BES-III distributed computing status

Presented by Alexander UZHINSKIY

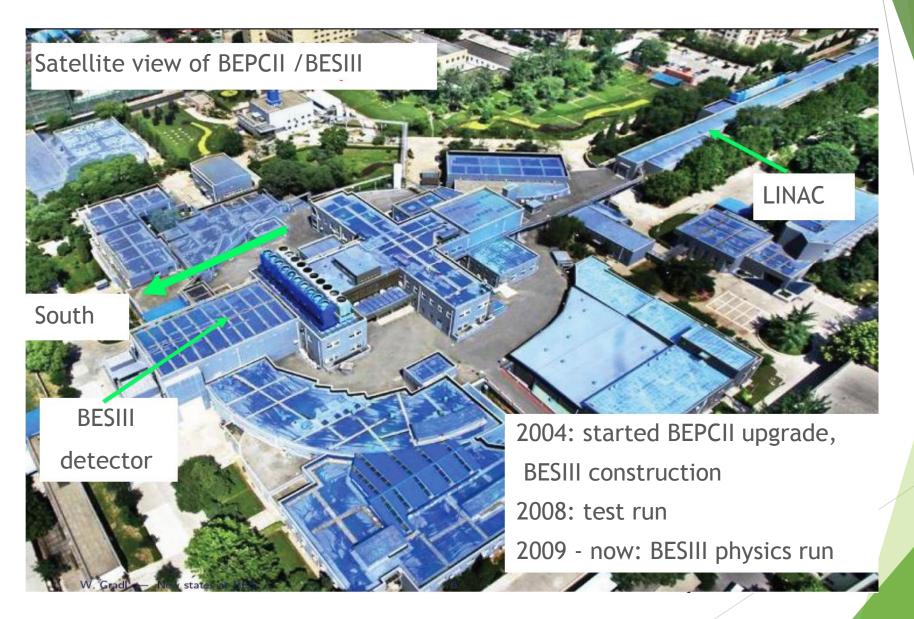
Authors: Sergey BELOV¹, Igor PELEVANYUK ¹, Alexander UZHINSKIY ¹, Alexey ZHEMCHUGOV ¹, Ziyan DENG ², Xiaomei ZHANG ², Weidong LI ², Xianghu ZHAO ², Tian YAN ², Tao LIN ², Gang ZHANG ², Xiaofei YAN ²

1. Joint Institute for Nuclear Research

2. Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, China

Grid2014, Dubna, 3.07.2014

Beijing Electron Positron Collider (BEPC)



The BESIII Collaboration

US (6) Germany: Univ. of Bochum, Univ. of Giessen, GSI Univ. of Hawaii Univ. of Washington Helmholtz Ins. In Mainz Carnegie Mellon Univ. Univ. of Minnesota Univ. of Torino, Frascati Lab Univ. of Rochester Univ. of Indiana Uppsala Univ. **Turkey Accelerator Center** china 30 Pakistan (2 Univ. of Punjab COMSAT CIIT

~ 400 members 53 institutions from 11 coutries

Europe (12)

Univ. of Johannes Gutenberg : JINR Dubna; BINP Novosibirsk KVI/Univ. of Groningen

Korea (1

Seoul Nat. Univ.

Japan (1)=

Tokyo Univ.

IHEP, CCAST, GUCAS, Shandong Univ., Univ. of Sci. and Tech. of China Zhejiang Univ., Huangshan Coll. Huazhong Normal Univ., Wuhan Univ. Zhengzhou Univ., Henan Normal Univ. Peking Univ., Tsinghua Univ., Zhongshan Univ., Nankai Univ. Shanxi Univ., Sichuan Univ., Univ. of South China Hunan Univ., Liaoning Univ. Nanjing Univ., Nanjing Normal Univ. Guangxi Normal Univ., Guangxi Univ. Suzhou Univ., Hangzhou Normal Univ. Lanzhou Univ., Henan Sci. and Tech. Univ. Hong Kong Univ., Hong Kong Chinese Univ.

