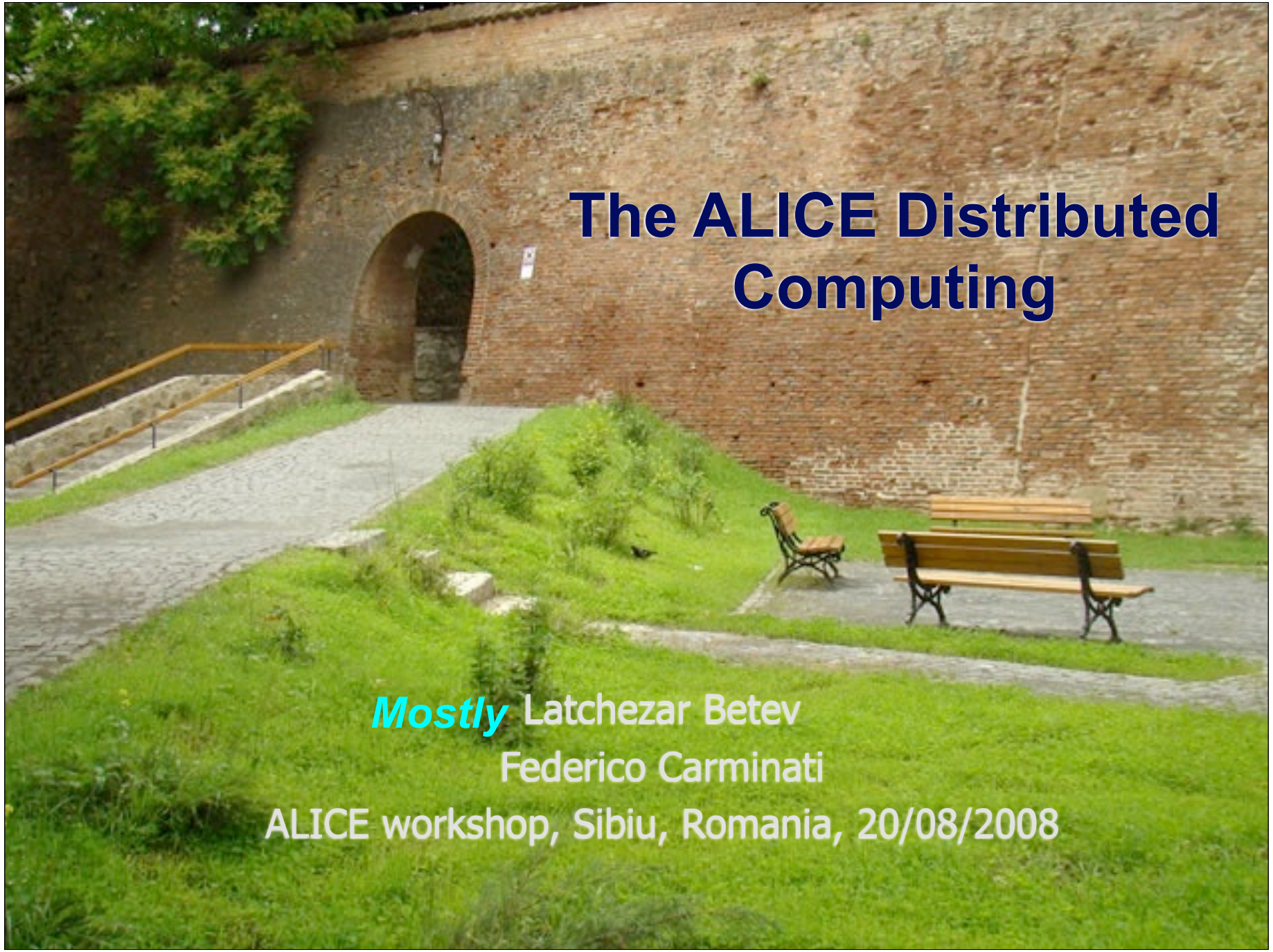
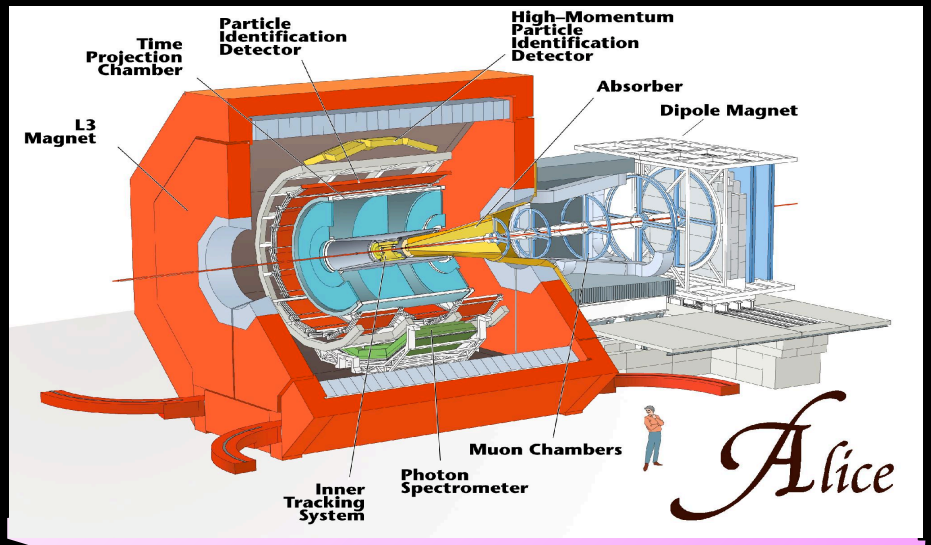

The ALICE Distributed Computing
Mostly Latchezar Betev
Federico Carminati
ALICE workshop, Sibiu, Romania, 20/08/2008

Perspectives…

- This workshop: exactly 30 days before the official start of LHC (2.6Mio seconds)

10 September 2008

- Is the ALICE Offline ready to register, process and provide the ALICE physicists with a platform to efficiently analyse the first data?

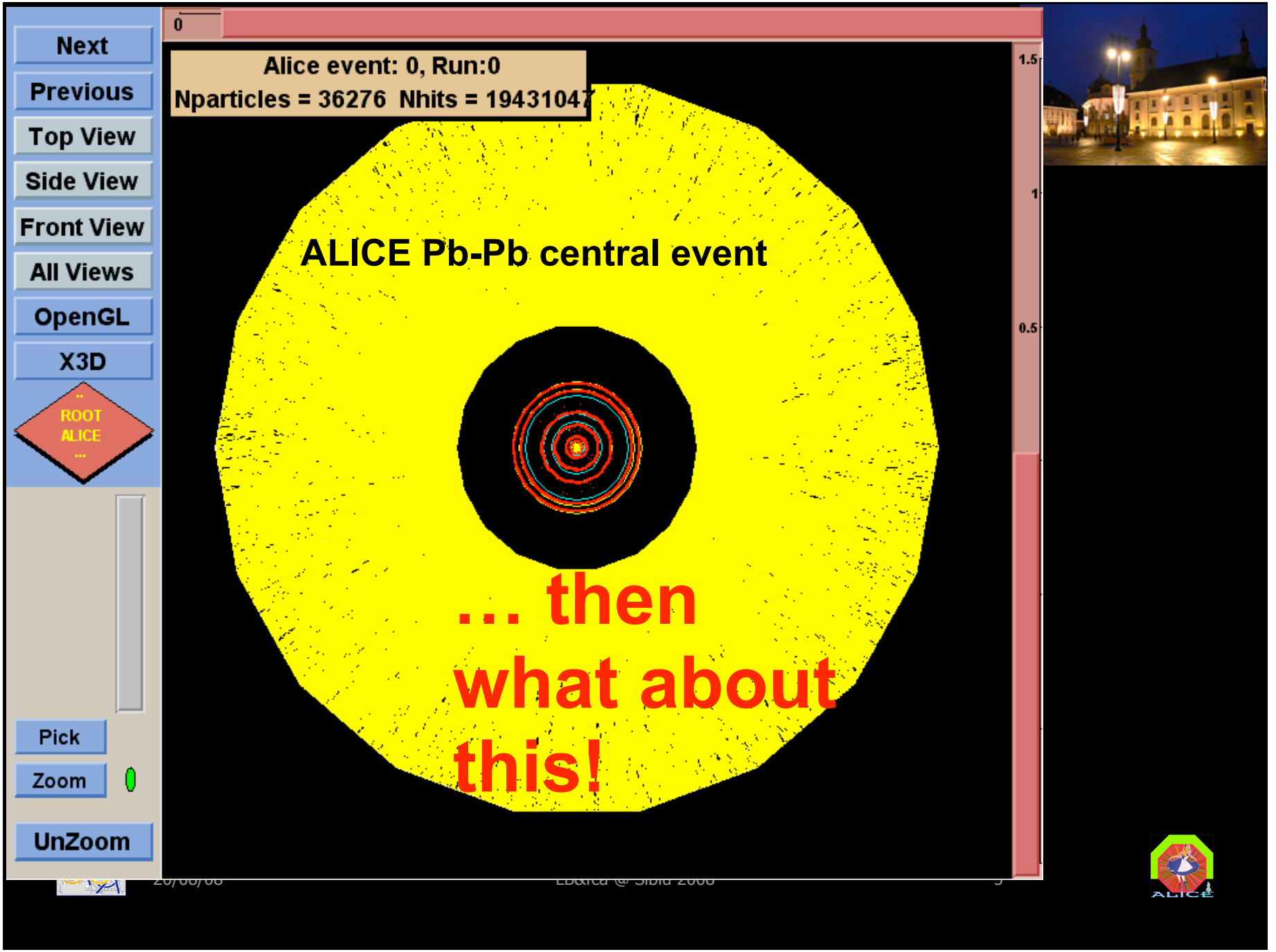


ALICE Collaboration
~ 1/2 ATLAS, CMS, ~ 2x LHCb
~1000 people, 30 countries, ~ 80 Institutes

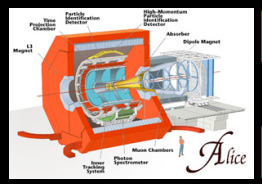
Total weight	10,000t
Overall diameter	16.00m
Overall length	25m
Magnetic Field	0.4Tesla

8 kHz (160 GB/sec)
level 0 - special hardware
200 Hz (4 GB/sec)
level 1 - embedded processors
30 Hz (2.5 GB/sec)
level 2 - PCs
30 Hz
(1.25 GB/sec)
data recording & offline analysis

If we thought this was difficult ...


Next
Previous
Top View
Side View
Front View
All Views
OpenGL
X3D
ROOT
ALICE
Pick
Zoom
UnZoom
0
Alice event: 0, Run:0
Nparticles = 36276 Nhits = 19431047
1.5
1
0.5
ALICE Pb-Pb central event
… then
what about
this!

Computing model – pp



Tape T0

Alien FC

To tape

To Grid FC

RAW

Calibration

Generation of calibration parameters

Disk buffer T0

First pass Reco

CERN T0

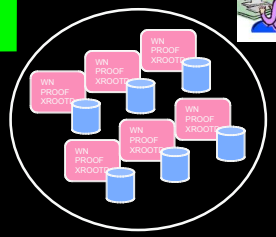
to T1s

CAF analysis

T2s

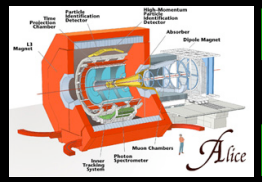
MC data

T1s



Computing model – AA

HI data taking



RAW

Calibration

Generation of calibration parameters

Disk buffer T0

From tape

Tape T0

To Grid FC

Alien FC

First pass Reco

CERN T0

LHC shutdown

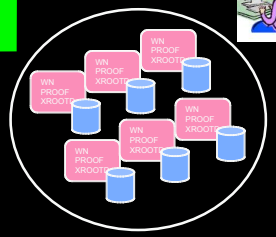
to T1s

CAF analysis

T2s

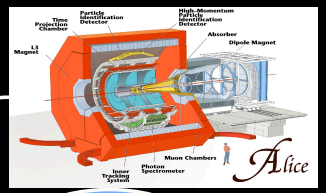
MC data

T1s



High-level dataflow

Online ⟵ ⟶ Offline



ALICE File Catalogue

lfn	guid	{se's}
lfn	guid	{se's}
lfn	guid	{se's}
lfn	guid	{se's}
lfn	guid	{se's}

DDL

LDC

Monit. Calib.

GDC

CAF

WN

Publish in AliEn

Data files

DAQ Network

Condition files

Publish agent

xrootd

Run info

HLT

DAQ Logbook DB

DAQ FXS

Data file

Castor cache

S R

FTS

S R

T1's

240TB

Condition files

HLT FXS

DCS

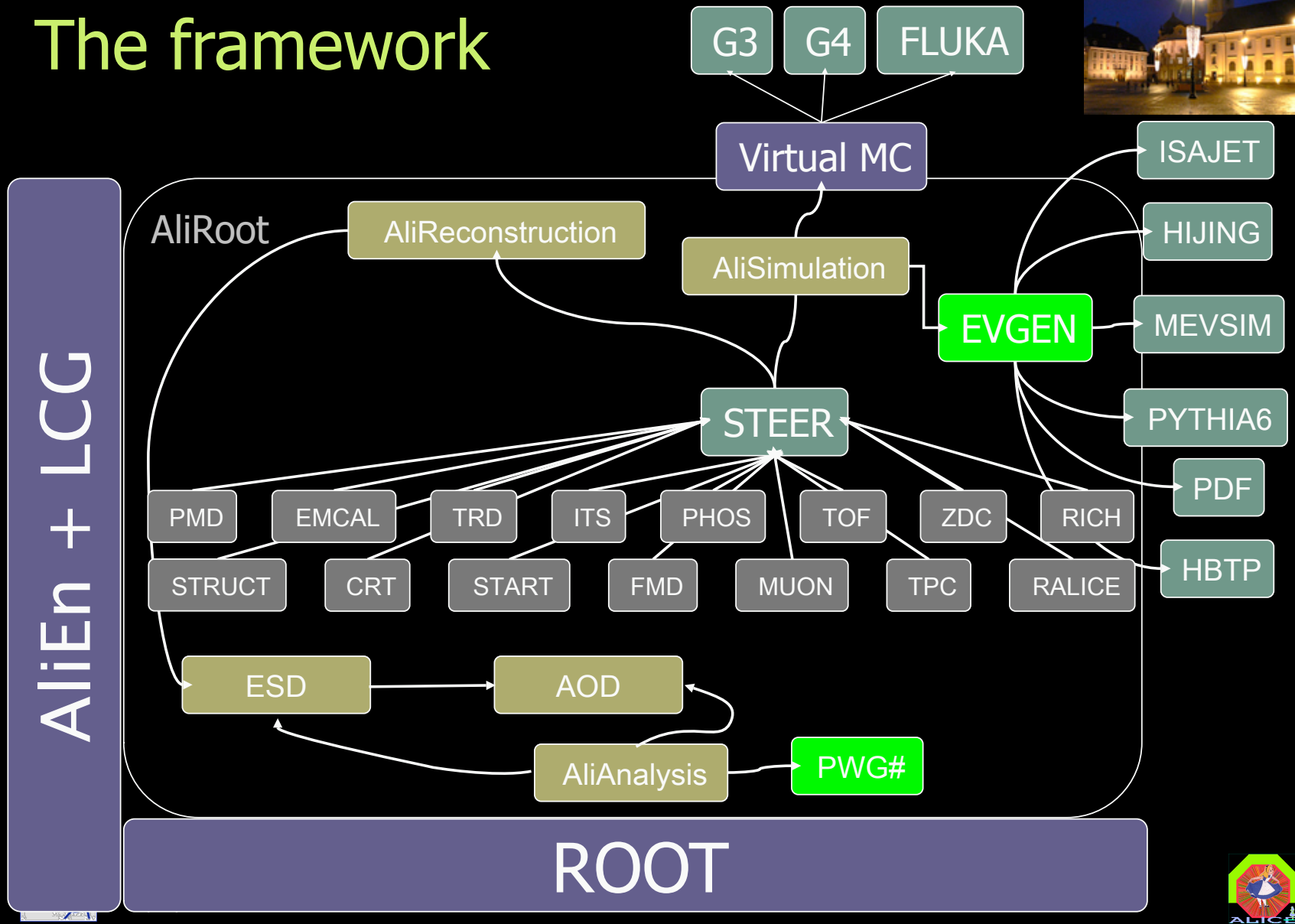
DCS FXS

DCS DB

Shuttle

CASTOR

The framework



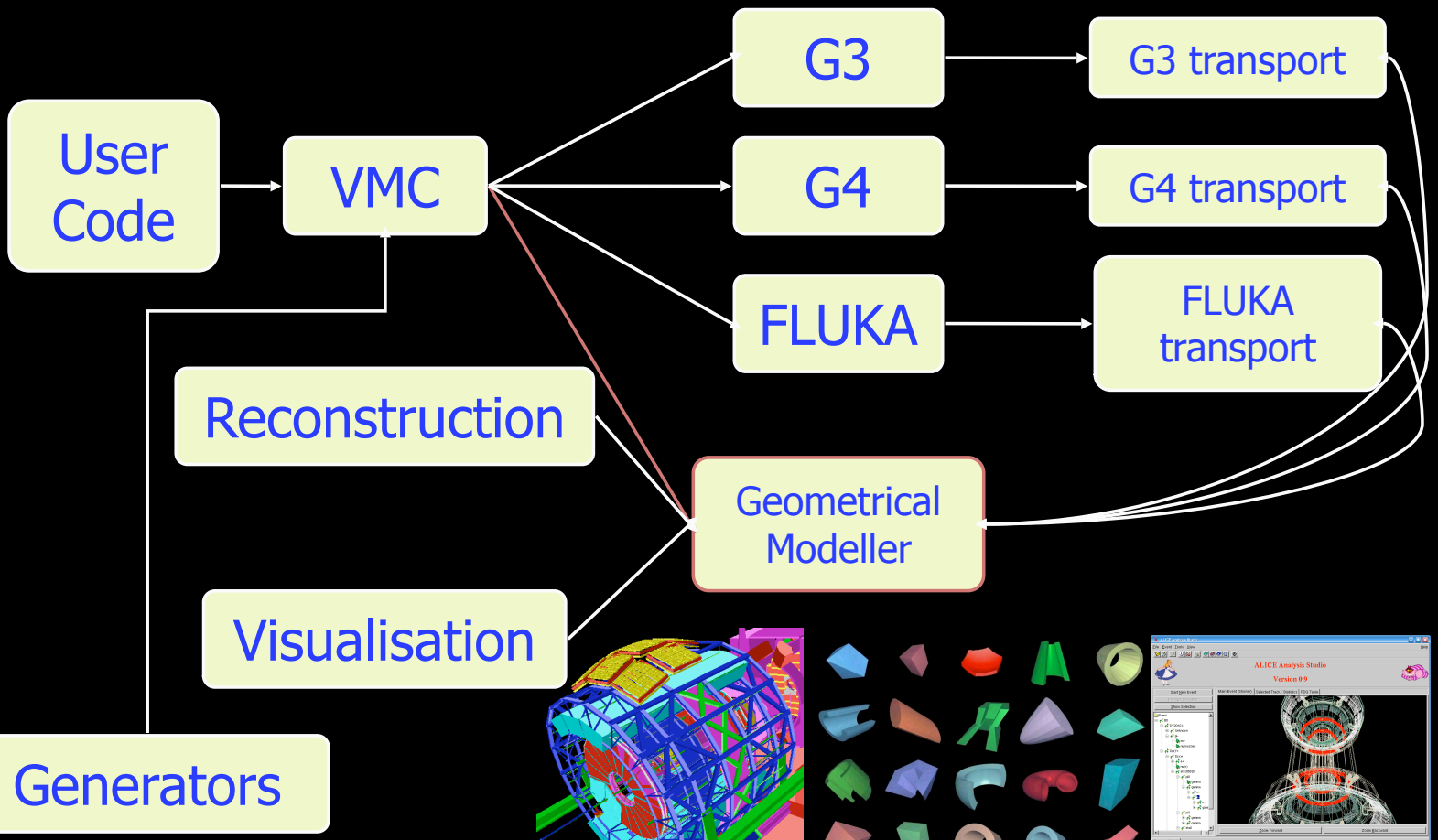
Software management

- Regular release schedule
  - Major release every two months, several minor releases (tag) / month
- Nightly produced UML diagrams, code listing, coding rule violations, build and tests, benchmarks, single repository with all the code
  - No version management software (we have very few packages!)
  - Very portable code (Linux 32/64, Mac OS X, Sun)
- Advanced code tools under development (with IRST/Trento)
  - Smell detection
  - Aspect oriented programming tools
  - Automated genetic testing
- Documentation for AliRoot and AliEn produced
- Intensive training ongoing

