



# The ALICE Online Data Storage System

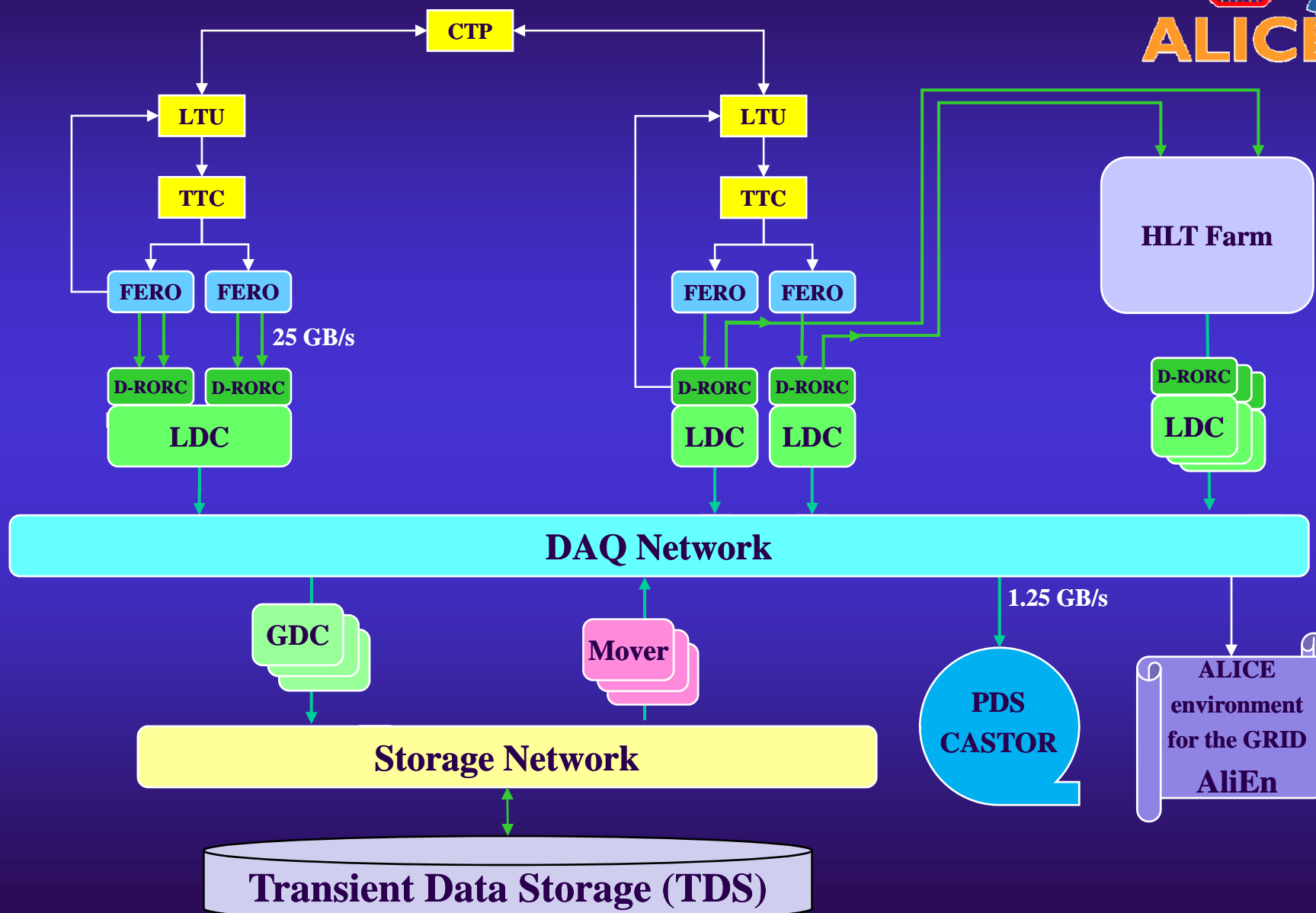
Roberto Divià (CERN), Ulrich Fuchs (CERN), Irina Makhlyueva (CERN), Pierre Vande Vyvre (CERN)

Valerio Altini (CERN), Franco Carena (CERN), Wisla Carena (CERN), Sylvain Chapeland (CERN), Vasco Chibante Barroso (CERN), Filippo Costa (CERN), Filimon Roukoutakis (CERN), Klaus Schossmaier (CERN), Csaba Soòs (CERN), Barthelemy Von Haller (CERN)

