supporting the requested software.

DDS – THE DYNAMIC DEPLOYMENT SYSTEM

Anar Manafov, Andrey Lebedev^a

GSI

E-mail: ^aandrey.lebedev@gsi.de

The Dynamic Deployment System (DDS) is a tool-set that automates and significantly simplifies a deployment of user-defined processes and their dependencies on any resource management system (RMS) using a given topology. DDS is a part of the ALFA framework. A number of basic concepts are taken into account in DDS. DDS implements a single responsibility principle command line tool-set and API. The system treats users' task as a black box – it can be an executable or a script. It also provides a watchdogging and a rule-based execution of tasks. DDS implements a plug-in system to abstract from RMS. Additionally it ships an SSH and a localhost plug-ins which can be used when no RMS is available. DDS doesn't require pre-installation and pre-configuration on the worker nodes. It deploys private facilities on demand with isolated sandboxes. The system provides a key-value property propagation engine. That engine can be used to configure tasks at runtime. DDS also provides a lightweight API for tasks to exchange messages, so-called, custom commands. In this report a detailed description, current status and future development plans of the DDS will be highlighted.

THE CONCEPT OF PROACTIVE PROTECTION IN A DISTRIBUTED COMPUTING SYSTEM

Anatoly Minzov¹, Pavel Osipov^{2,a}

¹National Research University "MEI"

²Dubna State University

E-mail: ^aosipov.pavel92@gmail.com

The paper considers modern problems of information security, the main stages of the implementation of attacks. The stages of implementing the concept of proactive protection, a technique for describing possible attacks are described, an example of a prototype of a proactive defense system is described.

COMPASS PRODUCTION SYSTEM: PROCESSING ON HPC Artem Petrosyan

Joint Institute for Nuclear Research

E-mail: artem.petrosyan@jinr.ru

Since the fall of 2017 COMPASS processes data on heterogeneous computing environment, which includes computing resources at CERN and JINR. Computing sites of the infrastructure work under management of workload management system called PanDA (Production and Distributed Analysis System). At the end of December 2017, integration of BlueWaters HPC to run COMPASS production jobs has begun. Despite an ordinary computing site, each HPC has many specific features, which make it unique, such as: hardware, batch system type, job submission and user policies, et cetera. That is why there is no ready solution out of the box for any HPC, development and adaptation is needed in each particular case. PanDA Pilot has a version for processing on HPCs, called Multi-Job Pilot, which was prepared to run simulation jobs for ATLAS. To run COMPASS production jobs, an extension of Multi-Job Pilot was performed. Details of the new computing resource integration into COMPASS Production System are described in the report.

PARTICIPATION OF RUSSIAN INSTITUTIONS IN THE PROCESSING AND STORAGE OF ALICE DATA

Andrey Zarochentsev

St Petersburg State University

E-mail: andrey.zar@gmail.com

The report presents the results of the work of Russian institutes in the processing of ALICE experiment data during the last 2 years of the LHC RUN2. The main issues and tasks facing both ALICE Grid Computing and its Russian segment before the start of RUN3 (2020 - onward) are considered. Plans and preparations for the operation of the LHC in the HL (High luminosity) mode are presented. The challenges of support and modernization of the existing resources and networking are discussed.

**SCIENTIFIC, INDUSTRY
AND BUSINESS APPLICATIONS
IN DISTRIBUTED COMPUTING
SYSTEMS**

USAGE OF THE DISTRIBUTED COMPUTING SYSTEM IN THE RECOVERY OF THE SPECTRAL DENSITY OF SEA WAVES

Ilya Busko

St Petersburg State University

E-mail: trassae95st@mail.ru

This article presents a task of the recovery of the spectral density of sea waves in the linear case. Creation of the onboard ship system giving the current information about sea state and weather forecast in the navigation area is one of the most urgent problem. Weather forecast can be based on the analysis of the sea waves spectral density change. Evaluation of the sea wave spectral density is solved on the basis of indirect dynamic measurements of vibrational motion of the marine dynamic object in a seaway. The first researcher to raise the wave parameter identification problem on the basis of object behavior was Y. Nechayev [1,2]. Over the past fifteen years, this problem has become rather popular and the works of Nielsen [3,4,5] and others are of the most significance. Nevertheless, despite of researches large number it is still impossible to speak of an acceptable effective solution to this problem. The recovery of the sea waves on the basis of the behavior of the marine dynamic object requires the analysis and processing of large amounts of information. To improve the accuracy of identification requires using different algorithm of recovery and a large number of test calculations. The calculations should be made in real time. The system should also store processed data and provide access at any time. The software should have the fault-tolerance property, i.e. the software should continue to work in the case of failure of one of the parts. All these requirements and features make us to use distributed computing system for developing software of the solution of the problem.

References

- [1] Nechaev Y. I., The collection of reports on the scientific and technical conference on experimental fluid mechanics (1990)
- [2] Nechaev Y. I., Navigation and Hydrography 3 (1996)
- [3] Nielsen U. D., Stredulinsky D. C., Proceedings of the 12th International Ship Stability Workshop, pp.61-67 (2011)
- [4] Simons A. N., Tannuri E. A., Sparano J. V., Matos V. L. F., Applied Ocean Research v.32, i.2, pp.191-208 (2010)
- [5] Pascoal [Pascoal R., C. Guedes Soares., Ocean Engineering v.36, i.6-7, pp.477-488 (2009)

SELECTION OF RATIONAL COMPOSITION OF IT-SERVICES OF INFORMATION SYSTEM WITH THE PURPOSE OF INCREASE OF EFFICIENCY OF TRANSPORT LOGISTICS COMPANIES FUNCTIONING

Alexander Degtyarev¹, Gennady Dik¹, Alexander Dik²

¹St Petersburg State University

²Saint Petersburg Electrotechnical University "LETI"

E-mail: ^ag.dick@prodick.ru

When the automation of transport logistics companies there is a problem with the gap between the existing business processes and means of automation. This circumstance makes it necessary to search for options for creating a software infrastructure (SI) for transport logistics enterprises. One of the solutions to this problem can be the transition of the enterprises to Service-Oriented Architecture (SOA). In essence, this means not considering the applied information system as a whole, but considering its individual functional components - IT services. The proposed SOA architecture is recommended to be formed from a set of business-oriented IT-services that collectively satisfy the tasks and business processes of the enterprise in transport logistics. When investigating the SI of the transport logistics system (TLS), it was established that the same function is often represented by different services as well as by different suppliers with different costs for providing access to services, speed of service provision, service availability, etc. Then the process of determining the necessary selection of services should be considered as a solution of the multicriteria problem of service composition for selected indicators. It is known that methods for solving multicriteria problems are divided into two groups. In this case, these groups are reduced to different strategies. The first strategy is based on the principle of the worst reaction of the external environment. The second strategy is based on the principle of equilibrium (the Nash principle). In addition, in most practical applications, the tasks of forming a rational composition of services have to be solved in conditions of significant uncertainty under the influence of the following factors: 1. The lack of a uniform and universally accepted methodology for the development and implementation of IT strategies at transport logistics enterprises. 2. Availability in the IT market of a big number of alternative IT solutions that implement similar functionality to automate the business processes of the enterprise. 3. The need to take into account the total costs associated with their acquisition and operation of IT services, etc. These factors form elements of uncertainty, which causes the need to use the mathematical apparatus of fuzzy sets. And then, the process of formation of a rational composition of IT services in the TLS infrastructure can be represented as a solution of multicriterial problem under fuzzy sets.

VIRTUAL TESTBED FOR NAVAL HYDRODYNAMIC PROBLEMS

**Alexander Degtyarev^{1a}, Yury Pylnev², Anatoly Eibozhenko²,
Alexander Bogdanov¹, Vladimir Korkhov¹, Ivan Gankevich¹**

¹*St Petersburg State University*

²*Engineering company "NEOTECH MARINE"*

E-mail: ^adeg@csa.ru

Complex modeling of the behavior of marine objects under the influence of real external excitation is the most important problem. At present, the accuracy of direct simulation of phenomena with known physics is comparable to the accuracy of the results obtained during the model experiment in towing tanks. Particularly relevant is the creation of such marine virtual testbed for full-featured simulators and when testing the knowledge base of onboard intelligent systems. Such integrated environment is a complex information object that combines the features of both the enterprise system and the high-performance modeling tool. Integrated environment based on these basic principles is designed to solve in real time the following problems: 1. Collection and analysis of information on the current state of dynamic object (DO) and the environment, remote monitoring of the state of objects. 2. Evaluation and coordination of joint actions of DOs, proceeding from current conditions, with the aim of optimally common problem solving. 3. Centralized support for decision-making by operators of DO control in non-standard situations, organization of information support for the interaction of decision-makers in the conduct of ongoing operations. 4. Computer modeling of possible scenarios of situation development with the aim of selecting the optimal control strategy. 5. Centralized control of technical means. 6. Cataloging and accumulation of information in dynamic databases. Modern architecture of computer systems (especially GP GPU) allows direct full-featured simulation of a marine object in real time. Efficient mapping to a hybrid architecture allows even the ability to render ahead of time under various scenarios. The report discusses general concept of high-performance virtual testbed development and the experience of creating on their basis full-featured simulators.

.NET CORE TECHNOLOGY IN SCIENTIFIC TASKS

Victor Dorokhin

Dubna State University

E-mail: victor.doroh@gmail.com

Today we have an established stack of tools to develop applications, systems and services for scientific purposes. However, not so long ago a new technology, called .NET Core appeared. It's supervised by Microsoft, but its open-source and developed by a wide community of programmers and engineers. That technology has a lot of advantages like high performance, simple and productive parallel programming abilities, support of high-level programming languages (C# 7.0, F#) and so on. But the main advantage in comparison with its predecessors is cross-platform abilities. Once written code can be natively compiled on a large number of platforms and hardware systems. It can be used on Windows, Linux, Mac and all Unix-based

operation systems. It supports different hardware components and processors, for example, it can be used on ARM processors. The technology supports containerization and can be used with Docker or Kubernetes. It gives an ability to develop applications with micro-service architecture from the box and provides convenient deployment tools. Now, in a set of tasks .NET Core surpasses currently used tools in scientific sphere tools like Python, Go, Ruby, Java, Node.JS and others. Also, it can be used to develop native desktop applications with HTML-based GUI for administration and monitoring purposes.

DIFFERENT APPROACHES FOR ELASTIC IMAGING USING MULTIPROCESSOR COMPUTING SYSTEMS

Vasily Golubev^a, Alena Favorskaya

Moscow Institute of Physics and Technology

E-mail: ^aw.golubev@mail.ru

At present, oil and natural gas form the basis of energy throughout the world. In view of significant depletion of reserves, the task of prospecting and exploration of new deposits is becoming increasingly important. In the industry the specific migration procedure is used to find contrast interfaces between geological layers with different properties. It should be noted that algorithms developed to date are constructed in the acoustic approximation of the medium, which leads to defects in migration images. In particular, subvertical boundaries are practically not restored. Prospective in terms of overcoming these shortcomings is the transition to a full elastic formulation of the tasks of seismic survey process. Despite rising computational complexity of the problem the goal can be achieved with the usage of modern HPC systems. The work is the continuation of the previously reported research about the usage of Born approximation as a new elastic imaging method. In this study different fundamental approaches were used. Born approximation and Kirchhoff approach were adopted to 2D and 3D elastic problems in the case of homogeneous background medium. The research software in C++/Mathematica was developed, and a set of calculations for simple geological models were carried out using multicore shared memory system. The assessment of the scalability shows high effectiveness.

The research was supported by the grant of the President of the Russian Federation No MK- 1831.2017.9.

The research software based on the new elastic imaging methods using different approaches: Born approximation, Kirchhoff approximation and full-wave simulation method was developed. The assessment of the scalability on multi-cores shared memory system shows approximately 90 % of effectiveness.

CACHE-FRIENDLY MEMORY TRAVERSAL TO IMPROVE PERFORMANCE OF GRID- CHARACTERISTIC METHOD

Andrey Ivanov^a, Nikolay Khokhlov

Moscow Institute of Physics and Technology

E-mail: ^aip-e@mail.ru

We consider well known cache optimization techniques to find the most efficient one when applied to grid-characteristic method. Grid-characteristic method is used to solve elastic wave equation. Elastic wave equation is hyperbolic system of equations inferred from model of linear elastic material, which describes propagation of elastic waves in deformable rigid bodies. The solution to this problem is important in seismic tomography and exploration geophysics. In this work only 2D scenario is studied, but it has an extension to 3D. Grid-characteristic method consists of a bunch of iterations over the array representing nodes of computational grid. The increase of performance due to change of array traversal order is connected with the spatial and time locality of memory accesses. The following 3 techniques are evaluated: bypassing of memory in rectangular blocks (block tiling), in blocks of diamond shape (diamond tiling), and in recurrently nested tiles of smaller size (hierarchical tiling). In the case of block tiling we achieved highest performance gain (about 15%). In contrast, performance with diamond tiling is declined by 6% and with hierarchical tiling is dropped by 13%. We assume that last two methods degrade performance because the amount of memory for single grid node is too large. Therefore all the necessary local nodes for block and hierarchical tiling can't fit simultaneously in L1 cache. We have concluded that block tiling is the most appropriate technique for grid-characteristic method optimization.

The reported study was funded by RFBR according to the research project № 18-07-00914A.

SCALABLE SEMANTIC VIRTUAL MACHINE FRAMEWORK FOR LANGUAGE-AGNOSTIC STATIC ANALYSIS

Maxim Menshchikov

St Petersburg State University

E-mail: maximmenshchikov@gmail.com

The more static program analysis spreads in the industry, the clearer it becomes that its real-world usage requires something more than just optimized algorithms to make execution fast. Distribution of computations to cluster nodes is one of the viable ways for speeding up the process. The research demonstrates one of the approaches for organizing distributed static analysis - "Semantic Virtual Machines", which is based upon symbolic execution of language-agnostic intermediate codes of programs. These codes belong to the developed language for program representation with semantic objects behind. As objects are fully serializable, they can be saved to and later retrieved from a distributed database. That opens up a possibility for making an analyzer spread to many cluster nodes. Semantic Hypervisor is the core

of the system which acts as a master node for managing runnable virtual machines and also a relay between them and a semantics storage. Main efforts are put into developing of intercommunication mechanisms which can be effective in most of the scenarios, as well as into integration of the approach in existing static analyzer. The study shows the architecture, describes its advantages and disadvantages, suggests solutions for the biggest problems observed during research. The empirical study shows improved performance compared to single node run, and the result is almost linearly scalable due to the high locality of data. The main result is the created scalable and unified language-agnostic architecture for static analysis with semantics data in a distributed storage behind of it. The significance of the research is in high achieved performance level alongside with a clean architecture.

ENABLING BIOLOGY, CHEMISTRY AND OTHER SCIENCES ON TITAN THROUGH BIGPANDA

**Alexei Klimentov^{1,a}, Kaushik De², Danila Oleynik³,
Ruslan Mashinistov⁴, Pavlo Svirin¹, Jack Wells⁵, Sergey Panitkin¹**

¹Brookhaven National Lab

²University of Texas at Arlington

³Joint Institute for Nuclear Research

⁴NRC "Kurchatov Institute"

⁵Oak Ridge National Laboratory

E-mail: ^aalexei.klimentov@cern.ch

The Oak Ridge Leadership Computing Facility (OLCF) is one of the most powerful HPC centers available to researchers from different scientific fields to solve some of the world's most challenging scientific problems. Small scientific groups often need to develop expertise to optimize their applications for running on Titan, and to fit the usage policies of such big machines. We have installed the BigPanDA workload management system at OLCF to simplify the submission of user tasks to Titan. In this talk we will present results of an R&D project to execute workloads from different scientific groups at OLCF. We will describe all steps: starting from deployment of PanDA server as service on demand at OLCF in OpenShift containers, to the adaptation of PanDA client tools for new users. Examples from some of the different scientific fields using this service will include biology/genomics, molecular dynamics, LQCD, solid-state and neutrino physics, and different data science experiments: nEDM, LSST, and IceCube. In more details we will address a "proof of concept" project with BlueBrain. It was conducted jointly by the BigPanDA team and the Blue Brain Project (BBP) of the Ecole Polytechnique Federal de Lausanne (EPFL). This proof of concept project showed the efficient application of the BigPanDA system to support the complex scientific workflow of the BBP using a mix of desktop, cluster and supercomputers to reconstruct and simulate accurate models of brain tissue.

MOLECULAR DYNAMIC SIMULATION OF WATER VAPOR INTERACTION WITH VARIOUS TYPES OF PORES USING HYBRID COMPUTING STRUCTURES

Eduard Nikonov^{1a}, Vladimir Korenkov¹, Maria Popovičová²

¹*Joint Institute for Nuclear Research*

²*University of Prešov, Slovakia*

E-mail: ^ae.nikonov@jinr.ru

Theoretical and experimental investigations of water vapor interaction with porous materials are very needful for various fields of science and technology. Not only studies of the interaction of water vapor and porous material as a continuous medium, but also the study of the interaction of water vapor with individual pore is very important in these researches. Mathematical modeling occupies an important place in these investigations. Conventional approaches to solve problems of mathematical research of the processes of interaction of water vapor with individual pore are the following. The first approach is based on the use of diffusion equation for description of interaction of water vapor with a pore. It is so called macro approach. The second approach is based on various particle methods like, for example, molecular dynamics (MD). These methods essentially consider the microstructure of the investigated system consisting of water vapor and a pore. This second approach can be called a micro approach. At the macro level, the influence of the arrangement structure of individual pores on the processes of water vapor interaction with porous material as a continuous medium is studied. At the micro level, it is very interesting to investigate the dependence of the characteristics of the water vapor interaction with porous media on the geometry and dimensions of the individual pore. Both approaches require the most efficient calculation methods as far as possible with the current level of development of computational technologies. Usage of efficient calculation methods is necessary because the degree of approximation for simulating system is largely determined by the dimensionality of the system of equation being solved at every time step. Number of time steps is also quite large. In this work, a study of efficiency of various implementations algorithms for MD simulation of water vapor interaction with individual pore is carried out. A great disadvantage of MD is its requirement of a relatively large computational effort and long time in simulations. These problems can be drastically reduced by parallel calculations. In this work we investigate dependence of time required for simulations on different parameters, like number of particles in the system, shape of pores, and so on. The results of parallel calculations are compared with the results obtained by serial calculations.

This work was supported by the JINR project No. 05-6-1118-2014/2019, protocol No. 4596-6-17/19, and used HybriLIT resources.

USING EXTENDED REALITY TECHNOLOGIES IN DISTRIBUTED COMPUTER SYSTEMS

Nadezhda Vozdvizhenskaya

Dubna State University

E-mail: n.vozdvizhenskay@gmail.com

Over the past few decades, web technologies have proven to be a fast, convenient and easy-to-access tool for retrieving information and sharing a large amount of heterogeneous data. The technologies used there, and in particular the HTML technology - have played a major role in the development of the Internet as it is. This is all due to the standardization and creation of a single tool to create network content. The goal of this work is to create a standard and a language for developing extended reality applications and interfaces built into existing applications. With this language, developers who are familiar with web technologies can quickly and with minimal effort move on to new technology and fill it with content, the lack of which is now the main problem of all extended reality technologies. With usage of the developed system it will be possible to combine the formed community of web developers and perspective technology on the basis of a standardized set of tools that will positively affect both developers and the pace of technology development. In the future that technology can be used in different spheres, like education, business, advertisement, etc. Also it can be used in global distributed computer systems, what will give an ability to make a global network of virtual objects referred to real-life points.

РАЗРАБОТКА ДЕЦЕНТРАЛИЗОВАННОЙ ПЛАТЕЖНОЙ СИСТЕМЫ НА ОСНОВЕ ТЕХНОЛОГИИ BLOCKCHAIN С УЧЕТОМ СПЕЦИФИКИ МОБИЛЬНЫХ ПЛАТФОРМ

Андрей Илюхин^{1а}, Эдуард Никонов²

¹Государственный университет «Дубна»

²Объединенный институт ядерных исследований

E-mail: ^аa.iluhin@nordavind.ru

Интернет находится в центре революций: централизованные проприетарные сервисы подвергаются замене на их децентрализованные аналоги со свободными лицензиями; доверенные третьи стороны юридических и финансовых договоров заменяются верифицируемыми вычислениями; неэффективные монолитные сервисы уступают место одноранговым алгоритмическим рынкам. Bitcoin, Ethereum и другие сети, фундаментом которых является технология blockchain, доказали полезность децентрализованных регистров транзакций. Имея в основе децентрализованные, открытые базы данных, они поддерживают выполнение сложных «умных» контрактов (smart contracts) и обслуживают крипто-активы стоимостью десятки миллиардов долларов. Эти системы являются первыми экземплярами межсетевых открытых сервисов, в которых участники образуют децентрализованную сеть, предоставляющую

полезные услуги для коммерции, без централизованного управления или доверенных лиц. Парадигма открытости и децентрализации коснулась не только мира коммерции, но и систем хранения и обработки больших объемов данных. InterPlanetary File System показала полезность адресации контента путем децентрализации самой всемирной паутины, обслуживая миллиарды файлов, используемых в глобальной одноранговой сети. Но достижение децентрализации и ухода от доверенных третьих лиц обернулось высокими требованиями к ресурсоемкости узлов сети и потерей масштабируемости, что препятствует массовой адаптации данных систем. В особенности данная проблема проявляется в виде обхода стороной блокчейн-технологиями мобильных платформ. Между тем, в октябре 2016 использование интернета мобильными и планшетными устройствами впервые превысило ПК по всему миру в соответствии с информацией от независимой веб-аналитической компании StatCounter¹. В дальнейшем тенденция роста числа мобильных узлов в сети будет сохраняться. Проблема масштабируемости не дает возможностей конкурировать с централизованными системами электронных платежей, таких как Visa, которая обеспечивает обработку порядка 65000 транзакций в секунду². В частности, это ограничивает интеграцию с «интернетом вещей» — перспективного направления цифровой экономики.

Целью данной работы является разработка распределенной сети на основе blockchain для мобильных платформ. Выдвигается концепт для преодоления вышеуказанных ограничений текущих blockchain-проектов. Уход от таких механизмов верификации как в Ethereum и Bitcoin, использующих для достижения консенсуса между участниками сети сложные вычислительные алгоритмы. В качестве их замены выступают ресурсоэффективные алгоритмы консенсуса Proof-of-Stack. Проблемы бесконечного роста цепочки блоков находится в плоскости организации распределенного хранения данных: эффективного алгоритма выбора массива блоков для хранения узлом с учетом необходимого коэффициента репликации. Подсети на основе системы каналов «узел-узел» для микроплатежей призваны решить проблему масштабируемости.

