CMS Tier-0 data processing during the detector commissioning in Run-3

CHEP2023: 26th International Conference on Computing in High Energy & Nuclear Physics

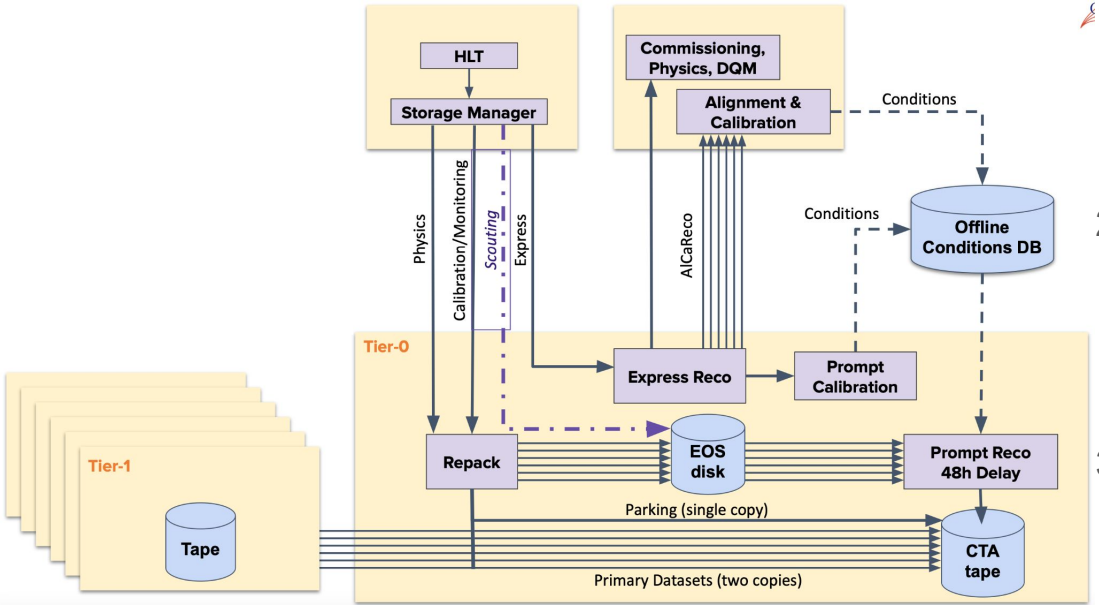
Jhonatan Amado
On behalf of the CMS Tier-0

