

Distributed Computing R&D

progress report

Fabio HERNANDEZ *on behalf of*

IHEP: CHEN Gang, LI Weidong, QI Fazhi, WANG Lu, ZHANG Xiaomei, CHEN Yaodong, YAN Tian, WANG Cong

CC-IN2P3: Ghita RAHAL, Vanessa HAMAR, Laurent CAILLAT-VALLET, Frédéric SUTER

CPPM: Andreï TSAREGORODTSEV



9th FCPPL Workshop — Strasbourg (France), March 30th, 2016

Background

- Guiding principle

explore technologies of potential interest for the data processing needs of HEP experiments

- Partners

CCPM, IN2P3 computing center, IHEP computing center

- Topics

DIRAC-based computing platform for IHEP experiments

Identification and analysis of I/O patterns

Cloud-based storage for remote direct data access

Explorations of software-defined networking

Background (cont.)

- Funding: several sources this year

IHEP and IN2P3 through the FCPPL 2015 call

CNRS-NSFC joint program for international collaboration

CAS President's International Fellowship Initiative

CAS fellowships for PhD students staying at CNRS labs

DIRAC-based platform

DIRAC-based platform

5

- BES III has been using DIRAC in production for a few years now

14 sites: China (9), Italy (3), US (1), Rusia (1)

both physical and virtual machines: about 1700 CPU cores, 500 TB of storage

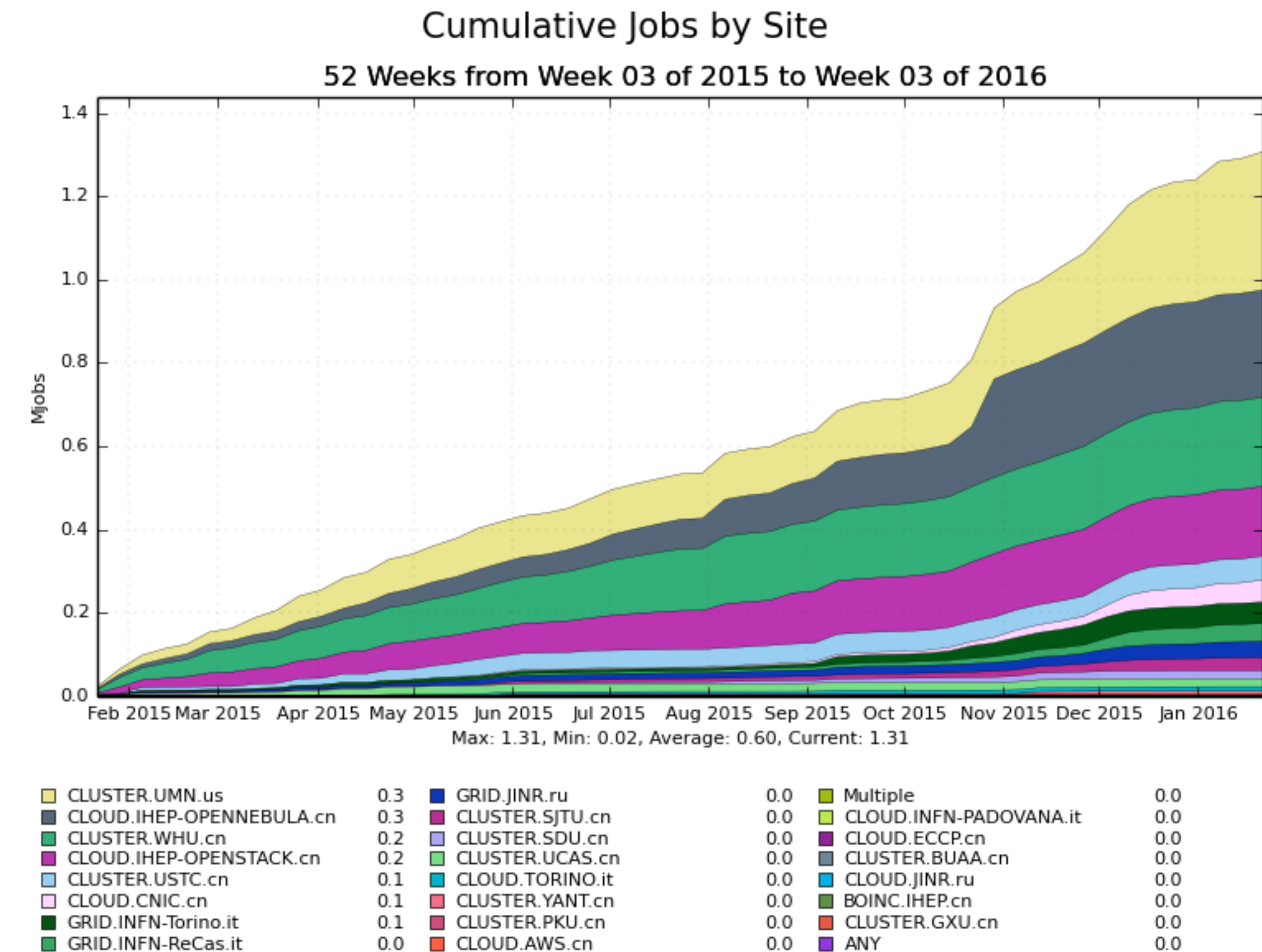
1.2 M jobs handled by this platform in 2015

