Infrastructures for Distributed Computing: the case of BESIII

J Pellegrino
Universit`a degli Studi di Torino, Dipartimento di Fisica, Via Pietro Giuria 1, Torino, 10125, IT
E-mail: jacopo.pellegrino@to.infn.it

Abstract. The BESIII is an electron-positron collision experiment hosted at BEPCII in Beijing and aimed to investigate Tau-Charm physics. Now BESIII has been running for several years and gathered more than 1PB raw data. In order to analyze these data and perform massive Monte Carlo simulations, a large amount of computing and storage resources is needed. The distributed computing system is based up on DIRAC and it is in production since 2012. It integrates computing and storage resources from different institutes and a variety of resource types such as cluster, grid, cloud or volunteer computing. About 15 sites from BESIII Collaboration from all over the world joined this distributed computing infrastructure, giving a significant contribution to the IHEP computing facility. Nowadays cloud computing is playing a key role in the HEP computing field, due to its scalability and elasticity. Cloud infrastructures take advantages of several tools, such as VMDirac, to manage virtual machines through cloud managers according to the job requirements. With the virtually unlimited resources from commercial clouds, the computing capacity could scale accordingly in order to deal with any burst demands. General computing models have been discussed in the talk and are addressed herewith, with particular focus on the BESIII infrastructure. Moreover new computing tools and upcoming infrastructures will be addressed.

1. Computing for HEP
1.1. Brief Considerations
High Energy Physics (HEP) community is involved in more and more experiments to investigate various subjects. The continuous increase of HEP experiments, together with the progress of measurement tools, leads to a significant growth of the collected data volume and therefore of the amount of computational power required to analyze them within computing sites distributed all over the world.

This trend has led as well to a significant progress in the computing infrastructures required to process, analyze and store large amounts of data. In this work, the focus is on distributed computing, an approach adopted by multiple sites (research labs, universities, ect) which allows for the reduction of the computational pressure on a single site by means of an opportunistic usage of resources made available by other sites. Such process is possible thanks to heterogeneous ways of providing and accessing computational resources between different sites. The term distributed computing requires multiple aspects that a site has to accomplish in order to share its computational resources. These tasks span from the management of the physical devices (hardware) to the definition of public and private networks, storage, and replication services in order to provide a reliable infrastructure to users.
Figure 1. Example of services provided by a computing center.[1]

Computing centers provide and maintain multiple services, the most common ones are depicted in Figure 1.

In order to maintain the sites up to date it is important, on one hand to spread new computing technologies to multiple sites developing tool to optimize the usage of resources, on the other hand it is necessary to think about and build new and more efficient computing centers. Examples of efforts in this direction will be provided respectively in Section 3.1 and 3.2.

1.2. The case of BESIII
BESIII is a spectrometer hosted on the BEPCII $e^+e^-$ collider, situated at the Institute of High Energy Physics (IHEP) of Beijing, and dedicated to the investigation of the Tau-Charm Physics. BESIII has collected around 3PB of data in the last five years. In order to face the computational effort Distributed Computing (DC) was first introduced at IHEP in 2012 and entered in production in 2014, adding external sites (80% batch computing farm, 20% grid). Later in 2015 cloud technology has been adopted as well and gained a significant role involving nowadays around 25% of the resources (the remaining 75% being composed of 10% grid and 65% batch)[3]. Moreover, IHEP computing resources are also provided to other experiments such as JUNO[4], LHAASO[5], and CEPC[6].

The distributed computing of BESIII faced an increasing amount of jobs since its adoption:
- 340k in 2014;
- 665K in 2015;
- 728K in 2016 as shown in Figure 2,

with a maximum number of running jobs of about 2k in 2015. The amount of job data exchange between the various sites is about 300TB/year. These numbers indicate how efficient the DC of BESIII is and they offer an idea of the effort needed to set up and maintain such an complex and distributed infrastructure.

