The ALICE Computing and Data Model

P. Cerello (INFN – Torino)

T0/1 Network Meeting

Amsterdam

January 20/21, 2005

The ALICE Computing Model

- Objective:
  - Reconstruct and analyze real pp and heavy-ion data
  - Produce, reconstruct and analyze Monte-Carlo data

- Requirements/Boundary Conditions:
  - Serve a large community of users (~1000) distributed around the world (30 countries, 80 institutes)
  - Process an enormous amount of data (several PB/year)

- Solution:
  - Exploit resources distributed worldwide
  - Access these resources within a GRID environment

Latest updates (more will come… ☺)

- Dec. 9-10: draft computing model and projected needs discussed at an ALICE workshop
- Dec. 14: presentation to the ALICE Management Board
- Jan. 18: presentation to the LHCC
- The evolution will depend on:
  - Improved knowledge of the physics (particle multiplicity density) gained from RHIC + theory
  - Continuous optimization of required processing power and produced objects size (ESD, AOD)
  - Lessons learned from the Physics Data Challenges

The ALICE Computing TDR

- ALICE Computing TDR
  - Elements of the early draft provided to LHCC on Dec. 17, 2004
  - Draft will be presented during the ALICE/offline week in Feb. 2005
  - Approval foreseen during the ALICE/offline week in Jun. 2005

- Parameters
  - Data format, model and handling
  - Analysis requirements and model

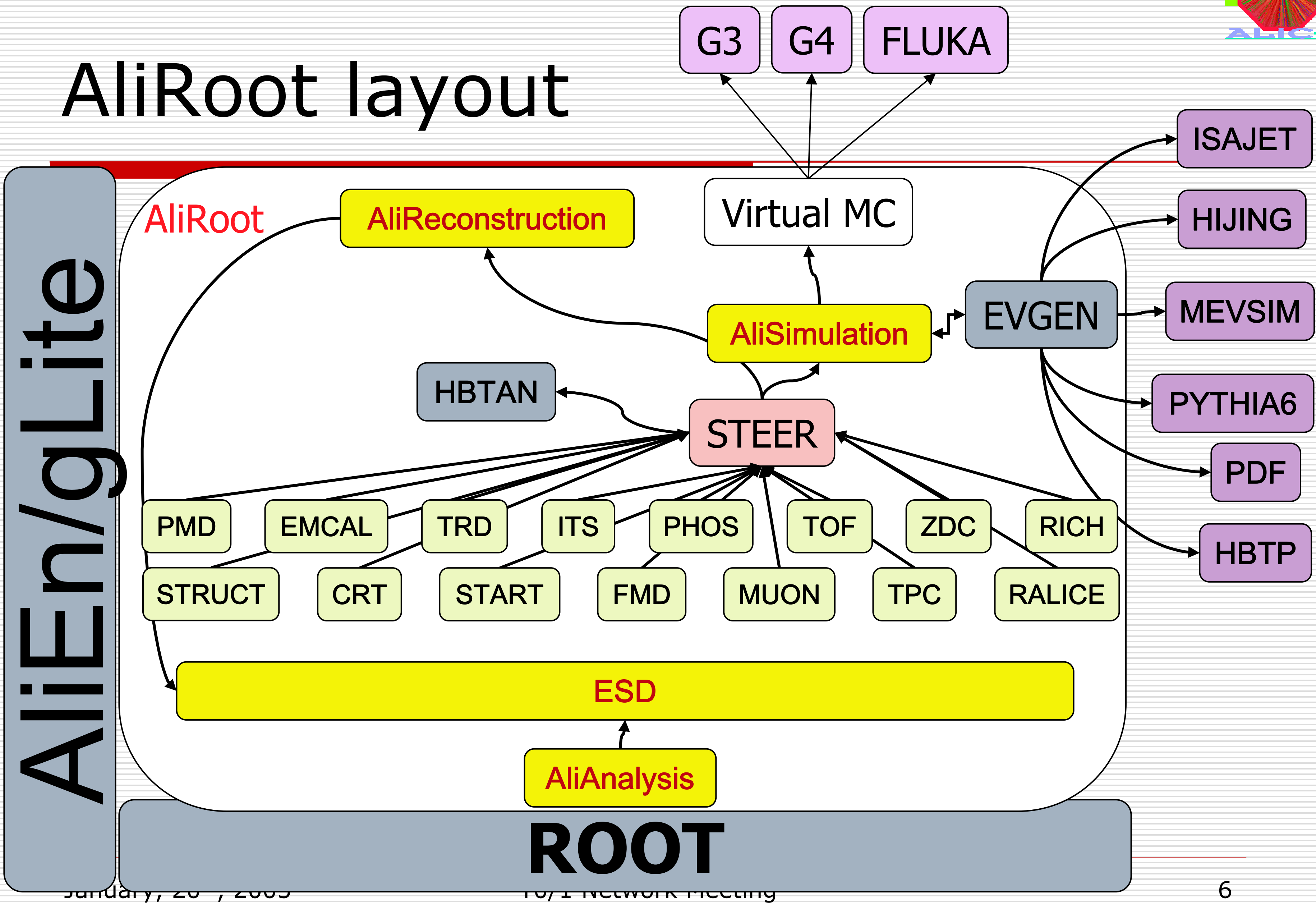
- Computing framework
  - Framework for simulation, reconstruction, analysis
- Distributed computing and Grid
  - T0, T1’s, T2’s, networks
  - Distributed computing model, MW requirements
- Project Organisation and planning
  - Computing organisation, plans, milestones
  - Size and costs: manpower

- Resources needed
  - CPU, disk, tape, network, services
  - Overview of pledged resources

Outline of the presentation

- The computing/data model
  - Framework (quickly)
- Experience with Data Challenge 2004
  - Configuration
  - Results
  - Lessons learnt
- Computing/Storage/Network needs
  - Data Handling model & issues
  - Data Flow (with numbers)

AliRoot layout



Physics Data Challenges

- We need:
  - Simulated events to exercise physics reconstruction and analysis
  - To exercise the code and the computing infrastructure to define the parameters of the computing model
  - A serious evaluation of the Grid infrastructure
  - To exercise the collaboration readiness to take and analyse data
- Physics Data Challenges are one of the major inputs for our Computing Model and our requirements on the Grid Middleware

ALICE Physics Data Challenges

Period (milestone)	Fraction of the final capacity (%)	Physics Objective
06/01-12/01	1%	pp studies, reconstruction of TPC and ITS
06/02-12/02	5%	• First test of the complete chain from simulation to reconstruction for the PPR • Simple analysis tools • Digits in ROOT format
01/04-06/04	10%	• Complete chain used for trigger studies • Prototype of the analysis tools • Comparison with parameterised MonteCarlo • Simulated raw data
NEW 05/05-07/05	TBD	• Test of condition infrastructure and FLUKA • Test of gLite and CASTOR • Speed test of distributing data from CERN NEW
01/06-06/06	20%	• Test of the final system for reconstruction and analysis

Experience from PDC'04

MC data
simulation,
reconstruction
(and analysis)

Do it all on the GRID(s)

