Mock Data Challenge for the JUNO experiment

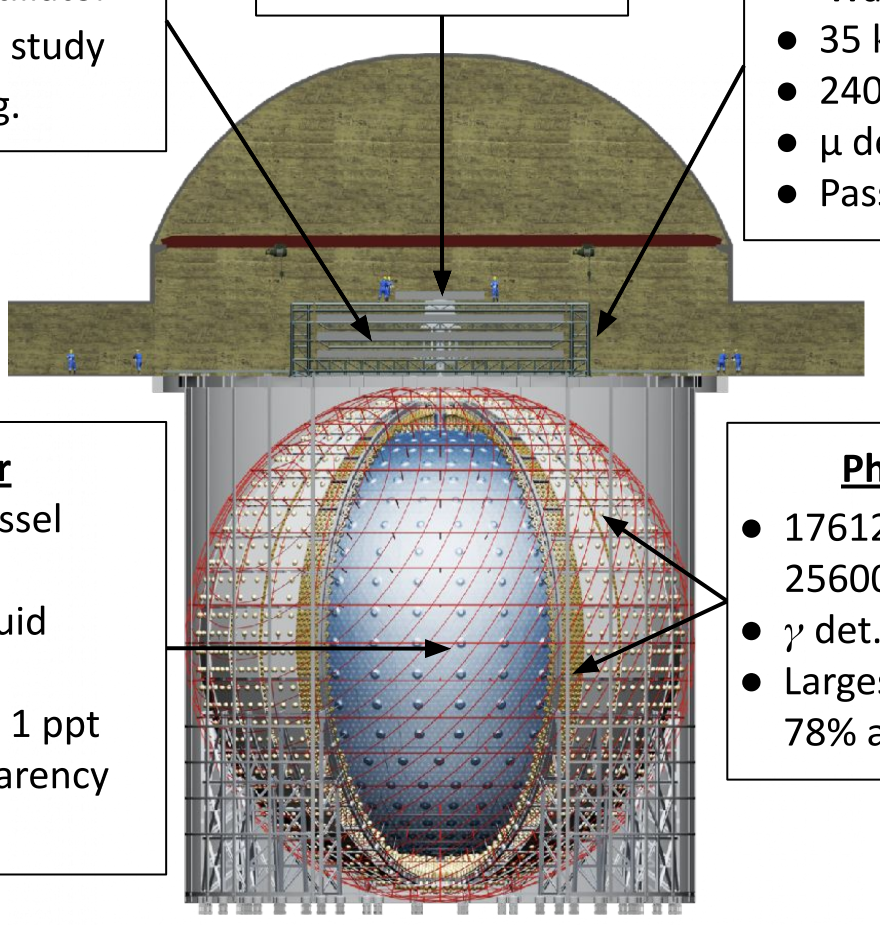
Alessandra Carlotta Re[1], C. Jiang[2], T. Lin[3], R. Li[4], J. Xu[3], Z. You[5] on behalf of the JUNO collaboration
1. Università degli Studi & INFN, Milano (Italy); 2. Nanjing University, Nanjing (China); 3. Institute of High Energy Physics, Beijing (China);
4. Shanghai Jiao Tong University, Shangai (China); 5. Sun Yat-sen University, Guangzhou (China).

The Jiangmen Underground Neutrino Observatory (JUNO) is a large liquid scintillator detector located in China and it is expected to start data taking in 2024. JUNO's ambitious design primarily aims to the determination of the neutrino mass ordering at high statistical significance (3-4 σ) in about 6 years of data taking, by measuring the oscillation pattern of electron antineutrinos generated by two nuclear power plants, on a ~53 km baseline from the experimental site. Moreover, JUNO aspires to the sub % determination of the neutrino oscillation parameters as well as to the measurement of atmospheric neutrinos, to solar neutrino precision spectroscopy, and to the detection of low energy neutrinos coming from supernovae and geo-neutrinos.
JUNO thus will be a unique facility for particle and astroparticle physics.

Top Tracker
- 3 layers of plastic scintillator
- Precise µ tracking and study of the cosmogenic bkg.

Calibration house

Water Pool
Water Cherenkov detector
- 35 kton ultrapure water
- 2400 20'' PMTs
- µ detection eff. > 99.5%
- Passive shield.

