# LHCb report (WLCG workshop)

***26/11/2013***

**Aresh Vedaee**

dapnia
cea
saclay

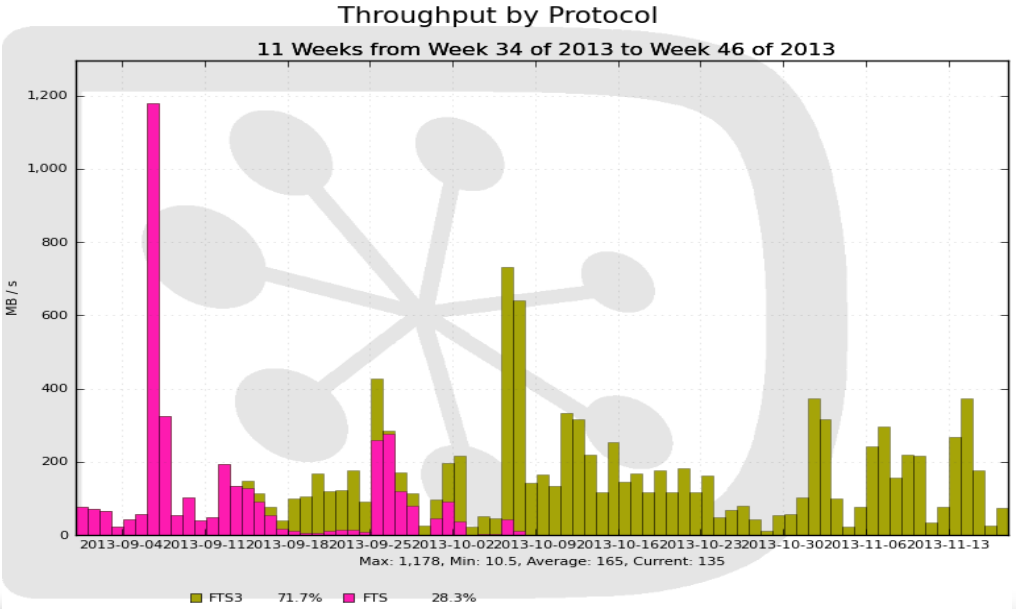
CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

# Summary

- Current Status
- Agenda 2014
- Computing Model Evolution
  - Processing Model
  - Data Management Model
  - Cloud & Opportunistic computing

# LHCb Current Status (1/2)

- **Production**
  - Stripping campaign ongoing: at CC closed (started at the beginning October and ended 10-Nov-2013)
- **First Tier2D sites**
- **CVMFS lhcb-conddb** mount point removed
- **FTS3** for WAN transfers
  - Since beginning October (CERN instance, RAL backup)



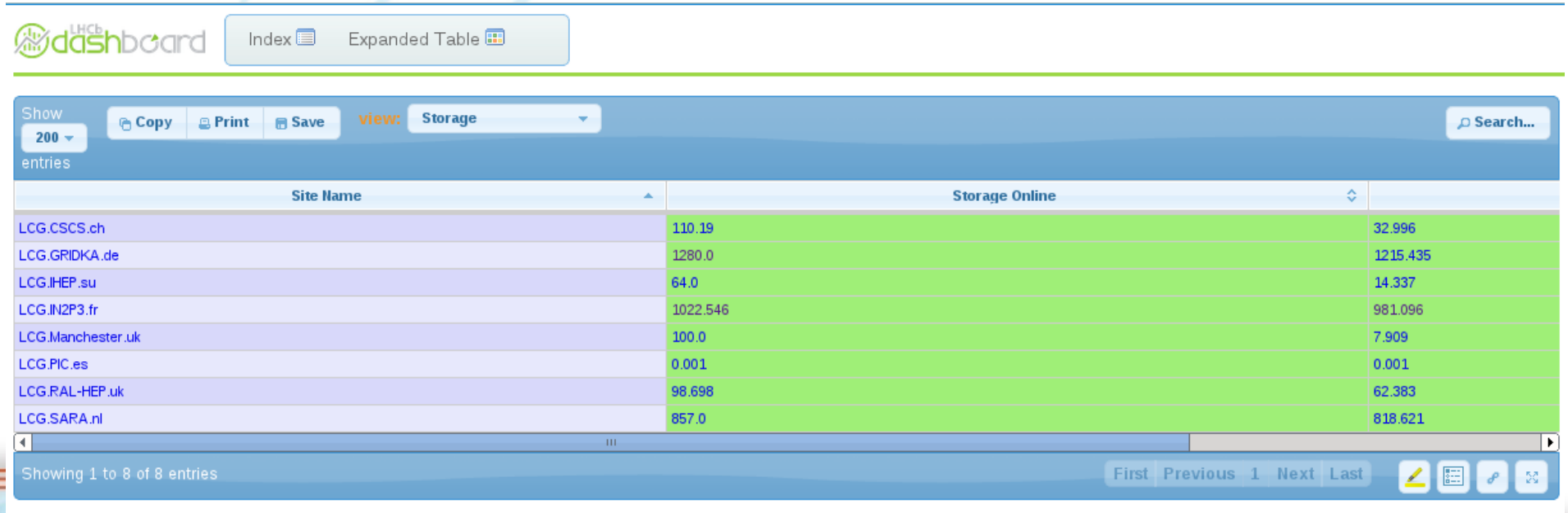
# LHCb Current Status (2/2)

