

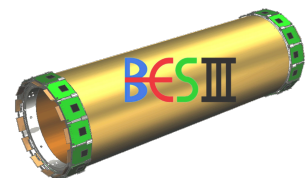
Infrastructures for Distributed Computing: the case of BESIII

Fairness 2017, Sitges, 31 May 2017

Jacopo Pellegrino



- Brief Considerations
- The case of BESIII



- Tools: Cloud Toy
- Facilities: Green Cube



Computing
for HEP

Cloud
Computing in
BESIII

Other
Computing
Efforts

Conclusion

- An Introduction on Cloud
- Cloud Computing in BESIII
- Data Flow
- Job Submission

BESIII



Computing for HEP

- Brief Considerations
- The case of BESIII

Computing in HEP

The number of experiments and the amount of data is increasing in HEP.

International collaborations can take advantage of multiple distributed computing sites:

- Reduce pressure on single site.
- Opportunistic usage of resources.
- Heterogeneous resources.



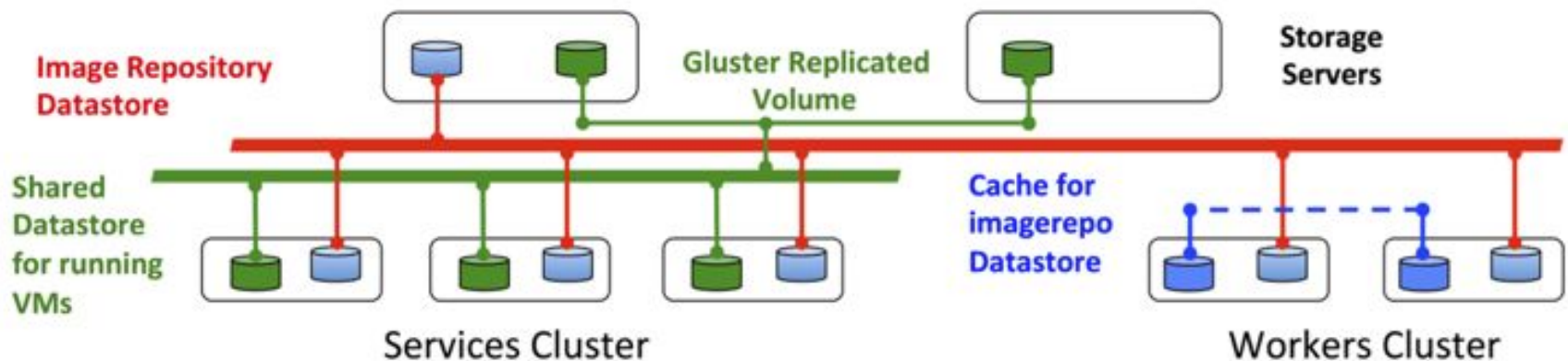
Spread new computing technologies to different sites:

- Develop new tools to optimize the usage of resources.
- Create efficient computing centres.

Computing in HEP

Tasks of computing span over multiple fields:

- Metal management.
- Network definition.
- Replication services.
- Datastore creation.



All these tasks (and more) are performed behind the scenes to provide computational resources to users.

Computing in HEP - BESIII

BESIII data production rate: ~3PB/5year.



To face such computational power requirements:

- Distributed Computing (DC) in IHEP was first built in 2012.
- Put into production in 2014 adding external grid sites:
80% Batch, 20% Grid.
- Cloud adoption in 2015:
now: 10% Grid, 65% Batch, 25% Clouds.

IHEP resources are devoted to multiple experiments:
BESIII, JUNO, LHAASO, CEPC.

