



# Preparing the computing solutions for the Large Hadron Collider (LHC) at CERN

**Sverre Jarp, openlab CTO  
IT Department, CERN**

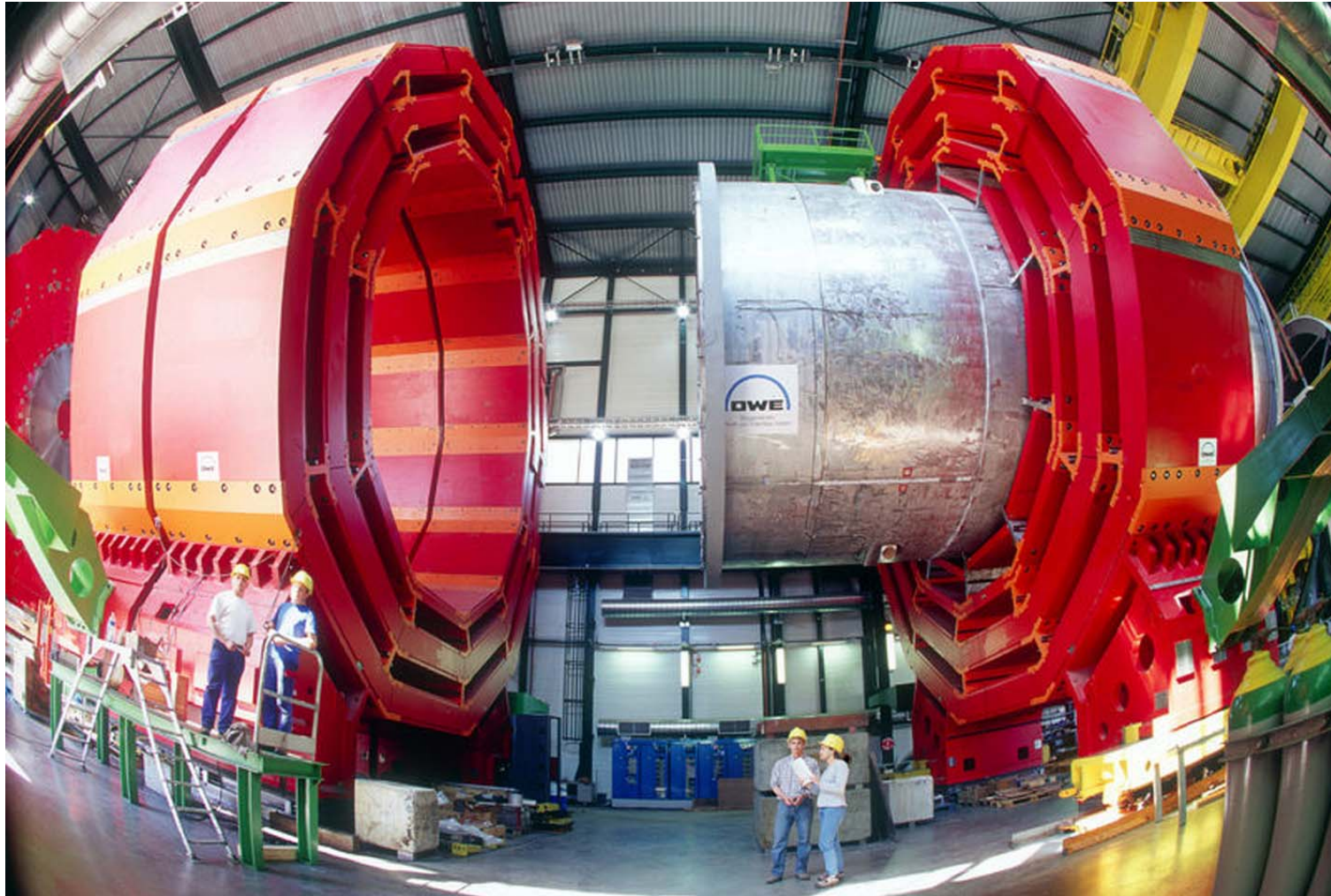
**Gelato meeting  
Champaign-Urbana (24.5.2004)**

