

CERN

European Organization for Nuclear Research

Organisation Européenne pour la Recherche Nucléaire

The LHC Computing Challenge: Preparation, Reality and Future Outlook

Tony Cass

Leader, Database Services Group
Information Technology Department

10th November 2010



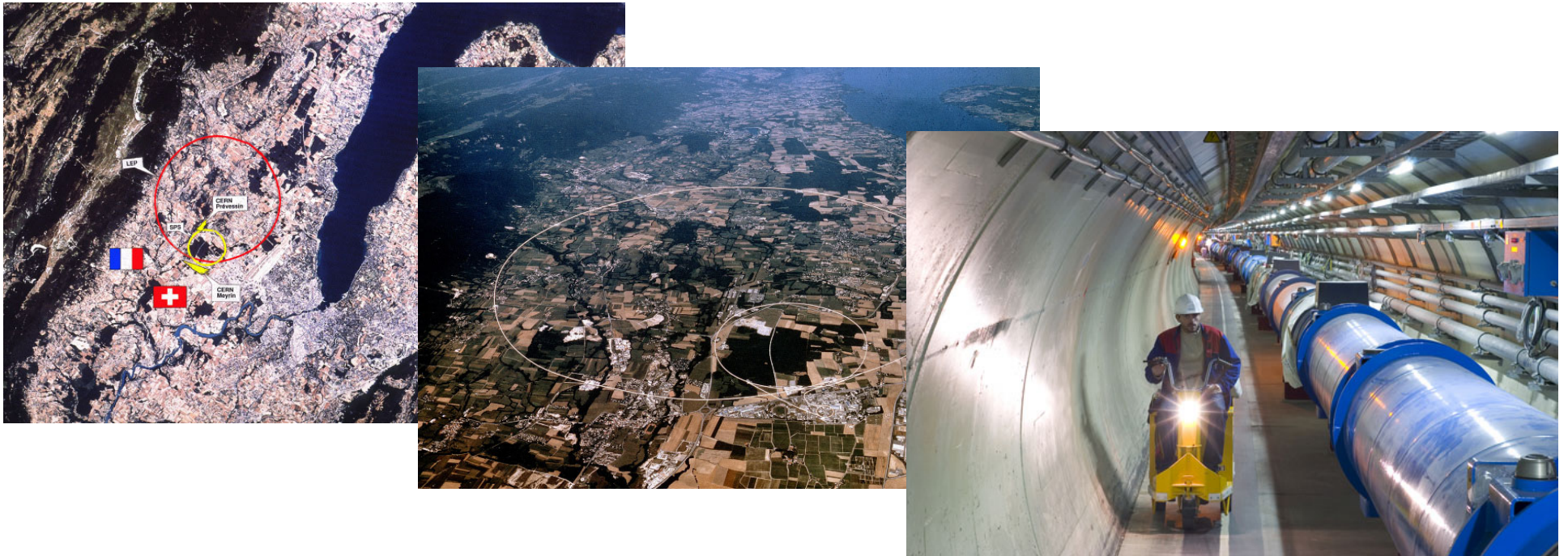
Outline

- Introduction to CERN, LHC and Experiments
- The LHC Computing Challenge
- Preparation
- Reality
- Future Outlook
- Summary/Conclusion



CERN

The fastest racetrack on the planet...

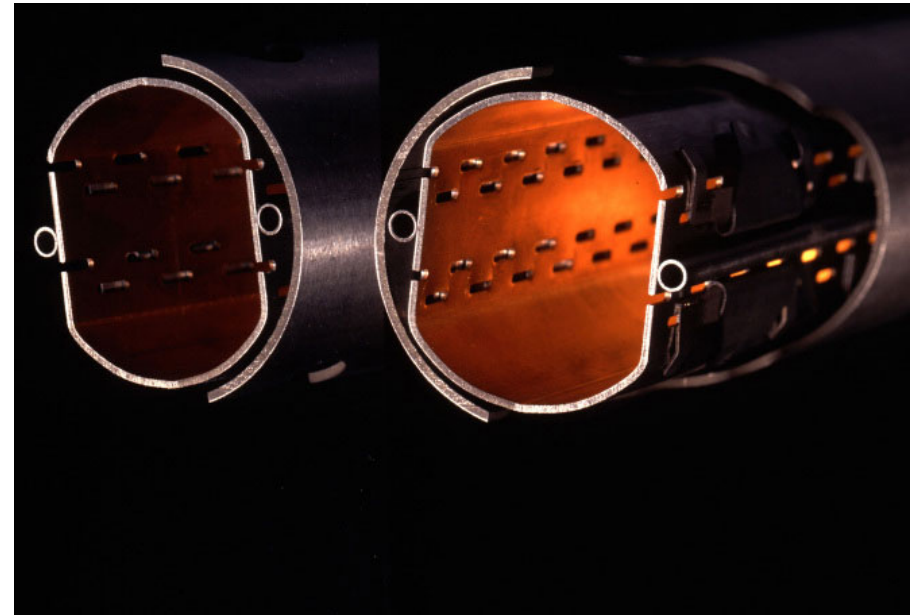
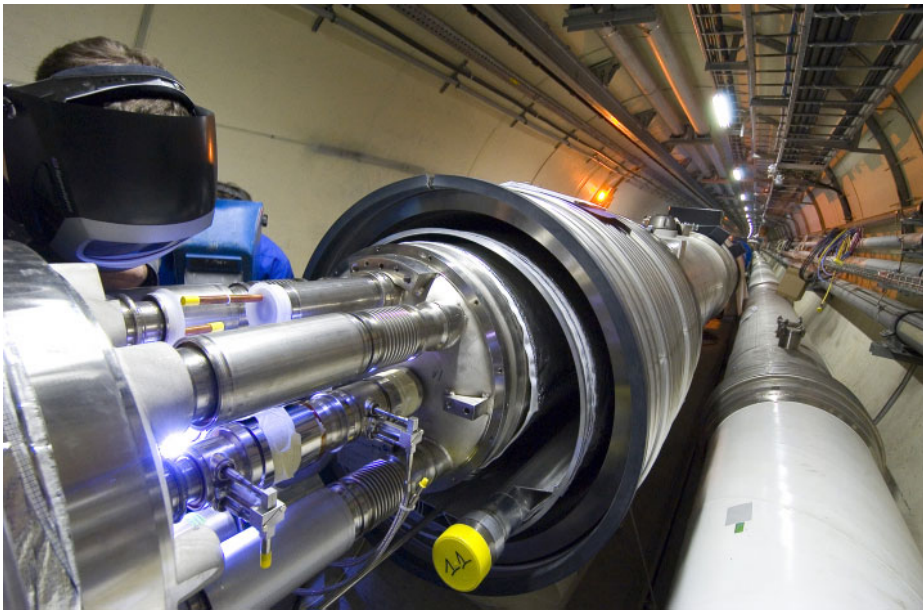


Trillions of protons will race around the 27km ring in opposite directions over 11,000 times a second, travelling at 99.999999991 per cent the speed of light.



CERN

The emptiest space in the solar system...



To accelerate protons to almost the speed of light requires a vacuum as empty as interplanetary space. There is 10 times more atmosphere on the moon than there will be in the LHC.



CERN

One of the coldest places in the universe...

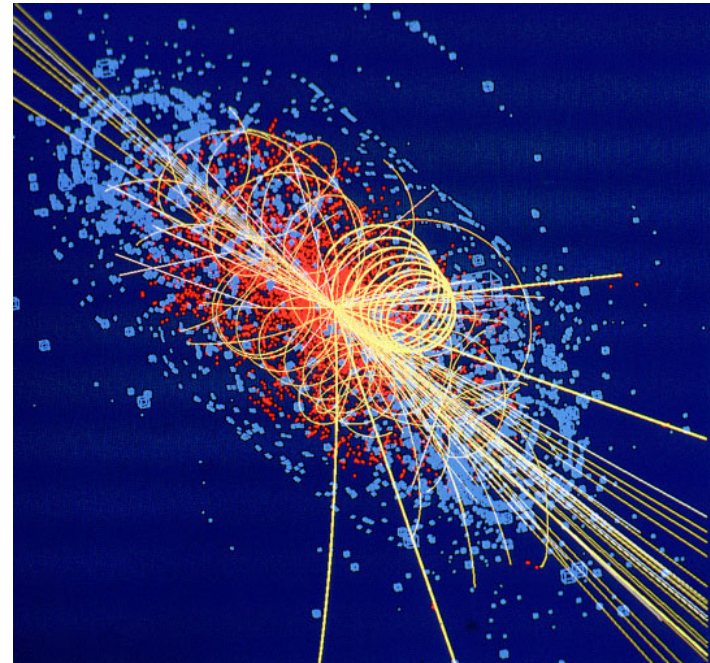
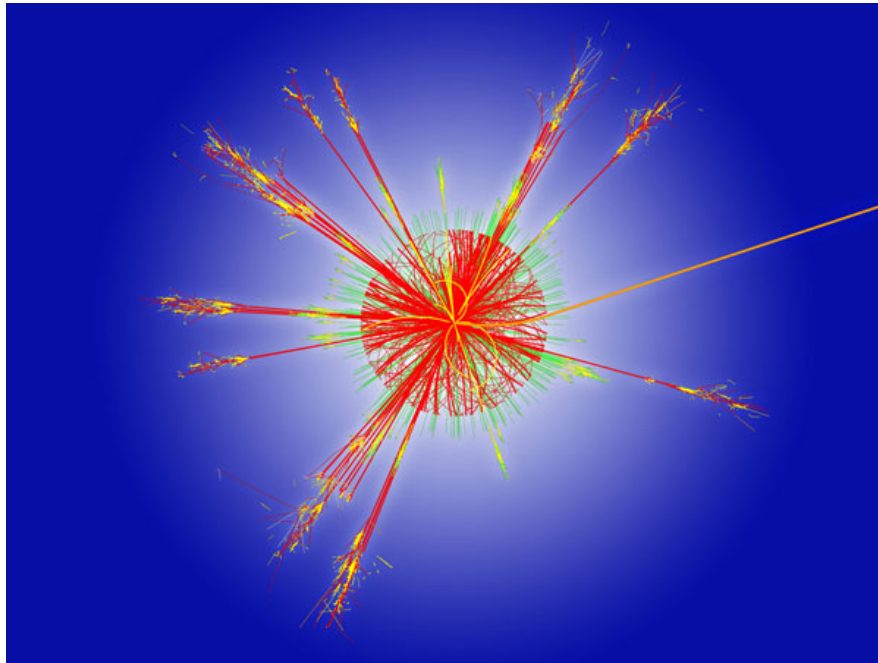


With an operating temperature of about -271 degrees Celsius, just 1.9 degrees above absolute zero, the LHC is colder than outer space.



CERN

The hottest spots in the galaxy...

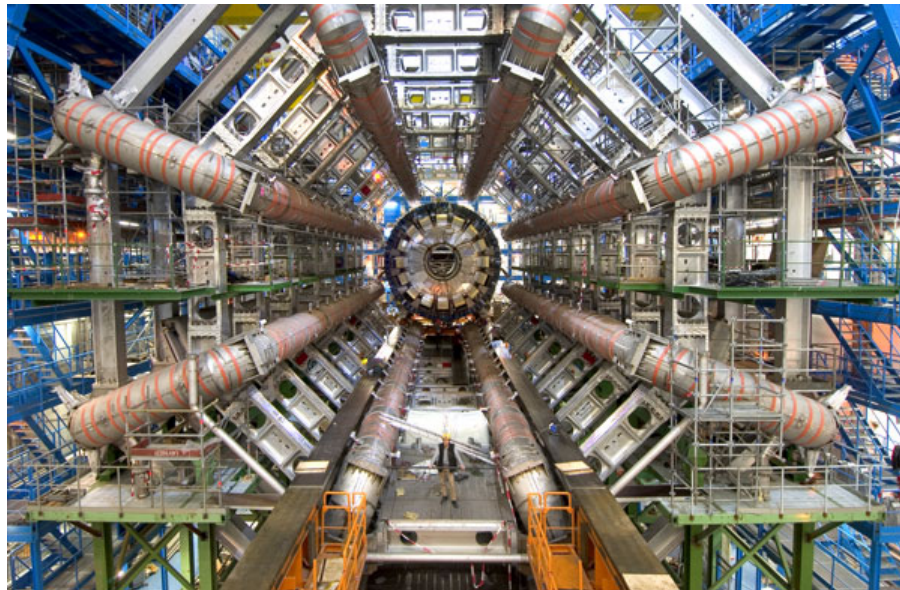


When two beams of protons collide, they will generate temperatures 1000 million times hotter than the heart of the sun, but in a minuscule space.



CERN

The biggest most sophisticated detectors ever built...

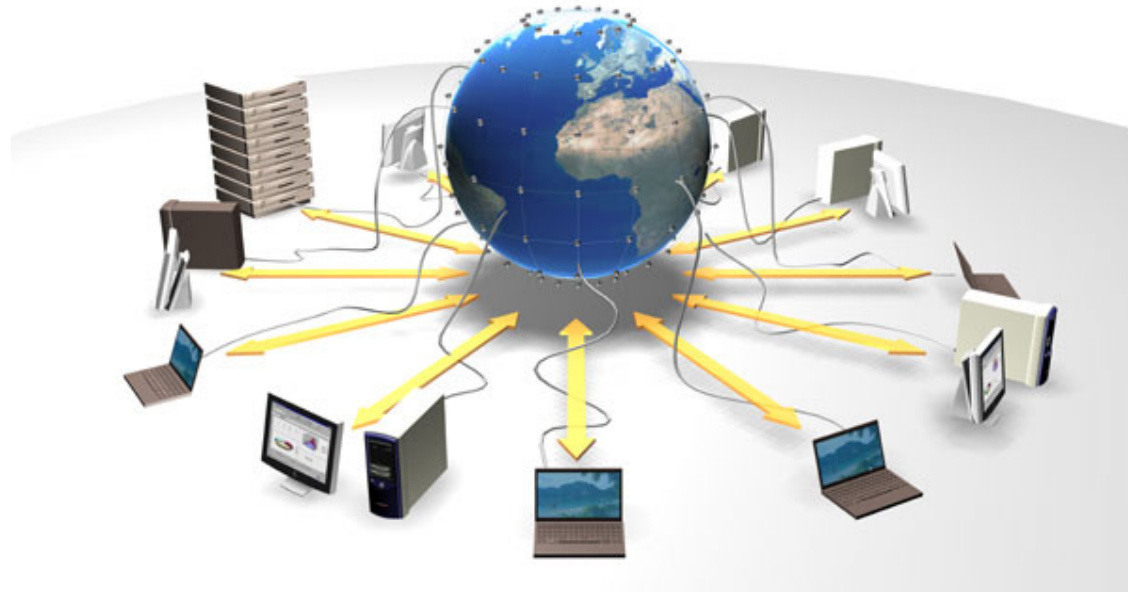


To sample and record the debris from up to 600 million proton collisions per second, scientists are building gargantuan devices that measure particles with micron precision.



CERN

The most extensive computer system in the world...



To analyse the data, tens of thousands of computers around the world are being harnessed in the Grid. The laboratory that gave the world the web, is now taking distributed computing a big step further.



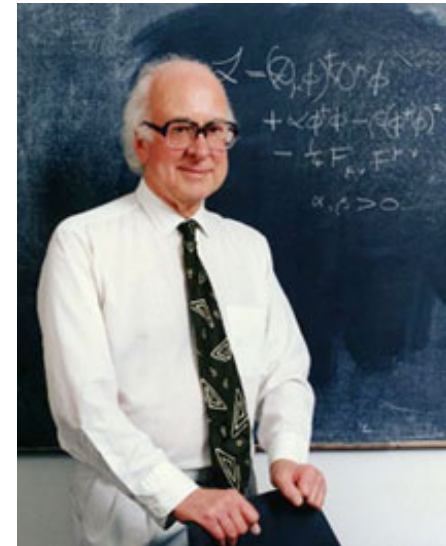
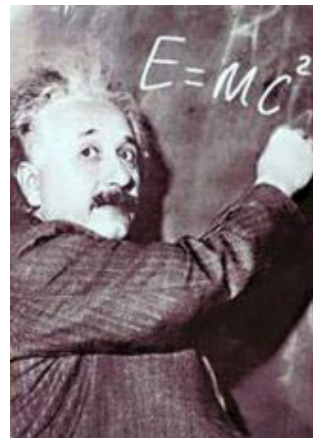
CERN

Why?



CERN

To push back the frontiers of knowledge...

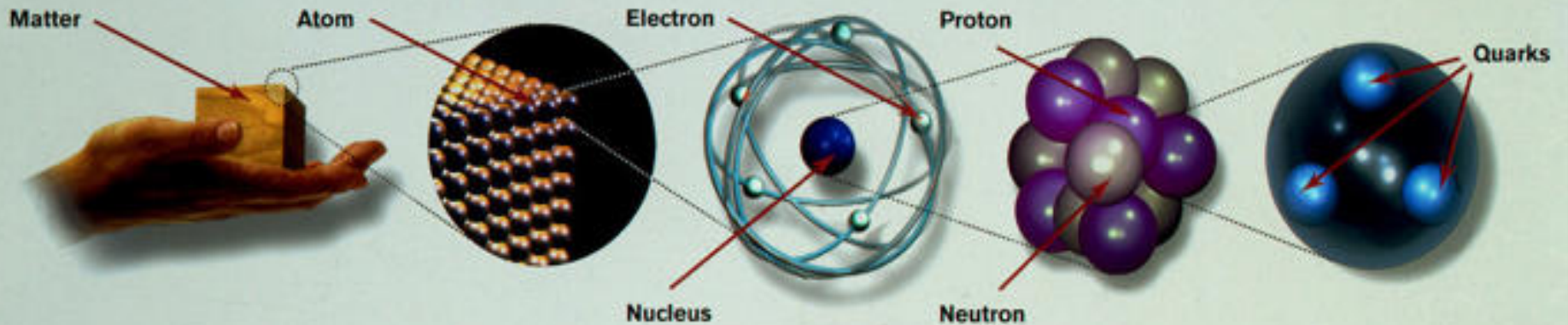


Newton's unfinished business... what is mass?

Science's little embarrassment... what is 96% of the Universe made of?

Nature's favouritism... why is there no more antimatter?

The secrets of the Big Bang... what was matter like within the first second of the Universe's life?



Matter particles
All ordinary particles belong to this group

These particles existed just after the Big Bang. Now they are found only in cosmic rays and accelerators

LEPTONS					
FIRST FAMILY	<table border="1"> <tr> <td> Electron Responsible for electricity and chemical reactions; it has a charge of -1 </td> <td></td> <td> Electron neutrino Particle with no electric charge, and possibly no mass; billions fly through your body every second </td> <td></td> </tr> </table>	Electron Responsible for electricity and chemical reactions; it has a charge of -1		Electron neutrino Particle with no electric charge, and possibly no mass; billions fly through your body every second	
Electron Responsible for electricity and chemical reactions; it has a charge of -1		Electron neutrino Particle with no electric charge, and possibly no mass; billions fly through your body every second			
SECOND FAMILY	<table border="1"> <tr> <td> Muon A heavier relative of the electron; it lives for two-millionths of a second </td> <td></td> <td> Muon neutrino Created along with muons when some particles decay </td> <td></td> </tr> </table>	Muon A heavier relative of the electron; it lives for two-millionths of a second		Muon neutrino Created along with muons when some particles decay	
Muon A heavier relative of the electron; it lives for two-millionths of a second		Muon neutrino Created along with muons when some particles decay			
THIRD FAMILY	<table border="1"> <tr> <td> Tau Heavier still; it is extremely unstable. It was discovered in 1975 </td> <td></td> <td> Tau neutrino not yet discovered but believed to exist </td> <td></td> </tr> </table>	Tau Heavier still; it is extremely unstable. It was discovered in 1975		Tau neutrino not yet discovered but believed to exist	
Tau Heavier still; it is extremely unstable. It was discovered in 1975		Tau neutrino not yet discovered but believed to exist			

QUARKS			
Up Has an electric charge of plus two-thirds; protons contain two, neutrons contain one		Down Has an electric charge of minus one-third; protons contain one, neutrons contain two	
Charm A heavier relative of the up; found in 1974		Strange A heavier relative of the down; found in 1964	
Top Heavier still		Bottom Heavier still; measuring bottom quarks is an important test of electroweak theory	

Force particles
These particles transmit the four fundamental forces of nature although gravitons have so far not been discovered

Gluons
Carriers of the **strong force** between quarks

Felt by: quarks

The explosive release of nuclear energy is the result of the **strong force**

Photons
Particles that make up light; they carry the **electromagnetic force**

Felt by: quarks and charged leptons

Electricity, magnetism and chemistry are all the results of **electro-magnetic force**

Intermediate vector bosons
Carriers of the **weak force**

Felt by: quarks and leptons

Some forms of radio-activity are the result of the **weak force**

Gravitons
Carriers of **gravity**

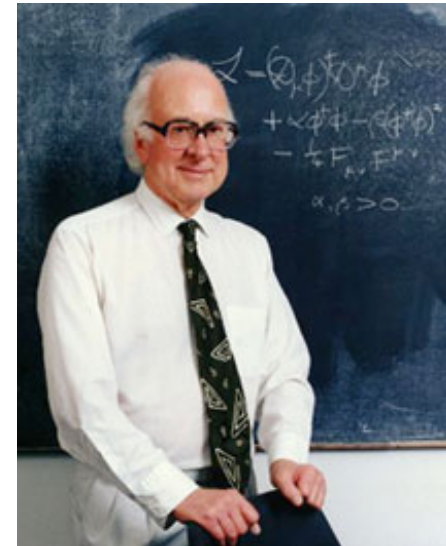
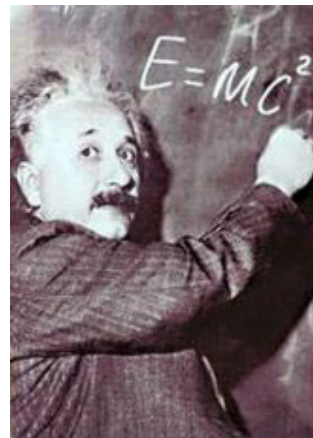
Felt by: all particles with mass

All the weight we experience is the result of the **gravitational force**



CERN

To push back the frontiers of knowledge...