125

BES-III Introduction

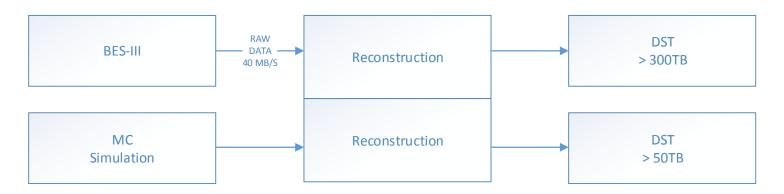
- The BES-III experiment in Beijing is a world best facility to test Standard Model and QCD with high precision in taucharm domain
- JINR participates in the experiment since 2005
- BES-III data volume ~0.5 PB/year \rightarrow 2 PB (2015), 2x109 CPU hours for data processing. Consolidation of resources is necesary so the BES-III Grid is being constructed
- LIT team is a key developer of the BES-III distributed computing system



Status of the project

- A prototype of BES-III Grid has been built (9 sites including IHEP CAS and JINR). Main developments have been done at IHEP and JINR. The Grid is based on DIRAC interware.
- First production (800 million J/ψ events) completed successfully. JINR have contributed ~10% of total resources. Current success rate 93%.
- Fully operational system should be setup by 2015
- The infrastructure can be used to process data in other joint JINR-China projects (Daya Bay, NICA (?) ...)

The BES-III data flow



Experimental data are taken from the BES-III detector and stored as raw to the tape storage managed by CASTOR.

The maximum data rate is about 40 MB/s.

DSTs are stored in a disk pool managed by Lustre and dCache and can be accessed only from internal IHEP network.

The total amount of DSTs currently is about 300 TB.

Both inclusive and exclusive Monte-Carlo simulation (MC) is made for each data sample. The total amount of MC DSTs is more than 50 TB now.

The BES-III offline software is based on the Gaudi framework and runs on Scientific Linux

The BES-III distributed computing system

Remote sites participate only in MC production and physics analysis, while all reconstruction of experimental and simulated data is done at IHEP.

Three operation models are considered, depending on the capabilities and priorities of each site: a) MC simulation runs at remote sites. The resulting

data are copied back to IHEP and then MC reconstruction runs there.

b) MC simulation and reconstruction runs at remote sites. The resulting data are copied back to IHEP;c) DSTs are copied from IHEP and other sites and analyzed using local resources.

Distributed analysis is postponed for later stage



The BES-III grid solution



DIRAC is BES-III grid solution because:

- DIRAC provides all the necessary components to build ad-hoc distributed computing infrastructures interconnecting resources of different types, allowing interoperability and simplifying interfaces.
- Dirac provide: job management, data management, information system, monitoring, security system
- Dirac is rather easy to install, configure and maintain
- DIRAC Supports grids which based on different middleware (gLite, EGI, VDT, ARC, etc)
- With DIRAC no grid middleware installation needed on site (accessed through an SSH tunnel)

The BES-III Job management



Job submission adopted to BES-III needs

BES-III CE's: 3 gLite-CREAM, 6 SSH-CE





Now main interest is in submitting of BES-III jobs in clouds

The BES-III Data management



DFC (Dirac file catalog) file operation, dataset operation

SE's: dCache, Bestman, Storm



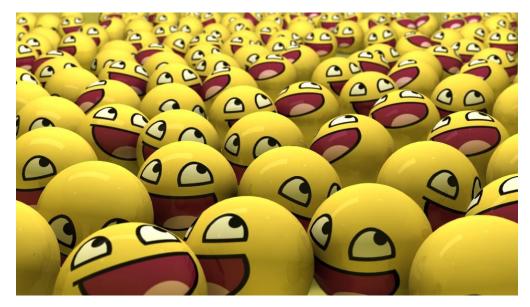


Data transfer - BES-III transfer system (FTS)

Solution on dCache-Lustre integration was provided for main data storage in IHEP



The BES-III Monitoring



The main goal of the monitoring to make more users happy:

- to decrease the number of failed jobs

- to understand the failure reasons Another goal is to lighten the administrator's work

- to show system malfunction before failure occures

- to control overall status of the grid
- to optimize data transfers
- to check storage (=data) availability
- to deploy new sites

First version of the BES-III grid monitoring is operational (http://vm162.jinr.ru)

• Basic site monitoring tools are there to facilitate commissioning of new sites and to improve production reliability

• Storage monitoring becomes a hot issue. Besides that, many more developments are necessary.