**CERN**

Openlab for DataGrid applications



# The LHC Challenge



May 2004

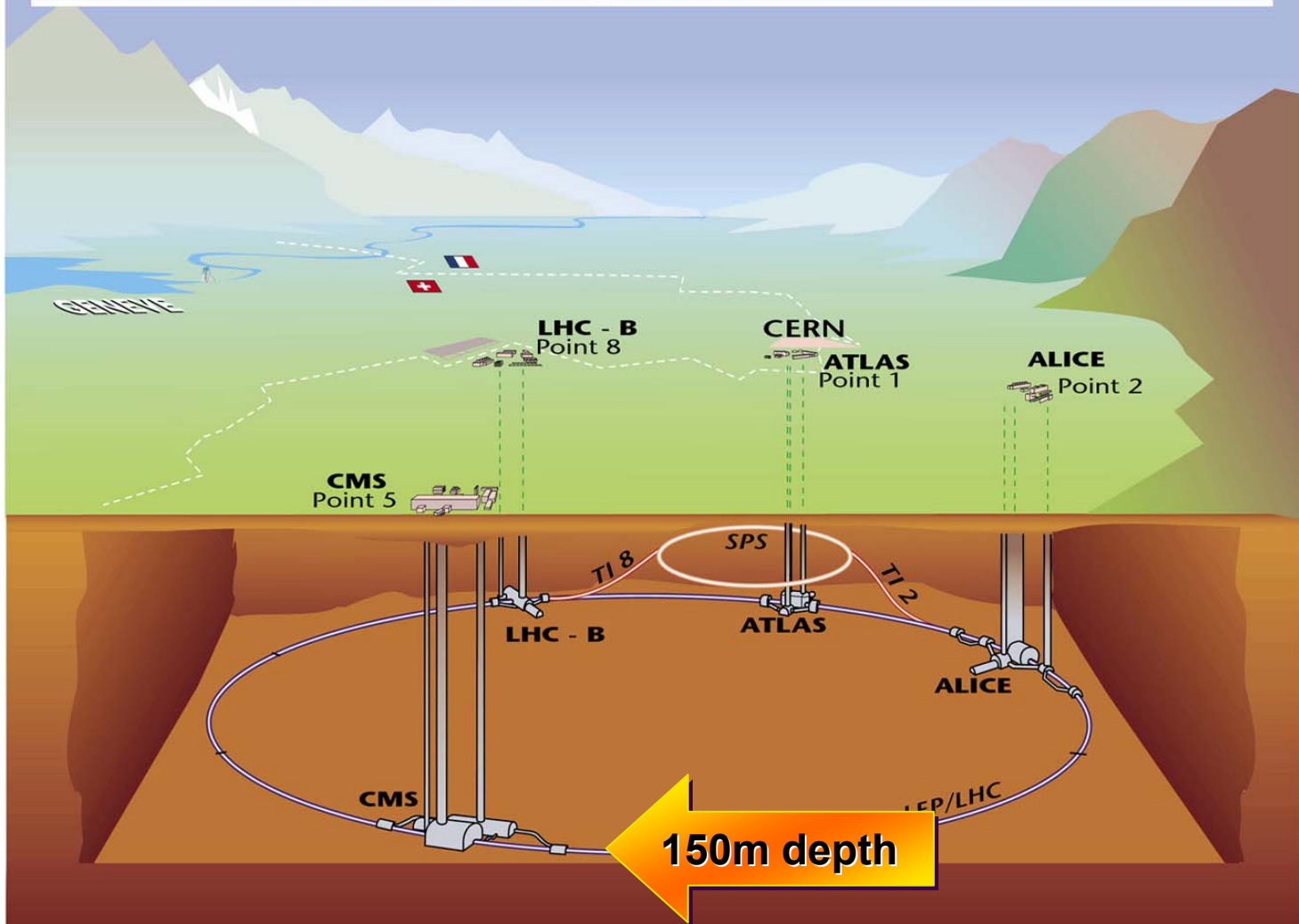
Sverre Iarn





# Accelerators and detectors in underground tunnels and caverns

## Overall view of the LHC experiments.



150m depth



# LHC will be equipped with four detectors

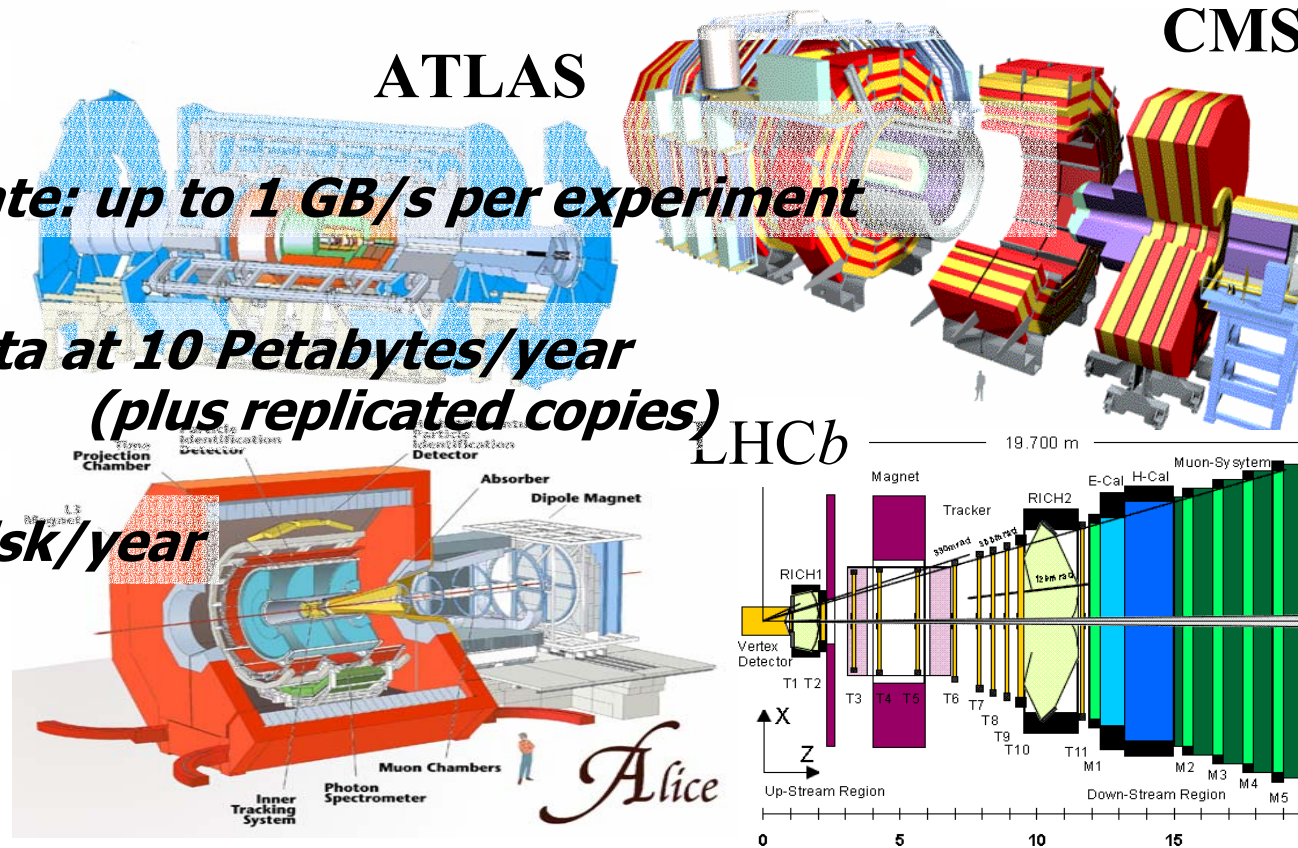
## Requirements:

### Storage –

**Raw recording rate: up to 1 GB/s per experiment**

**Accumulating data at 10 Petabytes/year  
(plus replicated copies)**

**2 Petabytes of disk/year**

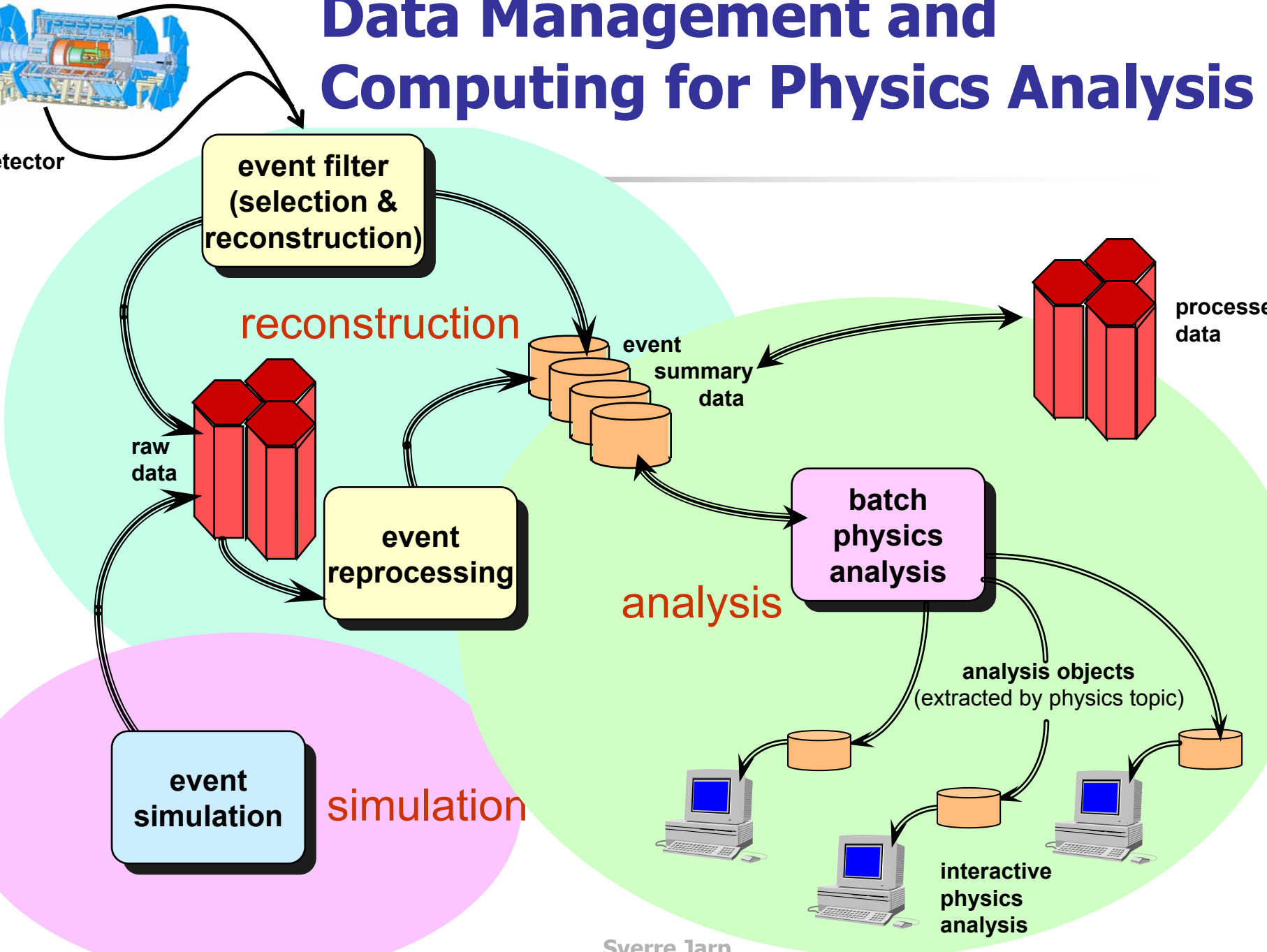




# Computing at CERN



# Data Management and Computing for Physics Analysis





# High Energy Physics Computing Characteristics

- **Independent events (collisions of particles)**
  - **trivial (read: pleasant) parallel processing**
- **Bulk of the data is read-only**
  - **versions rather than updates**
- **Meta-data in databases linking to “flat” files**
- **Compute power measured in SPECint (not SPECfp)**
  - **But good floating-point is important**
- **Very large aggregate requirements:**
  - **computation, data, input/output**
- **Chaotic workload –**
  - **research environment - physics extracted by iterative analysis, collaborating groups of physicists**
  - **Unpredictable** → **unlimited demand**