Newton's unfinished business... what is mass?

Science's little embarrassment... what is 96% of the Universe made of?

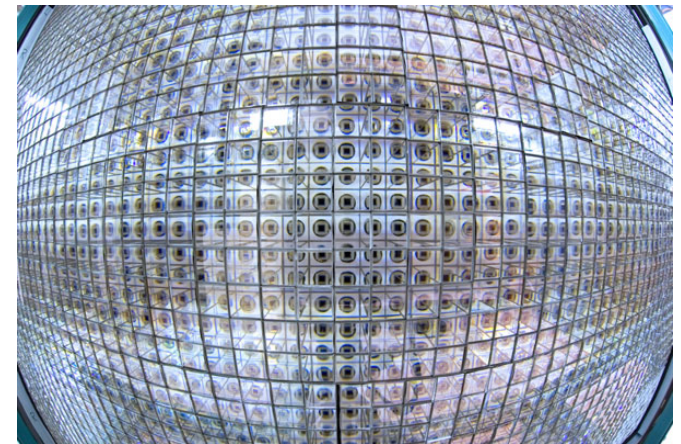
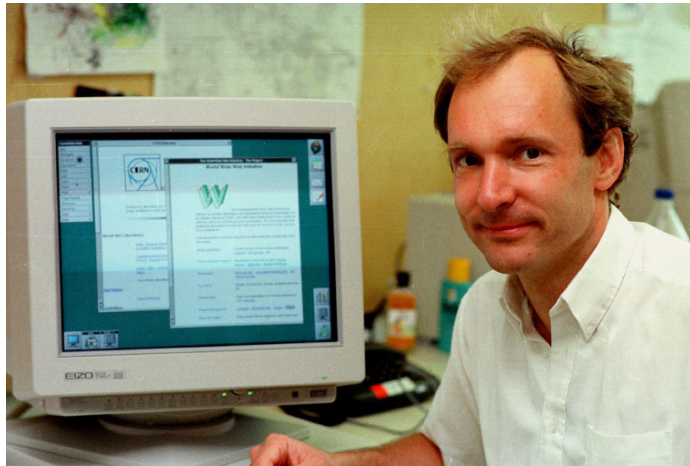
Nature's favouritism... why is there no more antimatter?

The secrets of the Big Bang... what was matter like within the first second of the Universe's life?



CERN

To develop new technologies...



Information technology - the Web and the Grid

Medicine - diagnosis and therapy

Security - scanning technologies for harbours and airports

Vacuum - new techniques for flat screen displays or solar energy devices



CERN

To unite people from different countries and cultures...



20 Member states

38 Countries with cooperation agreements

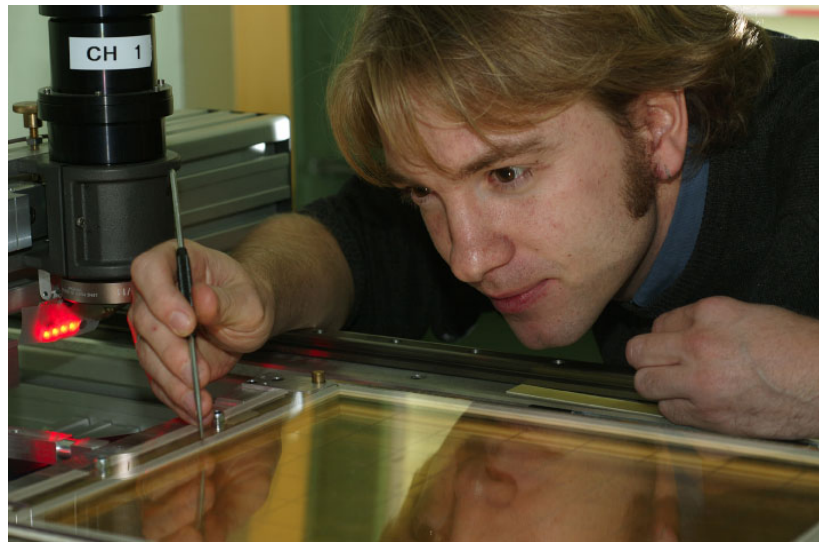
111 Nationalities

10000 People



CERN

To train the scientists and engineers of tomorrow...



From mini-Einstein workshops for five to sixes, through to professional schools in physics, accelerator science and IT, CERN plays a valuable role in building enthusiasm for science and providing formal training..



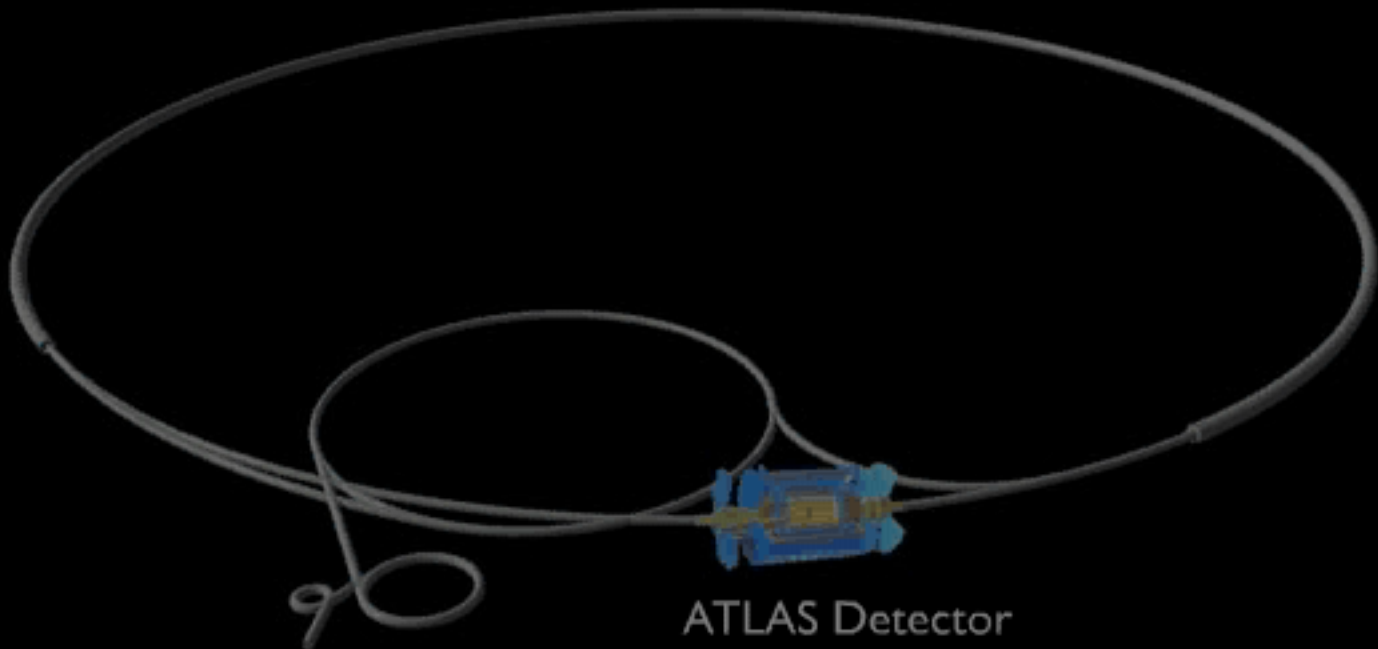
“Compact” Detectors!





PLAY ▶

Large Hadron Collider



ATLAS Detector

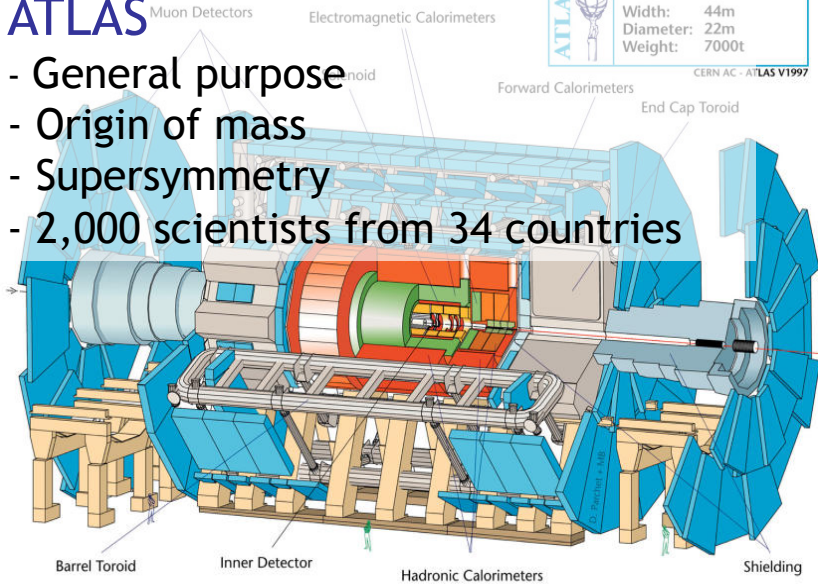


The Four LHC Experiments...

ATLAS

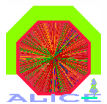
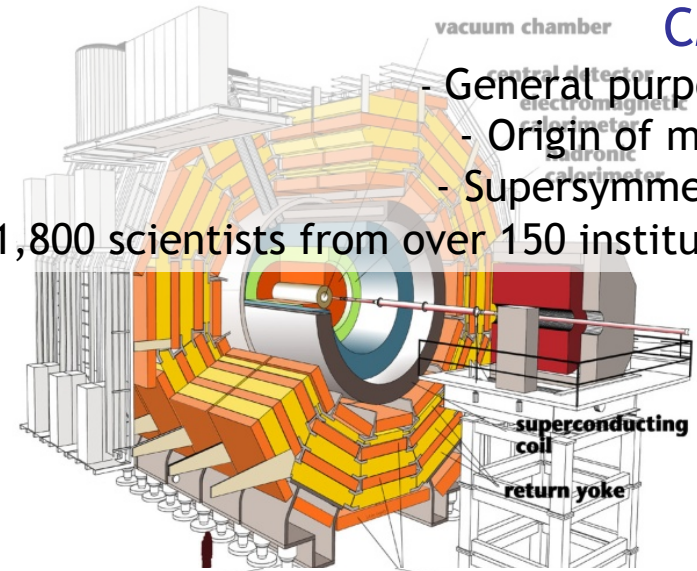
- General purpose
- Origin of mass
- Supersymmetry
- 2,000 scientists from 34 countries

Detector characteristics	
Width:	44m
Diameter:	22m
Weight:	7000t



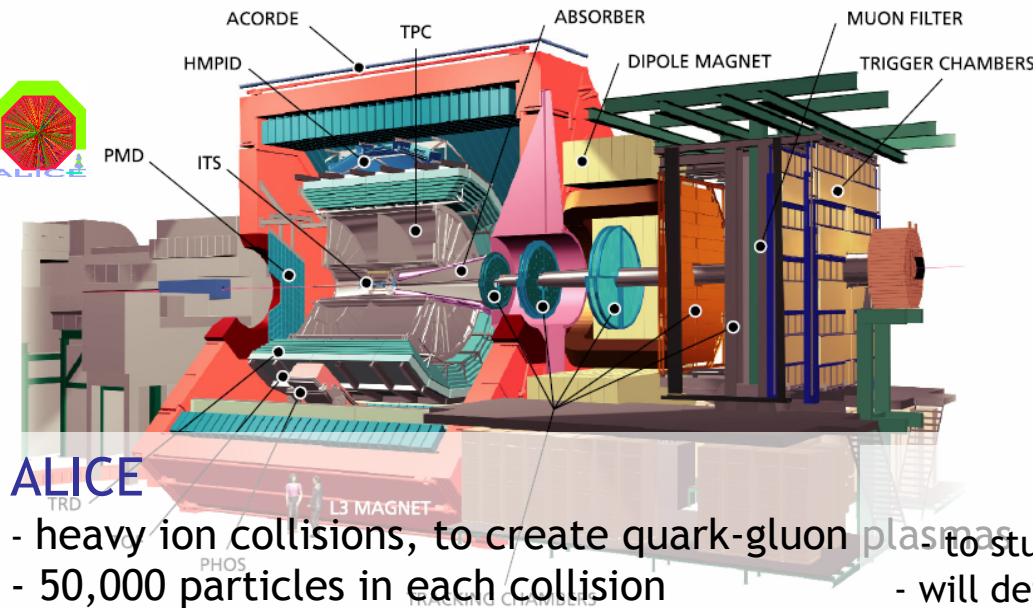
CMS

- General purpose
- Origin of mass
- Supersymmetry
- 1,800 scientists from over 150 institutes



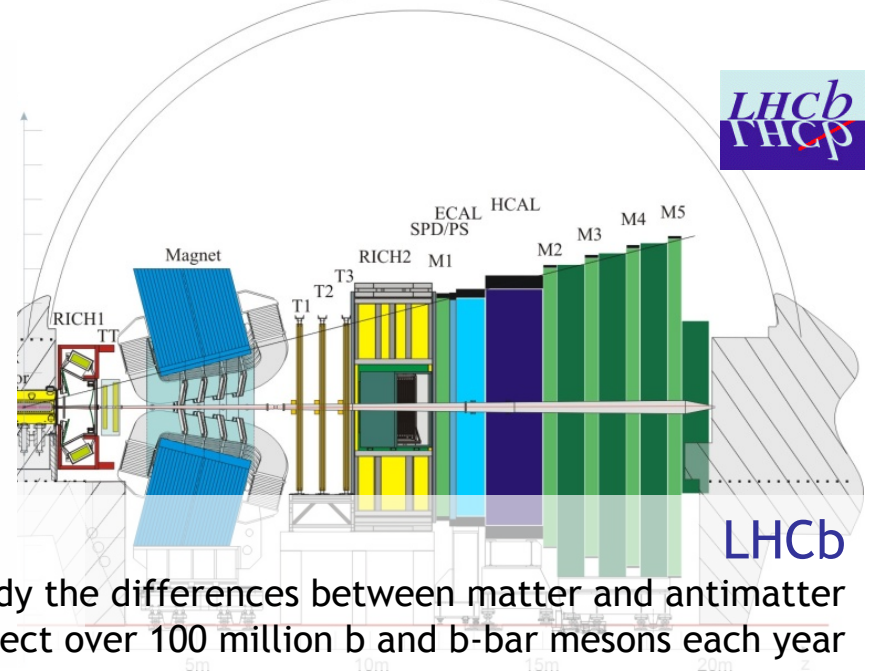
ALICE

- heavy ion collisions, to create quark-gluon plasmas
- 50,000 particles in each collision



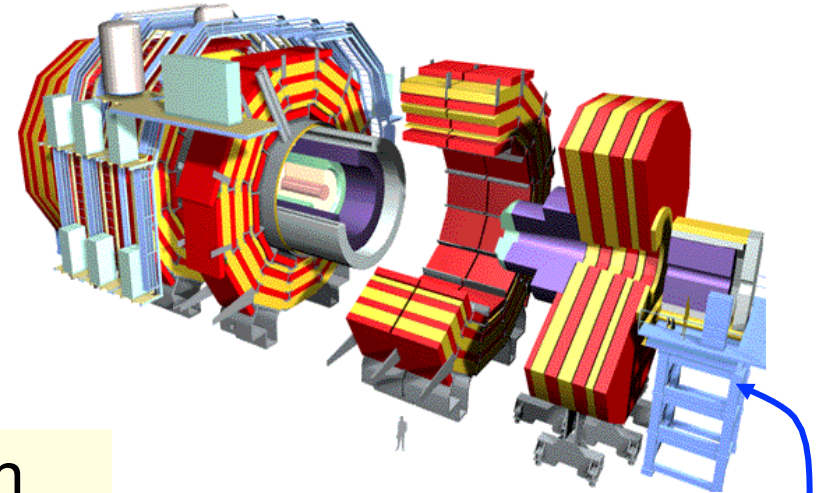
LHCb

- will detect over 100 million b and b-bar mesons each year





... generate lots of data ...



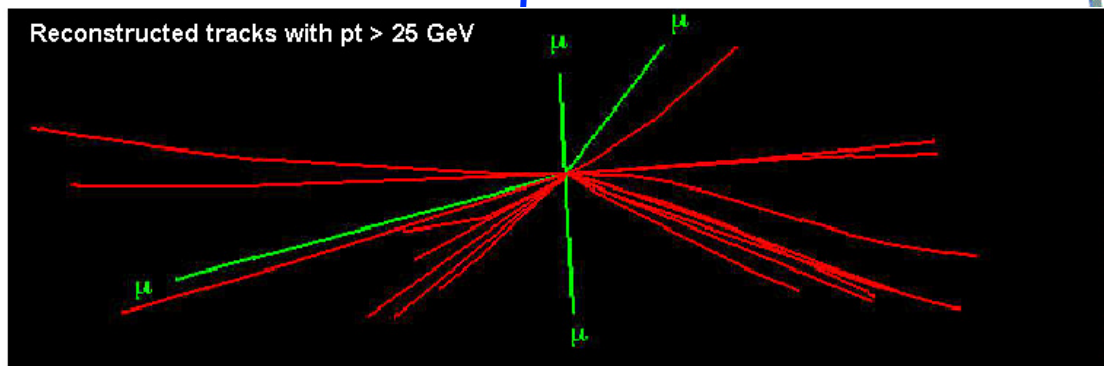
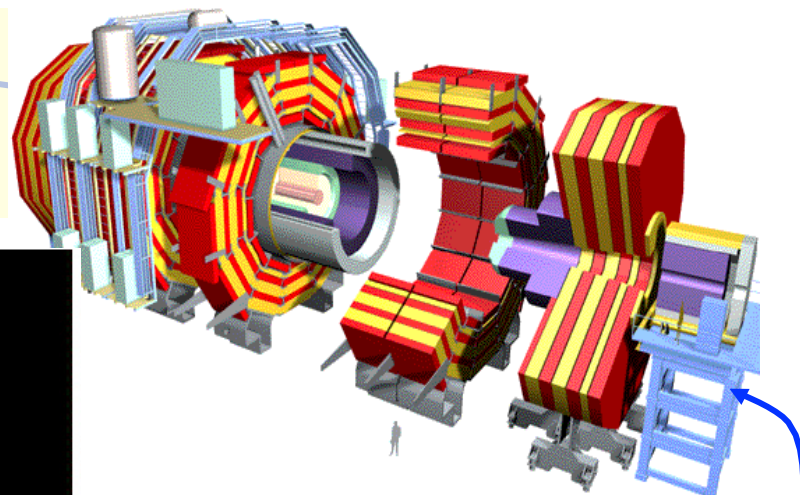
The accelerator generates 40 million particle collisions (events) every second at the centre of each of the four experiments' detectors





... generate lots of data ...

reduced by online computers to a few hundred “good” events per second.



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

- Current forecast ~ **23-25 PB / year**, 100-120M files / year
 - ~ 20-25K 1 TB tapes / year
- Archive will need to store **0.1 EB** in 2014, ~1Billion files in 2015





which is distributed worldwide

Tier-0 (CERN):

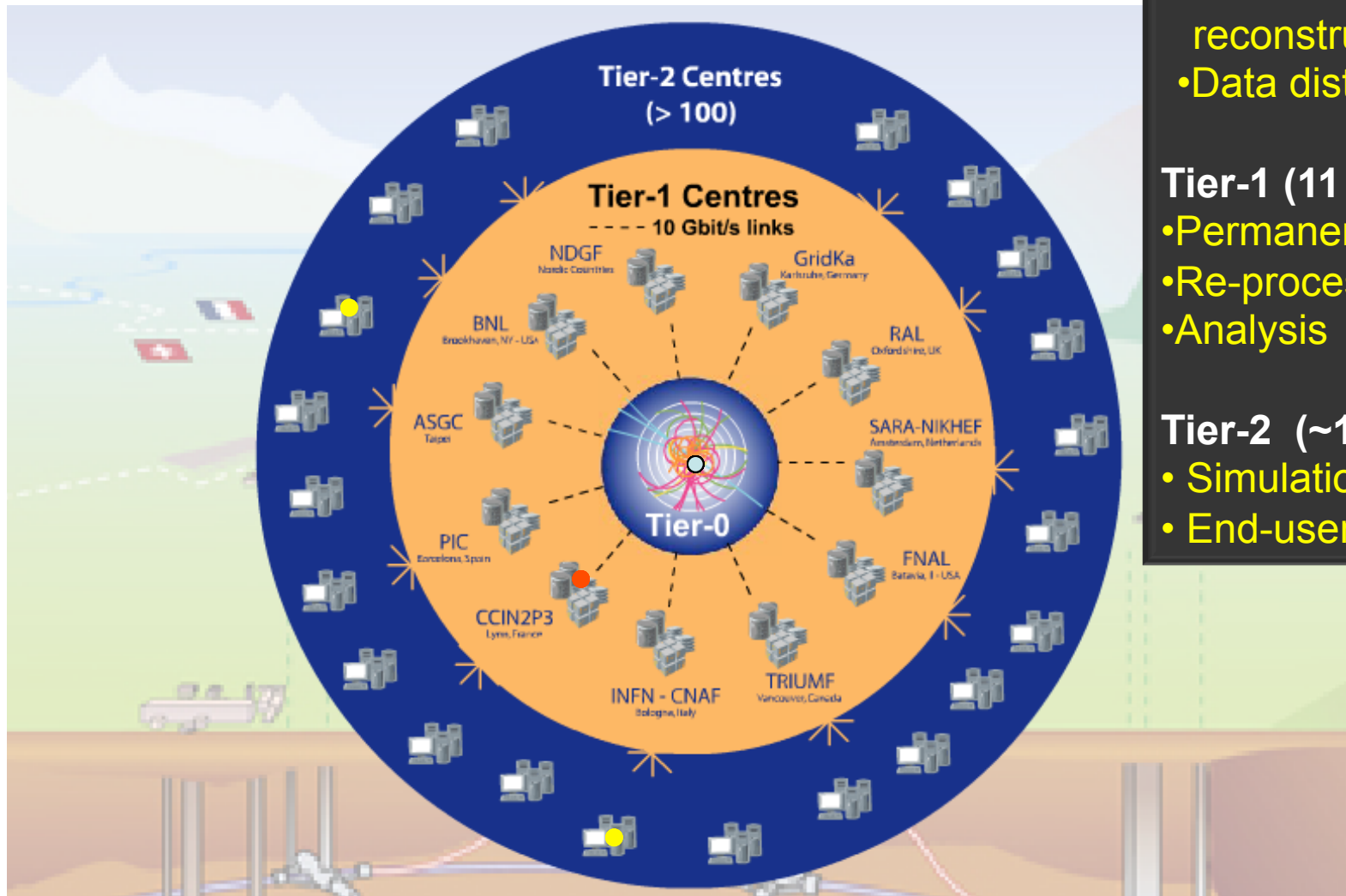
- Data recording
- Initial data reconstruction
- Data distribution

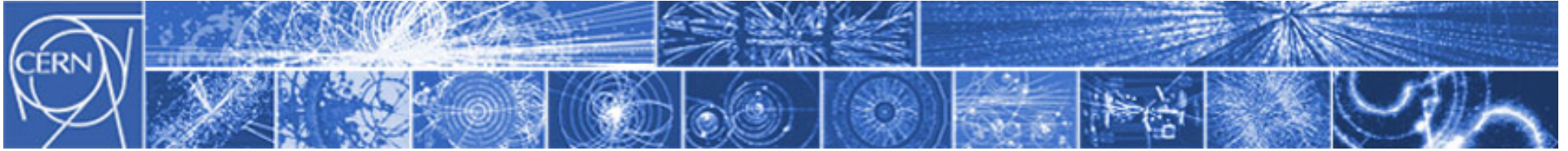
Tier-1 (11 centres):

- Permanent storage
- Re-processing
- Analysis

Tier-2 (~130 centres):

- Simulation
- End-user analysis





See

<http://dashb-earth.cern.ch/dashboard/doc/guides/service-monitor-gearth/html/user/setupSection.html>

For the Google Earth monitoring display

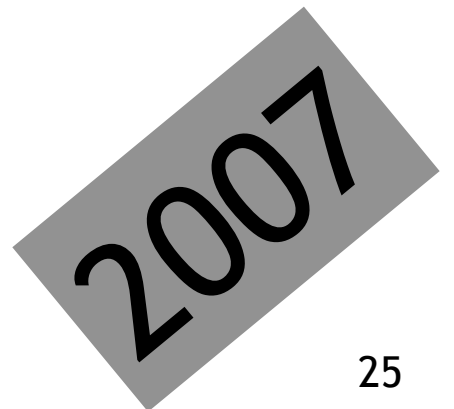


What were the challenges in 2007?



