LHC Grid Deployment Board

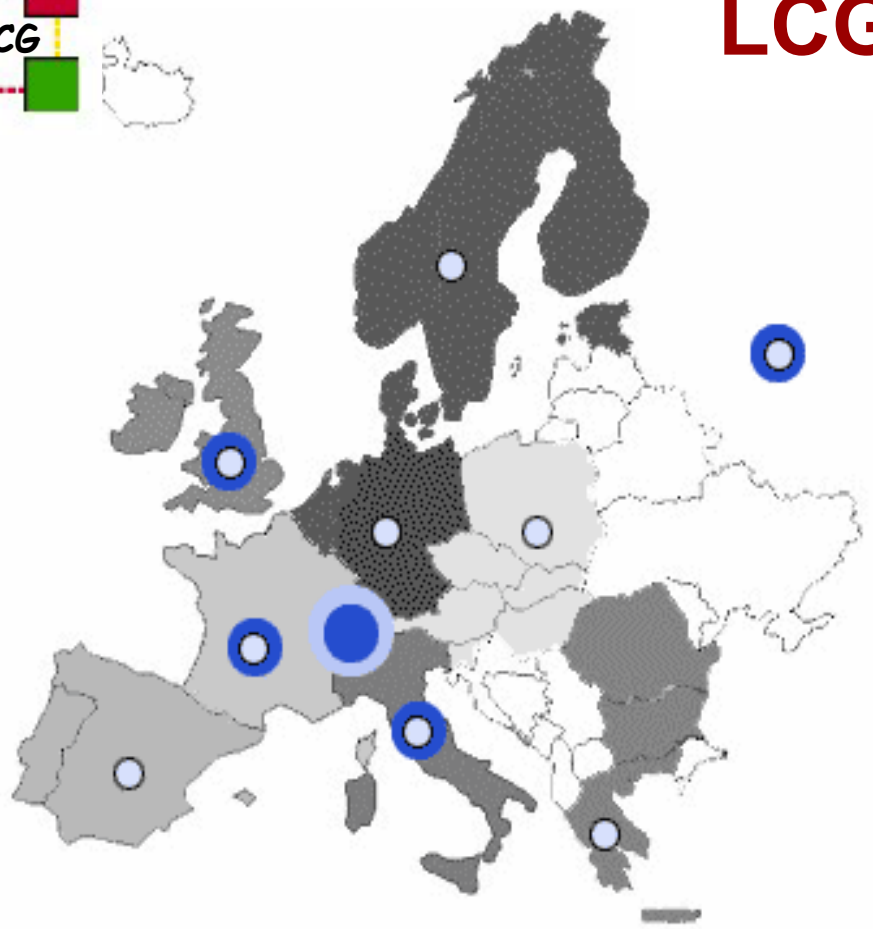
Regional Centers
Phase II Resource Planning
Service Challenges

LHCC Comprehensive Review
22-23 November 2004

Kors Bos, GDB Chair
NIKHEF, Amsterdam

Regional Centers

LCG → EGEE in Europe



- EGEE is hierarchical organised
- CERN & 4 countries & 4 Federations
- For HEP and non-HEP applications