ИМИТАЦИОННАЯ МОДЕЛЬ БРЛК С СИНТЕЗИРОВАННОЙ АПЕРТУРОЙ АНТЕННЫ В СЕТИ РАСПРЕДЕЛЕННЫХ ВЫЧИСЛЕНИЙ MARGRID

Владимир Безродный^а, Анатолий Леухин, Андрей Воронин

Марийский государственный университет

E-mail: ^аvova.bezrodny@gmail.com

Синтезирование апертуры представляет собой технический прием, позволяющий существенно повысить разрешающую способность радиолокатора в поперечном относительно направления полета направлении и получить детальное изображение

¹ <http://gs.statcounter.com/press/mobile-and-tablet-internet-usage-exceeds-desktop-for-first-time-worldwide>

² <https://usa.visa.com/dam/VCOM/global/about-visa/documents/visa-facts-figures-jan-2017.pdf>

радиолокационной карты местности, над которой совершает полет ЛА [1]. Для моделирования отражения сигнала от поверхности используется фацетная модель поверхности, представляющая поверхность в виде совокупности элементарных отражающих элементов, представляющих собой пластины конечных размеров, совпадающие с поверхностью крупномасштабных неровностей [2]. Отраженный сигнал от поверхности представляет собой сумму сигналов от всех облучаемых фацетов. Причем каждый парциальный сигнал имеет свою амплитуду, определяемую ориентацией локальной диаграммы обратного рассеяния, и свою случайную фазу. Сигнал на входе приемной антенны РЛС представляет собой сумму парциальных сигналов, отраженных от всех фацетов в облучаемой области [3]. При моделировании среды распространения сигнала используется явление рефракции, которое объясняется изменением диэлектрической проницаемости и, соответственно показателя преломления воздуха с высотой. Нижние слои воздуха рассматриваются как среда, диэлектрическая проницаемость которой изменяется с высотой вследствие разрежения воздуха. При прохождении луча наблюдается его отклонение от прямолинейного распространения, вызванное наличием градиента показателя преломления. Сложность моделирования заключается в большом объеме исходных данных. Например, для несущей частоты 10 ГГц (X – диапазон) плотность фацетов составит 44.4 эл./м². Для каждого фацета всей подстилающей поверхности на каждый момент излучения антенны требуется определить: принадлежность диаграмме направленности антенны, в тени или нет, дальность до антенны, угол падения луча, фазу, мощность сигнала, которая зависит от поляризации излучения и от взаимного направления поляризации при излучении и приеме. Для решения проблемы разработано клиент-серверное приложение для распределенного моделирования радиолокационной системы с синтезированной апертуры антенны. Поскольку, каждый фацет можно обрабатывать независимо используется параллелизм на уровне данных. Каждому клиенту соответствует определенная область подстилающей поверхности.

Список литературы

- [1] Антипов В.Н. Радиолокационные станции с цифровым синтезированием апертуры антенны / В.Н. Антипов и др.; Под ред. В.Т.Горяинова. – М.: Радио и связь, 1988. – 304 с.
- [2] Баскаков А.И. Локационные методы исследования объектов и сред: учебник для студ. Учреждений высш. проф. Образования / А.И. Баскаков, Т.С. Жутяева, Ю.И. Лукашенко; под. Ред. А.И. Баскакова. – М.: Издательский центр «Академия», 2011. – 384 с.
- [3] Карасев Д.В. Математическое моделирование процесса формирования радиолокационного изображения для полно-поляризационных радаров с синтезированной апертурой / Д.В. Карасев, В.А. Карпов, В.И. Безродный, Н.А. Кокоихина. - Сборник трудов XXIII Международной научно-технической конференции, 2017. – 874-880 с.

NRV WEB KNOWLEDGE BASE: SCIENTIFIC AND EDUCATIONAL APPLICATIONS

**Aleksandr Alekseev¹, Alexander Karpov², Andrey Denikin²,
Mikhail Naumenko², Viacheslav Samarin², Vladimir Rachkov^{2,a},
Vyacheslav Saiko²**

¹*Chuvash State Universit*

²*Joint Institute for Nuclear Research*

E-mail: ^arachkov@jinr.ru

The NRV web knowledge base on low-energy nuclear physics has been created in the Joint Institute for Nuclear Research. This knowledge base working through the Internet integrates a large amount of digitized experimental data on the properties of nuclei and nuclear reaction cross sections with a wide range of computational programs for modeling of nuclear properties and various processes of nuclear dynamics which run directly in the browser of a remote user. Today, the NRV knowledge base is both a powerful tool for nuclear physics research and an educational resource. The system is widely used, as evidenced by the large number of user queries to its resources and the number of references to the knowledge base in the articles published in scientific journals. The basic principles of the NRV knowledge base are covered, and a brief description of its structure is given. The practical usage of the NRV knowledge base for both scientific and educational applications is demonstrated in detail.

ESSENTIAL ASPECTS OF IT TRAINING TECHNOLOGY FOR PROCESSING, STORAGE AND INTELLECTUAL ANALYSIS OF THE BIG DATA USING THE VIRTUAL COMPUTER LAB

**Evgenia Cheremisina, Mikhail Belov^a, Nadezhda Tokareva,
Yury Kryukov**

Dubna State Univeristy

E-mail: ^abelov@uni-dubna.ru

Many new professions in the field of the virtual computer lab entered the labor market, corresponding the needs of the Russian economy. The professionals should have enough knowledge and skills to meet the current realities and needs of the business. Nowadays we face the primary task of supplying the software tools and technologies to provide the effective work in the classroom and at home, cause a true amazement and sincere admiration for the progress in math and computer science, making the students more self-confident and giving the reasons for the future actions.

The training of the "consumers" should be cut off in the process of the IT specialists' education, and we should spare no effort to training of the "creative doers". For this purpose, it is important not only to study the ways of creating the information systems from the scratch, paying attention to the configuring and adjustment of the equipment, connection and integration of all the necessary parts of the system without any help, and only after that to accomplish issue-oriented tasks.

The Virtual Computing Lab provides a set of software and hardware-based virtualization, containerization and management tools that enable the flexible and on-demand provision and use of computing resources, knowledge management system, theoretical materials and practical cookbooks in the form of cloud services for carrying out research projects, scientific computational calculations and tasks related to the development of complex corporate and other distributed information systems. The service also provides dedicated virtual servers for innovative projects that are carried out by students and staff at the Institute of System Analysis and Control of State Dubna University. One main distinguishing trait of the Virtual Computer Lab is its self-organizing principles, which make it possible to transition students from a rigid system of group security policies to a new system where each student can develop a sense of personal responsibility, respect for colleagues, and tolerance, which should provide a solid foundation for strengthening and developing basic civilizational values in the education environment.

Modern blade servers are the hardware components that support virtual computer labs. They are high-performance, high-capacity, but compact and allow the space in the server room to be used more efficiently. The software platform of the Virtual Computing Lab is implemented based on the VMware vSphere Software, which consists of vSphere ESXi hypervisors with some hand-made enhancements and optimizations for some specific hardware that handle all the computing work of the virtual machines as well as vCenter Server central management servers. A centralized management portal as well as a knowledge management system were created to improve productivity of work in the Virtual Computer Laboratory. The need to create such a system was conditioned by the fact that students are able to improve productivity of remote learning by themselves, so it is important to create a social network between all participants as well as to create an environment that allows pupils the opportunity to independently engage in such processes as the identification, acquisition, presentation, and use (distribution) of knowledge without the direct involvement of the instructor. Methods of use (propagation) are directly related to storage methods and, consequently, the technological tools that may be used for the transmission of formal knowledge include knowledge bases with various search functionality; blogs, wikis, and social networks; "Wiki Textbooks" that allow all participants to collaboratively create and update educational content and exchange practical problems (including from real companies); as well as user blogs, forums, and group chat systems. The new practice with containers is different compared to VMware case and effectively complements it for a wide range of practical tasks. For the underlying operating system kernel can be used for all containers. Current evolution of Virtual Computer Lab leads to development of design templates for both corporate IT deployment and students learning project.

It should also be emphasized that the Virtual Computer Lab has helped us provide an optimal and sustainable technological, educational-organizational, scientific-methodological, and regulatory-administrative environment for supporting innovative approaches to computer education.

CLOUD COMPUTING, VIRTUALIZATION

ABOUT SOME BLOCK CHAIN PROBLEMS

**Alexander Bogdanov^{1,a}, Alexander Degtyarev¹,
Magdalyne Kamande², Oleg Iakushkin¹, Vladimir Korkhov¹**

¹St Petersburg State University

²Saint Petersburg Electrotechnical University "LETI"

E-mail: ^abogdanov@csa.ru

Cloud Computing stands out from any other distributed computing paradigm by offering services on-demand basis which are not limited to any geographical restrictions. Consequently, this has revolutionized the computing by providing services to wide scope of customers starting from casual users to highly business oriented Industries. Despite of its capabilities, Cloud Computing still faces challenges in handling a wide array of faults, which could causes loss of credibility to Cloud Computing. Among those faults Byzantine faults offers serious challenge to fault tolerance mechanism, because it often goes undetected at the initial stages and it could easily propagate to other VMs before detection is made. Consequently, some of the mission critical applications such as air traffic control, online baking etc still avoid using the use of cloud computing for such reasons. Moreover, if a Byzantine faults is not detected and tolerated at initial stage then applications such as big data analytics can go completely wrong in spite of hours of computations performed by the entire cloud. Therefore, in the previous work a fool-proof Byzantine fault detection has been proposed, as a continuation this work designs a scheduling algorithm (WSSS) and checkpoint optimization algorithm (TCC) to tolerate and eliminate the Byzantine faults before it makes any impact. The WSSS algorithm keeps track of server performance which is part of Virtual Clusters to help allocate best performing server to mission critical application. WSSS therefore ranks the servers based on a counter which monitors every Virtual Nodes (VN) for time and performance failures. The TCC algorithm works to generalize the possible Byzantine error prone region through monitoring delay variation to start new VNs with previous checkpointing. Moreover, it can stretch the state interval for performing and error free VNs in an effect to minimize the space, time and cost overheads caused by checkpointing. The analysis is performed with plotting state transition and CloudSim based simulation. The result shows TCC reduces fault tolerance overhead exponentially and the WSSS allots virtual resources effectively.

EXPERIMENTS WITH JUPYTERHUB AT THE SAINT PETERSBURG STATE UNIVERSITY

Andrey Erokhin^a, Andrey Zarochentsev

St Petersburg State University

E-mail: ^aandrey.erokhin@cern.ch

The talk focuses on our experience with JupyterHub and JupyterLab, the ways we extend Jupyter and how we abuse JupyterHub to spawn something other than single-user Jupyter notebook servers. Our project has started as a copy of CERN SWAN environment, but it evolves independently. However, we are still using CVMFS to load Jupyter kernels and other software and EOS to store user home directories.

CLOUD META-SCHEDULER FOR DYNAMIC VM REALLOCATION

Nikita Balashov

Joint Institute for Nuclear Research

E-mail: balashov.nikita@gmail.com

Clouds gave us a more flexible way of sharing computing resources between users and combine computation-intensive workloads with other types of workloads. Due to the variety of workloads in such environments and to their dynamic nature, the hosts are often underloaded. In this talk we give a review of an approach to improve hardware utilization in IaaS clouds through dynamic reallocation of VMs (enabled by live-migration technology) and overcommitment. The software framework presented would allow one to use it as a meta-scheduler with the built-in simple algorithms for optimizing cloud workloads distribution or to implement custom schemes of dynamic reallocation and consolidation of virtual machines.

APPROACHES TO THE AUTOMATED DEPLOYMENT OF THE CLOUD INFRASTRUCTURE OF GEOGRAPHICALLY DISTRIBUTED DATA CENTERS

**Petr Fedchenkov¹, Nikita Samokhin¹, Sergey Khoruzhnikov¹,
Oleg Lazo¹, Andrey Shevel^{1,2,a}**

¹*ITMO University*

²*NRC "Kurchatov Institute" PNPI*

E-mail: ^ashevel_aye@pnpi.nrcki.ru

University ITMO (ifmo.ru) is developing the cloud of geographically distributed data centers under OpenStack. The article describes the deployment cycle for the developed multi-agent system on physical and virtual machines using the developed software product repository. The cloud system of distributed data centers assumes the existence of a number of subsystems, or independent agents. And each agent is aimed at implementing the creation, management and provision of certain services with a defined SLA: virtual machines, long-term data storage, communication channels with network functions. At the same time, it is necessary to perform both the initial deployment of the system and to provide the means that perform the automated modification of the hardware or software infrastructure, that is, the complete agent life cycle. The approaches suggested in the article allow us to use the minimum number of pre-installed components for the purpose of deploying and managing an array of physical and virtual entities of data centers, including deployment using Xcat, configuring using Salt and monitoring the infrastructure using Zabbix and Grafana.

KUBERNETES TESTBED CLUSTER FOR THE LIGHTWEIGHT SITES PROJECT

Iuliia Gavrilenko

Plekhanov Russian University of Economics

E-mail: yulya952@rambler.ru

The Worldwide LHC Computing Grid (WLCG) is a global collaboration of more than 170 computing centres in 42 countries and the number is expected to grow in the coming years. However, provisioning resources (compute, network, storage) at new sites to support WLCG workloads is still no straightforward task and often requires significant assistance from WLCG experts. Recently, the WLCG community has initiated steps towards reducing such overheads through the use of prefab Docker containers or OpenStack VM images, along with the adoption of popular tools like Puppet for configuration. In 2017, the Lightweight Sites project was initiated to construct shared community repositories providing such building blocks. These repositories are governed by a single Lightweight Site Specification Document which describes a modular way to define site components such as Batch Systems, Compute Elements, Worker Nodes, Networks etc. Implementation of the specification is based on a popular orchestration technology – Kubernetes. Here it is discussed the testbed cluster for deploying Lightweight grid sites. The research is mainly focused on the controlling lifecycle of containers for compute element, batch system and worker. Also some parameters for benchmarking and evaluation of the performance of different implementations were introduced.

DESIGN AND IMPLEMENTATION OF A SERVICE FOR PERFORMING HPC COMPUTATIONS IN CLOUD ENVIRONMENT

Ruslan Kuchumov^a, Vladimir Korkhov

St Petersburg State University

E-mail: ^akuchumovri@gmail.com

Cloud computing became a routine tool for scientists in many domains. In order to speed up an achievement of scientific results a cloud service for execution of distributed applications was developed. It obviates users from manually creating and configuring virtual cluster environment or using batch scheduler and allows them only to specify input parameters to perform their computations. One of the key parameters that this service aims to help users with is virtual cluster configuration. For most applications it is difficult to tell the optimal number of cluster nodes, amounts of their threads per node and memory so that application would have a minimal execution time. In this work an approach to optimization of cluster configuration has been proposed and software system for launching HPC application in a cloud has been presented.

NEW FEATURES OF THE JINR CLOUD

**Nikolay Kutovskiy^a, Nikita Balashov, Alexandr Baranov,
Yelena Mazhitova, Roman Semenov**

Joint Institute for Nuclear Research

E-mail: ^akut@jinr.ru

The report covers details on such aspects of the JINR cloud development as migration to RAFT-based algorithm high availability setup, Ceph-based storage back-end for VM images and DIRAC-based grid platform for external partner clouds integration into distributed computational cloud environment.

IMPROVING THE LOAD OF SUPERCOMPUTERS BASED ON JOB MIGRATION USING CONTAINER VIRTUALIZATION

Stanislav Polyakov^a, Minh-Duc Nguyen, Elena Petrova

Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University

E-mail: ^as.p.polyakov@gmail.com

Modern supercomputers employ schedulers to distribute the queued jobs between the computational nodes and maximize their load. The jobs can vary in size (number of required computational units, planned time) and it is impossible to schedule them with 100% efficiency. Additionally, some jobs are completed ahead of time which can also decrease the load. In practice the average load of supercomputers seldom exceeds 90% and often is below 80%. Theoretically, any slots left empty by the scheduler can be used by jobs that require a single scheduling CPU slot and sufficiently short execution time. In practice, there are numerous computational problems that do not require parallelization and therefore need a single CPU slot but have significant time requirements. A queue of jobs solving these problems can be used to fill short schedule slots if the jobs are migrated to different nodes or saved and returned to the queue before the end of the allotted time. It is possible with the use of container virtualization that allows to start, stop, and migrate containers relatively quickly. We develop a simulation model of a system intended to increase the load of supercomputers using an additional queue of non-parallel jobs packed in containers. Here we present an estimation of the load of supercomputers using different scheduler algorithms. The estimation is based on simulation with a simulated queue consisting of jobs with randomly generated parameters (number of CPUs, planned time, actual completion time). The chosen distribution of the parameters is based on typical data of supercomputers. We also present an estimation of the expected load increase resulting from the use of container migration.

CREATION OF CLOUD INFRASTRUCTURE OF INP'S ASTANA BRANCH - PRIVATE ESTABLISHMENT "NULITS" AND ITS INTEGRATION WITH THE DISTRIBUTED JINR CLOUD INFRASTRUCTURE

**Boris Potapchuk¹, Mikhail Mazhitov^{1,a}, Sergey Belov¹,
Yelena Mazhitova^{2,3}**

¹*Private establishment "Nazarbayev University Library and IT Services"*

²*Joint Institute for Nuclear Research*

³*Institute of Nuclear Physics, Republic of Kazakhstan*

E-mail: ^ammazhitov@nu.edu.kz

The article is devoted to the project of creating the cloud infrastructure of the Astana branch of the Institute of Nuclear Physics and the private establishment "Nazarbayev University Library and IT Services" (Republic of Kazakhstan, Astana) on the basis of the resources of both organizations, its integration with the distributed cloud infrastructure of the Joint Institute for Nuclear Research (Russian Federation, Dubna). Investigates the motivation and implementation of the cloud infrastructure, discusses various mechanisms for cloud integration, and outlines plans for using created infrastructure.

CLOUDS OF JINR, UNIVERSITY OF SOFIA AND INRNE - CURRENT STATE OF THE PROJECT

**Hristov Svetoslav¹, Kouzma Kouzmov², Nikita Balashov³,
Nikolay Kutovskiy³, Radoslava Hristova², Vladimir Dimitrov^{2,a},
Vladimir Korenkov³**

¹*INRNE*

²*University of Sofia*

³*Joint Institute for Nuclear Research*

E-mail: ^acht@fmi.uni-sofia.bg

JINR established a cloud based on OpenNebula. It is open for integration with the clouds from the member states. The paper presents current (first year) state of the 3 years project that aims to create a cloud backbone in Bulgaria. University of Sofia and INRNE participate in that initiative. This is a target project funded by JINR based on the research plan of the institute.

THE SERVICE FOR PARALLEL APPLICATIONS BASED ON THE JINR CLOUD AND HYBRILIT RESOURCES

**Elena Zemlyanaya¹, Pavel Goncharov², Andrey Nechaevskiy¹,
Ruslan Kuchumov³, Oksana Streltsova¹, Nikolay Kutovskiy¹,
Maksim Bashashin¹, Nikita Balashov¹, Gennady Ososkov¹,
Ivan Sokolov¹**

¹*Joint Institute for Nuclear Research*

²*Sukhoi State Technical University of Gomel*

³*St Petersburg State University*

E-mail: isokolov@jinr.ru

Cloud computing became a routine tool for scientists in many domains. The JINR cloud infrastructure provides JINR users computational resources for performing various scientific calculations. In order to speed up achievements of scientific results the JINR cloud service for parallel applications was developed. It is a application-specific web-interface. It consists of several components and implements a flexible and modular architecture which allows to utilize both more applications and various types of resources, as computational backends. Besides this architecture increases the utilization of cloud idle resources. This service allows scientist to focus on his research domain by interacting with the service in a convenient way via browser and abstracting away from underlying infrastructure as well as its maintenance. A user just set a required values for his job via web-interface and specify a location for uploading a result. The computational workload are done on the VMs deployed in the JINR cloud infrastructure. But It is planned in the nearest future to add a HybriLIT heterogeneous cluster as one more computational back-end of the JINR SaaS service. An example of using the Cloud&HybriLIT resources in the scientific computing is the study of superconducting processes in the stacked long Josephson junctions (LJJ). LJJ systems are undergone the intensive research because of a perspective of practical applications in nano-electronics and quantum computing. Respective mathematical model is described by a system of the sine-Gordon type partial differential equations where the spatial derivatives are approximated with help of standard finite difference formulas and the resulting system of ODEs is numerically solved by means of the 4th order Runge-Kutta procedure. Parallel MPI-implementation of the numerical algorithm was developed.

DESKTOP GRID TECHNOLOGIES AND VOLUNTEER COMPUTING

INTEGRATION OF THE BOINC SYSTEM AND ADDITIONAL SOFTWARE PACKAGES

Nikolay Khrapov

Institute for Information Transmission Problems RAS

E-mail: nkhrapov@gmail.com

Currently, the BOINC system [1] is the most well-known voluntary computing system. Many researchers use BOINC to solve scientific problems. The BOINC software allows them to automate the process of sending tasks to the computing node, starting it and returning the results. To solve many scientific problems, the BOINC system requires additional software. The need for integration of the BOINC system and additional software arises in the following cases: a) generation of computing tasks on the server side; b) processing of results on the server side; c) running additional software components on the side of the compute node; d) interface for visualization and display of results. The report considers various aspects of the integration of the BOINC system and additional software. Describes the approaches used for integration. Projects USPEX@HOME [2] and XANSONS4COD@HOME [3] are considered too. The practice of applying standard approaches is discussed and new ideas are proposed.

References

[1] D. P. Anderson "BOINC: A system for public-resource computing and storage. In Grid Computing", Proceedings. Fifth IEEE/ACM International Workshop, 4--10 (IEEE, November 2004).

[2] Nikolay P. Khrapov, Valery V. Rozen, Artem I. Samtsevich, Mikhail A. Posypkin, Vladimir A. Sukhomlin, Artem R. Oganov. Using virtualization to protect the proprietary material science applications in volunteer computing. Open Eng. 2018, v.8, pp. 57-60.

[3] Vladislav S. Neverov, Nikolay P. Khrapov. "XANSONS for COD": a new small BOINC project in crystallography. Open Eng. 2018, v.8, pp. 102-108.

THE ACTIVITY OF RUSSIAN CHAPTER OF INTERNATIONAL DESKTOP GRID FEDERATION

Ilya Kurochkin

Institute for Information Transmission Problems RAS

E-mail: qurochkin@gmail.com

Results of activity of the Russian Chapter of International Desktop Grid Federation (IDGF) are considered. Including interaction with community of the russian volunteers (crunchers), start new and support of the existing projects of the volunteer distributed computing.

In this report the results of activity of the Russian Chapter of International Desktop Grid Federation (IDGF) for 2 years are summed up. Continuous interaction with community of the russian crunchers (volunteers) was carried out.

Cooperation was conducted in several directions:

- Search of scientific groups among the russian scientists interested in creation of new projects;
- Carrying out sociological researches;
- Development of the existing projects of the voluntary distributed calculations;
- Joint publications and reports at scientific conferences;
- Support of the crowdfunding project on creations of the specialized computing server;
- Support of creation of a cluster with the improved relation the cost-performance.

By results of interaction of several scientific groups and community of the russian crunchers it is possible to tell that the group of crunchers which, at the high level helps scientific groups in a number of researches was created. It is possible to call such phenomenon the term "citizen science". Interaction of members of the Russian Chapter of IDGF happens on a constant basis that is expressed in a large number of joint publications.

MODELING OF TASK SCHEDULING IN DESKTOP GRID SYSTEMS AT THE INITIAL STAGE OF DEVELOPMENT

Ilya Kurochkin^{1,a}, Gerk Evgeny²

¹*Institute for Information Transmission Problems RAS*

²*NUST MISIS*

E-mail: ^aqurochkin@gmail.com

The paper presents an overview of modern methods for task scheduling in desktop grid systems, estimates of the quality of methods, including: the time of execution of all tasks, the level of resource utilization. Heuristic approach to task scheduling is considered, which allows ensuring high performance and reliability of such systems at the early stages of development. A comparative analysis of the results of computational experiments performed with the help of the GridSim high performance computing simulation tool for various desktop grid system configurations is carried out.