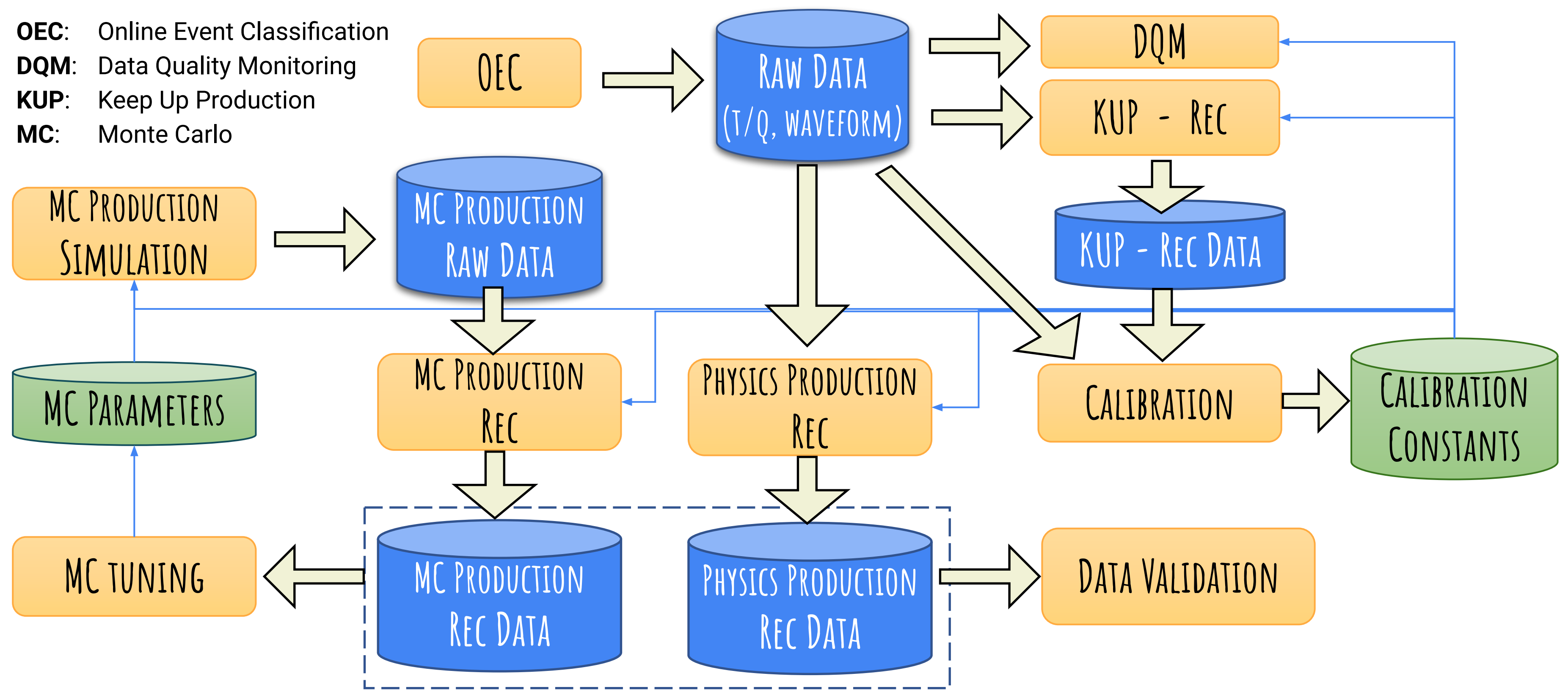


Central Detector
Acrylic spherical vessel
- 35.4 m diameter
- 20 kt of LAB based liquid scintillator
- Radiopurity: U/Th/K < 1 ppt
- Expected Light transparency in LS > 96%.

Photomultipliers Tubes
- 17612 - 20'' PMTs
  25600 - 3'' PMTs
- γ det. efficiency ~ 25-30%
- Largest PMT coverage to date: 78% active surface.

The JUNO Data Production workflow and the Mock Data Challenge

While awaiting for the JUNO detector real data, the Mock Data Challenge wants to qualitatively test the goodness of the JUNO's Data Production chain as for CPU time/disk-space.