**For the ALICE collaboration**

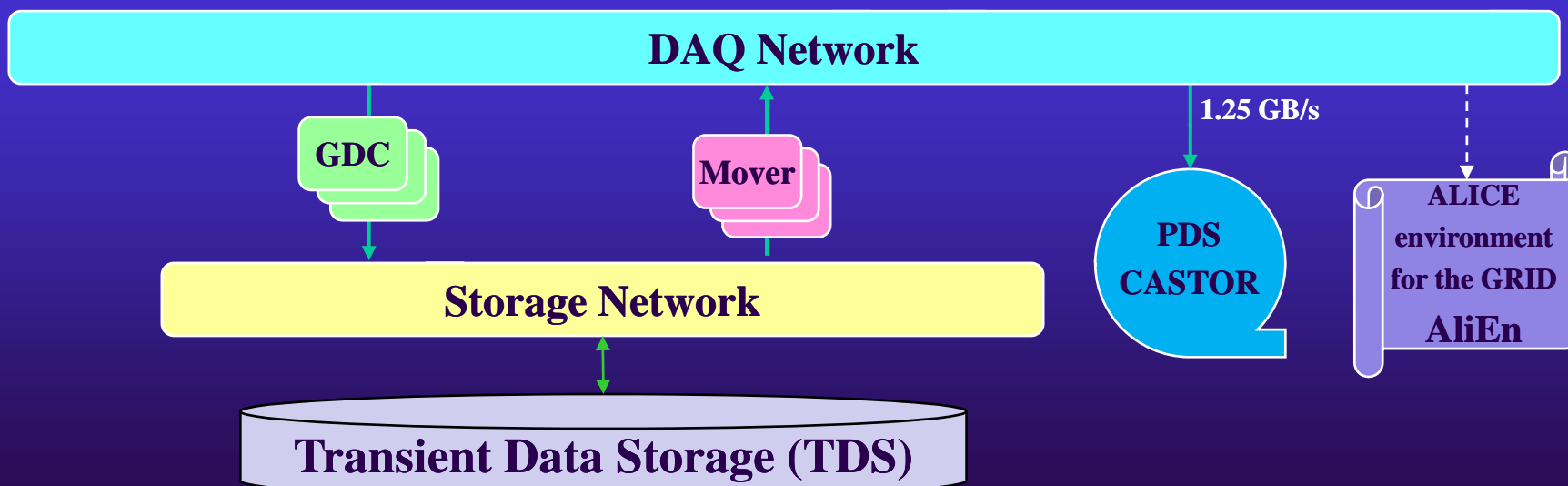


# ALICE trigger, DAQ & HLT





# ALICE trigger, DAQ & HLT

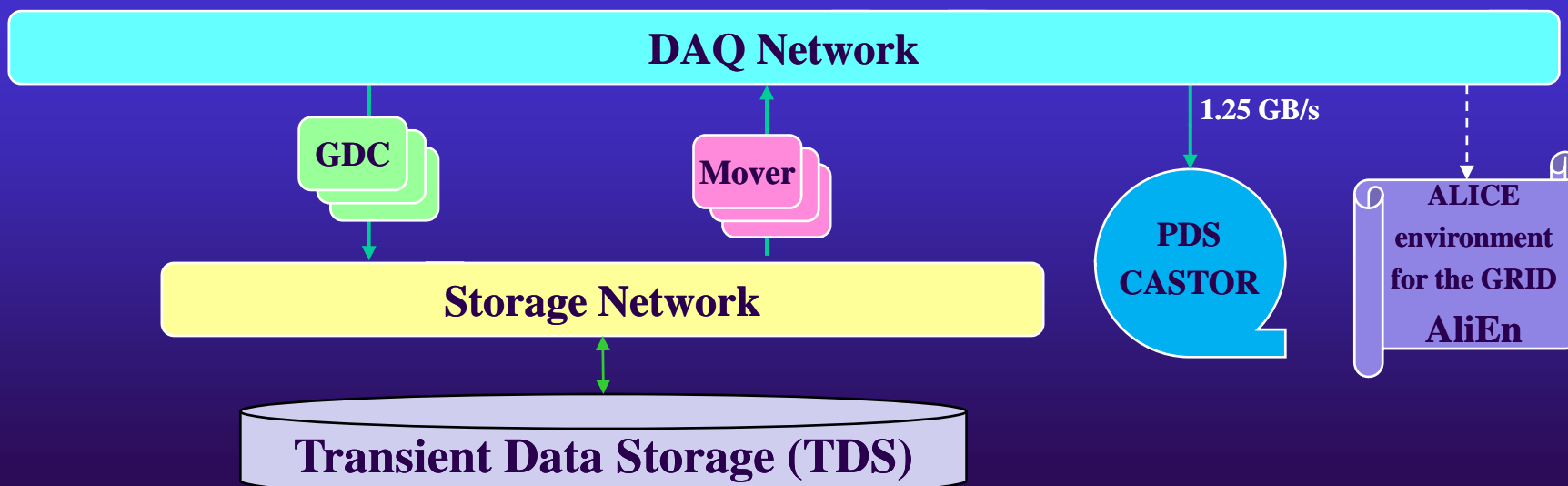




# Our objectives



- Ensure steady and reliable data flow up to the design specs
- Avoid stalling the detectors with data flow slowdowns
- Give sufficient resources for online objectification in ROOT format via AliROOT
  - very CPU-intensive procedure
- Satisfy needs from ALICE parallel runs and from multiple detectors commissioning
- Allow a staged deployment of the DAQ/TDS hardware
- Provide sufficient storage for a complete LHC spill in case the transfer between the experiment and the CERN Computer Center does not progress





# Current TDS architecture



5 \* ( 6 \* GDCs + 2 \* Movers )

CVFS over IP



5 switches Qlogic SANBox 5602:

- FC 4 Gb: equipment, PCs, storage
- FC 10 Gb: inter-switches connections

- 5 \* 5 Disk Arrays Infotrend A16F models G2422 & G2430
- 5 \* 15 disk volumes
- Total maximum space: 59 TB
- CVFS: StorNext 3.1.2
- Handled by the Transient Data Storage Manager (TDSM)