Goals, structure and tasks

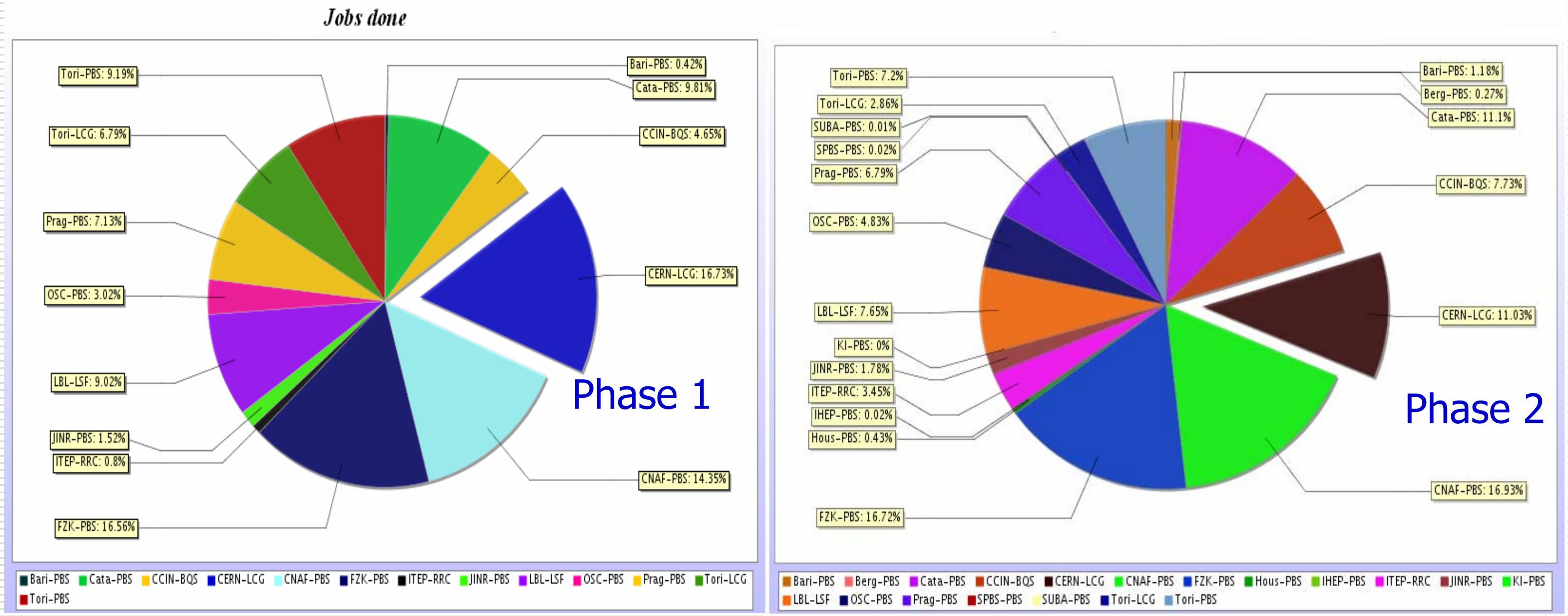
- Structure – logically divided in three phases:
  - Phase 1 - Production of underlying Pb+Pb events with different centralities (impact parameters) + production of p+p events
    - COMPLETED JUNE 2004
  - Phase 2 - Mixing of signal events with different physics content into the underlying Pb+Pb events (underlying events reused up to 50 times)
    - COMPLETED SEPTEMBER 2004
  - Phase 3 – Distributed analysis: to be started

Global PDC2004 statistics

- Job, storage, data volumes and CPU work:
  - Number and duration:
    - 400 K jobs
    - 6 hours/job
  - Number of files:
    - AliEn file catalogue: 9 M entries
    - 4 M physical files distributes at the AliEn SE's of 20 computing centres world-wide
  - Data volume:
    - 30 TB stored at CERN CASTOR
    - 10 TB stored at remote AliEn SEs + 10 TB backup at CERN
    - 200 TB network transfer CERN (T0) –> (T1/T2)
  - CPU work:
    - 750 MSi2K hours

Job repartition

- Jobs (AliEn/LCG): Phase 1 - 75/25%, Phase 2 – 89/11%
- More sites added to the ALICE GRID as PDC progressed



- 17 permanent sites (33 total) under AliEn direct control and additional resources through GRID federation (LCG)

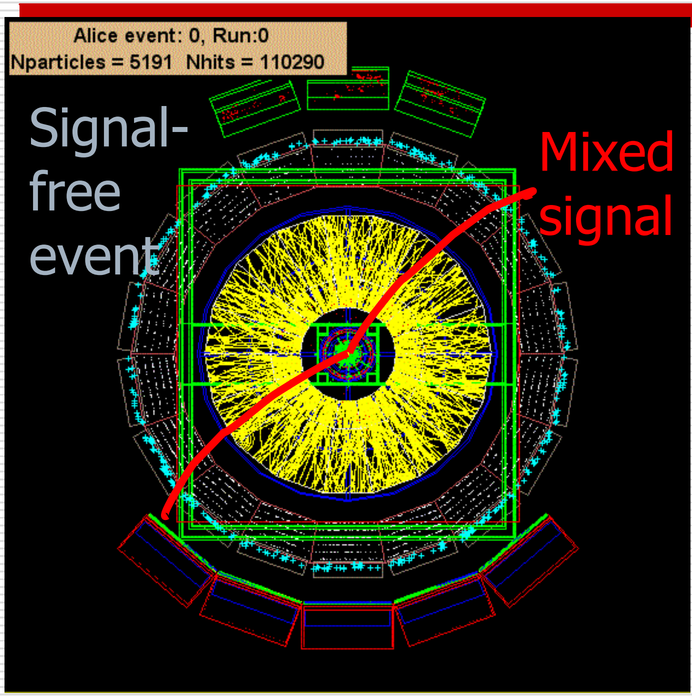
GRID efficiencies

- Network
  - Network utilization – minimized by the configuration of the PDC, have not seen any latency problems
- AliEn job failure rates calculations based on the job history
  - Major contributions:
    - 1% - internal AliEn errors, 8% - various errors at the CEs and SEs
    - The external errors are mostly spurious
    - The situation kept improving as the exercise advanced
- LCG job failures:
  - Calculation method – jobs are submitted to the LCG RB and expected to deliver the output (same as for AliEn)
  - Major contributors:
    - Phase 1 – jobs 'disappear' and no trace back is possible
    - Phase 2 – close/local SE failures – unable to save the output
    - Total job failure rate – 25-40%, mostly in Phase 2
    - Detailed information on the LCG GRID behaviour is available in the GAG document at

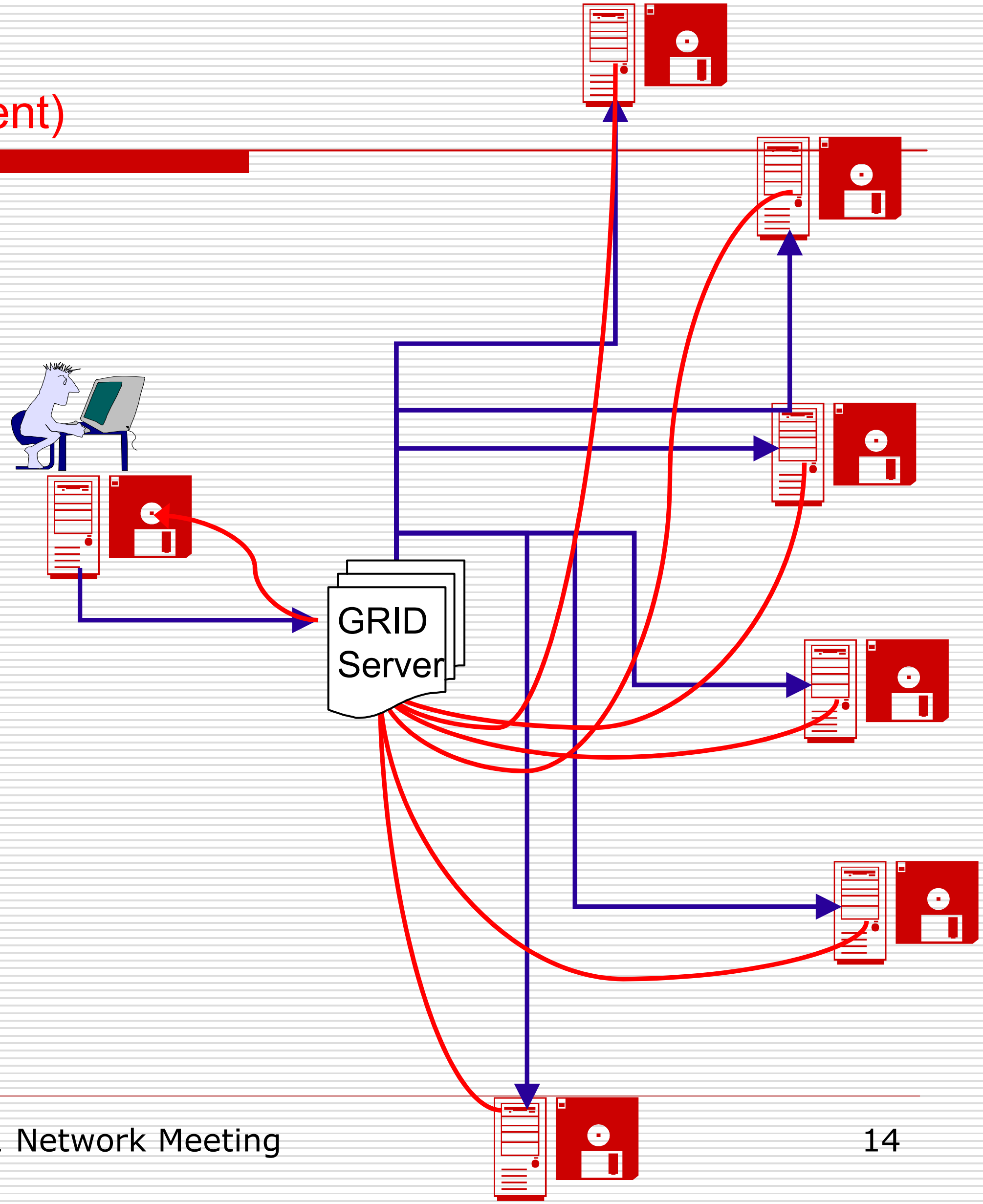
http://project-lcg-gag.web.cern.ch/project-lcg-gag/LCG_GAG_Docs_Public.htm

