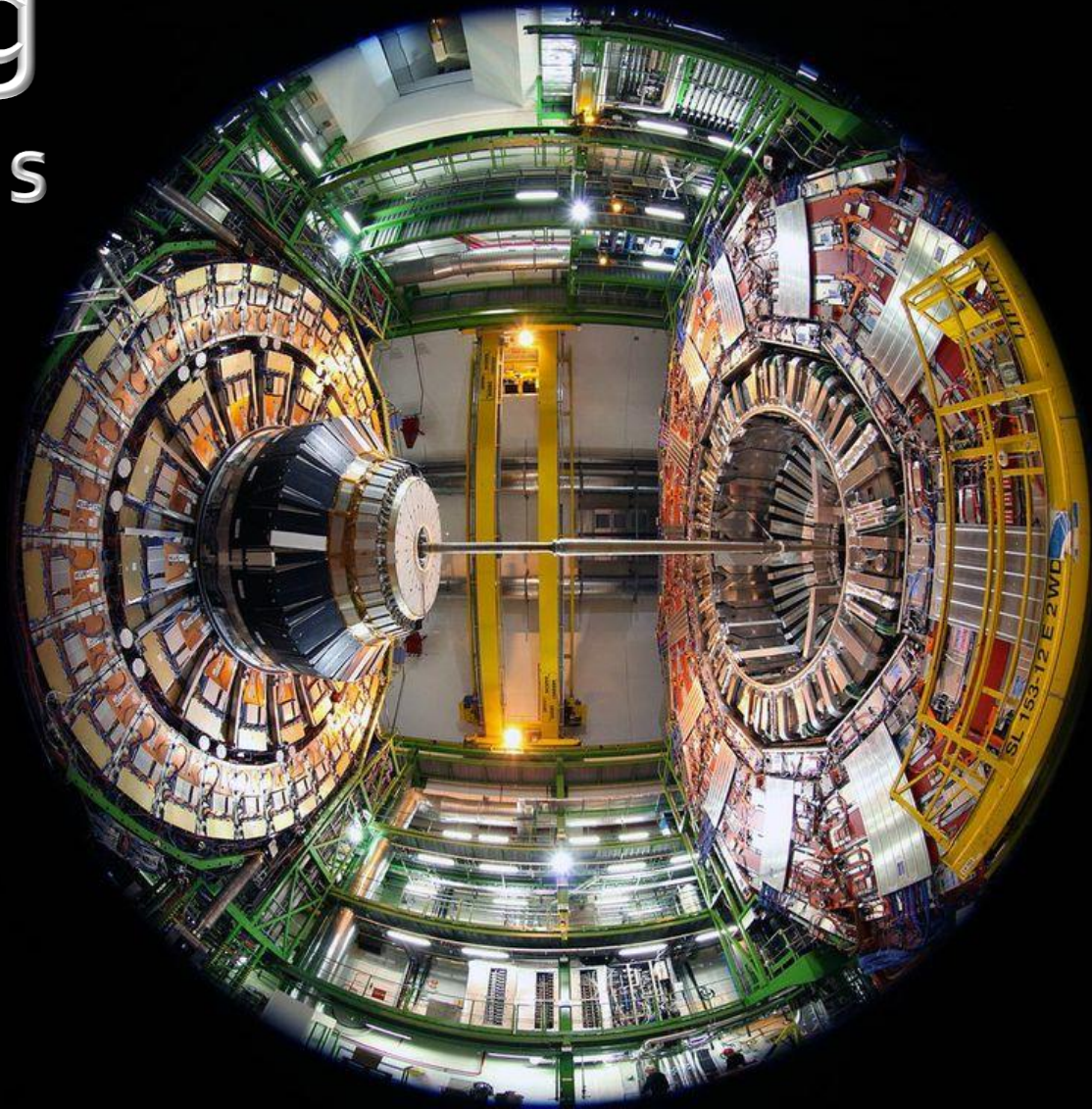# CMS Computing
Status and Plans

Maria Girone
LHCC Meeting
3rd June 2014

Outline

- **Q1 2014 WLCG Resource Utilization**
  - Commissioning the HLT for data reprocessing and MC production