# Future upgrades

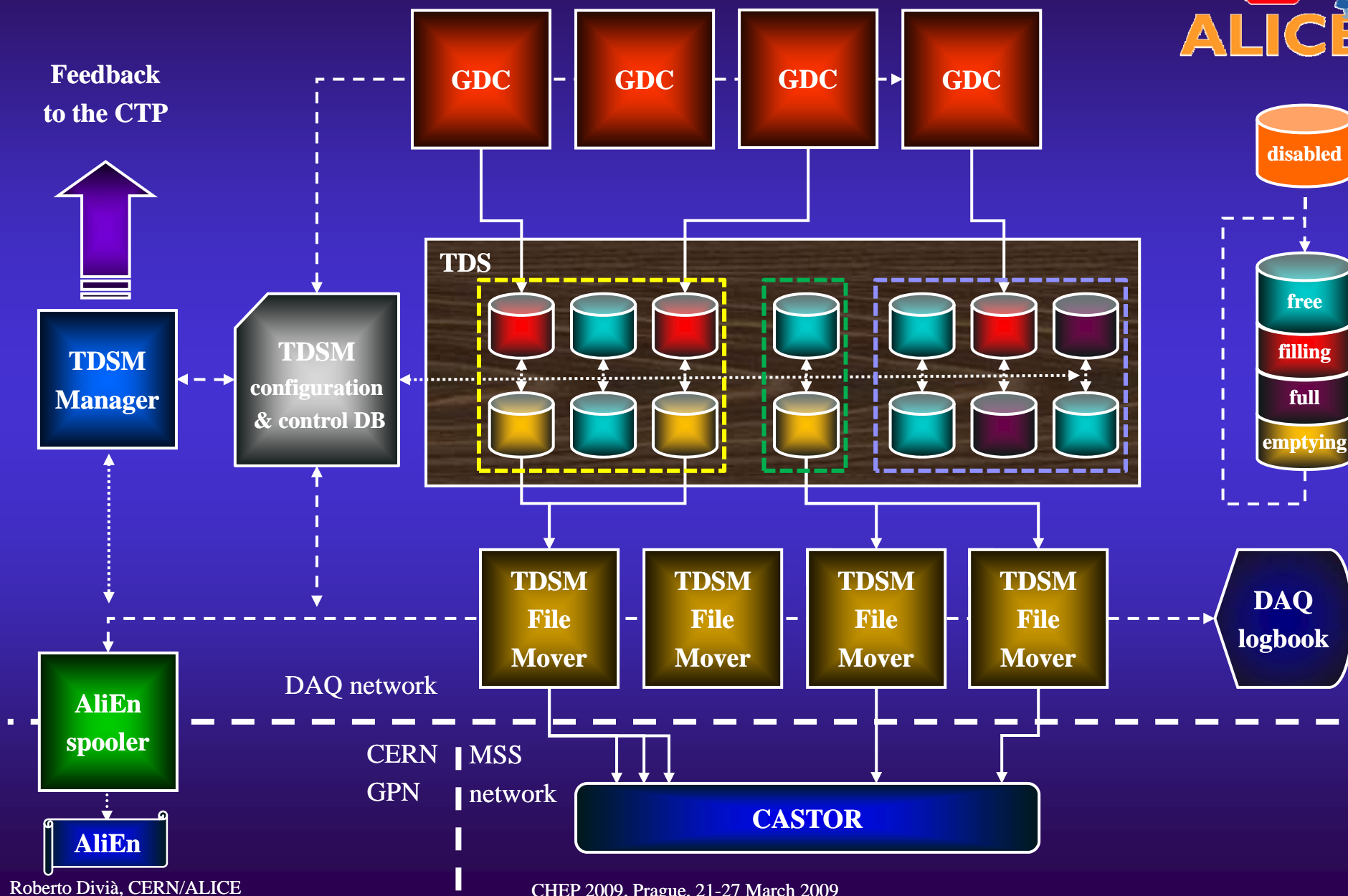


Two switches SANBox 9000, 8 blades maximum each:

- 9 \* Blades with 16 ports FC 4 Gb: equipment, hosts, storage
- 2 \* Blades with 4 ports FC 10 Gb: inter-switches connections



# TDSM architecture







# Monitoring



TDSM DB to get the status and history of the TDS/TDSM

Stats [2009-02-27 20:21:57 .. 2009-03-12 08:37:00]:  
 - Moved OK:96.8% ERR:3.2%  
 - Same volume/different volume transfers: 40.2% / 59.8%  
 - Remove: [0 .. 0.2 .. 4] seconds

	10'										30'									
	Num ops	Throughput	Mv OK	Mv Err	MD5 OK	MD5 Err	RM OK	RM Err	RM avgT	RM maxT	Num ops	Throughput	Mv OK	Mv Err	MD5 OK	MD5 Err	RM OK	RM Err	RM avgT	RM maxT
By TDSM mover																				
pcald48	3	1.63 MB/s	3				3		0 s	0 s	7	2.05 MB/s	7				7		0 s	0 s
pcald49	3	1.56 MB/s	3				3		0.33 s	1 s	7	903.11 KB/s	7				7		0.14 s	1 s
pcald50	6	1.99 MB/s	6				6		0.50 s	1 s	14	2.07 MB/s	14				14		0.21 s	1 s
pcald51	3	1.80 MB/s	3				3		0.33 s	1 s	7	1.70 MB/s	7				7		0.43 s	2 s
pcald52	0										0									
By TDSM mover group																				
