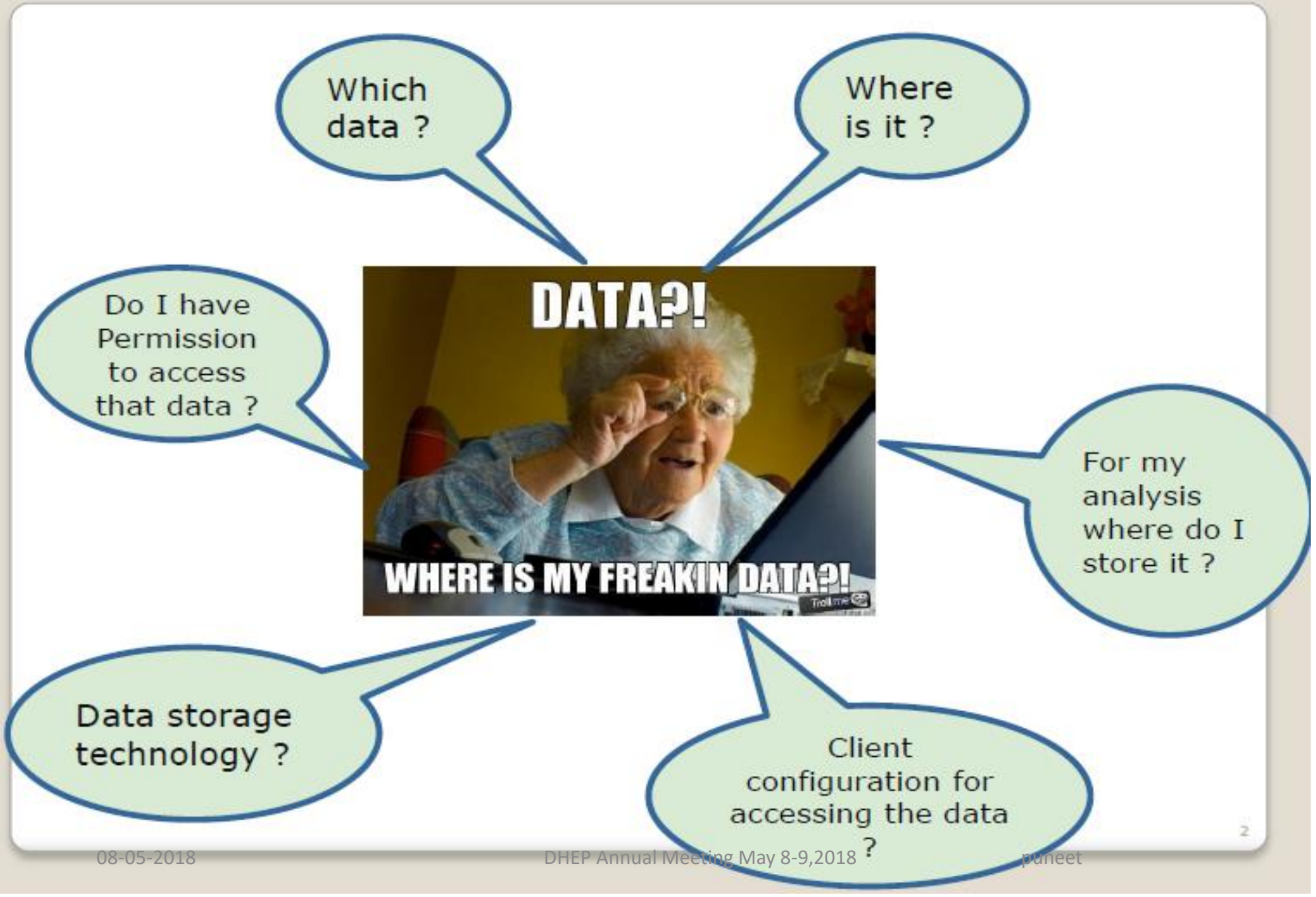
Tata Institute of Fundamental Research
टाटा मूलभूत अनुसंधान संस्थान

# CMS Grid Computing Facility at TIFR

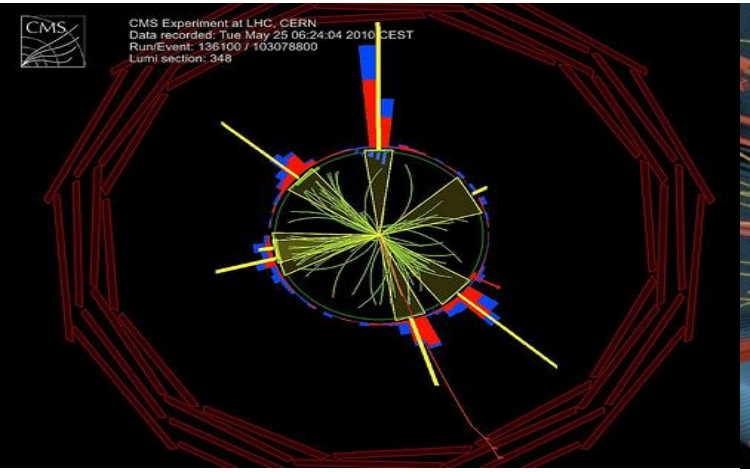
Puneet Patel, Brij Kishor Jashal,
Kajari Mazumdar, Gobinda Majumder

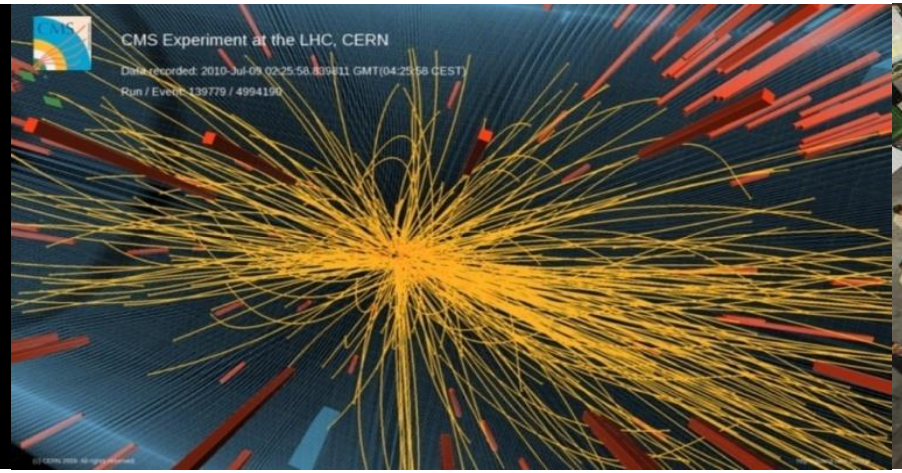



Grid provides the solution to these problems

# The GRID Computing Goal

- Science without borders.
- Provide **Resources** and **Services** to store/serve O(10) PB data/year
- Provide **access** to all interesting physics events to ~4000 collaborators