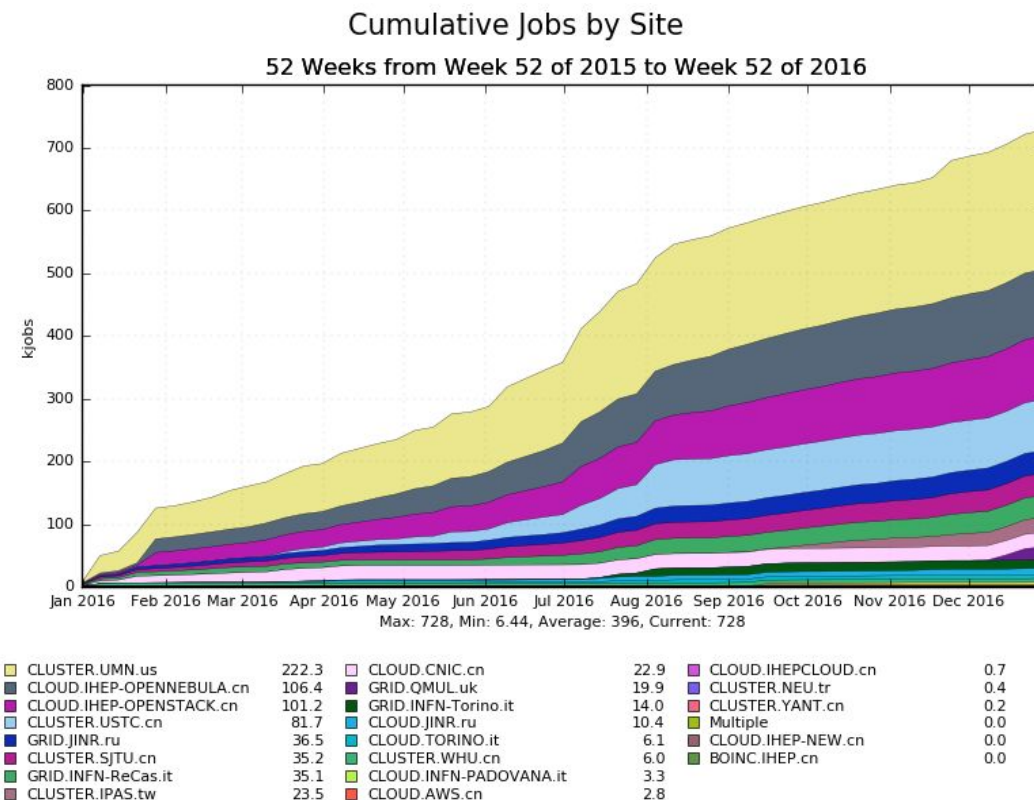
Computing in HEP - BESIII Production

Amount of jobs:

- 728K in 2016.
- 665K in 2015.
- 340K in 2014.

Max running job number:
2K (in 2015).

Job data exchange:
~300TB/year.



Generated on 2017-04-21 07:42:09 UTC

(Xiaomei Zhang, Distributed Computing in IHEP, 2017
Spring Hepix , Budapest)

Cloud Computing in BESIII

- An Introduction on Cloud
- Cloud Computing in BESIII
- Data Flow
- Job Submission

Cloud Computing

Cloud computing is becoming the reference model for HEP computing:

- Is suitable for different collaborations.
- Allows for scalability and elasticity.

The cloud computing paradigm provides access to a shared pool of computational and storage resources.

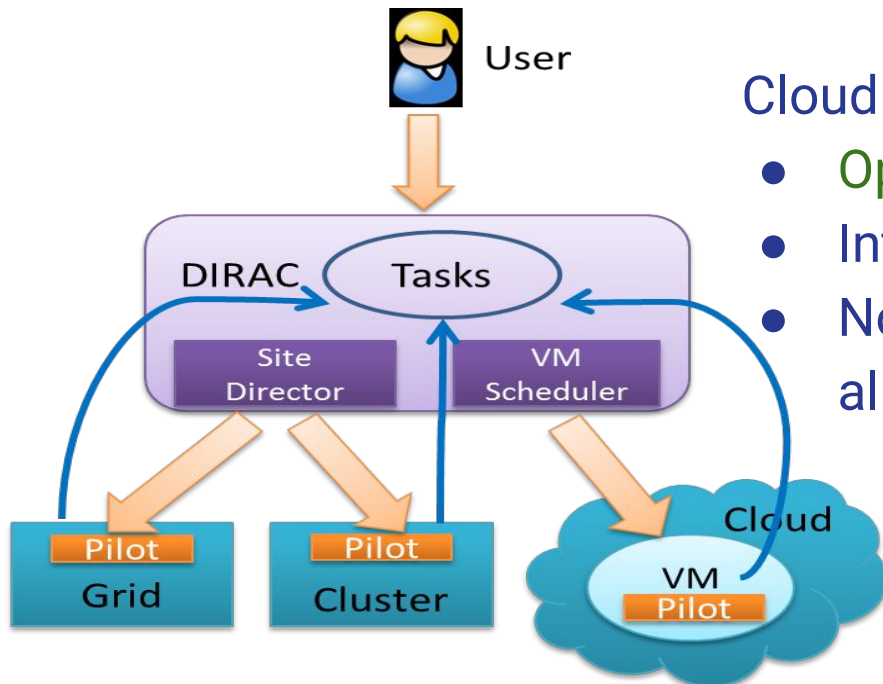
Several BESIII sites are running cloud infrastructures (Torino with OpenNebula).