The BES-III Monitoring

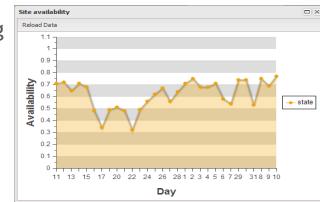
Available tests

- Network ping test
- WMS test (sending simple job)
- Simple BOSS job (full simulation of 50 events)
- combined test of CVMFS, environment and resources availability
- CPUlimit test

- a number of jobs failed at remote sites because the wrong CPU limit was set ata few WNs

- Host failure accounting
- analyzes failures per individual WN during one week and identifies problematic hosts

CE availability	monitoring						
Site	Service	Test	Result	Description	24H Reliability	48H Reliability	Week Reliability
BES.UCAS.cn	WMS	WMS_send_test	Fail	Failed after 30	0.00	0.00	0.00
BES.IHEP-PBS.cn	WMS	WMS_send_test	Success	Remote call	0.17	0.09	0.02
BES.JINR.ru	WMS	WMS_send_test	Success	Remote call	0.03	0.01	0.00
BES.PKU.cn	WMS	WMS_send_test	Fail	Failed after 30	0.07	0.03	0.01
BES.UMN.us	WMS	WMS_send_test	Success	Remote call	0.14	0.07	0.02
BES.USTC.on	WMS	WMS_send_test	Success	Remote call	0.24	0.13	0.04
BES.WHU.cn	WMS	WMS_send_test	Success	Remote call	0.21	0.11	0.03
BES.INFN-Torin	WMS	WMS_send_test	Success	Remote call	0.20	0.11	0.03
BES.IHEP-VM.cn	WMS	WMS_send_test	Fail	Failed after 30	0.00	0.00	0.00
BES.UCAS.cn	WMS	BOSS_work_test	Fail	Failed after 90	0.00	0.00	0.00
BES.IHEP-PBS.on	WMS	BOSS_work_test	Success	Success	0.20	0.11	0.03
BES.JINR.ru	WMS	BOSS_work_test	Success	Success	0.18	0.11	0.03
BES.PKU.cn	WMS	BOSS_work_test	Fail	Failed after 90	0.13	0.07	0.02
BES.UMN.us	WMS	BOSS_work_test	Success	Success	0.20	0.11	0.03
BES.USTC.on	WMS	BOSS_work_test	Fail	boss.exe not fo	0.00	0.00	0.00
BES.WHU.cn	WMS	BOSS_work_test	Success	Success	0.26	0.14	0.04
BES.INFN-Torin	WMS	BOSS_work_test	Success	Success	0.26	0.14	0.04
BES.IHEP-VM.on	WMS	BOSS_work_test	Fail	Failed after 90	0.00	0.00	0.00
BES.UCAS.cn	WMS	CPU_limit_test	Fail	Failed after 30	0.00	0.00	0.00
BES.IHEP-PBS.cn	WMS	CPU_limit_test	Success	Success	0.21	0.11	0.03
BES.JINR.ru	WMS	CPU_limit_test	Success	Success	0.03	0.01	0.00
BES.PKU.cn	WMS	CPU_limit_test	Fail	Failed after 30	0.06	0.03	0.01
BES.UMN.us	WMS	CPU_limit_test	Success	Success	0.17	0.09	0.02
BES.USTC.on	WMS	CPU_limit_test	Success	Success	0.20	0.11	0.03
BES.WHU.cn	WMS	CPU_limit_test	Success	Success	0.20	0.11	0.03
BES.INFN-Torin	WMS	CPU_limit_test	Success	Success	0.21	0.11	0.03



📰 Network moni					
Site	Host	CEType	AverageTime	Passed	Description
BES.IHEP-PBS.cn	lxslc5.ihep.ac.cn	SSHTorque	0.23	1.00	
BES.GUCAS.cn	gucasfarm0.ihe	SSHTorque	0.342	1.00	
BES.IHEP-LCG.cn	cce.ihep.ac.cn	CREAM	0.382	1.00	
BES.JINR.ru	lcgce 12.jinr.ru	CREAM	224.387	1.00	
BES.JINR.ru	lcgce21.jinr.ru	CREAM	224.384	1.00	
BES.JINR.ru	lgdce01.jinr.ru	CREAM	224.431	1.00	
BES.INFN-Torin	t2-ce-02.to.infn.it	CREAM	-1	0.00	Packets filtered
BES.USTC.cn	ui04.lcg.ustc.ed	SSHTorque	28.926	1.00	
BES.USTC.cn	ui01.lcg.ustc.ed	SSHTorque	31.501	1.00	
BES.UMN.us	bes3s1.spa.um	SSHGE	210.787	1.00	
BES.PKU.cn	hepfarm02.phy	SSHTorque	31.892	1.00	
BES.SDU.cn	sl03.hepg.sdu.e	SSHTorque	-1	0.00	
BES.WHU.cn	202.114.78.124	SSHTorque	19.949	1.00	
BES.NSCCSZ.cn	183.62.232.132	SSHTorque	38.928	1.00	