Phase III – (Interactive) Analysis

Large **distributed** input (2 MB/event)



Fairly small **merged** output



The distributed analysis – phase III

- Simplified view of the **ARDA E2E ALICE** analysis prototype:
  - ALICE experiment provides the UI (ROOT) and the analysis application
  - GRID middleware provides all the rest

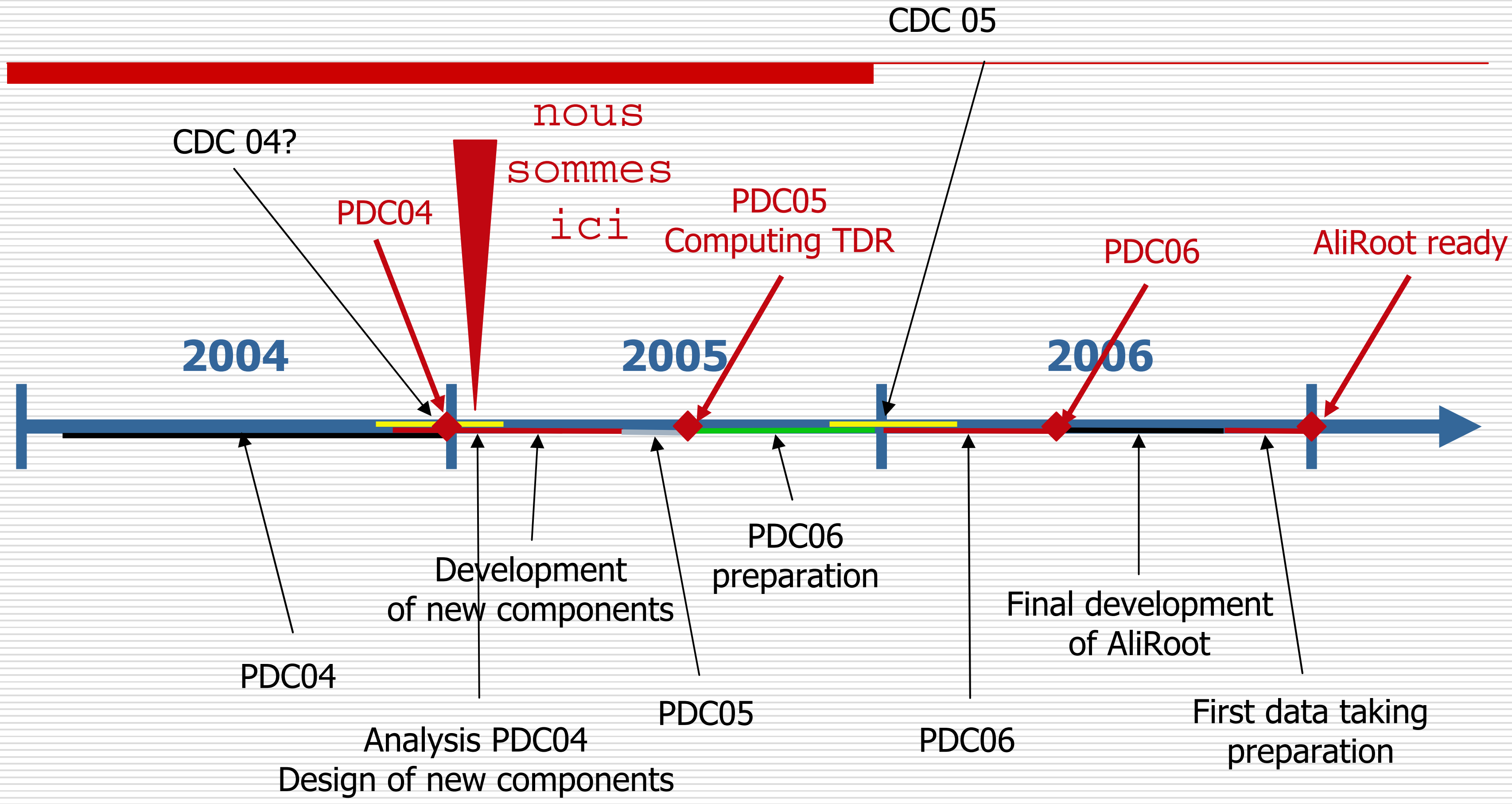
- Analysis possibilities:
  - interactive analysis mode: PROOF
  - batch analysis mode

Phase III - Lay...

Demo on SC2004, 3-rd ARDA workshop at CERN (October) and 2-nd EGEE conference at Hague (November)

...arallel, where they are stored
Ne... minimized by the configuration

ALICE Offline Timeline



The Computing Strategy

Boundary conditions
Processing strategy

ALICE computing model

- Static vs. Dynamic
  - Strict hierarchy of computing sites to which well defined tasks are assigned: Tier0, Tier1, Tier2,…

    vs.

  - Any task can be assigned to (taken by) sites with adequate free resources
- The GRID middleware selected implementation might intrinsically make a decision…
  - We assume a ‘cloud’ model: T2->T1 not strict

ALICE computing model/Assumptions

- We assume the latest schedule for LHC (peak L):
  - 2007 100d pp $5x10^6$s@$5x10^{32}$
  - 2008 200d pp 10^7s@$2x10^{33}$ 20d HI 10^6s@$5x10^{25}$
  - 2009 200d pp 10^7s@$2x10^{33}$ 20d HI 10^6s@$5x10^{26}$
  - 2010 200d pp 10^7s@10^{34} 20d HI 10^6s@$5x10^{26}$
- Staging of resources deployment during the initial period (cost reduction 40%/year):
  - 2007 20%;
  - 2008 40%;
  - 2009 100%.
- Reconstruction and simulation: scheduled tasks (PhysicsWorkingGroups, PhysicsBoard)
- Analysis: chaotic task eventually prioritized within PWG

Data format/flow

- RAW
  - Lightweight ROOT format tested in data challenges
  - No streaming (this might still change)

- Reconstruction produces ESD
  - Reconstructed objects (tracks, vertices, etc.)
  - Early/Detailed Analysis

- ESD are filtered into AOD, several streams for different analysis
  - Analysis specific reconstructed objects

- TAG are short summaries for every event with the event reference
  - Externalisable pointers
  - Summary information and event-level metadata

- Ion-Ion MC events are large due to embedded debugging information

Processing strategy

- For pp similar to the other experiments
  - Quasi-online reconstruction first pass at T0, further reconstruction passes at T1's
  - Quasi-online data distribution
- For AA different model
  - Calibration, alignment and pilot reconstructions during data taking
  - First reconstruction during the four months after AA run (shutdown) at T0, second and third pass distributed at T1's
  - Distribution of AA data during the four months after AA run
- we assume the Grid that can optimise the workload

Processing strategy

- Tier0
  - Computing: performs first reconstruction pass
  - Storage (permanent): one full copy of raw data, a share of ESD
- Tier1
  - Computing:
    - perform additional reconstruction passes (2 & 3)
    - Reconstruction on MC data
  - Storage (permanent): a share of the raw & MC data copy, ESDs
- Tier2
  - Computing: simulate and analyse Monte-Carlo data, analyse real data
  - Storage (permanent): shares of ESDs & AODs

Processing strategy / Network

- Tier0
  - Network:
    - OUT: 1 copy of raw data to Tier1
- Tier1
  - Network:
    - IN: 1 copy of raw data from Tier0
    - OUT: 1 copy of ESDs to Tier2 (x 2 times)
    - IN: 1 copy of MC raw data from Tier2
    - OUT: 1 copy of MC ESDs to Tier2
- Tier2
  - Network:
    - IN: 1 copy of ESDs from Tier1 (x 2 times)
    - OUT: 1 copy of MC raw data to Tier1
    - IN: 1 copy of MC ESDs from Tier1

Networking Numbers

- Most difficult to predict in absence of a precise (i.e. , tested) analysis model
- Net traffic T0⇒T1 can be calculated
  - Service data challenges will help here
- Traffic T1⇔T2 can also be calculated from the model, but it depends on Grid efficiency and analysis model
- Traffic T1⇔T1 & T2⇔T2 depends also on the Grid ability to use non local files and on the size of the disk cache available
  - A valid model for this does not exist (yet)

Uncertainties in the model

- No clear estimates of calibration and alignment needs
- No experience with analysis data access patterns
  - We will probably see "real" patterns only after 2007!
- We never tried to "push out" the data from T0 at the required speed
  - This will be done in the LCG service challenges
- We are still uncertain on the event size
  - In particular the pile-up in pp
  - ESD and AOD are still evolving

- We need to keep options open!