- **LHCb Monitoring and WLCG Mon**
  - DIRAC able to feed workload info into WLCG mon:

http://dashb-lhcb-ssb.cern.ch/dashboard/request.py/siteview?view=By%20tier#currentView=By%2520tier

  - WLCG Site Status Board resurrected:

http://dashb-lhcb-ssb.cern.ch/dashboard/request.py/siteview?view=Storage#currentView=Storage



# LHCb Agenda 2014 (1/2)

- **Production**
  - 2 more stripping campaigns (probably February and September) on datasets reconstructed in 2011/12
- **SL6 binaries**
  - Most workflows can run on SL5/6, but starting from Jan 2014 only SL6 binaries without backward compatibility
- **WMS submission** to be suppressed for last sites
  - ~10% of sites (~20) to be moved gradually to direct submission

# LHCb Agenda 2014 (2/2)

- **Implementing Gaudi Storage federation**
  - Get local replica first, if it fails, try to access remote replicas (LHCbDirac provides an ordered list of tURLs)
  - Xrootd the only data streaming access protocol
  - Gsiftp for transfers (may look at http in the future)
- **Dirac File Catalog**
  - LHCb Bookkeeping Catalog (Oracle based)
    - Full info about files provenance
  - DFC (MySQL based) will replace CERN LFC catalog
    - Much lighter and designed to fit exactly LHCb needs
    - Integrated “du” possibility and metadata queries

# LHCb Computing Model (TDR)

- **Tier1**:
  - RAW data: 1 copy at CERN, 1 copy distributed (6 Tier1s)
  - First pass reconstruction (for monitoring and calibration) at CERN+Tier1s
  - End of year reprocessing of complete year's dataset.
  - Each reconstruction followed by "stripping" pass -> Stripped DSTs distributed to CERN and all 6 Tier1s
- **Tier2**:
  - Tier2s dedicated to simulation (and analysis jobs with no input data) -> Simulation DSTs copied back to CERN and 3 Tier1s
  - All disks located at CERN and Tier1s

# LHCb TDR Issues

- **Problems with TDR model**
  - Inflexible use of CPU resources and pressure on storage space
    - Large peaks, increasing with accumulated luminosity. Only simulation can run anywhere
    - All sites treated equally, regardless of available space
  - Challenges of Run1 (2012) and Run2 (2015)

| | Design (TDR) | 2012 actual | 2015 design |
|---|---|---|---|
| Instantaneous luminosity ($cm^{-2}s^{-1}$) | $2 \times 10^{32}$ | $4 \times 10^{32}$ | $4 \times 10^{32}$ |
| Mean visible p-p interactions/crossing ($\mu$) | 0.4 | 1.7 | ~1.0 (25ns) |
| Raw event size (kB) | 25 | 60 | 60 (25ns) |
| HLT output rate (kHz) | 2 | 5 | 12.5 |

# Processing Model Evolution (1/2)

- **Reprocessing campaign**
  - First pass on <30% of RAW data for calibration within 2-4 weeks
  - Full data reco in 2-4 delays without end-of-year reprocessing campaign (2015 automated online validation of calibration for offline use due to major redesign of LHCb HLT system)
- **Reconstruction at Tier2 sites**
  - Reconstruction at selected T2s (45% of reconstruction CPU time on 44 new T2 sites but also outside WLCG e.g. Yandex) -> Download RAW file (3GB) from a T1 storage, run reco job (~ 24 hours), upload Reco output file to the same T1 storage
- **Analysis at 5 Tier2D**
  - Before 2013 no T2 sites provided disk for LHCb

