CASTOR2 performance and reliability

T0 tasks for Castor2

- **Central Data recording**
  → RAW data made available on a large disk cache (~1 week buffer) and stored on tape

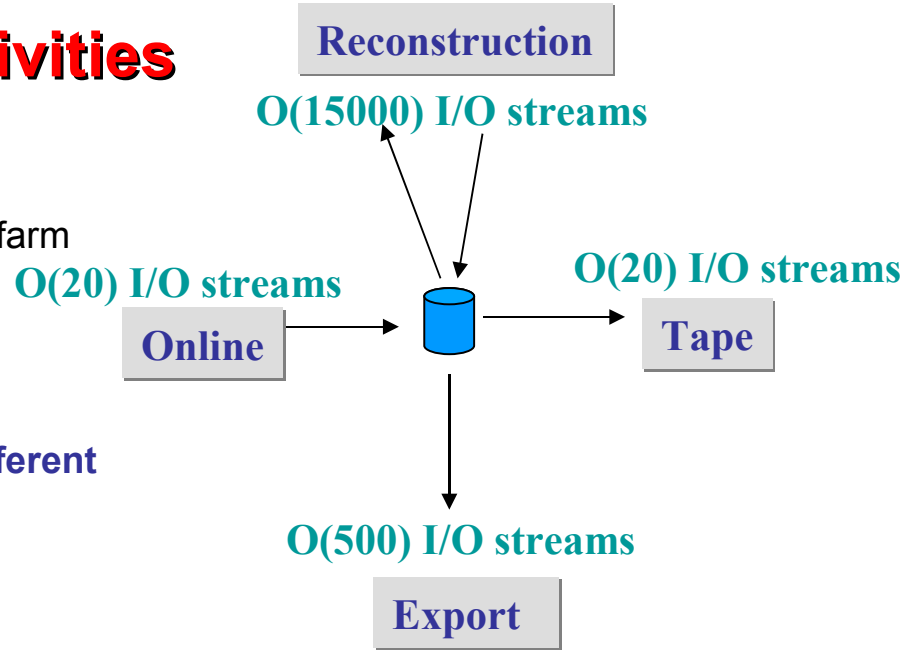
- **First-pass reconstruction of the RAW data plus storage of the derived ESD+AOD data on a disk cache and on tape**

- **Calibration and alignment of the detector data as a prerequisite for the RAW data reconstruction (mixed with the CAF)**
  → focus in 2007/08

- **Export of the RAW+ESD+AOD data to the T1 centers.**
  One copy of the RAW data, ~2 copies of the ESD data, ~10 copies of the AOD data

The focus of the 2006 high performance and throughput tests with Castor2 were on the anticipated T0 needs

Four different concurrent T0 activities



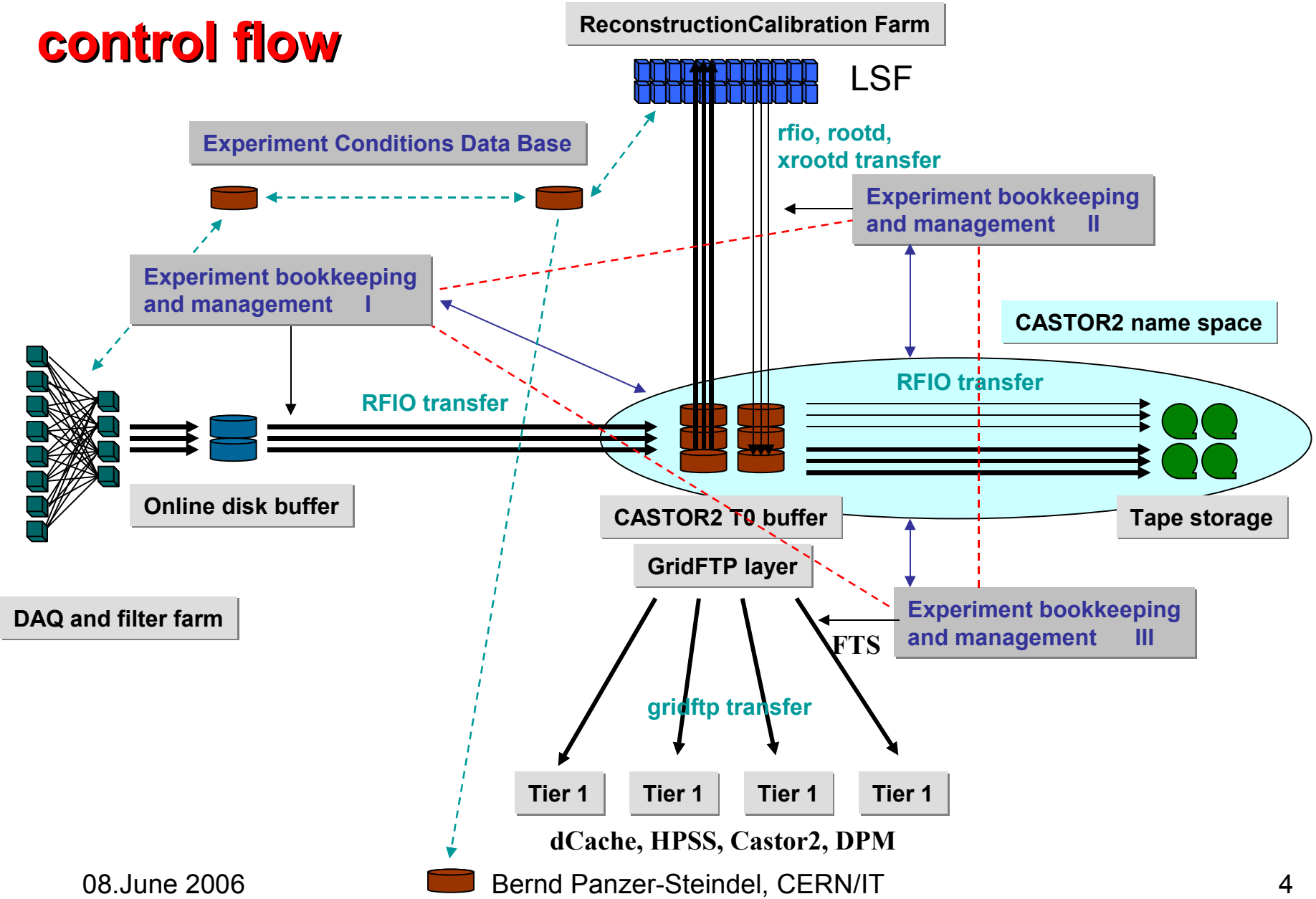
3. data transfer from the 'DAQ' buffer to the T0 buffer
4. data transfers from the T0 buffer to the reconstruction farm and derived data back to the T0 buffer
6 data migration from the T0 buffer to the tape system
7 data export from the T0 buffer to the T1 sites

each of the flows has its own characteristics and is different for each of the 4 experiments

