

Grid storage-types and user jobs in the AliEn Task Queue

Lucia Jancurova^{1,2}, M. Vala^{2,3}

¹*(Technical University, Faculty of Electrical Engineering and Informatics in Košice),*

²*(Laboratory of Information Technologies, Joint Institute for Nuclear Research, Dubna, Russia),*

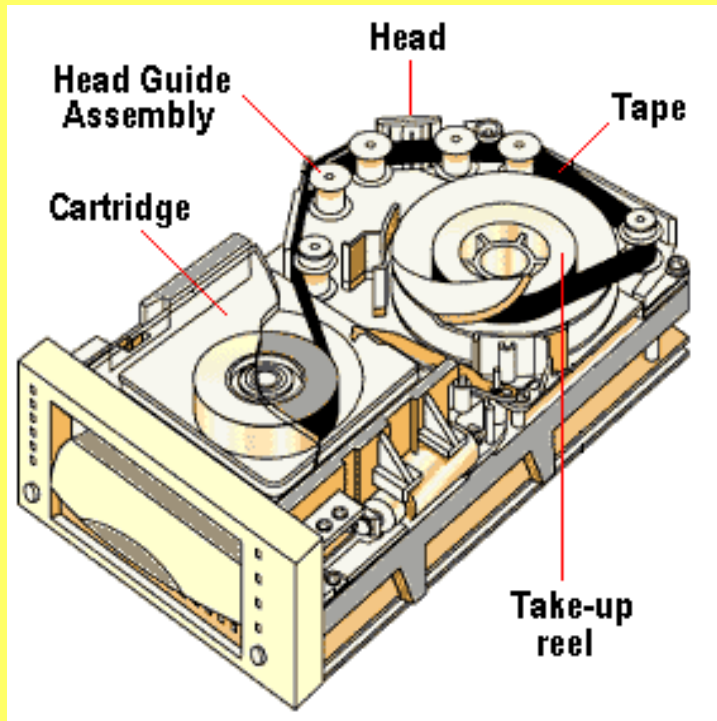
³*(Pavol Jozef Safarik, University in Kosice)*

Introduction

- LHC
- ALICE
- AliEn
- Storage elements



Storage elements



- Tapes



- Disks

Grid storage types

- Mass storage System (MSS) – all data written to this type of storage goes to tape
 - Available only at the large T1 centres
 - Very complex internal structure
- +
 - Configured to store very large amounts of data (multi-PB)
 - Still (slightly) cheaper than disk-only storage
 - Safer
- -
 - Data is recalled slowly from tape
 - Easy to fall victim to a race condition-multiple users reading different data sample

Grid storage types (2)

- Storage software types
 - dCache - developed at DESY/FNAL
 - CASTOR2 - developed at CERN
- In ALICE
 - CASTOR2 - CNAF, RAL, CERN
 - dCache - CCIN2F3, FZK, NL-T1, NDGF
- Both dCache/CASTOR2 implement reading/writing through the xrootd protocol
 - CASTOR2 – plug-in
 - dCache – protocol emulation

Grid storage types - Disk

- Disk – all data written to this type of storage stays on disk
 - Available everywhere, T0, T1 and T2 centres
 - Simple internal structure
- +
 - Fast data access
 - Prices per TB quickly falling
 - Very safe (if properly configured – RAID)
 - PB size disk storage can be easily build today
- -
 - Not really ideal storage type

Grid storage types – Disk (2)

- Storage types
 - dcache – developed at DESY/FNAL
 - DPM - developed at CERN
 - xrootd – developed at SLAC and INFN
- In ALICE
 - All T2 computing centres are/should deploy xrootd or xrootd-enabled storage
- Both dCache/DPM implement reading/writing through the xrootd protocol
 - DPM – plug-in
 - dCache – protocol emulation

