

# Data Production AFG: Introduction and Mock Data Challenge

Alessandra Carlotta Re

alessandra.re@mi.infn.it

Università degli Studi & INFN Milano (Italy)

05 May 2022

## **Data Production AFG - Scope**

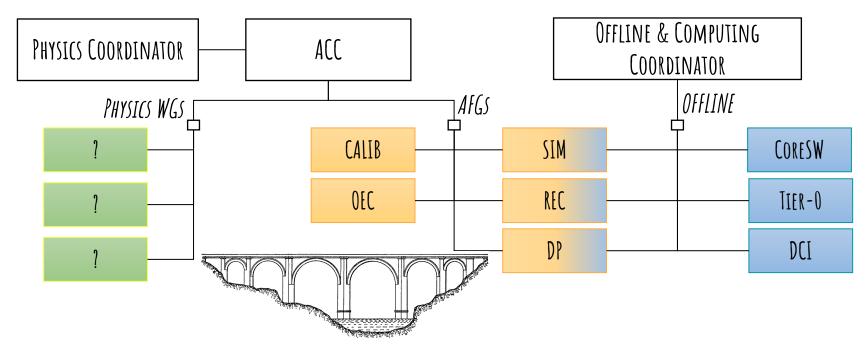
The **Data Production Analysis Foundation Group** (DP-AFG) was established by the Analysis Coordination Committee (ACC) on 18/11/2021.

https://juno.ihep.ac.cn/mediawiki/index.php/Analysis Foundation Groups/Data production

### Work scope:

- Design the production data flow and provide necessary software tools;
- Manage the Monte Carlo generation and real data production, as well as the data access;
- Data quality check, including the onsite data quality monitoring and the offline data monitoring;
- Cooperate with TierO computing and the distributed computing infrastructure (DCI) groups to manage the computing resources;
- Plan the production with ACC;
- Provide efficient and user-friendly web service for users to look up data and information of production (DataBase!).

### AFG? PWG? OFFLINE? What is their relation?\*



The Data production AFG should be the bridge between Physics analysis & Computing as well as a shared group between AFGs and Offline

# The Mock Data Challenge

<u>PURPOSE</u>: the Mock Data Challenge wants to test the goodness of all the JUNO's software chain from a qualitatively/time side.

For each species of interest (backgrounds/signals) we already converged on the number of events to be generated.

**STATUS:** The Mock Data Challenge has already started (backgrounds only) but there isn't an "official" timeline.

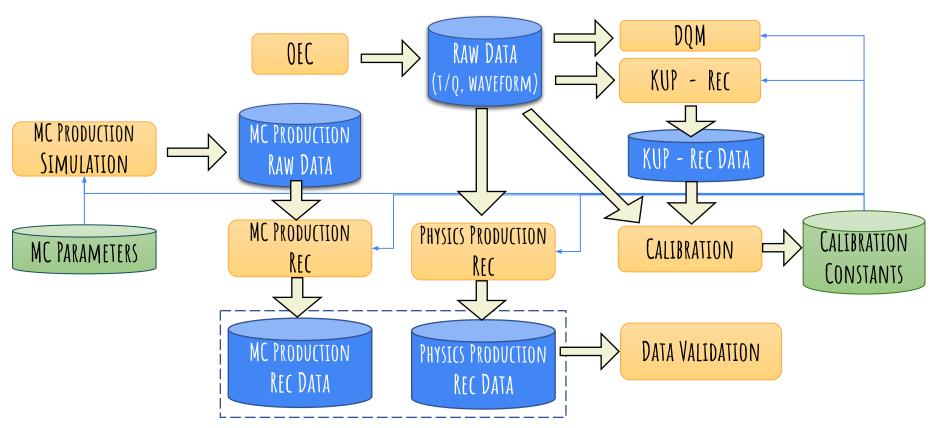
It is divided into several stages, each consistent with a software release.

The J22.1.0-rc0 focus on the update of detector simulation: it's ready and it's beeing used to produce the detector simulation samples.

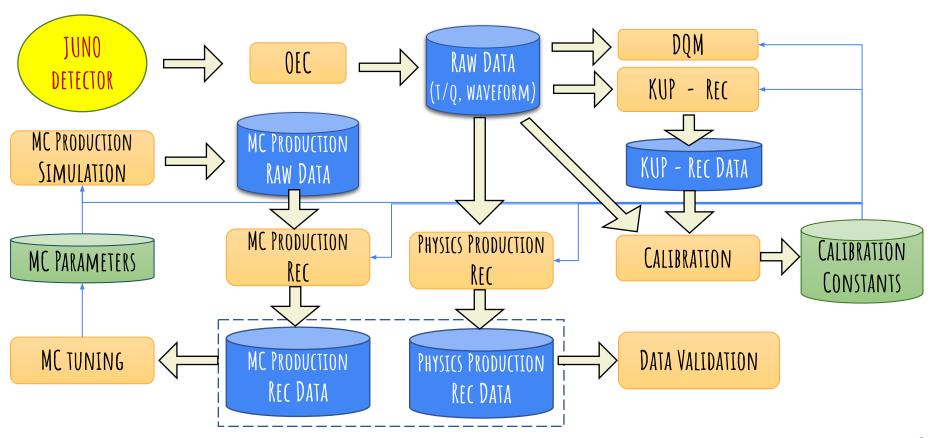
The new release (rc1) will focus on the update of EDM, electronics simulation and OEC. As there are still some problems found, the production of this stage is not started yet. The next-new release (rc2) will focus on the update of reconstruction algorithm.

https://juno.ihep.ac.cn/mediawiki/index.php/Analysis Foundation Groups/Data production/Mock Data Challeng

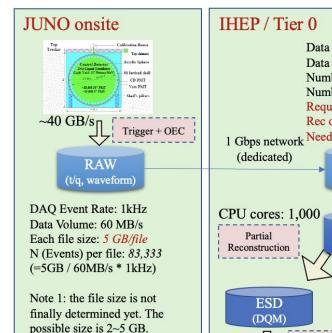
# The Mock Data Challenge



### The full chain of the Data Production



# **Data Production: Real Processing**



Note 3: Onsite mini DQM as an alternative if transfer lost

Note 2: the calibration data is

not considered here.

Data Volume per day: 5.2 TB/day Data Volume per year: 2 PB/year Number of files per day: ~1,000 Number of files per year: ~400,000 Requirements of rec speed: < 5s/event Rec of one file takes: ~115 hours (5 days) 1 Gbps network Need multi-threading to reduce the time. Baseline: 8 cores **RAW** (t/q, waveform) CPU cores: 5,000 **RTRAW** Reconstruction (old calibration) **ESD** IAD (KUP) (KUP) Offline Data Data Quality Monitoring Monitoring Delay: ~hour Delay: ~half day

JUNO DCI / Tier 1 Requirements: one year data needs to be reconstructed within 2 months. RAW (t/q, waveform) Reconstruction (new calibration) Indexing **IAD ESD** (PP) (PP)

JUNO Data Monitoring Web

Physics Analysis

# Manpower

#### **CONVENERS:**

- Tao Lin, IHEP
- Zhengyun You, SYSU
- Alessandra Carlotta Re, INFN-MI

#### **OTHER MEMBERS:**

- Benda Xu (Faculty, Tsinghua) Monte Carlo data validation
- Sicheng Yuan (PhD student, IHEP) Keep-up data production
- Kaixuan Huang (PhD student, SYSU) Data Quality Monitoring
- Cailian Jiang, Jilei and Rui (MasterStudents?)
- Someone from Livia's group