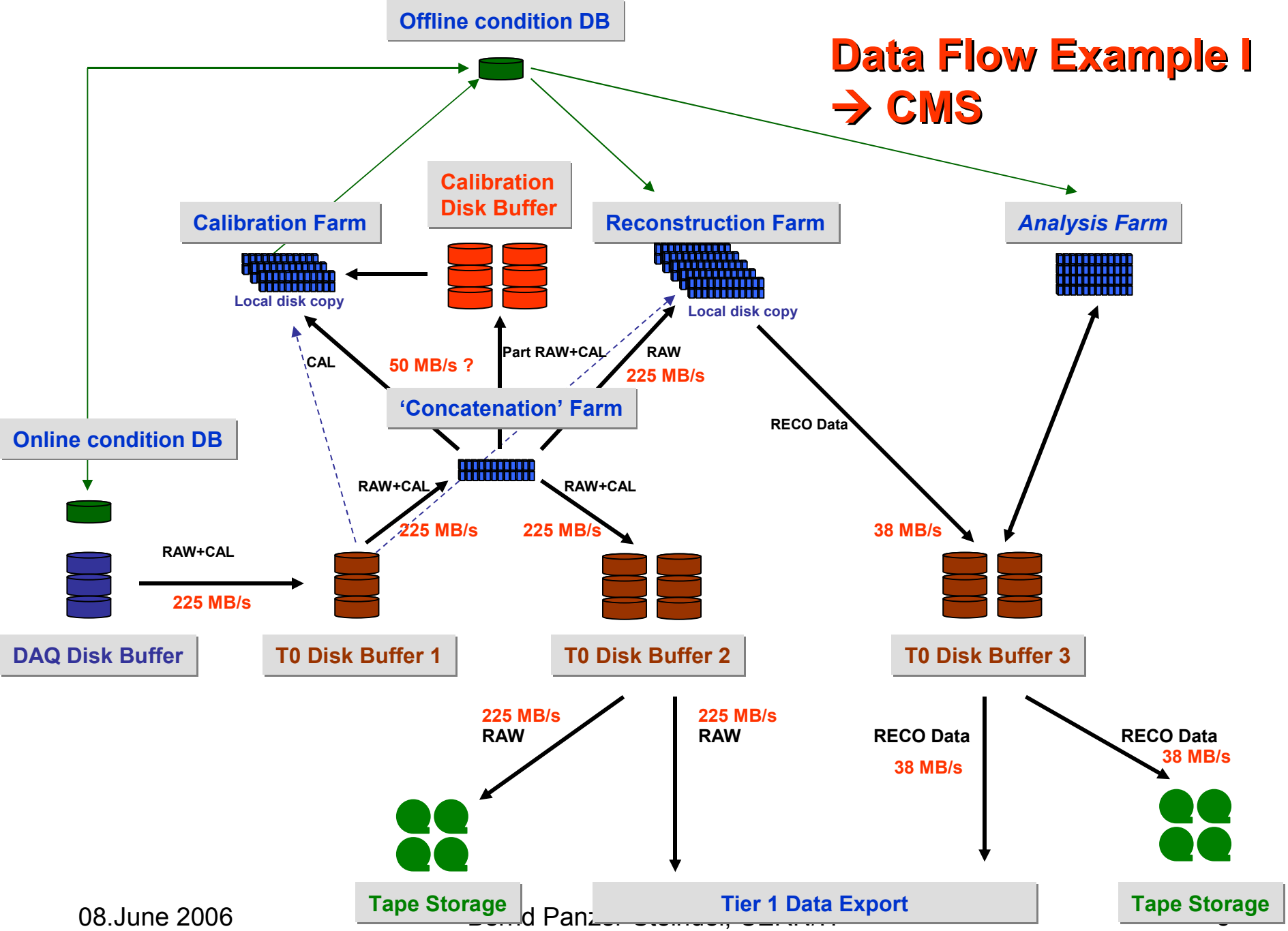
multi-dimensional 'impedance' matching problem

	DAQ [MB/s]	Tape [MB/s]	Reconstr. [MB/s]	Export [MB/s]	Total [MB/s]
ALICE HI	**1250**	**1250**	**300**	**300**	**~ 3000 HI ~600 pp**
ATLAS	**320**	**440**	**540**	**780**	**~ 2100**
CMS	**225**	**270**	**270**	**315**	**~ 1100**
LHCb	**60**	**40**	**35**	**35**	**~170**