Outline

- Introduction to CERN and Experiments
- LHC Computing
- **Challenges**
 - Capacity Provision
 - Box Management
 - Data Management and Distribution
 - What's Going On?
- Summary/Conclusion





Outline

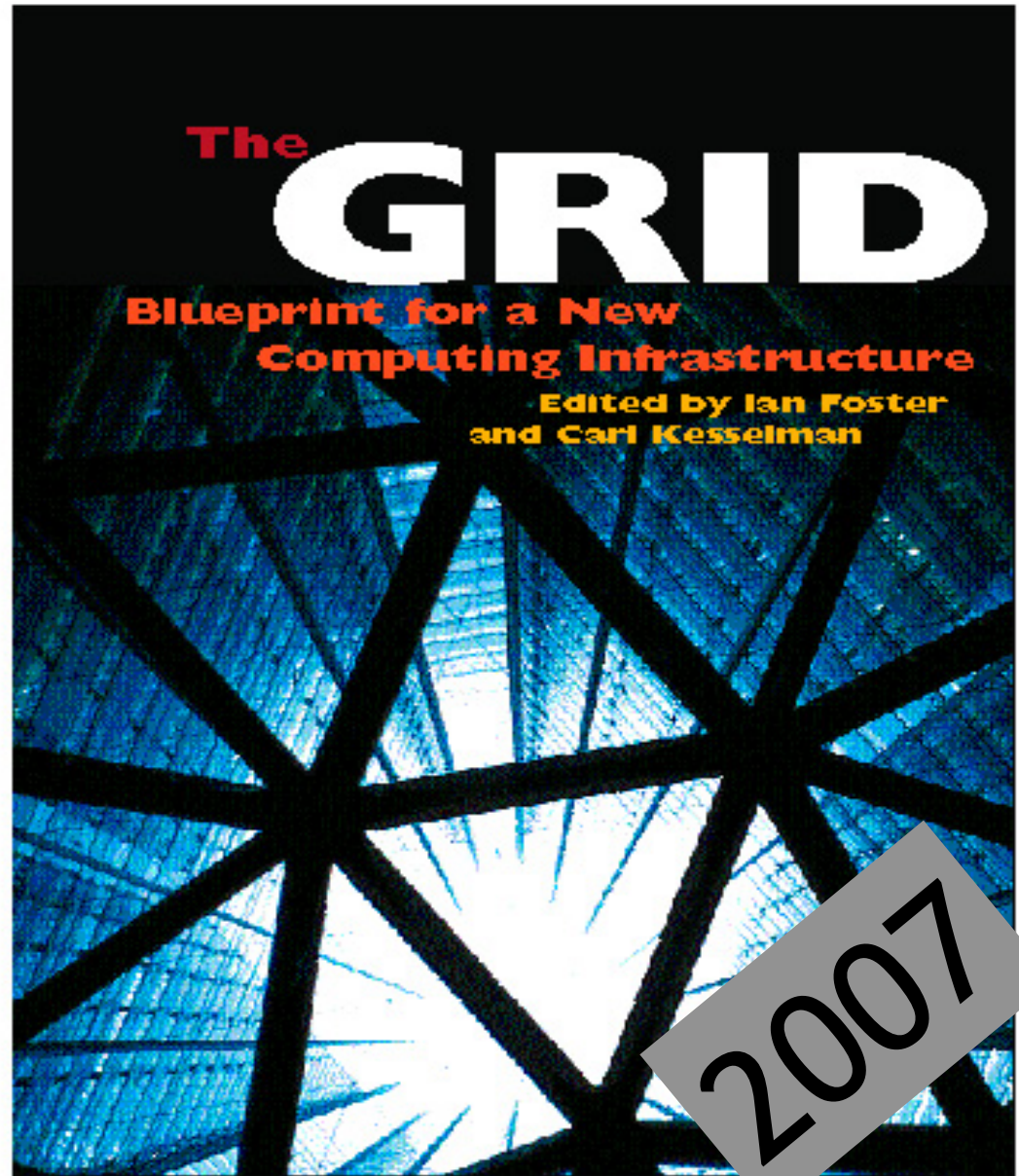
- Introduction to CERN and Experiments
- LHC Computing
- **Challenges**
 - Capacity Provision
 - Box Management
 - Data Management and Distribution
 - What's Going On?
- Summary/Conclusion

2007



The Grid

- Timely Technology!
- **Deploy** to meet LHC computing needs.
- Challenges for the **Worldwide LHC Computing Grid Project** due to
 - worldwide nature
 - competing middleware...
 - newness of technology
 - competing middleware...
 - scale
 - ...





TECHNOLOGY

Infrastructure

Network

Security

Client

Server

Mobile

Operating Systems

Data Center

Applications

Development

Architecture

LEADERSHIP

CIO Resource Alerts GET NOTIFIED!

NEWSLETTERS

CIO.com updates, insights and advice on technology, management and your career.

Advice and Opinion

CIO Consumer IT

CIO Leader

CIO Enterprise

CIO Insider

More Newsletters | Edit Profile

enter e-mail

SIGN-UP

Feature

Seven Wonders of the IT World

The fastest supercomputer. The most intriguing data center. The constantly changing core at the heart of Linux. Take a tour of the most impressive and most unusual marvels of the IT world.

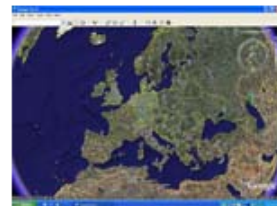
Leave a comment (65)

By C.G. Lynch

PAGE 4

World's largest scientific grid computing project: The E-science II (EGEE-II) project

Launched: September 2006, [for use by](#) scientists around the world.



A Google Earth view of European sites hooked into the EGEE grid computing project

Helps power: Large-scale scientific research projects in fields from geology to chemistry—for example, will analyze data from CERN's Large Hadron Collider, a particle accelerator being built to help investigate details around the Big Bang and related physics questions.

Amount of work it does: 98,000 jobs a day, more than 1 million per month.

Juggling ability: Runs about 30,000 jobs concurrently, on average.

RELATED SOLUTIONS

How-To »

- How to Lay Dead Technologies to Rest
- Ghostly Gear: Technology Tools for Paranormal Investigations

Research & Analysis »

- Lax Laptop Security Can Be Dangerous...and Expensive
- Reduce Information Technology Complexity, Costs with Consolidation

Advice & Opinion »

- Join the Conversation!
- Ask, Answer and Interact with the CIO Community

News »

- Nokia Makes the Ferrari of Phones...Literally
- Holiday Gift Guide 2007: Best Technology Bling

VIDEO

Business Innovation Video Series



IT Leaders are doing it. They're becoming part of the cycle.

Watch the videos »

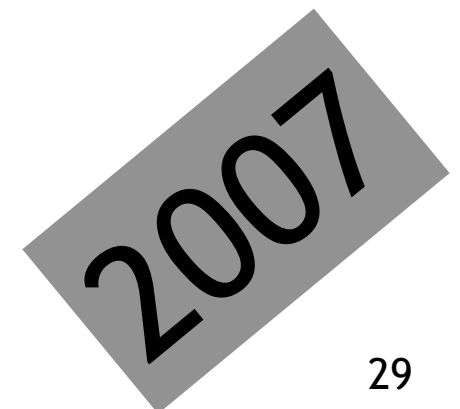
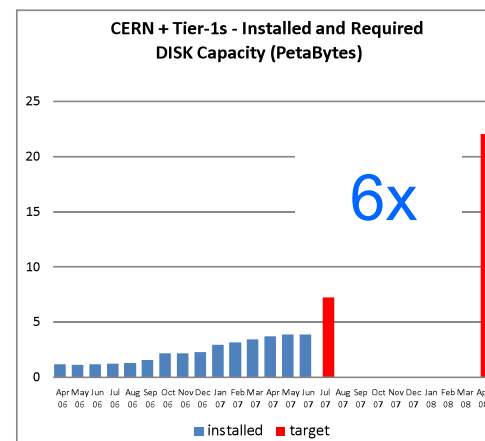
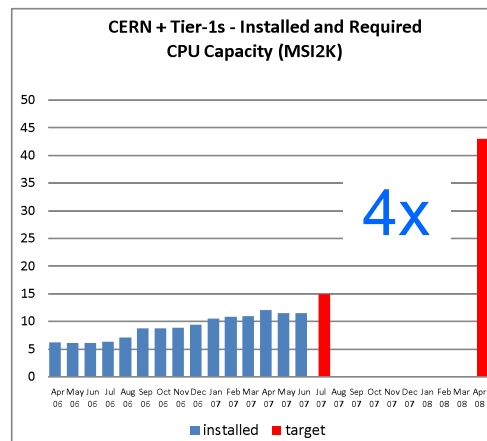




Remaining Challenges

- Creating a working Grid service across multiple infrastructure is clearly a success, but challenges remain
 - Reliability
 - Ramp-up
 - Collaboration
 - From computer centre empires to a federation
 - consensus rather than control

- ...





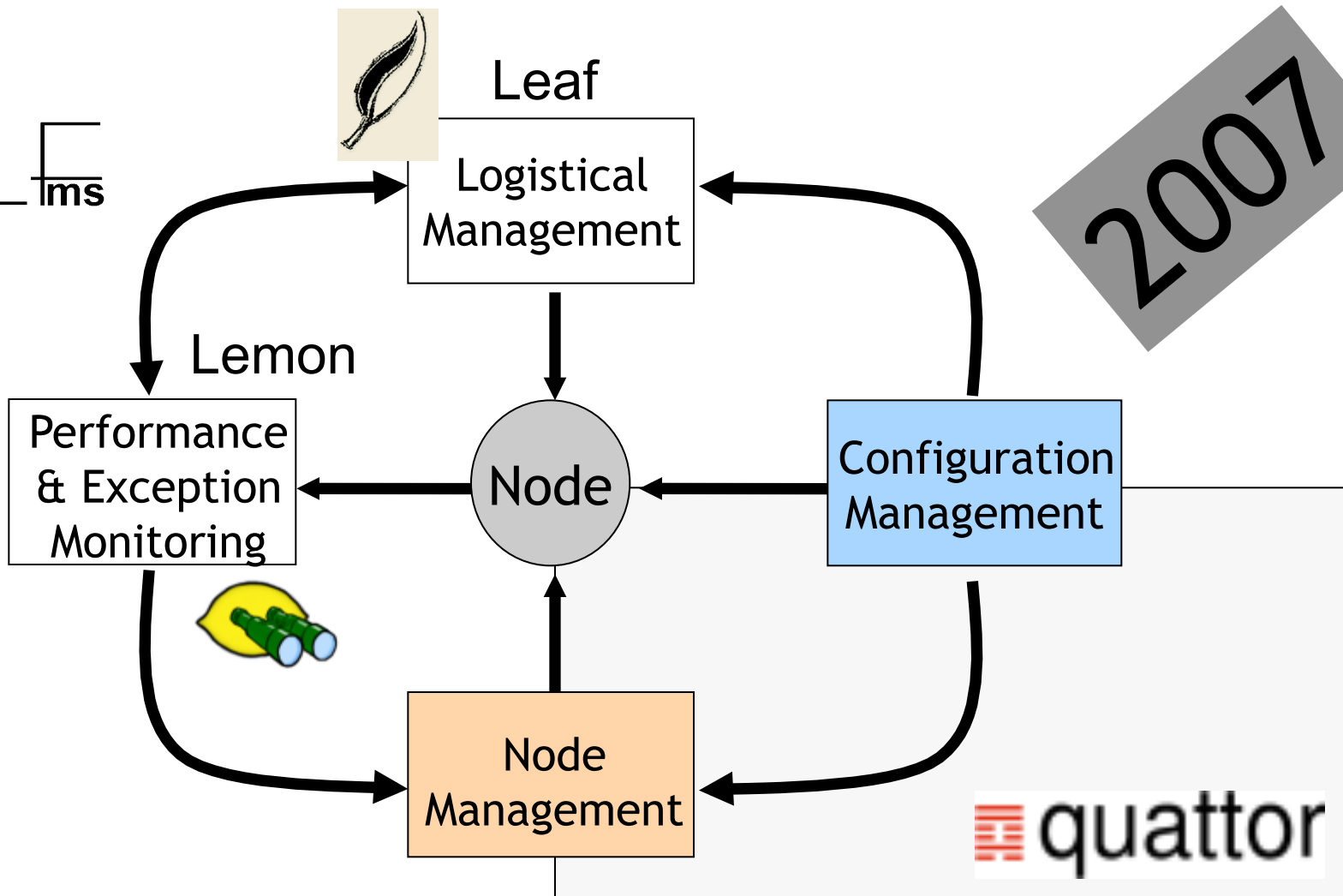
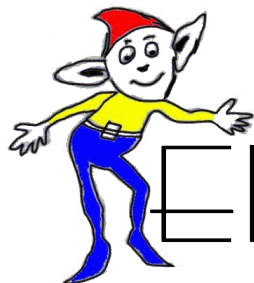
Outline

- Introduction to CERN and Experiments
- LHC Computing
- **Challenges**
 - Capacity Provision
 - **Box Management**
 - Installation & Configuration
 - Monitoring
 - Workflow
 - Data Management and Distribution
 - What's Going On?
- Summary/Conclusion

2007



ELFms Vision



Toolkit developed by CERN in collaboration with many HEP sites and as part of the European DataGrid Project. See <http://cern.ch/ELFms>



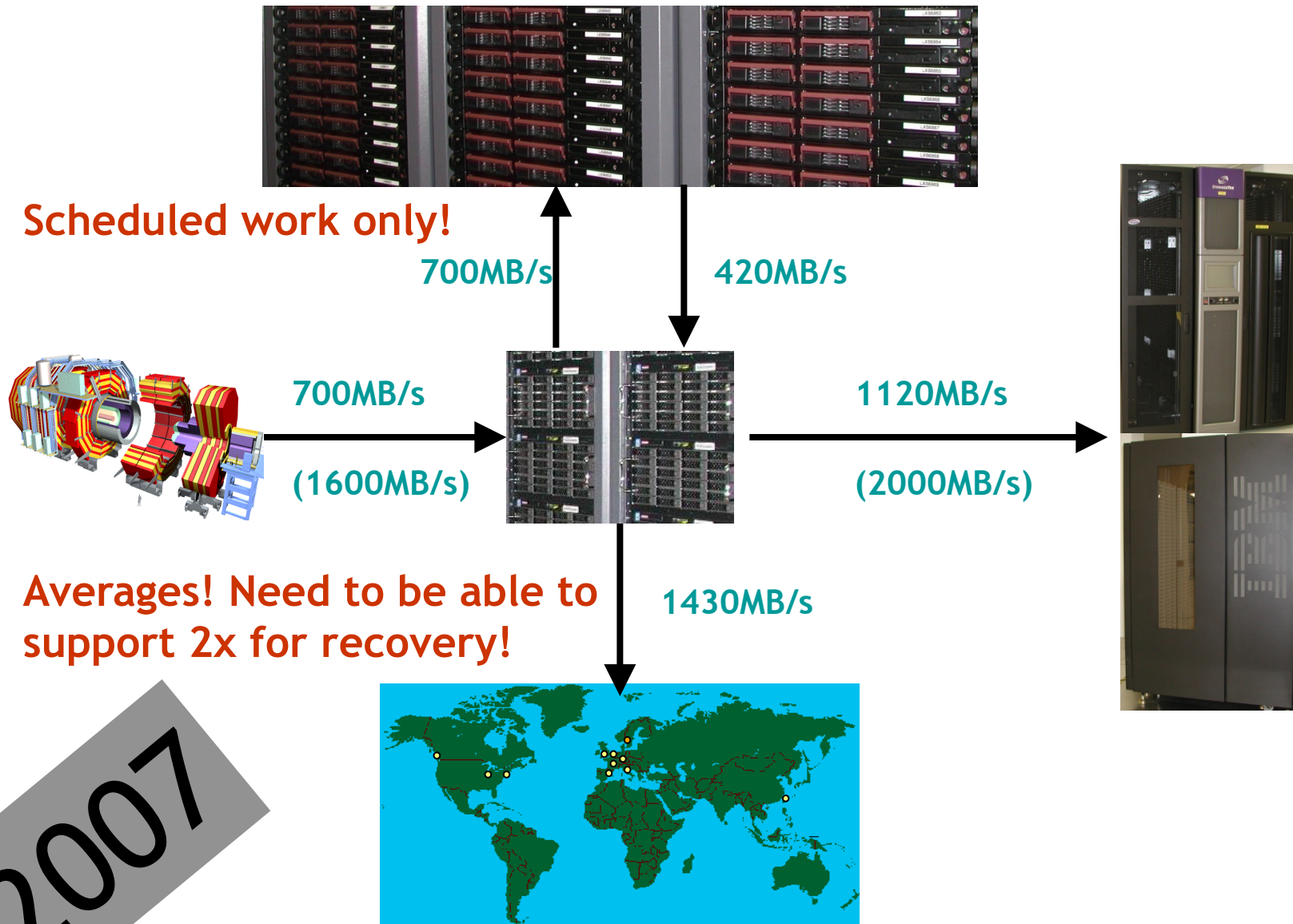
Outline

- Introduction to CERN and Experiments
- LHC Computing
- **Challenges**
 - Capacity Provision
 - Box Management
 - **Data Management and Distribution**
 - What's Going On?
- Summary/Conclusion

2007



Dataflows and rates



2007



Volumes & Rates

- 15PB/year. Peak rate to tape >2GB/s
 - 3 full SL8500 robots/year
- Requirement in first 5 years to reread all past data between runs
 - 60PB in 4 months: 6GB/s
- Can run drives at sustained 80MB/s
 - 75 drives flat out merely for controlled access
- Data Volume has interesting impact on choice of technology
 - Media use is advantageous: high-end technology (3592, T10K) favoured over LTO.

2007



Outline

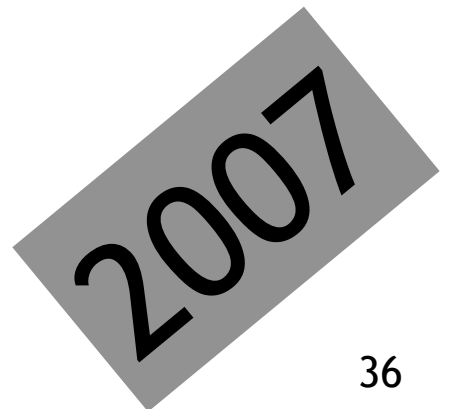
- Introduction to CERN and Experiments
- LHC Computing
- **Challenges**
 - Capacity Provision
 - Box Management
 - Data Management and Distribution
 - **What's Going On?**
- Summary/Conclusion

2007



A Complex Overall Service

- Site managers understand systems (we hope!).
But do they understand the service?
 - and do the users?
 - and what about cross site issues?
 - Are things working?
 - If not, just where is the problem?
 - how many different software components, systems and network service providers are involved in a data transfer site X to site Y?