rack1	9	1.92 MB/s	9				9		0.44 s	1 s	21	1.92 MB/s	21				21		0.29 s	2 s
rack2	6	1.60 MB/s	6				6		0.17 s	1 s	14	1.32 MB/s	14				14		0.07 s	1 s
By volume																				
/localmd/rack1/vol1	0										0									
/localmd/rack1/vol2	0										0									
/localmd/rack1/vol3	0										0									
/localmd/rack1/vol4	0										0									
/localmd/rack1/vol5	0										0									
/localmd/rack1/vol6	0										0									
/localmd/rack1/vol7	0										0									
/localmd/rack1/vol8	0										0									
/localmd/rack2/vol1	5	1.31 MB/s	5				5		0.20 s	1 s	10	1.60 MB/s	10				10		0.30 s	2 s
/localmd/rack2/vol2	5	2.66 MB/s	5				5		0.40 s	1 s	10	1.64 MB/s	10				10		0.20 s	1 s
/localmd/rack2/vol3	0										5	2.32 MB/s	5				5		0 s	0 s
/localmd/rack2/vol4	5	1.82 MB/s	5				5		0.40 s	1 s	5	1.82 MB/s	5				5		0.40 s	1 s
/localmd/rack2/vol5	0										0									
/localmd/rack2/vol6	0										0									
/localmd/rack2/vol7	0										0									
/localmd/rack2/vol8	0										5	1.24 MB/s	5				5		0 s	0 s
By volume group																				
rack1	0										0									
rack2	15	1.78 MB/s	15				15		0.33 s	1 s	35	1.65 MB/s	35				35		0.20 s	2 s