… now the numbers

ALICE computing model/Parameters

- Event statistics
  - Recoding rate: 100 Hz
  - MC: merge signal into reusable background
  - Same statistics for MC data as for real data

		pp	AA
Real data	(events/year)	1e9	1e8
MC data	background (events/year)	1e9	1e7
	Signal/background	-	10

ALICE computing model/Parameters

- Event size & Total size/year: 5.65 PB
  - Raw data: depends on
    - Particle multiplicity: unknown, assume dN/dy=4000
    - Centrality: take average between central and peripheral
    - Compression factor: take 2
  - MC: we know

	pp	AA
Real data (MB/event)	1	12.5
PB/year	1	1.25
MC data (MB/event)	0.4	300
PB/year	0.4	3.0

ALICE computing model/Parameters

- Reconstructed objects
  - Real data: we assume
    - ESD: 20% of raw size: 0.45 PB/year
    - AOD: 10% of ESD: 0.045 PB/year
  - MC: we know what we want to achieve

		pp	AA
Real data (MB/ev)	ESD	0.20	2.50
	AOD	0.050	0.250
	Event catalog	0.010	0.010
MC data (MB/ev)	ESD	0.04	2.14
PB/year		0.04	0.214

ALICE computing model/Parameters

- CPU power
  - Known for simulation and reconstruction, including future optimization
  - Guessed for calibration + alignment and for analysis

		pp	AA
CPU power (KSI2K $\times$ s / event)	Simulation	3.5E1	1.5E4
	Reconstruction	5.40	6.75E2
	Cal&Al	0.5	6E1
	Analysis	3	4E2

ALICE computing model/Parameters

- Repetition
  - 3 reconstruction passes
  - 23 analysis passes: 15 physicists analyze 10 times 1% of the data + 3 times full set, one per reconstruction pass
- Permanent data storage
  - Raw data: original at CERN + 1 copy distributed
  - Reconstructed and simulated: 1 set distributed
- Transient data storage (depends a lot on GRID)
  - Raw data: 2% at CERN, 10% at each Tier1, 24h buffer for export
  - Reconstructed data: 2 copies of one reconstruction pass distributed
  - MC data: 20% of everything distributed in Tier1s and Tier2s

ALICE computing model/Parameters

- Efficiency factors: adopted

Scheduled CPU	0.85
Chaotic CPU	0.60
Disk	0.70

ALICE computing model

- Total of CPU resources required per year:

		Tier0	Tier1	Tier2	Total
CPU (MSI2K)	Peak	7.5 22%	10.7 31%	15.8 47%	34.0
	Average	4.5 17%	10.6 41%	10.9 42%	26.0

ALICE computing model

Summary of Computing Capacities required by ALICE

	Tier0	Tier1	Tier2	Total
CPU (MSI2K)	4.5 17%	10.6 41%	10.9 42%	26.0
DisK (Pbytes)	0.5 5%	6.3 75%	1.7 20%	8.5
MS (Pbytes/year)	2.7 23%	8.7 77%	-	11.4

ALICE computing model

- Average capacity in T1 and T2 assuming:
  - 6 T1s:Lyon, CNAF, RAL, Nordic Countries, FZK, NIKHEF
  - 21 T2s

	Tier1	Tier2
CPU (MSI2K)	1.77	0.52
DisK (Pbytes)	1.05	0.08
MS (Pbytes/year)	1.3	-

ALICE computing model

- Network:
  - T0
    - IN: condition and raw data from DAQ
      - pp: 100 MB/s, 7 months, AA: 1.25 GB/s, 1 month, 24h disk buffer
    - OUT: condition and raw data and first pass ESD export to T1s
      - pp: 68 MB/s over 7 months, AA: 120 (600) MB/s, over 5(1) month(s), 24h disk buffer
  - T1
    - IN: condition and raw data and first pass ESD import, MC data from T2s: 22 MB/s, 12 months
    - OUT: ESD to T2s: 37 MB/s, 12 months
  - T2
    - IN: ESD from T1: 10-12 MB/s, 12 months
    - OUT: MC data to T1: 6-7 MB/s, 12 months

ALICE computing model

- Network total: averaged performance (rounded)

	T0	T1	T2
Network IN (Gb/s)	1.60	0.3 (1.0)	0.1
Network OUT (Gb/s)	1.0 (5.0)	0.3	0.05

Open issues

- Balance local-remote processing at T1's
  - We assume the Grid will be clever enough to send a job to a *free* T1 even if the RAW is not resident there
- Balance tape-disk at T1's
  - Will affect mostly analysis performance
- Storage of Simulation
  - Assumed to be at T1's
  - Difficult to estimate the load on the network
- Ramp-up
  - Our figures are calculated for a *standard year:* we need to work-out with LCG a *ramp-up* scenario
- T2's are supposed to *fail-over* to T1's for simulation and analysis
  - But again we suppose the Grid does this!

Conclusions

- ALICE choices for the Computing framework have been validated by experience
  - The Offline development is on schedule
- ALICE developed a Grid solution adequate to its needs
  - it future evolution is now uncertain, as a common project
  - this is a (non-technical) ***high-risk*** factor for ALICE computing
- ALICE developed a computing model from which predictions of the needed resources can be derived with reasonable confidence
- Numbers for CPU & Network might significantly change

Scope of the presentation

- Describe the current status of the ALICE Computing Model
- Describe the assumptions leading to the stated needs
- Give an overview of the future evolution of the ALICE Computing Project

Workplan in 2005

- Development of Alignment & Calibration framework
- Change of MC
- Continued collaboration with DAQ and HLT
- Continued AliRoot evolution
- Development of analysis environment
- Development of MetaData
- Development of visualisation
- Revision of detector geometry and simulation
- Migration to new Grid software
- Physics and computing challenge 2005
- Organisation of computing resources
- Writing of the computing TDR

Event statistics

Underlying events (Phase 1)

- 120 K events (30 TB of data) stored in CASTOR at CERN

Centrality name	Impact parameter value [fm]	Produced events
Cent1	0 - 5	20K
Per1	5 - 8.6	"
Per2	8.6 - 11.2	"
Per3	11.2 - 13.2	"
Per4	13.2 - 15	"
Per5	> 15	"

Phase 2 physics signals:

- 37 different signal conditions, necessary for the physics studies for the ALICE PPR.
- 1.5 M events, 340K jobs

Signal	No.of signal events per underlying	Number of jobs
Jets (un- and quenched) cent 1		
Jets PT 20-24 GeV/c	5	1666
Jets PT 24-29 GeV/c	5	1666
Jets PT 29-35 GeV/c	5	1666
Jets PT 35-42 GeV/c	5	1666
Jets PT 42-50 GeV/c	5	1666
Jets PT 50-60 GeV/c	5	1666
Jets PT 60-72 GeV/c	5	1666
Jets PT 72-86 GeV/c	5	1666
Jets PT 86-104 Gev/c	5	1666
Jets PT 104-125 GeV/c	5	1666
Jets PT 125-150 GeV/c	5	1666
Jets PT 150-180 GeV/c	5	1666
Total signal	399840	39984
Jets (un- and quenched) per 1		
Jets PT 20-24 GeV/c	5	1666
Jets PT 24-29 GeV/c	5	1666
Jets PT 29-35 GeV/c	5	1666
Jets PT 35-42 GeV/c	5	1666
Jets PT 42-50 GeV/c	5	1666
Jets PT 50-60 GeV/c	5	1666
Jets PT 60-72 GeV/c	5	1666
Jets PT 72-86 GeV/c	5	1666
Jets PT 86-104 Gev/c	5	1666
Jets PT 104-125 GeV/c	5	1666
Jets PT 125-150 GeV/c	5	1666
Jets PT 150-180 GeV/c	5	1666
Total signal	399840	39984