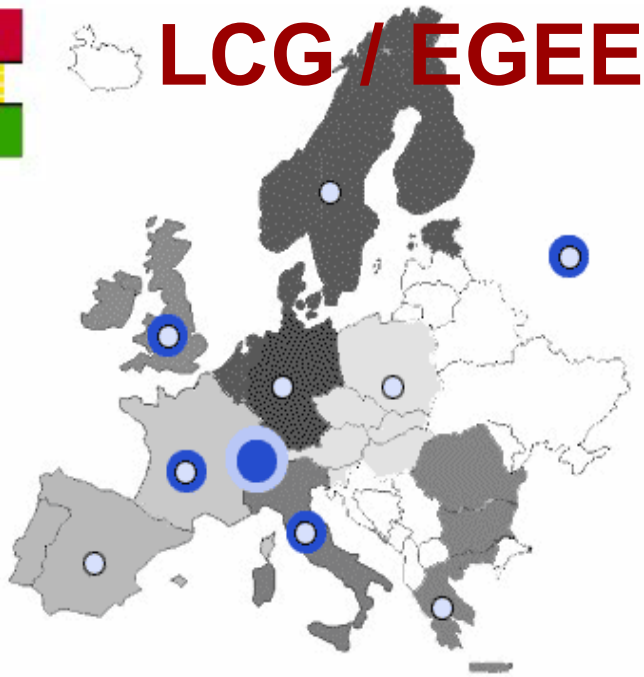
Operations Management Centre

Core Infrastructure Centre

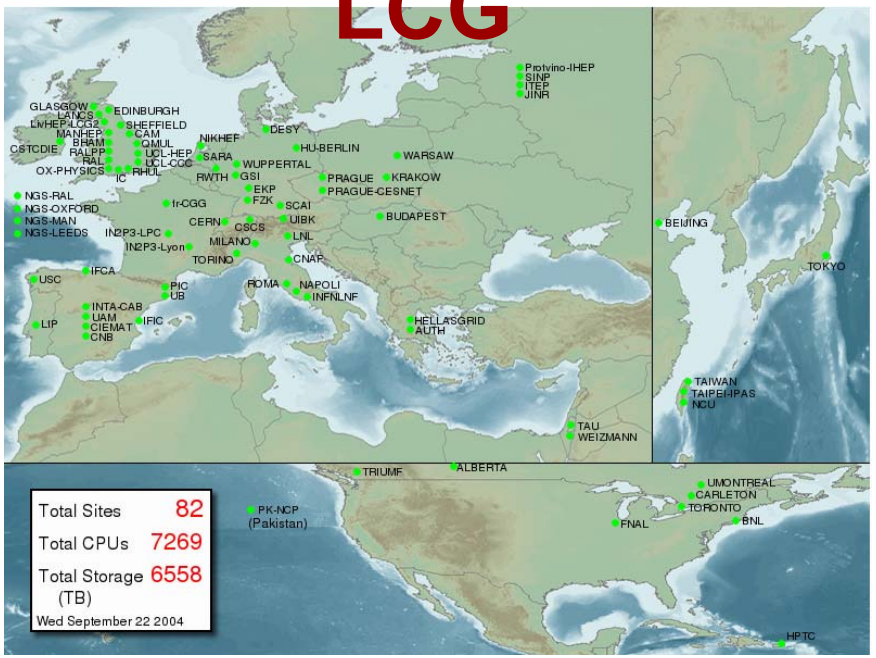
Regional Operations Centre

- *OMC*
  - Only CERN
- *CIC*
  - CERN, RAL, CCIN2P3, CNAF
  - *Provide central grid services like VO, Monitoring, Accounting, RB, BDII, etc*
- *ROC*
  - *RAL, CCIN2P3, CNAF, Stockholm & SARA Amsterdam, FZK Karlsruhe, Athens, PIC Barcelona, Moscow*
  - *Responsible for a region*

LCG / EGEE



LCG



- **OMC**
  - Only CERN
- **CIC 4**
  - CERN, RAL, CCIN2P3, CNAF
  - *Provide central grid services like VO, Monitoring, Accounting, RB, BDII, etc*
- **ROC 9**
  - *Stockholm, Amsterdam, Karlsruhe, Athens, Barcelona, Lyon, Bologna, Moscow, Didcot*
  - *Responsible for a region*

- **Tier-0**
  - Only CERN
- **Tier-1 ~10**
  - RAL,CCIN2P3, CNAF, GridKa, NL, Nordic, PIC, BNL, FNAL, Triumf, ASCC
  - *Provide central grid services like VO, Monitoring, Accounting, RB, BDII, etc*
  - *Data archive, re-processing*
- **Tier-2 ~100**
  - *No data archive, Monte Carlo, analysis*
  - *Depending on a Tier-1*