Host monitori	ng										8		x
Site	Host	24H :	24H *	24H I	24H Ra	48H :	48H ⁻	48H I	48H Ra	Weel	Weel	Weel Week F	
BES.IHEP-PBS.cn	gridtb002.ihep.ac.cn	2	2		1.00	2	2		1.00	2	2	1.00	_
BES.UMN.us	twins-a04.spa.umn.edu	1	1		1.00	1	1		1.00	1	1	1.00	
BES.JINR.ru	wn362.jinr.ru					1	1		1.00	1	1	1.00	
BES.IHEP-CLOU	diraccloudinit1403249980									7	7	1.00	
BES.IHEP-CLOU	diraccloudinit1403250760									1	1	1.00	
BES.UMN.us	twins-b14.spa.umn.edu	1	1		1.00	1	1		1.00	1	1	1.00	
BES.UMN.us	twins-a24.spa.umn.edu					1	1		1.00	1	1	1.00	
BES.IHEP-CLOU	diraccloudinit1403250400									4	4	1.00	
ES.JINR.ru	wni000.jinr.ru					1	1		1.00	1	1	1.00	
ES.JINR.ru	wn400.jinr.ru					1	1		1.00	1	1	1.00	
ES.JINR.ru	wn323.jinr.ru	1	1		1.00	1	1		1.00	1	1	1.00	
ES.IHEP-CLOU	diraccloudinit1403490272	1	1		1.00	14	14		1.00	14	14	1.00	
ES.WHU.cn	cu33	6	6		1.00	6	6		1.00	6	6	1.00	
BES.UMN.us	twins-b03.spa.umn.edu					1	1		1.00	1	1	1.00	
ES.IHEP-CLOU	diraccloudinit1403254687					5	5		1.00	5	5	1.00	
ES.IHEP-CLOU	diraccloudinit1403495687	2	2		1.00	7	7		1.00	7	7	1.00	
ES.JINR.ru	wn324.jinr.ru					1	1		1.00	1	1	1.00	*