RADIOACTIVE BACKGROUND: SIMULATED EVENTS and SPECIES assuming 1h of data-taking

Dataset Name	Number of Events	Job Sets		File Size	
		Numbers	Events	single file	Total
U238@LS(x13)	13,000,000 events	20	65000	2.45G/20.6M	444.34 G
Th232@LS(x9)	10,000,000 events	20	50000	2.40G/16.94M	412.18 G
K40@LS	1,000,000 events	20	5000	198M/1.57M	37.59 G
Pb210@LS(x3)	3,000,000 events	30	10000	185M/3.24M	50.65 G
C14@LS	1,000,000,000 events	2000	50000	146M/13.9M	2.65 T
Kr85@LS	1,000,000 events	20	5000	72M/1.53M	14.83 G
U238@Acrylic(x13)	130,000,000 events	200	65000	135M/17.7M	257.81 G
Th232@Acrylic(x10)	100,000,000 events	200	50000	190M/15M	322.94 G
K40@Acrylic	10,000,000 events	20	50000	130M/12.3M	24.82 G
U238@node/bar(x13)	1,300,000,000 events	2000	65000	23M/17M	776.06 G
Th232@node/bar(x10)	1,000,000,000 events	2000	50000	25M/14M	756.09 G
K40@node/bar	100,000,000 events	200	50000	19M/11.8M	56.93 G
Co60@node/bar	100,000,000 events	200	50000	131M/13.4M	263.47 G
U238@PMTGlass(x12)	13,000,000,000 events	20000	65000	17.7M/17M	6799.18 G
Th232@PMTGlass(x10)	10,000,000,000 events	20000	50000	14M/14M	5304.71 G
K40@PMTGlass	1,000,000,000 events	2000	50000	17M/12M	574.13 G
Rn222@WaterRadon	100,000,000 events	2000	35000	17M/9.50M	523.60 G

Besides the radioactive background simulations size (~ 19TB), we also expect about 200 TB from muon simulations.

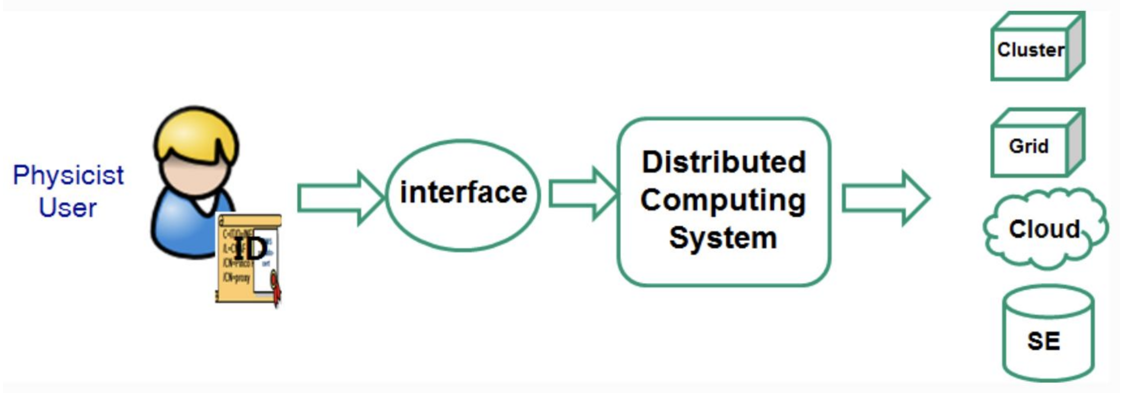
The size and the number of files of the simulated data are quite large to handle: there are too many input samples for the reconstruction and OEC.

We need a strategy:

1. **MERGE** We merge the files for a species
2. **PRE-MIX** We pre-mix background samples of different species
3. **RECONSTRUCT** OEC of the pre-mixed sample

The JUNO Distributed Computing Infrastructure

The best strategy to successfully deal with the computing needs of the JUNO experiment, is to rely on a Distributed Computing Infrastructure (DCI). Within the JUNO collaboration, computing resources are currently provided from several data centers around the world: IN2P3 (France), IHEP (China), JINR (Russia), CNAF (Italy), and SDU (China). The DCI integrates heterogeneous resources, hide complexity from users, and provides a simple way for users to use the computing resources.



JUNO COMPUTING EXPECTED NEEDS

Data flow: 60MB/s
DAQ event rate: 1 kHz
Raw data file size (avg, on 80 k events): 5 GB
Data volume per day: ~ 5 TB
Data volume per year: ~ 2PB
Number of files per day: ~ 1000
Number of files per year: ~ 400000

ACAT 2022
The 21st International Workshop on Advanced Computing and Analysis Techniques in Physics Research
24-28 October, 2022 – Bari (Italy).