**CERN**



openlab for DataGrid applications

---

# CERN openlab





# Gelato update

## ■ What is new concerning our Itanium cluster?

- New physical location
- New interconnect
- New systems
  - 2-way

**HP joined the LHC Computing Grid!**

- New CERN linux distribution (CEL3)
- New for IPF: Grid software

## Industrial Collaboration:

- **Enterasys, HP, IBM, and Intel and Oracle are our partners**
- **Voltaire (with a 96-way Infiniband switch) just joined**
- **Technology aimed at the LHC era:**
  - **Network switches at 10 Gigabits**
  - **~ 100 rack-mounted HP servers**
  - **64-bit computing: Itanium-2 processors**
  - **StorageTank storage system**
    - **w/28 TB**
    - **~1 GB/s throughput**

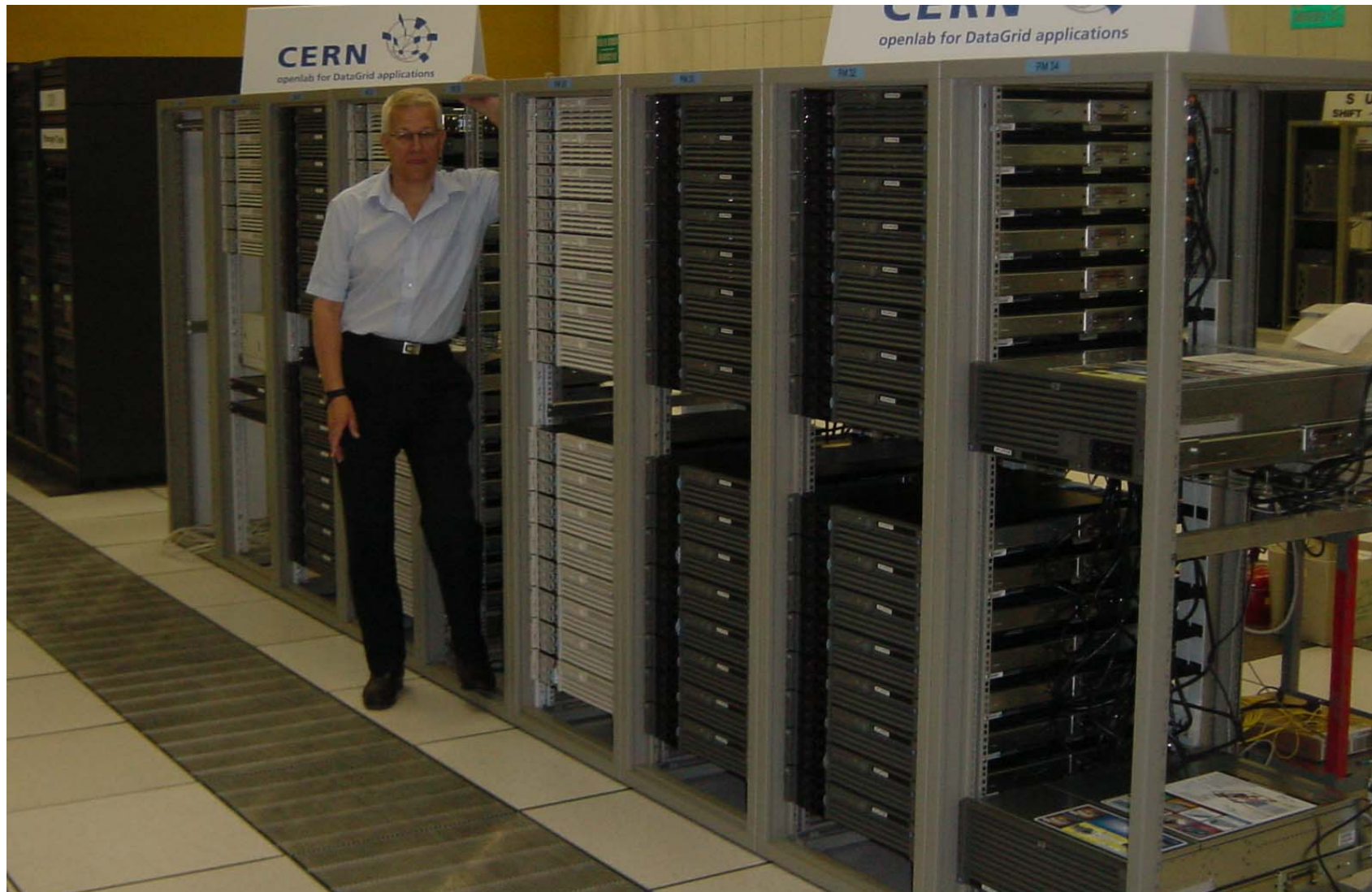


# CERN



openlab for DataGrid applications

# Front "seat" in the Computer Center





# Bench-1: ROOT tests/ecc

Reported at Gelato meeting in Stockholm

Benchmark "stress" (Bigger is better)	O2	O3	O2, IPO	O2, ftz, IPO, PGO
Intel 7.0	751	718	813	867
Intel 7.1	752	735	810	<b>887</b> (118% of O2 result)
Intel 8.0 beta	Better	Better	Better	Better