- **Preparing for Run II**
  - Data Management, Analysis Tools evolution
    - Progress on the Tier-0 and AI
    - Data Federation
    - Dynamic Data Placement
    - CRAB3

- **The CMS summer 2014 Computing, Software and Analysis challenge (CSA14)**

- **Outlook**

Q1 2014 Resource Utilization

- Resources fully used so far, with high-priority requests for selected physics results (Heavy Ion, precision W mass measurements, ..) CSA14 preparation samples and detector upgrade studies
  - Tier-1 pledge utilization has been close to 125%
  - 89% CPU efficiency for organized processing activities

CMS Utilization

- Tier-2s are still the bulk of the first step of simulation production, and the primary analysis resource
  - 99.8% pledge utilization
  - 80% average CPU efficiency, factoring in the analysis efficiency as well

- The opening the Tier-0 for production tasks like a Tier-1, and analysis activity like a Tier-2 have given CMS the opportunity to start commissioning the CERN AI infrastructure, building on the success of the HLT farm commissioning

Maria Girone, CERN
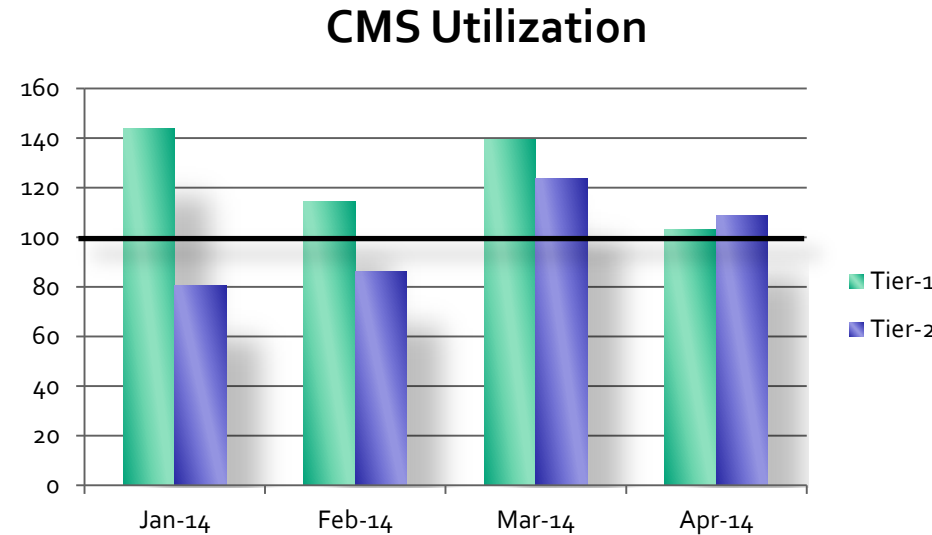
Production Activities

- Computing is utilizing the resources for the high priority tasks of CMS
  - Simulation is divided into GEN-SIM, which is concentrated at Tier-2s, and reconstruction (AOD), which is currently Tier-1
  - Events may be reconstructed more than once for different pile-up conditions