# Monitoring



Logbook to monitor the status of the migration

**ALICE Electronic Logbook** v1.38

Welcome Roberto Divia (PH/AID)

Logbook ▾ Runs ▾ Actions ▾ Admin ▾ Links ▾ Logout

**Run Details - 56001**

◀ 56001 ▶ Quick Access  Print

Run Conditions Run Statistics Trigger Clusters Info Run Quality LDCs Statistics GDCs Statistics Shuttle Info **File Info** Log Entries

**File Status**

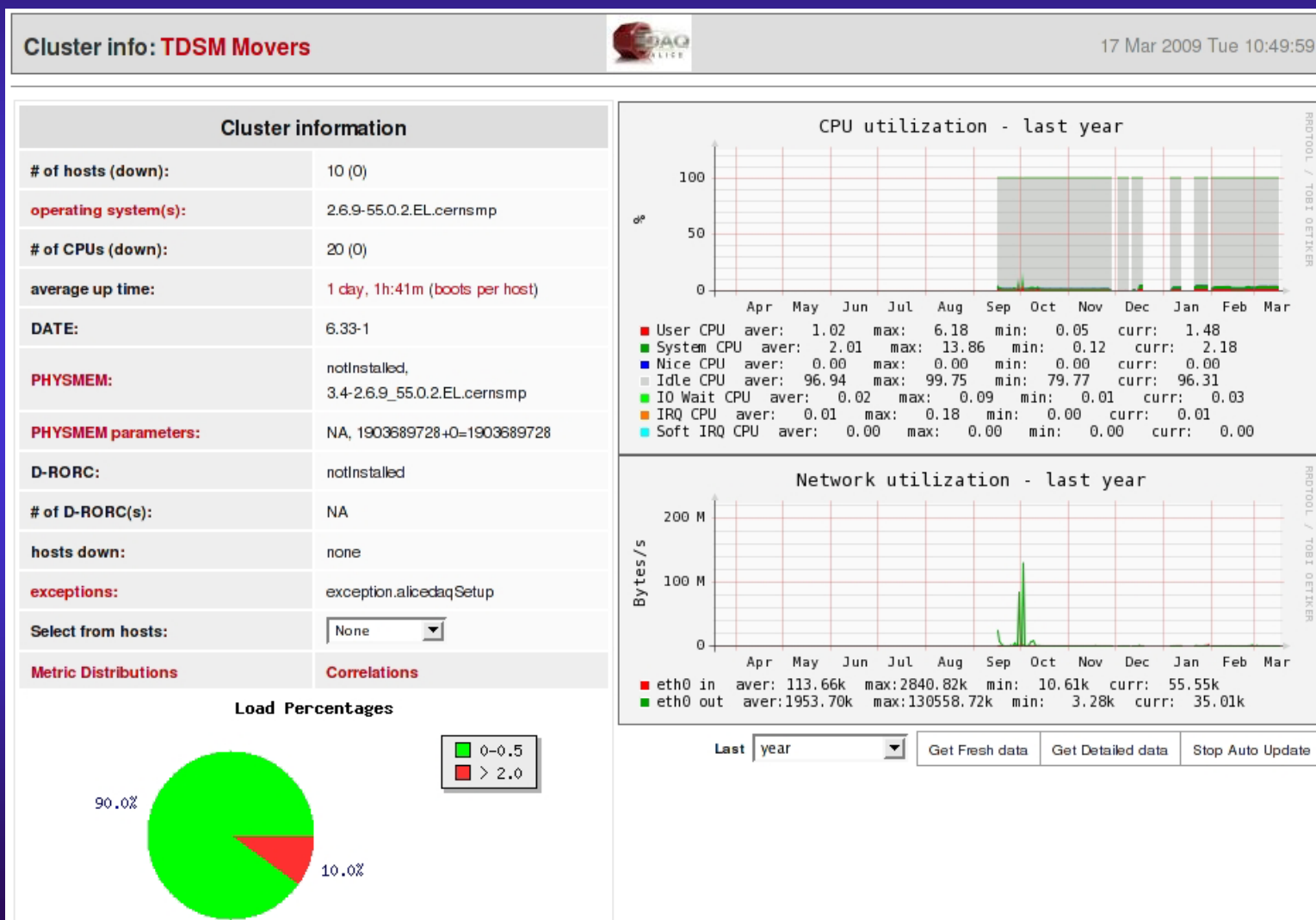
Status	# Files	Size	# Events
Writing	0	0 KB	0
Waiting migration request	0	0 KB	0
Migration requested	0	0 KB	0
Migrating	0	0 KB	0
Migrated	85	736.3 GB	7 909 475



