united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

ICTP 40th Anniversary

H4.SMR/1574-23

"VII School on Non-Accelerator Astroparticle Physics"

26 July - 6 August 2004

Large Scale Computing

Paolo Capiluppi

Dept. of Physics of the University
and INFN Section
Bologna, italy

Large Scale Computing

Paolo Capiluppi

Dept. of Physics of the University
and INFN Section

Bologna - Italy

Outline

- **1st lesson**
  - The problem of Large scale Computing
  - Applications that need Large scale Computing: an example (LHC)
  - Complexities and data Management
  - Possible solution(s): Distributed Computing and Data Access
    - Is a viable solution?
  - Grid Computing, a component of a possible solution
- **2nd lesson**
  - How and why building a Computing Model
  - Measurements of the "Model": Data Challenges (LHC)
  - Where we are: results of LHC Experiments Data (and Physics) Challenges
  - What is still missing? And how much time is left for a "solution"?
  - Conclusions

“Large” Scale

◆Large because of:

- Data amount (> several PetaBytes*)
- Data distribution (> 100 sites)
- Computing power needed (> tens of MSI2000⁺)
- Number of users (> 5000)
- Complexity of algorithms (>~ 500 k lines of code)
- Chaotic access (~ thousands of independent access per user)
- Coordination of resources infrastructure (> hundreds of Million Euro)
- Heterogeneity of resources (~tens/hundreds of different systems

*One PetaByte = 1000 TBytes >~ 500.000 movies DVDs

⁺One “modern” PC’s CPU ~ 1000 SI2000

Do we have "large scale" Applications?

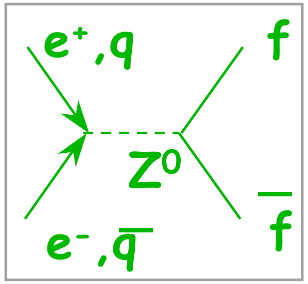
- Yes, we have (or we will shortly have)

- Experiments with particle accelerators (LHC) for example →

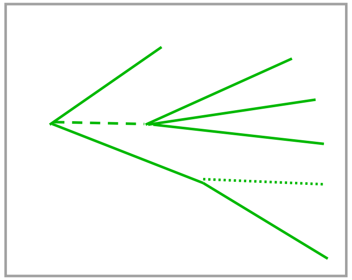
- But also non-LHC Experiments (or other social activities)
  - Earth observation (satellite)
  - Cosmic rays HEP experiments
  - Astroparticle experiments (also with earth orbit apparata)

  - World stock market
  - Prime elements availability
  - Whether forecasts
  - WEB mining
  - Etc

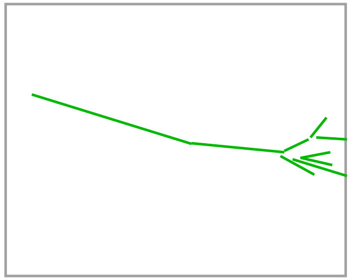
From Physics to Raw Data



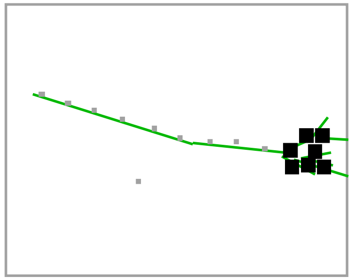
Basic physics



Fragmentation,
Decay



Interaction with detector material
Multiple scattering, interactions



Detector response
Noise, pile-up, cross-talk, inefficiency, ambiguity, resolution, response function, alignment, temperature

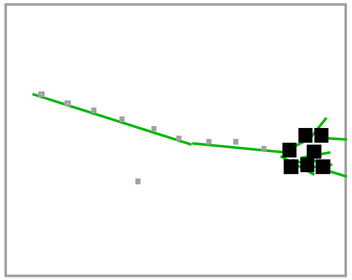
2037 2446 1733 1699
4003 3611 952 1328
2132 1870 2093 3271
4732 1102 2491 3216
2421 1211 2319 2133
3451 1942 1121 3429
3742 1288 2343 7142

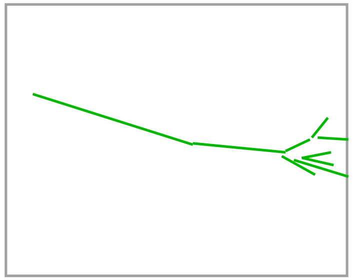
Raw data
(Bytes)

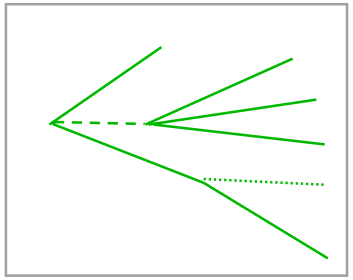
Read-out addresses, ADC, TDC values, Bit patterns

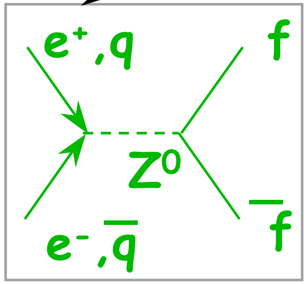
From Raw Data to Physics

2037 2446 1733 1699
4003 3611 952 1328
2132 1870 2093 3271
4732 1102 2491 3216
2421 1211 2319 2133
3451 1942 1121 3429
3742 1288 2343 7142









Raw data

Convert to physics quantities

Detector response

apply calibration, alignment,

Interaction with detector material

Pattern, recognition, Particle identification

Fragmentation, Decay

Physics analysis

Basic physics

Results

Analysis

Reconstruction

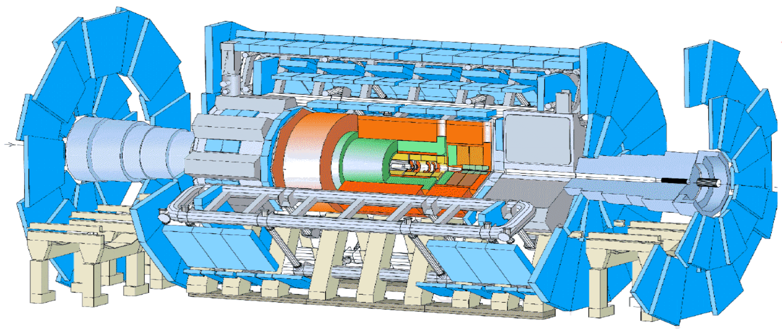
Simulation (Monte-Carlo)

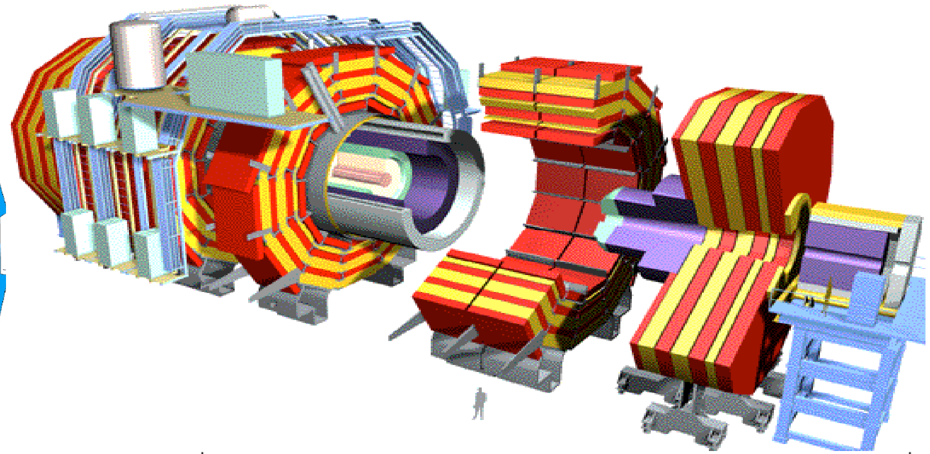
VIRGO

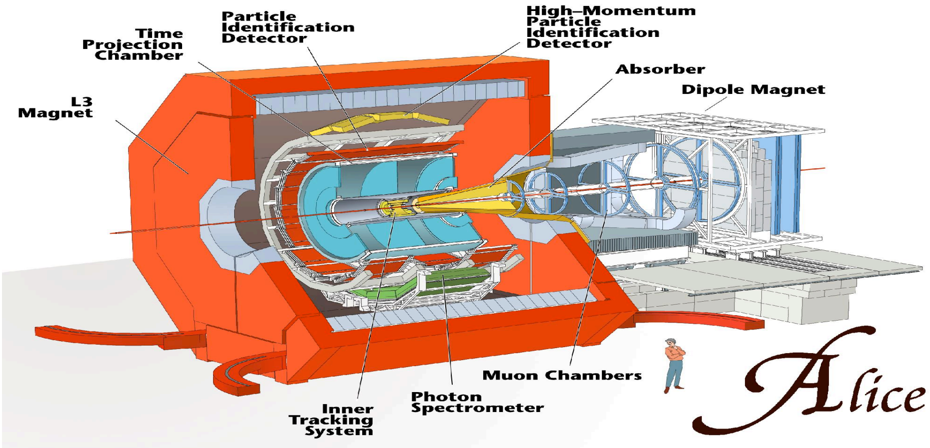
LHC Experiments

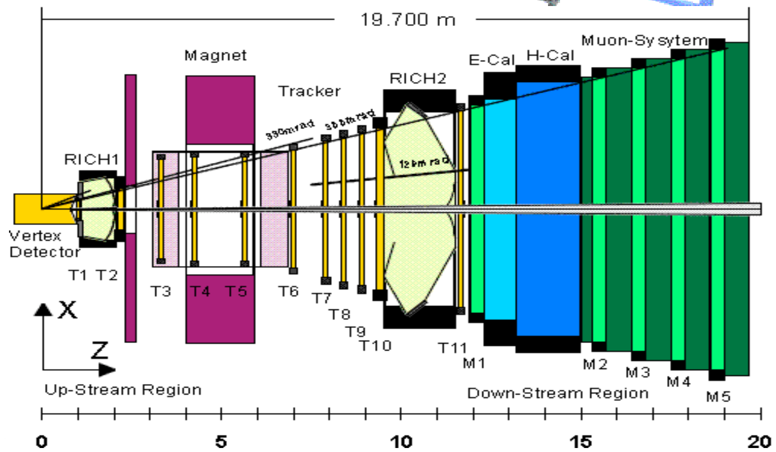
ATLAS, CMS, ALICE, LHCb

Higgs and New particles; Quark-Gluon Plasma; CP Violation

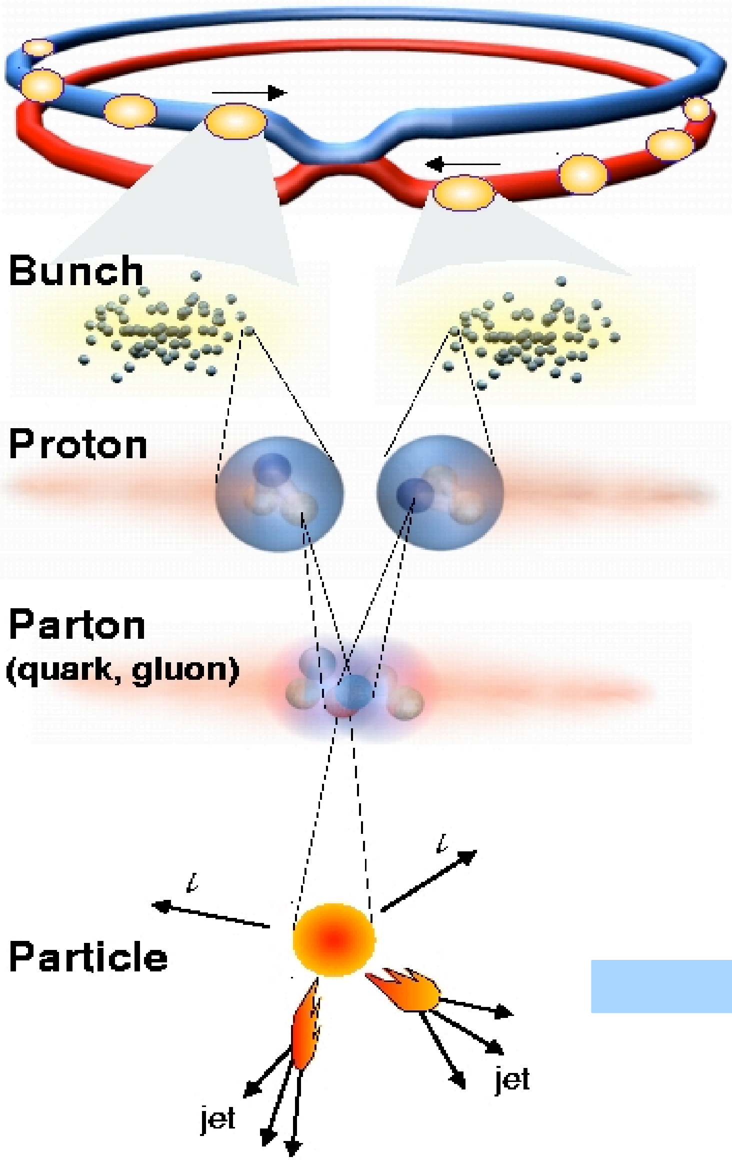








Large Hadron Collider LHC



Proton - Proton Collision

Beam energy : 7 TeV

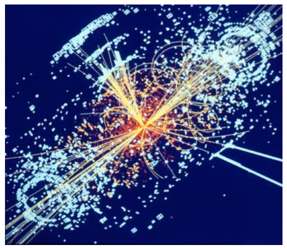
Luminosity : 10^{34} cm^{-2} s^{-1}

Data taking : > 2007

bunch-crossing rate: 40 MHz

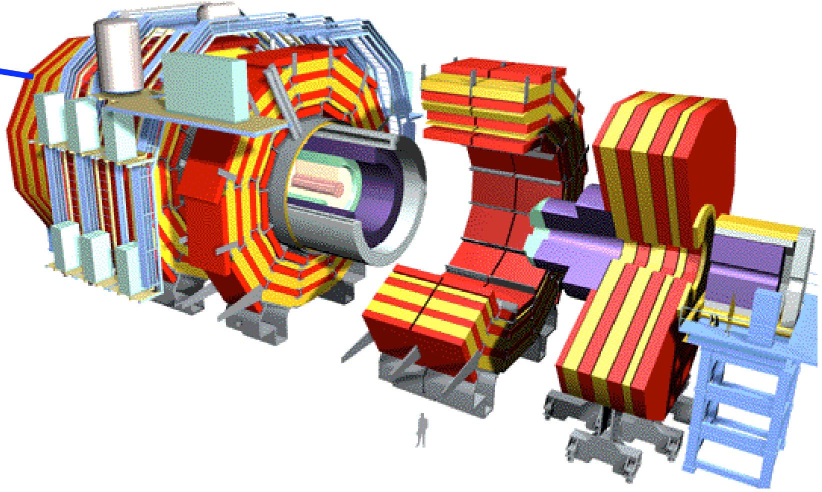
~20 p-p collisions for each bunch-crossing

p-p collisions ≈ 10^9 evt/s (Hz)

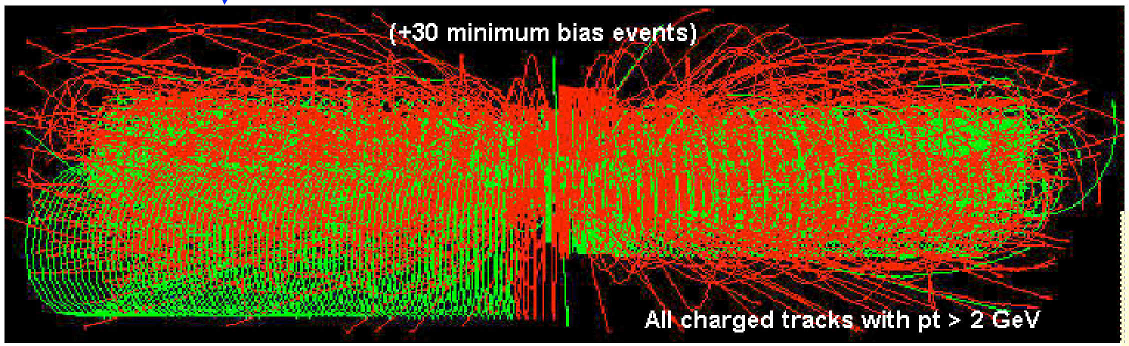


LHC DATA

The LHC Accelerator



This is reduced by online computers that filter out a few hundred "good" events per sec.

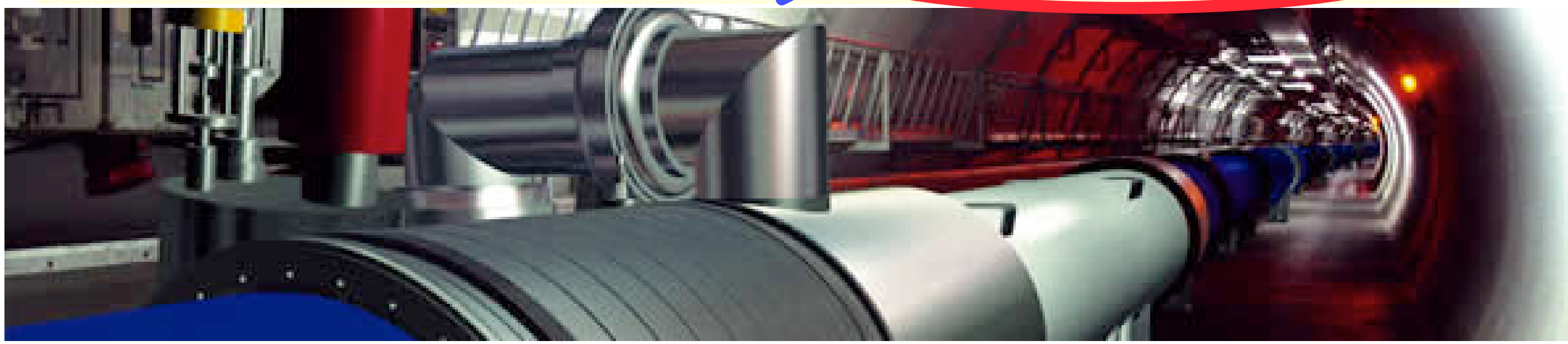


The accelerator generates 40 million particle collisions (events) every second at the centre of each of the four experiments' detectors

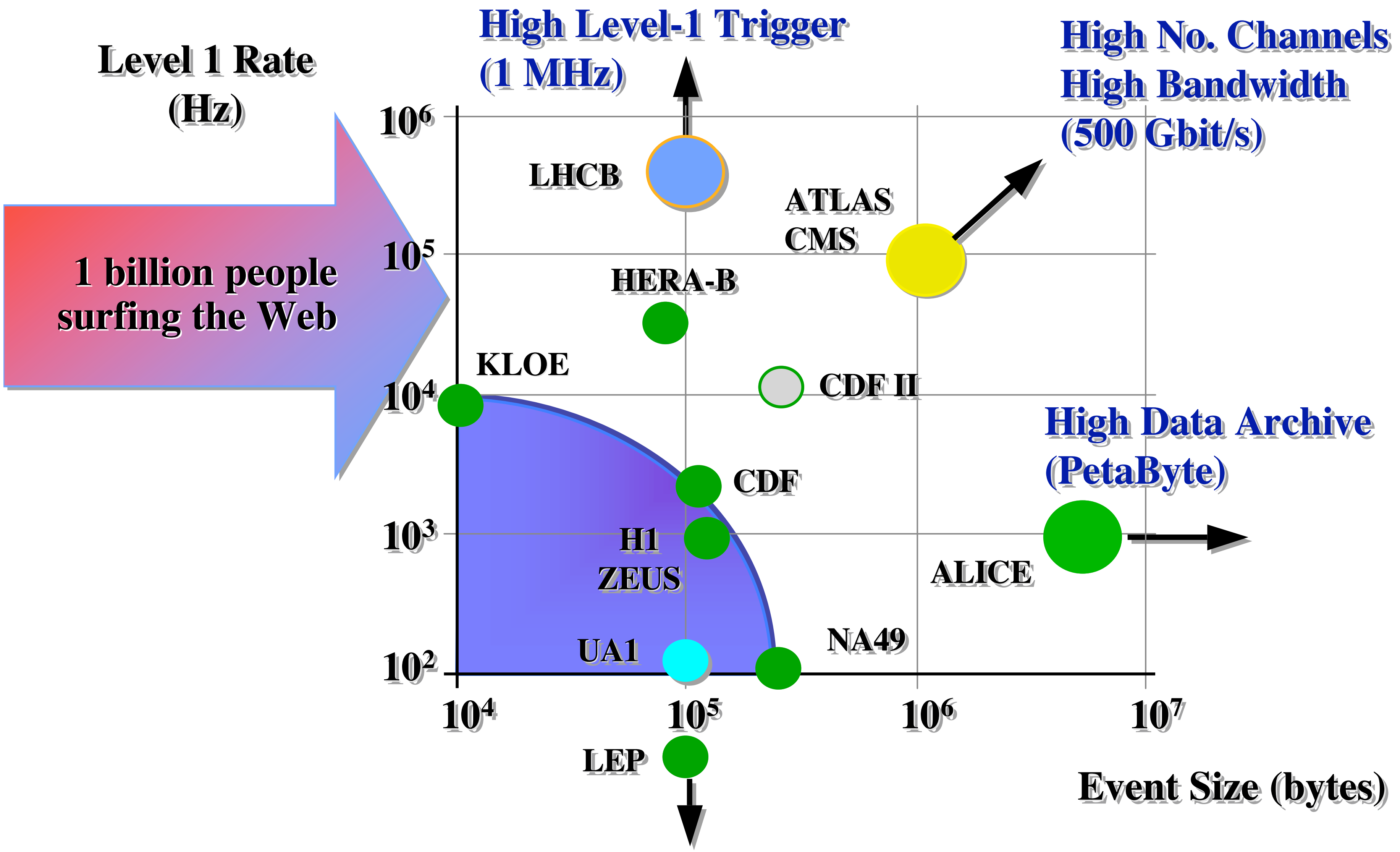
The LHC accelerator –

- the largest superconducting installation in the world
- 27 kilometres of magnets cooled to – 300º C
- colliding proton beams at an energy of 14 TeV

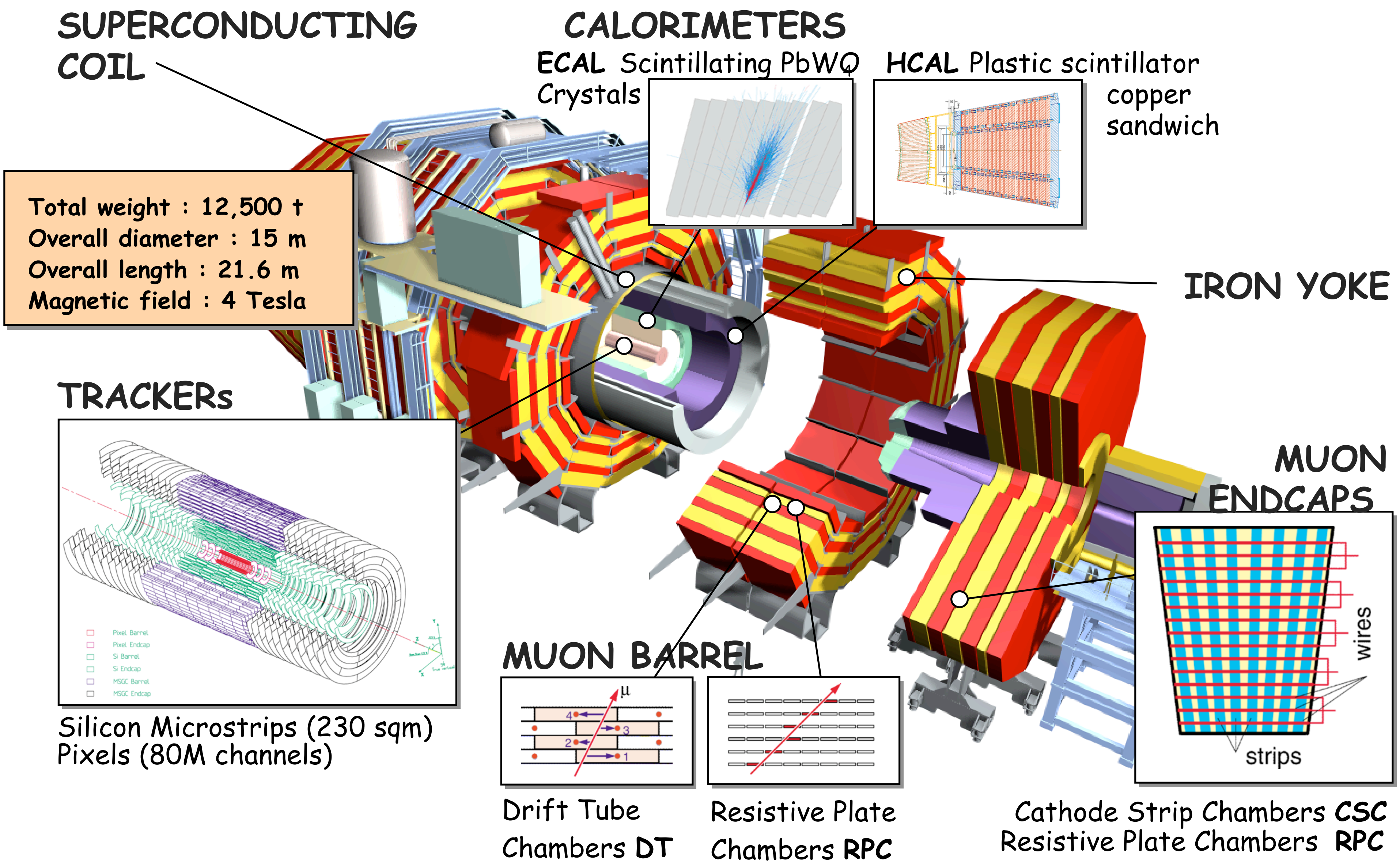
Which are recorded on disk and magnetic tape at 100-1,000 MegaBytes/sec → ~15 PetaBytes per year



How Much Data is Involved?