May 8th - 12th 2023

Overview

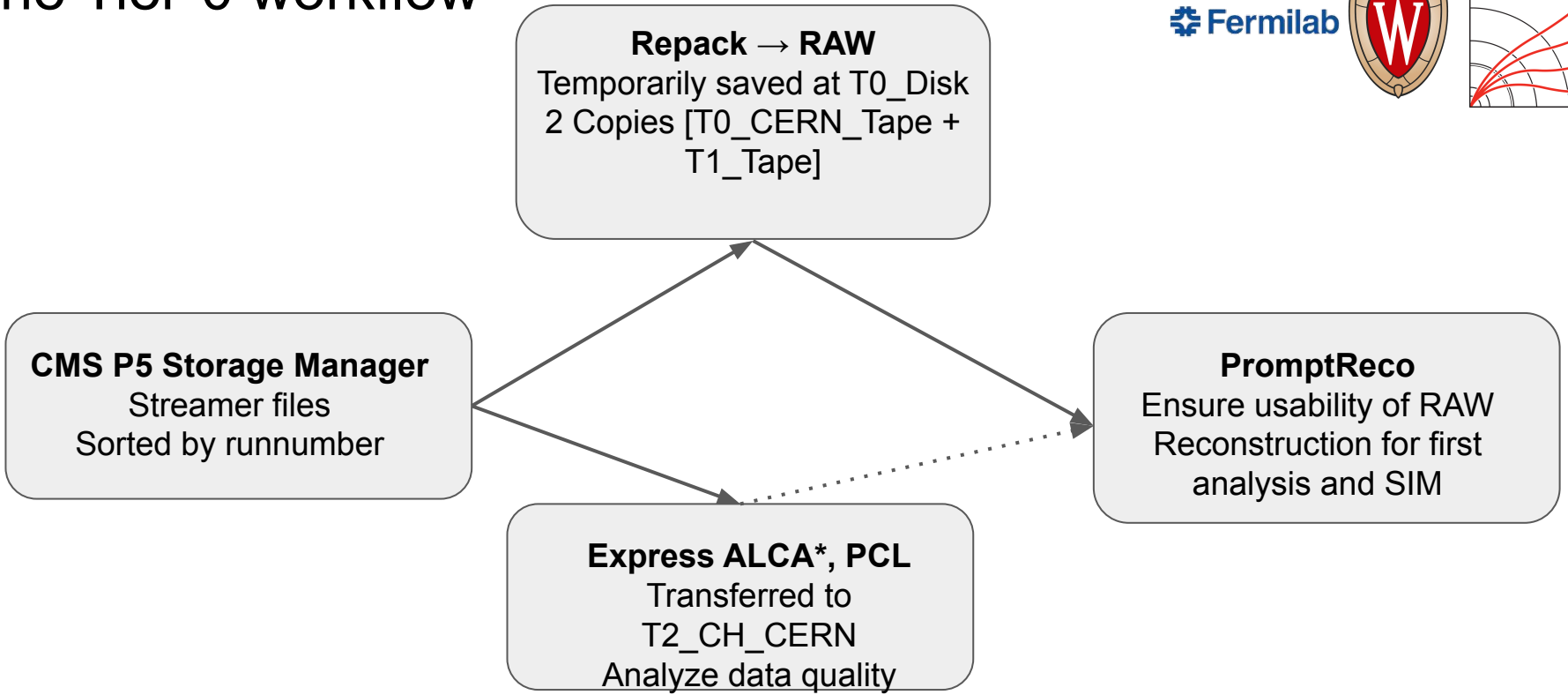
1. Operation and services of CMS Tier-0.
   a. What is CMS Tier-0 and the services we provide.
2. Tier-0 Upgrades during the LHC long shutdown 2.
   a. Hardware/software upgrades.
3. Tests during commissioning activities and real data taking stats
   a. Input/output data from Tier0.

The Tier-0 workflow



1. Repack streamer files
   a. CMS data primary datatier is RAW.
   b. Transform data from the HLT in columnar ROOT data
   c. Enforce 2 copies of RAW data on Tape (CERN_Tape + T1_Tape).
2. Express processing
   a. Calibrate the detector and provide info of data quality.
   b. Expected to be delivered shortly after data taking.
   c. Transferred to T2_CH_CERN (Will talk later [slide 6]).
3. Prompt Reconstruction
   a. Reconstruct all new data to ensure that RAW data is usable and deliver first data for analysis and simulations.
   b. 48h delay after RAW data is produced. Can be delayed if its needed.
   c. Transferred to T1_Disk and T1_Tape in function of Primary Datasets.

The Tier-0 workflow



- Streamer files are kept until RAW is generated + delay in deletion 7d / function of rates delivered by LHC.
- RAW is kept on T0_CERN_Disk until 2 tape copies are secured + Prompt has finished.
- PromptReco data is transferred to T1_Disk/Tape. PromptReco data is kept on T0_CERN_Disk until tape/disk copies are secured.