2. Cloud Computing in BESIII
As discussed in Section 1.2, cloud computing is becoming the reference model for HEP computing. The rationale behind this trend can be understood by pointing out a couple of features of this approach: first of all, cloud computing is particularly suitable for different collaborations since, as it will be further discussed in Section 2.1, it allows for provisioning heterogeneous resources in a transparent way from the users’ point of view. Moreover it allows for scalability and elasticity according to the actual needs of a site providing access to a shared
pool of computational and storage resources. Please note, setting a cloud infrastructure up may be not trivial by the providers point of view. A possible solution will be discussed in Section 3.1.

The adoption of cloud computing in BESIII not only involved IHEP but also six other sites such as the JINR in Dubna (Russia) or Torino (Italy) where an OpenNebula-based[7] cloud infrastructure has been prepared. The integration of the cloud approach with preexisting infrastructures has been implemented in an elastic way by means of VMDirac[8], a tool to integrate federated clouds in a transparent manner to the user. Thanks to this system jobs are sent to a VMDirac instance that checks if there are enough resources to create a new VM. If so, a pilot job is sent to switch the VM on, which will execute the job and, after that, can be eventually powered off if no further jobs are submitted. The output is anyway transferred the user before the VM is powered off.

Overall, more than 700K jobs have been processed in the past two years by the cloud sites of BESIII, thus indicating the relevance of cloud computing in the BESIII DC.

At the IHEP site three cloud managers are currently supported: OpenStack[9], OpenNebula, and AWS (Amazon Web Service)[10]. Among the others libcloud[11] and rOCCI[12] are the interfaces used to bind the different cloud infrastructures to the submitting infrastructure in a transparent way. Nevertheless finding a purpose middle-ware to meet all the different cloud-specific requirements remains a non trivial task. While OpenStack and OpenNebula are set up and maintained in site, AWS has been adopted as an attempt of elastic extension to commercial clouds, with the support of Amazon AWS China. The BESIII images were created and uploaded to AWS and VMDirac has been used in combination with AWS API for elastic scheduling.

---

**Figure 2.** Cumulative jobs by site in 2016[3].
Several tests revealed that, at the moment, self-maintenance is more convenient compared to the adoption of AWS[3].

2.1. Data Flow and Job Submission
As anticipated in Section 2, several sites provide a cloud infrastructure to BESIII. Among them the IHEP acts as central site providing tools for raw data processing, bulk reconstruction and data analysis. Remote sites are mainly dedicated to MC (Monte Carlo) production (simulation and reconstruction) and, if possible, analysis of data retrieved from IHEP Central Storage. Once the MC or the analysis has been performed, the remote sites send back to IHEP data for backup purposes[3].

The exchange of data also involves non-cloud sites for a total of fifteen sites: eight of them from China Universities and the remaining ones from USA, Italy, Russia, Turkey and Taiwan. Such a data traffic is hosted on the through 10Gb/s network linking either USA, Europe and Taiwan to China.

As far as job submission is concerned, either IHEP or remote cloud sites provide their resources. This tool is the DIRAC extension designed in order to integrate cloud infrastructures to preexisting infrastructures, and allows to perform remote operations on VMs. It acts as a separation layer between heterogeneous resources and users. Among the other tools it is worth to mention GANGA and JSUB for massive job submission, CVMFS to deploy experiment software on remote sites and StoRM for management of the Storage Elements (SE).

3. Developments in Computing
Apart from the BESIII DC, it is important to mention a couple of examples that indicate how relevant distributed and cloud computing is in the HEP context. A tool for cloud infrastructures named Cloud Toy[15] and a new computing facility named Green Cube[16] will be discussed in the following paragraphs.

3.1. Tools: Cloud Toy
The main idea behind this work is to address two common issues of cloud infrastructures: usability by simplifying the setup and installation process of an OpenNebula hypervisor minimizing the user interaction, and efficiency by making the usage of resources more dynamic, flexible and efficient allowing for an optimal and dynamic use of resources according to actual needs. The aim is to improve existing cloud infrastructures and give sites with limited manpower or cloud specific knowledge an easier access to cloud technologies[15].

The automatic installation takes place through a set of steps as follows: first of all the so-called kickstart file is created where configuration parameters, software packages, disks partitioning information are located. Then, a customized ISO file is prepared starting from a standard netinstall file in order to make it look for the kickstart at boot time. The ISO file is then burnt into a usb drive and installed on the server, which is ready in about one hour.