Current SE deployment status

- User-accessible storage

http://aliceinfo.cern.ch/Offline/Analysis/GRID_status.html

- The local support needs some **improvements**, however the stability is very reasonable

| Storage elements | | | | | | | |
|---------------------|------------------------|--|-----------|-----------------|----------------|----------|--------|
| SE Name | AliEn name | Description | SE Status | Size | Used | Free | Usage |
| 1. Prague - Disk2 | ALICE::Prague::Disk2 | xrootd (disk), general use | OK | 13.6 TB | 9.265 TB | 4.334 TB | 68.13% |
| 2. Subatech - DPM | ALICE::Subatech::DPM | DPM (disk), general use | OK | 11.64 TB | 4.613 TB | 7.029 TB | 39.62% |
| 3. SPbSU - DPM | ALICE::SPbSU::DPM | DPM (disk), general use | OK | 5.402 TB | 8.593 GB | 5.393 TB | 0.155% |
| 4. Catania - DPM | ALICE::Catania::DPM | DPM (disk), general use | OK | 45.63 TB | 14.92 TB | 30.72 TB | 32.69% |
| 5. Bari - dCache | ALICE::Bari::dCache | dCache (disk), general use | OK | 4.005 TB | 1.206 TB | 2.798 TB | 30.12% |
| 6. CERN - se | ALICE::CERN::se | xrootd (disk), OCDB master, application packages | OK | 2 TB | 1.279 TB | 738.8 GB | 63.93% |
| 7. GSI - se | ALICE::GSI::se | xrootd (disk), general use | OK | 27.94 TB | 20.07 TB | 7.87 TB | 71.83% |
| 8. Legnaro - dCache | ALICE::Legnaro::dCache | dCache (disk), general use | OK | 13.04 TB | 12.36 TB | 696.2 GB | 94.79% |
| 9. NDGF - dcache | ALICE::NDGF::dcache | dCache (disk), general use | OK | 23.28 TB | 20.62 TB | 2.662 TB | 88.57% |
| 10. NIHAM - File | ALICE::NIHAM::File | xrootd (disk), general use | OK | 39.12 TB | 4.026 TB | 35.09 TB | 10.29% |
| 11. Prague - Disk | ALICE::Prague::Disk | xrootd (disk), general use | OK | 1.267 TB | 1.267 TB | 1 KB | 100% |
| 12. Torino - DPM | ALICE::Torino::DPM | DPM (disk), general use | OK | 16.78 TB | 1.073 TB | 15.71 TB | 6.397% |
| Total | | | 12 | 203.7 TB | 90.7 TB | | |

User jobs in the AliEn in Task Queue (TQ)