- **Good news with 8.0:**
  - **O3, IPO, PGO** work together
- **Intel compiler well ahead of gcc**
  - Nice improvement with PGO



# ROOT 4.00.03 benchmarks

## icc 8.0.061 – 1.5 GHz Itanium-2

Results (Bigger is better )	02, ansi_alias, ipo	02, ansi_alias, ipo, pgo (except geom)
stress -b -q	948	1080
stressLinear	784	870
stressgeom -b -q	889	952
bench -b -q	919	1041
Geometric mean	883	982

- Aggressive tuning helps
- Will get 6% more with 1.6 GHz
- Plans (hopes?) for version 8.1:
  - 1000 RM with single compilations
  - 1200 with aggressive tuning (e.g. pgo)

# WAN connectivity



Organisation Européenne pour la Recherche Nucléaire  
European Organization for Nuclear Research

PR15.03  
15.10.2003

## CERN and Caltech join forces to smash Internet speed record

**Titanium-2 single and  
multiple streams:  
ipv4 @ 5.44 Gbps  
(also ipv6 results)**

**Microsoft enters the stage:  
Multiple streams (only): 6.25 Gbps  
2 February 2004**

CERN and the California Institute of Technology (Caltech) will tomorrow receive an award for transferring data across 7,000 km of network at 5.44 gigabits per second (Gbps), smashing the old record achieved in February between CERN in Geneva and Sunnyvale in California by a Los Alamos National Laboratory and Stanford Linear Accelerator Center team.

The Caltech team set this new Internet2® Land Speed Record on 1 October 2003 by transferring 100 Gb of data in less than 30 minutes, corresponding to 38,420.54 petabit-metres per second. The record rate of 5.44 Gbps is more than 20,000 times faster than a typical home broadband connection and is equivalent to transferring a full CD in 1 second or a full length DVD movie in approximately 7 seconds. The award will be made to Olivier Martin of CERN and Harvey Newman of Caltech on the Lake Geneva Region Stand at the [ITU Telecom World event](#) in Geneva live from the Internet2 conference in Indianapolis at [17:30 CET on Thursday 16 October](#).

**For LHC we now have to get from an R&D project (DATATAG)  
to a sustained, reliable service – GEANT, ESNET, ..**

... 44 gigabits a second (Gbps) to a lab at the California Institute of Technology, or Caltech, on October 1. This is more than 20,000 times faster than a typical home broadband connection, and is also equivalent to transferring a 90-minute compact disc within one second -- an operation that takes around eight minutes on standard broadband. Using current technology, a DVD -- or digital video disc -- film of some 90 minutes length takes some 15 minutes to download from the Internet.