GEN-SIM

650M evts/month

AOD

1B evts/month
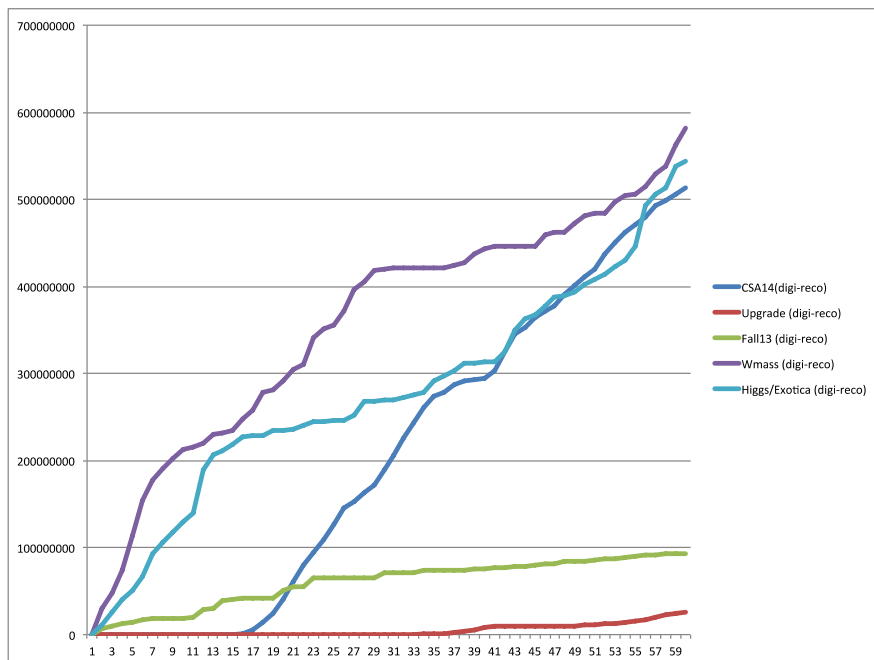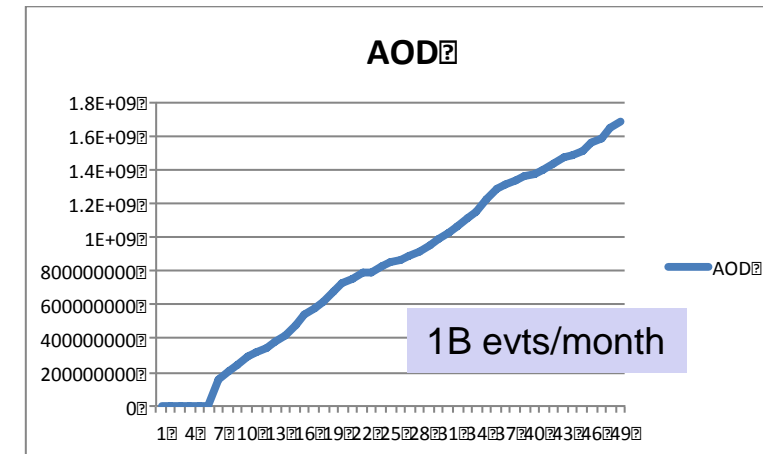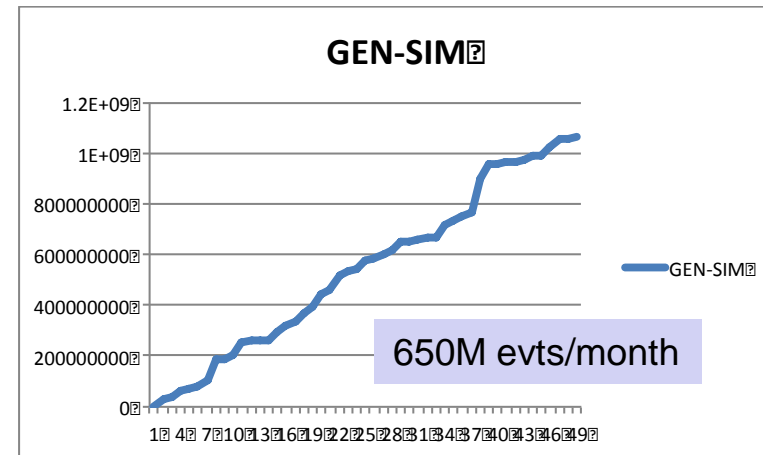
- System extremely busy with many production tasks running, all high priority!
  - Upgrade, W mass measurements, CSA14 sample preparation, and Higgs and Exotica Studies

CSA14(digi-reco)
Upgrade (digi-reco)
Fall13 (digi-reco)
Wmass (digi-reco)
Higgs/Exotica (digi-reco)

Maria Girone, CERN

The HLT and AI for HI data reprocessing

- The HLT, with an AI contribution, has been very successfully used for a high-priority data reprocessing of the 2011 Heavy Ion data
  - Made a very tight schedule of one month before Quark Matter 2014
  - Whole node cloud provisioning allows to provide **higher memory** slots for HI
  - Upgraded network allows high CPU efficiency for data reprocessing using the whole farm
    - No bottleneck reading from EOS