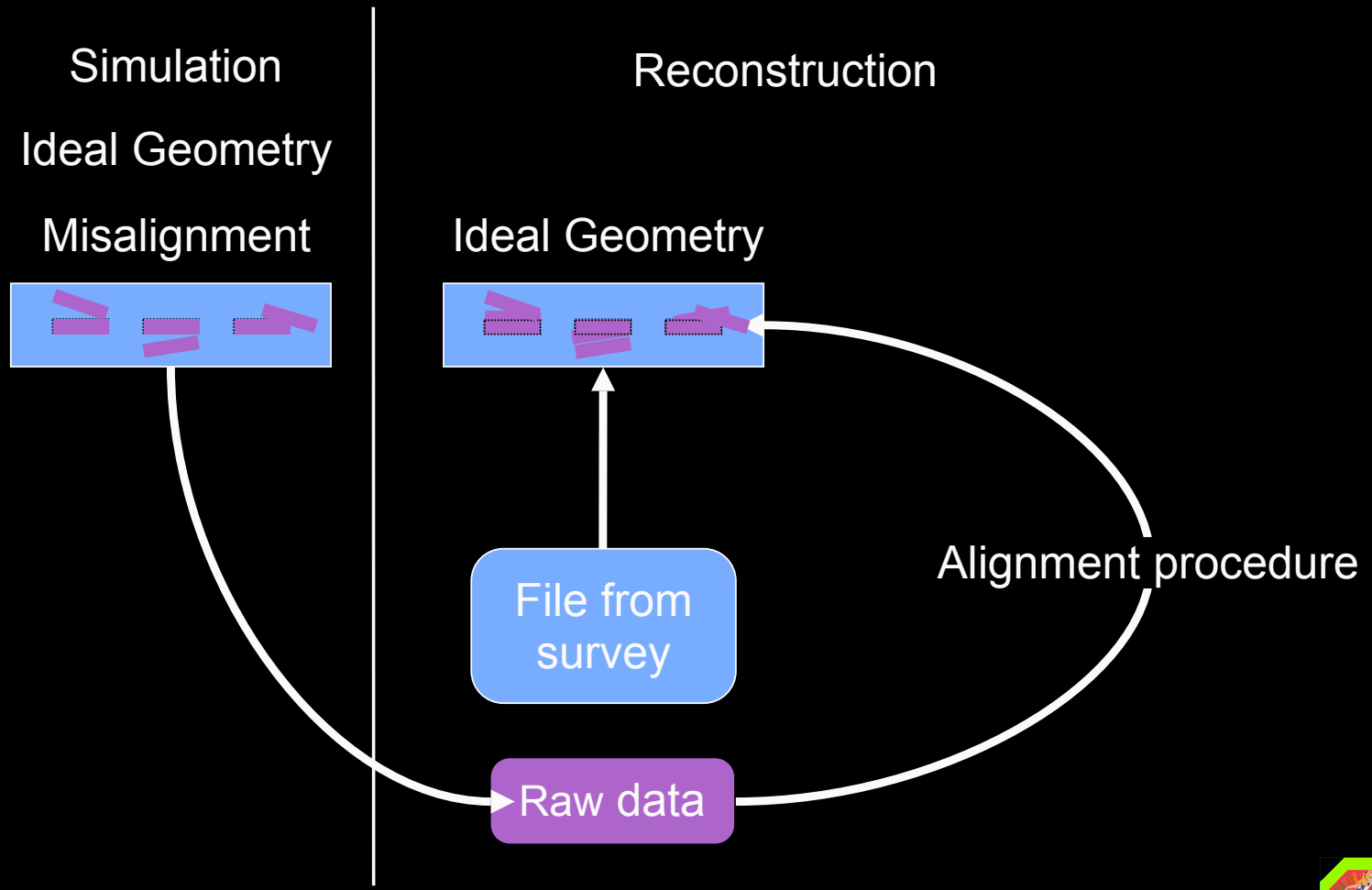
Software management

- Strict weekly release schedule for AliRoot
  - Coordinated with detector groups
  - Essential for integration and validation of detector algorithms in DAQ/HLT systems (shared code)
  - Some effort to explain to the community that "a release is a release"
- Simultaneous deployment on the Grid for RAW production and ESD analysis
- Titanic effort to remain below 2GB is paying off

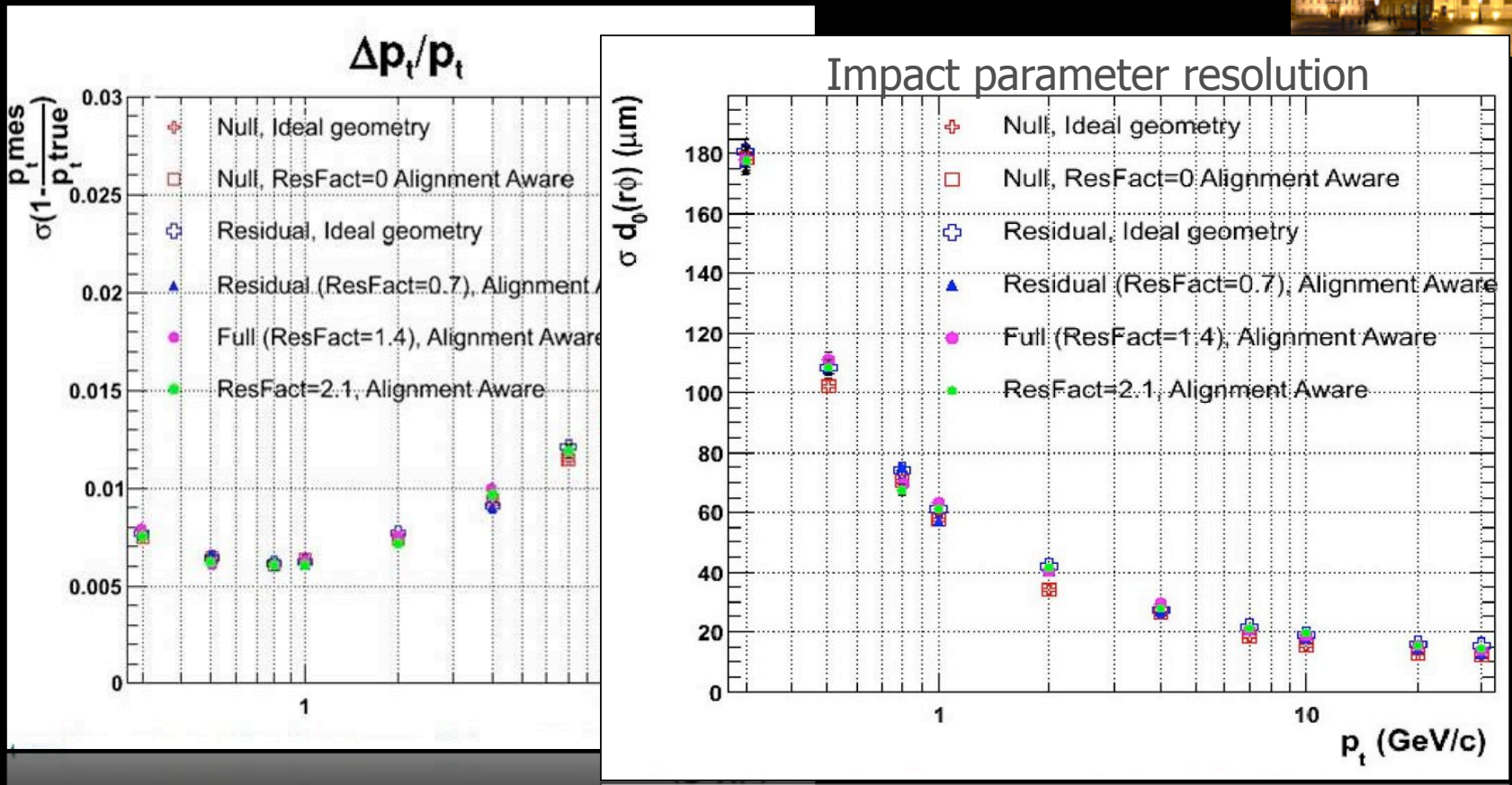
The Simulation



Alignment


Simulation
Ideal Geometry
Misalignment
Reconstruction
Ideal Geometry
File from survey
Alignment procedure
Raw data

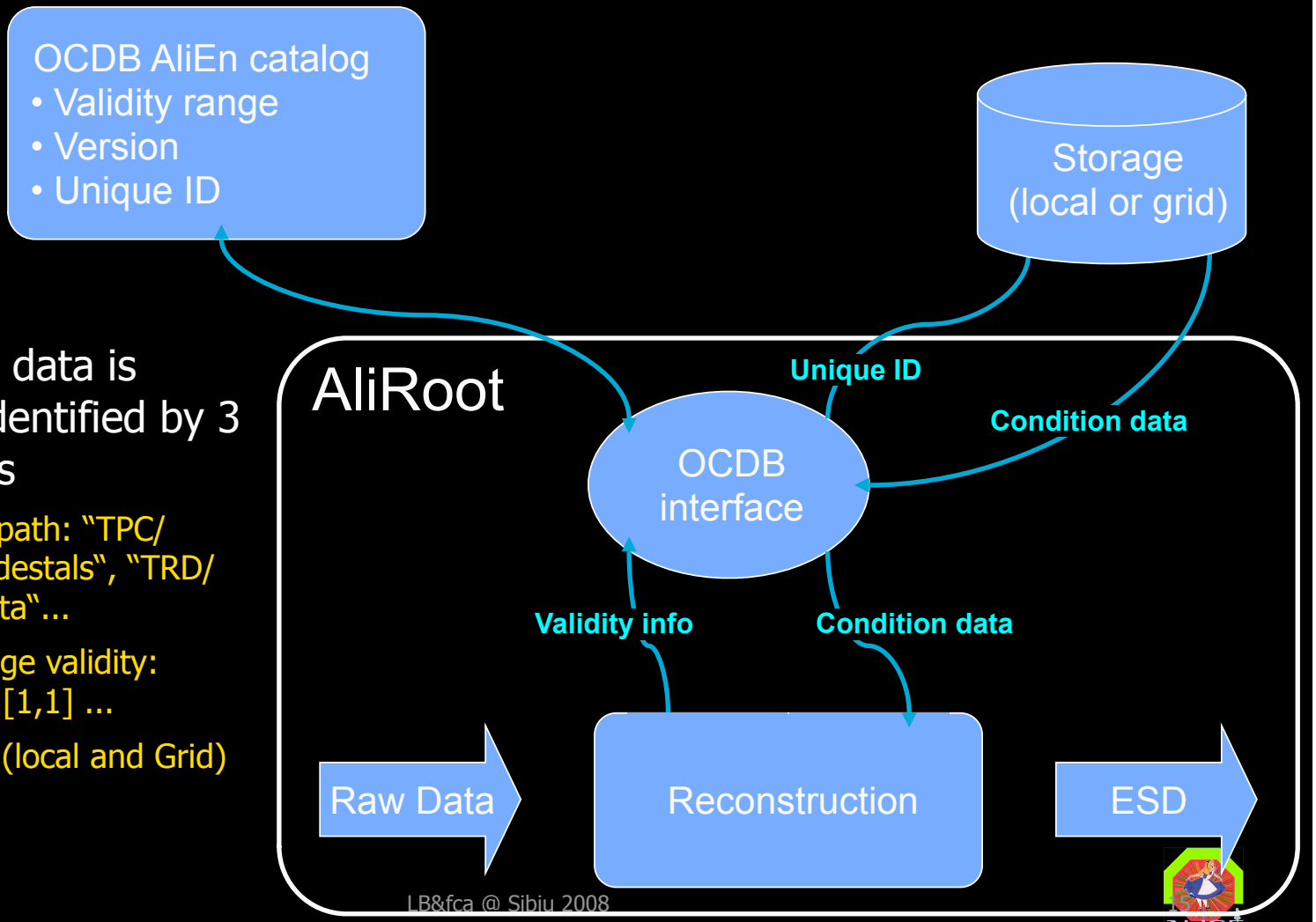
Alignment-aware ITS reconstruction



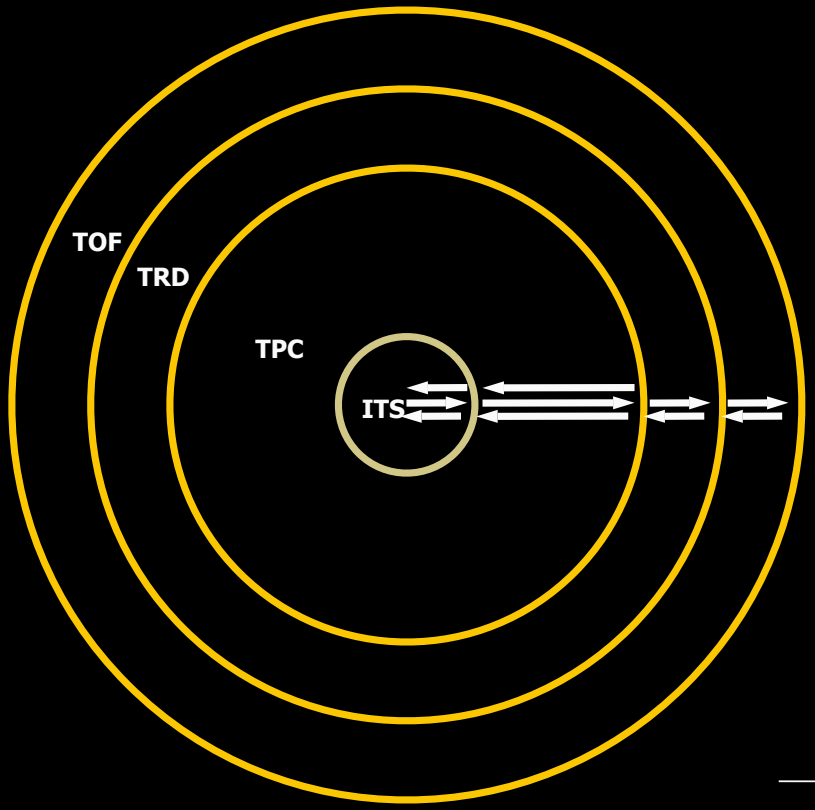
The momentum and impact parameter resolutions are defined by the "quality" of the provided alignment objects (ideal AliAlignObjs → ideal resolutions).

Calibration in the offline world

- Conditions data is uniquely identified by 3 parameters
  - Logical path: "TPC/Calib/Pedestals", "TRD/Align/Data"...
  - Run range validity: [0,100], [1,1] ...
  - Version (local and Grid)

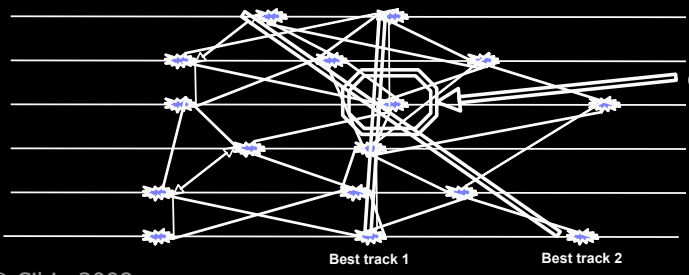


The reconstruction

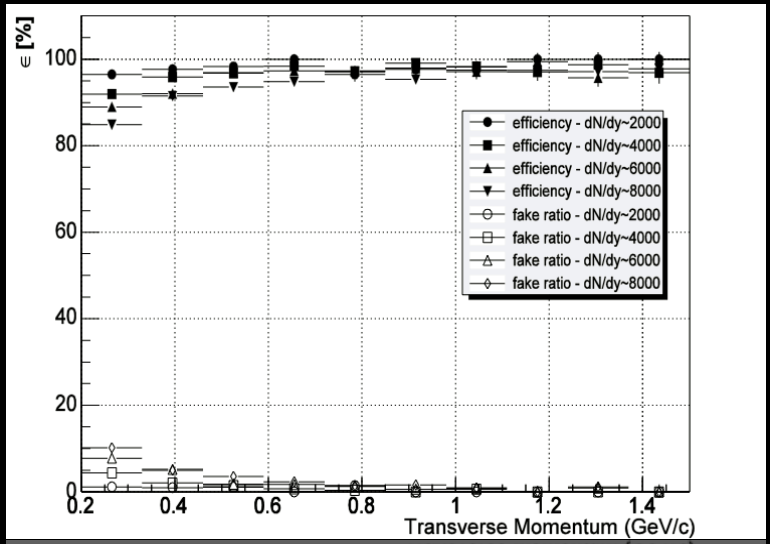


- Incremental process
  - Forward propagation towards to the vertex TPC⇒ITS
  - Back propagation ITS⇒TPC⇒TRD⇒TOF
  - Refit inward TOF⇒TRD⇒TPC⇒ITS
- Continuous seeding
  - Track segment finding in all detectors

- Combinatorial tracking in ITS
  - Weighted two-tracks χ^2 calculated
  - Effective probability of cluster sharing
  - Probability not to cross given layer for secondary particles