Setting a cloud infrastructure up may not be trivial by the providers' point of view.

BESIII Cloud - Integration

Cloud **integration** implemented in an **elastic way**:

- Extension VMDIRAC with VM scheduler was introduced.
- VM booted with “pilot” to pull jobs from TaskQueue.
- Cloud-init for VM **contextualization**.



Cloud types supported:

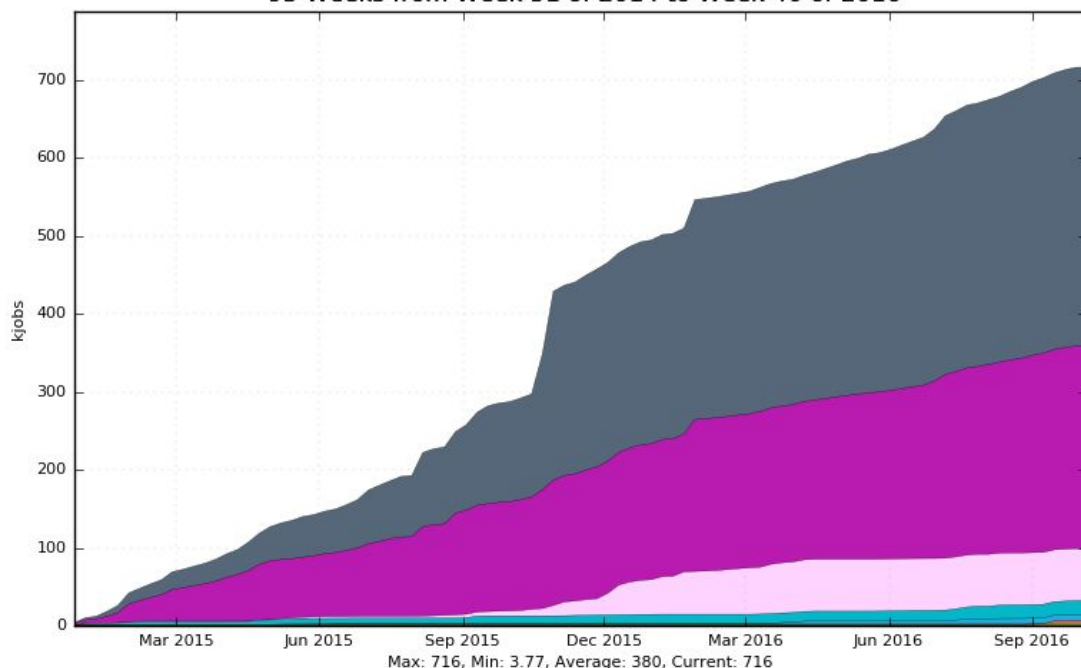
- **OpenStack, OpenNebula, AWS.**
- Interface used: libcloud, rOCCI.
- Not easy to find a general layer to meet all requirements.

Cloud Computing - Integration

Cloud infrastructures play an important role in IHEP DC:

- 6 cloud sites are available from Italy, China, Russia.
- More than 700K jobs have been processed in the past 2 years.