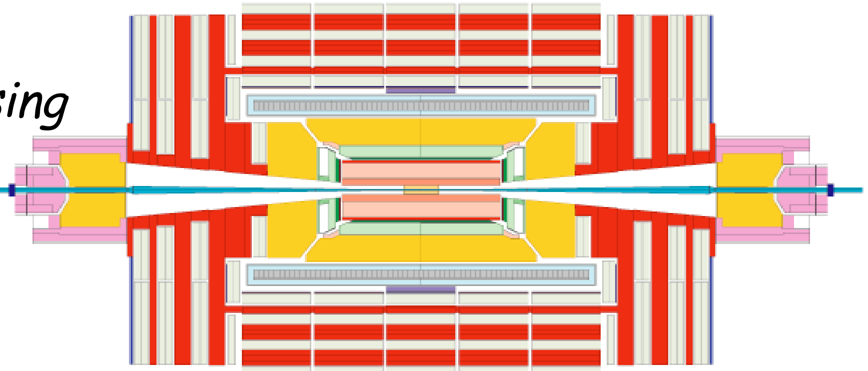
The Compact Muon Solenoid (CMS)



CMS Data Acquisition

1event is ~ 1MB in size

Bunch crossing 40 MHz



~ *GHz* *(~ PB/sec)*

Level 1 Trigger - special hardware

75 KHz *(75 GB/sec)*

High Level Trigger - PCs

100 Hz

(100 MB/sec)

data recording

data

Online system

multi-level trigger to:

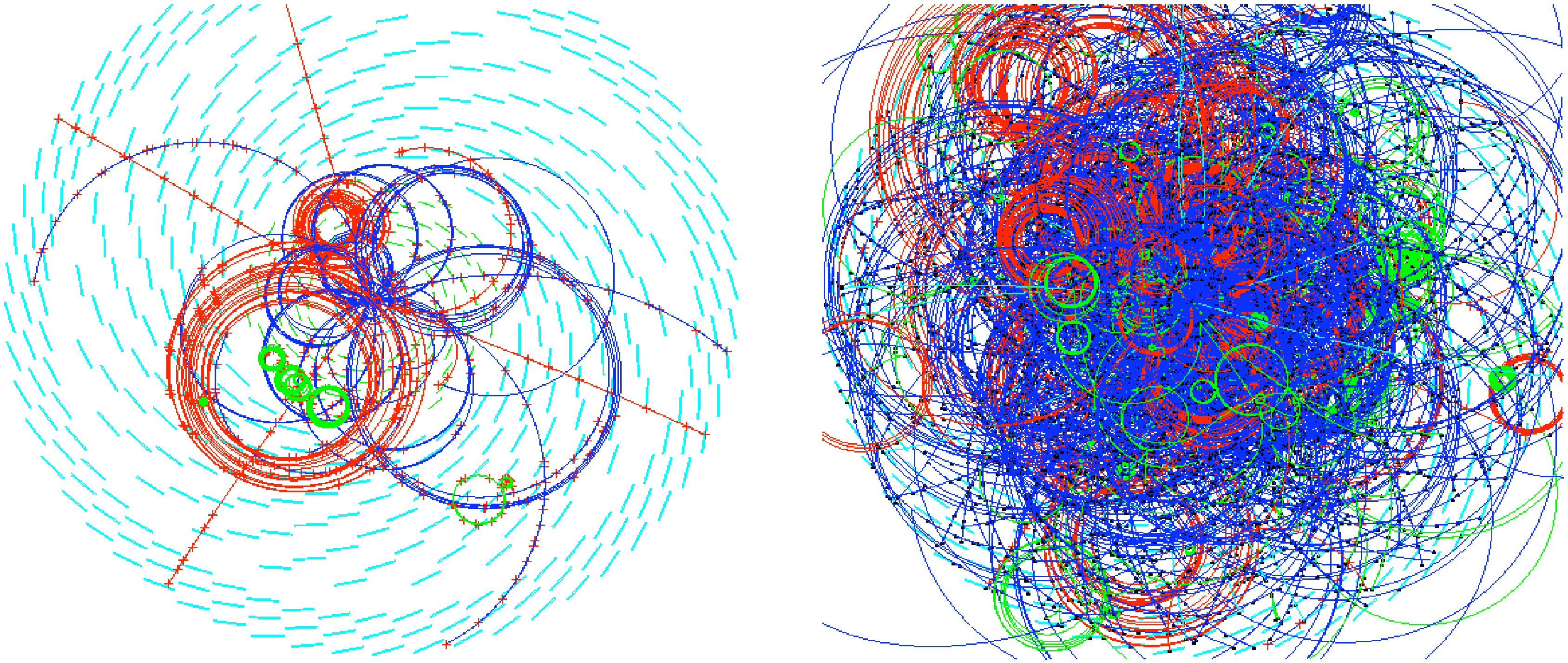
- filter out not interesting events
- reduce data volume

Offline analysis

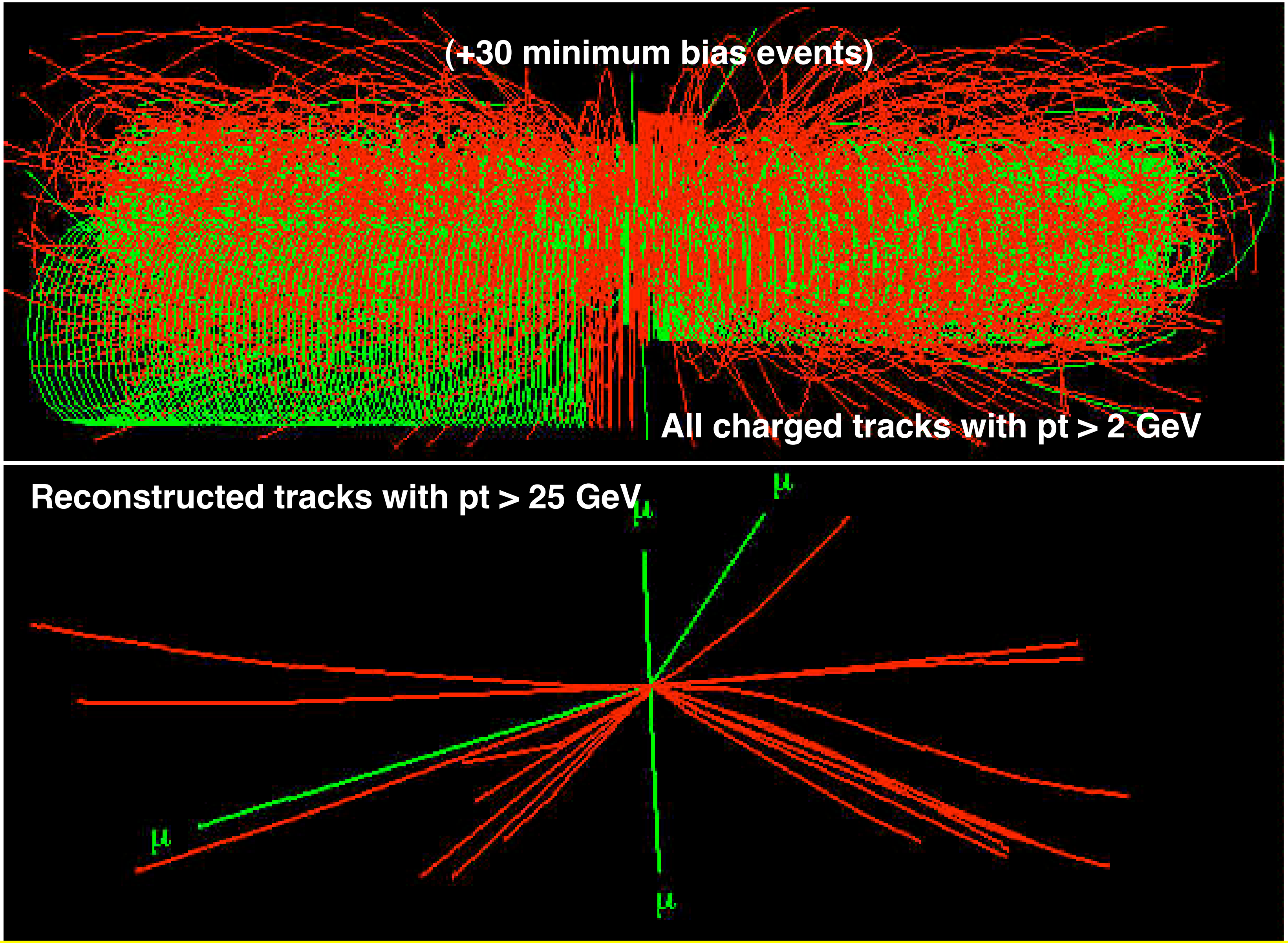
LHC Challenges: Complexity

Events:

- Bunch crossing time of 25 ns is so short that (parts of) events from different crossings overlap
- Signal event is obscured by 20 overlapping uninteresting collisions in same crossing
- Track reconstruction time at 10^{34} Luminosity several times 10^{33}



Higgs decay into 4 muons (tracker only)



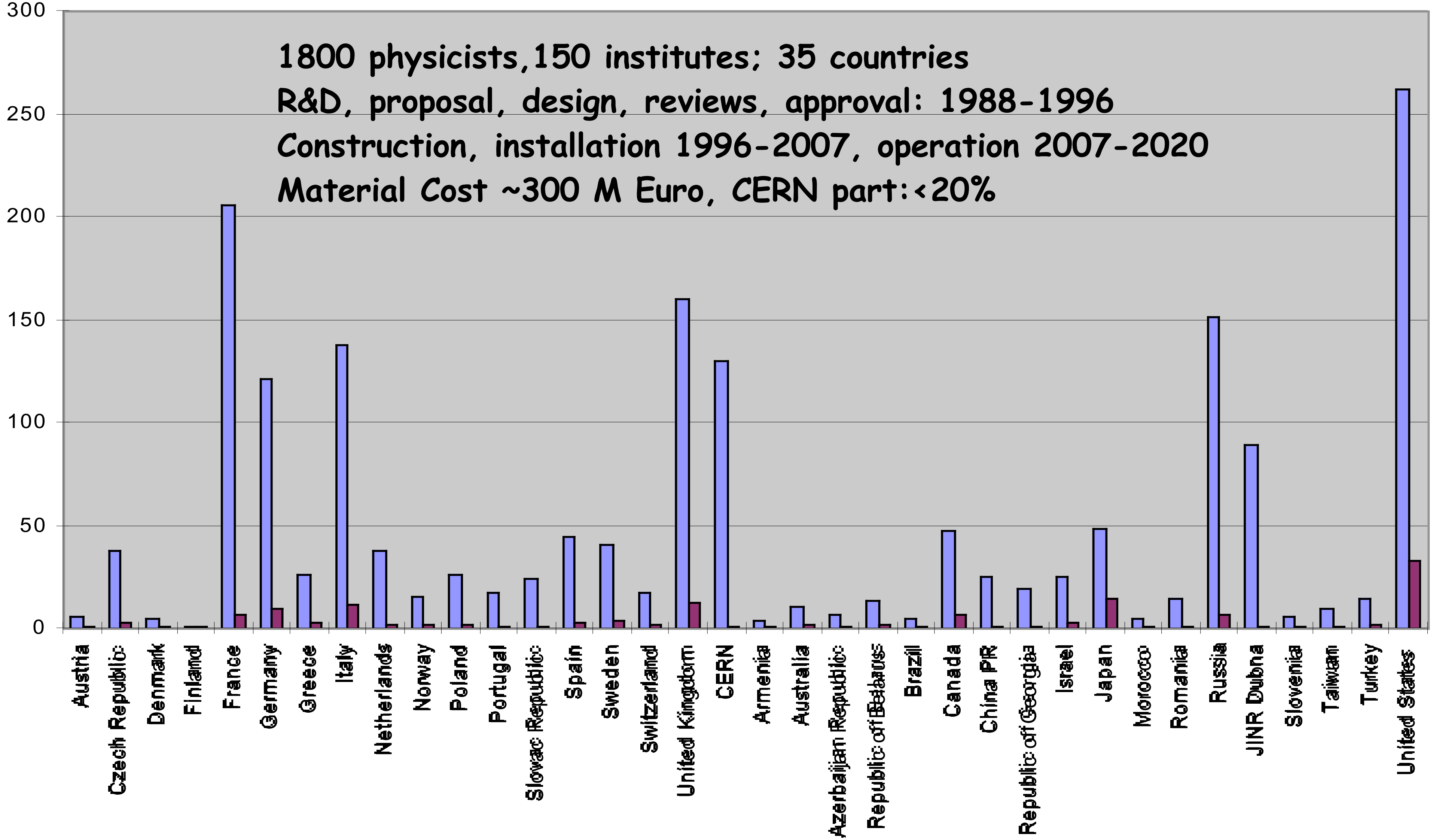
10^9 events/sec, selectivity: 1 in 10^{13} (1 person in a thousand world populations)

Dispersion of Actors

- Users just sit at their own home (Institution)
  - And cannot "go" near to the data (too much Money!)
- Moreover, "Funding Agencies" want (prefer) to invest at their own Country
- Therefore Actors are distributed worldwide

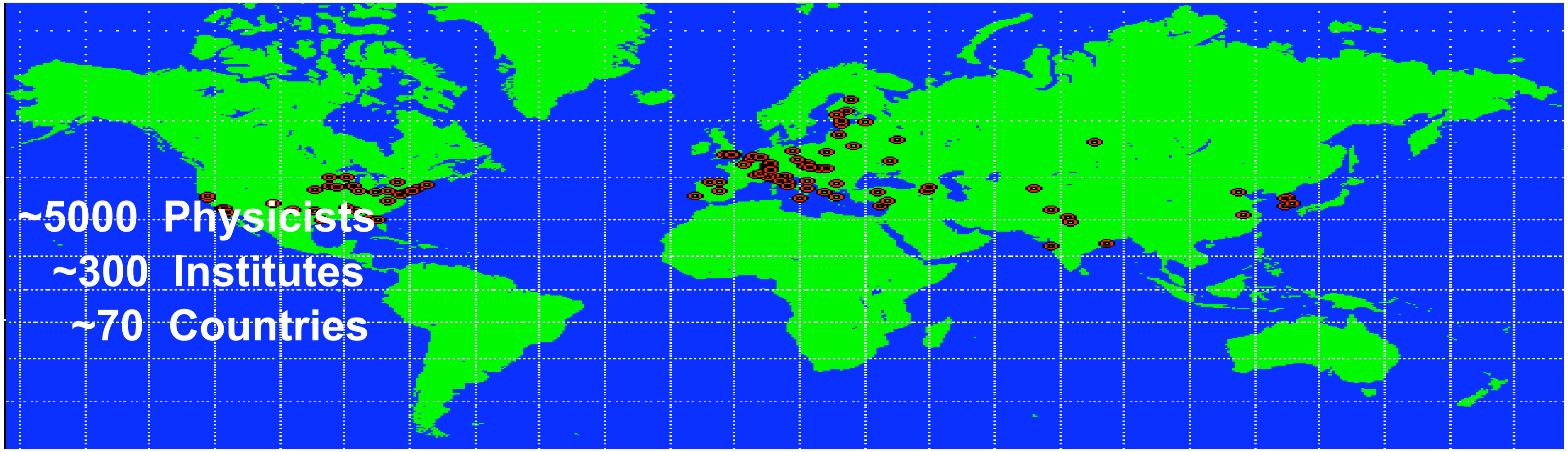
- The Solution, if possible, has to cope with location of users, … and Resources!
- And has to foster local ability to gain access to local resources (Computing, humans, organization, infrastructure,…)

ATLAS Collaboration



LHC Computing: *Different* from Previous Experiment Generations

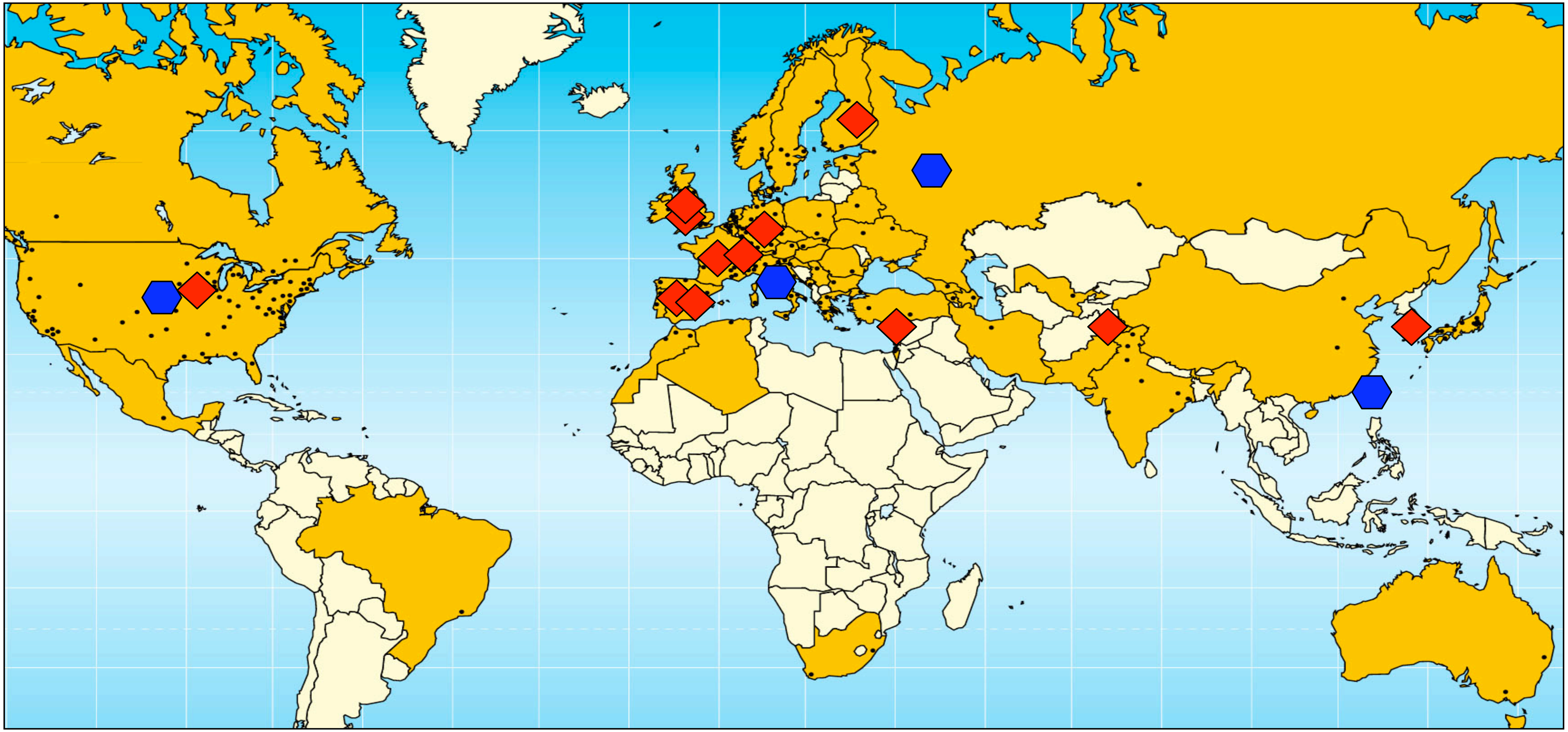
- Geographical dispersion: of people and resources
- Complexity: the detector and the LHC environment
- Scale: Petabytes per year of data
- **Technology: Software (Object Oriented) & Hardware (Commodity)**



Major challenges associated with:

- **Coordinated Use of Distributed Computing Resources**
- **Remote software development and physics analysis**
- **Communication and collaboration at a distance**

CMS World-wide Distributed Productions



CMS Production Regional Centre

CMS Distributed Production Regional Centre

Data Management Complexity

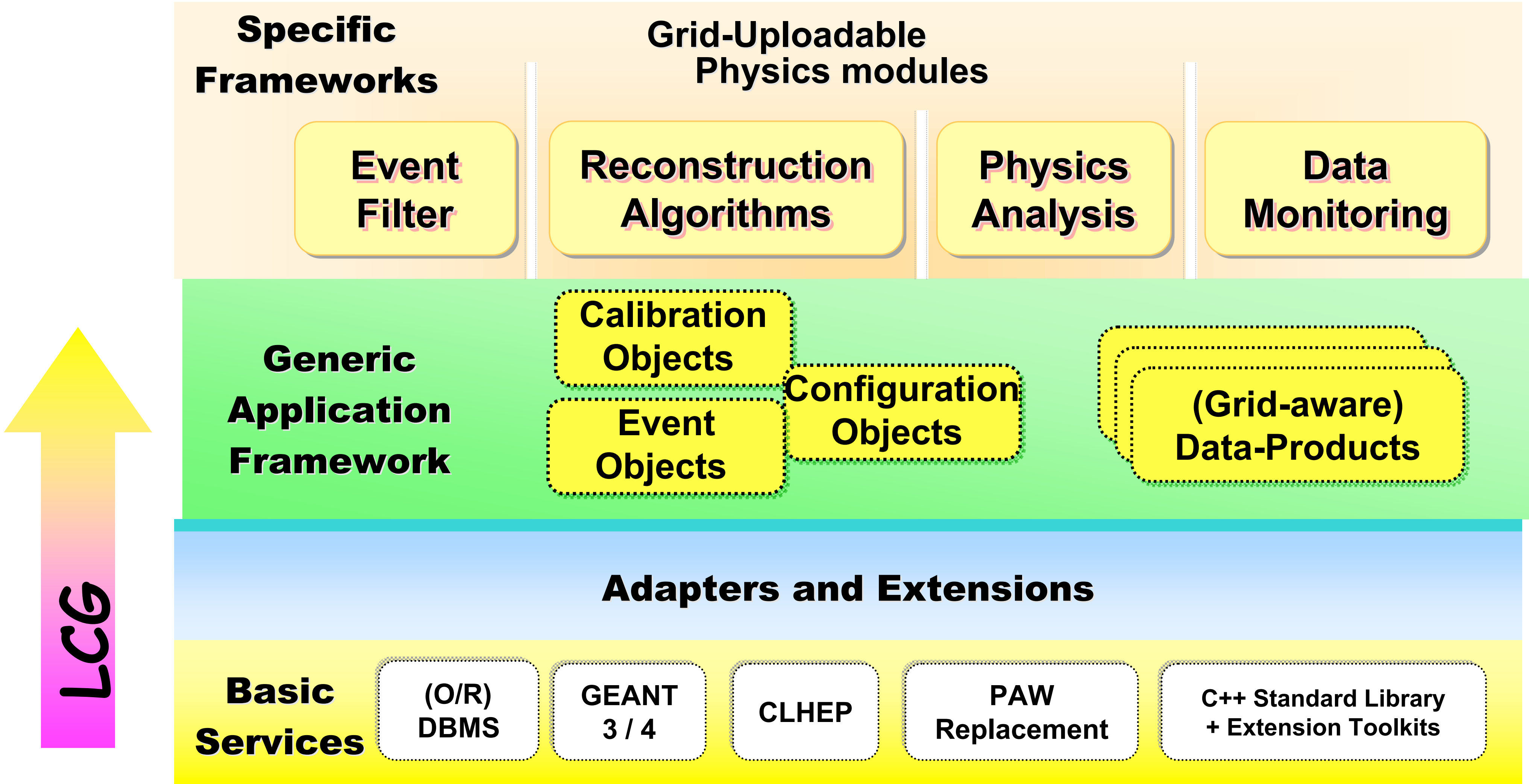
- Information are dispersed
- Bit of information of interest is hidden
- Objects are the natural (?) piece of information
- Access to objects is a possible solution
- But replication at different sites has to guarantee consistency (at the bit-wise level)
- Access to the "same" information in different sites must be "transparent"
  - Catalogs and Data bases issue: both relational and Object oriented
- Everyone has to be guaranteed of the same data access quality (not performance …)

Data access decomposition

- **Vincenzo?**
  - (Vincenzo Innocente is the software Architect of CMS!, not the only one!)

