PROCEEDINGS OF SCIENCE

Jiangmen Underground Neutrino Observatory computing requirements and infrastructure

Giuseppe Andronico*
INFN - Sezione di Catania, IT
E-mail: giuseppe.andronico@ct.infn.it

Zhang Xiaomei
IHEP, CN
E-mail: zhangxm@ihep.ac.cn

Li Weidong
IHEP, CN
E-mail: liwd@ihep.ac.cn

The Jiangmen Underground Neutrino Observatory (JUNO) is an underground 20 kton liquid scintillator detector being built in the south of China and expected to be completed by the end of 2021. The JUNO physics program is focused on exploring neutrino properties, by means of electron anti-neutrinos emitted from two nuclear power plants at a baseline of about 53 km. Targeting an unprecedented relative energy resolution of 3% at 1 MeV, JUNO will be able to study neutrino oscillation phenomena and determine neutrino mass ordering with a statistical significance of 3-4 sigma within six years running time. These physics challenges are addressed by a large Collaboration localized in three continents. Different groups of the Collaboration, as simulation and offline groups, have started the evaluation of the requirements of the experiment for computing and the related resources. In this context, key to the success of JUNO will be the realization of a distributed computing infrastructure, which will satisfy its predicted computing needs. Upon its establishment, it is expected to deliver not more than 3 PB of data per year, to be stored in at least four data centers in China and Europe. Data analysis activities will be distributed in a joint effort. This contribution is meant to report how the JUNO computing infrastructure is going to be designed and which will be its main characteristics.

XXIX International Symposium on Lepton Photon Interactions at High Energies - LeptonPhoton2019
August 5-10, 2019
Toronto, Canada

*Speaker.

PoS(LeptonPhoton2019)122

	Description	Data producer
Central detector	Design, material choice, optimization, deployment	No
Veto detector	Set of detectors to exclude uninteresting signals	Yes
Liquid scintillator	Design, production and management of liquid scintillator	No
Calibration system	JUNO calibration system, several sources, mechanical part design and implementation	Yes
Large PMT Electronics	Design, test and production of LPMT and electronics Wait trigger signal to write to disk	Yes
Small PMT system	Design, test and production of SPMT and electroncs Wait trigger signal to write to disk	Yes
Trigger	Select meaningfull physical events	No
Online event reconstruction	Reconstruct event from data passed trough the firewall and select events to write on disk	No
Offline	Framework, reconstruction, analysis, and all it is offline	No

(a) Which group in JUNO is going to produce data.

Raw data	upper limit 2 PB/year
Calibration data	0.6 PB/year
Rec data	order of 200 TB/year
Sim data	order of 100 TB/year
Analysis data	order of TB/year
Total upper limit	3 PB/year

(b) Several items contributing to JUNO yearly data volume.

Table 1: JUNO data producers and volumes.

JUNO experiment

The JUNO collaboration is made from 77 partners in 17 countries covering 3 continents. Collaboration's partners take part in several groups, working on different aspects of the challenging experiment, and interacting between them. In Table 1a is a simplified summary, able to put in evidence the components of the experiment that are going to producing data to be analyzed.

Computing model

The starting point of our computing model is to estimate the amount of data produced from the experiment. JUNO, as told by its name, is an observatory where several types of physics are studied. The main types of physics and the event rate expected in JUNO are summarized in Figure 1. Correlating the expected events number with the expected JUNO behaviour, it has been

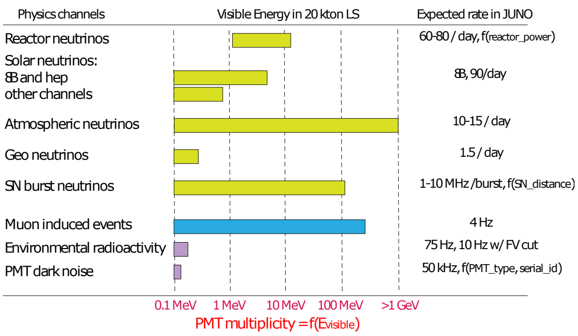


Figure 1: Event rate for some physics activities. SN burst is just in case of supernova happening.

possible to estimate the data rate. At this point, from the estimate about trigger and event filtering efficiency, we get an indication on how much data we will need to store, as reported in Table 1b. In total, this means a yearly data production of the order of 3 PB. The estimated computing power to handle this amount of data is about 12,000 cores. The data will be produced in Kaiping, the

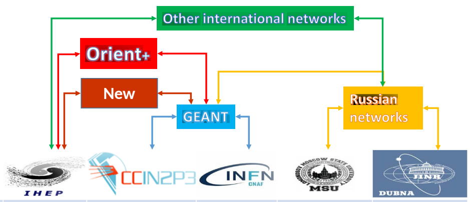
JUNO experimental site, and transferred through a 1 Gb/s devoted network connection to IHEP, in Beijing, where they will be stored. To ensure data safety at least a backup is required, and some European partners candidates to host it. Then, making use of international connection between National Research Networks, data are to be copied from IHEP to European data centers. It is required to have a file catalog and a book keeping system to keep trace of files copies.

Simulation, reconstruction and analysis are based on the software framework. The JUNO framework is based on standard libraries as ROOT, Geant4, CLHEP and on SNiPER [6], a software framework developed at IHEP.

Distributed infrastructure

Given the distributed nature of JUNO collaboration, it is quite natural to implement a distributed infrastructure to fulfill the computing model. At the moment, the IHEP computing center is a sort of T0 for JUNO experiment, receiving directly the data from Kaiping.