# Data Management Evolution (1/2)

- **Reduced archival for Tape Storage**
  - Archives of all derived data exist as single tape copy
- **T2Ds for Disk Storage**
  - https://twiki.cern.ch/twiki/bin/view/LHCb/T2DCommissioning
    - Disk storage (dCache, DPM or other): SRM service + 300 TB minimum per site (replicas necessary for analysis)
    - LHCb-Disk SRM space tokens (LHCb-USER is optional)
    - xrootd access with WAN access (for federate storage elements)

# Data Management Evolution (2/2)

- **Data format for analysis**
  - from DST, contains copy of RAW and full Reco information (~120kB/event) to microDST (~13kB/event)
- **Dynamic data placement** based on free space availability:
  - Automated selection of disk replication sites: random choice weighted by the free space
- **Replica removal and data popularity** (soon to be automated):
  - For processing n-1: reduce to 2 disk replicas
  - For processing n-2: only keep archive
  - Replicas data popularity: table per dataset, last access date, number of accesses in last (n) months, its size.

# Processing Model Evolution (2/2)

**Distinction between Tiers for different types of processing activities is becoming blurred.**

**Prod managers already attach/detach sites manually to different production activities in DIRAC config system**

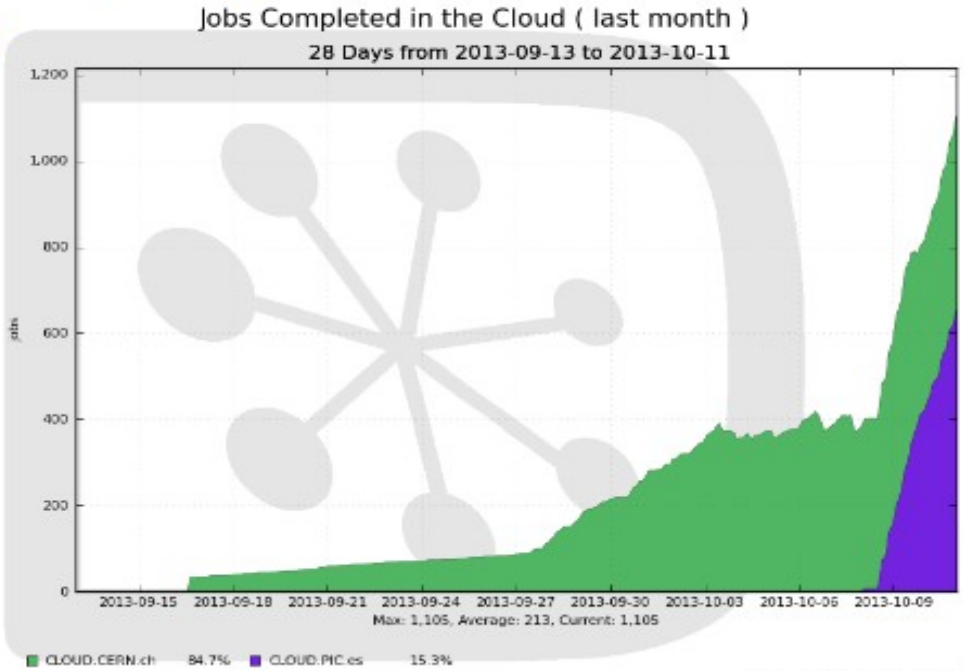
**In the future sites declare their availability for a given activity and provide the corresponding computing resources. Besides DIRAC allows easy integration of non-WLCG resources**

# LHCb Cloud integration (1/4)

- **What for?**
  - Running production in a more flexible way -> Pro/Con: lots of flexibility for a VO but at the price of becoming sysadmins rather than just users (reliability, security and monitoring)
  - Extremely handy testing and validating SW on different OS/SW/HW configurations to certify Grid middleware.
  - LHCbDIRAC Server Infrastructure: critical ratio components/resources (16 machines, ~40 different services, ~40 different agents, ~20 MySQL DBs & 1 Oracle DB) if a machine has issues, many services will stop working

# LHCb Cloud integration (2/4)

- No big numbers. In contrast, very solid infrastructure.
- Running on *Production* at:
  - CERN ( OpenStack ):
    - CLOUD.CERN.ch,
    - CLOUD.CERNMP.ch.
  - PIC ( OpenNebula ):
    - CLOUD.PIC.es.
  - RAL ( StratusLab ):
    - VERY SOON !
- Jobs we run:
  - MC,
  - Data processing.