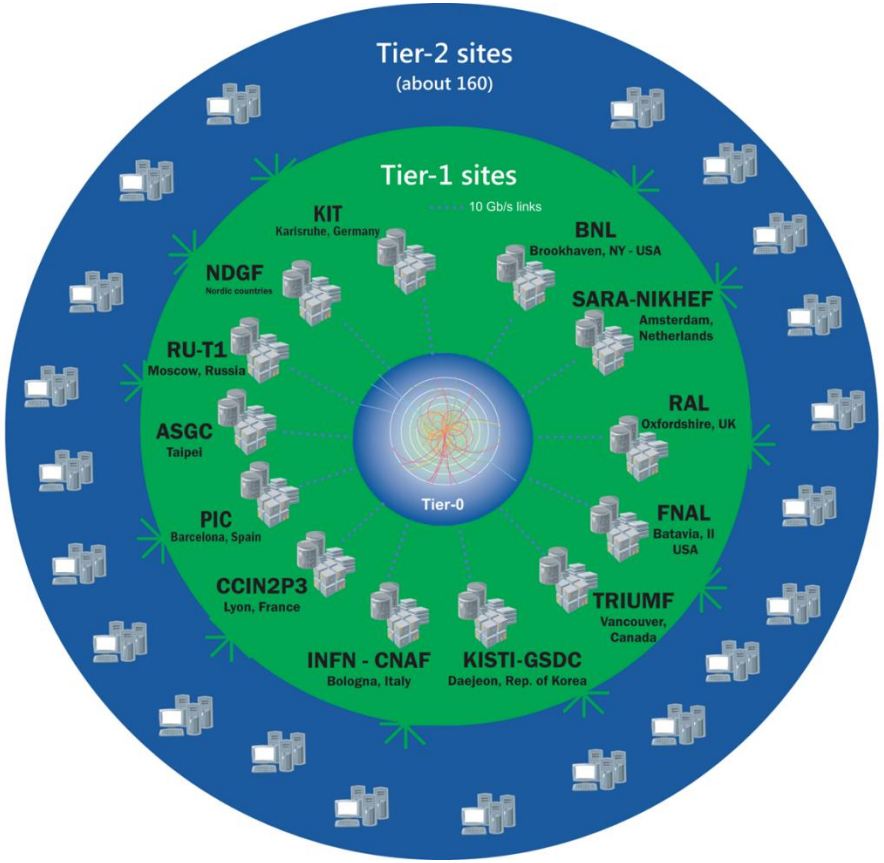

- Minimize constraints due to user localisation and resource variety
- Decentralize control and costs of computing infrastructure

➔ **Solution through LHC Computing GRID**

➔ **Much faster delivery of physics**

# WLCG Grid Model

- Four layer or level or “tiers”; 0, 1, 2, 3
- **Tire-0** is CERN Data Centre, responsible for the safe-keeping of the raw data (first copy), first pass reconstruction, distribution of raw data and reconstruction output to the Tier 1s, and reprocessing of data during LHC down-times.




- **Tier-1** sufficient storage capacity and round-the-clock support for the Grid. Safe-keeping of a proportional share or raw data and reconstructed data, large-scale reprocessing and safe-keeping of corresponding output, distribution of data to Tier 2s.

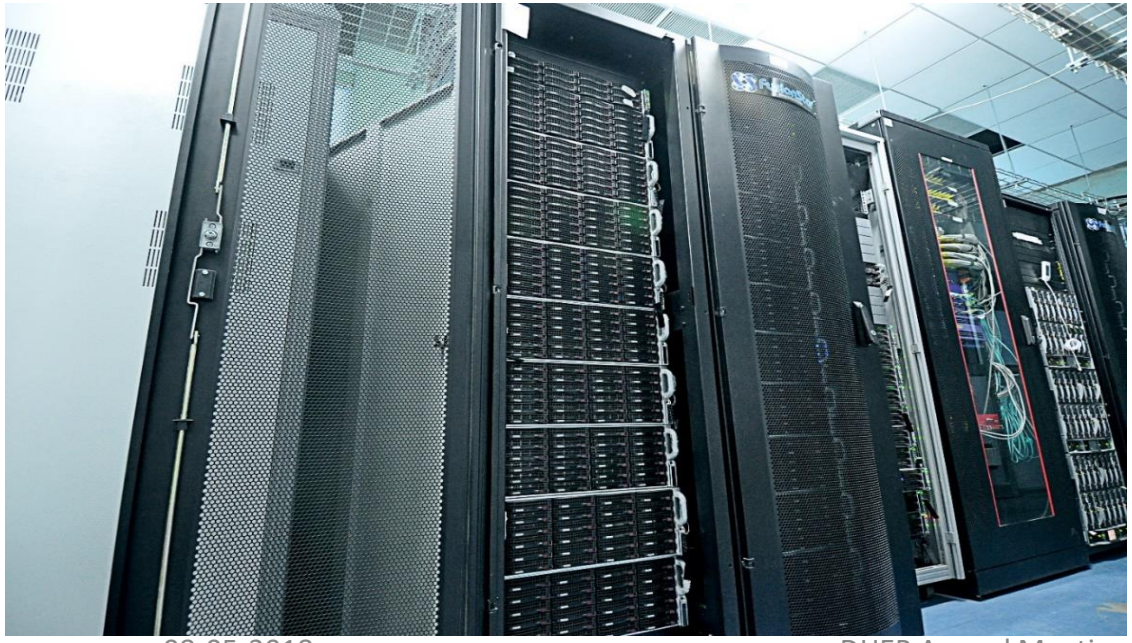
- **Tier-2** are typically universities and other scientific institutes, store sufficient data and provide adequate computing power for specific analysis tasks. Over the globe 160 centres.

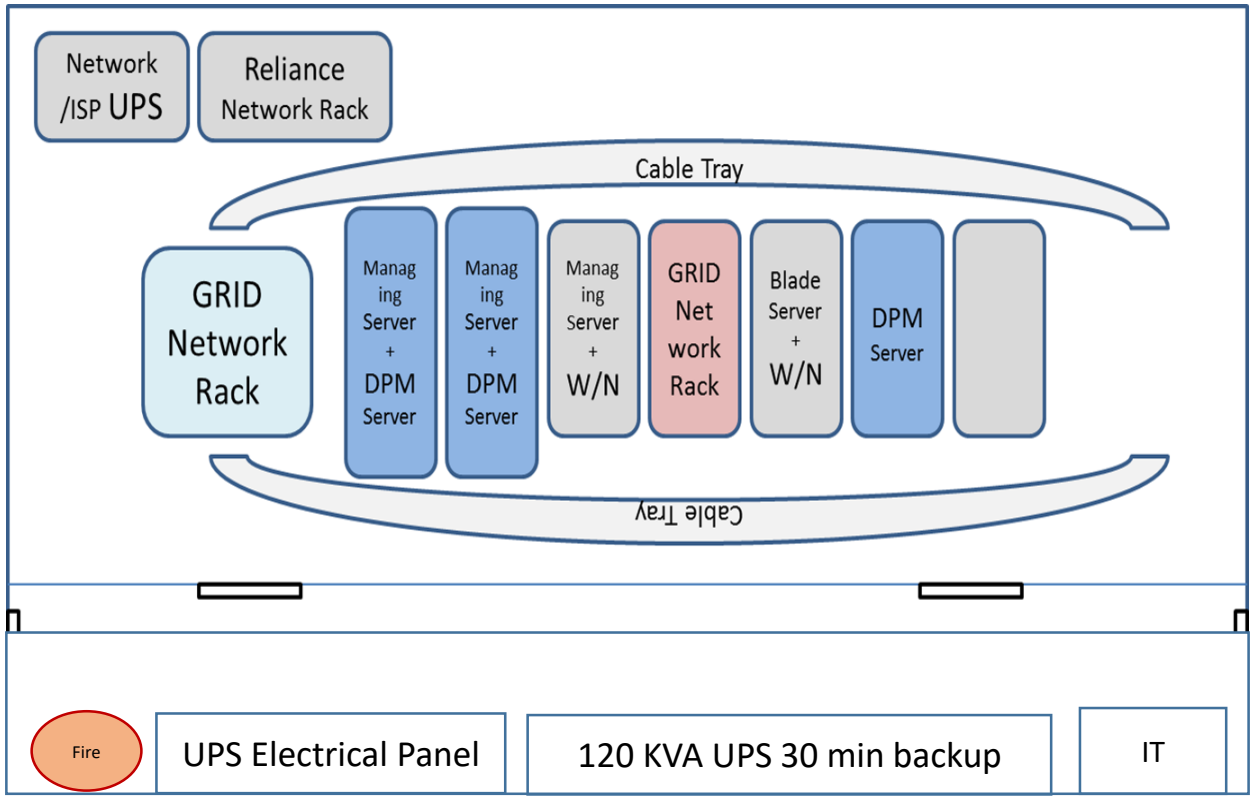
- **Tier-3** are local workstations or clusters, no formal engagement between WLCG and Tier 3 resources