Other four computing centers are available in Europe: CC-IN2P3 from France, INFN-CNAF from Italy, Moscow State University and the computing center at JINR, in Dubna. In Figure 2 there are summarized details on the data centers participating and on networks connecting them.



Kaiping (JUNO)	1 Gb/s				
Generic	10 Gb/s	10 Gb/s	20 Gb/s	10 Gb/s	10 Gb/s
LHCOne	yes	Yes	100 Gb/s	Yes	10 Gb/s
DC Storage (PB)	5+15	70+30	80+33	n.a.	11+12
DC Computing (core)	18,000	38,000	40,000	n.a.	17,000
Storage (disk)	500 TB	8 TB	230 TB	60 TB	25 TB
cores	888	230	250	32	10

Figure 2: Data centers participating in JUNO distributed infrastructure. In the table, blue rows are relative to network, green row is relative to entire data center, orange rows are resources devoted to JUNO.

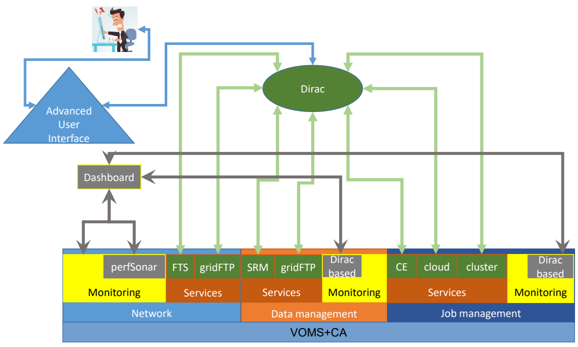


Figure 3: A simplified constituent block view of services for JUNO distributed infrastructure.

To share these resources and balance the load in order to form a distributed infrastructure, we developed a design and choose a set of services. In Figure 3 a component view description of our design is reported. The basic requirement is to be able to identify and authorize users. In our design we rely on the standard solution of Virtual Organization (VO) and of Virtual Organization Membership Service (VOMS)[1], that make use of digital certificates issued from trusted Certification Authorities. In VOMS it is possible to define groups and roles so that services and data are made accessible only to allowed people.

To ensure that the infrastructure is working properly a monitoring system with a dashboard is needed. Of fundamental importance, to ensure distributed infrastructure is working properly, are the network connections between the sites. On top of this we can find the services to replicate and move data around: **gridftp**[2], an enhanced version of FTP enabling security and parallel streams and **File Transfer Service (FTS)**[4], able to handle file transfer requests and properly schedule them.

What is required to the distributed infrastructure is to manage our data, replicate them around and analyze. This is done by means of the other two blocks in the design, Data management and Job management.

To distribute JUNO software, the infrastructure relies on **CernVM File System** (CernVM-FS) [3] providing a scalable, reliable and low-maintenance software distribution service. Data management is a set of services and protocols interacting with the storage resources. One of the key services is the Storage Resource Manager, in charge of managing files, as locating and retrieving copy of files.

Job management, instead, is a set of services aimed at submitting jobs, as data analysis, and managing these jobs.

In JUNO design an important role is delegated to **Distributed Infrastructure with Remote Agent Control** (DIRAC) [5], a software framework for distributed computing. In JUNO design, DIRAC provides user interface, some data management services, as File Catalog[1], and job management services.

Distributed infrastructure status

From January 2019, a working group was established. It is composed by JUNO members and representative of data centers involved to work on the distributed infrastructure design and implementation. Till today several parts were installed and tested:

monitoring perfSONAR installed and dashboard operational

VO JUNO VO created, VOMS installed and configurated, VOMS replica deployed

SRM data centers SRM configured for JUNO and some test data transfer already performed

job submission first test successfully performed.

Soon, the infrastructure will be ready for real functionality tests.

[1] File Catalog provides interface to locate files and their copies

References

[1] R. Alfieri, R. Cecchini, V. Ciaschini, L. dell'Agnello, Á. Frohner, A. Gianoli, K. Lõrentey, and F. Spataro. Voms, an authorization system for virtual organizations. In Francisco Fernández Rivera, Marian Bubak, Andrés Gómez Tato, and Ramón Doallo, editors, *Grid Computing*, pages 33–40, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[2] William E. Allcock. Gridftp protocol specification. 2002.

[3] Jakob Blomer, Carlos Aguado-Sánchez, Predrag Buncic, and Artem Harutyunyan. Distributing LHC application software and conditions databases using the CernVM file system. *Journal of Physics: Conference Series*, 331(4):042003, dec 2011.

[4] Ákos Frohner, Jean-Philippe Baud, Rosa Maria Garcia Rioja, Gilbert Grosdidier, Rémi Mollon, David Smith, and Paolo Tedesco. Data management in EGEE. *Journal of Physics: Conference Series*, 219(6):062012, apr 2010.

[5] A Tsaregorodtsev, N Brook, A Casajus Ramo, Ph Charpentier, J Closier, G Cowan, R Graciani Diaz, E Lanciotti, Z Mathe, R Nandakumar, S Paterson, V Romanovsky, R Santinelli, M Sapunov, A C Smith, M Seco Miguelez, and A Zhelezov. DIRAC3 – the new generation of the LHCb grid software. *Journal of Physics: Conference Series*, 219(6):062029, apr 2010.

[6] Jiaheng Zou, Xingtao Huang, Weidong Li, Tao Lin, Teng Li, Kun Zhang, Ziyan Deng, and Guofu Cao. Sniper: an offline software framework for non-collider physics experiments. *Journal of Physics: Conference Series*, 664:072053, 12 2015.

PoS(LeptonPhoton2019)122