ALICE update

F.Carminati

GDB, July 20, 2005

Content

- Overview of SC3 (aka PDC05)
- Plans for SC3 sample jobs
- SC3 resource needs
- Planning updates
- Problems and issues (SC3 and LCG)

- Overview of SC3 (aka PDC05)
- Plans for SC3 sample jobs
- SC3 resource needs
- Planning updates
- Problems and issues (SC3 and LCG)

ALICE PDC'05

- Simulate till September 2005
- Register events in the FC
- Store at CASTOR@CERN
- Coordinate with SC3 framework

Goals

- Use the LCG SC3 infrastructure
- Test performance of SC3 data transfer and storage services
- Test of distributed reconstruction and calibration model
- Integrate all available resources within one single VO interfaced to different Grids
- Analysis (batch and interactive) of reconstructed data
- Mimic our data (MC data in raw-like format) flow

Success

- Meet stated goals
  - Failure: miss (any of) the goals
- Metrics
  - Efficiency (global and per component)
  - Stability as a function of time
  - Fraction of resources coming from centres hosting multiple VO's

- Overview of SC3 (aka PDC05)
- Plans for SC3 sample jobs
- SC3 resource needs
- Planning updates
- Problems and issues (SC3 and LCG)

UC1 Simulation

- Simulate events at T2's
  - Transfer Data to supporting T1's
- Simulate events anywhere
  - Transfer Data to T0

UC2 Reconstruction

- Get RAW events stored at T0 from catalogue
- First reconstruct pass at T0
- Ship from T0 to T1's
  - 500 MB/s out of T0
- Reconstruct at T1 with calibration data
- Store & catalogue the output in the global FC

- Overview of SC3 (aka PDC05)
- Plans for SC3 sample jobs
- SC3 resource needs
- Planning updates
- Problems and issues (SC3 and LCG)

CPU requirements

- Simulation 12h•85,000 central → 1Mh
  - 30,000 Pb-Pb (~ 24,000 central)
  - 100,000 Pb-Pb (~60,000 central)
  - 100,000 p-p (~1,000 central)
- Reconstruction → 15kh
- Assume 1,000 CPUs:
  - Reconstruction: 1 day
  - Simulation: 1 kh each (25-50 days occupancy)

Storage requirements

- Simulation & Reconstruction
  - 85,000 equivalent Pb-Pb central events @ 0.8GB
    - 68 TB (T0) + 68 TB (integral over T1s) = 136 TB
    - LCG fraction: to be defined, aiming at >= 80%
- Flow analysis
  - Simulation: 0.8 GB/event • 4,000 events/day = 3.2 TB/day@T1s
  - Reconstruction 22 MB/event • 3,600 events/h = 80 GB/h = 1.92 TB/day@T1s

Network flow

- Simulation
  - Assume 1000 CPUs at T2s and central Pb-Pb
  - 1 Job = 1 Event
  - In: few kB of configuration – Out: 0.8GB on T1 SE
  - Job duration: 6 h -> 4000 Jobs/day -> 3.2 TB/day (40 MB/s) from T2s -> T1s
- Remark
  - This will cover only a part of our simulation needs
  - The rest will be done also at T1's