# IndiaCMS Grid Tier2 facility, TIFR

- **Facility:**
  - 12 server racks
  - 120 KVA UPS + Isolation transformer

  (recently installed new UPS and Isolation Transformer, Electrical Panel)
  - Fire system: FM 200
  - Cooling : 17K CFM
  - Networking – 10G + 10G WAN Links



# Resources: current and recent past

| Year | Site Name | Resources Pledged | |
|---|---|---|---|
| | | CPU(HEPSPEC06) | Disk (TB) |
| 2014 | T2_IN_TIFR | 2,800 (~300 cores) | 940 |
| 2015 | T2_IN_TIFR | 6,150 (~600 cores) | 970 |
| 2016 | T2_IN_TIFR | 12,288 (~1200 cores) | 1,980 |
| 2017 | T2_IN_TIFR | ~20,000 (~1400 cores) | 2,000 |
| | T3_IN_TIFRCloud | 280 K | - |
| 2018 | T2_IN_TIFR | ~25,000 (~2500 cores) | 3,000 |
| | T3_IN_TIFRCloud | 280 K | - |

- **Two commissioned sites for CMS**
  - **T2_IN_TIFR**
    - Torque/PBS/Cream-CE
    - DPM (Disk Pool Manager)
  - **T3_IN_TIFRCloud** (Dynamic resources site)
    - HTCondor
    - MS Azure (Grid ASCII Helper Protocol) GAHP
  - **Local T3 cluster**
    - 80 cores
    - HTCondor
    - 100 TB dedicated user storage
    - NFS

- GRID is distributed computing → uses off-the-shelf hardware
- **Cannot be compared with a HPC**

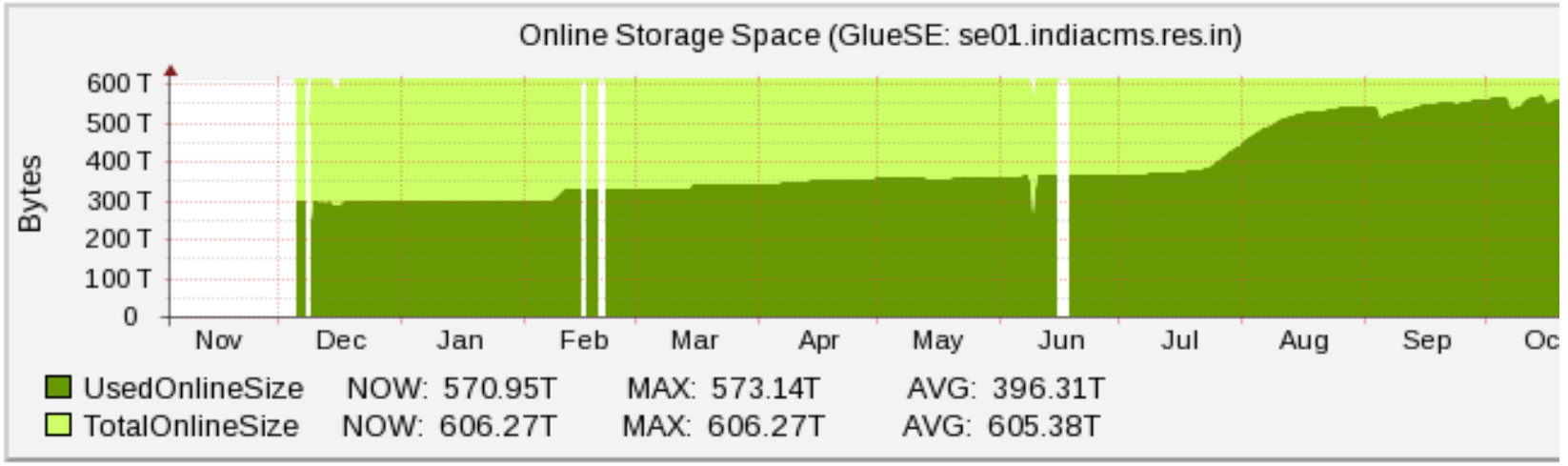
# Components and services

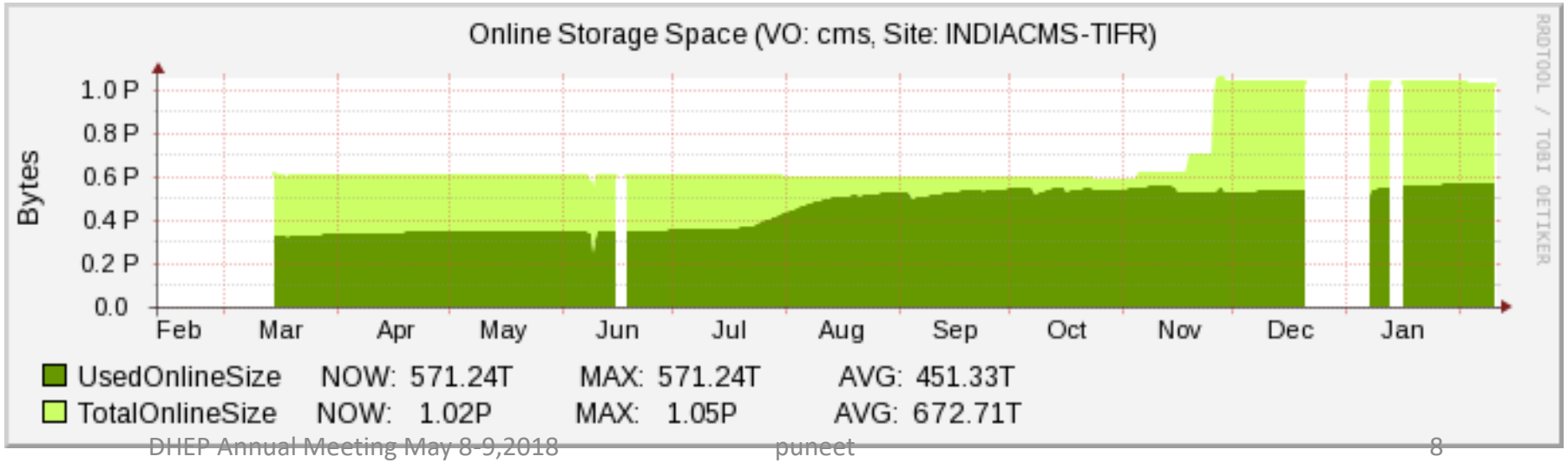
- **Storage:**
  - 34 – 4U nodes – RAID6
  - < 12% total storage overhead (16+2 RAID 6)
  - Disk pool manager without replication
  - Regional xrootd redirector
  - 3000 TB online
  - T3 user-space 100 TB
  - GridFTP and xrootd on all the nodes
- **Computing:**
  - 92 – half-width 1U nodes – RAID 1
  - Provides 2500 logical cores
  - 1/10G connectivity

- **Authentication:**
  - Argus
  - GSI ( x.509)
- **Accounting:**
  - APEL
- **Configuration & Orchestration**
  - Puppet
- **Monitoring and management**
  - IPMI on all the nodes
  - Scripts
    - Disk failure notification
    - Xootd transfer tests on all nodes
    - Storage head node backup
    - Auto-shutdown (UPS or Temperature)