Heavy Ion 2011 Data

HLT CPU Efficiency
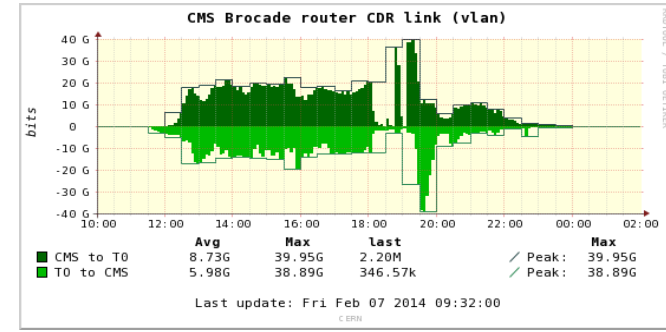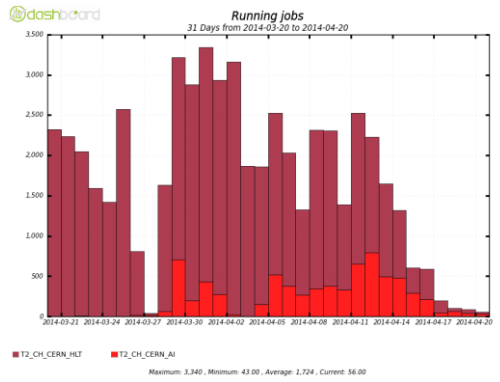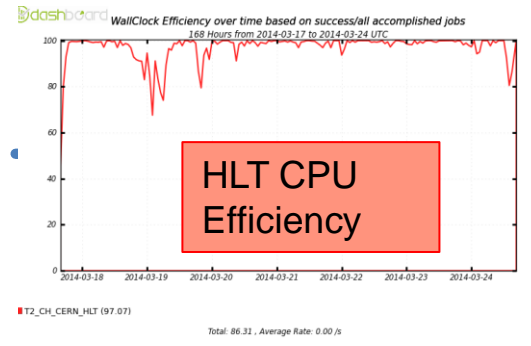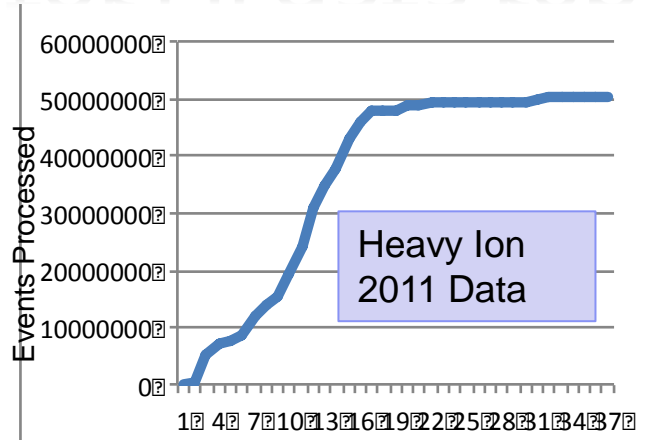
- **Excellent test of a thorough commissioning work of 6000 virtual cores set up as Openstack Cloud resources over 8 months**
  - during the winter shutdown the HLT farm will roughly increase the Tier-1 reprocessing resources by 40%
  - Able to start all virtual machines on the HLT quickly compatibly with usage in inter-fill periods, which can provide some contingency with additional resources during the year