Regional Centers & LCG Tier-1 Sites

				ALICE	ATLAS	CMS	LHCb	
1	GridKa	Karlsruhe	Germany	X	X	X	X	4
2	CCIN2P3	Lyon	France	X	X	X	X	4
3	CNAF	Bologna	Italy	X	X	X	X	4
4	NIKHEF/SARA	Amsterdam	Netherlands	X	X		X	3
5	Nordic	Distributed	Dk, No, Fi, Se		X			1
6	PIC	Barcelona	Spain		X	X	X	3
7	RAL	Didcot	UK	X	X	X	X	4
8	Triumf	Vancouver	Canada		X			1
9	BNL	Brookhaven	US		X			1
10	FNAL	Batavia, Ill.	US			X		1
11	ASCC	Taipei	Taiwan		X	X		2
				5	10	7	6	28

Grid Deployment Board

- National representation of countries in LCG
  - *Doesn't follow T0/1/2 or EGEE hierarchy*
  - *Reps from all countries with T1 centers*
  - *Reps from countries with T2 centers but no T1's*
  - *Reps from LHC experiments (comp. coordinators)*
- Meets every month
  - *Normally at CERN (twice a year outside CERN)*
- Reports to the LCG Proj.Exec.Board
- Standing working groups for
  - *Security (same group also serves LCG and EGEE)*
  - *Software Installation Tools (Quattor)*
  - *Network Coordination (not yet)*
- Only official way for centers to influence LCG
- Plays an important role in Data and Service Challenges

Phase 2 Resources in Regional Centers

Phase 2 Planning Group

- Ad hoc group to discuss LCG resources (3/04)
- Expanded to include representatives from major T1 and T2 centres and experiments and project management)
- Not quite clear what Phase 2 is: up to start-of-LHC (?)
- Collected resource planning data from most T1 centres for Phase 2
- But very little information available yet for T2 resources
- Probable/possible breakdown of resources between experiments is not yet available from all sites – essential to complete the planning round
- Fair uncertainty in all numbers
- Regular meetings and reports

Preliminary Tier-1 planning data

Experiment requirements and models still under development
Two potential Tier-1 centres missing

Total resources required and planned in all Tier-1 Centres (except CERN)

First full year of data taking (2008)

All data is preliminary

Resource type	***Estimated requirements*** *(note 1)*					***Total resources planned at Tier-1 centres*** *(note 2)*	***Balance***
	ALICE	***ATLAS***	***CMS***	***LHCb***	***Total***		
CPU (MSI2K)	9.1	16.6	12.6	9.5	47.8	**44.7**	**- 6%**
Disk (PBytes)	3.0	9.2	8.7	1.3	22.2	**10.6**	**- 52%**
Tape (PBytes)	3.6	6	6.6	0.4	16.6	**19.8**	**19%**

Notes

1. Requirements will be reviewed by LHCC in January 2005
2. Current planning includes estimates of resources for which funding has not yet been secured

Some conclusions

- The T1 centers will have the cpu resources for their tasks
- Variation of a factor 5 in cpu power of T1 centers
- Not clear how much cpu resources will be in the T2 centers

- Shortage of disk space at T1's compared to what experiments expect

- Enough resources for data archiving but not clear what archive speed is needed and if this will be met

- Network technology is there but coordination is needed
- End-to-end service is more important than just bandwidth

Further Planning for Phase 2

- end November 2004: complete Tier-1 resource plan
- first quarter 2005:
  - *assemble resource planning data for major Tier-2s*
  - *understand the probable Tier-2/Tier-1 relationships*
  - *initial plan for Tier-0/1/2 networking*
- Developing a plan for ramping up the services for Phase 2
  - *set milestones for the Service Challenges*
  - *See slides on Service Challenges*
- Develop a plan for Experiment Computing Challenges
  - *checking out the computing model and the software readiness*
  - ***not** linked to **experiment data challenges** – which should use the regular, permanent grid service!*
- TDR editorial board established → TDR due July 2005