SUPPORTING EFFICIENT EXECUTION OF MANY-TASK APPLICATIONS WITH EVEREST

Oleg Sukhoroslov^a, Sergey Smirnov, Sergey Volkov

Institute for Information Transmission Problems RAS

E-mail: ^aoleg.sukhoroslov@gmail.com

Distributed computing systems are widely used for execution of loosely coupled many-task applications. There are two important classes of such applications. Bag-of-tasks applications, e.g., parameter sweeps or Monte Carlo simulations, represent a set of independent tasks. Workflows, which are used for automation of complex computational and data processing pipelines, consist of multiple tasks with control or data dependencies. The report discusses the common problems related to the efficient execution of such applications on distributed computing resources and the relevant solutions implemented within the Everest platform.

Everest [1-3] is a web-based distributed computing platform which provides users with tools to publish and share computing applications as web services. The platform also manages the execution of applications on remote computing resources. Everest implements the PaaS model by providing its functionality via remote web and REST interfaces. A single instance of the platform can be accessed by many users in order to create, run and share applications with each other. Instead of using a dedicated computing infrastructure, Everest performs the execution of applications on external resources attached by users. The platform supports integration with standalone servers, clusters, grid infrastructures, desktop grids and clouds. A user can specify multiple resources, possibly of different type, for running an application.

Everest provides multiple tools for execution of many-task applications. First, it includes a general-purpose service for execution of bag-of-tasks applications such as parameter sweeps. The application tasks are described using a simple declarative notation. Second, it is possible to dynamically add new tasks or invoke other applications from a running application via the REST API. This allows users to run complex many-task applications such as workflows. In this case, the dependencies between tasks are managed internally by a user application. While this approach provides maximum flexibility, it does not allow passing the complete task graph to the platform to enable scheduling optimizations. To overcome this limitation, a new interface for submitting workflows has been added recently.

The application tasks are executed by Everest on computing resources specified by a user. The efficiency of application execution, i.e. the execution time, critically depends on the methods used for task scheduling [4]. Everest implements a two-level scheduling mechanism that allows to plug-in different scheduling algorithms. First, the available resources are fairly distributed among the running applications. Then, the application-level scheduler selects tasks for running on provided resources. The separate schedulers are implemented for bags-of-tasks and workflows, which are based on MaxMin and DLS algorithms respectively. The used algorithms require the estimates of task execution and data transfer times. Currently, these estimates are computed based on the statistics from previous task and application executions.

The other features that are essential for efficient execution of many-task applications include accounting for local resource policies and automatic recovery of failed tasks. For example, the limit on the maximum number of jobs per user imposed by an HPC cluster administrators

may not allow to fully utilize the resource when running a single job per Everest task. An advanced adapter for Slurm manager has been developed which allows to solve this problem by submitting complex jobs consisting of multiple tasks. When dealing with failed tasks, Everest distinguishes between critical and recoverable faults. In the latter case, the task is retried multiple times, and the resources with many failures are blacklisted. To account for temporary network failures between Everest and resources, the tasks running on the disconnected resource are not rescheduled immediately to avoid wasting compute time.

References

- [1] Everest. <http://everest.distcomp.org/>
- [2] Sukhoroslov O., Volkov S., Afanasiev A. A Web-Based Platform for Publication and Distributed Execution of Computing Applications // 14th International Symposium on Parallel and Distributed Computing (ISPDC). IEEE, 2015, pp. 175-184.
- [3] Sergey Smirnov, Oleg Sukhoroslov, and Sergey Volkov. Integration and Combined Use of Distributed Computing Resources with Everest // *Procedia Computer Science*, Volume 101, 2016, pp. 359-368.
- [4] Nazarenko A., Sukhoroslov O. An Experimental Study of Workflow Scheduling Algorithms for Heterogeneous Systems. In: Malyshkin V. (eds) *Parallel Computing Technologies. PaCT 2017. Lecture Notes in Computer Science*, vol 10421. Springer, Cham, 2017, pp. 327-341.

THE GAME CHARACTER OF COLLABORATION IN VOLUNTEER COMPUTING COMMUNITY

Victor Tishchenko

RCF "Computer Science and Control" RAS

E-mail: vtichenko@mail.ru

The implementation of VC project directly depends on the number of participants, PC (other machines) and time of their work. So project managers/organizers in search of a mechanism to encourage participation in VC-projects use conditional points accrual mechanism; the number of these points ("credits") depends on the provided capacities, the time of participation in projects, and other characteristics of the activity of volunteers and their teams. The availability of constant statistics for all projects, in addition to tracking various ratings, provokes various virtual competitions ("challenges") between participants and teams. And many of the projects create an environment for the competitors by volume computations are done, both individually and in the team event. Thus modus operandi of VC – is spirit of competition. If volunteers are members of a team, they are simultaneously competing with the other teams on the project on the more immediate goal of racking up the most contributions and coming out on top of the table of statistics documenting contributions. This form of cooperation and competition demonstrates a new type of online scientific collaborative network. Emerging from individual participation at the first stages of VC, nowadays VCvolunteers form a virtual scientific community which working towards the common goal of scientific research. VC projects are by their nature decentralized and so, volunteer's

communication through online form a communal virtual place. As for social underpinning of VC-projects, we should note a nontraditional form of VC social structure. And the VC organizational framework consists of projects; individuals; VC teams; Internet resources – project websites with statistics in rank tables, discussion forums, etc. The revealed differences in the distribution of volunteers underscore the need for a detailed study of the motives for participating in such projects. However, the researchers and organizers of VC projects do not have an unambiguous answer to the question: how can a significant scientific project requiring large-scale calculations form an environment that will stimulate the contribution of resources by many volunteers? Most sociological studies of the motives of volunteers' behavior were based on the representation of the community as a set of participants and as a result, the research focuses only on the motivation of individual participants [Holohan A. and Garg A., 2006; Nov O., et al., 2014; Andreev. A., 2014; Kurochkin et al., 2015]. The background of such view is based on an idea that VC participants provide something for others (processing data for scientific projects) at their own cost (time, energy, opportunity costs, use of PC resources). E.g., factors which need for self-oriented motivations. But we think that a motive of self-actualization is not enough to explain why millions of common people regular (day by day) participate in scientific computation. Our research indicates that the answer lies at the intersection of self-oriented motivation and the interactional and organizational possibilities emerging through the Internet. We suggest: online collaboration can capture people's motivation better than only intrinsic motives. In order to highlight traces of collaborative networks we will allocate links between the active volunteers in a whole Boinc.ru community. For determining an active group of participants in VCprojects in Boinc.ru community we have used the accrued credits, assumed that a person took part in the project if he collected at least 50,000 points in it. Having constructed a connectivity table, where a_{ij} – the number of general projects for two volunteers we construct a graph. So it appeared – 200 volunteers, who participated in 2176 projects. It means that almost all active users are connected with some projects. The most active and productive participants in VC exemplify “co-competition” (Brandenburger & Nalebuff, 1996) – collaboration within the network organization of a VC project. Our research indicates that the motivation of participation of millions of unskilled volunteers in VC-projects lies at the intersection of intrinsic motivation and the organizational possibilities emerging through the collaboration. In providing the means for channeling participants' motivations to compete and cooperate, VC-projects provide powerful insights into a new type of collaborative network.

ORTHOGONALITY-BASED CLASSIFICATION OF DIAGONAL LATIN SQUARES OF ORDER 10

**Eduard Vatutin^{1a}, Vitaly Titov¹, Oleg Zaikin², Stepan Kochemazov²,
Maxim Manzyuk³, Natalia Nikitina⁴**

¹*Southwest State University*

²*Matrosov Institute for System Dynamics and Control Theory SB RAS*

³*BOINC.ru*

⁴*Institute of Applied Mathematical Research KRC RAS*

E-mail: ^aevatutin@rambler.ru

The search for pairs of orthogonal diagonal Latin squares (ODLS) is a hard-combinatorial problem [1]. According to the Euler-Parker approach, a set of diagonal transversals is constructed for a given DLS of order N . If a subset of N non-overlapping transversals is found, then an orthogonal mate for the DLS can be easily constructed. According to some estimations, only 1 DLS of order 10 out of 32 millions has an orthogonal mate. Authors of the volunteer computing project Gerasim@home and SAT@home maintain the collection of pairs of ODLS of order 10. It contains more than 580 000 canonical forms (isotopy classes) of DLS of order 10 as for June 2018.

DLSs from the collection can be classified by the number of their orthogonal mates. According to this classification, about 550 000 of DLSs are bachelor – i.e. each of them has exactly one orthogonal mate. About 7 500 of DLSs are line-2 – i.e. that each of them has exactly two orthogonal mates. There are also 63 line-3, 283 fours, 2 fives, 9 sixes, 1 sevens, 7 eights and 1 ten (see Fig. 1). This classification can be expanded.

Figure 1 contains examples of DLSs of order 10 that are part of found combinatorial structures. These DLSs were constructed during several computational experiments: random search for DLSs with consequent attempt to construct their orthogonal mates; comprehensive search for DLSs that are symmetric according to some plane; comprehensive search for general symmetric DLSs; random search for partially symmetric DLSs.

The found combinatorial structures (graphs from DLSs on the orthogonality binary relation set) are new and were not published before. Due to their simplicity they allow a trivial classification based on a vector of degrees of vertices which is sorted in ascending order. In fact, in this case a degree of a vertex is the number of ODLS for the chosen DLS.

The research was partially supported by Russian Foundation for Basic Research (grants 16-07-00155-a, 17-07-00317-a, 18-07-00628-a, 18-37-00094-mol-a) and by Council for Grants of the President of the Russian Federation (stipend SP-1829.2016.5). Authors thank citerra [Russia Team] from the internet portal BOINC.ru for his help in the development and implementation of some algorithms. Also authors thank all the volunteers of SAT@home and Gerasim@home for their participation.

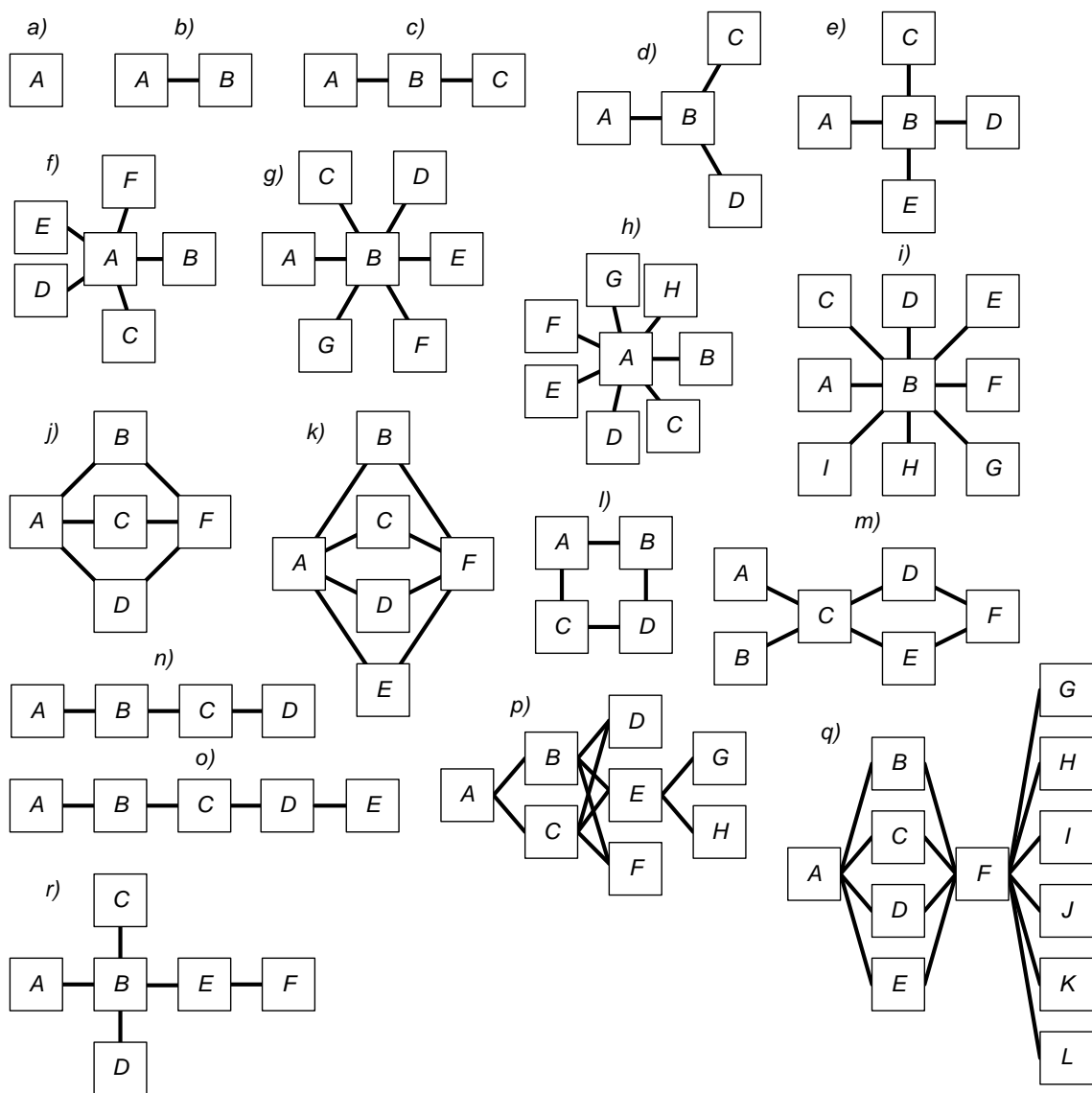


Figure 1. Combinatorial structures based on the orthogonality of DLSs of order 10: a – square with no mate (bachelor); b – line-2; c – line-3; d – triple; e – four; f – five; g – six; h – seven; i – eight; j – rhomb-3; k – rhomb-4; l – loop-4; m – fish; n – line-4; o – line-5; p – flyer; q – ten; r – cross.

References

[1] Colbourn C.J., Dinitz J.H. Handbook of Combinatorial Designs. Second Edition. Chapman&Hall, 2006. 984 p.

BOINC-BASED COMPARISON OF THE GEOACOUSTIC INVERSION ALGORITHMS EFFICIENCY

Oleg Zaikin^{1a}, Pavel Petrov², Ilya Kurochkin³

¹*Matrosov Institute for System Dynamics and Control Theory SB RAS*

²*V. I. Il'ichev Pacific Oceanological Institute FEB RAS*

³*Institute for Information Transmission Problems RAS*

E-mail: ^azaikin.icc@gmail.com

The BOINC-based volunteer computing project Acoustics@home was employed to study the accuracy of the sound speed profile reconstruction in a shallow-water waveguide using a dispersion-based geoacoustic inversion scheme. This problem was transformed into a problem of black-box minimization of a certain mismatch function. According to the first approach, a sound speed profile is considered a piecewise-linear function with fixed uniformly-spaced nodes. At these nodes, the values of sound speed are obtained in the course of inversion. In the second approach the depths of the sound speed profile nodes are also considered inversion parameters, however, their number must be smaller than in the first approach due to the computational complexity limitation. Several large-scale computational experiments reveal that for the considered problem the second approach leads to a more accurate sound speed profile estimation.

This study was supported by the Council for Grants of the President of the Russian Federation (grant No. MK-2262.2017.5), the Russian Foundation for Basic research (grants No. 16-05-01074-a, No. 16-07-00155-a), and the POI FEB RAS Program 'Nonlinear dynamical processes in the ocean and atmosphere'.

РЕАЛИЗАЦИЯ ВЫЧИСЛЕНИЙ С ДИНАМИЧЕСКИМИ ЗАВИСИМОСТЯМИ ЗАДАЧ В СРЕДЕ ДЕСКТОП ГРИД С ИСПОЛЬЗОВАНИЕМ EVEREST И TEMPLET WEB

**Востокин Сергей^{1a}, Сухорослов Олег², Бобылева Ирина¹,
Стефан Попов¹**

¹*Самарский национальный исследовательский университет*

²*Институт проблем передачи информации РАН*

E-mail: ^aeasts@mail.ru

Целью исследования была экспериментальная проверка технологии автоматизированной разработки приложений с динамически формируемым графом зависимостей между задачами для вычислений в грид-среде настольных компьютеров организации. Данный тип вычислений привлекателен с точки зрения минимума

аппаратных затрат, но остается сложным как для программирования, так и для развертывания. Специфические требования, которые учитывались при разработке тестового приложения, включают: (1) использование простаивающих компьютеров; (2) исполнение на гетерогенном оборудовании; (3) простоту и оперативность развертывания компонентов приложения; (4) организацию длительных вычислений, устойчивых к отказу; (5) организацию вычислений с большим количеством задач и сложными зависимостями между ними. Нами было разработано приложение блочной сортировки большого массива экспериментальных данных в грид-среде настольных компьютеров, в котором были учтены заявленные требования. Протестировано использование ноутбуков, рабочих станций, виртуальных машин, имитирующих варианты доступных простаивающих вычислительных ресурсов организации. Компоненты приложения исполнялись под управлением операционной системы Linux (оркестратор вычислений), а также под управлением ОС Windows (сортировщики и мерджеры блоков данных). Простота развертывания обеспечивалась за счет использования платформ Everest [1], Templet Web [2] и программ-агентов с несложной процедурой установки. Отказоустойчивость вычислений задач обеспечивалась внутренними механизмами платформы Everest. Оркестратор в целях отказоустойчивости развертывался из сервиса Templet Web на виртуальную машину под управлением VMware. Код оркестратора сортировки написан на языке C++ с использованием специально разработанного акторо-подобного фреймворка, что позволило формировать задачи сортировки и объединения блоков динамически в зависимости от результатов исполнения ранее запущенных задач. В дальнейшем планируется расширить код фреймворка для прозрачного взаимодействия между платформой Everest и Templet Web, что позволит реализовывать код оркестратора непосредственно исследователем, без участия системного программиста.

References

- [1] Sukhoroslov O. A Web-Based Platform for Publication and Distributed Execution of Computing Applications [Text] / Sukhoroslov O., Volkov S., Afanasiev A. // IEEE Xplore. – 2015. – Vol. 14. – P. 175-184.
- [2] Vostokin S.V. Templet Web: the use of volunteer computing approach in PaaS-style cloud [Text] / S.V. Vostokin, Y.S. Artamonov, D.A. Tsaryov // Open Engineering. – 2018. – Vol. 8(1). – P. 50-56.

**HIGH PERFORMANCE
COMPUTING, CPU
ARCHITECTURES, GPU, FPGA**

A SOFTWARE PACKAGE FOR STUDYING THE SYSTEM OF LONG JOSEPHSON JUNCTIONS ON HYBRID COMPUTING ARCHITECTURES

**Maksim Bashashin^{1,2}, Ksenia Erofeeva², Oksana Streltsova¹,
Elena Zemlyanaya^{1,2}, Maksim Zuev^{1,a}**

*¹Joint Institute for Nuclear Research
²Dubna State University*

E-mail: ^azuevmax@jinr.ru

The report presents the work on developing of a software package for investigating the system of long Josephson junctions. It makes possible to perform computations on heterogeneous computation architectures: Intel processors (CPU), Intel Xeon Phi processors (KNL), NVIDIA graphics processors (GPU). A comparative analysis of the acceleration and efficiency of the developed parallel implementations depending on the task parameters and the parallelization scheme was performed; analysis of the effectiveness of the developed parallel implementations using OpenMP, MPI and CUDA technologies for a single Josephson junction has been carried out in order to select the optimal computing architecture for the solution of the task. The calculations were carried out on a heterogeneous platform HybriLIT (LIT JINR).

The work is supported by RFBR grant No 15-29-01217.

NETWORK INFRASTRUCTURE OF HETEROGENEOUS PLATFORM “HYBRILIT”

**Dmitry Belyakov^{1,a}, Yurii Butenko¹, Andrey Dolbilov¹,
Mikhail Matveev¹, Vladislav Moiseev^{2,3}, Martin Vala³**

*¹Joint Institute for Nuclear Research
²Dubna State University*

³Institute for Theoretical and Experimental Physics, Košice, Slovak Republic

E-mail: ^admitry@jinr.ru

The heterogeneous platform “HybriLIT” provides users with ample opportunities for conducting parallel calculations and modeling of new experiments in the field of high energy physics. The platform consists of two clusters: a training and test polygon “HybriLIT” and supercomputer named after Govorun N.N. High performance servers with Intel Xeon series E5, Intel Xeon series Gold (Skylake) and Intel Xeon Phi 7290 (KNL), servers with graphics accelerators Nvidia Tesla K40, K80 and Nvidia Tesla V100 (Volta) are available for users.

The network infrastructure of the heterogeneous platform is build on basis of Ethernet 10 Gbit/sec technology, fabric Mellanox InfiniBand 100 Gbit/sec and fabric Intel Omni-Path 100 Gbit/sec are used as high-speed and low-latency backbone between the servers.

In this paper the network infrastructure of the heterogeneous platform “HybriLIT” is described, the results of testing the high-speed and low-latency backbone are given, the comparative analysis of fabric Mellanox InfiniBand 100 Gbit/sec and fabric Intel Omni-Path 100 Gbit/sec for different class of network tasks is performed.

ON PORTING OF APPLICATIONS TO NEW HETEROGENEOUS SYSTEMS

**Alexander Bogdanov^a, Vladimir Mareev, Nikita Storublevtzev,
Denis Manyashin**

St Petersburg State University

E-mail: ^abogdanov@csa.ru

This work is devoted to the development of guidelines for the porting of existing applications to GPGPU. The paper provides an overview of the current state of the parallel computation area with respect to GPGPU. Various approaches to the organization of parallel computations are considered, and their effectiveness is evaluated in relation to the application under study. Special attention is given to delicate relation between vectorization (done on the level of most internal loops of code) and parallelization (done on the external computational tasks). The proper combination of this makes it possible to get optimal speed-up. But in reality it can be too ideal point of view because of two principle limitations – memory of the GPGPU and the link between CPU and GPGPU. We argue that due to those limitations it is impossible to work out general strategy of porting applications to any GPGPU. Anyway for particular codes and special GPU’s the proposed approach makes it possible to get speed-up’s up to hundreds. This becomes even more effective when combined with virtualization of GPGPU to provide the balance between the size of computing core and rate of data transfer to it. We illustrate our approach on the examples of OpenFoam and DSMC porting to P100 GPU. It is clear, that only combination of all proposed measures makes it possible to get necessary speed-up. As a result, a strategy has been developed for migrating the application to a heterogeneous system. The results of the work can be applied when transferring similar applications to GPGPU or modified to transfer other types of applications.

ALGORITHMS FOR THE CALCULATION OF NONLINEAR PROCESSES ON HYBRID ARCHITECTURE CLUSTERS

Alexander Bogdanov^a, Vladimir Mareev, Nikita Storublevtzev

St Petersburg State University

E-mail: ^abogdanov@csa.ru

The problem of porting programs from one hardware platform to another has not ceased to be less relevant and simpler with time. The need to transfer programs to platforms with different architectures can have different roots. One of them is to increase the efficiency of executing programs for mass calculations on multiprocessor systems and clusters. The purpose of our work is to identify the key features of algorithms in porting codes for calculating of essentially nonlinear processes to a modern cluster of hybrid architecture that includes both CPUs (Intel Xeon) and GPU (NVIDIA TESLA) processors. In order to increase cluster productivity by the well-known Amdahl law, it is necessary to achieve heterogeneity of computational nodes. As a test problem for studying the process of porting a code to a cluster of hybrid architecture, the KPI equation of Kadomtsev-Petviashvili was chosen, written in integro-differential form [1]. As a result of the work, the procedure for porting a simulation code for a two-dimensional nonstationary model problem to a hybrid system is proposed. The features of such a transition are revealed.