# LHC Computing Grid

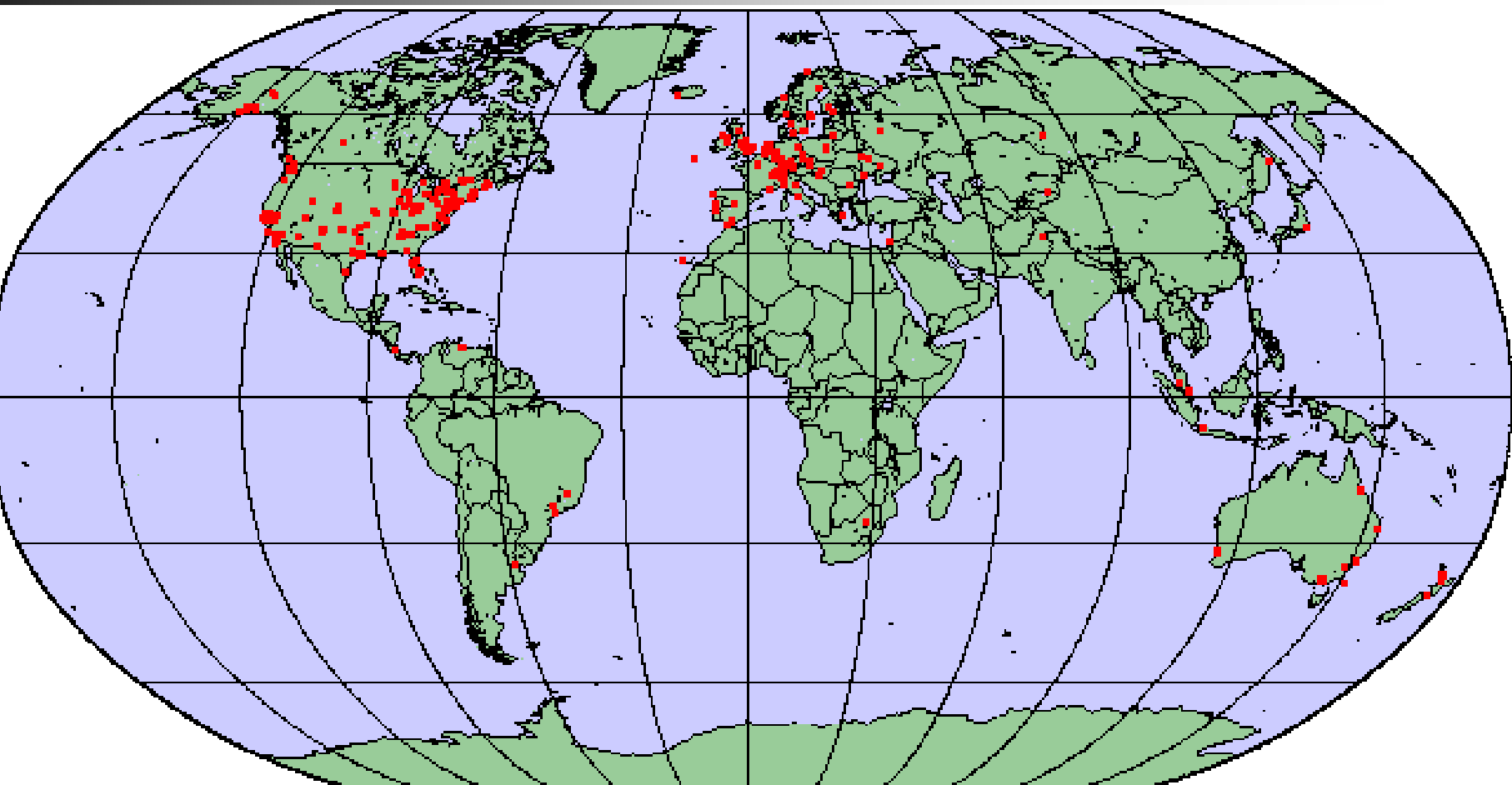
# Ingredients for a successful GRID

- **As I see them:**
  - **Geographically distributed:**
    - Funding
    - Computer Centers
    - Users
  - **A REAL deadline for a HUGE project**
  - **Clear computing model**
  - **Adequate funding for development**
  - **Good and motivated people**

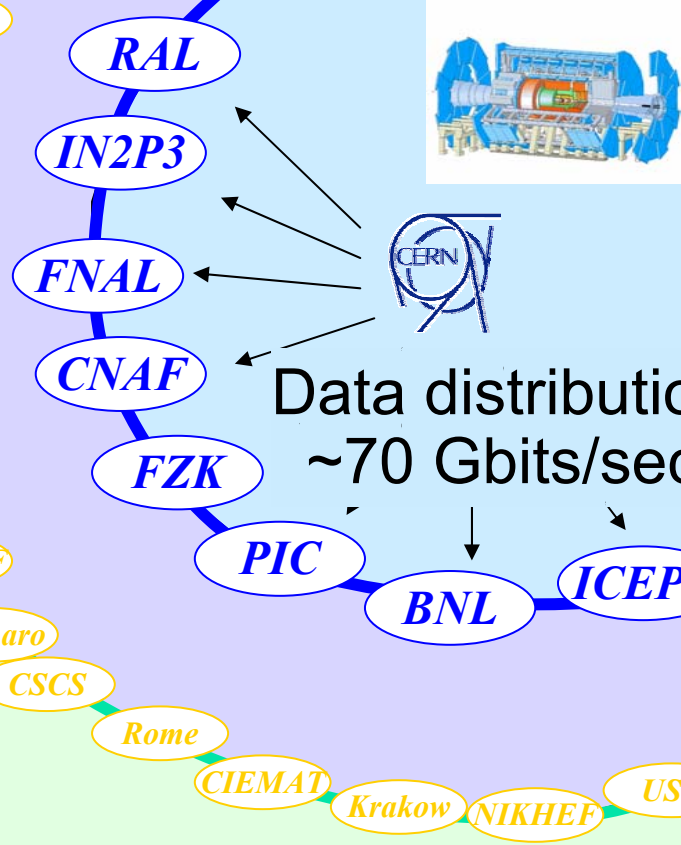
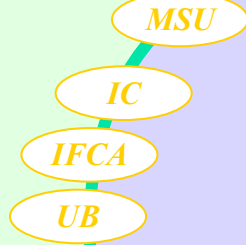




# CERN User Community



**Europe: 267 institutes, 4603 users**  
**Elsewhere: 208 institutes, 1632 users**



Data distribution  
~70 Gbits/sec

# 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
<b>Totals</b>	<b>105</b>	<b>37</b>	<b>43</b>



# LHC Computing Grid (LCG) Project

<http://lcg.web.cern.ch/lcg>

## Goal of the project:

To prepare, deploy and operate the computing environment for the experiment and analyze the data from the LHC detectors (as of 2007)

Phase 1 – 2002-05

development of common applications, libraries, networks, prototyping of the environment, operation of a pilot computing