Tier-0 services during daily operations

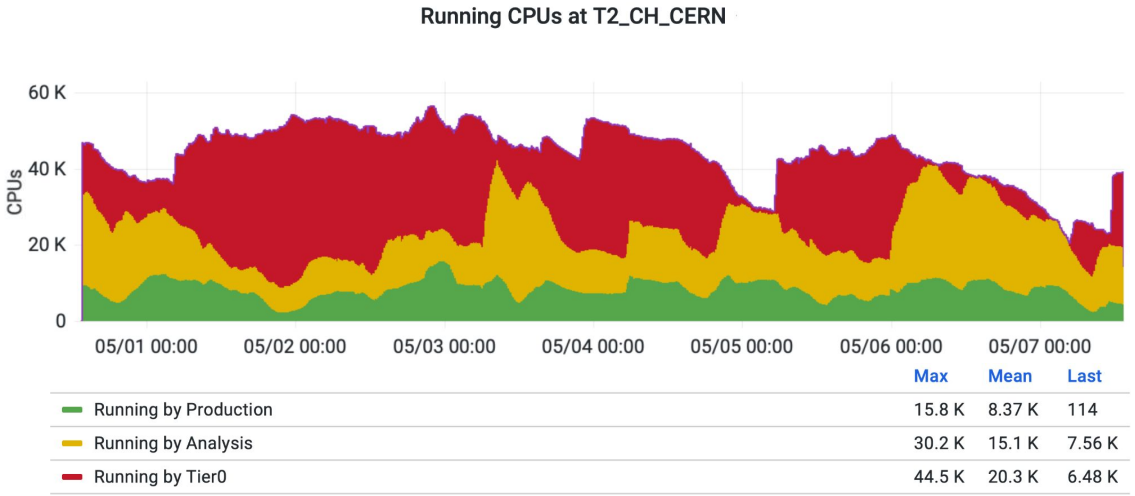
- **WMCore, WMAgent** Workflow Management System
  - Collection of libraries and components that provide a grid workload management system.
- **SI** Submission Infrastructure
  - Manage the major part of CMS CPU resources (HTCondor).
- **CMSSW**.
  - CMS Software.
- **Data Management**
  - Group in charge of control/distribute all the data between the different CMS sites.
- **StorageManager**
  - System which write data to the disk at P5.
- **Alignment and calibration (ALCA)**
  - For calibration of physics stream. Updated condition for a give run.

Tier-0 upgrades during LS2

- **Resource re-organization at CERN**
  - Tier-0 shares storage and computational resources with CERN_Disk.

    Run-3 CERN_Disk Storage = (T0_CERN_Disk+T2_CH_CERN)

  - Not having control over Tier-0 storage resources is a source of risk for T0 operations. (data loss risk).
  - Creation of T0_CERN_Disk with site configuration according to T0 needs.
  - Still share computational resources with T2_CH_CERN.
    - T0 has priority submitting jobs. Total ~ 40k CPUs. Used ~ 20k during commissioning and early Run-3
  - T0 manage quota at CERN.