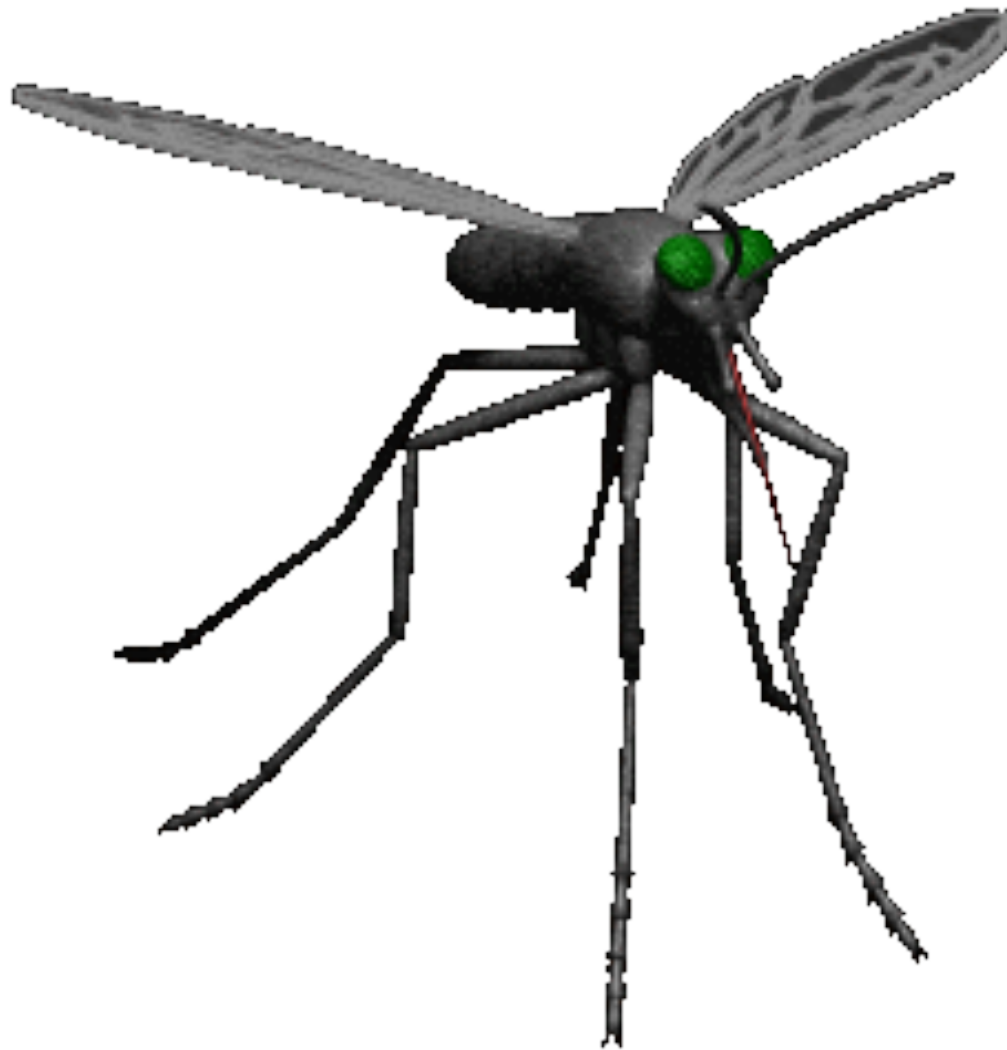




And here's a couple
more...

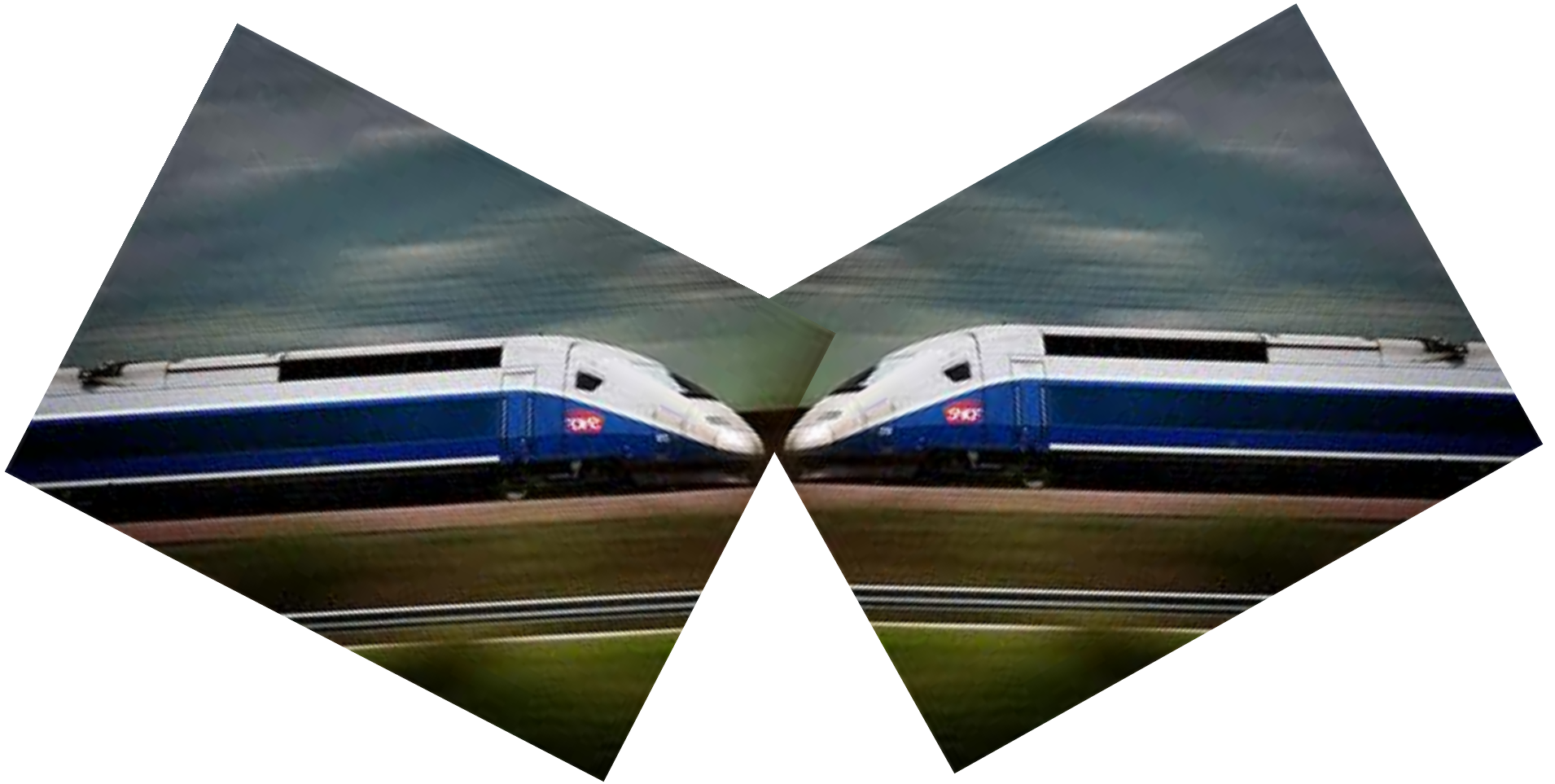


Energy of a 1TeV Proton

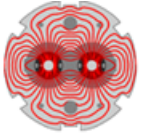




Energy of 7TeV Beams...



Two nominal beams together can melt ~1,000kg of copper.
Current beams: ~100kg of copper.



TT40 Damage during 2004 High Intensity SPS Extraction / [Goddard, B ; Kain, V ; Mertens, V ; Uythoven, J ; Wenninger, J](#)

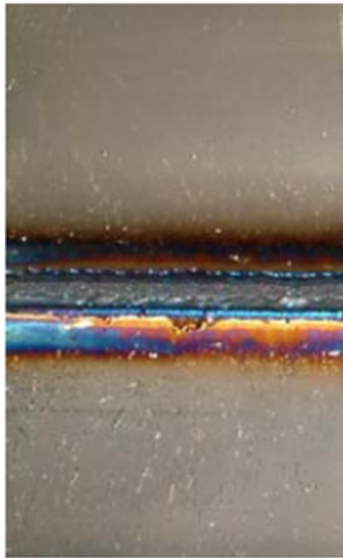
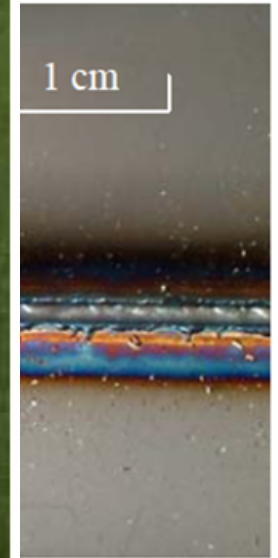
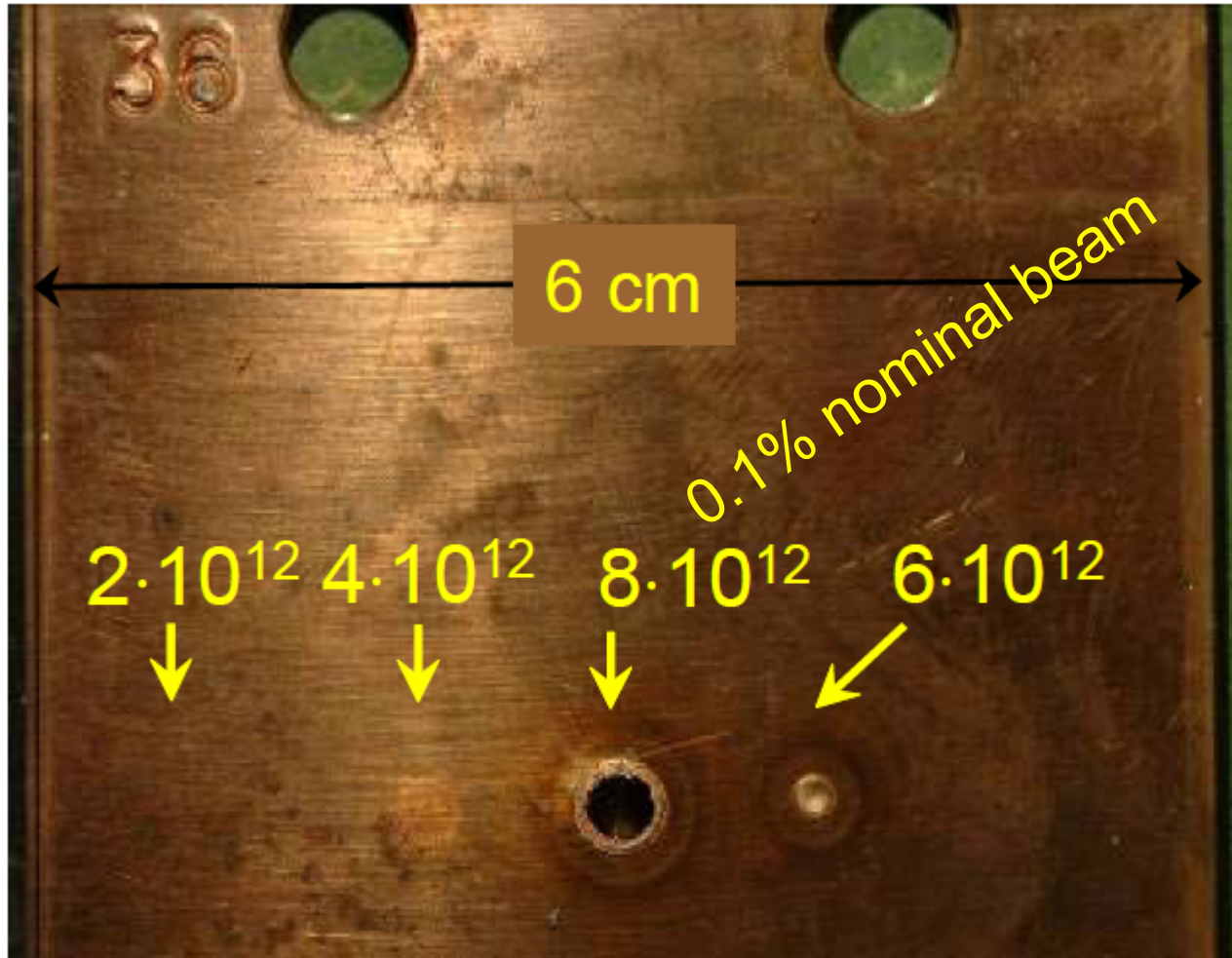


Figure 4. Damage approximately 110 cm from the entrance.

During high vacuum



side. A groove 10 cm from the

ch the ed.

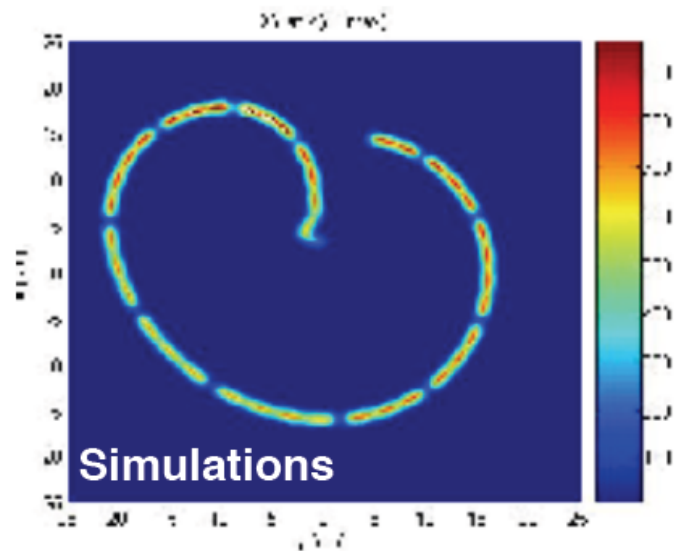
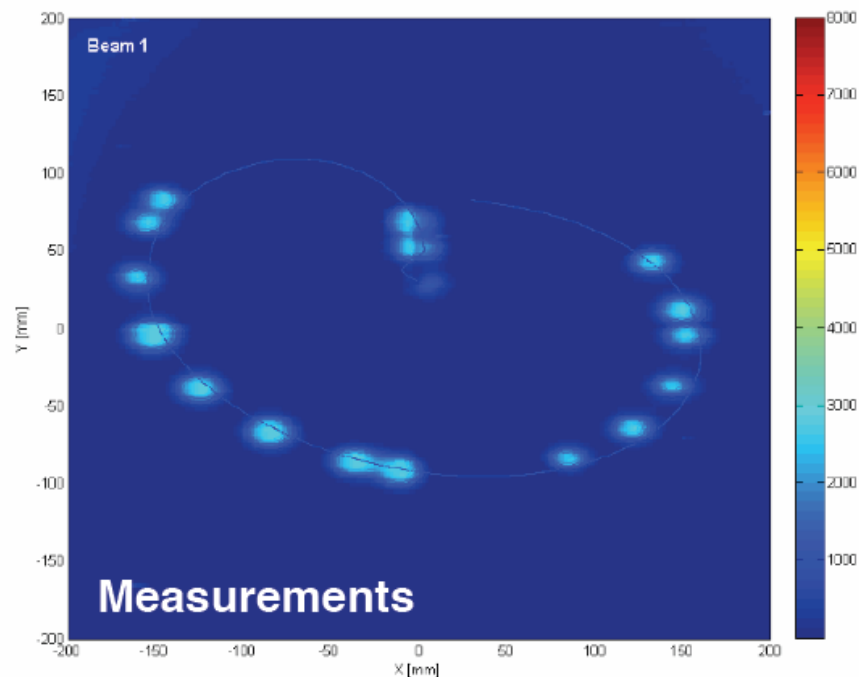
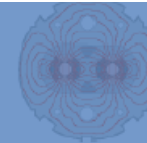
The beam was a 450 GeV full LHC injection batch of $5.4 \cdot 10^{12}$ p+ in 200 bunches, and was extracted from SPS LSS4 with the wrong trajectory

4.4 e12 at 3.5 TeV

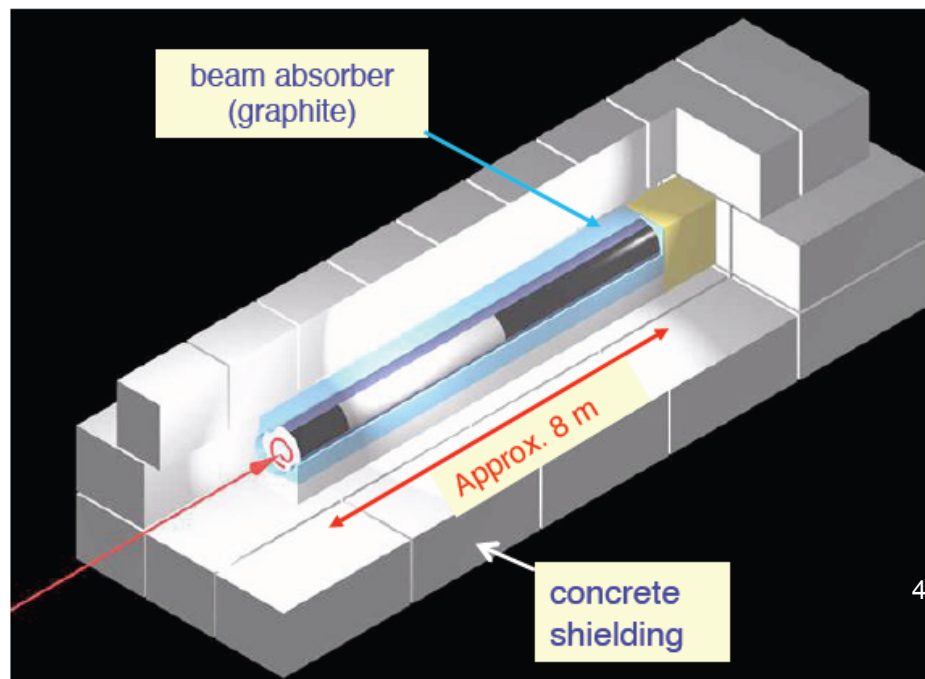


Accelerator “fly by Oracle”

- Three accelerator database applications:
 - Short term settings and control configuration
 - Considered as “any other active component necessary for beam operation”.
 - **No database: no beam**
 - **Lose database: lose beam (controlled!)**
 - Short term (7-day) real-time measurement log
 - Long term (20 yr+) archive of log subset

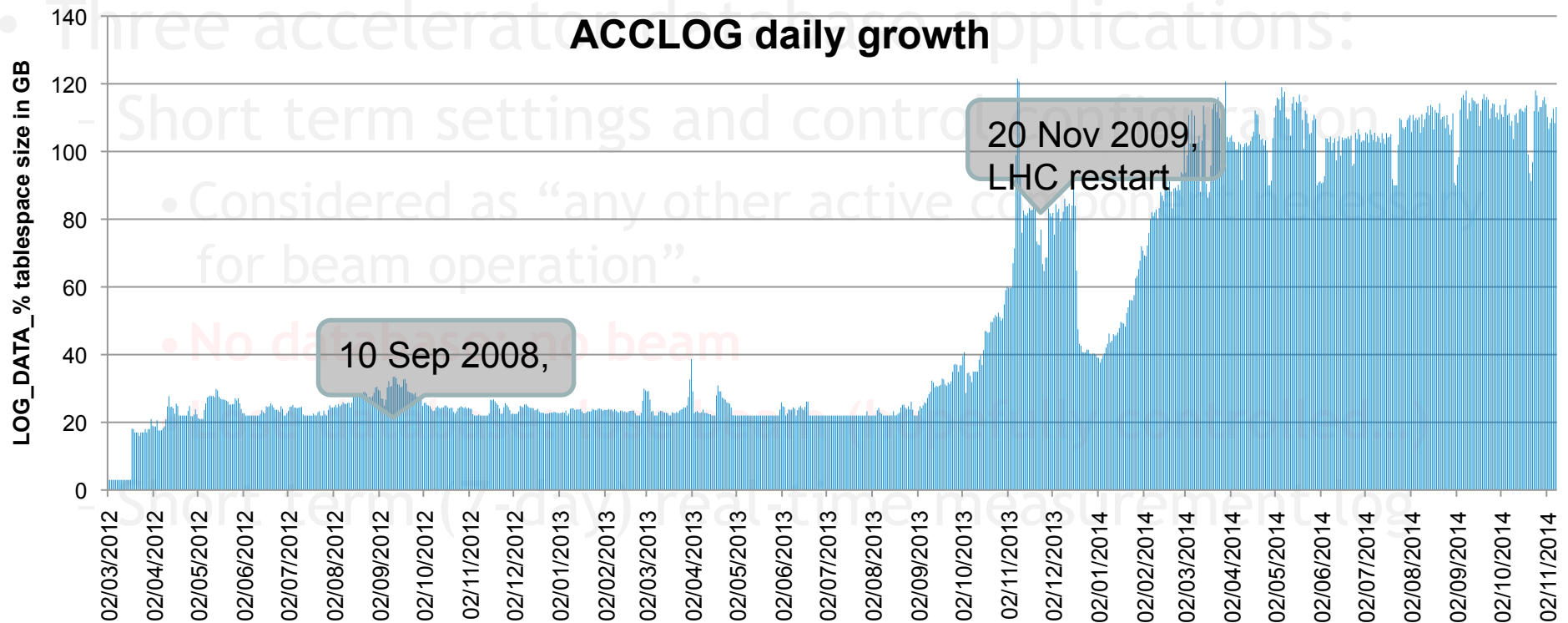


This is the **ONLY** element in the LHC that can withstand the impact of the full 7 TeV beam !
Nevertheless, the dumped beam must be painted to keep the peak energy densities at a tolerable level !