Phase 2 – 2006-08

acquire, build and operate the LHC computing service

**The Grid is just a tool towards achieving this goal**

**LGC Production status**

1	<a href="#">CERN-LCG2</a>	pro,ext	Pass	912	111	1
2	<a href="#">CNAF-LCG2</a>	pro,ext	Pass	816	331	424
3	<a href="#">RAL-LCG2</a>	pro,ext	Pass	144	132	10
4	<a href="#">nikhef.nl</a>	pro,ext	Pass	244	143	101
5	<a href="#">FZK-LCG2</a>	pro,ext	Pass	214	211	3
6	<a href="#">FNAL-LCG2</a>	pro,ext	Pass	18	17	2
7	<a href="#">Taiwan-LCG2</a>	pro,ext	Pass	100	98	2
8	<a href="#">PIC-LCG2</a>	pro,ext	Pass	142	50	0
9	<a href="#">TRIUMF</a>	pro,ext	Pass	10	9	1
10	<a href="#">NCU-TAIWAN</a>	pro,ext	Pass	8	7	1
11	<a href="#">USC-LCG2</a>	pro,ext	Pass	8	8	0
12	<a href="#">IMPERIAL</a>	pro,ext	Pass	44	3	41
13	<a href="#">CAVN</a>	pro,ext	Pass	6	5	1
14	<a href="#">IFIC</a>	pro,ext	Pass	95	90	5
15	<a href="#">KRAKOW</a>	pro,ext	Pass	10	9	1
16	<a href="#">INFN-TORINO</a>	pro,ext	Pass	24	5	19
17	<a href="#">INFN-MILANO</a>	pro,ext	Pass	38	33	5
18	<a href="#">INFN-LEGNARO</a>	pro,ext	Pass	142	110	3
19	<a href="#">IFCA</a>	pro,ext	Pass	2	2	0
20	<a href="#">CTEMAT</a>	pro,ext	Pass	2	2	0

First 2 weeks in May: from 28 to 42 sites, from 2200 to 3340 CPUs

23	<a href="#">PRAGUE</a>	ext	Pass	0	0	0
24	<a href="#">UB-BARCELONA</a>	ext	Pass	3	3	0
25	<a href="#">HG-01-GRNET</a>	ext	Pass	48	48	0
26	<a href="#">WUPPERTAL</a>	ext	Pass	2	2	0
27	<a href="#">HEPHYUIBK</a>	ext	Pass	6	5	1
28	<a href="#">UAM</a>	ext	Fail	50	50	0
29	<a href="#">SHEF</a>	ext	Pass	16	14	2
30	<a href="#">LIP</a>	ext	Pass	4	4	0
31	<a href="#">LANC</a>	ext	Pass	26	26	0
32	<a href="#">BUDAPEST</a>	ext	Pass	53	20	19
33	<a href="#">PRAGUE-CESNET</a>	ext	Pass	28	28	0
34	<a href="#">#WEIZMANN</a>	ext	Fail	24	24	0
35	<a href="#">MONTREAL</a>	ext	Pass	10	10	0
36	<a href="#">ALBERTA</a>	ext	Pass	2	2	0
37	<a href="#">QMUL</a>	ext	Pass	2	2	0
38	<a href="#">MANCHESTER</a>	ext	Pass	4	4	0
39	<a href="#">INFN-ROMA1</a>	ext	Pass	56	44	12
40	<a href="#">AACHEN</a>	ext	Pass	2	2	0
41	<a href="#">INFN-NAPOLI</a>	ext	Pass	12	12	0
42	<a href="#">UCL</a>	ext	Fail	2	2	0



**The initial LCG (plus EDG)  
software was created in  
an x86-centric world**



# Porting the LCG code to Itanium

**Key people:****Stephen Eccles (PhD student; now at Lancaster Uni.)****Andreas Unterkircher (CERN fellow)****Report:****<http://openlab-mu-internal.web.cern.ch/openlab-mu-internal/Documents/Reports/Technical/porting LCG to IA64 v2.doc>**



# LCG code

- **Virtual Data Toolkit (VDT):**
  - Globus, Condor-G, MyProxy
- **European Data Grid:**
  - Work Packages 1 (job scheduling) & 2 (data mgmt)
- **DataTAG:**
  - GLUE, GridICE
- **LCG:**
  - Global File Access Library (GFAL), LCG's BDII (Berkeley Database Info. Index), experiment specific SW
- **DESY/FNAL:**
  - dCache (Mass Storage Management system)
- **External:**
  - ant, tomcat,...



# Porting to Itanium

- **We had to rebuild everything from scratch.**
- **LCFG (the basic configuration tool) was not available for IPF, so we had to install and configure everything manually.**
  - **Quote from EDG Installation Guide**  
“...manual configuration of the services without the benefit of the LCFG configuration components is extremely difficult (...) for now, using this method is strongly discouraged.”



# Our strategy

- 1. Initially only port a “minimal grid node”:**
  - 1. Port the code necessary for a Worker Node and a Compute Element.**
  - 2. Try to reproduce the x86 rpms provided by LCG as close as possible.**
- 2. Demonstrate that such an Itanium node would actually work.**
- 3. Try to get Itanium-specific changes into the CVS code repository.**
- 4. Help with porting experiments’ software.**
- 5. Port more of LCG to IPF if needed.**
- 6. Make IPF a fully supported LCG platform.**



# Porting obstacles

- **Source RPMs were often difficult to find.**
- **Complicated dependencies.**
- **Documentation was scarce and knowledge was spread amongst different groups.**
- **EDG initially was not targeted for porting.**
- **Reverse engineering of build procedures is always extremely time consuming.**



# Porting observations

- **Most of the problems encountered were due to build or configuration issues.**
  - The code itself is mostly “64-bit ready”.
- **Different languages:**
  - C, C++, Java, Perl, Python, Swig
- **Keeping track with new LCG tags was difficult for us.**
  - Getting the IPF extensions into the CVS tree was essential.