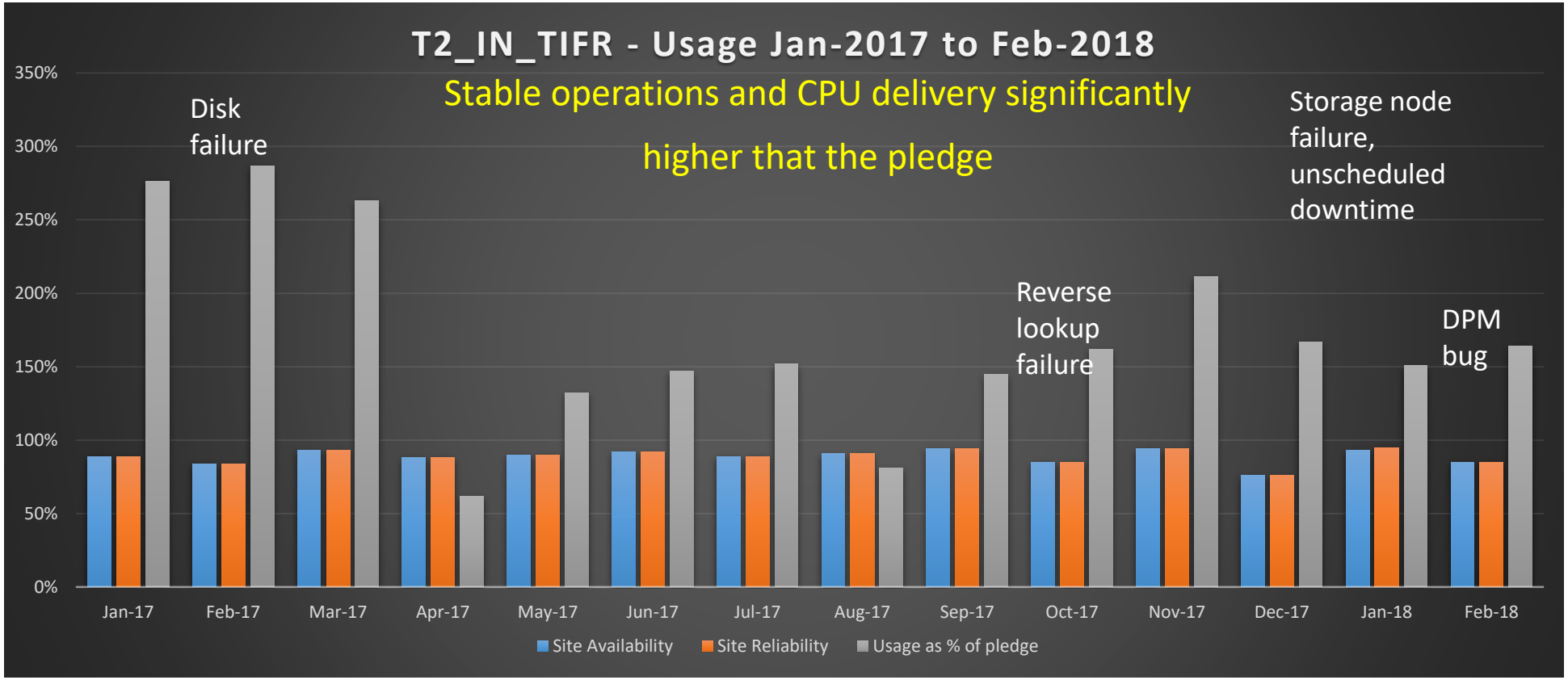
# Why do we need to continuously add resources?