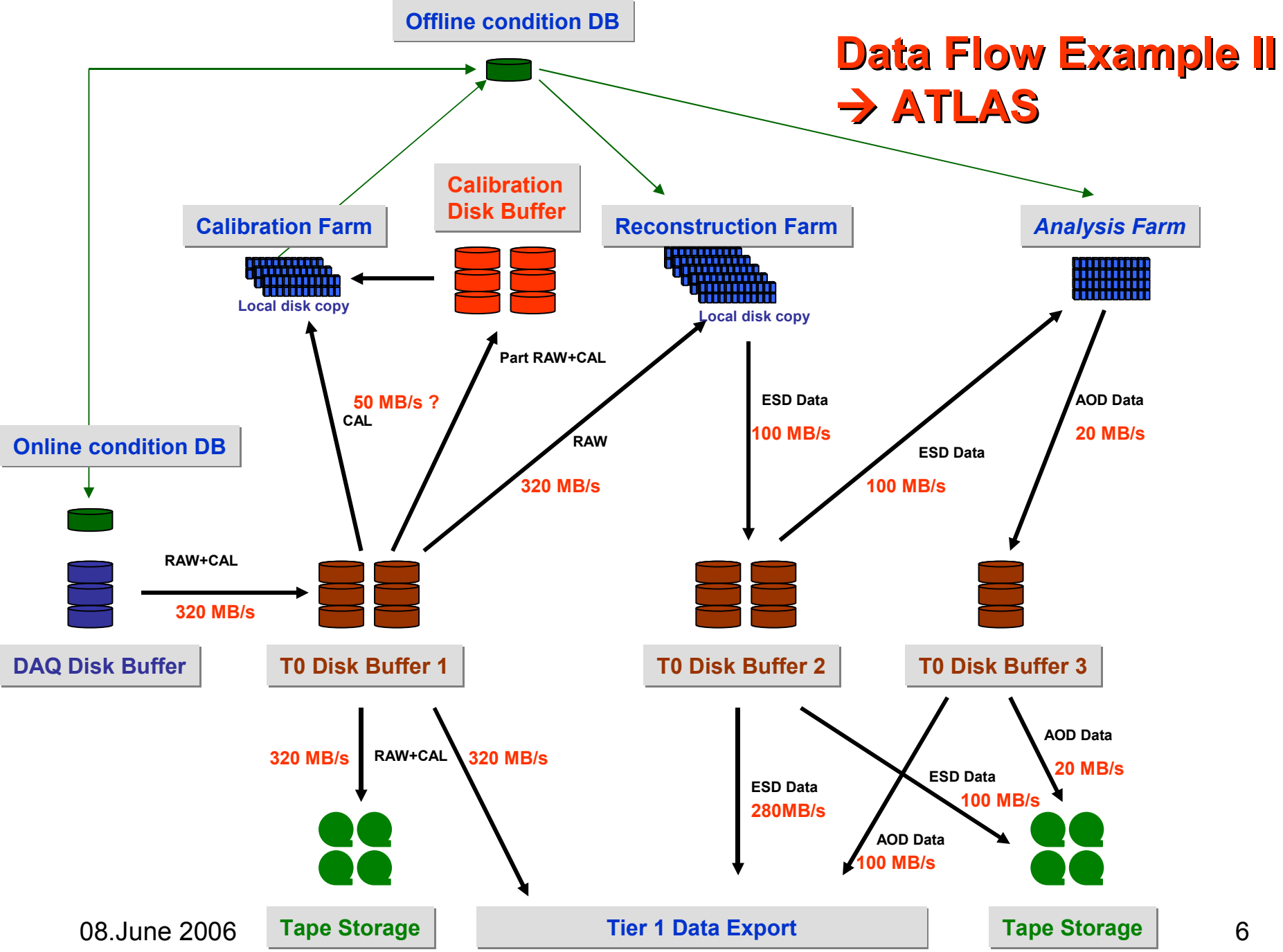
General T0 data and control flow



Data Flow Example I → CMS



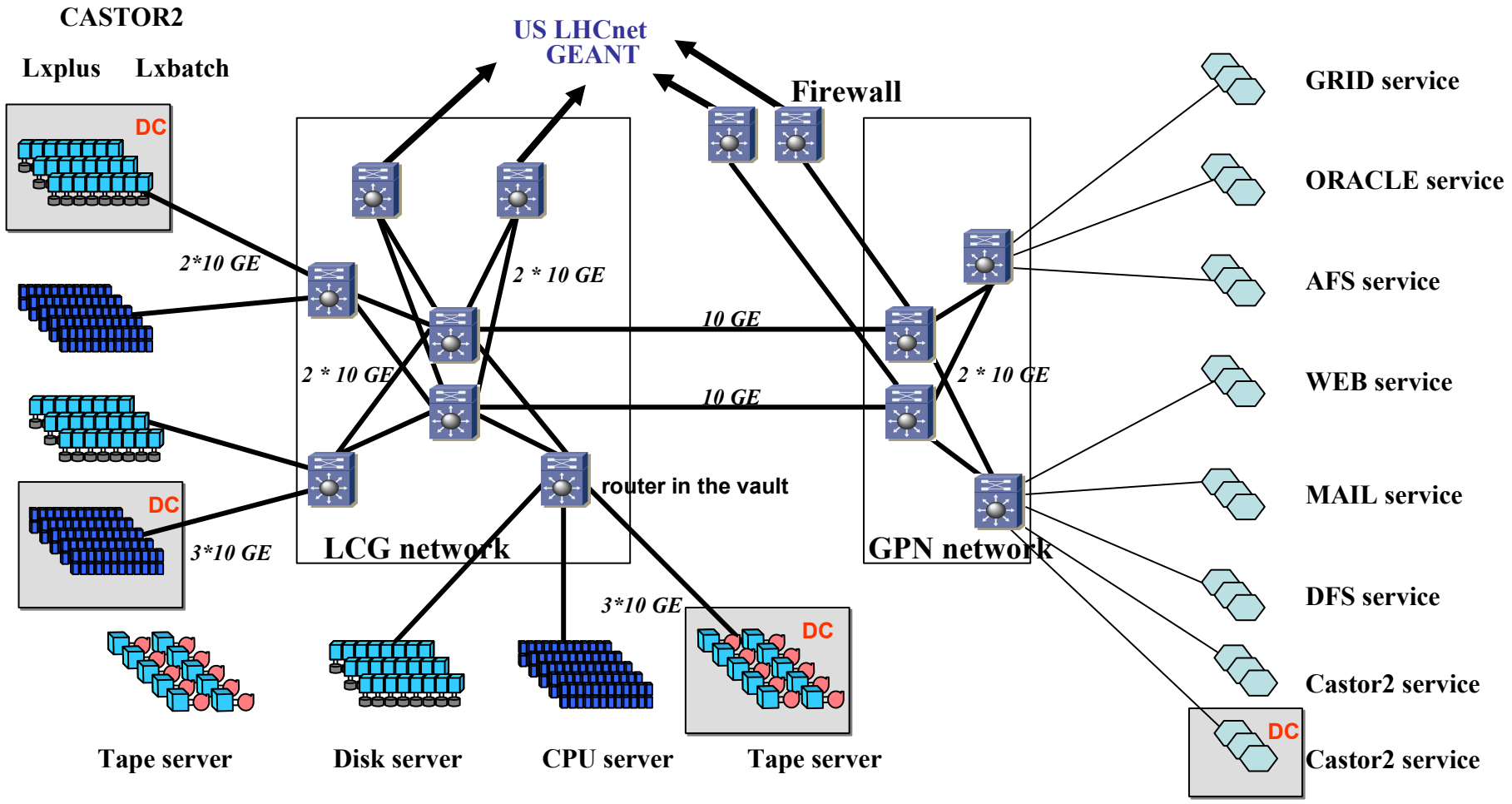
Data Flow Example II → ATLAS



Test setup

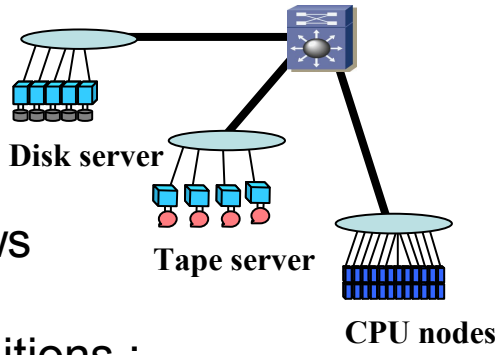
- dedicated Castor2 instance → ITDC
- one stager
- same name server as the production system
- reuse of tapes after ~24h
- ~ 40 disk server distributed over three different network switches
- ~ 30 – 40 tape drives, IBM 3592B, STK T10K, LTO-3
- ~ 200 CPU nodes as data sources and sinks
- emulation programs for the 4 different data streams (IT and experiment)