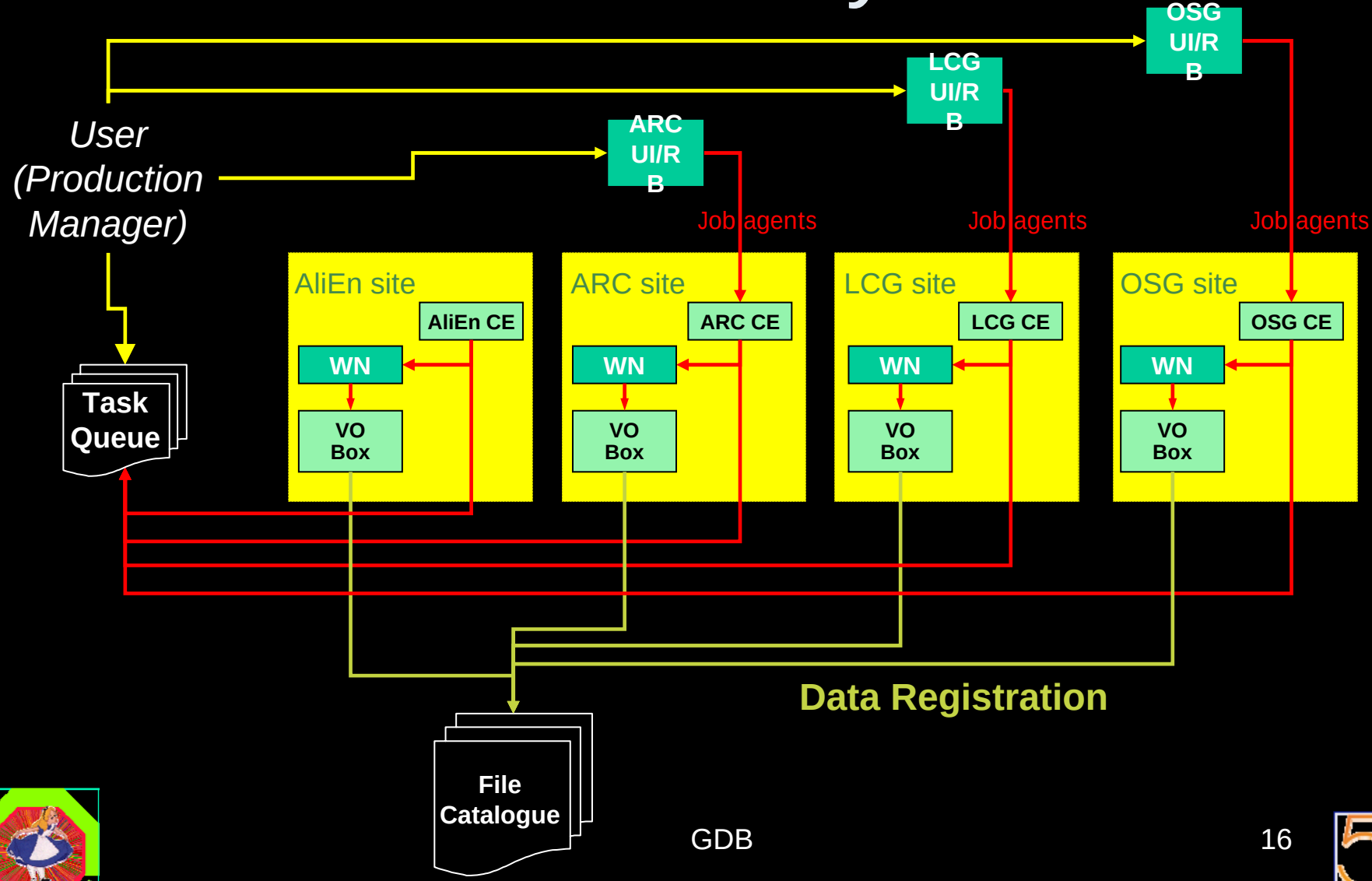
Reconstruction Network flow

A. Process@T0 & push to T1s
   - Central Pb-Pb events, 300 CPUs at T0
   - 1 Job = 1 Input event
   - In: 0.8 GB @ T0 SE – Out: 22 MB @ T0 SE
   - 10 min / job -> 1800 Jobs/h -> 1.4 TB/h (400 MB/s) from T0 -> T1s

B. Push to T1s & run at T1s
   - Central Pb-Pb events, 600 CPUs at T1s
   - 1 Job = 1 Input event
   - In: 0.8 GB @ T0 SE – Out: 22 MB @ T1 SE
   - 10 min/job -> 3600 Jobs/h -> 2.8 TB/h (800 MB/s) T0 -> T1s

- Overview of SC3 (aka PDC05)
- Plans for SC3 sample jobs
- SC3 resource needs
- Planning updates
- Problems and issues (SC3 and LCG)

ALICE SC3 layout



LCG interfacing

- gLite FTS service for data placement
  - Test in SC3
  - Implementation to allow choice of higher level components
- ALICE File Catalogue (central and local)
  - Interface to LCG LFC
  - Evaluate Fireman
- VOMS to manage VO
- gLite SE to access storage via SRM

LCG interfacing (cont)

- gLite RB to submit job agents to LCG
  - Options under study (# of RB, config…)
  - Jobs get execution information from the TQ via the ClusterMonitor on the WN
- Interface to gLite CE
  - Need outbound WN connection, possibly via proxy
- Monitoring via job agent and VO services