Cumulative Jobs by Site
93 Weeks from Week 51 of 2014 to Week 40 of 2016



~ 5% failure related with VM performance issue (not to the submission infrastructure).

(Xiaomei Zhang, Distributed Computing in IHEP, 2017 Spring Hepix , Budapest)

CLOUD.IHEP-OPENNEBULA.cn	357.2	CLOUD.TORINO.it	17.8	CLOUD.INFN-PADOVANA.it	3.0
CLOUD.IHEP-OPENSTACK.cn	260.0	CLOUD.JINR.ru	7.8	CLOUD.ECCP.cn	0.1
CLOUD.CNIC.cn	66.7	CLOUD.AWS.cn	3.4		

Cloud Computing - Elastic Extension

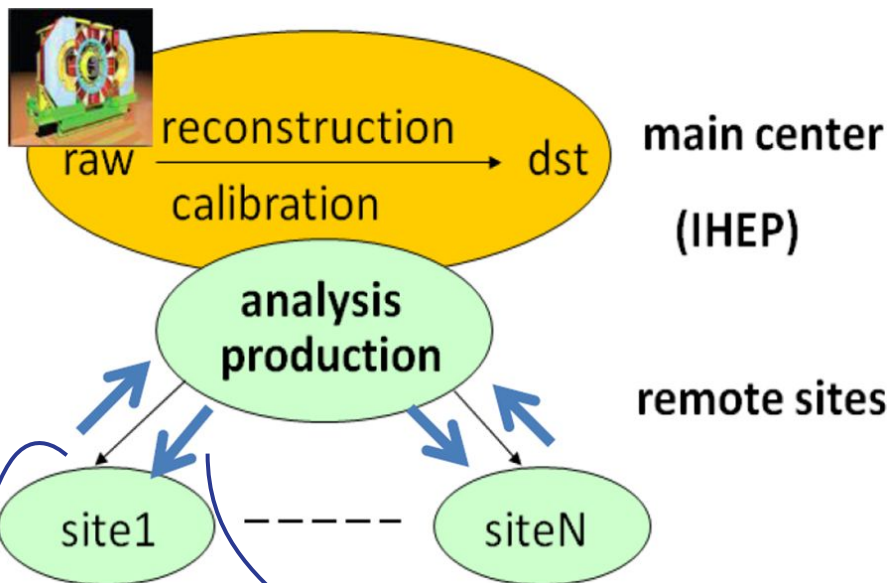
An attempt of elastic **extension to commercial clouds** has been done with the support of Amazon AWS China region:

- **BESIII image** created and uploaded to AWS.
- **VMDIRAC** used in combination with **AWS API** for elastic scheduling.

Tests and Price evaluation:

- 400K BESIII events simulated with 100% success rate with focus on CPU price.
- Proper CPU type need to be chosen: optimization of performance and price.
- **Prices are still not competitive** (Amazon market price in China) compared to self-maintenance.

BESIII Computing Model - Data Flow



IHEP acts as **central site** (raw data processing, bulk reconstruction, analysis).

Remote sites dedicated to MC production, analysis.

(Xiaomei Zhang, Distributed Computing in IHEP, 2017 Spring Hepix , Budapest)

Data are sent from IHEP Central Storage to remote sites for analysis.

Remote sites send back MC data for backup.

BESIII Computing Model - Data Flow

15 overall sites: 8 from China Universities + USA, Italy, Russia, Turkey, Taiwan.

Resources: ~3000 cores, ~500TB storage.



Network Performances:

- 10Gb/s to USA and Europe.
- 10Gb/s to Taiwan and China.