platform centrally managed and operated by IHEP

- Visit of IHEP by Andrei Tsaregorodtsev (CPPM) in May 2015

complete software upgrade and reconfiguration of back-end services and user-facing web application

now both JUNO and CEPC can use the same infrastructure



Generated on 2016-01-26 15:47:12 UTC

DIRAC-based platform (cont.)

○ Consolidation of storage infrastructure

using StoRM (backed by Lustre) as the storage element for CEPC: storage area accessible both from grid jobs and from local jobs running at IHEP

no need for data replication between storage elements

BES III configuration is expected to evolve in the same direction

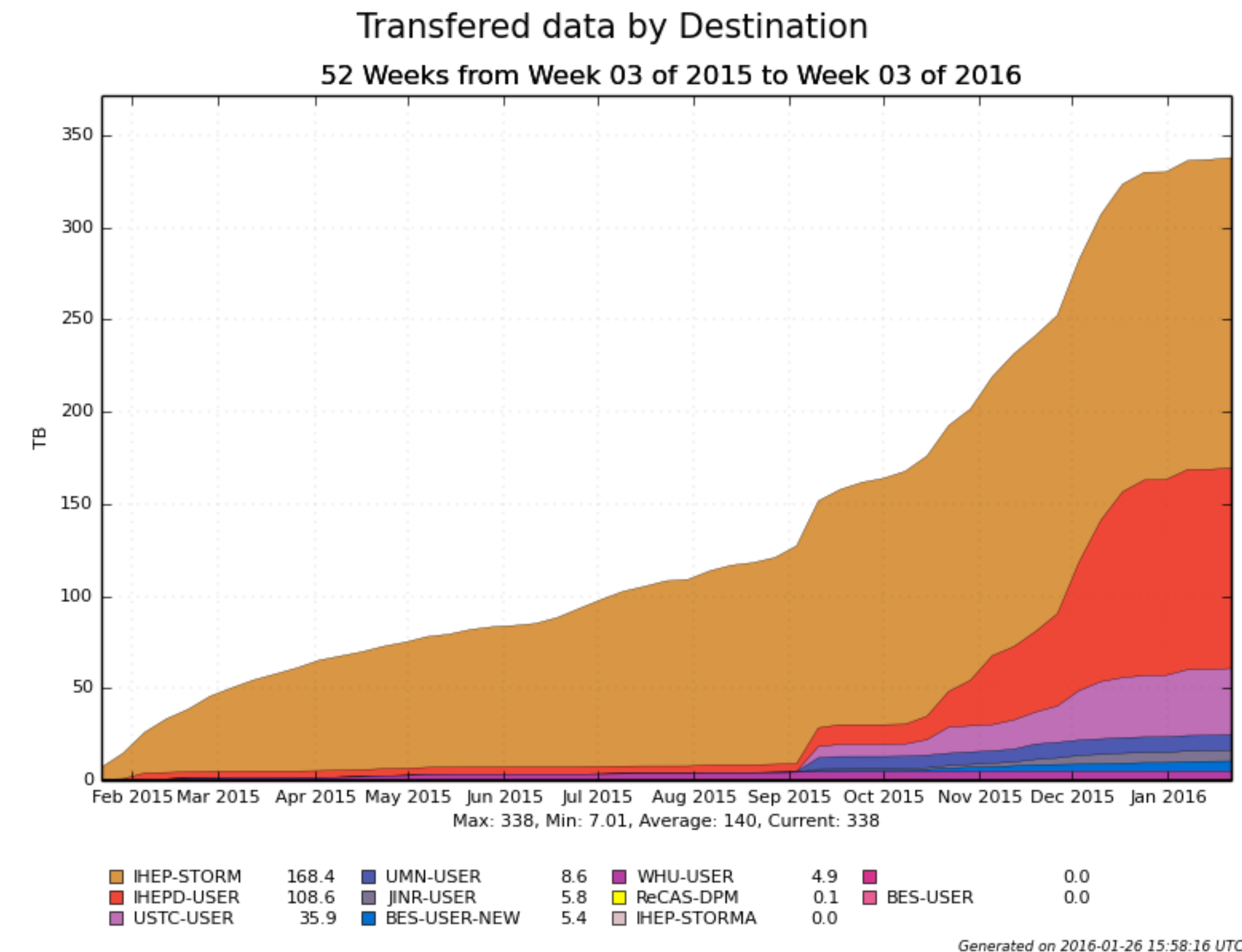
DIRAC file catalog now contains files from all configured virtual organisations

350 TB of data exchanged by the participating sites

○ Software distribution

generalised use of CernVM FS for software distribution for BES III, CEPC and JUNO

centrally controlled by IHEP, using servers at IHEP and at CERN



DIRAC-based platform (cont.)

- Development

effort provided by IHEP for developing tools for continuous central monitoring of the platform

development work on top of DIRAC API to extend its functionalities to submit and manage the lifecycle of tasks: groups of related jobs handled as a unit from the user point of view

- Exploration

IHEP got a 10.000 RMB grant from Amazon Web Services China

used to exercise transparent integration to DIRAC of virtual machines provided by commercial cloud operators

- People

ZHANG Xiaomei consistently attending and representing IHEP at the annual DIRAC users workshops

Analysis of I/O patterns

Analysis of I/O patterns

- Work performed in the framework of the CNRS-NSFC collaboration program
- Tools for collecting and analysing I/O activity observed by the Lustre file servers at IHEP

3 years-long grant obtained by WANG Lu (IHEP) and Fabio HERNANDEZ (CC-IN2P3)

instrumentation of the Linux kernel for collecting data on I/O activity

*development of supervised learning tool for automatic classification of the kind of application (e.g. simulation, reconstruction, analysis, etc.) based on the **observed** I/O profile*

neural network trained with data from 320.000 jobs showed a rate of 96% of precision

preliminary results published at CHEP 2015, software available on line

Analysis of I/O patterns (cont.)