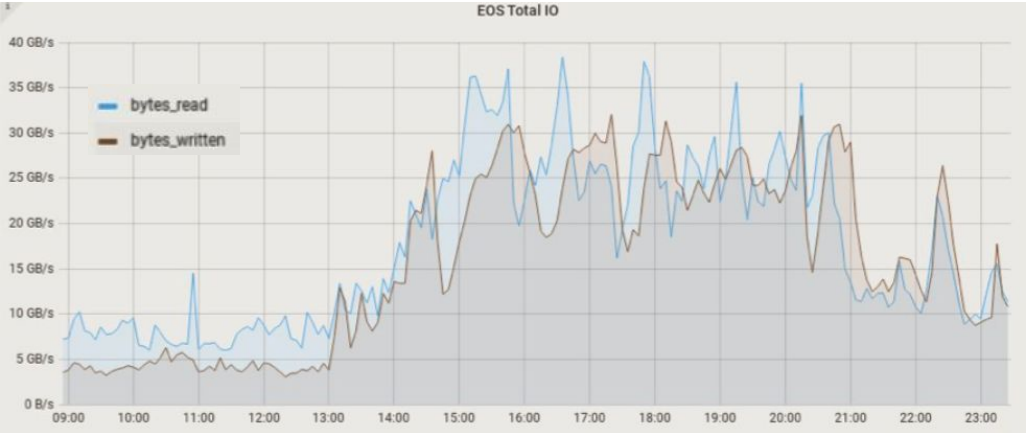
Tier-0 upgrades during LS2

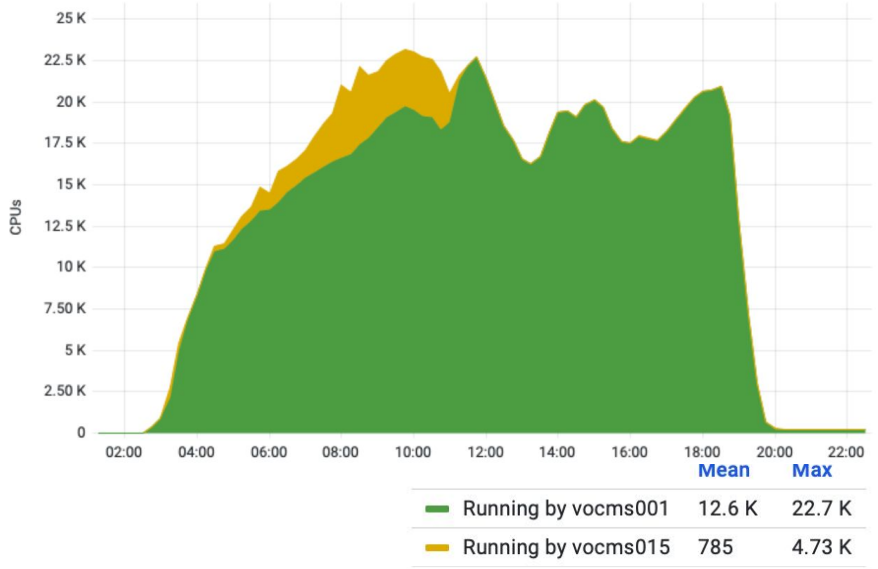
- **Python 3 migration**
  - **WMCore** WMAgent (Py2&Py3) was validated and run in production on Nov 2020.
  - **CMSSW** CMS Offline Software released CMSSW_12_0_0 on Sept 2021.
  - **Tier-0 (Py3)** released on Sept 2021.
    - Used in production on Oct 2021.
- **PhEDEx to Rucio**
  - Transition deadline was Nov 30th 2020. T0 started to use it on Oct12th 2020.
  - No downtime in production.
  - Dedicated reaper (deletion tool) for T0_CH_CERN_Disk.
- **Automated replays**
  - Replay. Integration test. (WMAgent-CMSSW-Tier-0)
  - During operations, T0 has to deploy an average of 4 replays per week.
  - In order to speed-up those integration tests, replays can be triggered by Github-PR.

Tier-0 upgrades during LS2

T2_CH_CERN batch system now using SSDs

- Increased performance running repack jobs
  - Use all worker nodes instead of 1 job per 8-core slot.
  - SSDs greatly improved performance of repacking steps
- Tier-0 stress Test
  - Total of ~ 400TB
  - ~ 20k CPUs
  - Time ~ 10h
  - EOS ~ 30GB/s
    - Conclusion: Fully saturated I/O bandwidth offered by EOS