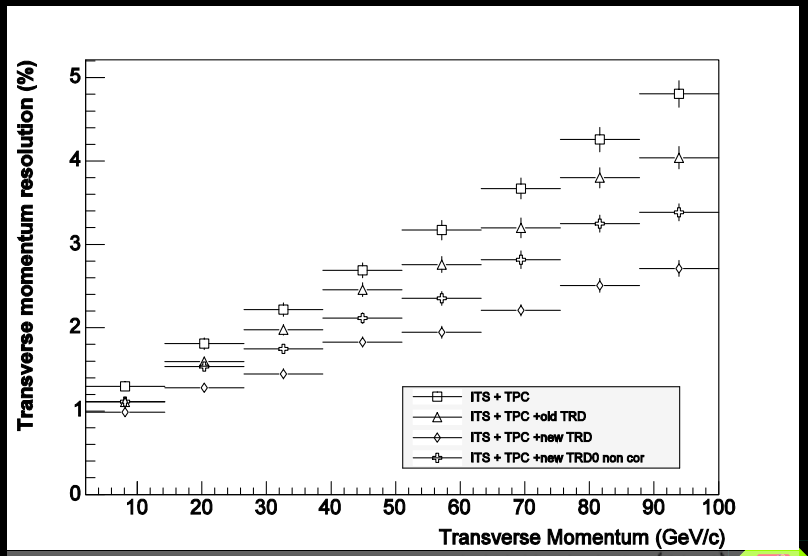


Combined ITS+TPC+TRD tracking

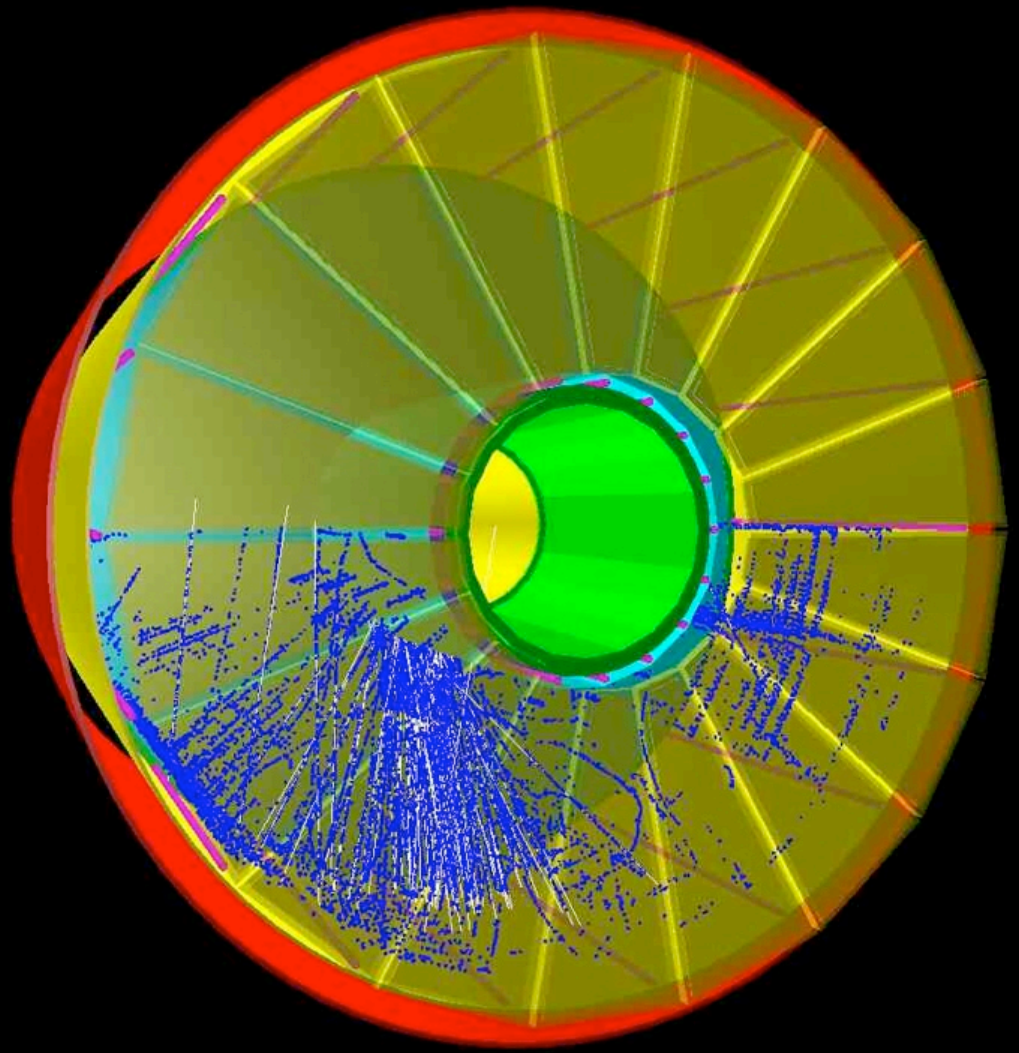


Efficiency:
~85% at 0.2 GeV/c
~95% above 2.0 GeV/c

Momentum resolution:
~1% at 1 GeV/c,
~4% at 100 GeV/c

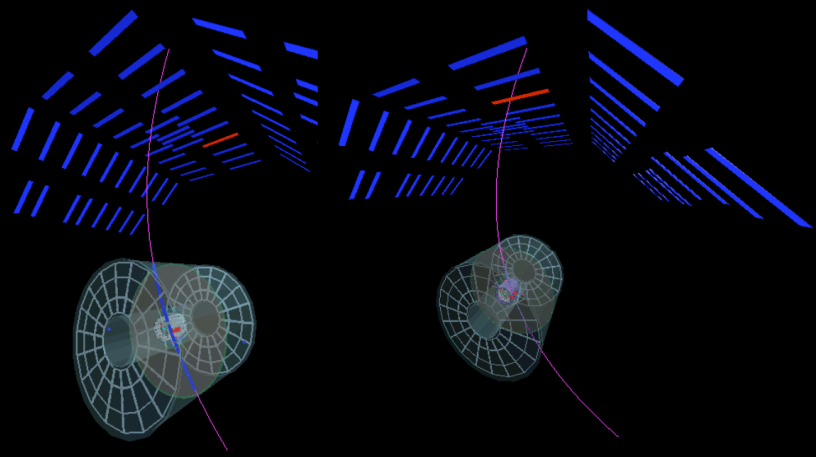


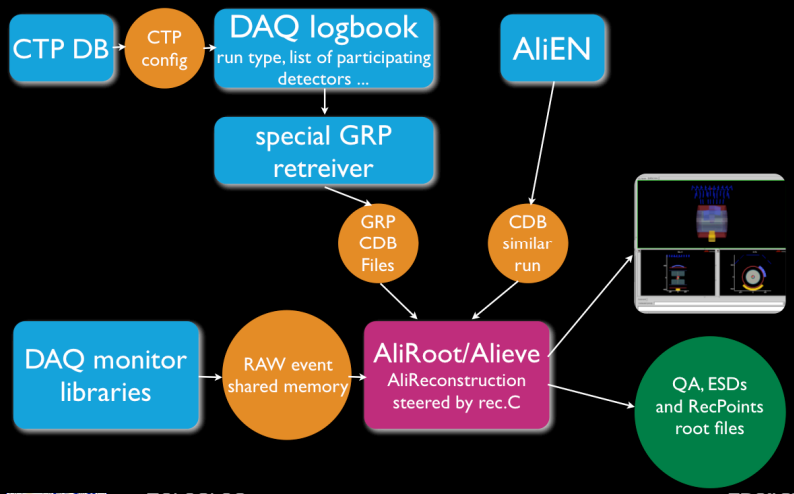
Visualisation

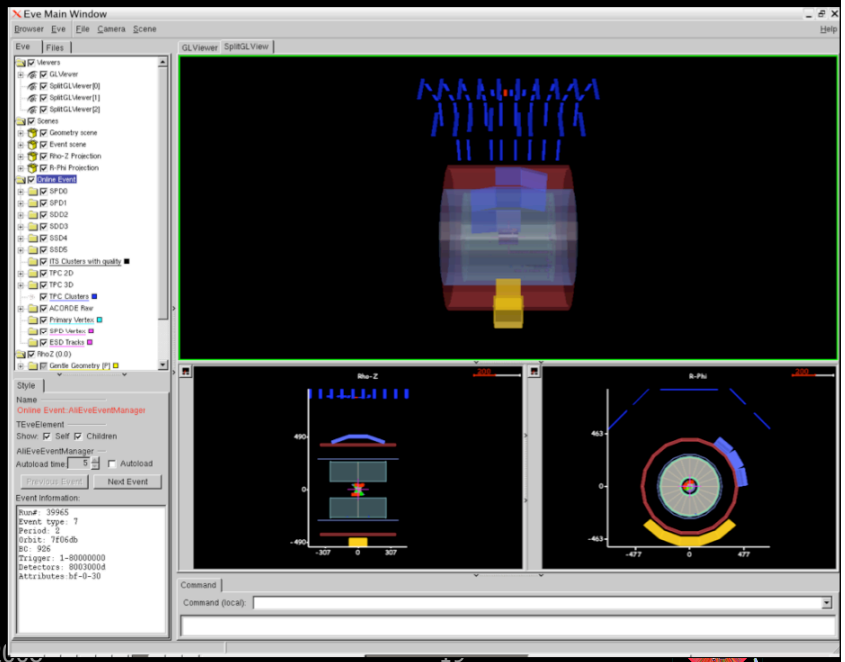


Quasi-online reco

- Recent development
- Very useful for high-level QA and debugging
- Integrated in the AliEVE event display
- Full Offline code sampling events directly from DAQ memory



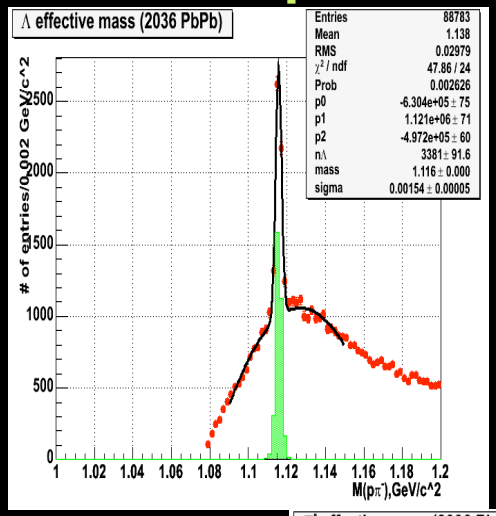


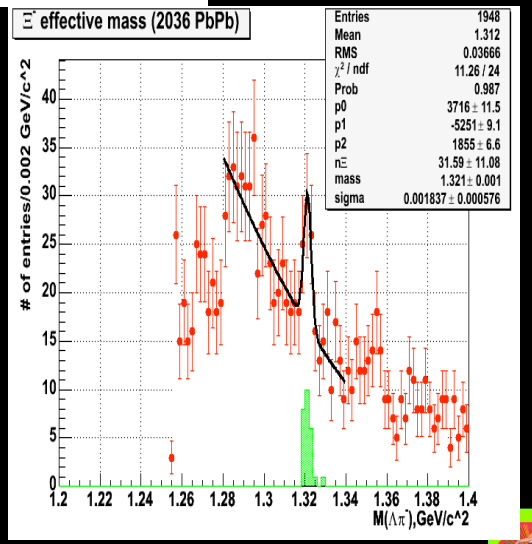


ALICE Analysis Basic Concepts

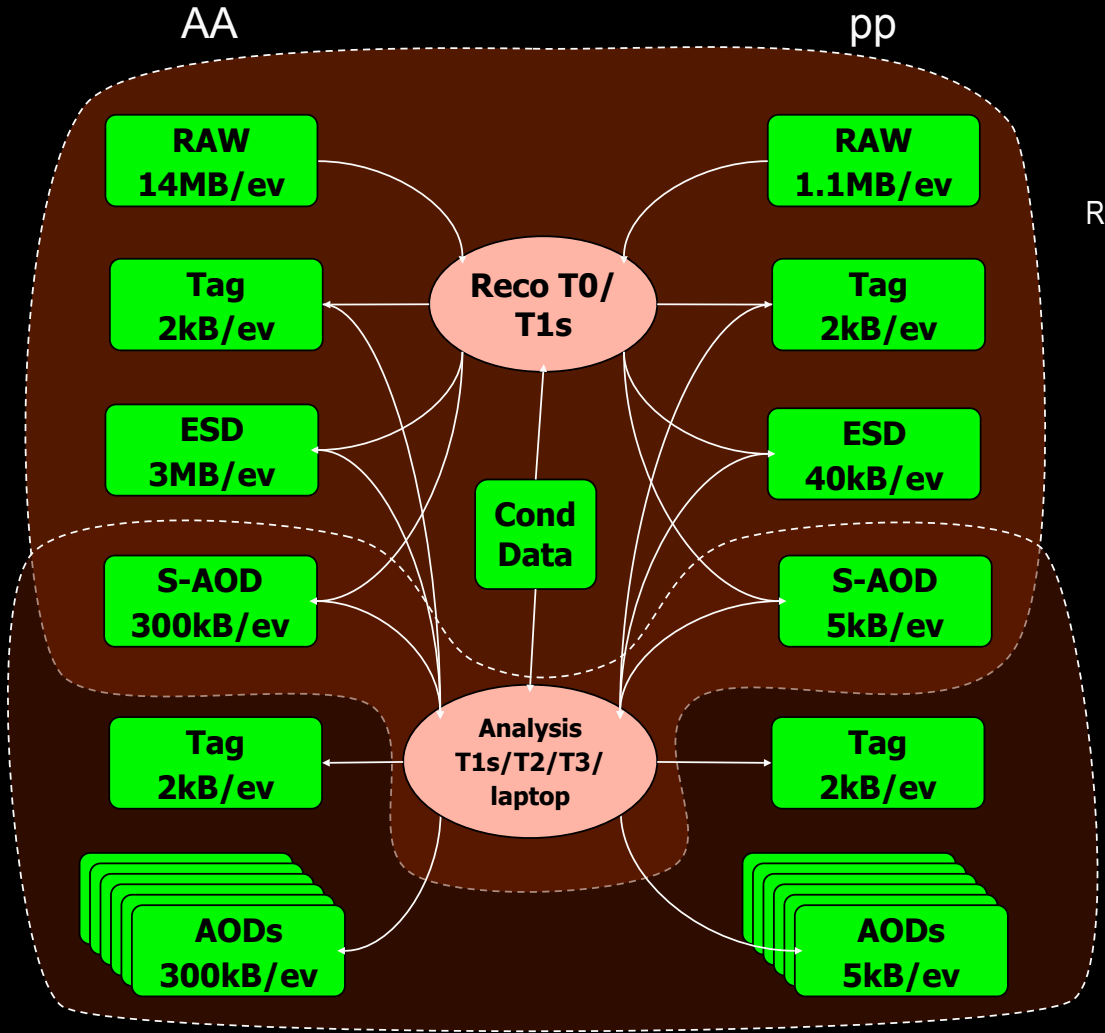
$\Xi \rightarrow \pi\Lambda \rightarrow p\pi$

- Analysis Models
  - Prompt data processing (calib, align, reco, analysis) @CERN with PROOF
  - Batch Analysis using GRID infrastructure
  - Local analysis
  - Interactive analysis PROOF+GRID
- User Interface
  - Access GRID via AliEn or ROOT UIs
- PROOF/ROOT
  - Enabling technology for CAF
  - GRID API class TAliEn
- Analysis Object Data contain only data needed for a particular analysis
- Analysis à la PAW
  - ROOT + at most a small library
- Work on the distributed infrastructure has been done by the ARDA project





Data reduction in ALICE



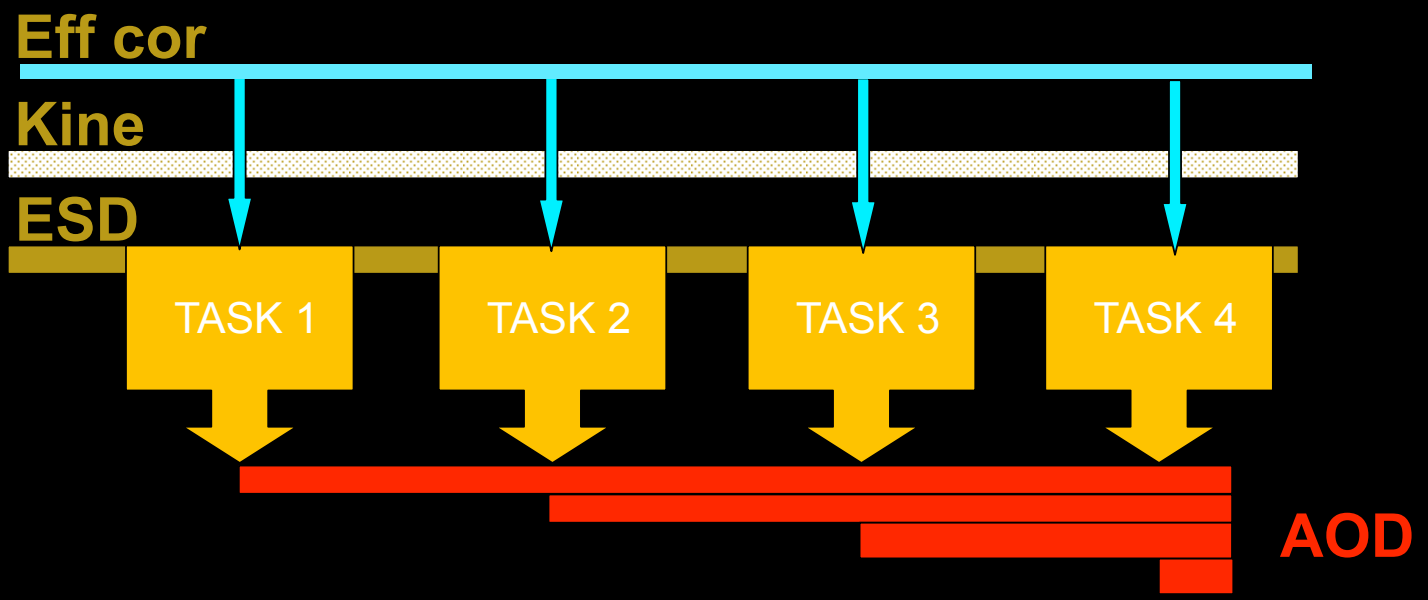
Requires AliRoot+Cond
+AliEn (once)

Has to run on a
disconnected laptop

Organised analysis

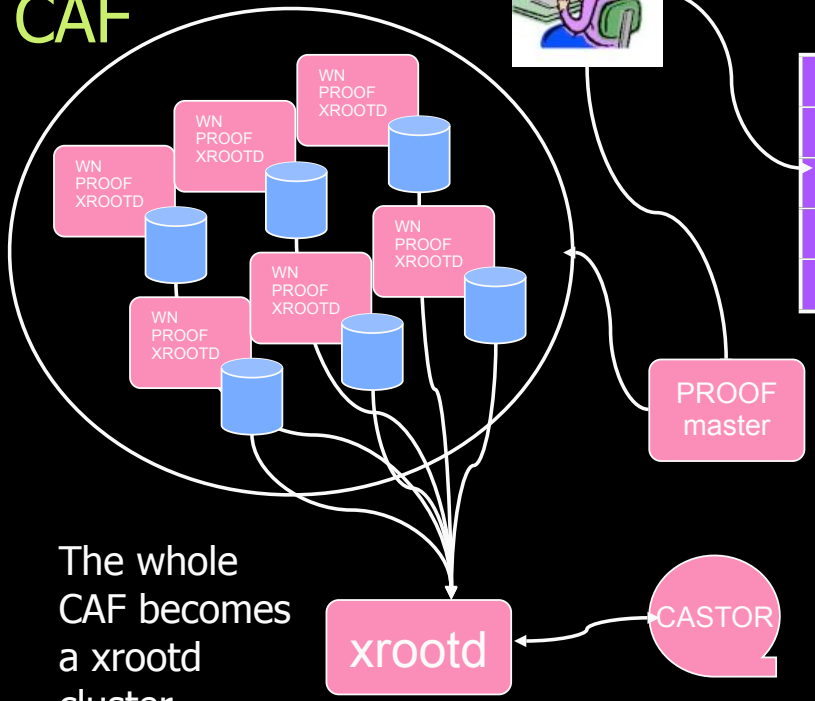
- Organised analysis is the most efficient way for many tasks to read and process the full data set
  - Optimise CPU/IO ratio for distributed resources
  - Common well tested framework
  - Common knowledge base and terminology
  - Document procedure
  - Makes results reproducible
- Will run "sanctified" algorithms and will assess global data quality

Analysis train

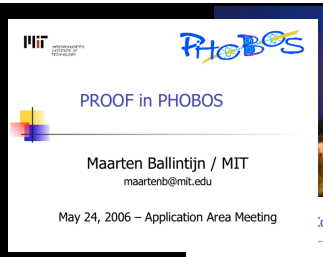


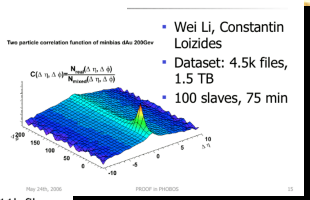
- AOD production will be organized in a 'train' of tasks
  - To maximize efficiency of full dataset processing
  - To optimize CPU/IO
  - Using the analysis framework

