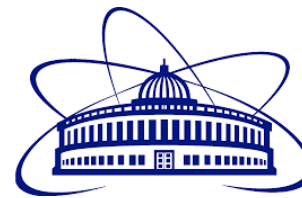


NEC'2019



XXVII International Symposium on Nuclear Electronics & Computing

Montenegro, Budva, Becici, 30 September - 4 October 2019



Heterogeneous IT platform “HybriLIT” for organizing the educational process on the basis of the International IT School “Data Science”



S. Belov^{1,3}, V. Korenkov^{1,2,3}, M. Matveev^{1,2,3}, D. Podgainy¹,
D. Priakhina^{1,2,3}, O. Streltsova^{1,2}, P. Zrellov^{1,3}

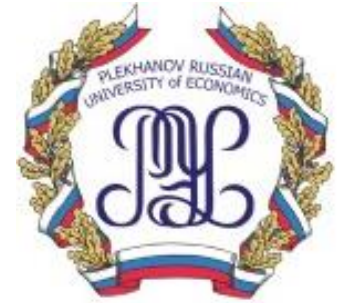
1. Joint Institute for Nuclear Research, Laboratory of Information Technology

2. Dubna State University

3. Plekhanov Russian University of Economics

Introduction

International IT School “Data Science”



DATA SCIENCE
IT SCHOOL

Training of IT specialists for:

- development of computing of megaprojects;
- Big Data analytics (Data Science);
- digital economy and other promising directions.

Introduction

International IT School “Data Science”



1st student enrollment
2018-2019 educational year



DATA SCIENCE
IT SCHOOL

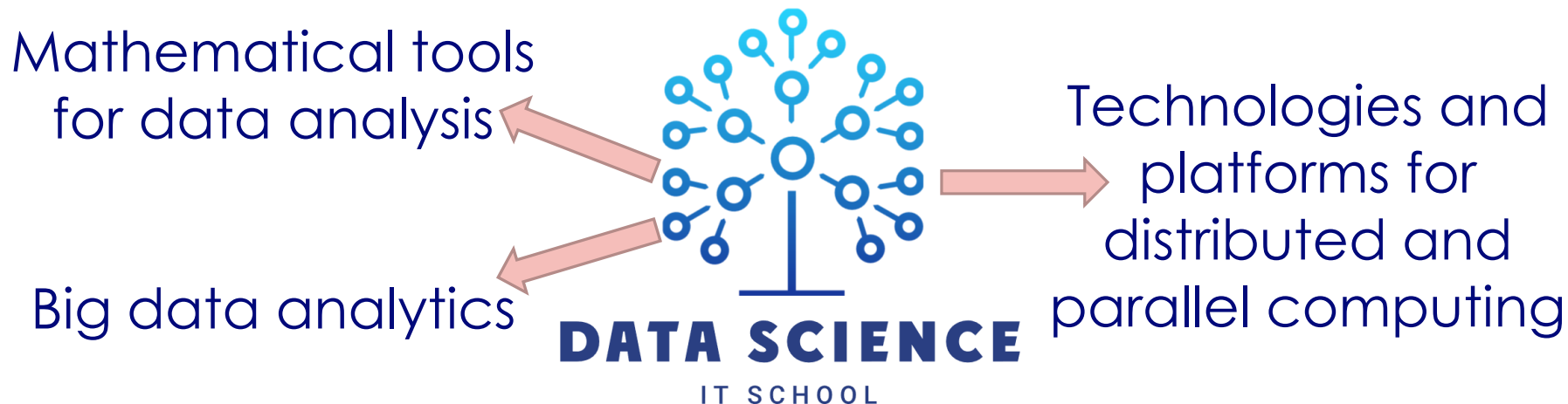
- Free education
- Admission based on the results of competitive selection
- The school program is mastered in parallel with the main educational programs of universities



Introduction

International IT School “Data Science”