Network Topology (March 2006)



- Dedicated resources for the Data Recording Challenges (CPU, disk, tape, Castor2)
- 'logical' separation between DC setup and production systems

Performance 'definition'



The idea of Castor2 is to 'couple' disk server, tape server and clients to enable efficient and high performance data flows

This needs to take into account all the existing boundary conditions :

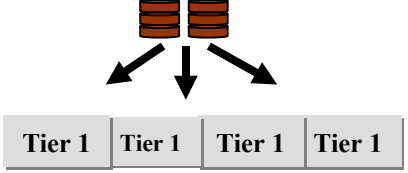
- disk server performance characteristics
  - → single disk , RAID5 , file system (ext3, xfs), Linux IO and virtual memory, hardware IO sub-system (disk controller, network interface, bus)
  - → 'footprint' and efficiency of the IO daemons (rfiod, rootd, xrootd, gridftp)
  - → **Castor2 daemons and programs on the server (permanent and per request)**

- tape server performance characteristics
  - → tape drive hardware and firmware
  - → tape server hardware (memory, network, Linux)
  - → **Castor2 tape software**

- **Castor2 stager nodes and name server performance** (e.g. Oracle)

Input parameter for the load balancing and scheduling policies

Results I : SC3 Throughput rerun



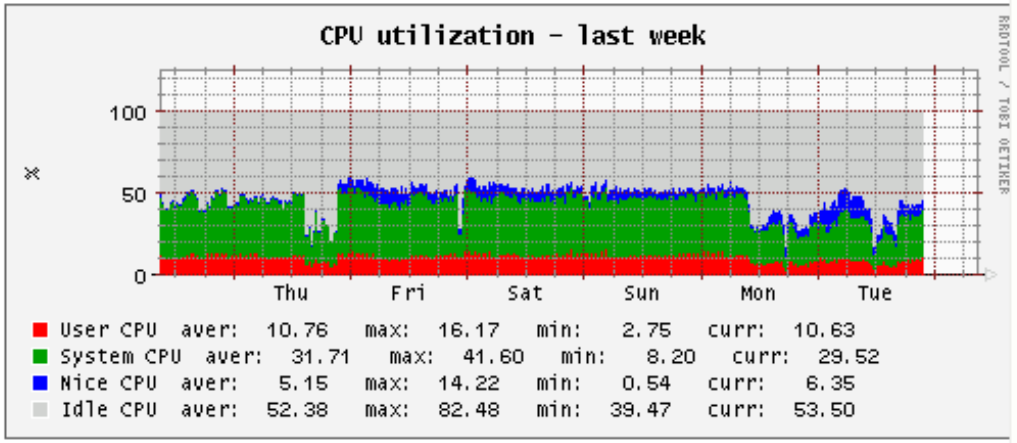
Cluster info: sc3 — 25 Jan 2006 Wed 08:43:33

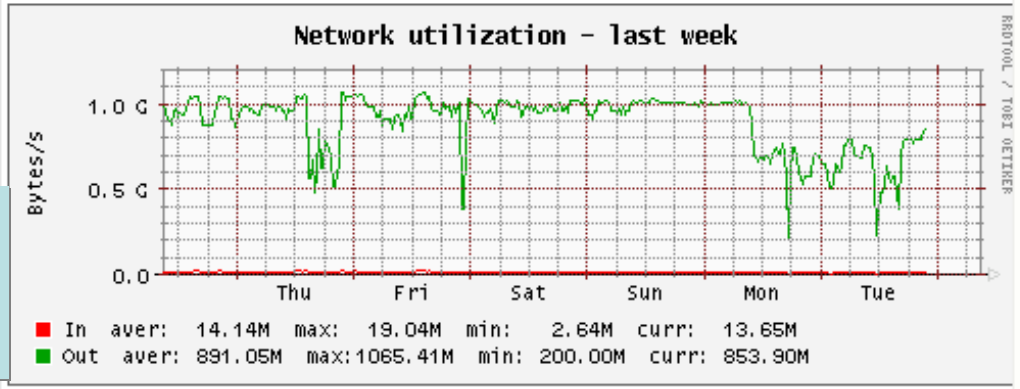
Unable to connect to the Oracle server. Cluster information will be skipped.

Cluster Information	
# of hosts (down):	16 (16)
operating system(s):	
# of CPUs (down):	0 (0)
average up time:	0h:00m (boots per host)
hosts down:	lxfsra2801, lxfsra2807
ITCM history	View template
Select from hosts:	None
Metric Distributions	Correlations

Load Percentages

down

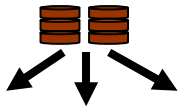




Last week | Get Fresh data | Get Detailed data | Auto Update

16 disk server, pre-staged data reused, SRM and GridFTP on the disk servers, many disk I/O streams