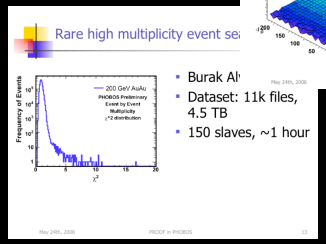
CAF

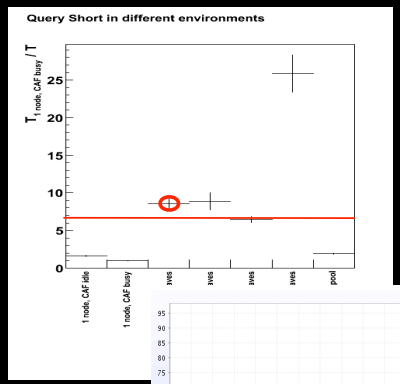


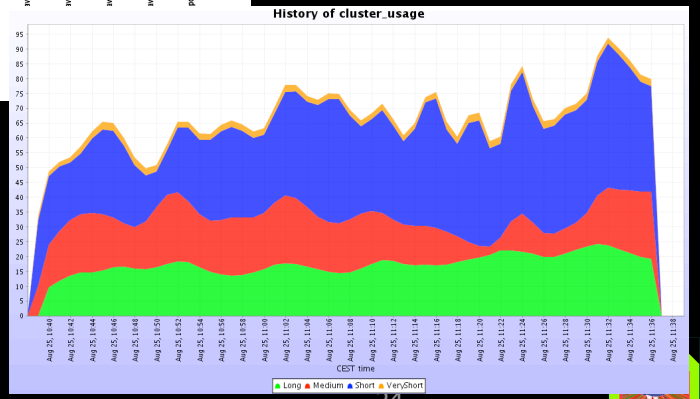
lfn	guid	{se's}
lfn	guid	{se's}
lfn	guid	{se's}
lfn	guid	{se's}
lfn	guid	{se's}









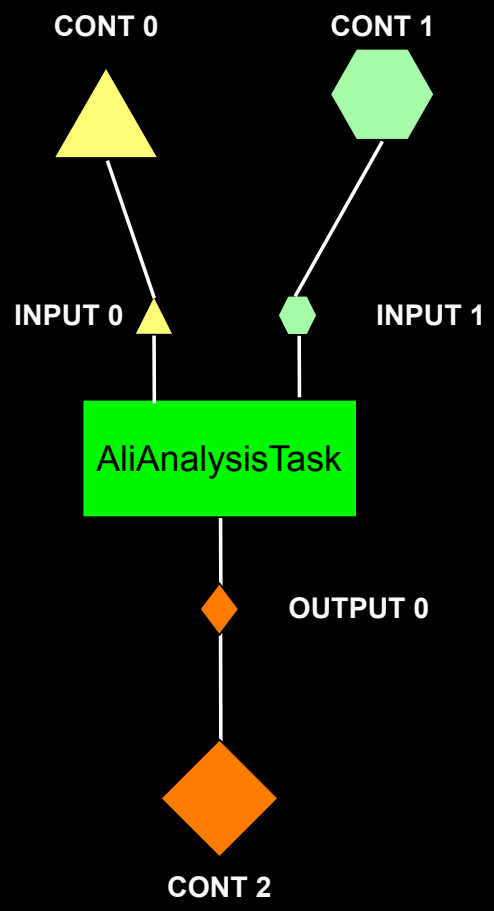


The whole CAF becomes a xrootd cluster

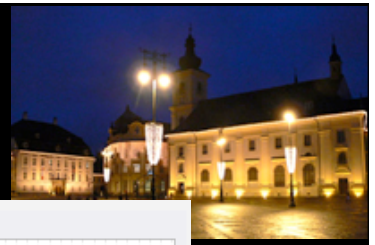
- Powerful and fast machinery – very popular with users
- Allows for any use pattern, however quite often leading to contention for resources
  - CPU and disk space quotas are a must

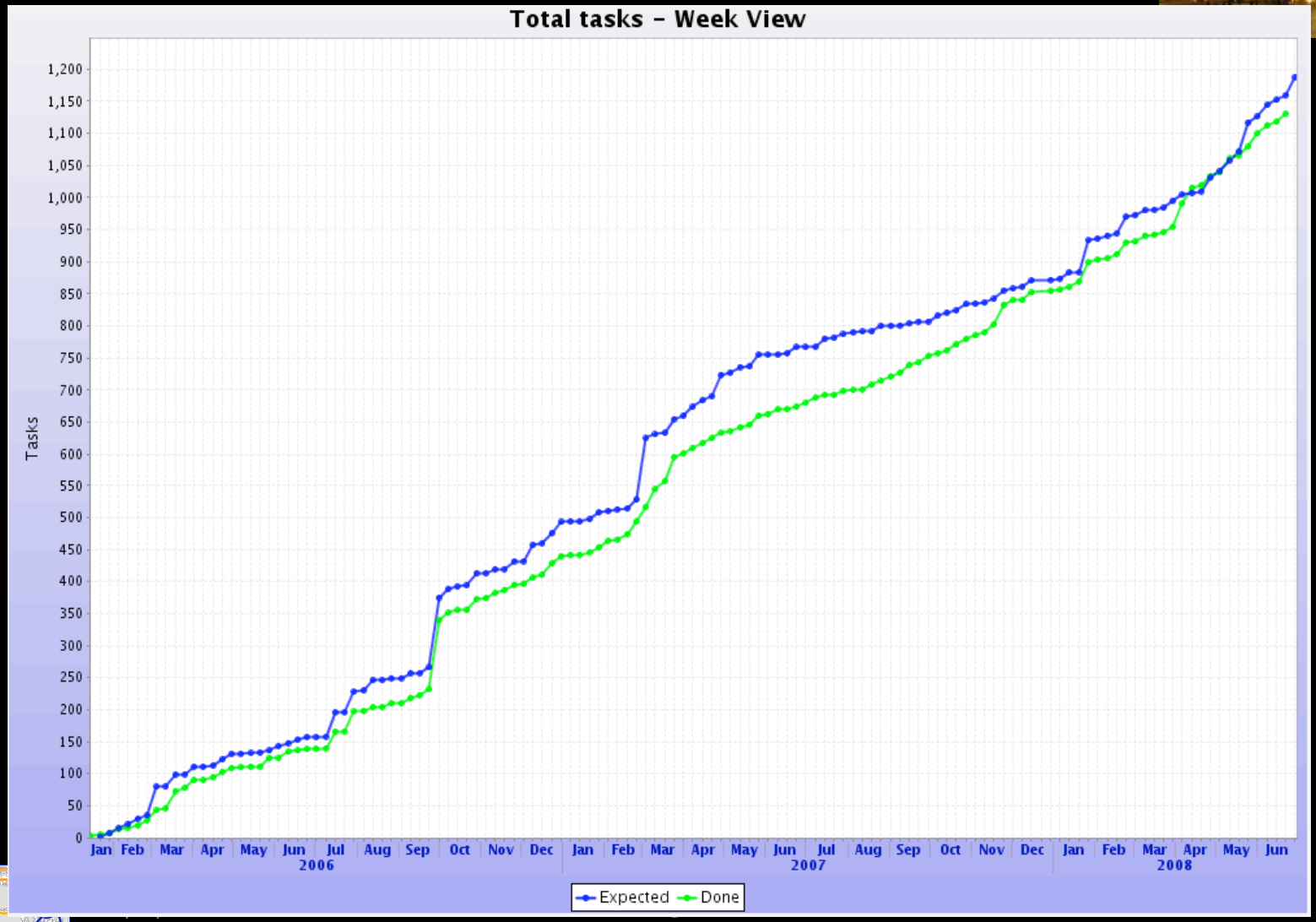
AliAnalysis Framework

- Data-oriented model composed of independent tasks
  - Task execution triggered by data readiness
- Parallel execution and event loop done via TSelector functionality
  - Mandatory for usage with PROOF
- Analysis execution performed on event-by-event basis
- Same code for interactive / parallel / local / grid analysis
- Adopted by the PWG's



Planning




Total tasks - Week View
Tasks
1,200
1,150
1,100
1,050
1,000
950
900
850
800
750
700
650
600
550
500
450
400
350
300
250
200
150
100
50
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun
2006
2007
2008
Expected
Done

Resource overview

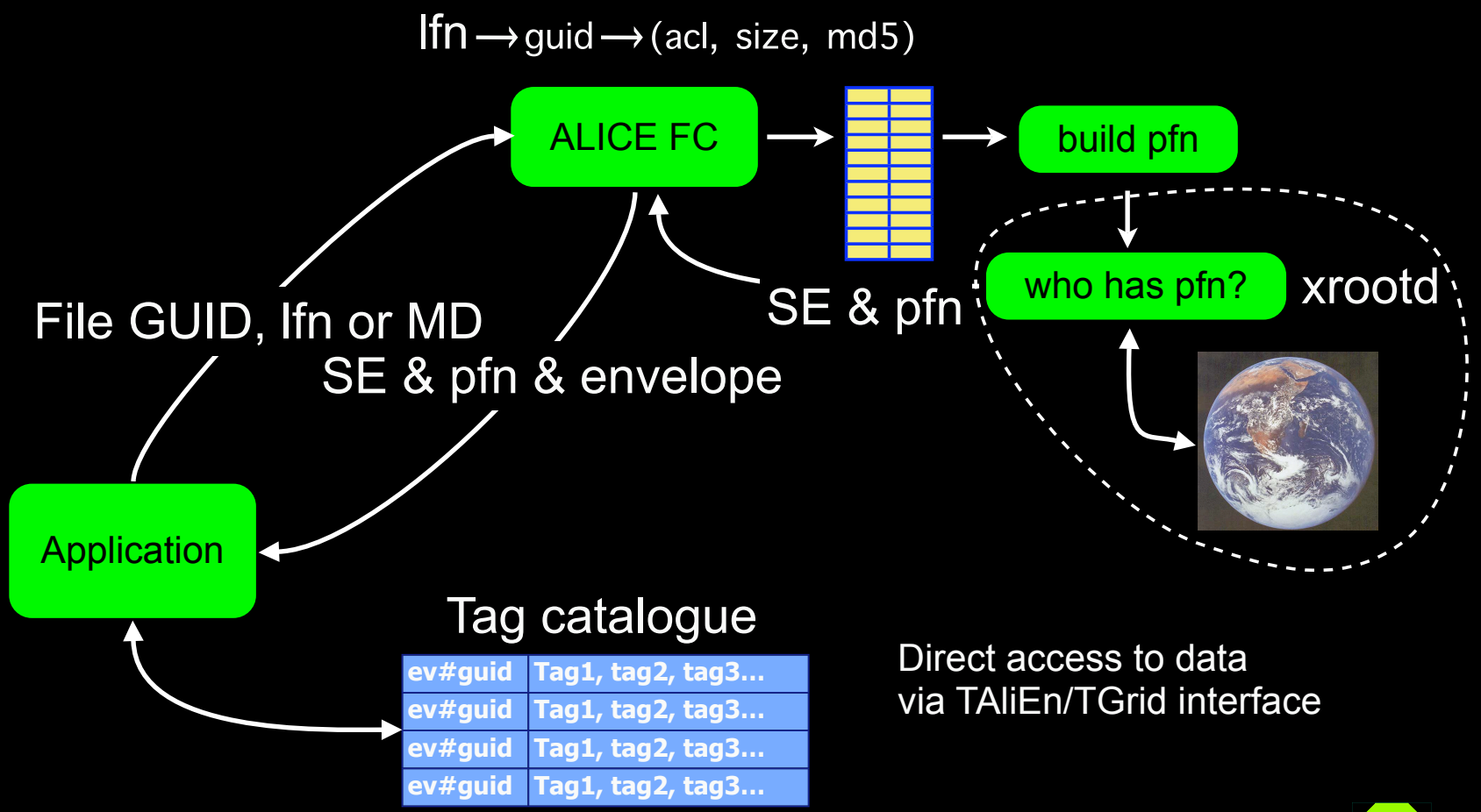
Parameter	Now	CTDR	Ratio
pp RAW	1.0MB	0.2MB	5*
Pb RAW	35MB	13.8MB	2.5
ESD pp	0.04MB	0.04MB	1.0
ESD Pb	6.3MB	3.0MB	2.1
AOD pp	5kB	16kB	0.3
AOD Pb	1.3MB	0.34MB	3.8
Reco pp	6.8s	6.5s	1.0
Reco Pb	800s	810s	1.0

No Root compression yet

* was 22!!

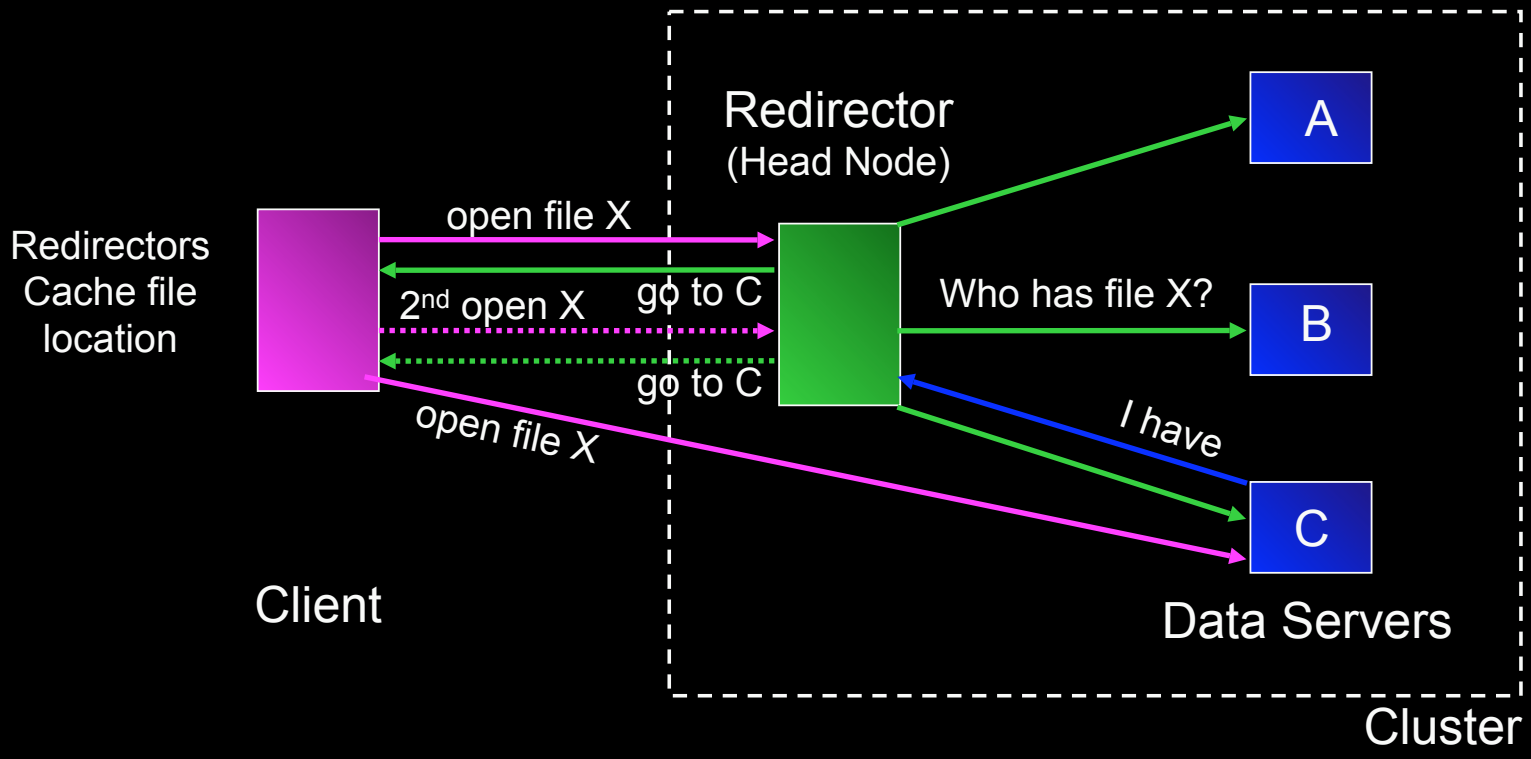
Pledged by external sites versus required (new LHC schedule) all		2008		2009		2010		2011		2012	
		T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
CPU	Requirement (MSI2K)	10.1	12.5	19.9	14.3	23.5	25.0	30.5	32.5	39.7	42.2
	Missing %	-33%	-28%	-37%	-7%	-32%	-27%	-48%	-44%	-60%	-57%
Disk	Requirement (PB)	3.9	1.7	6.8	4.0	12.0	4.3	16.8	5.6	22.7	7.3
	Missing %	-31%	6%	-30%	-10%	-35%	9%	-53%	-17%	-65%	-36%
MS	Requirement (PB)	5.7	-	12.4	-	19.8	-	27.3	-	34.1	-
	Missing %	-40%	-	-36%	-	-34%	-	-52%	-	-62%	-

The access to the data



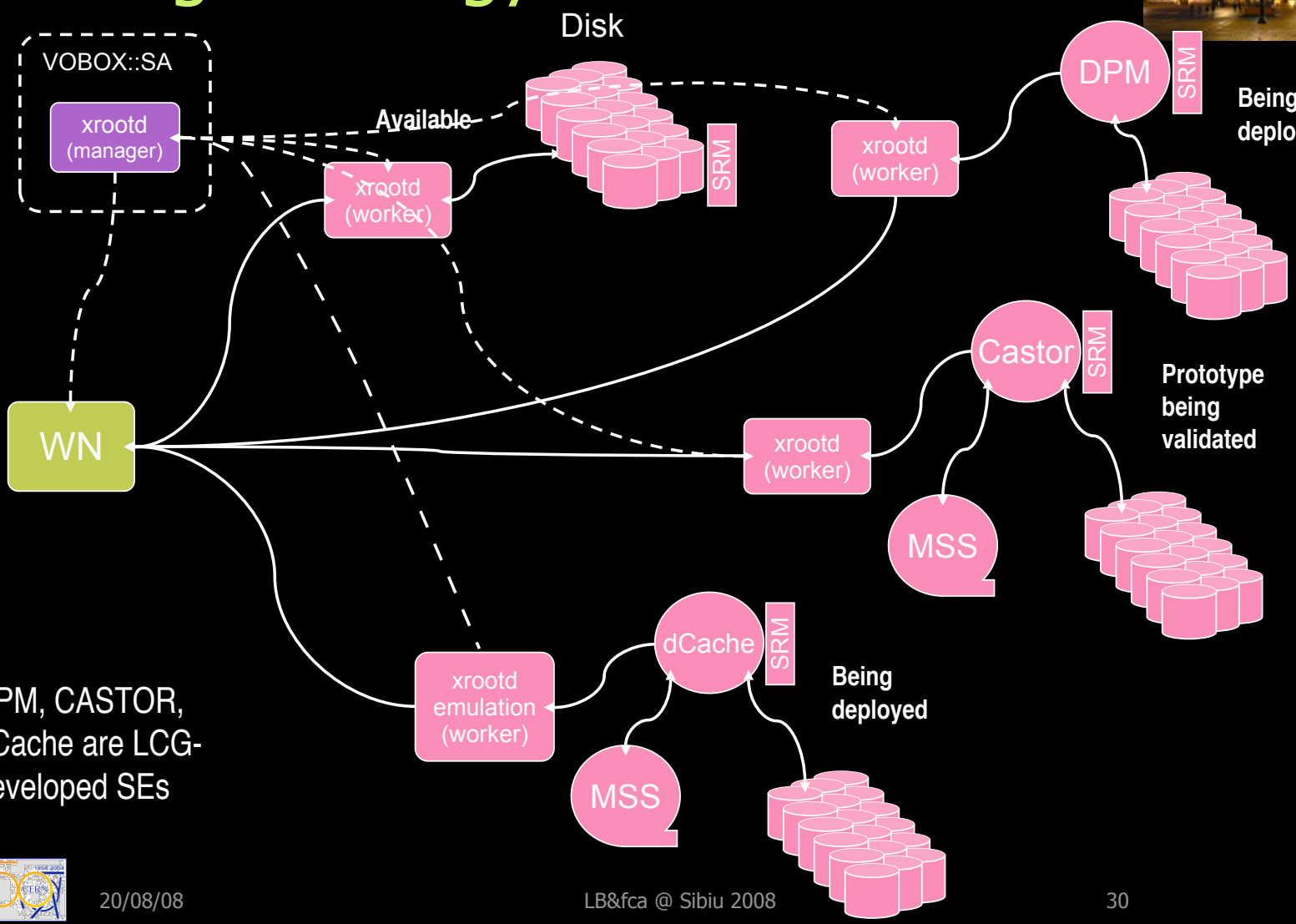
Direct access to data
via TAliEn/TGrid interface