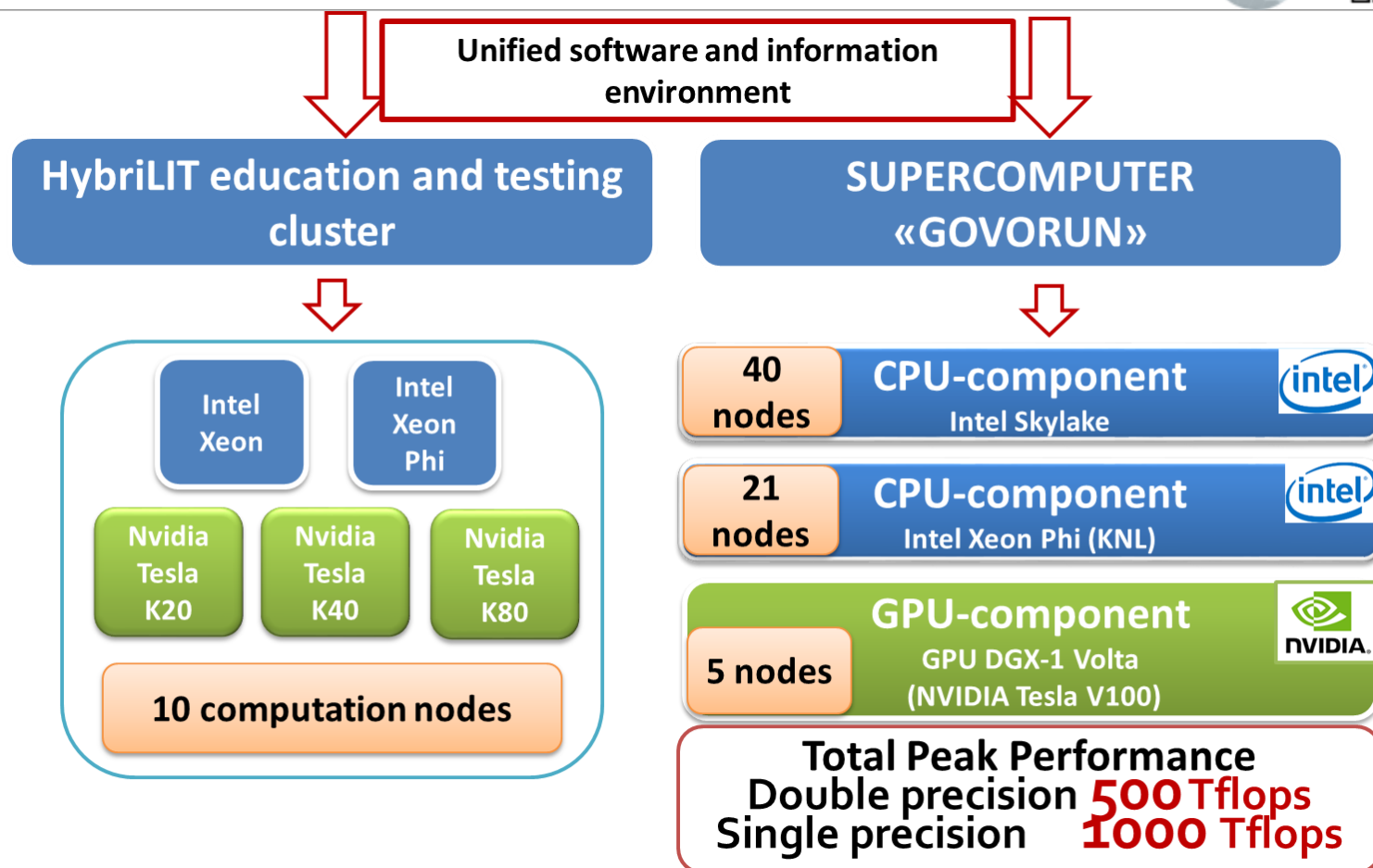
Some subjects



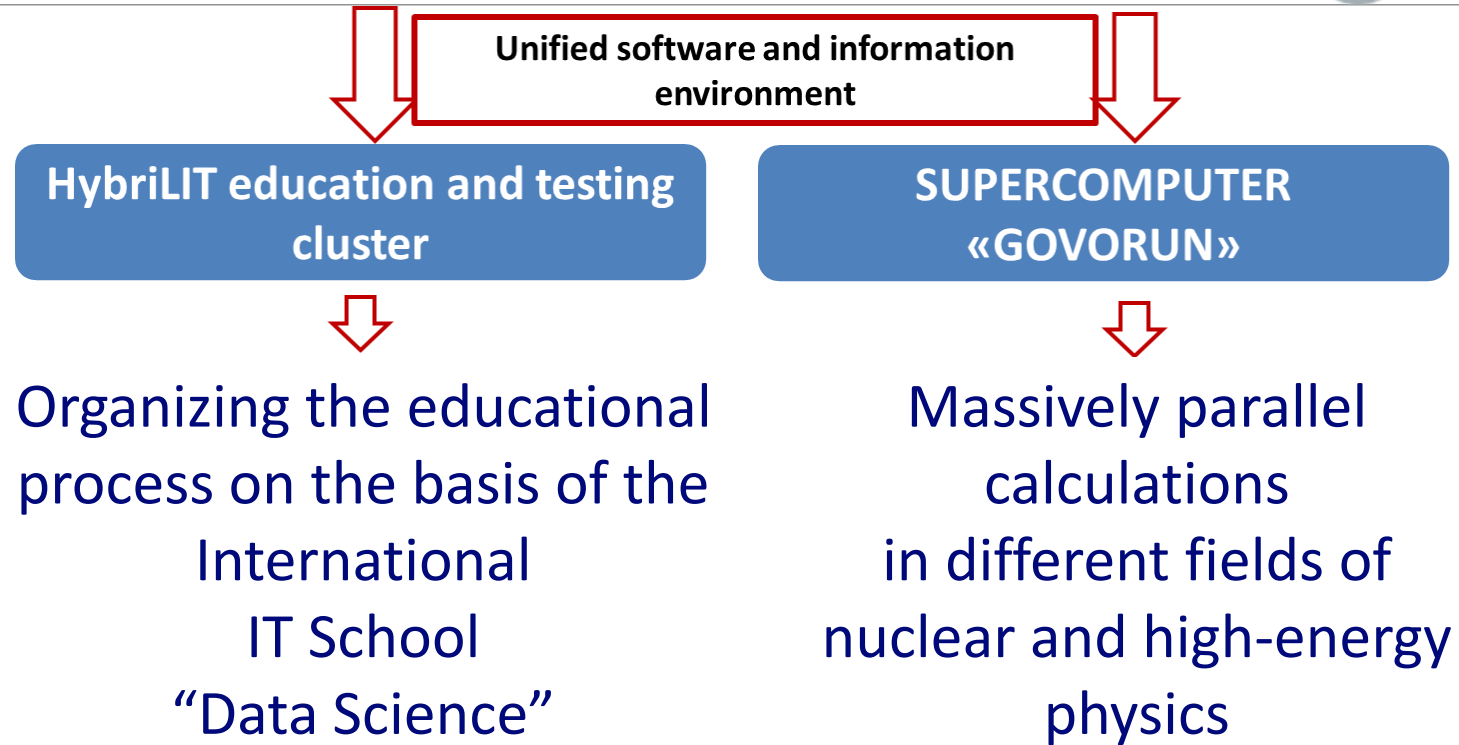
Training is held using the “HybriLIT” heterogeneous platform, which is part of the JINR Multifunctional Information and Computing Complex *



Heterogeneous platform “HybriLIT”



Heterogeneous platform “HybriLIT”



“HybriLIT” education and testing polygon in the educational process of IT School



«Technologies and platforms for
distributed and parallel computing»

Parallel programming technologies

- **OpenMP** – development on multiprocessor systems with shared memory (one node)
- **MPI** – development on multiprocessor systems with distributed memory (multiple nodes)
- **CUDA** – development using GPUs from NVIDIA
- **OpenCL** – development using CPUs and GPUs

OpenMP

MPI



Hybrid technologies

OpenMP
+
MPI

NVIDIA
CUDA
+
MPI

“HybriLIT” education and testing polygon in the educational process of IT School



«Mathematical tools for data analysis» &
«Big data analytics»

Ecosystem for ML/DL tasks and data analysis

Computation component

Servers with
NVIDIA Volta & Intel Xeon Gold

Dell Volta specs:

GPU: 4x Nvidia Volta V100-SXM2 *NVLink*
32Gb HBM2

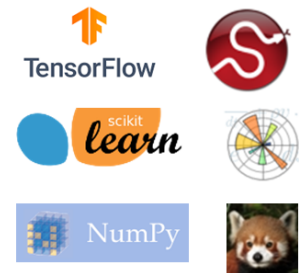
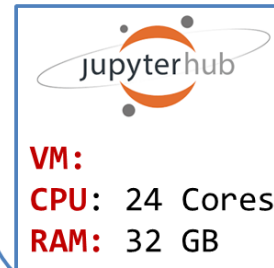
CPU: 2x Intel(R) Xeon(R) Gold 6148 CPU @
2.40GHz 20 Cores/40 Threads

RAM: 512 GB DDR4 2666MHz

SSD: 2*240 GB

Development component

VM with JupyterHub
<https://jhub.jinr.ru>



“HybriLIT” education and testing polygon in the educational process of IT School

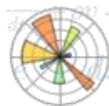


«Mathematical tools for data analysis»
Purposes

- Study of basic computational and statistical methods and tools for solving data analysis tasks
- Obtaining practical skills in developing software with programming languages used for scalable data processing (Python, R, Scala, C ++)

Libraries for studying Python, ML/DL methods and data analysis

➤ NumPy ➤ SciPy ➤ matplotlib ➤ scikit-learn ➤ pandas ➤ TensorFlow



“HybriLIT” education and testing polygon in the educational process of IT School



«Mathematical tools for data analysis»
Laboratory works

Jupyter Prac_tarainig_doc Last Checkpoint: 18 минут назад (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Run Code

Практическое занятие 2

1. Оформление отчетов:

все задания (лабораторные работы) оформляются в виде `.jpynb` файла, используя `Markdown Cell`:
! [[верхнее меню](#)] (Markdown.png).

Более подробная информация доступна [[по ссылке:](https://jupyter-notebook.readthedocs.io/en/stable/examples/Notebook/Working%20With%20Markdown%20Cells.html)] (<https://jupyter-notebook.readthedocs.io/en/stable/examples/Notebook/Working%20With%20Markdown%20Cells.html>) и [[Makddow](https://daringfireball.net/projects/markdown/syntax#link)] (<https://daringfireball.net/projects/markdown/syntax#link>).

Некоторые полезные элементы

Списки:
* утверждение 1
* утверждение 2
- утверждение 2.1

Например, выше определены заголовки разного уровня, блоки с кодом можно вставить так:

```
python
def myfunction(x)
    return (x+2)
```

Для формул можно использовать нотации `TeX/LaTeX`, например, формулы в тексте выделяются знаком `$`:
на новой строке формулы выделяются знаком `$$`:

```
$$ M[X] = \frac{1}{N} \sum_{i=1}^N x_i $$
```

После выполнения (`Alt+Enter` или кнопка `Run` верхнего меню) все будет выглядеть так:

Практическое занятие 2

1. Оформление отчетов:

все задания (лабораторные работы) оформляются в виде `.jpynb` файла, используя `Markdown Cell`:

верхнее меню

Более подробная информация доступна [по ссылке](#); и [Markdown Style](#).

Некоторые полезные элементы

Списки:

- утверждение 1
- утверждение 2
 - утверждение 2.1

Например, выше определены заголовки разного уровня, блоки с кодом можно вставить так:

```
def myfunction(x)
    return (x+2)
```

Для формул можно использовать нотации `TeX/LaTeX`, например, формулы в тексте выделяются знаком `$`: $e^x = 0$, а для формул на новой строке формулы выделяются знаком `$$`:

$$M[X] = \frac{1}{N} \sum_{i=1}^N x_i$$

После выполнения (`Alt+Enter` или кнопка `Run` верхнего меню) все будет выглядеть так:

Jupyter Nonebook
Markdown style



.pdf document
to teachers for review

“HybriLIT” education and testing polygon in the educational process of IT School



«Big data analytics»
Purpose

- Study of modern tools and technologies used in data analysis



- Gaining experience with:

- Unix-like operating systems;
- cloud technologies * as a tool for the organization of software-managed computing infrastructures
- collaborative software development tool;