In this way a ready to use cloud infrastructure could be installed with a very limited user interaction. Such tool is currently under test by BelleII[13] groups and soon will be extended to BESIII[14] Distributed Computing as well.

3.2. Facilities: Green Cube
The Green IT Cube, hosted at the GSI Helmholtz Centre for Heavy Ion Research, is the result of a large investment to build new high-performance computing center. It is a 27 x 30 x 22 meter cube-shaped building on six floors side by side which hosts 768 computer racks with a water-cooled system for an overall of around 300000 CPUs and 100 PB of storage, capable to face a data I/O of over one TB per second[16]. Moreover GPUs will be widely adopted to
improve parallel processing, required especially when treating heavy ions collisions (i.e. data from CBM experiment at FAIR). The cooling system is the flagship of the infrastructure: due to its features the energy requirement for cooling is between 5 and 7% of the power required for computing[17].

4. Conclusion
This work aimed to give a brief report of the computing activities of BESIII at IHEP and in remote sites. The overall infrastructure has been presented with a particular focus on the cloud technology and its integration within the preexisting infrastructure. An overview about the data flow process and the job submission system has been given as well.

Among the others, two relevant computing activities have been presented: the Cloud Toy, a tool for spreading cloud technologies, and the Green IT Cube, a futuristic infrastructure, now under construction at FAIR.

5. Acknowledgement
The research activity of this article is funded by the EU and INFN by means of the project BESIICGEM RISE 645664 within the call H2020-MSCA-RISE-2014, and is supported in part by the NSFC project 11375221.

References
[1] Bagnasco S, Berzano D, Brunetti R, Lusso S, Vallero S, Managing a tier-2 computer centre with a private cloud infrastructure 2014 J. Phys: Conf. Series 523 012012
[2] BesIII Experiment, webpage http://bes3.ihep.ac.cn
[3] Zhang XM, Distributed Computing in IHEP 2017 Spring Hepix (Budapest, Hungary)
[4] JUNO Jiangmen Underground Neutrino Observatory, webpage http://juno.ihep.ac.cn
[5] ZHEN C, A future project at tibet: the large high altitude air shower observatory (LHAASO) 2010 Chinese physics C 34.2: 249
[6] Circular Electron Positron Collider, webpage http://cepc.ihep.ac.cn
[7] Montero S, Moreno-Vozmediano R and Llorente I M 2012 IaaS Cloud Architecture: From Virtualized Datacenters to Federated Cloud Infrastructures Computer, vol. 45, no. 12, pp. 65-72 doi:10.1109/MC.2012.76
[8] VMDirac, webpage https://github.com/DIRACGrid/VMDIRAC/wiki
[9] Sefraoui O, Aissaoui M, Eleuldj M OpenStack: toward an open-source solution for cloud computing 2012 Int. J. of Computer Applications 55.3
[10] AMAZON, E. C. Amazon web services, website: http://aws.amazon.com
[11] Apache Libcloud, website https://libcloud.apache.org
[12] Edmonds A, Metsch T and Papaspyrou A Open Cloud Computing Interface in Data Management-related Setups 2011 Springer Grid and Cloud Database Management pp. 127
[13] Grzymkowski R, Hara T and computing group B 2015 J. Phys: Conf. Series 664 022021
[14] Zhang XM, Yan T, Zhao X H, Ma Z T, Yan X F, Lin T, Deng Z Y, Li W D, Belov S, Pelevanyuk I, Zhemchugov A and Cai H 2015 J. Phys: Conf. Series 664 032036
[15] Astorino F, Amoroso A, Bagnasco S, Balashov N A, Bianchi F, Destefanis M, Maggiora M, Pellegrino J, Yan L, Yan T, Zhang X and Zhao X, A modular (almost) automatic set-up for elastic multi-tenants cloud (micro)infrastructures, Proc. Int. Conf. on Computing in High Energy and Nuclear Physics, (San Francisco, USA)
[16] Green IT Cube, webpage, http://www.fair-center.eu/news-events/news-view/article/green-it-cube.html
[17] Kollegger T, Green Cube GSI FAIR Tier 0 2015 AIME Big Data (Budapest, Hungary)