Service Challenge for Robust File Transfer

Service Challenges

- Expts ←→Tier-0 ←→ Tier-1 ← →Tier-2 is a complex engine
- Experiments DCs mainly driven by production of many MC events
- Distributed computing better tested than data distribution
- Not well tested:
  - Tier 0/1/2 model
  - Data distribution
  - Security
  - Operations and support

- Specific service Challenge for
  - Robust file transfer ⇐ **I will now only talk about this**
  - Security
  - Operations
  - User Support

Example: ATLAS

- Trigger rate = 200 Hz
- Event size = 1.6 MByte
- 10 T1 centers

} To each T1 32 MByte/sec = 256 Mbit/sec

Result of first pass reconstruction is called ESD data

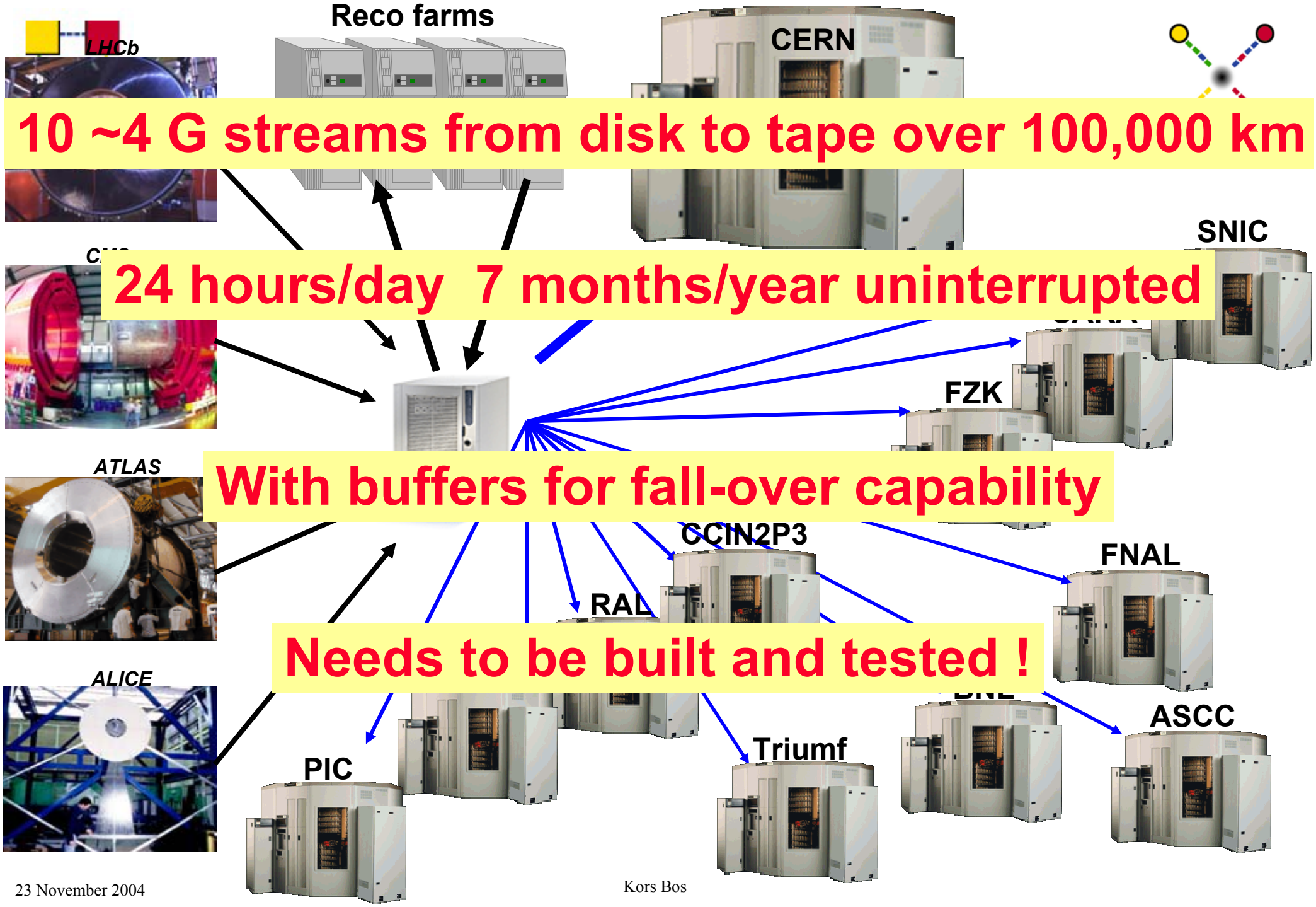
- ESD = 0.5 MByte
- Trigger rate = 200 Hz
- 2 copies at T1's