References

[1] A.V. Bogdanov, V.V. Mareev. Numerical Simulation KPI Equation. Proceedings of the 15th International Ship Stability Workshop, 13-15 June 2016, Stockholm, Sweden. pp. 115-117.

DISTRIBUTED VIRTUAL CLUSTER MANAGEMENT SYSTEM

**Amissi Cubahiro¹, Vladimir Gaiduchok^{1a}, Magdalyne Kamande¹,
Vladimir Korkhov², Alexander Degtyarev²**

¹*Saint Petersburg Electrotechnical University "LETI"*

²*St Petersburg State University*

E-mail: ^agvladimiru@gmail.com

An effective cluster management system is the key to solving many problems that arise in the field of distributed computing. The wide spread of computing clusters led to the active development of task management systems, however, many issues, in particular, migration issues, have not yet been resolved. In this paper, we consider the organization of a virtual cluster created with virtualization at the OS level, as well as issues related to the dynamic migration of individual processes and containers. The complexity of this task within the cluster is determined by stringent requirements, a large set of process and container state

parameters, and the possibility of using specialized equipment. The ability to restore the state of a container to another node is a complex task that requires the development and implementation of multiple subsystems. Migration of containers and processes is much more difficult than migration of virtual machines because of close integration into the OS and the ability to work with individual components of equipment directly: you need to restore the state of individual subsystems, while in the case of a traditional virtual machine, the VM works with virtual equipment, provided by the hypervisor, and the state of the guest OS is inside the VM itself. Migration of processes and containers is an actively developing direction at present. We will present and discuss a technique for managing distributed heterogeneous virtual clusters using virtualization at the OS level; a technique of ensuring reliability, fault tolerance and load balancing of computing clusters due to the dynamic migration of tasks within a virtual cluster; an architecture of a virtual computer network for a computing cluster, minimizing the overhead associated with data exchange, for a specific task.

THE PROBLEM OF SYMBOLIC-NUMERIC COMPUTATION OF THE EIGENVALUES AND EIGENFUNCTIONS OF THE LEAKY MODES IN A REGULAR HOMOGENEOUS OPEN WAVEGUIDE

Andrey Drevitskiy^a, Dmitriy Divakov

RUDN University

E-mail: ^aadrevitskiy@gmail.com

In this paper the algorithm of finding eigenvalues and eigenfunctions for the leaky modes in a three-layer planar dielectric waveguide is considered. The problem on the eigenmodes of open three-layer waveguides is formulated as the Sturm-Liouville problem with the corresponding boundary and asymptotic conditions. In the case of guided and radiation modes of open waveguides, the Sturm-Liouville problem is formulated for self-adjoint second-order operators on the axis and the corresponding eigenvalues are real quantities for dielectric media. The search for eigenvalues and eigenfunctions corresponding to the leaky modes involves several difficulties: the boundary conditions for the leaky modes are not self-adjoint, so that the eigenvalues can turn out to be complex quantities. The problem of finding eigenvalues and eigenfunctions will be associated with finding the complex roots of the nonlinear dispersion equation. In the present paper, an original scheme based on the method of finding the minimum of a function of several variables is used to find the eigenvalues. The paper describes the algorithm for searching for eigenvalues, the algorithm uses both symbolic transformations and numerical calculations. Based on the developed algorithm, the dispersion relation for the slowly leaky mode of a three-layer open waveguide was calculated in the Maple computer algebra system using CUDA(R) technology to accelerate certain routines.

REAL-TIME VISUALIZATION OF SHIP AND WAVY SURFACE MOTIONS BASED ON GPGPU COMPUTATIONS

Anton Gavrikov^a, Andrei Ivashchenko, Ivan Gankevich, Nataliia Kulabukhova, Alexander Bogdanov, Alexander Degtyarev, Vladimir Rukovchuk

St Petersburg State University

E-mail: ^agavrikovantonkapi@gmail.com

One of the key stages in ship design process is the modeling of its behavior on the wavy sea surface, carried out with the expected operational characteristics taken into account. Similar modeling process could be done within real conditions at virtual testbed, which allows to monitor the influence of external disturbances on ship's running characteristics in real time with sensors installed onboard. Visualization of the results for such modeling process allows the researcher to correctly and holistically perceive occurring events, as well as to predict and timely respond to emerging dangerous situations. If we are using GPGPU technology for computation purposes, results of modeling will be already placed in GPU memory after process completion. This fact can be regarded as an opportunity to optimize the visualization process by converting the raw simulation data into graphic objects directly on the GPU, and interaction mechanisms between OpenGL and OpenCL could be used here. In this article we demonstrate the effectiveness of this technique on the example of ship behaviour visualization on a wavy sea surface, as well as forces acting on the ship's hull.

THE USAGE OF HPC SYSTEMS FOR SIMULATION OF DYNAMIC EARTHQUAKE PROCESS

Yulia Golubeva, Vasily Golubev^a

MIPT

E-mail: ^aw.golubev@mail.ru

Nowadays the HPC systems are very widespread in the world. Due to their computational power it is possible to simulate with a high precision a lot of phenomena: drugs development, seismic survey process, hydraulic fracturing and multi-component fluid flow, human-human interaction, high-speed collisions in open space, tsunami and earthquake initiation. That is why the development of modern applied research software for multi-processors systems are important. In the current work seismic waves generated during the earthquake process are considered. To describe precisely the dynamic behavior of the heterogeneous geological medium the 2D/3D full-wave system of elastic equations was used. Unfortunately, the analytical solution is available only for simple source and geometry of the area of interest. The grid-characteristic numerical method on curvilinear structured meshes was successfully applied. To achieve enough computational speed on large grids the research software designed by Khokhlov N.I. at MIPT was used. It is parallelized with OpenMP and MPI technologies with a good scalability up to thousands of CPU cores. A low-parameteric numerical model of

hypocenter was introduced. As a verification a set of calculations for simple geological models in 3D were carried out. In 2D/3D cases the process of earthquake initiation at shelf was simulated. The contact between water (acoustic approximation) and geological bottom of the sea (full-wave elastic approximation) was explicitly taken into account. The magnitude at hypocenter was estimated with the Richter scale. The obtained time-spatial distribution of elastic stresses may be subsequently used in problems of strength of structures.

The reported study was funded by RFBR according to the research project № 18-37-00127.

OPENFOAM WAVE MODELLING OPTIMIZATION WITH HETEROGENEOUS SYSTEMS APPLICATION PORTING

Nikita Nizovtsov^a, Alexander Degtyarev

Saint-Petersburg State University

E-mail: ^avostatin@gmail.com

In the report application of porting optimization on heterogeneous systems in the field of wave propagation modelling is discussed. Also, reviews of organization of computing in frame of OpenFOAM package and estimations of effectiveness of application porting on heterogeneous systems are given for the wave propagation problem in fluid. Evaluations of the difficulty and time required for implementations of these approaches in relation to performance improvements are considered.

ACCELERATING REAL-TIME SHIP MOTION SIMULATIONS USING GENERAL PURPOSE GPU COMPUTATIONS

**Ivan Petriakov^a, Ivan Gankevich, Alexander Degtyarev,
Vladimir Korkhov**

St Petersburg State University

E-mail: ^ast049350@student.spbu.ru

Software suites for ship simulations are typically used for statistical studies of ship dynamics, but also as a simulator for training ship crew in dangerous situations. One problem that arises during training is speeding-up a part of the session which does not involve actions from the crew. The aim of the study reported here is to accelerate solution of ship motions equations using general purpose computations on GPU. These equations describe dynamics of ship manoeuvring in wavy sea surface, and are central to the simulator programme. The equations are solved numerically via Runge—Kutta—Fehlberg method. Due to high number of floating point operations, computation on GPU achieves considerable speed-up over CPU. High performance solution allows to shorten training sessions and make them more efficient, but also beneficial for statistical studies as it reduces simulation time.

GPGPU IMPLEMENTATION OF SCHRÖDINGER'S SMOKE FOR UNITY3D

Oleg Iakushkin^a, Anastasia Iashnikova, Olga Sedova

Saint-Petersburg State University

E-mail: ^aoleg.jakushkin@gmail.com

The paper describes an algorithm for Eulerian simulation of incompressible fluids—Schrödinger's Smoke. The algorithm is based on representing models as a system of particles. Each particle represents a small portion of a fluid or amorphous material. A particle has a certain 'lifespan', during which it may undergo various changes. The Schrödinger's Smoke solver algorithm was implemented in Unity3D environment. We used Particle System and Compute Shader techniques to transfer the bulk of computational load relating to simulation of physical processes to GPGPU — it allowed real-time interaction with the model. The solution we developed allows to model such effects as interactions between vortex rings — i.e., their collisions and overlapping — with a high degree of physical accuracy.

RSC BASIS PLATFORM - MICRO-AGENT PLATFORM TO MANGE COMPUTER CLUSTERS

Pavel Lavrenko^a, Mikhail Malkov

RSC Group

E-mail: ^ap.lavrenko@rsc-tech.ru

Nowadays there are a lot of loosely connected layers in datacenter management systems such as: datacenter facilities, computing nodes, storage systems, networks or job schedulers. We see the need to have a comprehensive approach to manage these resources as a whole. The talk covers our search for optimal methods and tools to manage data processing infrastructure - to incorporate best practices from different industries into a comprehensive yet flexible platform and to build a community around it. Thus, we would like to demonstrate our RSC BasIS Platform for data center automation, running Govorun cluster in JINR and many others. The platform is under active development, and not only does it provide its users with a way to effectively manage all levels of their HPC datacenters but makes independent systems work together, sharing their resources and users. Operating experience and plans for further development will be discussed. Storage-on-demand: storage and compute united. RSC Tornado hyper-converged solution for data processing

STORAGE-ON-DEMAND: STORAGE AND COMPUTE UNITED. RSC TORNADO HYPER-CONVERGED SOLUTION FOR DATA PROCESSING

Pavel Lavrenko

RSC Group

E-mail: p.lavrenko@rsc-tech.ru

A continuous growth of compute power and amount of data for processing demands a proportional growth of storage system performance and capacity. However traditional ways to scale storage systems are expensive and inflexible, thus we need to look for new approaches. Upon rethinking the datacenter architecture for data processing RSC came up with an idea to converge compute and storage systems. We present our hyper-converged system, a unified solution that is adaptive to storage and compute requirements, which provides its users with top rank performance and flexibility. We will talk about our new hardware appliances with an integrated storage subsystem and software-defined methods to build and manage high performance clusters on demand. The efficiency of the system enables the new JINR supercomputer to rank 9th in the io500 list.

WAYS TO IMPROVE THE PRODUCTIVITY OF FIRE SIMULATION TOOLS ON MODERN EQUIPMENT

Victor Smirnov

St Petersburg State University

E-mail: ariox41@gmail.com

One of the problems for all countries of the world are fires, in particular, fires in the premises. For the creation and effective use of firefighting means, it is necessary to calculate possible scenarios for the development of fires in specific conditions. At the present time, there are various tools for computer modeling of fires, but they have disadvantages - they have either a large error or a low performance. In this paper, mathematical models of fires and possible ways to improve fire modeling tools are considered, in particular, parallelization on GPU and distribution to multiple computers.

INFORMATION-SOFTWARE ENVIRONMENT FOR THE GOVORUN SUPERCOMUTER

**Shushanik Torosyan^{1,a}, Oksana Streltsova¹, Dmitry Podgainy¹,
Martin Vala², Margarit Kirakosya¹**

¹*Joint Institute for Nuclear Research*

²*Institute for Theoretical and Experimental Physics, Košice, Slovak Republic*

E-mail: ^ashusha.torosyan@gmail.com

In order to increase the efficiency of the application development and of the computations using the resources of the GOVORUN supercomputer, an IT-ecosystem that includes a set of services for users is being actively developed. The maintenance of the current services and the development of new ones is a key prerequisite since they secure to the users modern tools for efficient organization of their work under rapidly evolving technologies. This article describes new additions to the information-software environment of the GOVORUN supercomputer.

MERGING MULTIDIMENSIONAL HISTOGRAMS VIA HYPERCUBE ALGORITHM

Martin Vala^{1,2}, Andrey Bulatov^{3,a}, Yurii Butenko¹

¹*Joint Institute for Nuclear Research*

²*Institute for Theoretical and Experimental Physics, Košice, Slovak Republic*

³*Dubna State University*

E-mail: ^aandrey.bulatov20@gmail.com

Scientists in high energy physics produce their output mostly in form of histograms. Set of histograms are saved in output file for each grid job. As the next step is to merge these files/histograms to one file where scientist can produce final plots for publication. Merging of these out files may be done sequentially as one job or do it in parallel via binary tree algorithm as it is done by many users. Using histogram with low dimensions (1D or 2D) one can fit in memory with final merged objects. On the other side, if dimensions or binning of histograms are increased, sparse implementation of histogram has to be used in analysis and final object might grow so much that user will not be able to merge or open final merged object because it will not fit in memory at some point. Our task is merge these multidimensional histograms to N independent objects to multiple files, where each file will contain unique part of merged object sorted by some axis in histogram dimension. For optimization reasons hypercube algorithm is used.

SALSA — SCALABLE ADAPTIVE LARGE STRUCTURES ANALYSIS

Martin Vala^{1,2,a}, Yurii Butenko¹, Branislav Beke³, Fedor Matej³

¹Joint Institute for Nuclear Research

²Institute for Theoretical and Experimental Physics, Košice, Slovak Republic

³SPSEKE, Kosice, Slovakia

Data environments are growing exponentially and the complexity of data analysis is becoming critical issue. The goal of SALSA project is to provide tools to make connection between human and computer to understand and learn from each other. Analysis of different parameters in N-dimensional space should be made easy and intuitive. Task distribution system has to be adaptive to the environment where analysis is done and has provide easy access and interactivity to the user. SALSA contains distribution network system that can constructed at level of clusters, nodes, processes and threads and will be able to build any tree structure. User interface is implemented as web service that can connect to SALSA network and distribute tasks to workers. Web application is using latest web technologies like ReactJS, WebSockets to provide interactivity and dynamism. JavaScript ROOT (JSROOT) package is used as analysis interface. EOS storage support with JSROOT is included to provide possibility to browse files and view results on web browser. Users can create, delete, start and stop tasks. The web application has several templates for different types of user tasks that makes it possible to quickly create new task and submit it to the SALSA network.

HYBRILIT MONITORING SYSTEM

Martin Vala¹, Ivan Kashunin², Yurii Butenko^{2,a}, Dmitry Belyakov²

¹Institute for Theoretical and Experimental Physics, Košice, Slovak Republic

²Joint Institute for Nuclear Research

E-mail: ^agohas94@gmail.com

The heterogeneous cluster HybriLIT and Supercomputer Govorun are designed for the development of parallel applications and for carrying out parallel computations asked by a wide range of tasks arising in the scientific and applied research conducted by JINR. The efficient work on the cluster needs the implementation of service of statistics provided to the users. Even though tasks of monitoring of distributed computing and gathering its statistics are encountered more and more frequently, there is not so many well-known methods to do this. We developing web-service for hybrid heterogeneous cluster “HybriLIT”, that solves that task using Node.JS as it’s server and Angular for a presentment of data. Monitoring itself carried out by a sensor written on C++ with the using of libgtop library. At the moment functions of monitoring CPU load, memory load, network and GPU load of the computing node and browsing that data in both table and graphical form are already implemented. Also, there are diagrams of usage for different laboratories and users, information about currently running jobs and an archive table for a jobs that was computed on a cluster.

GOVORUN SUPERCOMPUTER ENGINEERING INFRASTRUCTURE. MONITORING SYSTEM OF ENGINEERING INFRASTRUCTURE.

**Alexey Vorontsov^a, Andrey Dolbilov, Dmitry Podgainy,
Oksana Streltsova, Vladimir Korenkov, Tatiana Strizh**

Joint Institute for Nuclear Research

E-mail: ^avorontsov@jinr.ru

A complex engineering infrastructure has been developed to support Govorun supercomputer that is expansion of the HybriLIT heterogeneous cluster. This infrastructure combines integration of two solutions on cooling systems: air cooling system for the GPU-component and water cooling system for the CPU-component based on the solution of the RSC Group. The report provides a review of the engineering infrastructure of the supercomputer, and it is important to note that a special emphasis is put on the water cooling system. Review on the monitoring system based on the solution of the "RSC BazIS" which allows managing both separate nodes and all nodes of the infrastructure component will be also presented in the report.

БИБЛИОТЕКИ И ПАКЕТЫ ПРИКЛАДНЫХ ПРОГРАММ, ДОСТУПНЫЕ ПОЛЬЗОВАТЕЛЯМ ЭВМ ОИЯИ

**Людмила Попкова, Александр Сапожников,
Татьяна Сапожникова^a**

Объединенный институт ядерных исследований

E-mail: ^atsap@jinr.ru

Информация о библиотеках и пакетах прикладных программ, поддерживаемых и сопровождаемых в ЛИТ ОИЯИ — JINRLIB, CERNLIB, CPCLIB — размещена на сайте <http://wwwinfo.jinr.ru/programs/>. JINRLIB (<http://www.jinr.ru/programs/jinrlib/>) — библиотека программ, предназначенных для решения широкого круга математических и физических задач. Пополнение библиотеки происходит новыми программами, создаваемыми сотрудниками ЛИТ ОИЯИ и их коллаборантами. В зависимости от способа сопровождения и распространения библиотека делится на две части: одна часть распространяется в виде объектных модулей, другая - в виде автономных пакетов прикладных программ. Библиотеки объектных модулей готовятся на компьютерах Многофункционального информационно - вычислительного комплекса ОИЯИ с ОС Linux, а также на компьютерах с ОС Windows для всех доступных фортранных трансляторов. Программы, которые по разным причинам не могут распространяться в виде библиотек объектных модулей, также размещаются в JINRLIB. Вся информация, предоставленная автором программы, помещается на WWW-сайт. В настоящий момент насчитывается более 60 программных пакетов, большинство которых решает задачи автоматизации обработки экспериментальных данных и вычислительной

математики. В последнее время происходит бурное развитие технологий программирования параллельных вычислений, в частности, MPI. Эта тенденция нашла свое отражение и в библиотеке JINRLIB, где создан раздел для программ с использованием технологии MPI. Специализированный WWW — сайт обеспечивает электронный доступ к библиотеке JINRLIB, где можно найти описания программ и программных пакетов, исходные тексты, библиотеки объектных модулей. Ведется каталог вновь поступивших программ и программных пакетов. Для получения статистики использования программ заведены счетчик посещений страницы и счетчик количества скачиваний текстов программ. По каталогу JINRLIB создана база данных авторов программ библиотеки, на основании которой строится таблица авторов со списком программ, имеется возможность получить список программ нужного автора. Библиотека CERNLIB (<http://wwwinfo.cern.ch/asd/index.html>) — большая коллекция программ, поддерживаемых и распространяемых на исходном языке, в объектном коде и в виде готовых программ. Большинство этих программ разработано в CERN и ориентировано на решение физических и математических проблем. Библиотека CERNLIB в CERN сейчас не поддерживается, но, учитывая интерес к ней FORTRAN-ориентированных пользователей, была выполнена пересборка наиболее востребованных программ библиотеки MATHLIB для OS Windows. CPCPL — международная библиотека программ журнала Computer Physics Communications (CPC) — в настоящее время является одним из самых представительных и хорошо организованных банков программ, решающих задачи физики, математики, химии и других смежных областей знаний. ОИЯИ имеет подписку на журнал CPC и библиотеку программ, и по лицензионному соглашению сотрудники ОИЯИ имеют доступ к общей информации и к программам библиотеки.

О СОГЛАСОВАНИИ ВЫЧИСЛИТЕЛЬНОГО ЭКСПЕРИМЕНТА ПРИ ИНТЕРАКТИВНОМ МОДЕЛИРОВАНИИ ГИДРОМЕХАНИКИ КОРАБЛЯ В ШТОРМОВОМ МОРЕ

**Василий Храмушин^а, Александр Дегтярев,
Александр Богданов, Иван Ганкевич**

Санкт-Петербургский государственный университет

E-mail: ^аv.khram@gmail.com

При реализации сложных прикладных вычислительных экспериментов, в проектировании, разработке и построении программных комплексов востребуется особая логика синтеза числовых структур для описания физических явлений в тесной связи с требованиями эффективного применения моделирующих операций гидромеханики на фоне непрерывной графической визуализации всех пространственных процессов. В конкретной задаче формулируются требования к декларативному представлению трехмерной геометрии формы корпуса корабля, контурные линии которого описываются непрерывными, но неоднозначными функциями. Корпус подвергается вынужденным кинематическим перемещениям без деформации, под воздействием динамически нестабильного морского волнения, подверженного непрерывной трансформации в рамках трохойдальной теории групповых структур ветровых волн и зыби, параметры которых задаются по типовым

записям судовых метеопостов. Условно статическое или явное описание динамики корабля и морского волнения приводит к функциональным методам реализации вычислительных операций, моделирующих нестационарную механику взаимодействия локальных фрагментов судовой обшивки с гребнями обрушающихся морских волн, как волн теоретически предельной высоты. Проектная взаимосвязь геометрических объектов, физических явлений и нестационарных процессов гидромеханики синтезируется с помощью троичной матрицы [1]: Корабль в шторм - корпус - волнение - визуализация: 1. геометрия: теоретический чертеж - группы трохoidalных волн - 3D графика OpenGL; 2. гидростатика: остойчивость - силовое воздействие волн - несвободная динамика и качка; 3. механика взаимодействия: излучение волн корпусом корабля - трансформация волн - механика волнообразования

Таким образом, по строкам логической матрицы располагаются геометрические объекты и физические явления в качестве независимых «существительных» или структур данных со связанным набором алгоритмов – методов быстрого доступа к данным. По столбцам формализуются операции гидромеханики, конкретное построение которых существенно зависит от текущего состояния гидромеханики корпуса, морского волнения и их взаимодействия. При этом, в последнем столбце концентрируются результирующие материалы, предназначенные для детальной визуализации и представления морскому или корабельному инженеру-исследователю, изучающему мореходные качества или создающему проекты перспективных кораблей для конкретных географических и навигационных условий эксплуатации.

Список литературы

[1] Храмушин В. Н. Трехмерная тензорная математика вычислительных экспериментов в гидромеханике. Владивосток: ДВО РАН, 2005. 212 с. Shipdesign.ru/Khram/Tensor.html

КОНТЕКСТНАЯ ГРАФИЧЕСКАЯ СРЕДА ПРОСТРАНСТВЕННОЙ ВИЗУАЛИЗАЦИИ РЕЗУЛЬТАТОВ ВЫЧИСЛИТЕЛЬНЫХ ЭКСПЕРИМЕНТОВ В МЕХАНИКЕ СПЛОШНЫХ СРЕД

Василий Храмушин

Санкт-Петербургский государственный университет

E-mail: v.khram@gmail.com

В исследовательском проектировании, построении и последующем анализе достоверности моделируемых процессов в ресурсоемких вычислительных экспериментах, что особо востребуется при изучении нестационарных процессов в механике сплошных сред, становится весьма актуальным использование открытой и легко модифицируемой программной среды для пространственной визуализации быстротекущих физических явлений непосредственно в ходе суперкомпьютерных расчетов. Важным условием такой визуализации является минимальность воздействия на вычислительные процессы, с возможностью внешнего влияния на реологические параметры моделируемой физической среды и критерии динамической или гибридной перестройки вычислительных процессов. Практически все современные вычислительные комплексы обладают встроенными графическими средствами, обеспечивающими быструю визуализацию пространственных геометрических объектов с использованием независимых многоядерных процессоров, которые в полной мере способны обеспечивать решение сформулированной задачи для параллельной визуализации текущих результатов без существенного влияния на основные вычислительные процессы. В настоящем исследовании рассматривается вариант построения программного комплекса на базе графической среды программирования OpenGL, окружаемой инструментальными средствами для работы со временем и интервальными таймерами, устройствами ввода информации и представления текстовых данных на предельно низком уровне прямого ввода/вывода информации и обработки прерываний в OS Windows.