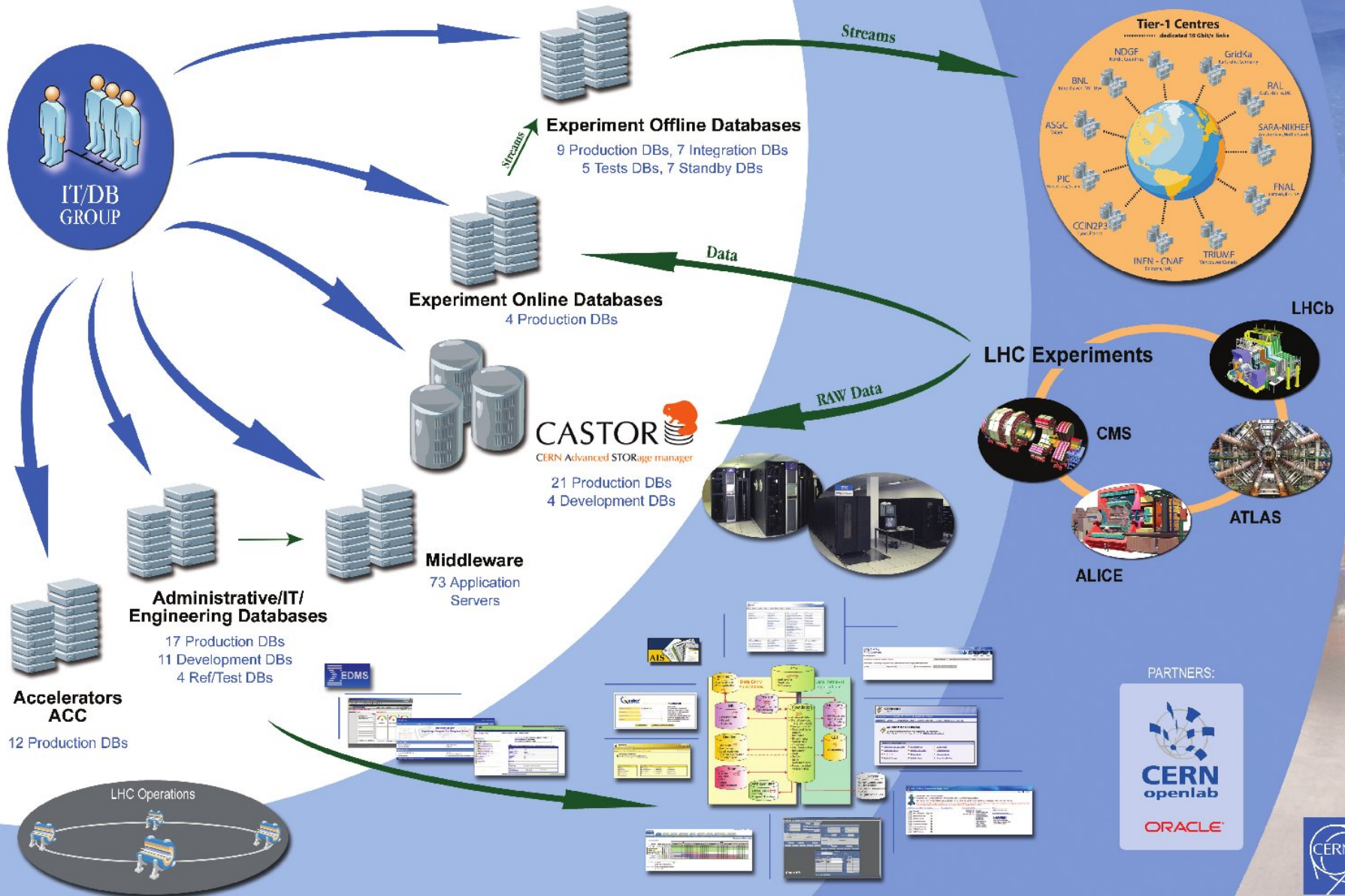




Accelerator “fly by Oracle”



- Long term (20 yr+) archive of log subset
 - ~2,000,000,000,000 rows; ~4,000,000,000/day





Responsibilities & Requirements

A non sleeping 24hr/day 365d/year running system

- ❑ Ensure safe detector operation
 - ❑ anticipating the Detector Safety System (DSS) actions, triggering protection mechanisms on adverse conditions (high temperatures, high humidity, overcurrents, water leaks, electrical trips...)
 - ❑ preventing potentially dangerous actions
 - ❑ issuing alert notifications (alert screen, SMS, control room voice alerts)

- ❑ Provide efficient detector operation
 - ❑ making sure that voltages are present whenever the accelerator conditions allow for physics data taking
 - ❑ guaranteeing that the controlled parameters are stable within their calibrated operating ranges



Control system size

~ 10⁶ control system parameters

System Name	Number of PCs	Monitored Parameters	Controlled Parameters
Tracker	14	350k	20k
Calorimeter	14	115k	2k
Muon	30	435k	30k
Trigger DCS	2	1k	0.5k
Alignment	3	3k	0.5k
Services	35	20k	1k
Total	98	934k	34k

PVSS by ETM (now owned by Siemens)



Main supervisor panel

CMS NOT READY

AUTOMATIC ACTIONS ENABLED

COMPONENT	MODE	PROTECTED	PREPARE BEAM	READY
PIXEL	IN CENTRAL	PROTECTED	PROTECTED	SAFE
STRIPS	IN CENTRAL	PROTECTED	PROTECTED	SAFE
ECAL	IN CENTRAL			NOT READY
HCAL	IN CENTRAL			NOT READY
ZDC	IN CENTRAL	PROTECTED	PROTECTED	SAFE
CASTOR	IN CENTRAL			PHYSICS
DT	IN CENTRAL	PROTECTED	PROTECTED	SAFE
RPC	IN CENTRAL	PROTECTED	PROTECTED	SAFE
CSC	IN CENTRAL			PHYSICS

MACHINE MODE PROTON PHYSICS

MODE	STATUS	READY	
B PIX	0.00%	IN CENTRAL	STANDBY
F PIX	0.00%	IN CENTRAL	STANDBY
TIB/TID	0.00%	IN CENTRAL	STANDBY
TOB	0.00%	IN CENTRAL	STANDBY
TEC+	0.00%	IN CENTRAL	STANDBY
TEC-	0.00%	IN CENTRAL	STANDBY
EE-	100%	READY FOR PHYSICS	ON
EB-	100%	READY FOR PHYSICS	ON
EB+	100%	READY FOR PHYSICS	ON
EE+	100%	READY FOR PHYSICS	ON
ES-	0.00%	STANDBY	STANDBY
ES+	0.00%	STANDBY	STANDBY
HF	100%	IN CENTRAL	ON
HO	100%	IN CENTRAL	OFF
HEHBa	100%	READY FOR PHYSICS	ON
HEHBb	100%	READY FOR PHYSICS	ON
HEHBc	100%	READY FOR PHYSICS	ON
ZDC	??	IN CENTRAL	STANDBY
CASTOR	100%	READY FOR PHYSICS	ON
DT+	0.00%	IN CENTRAL	STANDBY
DTO	0.00%	IN CENTRAL	STANDBY
DT-	0.00%	IN CENTRAL	STANDBY
RPC	0.00%	IN CENTRAL	STANDBY
CSC-	100%	READY FOR PHYSICS	ON
CSC+	100%	READY FOR PHYSICS	ON

LHC HANDSHAKE (LHS)

HANDSHAKE MODE AUTOMATIC

Reaction during handshake for: injection, adjust, dump

No problem reported to LHC

REPORT PROBLEM TO LHC

CLEAR REPORTED PROBLEM

BEAM MODE DIAGRAM

DCS EVENT LOG

10/12/2010 10:55:54 PM - Including CASTOR in CENTRAL

10/12/2010 11:10:03 PM - Machine mode changes to BEAM SETUP

10/12/2010 11:10:06 PM - Beam mode changes to CYCLING

10/12/2010 11:17:10 PM - Central shifter gives command GO_TO_STANDBY to DT

10/12/2010 11:17:22 PM - Central shifter gives command GO_TO_STANDBY to RPC

10/12/2010 11:31:06 PM - Beam mode changes to SETUP

10/12/2010 11:31:36 PM - Handshake changes to INJECT_WARNING

10/12/2010 11:33:19 PM - All subdetectors confirm to be ready for INJECT

10/12/2010 11:33:19 PM - Waiting for the injection inhibit button to be released

10/12/2010 11:34:06 PM - Handshake changes to INJECT_READY (all experiments are ready for INJECT)

10/12/2010 11:41:55 PM - Beam mode changes to INJECTION PROBE BEAM

10/12/2010 11:43:55 PM - Machine mode changes to PROTON PHYSICS

10/12/2010 11:47:13 PM - Injection button moves to INJECTION ALLOWED

10/13/2010 12:06:07 AM - Beam mode changes to INJECTION PHYSICS BEAM

10/13/2010 12:53:45 AM - Beam mode changes to INJECTION PROBE BEAM

10/13/2010 1:34:19 AM - Beam mode changes to INJECTION PHYSICS BEAM

10/13/2010 1:49:53 AM - Rebeca Gonzalez Suarez logs out and releases the tree (from CMS-SCR-DCS-01)

10/13/2010 1:50:14 AM - David Michael Newbold logs in (from CMS-SCR-DCS-01)

10/13/2010 1:50:32 AM - David Michael Newbold logs in (from CMS-SCR-DCS-01)

10/13/2010 2:42:50 AM - David Michael Newbold logs out and releases the tree (from CMS-SCR-DCS-01)

10/13/2010 2:43:11 AM - Rebeca Gonzalez Suarez logs in (from CMS-SCR-DCS-01)

10/13/2010 2:43:26 AM - Rebeca Gonzalez Suarez logs in (from CMS-SCR-DCS-01)

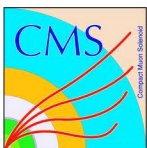
10/13/2010 6:05:11 AM - Beam mode changes to INJECTION PROBE BEAM

10/13/2010 6:52:26 AM - Rebeca Gonzalez Suarez logs out and releases the tree (from CMS-SCR-DCS-01)

10/13/2010 6:52:47 AM - Vladimir Talov logs in (from CMS-SCR-DCS-01)

10/13/2010 6:53:03 AM - Vladimir Talov logs in (from CMS-SCR-DCS-01)

10/13/2010 8:20:11 AM - Beam mode changes to INJECTION PHYSICS BEAM



Main supervisor panel

The image shows a screenshot of the CMS Main supervisor panel. A magnifying glass highlights a detailed view of the PROTON PHYSICS section. The main panel includes:

- CMS STATUS:** NOT READY
- AUTOMATIC ACTIONS:** ENABLED
- BEAM MODES:** CURRENT (INJECTION PHYSICS BEAM), NEXT (PREPARE RAMP), READY (PROTECTED FOR INJECTION PHYSICS BEAM)
- MACHINE MODE:** PROTON PHYSICS
- DESIRED STATE:** PROTECTED, PROTECTED, SAFE
- PIXEL:** IN CENTRAL, STANDBY
- STRIPS:** IN CENTRAL, STANDBY
- Machine Parameters:**

B PIX	0.00%	IN CENTRAL	STANDBY
F PIX	0.00%	IN CENTRAL	STANDBY
TIB/TID	0.00%	IN CENTRAL	STANDBY
- LHC HANDSHAKE (LHS):** Includes a flow diagram showing the process from injection warning to protection ready, and a status bar showing 'No problem reported to LHC'.
- USER ROLE:** CENTRAL_DCS

The magnified view shows the PROTON PHYSICS section in more detail:

- CMS STATUS:** NOT READY
- AUTOMATIC ACTIONS:** ENABLED
- BEAM MODES:** CURRENT (INJECTION PHYSICS BEAM), NEXT (PREPARE RAMP), READY (PROTECTED FOR INJECTION PHYSICS BEAM)
- MACHINE MODE:** PROTON PHYSICS
- DESIRED STATE:** PROTECTED, PROTECTED, SAFE
- PIXEL:** IN CENTRAL, STANDBY
- Machine Parameters:**

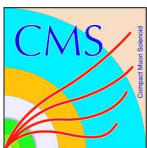
B PIX	0.00%	IN CENTRAL	STANDBY
F PIX	0.00%	IN CENTRAL	STANDBY

At the bottom, there is a table for other detector components and an event log.

Component	Current	Next	Ready	Machine Mode
ZDC	IN CENTRAL	STANDBY	PROTECTED	PROTON PHYSICS
CASTOR	IN CENTRAL	PHYSICS	PROTECTED	PROTON PHYSICS
DT	IN CENTRAL	STANDBY	PROTECTED	PROTON PHYSICS
RPC	IN CENTRAL	STANDBY	PROTECTED	PROTON PHYSICS
CSC	IN CENTRAL	PHYSICS	PROTECTED	PROTON PHYSICS
HEHBc	100%	READY FOR PHYSICS	ON	PROTON PHYSICS
ZDC	??	IN CENTRAL	STANDBY	PROTON PHYSICS
CASTOR	100%	READY FOR PHYSICS	ON	PROTON PHYSICS
DT+	0.00%	IN CENTRAL	STANDBY	PROTON PHYSICS
DT0	0.00%	IN CENTRAL	STANDBY	PROTON PHYSICS
DT-	0.00%	IN CENTRAL	STANDBY	PROTON PHYSICS
RPC	0.00%	IN CENTRAL	STANDBY	PROTON PHYSICS
CSC-	100%	READY FOR PHYSICS	ON	PROTON PHYSICS
CSC+	100%	READY FOR PHYSICS	ON	PROTON PHYSICS

DCS EVENT LOG:

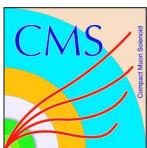
- 10/12/2010 10:55:54 PM - Including CASTOR in CENTRAL
- 10/12/2010 11:10:03 PM - Machine mode changes to BEAM SETUP
- 10/12/2010 11:10:06 PM - Beam mode changes to CYCLING
- 10/12/2010 11:17:10 PM - Central shifter gives command GO_TO_STANDBY to DT
- 10/12/2010 11:17:22 PM - Central shifter gives command GO_TO_STANDBY to RPC
- 10/12/2010 11:31:06 PM - Beam mode changes to SETUP
- 10/12/2010 11:31:36 PM - Handshake changes to INJECT_WARNING
- 10/12/2010 11:33:19 PM - All subdetectors confirm to be ready for INJECT
- 10/12/2010 11:33:19 PM - Waiting for the injection inhibit button to be released
- 10/12/2010 11:34:06 PM - Handshake changes to INJECT_READY (all experiments are ready for INJECT)
- 10/12/2010 11:41:55 PM - Beam mode changes to INJECTION PROBE BEAM
- 10/12/2010 11:43:55 PM - Machine mode changes to PROTON PHYSICS
- 10/12/2010 11:47:13 PM - Injection button moves to INJECTION ALLOWED
- 10/13/2010 12:06:07 AM - Beam mode changes to INJECTION PHYSICS BEAM
- 10/13/2010 12:53:45 AM - Beam mode changes to INJECTION PROBE BEAM
- 10/13/2010 1:34:19 AM - Beam mode changes to INJECTION PHYSICS BEAM
- 10/13/2010 1:49:53 AM - Rebeca Gonzalez Suarez logs out and releases the tree (from CMS-SCR-DCS-01)
- 10/13/2010 1:50:14 AM - David Michael Newbold logs in (from CMS-SCR-DCS-01)
- 10/13/2010 1:50:32 AM - David Michael Newbold logs in (from CMS-SCR-DCS-01)
- 10/13/2010 2:42:50 AM - David Michael Newbold logs out and releases the tree (from CMS-SCR-DCS-01)
- 10/13/2010 2:43:11 AM - Rebeca Gonzalez Suarez logs in (from CMS-SCR-DCS-01)
- 10/13/2010 2:43:26 AM - Rebeca Gonzalez Suarez logs in (from CMS-SCR-DCS-01)
- 10/13/2010 6:05:11 AM - Beam mode changes to INJECTION PROBE BEAM
- 10/13/2010 6:52:26 AM - Rebeca Gonzalez Suarez logs out and releases the tree (from CMS-SCR-DCS-01)
- 10/13/2010 6:52:47 AM - Vladimir Talov logs in (from CMS-SCR-DCS-01)
- 10/13/2010 6:53:03 AM - Vladimir Talov logs in (from CMS-SCR-DCS-01)
- 10/13/2010 8:20:11 AM - Beam mode changes to INJECTION PHYSICS BEAM



Main supervisor panel

The screenshot displays the main supervisor panel for the Compact Muon Solenoid (CMS) detector control system. It is divided into several functional areas:

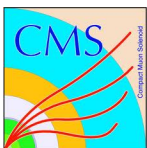
- Machine Status Table:** A large table on the left lists various detector components (PIXEL, STRIPS, ECAL, HCAL, ZDC, CASTOR, DT, RPC, CSC) and their status across different beam modes (CURRENT, NEXT, READY) and machine modes (PROTON PHYSICS).
- LHC Handshake (LHS) Flowchart:** A central flowchart illustrates the sequence of operations between the CMS and the LHC. It includes states like 'standby', 'injection warning', 'injection imminent', 'injection started', 'injection finished', 'protection ready', and 'All exper. ready'. A magnifying glass highlights this flowchart.
- Handshake Mode Panel:** A panel on the right allows the operator to select the handshake mode (currently 'AUTOMATIC') and provides buttons for 'injection', 'adjust', and 'dump'. It also includes a 'REPORT PROBLEM TO LHC' button and a 'CLEAR REPORTED PROBLEM' button.
- Log Window:** A bottom-right window displays a real-time log of system events, such as 'Handshake changes to INJECT_WARNING' and 'Waiting for the injection inhibit button to be released'.



Main supervisor panel

The screenshot displays the CMS Main supervisor panel, which is divided into several functional areas:

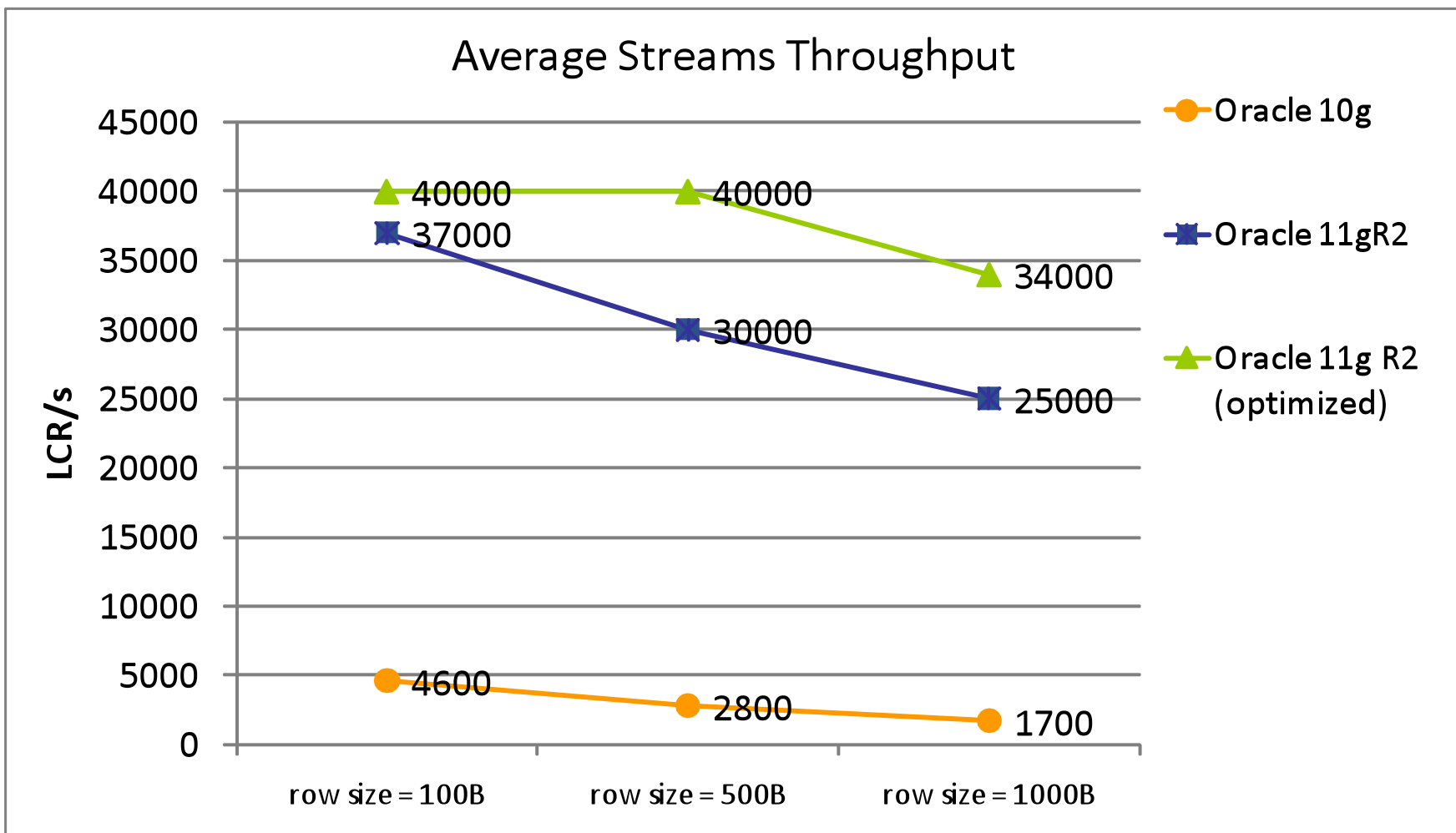
- Top Left:** CMS status (NOT READY) and AUTOMATIC ACTIONS (ENABLED).
- Top Center:** BEAM MODES (CURRENT, NEXT, READY) and MACHINE MODE (PROTON PHYSICS).
- Top Right:** USER ROLE (CENTRAL DCS) and LHC HANDSHAKE (LHS) status.
- Left Panel:** A list of detector components (PIXEL, STRIPS, ECAL, HCAL, ZDC, CASTOR, DT, RPC, CSC) with their current and desired states (e.g., IN CENTRAL, STANDBY, PROTECTED, SAFE).
- Center Panel:** A detailed status table for various machine components (B PIX, F PIX, TIB/TID, TOB, TEC+, EE-, EB-, EB+, EE+, ES-, ES+, HF, HO, HEHBa, HEHBb, HEHBc, ZDC, CASTOR, DT+, DTO, DT-, DT-, RPC, CSC-, CSC+), showing their current status and readiness for physics.
- Right Panel:** LHC HANDSHAKE (LHS) flowchart and BEAM MODE DIAGRAM. The flowchart shows the sequence from LHC standby to injection, protection, and finally to 'All exper. ready'. The BEAM MODE DIAGRAM illustrates transitions between states like ABORT, SETUP, RECOVERY, RAMP DOWN, CYCLING, UNSTABLE BEAMS, STABLE BEAMS, and NO BEAM.
- Bottom Right:** A log window showing system events and user actions, such as 'Rebecca Gonzalez Suarez logs out and releases the tree' and 'Beam mode changes to INJECTION PHYSICS BEAM'.