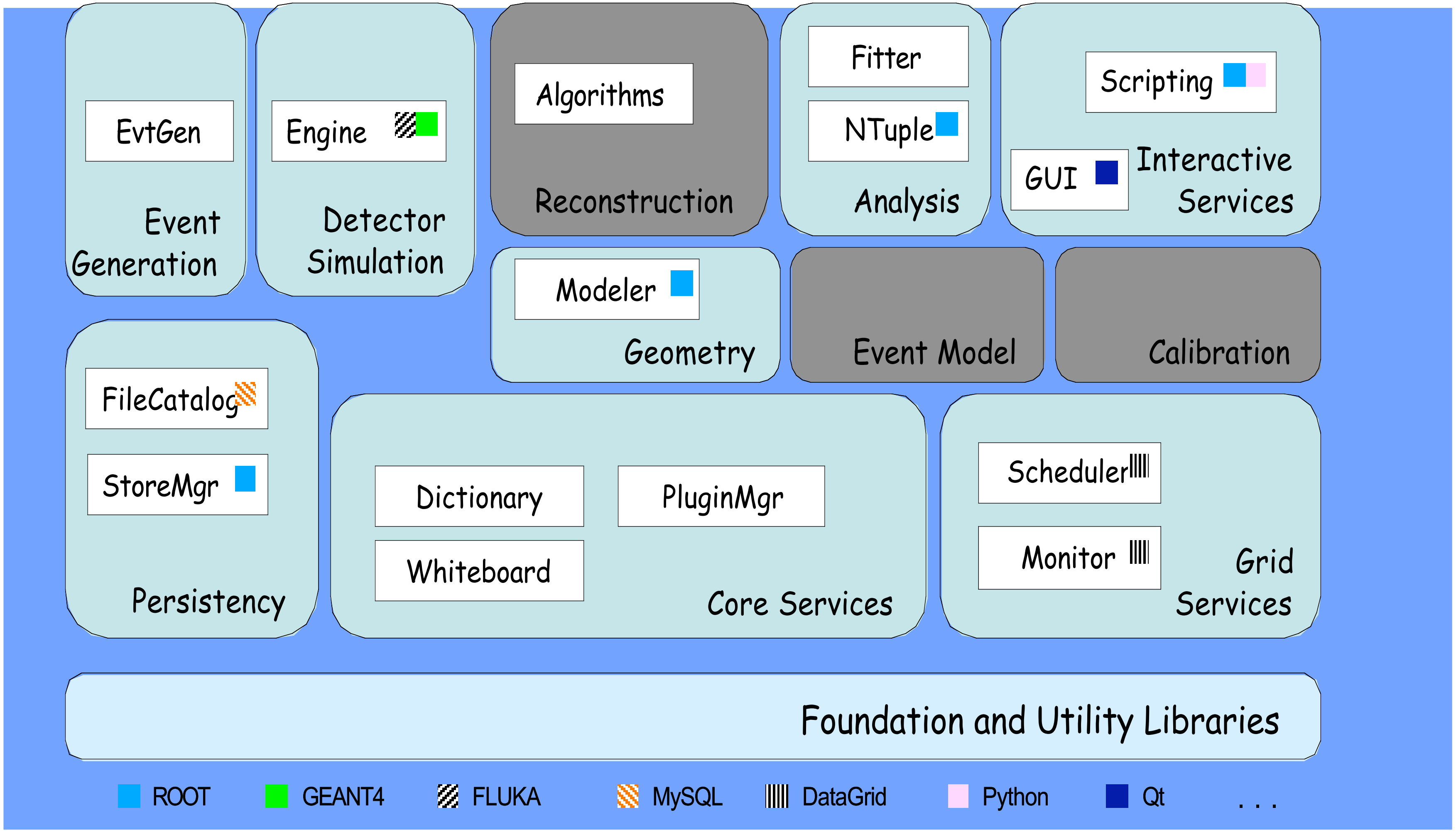
- **Well, I could not find the relevant slide…**
  - Apologizes
  - I'll try to say it in words …
    - And some related slides

Component Architecture Framework Layering



LCG Blueprint Software Decomposition

❖Building a Common Core Software Environment for LHC Experiments



The LCG Persistency Framework

- POOL is the LCG Persistency Framework
  - Pool of persistent objects for LHC
- Started in April '02
  - Common effort in which the experiments take a major share of the responsibility
    - for defining the system architecture
    - for development of POOL components
- The LCG Pool project provides a hybrid store integrating object streaming (eg Root I/O) with RDBMS technology (eg MySQL/Oracle) for consistent meta data handling
  - Strong emphasis on component decoupling and well defined communication/dependencies
  - Transparent cross-file and cross-technology object navigation via C++ smart pointers
  - Integration with Grid technology (via EDG-RLS)
    - but preserving networked and grid-decoupled working model

CMS software and LCG/POOL

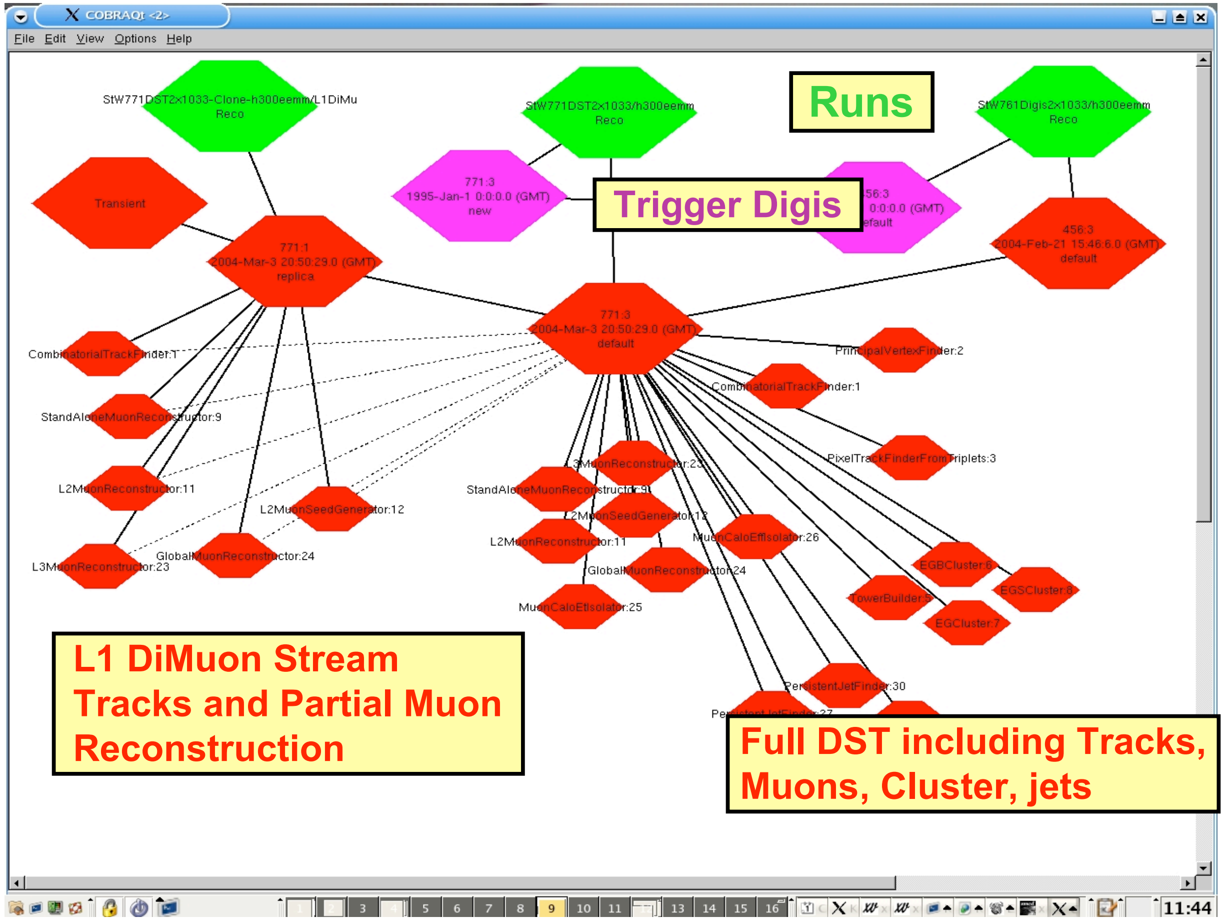
Reconstruction and analysis: using ORCA

DST have links to raw data but may be processed without raw data

Event streams operational

Persistency through POOL

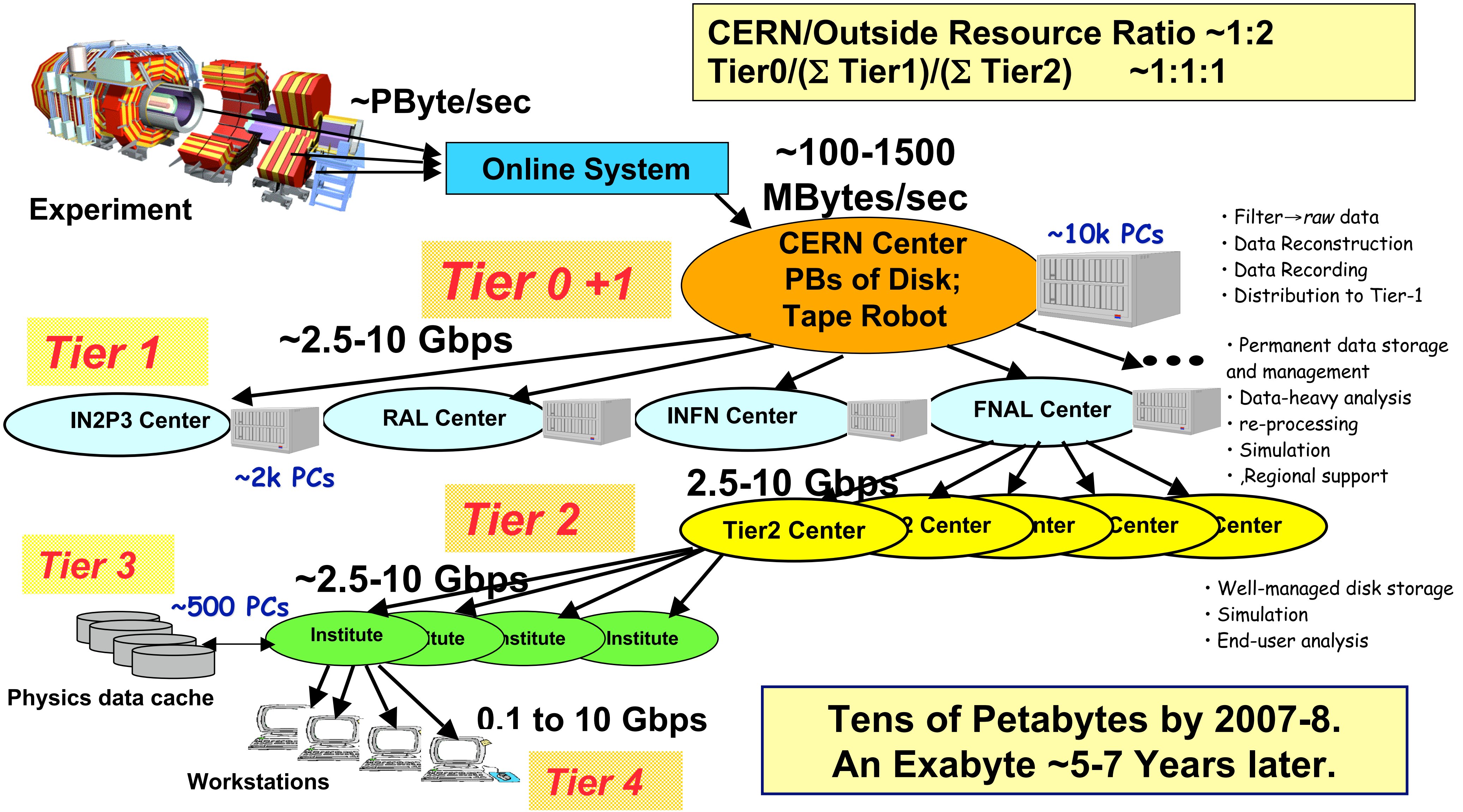
- All jobs use local XML catalogues
- Updates to central RLS catalogue only done for successful jobs



Well Known Solution(s)?

- Distributed Computing is a CS well known problem since time
- Distributed data access is something quite new also for CS
  - Web servers or "Web services"(?)
  - HEP is facing the data distribution problem since time, and had some partial solutions, but not the solution …until now?
- The paradigm of using Computers and Networks in a coherent design is a challenge for HEP (and CS)
- What's help here:
  - Data mining of info in different DBs is growing in knowledge
  - HEP applications are a bit simpler than other applications
    - So can drive the development
    - Executables (jobs) are quite similar even if highly variable in time and scope
    - Atomicity of HEP jobs is "one event" (also for many other Sciences), which facilitate the "decomposition" of the problem

LHC Data Grid Hierarchy



Emerging Vision: A Richly Structured, Global Dynamic System

The LHC Computing Grid Project LCG (2001 →)

Collaboration

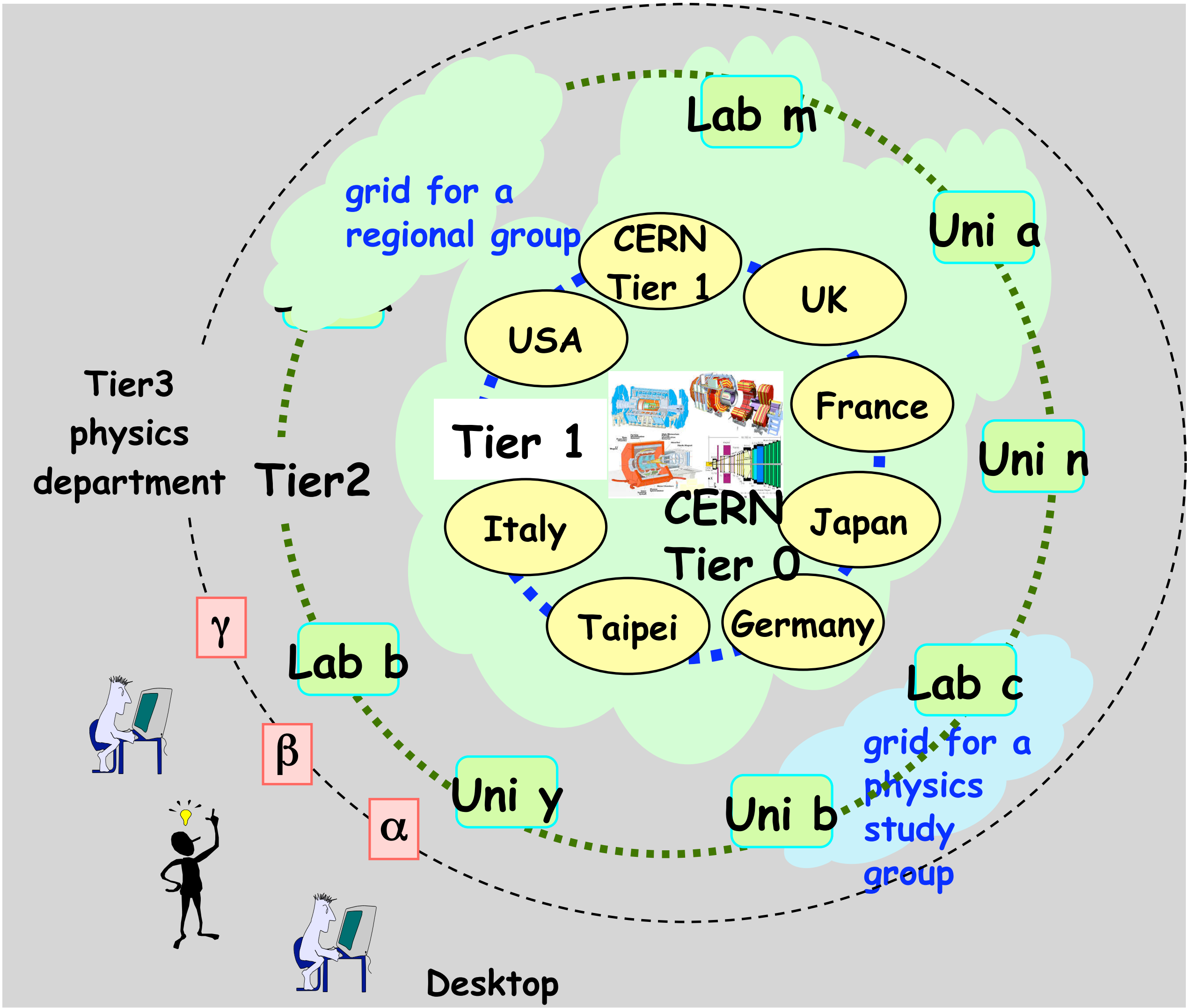
- LHC Experiments
- Grid projects: Europe, US
- Regional & national centres

Choices

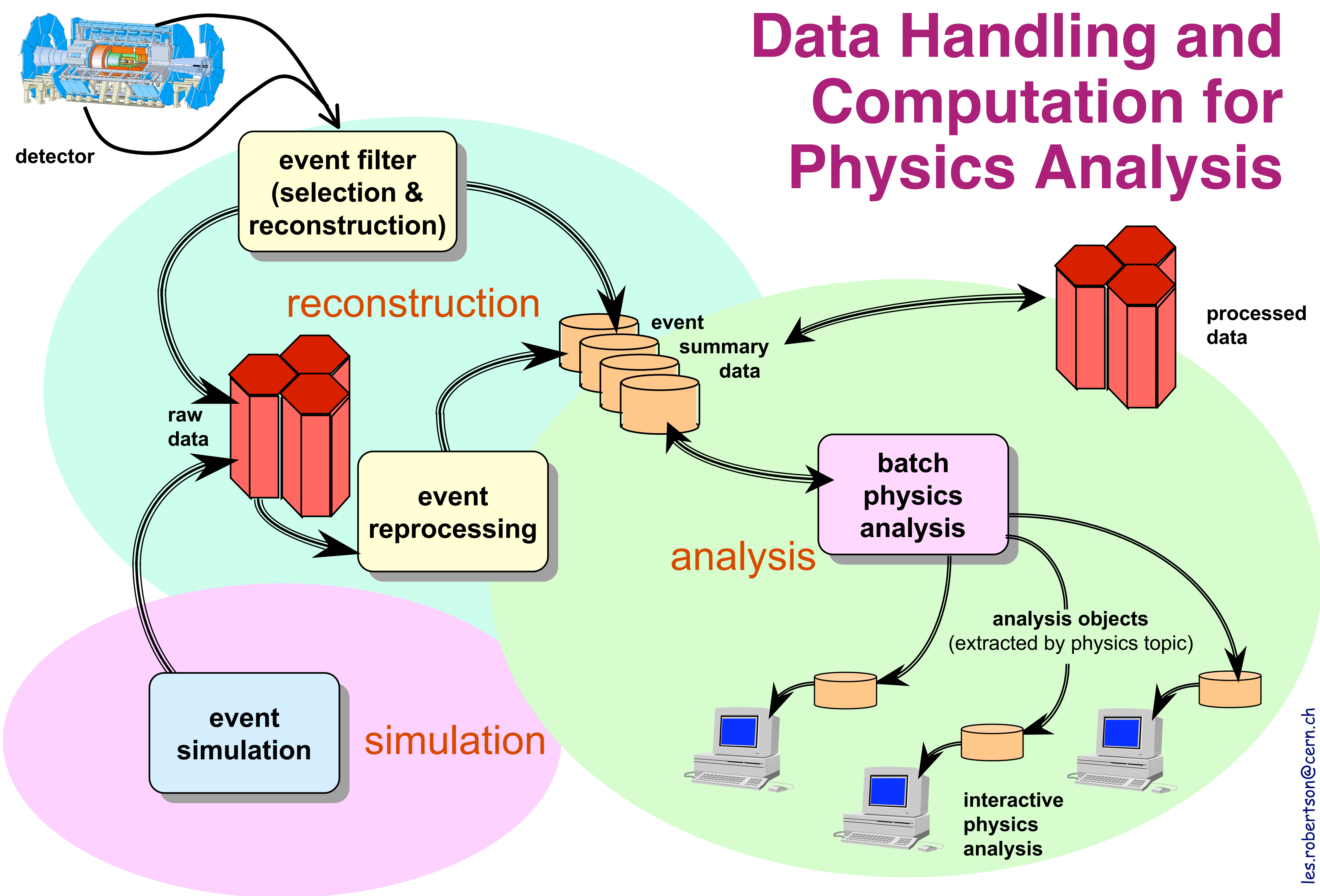
- Adopt Grid technology.
- Go for a "Tier" hierarchy.
- Use Intel CPUs in standard PCs
- Use LINUX operating system.

Goal

- Prepare and deploy the computing environment to help the experiments analyse the data from the LHC detectors.

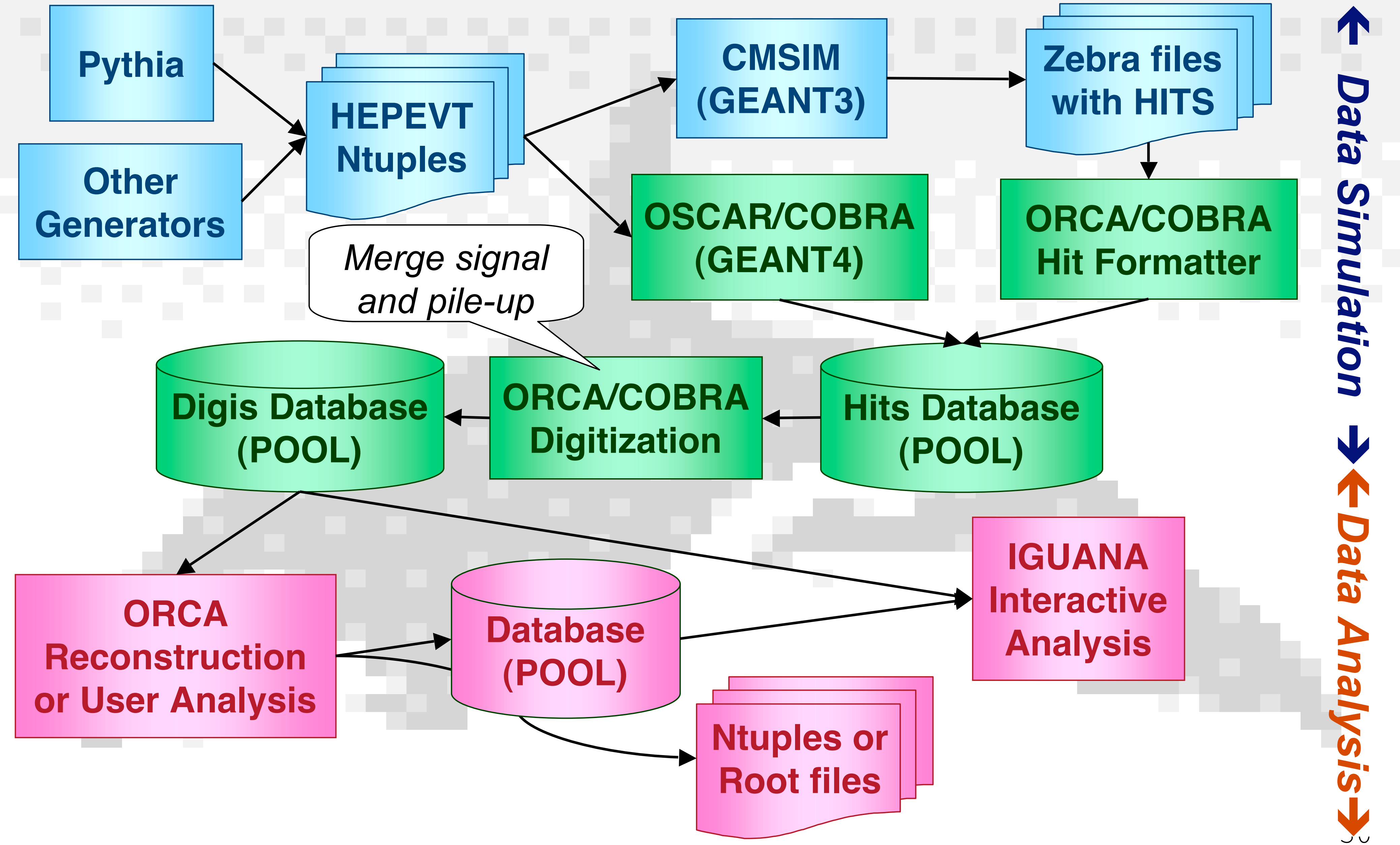


Data Handling and Computation for Physics Analysis

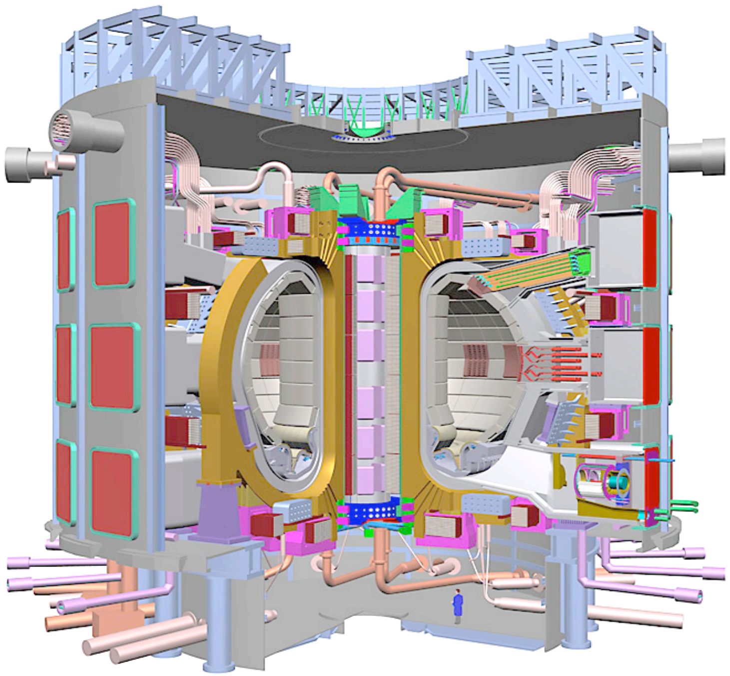


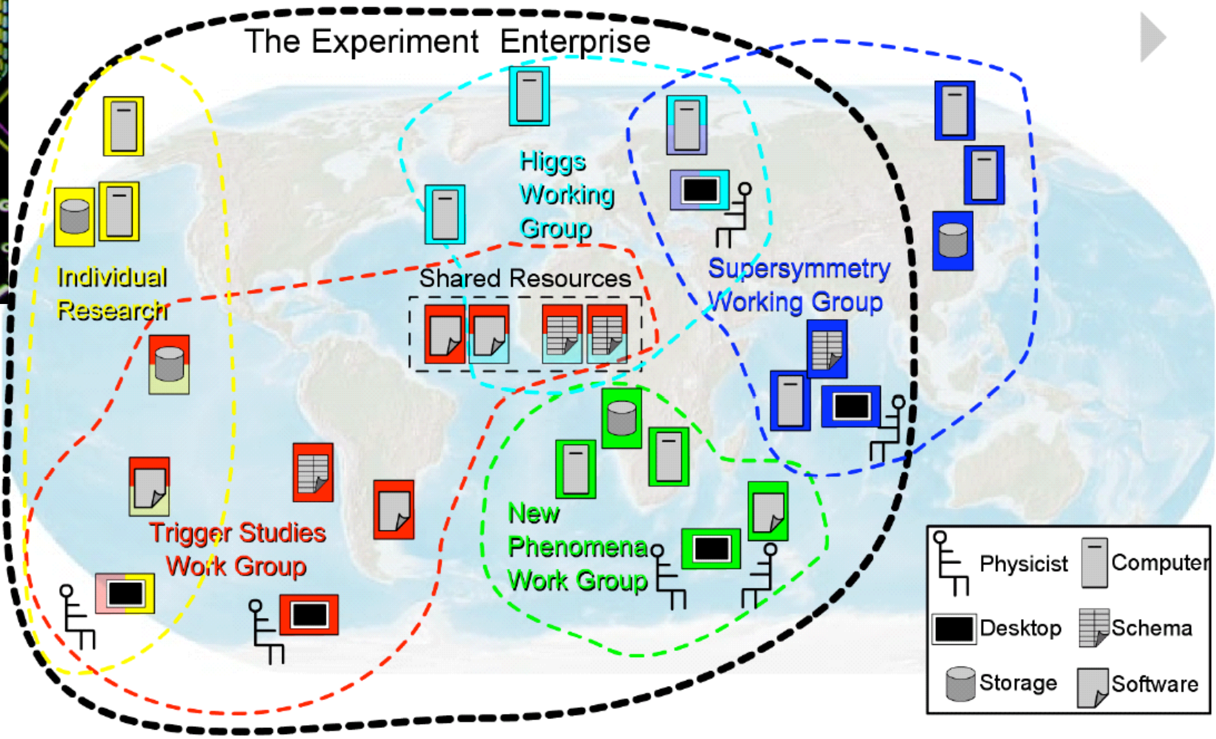
les.robertson@cern.ch

CMS software: ORCA & C.