В реализации унифицированного объектно-ориентированного комплекса программ, названного Window-Place, представляется многооконный интерфейс Window для работы с внешними устройствами и компьютерной аппаратурой, с наложением текстовых и графических страниц Place, непосредственно поддерживающих контекстно-зависимые примитивы и операции OpenGL. Встроенные виртуальные процедуры для масштабирования наложенных страниц, обработки программных и аппаратных прерываний и др., создают интуитивно понятное и визуально естественное поведение графических площадок и текстовых страниц. При этом полиморфизм производных классов для прикладных пользовательских вычислительных объектов, допускает подмену всех базовых виртуальных функций, что может быть полезным для ускорения вычислений за счет отмены или упрощения изначально встроенных в Window-Place процедур при управлении процессами визуализации непосредственно в ходе ресурсоемкого суперкомпьютерного эксперимента.

Храмушин В.Н. «Контекстная графика» – (Window-Place) Контекстно-зависимая среда построения трехмерной графики OpenGL с использованием виртуальных процедур C++ и многооконного интерфейса Windows со стековым наложением графических и текстовых фрагментов. Сертификат Роспатента, СахГУ № 2010615850 от 8 сентября 2010 г. Shipdesign.ru/SoftWare/2010615850.html

**CONSOLIDATION
AND INTEGRATION
OF DISTRIBUTED RESOURCES**

CURRENT STATUS OF DATA CENTER FOR COSMIC RAYS BASED ON KCDC

Dmitriy Kostunin¹, Victoria Tokareva^{2a}

¹*Karlsruhe Institute of Technology*

²*Joint Institute for Nuclear Research*

E-mail: ^atokareva@jinr.ru

We present a current status of data center based on KCDC (KASCADE Cosmic Ray Data Centre), which was originally designed for providing an open access to the events measured and analyzed by KASCADE-Grande, a cosmic-ray experiment located in KIT, Karlsruhe. In the frame of the Russian-German Astroparticle Data Life Cycle Initiative we extend KCDC in order to provide an access to different cosmic-ray experiments and make possible aggregation and joint querying of heterogeneous air-shower data. In the present talk we discuss the description of data and metadata structures, implementation of data querying and merging, and first results on including data of experiments located in Tunka, Russia, in this common data center.

THE INTEROPERABILITY PROBLEM DURING THE IMPLEMENTATION OF THE FOURTH PARADIGM

Alexander Oleynikov^a, Andrey Kamenshchikov^b

Kotelnikov Institute of Radioengineering and Electronics of RAS

E-mail: ^aolein@cplire.ru, ^bprostonau@mail.ru

As it is known, at the present time worldwide there is a transition to the so called FOURTH PARADIGM in the methods and means of scientific researches [1]. The essence of the FOURTH PARADIGM is the intensive use of information technologies for the simultaneous using of large amounts of experimental data, numerical simulation results and accumulated knowledge. It is obvious that this requires using of high-performance distributed environment, which includes individual supercomputers, clusters, GRID-systems, cloud computing systems and end-user personal computers. It is also quite obvious that in such a purely heterogeneous environment, which should be classified as a System of Systems (SOS), there is a problem of compatibility and interaction of heterogeneous software and hardware platforms, called "interoperability problem". The interoperability problem should be solved on the basis of the use of profiles - sets of ICT-standards, this problem is dealt with by many organizations and individual researchers around the world, but it is never solved until the end, due to the great complexity. In particular, therefore, the issue of interoperability and development of standards are included in the RAS Program of fundamental researches for 2013-2020 (clause 34). The authors investigate the problem of interoperability for more than 10 years and have developed several standards. The main result should be considered the proposed unified approach to ensuring interoperability for information systems (IS) of the widest class [2], which is subsequently issued in the form of GOST R 55062-2012. Further, the authors applied this

approach to specific areas: e-science, e-education, e-health, e-libraries, e-military. Research was also conducted for GRID and cloud computing systems. The authors have regularly reported their results at previous JINR conferences since 2010 [3]. All of the above classes of IS are components of the SoS. Therefore, using these developments and foreign experience [4], we have now started to solve the problem of interoperability in SoS. The problem is very difficult and we have only preliminary results.

This work is done with the support of the Program №. 27 of fundamental researches of the RAS Presidium.

References

- [1] The Fourth Paradigm Data-Intensive Scientific Discovery. Microsoft research Redmond, WA
- [2] Gulyaev Yu. V., Zhuravlev E. E., Oleynikov A.Ya. Methodology standardization to ensure interoperability of information systems wide class. Zhurnal Radioelektroniki - Journal of Radio Electronics. 2012. N2. Available at: <http://jre.cplire.ru/win/mar12/2/text.pdf>, accessed: 14.12.2017
- [3] S. V. Ivanov, A. Y. Oleynikov. Methodology and algorithm for selecting standards for interoperability profile in cloud computing. Proceedings of the 7th International conference " Distributed computing and Grid technology in science and education, Dubna, JINR, 4-9, July 2016 pp.264-268.
- [4] Information system Development: improving enterprise communication. (chapters 7 and 9). Proceedings from the 22nd annual meeting (ISD2013) held in Seville, Spain, from September 2 to 4, 2013

DIRAC AT JINR — PURPOSE, EXPERIENCE, FUTURE

Igor Plevanyuk^{1a}, Andrei Tsaregorodtsev²

¹Joint Institute for Nuclear Research

²Center for Particle Physics of Marseilles IN2P3 CNRS

E-mail: ^apelevanyuk@jinr.ru

The Joint Institute for Nuclear Research is an international intergovernmental organization. It is a large multidisciplinary scientific center incorporating fundamental research in many different areas. This implies extensive need in computing and storage resources. DIRAC is a general purpose system to provide common access to a number of heterogeneous resources. DIRAC have been installed and configured in JINR for several purposes. First, it is a possible solution for NICA computing. Monte-Carlo generation for MPD detector has been successfully performed on the JINR installation. Second, it was used to provide unified access to several clouds in the JINR Member States. Third, it is used by students to learn grid-technologies and distributed computing.

MODERN E - INFRASTRUCTURE FOR SCIENCE AND EDUCATION IN MOLDOVA BASED ON THE RENAM-GEANT PLATFORM

**Grigore Secrieru^a, Peter Bogatencov, Nicolai Iliuha,
Nichita Degteariov**

RENAM

E-mail: ^asecrieru@renam.md

The article is devoted to the analysis of approaches and solutions for the development and use of the electronic infrastructure for offering services to science and education in the Republic of Moldova. In the paper is considering of trends in the development of the electronic infrastructure and services in the national network RENAM, which provides effective information support for basic scientific and educational processes. The prospects of creating new optical CBF (Cross Border Fibers) links and other components of the electronic platform RENAM-GEANT development on the basis of the UE EaPConnect project are described. The strategy of development of modern regional e-Infrastructure resources and the provision of services that are focused on support of scientific and educational communities in the countries included in the European EaP program are considered.

DISCRETE AND GLOBAL OPTIMIZATION IN EVEREST DISTRIBUTED ENVIRONMENT BY LOOSELY COUPLED BRANCH-AND-BOUND SOLVERS

Vladimir Voloshinov^a, Sergey Smirnov

Institute for Information Transmission Problems RAS

E-mail: ^avladimir.voloshinov@gmail.com

The report presents a new approach to solving rather hard discrete and global optimization problems in Everest, <http://everest.distcomp.org>, a web-based distributed computing platform. The platform enables convenient access to heterogeneous resources (standalone servers, high performance clusters etc.) by means of domain-specific computational web services, development and execution of many-task applications, and pooling of multiple resources for running distributed computations. Rather generic Everest-application had been implemented for solving discrete and global optimization problems - so called DDBNB, Domain Decomposition Branch-and-Bound, <https://github.com/distcomp/ddbnb>. DDBNB implements a kind of coarse-grained parallelization of Branch-and-Bound (BNB) method. It supports two strategies (including combined usage of both): decomposition of feasible domain into a set of sub-problems; multisearch (or concurrent) solving the same problem with different settings of the BNB-method. In both cases several BNB-solver's processes exchange incumbents, best values of goal function on feasible solutions, they found. DDBNB uses

generic Everest messaging service and open source solvers : CBC, <https://projects.coin-or.org/Cbc>; SCIP, <http://scip.zib.de>. By now we got some experience in solving different optimization problems: Travelling Salesman Problem (TSP) as Mixed-Integer Linear Programming; global optimization problems with all continuous variables (so called Tammes and Thomson problems, both relate to sphere packing) and global optimization with continuous and binary variable (so called Flat Torus Packing problems). For TSP, computing experiments are presented, when DDBNB worked in different modes: Domain Decomposition only, multiseach only; combined mode. As to global optimization, all problems had been reduced to the form of mathematical programming problems with polynomial functions (solver SCIP supports solving of such class of problems). Here, only Domain Decomposition had been used and different methods of decomposition in the case of sphere's and torus's packing and corresponding computing experiments are presented. Formulation of original problems and decomposition strategies have been implemented via Pyomo, <http://www.pyomo.org>. All sub-problems have been passed to solvers as AMPL-stubs, (also known as NL-files). Usage of Everest platform enables to involve in experiments computing resources of Center for Distributed Computing, <http://distcomp.ru>, HPCs of NRC Kurchatov Institute and South Ural State University.

The reported study was funded by RFBR according to the research projects #18-07-01175 and #18-07-00956.

DATABASES, DISTRIBUTED STORAGE SYSTEMS, DATALAKES

GEOMETRY DATABASE FOR THE CBM EXPERIMENT AND ITS FIRST APPLICATION TO EXPERIMENTS IN THE NICA PROJECT

**Elena Akishina¹, Evgeny Alexandrov¹, Igor Alexandrov¹,
Irina Filozova^{1,a}, Volker Friese², Victor Ivanov^{1,3},
Oleg Rogachevskiy¹, Konstantin Gertsenberger¹**

¹*Joint Institute for Nuclear Research*

²*GSI Darmstadt*

³*National Research Nuclear University "MEPhI", Moscow, Russia*

E-mail: ^afia@jinr.ru

This paper is dedicated to the current state of the Geometry Database (Geometry DB) for the CBM experiment and a first result of using the Geometry DB for NICA project. Geometry DB is an information system that supports the CBM geometry. The main aims of Geometry DB are to provide the storage of the CBM geometry, to manage the geometry modules, to assemble various setups as combinations of geometry modules and additional files, to provide its support. The development takes into account the specifics of the workflow for simulation of particles transport through the setup. Both Graphical User Interface (GUI) and Application Programming Interface (API) are available for members of the CBM collaboration. In our approach, the details of the geometry modules are stored in the format of ROOT files. Such a technique allows using the Geometry DB in the NICA project: BM@N and MPD experiments.

A NEW APPROACH TO THE DEVELOPMENT OF PROVENANCE METADATA MANAGEMENT SYSTEMS FOR LARGE SCIENTIFIC EXPERIMENTS

Andrey Demichev^a, Alexander Kryukov

Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University

E-mail: ^ademichev@theory.sinp.msu.ru

Provenance metadata (PMD) contain key information that is necessary to determine the origin, authorship and quality of corresponding data, their proper storage, correct using, and for interpretation and confirmation of relevant scientific results. The need for PMD is especially essential when big data are jointly processed by several research teams, which is a very common practice in many scientific areas of late. This requires a wide and intensive exchange of data and programs for their processing and analysis, covering long periods of time, during which both the data sources and the algorithms for their processing may be modified. Although a number of projects have been implemented in recent years to create management systems for such metadata, but the vast majority of implemented solutions are centralized, which is poorly suited to current trends of working in distributed environments,

open data access models, and the use of metadata by organizationally unrelated or loosely coupled communities of researchers. We propose to solve this problem by using a new approach to creating a distributed registry of provenance metadata based on blockchain technology and smart contracts. In this work, the functional requirements for the PMD management system were formulated. Based on these requirements, we investigated the problem of the optimal choice of the type of blockchain for such a system, as well as the optimal choice of consensus algorithm for records ordering within the blockchain without participation of third-party trusted bodies. The architecture and algorithms of the system operation, as well as its interaction with the distributed storage resources management systems, are proposed. Specific use cases for the PMD management system are considered. A number of existing blockchain platforms are considered and the most preferable one is selected. The results of this work are of particular importance in the big data era, when a full analysis of the results of experiments is often not possible for one team, so that many independent teams take part in their analysis. The suggested approach is currently under implementation in SINP MSU in the framework of the project supported by the Russian Science Foundation (grant No 18-11-00075).

PROBLEMS OF DATE AND TIME DATA TYPES IN RELATIONAL MODEL OF DATA

Vladimir Dimitrov

University of Sofia

E-mail: cht@fmi.uni-sofia.bg

Several years after the initial announcement of the relational model of data, Codd published a review on the model, so called Version 2. This review is based on the experience of relational database systems implementation in the intermediate period. One of the main corrections is recommendation on date and time data types. This paper reinvestigate the topic from the nowadays point of view.

DATA KNOWLEDGE BASE FOR THE ATLAS COLLABORATION

Marina Golosova^a, Maria Grigorieva

National Research Center "Kurchatov Institute"

E-mail: ^agolosova.marina@gmail.com

ATLAS experiment at the CERN LHC is one of the most data-intensive modern scientific apparatus. To help managing all the experimental and modelling data, multiple information systems were created during the experiment's lifetime (more than 25 years). Each such system addresses one or several tasks of data and workload management, as well as information lookup, using specific sets of metadata (data about data). Growing data volumes and the

computing infrastructure complexity require from researchers more and more complicated integration of different bits of metadata from different systems using different conditions.

A common problem are multi-system join requests, which are not easy to implement in timely manner and, obviously, are less efficient than a request to a single system with integrated and pre-processed information would be. To address this issue, a joint team of researchers and developers from Kurchatov Institute and Tomsk Polytechnic University has initiated the Data Knowledge Base (DKB) R&D project in 2016. This project is aimed at knowledge acquisition and metadata integration, providing fast response for a variety of complicated requests, such as finding articles, based on same or similar data samples (search by links between objects), summary reports and monitoring tasks (aggregation requests), etc.

WLCG DATA LAKE PROTOTYPE FOR HL-LHC

**Ivan Kadochnikov^{1,a}, Ian Bird², Gavin McCance²,
Jaroslava Schovancova², Maria Girone², Simone Campana²,
Xavier Espinal Currul²**

¹Joint Institute for Nuclear Research

²CERN, the European Organization for Nuclear Research

E-mail: ^akadivas@jinr.ru

A critical challenge of high-luminosity Large Hadron Collider (HL-LHC), the next phase in LHC operation, is the increased computing requirements to process the experiment data. Coping with this demand with today's computing model would exceed a realistic funding level by an order of magnitude. Many architectural, organizational and technical changes are being investigated to address this challenge. This talk describes the prototype of a WLCG data lake, a storage service of geographically distributed data centers connected by a low-latency network. The architecture of a EOS data lake is presented, showing how it leverages economy of scale to decrease cost. The talk discusses first experiences with the prototype and benchmark jobs reading data from the lake.

NRC "KI" PARTICIPATION IN DATALAKE PROJECT

Andrey Kiryanov^{1,a}, Andrey Zarochentsev²

¹*NRC "Kurchatov Institute" PNPI*

²*St Petersburg State University*

E-mail: ^aglobus@pnpi.nw.ru

WLCG DataLake R&D project aims at exploring an evolution of distributed storage while bearing in mind very high demands of HL-LHC era. Its primary objective is to optimize hardware usage and operational costs of a storage system deployed across distributed centers connected by fat networks and operated as a single service. Such storage would host a large fraction of the WLCG data and optimize the cost, eliminating inefficiencies due to fragmentation. In this talk we will explain NRC "KI" role in the DataLake project with highlight on our goals, achievements and future plans.

INTEGRATING LEAF TO DATA MANAGEMENT WORKFLOW IN LHAASO

Haibo Li^a, Yaodong Cheng, Qi XU, Qiulan Huang

Institute of High Energy Physics, Chinese Academy of Sciences

E-mail: ^alihaibo@ihep.ac.cn

Nowadays, data storage and management in cloud computing environment has been very important in high energy physics field. The LHAASO(Large High Altitude Air Shower Observatory) experiment of IHEP will generate 2 PB per year in the future. This massive data processing faces many challenges in the distributed computing environment. For example, some sites may have no local HEP storage which made the distributed computing unavailable. Our goal is to make the data available for LHAASO in any remote sites. In our architecture, we use EOS as our local storage system, and use LEAF as the data federation system. LEAF is a data cache and access system across remote sites proposed by IHEP. LEAF can present one same file system view at local and the remote sites, supporting directly data access on demand. In this paper, we will present the whole data management architecture, data workflow and performance evaluation of LEAF in LHAASO.

A DISTRIBUTED DATA WAREHOUSE SYSTEM FOR ASTROPARTICLE PHYSICS

Minh Duc Nguyen^a

Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University

E-mail: ^anguyendmitri@gmail.com

A distributed data warehouse system is one of the actual issues in the field of astroparticle physics. Famous experiments, such as Tunka, Taiga, produce tens of terabytes of data measured by their instruments. It is critical to have a smart data warehouse system on-site to store the collected data for further distribution effectively. It is also vital to provide scientists with a handy and user-friendly interface to access the collected data with proper permissions not only on-site but also online. The latter case is handy when scientists need to combine data from different experiments for analysis. In this work, we describe an approach to implementing a distributed data warehouse system that allows scientists to acquire just the necessary data from different experiments via the Internet on demand. The implementation is based on the CERN CVMFS with additional components developed by us to search through the whole available data sets and deliver their subsets to users' computers.

EVOLUTION OF DATA STORAGE AND HANDLING AT ITALIAN WLCG DATA CENTER

**Vladimir Sapunenko^a, Luca Dell'Agnello, Gaetano Maron,
Daniele Cesini**

INFN

E-mail: ^avladimir.sapunenko@cnaif.infn.it

The Italian WLCG Data Center which is hosted and managed by National Center of INFN for Informatics and Telecommunications (CNAIF) located in Bologna, is the one of the biggest European computing centers for WLCG. Operational since 2005 it provides storage resources for all the Large Hadron Collider (LHC) experiments, as well as for many other scientific collaborations. In 12 years of operations with increasing demands and growing resources the data storage and handling services underwent some evolutionary changes. Here we will discuss our experience we have collected during these years and will present our vision for future developments in the Data Storage and Handling services for WLCG-oriented sites.

НОВЫЙ ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКИЙ ПОДХОД РАСЧЕТА ИНФОРМАЦИОННЫХ ПОТЕРЬ В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ ХРАНЕНИЯ И ОБРАБОТКИ ДАННЫХ ФИЗИЧЕСКИХ ЭКСПЕРИМЕНТОВ

**Андрей Нечаевский, Геннадий Ососков, Дарья Пряхина^а,
Владимир Трофимов**

Объединенный институт ядерных исследований

E-mail: ^аpryahinad@jinr.ru

В рамках работ по созданию компьютерной системы хранения и обработки данных установок ВМ@N и MPD, входящих в проект коллайдера NICA, возникает проблема выбора оптимальной конфигурации необходимого компьютерного и сетевого оборудования. Для решения этой проблемы требовалось разработать и исследовать модель перемещения данных внутри системы. Предыдущий опыт моделирования авторов настоящей статьи [1], показал, что описанные в литературе подходы моделирования процессов обработки потока заданий в распределенных и облачных системах [2,3], не подходят для анализа потоков данных, поскольку в библиотеках указанных моделирующих программ выполняется детализация потока данных до уровня пакета или файла, что приводит к сложной организации программ и большим вычислительным затратам. Поэтому нами предложен и реализован подход, рассматривающий процесс перемещения данных, как поток байтов, имеющий статистическую природу, без анализа отдельных частей этого потока. Для оценки различных конфигураций оборудования использовался вероятностно-статистический подход, при котором определяются вероятности потерь информации, поступающей с детекторов для каждой из этих конфигураций. В качестве причины потерь рассматривается переполнение буферов на одной из стадий накопления и передачи данных. Оптимальной конфигурацией считается та, что имеет минимальную стоимость при заданном допустимом уровне потерь. Для реализации этой схемы моделирования компьютерной системы хранения и обработки данных потребовалось, прежде всего, описать эту систему с помощью набора параметров, которые могут быть дефолтными или задаются пользователем. К параметрам относятся размеры дисковых буферов, количество потоков данных, пропускные способности каналов передачи и т.п. Для каждого параметра должны быть определены его граничные значения и шаг его изменения. На основе подготовленного набора параметров формируется поток независимых заданий расчёта потерь при передачах данных за время физического сеанса работы ускорителя. Таким образом, для осуществления процесса моделирования перемещения данных в системе и подсчета происходящих потерь потребовалось разработать два программных модуля. Первый записывает в базу данных сформированный набор параметров и автоматически строит и записывает в базу данных набор независимых заданий. Собственно расчёт выполняет второй модуль. Он выбирает из базы задание, строит конфигурацию оборудования, в соответствии с заданными параметрами, разыгрывает интенсивность потока данных на каждом шаге и рассчитывает процесс миграции данных в системе. Предложенная схема допускает параллельный расчёт вариантов, что позволяет анализировать значительное

количество вариантов (десятки тысяч). Как перспектива развития этого подхода рассматривается создание классов, которые позволяют гибко менять топологию системы хранения данных. В существующем варианте она ориентирована только на анализ потерь при работе установок $BM@N + MPD$. Полученные к настоящему времени результаты моделирования позволили вести с проектировщиками систем DAQ и триггеров содержательные и аргументированные дискуссии по поводу параметров потоков данных, способствующие принятию мотивированных решений.

BIG DATA ANALYTICS, MACHINE LEARNING

SEMANTIC INFORMATION MANAGEMENT: THE APPROACH TO SEMANTIC ASSETS DEVELOPMENT LIFECYCLE

Yury Akatkin^a, Elena Yasinovskaya, Michael Bich

Plekhanov Russian University of Economics

E-mail: ^auakatkin@semanticpro.org

The application of semantic integration methods meets challenges, which arise during collaboration between IT-specialists and domain experts at the model building stage. These challenges can affect correct formalization of the domain as well as the outcome of the integration in distributed information systems as a whole. The creation of a collaborative platform for semantic integration which provides the (re)use of semantic assets (SA) is suggested to overcome the lack of semantic interoperability. The analysis of the limitations existing in standard SA management leads the authors to propose the collaborative approach, based on an extended lifecycle of semantic assets. The authors consider the implementation of the platform based on the Asset Description Metadata schema extension to be a valid option.