# Achievements so far

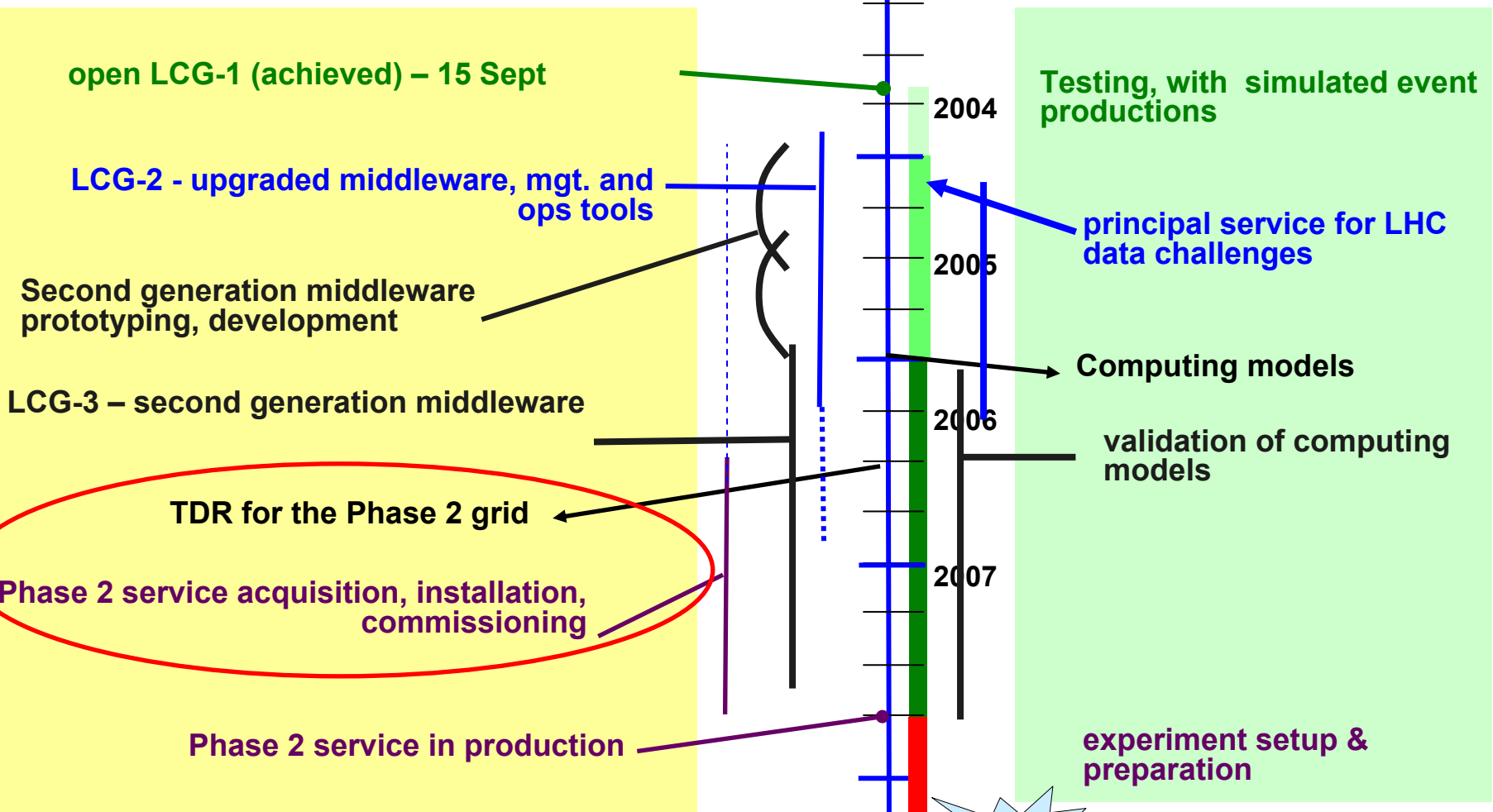
- **VDT will start to release an IPF distribution from the next release onwards.**
- **Itanium-specific changes get into the EDG CVS tree.**
- **Systematic recording of source locations or gathering of source files.**
- **Some tests of the Itanium Grid node have been conducted and so far there seem to be very few problems.**
- **The latest tag was brought to HP Puerto Rico for installation there.**



# LCG Time Line

**computing service**

**physics**



\* TDR – technical design report



# Itanium systems in LCG in 2007 ?

- **Priorities (as I see them):**
  - **Cost optimization**
    - DP servers w/cost-effective chip-sets, S-ATA disks, etc. (just like Xeon)
  - **SPEC optimization:**
    - Maximize SPEC-results/USD
  - **Heat optimization**
    - Minimize Watts/SPEC-result
  - **More chip variants**
  - **Better and more compilers**
    - OpenImpact is coming (even with C++ support)
    - But, gcc is still "limping" on IPF
    - What about HP's compiler suite (for Linux) ?
    - OR, even Pathscale
      - Which originates from the SGI compilers
        - AKA ORC