} To each T1 20 MByte/sec = 180 Mbit/sec

More refined datasets AOD and TAG add another few %
Total for ATLAS: To each T1 ~500 Mbit/sec

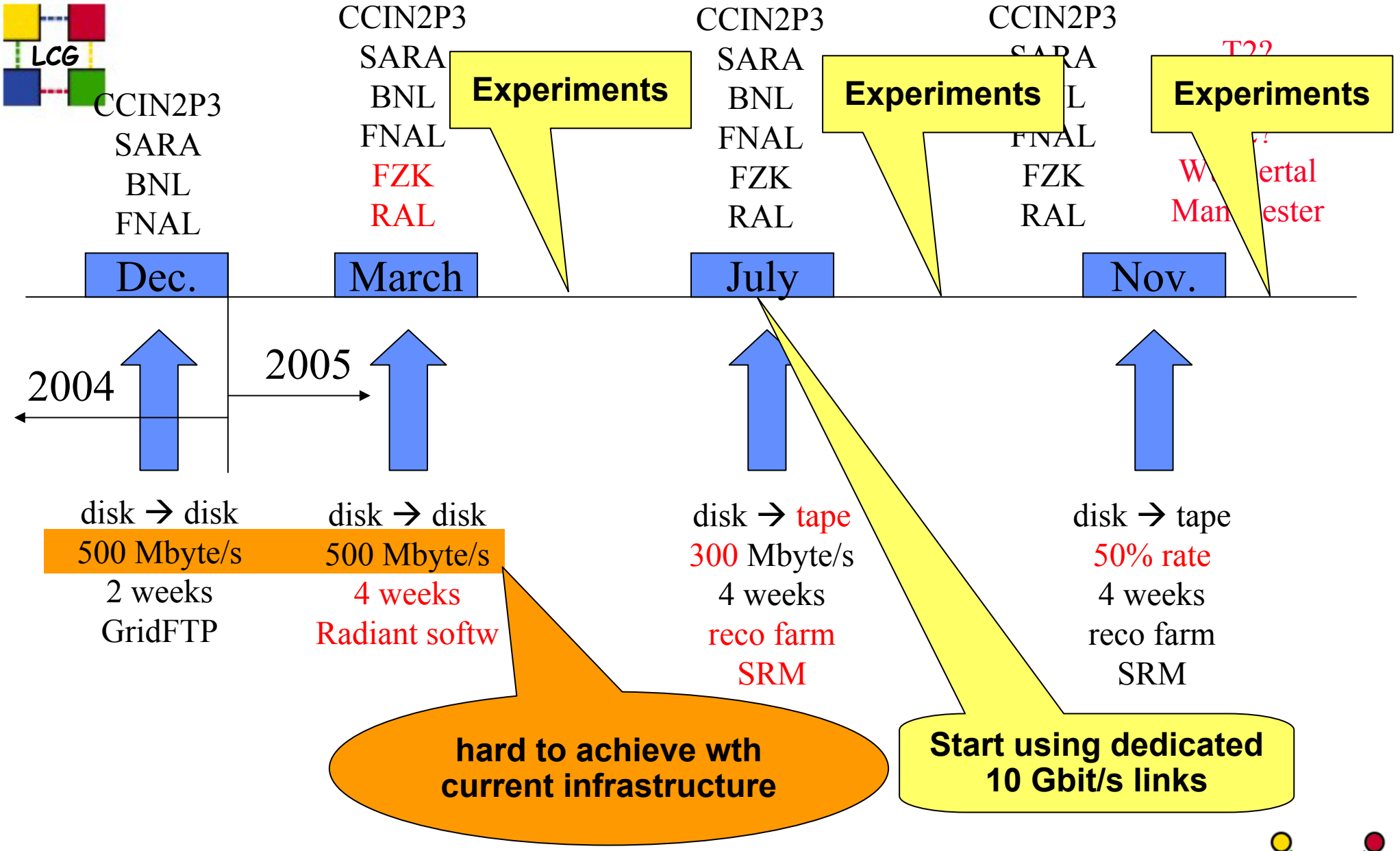
NB1 other experiments have fewer T1's
NB2 not all T1's support 4 experiments
NB3 Alice events are much bigger, but runs are shorter
NB4 Monte Carlo, Re-Processing and Analysis not mentioned