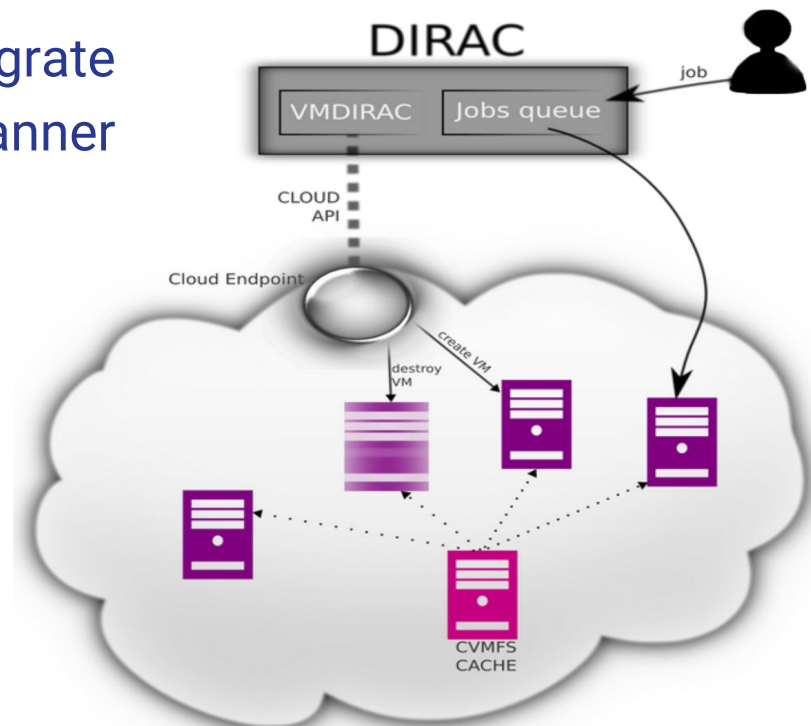
BESIII Computing Model - Job Submission

VMDIRAC is the DIRAC extension to integrate Cloud infrastructures in a transparent manner to the user.

It acts as a **separation layer** between users and **heterogeneous resources**.

Other tools:

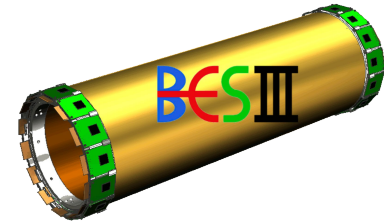
- GANGA and JSUB for massive job submission.
- **CVMFS** to deploy experiment software to remote sites.
- StoRM for management of SE.



(Grzymkowski, R, et al.. "Belle II public and private cloud management in VMDIRAC system." *Journal of Physics: Conference Series*. Vol. 664. No. 2. IOP Publishing, 2015.)

Other Computing Efforts

- Cloud Toy
- Green Cube



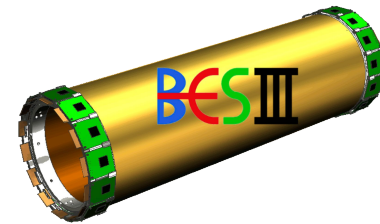
Cloud Toy - The Aim

The main idea behind this work is to address two common issues of cloud infrastructures:

- **Usability**: simplifying the setup and **installation process**. Setting up a **OpenNebula** hypervisor minimizing the user interaction.
- **Efficiency**: making the **usage of resources** more dynamic, flexible and efficient. Allow for a **dynamic** use of resources according to **actual needs**.

The aim is to **improve existing cloud infrastructures** and give sites with limited manpower/knowledge **easier access to cloud technologies**.