Main supervisor panel

The screenshot displays the CMS DCS supervisor panel. At the top, it shows the system status as 'NOT READY' and 'AUTOMATIC ACTIONS' as 'ENABLED'. The 'BEAM MODES' section includes 'CURRENT', 'NEXT', and 'READY' states for 'INJECTION PHYSICS BEAM' and 'PREPARE RAMP'. The 'MACHINE MODE' is set to 'PROTON PHYSICS'. A 'DCS EVENT LOG' window is open, showing a list of events with timestamps and descriptions, such as 'Machine mode changes to BEAM SETUP' and 'Waiting for the injection inhibit button to be released'. The 'LHC HANDSHAKE (LHS)' section shows a flowchart of the handshake process. The 'HANDSHAKE MODE' is set to 'AUTOMATIC'. The 'NO BEAM' section shows a flowchart of the beam status. The 'DCS EVENT LOG' at the bottom right shows a detailed list of events, including 'Injection button moves to INJECTION ALLOWED'.

DCS EVENT LOG
10/12/2010 10:53:54 PM - Including CASTOR in CENTRAL
10/12/2010 11:10:03 PM - Machine mode changes to BEAM SETUP
10/12/2010 11:10:06 PM - Beam mode changes to CYCLING
10/12/2010 11:17:10 PM - Central shifter gives command GO_TO_STANDBY to DT
10/12/2010 11:17:22 PM - Central shifter gives command GO_TO_STANDBY to RPC
10/12/2010 11:31:06 PM - Beam mode changes to SETUP
10/12/2010 11:31:36 PM - Handshake changes to INJECT_WARNING
10/12/2010 11:33:19 PM - All subdetectors confirm to be ready for INJECT
10/12/2010 11:33:19 PM - Waiting for the injection inhibit button to be released
10/12/2010 11:34:06 PM - Handshake changes to INJECT_READY (all experiments are ready for INJECT)
10/12/2010 11:41:55 PM - Beam mode changes to INJECTION PROBE BEAM



database code
hardware configuration

3. understand why time is spent on this event

Figure 2. Iterative methodology for performance tuning.

x 150

select ... from temp

x 6

Figure 4. description of the architecture changes. Test setup with a 6 nodes RAC and 150 clients (kplus) each generating 1 000 changes.

Sallar (CERN IT-OO), Chris Lambert, Nilo Segura Chinchilla (CERN IT-DES), Luca Canali (CERN IT-PSS), Svetozar Kapusta (ALICE experiment), Bjorn Engsig and Lohar Platz (Oracle).



Outline

- Introduction to CERN, LHC and Experiments
- The LHC Computing Challenge
- Preparation
- **Reality**
- Future Outlook
- Summary/Conclusion



Praise for a Sys Admin?

[Silence]

An impressive start

- LHC commissioning process in an unprecedented pace
- Experiments showed their readiness for the exploitation of the 7 TeV data...
- ...ready to follow with more complex triggers in the face of the increase of luminosity.
- Analyses proceeding very rapidly in all experiments and results being submitted for publication within days
- **Brilliant performances of the WLCG a key factor in the spectacular startup.**

CERN Research Director
in report to
CERN's Member State Representatives



2004

GRID:

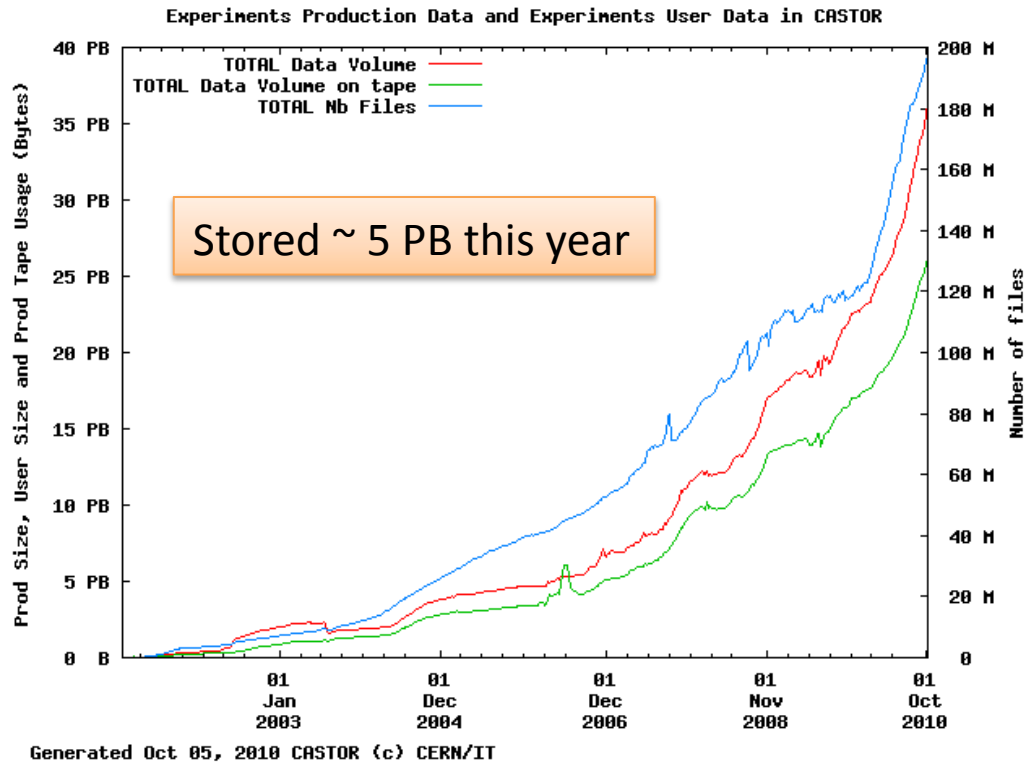
- should allow everybody to do physics anywhere anytime
- should not slow down delivery of physics results
- must be transparent to the users
- user support must be available 24 hours x 365 days x 15 years
- basic tools (e.g. easy-to-understand diagnostic about failed jobs, job monitoring histos, status of job and resources) should be available to the end-users

Fabiola Gianotti (ATLAS spokesperson)

*More striking still is the **speed with which the raw data are being processed**. The freshest batch emerged from the LHC on July 18th and were moulded into meaningful results by July 21st, in time for the Paris conference. Not long ago this process would have taken weeks, says **Fabiola Gianotti**, the spokeswoman for ATLAS, one of the four main LHC experiments. One reason is the development of the Grid, a computing network CERN hopes will prove a worthy successor to its previous invention, the World Wide Web. **The Grid lets centres around the world crunch the numbers as soon as they come out of the machine.***

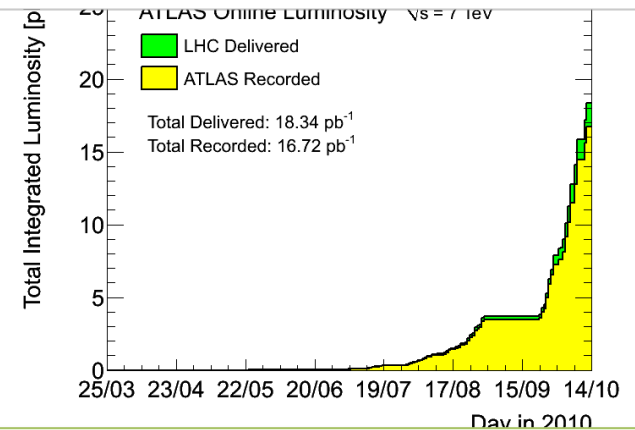
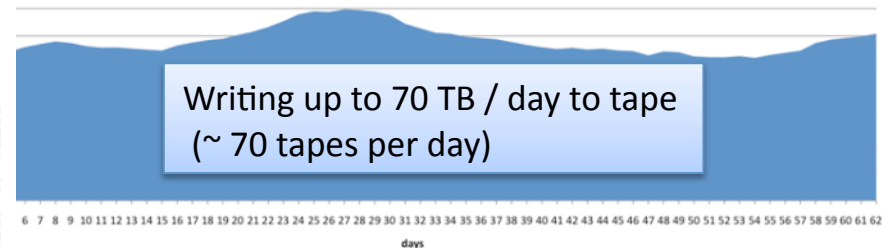
The Economist, 29th July

6 months of LHC data

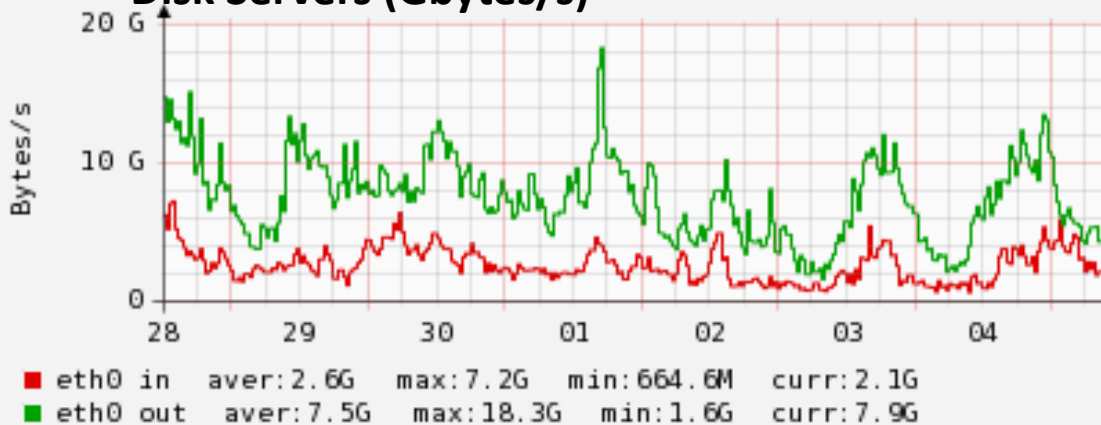


Data written to tape (Gbytes/day)

Rate of tape consumption (GiB/day), last 2 months



Disk Servers (Gbytes/s)

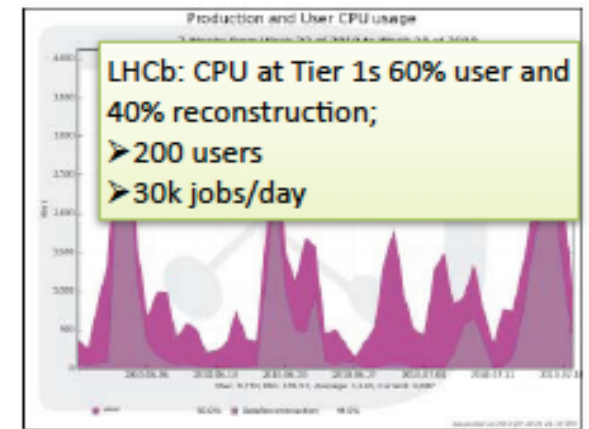
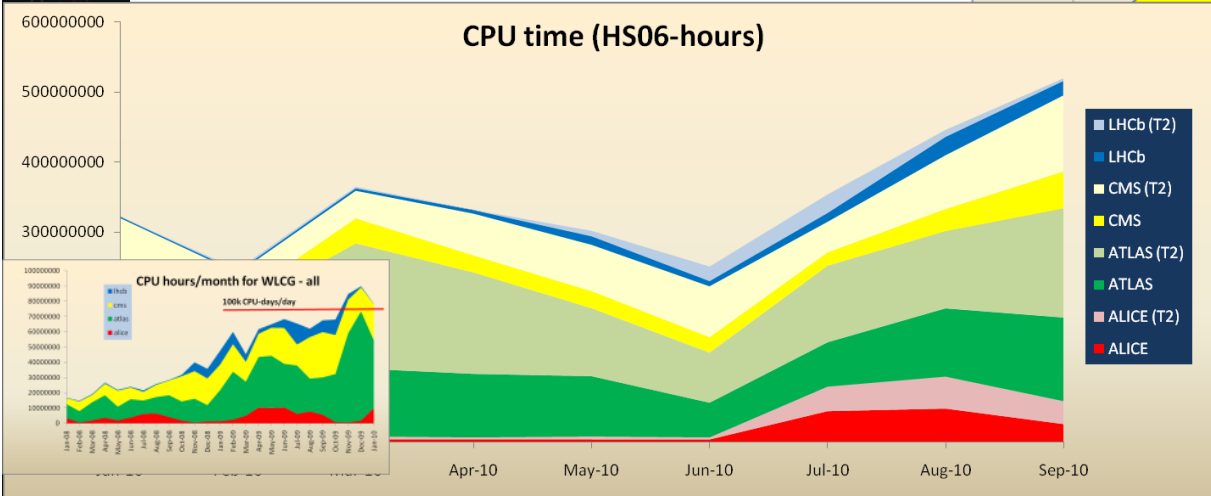
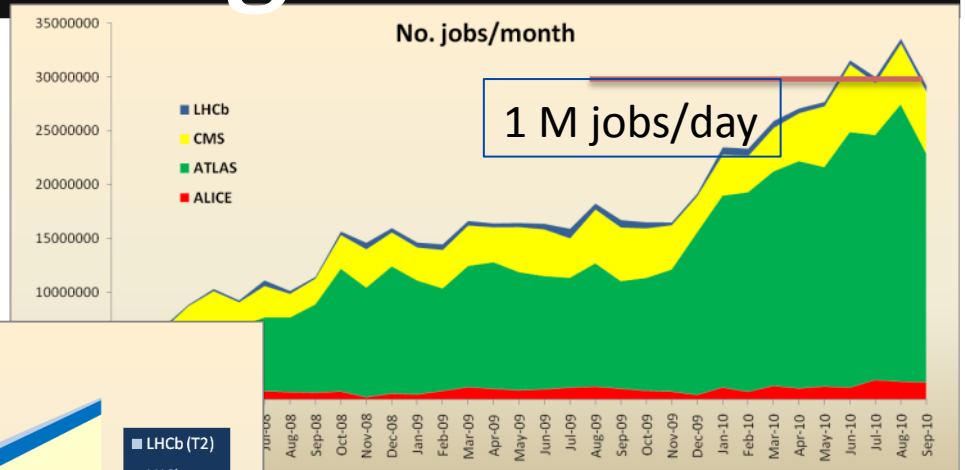


Tier 0 storage:

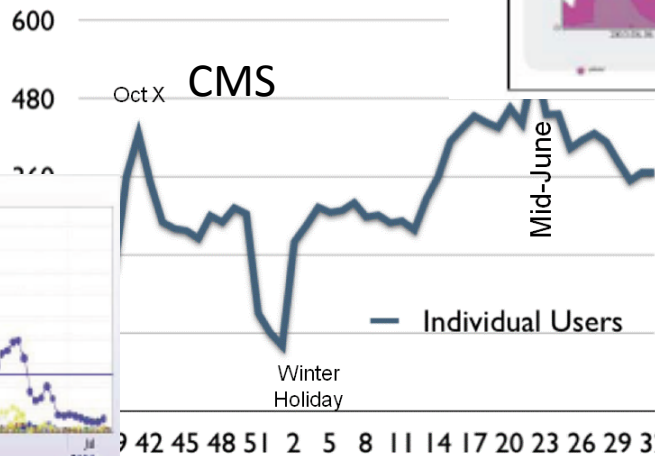
- Accepts data at average of 2.6 GB/s; peaks > 7 GB/s
- Serves data at average of 7 GB/s; peaks > 18 GB/s
- **CERN Tier 0 moves ~ 1 PB data per day**

WLCG Usage

- Use remains consistently high
 - 1 M jobs/day; >>100k CPU-days/day
 - Actually much more inside pilot jobs



As well as LHC data, large simulation productions ongoing



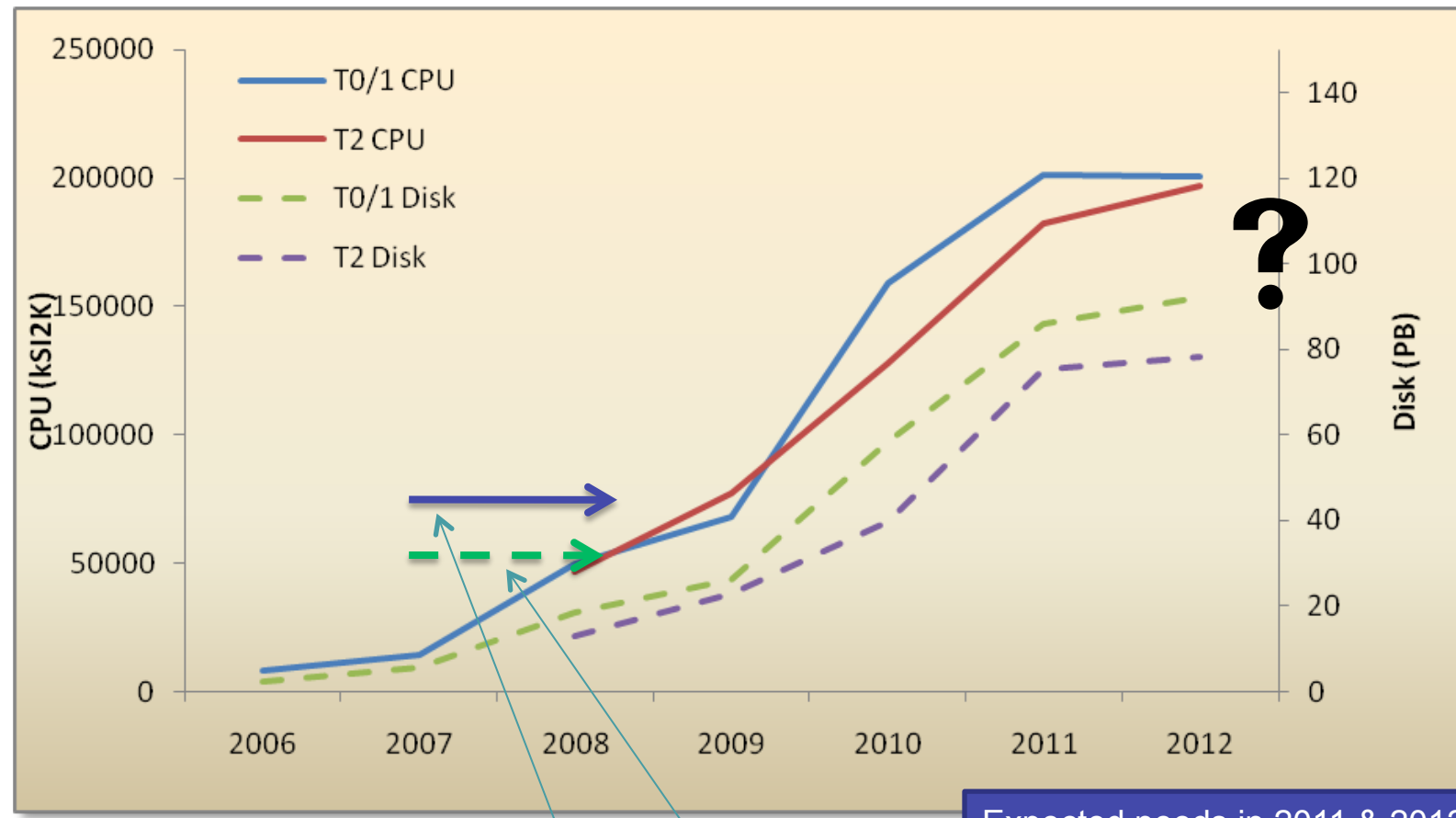
Large numbers of analysis users
 CMS ~800,
 ATLAS ~1000,
 LHCb/ALICE ~200





The 2007 challenges

- Ramp-up



Need foreseen @ TDR for T0+1 CPU and Disk for 1st nominal year

Expected needs in 2011 & 2012



Hardware ramp-up

Summary of Computing Resource Requirements

All experiments - 2008

From LCG TDR - June 2005

	<i>CERN</i>	<i>All Tier-1s</i>	<i>All Tier-2s</i>	<i>Total</i>
CPU (MSPECint2000s)	25	56	61	142
Disk (PetaBytes)	7	31	19	57
Tape (PetaBytes)	18	35		53

~2,500 PCs

Another ~1,500 boxes

Country	Federation	Physical CPU	Logical CPU	HEPSPEC06	CPU Pledge	Total Online Storage (GB)	Disk Pledge	Total Nearline Storage (GB)	Tape Pledge
Switzerland	CH-CERN	4,496	17,644	197,308	233,400	18,181,259	14,790,000	30,957,137	31,600,000
Total		4,496	17,644	197,308	233,400	18,181,259	14,790,000	30,957,137	31,600,000
All Tier-2s									
Canada	CA-TRIUMF	477	1,260	13,975	10,800	1,100,000	1,095,000	750,000	710,000
France	FR-CCIN2P3	1,614	9,072	78,019	44,186	5,608,846	5,109,000	22,534,710	5,300,000
Germany	DE-KIT	2,626	9,796	89,690	58,730	6,855,040	6,924,000	9,292,000	8,932,000
Italy	IT-INFN-CNAF	2,252	8,192	85,516	44,000	12,796,781	5,300,000	9,510,000	5,450,000
Netherlands	NL-T1	941	4,512	48,241	47,296	4,073,412	4,186,000	1,405,702	3,629,000
Nordic	NDGF	10,660	10,660	60,890	13,710	3,166,358	1,720,000	1,464,000	1,770,000
Spain	ES-PIC	592	2,368	24,144	17,238	3,106,092	1,968,000	3,804,697	2,136,000
Taiwan	TW-ASGC	992	3,968	40,873	28,000	4,800,000	3,500,000	4,266,000	3,500,000
UK	UK-T1-RAL	1,300	5,200	49,140	44,376	5,876,473	4,638,000	4,085,877	4,877,000
USA	US-FNAL-CMS	1,692	6,768	44,400	44,400	4,100,000	4,100,000	11,000,000	11,000,000
USA	US-T1-BNL	0	0	58,000	49,680	6,100,000	5,037,000	4,000,000	3,266,000
Total		23,146	61,796	592,888	402,416	57,583,002	43,577,000	72,112,986	50,570,000