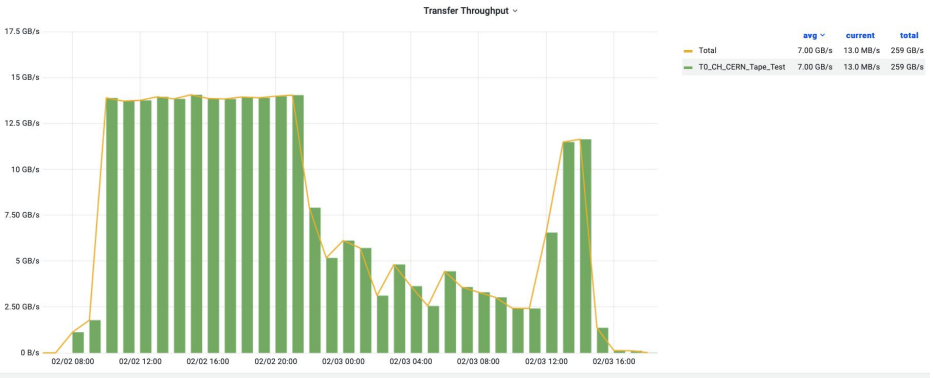




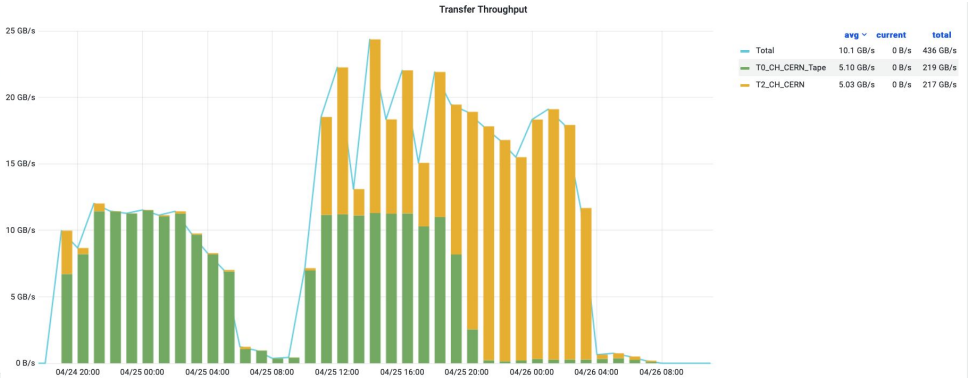
Tier-0 upgrades during LS2

Increased in network capacity at CERN

- During tape performance test (36h). Max 14GB/s.
- During collisions 2023. Overall Max ~ 25GB/s.