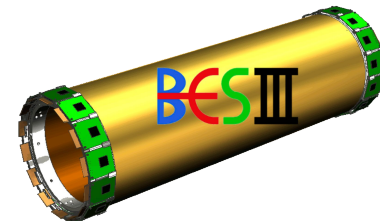




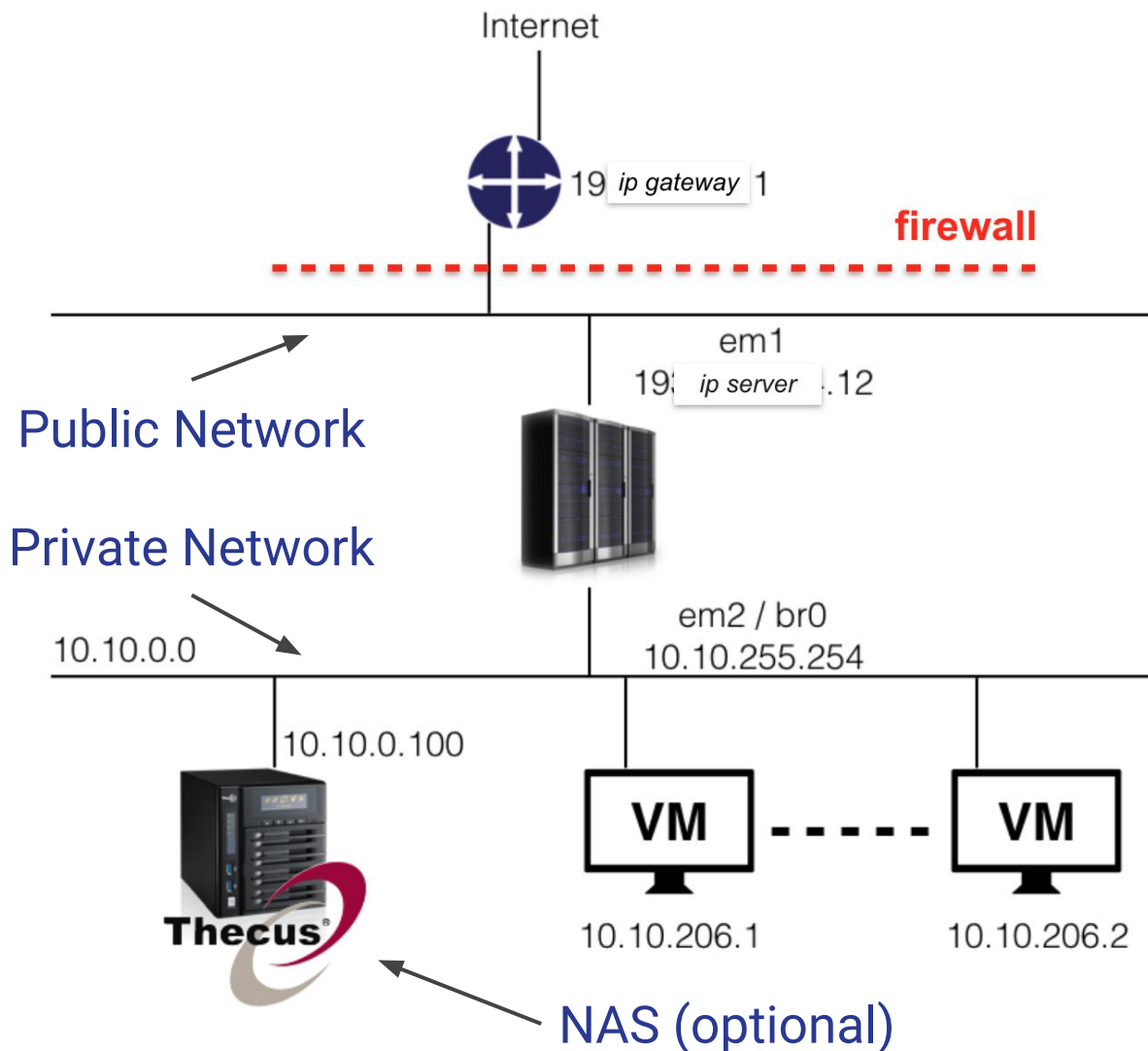
Cloud Toy - Automatic Installation

- Creation of kickstart file: configuration parameters, software packages, disks partitioning, and the like...
- Preparation of customized ISO: start from a standard netinstall iso file, make it look for the kickstart at boot time.
- Preparation of bootable usb drive: burn the iso into a usb drive so that it is possible to boot from it.
- Installation on server: plug the usb drive and reboot.
- Test of the installation: the server is ready in about 1 hour.