Conclusion 1 ~10 Gbit/sec network needed between CERN and all T1's

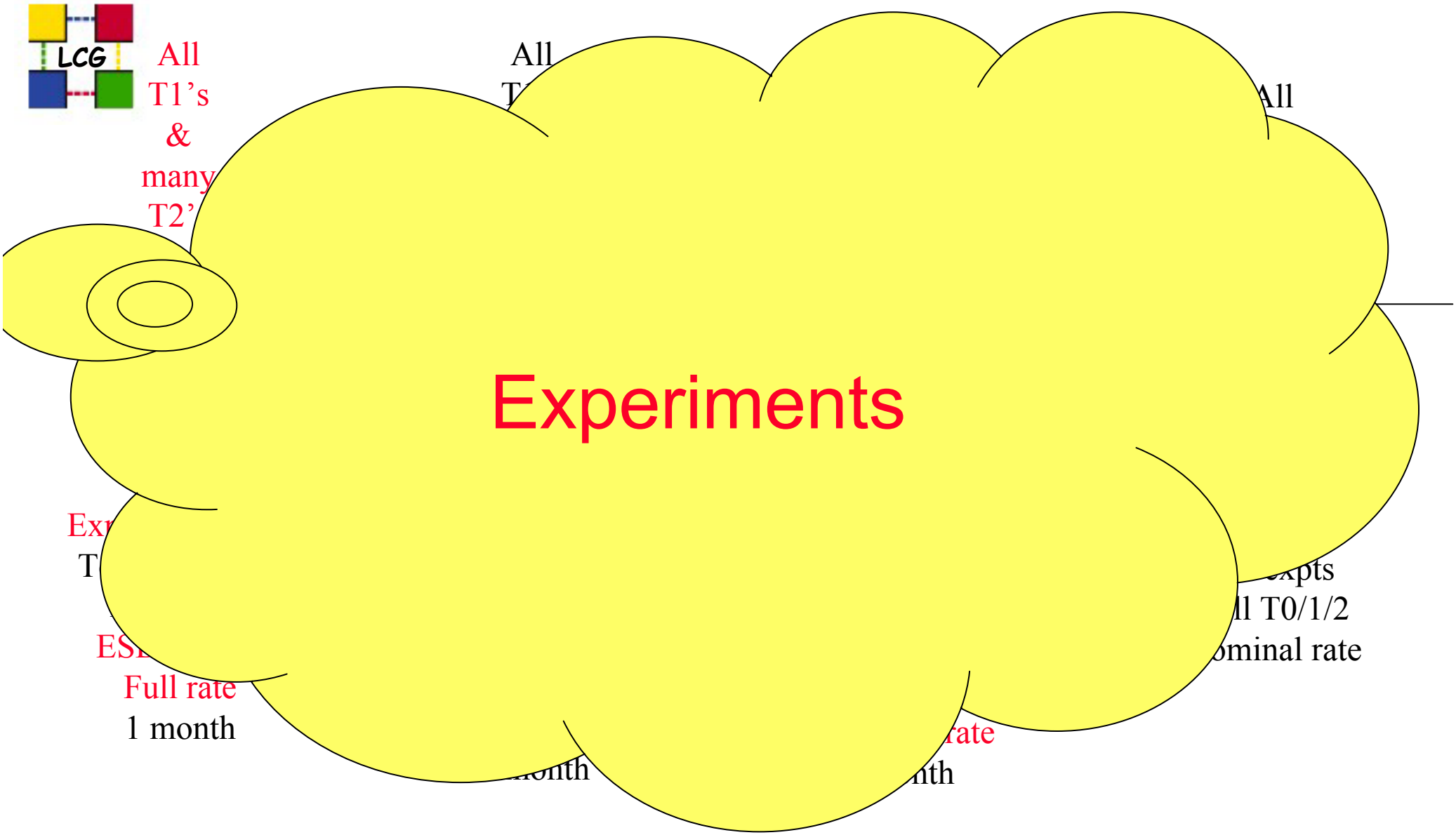

LHCb
Reco farms
CERN
10 ~4 G streams from disk to tape over 100,000 km
SNIC
24 hours/day 7 months/year uninterrupted
FZK
ATLAS
With buffers for fall-over capability
CCIN2P3
RAL
FNAL
ALICE
Needs to be built and tested !
PIC
Triumf
ASCC

Principles for Service Challenges

- Not a network bandwidth race
  - *In 2004 10 Gbit/sec has already been proven to be possible*
- International network topology is important
  - All T0-T1 dedicated links, dark fibers, redundancy, coordination
- End-to-end application: from the exp. DAQ to remote tape robot
  - *Progress to be made in steps by adding more components each step*
- Sustainability is a challenge
  - *24 hours/day for 7 months in a row*
- Redundancy and fall-over capability
  - *Data buffers for non-stop operation*
  - *If one site fails other sites must be able to take more*
- Performance must include grid software
  - *Not only bare GridFTP but SRM and Catalogs*
- Performance must include experiment specific hard/soft/people-ware
  - *Concentrate on generic issues first*
- Schedule/synchronise related service and computing challenges
  - *Must be able to all run concurrently*



Milestones 1-4



Milestones 5-8

Planning for Service Challenges

Role of GDB