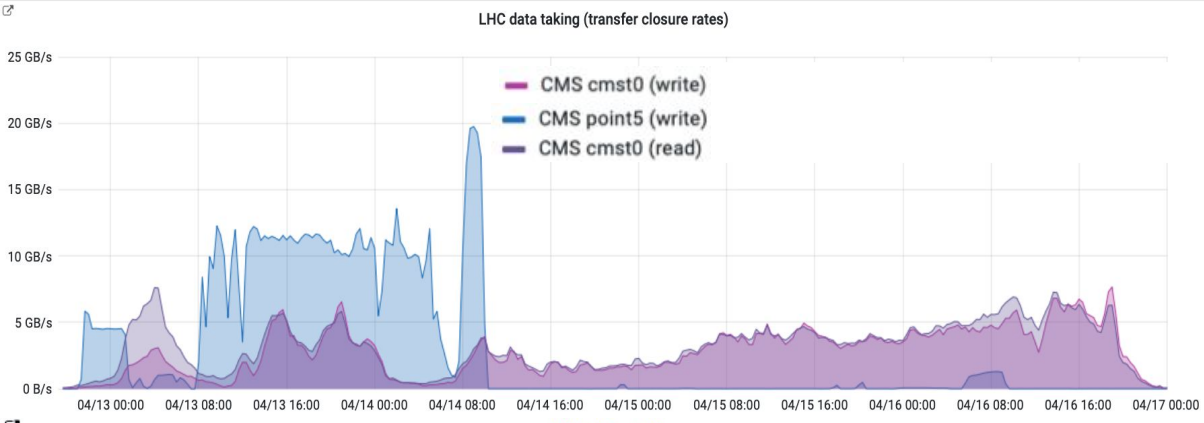
Test performance

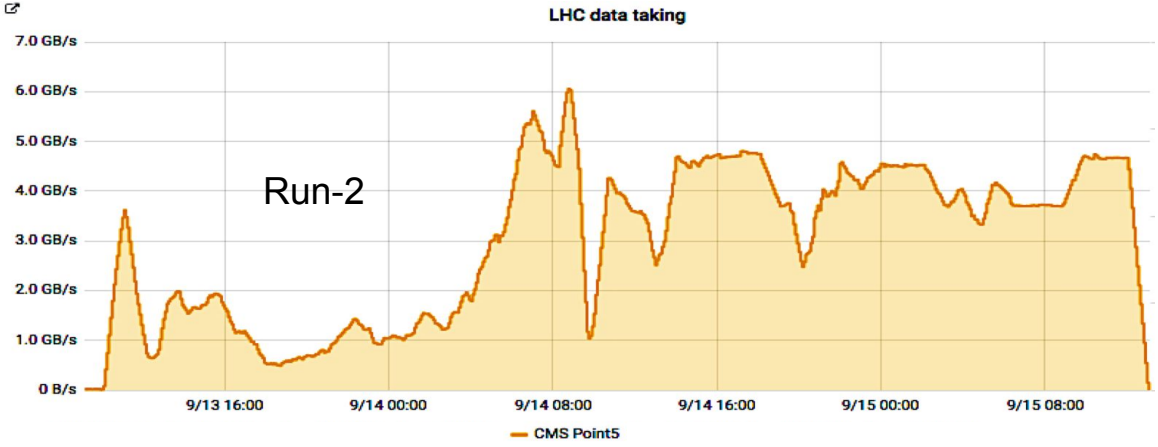


Collisions 2023

Tier-0 upgrades during LS2

Run-3

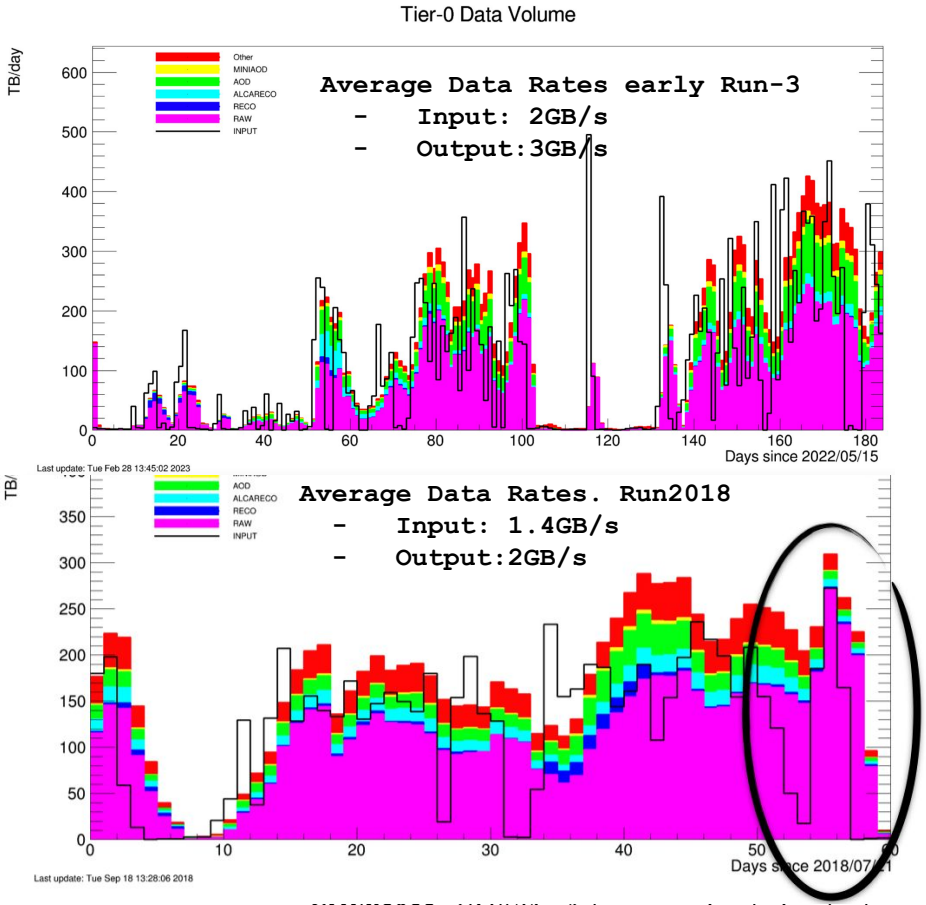




Improvement in storage manager

- New storage hardware for CDAQ
  - Higher data-taking rates.
- Commissioning outline
  - Take data up to 13GB/s without delays in transfers.
  - Take data up to 17GB/s. Transfers will not keep up and backlog is transferred during the interfills.

Performance in early Run-3 vs Run-2 (2018)



1. Average data rates during commissioning activities before Run-3.
   a. Input 2GB/s Output 3GB/s.
   b. Weekly rates during commissioning were 1.5x compared to Run-2

2. Peak during test activities

Summary

- Tier-0 operation.
  - Operates 24/7, with peaks of demand during continuous data taking. 24/5 human support.
  - No delays in initiating data processing for new runs. Downtimes are coordinated with Run coordination.
  - Overall transfers performance and capacity to reconstruct all the data is adequate.
- Upgrades.
  - All hardware upgrades improved overall the system and reliability of Tier-0.
  - Software upgrades took place without downtime in production.
  - We can process data from P5 at higher rates compared to Run-2 (2kHz at peak VS 5kHz sustained)
- Experience after commissioning for Run-3.
  - Weekly data rates are x1.5 times respect Run-2 and T0 still able to repack,reconstruct and deliver good quality data with no delays.

Thanks