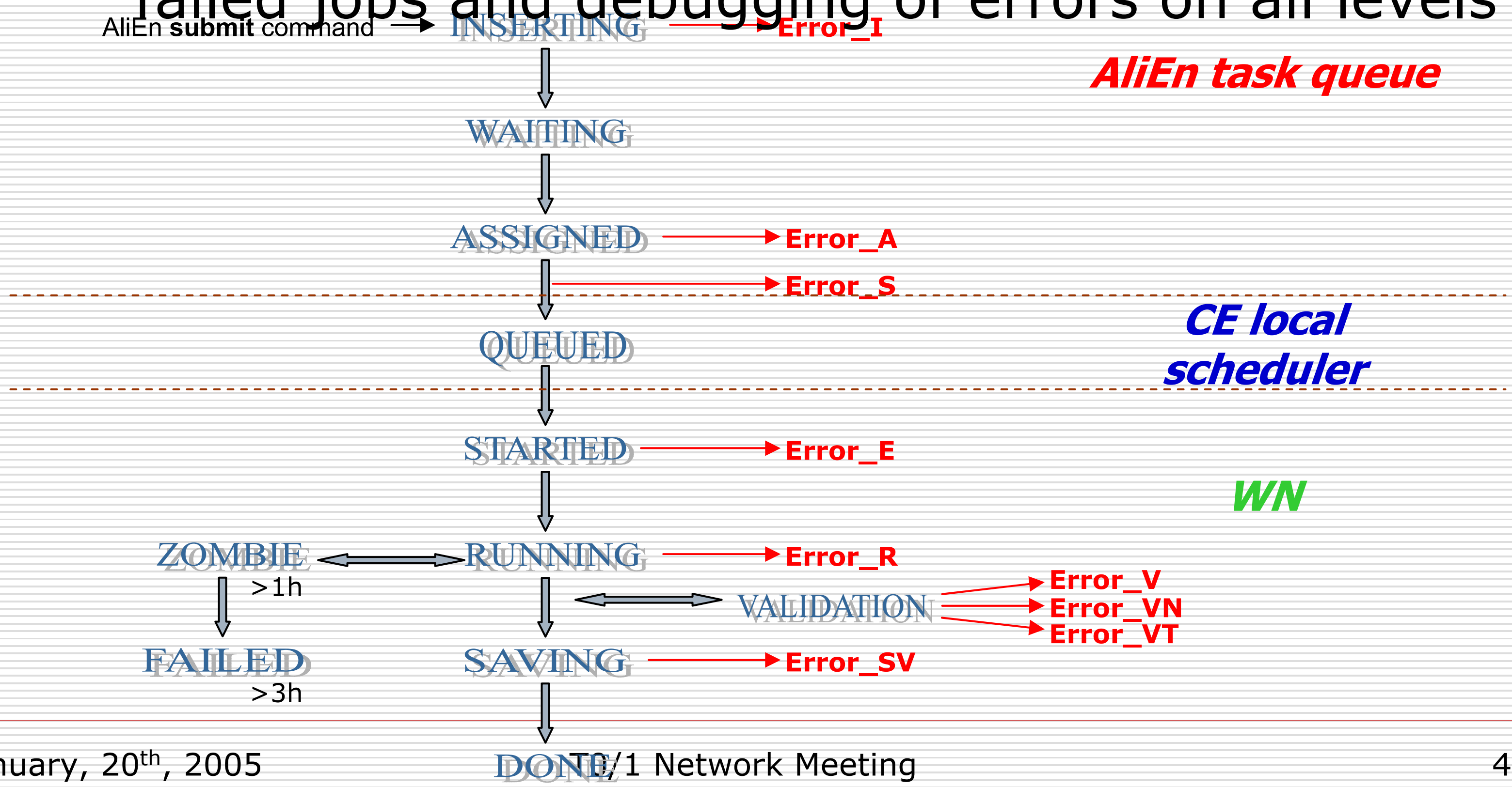
PHOS	**cent 1**	
Jet-Jet PHOS	1	20000
Gamma-jet PHOS	1	20000
Total signal	40000	40000
D0	**cent 1**	
D0	5	20000
Total signal	100000	20000
Charm & Beauty	**cent 1**	
Charm (semi-e) + J/psi	5	20000
Beauty (semi-e) + Y	5	20000
Total signal	200000	40000
MUON	**cent 1**	
Muon coctail cent1	100	20000
Muon coctail HighPT	100	20000
Muon coctail single	100	20000
Total signal	6000000	60000
MUON	**per 1**	
Muon coctail per1	100	20000
Muon coctail HighPT	100	20000
Muon coctail single	100	20000
Total signal	6000000	60000
MUON	**per 4**	
Muon coctail per4	5	20000
Muon coctail single	100	20000
Total signal	2100000	40000
Grand total	15239680	339968

Principles and platforms

- True GRID data production and analysis: all jobs are run on the GRID, using only ***AliEn*** for access and control of native computing resources
- LCG GRID resources: access through AliEn-LCG interface
- In phase 3: gLite+***PROOF*** ***with ARDA E2E Prototype for ALICE***
- Reconstruction and analysis software distributed remotely by AliEn: AliRoot/GEANT3/ROOT/gcc3.2 libraries:
  - The AliROOT code was kept backward compatible throughout the exercise
- Heterogeneous platforms:
  - Various types of scheduling systems: LSF, BQS, PBS, SGE, Condor, Fork
  - Multitude of storage element types: NFS, CASTOR, HPSS, dCache (untested)
  - GCC 3.2 + ia32-bit Cluster
  - **GCC 3.3 + ia64 Itanium Cluster**

Monitoring – AliEn

- Sophisticated monitoring system:
  - Job tracking from submission to finish – 11 different states with 9 possible error conditions
  - Essential for the operation, resubmission of failed jobs and debugging of errors on all levels



Software management

- Regular release schedule
  - Major release every six months, minor release (tag) every month
- Emphasis on delivering production code
  - Corrections, protections, code cleaning, geometry
- Nightly produced UML diagrams, code listing, coding rule violations, build and tests , single repository with all the code
  - No version management software (we have only two packages!)
- Advanced code tools under development (collaboration with IRST)
  - Aspect oriented programming
  - Smell detection
  - Automated testing

Condition DataBases

- Information comes from heterogeneous sources
- All sources are periodically polled and ROOT files with condition information are created
- These files are published on the Grid and distributed as needed by the Grid DMS
- Files contain validity information and are identified via DMS metadata
- No need for a distributed DBMS
- Reuse of the existing Grid services

Operation methods and groups

- Phase 1 and 2:
  - Central job submission – one person in charge of everything
- Phase 3:
  - Many users with centralized user support
- 2 ALICE experts responsible for:
  - The operation of the core AliEn services
  - Monitoring of jobs, remote CEs and SEs
- CERN storage and networking: IT/FIO, IT/ADC
- LCG operation: IT Grid Deployment Team
- Local CE/SE: one local expert (typically the site administrator)
- The above structure was/is working very well:
  - Regular task-oriented group meetings
  - Direct consultations and error reporting to the experts at the CEs
  - LCG Savannah, Global Grid User Support at FZK

Experiences – duration of PDC'04

- Many of the challenges we encountered would not have shown in a short DC:
  - Particularities of operating the GRID and CE machinery for extended periods of time
  - Keeping a backward compatibility of the software, which is constantly under development
  - Need for a stable and Grid-aware personnel, especially at the T2 type computing centres
  - Keeping the pledged amount of computing resources throughout the exercise at the CEs
    - Once committed, the local resources cannot be 'taken away'
  - Steady utilization of the available resources to their maximum capacity
    - Not always possible – breaks were needed to do software development and fixes (intrinsic property of a Data Challenge)