- Complementary approach: synthesised file system to record I/O activity induced by applications

developing a toolbox for tracing I/O activity without superuser privileges, able to run on personal computers or file servers

developed clueFS to collect the set of I/O events generated on the file system by an application and associated tools for exploring the collected data

being used to understand I/O load induced by the data access component of the Large Synoptic Survey Telescope (LSST) software framework

developed a proof-of-concept tool on top of SimGrid to replay the I/O traces in a simulation environment

end-goal: to perform repeated experimentation with multiple parameters of the underlying storage platform, without actually deploying hardware nor even executing the application

work being performed by WANG Cong, IHEP PhD student based at CC-IN2P3 since April 2015 for one year

Cloud-storage

Cloud storage

- Visit of Fabio HERNANDEZ to IHEP in January 2016
- Reinstallation of IHEP's OpenStack Swift evaluation platform
 - latest stable release deployed and configured to use the built-in authorisation component (Keystone)*
- Two target use cases
 - allow applications remote access to IHEP's central data repository, initially in read-only mode*
 - use object stores as emission/reception buffers for bulk inter-site file exchange*
- Transparent remote access
 - transparent access for the application, no matter where it runs: no need for modifications*
 - avoid whole file download to the job execution site: download only the chunks requested by the application*
 - demonstrated that our solution is realistic but needs additional development work for implementing caching strategies*

Cloud storage (cont.)

- Inter-site bulk file exchange

object stores such as Swift to be used as emission/reception buffers for inter-site exchange

to use standard HTTP2 as transport protocol

use case of interest for LSST: 15 TB of raw data per day to be sent from US to France, over 10 years

lessons learned could also be of interest to JUNO

Software-defined networking

Exploration of software-defined networking

- Implement a centrally-managed adaptive virtual private network to exploit available IPv4 and IPv6 network links for data exchange

3 sites testbed: IHEP, Shanghai Jiaotong Univ., Shandong Univ.

China Central Normal Univ. to join soon

commercial partner: Ruijie Networks

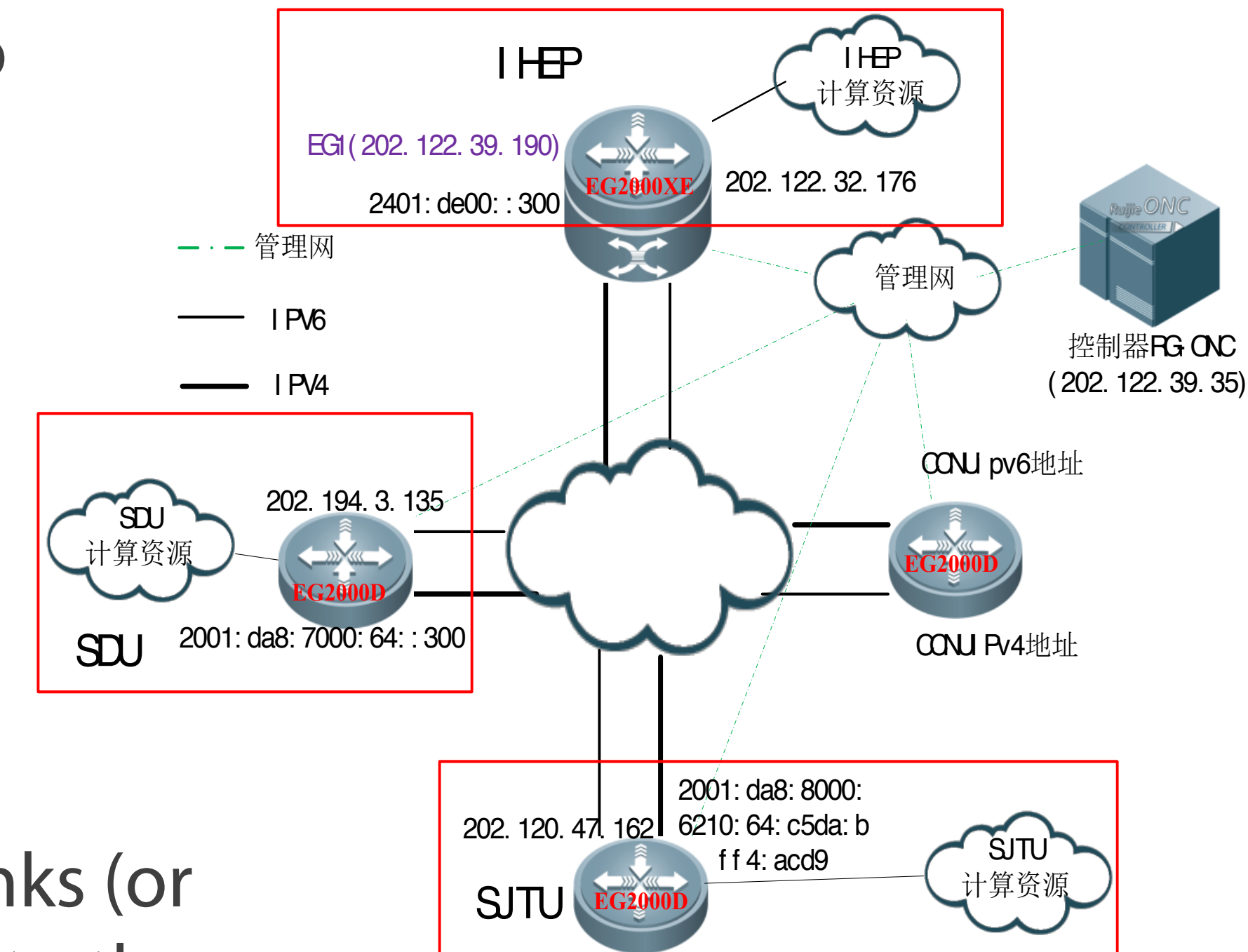
- Chinese universities incentivised to use IPv6 links