Maria Girone, CERN

Preparing for Run2

- **CMS computing has evolved around two concepts**
  - **Flexibility in where resources are used**
    - We are developing the ability to share workflows across sites and access the input from remote storage
      - Data Federation and disk/tape separation
    - Dedicated cloud resources from AI (Tier-0) and HLT
    - Opportunistic Computing
      - Bootstrapping CMS environment on any resource using pilots, utilizing allocations on super computers and leadership class machines
      - Academic Cloud Provisioned resources
        - Commissioned new opportunistic cloud resource in Taiwan
  - **Efficiency in how resources are used**
    - Dynamic data placement and cache release
    - Single global queue for analysis and production with Glide-In WMS
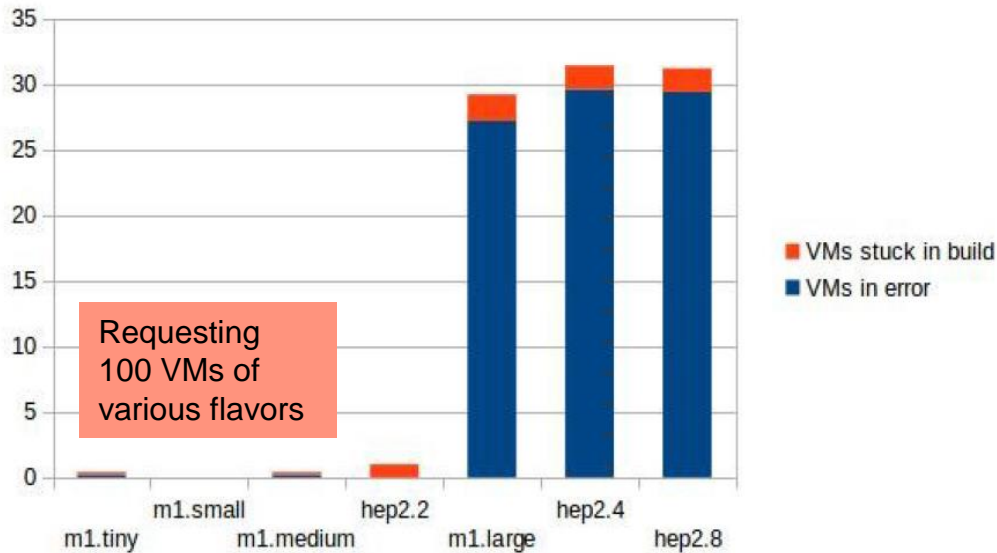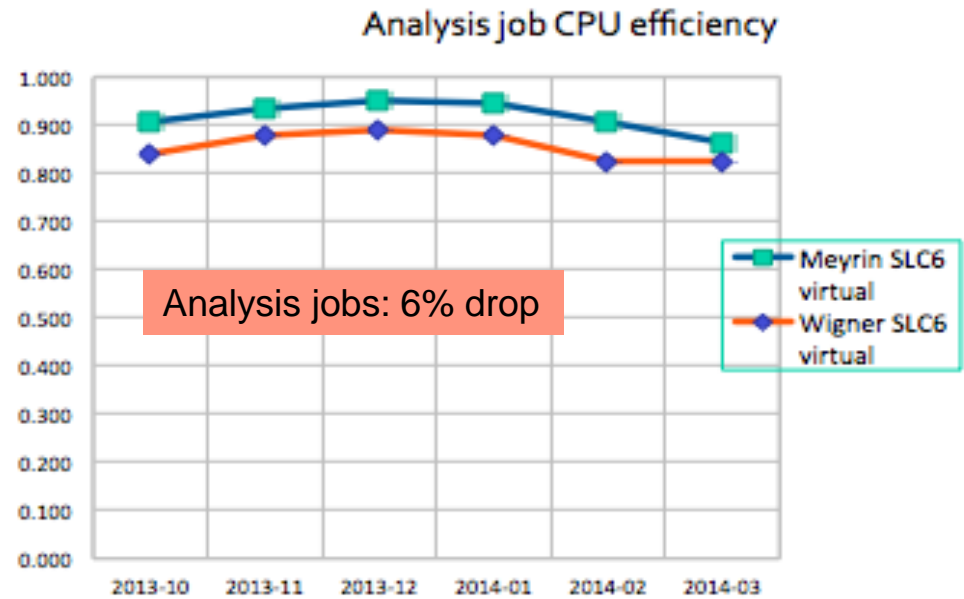   - New distributed analysis tools for users: CRAB3

2014 Milestones

✓ **Data Management milestone: 30 April 2014**

- Demonstrate the functionality and scale of the improved Data Management system
  - Disk and tape separated to manage Tier-1 disk resources and control tape access
  - Data federation and data access (AAA)