Xrootd architecture



Client sees all servers as xrootd data servers

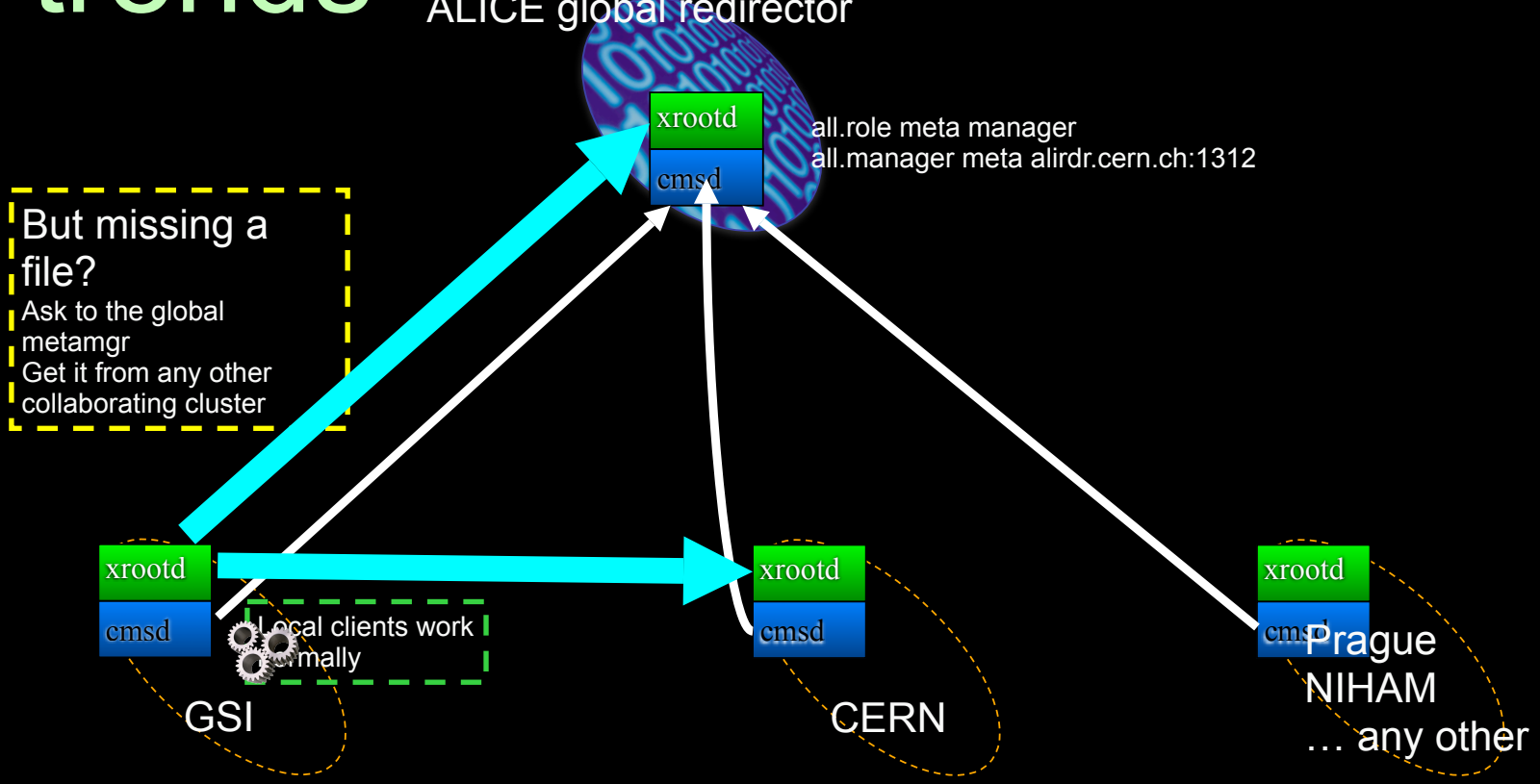
Storage strategy



DPM, CASTOR, dCache are LCG-developed SEs

Storage tech: new trends

ALICE global redirector

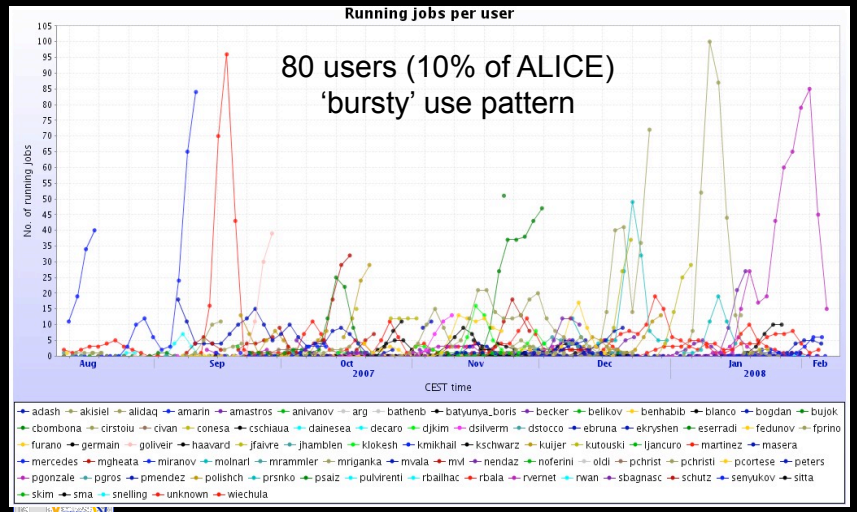


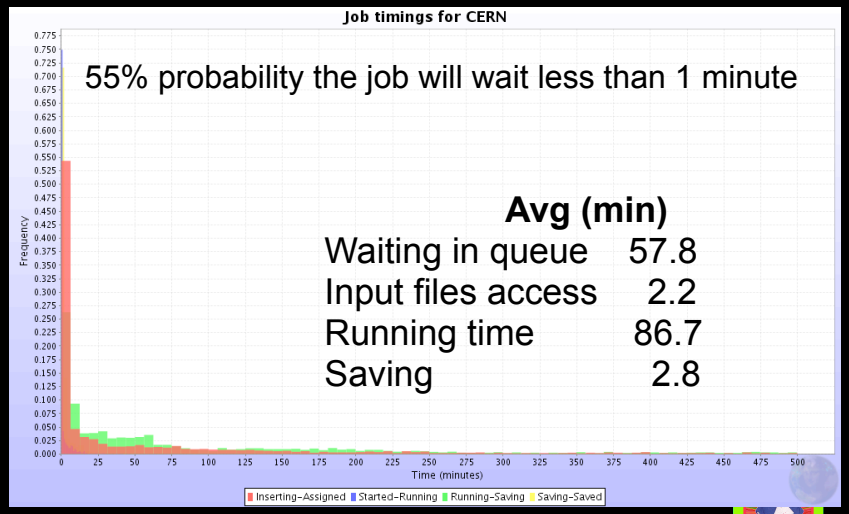
Analysis on the Grid - challenges

- Simplify the Grid job submission language
  - Introduction of storage types – removal of explicit storage elements
  - Already done for job management
- Provide stronger job optimization machinery
  - In particular fairshare between users
  - Remove entirely the need to specify any grid components
  - Automatic methods for data retrieval from MSS 'in advance' of the job landing on a CPU
    - This is the hardest part – currently effective analysis is only possible on pre-staged datasets

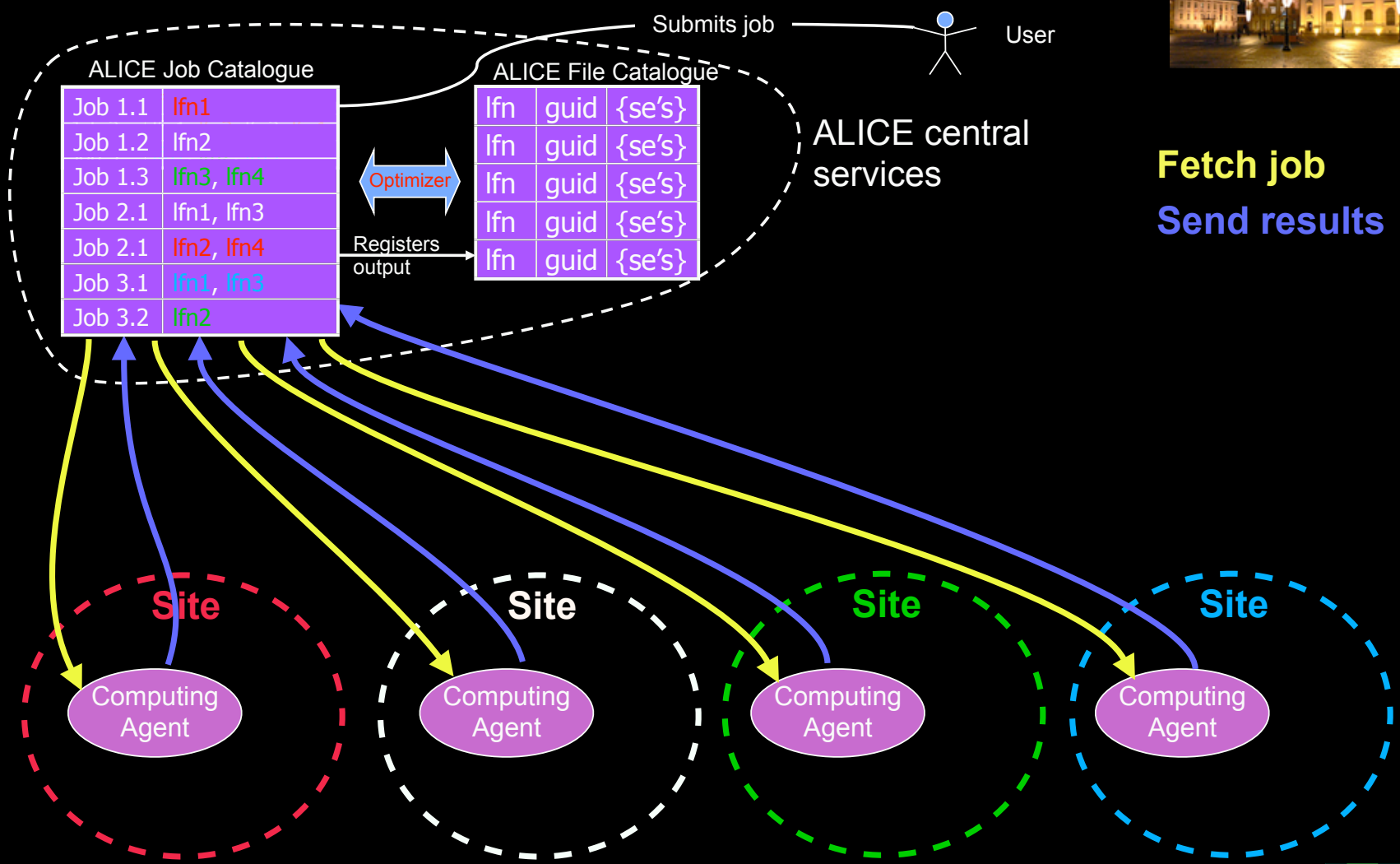
Chaotic analysis

- The most difficult activity to "tame"
  - We don't know how it looks like but will tell it when we see it
  - Can only be "simulated" till the users are <u>really</u> there (and then is too late!)
  - Response time and flexibility of the system are essential
    - NOT the strong point of today's Grid!





Sending jobs to data



Fetch job

Send results

The ALICE Grid

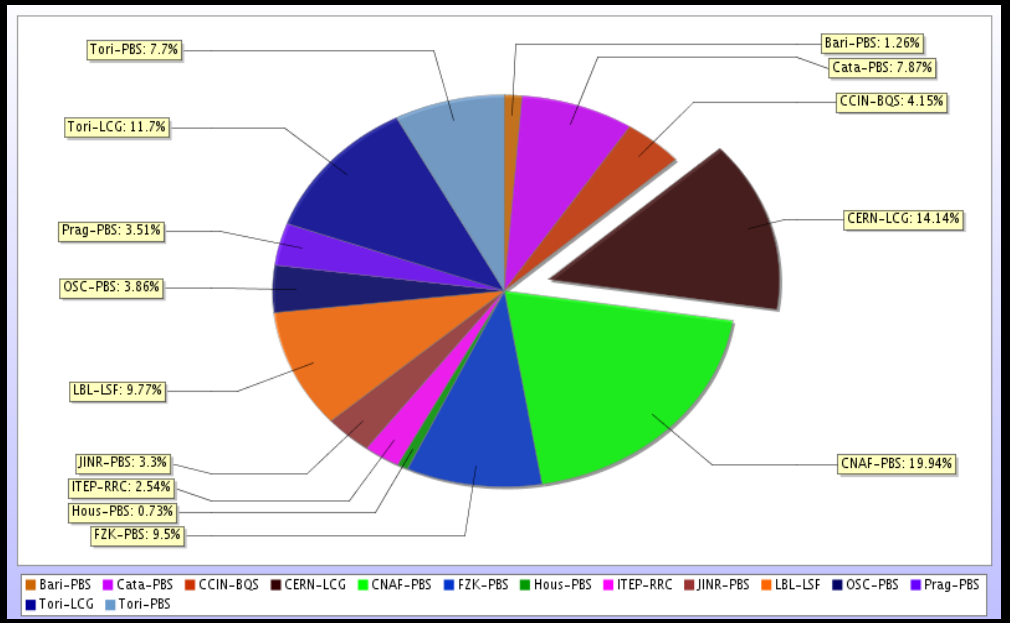
A short history of the ALICE Grid

- Working prototype in 2002
  - The ALICE Grid is born
- The Vision From the Very Beginning
  - Single interface to distributed computing for all ALICE physicists
  - File catalogue, job submission and control, application software management, end user analysis
  - And this is….
  - AliEn – Alice Environment

Toddler years 2003-2004

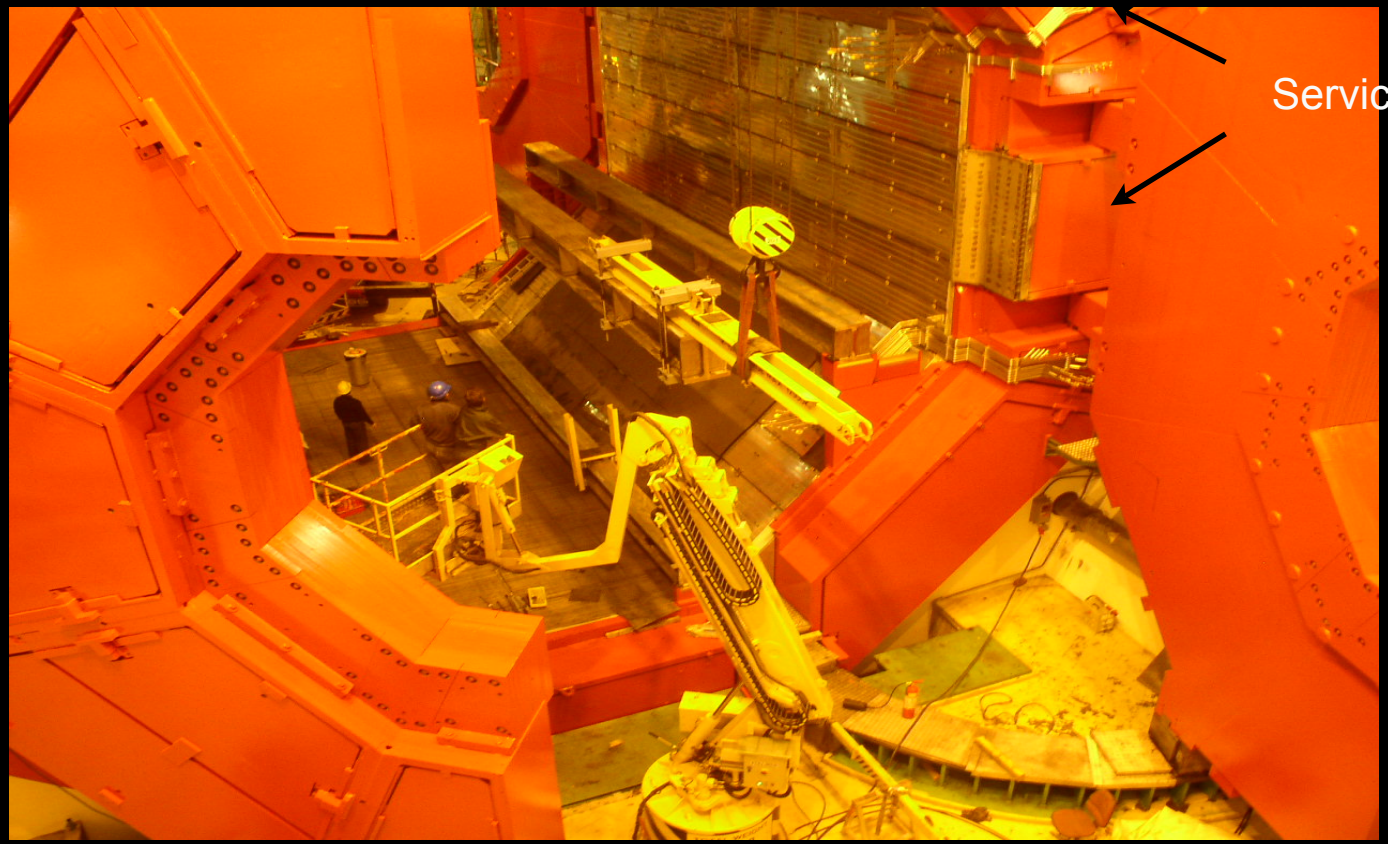
- First MC productions…
- Full vertical Grid – interfaced down to local batch system level and any type of local storage the site provides (capability retained and refined up to today)
- Few hundred CPUs at 13 sites
- Very ambitious goals – validation of the entire ALICE Computing model

Toddler years (2)



CPU work: 285 MSI-2K hours (one 2.8 GHz PC working for 35 years)

At that time ALICE was…

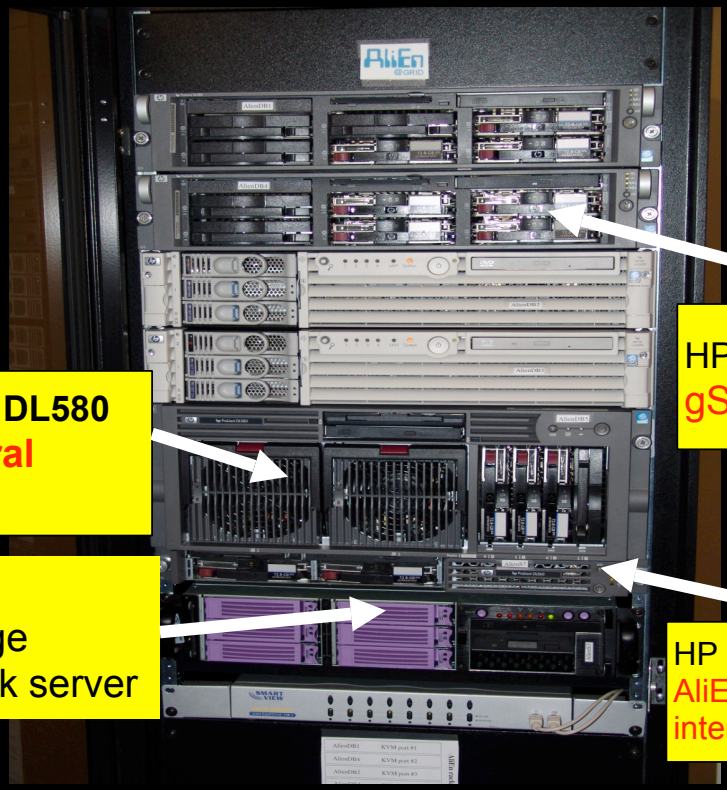


C.W.Fabjan

Terrible twos (and a bit beyond)… 2005-2006

- Increased sophistication of the central AliEn services
- Full integration with the LCG services
  - AliEn is moving toward 'Top level Grid management'
- Beginning of the story with storage (still ongoing)
- Increase the number of participating sites and CPUs used
- And more data challenges…this time with a solid backup of the Computing TDR
  - Published in June 2005