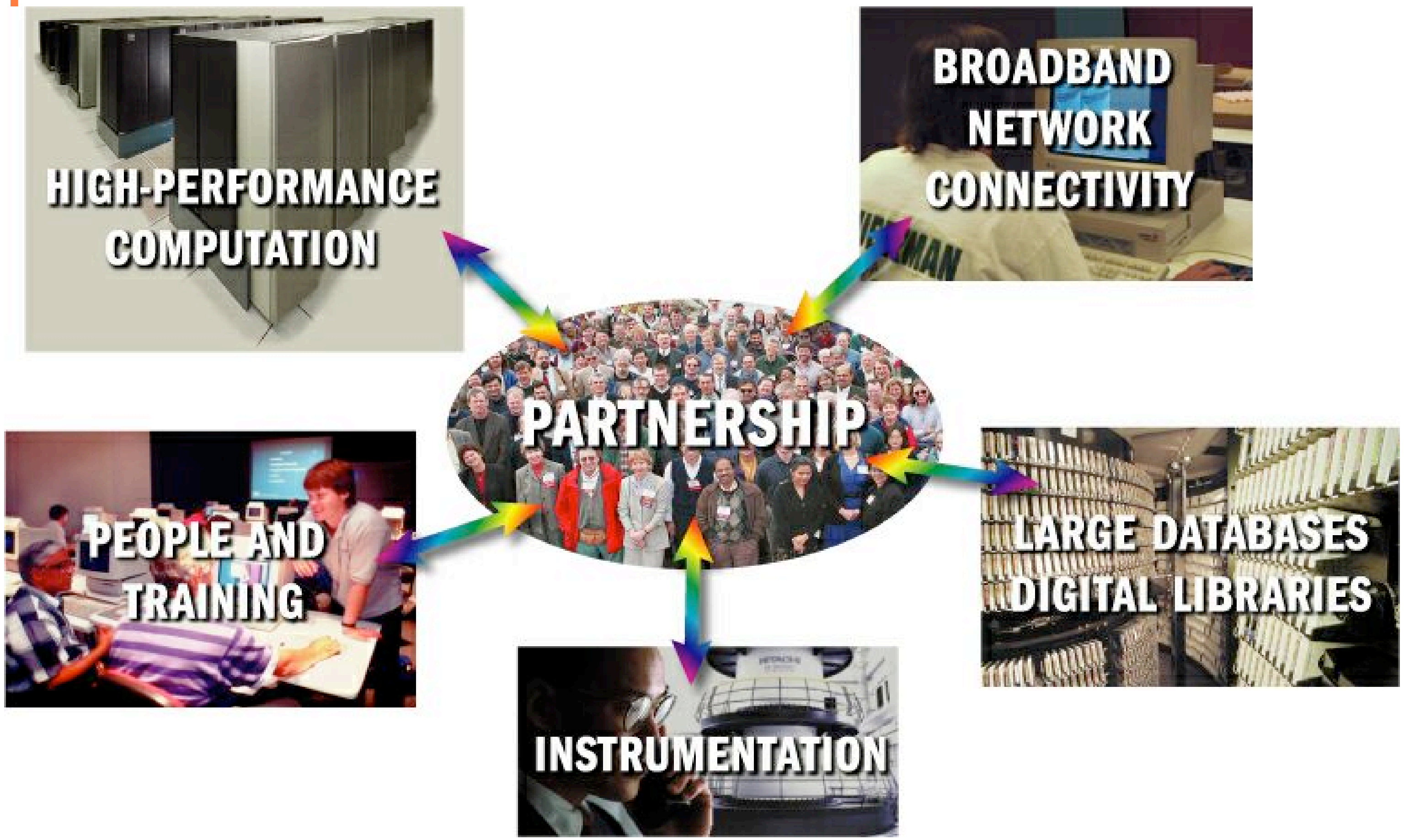


Science Today is a Team Sport





We Must be able to Assemble Required Expertise & Resources When Needed!



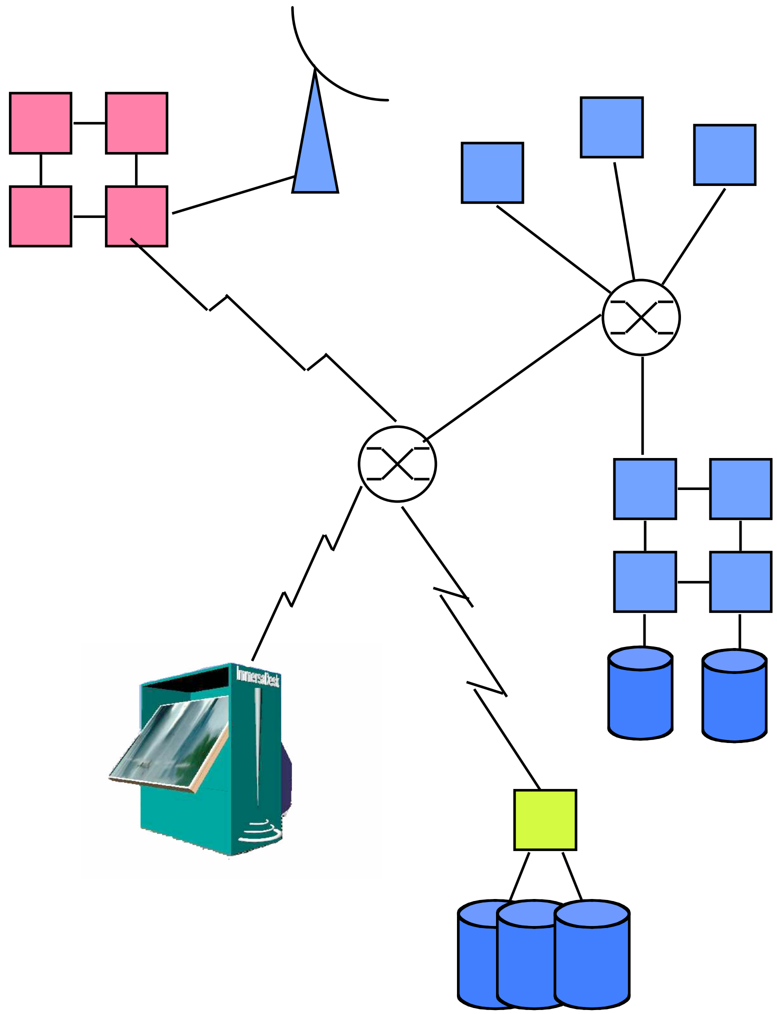
Transform resources into on-demand services accessible to any individual or team

"Grid Computing": a solution?

- What's Grid Computing
  - Succesor of Web?
- Is it a new paradigm of CS?
  - Yes, in the sense that try to build a new "standard": middleware
    - What's middleware?
- When and where is born?
  - ~1998/99: Globus/Condor in USA, DataGrid in EU
    - INFN special Project "INFN-Grid" since beginning of 2000
- Why HEP coming Experiments are building on it?
  - It's a possible solution for some of the Computing Models components (as others, like databases, web services and tools, networks both local and wide area, information systems, authorization systems, etc.)
- Grid has an architecture (middleware and layered models)

The Grid

"Dependable, consistent, pervasive access to [high-end] resources"



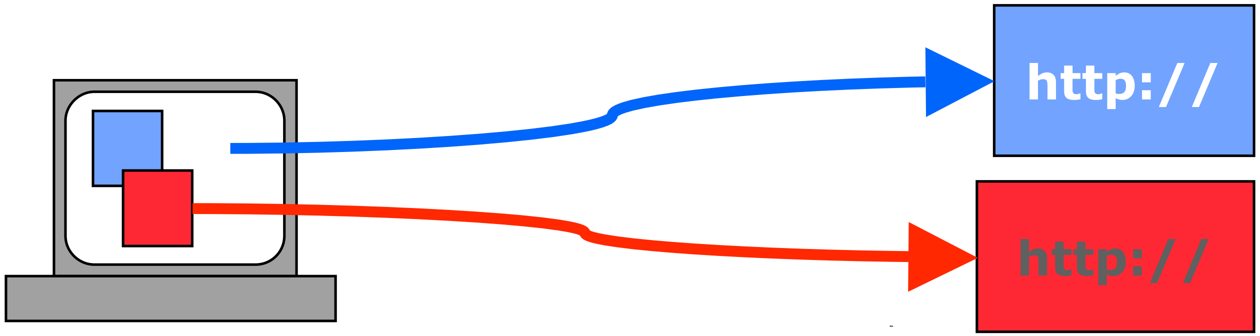
➡**Dependable:** Can provide performance and functionality guarantees

➡**Consistent:** Uniform interfaces to a wide variety of resources

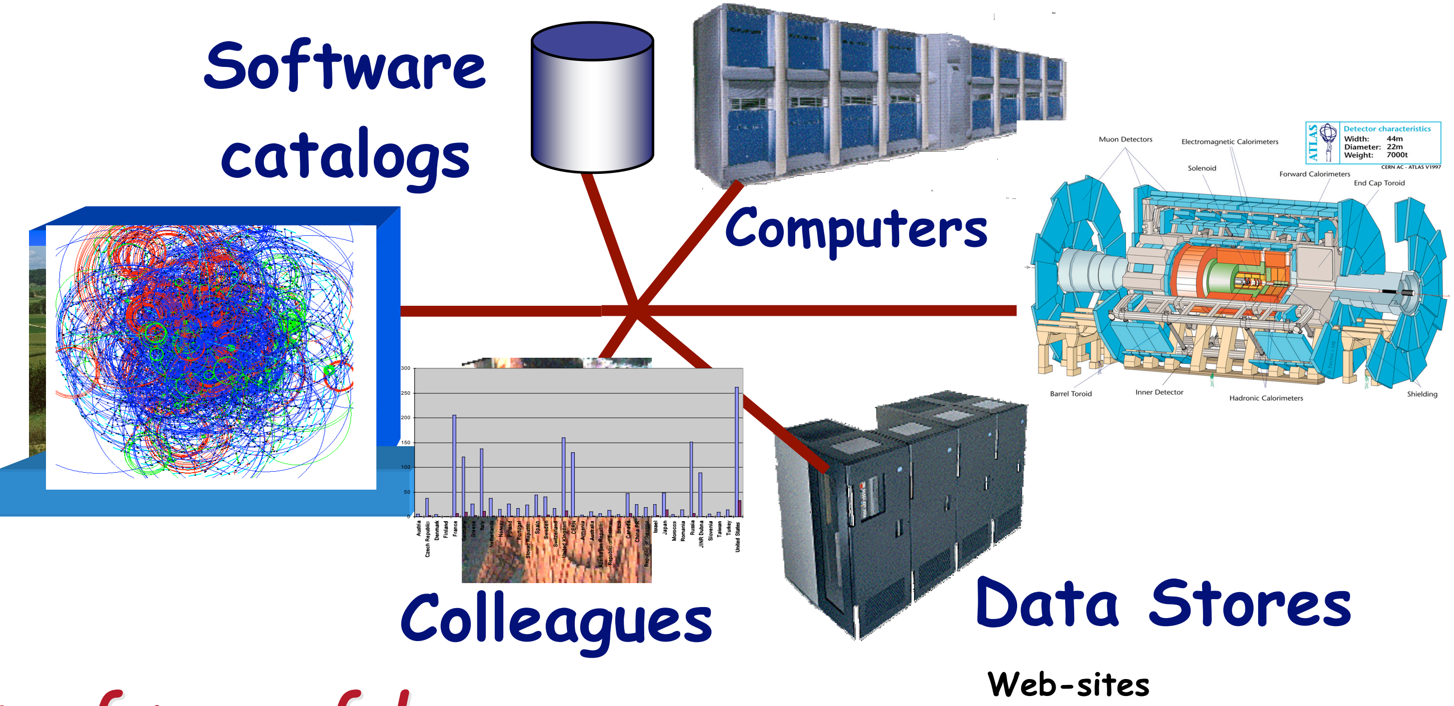
➡**Pervasive:** Ability to "plug in" from anywhere

Grids:
Next Generation Computing

Web: Uniform access to HTML documents



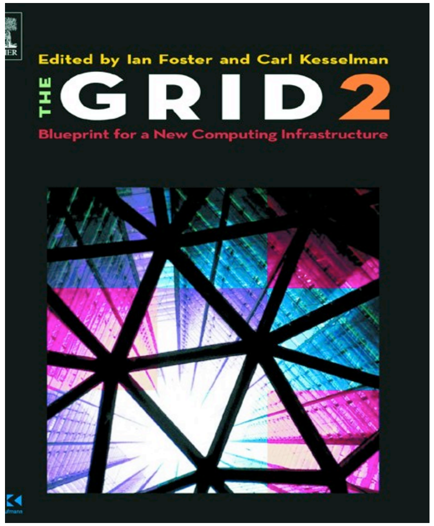
Grid: Flexible, high-performance access to all significant resources



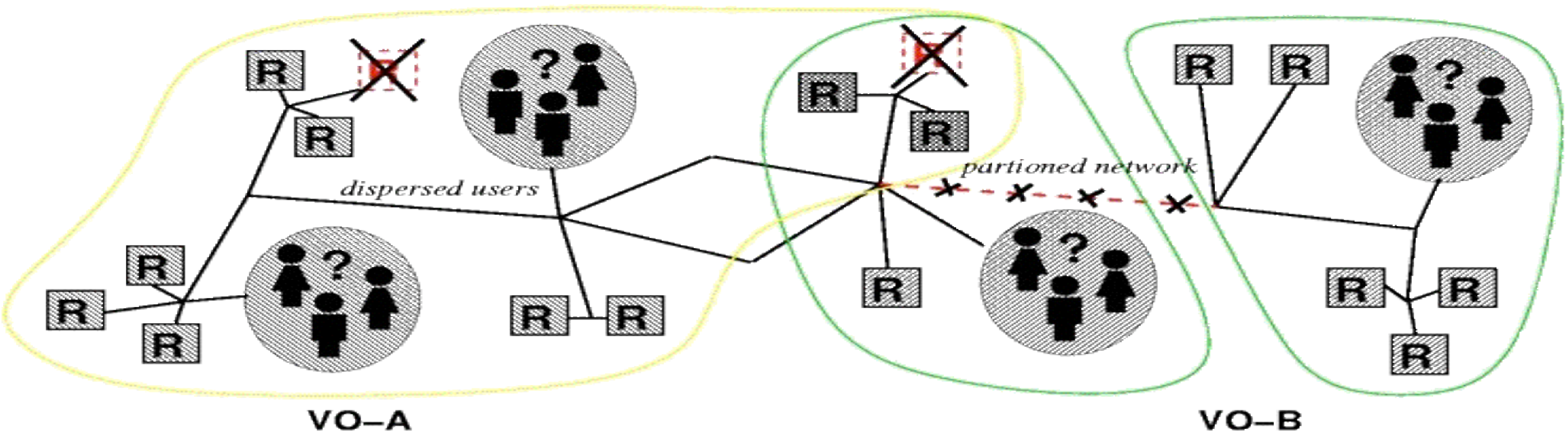
On-demand creation of powerful virtual computing and data systems

the globus alliance
www.globus.org

A Unifying Concept: The Grid



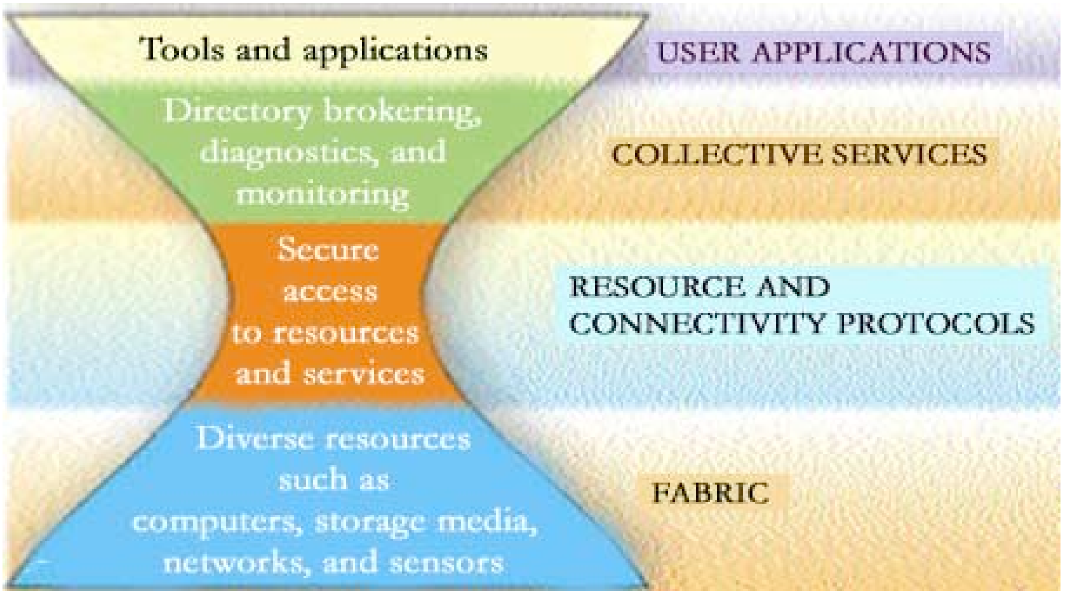
"Resource sharing & coordinated problem solving in dynamic, multi-institutional virtual organizations"



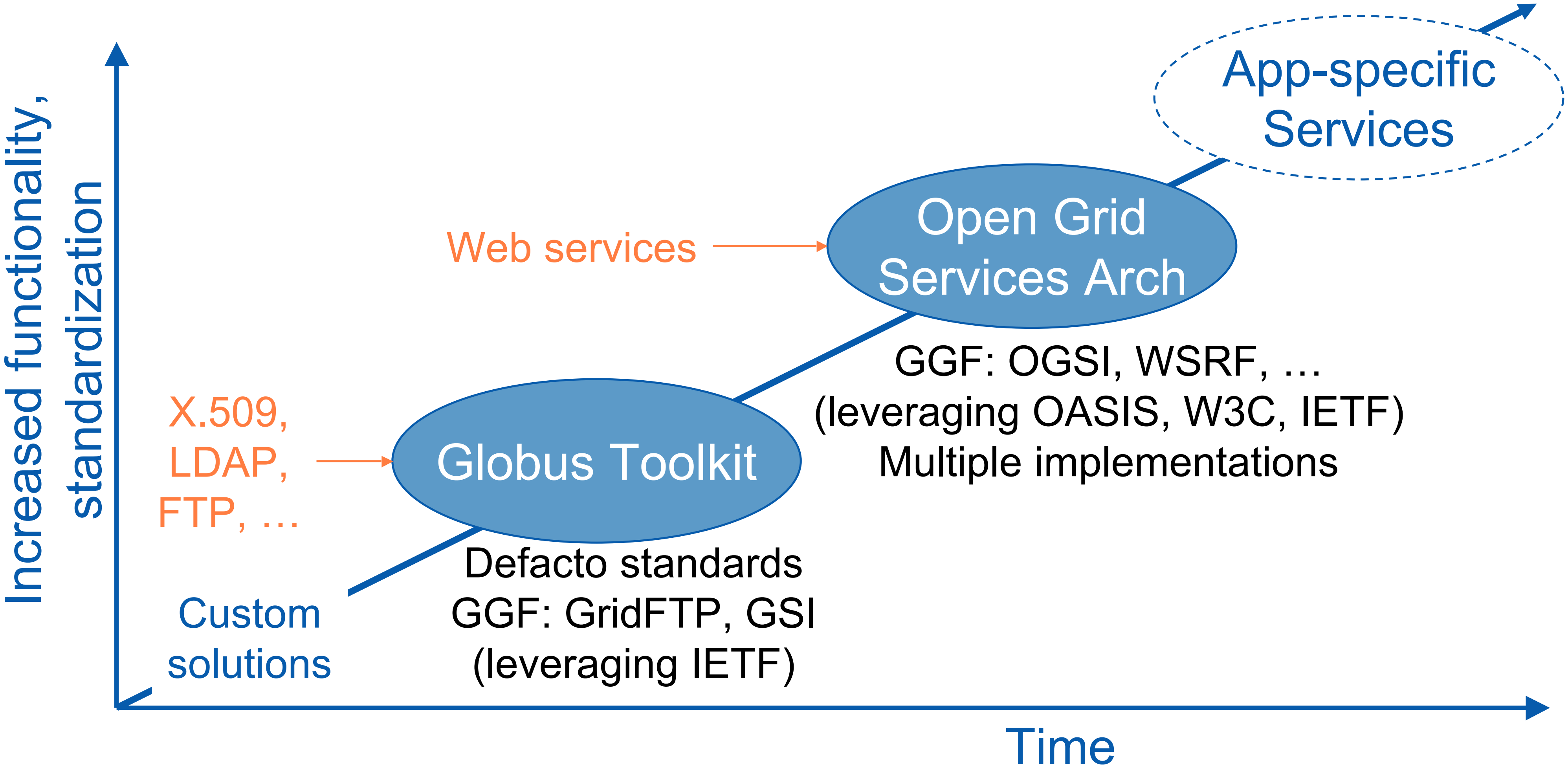
1. Enable integration of distributed resources
2. Using general-purpose protocols & infrastructure
3. To achieve better-than-best-effort service

Forget Homogeneity!

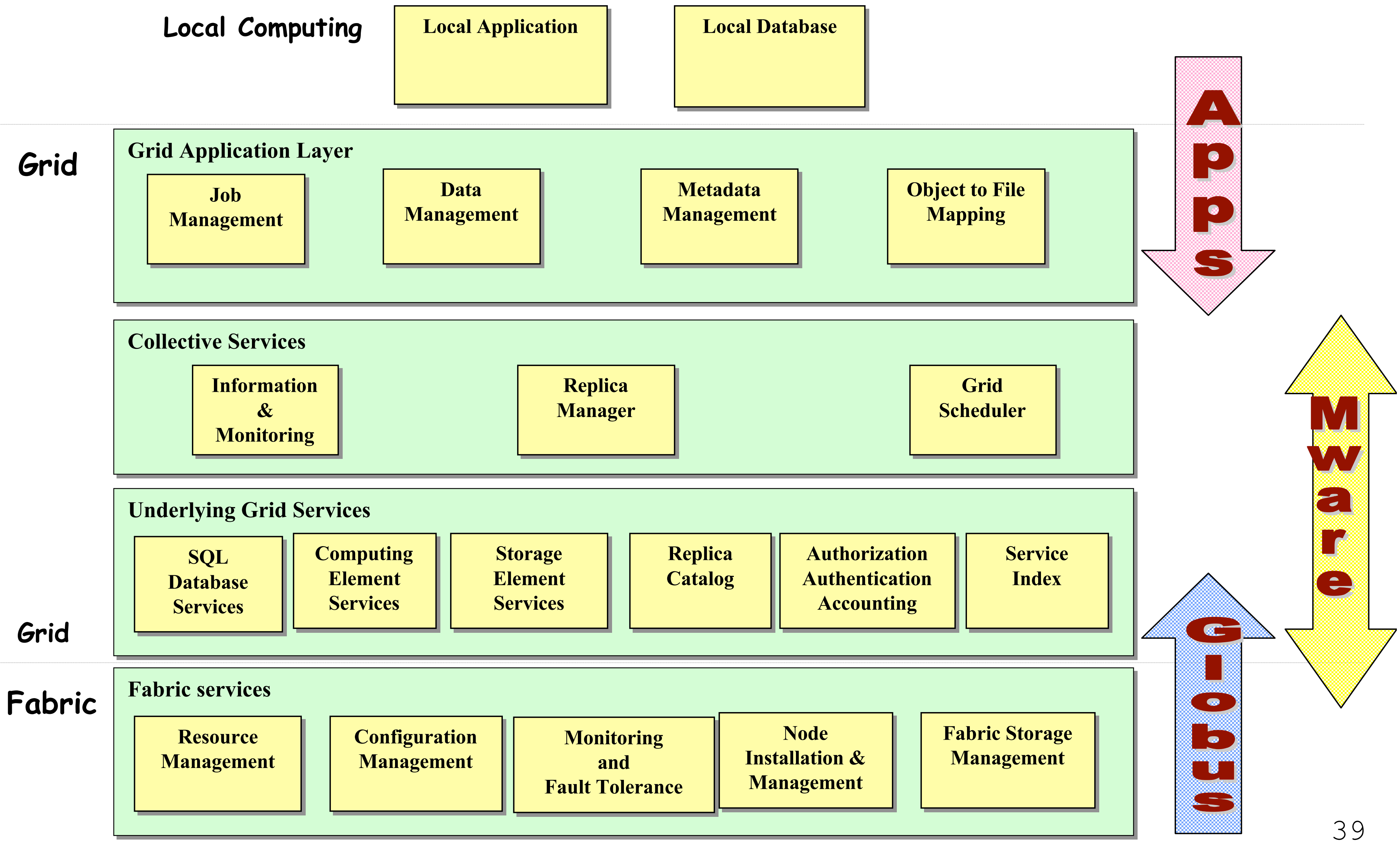
- Trying to force homogeneity on users is futile. Everyone has their own preferences, sometimes even *dogma*.
- The Internet provides the model…



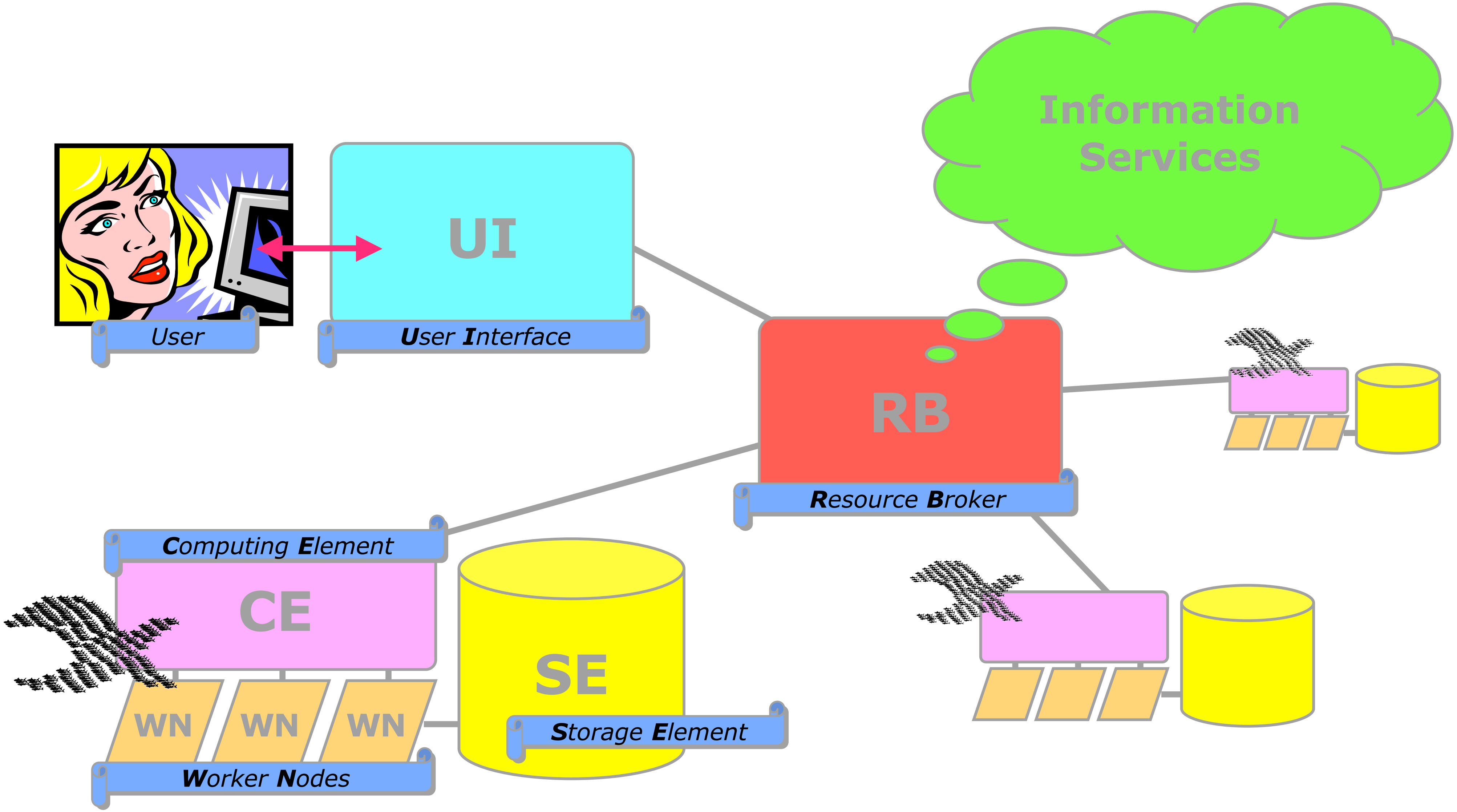
The "Grid Ecosystem"