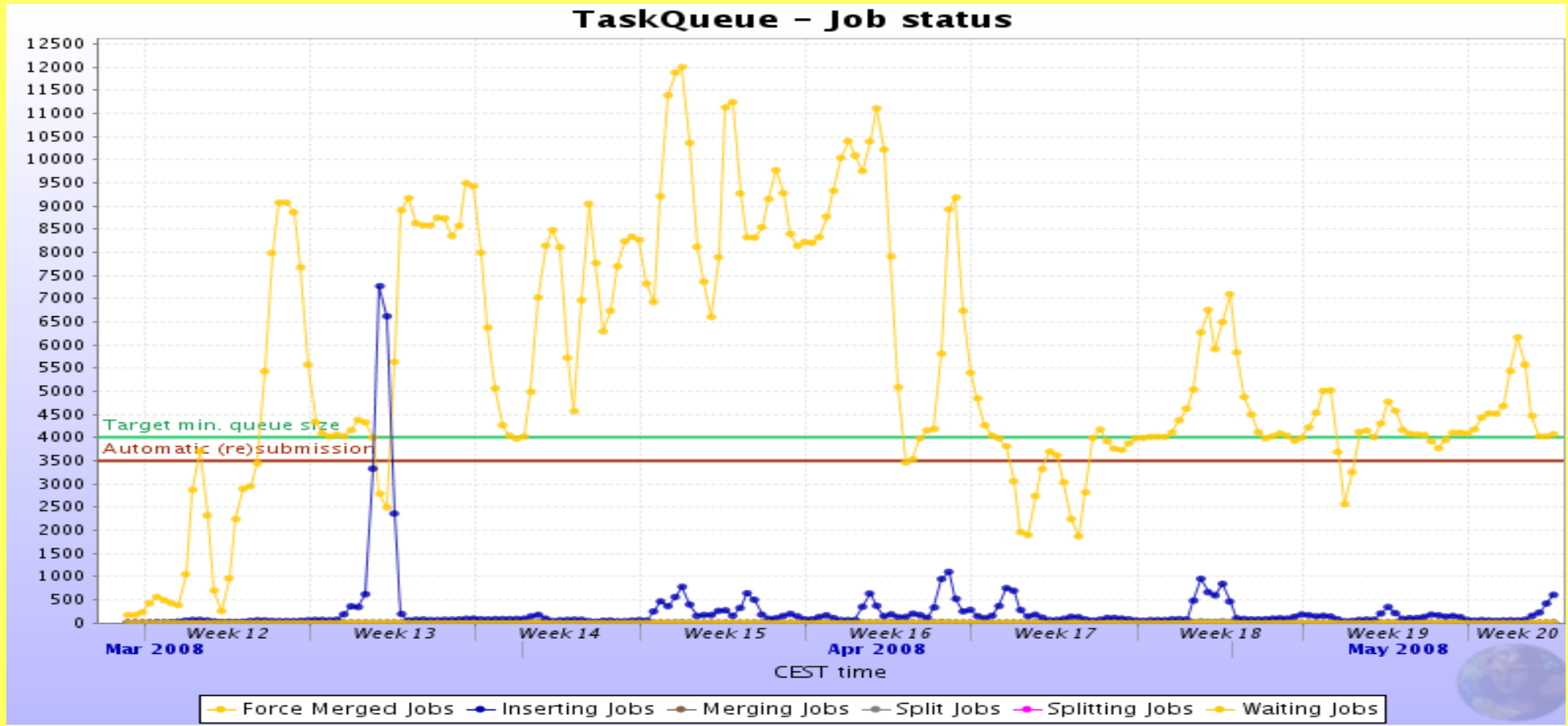
- The common AliEn tasks queue is holding all jobs submitted to the Grid
 - Monte Carlo (MC) production: 4K running + 4K jobs waiting at all times
 - RAW data production – at present, in principle same as MC production
 - User jobs: chaotic submission, executed with high priority

User jobs in the TQ (3)

- AliEn can process ~8 000 user jobs/day
- So far, we do not observe a critical overload of the TQ and services (JobOptimizer)
- For comparison at CERN (entire cluster)
 - 5 000 running jobs (comparable to AliEn)
 - 10 000 waiting jobs (less than in AliEn)

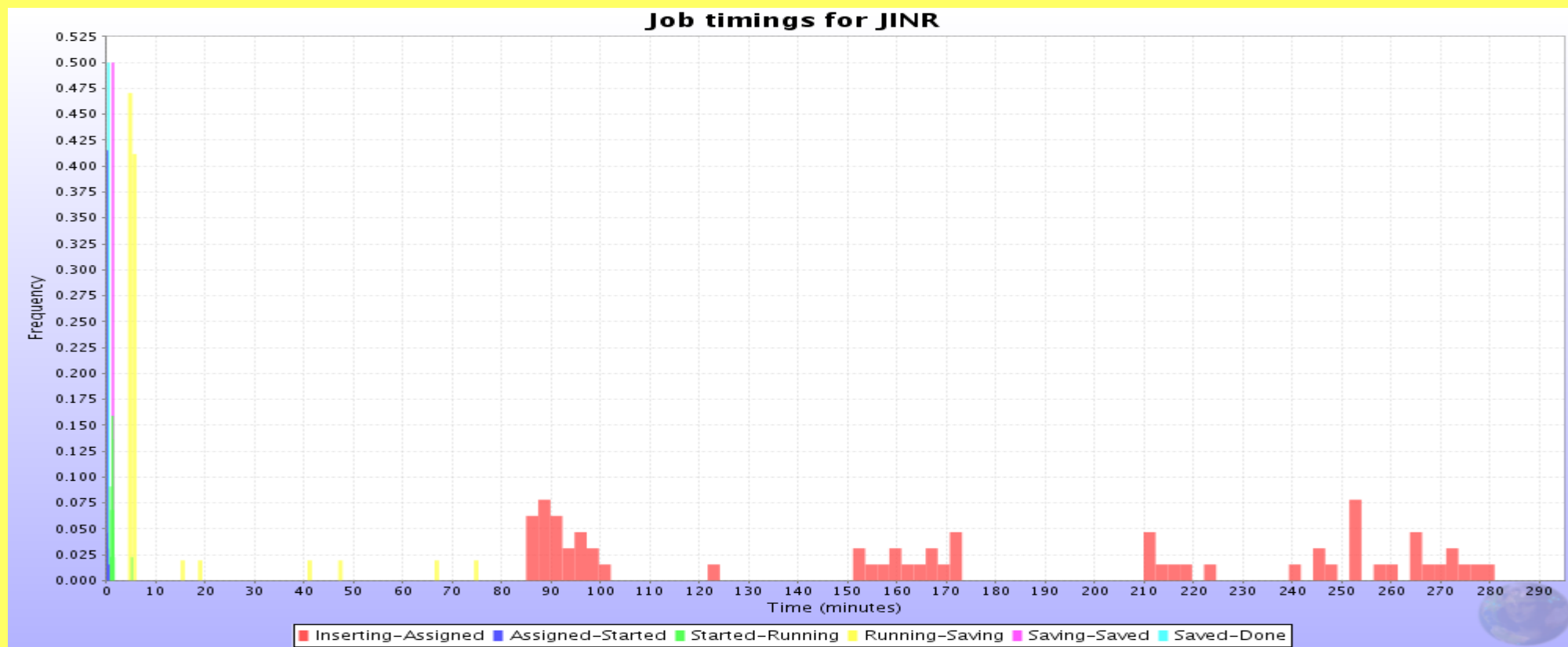
User jobs in the TQ (2)