AliEn services in 2005



HP ProLiant DL580
AliEn central services

Disk SE
log files storage
1TB SATA Disk server

HP ProLiant DL380
gShell API service

HP ProLiant DL360
AliEn to CASTOR (MSS) interface

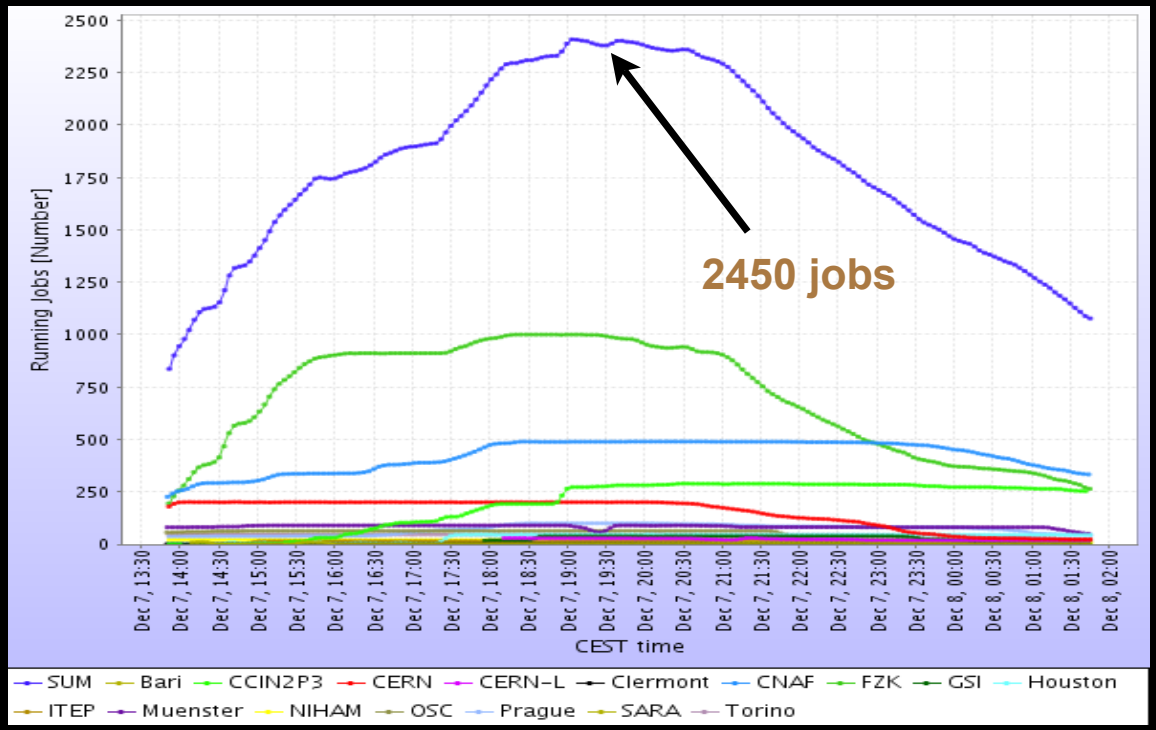
PDC results@end of 2005

- Data Challenge IV with LCG SC3: distributed MC production works routinely
- Monitoring: 25 sites operational, up to 1800 concurrent jobs, all using LCG resources, LCG services and ALICE services, thanks to our experience with AliEn

Running jobs (8 November)

Farm	Min	Avg	Max
Sum	1160	1651	1771
CCIN2P3	134	210	231
CERN-L	268	286	304
CNAF	255	362	394
FZK	0	531	600
Houston	0	3	14
Münster	2	58	81
Prague	43	61	71
Sejong	2	2	2
Torino	33	41	43

First records – Dec. 2005

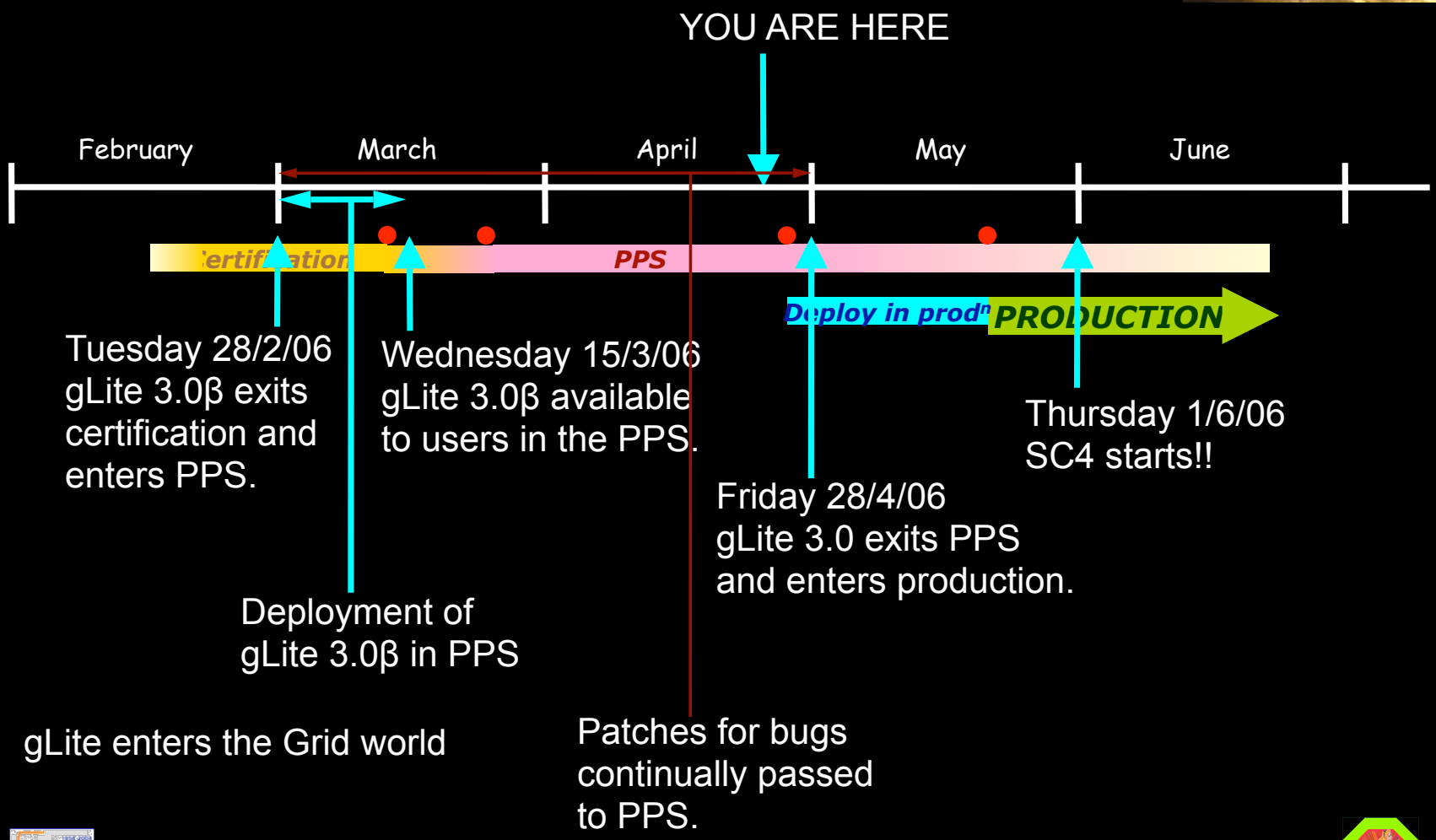


At that time ALICE was…

Beginning of the endless DC

- April 2006 – decision to run in a quasi-permanent mode
- Integration of new sites
  - Workload and storage
- Improving the robustness of central AliEn services
- Operational experience
- Test of AliRoot, new MC productions

PPS Schedule for gLite 3.0

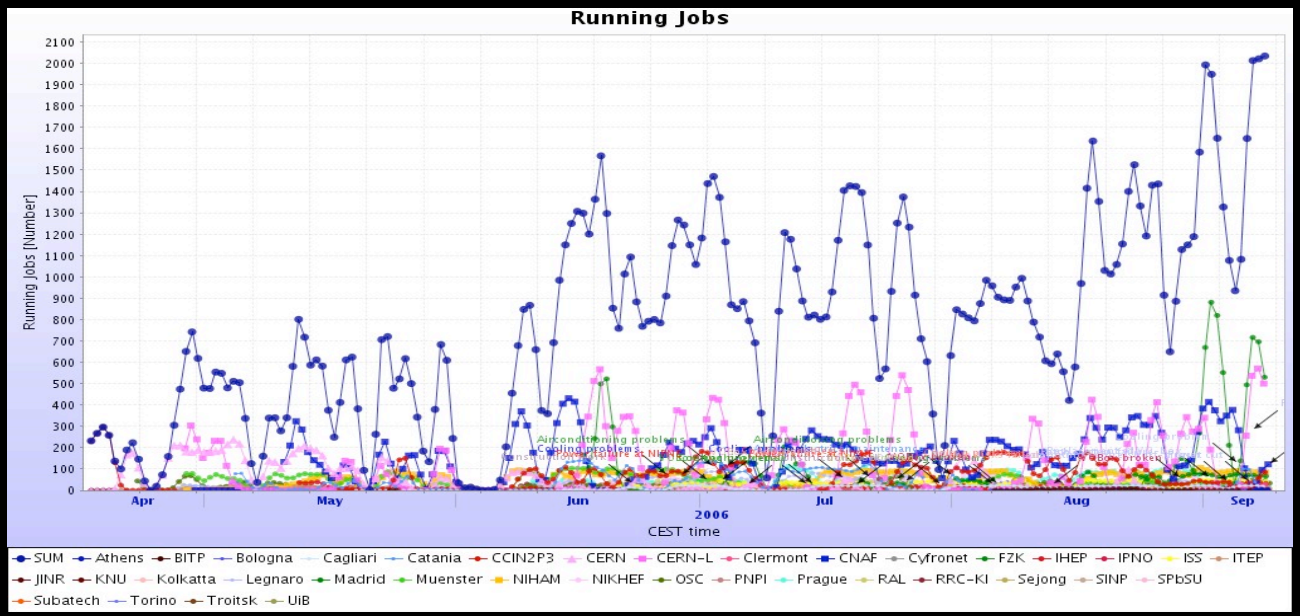


April 2006 - xrootd is adopted

- xrootd with MSS backend is operational at CERN
- Similar setup is being brought online at Lyon
- For other sites – experts are preparing the software and instructions for site experts how to install xrootd on the local storage servers
- Discussion of interoperation of various storage solutions (CASTOR2, DPM, dCache and xrootd ongoing
- SRM standard is still being discussed
- ALICE will use xrootd based storage and is actively pursuing its inclusion in the standard LCG package

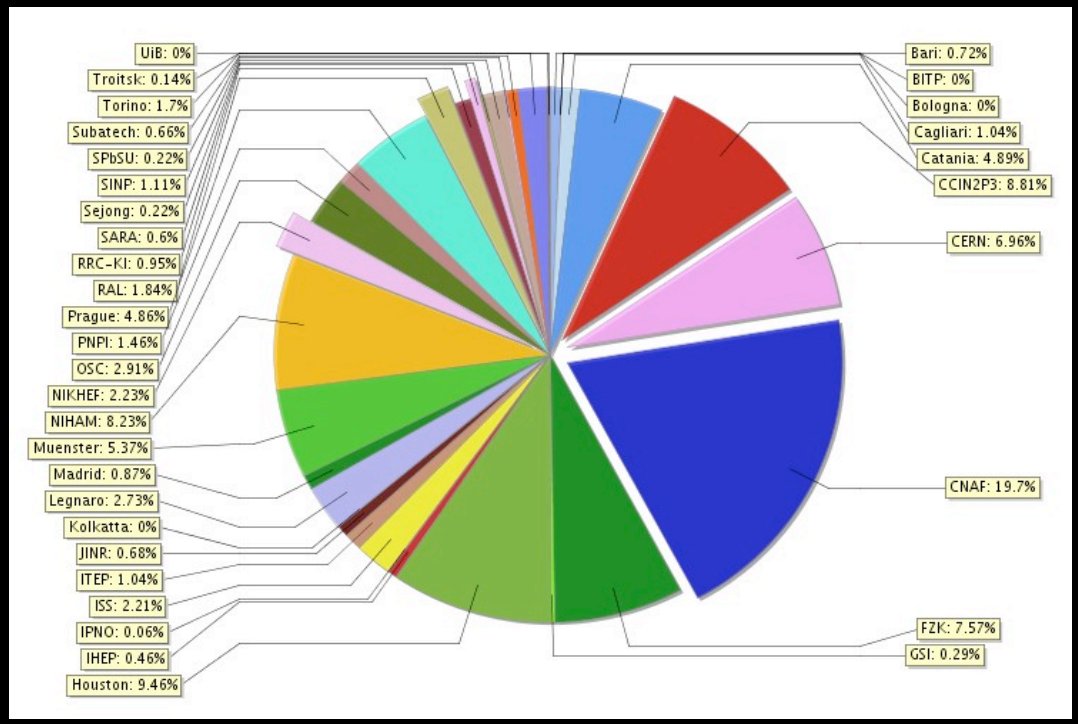
History of PDC'06

- Continuous running since Aplil 2006
  - Test jobs, allowing to debug all site services and test the stability
  - From July – production and reconstruction of p+p MC events



Resources statistics PDC'06

- Resources contribution (normalized Si2K units): 50% from T1s, 50% from T2s



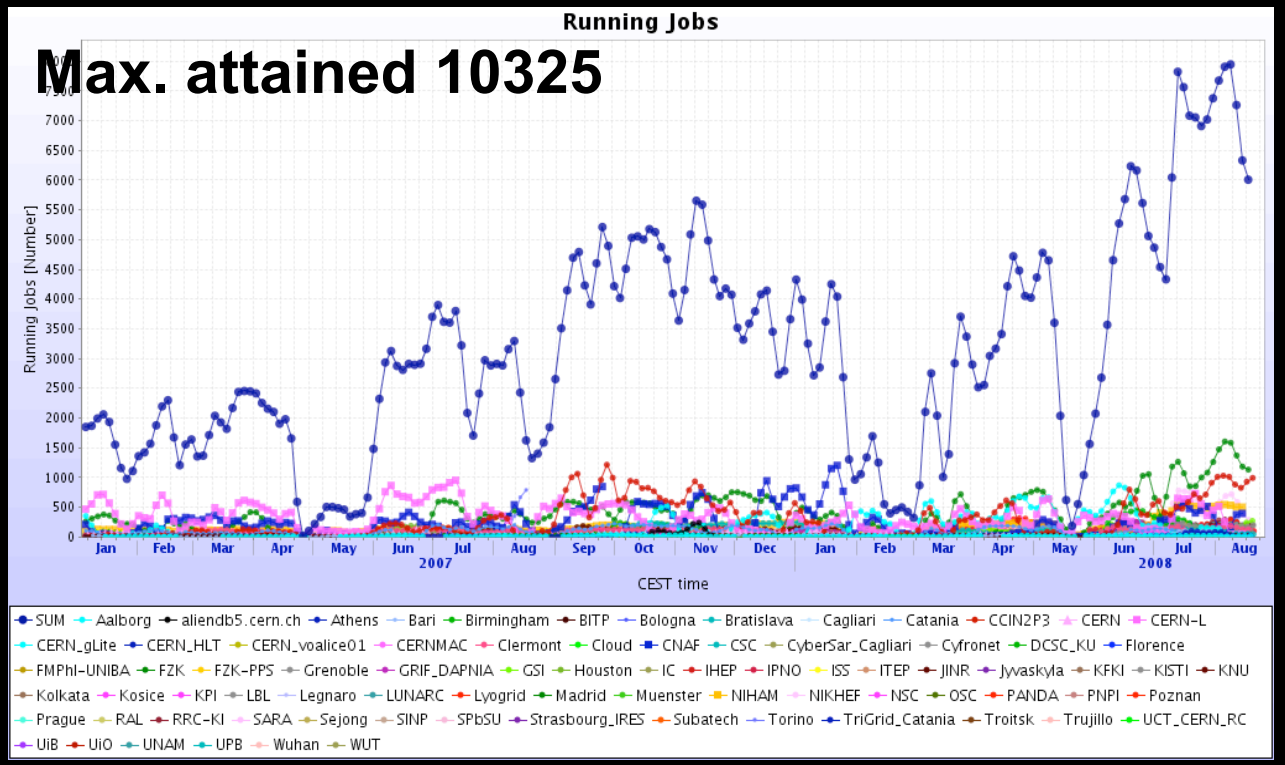
6 T1s, 30 T2s

At that time ALICE was…

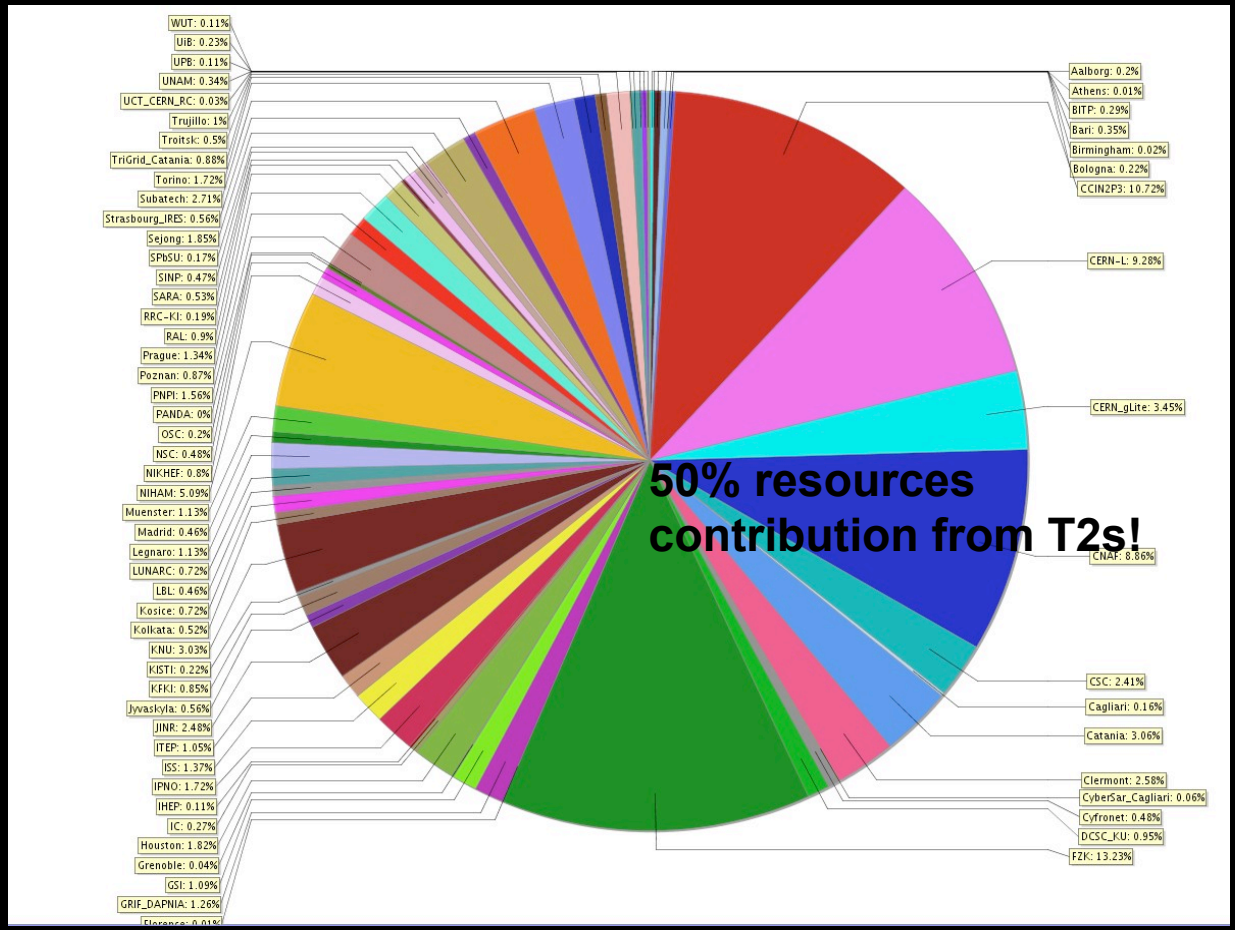
Kindergarten 2007-today

- Toward fully redundant central services
- Integration of gLite transfer and storage solutions
- Automatic site and central services management tools 'auto-pilot mode'
- Automatic production tools
- Integration of more sites, deployment of storage
- MC and RAW data production