* <https://cloud.jinr.ru/>

“HybriLIT” information environment in the educational process of IT School



«Indico» service



➤ Platform for publishing schedules

	1 Mar 2019	6 Mar 2019	11 Mar 2019	12 Mar 2019	13 Mar 2019	15 Mar 2019	18 Mar 2019
AM							
PM	17:20 Лекция 1. Основные тренды и вызовы в ИТ, концепция Международной школы по информационным технологиям «Аналитика больших данных» - Владимир Кореньков (1-122)	17:50 Лекция 1. Экосистема для анализа данных. Задачи машинного обучения - Oksana Streltsova (1-118)	14:30 Семинар 1. Группа Б-2. Математический аппарат и инструментарий для анализа данных - Oksana Streltsova (1-213) 14:30 Семинар 1. Группа М-1. Введение в OC UNIX - Nikita Balashov (1-211) 16:10 Семинар 1. Группа Б-2. Введение в OC UNIX - Nikita Balashov (1-318) 17:50 Семинар 1. Группа Б-1. Введение в OC UNIX - Nikita Balashov (1-318)	17:50 Семинар 1. Группа Б-1. Математический аппарат и инструментарий для анализа данных - Darya Pryahina (JINR) (1-318)	16:10 Семинар 1. Группа М-1. Математический аппарат и инструментарий для анализа данных - Petr Sychov (1-319) 17:50 Лекция 2. Введение в Python - Petr Sychov (1-118)	17:50 Лекция 2. Введение в аналитику Больших данных - Петр Валентинович Зрепов (1-118)	14:30 Семинар 2. Группа Б-2. Математический аппарат и инструментарий для анализа данных - Oksana Streltsova (1-213) 14:30 Семинар 2. Группа М-1. Введение в OC UNIX - Nikita Balashov (1-211) 16:10 Семинар 2. Группа Б-2. Введение в OC UNIX - Nikita Balashov (1-211) 17:50 Семинар 2. Группа Б-1. Введение в OC UNIX - Nikita Balashov (1-211)

“HybriLIT” information environment in the educational process of IT School



«Indico» service



➤ Platform for publishing lectures and workshops

Математический аппарат и инструментарий для анализа данных

1 March 2019 to 1 June 2019
Университет “Дубна”
Europe/Moscow timezone

Описание

Расписание

Программа курса

Список студентов

Поддержка

pryahina@jinr.ru

Целью курса является освоение основных вычислительных и статистических методов и инструментов для решения широкого круга задач в сфере анализа данных (*Data Science*), практических навыков по разработке программного обеспечения на языках программирования, применяемых для масштабируемой обработки данных: *Python, R, Scala, C++*.

Основные модули дисциплины:

- Дополнительные главы математики
- Языки программирования для анализа данных

Разработка будет проходить на VM гетерогенной платформы HYbriLIT.

Для начала работы необходимо:

1. Войти под своей учетной записью в GitLab:
<https://gitlab-hybrilit.jinr.ru/>
2. Войти на VM:
<https://hub.jinr.ru>

(авторизация через GitLab)

Для установки *Jupyter Notebook* на персональный компьютер необходимо:

1. Скачать дистрибутив *Anaconda* по [ссылке](#), выбрав *Python 3.7 version*.
2. Запустить скаченный файл и установить *Jupyter Notebook*.

Математический аппарат и инструментарий для анализа данных

1 March 2019 to 1 June 2019
Университет “Дубна”
Europe/Moscow timezone

Описание

Расписание

Программа курса

Список студентов

Поддержка

pryahina@jinr.ru

Программа курса

33 / 33

Enter #id or search string

🔍 📄 🔗

Лекция 1. Экосистема для анализа данных. Задачи машинного обучения
👤 Oksana Streltsova
🕒 06/03/2019, 17:50

В первой лекции представлена экосистема для проведения практических занятий: набор библиотек Python, которыми будем пользоваться для решения задач в области анализа данных. На примере решения прикладной

Семинар 1. Группа Б-2. Математический аппарат и инструментарий для анализа данных
👤 Oksana Streltsova
🕒 11/03/2019, 14:30

На первом практическом занятии будет рассмотрена работа с IPython и Jupyter Notebook, ключевые составляющие языка Python, начало работы с библиотечкой matplotlib.