# Monitoring



Lemon to monitor the system setup and metrics





# Validation & testing



- ◆ Small “lab style” test setups
  - Special running mode where:
    - a single LDC can inject several real events
    - the Event builder unpacks it as for the original event
  - Dedicated “write and forget” CASTOR pool
  - Ad-hoc “black hole” AliEn registration service
- ◆ Profiling during detectors commissioning and cosmic runs
- ◆ ALICE Data Challenges
  - Run between 1999 and 2006
  - Periodic full-chain tests (ALICE DAQ/Offline + IT department)



# ALICE Data Challenges



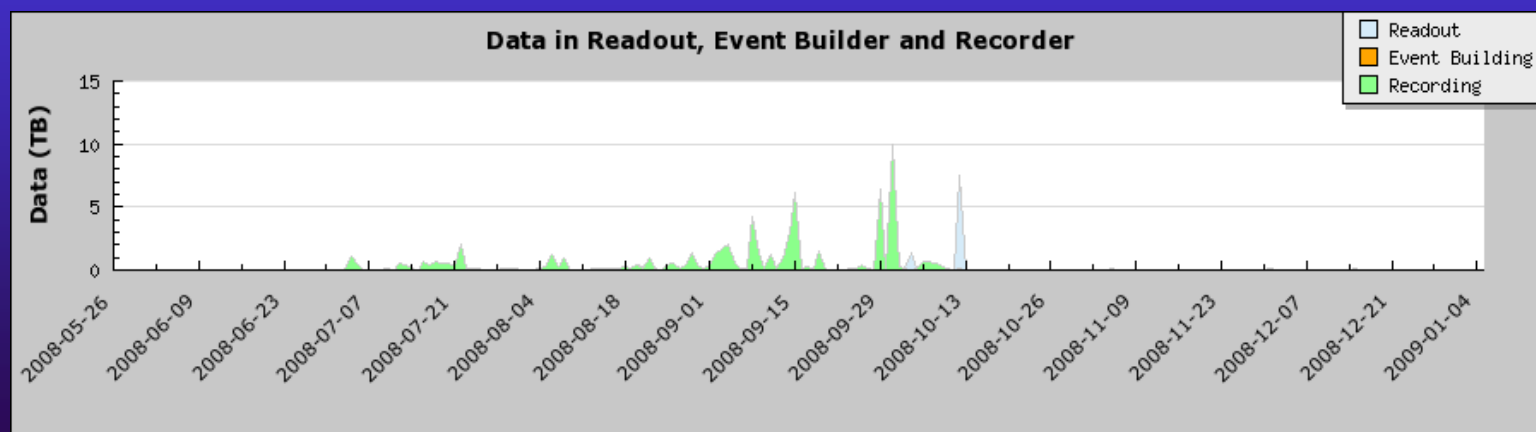
- ◆ Define an architecture (HW and SW)
  - Re-use existing idling components (IT and ALICE)
  - If needed, add some glue here and there
    - Earlier ADCs: lots of glue!
- ◆ Evaluate and profile the individual components
- ◆ Put them together and check the result
- ◆ Do short sustained tests (hours)
- ◆ Run the final Challenge (7 days) with two targets:
  - Sustained overall data rate
  - Amount of data to PDS
- ◆ Repeat the exercise year after year with more challenging objectives
- ◆ Achieve quasi-ALICE results with minimum glue right before ALICE commissioning



# The TDS in 2008



- ◆ 25 February to 9 March 2008: ALICE Cosmic runs
  - 1500 runs
  - 340 hours
  - 70 TB
- ◆ 3+4Q08:
  - 6800 runs
  - 3300 hours
  - 108 TB





# In conclusion...



- ◆ Continuous evaluation of HW & SW components proved the feasibility of the TDS/TDSM architecture
- ◆ All components validated and profiled
- ◆ ADCs gave highly valuable information for the R&D process
  - Additional ADCs added to the ALICE DAQ planning for 2009
- ◆ Detector commissioning went smoothly & all objectives were met
- ◆ No problems during cosmic and preparation runs
- ◆ Staged commissioning on its way
- ◆ Global tuning in progress

We are ready for LHC startup