- Job profile



User job timing

- User jobs are executed in average 120 mins after submission



Data availability for analysis

- Primary copy of all recent productions is on disk (at T2s)

| Production | Description | Status | Run Range | Events Count | Comments |
|---------------|---|-----------|-----------------|--------------|---------------------------------------|
| PDC 08/LHC08x | p+p, charm, forced had.decays | Completed | 180001 - 180066 | 6,331,700 | All runs staged |
| PDC 08/LHC08w | p+p, beauty, forced had.charm decay, PYTHIA | Completed | 290001 - 290017 | 786,500 | Residual misalignment, all run staged |
| PDC 08/LHC08v | jet-jet pp, PYTHIA, 15 GeV/c < Pt hard < 50 GeV/c | Completed | 280001 - 280043 | 4,316,900 | All runs staged |
| PDC 08/LHC08u | gamma-jet pp (2), PYTHIA, no quenching | Completed | 260007 - 260036 | 3,036,000 | All runs staged |
| PDC 08/LHC08t | MUON Cocktail pp, MB | Completed | 170001 - 170444 | 191,858,500 | Ideal alignment, all runs are staged |
| PDC 08/LHC08s | p+p, beauty, with B->J/psi->ee decay | Completed | 200001 - 200003 | 197,400 | All runs staged |
| PDC 08/LHC08r | jet-jet pp, PYTHIA, Pt hard > 50 GeV/c | Completed | 270001 - 270028 | 2,900,000 | All runs staged |
| PDC 08/LHC08q | jet-jet pp, PYTHIA, hard > 100 GeV/c | Completed | 230002 - 230010 | 878,400 | All runs staged |
| PDC 08/LHC08p | gamma-jet pp, PYTHIA, quenching | Completed | 220001 - 260006 | 4,267,800 | All runs staged |

SLC4 and AliEn v.2-15

- We are migrating all build server and site installations to SLC4 (and gcc 3.4.6)
- This will solve a number of issues related to the use of the code compiled with gcc 3.2 on new systems
- In addition, we are trying to have a coherent and matching installation (32 or 64-bit) of the services, AliRoot/ Root /Geant, and the worker nodes in the local batch systems

Use action

- Supported version

AliRoot::v4-11-Rev-03, ROOT::v5-18-00b, GEANT3::v1-9-2

AliRoot::v4-12-Rev-02, ROOT::v5-19-02a, GEANT3::v1-9-3

AliRoot::v4-13-Rev-02, ROOT::v5-19-04, GEANT3::v1-9-4

AliRoot::v4-13-Rev-03, ROOT::v5-19-04, GEANT3::v1-9-4

Thank You