- *Planning and coordination*
- *Monthly reporting*
- *Network coordination: dedicated GDB working group*

Service challenge meetings:

- *Oct 12 2004 -- Amsterdam*
- *Dec 2 2004 -- GridKa, Karlsruhe*
- *Jan 2005 – RAL, Abingdon*
- *March 2005 – CCIN2P3, Lyon*
- *April 2005– ASTW, Taipei*

Dedicated Network Coordination Meeting

- *Jan 11-12 – CERN: T1 reps + NRENs*

Milestones Document : end January 2005

END

Role of Tier-1 Centers

- Archive of raw data
  - *1/n fraction of the data for experiment A where n is the number of T1 centers supporting experiment A*
  - *Or an otherwise agreed fraction (MoU)*
- Archive reconstructed data (ESD, etc)
- Large disk space for keeping raw and derived data
- Regularly re-processing of the raw data and storing new versions of the derived data
- Operations coordination for a region (T2 centers)
- Support coordination for a region
- Archiving of data from the T2 centers in its region

Tier-2 Centers

- Unclear how many there will be
  - *Less than 100, depends on definition*
- Role for T2 centers:
  - *Data analysis*
  - *Monte Carlo simulation*
- In principle no data archiving
  - *no raw data archiving*
  - *Possibly derived data or MC data archiving*
- Resides in a region with a T1 center
  - *Not clear to what extend this picture holds*
  - *A well working grid doesn't have much hierarchy*