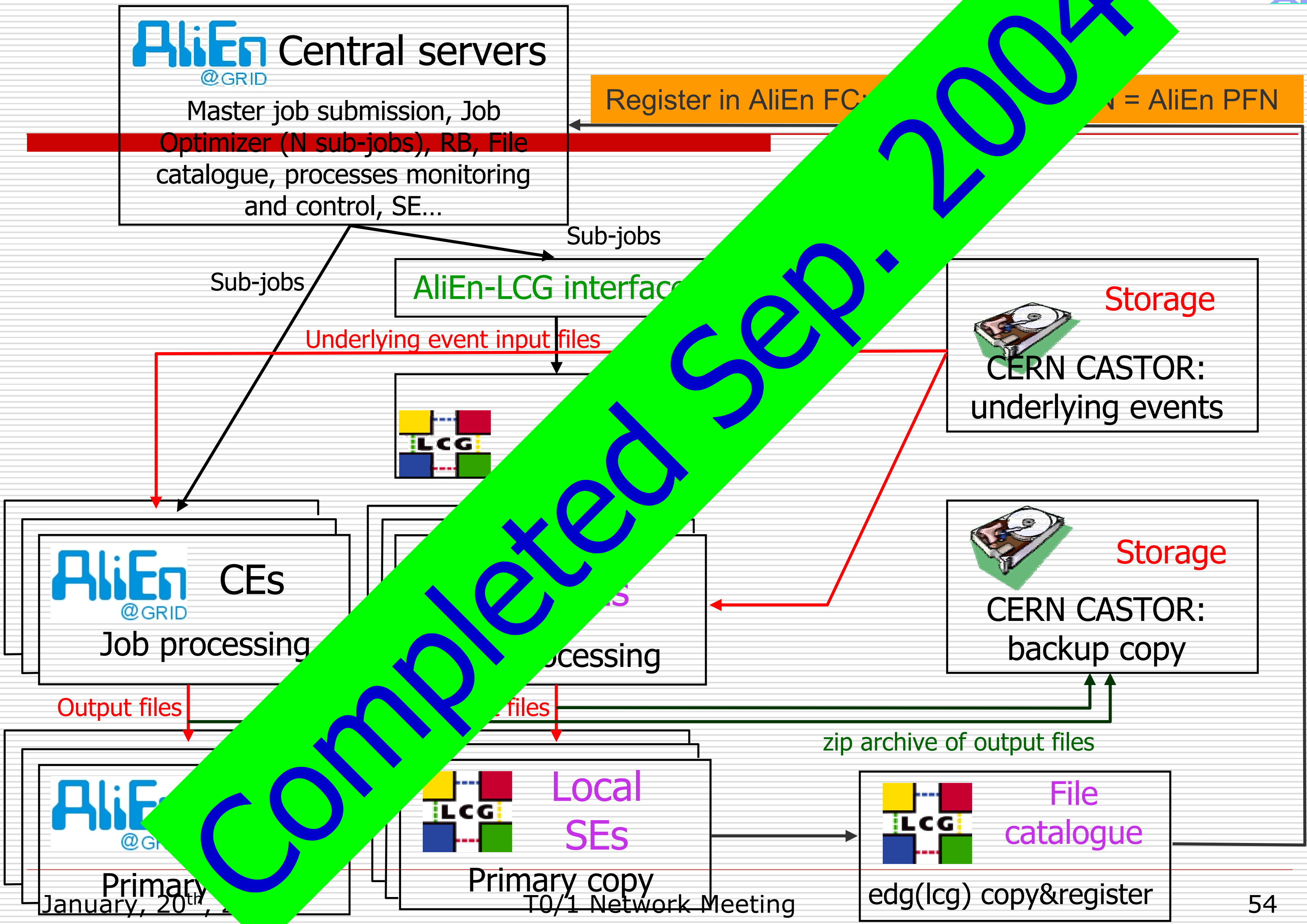
Experiences
– operation and computing resources

- Phase 1:
  - Slow ramp-up and steady progress afterwards
  - Hit the limitations of the CASTOR MSS stager (being reworked)
  - Limiting factor – number of CPUs available at the ALICE controlled computing centres and through LCG
- Phase 2:
  - Difficulty to achieve planned number of CPUs and uniform job distribution at the LCG sites:
    - Competition for resources with the other LHC data challenges – partially alleviated by introducing dedicated ALICE queues at the LCG sites and more instances of the LCG RB
  - Instability and frequent failures of the LCG SEs
- Phase 3 (anticipated):
  - Need for extensive user support for analysis on the GRID

Experiences - future

- As expected – the most challenging part is the multi-user operation during phase 3:
  - To execute it properly, we need the AliEn components in gLite, which have been tested by ARDA for ALICE
  - The lost momentum should be regained once we deploy the middleware – the computing resources are on stand-by
  - In the case we cannot deploy the new middleware within weeks – we have to scale down the planned Phase 3 scope and limit it to expert users

Phase 2 job structure



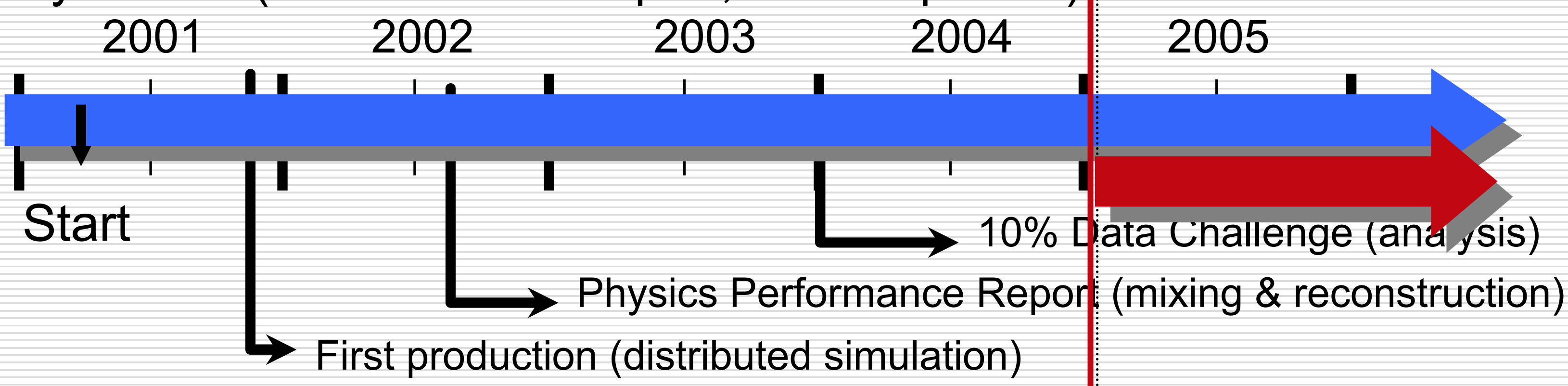
Summary on PDC'04

- Computing resources:
  - It took some effort to 'tune' the resources at the remote computing centres to meet the expectations and demands of the GRID software
  - By and large, the outside response to the exercise was very positive – more CPU and storage capacity was made available as the PDC progressed
- Middleware:
  - AliEn proved to be fully capable of routinely executing jobs with high complexity (Phase 1 and 2 like) and exercising control over large amounts of computing resources
  - Its functionality needed for Phase 3 has been demonstrated, but due to the 'frozen' status and support issues, cannot be released to the ALICE physics community
  - The LCG middleware proved adequate for Phase 1-type tasks, but below average for Phase 2-type tasks and in a competitive environment
  - It cannot provide the additional functionality needed for Phase 3-type jobs (f.e. reliable handling of hundreds of parallel analysis jobs, fair sharing of resources)

The ALICE Grid strategy

- There are millions lines of code in OS dealing with GRID issues
- Why not using them to build the minimal GRID that does the job?
  - Fast development (cycle) of a prototype
  - Quick (Immediate) adoption of emerging standards
- AliEn by ALICE (5% code developed, 95% imported)

gLite



Functionality + Simulation → Interoperability + Reconstruction → Performance, Scalability, Standards + Analysis

ALICE requirements on MiddleWare

- ALICE assumes that a MW with the same quality and functionality that AliEn would have had in two years from now will be deployable on the LCG computing infrastructure
- All users should work in a pervasive Grid environment
- This would be best achieved via a common project, and ALICE still hopes that the EGEE MW will provide this
- If this cannot be done via a common project, then it could still be achieved continuing the development of the AliEn-derived components of gLite
  - But then few key developers should support ALICE
- Should this turn out to be impossible (but why?), the Computing Model would have to be changed
  - More human [O(20) FTE/y] and hardware resources [O(+25%)] will be needed for the analysis of the ALICE data

Phase III – new middleware strategy

- Change of middleware - reasons:
  - The status of LCG DMS is not brilliant
  - Phase 3 functionality is existing and adequate in AliEn but…
  - ***All AliEn developers/maintainers working now in EGEE and ARDA***
- **Obvious choice is to do Phase 3 with the next generation of middleware – gLite with the AliEn components imported and improved**
- Advantages
  - Uniform configuration: gLite on EGEE/LCG-managed sites & on ALICE-managed sites
  - If we have to go that way, the sooner the better
- Disadvantages
  - It introduces a delay with respect to the original plan – ***proved to be considerably longer than anticipated***

Summary on PDC'04 (2)

- ALICE computing model validation:
  - AliRoot – all parts of the code successfully tested
  - AliEn – full functionality tests in Phases 1 and 2 and demonstrated for Phase 3
  - Computing elements configuration:
    - Need for a performing MSS shown
    - The Phase 2 distributed data storage schema proved very robust and fast
    - Network utilization – minimized by the configuration of the PDC, have not seen any latency problems (also the AliEn built-in protection helped)
  - Data analysis – the planned execution of this phase is contingent on the availability of the tested AliEn components in gLite