Cloud Toy - Automatic Installation



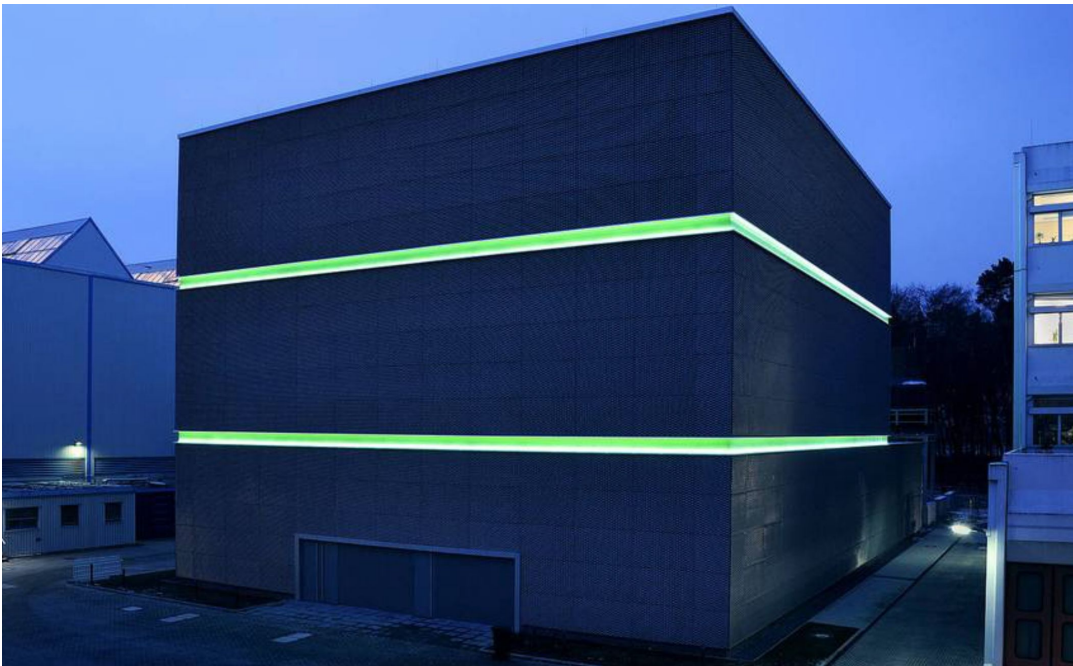
Schematic of the infrastructure resulting from the automatic installation.

Open Ports:

9869 Sunstone.
4567 OCCl.
11443 rOCCl.
29876 proxy VNC.

Computing - Green IT Cube

Large investment (~ 12 million EUR) to build new **high-performance computing centre** Green IT Cube at the GSI Helmholtz Centre for Heavy Ion Research. 27 x 30 x 22 metre cube-shaped building on six floors side by side.



(The Green IT Cube at dawn. (Photo: Gaby Otto, GSI))



(Thorsten Kollegger, Green Cube GSI FAIR Tier 0 2015 AIME Big Data, Budapest)

Green IT Cube

The Green IT Cube received:

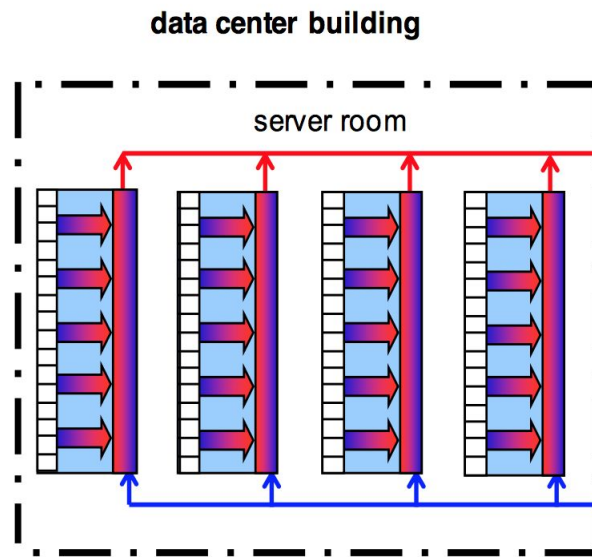
- an international award for the most innovative computing centre at Datacloud 2015.
- *Datacloud Enterprise Datacentre Award*, in recognition of outstanding technical developments.