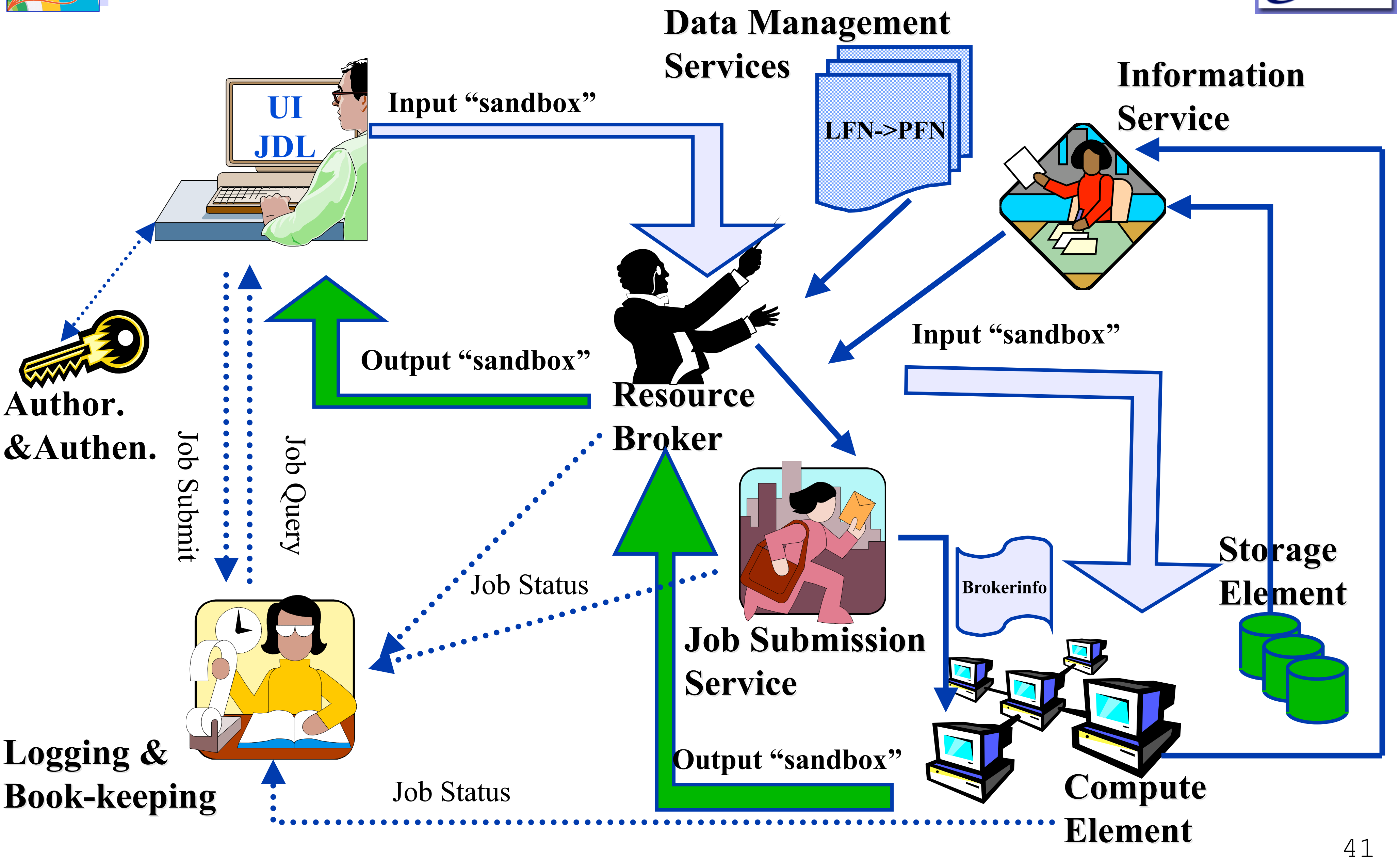
DataGrid Architecture: High level Implementation



A Grid Layout



A Job Submission example (EDG)

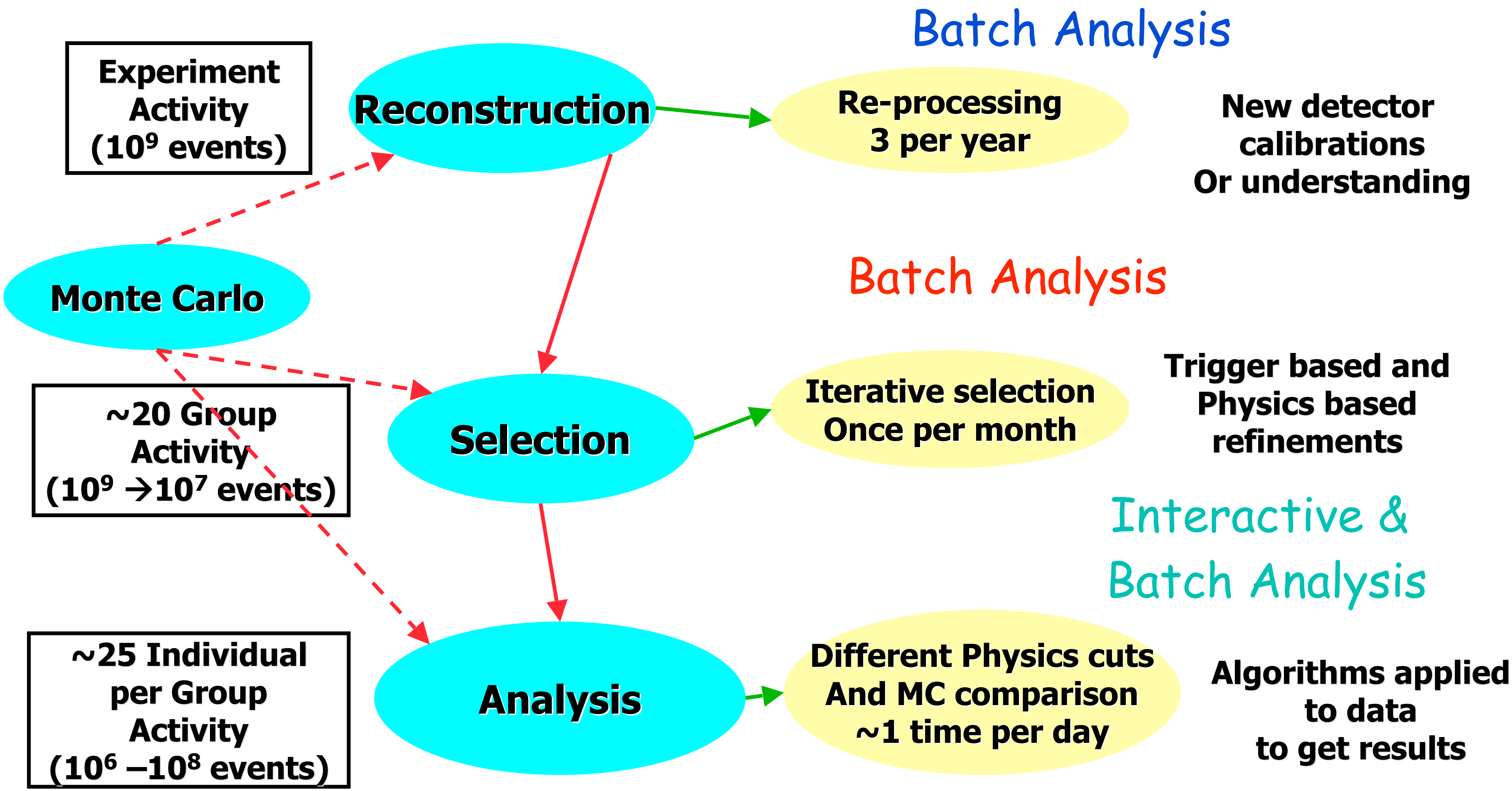


Second lesson

Computing Models (Every HEP Experiment has one, but also other Applications have one!)

- **Why a Computing Model?**
  - Complexity requires to state it beforehand
- **What's a Computing model?**
  - Components, infrastructure, application software, hardware resources, organization, user interfaces and … System Architecture
- **How can be built, or at least designed?**
  - It's a distributed effort, by definition
  - Has to cope with hierarchical dependencies to control the complexity
  - Has to allow for direct communications to mitigate the hierarchy (and control chaotic of user access and initiatives)
  - Need formal and real agreement of cooperation among Institutes for support (fair share stated with MoUs)
    - Need delegation of trust among the actors
  - Do not forget:
    - Time zones; really funny when considering a worldwide application
    - Data Model!; data access and format is mandatory to choose an implementation

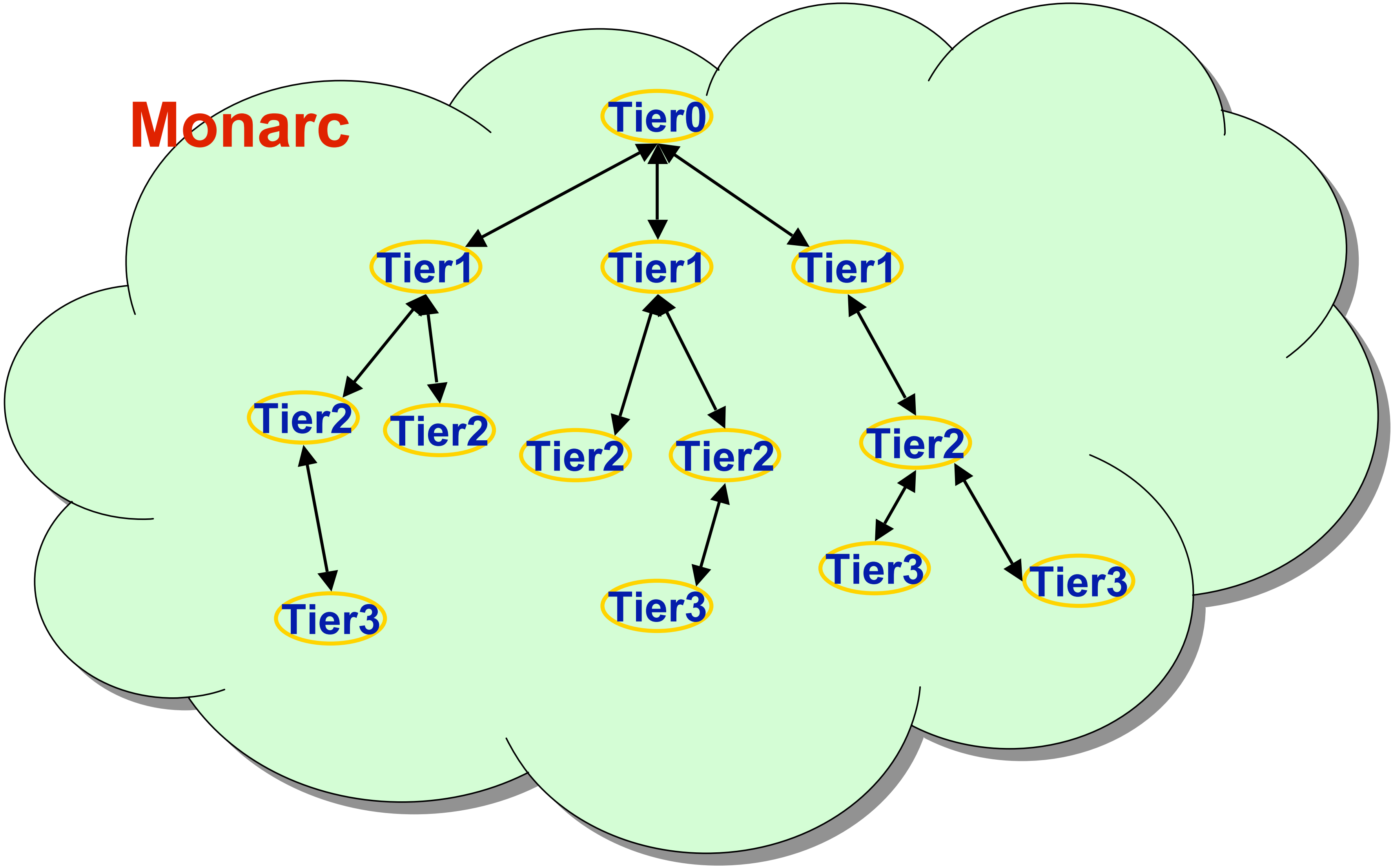
Hierarchy of Processes
(Experiment, Analysis Groups, Individuals)



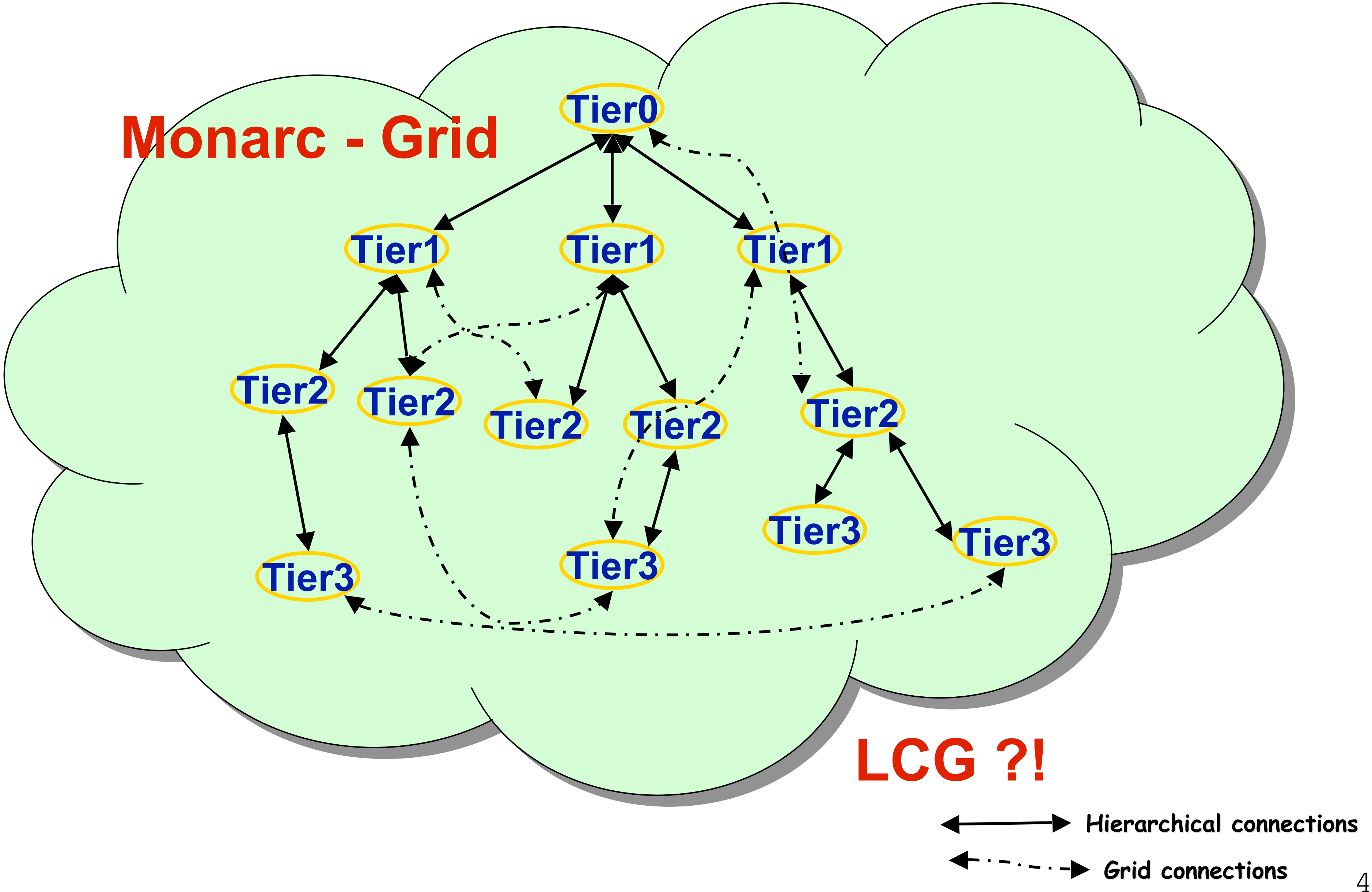
CMS Model: a remind (?)

- Scope and roles of the Tiers
  - Tier0: Central recording and "first" treatment of data
  - Tier1s: Computing support for the CMS Collaboration and the Analysis Groups
  - Tier2s: Analysis support and specific (identified) problems task-forces
  - Tier3s: Analysis dedicated and focused issues on particular tasks
  - Lower level Tiers: Local agreed activities and personal (users') tasks
- Scope and roles of the Regional Centers (RCs) in the "Grid"
  - Local RCs: User Interfaces and personal DBs
  - Distributed RCs: Ad-hoc resources for particular tasks and test services
  - Dedicated RCs: Analysis-dedicated resources and common (CMS) DBs
  - Common RCs: Grid Services (both common and CMS-specific) and DBs repositories
- Dynamically de-localized commitments and resources
  - Mostly person-power- & knowledge-based on specific problems
    - Both for computing and Physics skills
  - Re-allocation of tasks within a:
    - Virtual Organization (Grid VOs)
    - Country Organization (e.g. INFN coordination, hierarchy of Centers)
    - Analysis Organization (CMS coordination, hierarchy of Roles)

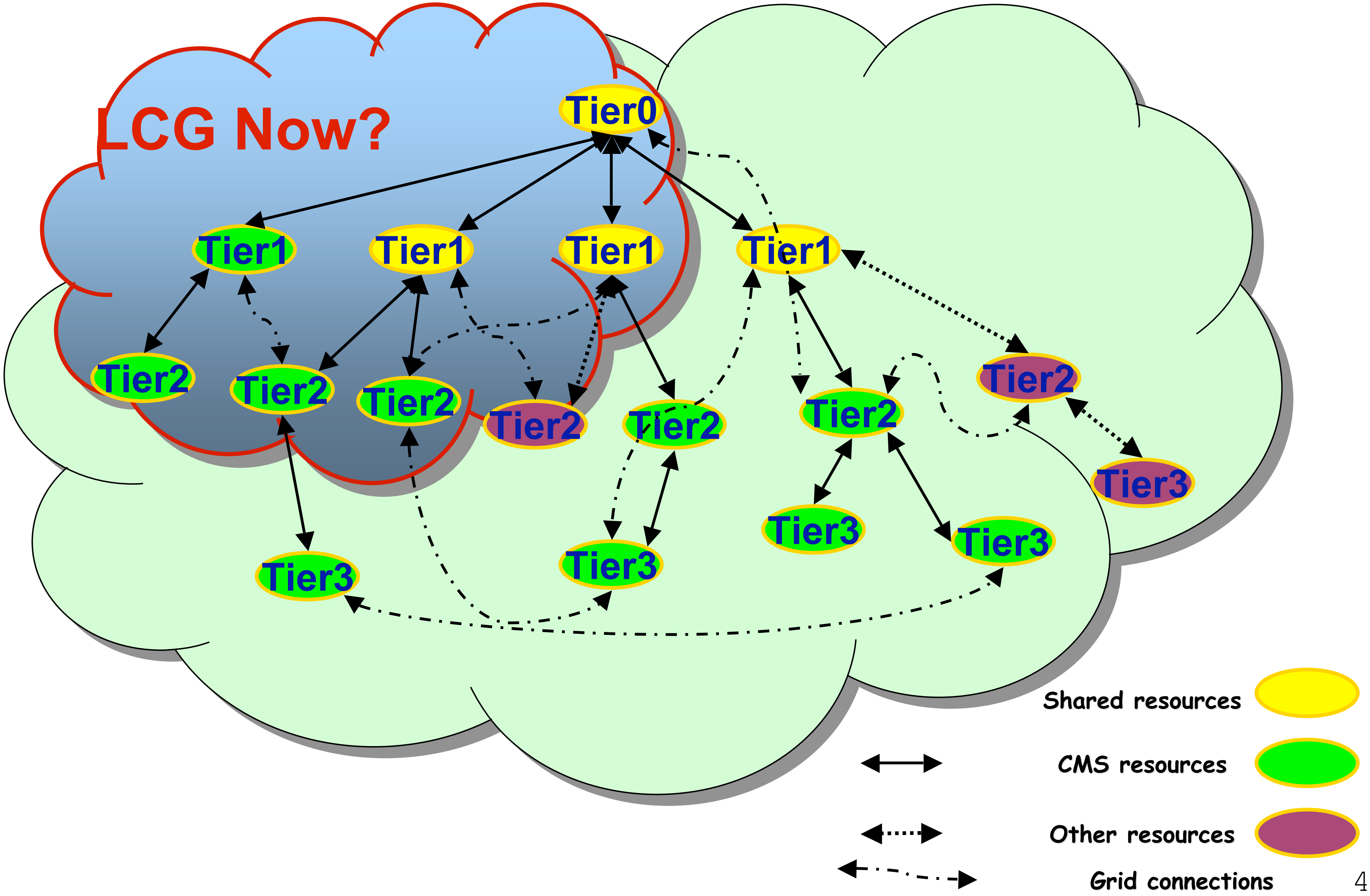
“Distributed” Models



“Distributed” Models



"Distributed" Models



The “dimension” of CMS Tiers: setting the requirements

The CERN T0 (Capacity available)	2006	2007	2008	
CPU scheduled	693	1485	3176	kSI2K
Disk	327	322	309	Tbytes
Active tape	2367	4576	6738	Tbytes
Tape I/O	282	508	800	MB/s
Number of bi-CPU boxes	340	561	882	

The Capacity available in a single T1	2006	2007	2008	
CPU scheduled	97	162	290	kSI2K
CPU analysis	616	1024	1834	kSI2K
Total CPU	713	1186	2124	kSI2K
Disk	508	832	1454	Tbytes
Active tape	1072	1950	2769	Tbytes
Tape I/O	183	282	400	MB/s
Number of CPU boxes	392	492	590	

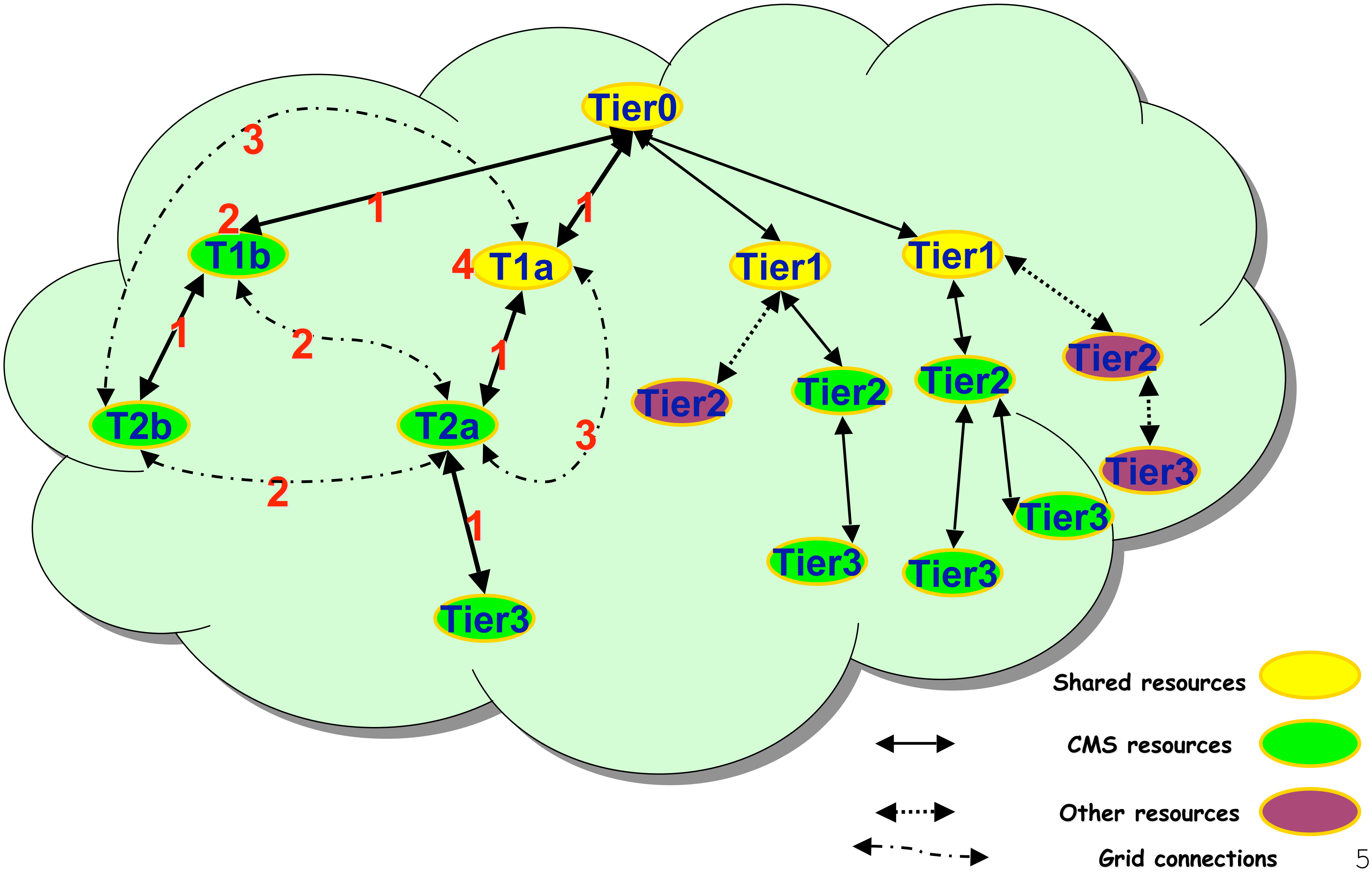
The Capacity available in a single T2	2006	2007	2008	
CPU scheduled	47	78	139	kSI2K
CPU analysis	59	97	174	kSI2K
Total CPU	105	175	313	kSI2K
Disk	64	105	183	Tbytes
Archive tape	133	242	345	Tbytes
Tape I/O	46	71	100	MB/s
Number of CPU boxes	58	73	87	

An Analysis scenario at T2

1. User jobs run locally on local input sample (T2a)
2. Until a collaborator re-calculate new partial sample (at T2b or T1b):
   a) User jobs are split to run locally on local partial sample and remotely on the new re-calculated data
   b) At a certain point the user (or the system) decides that would be better to have the remote “new” part of the data locally (replication of data)
3. It may happen that the CPU resources of the the T2a and T2b are already committed for other tasks
   a) User jobs can run on remote resources with “remote” data access (either a CMS T1 or T2, or even a non-CMS Tier)
4. The user decide to run on a larger sample (requiring also a consistent CPU power)
   a) User jobs go to the T1 on which the T2 user depend (T1a) or to the T1 of the remote collaborator (T1b), or ? (don’t think it can run on a non-CMS T1)