Related documents

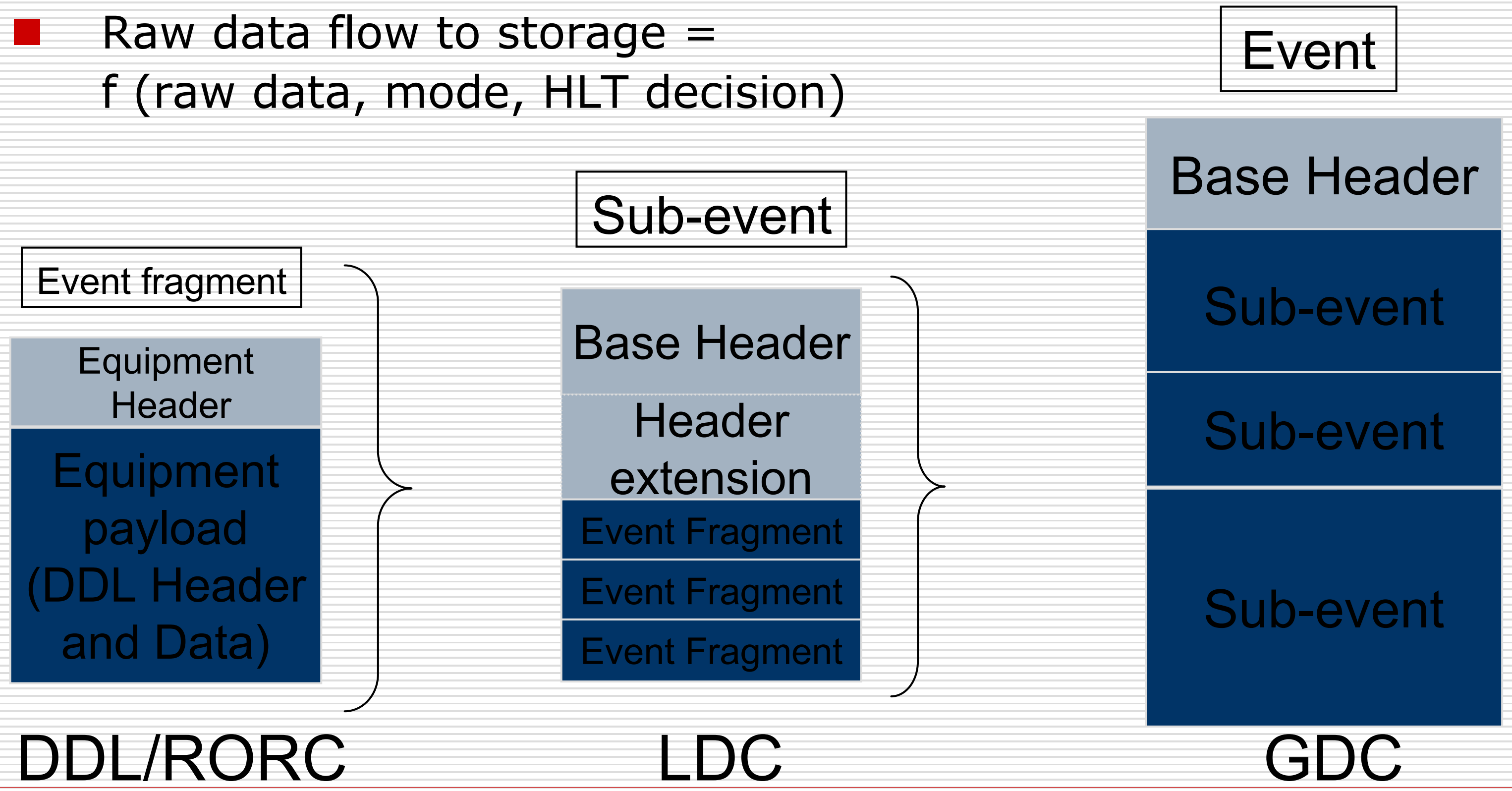
- Computing MOU
  - Distributed to the Collaboration for feedback on October 1, 2004
  - Provide the C-RRB with documents to be approved at its April 2005 meeting
  - Subsequently distributed for signature
- ALICE Computing TDR
  - Elements of the early draft given to LHCC on December 17, 2004
  - Draft will be presented during the ALICE/offline week in February 2005
  - Approval during the ALICE/offline week in June 2005

Metadata

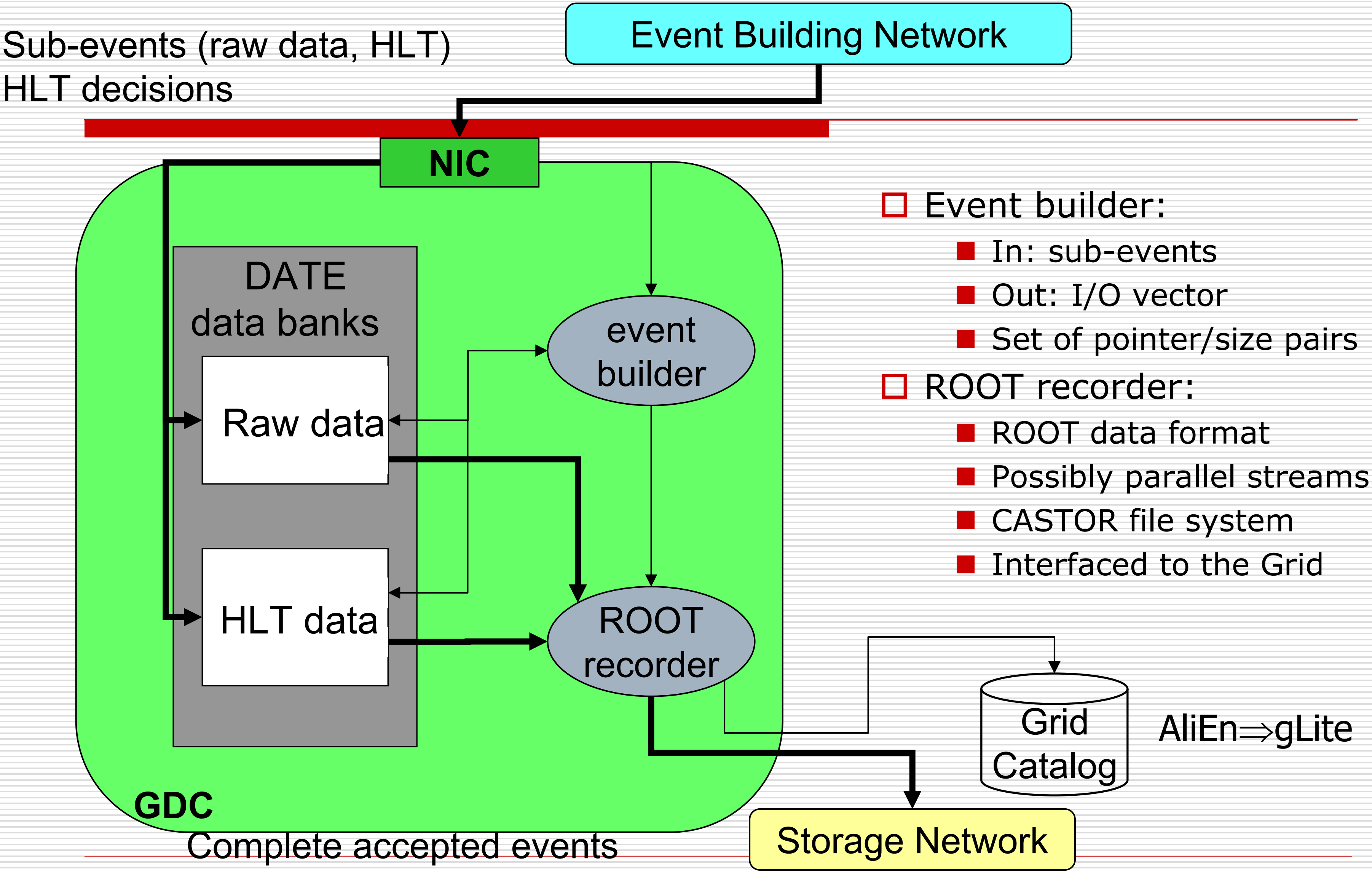
- MetaData are essential for the selection of events
- We hope to be able to use the Grid file catalogue for one part of the MetaData
  - During the Data Challenge we used the AliEn file catalogue for storing part of the MetaData
  - However these are file-level MetaData
- We will need an additional catalogue for event-level MetaData
  - This can be simply the TAG catalogue with externalisable references
- We will take a decision in 2005, hoping that the Grid scenario will be clearer

Online Framework: Data Format

- Physics data:
  - Raw data flow to DAQ/HLT = f (interaction, Triggers L0 L1 L2)
  - Raw data flow to storage =
    f (raw data, mode, HLT decision)