Analysis Milestone: 30 June 2014

- Demonstrate the full production scale of the new **CRAB3** distributed analysis tools
  - Profiting from the refactoring in the Common Analysis Framework project
  - Main changes are reducing the job failures in handling of user data products, improved job tracking and automatic resubmission

Organized Production Milestone: 31 October 2014

- Exercise the full system for organized production
  - Utilize the Agile Infrastructure (IT-CC and Wigner) at 12k cores scale for the Tier-0
  - Run with multi-core at Tier-0 and Tier-1 for data reconstruction
  - Demonstrate shared workflows to improve latency and speed

Progress on Tier-0 and AI

- We are planning to provision the Tier-0 resources directly on OpenStack through glide-In WMS
  - We currently cannot reach the same ramp-up performance as the HLT due to a high rate of VMs failures on AI
    - Discussion planned with AI experts

Analysis job CPU efficiency

Analysis jobs: 6% drop

- Meyrin SLC6 virtual
- Wigner SLC6 virtual

- VMs stuck in build
- VMs in error

Requesting 100 VMs of various flavors

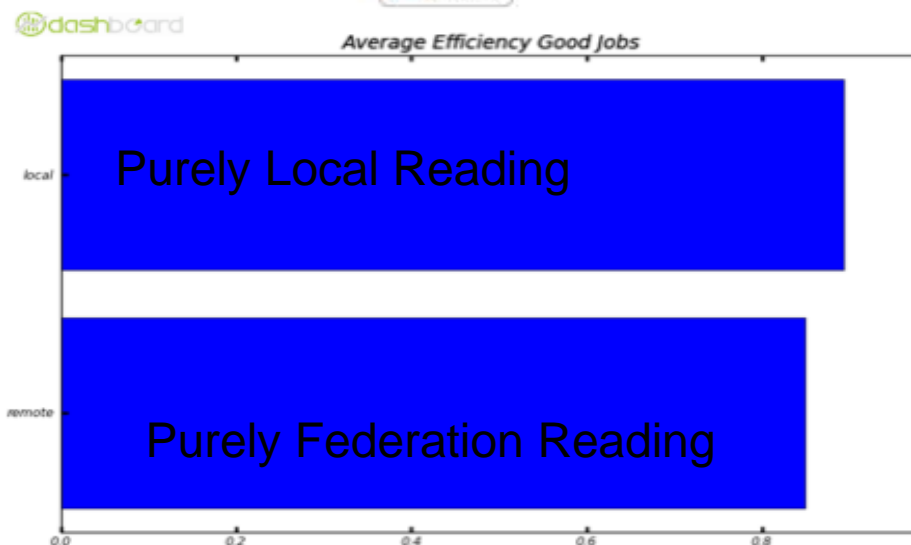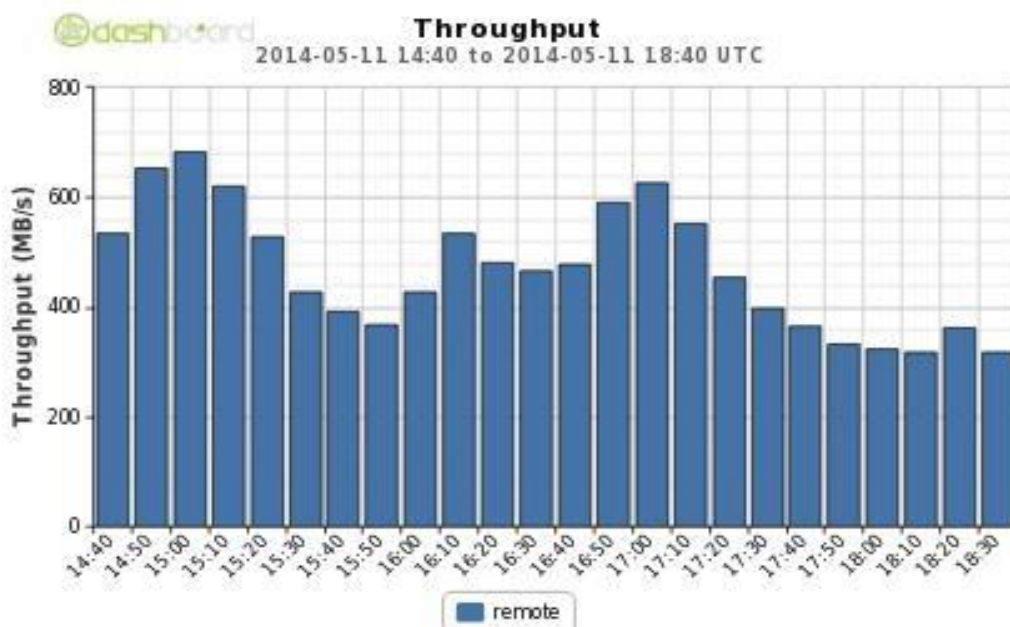Wigner CPU efficiency has been measured for both CMS production and analysis jobs