**Improvement in online storage space in Q4 of 2014**



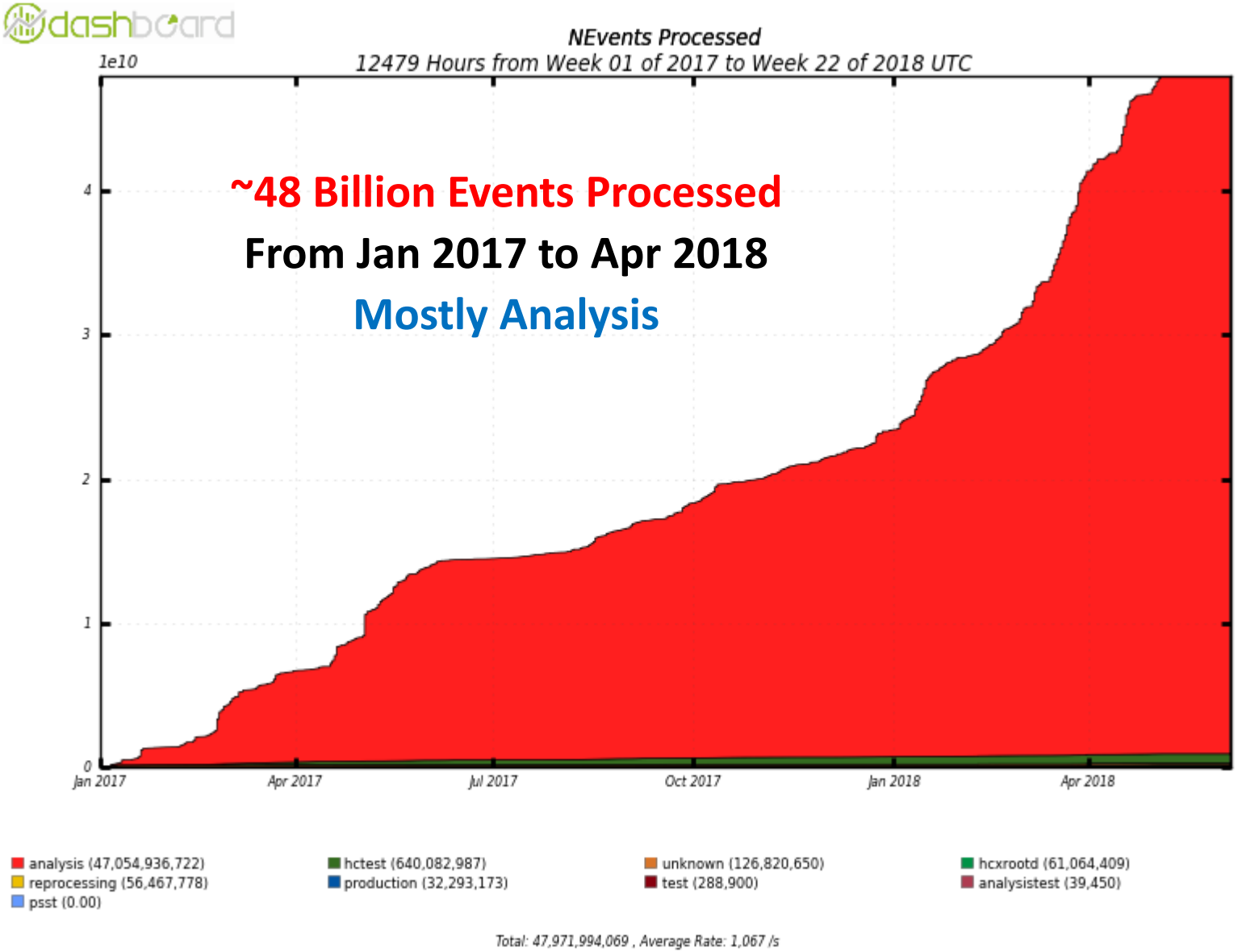



# Performance



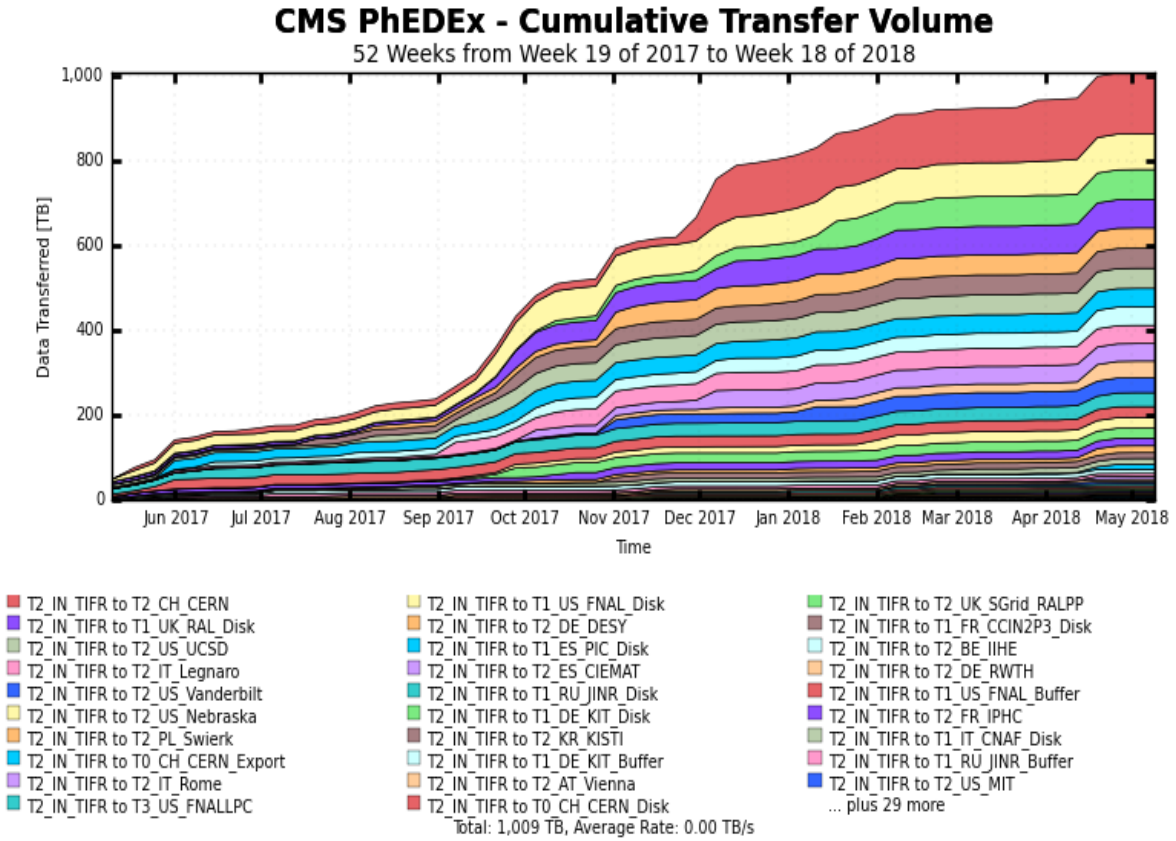

| | Jan-17 | Feb-17 | Mar-17 | Apr-17 | May-17 | Jun-17 | Jul-17 | Aug-17 | Sep-17 | Oct-17 | Nov-17 | Dec-17 | Jan-18 | Feb-18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Site Availability | 89% | 84% | 93% | 88% | 90% | 92% | 89% | 91% | 94% | 85% | 94% | 76% | 93% | 85% |
| Site Reliability | 89% | 84% | 93% | 88% | 90% | 92% | 89% | 93% | 94% | 89% | 94% | 76% | 93% | 85% |
| Usage as % of pledge | 277% | 287% | 263% | 62% | 132.00% | 146.83% | 151.92% | 80.79% | 144.78% | 161.84% | 211.41% | 166.82% | 150.73% | 163.95% |

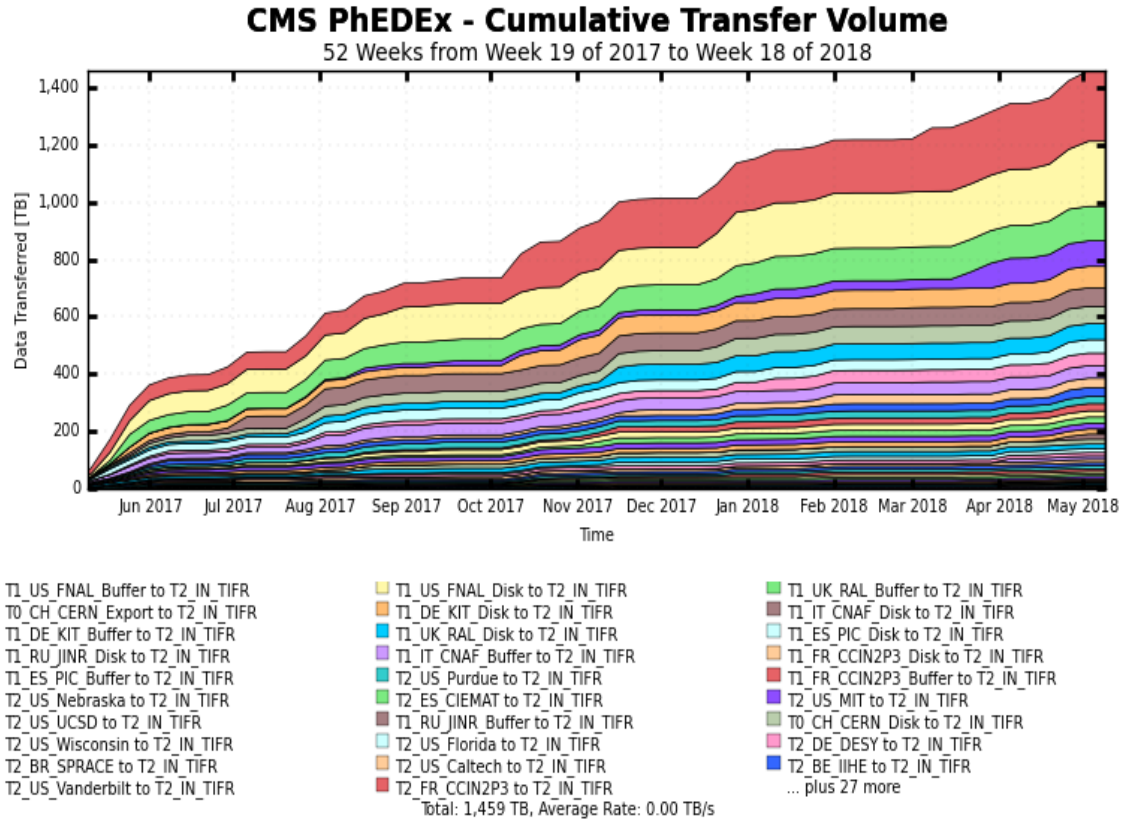
# Computing


dashboard
NEvents Processed
12479 Hours from Week 01 of 2017 to Week 22 of 2018 UTC
1e10
~48 Billion Events Processed
From Jan 2017 to Apr 2018
Mostly Analysis
0
1
2
3
4
Jan 2017
Apr 2017
Jul 2017
Oct 2017
Jan 2018
Apr 2018
analysis (47,054,936,722)
reprocessing (56,467,778)
psst (0.00)
hctest (640,082,987)
production (32,293,173)
unknown (126,820,650)
test (288,900)
hcxrootd (61,064,409)
analysistest (39,450)
Total: 47,971,994,069 , Average Rate: 1,067 /s