Семинар 1. Группа Б-1. Математический аппарат и инструментарий для анализа данных
👤 Darya Pryahina
🕒 12/03/2019, 17:50

На первом практическом занятии будет рассмотрена работа с IPython и Jupyter Notebook, ключевые составляющие языка Python.

Семинар 1. Группа М-1. Математический аппарат и инструментарий для анализа данных
👤 Dr Peter Syrov (Dubna University)
🕒 13/03/2019, 16:10

Лекция 2. Введение в Python
👤 Dr Peter Syrov (Dubna University)
🕒 13/03/2019, 17:50

“HybriLIT” information environment in the educational process of IT School



«Indico» service



➤ Platform for publishing lectures and workshops

Математический аппарат и инструментарий для анализа данных

1 March 2019 to 1 June 2019
Университет "Дубна"
Europe/Moscow timezone

Описание

Расписание

Программа курса

Список студентов

Поддержка

✉ pryahinad@jlnr.ru

Лекция 1. Экосистема для анализа данных. Задачи машинного обучения

6 Mar 2019, 17:50
1h 30m
1-118 (Университет "Дубна")

Speaker
Oksana Streltsova

Description
В первой лекции представлена экосистема для проведения практических занятий: набор библиотек Python, которыми будем пользоваться для решения задач в области анализа данных. На примере решения прикладной задачи разделение и идентификация событий, соответствующих нейтронам и гамма-квантам, детектора DEMON, будет показан весь путь от сбора данных, постановки задачи, выбора методов, исследования их эффективности до получения конечного программного продукта.
Материалы
1. The official home of the Python Programming Language: <https://www.python.org/>
2. М. Саммерфилд. Программирование на Python 3. Подробное руководство.
-СПб.: Символ-Плюс. 2015.-608 с.

Presentation Materials

Lecture_1_MAT&INSTR_06_03_2019.pdf

Математический аппарат и инструментарий для анализа данных

1 March 2019 to 1 June 2019
Университет "Дубна"
Europe/Moscow timezone

Описание

Расписание

Программа курса

Список студентов

Поддержка

✉ pryahinad@jlnr.ru

Семинар 1. Группа Б-2. Математический аппарат и инструментарий для анализа данных

11 Mar 2019, 14:30
1h 30m
1-213 (Университет "Дубна")

Speaker
Oksana Streltsova

Description
На первом практическом занятии будет рассмотрена работа с Python и Jupyter Notebook, ключевые составляющие языка Python, начало работы с библиотекой matplotlib.
Материалы:
1. The official home of the Python Programming Language: <https://www.python.org/>
2. М. Саммерфилд. Программирование на Python 3. Подробное руководство.
-СПб.: Символ-Плюс. 2015.-608 с.

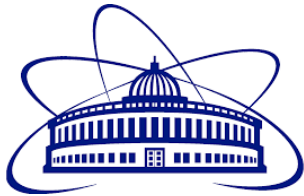
Presentation Materials

Python_practical_training-1.pdf

Conclusions



- Using the “HybriLIT” education and testing polygon allows students to master novel IT solutions and technologies that will be included in the curricula of higher educational institutions only in the future.



- Using the “HybriLIT” information environment allows teachers to increase the effectiveness of interaction with students.

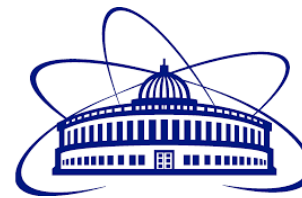


NEC'2019



XXVII International Symposium on Nuclear Electronics & Computing

Montenegro, Budva, Becici, 30 September - 4 October 2019



Thank you for your attention!



**Heterogeneous IT platform “HybriLIT”
for organizing the educational process on the basis of the
International IT School “Data Science”**

P. Zrellov^{1,3}, S. Belov^{1,3}, V. Korenkov^{1,2,3}, M. Matveev^{1,2,3},
D. Podgainy¹, **D. Priakhina**^{1,2,3}, O. Streltsova^{1,2}

1. Joint Institute for Nuclear Research, Laboratory of Information Technology

2. Dubna State University

3. Plekhanov Russian University of Economics