Provided that we know frequencies of jobs and data dimensions, the load on CPUs, storage and network can be derived

A "use case"

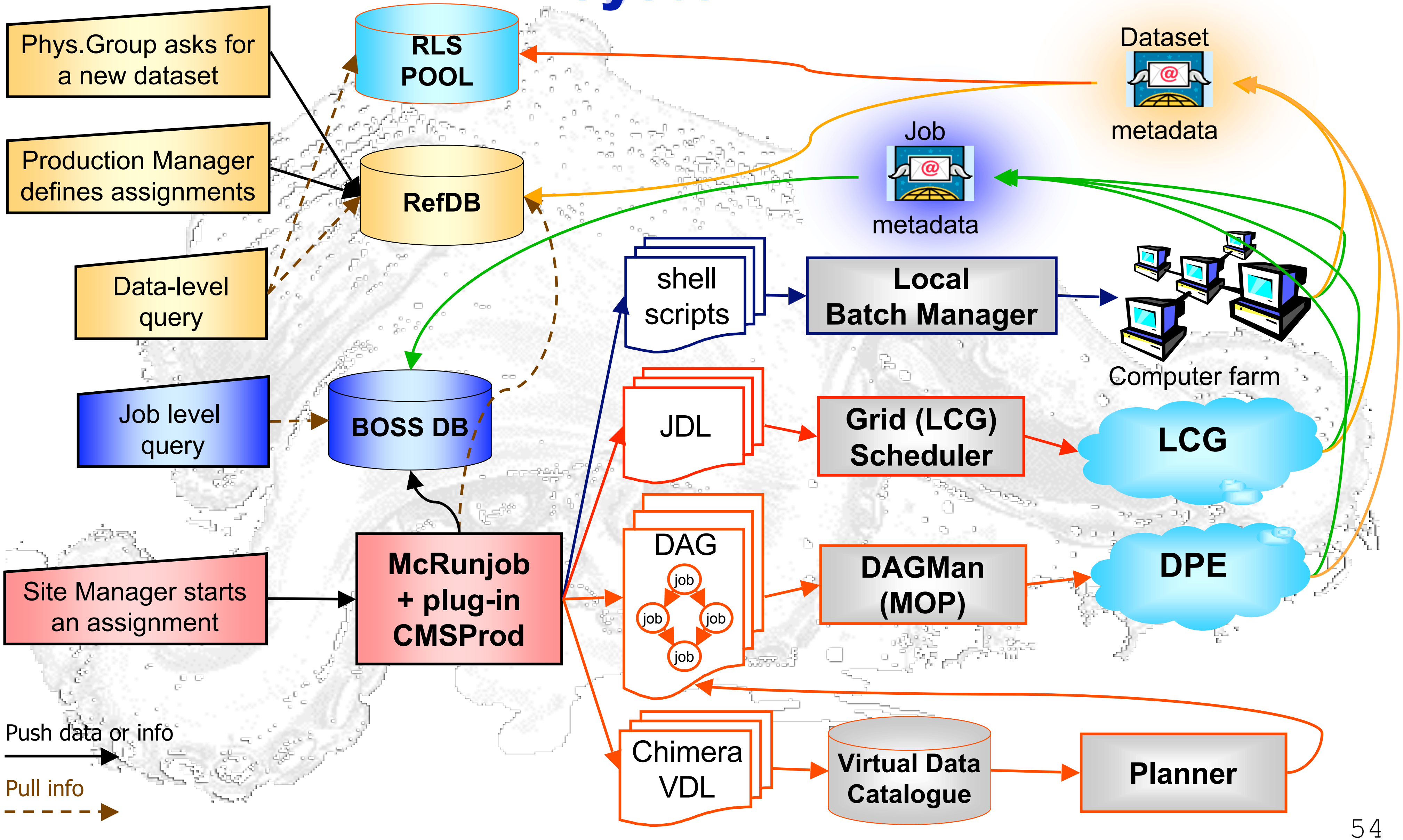

Tier0
1
1
2
T1b
4
T1a
Tier1
Tier1
3
1
2
1
T2b
T2a
3
2
1
Tier2
Tier2
Tier2
Tier2
Tier3
Tier3
Tier3
Tier3
Tier3
Shared resources
CMS resources
Other resources
Grid connections

But avoid ... well known symptoms

Building and Measuring the Systems (Models)

- Data and Physics Challenges
  - Better, Experiment's Challenge
- What's a "challenge" of Large Scale Computing for LHC Experiments?
  - Examples follow

CMS OCTOPUS Data Production System

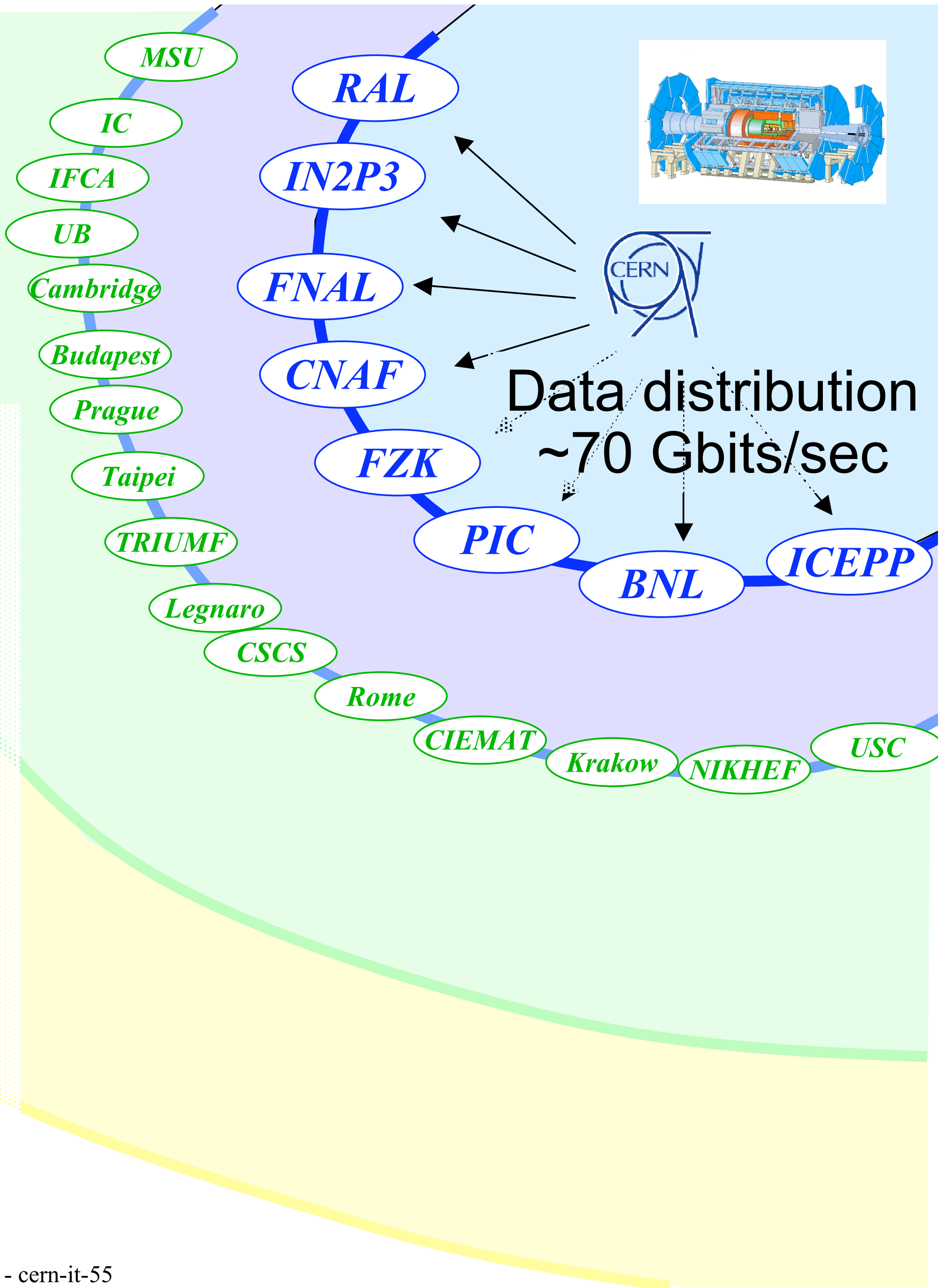


Current estimates of Computing Resources needed at Major LHC Centres

First full year of data - 2008

	*Processing M SI2000***	*Disk PetaBytes*	*Mass Storage PetaBytes*
CERN	20	5	20
Major data handling centres (Tier 1)	45	20	18
Other large centres (Tier 2)	40	12	5
Totals	**105**	**37**	**43**

*** Current fast processor ~1K SI2000*



Regional centres connected to the LCG grid

country	*centre*	*country*	*centre*
Austria	**UIBK**	**Portugal**	**LIP, Lisbon**
Canada	**TRIUMF, Vancouver**	**Russia**	**SINP, Moscow**
	Univ. Montreal	**Spain**	**PIC, Barcelona**
	Univ. Alberta		**IFIC, Valencia**
Czech Republic	**CESNET, Prague**		**IFCA, Santander**
	University of Prague		**University of Barcelona**
France	IN2P3, Lyon**		**Uni. Santiago de Compostela**
Germany	**FZK, Karlsruhe**		**CIEMAT, Madrid**
	DESY		**UAM, Madrid**
	University of Aachen	**Switzerland**	**CERN**
	University of Wuppertal		CSCS, Manno**
Greece	**GRNET, Athens**	**Taiwan**	**Academia Sinica, Taipei**
Holland	**NIKHEF, Amsterdam**		**NCU, Taipei**
Hungary	**KFKI, Budapest**	**UK**	**RAL**
Israel	Tel Aviv University**		**Cavendish, Cambridge**
	Weizmann Institute		**Imperial, London**
Italy	**CNAF, Bologna**		**Lancaster University**
	INFN, Torino		**Manchester University**
	INFN, Milano		**Sheffield University**
	INFN, Roma		**QMUL, London**
	INFN, Legnaro	**USA**	**FNAL**
Japan	ICEPP, Tokyo**		BNL**
Poland	**Cyfronet, Krakow**		

** not yet in LCG-2

Centres in process of being connected

country	*centre*
China	**IHEP, Beijing**
India	**TIFR, Mumbai**
Pakistan	**NCP, Islamabad**

Hewlett Packard to provide "Tier 2-like" services for LCG, initially in Puerto Rico

> 40 sites; > 3,100 CPUs

LCG for the 2004 Data Challenges

LCG-2 target
– the 2004 "*LHC Data Challenges*"

- Large-scale tests of the experiments' computing models, processing chains, grid technology readiness, operating infrastructure
- ALICE and CMS data challenges started at the beginning of March
- LHCb and ATLAS – started in May
- The big challenge for this year - data –
  - file catalogue,
  - replica management,
  - database access,
  - integrating mass storage

Grid Operations Centre at RAL

CCLRC

User Support Centre at FZK

GridKa

Planning for a second operations & support centre in Taipei

ASCC
ACADEMIA Sinica Computing Centre
中央研究院計算中心

CMS DC04 Data Challenge

T0 at CERN in DC04

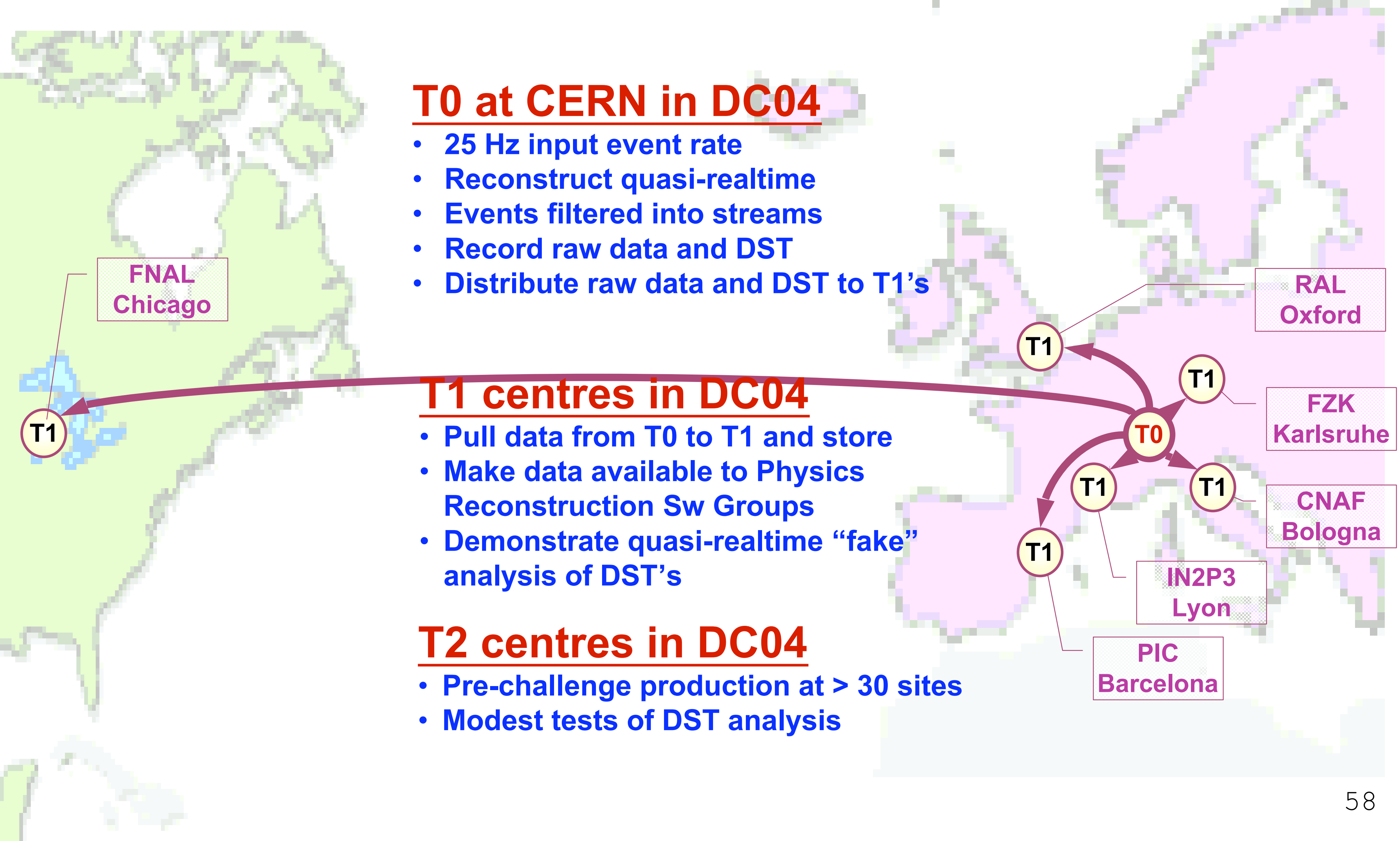
- 25 Hz input event rate
- Reconstruct quasi-realtime
- Events filtered into streams
- Record raw data and DST
- Distribute raw data and DST to T1's

T1 centres in DC04

- Pull data from T0 to T1 and store
- Make data available to Physics Reconstruction Sw Groups
- Demonstrate quasi-realtime "fake" analysis of DST's

T2 centres in DC04

- Pre-challenge production at > 30 sites
- Modest tests of DST analysis



Scope of CMS Data Challenge 04

Aim of DC04:

- reach a sustained 25Hz reconstruction rate in the Tier-0 farm (25% of the target conditions for LHC startup)
- register data and metadata to a catalogue
- transfer the reconstructed data to all Tier-1 centers
- analyze the reconstructed data at the Tier-1's as they arrive
- publicize to the community the data produced at Tier-1's
- monitor and archive of performance criteria of the ensemble of activities for debugging and post-mortem analysis

Not a CPU challenge, but a full chain demonstration!

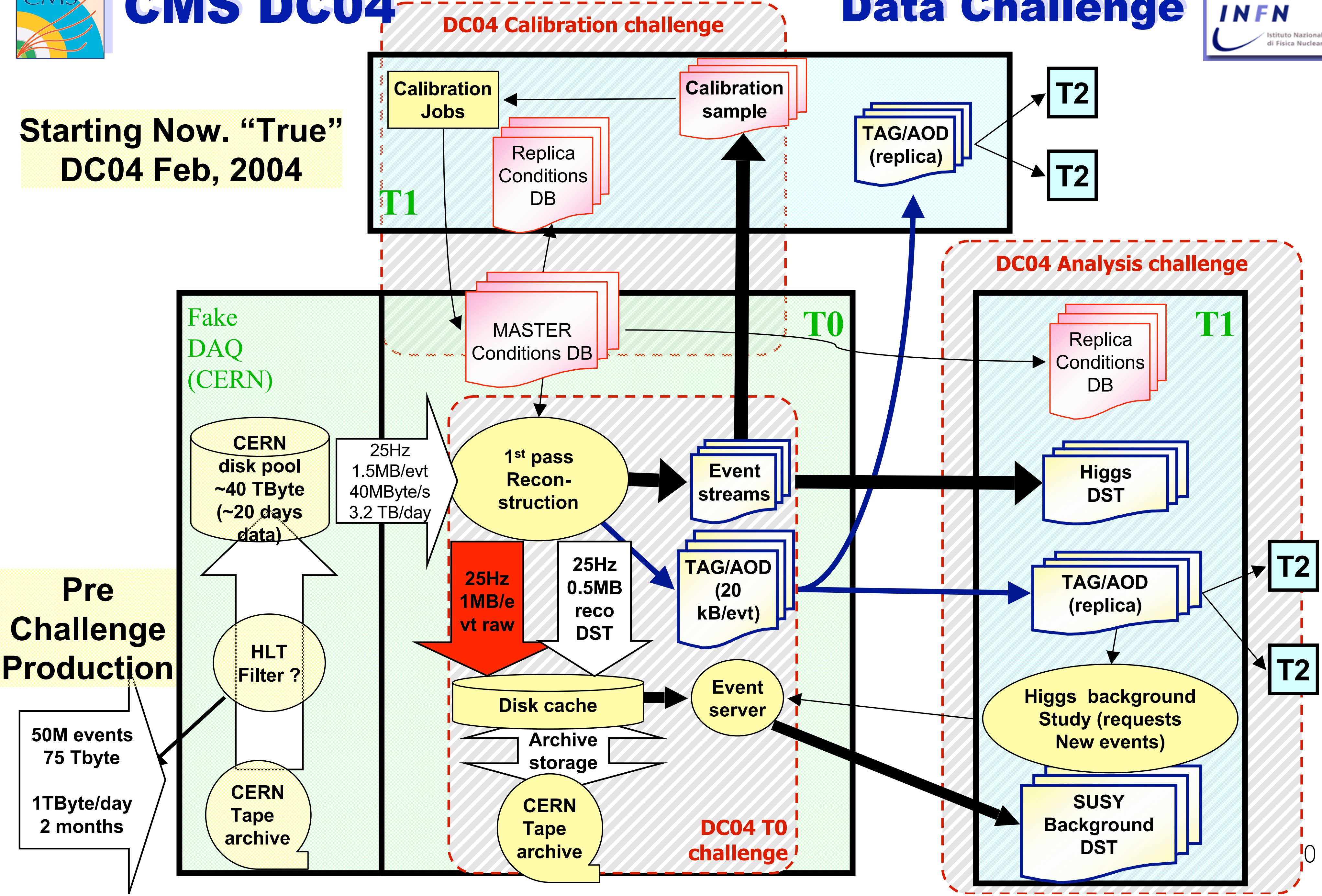
Pre-challenge production in 2003/04

- 70M Monte Carlo events (30M with Geant-4) produced
- Classic and grid (CMS/LCG-0, LCG-1, Grid3) productions

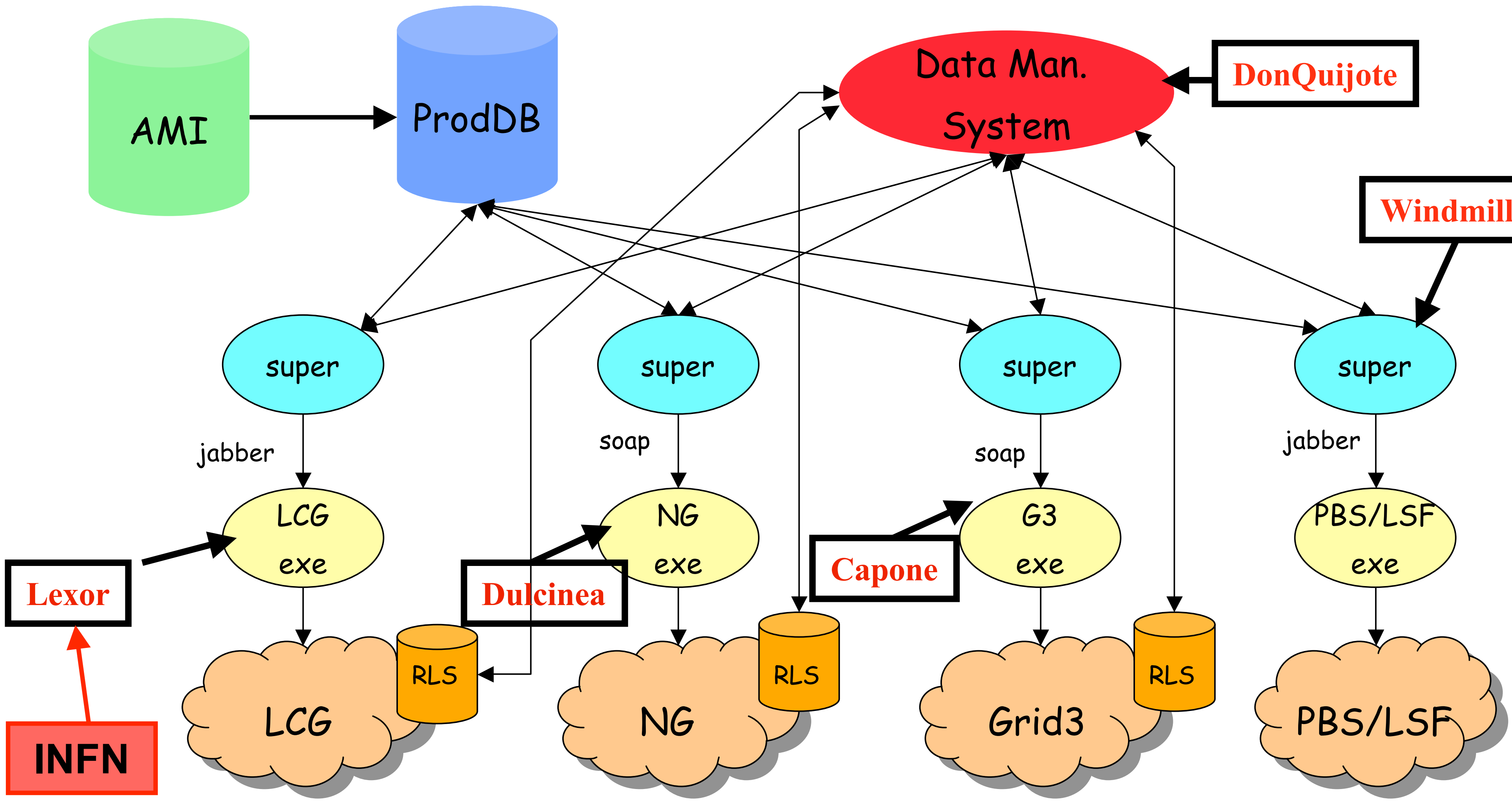
Was a "challenge", and everytime we found a scalability limit of a component, was a Success!