Results II : SC4 throughput challenge

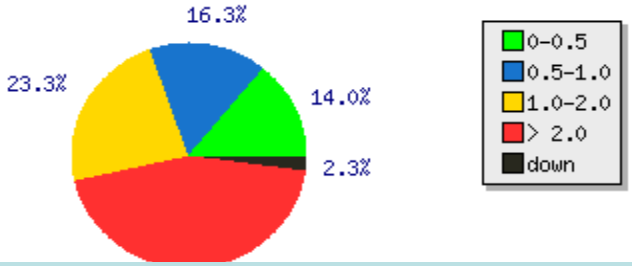


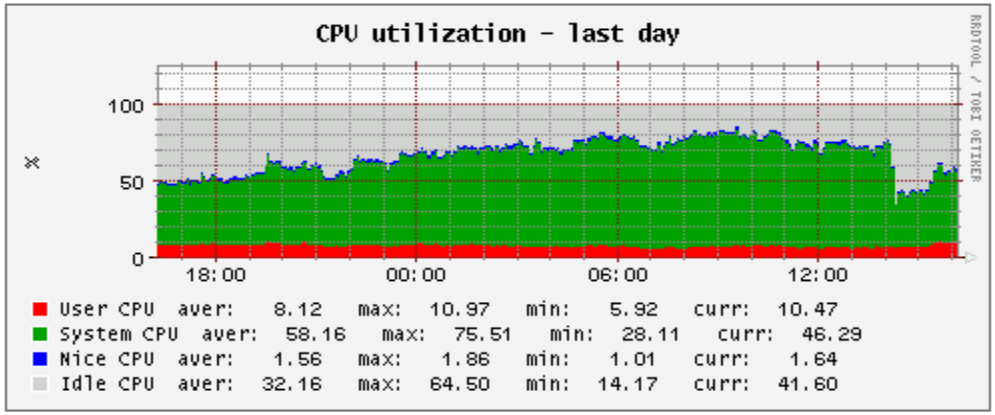
Tier 1 | Tier 1 | Tier 1 | Tier 1

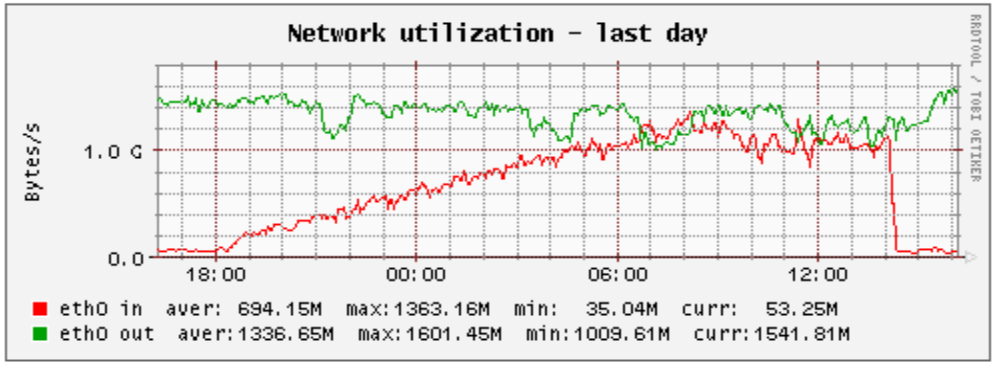
Cluster info: c2sc4 subcluster wan — 19 Apr 2006 Wed 16:14:27

Cluster Information	
# of hosts (down):	43 (1)
operating system(s):	2.6.9-34.EL.cernsmp
# of CPUs (down):	51 (2)
average up time:	20 days, 18h:50m (boots per host)
hosts down:	lxfsra3004
exceptions:	RPC_STATD_WRONG, MIRROR_BROKEN,
ITCM history	View template
Select from hosts:	None
Metric Distributions	Correlations

Load Percentages







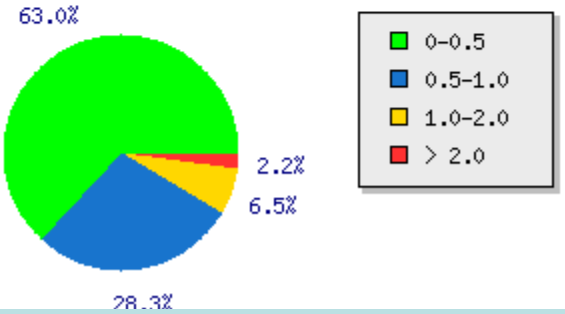
40 disk server, pre-staged data reused, SRM and GridFTP on the disk servers, many disk I/O streams
added emulation of DAQ input streams (~100)

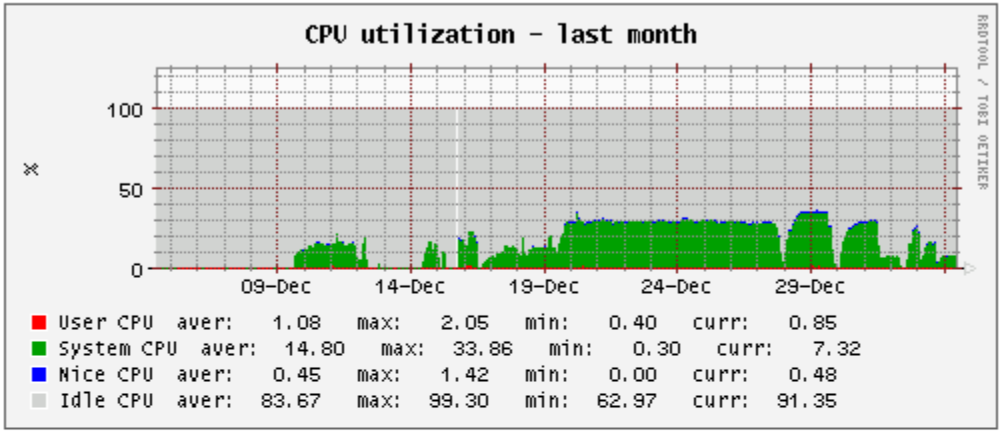
Results III : DAQ → T0 buffer → Tape Data Challenge I

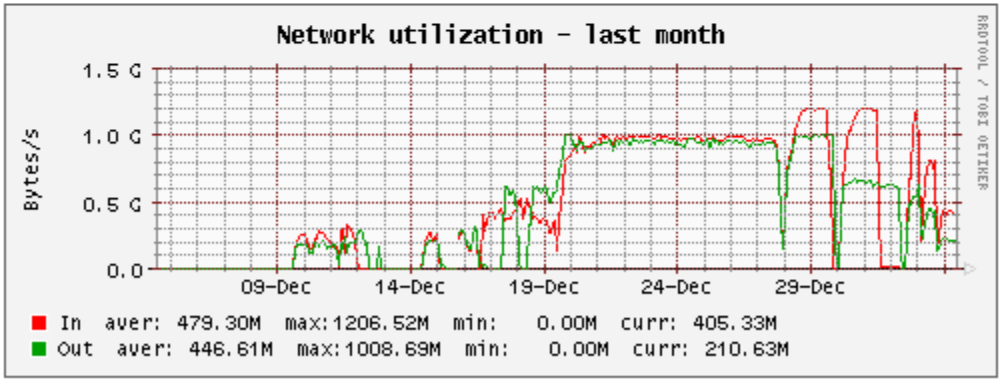
Cluster info: ITDC — 03 Jan 2006 Tue 10:19:12

Cluster Information	
# of hosts (down):	46 (0)
operating system(s):	2.4.21-37.EL.cernsmp
# of CPUs (down):	62 (0)
average up time:	47 days, 14h:14m (boots per host)
hosts down:	none
exceptions:	FILESYSTEM_ERROR
ITCM history	View template
Select from hosts:	None
Metric Distributions	Correlations