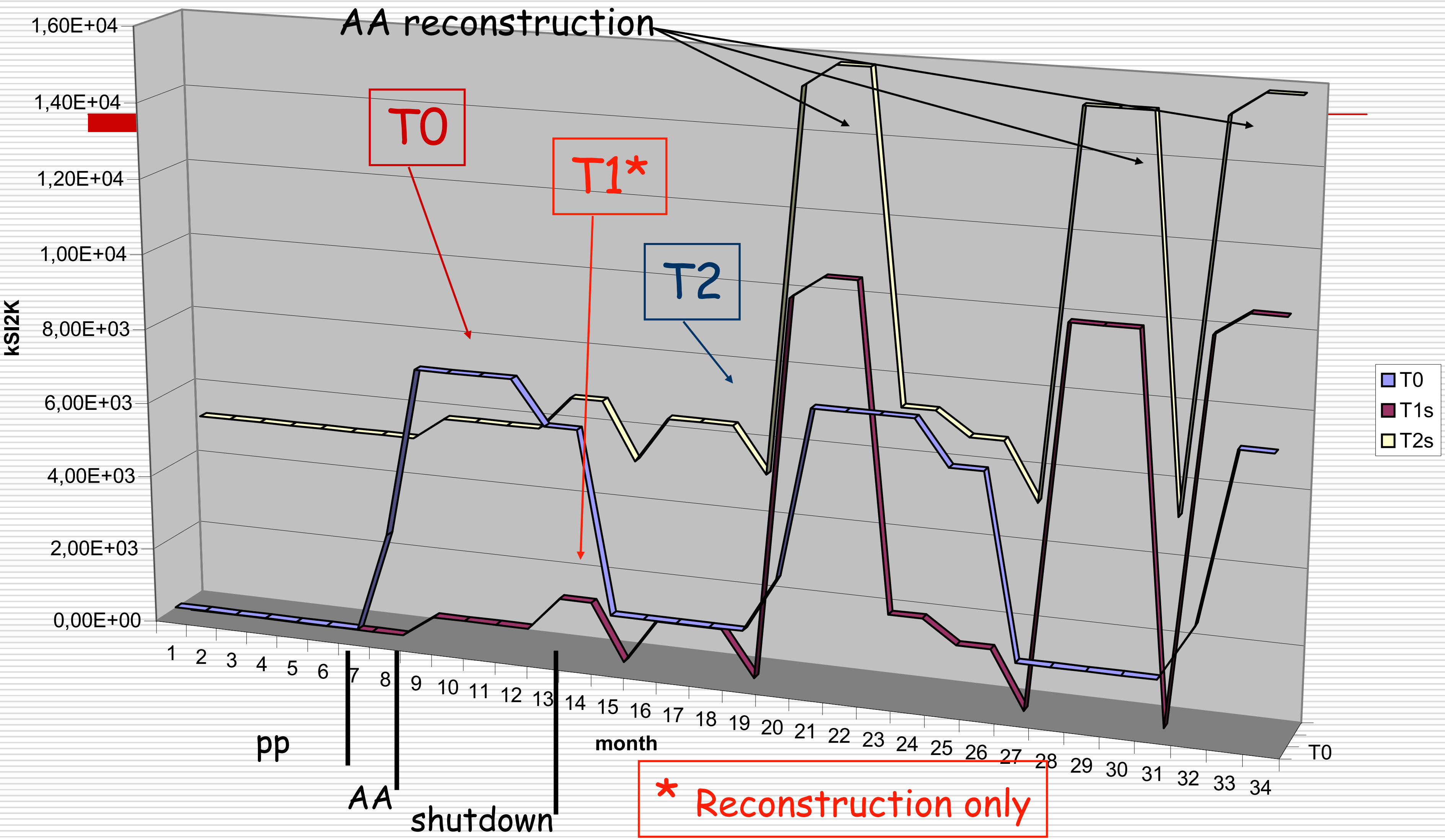


Event building and data recording in GDCs

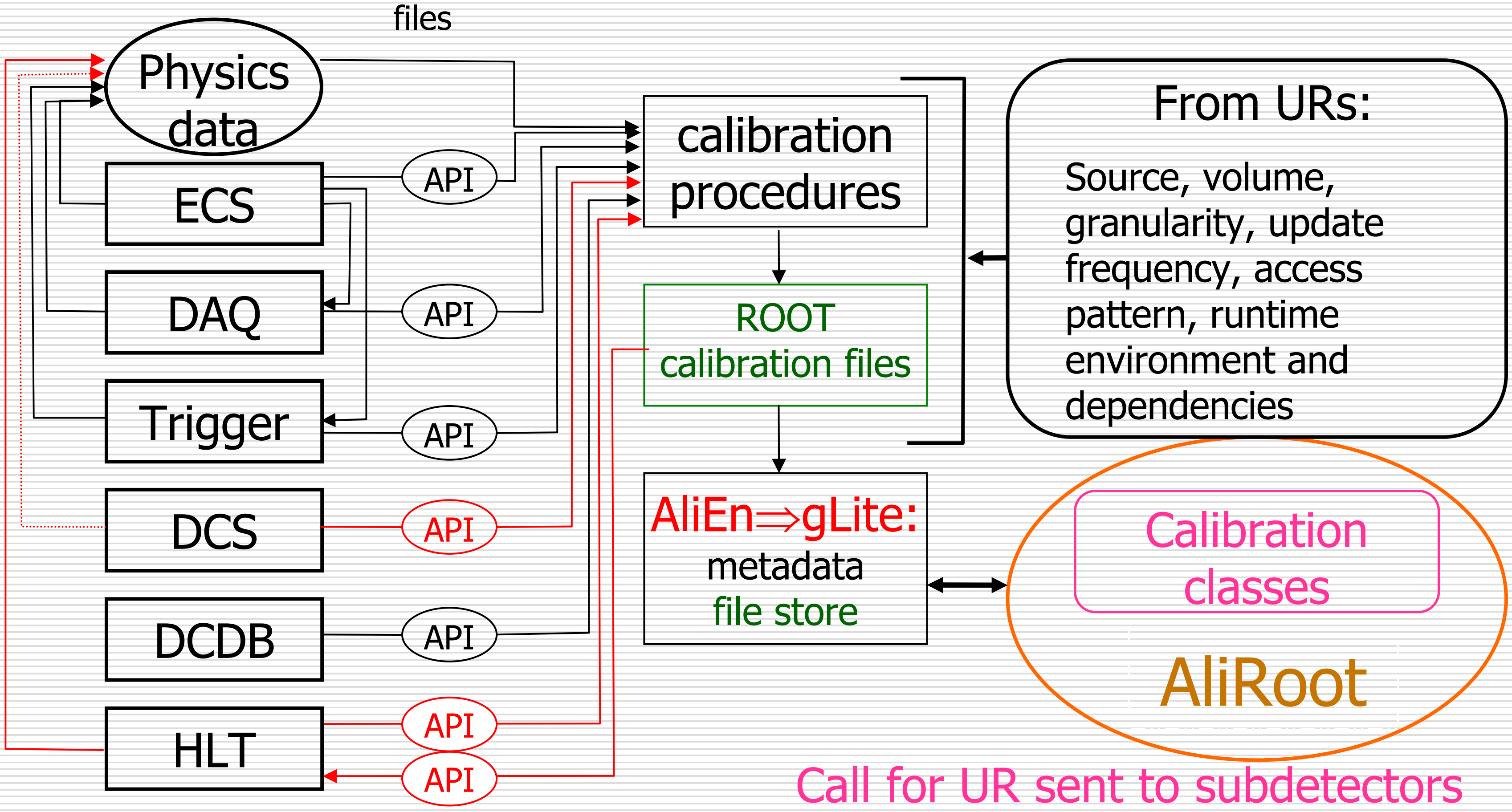


- Event builder:
  - In: sub-events
  - Out: I/O vector
  - Set of pointer/size pairs
- ROOT recorder:
  - ROOT data format
  - Possibly parallel streams
  - CASTOR file system
  - Interfaced to the Grid

Computing Resources profile



External relations and DB connectivity



Call for UR sent to subdetectors

API – Application Program Interface

The Offline Framework

- AliRoot in development since 1998
  - Entirely based on ROOT
  - Used for the detector TDR's and the PPR
- Two packages to install (ROOT and AliRoot)
  - Plus transport MC's
- Ported on several architectures (Linux IA32, IA64 and AMD, Mac OS X, Digital True64, SunOS…)
- Distributed development
  - Over 50 developers and a single cvs repository
- Tight integration with DAQ (data recorder) and HLT (same code-base)

Development of Analysis

- Analysis Object Data designed for efficiency
  - Contain only data needed for a particular analysis
- Analysis *à la PAW*
  - ROOT + at most a small library

- Batch analysis infrastructure
  - Prototype published at the end of 2004 based on AliEn
- Interactive analysis infrastructure
  - Demonstration performed at the end 2004 with AliEn⇒gLite

- Waiting now for the deployment of gLite MW to analyse the data of PDC04
- Physics working groups are just starting now, so timing is right to receive requirements and feedback

Production history

- ALICE repository – history of the entire DC
- ~ 1 000 monitored parameters:
  - Running, completed processes
  - Job status and error conditions
  - Network traffic
  - Site status, central services monitoring
- 7 GB data
- 24 million records with 1 minute granularity – these are being analysed with the goal of improving the GRID performance