CMS DC04 Data Challenge

Starting Now. "True" DC04 Feb, 2004



ATLAS New Production System



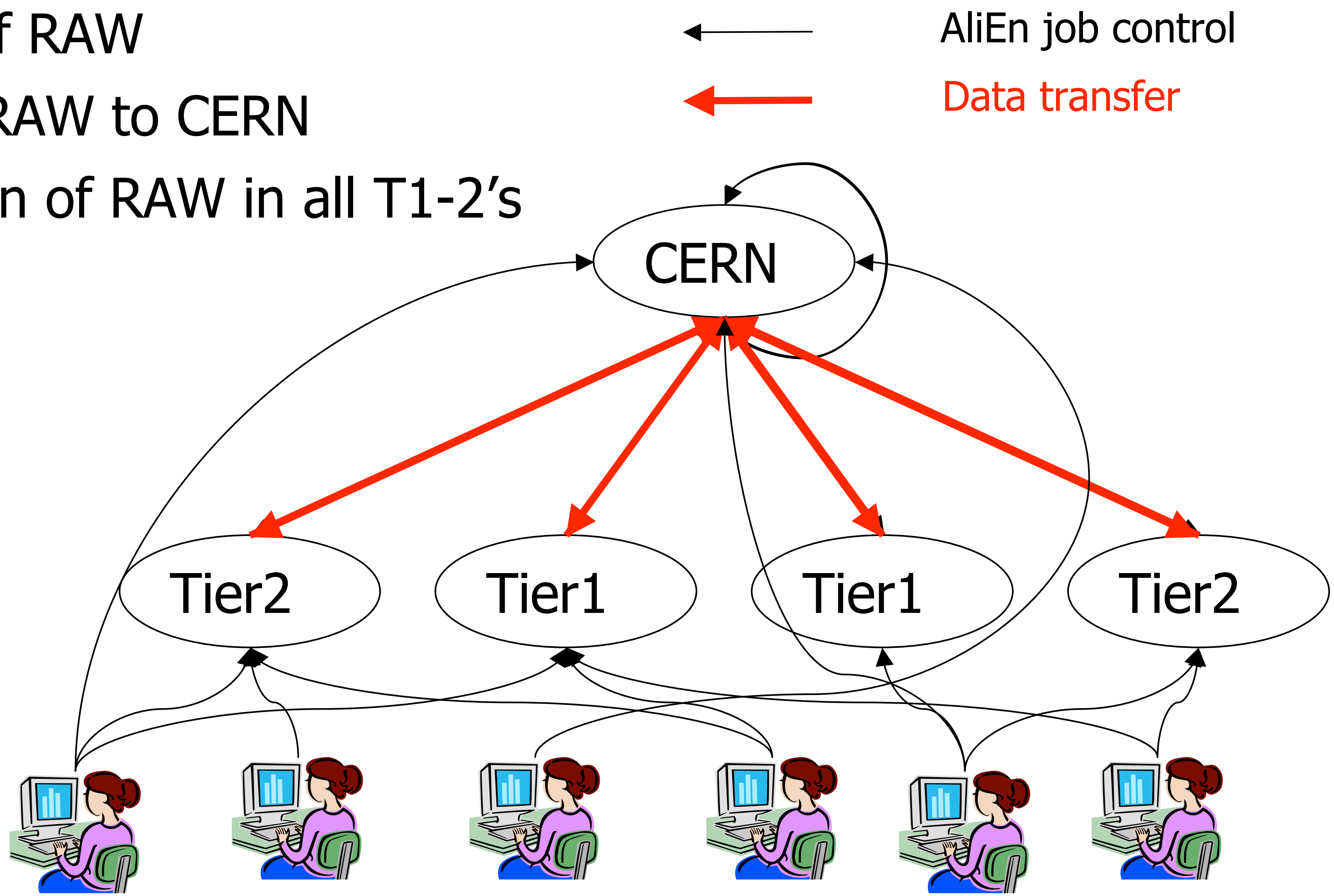
ALICE PDC 3 schema

Production of RAW

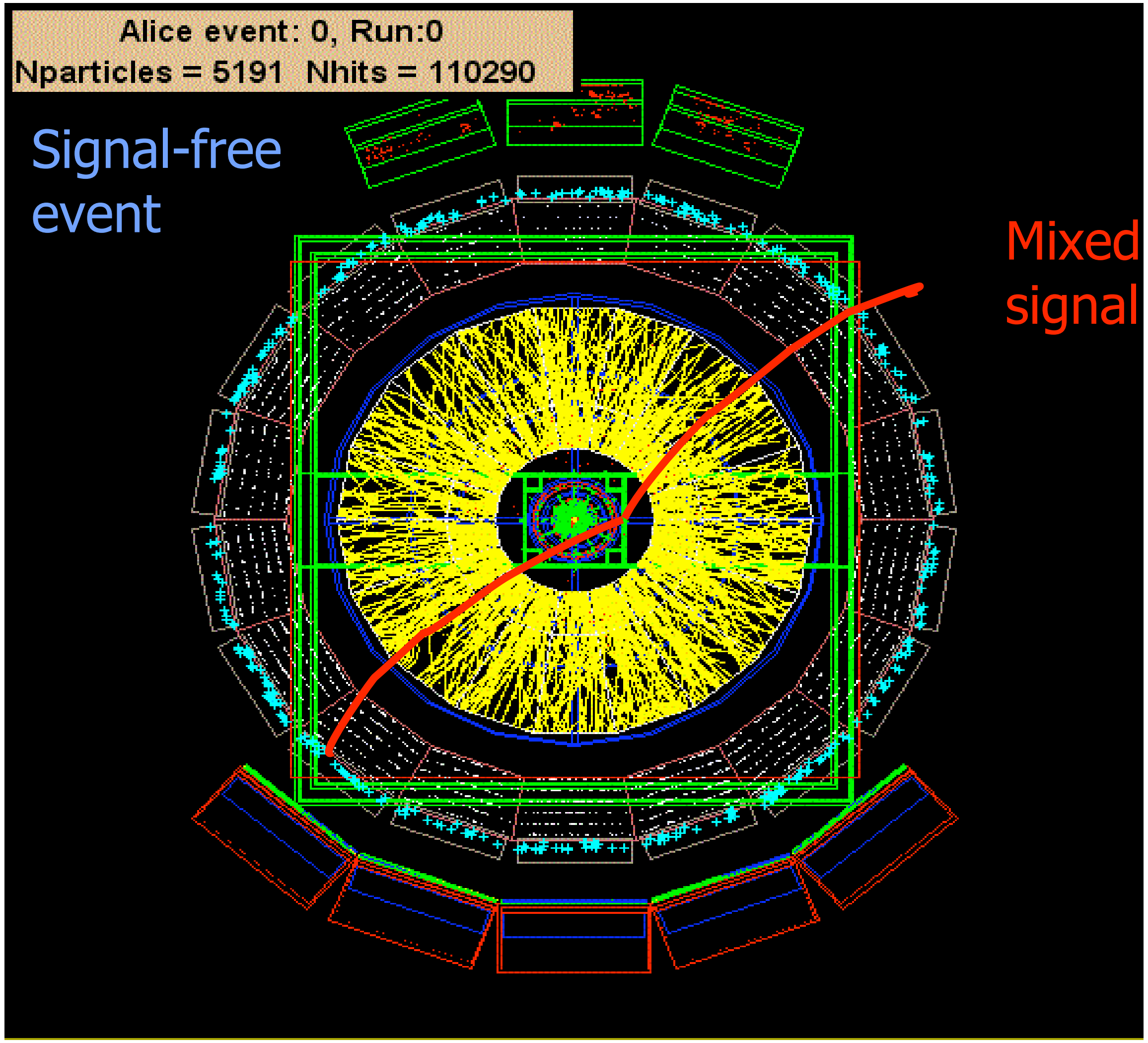
Shipment of RAW to CERN

Reconstruction of RAW in all T1-2's

Analysis



Alice Merging of one event in DC

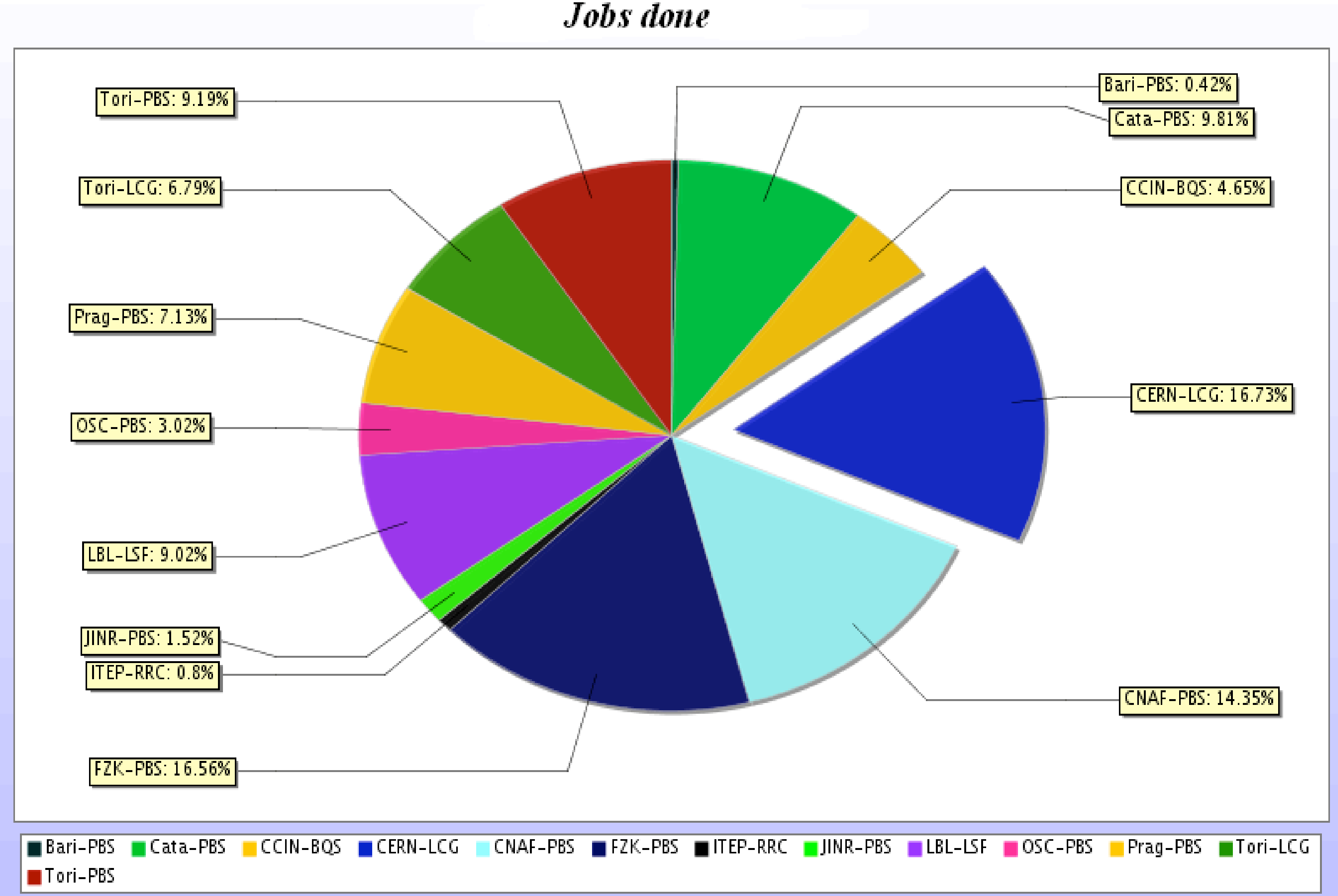


Where we are

Results of LHC Experiments' Data Challenges

Alice Data Challenge Phase 1 resource statistics:

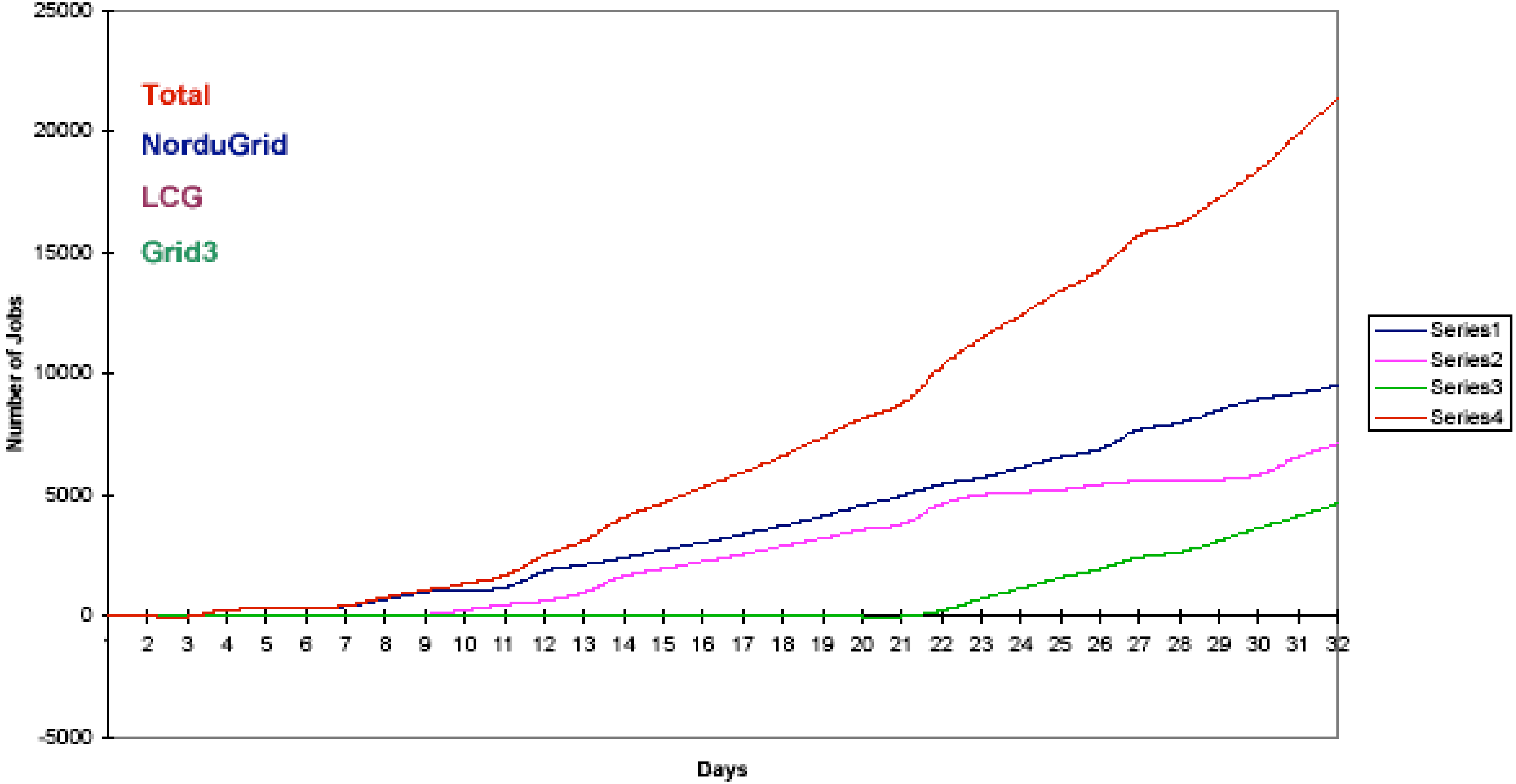
- 27 production centres, 12 major producers, no single site dominating the



Statistics for phase 1 of ALICE PDC 2004

- **Number of jobs:**
  - Central 1 (long, 12 hours) – 20 K
  - Peripheral 1 (medium – 6 hours) – 20 K
  - Peripheral 2 to 5 (short – 1 to 3 hours) – 16 K
- **Number of files:**
  - AliEn file catalogue: 3.8 million (no degradation in performance observed)
  - CERN Castor: 1.3 million
- **File size:**
  - Total: 26 TB
- **CPU work:**
  - Total: 285 MSI-2K hours
  - LCG: 67 MSI-2K hours

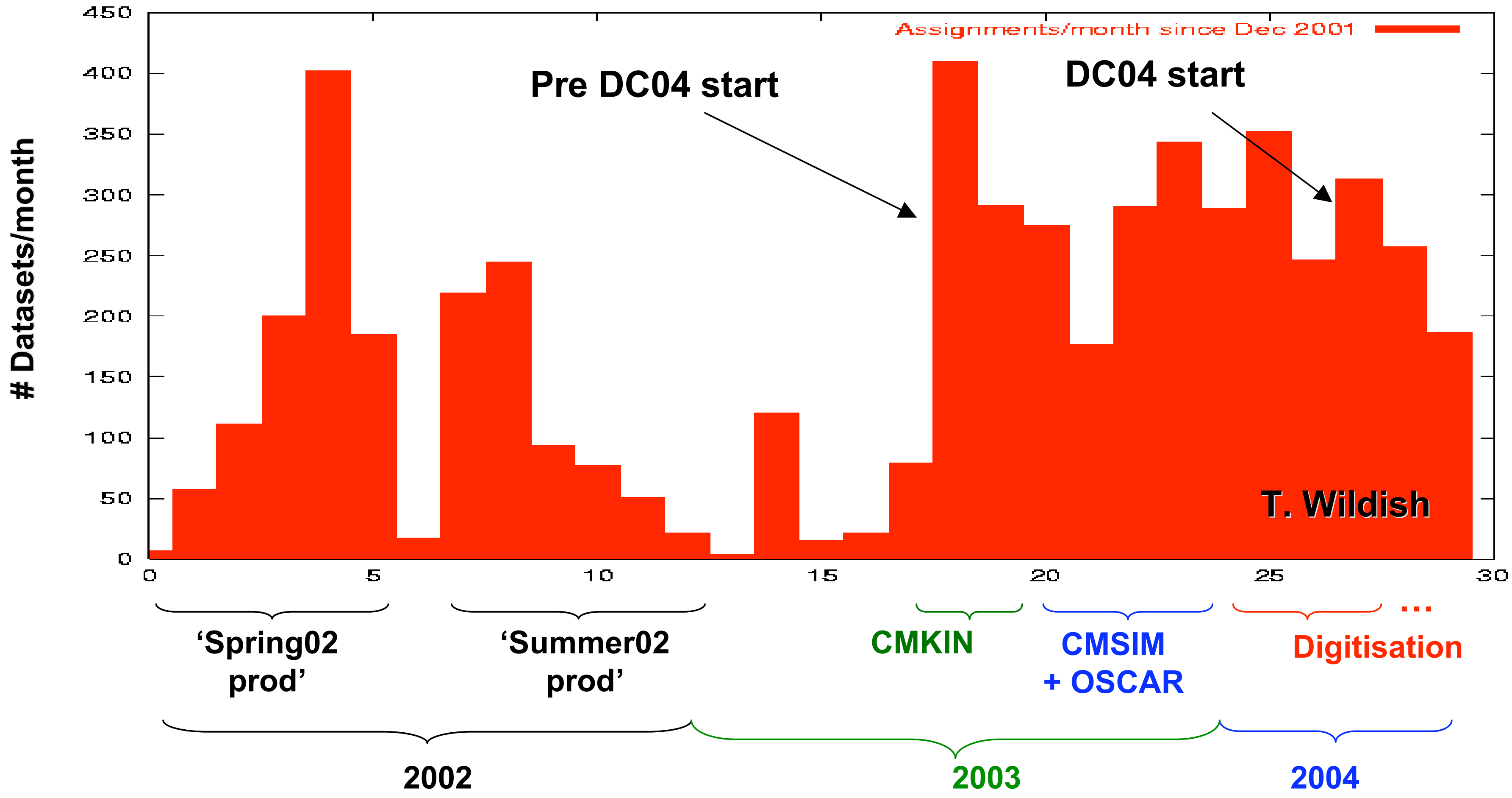
Successful Atlas DC2 Geant4 Jobs (20/7/04)



Atlas "Tiers" in DC2

Country	"Tier-1"	Sites	Grid	kSI2k
Australia			NG	12
Austria			LCG	7
Canada	TRIUMF	7	LCG	331
CERN	CERN	1	LCG	700
China				30
Czech Republic			LCG	25
France	CCIN2P3	1	LCG	~ 140
Germany	GridKa	3	LCG+NG	90
Greece			LCG	10
Israel		2	LCG	23
Italy	CNAF	5	LCG	200
Japan	Tokyo	1	LCG	127
Netherlands	NIKHEF	1	LCG	75
NorduGrid	NG	30	NG	380
Poland			LCG	80
Russia			LCG	~ 70
Slovakia			LCG	
Slovenia			NG	
Spain	PIC	4	LCG	50
Switzerland			LCG	18
Taiwan	ASTW	1	LCG	78
UK	RAL	8	LCG	~ 1000
US	BNL	28	Grid3/LCG	~ 1000
Total				~ 4500

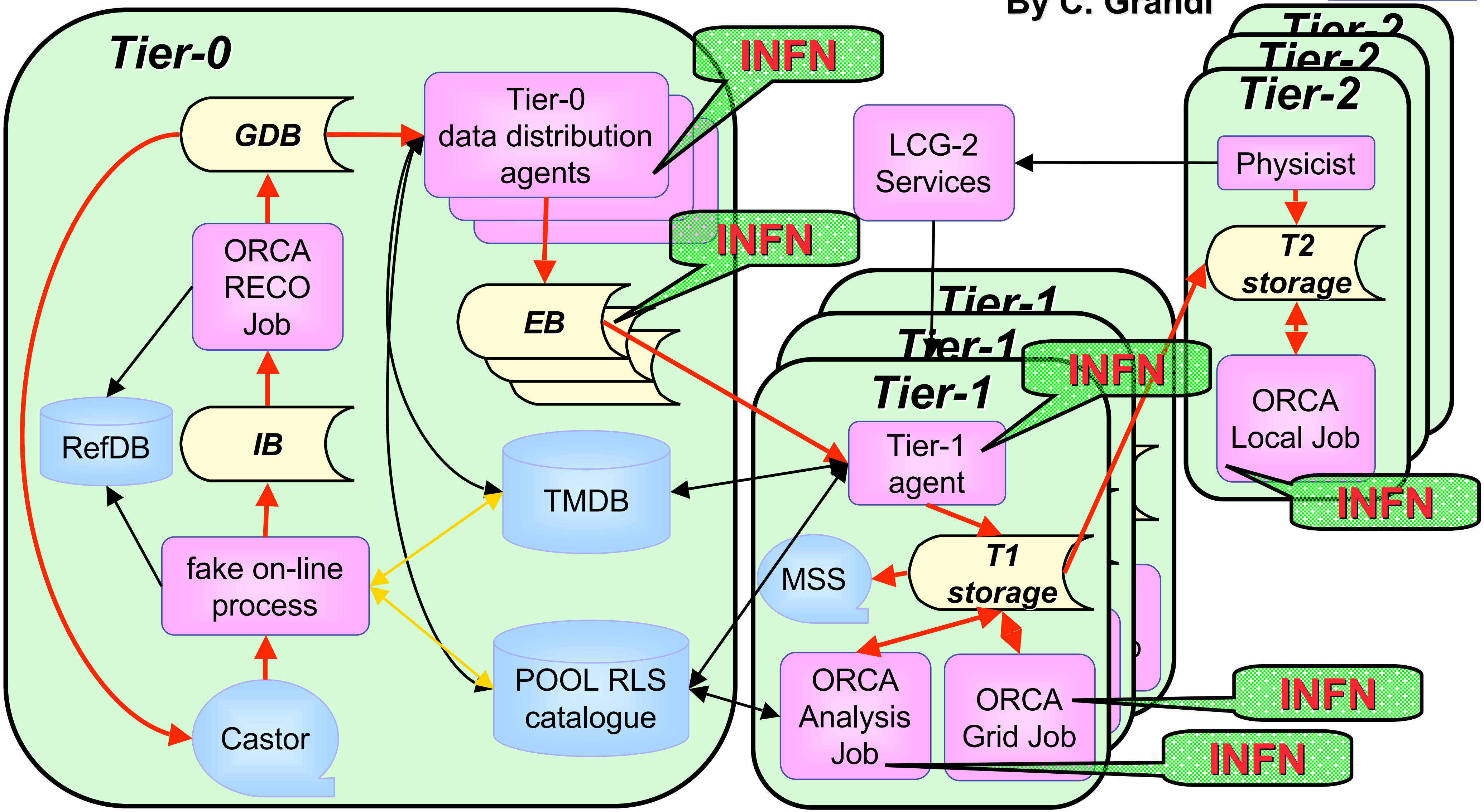
CMS 'permanent' production



The system is evolving into a *permanent* production effort…

CMS Data Challenge 04: layout

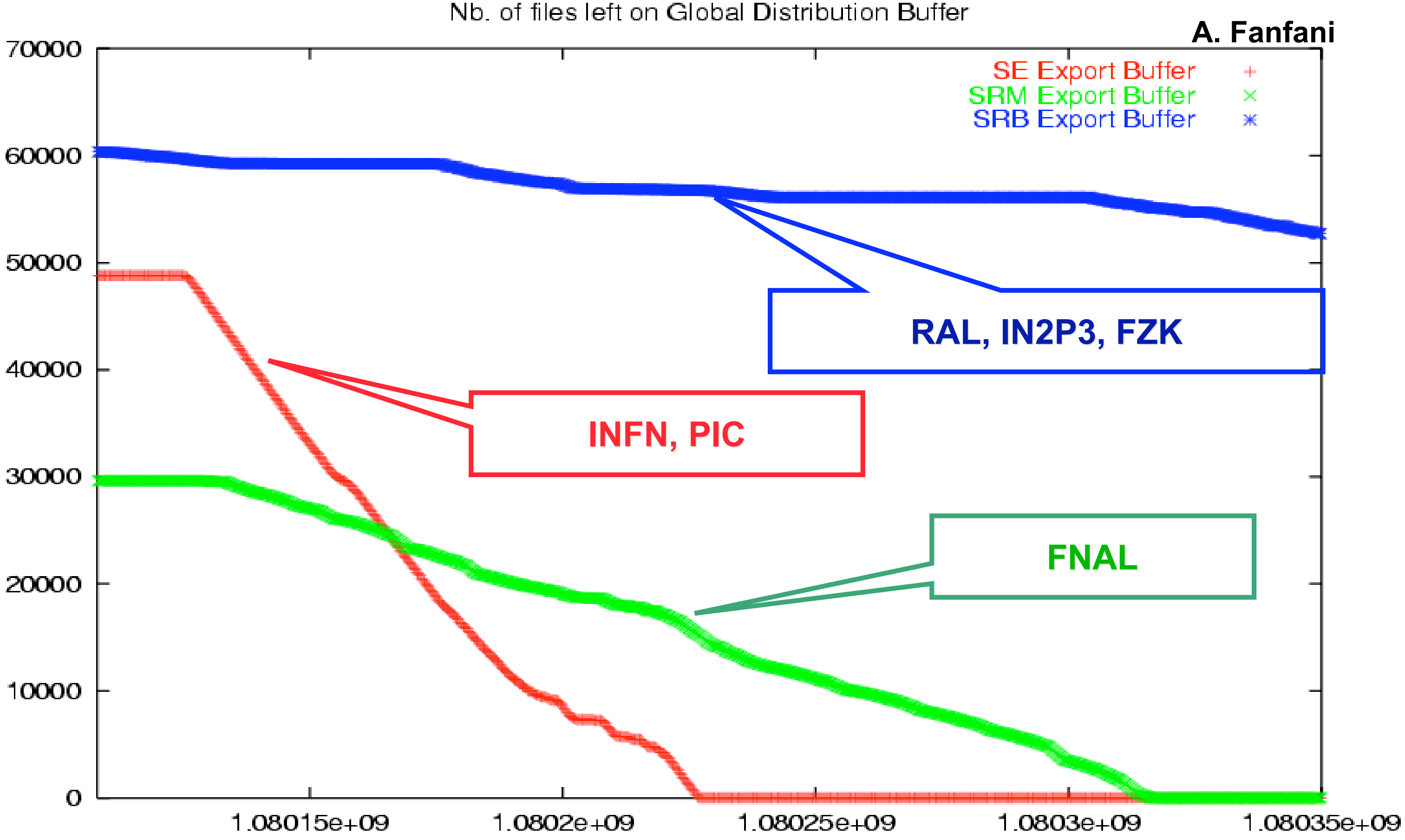
By C. Grandi



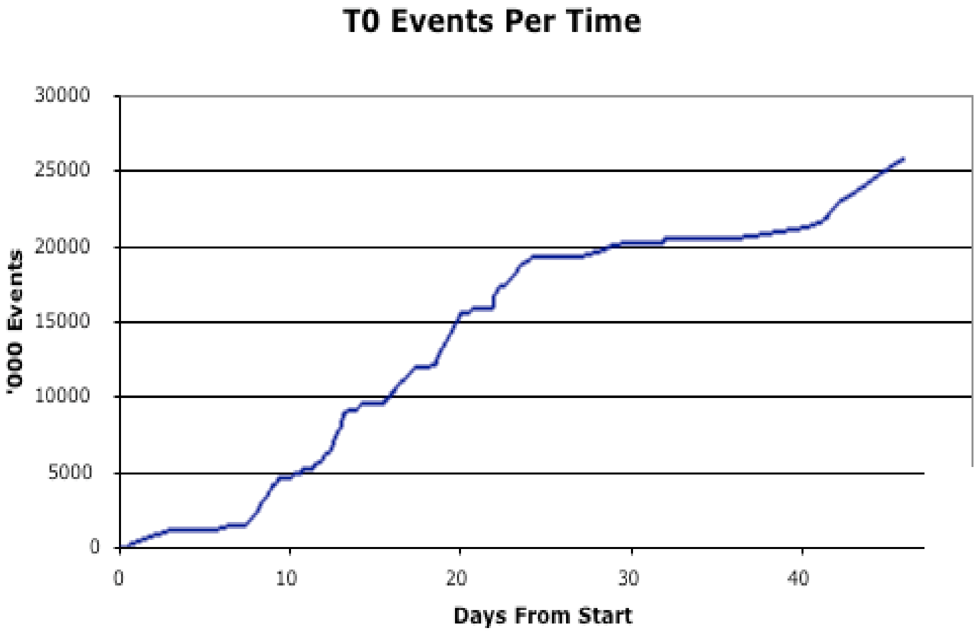
Full chain (but the Tier-0 reconstruction) done in LCG-2, but only for INFN and PIC

Not without pain...