CHECKING FOREIGN COUNTERPARTY COMPANIES USING BIG DATA

Lazar Badalov¹, Sergey Belov^{2,a}, Ivan Kadochnikov²

¹*Plekhanov Russian University of Economics*

²*Joint Institute for Nuclear Research*

E-mail: ^abelov@jinr.ru

The project aims to create a database of companies and company data and an automated analytical system based on this data. The development of the system will allow credit institutions to obtain information about the links between companies, to carry out a policy of "Know your customer" - to identify the final beneficiaries, to assess risks, to identify relationships between customers. For the moment, there are some projects like OpenCorporates having global databases of companies collected from a large number of jurisdictions. But at the same they cover neither all the national registries, nor other useful data sources (courts, customs, press, etc.). Also the existing services have rather sketchy abilities on searching for relations between companies, which are not always direct. The project we present is about to overcome main of these deficiencies. Number of companies worldwide is more than 150 millions. Having company information from many sources, there is no other reasonable way to process it using Big Data technologies. In the research we use such technologies along with machine learning and graph databases.

LABOUR MARKET MONITORING SYSTEM

**Sergey Belov^{1,a}, Irina Filozova¹, Ivan Kadochnikov¹,
Vladimir Korenkov¹, Roman Semenov¹, Petr Zrelov¹, Pavel Smelov²**

¹*Joint Institute for Nuclear Research*

²*Plekhanov Russian University of Economics*

E-mail: ^abelov@jinr.ru

Last years, the prospects for digital transformation of economic processes were actively discussed. It is quite a complex problem having no solution with traditional methods. Opportunities of the qualitative development of the transformation are illustrated by the example of use of Big Data analytics, in particular intellectual text analysis, for the assessment of the needs of regional labour markets in the man-power. The problem is solved using the developed by the authors the automated information system of monitoring of matching the staffing needs of employers with the training level. The system presented use the information gathering from open data sources and provides additional opportunities to identify qualitative and quantitative interrelation between the education and the labour market. The system is targeted at a wide range of users: authorities and management of regions and municipalities; the management of universities, companies, recruitment agencies; graduates and prospective students.

TEXT SEGMENTATION ON PHOTOREALISTIC IMAGES

Valery Grishkin^a, Aleksaner Ebral, Nikolai Stepenko, Jean Sene

St.Petersburg State University

E-mail: ^avalery-grishkin@yandex.ru

The paper proposes an algorithm for segmentation of text, applied or presented in photorealistic images, characterized by a complex background. Because of its application, the exact location of image regions containing text is determined. The algorithm implements the method for semantic segmentation of images, while the text symbols serve as detectable objects. The original images are pre-processed and fed to the input of the pre-trained convolutional neural network. The paper proposes a network architecture for text segmentation, describes the procedure for the formation of the training set, and considers the algorithm for pre-processing images, reducing the amount of processed data and simplifying the segmentation of the object "background". The network architecture is a modification of well-known ResNet network and takes into account the specifics of text character images. The convolutional neural network is implemented using CUDA parallel computing technology at the GPU. The experimental results for evaluating quality of the text segmentation IoU (Intersection over Union) criterion have proved effectiveness of the proposed method.

DEEP LEARNING METHODOLOGY FOR PREDICTION OF LONG-TERM DYNAMICS OF FINANCIAL DERIVATIVES

Alexander Bogdanov^{1,a}, Vladimir Rukovchuk², Kirill Lysov¹

¹*St.Petersburg State University*

²*CSA*

E-mail: ^abogdanov@csa.ru

Algorithms for predicting the dynamics of stock options and other assets derivatives for both small times (where one plays on market fluctuations), and medium ones (where trade is stressed at the beginning and closing moments) are well developed, and trading robots are actively used for these purposes. Analysis of the dynamics of assets for very long time-frames (of several months order) is still beyond the scope of analysts as it is expensively prohibited, although this issue is extremely important for hedging the investments portfolios. In the paper the dynamic processes in the stock market in long-term periods are considered. Pricing of portfolio investments dynamics is made on the basis of neural networks using the deep learning and soft computing methodology. It does not require heavy computational resources, and their relatively low accuracy is not a disadvantage in tasks where only trends are subject to consideration. Operation with two and three layers neural networks produced until recently still unfitting results. However, emergence of the suggested approaches with specialized processors and software for learning the multi-layer networks has changed the situation. The most important factor is a high quality trained artificial neural network and its ability to predict for a long time-frame without retraining. The number of layers in experiments reached 250. For network input data the real S&P500 price series was taken dated from 1950 till 2017 with several one-day steps. Model predictions vs true S&P500 price performance has demonstrated practically acceptable compliance.

CLUSTER ANALYSIS OF SCIENTIFIC PAYLOAD TO EXECUTE IT EFFICIENTLY IN DISTRIBUTED COMPUTING ENVIRONMENT

Maksim Gubin^{1,a}, Mikhail Titov², Maria Grigorieva²

¹*National Research Tomsk Polytechnic University*

²*National Research Center "Kurchatov Institute"*

E-mail: ^agubin.m.u@gmail.com

Every modern scientific experiment deals with the processing of large amounts of experimental data (up to exabytes) employing millions of computing processes and delivering corresponding scientific payloads. It raises the task of the efficient management of the batch processing of payloads, that should consider the well-defined grouping mechanism. Understanding naturally occurring groupings of payloads would increase the processing rate and improve the scheduling. The automated discovery of the payloads groups should consider

not only the descriptive parameters of the payload itself but the characteristics of its interaction processes with computing resources and the computing environment. Our work is focused on applying machine learning methods to solve the stated problem, and particularly to evaluate and to use the approach of cluster analysis. Besides the ultimate goal, it will benefit the other related analytical services and processes aimed at analyzing particular payload parameters (e.g., prediction process) and/or set of parameters (e.g., correlations discovery).

RUSSIAN-LANGUAGE SPEECH RECOGNITION SYSTEM BASED ON DEEPSPEECH

**Oleg Iakushkin^a, George Fedoseev, Anna Shaleva, Olga Sedova,
Alexander Degtyarev**

St.Petersburg State University

E-mail: ^aoleg.jakushkin@gmail.com

The paper examines the practical issues in developing a speech-to-text system using deep neural networks. The development of a Russian-language speech recognition system based on DeepSpeech architecture is described. The Mozilla company's open source implementation of DeepSpeech for the English language was used as a starting point. The system was trained in a containerized environment using the Docker technology. It allowed to describe the entire process of component assembly from the source code, including a number of optimization techniques for CPU and GPU. Docker also allows to easily reproduce computation optimization tests on alternative infrastructures. We examined the use of TensorFlow XLA technology that optimizes linear algebra computations in the course of neural network training. The number of nodes in the internal layers of neural network was optimized based on the word error rate (WER) obtained on a test data set, having regard to GPU memory limitations. We studied the use of probabilistic language models with various maximum lengths of word sequences and selected the model that shows the best WER. Our study resulted in a Russian-language acoustic model having been trained based on a data set comprising audio and subtitles from YouTube video clips. The language model was built based on the texts of subtitles and publicly available Russian-language corpus of Wikipedia's popular articles. The resulting system was tested on a data set consisting of audio recordings of Russian literature available on voxforge.com — the best WER demonstrated by the system was 18%.

TEXTURE GENERATION FOR ARCHAEOLOGICAL RECONSTRUCTIONS

**Oleg Iakushkin¹, Anna Fatkina², Dmitry Selivanov^{1,a},
Alexander Degtyarev¹**

¹St.Petersburg State University

²Joint Institute for Nuclear Research

E-mail: ^admitry@batbox.ru

The paper describes a solution that reconstructs the texture in 3D models of archeological monuments and performs their visualization. The software we have developed allows to model the outward surface of objects in various states of preservation. Drawings and photographs of preserved wall fragments and stonework elements are used in the modelling process. Our work resulted in development of a texturing system that reconstructs textures of a given object based on photographs and fragments of drawings. The major distinguishing feature of the system is that it can reconstruct textures using limited and low-quality input data. For instance, the input data fed to the system may consist of photographs of an object taken with an ordinary camera (e.g., with a smartphone). In developing the system, we used OpenCV, CGAL and AwesomeBump open source computer vision packages.

MACHINE LEARNING FOR NATURAL LANGUAGE PROCESSING TASKS

Aleksey Kulnevich^a, Vladislav Radishevskiy

National Research Tomsk Polytechnic University

E-mail: ^akulnevich94@mail.ru

There are two popular algorithms for text vector extraction: bag of words and skip-gram. The intuition behind it is that a word can be predicted by context and context can be predicted from a word. The vector size of a word is the number of neurons in the hidden layer. The task of named entity recognition can be solved by using LSTM neural networks. The features for every word can be word-embeddings (skip-gram or bag of words model), char-embeddings features, and additional features, for example, morphological. To solve this task, we used a tagged dataset (where a human choose which words are entities like a Person, Organization, Location or Product type). We used the softmax function in a neural network for classification. Also, is possible to use other approaches like CRF. There are many neural architectures for the problem of named entity recognition. After that, it is possible to teach our model to predict the entities of predefined types. There are many approaches for text classification, and for vectorization it is possible to use document-embeddings (doc2vec model) or TF-IDF. After this, it is possible to use classification algorithms like an SVM or Random Forest model. To verify the classification task, it is possible to use the most important words in class (for example 20-30 most important words can include the terms which characterize the class).

BUILDING CORPORA OF TRANSCRIBED SPEECH FROM OPEN ACCESS SOURCES

Oleg Iakushkin^a, George. Fedoseev, Anna Shaleva

St Petersburg State University

E-mail: ^aoleg.jakushkin@gmail.com

Currently there are hardly any open access corpora of transcribed speech in Russian that can be effectively used to train those speech recognition systems that are based on deep neural networks—e.g., DeepSpeech. This paper examines the methods to automatically build massive corpora of transcribed speech from open access sources in the internet, such as radio transcripts and subtitles to video clips. Our study is focused on a method to build a speech corpus using the materials extracted from the YouTube video hosting. YouTube provides two types of subtitles: those uploaded by a video’s author and those obtained through automatic recognition by speech recognition algorithms. Both have their specifics: author subtitles may have timing inaccuracies, while automatically recognized subtitles may have recognition errors. We used the YouTube Search API to obtain the links to various Russian-language video clips with subtitles available—words from a Russian dictionary served as an input. We examined two strategies to extract audio recordings with transcripts corresponding to them: by using both types of subtitles or only those that were produced through automatic recognition. The voice activity detector algorithm was employed to automatically separate the segments. Our study resulted in creating transcribed speech corpora in Russian containing 1000 hours of audio recordings. We also assessed the quality of obtained data by using a part of it to train a Russian-language automatic speech recognition system based on DeepSpeech architecture. Upon training, the system was tested on a data set consisting of audio recordings of Russian literature available on voxforge.com—the best WER demonstrated by the system was 18%.

OPTIMIZATION OF NEURAL NETWORKS TRAINING WITH VECTOR-FREE HEURISTIC ON APACHE SPARK

Kamil Khamitov^a, Nina Popova

Lomonosov Moscow State University

E-mail: ^aberserq0123@gmail.com

One of the most computation complicated tasks during neural networks development is a training process. It could be considered as high dimensional numerical optimization tasks. In the modern MapReduce systems, like Apache Spark, it's hard to efficiently implement traditional for Neural networks training gradient-based algorithms or Quasi-newton L-BFGS method, because there are too many non-linear or memory bound operations, could dramatically decrease the performance of your cluster and scalability of your training task. It's known that for L-BFGS methods there are Vector-Free heuristics which allows to reduce task complexity in terms of an amount of Map and Reduce operations, in the application for

large-scale logistic regression tasks. Also it's unclear in which types of neural networks these approaches are applicable. In this research, we applied the heuristics which reduces the amount of the memory-bound and nonlinear operations, Vector-Free heuristic, to the modern numerical optimization algorithms, like L-BFGS, Adam, AdaGrad on Spark cluster. We tested modified versions of algorithms on the different types of NNs architectures: MLP, VGG-16, LSTM, which covers popular neural network types and particular tasks for them. Also to provide efficient and usable environment for computational experiment we developed a software system which could in a semi-automatic way perform a testing of these methods. It allows a researcher to measure the effect of Vector-Free or other heuristics on different platforms and neural networks architectures. Also, this system supports comparing with external data, so researcher is able to compare effectiveness and speedup with other system types like GPU's versions of the methods above. In this research, we only applied this type of heuristic to this algorithms without taking any consideration which thing results in bad performance of modified methods, and don't provide any empirical boundaries for the errors or convergence. All experiments have been performed on the Microsoft Azure cloud platform, with 16 HD12v2 nodes.

PARTICLE IDENTIFICATION IN GROUND-BASED GAMMA-RAY ASTRONOMY USING CONVOLUTIONAL NEURAL NETWORKS

**Evgeny Postnikov^{1,a}, Alexander Kryukov¹, Stanislav Polyakov¹,
Dmitry Shipilov², Dmitriy Zhurov²**

¹*Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University*

²*Irkutsk State University*

E-mail: ^aevgeny.post@gmail.com

Modern detectors of cosmic gamma rays are a special type of imaging telescopes (Cherenkov telescopes) supplied with cameras with relatively large number of photomultiplier-based pixels. For example, the camera of the TAIGA telescope has 560 pixels of hexagonal structure. Images in such cameras can be analyzed by various deep learning techniques to extract numerous physical and geometrical parameters and/or for incoming particle identification. We implement for this purpose the most powerful deep learning technique for image analysis, the so called convolutional neural networks (CNN). In this work we present the results of tests with two open source machine learning libraries, PyTorch and TensorFlow, as possible platforms for particle identification in imaging Cherenkov telescopes. Monte Carlo simulation was performed to analyze images of gamma rays and other (background) particles as well as estimate identification accuracy. Further steps of implementation and improvement of this technique are discussed.

BOTNET IN PYPY TO SPEED UP THE WORK OF THE EARLEY PARSER

**Vladislav Radishevskiy^a, Aleksey Kulnevich, Nikolai Stepenko,
Jean Sene**

National Research Tomsk Polytechnic University

E-mail: ^avladrad95@mail.ru

Extraction of information from texts is a crucial task in the area of Natural Language Processing. It includes such tasks as named-entity recognition, relationship extraction, coreference resolution, etc. These problems are being resolved using two approaches. The first one is the rules-based approach and the second one is machine learning. Solutions based on machine learning are currently very popular but work well only with frequently used entities such as person, company, or organization. They require the presence of a large tagged dataset. With attributes from narrow subject areas and facts, machine learning works much worse. It is better to use this approach in writing rules for context-free grammars. The problem here is that the more grammars there are, the slower the analyzer works. Often the speed of the algorithm is less than 1 KB of text per second. For my project concerned with collecting information and statistics in the subject area of oil and gas geology, I created a system that includes a botnet using the Selenium library, in which one computer generates search queries from a list of objects and collects the resulting links. Then, the resulting links send the tasks to other computers via the REST service based on the asynchronous queue implemented in the Flask framework. To avoid link duplication, hashing was used with Redis. Next, the task is occupied by a botnet consisting of various computers. They take the task and depending on the content of the link carry out the following: if this is an ordinary site, then its contents are parsed; if this is a doc/docx/pdf document, then it is downloaded and then text is extracted from it using the textract library. After saving the text, algorithms are used to extract entities and thematic attributes using the GRL parser on context-free grammars. The first step to accelerate is the use of distributed computing as described above. The second step in the acceleration that was undertaken in this study is the use of the PyPy interpreter for Python, which compiles the code in the C language. It accelerated the work of the algorithm 4 times on average, but the consumption of RAM increased by ~ 25%. The calculations involved 8 computers with 4 threads each. Thus, the use of distributed computations together with the replacement of the standard Python interpreter with PyPy allowed to increase the speed of the extraction of facts increased ~ 128 times.

IMPROVING THE EFFICIENCY OF SMART GRIDS OF ENERGY CONSUMPTION WITH APPLICATION OF SYSTEMS OF ARTIFICIAL INTELLECT

Eugene Shchetinin^{1,a}, Mikchail Berezhkov²

¹Financial University under the Government of the Russian Federation

²Stankin Moscow State Technological University

E-mail: ^ariviera-molto@mail.ru

Clustering is a well-known machine learning algorithm which enables the determination of underlying groups in datasets. In electric power systems it has been traditionally utilized for different purposes like defining consumer individual profiles, tariff designs and improving load forecasting. A new age in power systems structure such as smart grids determined the wide investigations of applications and benefits of clustering methods for smart meter data analysis. This paper presents an improvement of energy consumption forecasting methods by performing cluster analysis. For clustering the centroid based method K-means with K-means++ centroids was used. Various forecasting methods were applied to find the most effective ones with clustering procedure application. Used smart meter data have an hourly measurements of energy consumption time series of Russian central region customers. In our computer modeling investigations we have obtained significant improvement due to carrying out the cluster analysis for consumption forecasting.

USING TENSORFLOW TO SOLVE THE PROBLEMS OF FINANCIAL FORECASTING FOR HIGH-FREQUENCY TRADING

Alexey Stankus

St Petersburg State University

E-mail: alexey@stankus.ru

The use of neural networks significantly expands the possibilities of analyzing financial data and improves the quality indicators of the financial market. In article we examine various aspects of working with neural networks and Frame work TensorFlow, such as choosing the type of neural networks, preparing data and analyzing the results. The work was carried out on the real data of the financial instrument Si-6.16 (futures contract on the US dollar rate).

COMPARISON OF EXPLICIT AND NOT EXPLICIT MATHEMATICAL METHODS OF FINANCIAL FORECASTING

Alexey Stankus

St Petersburg State University

E-mail: alexey@stankus.ru

In most cases, explicit methods are used for the tasks of financial forecasting. The modern possibilities of computer technology already allow the use of neural networks for such problems, but the volumes of their application for forecasting are still not large. This article compares explicit methods, such as the Capital Asset Pricing Model (CAPM) and linear time series, with the results of forecasting obtained as a result of the application of neural networks.

AGENT TECHNOLOGY SITUATIONAL EXPRESS ANALYSIS IN ASSESSMENT OF TECHNOLOGICAL DEVELOPMENT LEVEL OF THE BRICS COUNTRIES

**Evgeny Tretyakov¹, Diana Koshlan^{2,a}, Vladimir Korenkov²,
Boris Onykij¹, Alexey Artamonov¹**

¹*NRNU "MEPhI"*

²*Joint Institute for Nuclear Research*

E-mail: ^adkoshlan@yandex.ru

Stages of development and operation of specialized agent system concerning collection and analysis of the BRICS countries' scientific publications are considered in this paper. The data are extracted from more than 60 sources of authoritative publications in fields of Chemistry, Physics, Genetics, Biochemistry, Ecology, Geology etc. Algorithms for data analysis used in the system directed to reveal scientometric indicators and factographic information. The fact analyzed scientific publications are indexed by a referential database Web of Science indicates credibility level of the material. However, the form of Web of Science providing information imposes its limitations that can be overcome with the help of specialized agents in the inner loop of the system. Aggregation of the material is done in a centralized database. However, there is also a mechanism using prepared SQL queries and a separate function for forming tables of the proper format for data output as MS Excel format that is appropriate for an end user. The work result let to assess the development level of certain technologies and research in the BRICS countries in a short time. And since the system has a certain degree of autonomy a constant monitoring of scientific and technical activities in the BRICS countries is possible. It is concluded that the use of agent technologies for collection and processing of materials in this field significantly accelerates the analysis of scientific and technical publications in comparison with manual mode, and also have a high degree concretization of particular indicators for scientific activity in the analyzed field of publication activity.

COMBINING SATELLITE IMAGERY AND MACHINE LEARNING TO PREDICT ATMOSPHERIC HEAVY METAL CONTAMINATION

**Alexander Uzhinskiy^{1,a}, Gennady Ososkov¹, Pavel Goncharov²,
Marina Frontsyeva¹**

¹Joint Institute for Nuclear Research

²Sukhoi State Technical University of Gomel

E-mail: ^azalexandr@list.ru

We present an approach to predict atmospheric heavy metals contamination by statistical models and machine learning algorithms. The source of the field contamination data is ICP Vegetation Data Management System (DMS). DMS is a cloud platform developed at the Joint Institute of Nuclear Research (JINR) to manage ICP Vegetation data. The aim of the UNECE International Cooperative Program (ICP) Vegetation in the framework of the United Nations Convention on Long-Range Transboundary Air Pollution (CLRTAP) is to identify the main polluted areas of Europe and Asia. Currently there are 6000 sampling sites from 40 regions of different countries presented at the DMS now. The source of the satellite imagery is Google Earth Engine platform (GEE). There are more than 100 satellite programs and modeled datasets at GEE. We are taking data from GEE together with sampling data from DMS to train our deep neural models, but then on the next inference stage we apply the trained neural net to only data from GEE to predict atmospheric contamination by some of heavy metals. Correlation between the satellite imagery data and the heavy metals contamination, considered statistical models and modeling results are presented.

ARCHITECTURE AND BASIC PRINCIPLES OF THE MULTIFUNCTIONAL PLATFORM FOR PLANT DISEASE DETECTION

**Alexander Uzhinskiy^{1,a}, Pavel Goncharov², Gennady Ososkov¹,
Andrey Nechaevskiy¹**

¹Joint Institute for Nuclear Research

²Sukhoi State Technical University of Gomel

E-mail: ^azalexandr@list.ru

The aim of our research is to facilitate the detection and preventing diseases of agricultural plants by combining deep learning and programming services. The idea is to develop multifunctional platform for plant disease detection (PDD) that will use modern organization and deep learning technologies to provide new level of service to farmer's community. Web-platform for PDD consists of a set of interconnected services and tools developed, deployed and hosted in the JINR cloud infrastructure. We are going to use the only free software and

to provide open access to our image database. The platform will include web-interface and a mobile application allowing users to send photos of sick plants with some accompanying text and then obtain a reply explaining a possible cause of the illness. PDD is intended for the data management of various crop data bases needed to train and test corresponding deep neural model. The platform will also provide effective, reliable, secure and convenient tools for storing, transferring and mining of the farmer's text and photo materials. We considered several models to identify the most appropriate type and architecture of deep neural network. The PDD basic principles together with results of comparative study of various deep neural models and their architecture are presented. Up to now we reached promising accuracy result on the level over 90% in the detection of three concrete diseases on the dataset of images of grape leaves.

The reported study was funded by RFBR according to the research project № 18-07-00829 A.

COMPARISON OF DIFFERENT CONVOLUTION NEURAL NETWORK ARCHITECTURES FOR THE SOLUTION OF THE PROBLEM OF EMOTION RECOGNITION BY FACIAL EXPRESSION

Anton Vorontsov^{1,a}, Alexey Averkin²

¹Dubna State University

²Informatics and Control Research Institute RAS

E-mail: ^adealwithbotalfred@gmail.com

In this paper the usage of convolution neural networks considers for solving the problem of emotion recognition by face expression images. Emotion recognition is a complex task and the result of recognition is highly dependent on the choice of the neural network architecture. In this paper various architectures of convolutional neural networks were reviewed and there were selected the most prospective architectures. The training experiments were conducted on selected neural networks. The proposed neural network architectures were trained on the AffectNet dataset, widely used for emotion recognition experiments. A comparison of the proposed neural network architectures was made using the following metrics: accuracy, precision, recall and training speed. At the end of this paper the comparative analysis was made and obtained results were overviewed.