- We see only a small decrease in CPU efficiency in I/O bound applications reading data from EOS Meyrin
  - CMSSW is optimized for high latency I/O performance using advanced TTree secondary caching techniques
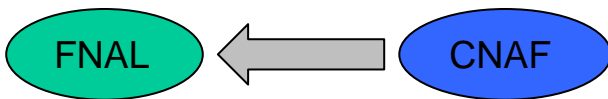
Maria Girone, CERN

Progress on Data Federation

- Provide access to all CMS data for analysis over the wide area through Xrootd
  - access 20% of data across wide area, 60k files/day, O(100TB)/day
- File-opening test complete
  - Sustained rate is 200Hz – much higher than what we expect in production use
- File-reading test: we see about 400MB/s (35TB/day) read over the wide area
  - Thousands of active transfers
- Small penalty in CPU efficiency for analysis jobs that read data remotely using data federation

Throughput
2014-05-11 14:40 to 2014-05-11 18:40 UTC

remote

Average Efficiency Good Jobs

Purely Local Reading

Purely Federation Reading

Maria Girone,

Data Federation for Reprocessing

Throughput
2014-01-31 00:00 to 2014-02-07 00:00 UTC
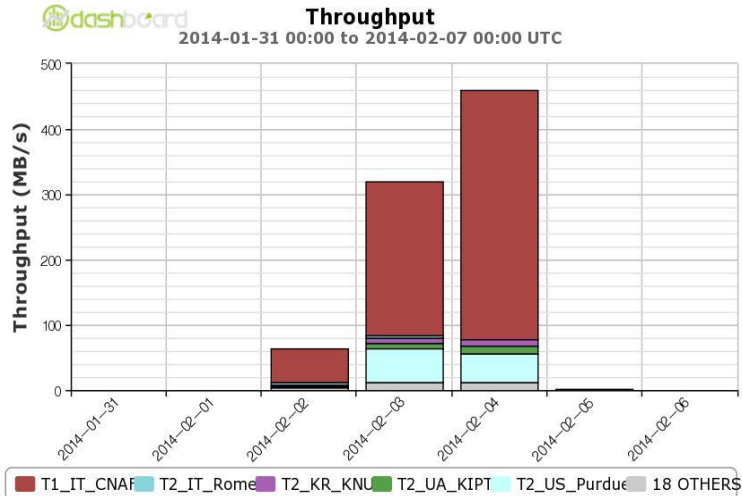
- ▪ CMS would like to make use of data federations for reprocessing use cases

  - ▪ Allowing shared production workflows to use CPU at multiple sites and data served over the data federation

- ▪ We had an unintended proof of principle. FNAL accidently pointed to an obsolete DB and production traffic failed over to Xrootd
  - ▪ Data was read from CNAF disk at 400MB/s for 2 days

- ▪ Regular tests have been added to our 2014 schedule, also during CSA14; larger reprocessing campaigns will be executed in Q3-Q4 2014

Dynamic Data Placement

- The development of dynamic data placement and clean-up is progressing
  - Prototype placement and clean-up algorithms are in place
    - The current CSA14 and upgrade samples are placed at Tier-2 sites automatically and the centrally controlled space is under automatic management