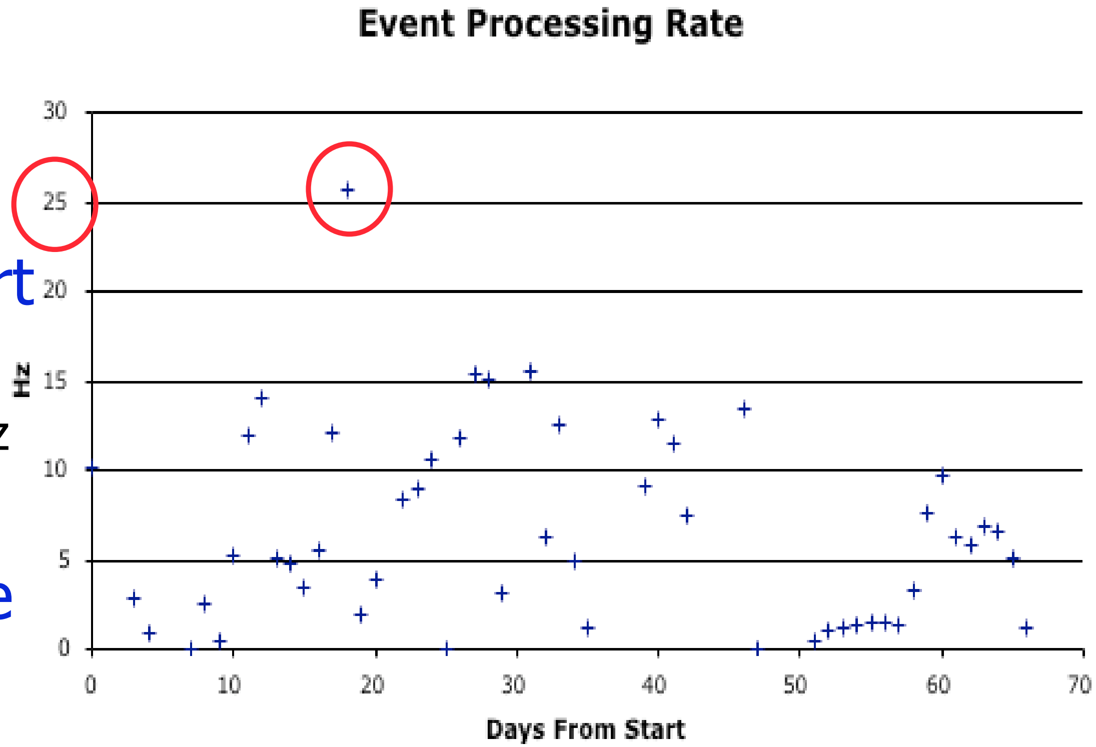
30 Mar 04 – Rates from GDB to EBs



Data Challenge 04 Processing Rate



- ❖Processed about 30M events
  - ◆ But DST "errors" make this pass not useful for analysis
- ❖Generally kept up at T1's in CNAF, FNAL, PIC



- ❖Got above 25Hz on many short occasions
  - ◆ But only one full day above 25Hz with full system
- ❖Working now to document the many different problems

What is missing?

- **In the LHC Experiments Computing Models**
  - Analysis primarily! It's still missing an estimate of the Worldwide load on resources.
- **In the Grid Projects**
  - Services stability
  - Design and architecture (components)
- **And how much time is still allowed for a "solution"?**
  - Not really much!

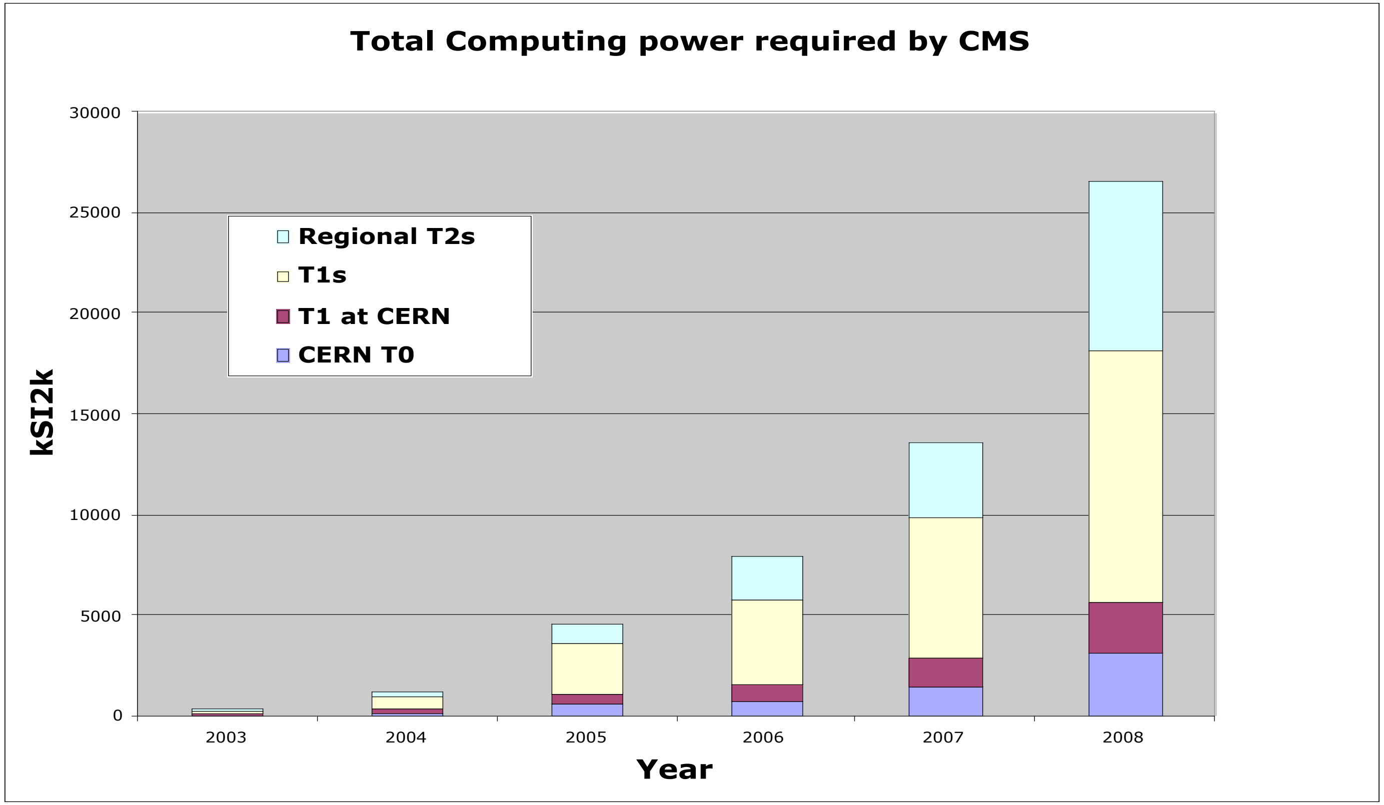
(Some CMS) Guiding Principles for LHC Computing

- Access to Data is more of a bottleneck than access to CPU
  - Make multiple distributed copies as early as possible
- Experiment needs to be able to enact Priority Policy
  - Stream data from Raw onwards
    - Some overlap allowed
  - Partition CPU according to experiment priorities
- Initial detailed analysis steps will be run at the T1's
  - Need access to large data samples
- T2's have (by definition?) more limited Disk/Network than the T1's
  - Good for final analysis, small (TB) samples
    - Make sure there is rapid access to locally replicate these
  - Perfect for Monte-Carlo Production
- User Analysis tasks are equal in magnitude to Production tasks
  - 50% Resources for each
  - Self correcting fraction
    - (When it gets to big strong motivation to make the user task a common production task)

Scheduled Computing

◆Organized, Scheduled, Simulation and Large-Scale Event Reconstruction is a task we understand "well"

- We can make reasonably accurate estimates of the computing required
- We can perform simple optimizations to share the work between the large computing centers



Chaotic Computing

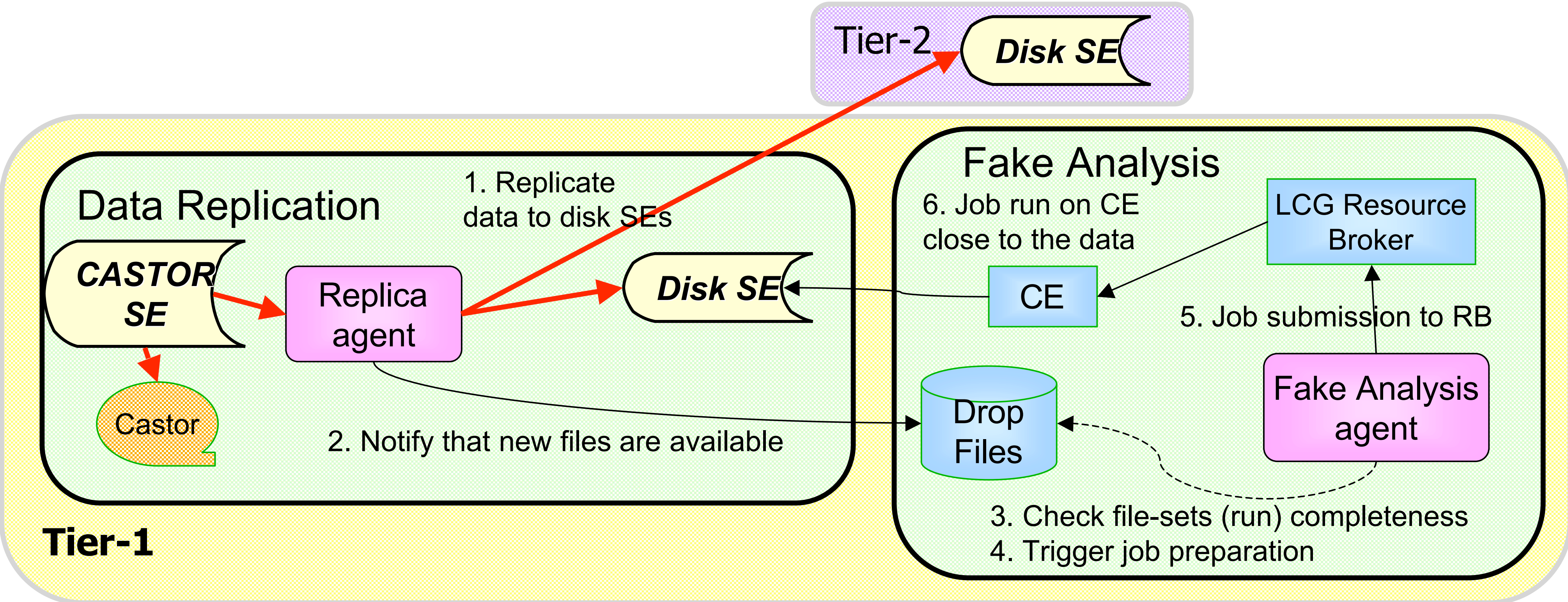
- Data Analysis is a "Feeding Frenzy"
  - Data is widely dispersed, may be geographically mismatched to available CPU
  - Choosing between data and job movement?
    - How/When will we have the information to motivate those choices?

- Move Data to Job
  - Moving only those parts of the data that the user really needs
    - All of some events, or some parts of some events?
      Very different resource requirements
  - Web-Services/ Web-Caching may be the right technologies here

- Move Job to Data
  - Information required to describe the data requirements can (will) be complex and poorly described
    - Difficult for a resource broker to make good scheduling choices
    - Current Resource Brokers are quite primitive

- Balancing the many priorities internal to an experiment is essential
  - Completing the a-priori defined critical physics as quickly and correctly as possible
  - Enabling the collaboration to explore the full Physics richness
- Build a Flexible System, Avoid Optimizations now (2004)

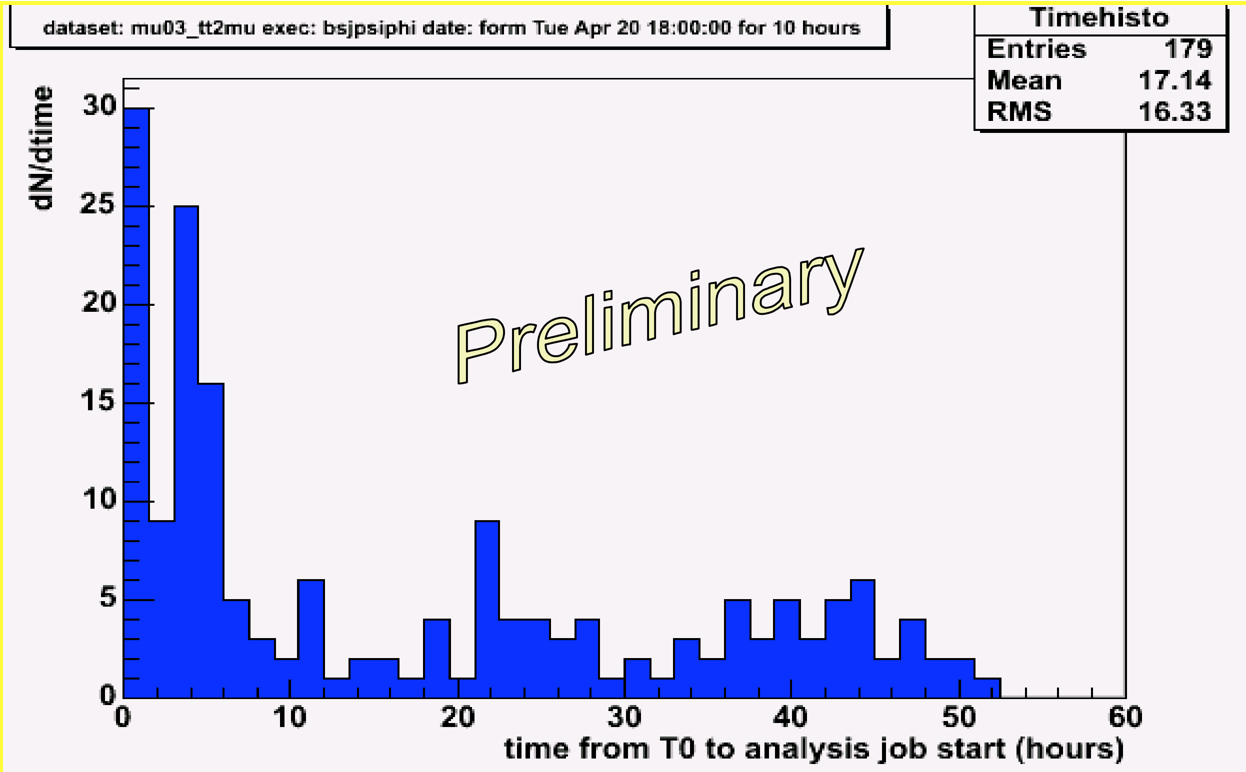
CMS Real-time Analysis Architecture



- Replication Agent make data available for analysis (on disk) and notify that
- Fake Analysis agent:
  - trigger job preparation when all files of a given file set are available
  - job submission to the LCG Resource Broker

CMS Real-time DC04 analysis: Turn-around time from T0

- ❑ The minimum time from T0 to T1 analysis was 10 minutes
- ❑ Different problems contributed to the time spread:

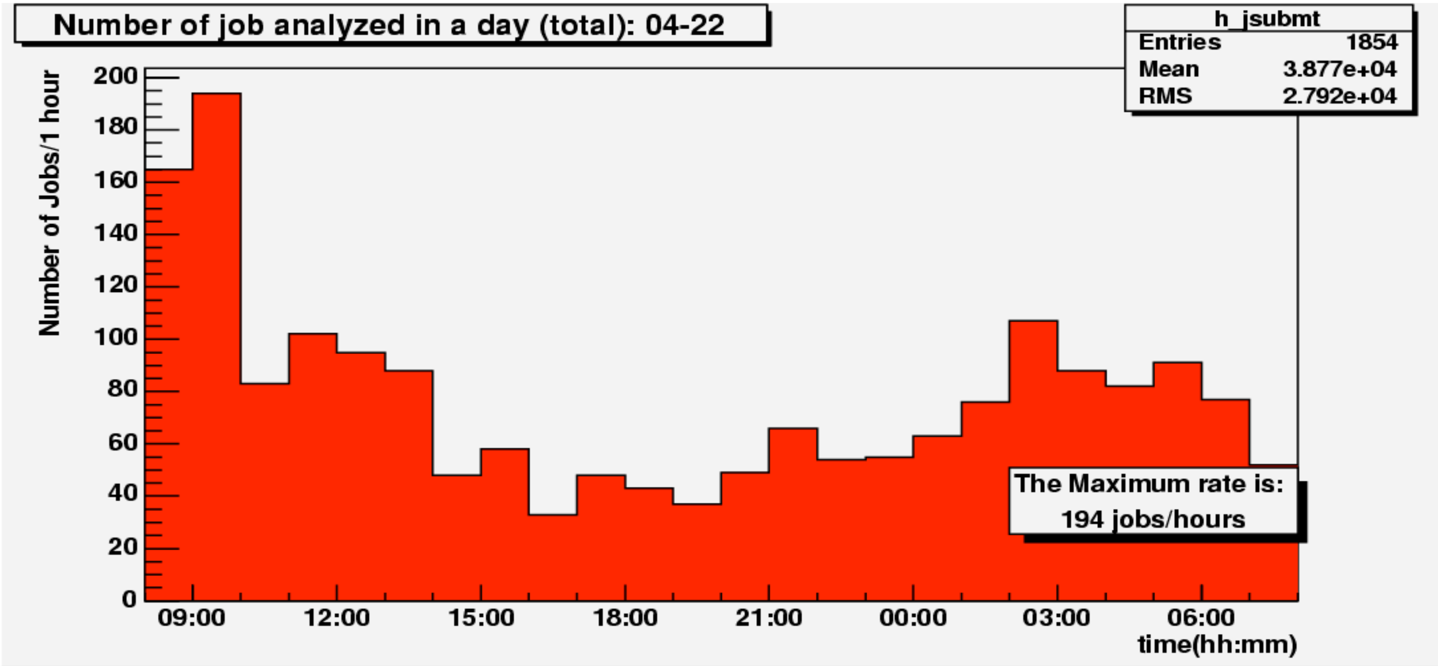


- the dataset-oriented analysis made the results dependent on which dataset were sent in real time from CERN
- Tuning of the Tier-1 Replica Agent
- Replica Agent operation affected by CASTOR problem
- Analysis Agents were not always up due to debugging
- for 1 dataset Zipped Metadata were late with respect to data
- few problems with submission

N. De Filippis, A. Fanfani, F. Fanzago

CMS DC04 Real-time Analysis

- Maximum rate of analysis jobs: **194 jobs/hour**
- Maximum rate of analysed events: **26 Hz**
- Total of **~15000** analysis jobs via **Grid** tools in ~2 weeks **(95-99% efficiency)**



- Datasets examples:
  - $B^0_s \rightarrow J/\psi\ \varphi$

    Bkg: mu03_tt2mu, mu03_DY2mu
  - tTH, $H \rightarrow$ bbbar $t \rightarrow Wb$ $W \rightarrow l\nu$ $T \rightarrow Wb$ $W \rightarrow$ had.

    Bkg: bt03_ttbb_tth

    Bkg: bt03_qcd170_tth

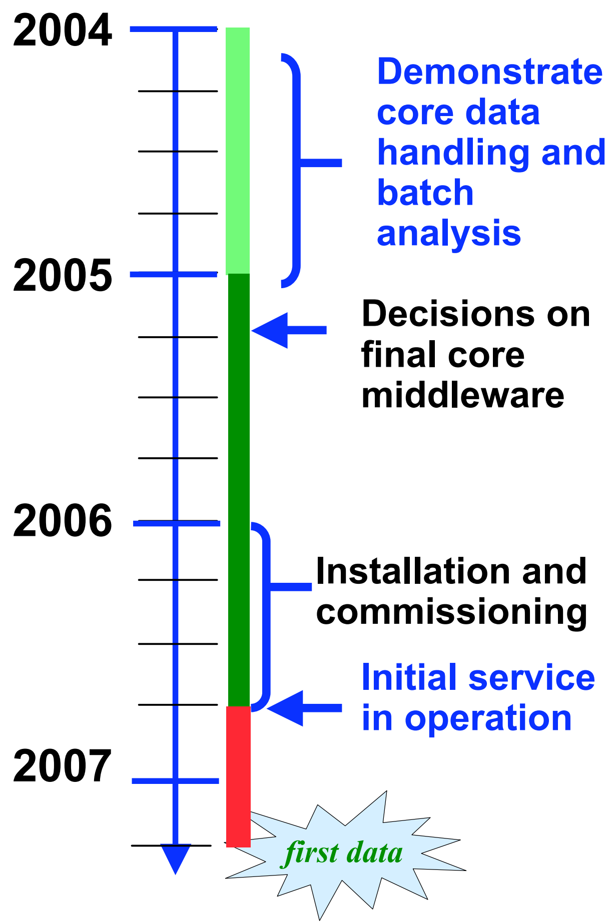
    Bkg: mu03_W1mu
  - $H \rightarrow WW \rightarrow 2\mu\ 2\nu$

    Bkg: mu03_tt2mu, mu03_DY2mu

N. De Filippis, A. Fanfani, F. Fanzago

LCG Timescale

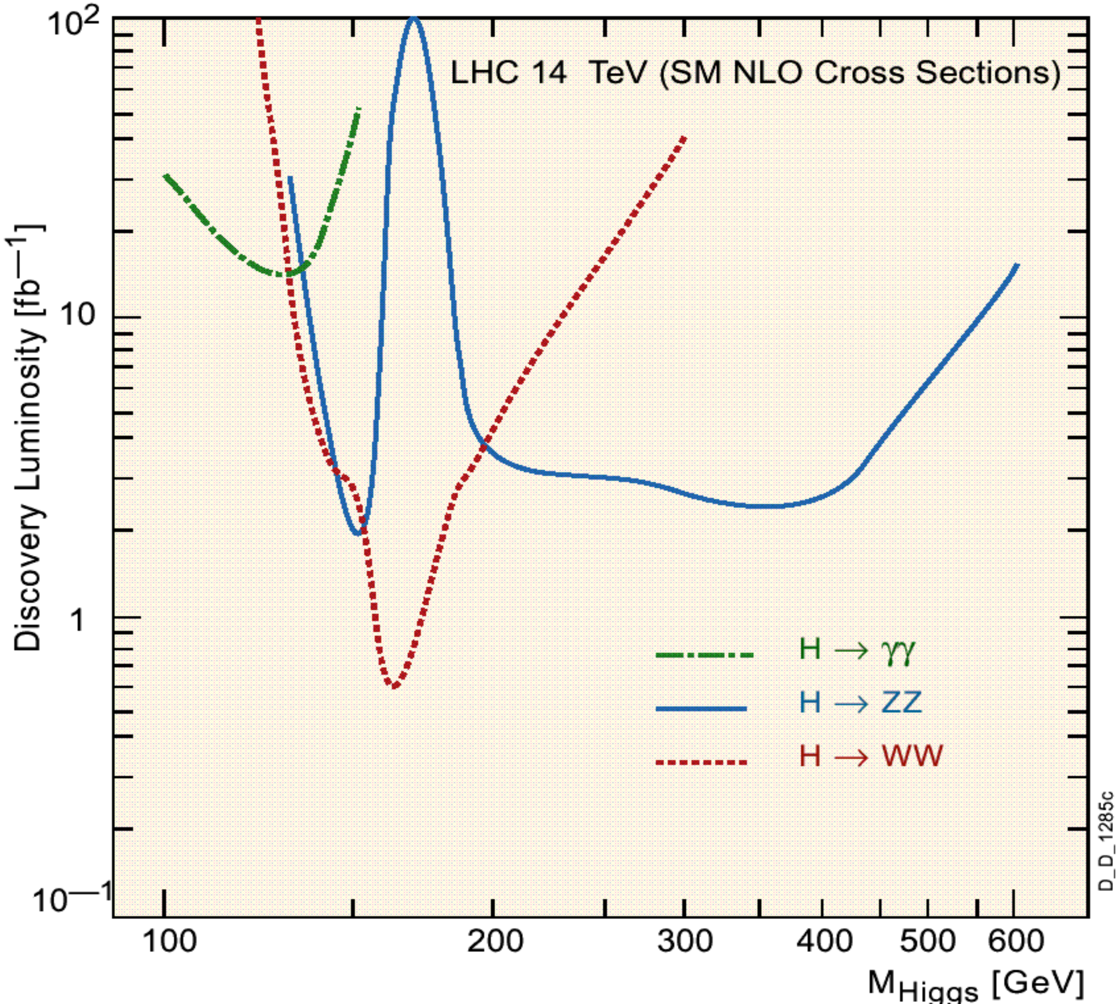
- Still early days for operational grids
- There are still many questions about grids & data handling
- EGEE provides LCG with opportunities -
  - to develop an operational grid in an international multi-science context
  - to influence the evolution of a generic middleware package
  - maybe leading to a general science grid infrastructure
- But the LHC clock is ticking - deadlines will dictate simplicity and pragmatism
- LCG has long-term requirements – and at present EGEE is a two-year project
- LCG must encompass non-European resources and grids



- No shortage of challenges and opportunities

The Goal is the Physics, not the Computing…

- Motivation: at $L_0=10^{33}$ $cm^{-2}s^{-1}$,
  - 1 fill (6hrs) ~ 13 pb^{-1}
  - 1 day ~ 30 pb^{-1}
  - 1 month ~ 1 fb^{-1}
  - 1 year ~ 10 fb^{-1}



- Most of Standard-Model Higgs can be probed within a few months
  - Ditto for SUSY

- Turn-on for detector + computing and software will be crucial

Conclusion

- **A lot still to do!**
  - Quickly !!! Very quickly.
  - But also long term solutions and ideas are needed and welcomed
    - There's a lot of room for them
    - Not only for High Energy Physics, even if it's driving the effort
- **Need your help !**
  - I'll not be there, but hopefully elsewhere

Links and references

- http://public.web.cern.ch/public/
- http://public.web.cern.ch/public/about/aboutCERN.html
- http://lhc-new-homepage.web.cern.ch/lhc-new-homepage/
- http://alice.web.cern.ch/Alice/AliceNew/
- http://atlas.web.cern.ch/Atlas/
- http://cmsinfo.cern.ch/Welcome.html/
- http://lhcb.web.cern.ch/lhcb/
- http://lcg.web.cern.ch/LCG/
- http://egee-intranet.web.cern.ch/egee-intranet/gateway.html
- http://www.ivdgl.org/grid2003/
- http://grid-it.cnaf.infn.it/
- http://grid.infn.it/