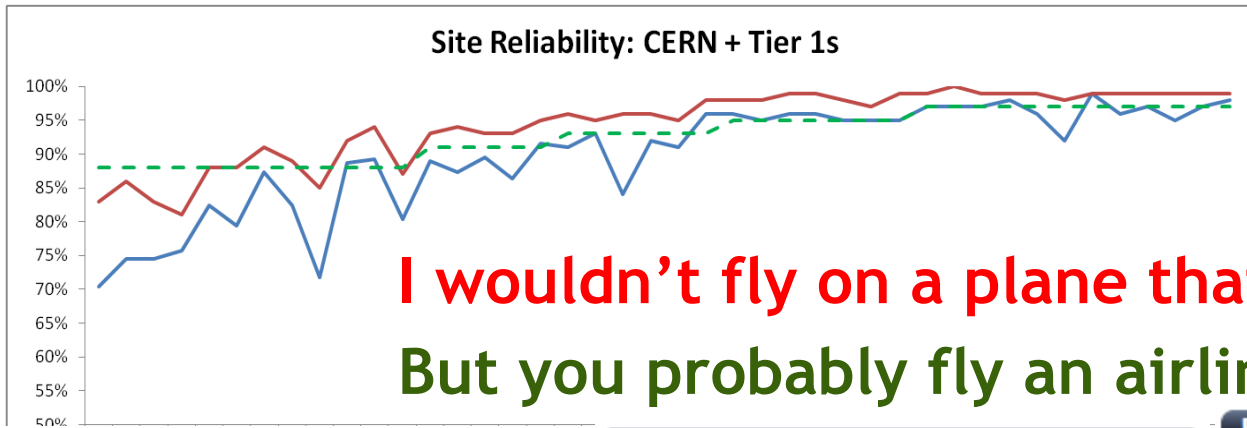
4,000HS06 = 1MSPECint2000

00

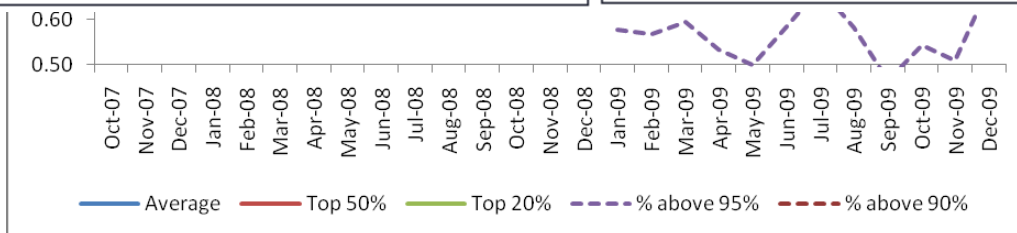
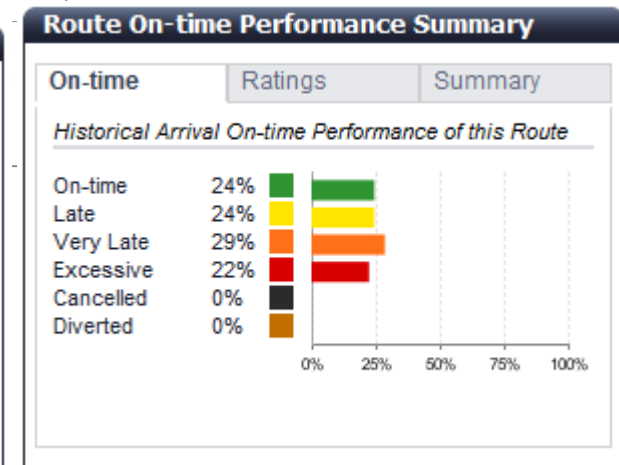
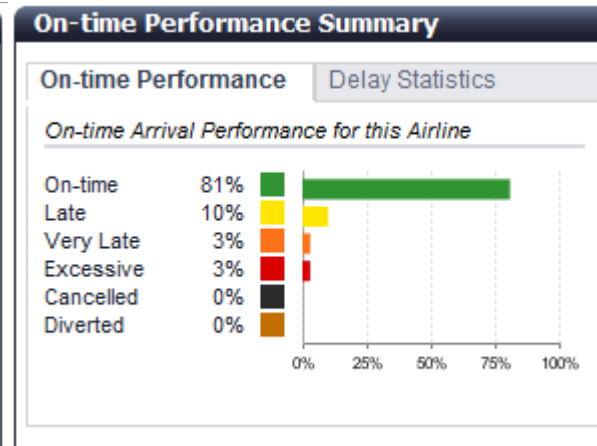
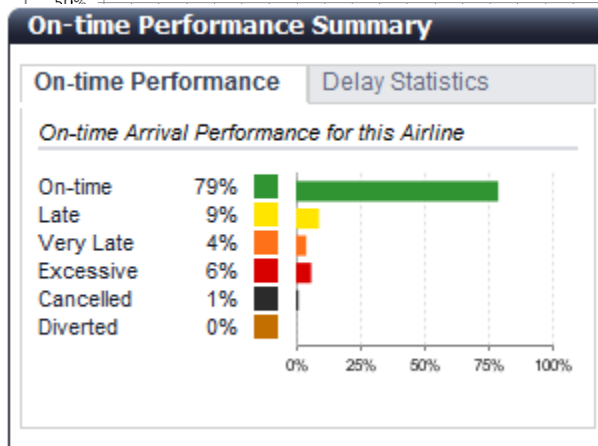


The 2007 challenges

• Reliability



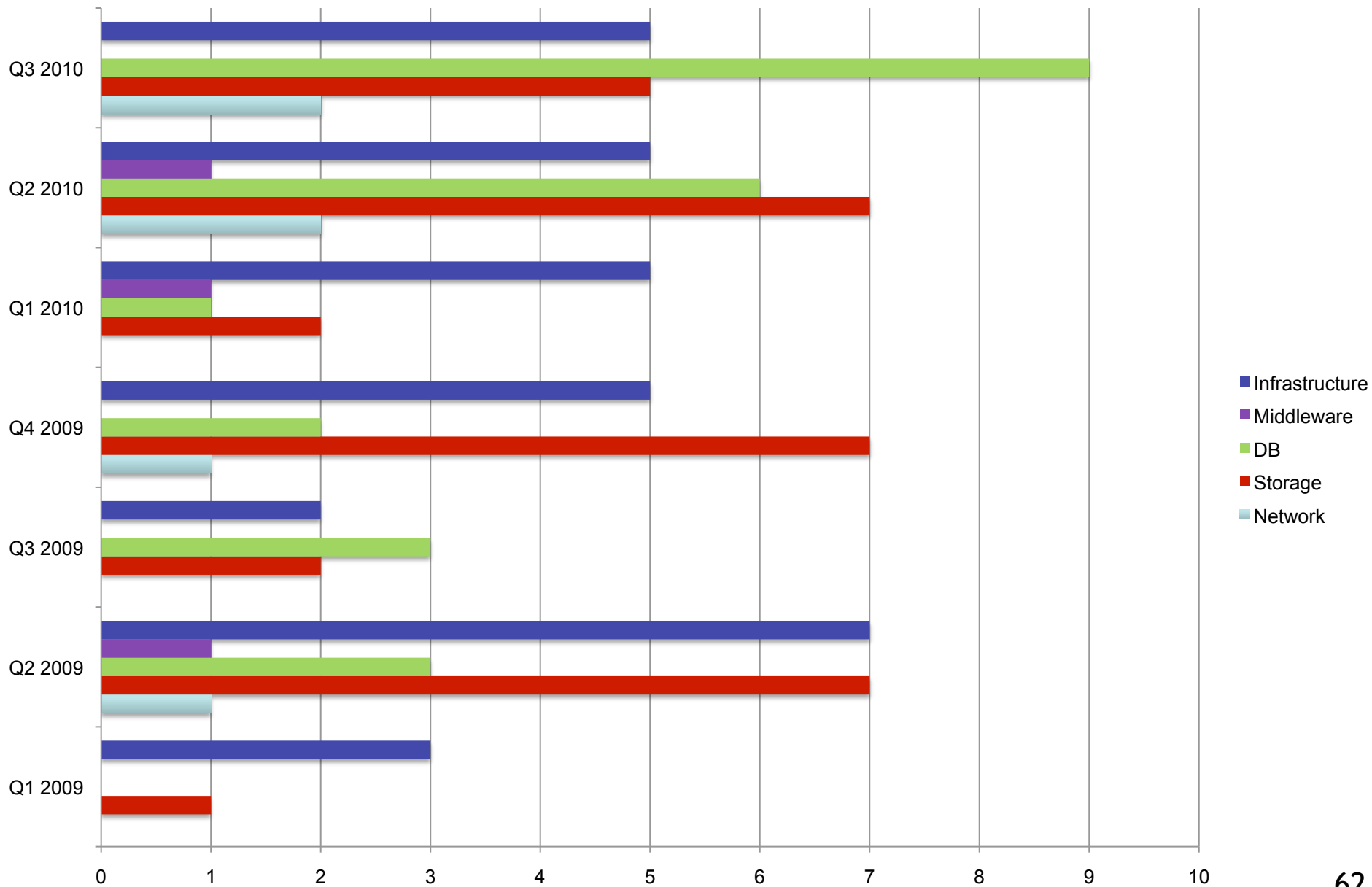
I wouldn't fly on a plane that was 98% reliable!!!!
But you probably fly an airline that is...



Punctuality details from flightstats.com

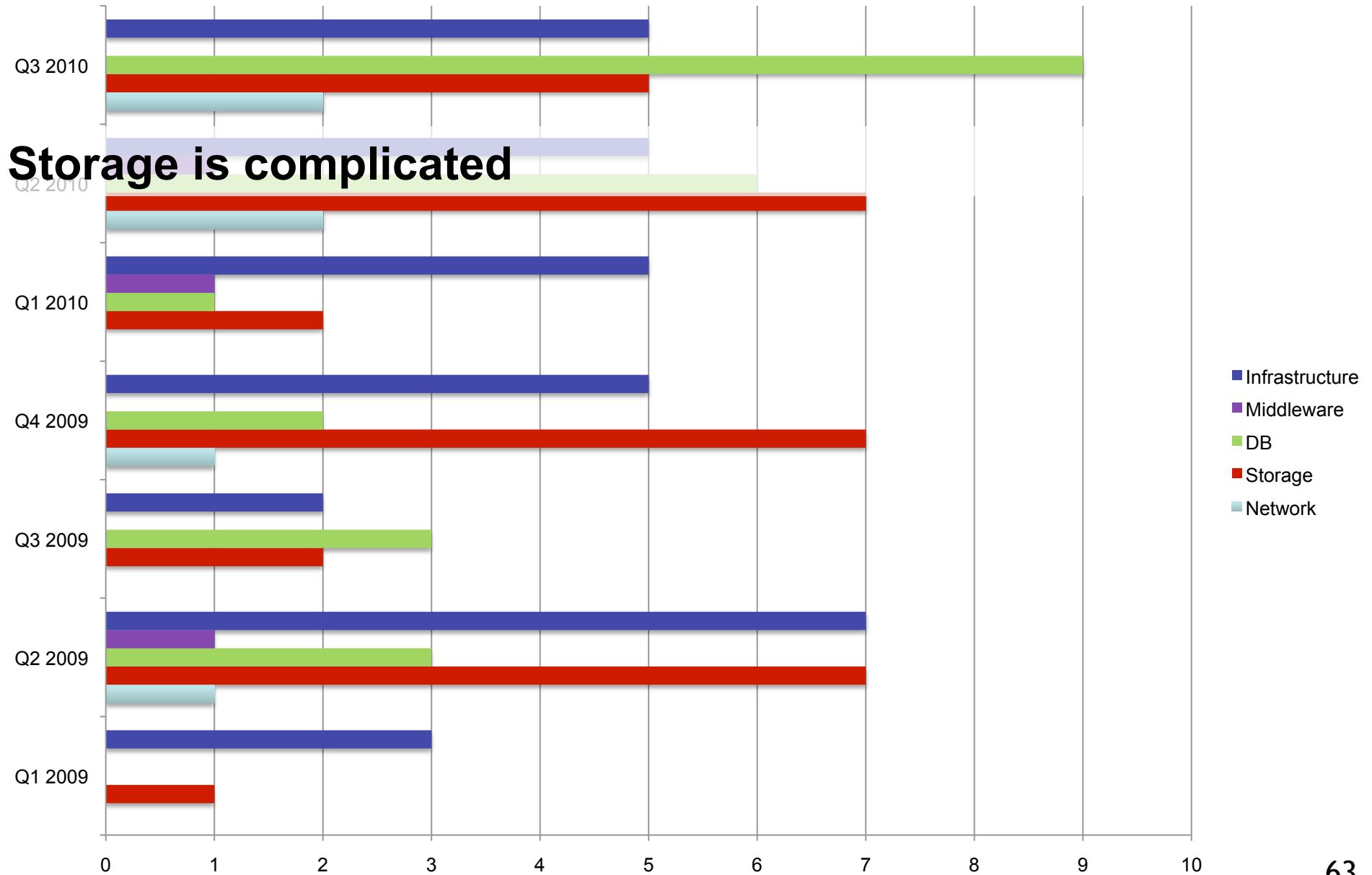


Operational Issues - I



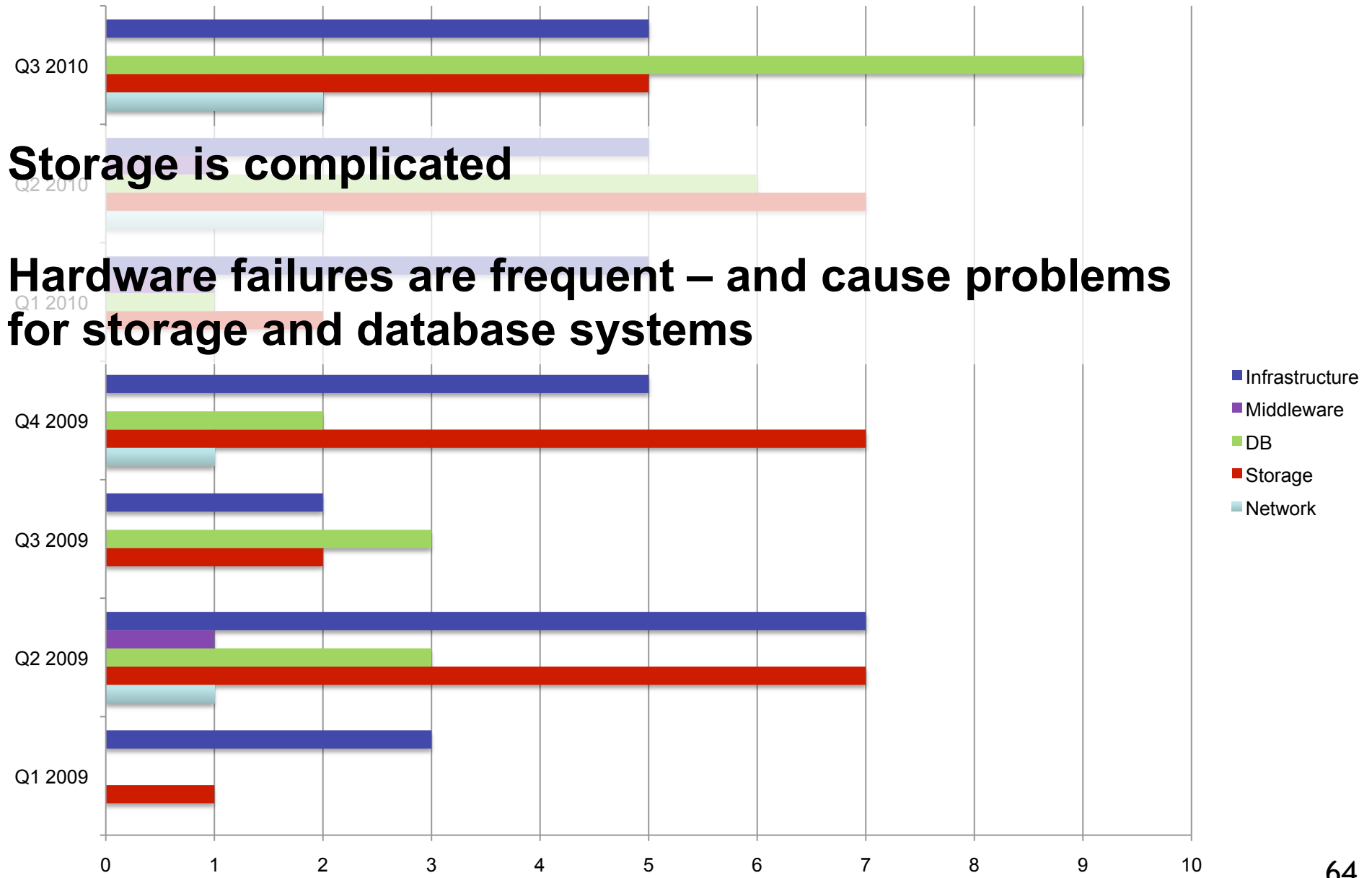


Operational Issues - I





Operational Issues - I



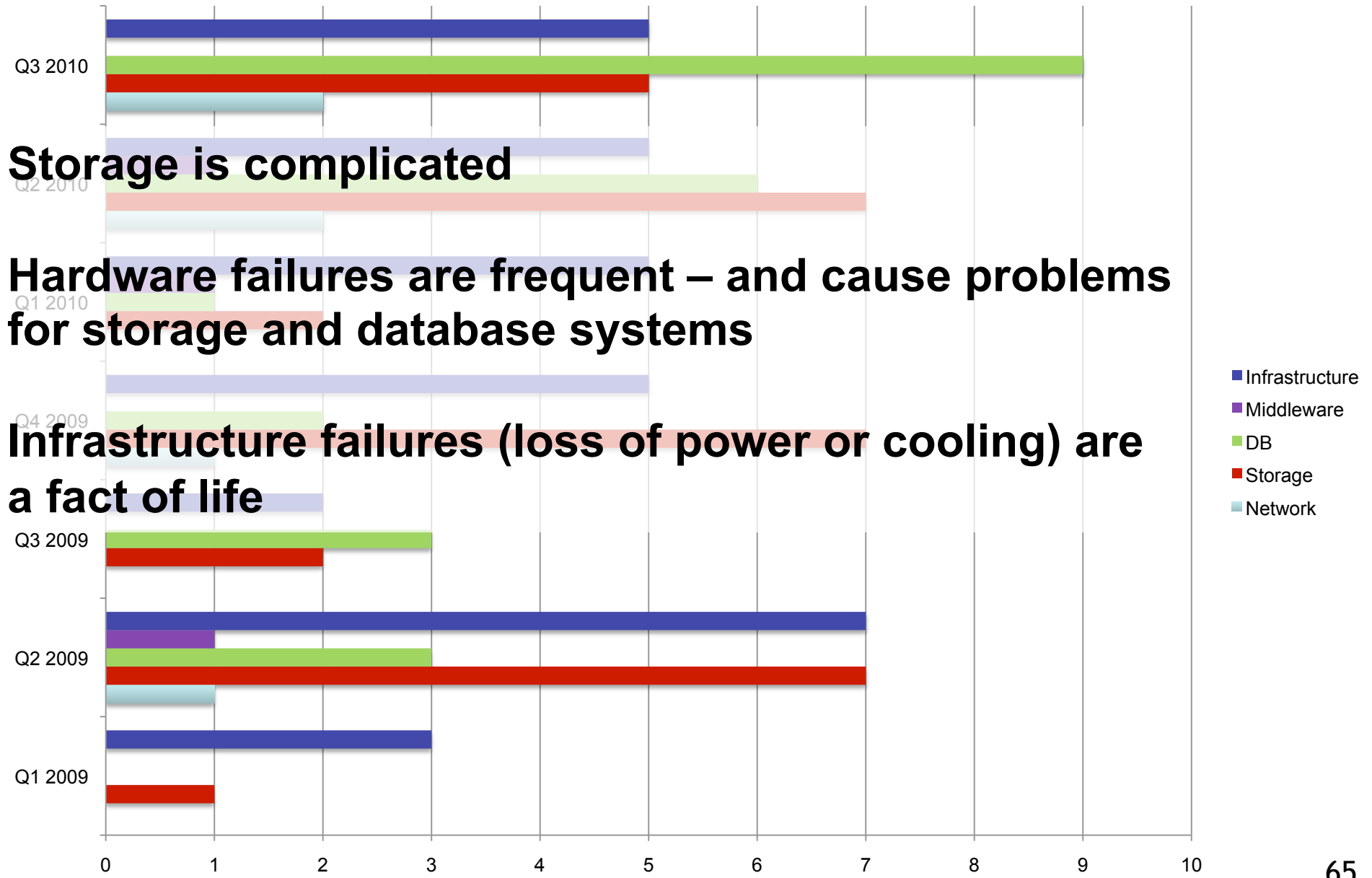
Storage is complicated

Hardware failures are frequent – and cause problems for storage and database systems