Load Percentages

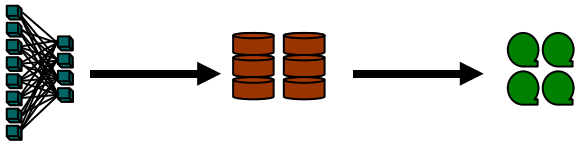






34 CPU data sources + 46 disk server (~230 TB) + 18 tape drives (3 different types)
→ stable running for >= one week at 950 MB/s , input data rate > 1 GByte/s

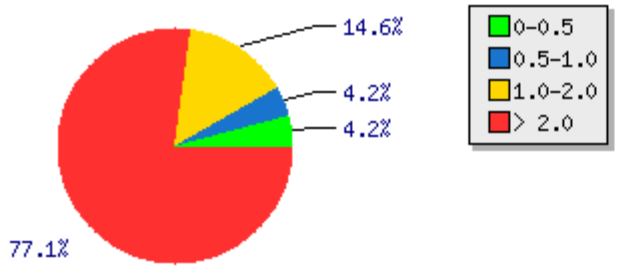
Results IV : DAQ → T0 buffer → Tape data Challenge II

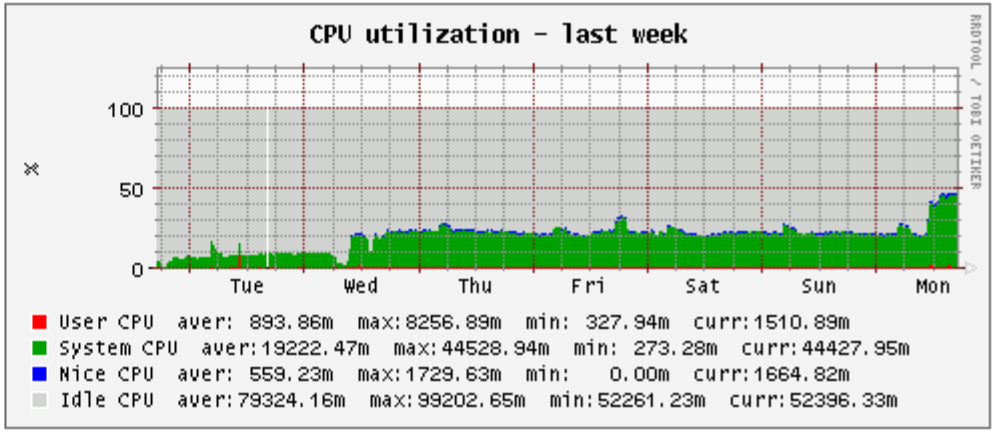


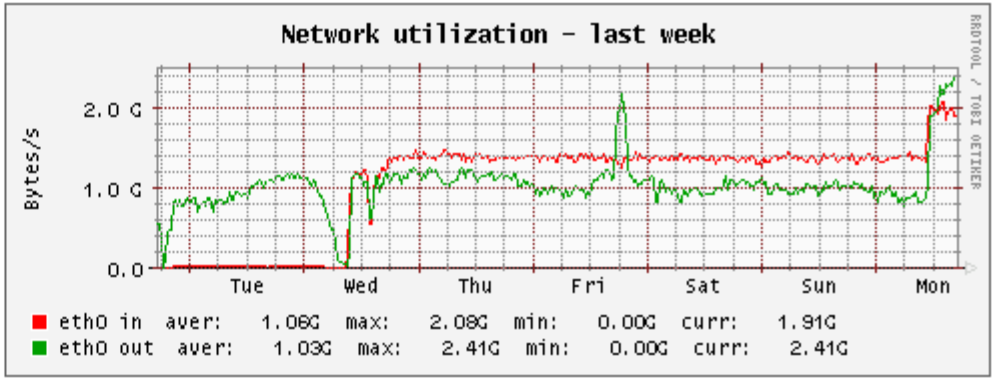
Cluster info: castor2 subcluster ITDC — 20 Feb 2006 Mon 17:00:40

Cluster Information	
# of hosts (down):	48 (0)
operating system(s):	2.4.21-37.EL.cernsmp, 2.4.21-37.0.1.EL.cernsmp
# of CPUs (down):	88 (0)
average up time:	25 days, 20h:44m (boots per host)
hosts down:	none
ITCM history	View template
Select from hosts:	None
Metric Distributions	Correlations