| Cloud Site | Count | RAM * | Disk * | VCPUs |
|---|---|---|---|---|
| CLOUD.CERN.ch | 26 | 4 | 40 | 2 |
| CLOUD.CERNMP.ch | 4 | 8 | 80 | 4 |
| CLOUD.PIC.es | 75 | 2 | 10 | 1 |
| ΣTotal | 105 | 286 | 2.11K | 143 |

* ( GB )

# LHCb Cloud integration (3/4)

- **Basic VM settings:**
  - OS: CERNVM
  - SW distro: CVMFS
  - LHCb Config: Amiconfig (not CloudInit)

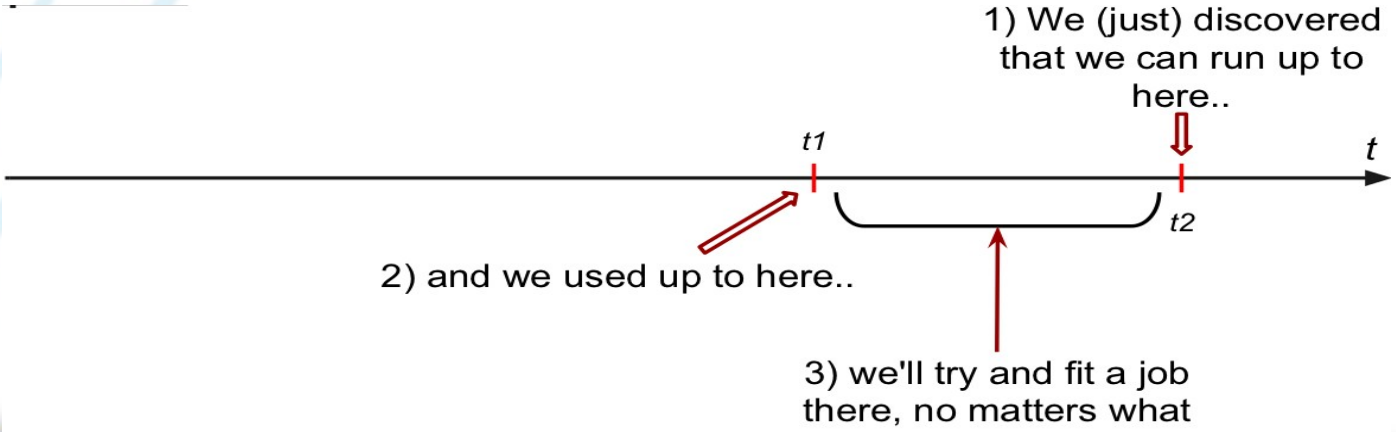
- **Models:**
  - Cloud + Batch System
  - Cloud + VMDirac (VM Registry, Scheduler, Monitor)
  - Vacuum

# LHCb Cloud integration (4/4)

- **Critical questions a VOs/Experiments:**
  - How to deal with authentication/authorization and host certificates?
  - What contextualisation? What VM image? Custom or per VO?
  - How to interface with hypervisor/host? How do we implement sharing machanisms?
  - HEPiX files? Shutdown procedure (VMs lifetime: run forever, load based, credit based, shutdown-on-demand)?
  - Accounting: just use wall clock time weighted by HEP06?
  - Outages announced via GocDB? Automatic IaaS discovery (BDII ?)
  - Isn't useful a common marketplace for all the clouds we use?
  - How does storage fit into all this?

# Miscellanea

- **Opportunistic computing:**
  - CERN HLT Farm: no virtualization, no pilot sub, PVSS driven (17% of the LHCb CPU during LS1). LHCb is not planning to use HLT farm for offline work during inter-fill gaps after LS1
  - Volunteer Computing (BOINC)
- **Job masonry**:
  - With flexible MC jobs (n° of events adapted to resources available)



# LHCb Operation Summary

- Put in production
  - First 5 Tier2D sites
  - Switch to FTS3
- Currently Ongoing
  - Implement Storage federation
  - WLCG/SSB for site admins
  - WMS decommissioning for last sites
- Future
  - Move to Dirac File Catalog
  - Only sl6 binaries from Jan ‘14
  - LHCb Dirac and Cloud/Opportunistic Computing integration