usable bandwidth varies greatly, correlated to students activities

- Demonstrator of an adaptive network able to exploit IPv4 or IPv6 links (or both simultaneously) by routing traffic among the sites according to the observed conditions

*decisions on how to route traffic are taken by a **controller application** based on the measured activity of the links*

*controller application dynamically **reprogram** the network equipment to switch from an IPv4 to an IPv6 link according to the congestion of a particular path*



Exploration of software-defined networking (cont.)

- Initial version of a web-based administration interface developed
- Next steps
 - improve path selection algorithm*
 - add intelligence to account for the behaviour of protocol-specific flows (i.e. SSH, HTTP, etc.)*
 - simultaneously exploit the capacity of IPv4 and IPv6 links*
- Scheduled visit of Fazhi QI to CC-IN2P3 to present this project in April

Perspectives

Perspectives

- Project submitted to FCPPL 2016 call
- Ongoing explorations to be continued
- New topics of common interest emerging

exploitation of high-performance computing platforms, involving GPUs, accelerators and low latency interconnection

understanding future data processing needs for JUNO

References

References

X M Zhang, T Yan, X H Zhao, et al, BESIII production with distributed computing, J. Phys.: Conf. Ser. 664, 032036 (2015), <http://iopscience.iop.org/article/10.1088/1742-6596/664/3/032036>

T Yan, B Suo, X H Zhao, X M Zhang, et al, Multi-VO support in IHEP's distributed computing environment, J. Phys.: Conf. Ser. 664, 062068 (2015), <http://iopscience.iop.org/article/10.1088/1742-6596/664/6/062068>

L Wang, J Shi and X Yan , Applying deep neural networks to HEP job statistics, <http://indico.cern.ch/event/304944/session/9/contribution/211>

L Wang, J Shi and X Yan, Applying deep neural networks to HEP job classification, J. Phys.: Conf. Ser. . Proceedings of CHEP2015. 664 (2015), <http://iopscience.iop.org/article/10.1088/1742-6596/664/5/052042>

A Classifier of High Energy Physics Cluster Jobs, https://github.com/wang-lu/HEP_Job_Classifier

clueFS, <https://github.com/airnandez/cluefs>

clueFS tools, <https://github.com/airnandez/cluefs-tools>