Number of jobs evolution



Sites contribution



The ALICE Grid in numbers

- 73 participating sites
  - 1 T0 (CERN/Switzerland)
  - 6 T1s (France, Germany, Italy, The Netherlands, Nordic DataGrid Facility, UK)
  - 66 T2s spread over 4 continents
- As of today the ALICE share is some 7000 (out of ~30000 total Grid) CPUs and 1.5 PB of distributed storage
- In ½ year ~15K CPUs, x2 storage

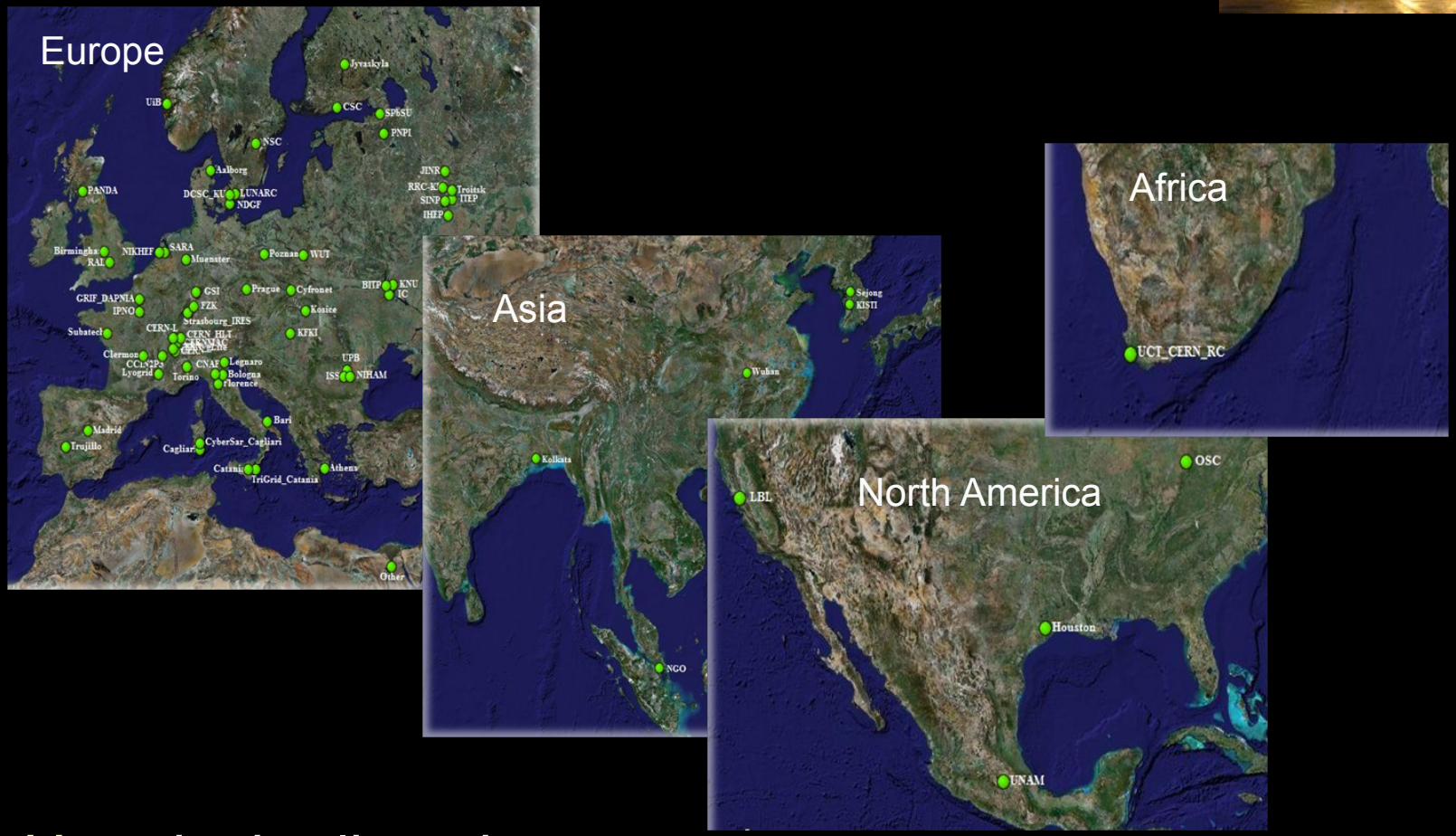
And Romania?

- Since 2001 NIHAM is participating to the ALICE Grid
  - One of the most stable and efficient ALICE T2s
- In 2007 NIHAM has deployed the first SE outside CERN
  - It provided the model for the other ALICE SEs
  - NIHAM is currently operating the largest ALICE SE@T2 (120TB)
  - A test site for new software and procedures
- In 2007 ISS also deployed an efficient and well-maintained T2
- NIHAM is the entry point for the ALICE usage of the Romania ALICE official LCG site (RO-07-NIPNE)

And Romania?

- According to LCG accounting NIHAM delivered ~60% of all CPU resources delivered by Romanian Tier2 Federation (ROT2F) to LHC
  - This represents ~74% of the CPU time delivered by ROT2F to ALICE
  - For the time being the NIHAM centre is not supported by Romania LCG
- NIHAM alone delivered 167% of the resources pledged to ALICE.
  - ISS and RO-07-NIPNE faired also well
- ROT2F delivered ~67% of the pledges to LCG (including NIHAM over-performance).
- The overall Romanian contribution to ALICE computing is quantitatively around 7%, and it is qualitatively invaluable

The ALICE Grid Map



Here is the live picture

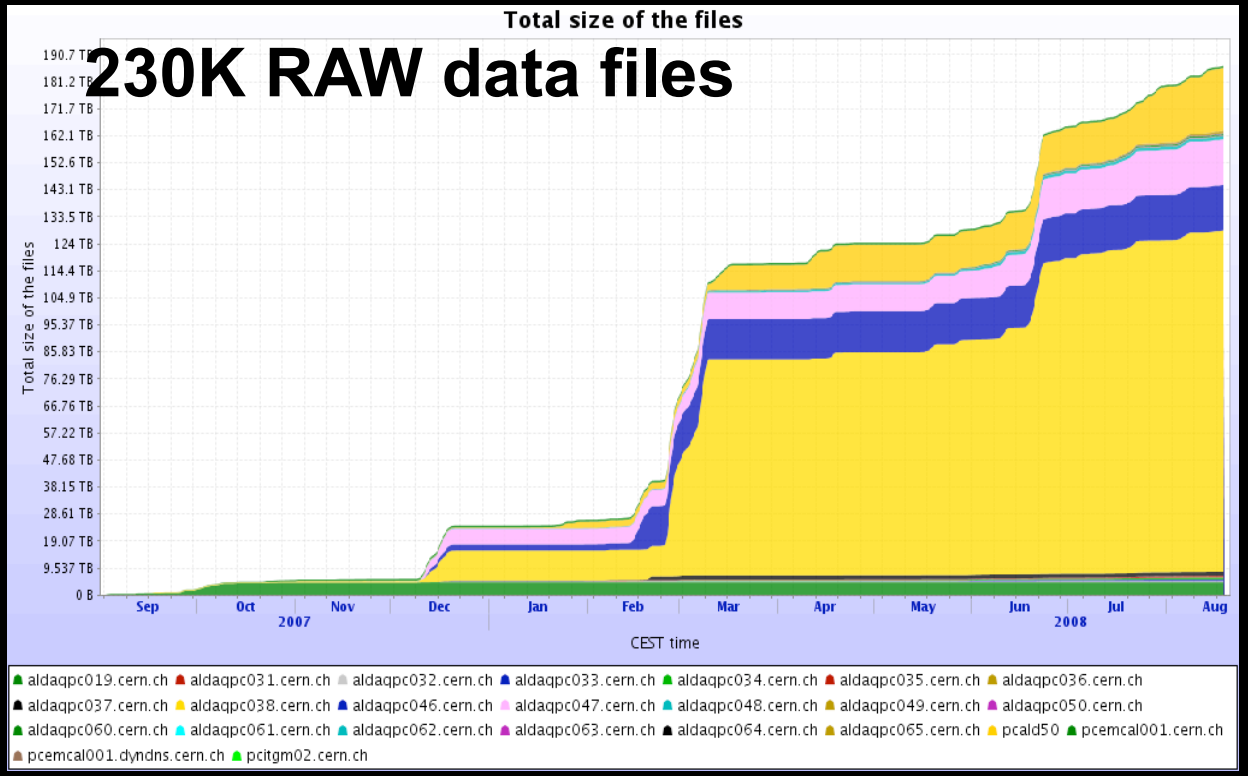
Control of the ALICE Grid

- Fully redundant
  - DB (MySQL) master-slave structure and backup
  - All central services run multiple instances
- Build servers for
  - I686, x86_64, ia63, MacOS 32- and 64-bit

ALICE Grid task list today

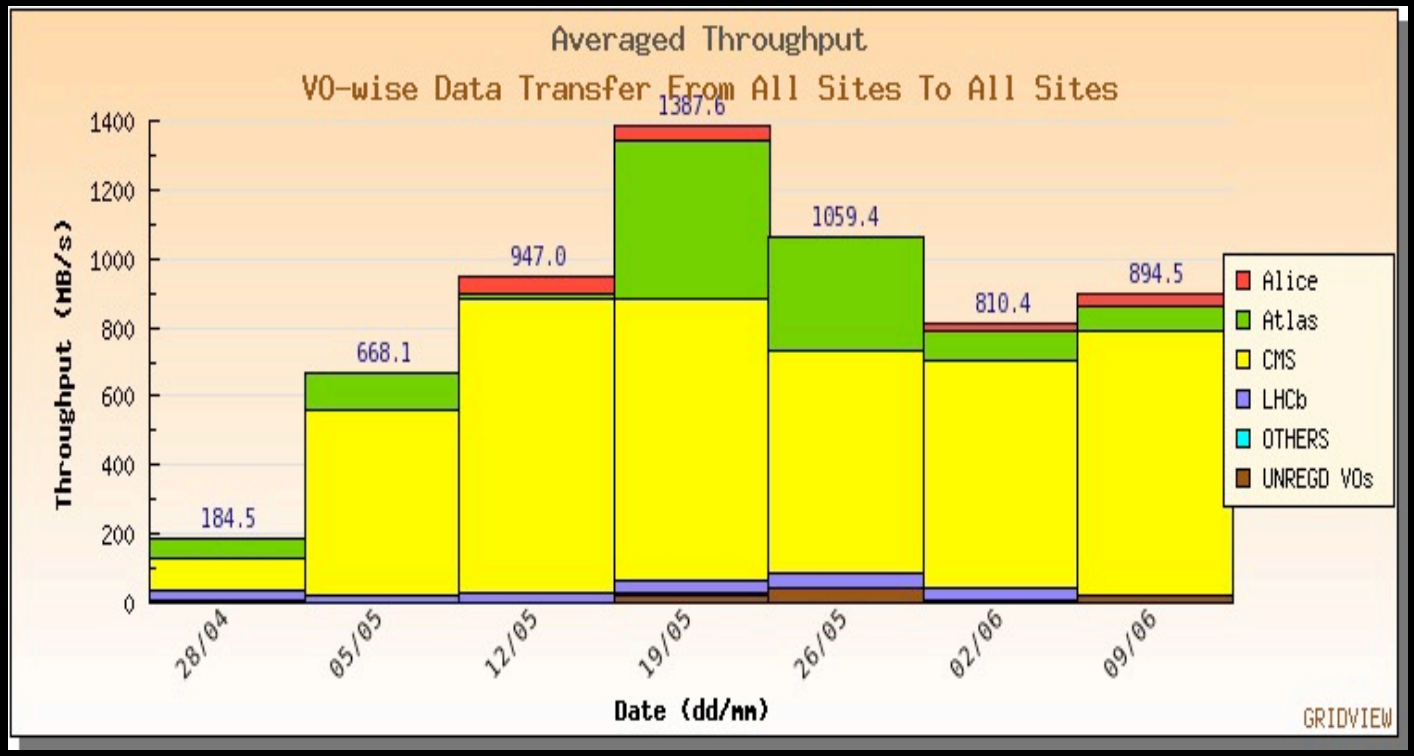
- Registration of data at MSS T0 and on the GRID
- Replication T0->T1
- Quasi-online reconstruction
  - Pass 1 at T0
  - Pass 2 at T1s
- MC production and user analysis

RAW data from cosmic tests



Replication of RAW

60MB/sec rate (p+p data taking scenario)
Using gLite transfer tools (FTS), operated through AliEn FTD



Data and MC production - emphasis

- Fast MC production for first physics with various LHC startup scenarios
  - And analysis for first publication
- Already 2 cycles made
- Fast analysis of detector calibration data
  - Essentially immediately after data taking (same day/night)
- Crucial for feedback to detector experts

Production of RAW

- Major activity in the past 6 months, very successful despite rapidly changing conditions
  - Both in the code and detector operation
  - In total some 120 TB of RAW passed through the reconstruction

Production	Description	Status	Run Range	Recorded chunks	Processed chunks	Comments
PDC 08/LHC08c_TPC	LHC08c_TPC partition	Running	47024 - 47032	36	0	RUN III cosmics - TPC
PDC 08/LHC08c	LHC08c global partition	Running	45321 - 48868	1,600	1,449	LHC08c global partition
PDC 08/LHC08b	LHC08b global partition	Completed	34784 - 42384	57,343	55,804	RUN III cosmics
PDC 08/LHC08a	LHC08a global partition - MUON reco	Completed	21392 - 26024	8,336	7,412	RUN II cosmics (Muon)
PDC 08/LHC08a	LHC08a global partition	Completed	18047 - 26042	43,464	39,003	RUN II cosmics

MC production since 2006

- Some 330Mio events with various physics content

Production	Description	Status	Run Range	Events Count	C
PDC 08/LHC08x	p+p, charm, forced had.decays	Completed	180001 - 180066	6,382,200	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	194,811,500	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
PDC 08/LHC08c9	First physics (stage 2) pp, Phojet, No field, 900GeV	Completed	377000 - 377002	247,950	Full misalignment/decalibration
PDC 08/LHC08c8	First physics (stage 2) pp, Phojet, 5kG, 900GeV	Completed	376000 - 376002	353,250	Full misalignment/decalibration
PDC 08/LHC08c7	First physics (stage 2) pp, Phojet, No field, 10TeV	Completed	375000 - 375002	257,700	Full misalignment/decalibration
PDC 08/LHC08c6	First physics (stage 2) pp, Pythia6, No field, 900GeV	Completed	374000 - 374002	256,050	Full misalignment/decalibration
PDC 08/LHC08c5	First physics (stage 2) pp, Pythia6, No field, 10TeV	Completed	373001 - 373002	226,950	Full misalignment/decalibration
PDC 08/LHC08c4	First physics (stage 2) pp, Phojet, 5kG, 10TeV	Completed	372000 - 372001	305,250	Full misalignment/decalibration
PDC 08/LHC08c3	First physics (stage 2) pp, Pythia6, 5kG, 900GeV	Completed	371000 - 371001	265,500	Full misalignment/decalibration
PDC 08/LHC08c2	First physics (stage 2) pp, Pythia6, 5kG, 10TeV	Completed	370000 - 370001	245,850	Full misalignment/decalibration
PDC 08/LHC08c10	T0+V0 First physics pp, 14TeV	Completed	470001 - 470021	2,950,350	Ideal geometry, V0 code buggy
PDC 08/LHC08c1	Therminator Pb+Pb, Cent.20-30%, 5.5 TeV	Pending	410002 - 410008	3,213	Full misalignment/decalibration
PDC 08/LHC08b6	First physics pp, Phojet, No field, 10 TeV	Completed	350000 - 350002	328,300	Full misalignment/decalibration
PDC 08/LHC08b5	First physics pp, Pythia6, No field, 900GeV	Completed	340000 - 340001	217,400	Full misalignment/decalibration
PDC 08/LHC08b4	First physics pp, Pythia6, No field, 10TeV	Completed	330000 - 330001	215,700	Full misalignment/decalibration
PDC 08/LHC08b3	First physics pp, Phojet, 5kG, 10TeV	Completed	320000 - 320001	219,000	Full misalignment/decalibration
PDC 08/LHC08b2	First physics pp, Pythia6, 5kG, 900GeV	Completed	310000 - 310001	429,400	Full misalignment/decalibration
PDC 08/LHC08b1	First physics pp, Pythia6, 5kG, 10TeV	Completed	300000 - 300006	431,300	Full misalignment/decalibration
PDC 07/LHC07g	MC pp, di-muon cocktail	Completed	200007 - 200317	25,446,300	Ideal geometry
PDC 07/LHC07f	MC pp min. bias for V0 studies	Completed	160000 - 160300	3,286,500	V0, Residual misalignment
PDC 07/LHC07e	MC PbPb central events, HUING	Completed	15000 - 15132	118,648	Residual misalignment
PDC 07/LHC07c	MC pp min.bias, 900 GeV	Completed	8000 - 8268	20,661,800	Ideal geometry
PDC 07/LHC07b	MC pp min.bias	Completed	6000 - 6011	1,128,300	Ideal geometry
PDC 07/LHC07a	MC pp min.bias	Completed	5000 - 5380	30,446,400	Residual misalignment and decalibration
PDC 06/06	MC pp min.bias	Completed	2000 - 2010	194,200	Vertex displacement in x 0.5 and 1 cm
PDC 06/05	MC pp min.bias	Completed	1000 - 1000	95,500	Full misalignment
PDC 06/04	MC pp min.bias	Completed	100 - 5380	51,404,500	Residual misalignment and decalibration
PDC 06/03	MC single-muon events	Completed	700 - 777	12,202,450	Residual misalignment and decalibration
PDC 06/02	MC di-muon events	Completed	503 - 596	4,882,200	Residual misalignment and decalibration
PDC 06/01	MC jet kinematics	Completed	1 - 9	1,800	42-50 GeV