- Next step is to more aggressively clear the caches of underutilized samples
  - Will be exercised during the CSA14 challenge and monitor that we improve the storage utilization efficiency

- Due to effort redirection in CERN-IT, CMS has been building expertize
  - Exploring with ATLAS on expanding the scope of central operations

Distributed Analysis Tools

- **CRAB3 is currently in integration and validation tests before being handed to users for CSA14**
  - Improved elements for job resubmission and task completion, output data handling and monitoring
    - 20k running cores have been achieved in the validation, production target is 50% higher
    - Physics power users are expected to give feedback this month

- **Underlying services have also been updated**
  - FTS3 for output file handling
  - Central Condor queue for scheduling

Running jobs
191 Hours from 2014-05-01 to 2014-05-08 UTC

Maria Girone, CERN

CSA14
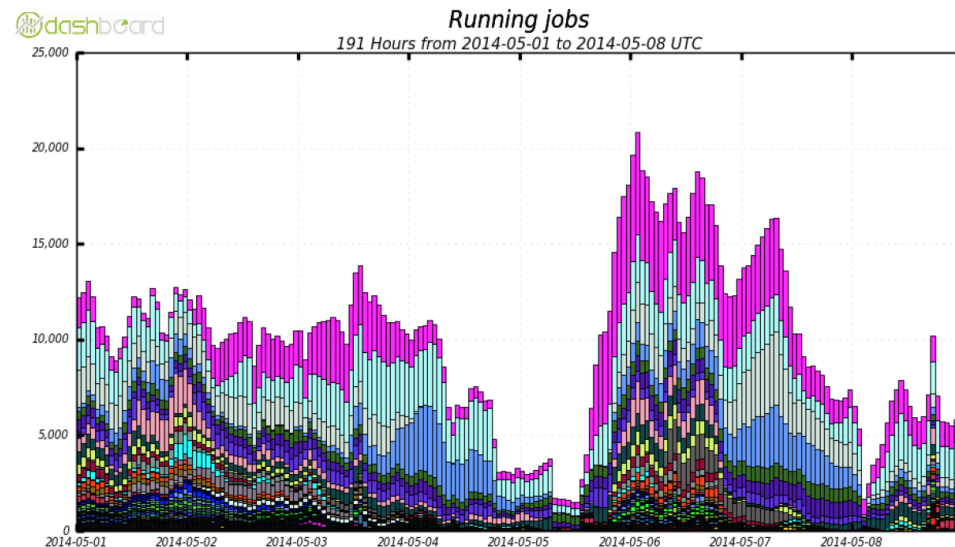
- CMS is preparing for nominal LHC beam conditions of Run II

  - Unique opportunity for early discovery ⟶ we need to make sure that all the key components will be in place and working since the very beginning, to ensure this CMS is performing a Computing Software and Analysis Challenge in July and August
    - Users will perform analysis exclusively using CRAB3 and Data Federation and Dynamic Data techniques
    - Production will exercise access through data federation
    - System performance will be monitored

- Ongoing preparatory work, including samples preparation for 2B events at 50ns and <PU>=40 & 20 and 25ns and <PU>=20

- We believe we are on schedule for the improvement work to allow CMS to function in the challenging 2015 environment and within the resources requested
  - The schedule had little contingency, but we are making steady progress

- The CSA14 challenge will be important to expose the users to new data access techniques and new analysis tools

- Effort to sustain the progress on improvements continues to be an issue

- CMS has a brainstorming workshop in mid-June to discuss Upgrade activities
  - There are innovative ideas, but a coherent direction is still forming

Maria Girone, CERN

Backup Slides

Resource Summary Table

Resource	Site	2014 CMS	2014 CRSG	2015 CMS	2015 CRSG
CPU (kHS06)	T0+CAF	121	121	271	271
	T1	175	175	300	300
	T2	390	390	500	500
Disk (PB)	T0+CAF	7	7	3+12	15
	T1	26	26	27	26
	T2	27	27	31	29
Tape (PB)	T0+CAF	26	26	31+4	35
	T1	55	55	74	74

Table 13 CMS resources request and the CRSG recommendations.

Maria Girone, CERN