IHEPO-USER Link (J. C.) IHEPO-USER LINR-USER 2,673 IHEPO-USER JINR-USER 16,316 IHEPO-USER USTC-USER 15,932 IHEPO-USER USTC-USER 15,932 JINR-USER IHEPO-USER 6,728 JINR-USER IHEPO-USER 14,322 JINR-USER JINR-USER 14,24 JINR-USER USTC-USER 14,825 JINR-USER WHU-USER 8,516 USTC-USER JINR-USER 17,855 USTC-USER JINR-USER 2,746 USTC-USER IHEPO-USER 5,727 WHU-USER IHEPO-USER 5,227 WHU-USER JINR-USER 20,227 WHU-USER USTC-USER 9,199	SE latency n	nonitoring	
IHEPD-USER JINR-USER 16.316 IHEPD-USER USTC-USER 15.932 IHEPD-USER USTC-USER 14.322 JINR-USER JINR-USER 14.322 JINR-USER JINR-USER 14.24 JINR-USER USTC-USER 14.827 JINR-USER WHU-USER 14.827 JINR-USER WHU-USER 14.827 JINR-USER USTC-USER 14.827 JINR-USER WHU-USER 8.516 USTC-USER JINR-USER 17.855 USTC-USER USTC-USER 2.746 USTC-USER HEPO-USER 5.727 WHU-USER IHEPO-USER 5.727 WHU-USER JINR-USER 20.227 WHU-USER USTC-USER 9.199	Source	Destination	Latency(sec)
IHEPD-USER USTC-USER 15.932 IHEPD-USER WHU-USER 6.728 JINR-USER IHEPD-USER 14.322 JINR-USER JINR-USER 14.24 JINR-USER WHU-USER 14.827 JINR-USER WHU-USER 8.516 USTC-USER IHEPD-USER 3.677 USTC-USER JINR-USER 2.746 USTC-USER WHU-USER 6.24.375 WHU-USER IHEPD-USER 5.727 WHU-USER IHEPO-USER 5.227 WHU-USER JINR-USER 20.227 WHU-USER USTC-USER 9.199	IHEPD-USER	IHEPD-USER	2.678
IHEPD-USER WHU-USER 6.728 JIINR-USER IHEPD-USER 14.322 JIINR-USER JINR-USER 14.24 JIINR-USER JINR-USER 14.827 JIINR-USER WHU-USER 8.516 JISTC-USER HEPD-USER 3.677 USTC-USER JIINR-USER 7.465 USTC-USER WHU-USER 2.746 USTC-USER WHU-USER 624.375 WHU-USER JIINR-USER 20.227 WHU-USER JIINR-USER 20.227 WHU-USER USTC-USER 9.199	IHEPD-USER	JINR-USER	16.316
INR-USER IHEPO-USER 14.322 JINR-USER JINR-USER 14.24 JINR-USER USTC-USER 14.827 JINR-USER USTC-USER 14.827 JINR-USER WHU-USER 8.516 USTC-USER IHEPO-USER 3.677 USTC-USER JINR-USER 17.855 USTC-USER WHU-USER 2.746 USTC-USER WHU-USER 624.375 WHU-USER JINR-USER 5.727 WHU-USER JINR-USER 20.227 WHU-USER USTC-USER 9.199	IHEPD-USER	USTC-USER	15.932
IJINR-USER JINR-USER 11.4.2 JIJNR-USER JINR-USER 14.827 JINR-USER USTC-USER 14.827 JINR-USER WHU-USER 8.516 USTC-USER JHEPD-USER 3.677 USTC-USER JINR-USER 17.855 USTC-USER USTC-USER 2.746 USTC-USER WHU-USER 624.375 WHU-USER IHEPO-USER 5.727 WHU-USER JINR-USER 20.227 WHU-USER USTC-USER 9.199	IHEPD-USER	WHU-USER	6.728
INR-USER USTC-USER 14.827 JINR-USER WHU-USER 8.516 USTC-USER IHEPD-USER 3.677 USTC-USER JINR-USER 17.855 USTC-USER USTC-USER 2.746 USTC-USER WHU-USER 624.375 WHU-USER IHEPO-USER 5.727 WHU-USER JINR-USER 20.227 WHU-USER USTC-USER 9.199	JINR-USER	IHEPD-USER	14.322
INR-USER WHU-USER 8.516 USTC-USER IHEPD-USER 3.677 USTC-USER IJINR-USER 17.855 USTC-USER USTC-USER 2.746 USTC-USER WHU-USER 624.375 WHU-USER IHEPO-USER 5.727 WHU-USER JINR-USER 20.227 WHU-USER USTC-USER 9.199	JINR-USER	JINR-USER	14.24
USTC-USER IHEPD-USER 3.677 USTC-USER JINR-USER 17.855 USTC-USER USTC-USER 2.746 USTC-USER WHU-USER 624.375 WHU-USER IHEPD-USER 5.727 WHU-USER JINR-USER 20.227 WHU-USER USTC-USER 9.199	JINR-USER	USTC-USER	14.827
USTC-USER JINR-USER 17.855 USTC-USER USTC-USER 2.746 USTC-USER WHU-USER 624.375 WHU-USER IHEPD-USER 5.727 WHU-USER JINR-USER 20.227 WHU-USER USTC-USER 9.199	JINR-USER	WHU-USER	8.516
USTC-USER USTC-USER 2.746 USTC-USER WHU-USER 624.375 WHU-USER IHEPD-USER 5.727 WHU-USER JINR-USER 20.227 WHU-USER USTC-USER 9.199	USTC-USER	IHEPD-USER	3.677
USTC-USER WHU-USER 624.375 WHU-USER IHEPD-USER 5.727 WHU-USER JINR-USER 20.227 WHU-USER USTC-USER 9.199	USTC-USER	JINR-USER	17.855
WHU-USER IHEPD-USER 5.727 WHU-USER JINR-USER 20.227 WHU-USER USTC-USER 9.199	USTC-USER	USTC-USER	2.746
WHU-USER JINR-USER 20.227 WHU-USER USTC-USER 9.199	USTC-USER	WHU-USER	624.375
WHU-USER USTC-USER 9.199	WHU-USER	IHEPD-USER	5.727
	WHU-USER	JINR-USER	20.227
WHU-USER WHU-USER 3.092	WHU-USER	USTC-USER	9.199
	WHU-USER	WHU-USER	3.092

Summary & plans

The BES-III computing is operational since 2013 about 350000 jobs executed about 250 TB managed disk space

More development needed at:

- Data set management
- Job management and data management integration (random trigger data)
- Monitoring & accounting
- FTS migration
- Clouds integration

JINR team participates in all tasks of BES-III grid development

•

•

Thank you!