Storage evolution

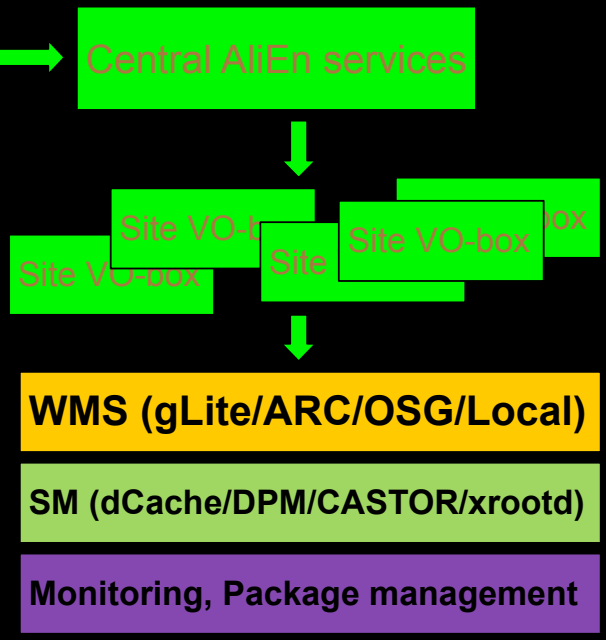
- Subject of intense development and deployment
  - For ALICE – xrootd as single access protocol
- 4 de-facto storage solutions
  - Pure MSS – CASTOR2
  - Hybrid disk+MSS – dCache
  - Disk – DPM, xrootd
  - Critical for analysis is the disk-based storage
- Good overall deployment progress and stability of all storage types
- Secure access through the ALICE security envelope – every operation is authorized through unique set of encrypted keys

Storage deployment

38 storage endpoints over 21 sites, ~1.5 PB

		Storage elements				
SE Name	AliEn name	SE Status	Size	Used	Free	Usage
1. Bari - dCache	ALICE::Bari::dCache	OK	9.77 TB	1.216 TB	8.553 TB	12.45%
2. Catania - DPM	ALICE::Catania::DPM	OK	45.63 TB	16.55 TB	29.08 TB	36.27%
3. CCIN2P3 - dCache	ALICE::CCIN2P3::dCache	OK	90.84 TB	18.65 TB	72.19 TB	20.53%
4. CCIN2P3 - dCache_sink	ALICE::CCIN2P3::dCache_sink	OK	838.2 TB	33.26 TB	804.9 TB	3.968%
5. CCIN2P3 - dCache_tape	ALICE::CCIN2P3::dCache_tape	OK	838.2 TB	32.21 TB	806 TB	3.842%
6. CERN - C2PPS	ALICE::CERN::C2PPS	OK	-	5.855 GB	-	-
7. CERN - Castor2	ALICE::CERN::Castor2	OK	931.3 TB	496.8 TB	434.5 TB	53.35%
8. CERN - Castor2X	ALICE::CERN::Castor2X	OK	931.3 TB	33.76 GB	931.3 TB	0.004%
9. CERN - DPM	ALICE::CERN::DPM	Not responding	-	-	-	-
10. CERN - se	ALICE::CERN::se	OK	5.588 TB	1.588 TB	4 TB	28.41%
11. CNAF - Castor2	ALICE::CNAF::Castor2	OK	30.73 TB	25.01 TB	5.725 TB	81.37%
12. CNAF - CASTOR2_sink	ALICE::CNAF::CASTOR2_sink	OK	931.3 TB	106.7 TB	824.6 TB	11.46%
13. FZK - dCache	ALICE::FZK::dCache	OK	423.8 TB	2.532 TB	421.2 TB	0.598%
14. FZK - dCache_sink	ALICE::FZK::dCache_sink	OK	2.728 PB	104.5 TB	2.626 PB	3.742%
15. FZK - dCache_tape	ALICE::FZK::dCache_tape	OK	2.728 PB	51.43 TB	2.678 PB	1.841%
16. GSI - dCache	ALICE::GSI::dCache	OK	953.7 GB	953.7 GB	9.069 MB	100%
17. GSI - se	ALICE::GSI::se	OK	35.01 TB	23.93 TB	11.08 TB	68.36%
18. GSI - se_tactical	ALICE::GSI::se_tactical	OK	27.94 TB	200.4 GB	27.74 TB	0.7%
19. ISS - File	ALICE::ISS::File	OK	4.581 TB	1.603 TB	2.978 TB	35%
20. ITEP - DPM	ALICE::ITEP::DPM	OK	23.44 TB	2.942 GB	23.43 TB	0.012%
21. JINR - dCache	ALICE::JINR::dCache	OK	51.76 TB	779.9 GB	51 TB	1.471%
22. Legnaro - dCache	ALICE::Legnaro::dCache		-	-	-	-
23. NDGF - dcache	ALICE::NDGF::dcache	OK	68.36 TB	21.73 TB	46.63 TB	31.79%
24. NDGF - dCache_sink	ALICE::NDGF::dCache_sink	OK	838.2 TB	25.3 TB	812.9 TB	3.018%
25. NDGF - dCache_tape	ALICE::NDGF::dCache_tape	OK	24.21 TB	24.38 GB	24.19 TB	0.098%
26. NIHAM - File	ALICE::NIHAM::File	OK	39.12 TB	8.38 TB	30.74 TB	21.42%
27. PNPI - DPM	ALICE::PNPI::DPM	OK	27.34 TB	5.588 GB	27.34 TB	0.02%
28. Prague - Disk	ALICE::Prague::Disk	OK	1.267 TB	1.267 TB	1 KB	100%
29. Prague - Disk2	ALICE::Prague::Disk2	OK	19.4 TB	9.318 TB	10.08 TB	48.03%
30. RAL - Castor2	ALICE::RAL::Castor2	OK	931.3 TB	0.341 GB	931.3 TB	0%
31. RAL - Castor2_sink	ALICE::RAL::Castor2_sink	OK	90.95 PB	0.351 GB	90.95 PB	0%
32. RRC-KI - DPM	ALICE::RRC-KI::DPM	OK	113.3 TB	2.971 GB	113.3 TB	0.003%
33. SARA - dcache	ALICE::SARA::dcache	OK	0	0	0	-
34. SARA - dCache_sink	ALICE::SARA::dCache_sink	OK	0	0	0	-
35. SARA - dCache_tape	ALICE::SARA::dCache_tape	OK	0	0	0	-
36. SPbSU - DPM	ALICE::SPbSU::DPM	OK	5.402 TB	31.22 GB	5.371 TB	0.564%
37. Subatech - DPM	ALICE::Subatech::DPM	OK	40.04 TB	4.715 TB	35.32 TB	11.78%
38. Torino - DPM	ALICE::Torino::DPM	OK	16.78 TB	1.083 TB	15.7 TB	6.454%

Current operation principle



- The VO-box system (very controversial in the beginning)
  - Has been extensively tested
  - Allows for site services scaling
  - Is a simple isolation layer for the VO in case of troubles

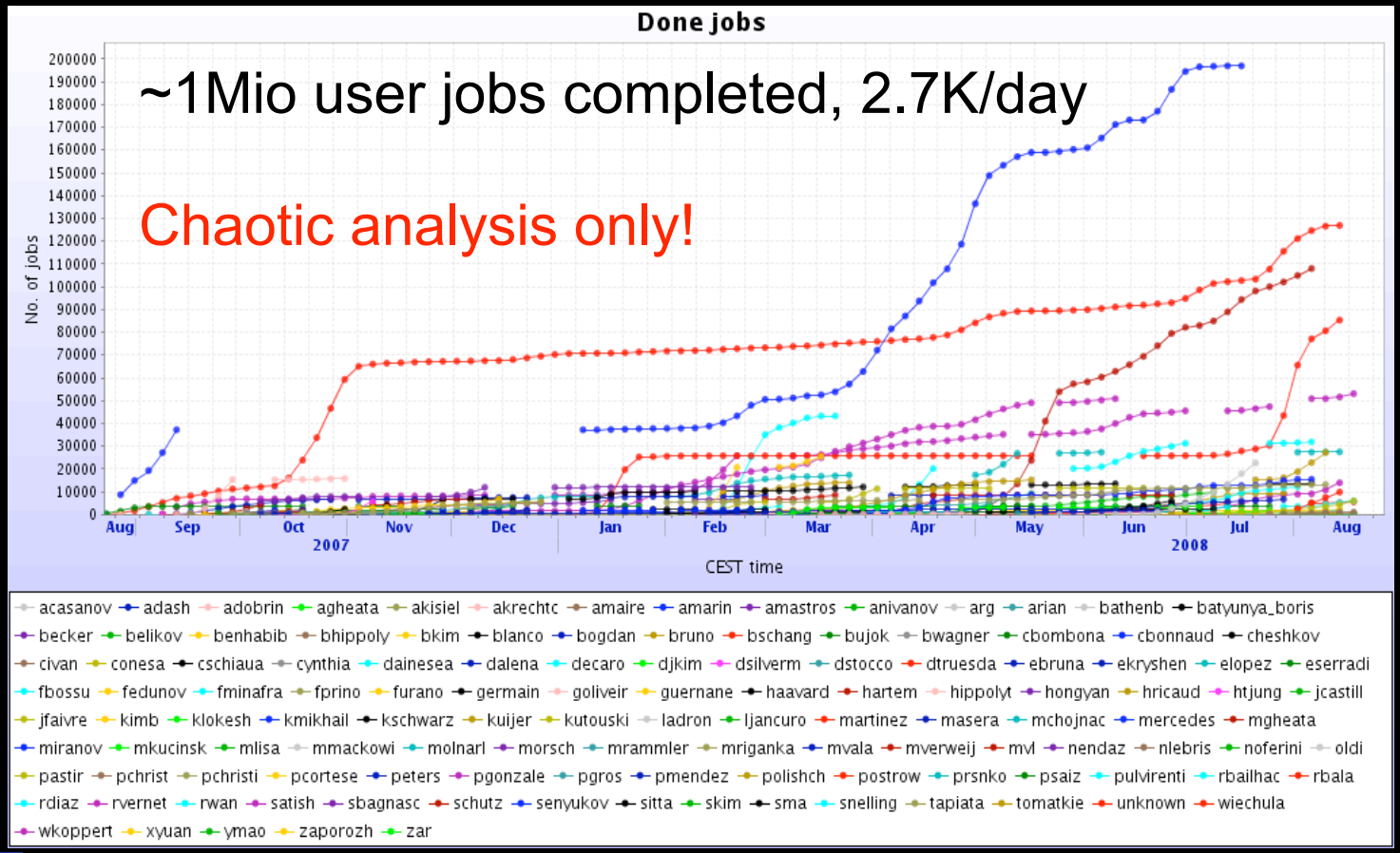
Operation – central/site support

- Central services support (2 FTEs equivalent)
  - There are no experts which do exclusively support – there are 6 highly-qualified experts doing development/ support
- Site services support - handled by 'regional experts' (one per country) in collaboration with local cluster administrators
  - Extremely important part of the system
  - In normal operation ~0.2FTEs/regions
- Regular weekly discussions and active all-activities mailing lists

User analysis activities

- Generally successful
- User jobs priorities are well mastered in the AliEn system
- Simple priority scheduling seem to work well, will be expanded soon to 'pay for what you use' principles
- Storage remains a weak point
  - Only the lack of it – available amount does not allow full (as per computing model) replication of data to be analysed
- As of today, 180 registered, 110 active users on the Grid
  - Not counting the MC/RAW production and CAF

User analysis statistics, one year



Monitoring, monitoring, monitoring

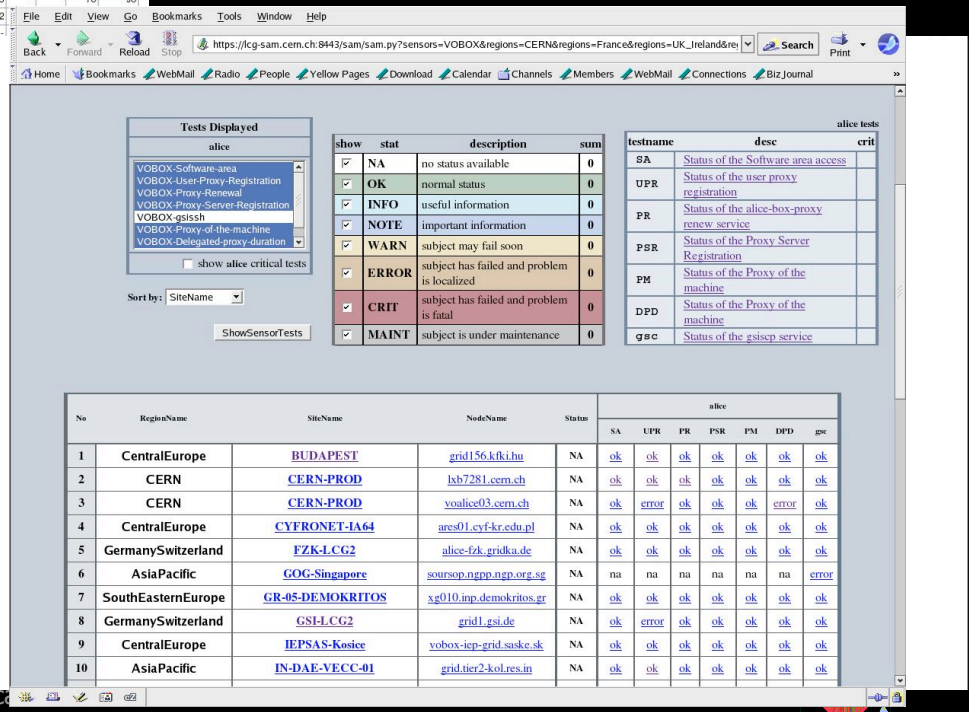
- Standard SAM tests to check LCG services availability are incorporated in the VO-box
- Available to Grid Support and ALICE (via ML)

VO Box machine status

What is this about?

Machine status (last hour average values)

		CPU										Mem [% MB]		Swap [% MB]		Eth0 [KB/s]		Eth1 [KB/s]		Eth2 [KB/s]		System	
Site name	Last see online	Load5	User	System	IOWait	Int	SoftInt	Nice	Steal	Idle	Cnt	MHz	Usage	Total	Usage	Total	In	Out	In	Out	In	Out	Procs	Socks
1. Aalborg	2007-02-01 09h	0.92	4.799	0.833	7.023	0.024	0.316	0	0	87	2	2793	77	2026	12.87	3121	8.181	8.038	55.49	2.278	-	-	169	144
2. Athens	2007-01-29 21h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3. Bari	2007-01-22 11h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4. Birmingham	2007-02-01 09h	0.201	7.804	1.719	0.166	0.014	0.29	0	-	90.01	2	800	55	1001	7.468	2047	13.54	34.98	-	-	-	-	94	39
5. BITP	2007-02-01 09h	0.411	1.026	0.431	12.29	0.018	0.215	0	-	86.03	4	3192	23	3999	0	8181	41.03	43.81	11.14	2.092	-	-	94	91
6. Bologna	2007-02-01 09h	0.063	1.208	0.479	0.077	0.003	0.114	0	-	98.12	4	3067	19	4005	0	2000	-	-	17.35	7.977	-	-	94	43
7. Cagliari	2007-02-01 09h	0.058	2.867	1.089	0.092	0.015	0.273	0	-	95.66	2	3199	34	2007	0	2000	22.28	11.75	-	-	-	-	90	55
8. Catania	2007-02-01 09h	0.202	3.447	0.965	0.725	0.017	0.17	0	-	94.68	4	2799	38	2006	0	4094	32.47	18.77	-	-	-	-	85	52
9. CCIN2P3	2007-02-01 09h	16.31	24.77	54.19	1.643	0.035	0.254	0	-	19.11	4	3000	51	2007	12.41	2047	1.683	0.093	32.87	21.05	-	-	250	229
10. CERN	2007-02-01 09h	1.019	16.86	4.55	1.106	0.037	0.893	0	-	76.56	4	2388	52	5788	2.959	8000	91.44	62	-	-	-	-	539	1179
11. CERN_gLite	2007-02-01 09h	0.099	3.849	1.337	0.388	0.005	0.164	0	-	94.26	2	3000	17	3995	0	4094	9.95	1.712	-	-	-	-	96	35
12. CERN-L	2007-02-01 09h	0.731	12.26	3.18	1.942	0.04	0.27	0	-	82.3	2	2793	22	3991	0	4094	34.33	14.08	-	-	-	-	119	55
13. CERNMAC	2007-02-01 09h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14. Clermont	2007-02-01 09h	0.081	5.753	1.057	0.012	0.167	0.164	0	-	92.85	1	2007	18	3013	0.007	8189	18.43	9.165	-	-	-	-	109	56
15. CNAF	2007-02-01 09h	0.023	1.519	0.687	0.071	0.011	0.164	0	-	97.55	2	3067	9	4005	0	4000	14.37	5.678	-	-	-	-	89	35
16. Cyfronet	2007-02-01 09h	0.037	1.708	0.296	0.288	0.011	0.073	0	-	97.62	2	1300	34	1982	0.186	1000	-	-	16.36	11.06	-	-	70	30
17. FZK	2007-02-01 09h	3.023	14.71	5.382	27.56	0.02	0.278	0	-	52.06	4	3000	67	2007	8.054	3827	23.81	15.24	51.01	84.12	-	-	-	-
18. GRIF_DAPNIA	2007-02-01 09h	0.118	6.042	1.096	0.133	0.053	0.105	0	-	92.57	1	2793	25	2001	0	2047	19.4	12.04	-	-	-	-	-	-
19. GSI	2007-02-01 09h	0.182	8.336	1.751	0.07	0.023	0.253	0	0	89.57	1	2667	58	820.9	0	512	42.78	23.17	-	-	-	-	-	-
20. Houston	2007-02-01 09h	0.05	2.898	0.578	0.049	0.011	0.347	0	-	96.12	1	1396	29	4014	13.04	4095	0.015	0.034	-	-	-	-	-	-
21. IHEP	2007-01-31 10h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22. IPNO	2007-02-01 09h	0.072	1.824	0.727	0.06	0.003	0.116	0	-	97.27	2	2394	15	3994	0	1992	16.51	8.048	-	-	-	-	-	-
23. ISS	2007-02-01 09h	1.943	4.217	3.768	15.06	0.16	0	0	-	76.8	4	2392	67	1009	49.62	996.2	2575	597.5	-	-	-	-	-	-
24. ITEP	2007-02-01 09h	0.073	2.831	0.892	0.464	0.008	0.161	0	-	95.65	2	2999	60	1000	3.843	1992	18.25	10.2	-	-	-	-	-	-
25. JINR	2007-02-01 09h	0.065	2.603	0.864	0.228	0.052	0.261	0	-	95.99	2	2793	47	2005	0	2047	20.68	243.4	-	-	-	-	-	-
26. Jyvaskyla	2006-12-12 14h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
27. KFKI	2007-02-01 09h	0.071	2.479	0.908	0.089	0.013	0	0	-	96.51	2	3392	15	4052	0	4094	16.3	8.323	-	-	-	-	-	-
28. KISTI	2007-01-25 18h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29. KNU	2007-02-01 09h	1.585	34.8	18.96	-	-	-	0	-	46.24	2	1000	54	1008	4.4	1913	32.55	63.95	-	-	-	-	-	-
30. Kolkata	2007-02-01 09h	0.215	2.486	2.308	4.129	0.021	0.273	0	-	90.78	4	2399	25	3829	0	2000	0.541	0.425	-	-	-	-	-	-
31. Kosice	2007-02-01 09h	0.091	6.247	1.283	0.176	0.124	0.086	0	-	92.08	1	3207	27	2009	0.133	2047	0.104	0.081	-	-	-	-	-	-
32. LBL	2007-01-30 11h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-



- Status of the VOBOX, ALICE and WLCG services are monitored through ML
- Sites are encouraged to check the status through these pages
- Alarm system established

Today, ALICE is ready…

Getting ready for the first school day

- In 6 short years the ALICE Grid has been transformed from a 'proof-of-concept' into a fully operational computational platform for ALICE offline work
- The ALICE computing model has been validated thoroughly through a series of data challenges with increasing complexity and scope
- The ALICE Grid today uses CPU and storage resources at some 73 computing centres, fully integrated into the local fabric and services
- The MC and RAW data production is a routine exercise, moving rapidly into an automated mode of operation
- User analysis is a routine exercise too…
- Services support on all levels is well understood
- The ALICE Grid is ready for its first day in school, coming, as usual, at the beginning of September


"That's
all
folks!"