Load Percentages



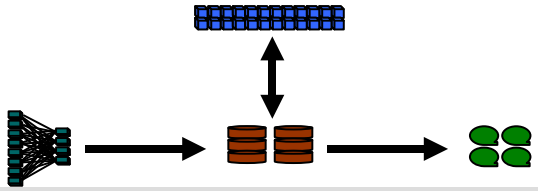




Auto Update

100 CPU nodes + 48 disk server + 28 IBM tape drives, reached 1.0 GByte/s for one week and peaks of 1.5 GB/s

Results V : ATLAS T0 test

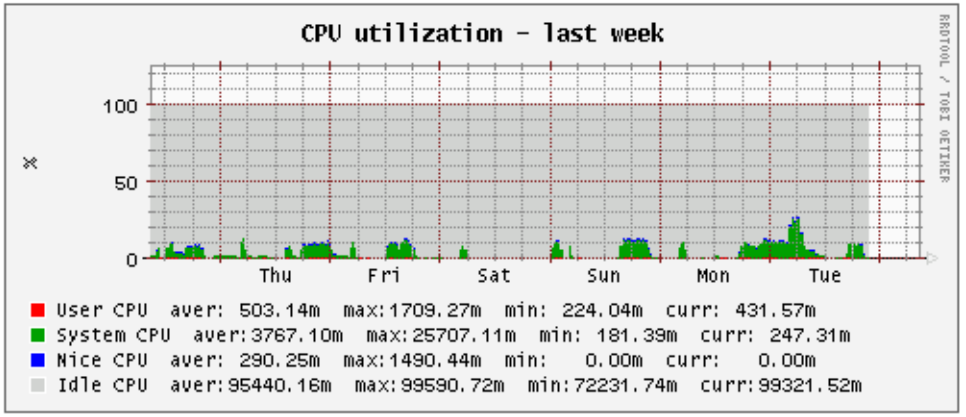


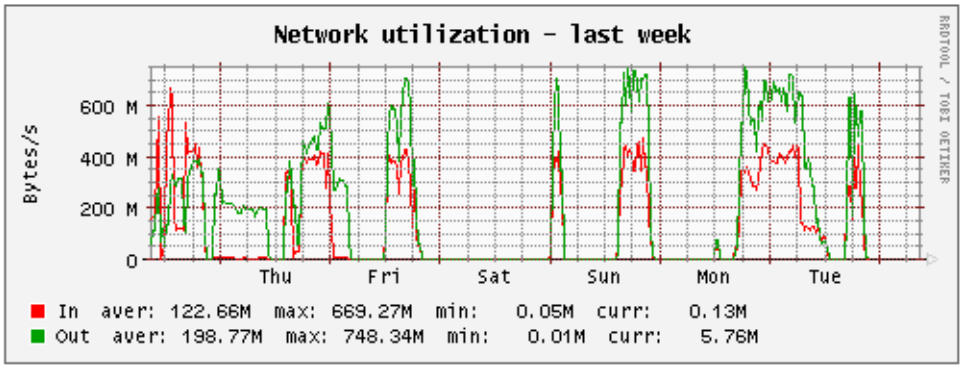
Cluster info: ITDC — 25 Jan 2006 Wed 08:39:12

Unable to connect to the Oracle server. Cluster information will be skipped.

Cluster Information	
# of hosts (down):	48 (48)
operating system(s):	
# of CPUs (down):	0 (0)
average up time:	0h:00m (boots per host)
hosts down:	lxfsra1201, lxfsra1202
ITCM history	View template
Select from hosts:	None
Metric Distributions	Correlations

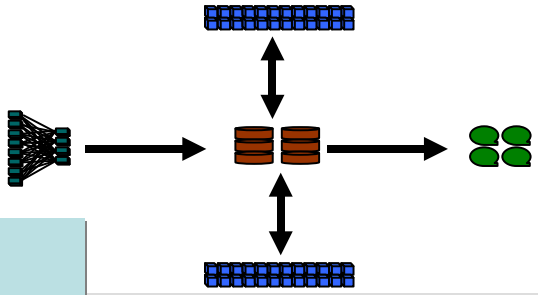
Load Percentages





ATLAS T0 tests with Castor2, 220 CPU nodes + 24 disk server + 12 tape drives up to 3000 batch jobs, reached more than nominal ATLAS speed (>320 MB/s*3) (without T1 export)

Results VI : Castor2, full T0 tests

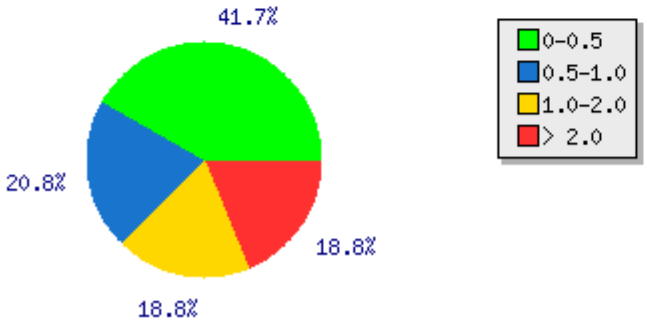


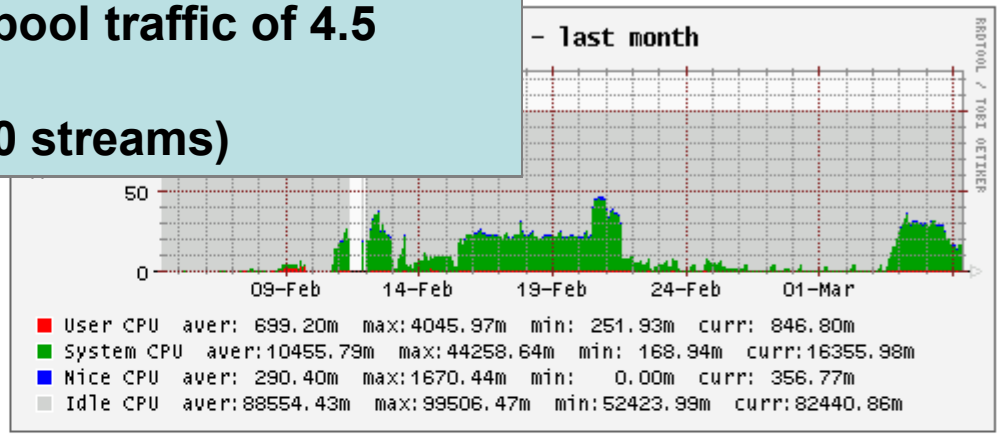
120 CPU nodes + 48 disk server + 28 IBM tape drives
2.1 GByte/s input data rate + 2.2 GByte/s output data rate
emulating DAQ + Recon + Export + Tape writing
(expect for pp running internal disk pool traffic of 4.5 Gbytes/s
→ reached now 4.3 GByte/s with 250 streams)

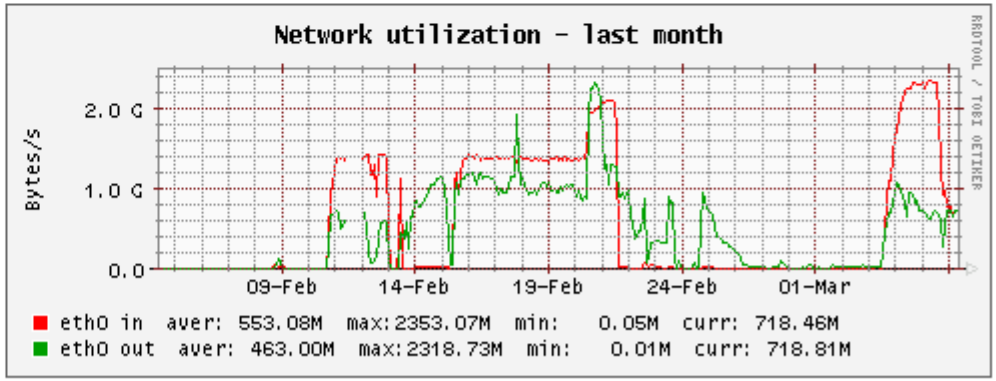
06 Mar 2006 Mon 08:15:45

# of CPUs (down):	88 (0)
average up time:	38 days, 7h:43m (boots per host)
hosts down:	none
ITCM history	View template
Select from hosts:	None
Metric Distributions	Correlations