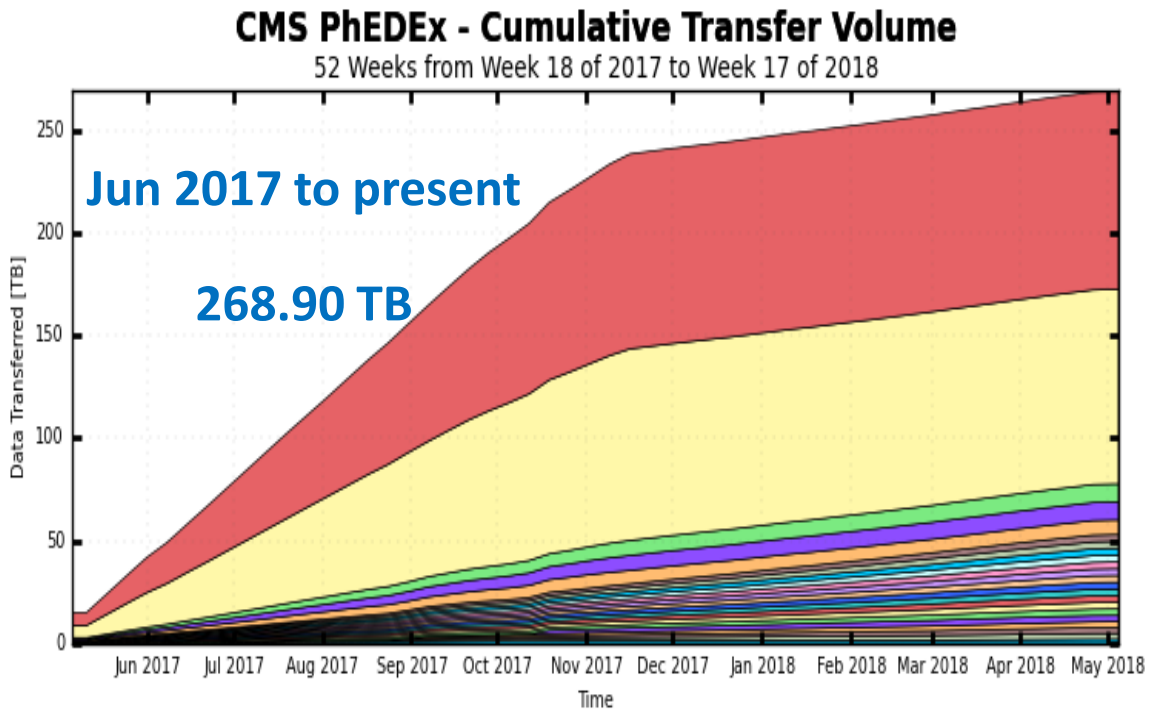
# Data Traffic

## 1PB upload + 1.4 PB download Productions Transfers

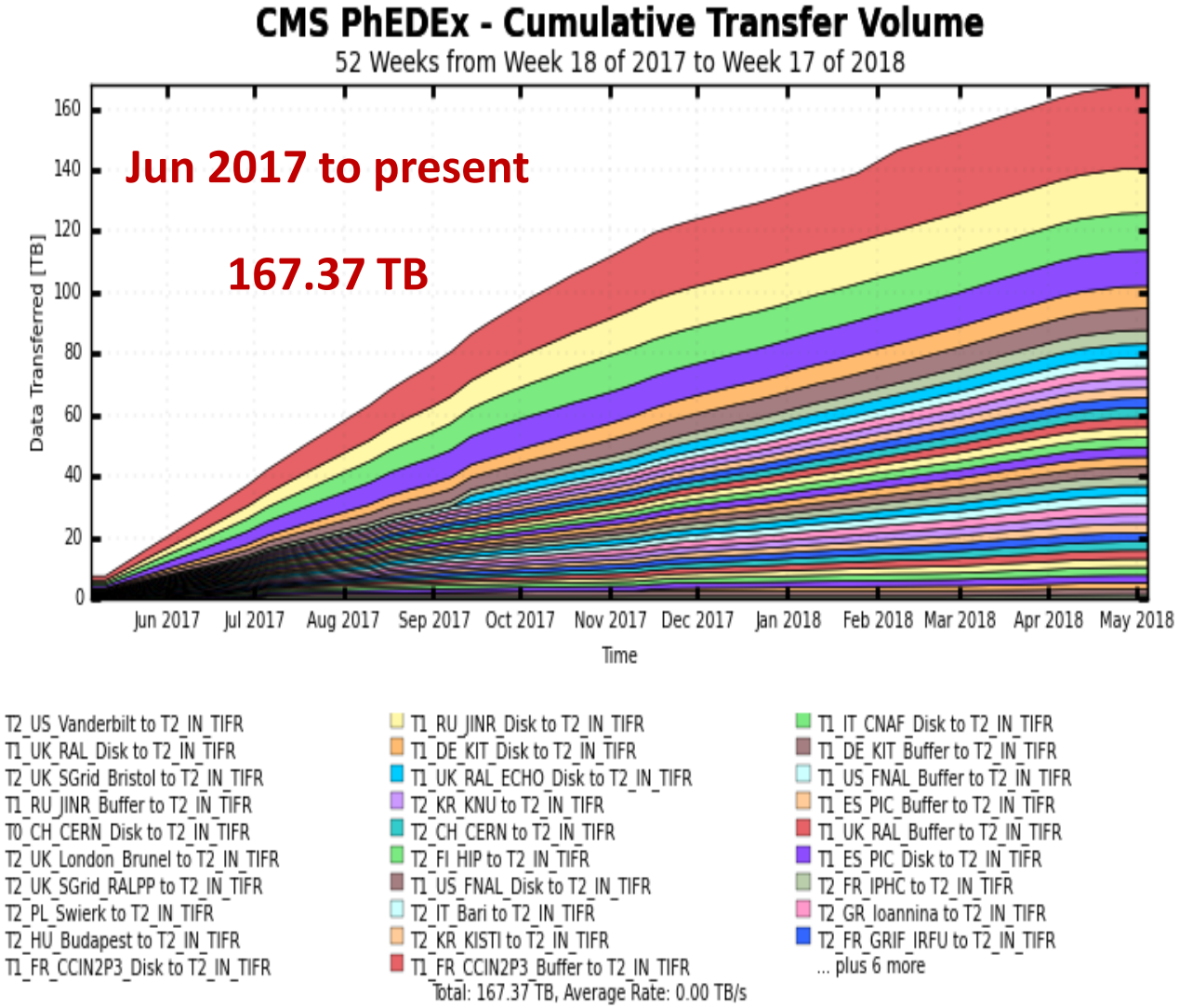


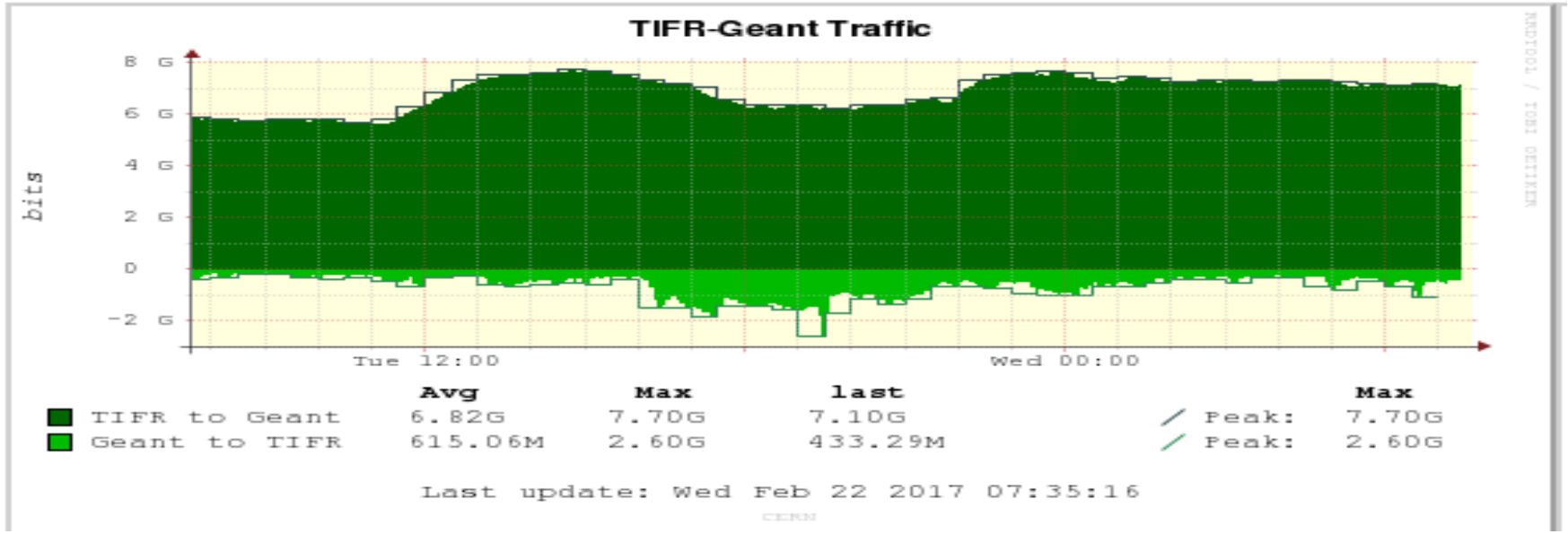
## TIFR to other T1/T2

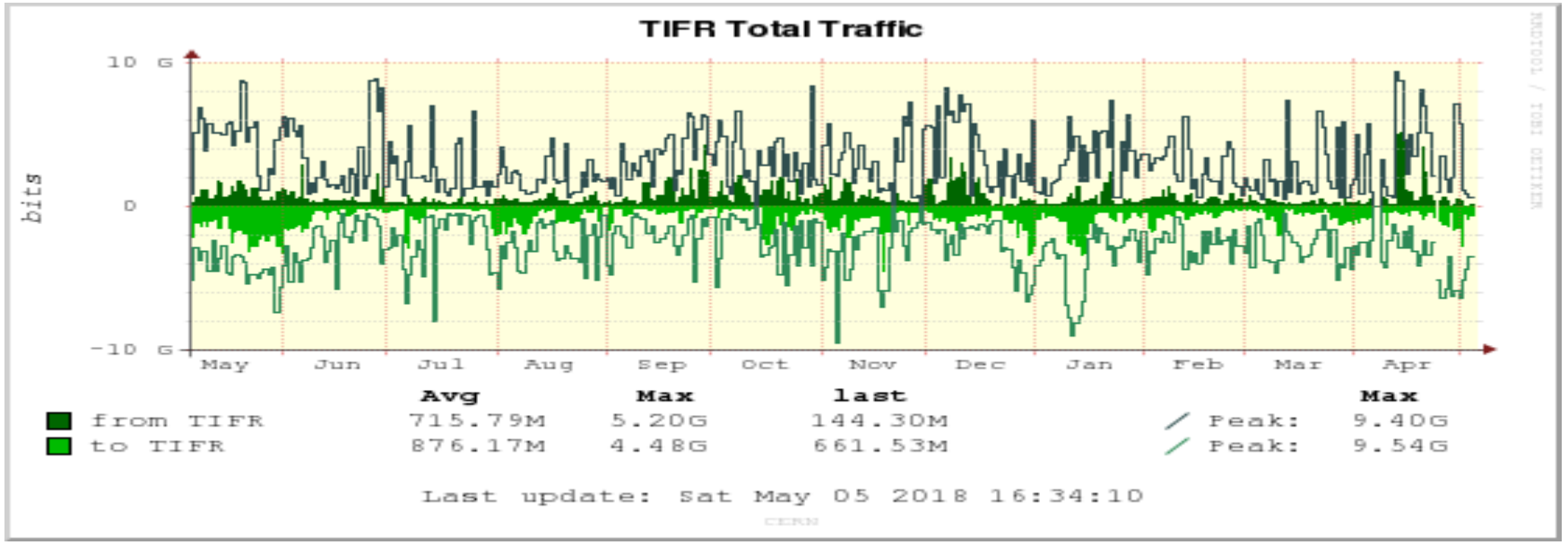



## Other T1/T2 to TIFR

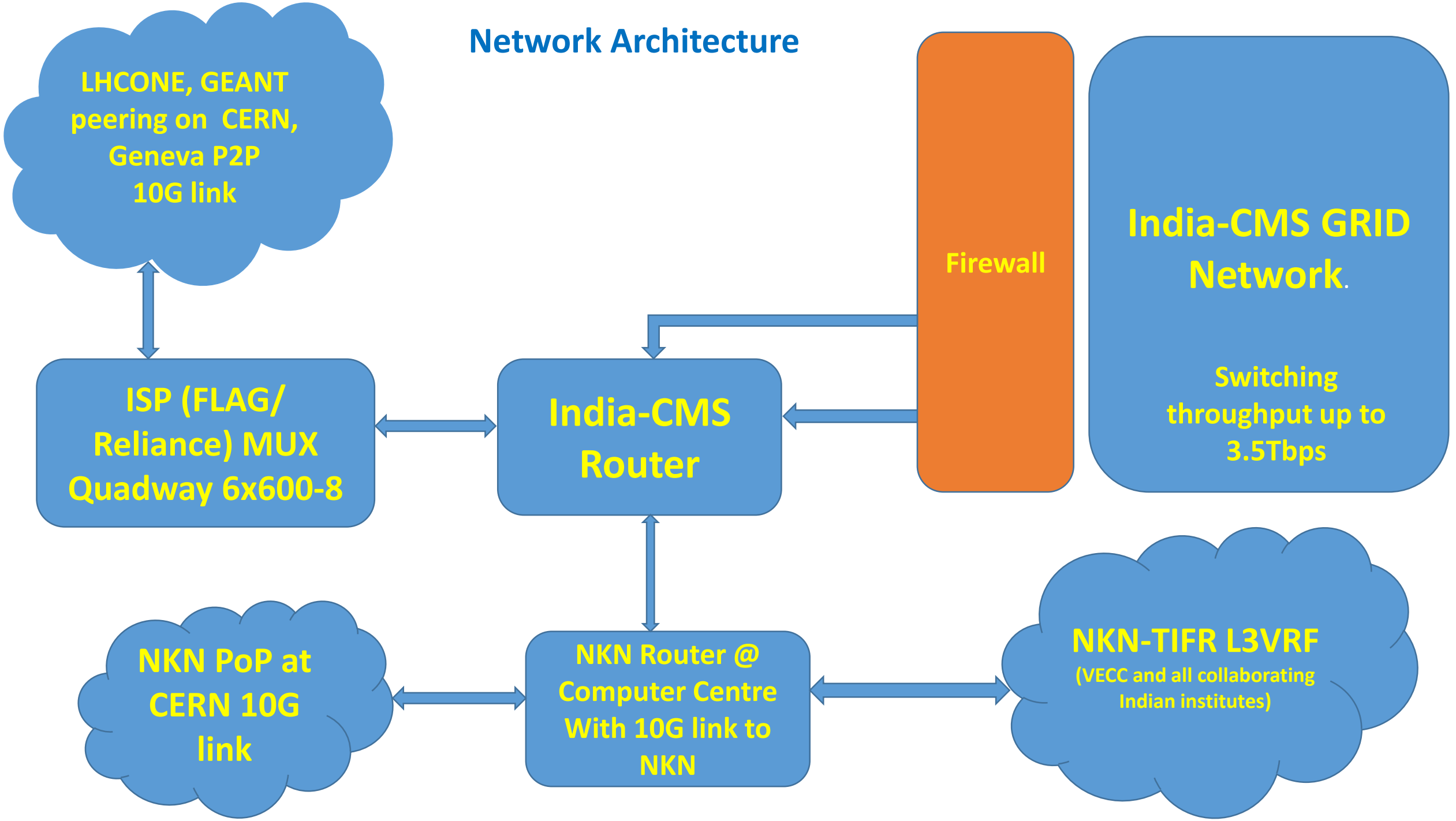


# TIFR-GRID Network Traffic



# Network Architecture



# Upgrade Plan for 2018

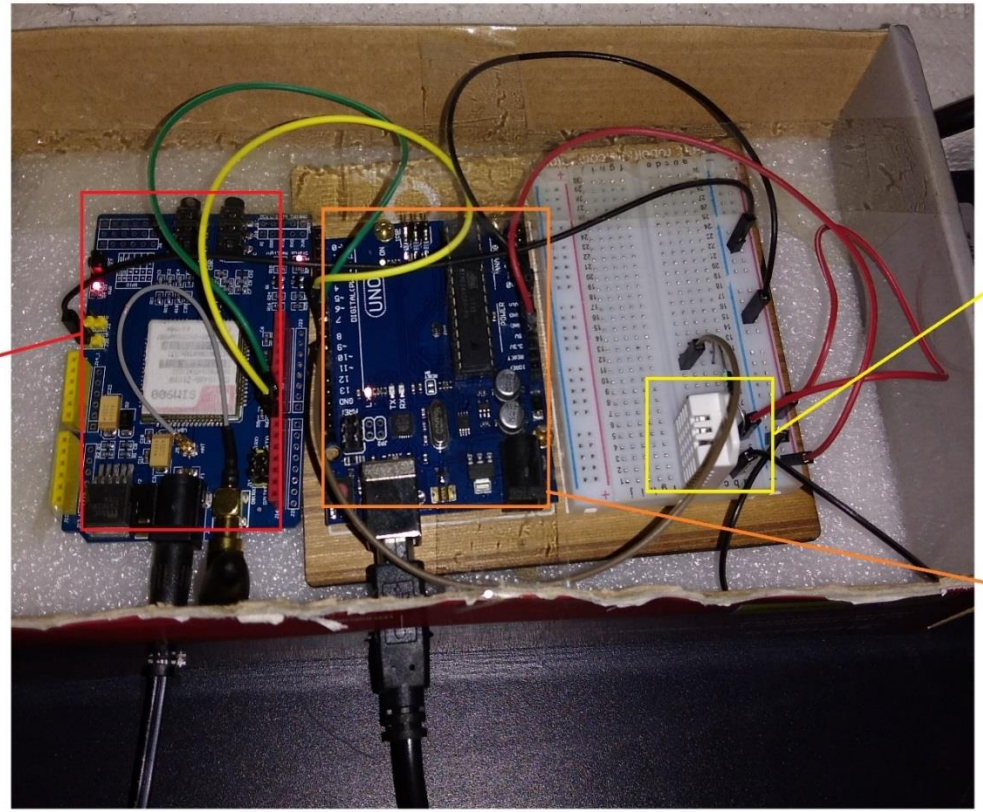