TIME SERIES AND DATA ANALYSIS BASED ON HYBRID MODELS OF DEEP NEURAL NETWORKS AND NEURO-FUZZY NETWORKS

Sergey Yarushev^a, Alexey Averkin^b

Plekhanov Russian University of Economics

E-mail: ^asergey.yarushev@icloud.com, ^baverkin2003@inbox.ru

In this paper we consider approach to data analysis and time series forecasting based on hybrid models. This models contains a Deep NN models and Neuro-Fuzzy networks. We are show an overview of new approaches for data science field - time series and data analysis. Also, we propose our models of DL and Neuro-Fuzzy Networks for this task. Finally we show possibility of using this models for data science tasks.

МНОГОМЕРНЫЙ АНАЛИЗ ДАННЫХ О ПРОДАЖАХ НА ОСНОВЕ ТЕХНОЛОГИИ OLAP

Сауле Сарсимбаева¹, Владимир Димитров^{2,a}

¹*Актюбинский региональный государственный университет
имени К. Жубанова*

²*Софийский университет «Св. Климент Охридски»*

E-mail: ^acht@fmi.uni-sofia.bg

В работе исследованы вопросы многомерного анализа данных на основе технологии семейства Business Intelligence и применение этой технологии для анализа продаж. Изучены OLAP технологии и требования к ним, способы реализации, на примере Business Intelligence. Рассмотрены основные положения технологии бизнес интеллекта в Visual Studio, внутренние интерфейсы Microsoft SQL Server. Разработана система управления базами данных для многомерного анализа данных в сфере продаж розничной сети бытовой и электронной техники, увеличивающая эффективность работы менеджеров, аналитиков компании.

СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ В СИСТЕМЕ СТЕРЕОЗРЕНИЯ МОБИЛЬНОГО РОБОТА

Сергей Ульянов, Кирилл Кошелев^а

Государственный университет «Дубна»

E-mail: ^аkirill_koshelev18@rambler.ru

Распознавание образов – научная дисциплина, целью которой является классификация объектов. Сами объекты называются образами или паттернами. Возможность распознавания опирается на схожесть однотипных объектов. Несмотря на то, что все предметы и ситуации уникальны в строгом смысле, между некоторыми из них всегда можно найти сходства по тому или иному признаку. Отсюда возникает понятие классификации – разбиения всего множества объектов на непересекающиеся подмножества – классы, элементы которых имеют некоторые схожие свойства, отличающие их от элементов других классов. И, таким образом, задачей распознавания является отнесение рассматриваемых объектов или явлений по их описанию к нужным классам. Высокие показатели качества распознавания образов достигаются за счет инвариантного распознавания. Несмотря на изменчивость образов, относящихся к одному и тому же классу, классификация нового образа при инвариантном распознавании может быть осуществлена правильно. Разработка и совершенствование методов компьютерного зрения позволяет расширить круг выполняемых компьютерами задач и сделать машинную переработку информации более интеллектуальной. Задача инвариантного распознавания образов на сегодняшний день остаётся важной нерешённой задачей, относящейся к задачам искусственного интеллекта. Когнитивные способности человека очень тяжело смоделировать на вычислительной технике, не существует единой и эффективной теории, которая бы объясняла, как человек способен с большой точностью распознавать объекты внешнего мира. Представляемая система распознавания образов базируется на технологии стереозрения. Модуль распознавания выделяет объекты и осуществляет слежение за ними. Однако стоит отметить, что при изменении условий окружающей среды (изменение освещения) наблюдается значительное снижение качества распознавания. Изобилие систем машинного зрения не устраняет главные недостатки систем распознавания – погрешность распознавания при изменении ракурса объекта, изменение освещения, чувствительность ПО и т.д. На сегодняшний день лучшие результаты в распознавании образов получают с помощью сверточных нейронных сетей (СНС). При достаточно большом размере СНС имеют небольшое количество настраиваемых параметров, довольно быстро обучаются. Именно поэтому на начальном этапе интеллектуализации системы распознавания было решено использовать сверточную нейронную сеть. В настоящее время довольно подробно описано множество алгоритмов и методик компьютерного зрения и распознавания образов. Данные алгоритмы и методы имеют определенные недостатки, которые перечислены выше. Поэтому разработка универсального и гораздо более эффективного алгоритма распознавания – первостепенная задача исследователя в области компьютерного зрения. На данный момент разрабатывается технология распознавания, базирующаяся на мягких и квантовых вычислениях. С помощью данной технологии представляется возможным повысить эффективность процесса распознавания.

О МЕТОДАХ И ТЕХНОЛОГИЯХ ИНТЕЛЛЕКТУАЛЬНОГО ЭНЕРГОСБЕРЕЖЕНИЯ В КОММЕРЧЕСКИХ ЗДАНИЯХ

Евгений Щетинин^а, Евгения Попова^б

Financial University under the Government of the Russian Federation

E-mail: ^аriviera-molto@mail.ru, ^бevgenia.popova1397@gmail.com

Интеллектуальные технологии энергосбережения и энергоэффективности являются современным масштабным мировым трендом не только в развитии энергетических систем, но и в строительном, девелоперском бизнесе. Спрос на «умные» здания растет не только в мире, но и в России, прежде всего на рынке строительства и эксплуатации, крупных бизнес-центров, торгово-развлекательных центров и др. строительных деловых проектов. Точные оценки экономии важны для продвижения строительных проектов в области энергоэффективности и демонстрации их экономической эффективности. Растущее количество современной измерительной инфраструктуры в коммерческих зданиях привело к повышению доступности данных высокой частоты. Эти данные можно использовать для обнаружения неисправностей и диагностики оборудования, отопления, вентиляции, и оптимизации кондиционирования воздуха. Это также обусловило применение современных и эффективных методов машинного обучения, которые предоставляют перспективные возможности для получения точных прогнозов базового энергопотребления здания, и, таким образом, точные оценки экономии. В настоящей работе для моделирования временных высокочастотных серий энергопотребления был применен алгоритм градиентного бустинга, мощный алгоритм машинного обучения в широком диапазоне применения в анализе больших данных. На его основе предложен метод моделирования дневного профиля энергопотребления и разработан численный алгоритм, его реализующий. Для оценки его эффективности были использованы данные о энергопотреблении 380 коммерческих зданий. Периоды обучения модели были различными, и для оценки эффективности модели использовались несколько показателей точности прогнозирования. Результаты показали, что использование модели градиентного бустинга улучшило точность прогнозирования более чем в 80 процентах случаев по сравнению с моделями промышленных зданий, использующих линейную регрессию и алгоритмом случайного леса.

BIOINFORMATICS

DIRECT SIMULATION OF THE CHARGE TRANSFER ALONG OLIGONUCLEOTIDES AT T=300K

Nadezhda Fialko^a, Victor Lakhno

IMPB RAS - the Branch of KIAM RAS

E-mail: fialka@impb.ru

At present, the attention of researchers is attracted to the possible mechanisms of charge transfer in quasy-1D biomacromolecules, such as DNA, in connection with the potential use of this nano-objects in nanobioelectronics. Biophysical experiments on the hole transfer from guanine G (donor) to guanine triplet GGG (acceptor), separated by adenine-thymine (A-T) bridges of various lengths, demonstrate that the rate of charge transfer between donor and acceptor decreases exponentially with increasing separation only if the guanines are separated by no more than three base pairs; if more bridging base pairs are present, the transfer rates exhibit only a weak distance dependence. We performed direct numerical experiment on the charge transfer from donor to the acceptor along bridge, consisting of homogeneous sites. The model is based on the semi-classical Holstein Hamiltonian. The Holstein polaron model is simple but relevant for explaining charge transfer in DNA. To take into account the temperature, Langevin thermostat is used. For computation we chose some parameter values as for DNA model: the charge donor is guanine G, acceptor - guanine triplet GGG, and the bridge consists of adenines A or thimines T with number of sites N. We modeled different length of bridge N from 1 to 26 sites (length of the whole chain is from 5 to 30 sites). For each N we calculated a set of 100 samples at temperature 300 K, and estimated time-dependencies averaged over the ensemble. The sample is a trajectory of system with its own initial data and pseudorandom time series, which modeling medium thermal fluctuations. Initial data for classical sites are chosen from Maxwell distribution corresponding to T = 300 K, and the charge at moment $t = 0$ is localized on the donor. We calculate the samples until the probabilities of charge distribution on the sites become similar to the thermodynamic equilibrium state, and estimate the time t_{TDE} to reach this state. Results of the simulation demonstrate that: for short chains ($N < 4$ for bridges of adenines and $N < 5$ for thymine bridges) the value of t_{TDE} increases exponentially with the increasing N; for big N, t_{TDE} values are almost the same. In suggestion that the charge transfer rate is a reciprocal of the time t_{TDE} , the results of computer modeling and data of biophysical experiments have a qualitative similarity. We are grateful to the HybriLIT group of JINR for computational resources.

The work is partially supported by the Russian Foundation for Basic Research, projects no. 16-07-00305, 17-07-00801, and Russian Science Foundation, grant 16-11-10163.

DATA CONSOLIDATION AND ANALYSIS SYSTEM FOR BRAIN RESEARCH

**Vladimir Korkhov¹, Alexander Bogdanov¹, Vladislav Volosnikov¹,
Andrey Vorontsov¹, Natalia Zalutskaya², Kirill Gribkov¹,
Nikolay Neznanov², Natalia Ananyeva², Alexander Degtyarev¹**

¹St. Petersburg State University

²National Research Medical Center for Psychiatry and Neurology

E-mail: ^avladimir@csa.ru

Comprehensive human studies, in particular studies in the field of brain pathology, require strong information support for the consolidation of clinical and biological data from different sources in order to allow processing and analysis of data. The heterogeneity of data sources, the variety of presentation formats and the resource-intensive nature of preprocessing make it difficult to conduct comprehensive interdisciplinary research. Combining data for each individual case is a time-consuming process that requires not only time, but profound knowledge in the field of information technology. To solve the problem of sharing heterogeneous sources of clinical and biological species in brain research, an information system with unified access to heterogeneous data is required. Effective implementation of such a system requires creating a model for combining disparate data into a single information environment and adapting preprocessing methods applied individually to each individual data type. The introduction of a model that solves the fundamental problem of consolidating medical and biological data in the form of a cloud service will solve the problem of organizing researchers' access to consolidation results, and equalizing the geographical distribution of research groups and equipment. We analyze the possibilities and methods of consolidation of clinical and biological data, build a model for the consolidation and interaction of heterogeneous data sources for brain research, programmatically implement the model as a cloud service, and provide an interface for supporting queries in a format encapsulating a complex consolidation architecture from the user. We present the design and implementation of an information system for the collection, consolidation and analysis of patient data; we show and discuss the results of the application of cluster analysis methods for the automatic processing of voxel based magnetic resonance imaging data to facilitate the early diagnosis of Alzheimer's disease. Our results show that a detailed study of the properties of cluster analysis data can significantly help neurophysiologists in the study of Alzheimer's disease, especially with the help of automated data processing provided by the proposed information system.

DEVELOPMENT OF SOFTWARE FOR FACE RETRIEVAL SYSTEMS MODELING

Nadezhda Shchegoleva, Varvara Petrova^a

Saint Petersburg Electrotechnical University "LETI"

E-mail: ^apetrova.varvara.0381@mail.ru

The development of software for face retrieval systems modeling is studied. An overview of the state of the problem is provided. Computer modeling is shown to be required to select the most appropriate system structure, set of modules and their parameters. The basic requirements for modern face retrieval systems are determined. It is found that they provided the concept of building a software complex for FaRetSys modeling, which formed the basis for a new Simulink library developed by the authors. Examples of solving practical problems of facial biometrics, structure, composition and parameters of blocks of used systems are shown. Compact models of computer experiments are presented.

ХАОТИЧЕСКАЯ ДИНАМИКА МГНОВЕННОГО СЕРДЕЧНОГО РИТМА И ЕГО ФАЗОВОЕ ПРОСТРАНСТВО.

**Виктор Цветков^{1a}, Алексей Кудинов¹, Александр Иванов²,
Илья Цветков¹, Сергей Михеев¹**

¹*Тверской государственной университет*

²*Тверской кардиологический медицинский центр*

E-mail: ^atsvetkov.vp@tversu.ru

В данном докладе на основе данных суточного холтеровского мониторинга построены фазовые пространства мгновенного сердечного ритма четырех пациентов Тверского областного клинического кардиологического диспансера. Эти пространства наиболее адекватно отражают такие важные свойства кардиоритмов, как хаотичность и самоподобие (фрактальность). Приведены методы вычисления фрактальной размерности D и D -мерного объема фазового пространства мгновенного сердечного ритма в наиболее удобном для практического применения виде. На основе созданного и реализованного авторами комплекса программ проведено вычисление таких параметров состояния мгновенного сердечного ритма, как значение фрактальной размерности D , фрактального фазового объема Γ , определяемого по покрытию фазовой траектории сеткой единичного размера. Показана близость фазовых пространств мгновенного сердечного ритма исследуемых пациентов к фракталам с точностью $4.53 \cdot 10^{-2}$ в C -метрике. При временах холтеровского мониторинга превышающих 6 часов эти параметры стремятся к постоянным значениям и могут быть использованы как маркеры состояния сердечно-сосудистой системы в кардиодиагностике.

ВИЗУАЛИЗАЦИЯ КВАНТОВОГО ФАЗОВОГО ПРОСТРАНСТВА МГНОВЕННОГО СЕРДЕЧНОГО РИТМА

Илья Цветков, Виктор Цветков^а, Сергей Михеев

Тверской государственной университет

E-mail: ^аtsvetkov.vp@tversu.ru

Данные суточного холтеровского мониторирования (ХМ) по кардиоинтервалам для анализа представляются в форме, сочетающей простоту и информативность. Показано, что это можно сделать, используя визуализацию массива данных по кардиоритмам на основе квантового фазового пространства мгновенного сердечного ритма. Под визуализацией квантового фазового пространства мгновенного сердечного ритма мы будем понимать способ представления цифровой информации о мгновенном сердечном ритме в виде, удобном для наблюдения и анализа. Сформулированное в докладе квантование фазового пространства мгновенного сердечного ритма приводит его к делению на ячейки конечной величины h и объема $\Delta\Gamma=h^2$. Информация о структуре фазового пространстве мгновенного сердечного ритма при этом будет определяться числами заполнения этих ячеек. Важнейшей задачей нашего подхода является визуализация точек квантового фазового пространства мгновенного сердечного ритма. Для этого его точкам приписали определённые значения цвета в зависимости от значений чисел заполнения состояний.

POSTERS

NUMERICAL SOLUTION OF DIFFRACTION PROBLEM ON THE JOINT OF TWO OPEN THREE- LAYERED WAVEGUIDES

Veniamin Chupritskiy^{1a}, Dmitriy Divakov²

¹*National Research Tomsk Polytechnic University*

²*RUDN University*

E-mail: ^avkchupr@gmail.com

This paper describes the algorithm for the numerical solution of the diffraction problem of waveguide modes at the joint point of two open planar waveguides. For the planar structures under consideration, we can formulate a scalar diffraction problem, which is a boundary value problem for the Helmholtz equation with a variable coefficient in two-dimensional space. The problem on the eigenmodes of an open three-layered waveguide is the Sturm-Liouville problem for a second-order operator with piecewise constant potential on the axis, where the potential is proportional to the refractive index. The described problem is singular and has a mixed spectrum: the discrete part of the spectrum corresponds to the guided waveguide modes, the continuous part of the spectrum to the radiative modes. The presence of a continuous part of the spectrum complicates the numerical solution of the diffraction problem, since the eigenfunctions from the region of the continuous spectrum do not integrate on the axis, and therefore Galerkin's method can not be used in this definition. One of the ways to adapt the Galerkin method for the problem solution is to limit artificially the area, which is equivalent to placing the open waveguide in question in a hollow closed waveguide whose boundaries are distanced from the real boundaries of the waveguide layer of the open waveguide. As a result of the described approach, we obtain a diffraction problem on a finite interval and with a discrete spectrum, which can be solved by the projection method. The described method is realized in the Maple computer algebra system using CUDA(R) technology to accelerate certain routines.

AN IMAGE VERIFICATION FRAMEWORK DEVELOPMENT

Ahmed Elaraby¹, Andrey Nechaevskiy^{2a}

¹*South Valley University*

²*Joint Institute for Nuclear Research*

E-mail: ^anechav@jinr.ru

An efficient representation and implementation of image are necessary, as a digital image is an approximation of some real situation, and carries some uncertainty. In order to deal with this uncertainty we need appropriate image model, which also enable image processing without losing the information regarding the uncertainty. Interval arithmetic techniques appear as a good option for handling the uncertainty. In this work we will discuss the extended of the classical notion of digital image, in the which each pixel has as degree of intensity an

exact value to the interval digital image one, where each pixel possesses an interval intensity that include lower and upper bound of every element of the image. The time consuming process of image data processing can be address using parallel computing techniques that provide an efficient and convenient way to address this issue. The paper concludes that considering the interval arithmetic in designing solutions for some applications may impact the performance of algorithms and the image processing tasks may benefit from an efficient image verification model.

EFFICIENCY MEASUREMENT SYSTEM FOR THE COMPUTING CLUSTER AT IHEP

Victoria Ezhova, Viktor Kotliar^a

NRC "Kurchatov Institute" IHEP

E-mail: ^avictor.kotlyar@ihep.ru

Every day IHEP central computing cluster produce thousands of calculations related to research activities, both IHEP and GRID experiments. A lot of machine resources are expended on this work. So, we can estimate the size of the spent resources used for all types of tasks, make decisions for changing cluster configuration and to do the forecast for the work of the computer center in general. In this work you can see the calculations of the efficiency index and the graphical representation of work of a cluster on the basis of account information. It is one of the main tasks within work on creation of system of uniform monitoring of computer center of IHEP.

COMBINED EXPLICIT-IMPLICIT TAYLOR SERIES METHODS

**Ivan Hristov^{1a}, Zafar Tukhliev², Stefka Dimova¹,
Radoslava Hristova¹, Igor Puzynin², Taisiya Puzynina², Nikolay
Shegunov¹, Zarif Sharipov²**

¹*Sofia University "St. Kliment Ohridski"*

²*Joint Institute for Nuclear Research,*

E-mail: ^aivanh@fmi.uni-sofia.bg

Hamiltonian systems arise in natural sciences and are used as mathematical models for many practical problems. Due to their wide applications, a large class of numerical methods, usually symplectic ones, has been developed. The most commonly used is Verlet method which is second order, fast, and simple to implement. Here we consider a novel idea proposed and developed in [1]. It is based on Taylor series expansion and produces a large class of methods of various orders of accuracy. The idea of the method is to combine Taylor expansions about the forward and current time levels. Construction of such methods is simple and moreover, they inherit the desired for Hamiltonian systems properties of symmetry and energy conservation. When high order of accuracy is needed, these new methods have lower

computational cost than Verlet method. In some problems of Computational Dynamics, at a given stage of the process, incorporation of a large set of initial conditions or a large set of parameters is required. This motivates us to parallelize the numerical algorithms. Here we consider instruction level parallelism, namely, vectorized instructions combined with OpenMP threads. A comparison between the classical Verlet method and the Combined Taylor Series Methods on some Hamiltonian systems has been made, with main focus on the time-accuracy diagrams. The results illustrate the strengths and the weaknesses of the two different approaches. The work was financially supported by RFBR grant No. 17-01-00661-a and by a grant of the Plenipotentiary Representative of the Republic of Bulgaria at the JINR. [1] Akishin, P. G., Puzynin, I. V., Vinitisky, S. I. (1997). A hybrid numerical method for analysis of dynamics of the classical Hamiltonian systems. *Computers & Mathematics with Applications*, 34(2-4), 45-73

DATA GATHERING AND WRANGLING FOR THE MONITORING OF THE RUSSIAN LABOUR MARKET

Javad Javadzade

Joint Institute for Nuclear Research

E-mail: jjavadzade@inbox.ru

This project is devoted to monitoring and analyzing the labour market based on the publicly available data on job offers, CVs and companies gathered from open data projects and recruitment agencies. The relevance of project is that some work areas have already overcrowded, some are outdated or some may have a little need for new workers, and some new and growing industries will offer good jobs. The result obtained at the end will allow one to have a look on the labor market on different levels starting from the local one. This information is useful not only for school graduates, students and people who is just looking for a better job for themselves, but also for the employers. It is also can be useful for universities to estimate the relevance of the educational programs they offer. One of the key tasks is the collection of job offers data from open sources and recruitment agencies. Before writing parsing-scripts, need to analyze existing open sources of vacancies and identify the final list from which the vacancy data will be downloaded. No less important task is data pre-processing, where the main task is to remove duplicate job offers appear from different sources. Because sophisticated comparison of more than a million vacancies requires significant time, this step was realized using Apache Spark on a cluster. Also, this step involves using of machine learning algorithms. For the job offers, the vector representation is constructed using gensim word2vec, then the closest ones are selected. For the moment, more than a million of vacancies from Headhunter, Superjob, Trudvsem recruitment agencies have been already collected and processed.

APPLICATION OF HUBZERO PLATFORM FOR THE EDUCATIONAL PROCESS IN ASTROPARTICLE PHYSICS

**Yuliya Kazarina¹, Oleg Fedorov¹, Andrey Mikhailov²,
Igor Bychkov², Andreas Haungs³, Alexander Kryukov⁴,
Dmitriy Kostunin³, Alexey Shigarov², Dmitriy Shipilov¹,
Daria Chernykh¹**

¹*Applied Physics Institute Irkutsk State University*

²*Matrosov Institute for System Dynamics and Control Theory SB RAS*

³*Karlsruhe Institute of Technology*

⁴*Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University*

E-mail: ^alutien777@mail.ru

In the frame of the Karlsruhe-Russian Astroparticle Data Life Cycle Initiative it was proposed to deploy an educational resource `astroparticle.online` for the training of students and graduate students in the field of astroparticle physics. This resource is based on HUBzero, which is an open-source software platform for building powerful websites, which supports scientific discovery, learning, and collaboration. HUBzero have been deployed on the servers of Matrosov Institute for System Dynamics and Control Theory. The educational resource `astroparticle.online` is being filled with the information covering cosmic messengers, astroparticle physics experiments and educational courses and schools on astroparticle physics. Furthermore, the educational resource `astroparticle.online` can be used for online collaboration. We present the current status of this project and our first experience of application of this service as a collaboration frame-work. This work was financially supported by Russian Science Foundation and Helmholtz Society, Grant No. 18-41-06003. The developed educational resources were freely deployed on the cloud infrastructure of the Shared Equipment Center of Integrated Information and Computing Network for Irkutsk Research and Educational Complex (<http://net.icc.ru>).

EVENT-DRIVEN AUTOMATION AND CHAT-OPS ON IHEP COMPUTING CLUSTER

Anna Kotliar, Viktor Kotliar^a

NRC "Kurchatov Institute" IHEP

E-mail: ^avictor.kotlyar@ihep.ru

Dealing with cluster-systems you have multiple ordinary situations which can be solved using automation tools. Stackstorm is a quite good event-driven system which helps to manage typical problems and to communicate with cluster via chat-ops extension. Just write a rule for such event and it will be triggered and solved. In the presented work will be shown an example of a real event-driven system on IHEP computing cluster which use Nagios, CheckMK, Stackstorm, Mattermost for routine work automation as a part of multicomponent cluster management system.