2004 Achievements for T0 → T1 Services

- Introduced at May 2004 GDB and HEPIX meeting
- Oct.5 2004 – PEB concluded
  - *Must be ready 6 months before data arrives: early 2007*
  - *Close relationship between service & experiment challenges*
    - *include experiment people in the service challenge team*
    - *use a.m.a.p. real applications – even if in canned form*
    - *experiment challenges are computing challenges – treat data challenges that physics groups depend on separately*
- Oct 12 2004 – service challenge meeting in Amsterdam
- Planned service challenge meetings:
  - *Dec 2 2004 GridKa, Karlsruhe*
  - *Jan 2005 – RAL, Abingdon*
  - *March 2005 – CCIN2P3, Lyon*
  - *April 2005– ASTW, Taipei*
- First Generation Hardware and Software in place at CERN
- Data transfers have started to Lyon, Amsterdam, Brookhaven and Chicago

Milestone I & II Proposal
Service Challenge 2004/2005

Dec04 - Service Challenge I complete

- *mass store (disk) -mass store (disk)*
- *3 T1s (Lyon, Amsterdam, Chicago)*
- *500 MB/sec (individually and aggregate) ← difficult !*
- *2 weeks sustained ← 18 December shutdown !*
- *Software; GridFTP plus some macro's*

Mar05 - Service Challenge II complete

- *Software: reliable file transfer service*
- *mass store (disk) - mass store (disk),*
- *5 T1's (also Karlsruhe, RAL, ..)*
- *500 MB/sec T0-T1 but also between T1's*
- *1 month sustained ← start mid February !*

Milestone III & IV Proposal
Service Challenge 2005

July05 - Service Challenge III complete

- *Data acquisition → disk pool → on tape at T0 and T1's*
- *Reconstruction Farm at CERN: ESD also to T1's*
- *Experiment involvement: DAQ, Reconstruction Software*
- *Software: real system software (SRM)*
- *5 T1s*
- *300 MB/sec including mass storage (disk and tape)*
- *1 month sustained: July !*

Nov05 - Service Challenge IV complete

- *ATLAS and/or CMS T1/T2 model verification*
- *At 50% of data rate T0→ T1 ← → T2 ← → T2*
- *Reconstruction scaled down to 2005 cpu capacity*
- *5 T1's and 5 T2's*
- *1 month sustained: November !*

Milestone V & VI Proposal
Service Challenge 2006

Apr06 - Service Challenge V complete

- *Data acquisition → disk pool → on tape at T0 and T1's*
- *Reconstruction Farm at CERN: ESD also to T1's*
- *ESD skimming, distribution to T1's and T2's*
- *Full target data rate*
- *Simulated traffic patterns*
- *To all T1s and T2's*
- *1 month sustained*

Aug06 - Service Challenge VI complete

- *All experiments (ALICE in proton mode)*
- *Full T0/1/2 model test*
- *100% nominal rate*
- *Reconstruction scaled down to 2006 cpu capacity*
- *1 month sustained*

Milestone VII & VIII Proposal
Service Challenge 2006/2007

Nov06 - Service Challenge VII complete

- ***Infrastructure ready*** *at T0 and all T1's and selected T2's*
- ***Twice the target data rates****, simulated traffic patterns*
- *1 month sustained T0/1/2 operation*

Feb07 - Service Challenge VIII complete

- *Ready for data taking*
- *All experiments*
- *Full T0/1/2 model test*
- *100% nominal rate*

Resources for Service Challenges

Cannot be achieved without significant investments in (initially)

- *Manpower: few fte per T1 and at CERN*
- *Hardware: dedicated data servers, disk space, network interfaces*
- *Software: SRM implementations*
- *Network: 10 Gb dedicated T0 – T1*

Role of GDB

- *Planning and coordination*
- *Monthly reporting*
- *Network coordination: dedicated GDB working group*

Concerns

- *T1 centers have not yet invested very much in it*
- *Also experiments have to take into their planning*
- *Dedicated network needs to be realised (coordination, finances, politics)*