**For Tier-2**

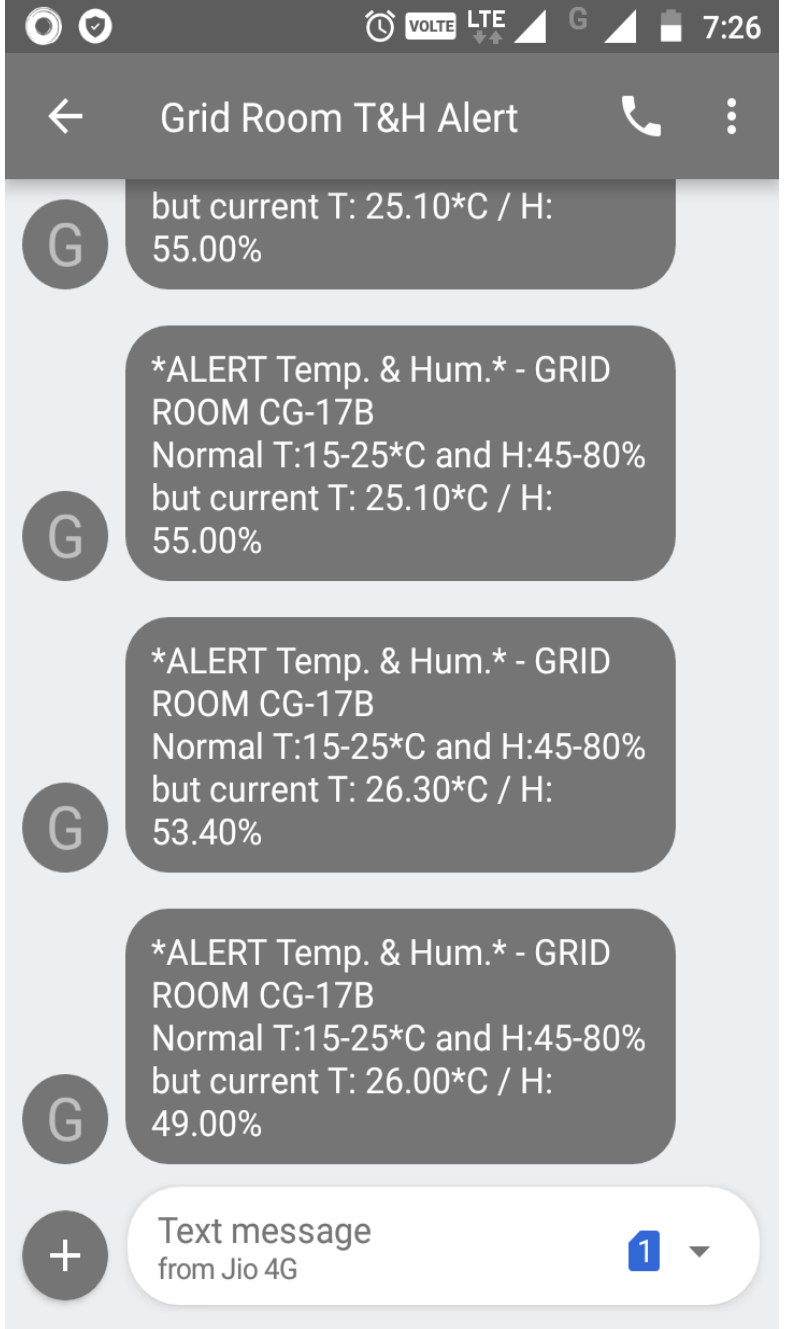
1 – Adding 2-petabyte additional storage
2 – Additional 1000 job slots. (procurement underway)
3 – Moving all the data nodes to 10G with new 40G capable switch and firewall
4 – Fully migrating to HTCondor-CE
5 – Migration to Cent-OS 7
6 – IPv6 implementation (CMS deadline Dec 2018)

**For Tier 3**

7 – Provisioning 400TB space for India-CMS users at T3
9 – T3 cluster cores to be increased to 200
10 – Adding cores from other clusters of India-CMS to T3 cluster.
11 – Providing Latest tools for users ( Brij talk )

# Recent tool development Environment Alert System




- Efficient and cost effective
- Real time notification and Call alert
- Software based policy (no limitations to add clients)
- Upgradation under progress

# Thank you!