Load Percentages







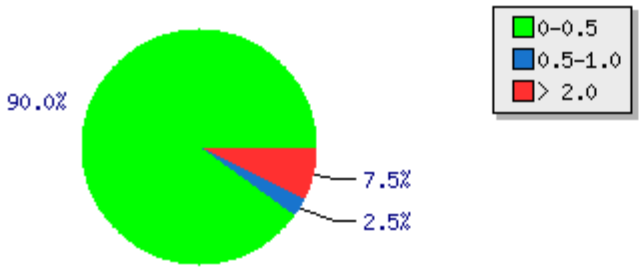
Last month | Get Fresh data | Get Detailed data | Auto Update

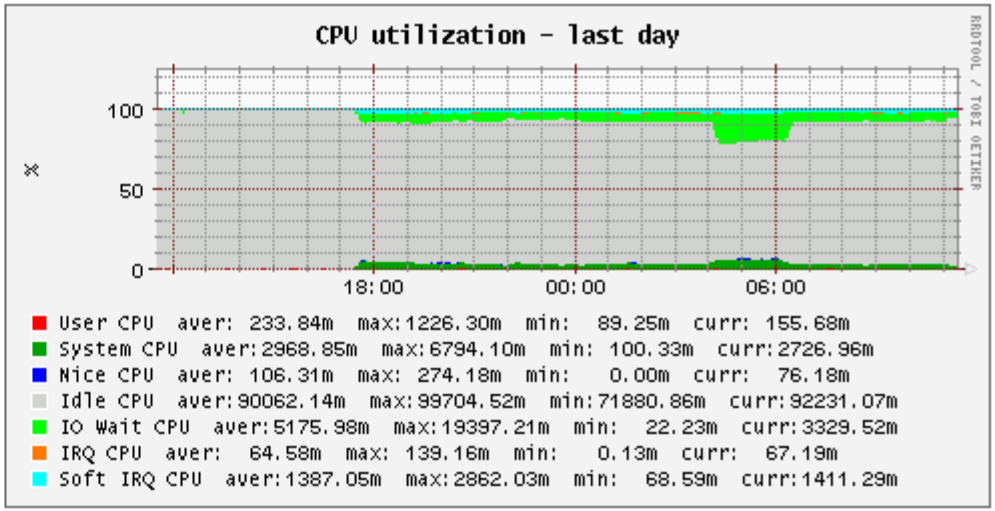
Instabilities , test run

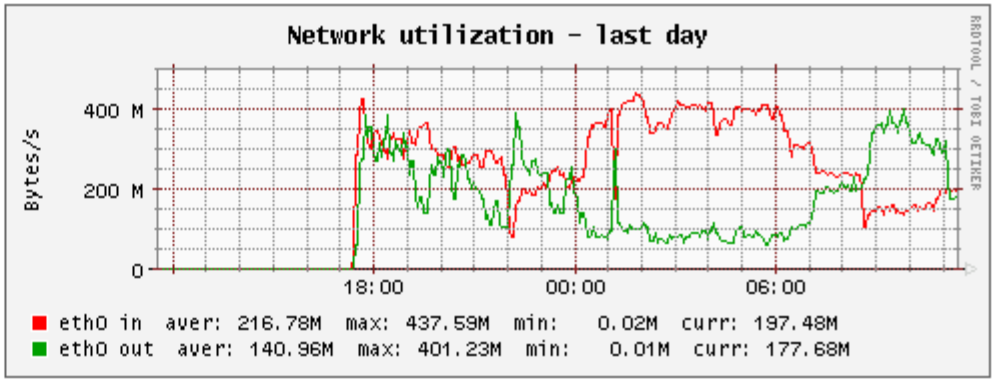
Cluster info: castor2 subcluster ITDC — 27 May 2006 Sat 11:30:27

Cluster Information	
# of hosts (down):	40 (0)
operating system(s):	2.4.21-40.EL.cernsmp
# of CPUs (down):	65 (0)
average up time:	10 days, 13h:47m (boots per host)
hosts down:	none
exceptions:	3WARE_CACHE_OFF, FILESYSTEM_ERROR,
ITCM history	View template
Select from hosts:	None
Metric Distributions	Correlations

Load Percentages







Last day | Get Fresh data | Get Detailed data | Auto Update

Stability and Problems I

- single tape drives and disk servers were regularly added or taken out on-the-fly, either scheduled or crashes
  → Castor2 behaved well, no side-effects, performance changed correspondingly

- several time the whole stager system (including DB) went into a state where a complete cleanup was necessary
  → fast fix, but limited debugging done

- interruptions due to software upgrades, network reconfiguration, cleanup after power-cuts
  → fast recovery and only minor problems

- instabilities of the IBM tape drive system caused by a combination of tries to enhance the tape drive efficiency and problems in the tape software
  → several weeks to resolve, difficult to debug, some problems with software version control

Stability and Problems II

- bugs and problems in the correct space allocation and limits in the disk pool, lead to strange IO patterns , long and slow migration queues, overload
  →few weeks to fix and deploy

- load balancing mechanism (disk server, streams) not really understood, does not work correctly, small changes in the setup can cause instabilities
  → ongoing since >6 weeks, hampers further tests

- the T0 operations have to be done under one stager, as the Castor2 architecture does not foresee a disk-to-disk copy between different stager instances. Thus the stager limits (e.g. number of operations per second) are important and the possibility to assign priorities to operations (e.g. DAQ versus export, calibration versus reconstruction, production user versus detector calibration user)
  → not yet fully tested and detailed scheduling not implemented yet

Time schedule, major data challenges

- **June-July** — ATLAS T0 tests at nominal speed for 3 weeks (320 MB/s DAQ-T0 + 440 MB/s reconstr. + 540 MB/s tape + 780 MB/s export)
- **July** — ALICE DC 7, Castor2 DAQ-T0-tape storage at 1 GB/s
- **Aug** — T0 DRC at 1.0 GB/s (internal 2 GB/s, many streams) (postponed from May)
- **Sept** — Castor2 tape storage DRC at 1.6 GB/s
- **Sept** — ATLAS T0 tests at nominal speed, increasing complexity, real DAQ system

Focus during 2006 **Full DAQ –T0 – T1 data flow and work flow tests**

Summary

- very good achievements for Castor2 during the last 6 month
- good performance and stability of the underlying hardware
  (disk and tape server)
- to reach the performance anticipated for 2008 is not a problem
  but
  there are problems to do this in an efficient and sustainable manner
  → load-balancing and prioritized scheduling are not working yet (correctly)
  this needs more priority
- bug fixes and problem identification take some time
  → noticed some overload in the teams
  → not always clear where the priorities are