USING BINARY FILE FORMAT DESCRIPTION LANGUAGES FOR VERIFYING RAW DATA IN ASTROPARTICLE PHYSICS EXPERIMENTS

**Alexander Kryukov¹, Elena Korosteleva^{1a}, Dmitriy Kostunin²,
Igor Bychkov³, Alexey Khmelnov³, Andrey Mikhailov³,
Oleg Fedorov⁴, Alexey Shigarov¹**

¹*Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University,*

²*Karlsruhe Institute of Technology*

³*Matrosov Institute for System Dynamics and Control Theory SB RAS*

⁴*Applied Physics Institute Irkutsk State University*

E-mail: ^aelkrs@yandex.ru

The exponential growth of the amount of astroparticle data is expected to happen in near future. This trend gives rise to a number of emerging issues of big data management. One of the important issues is how to describe and verify raw binary data to support their availability and reuse in future. The present work demonstrates a possible solution for this issue in application to data of TAIGA observatory, which consists of a number of facilities featuring five different formats of raw data. The long-term preservation of raw binary data as originally generated is essential for rerunning analyses and reproducing research results in future. In this case, the raw data should be well documented and accompanied by the parsing and verifying tools. There are some declarative languages for describing binary file format description (e.g. DFDL, FLEXT, KAITAI STRUCT). The present work shows the progress of the application of KAITAI STRUCT to specify, parse and verify raw data of the TAIGA binary file formats. The format specifications implemented using this framework allow us to generate program code for parsing and verifying the raw binary data in the target languages (C++, Java, Python, etc.). The libraries were tested on real data, have shown good performance and indicated the parts with corrupted data. This study can be interested in other experiments which raw binary data formats remain weakly documented. This work was financially supported by the Russian Scientific Foundation (grant 18-41-06003).

PSEUDO-RANDOM NUMBER GENERATOR BASED ON NEURAL NETWORK

Dmitry Kulyabov^{1,2}, Anna Korolkova¹, Migran Gevorkyan^{1a},
Anastasiya Demidova¹

¹*RUDN University,*

²*Joint Institute for Nuclear Research*

E-mail: ^amngevorkyan@sci.pfu.edu.ru

Pseudorandom uniform distributed number generators are used in many fields of science and technology [1]. It is very important to test the quality of pseudo-random sequence produced by an algorithm. An overview of a large number of criteria for testing the quality of the sequence produced by pseudo-random generators can be found in the third chapter of [2], as well as in the article [3]. One of the most robust software packages that implements such tests is the DieHarder utility [4]. Among the algorithms that show good results when passing the entire set of DieHarder tests, the following three groups of algorithms should be mentioned.

- The Mersenne Twister (MT — Mersenne Twister) [5] gives a very qualitative sequence, but is relatively complex. It is currently used as default for many modern programming languages.
- Algorithms xorshift, xorshift+ and xorshift* [6] pass all sorts of tests from the DieHarder package on a level with Mersenne Twister, but algorithmically they are more simple than MT, although they have a slightly shorter period.
- The family of KISS algorithms [5] (Keep It Simple Stupid), whose name indicates the extreme simplicity, are almost as good as Mersenne Twister and even simpler than xorshift. In this paper, we test our pseudo-random number generator based on the neural network, comparing it with the above algorithms. This comparison imposes strict requirements for the efficiency of the neural network, as each of these three algorithms has the potential for parallelization, requires for initialization from one to three initial seeds, uses only arithmetic and bitwise operations, and does not fail any DieHarder test, showing weak results in only 4 of them (of more than 100).

References

- [1] M.N. Gevorkyan, M. Hnatich, I.M. Gostev, A.V. Demidova, A.V. Korolkova, D.S. Kulyabov, L.A. Sevastianov, The Stochastic Processes Generation in OpenModelica, in: V.M. Vishnevskiy, K.E. Samouylov, D.V. Kozyrev (Eds.), DCCN 2016, Moscow, Russia, November 21-25, 2016, Revised Selected Papers, Springer International Publishing, 2016: pp. 538–552.
- [2] Knuth Donald E. The Art of Computer Programming, Volume 2 (3rd Ed.): Seminumerical Algorithms. — Boston, MA, USA : Addison-Wesley Longman Publishing Co., Inc., 1997.— Vol. 2.—ISBN: 0-201-89684-2
- [3] L'Ecuyer Pierre, Simard Richard. TestU01: A C library for empirical testing of random number generators // ACM Transactions on Mathematical Software (TOMS). — 2007. — Vol. 33, no. 4. — P. 22
- [4] Brown Robert G., Eddelbuettel Dirk, Bauer David. Dieharder: A Random Number Test Suite. — 2013. — Access mode: http://www.phy.duke.edu/~rgb/General/rand_rate.php .
- [5] Rose Greg. KISS: A Bit Too Simple. — 2011. — Access mode: <https://eprint.iacr.org/2011/007.pdf>
- [6] Marsaglia George. Xorshift RNGs // Journal of Statistical Software. — 2003. — Vol. 8, no. 1. — P. 1–6.

**METHODS & TOOLS OF THE RSC BASIS
DISTRIBUTED MICRO AGENT PLATFORM
FOR MANAGING COMPUTE, NETWORK
AND STORAGE RESOURCES TO EFFICIENTLY
PROCESS DATA**

Mikhail Malkov

RSC Group

E-mail: m.malkov@rsc-tech.ru

The poster covers methods and tools, developed within the RSC BasIS datacenter management platform: key benefits of a micro agent approach in solving the problem of the lack of connectivity between different layers in datacenter management and automation; architecture overview and use case examples. We would like to share our operating experience of RSC BasIS Platform and discuss plans for its further development.

**MODERN HYPER-CONVERGED PLATFORM
FOR COMPUTATIONAL- AND I/O-
HEAVY ENVIRONMENTS**

Mikhail Malkov

RSC Group

E-mail: m.malkov@rsc-tech.ru

This poster describes the benefits of using a hyperconverged approach to build next-gen high performance computer clusters. We will cover the need for storage and compute convergency and its advantages over existing architectures; present our results, use case scenarios and the achieved efficiency of the system, which enabled the new JINR supercomputer built with the RSC hyper-converged technology to rank 9th in the io500 list.

MODERNIZATION OF WEB SERVICE FOR THE DATA CENTER SIMULATION PROGRAM

Dmitry Marov^{1a}, Daria Priakhina^{2b}

¹Dubna State University

²Joint Institute for Nuclear Research

E-mail: ^admitriy.marov97@yandex.ru, ^bpry-darya@yandex.ru

The data storage and processing systems simulation program "SyMSim" was developed at the Laboratory of Information Technologies of the Joint Institute for Nuclear Research. The input parameters and the simulation results are stored in a database. A web service was developed to interact with the program, but it has some disadvantages. - there are no examples of program operation and guest mode; - user's personal cabinet isn't finished; - there is no visual output of the results; - there is no form for creating a model; - design of the web service isn't finished; The web service user interface modernization results are presented: - user manual, examples of using the program and guest mode are developed; - user's personal account is modified and has user-friendly interface; - the process of creating the simulated infrastructure and setting the equipment parameters is modified; - results output has been visualized.

PARALLEL CALCULATIONS OF GROUND STATES OF 6,7,9,11LI NUCLEI BY FEYNMAN'S CONTINUAL INTEGRALS METHOD

Mikhail Naumenko^a, Viacheslav Samarin

Joint Institute for Nuclear Research

E-mail: ^aanaumenko@jinr.ru

The structure of lithium isotopes and nuclear reactions with their participation are extensively studied both experimentally and theoretically. In this work, the wave functions of the ground states of few-body nuclei 6,7,9,11Li are calculated by Feynman's continual integrals method in Euclidean time. The algorithm of parallel calculations was implemented in C++ programming language using NVIDIA CUDA technology. Calculations were performed on the NVIDIA Tesla K40 accelerator installed within the heterogeneous cluster of the Laboratory of Information Technologies, Joint Institute for Nuclear Research, Dubna. The studied isotopes are considered as cluster nuclei with the following configurations: 6Li ($\alpha + n + p$), 7Li ($\alpha + n + n + p$), 9Li ($7\text{Li} + n + n$), and 11Li ($9\text{Li} + n + n$). The results of calculations for the studied nuclei are in good agreement with the experimental energies of separation into clusters and nucleons. The obtained probability densities may be used for the correct definition of the initial conditions in the time-dependent calculations of reactions with the considered nuclei. This work was supported by the Russian Science Foundation (RSF), research project 17-12-01170.

ANALYSIS OF THE FEATURES OF THE OPTIMAL LOGICAL STRUCTURE OF DISTRIBUTED DATABASES

Elena Nurmatova^{1a}, Victor Gusev², Viktor Kotliar²

¹*Dubna State University,*

²*NRC "Kurchatov Institute" IHEP*

E-mail: ^aev8@mail.ru

The questions of constructing optimal logical structure of a distributed database (DDB) are considered. Solving these issues will make it possible to increase the speed of processing requests in DDB in comparison with a traditional database. Optimal logical structure of DDB will ensure the efficiency of the information system on computational resources. The problem of constructing an optimal logical structure of DDB is reduced to the problem of quadratic integer programming. As a result of its solution, the local network of the DDB is decomposed into a number of clusters that have minimal information connectivity with each other. In particular, such tasks arise for the organization of systems for processing huge amounts of information from the Large Hadron Collider. In these systems various DDBs are used to store information about: the system of triggers of data collection from physical experimental installations (ATLAS, CMS, LHCb, Alice), the geometry and the operating conditions of the detector while collecting experimental data.

USING MULTIVARIATE QUANTILE FUNCTION FOR SOLVING BIOINFORMATICS PROBLEMS

Sergey Poluyan^{1a}, Nikolay Ershov²

¹*Dubna State University*

²*Moscow State University*

E-mail: ^asvpoluyan@gmail.com

In this work, we study evolutionary optimization algorithms for solving the problems in structural bioinformatics: prediction of three-dimensional peptide structure from amino acid sequence and peptide-protein docking. We provide a way of using evolutionary optimization algorithms based on using quantile functions. The used schemes for building and using of the quantile functions were described. The GPU-accelerated implementation of the presented schemes was carried out. We present the results of various numerical experiments.

A WAY OF ANOMALY DETECTION IN ENGINEERING EQUIPMENT CHARACTERISTICS OF SYMMETRA AT IHEP IT CENTER

Ekaterina Popova, Viktor Kotliar^a

NRC "Kurchatov Institute" IHEP

E-mail: ^avictor.kotlyar@ihep.ru

The information flow should be monitored on anomaly detection. It is important, because it allows you to see a possible problem in advance and prevent it from turning into a real one. A huge flow of diverse data within the modern computing center flows from everywhere. As a rule, these are time series - numerical characteristics that are consistently measured after some time intervals. At this work there was developed the way of analysis for engineering equipment characteristics in centralized system of uninterrupted power supply (Symmetra) at IHEP IT center. When tracking time series, extracted from the data processing and storage system, anomalies are detected using the Twitter AnomalyDetection package. The information on problem is provided to the engineering and operational staff.

SENSITIVITY ANALYSIS IN A PROBLEM OF REAXFF MOLECULAR-DYNAMIC FORCE FIELD OPTIMIZATION

Konstantin Shefov, Stepanova Margarita^a

St Petersburg State University

E-mail: ^amstep@mms.nw.ru

In a wide range of modern problems, it is required to estimate an influence of uncertainty of input parameters on uncertainty of an output value of a modeling function. In this contribution, we present algorithms for analyzing the sensitivity of a target function with respect to parameters in the problem of optimization of ReaxFF molecular-dynamic force field. In this particular case it allows one to effectively decrease the number of simultaneously optimized parameters. We compare the Sobol's global sensitivity indexes (SI) approach and the correlation analysis. Both methods are based on computations of the target function value on the set of pseudo- or quasi-randomly distributed points. The distribution derived is used for further computations of SI using Monte-Carlo technique and correlation coefficients. In the case of optimized ReaxFF force field one may spend up to several seconds to compute a value of the target function in a particular point. That is why it is important to perform calculations in parallel for multiple points. A parallel algorithm has been implemented in C++ using MPI. We compute Sobol's SI and coefficients of correlation of parameters variation and target function values variation while we optimize the force field for molecules and crystals of zinc hydroxide. We show that using of parameter set sorted by influence allows one to significantly increase convergence speed of the optimization algorithm and even completely exclude those parameters with relatively small influence.

OPTIMISATION OF TENSORFLOW APPLICATIONS ON THE WORKSTATION INTEL® XEON® PLATINUM

Svetlana Shikota^{1a}, Lev Shchur², Alexander Russkov¹

¹*Science Center in Chernogolovka,*

²*Higher School of Economics*

E-mail: ^asvetlana.shikota@gmail.com

Платформа TensorFlow является одним из наиболее развитых наборов программных продуктов с открытым кодом для задач машинного обучения. С другой стороны, рабочие станции на базе процессоров Intel® Xeon® Platinum представляются перспективным аппаратным решением для задач машинного обучения. Их отличительная черта состоит в комбинации из трех важных элементов. Во-первых, это большое число тяжелых ядер в одном CPU, более двух десятков. В нашем случае, это 26 ядер в каждом из двух процессоров 8164, плюс hyper-threading. Во-вторых, это наличие двух устройств AVX-512 (Advanced Vector Extension 512), которые дают возможность работы с 512-битными регистрами. Теоретически это позволяет ускорить вычисления на 32-битными числами в 16 раз. В третьих, это очень большой размер памяти, на одной материнской плате 1.5 TB высокоскоростной памяти DDR4, которая поддерживается большим кэшем второго уровня. Такое устройство одновременно предоставляет большую скорость вычислений и работу с данными большого объема в оперативной памяти. В перспективе, это позволяет проводить анализ сложных проблем с большим объемом данных. Мы обсуждаем результаты тестирования некоторых приложений с использованием платформы TensorFlow фирмы Google и библиотеки Intel® Math Kernel Libraries (Intel® MKL).

Work is supported by the grant of Russian Foundation for Basic Research 17-07-01377.

SCALABILITY OF THE PARALLEL STRONGIN ALGORITHM IN THE PROBLEM OF OPTIMIZING A MOLECULAR-DYNAMIC FORCE FIELD

Margarita Stepanova, Konstantin Shefov^a

St Petersburg State University

E-mail: ^ak.s.shefov@gmail.com

Strongin's multifactorial global search algorithm (MGSA) allows one to find an absolute minimum of a function of multiple variables on a mesh. In this contribution a parallel program is presented that implements the algorithm above applied to ReaxFF MD force field parameter search. The MGSA converges the faster, the greater number of mesh points have computed target function value in them. In case of ReaxFF optimization computation time of a target function value significantly exceeds time of data exchange between parallel processes. One

is able to speed up computation by obtaining not only one but several function values in various points simultaneously. Our software implements two levels of parallelism. To deal with function of multiple variables, one uses a scan for mapping a multidimensional domain of definition of a function into a one-dimensional segment. To decrease the effect of losing information of multidimensional points proximity, multiple scans are used, the number of which is denoted by N . The algorithm deals with each scan in parallel, computing function value in N different mesh points in a single iteration. This is the first level of parallelism. To define a mesh point of a next iteration, MGSA finds a subinterval with the most probable location of the minimum and computes a target function value in a certain point of this subinterval. If one computes a function value not only in the most probable subinterval but also in $(M-1)$ subintervals with less probability in parallel, one will be able to obtain function values in M different mesh points. This will accelerate the convergence, increasing the amount of data about the target function received every iteration. This is the second level of parallelism. Thus the two levels allow one to compute $M * N$ function values every iteration. In this contribution we research scalability of our MGSA implementation, namely, the dependence of the number of algorithm iterations it needs to converge on the number of CPU cores used, separately for each level of parallelism.

CONVOLUTIONAL NEURAL NETWORKS FOR SELF-DRIVING CARS ON GPU

Boris Tyulkin, Nataliia Kulabukhova^a

St Petersburg State University

E-mail: ^akulabukhova.nv@gmail.com

The challenge is to teach how to drive a vehicle without human with the help of deep learning power using visual data from the cameras installed on the machine. The problem is to process the amount of data in the real time. Convolutional neural networks (CNNs) are used for training data. And the idea of how to use CNNs on graphical processing units is described.

POSSIBLE APPLICATION AREAS OF MACHINE LEARNING TECHNIQUES AT MPD/NICA EXPERIMENT AND THEIR IMPLEMENTATION PROSPECTS IN DISTRIBUTED COMPUTING ENVIRONMENT

Dmitry Zinchenko^a, Alexander Zinchenko, Eduard Nikonov

Joint Institute for Nuclear Research

E-mail: ^azinchenk1994@gmail.com

At present, the accelerator complex NICA is being built at JINR (Dubna). It is intended for performing experiments to study interactions of relativistic nuclei and polarized particles (protons and deuterons). One of the experimental facilities MPD (MultiPurpose Detector) was designed to investigate nucleus-nucleus, proton-nucleus and proton-proton interactions. Preparation of the physics research program requires production of a large amount of simulated data, including high-multiplicity events of heavy-ion interactions with high energy. Realistic modelling of the detector response for such events can be significantly accelerated by making use of the generative models. Selection of rare physics processes traditionally utilizes machine learning based approaches. During the high luminosity accelerator operation for the proton-proton interaction research program it will be necessary to develop high-level trigger algorithms, based, among others, on machine learning methods. As the data taking proceeds, the tasks of the fast and efficient processing of experimental data and their storage in large volumes will become more and more important, requiring involvement of distributed computing resources. In this work these problems are considered in connection to the MPD/NICA experimental program preparation.

ПРИМЕНЕНИЕ ЭВОЛЮЦИОННЫХ И РОЕВЫХ АЛГОРИТМОВ ОПТИМИЗАЦИИ ДЛЯ РЕШЕНИЯ МОДЕЛЬНОЙ ЗАДАЧИ ПРЕДСКАЗАНИЯ СТРУКТУРЫ БЕЛКА

Максим Быстров^{1a}, Николай Ершов²

¹*Государственный университет «Дубна»*

²*Московский государственный университет*

E-mail: ^afar334_oxid@mail.ru

Работа посвящена проблемам прогнозирования пространственной структуры белковых молекул, полипептидов и их комплексов. Предлагаемый нами метод основан на решении задачи оптимизации, в которой целевой функцией является потенциальная энергия молекулы, а параметрами оптимизации – длины связей между атомами и углы вращения. Главными особенностями таких задач является большая размерность и высокая вычислительная сложность. Проведенные предварительные исследования

показали, что множество существующих алгоритмов оптимизации решают такую задачу неудовлетворительно. Типичное время вычисления задачи занимает от нескольких часов до нескольких дней. Кроме того, для корректного вычисления целевой функции требуется серьезное программное обеспечение. Эти факторы существенно усложняют процесс разработки эффективных алгоритмов для решения задачи предсказания структуры белков и их комплексов. В настоящей работе предлагается упрощенная, модельная задача укладки графа на плоскости, которая позволяет проводить расчеты быстрее и без использования специального программного обеспечения, разрабатывать новые и улучшать существующие алгоритмы оптимизации. В работе приводятся результаты численного исследования ряда алгоритмов роевой и эволюционной оптимизации при решении поставленной модельной задачи.

АНАЛИЗ ПАРАЛЛЕЛЬНОЙ СТРУКТУРЫ ПОПУЛЯЦИОННЫХ АЛГОРИТМОВ ОПТИМИЗАЦИИ

Николай Ершов¹, Сергей Полуян^{2а}

¹Московский государственный университет

²Государственный университет «Дубна»

E-mail: ^аsvpoluyan@gmail.com

Работа посвящена вопросам крупно-блочной параллельной реализации методов эволюционной и роевой оптимизации на примере решения задачи минимизации функций действительного аргумента. Рассматривается субпопуляционная схема распараллеливания. Предлагается классификация паттернов параллельного взаимодействия субпопуляций, выполненная на основе анализа ряда методов рассматриваемого класса. Описывается программная реализация в форме библиотеки шаблонных функций, реализующих данные паттерны, предлагающая пользователю высокоуровневое средство описания популяционных алгоритмов оптимизации. Рассматриваются вопросы параметризации реализуемых алгоритмов с целью исследования эффективности их распараллеливания.

ПРИМЕНЕНИЕ СЕТИ ХОПФИЛДА ДЛЯ АВТОМАТИЗИРОВАННОЙ ПОДБОРКИ КПЭ

Денис Кравченко^{1,а}, Эдуард Никонов²

¹*ООО Business Ability*

²*Объединенный институт ядерных исследований*

E-mail: ^аkradenis@gmail.com

Для решения задач распознавания образов применяют методы машинного обучения, одним из которых являются искусственные нейронные сети (ИНС). Их реализация была рассмотрена у природы и похожа на сети нервных клеток живого организма. Работа ИНС повторяет некоторые функции головного мозга и делится на два этапа: обучение нейронной сети, которое позволяет ей выстраивать собственные правила с помощью весовых коэффициентов и распознавание, построенное на основе собранных данных (опыта). Одной из сетей является автоассоциативная рекуррентная сеть Хопфилда, с помощью которой была произведена автоматизация подбора ключевых показателей эффективности для руководителей предприятий.

КОГНИТИВНО-ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА АДАПТАЦИИ И ОБУЧЕНИЯ ДЕТЕЙ-АУТИСТОВ

Алла Мамаева^а, Андрей Шевченко

Государственный университет «Дубна»

E-mail: ^аallamamaeva.d@gmail.com

Когнитивно-интеллектуальная система адаптации и обучения детей-аутистов предназначена для извлечения, обработки и формирования программы обучения на основе когнитивных процессов, в частности ЭЭГ сигналов, адаптации детей-аутистов к социуму и обучения базовым бытовым навыкам. Рассмотрена и проанализирована возможность программной реализации определения уровня эмоционального возбуждения. Проведенная работа демонстрирует оптимальную обучаемость системы, возможность создания БЗ на основе регистрируемого сигнала ЭЭГ и использования полученных результатов для распознавания эмоций. Использование интеллектуальной надстройки в виде ОБЗ, основанной на нечеткой логике, в распознавании эмоций является наиболее оптимальным решением по сравнению с использованием нейронных сетей, которые представляют из себя исключительно абстрактный математический аппарат.

КОГНИТИВНО ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ДИАГНОСТИКИ, АДАПТАЦИИ И ОБУЧЕНИЯ ДЕТЕЙ-АУТИСТОВ. МОДУЛЬ ОБРАБОТКИ ДАННЫХ

Андрей Шевченко^а, Алла Мамаева

Государственный университет «Дубна»

E-mail: ^аs13m@yandex.ru

Когнитивно-интеллектуальная система адаптации и обучения детей-аутистов предназначена для извлечения, обработки и формирования программы обучения на основе когнитивных процессов, в частности ЭЭГ сигналов, адаптации детей-аутистов к социуму и обучения базовым бытовым навыкам. Основной частью КИСДАОДА является модуль обработки данных. Модуль обработки данных представляет собой структуру взаимодействия ребёнка и программы обучения посредством применения нечеткой логики. Модуль обработки данных предназначен для извлечения ЭЭГ посредством когнитивного шлема, обработки и фильтрации полученного сигнала, формирования программы обучения на основе когнитивных процессов, диагностики проблем работы ребёнка с системой и оценки реакции оператора на задания, сформированные модулем обучения. Модуль обработки данных состоит из: 1. Обработчика ЭЭГ, который выводит необработанный сигнал ЭЭГ с датчиков когнитивного шлема, фильтрует и обрабатывает. 2. Модуля распознавания эмоций. 3. Шины, формирующей пакет данных для передачи их на вход оптимизатора баз знаний. 4. Интерпретатора полученного коэффициента от оптимизатора баз знаний. В результате работы модуля обработки данных в блок обучения передаётся сформированный параметр оценки задания. Таким образом происходит корректировка программы обучения и индивидуализация системы под конкретного ребёнка.