The infrastructure:

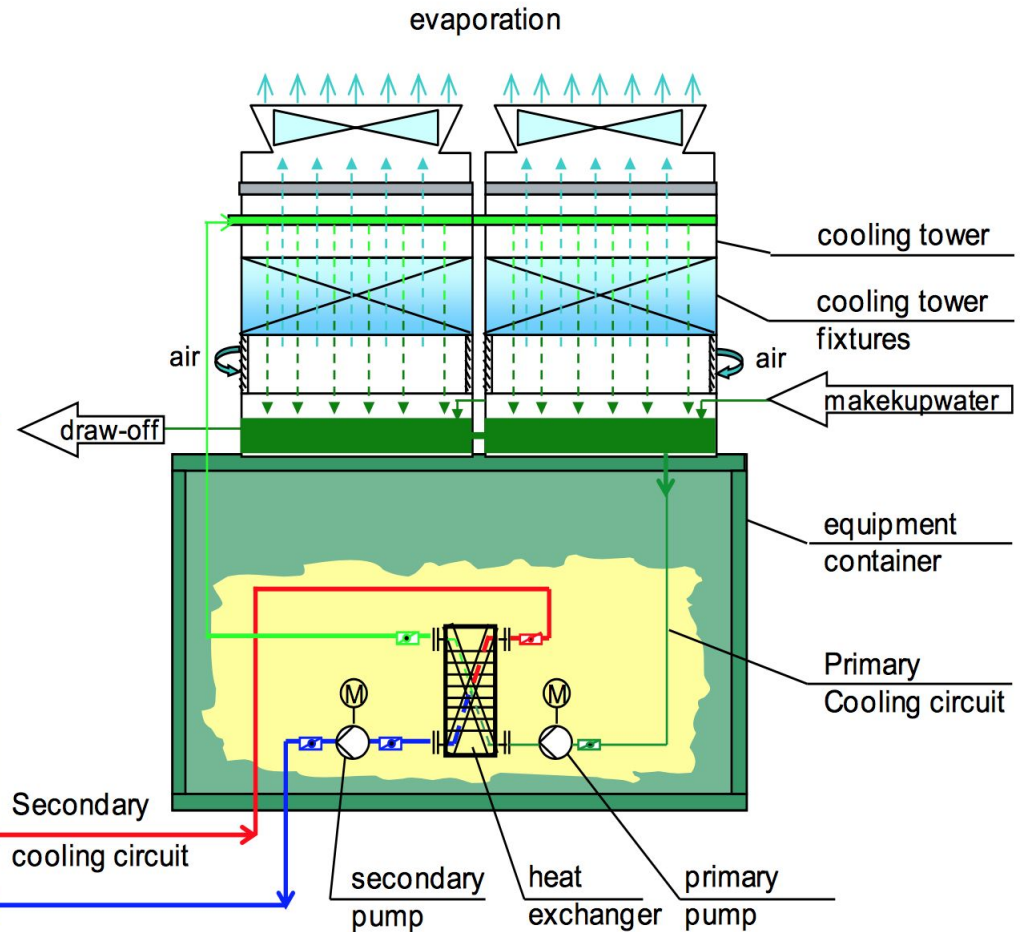
- 768 computer racks with water-cooled system.
- The energy requirement for cooling is ~5-7% of the power required for computing.
- Around 300,000 CPUs.
- 100 petabytes to store experiment data, speed of over one terabyte per second.
- Power Usage Effectiveness = 2 (typical PUE = 1,6 to 3,5).

Green IT Cube

Cooling Overhead
5% of computer power



Works with every
commercial computer



Green IT Cube

Such computing centre will manage data volumes acquired from experiments on the accelerator facilities at GSI and the future accelerator facility FAIR.



CPU + GPU adoption
for heavy ions
tracking.

(Thorsten Kollegger, Green Cube GSI FAIR Tier 0 2015 AIME Big Data, Budapest)

Conclusion

Conclusion

Some general consideration on **computing**: an important aspect for HEP.

An overview of **computing** activities of BESIII at IHEP:

- The overall infrastructure.
- The cloud integration.
- Data flow.
- Job submission system.



Other computing activities:

- **Cloud Toy**: a tool for **spreading** cloud technologies.
- **Green IT Cube**: a futuristic infrastructure.

Question Time