Services for SC3

VO-box deployment on SC3 sites	ALICE
ROOT/AliRoot deployment on SC3 sites	ALICE
AliEn Top Level Services	ALICE
UI(s) for submission to LCG/SC3	ALICE/LCG
File Catalog	ALICE/LCG
WMS (n RBs) + CE/SE Services on SC3	LCG
LFC instances on all SC3 sites, seen from WNs/VO-boxes	LCG
xrootd	ALICE/LCG
gliteFTS accessible from all WNs/VO-boxes	LCG
SRM (DPM) accessible from all WNs/VO-boxes	LCG
SC3 resources for ALICE (Computing/Storage)	LCG
Appropriate JDL files for the different tasks	ALICE

Status of deployment

- ✓ AliEn Central Services
- ✓ VO-box deployment on SC3 sites
  - ✓ OK in CERN (temporary) Catania, Torino and Bari
  - ✓ Allocated at: Lyon, CNAF, FZK, GSI
  - To be asked: NIKHEF, RAL
- ✓ AliROOT deployment: automated (AliEn PackMan)
- UI(s) for submission to LCG/SC3
  - ✓ Available: Catania, Torino
- WMS (n RBs) + CE/SE Services on SC3
- LFC instances on all SC3 sites, seen from WNs/VO-boxes
- gLiteFTS accessible from all WNs/VO-boxes
- SRM (DPM) accessible from all WNs/VO-boxes
- Deployment also on Itanium started

Early activity

- FTS server (fts-alice-test) for data transfers with
  - lcg-infosites
  - LFC
  - Some SRM endpoints available for storing data
- Test gLite RB & CE in Torino
  - test job submission, interaction with file catalogue and, gradually, other pieces of the framework
- Tests of storage and data management in Bari
  - dCache+SRM, FTS
  - Integrate with the Torino setup to build a full testbed

- Overview of SC3 (aka PDC05)
- Plans for SC3 sample jobs
- SC3 resource needs
- Planning updates
- Problems and issues (SC3 and LCG)

Issues

- The gLite UI command does not interact correctly with the VOMS.
  - known problem, fixed but the fix did not get through to the release (not even 1.1)
- Submission to the gLite RB fails with certificates mapped to a SGM (software manager) account (Savannah bug #8616)
- gLite RB interacts correctly with LCG 2.4.0 CEs but there are still some problems
- Responsiveness of EGEE developers has improved
- Support from LCG is good

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Courtesy of I. Bird, LCG PEB, Jun, 7th 2005

Baseline services

- Storage management services
  - Based on SRM as the interface
- Basic transfer services
  - gridFTP, srmCopy
- Reliable file transfer service
- Grid catalogue services
- Catalogue and data management tools
- Database services
  - Required at Tier1,2
- Compute Resource Services
- Workload management
- VO management services
  - Clear need for VOMS: roles, groups, subgroups
- POSIX-like I/O service
  - local files, and include links to catalogues
- Grid monitoring tools and services
  - Focussed on job monitoring
- VO agent framework
- Applications software installation service
- Reliable messaging service
- Information system

VO Box requirements

- at least one normal user account (no super-user privileges) with access via ssh
- optimal configuration: two accounts, belonging to the same group
- directory seen by the VO box shared among the WNs, min 5 GB disk space (it can be $VO_ALICE_SW_DIR)
- outbound connectivity
- inbound connectivity from CERN on one fixed network port
- inbound connectivity from World on two fixed network ports
- local tactical data buffer (local disk, LCG deployed Disk Pool Manager, NFS mounted disk) for intermediate input and output data storage of jobs. The buffer size is at least number of jobs slots on the site * 3GB. This buffer is not necessary if xrootd is running on the site storage element.
- Linux kernel 2.4 or higher, any Linux flavour on i386, ia64 or Opteron
- hardware: min. PIII 2GHz, 1024 MB RAM

Monitoring on the VO node

- Storage Element Service (SES) interface to local storage (via SRM or directly)
- File Transfer Daemon (FTD) scheduled file transfers agent (possibly using FTS implementation)
- xrootd – application file access
- Cluster Monitor (CM) – local queue monitoring
- MonALISA – general monitoring agent
- PackMan (PM) – application software distribution and management
- Computing Element (CE)