- Infrastructure
- Middleware
- DB
- Storage
- Network



Operational Issues - I



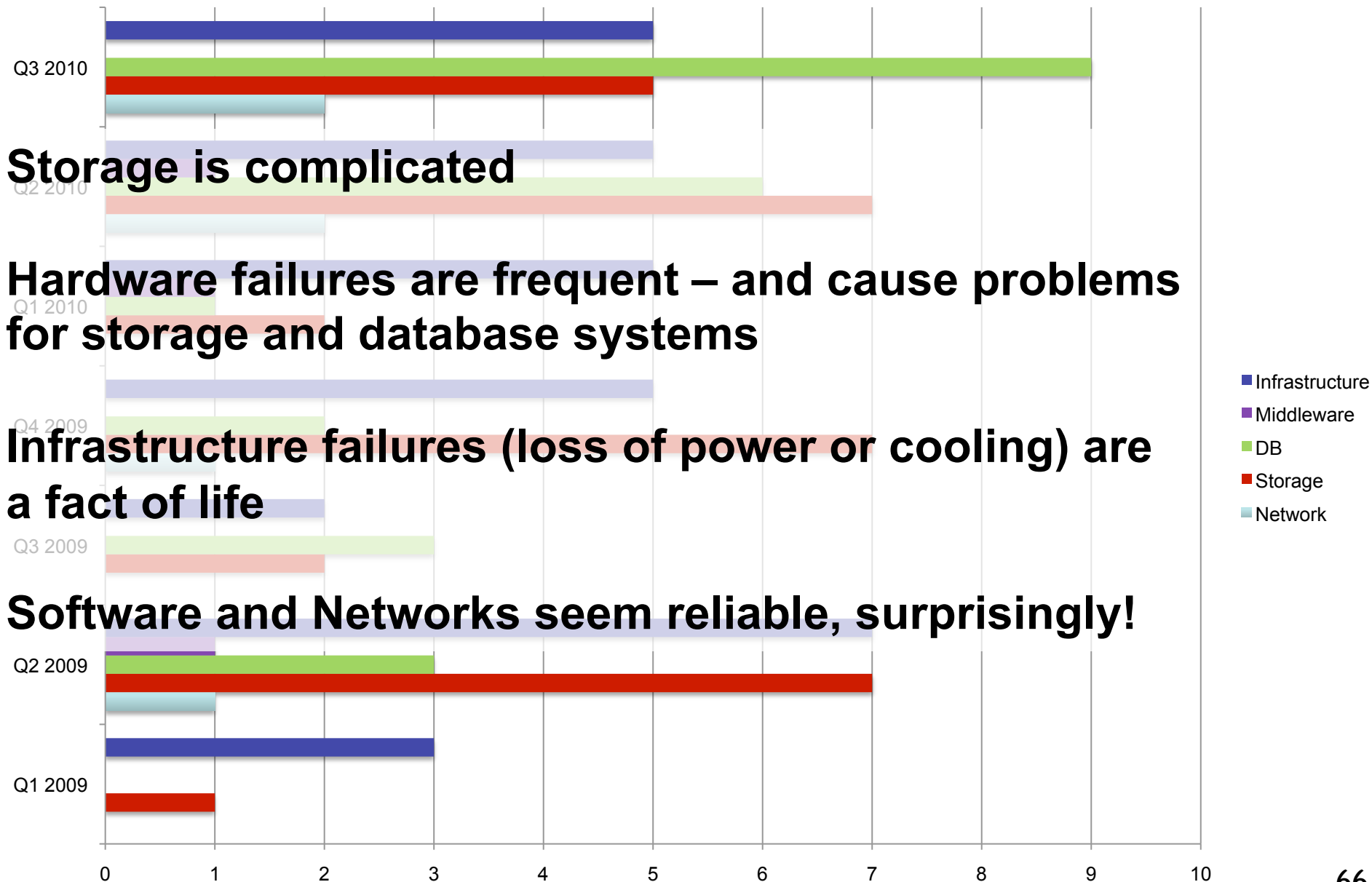
Storage is complicated

Hardware failures are frequent – and cause problems for storage and database systems

Infrastructure failures (loss of power or cooling) are a fact of life



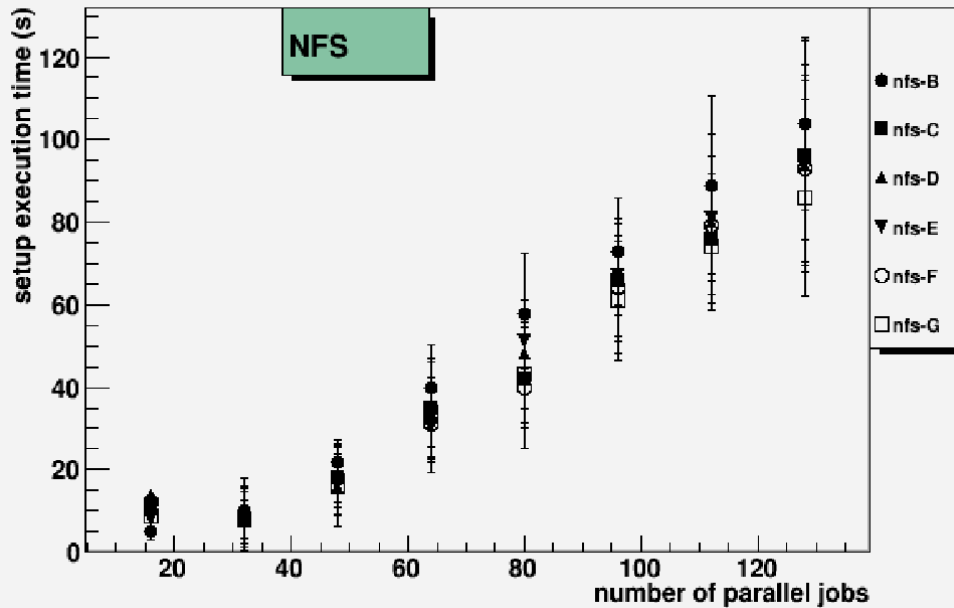
Operational Issues - I



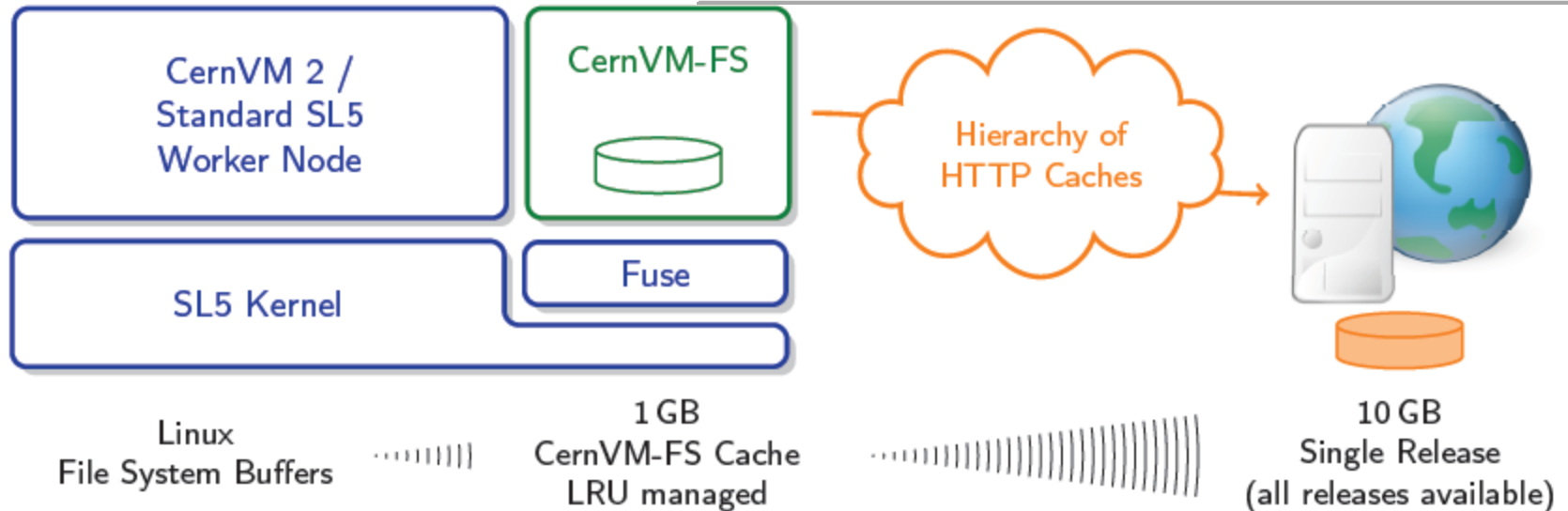
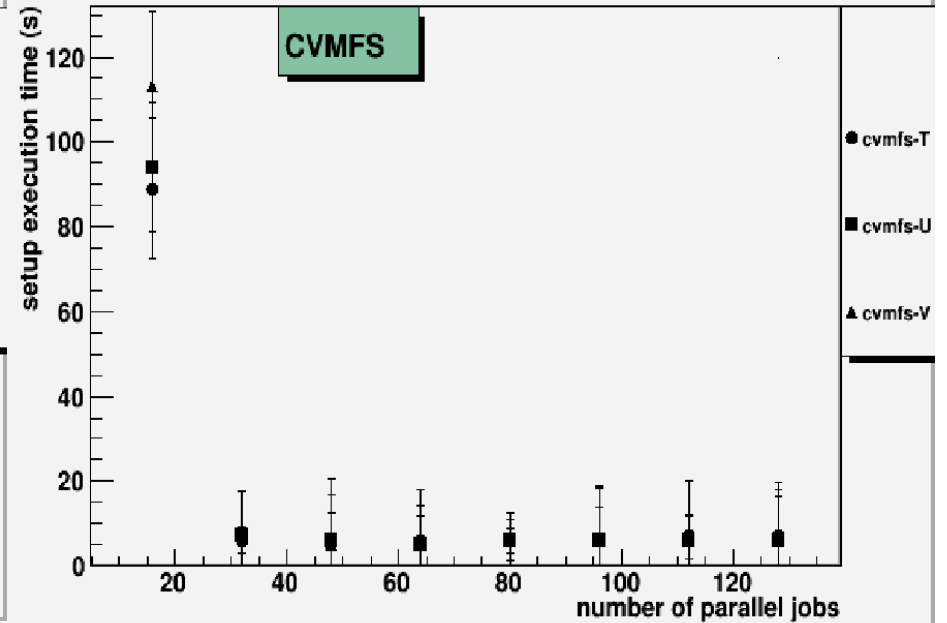


Shared filesystem bottlenecks

Setup time vs parallel jobs



Setup time vs parallel jobs



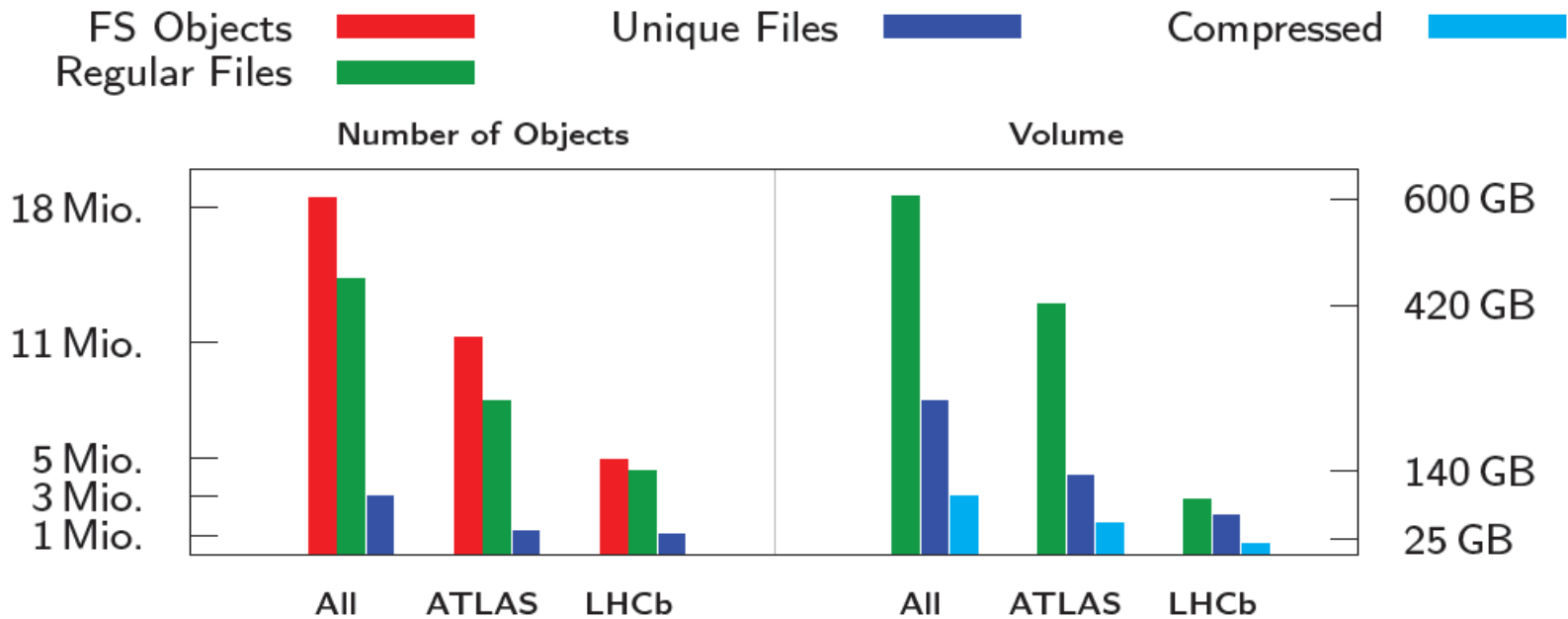


Shared filesystem bottlenecks

Repositories at CernVM:

ATLAS, CMS, LHCb, ALICE, LCD, NA61, H1, BOSS
HEPSOFT, Grid UI, LCG Externals

Ongoing: ATLAS Nightlies, ATLAS Conditions Database



Overall: 600 GB, 18.5 Mio. File System Objects
Repository Core: 97 GB (16%), 3.3 Mio. File System Objects (18%)
(+ 40 GB Archive Data)



The 2007 challenges

- Collaboration

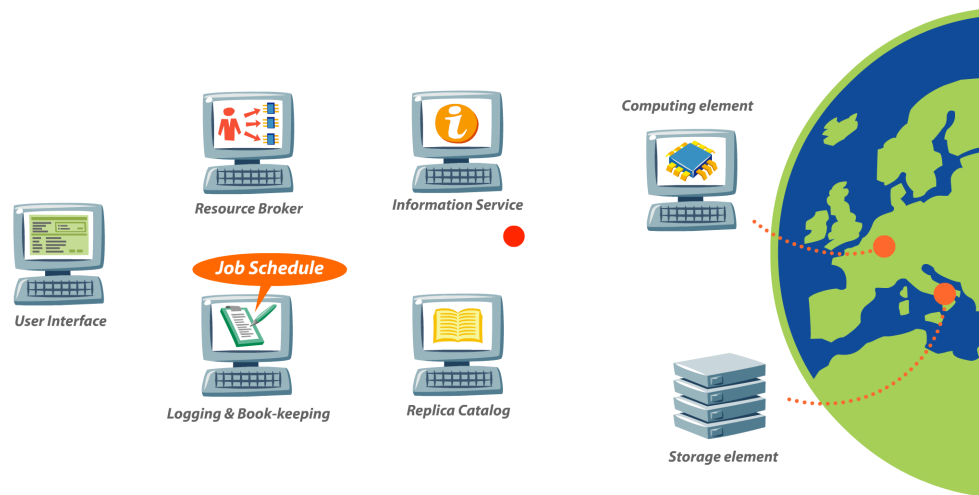
- From computer centre empires to a federation
- Consensus rather than control

This remains a challenge in 2010!

We reach consensus on most issues, but

- *Communication is a headache with so many sites and a changing population*
- *Site policies can be problematic in certain cases (e.g. installation of setuid software) especially for sites that are not 100% HEP.*
- *We reinvent the wheel: quattor & Lemon not widely adopted by Tier1s and Tier2s*
 - *although adopted after evaluation of various systems by a major financial institution with 10s of thousands of boxes.*

Pilot Jobs

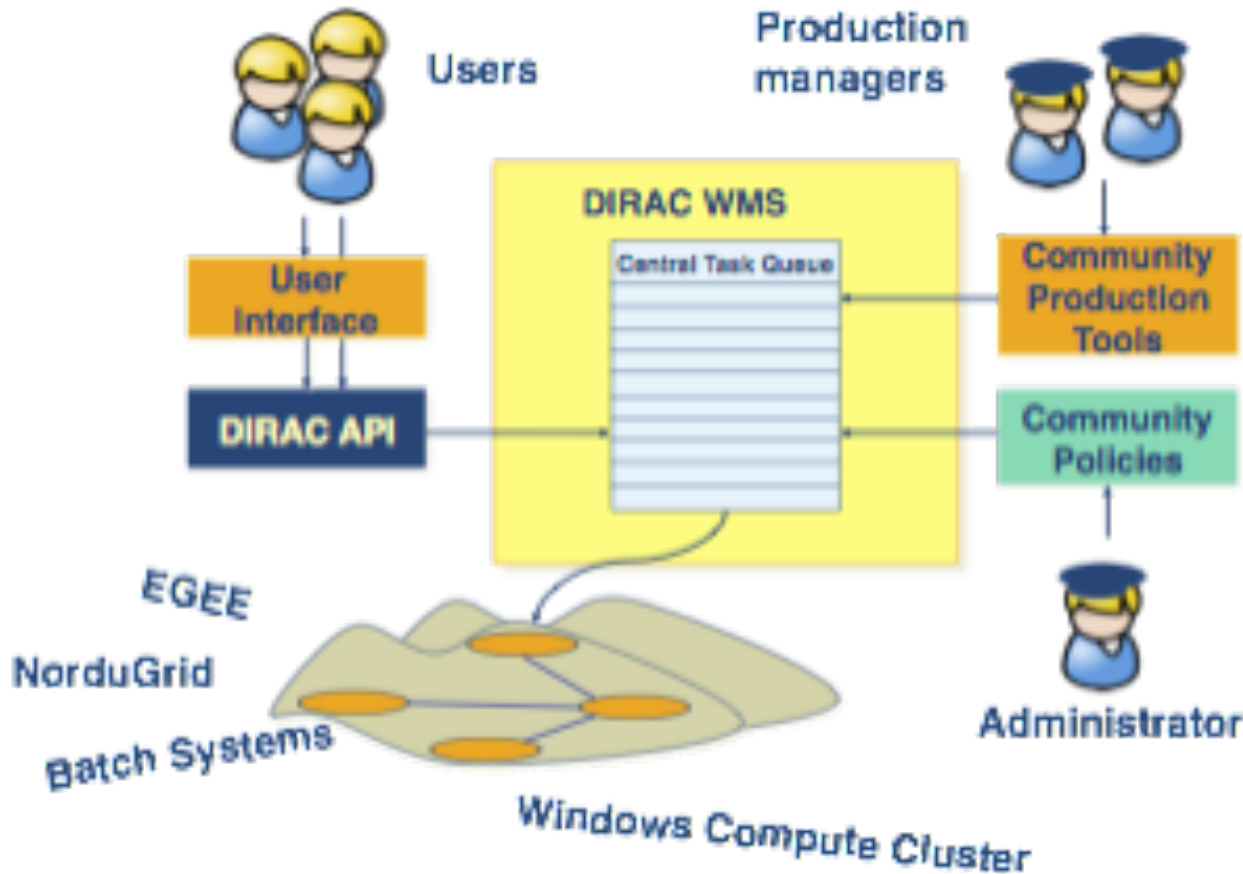


Grid sites generally want to maintain a high average CPU utilisation. Easiest to do this if there is a local queue of work to select from when another job ends.

Users are generally interested in turnround times as well as job throughput.

Turnround is reduced if jobs are held centrally until a processing slot is known to be free at a target site.

Pilot Jobs



Pilot job systems ensure “joblets” are sent to a host that will provide immediate execution.

Pilot job will check for correct s/w environment before loading “joblets”.

They also guarantee experiment control over job execution order. Low priority work can (will...) be pre-empted!

More of the “grid intelligence” is in per-VO software than was imagined at the start of the Grid adventure.



Data Issues



Scheduled work only!

700MB/s

420MB/s



700MB/s

~~(1600MB/s)~~
~3600MB/s



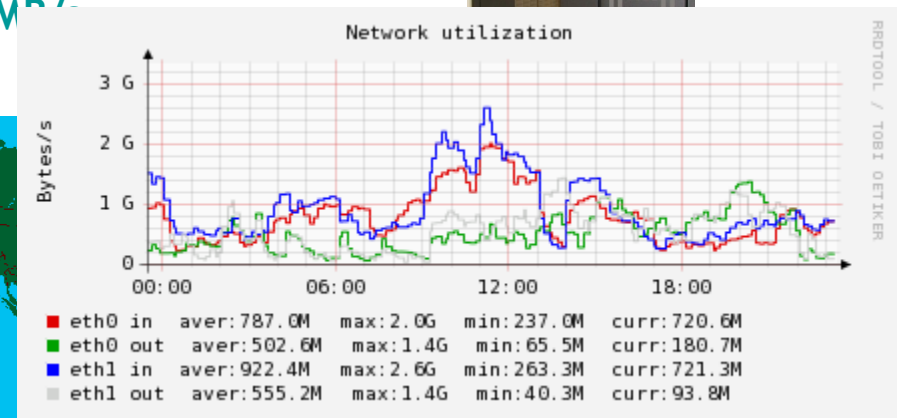
2600MB/s
~~1120MB/s~~

~~(2000MB/s)~~
?MB/s



Averages! Need to be able to support 2x for recovery!

1430MB/s





Data Reality

- Mass Storage systems have worked well for recording, export and retrieval of “production” data.
- But some features of the CASTOR system developed at CERN are unused or ill-adapted
 - experiments want to manage data availability
 - file sizes, file-placement policies and access patterns interact badly
 - alleviated by experiment management of data transfer between tape and disk...
 - analysis use favours low latency over guaranteed data rates
 - aggravated by experiment management of data; automated replication of busy datasets is disabled.



Outline

- Introduction to CERN, LHC and Experiments
- The LHC Computing Challenge
- Preparation
- Reality
- **Future Outlook**
 - Data Access
 - Virtualisation
- Summary/Conclusion



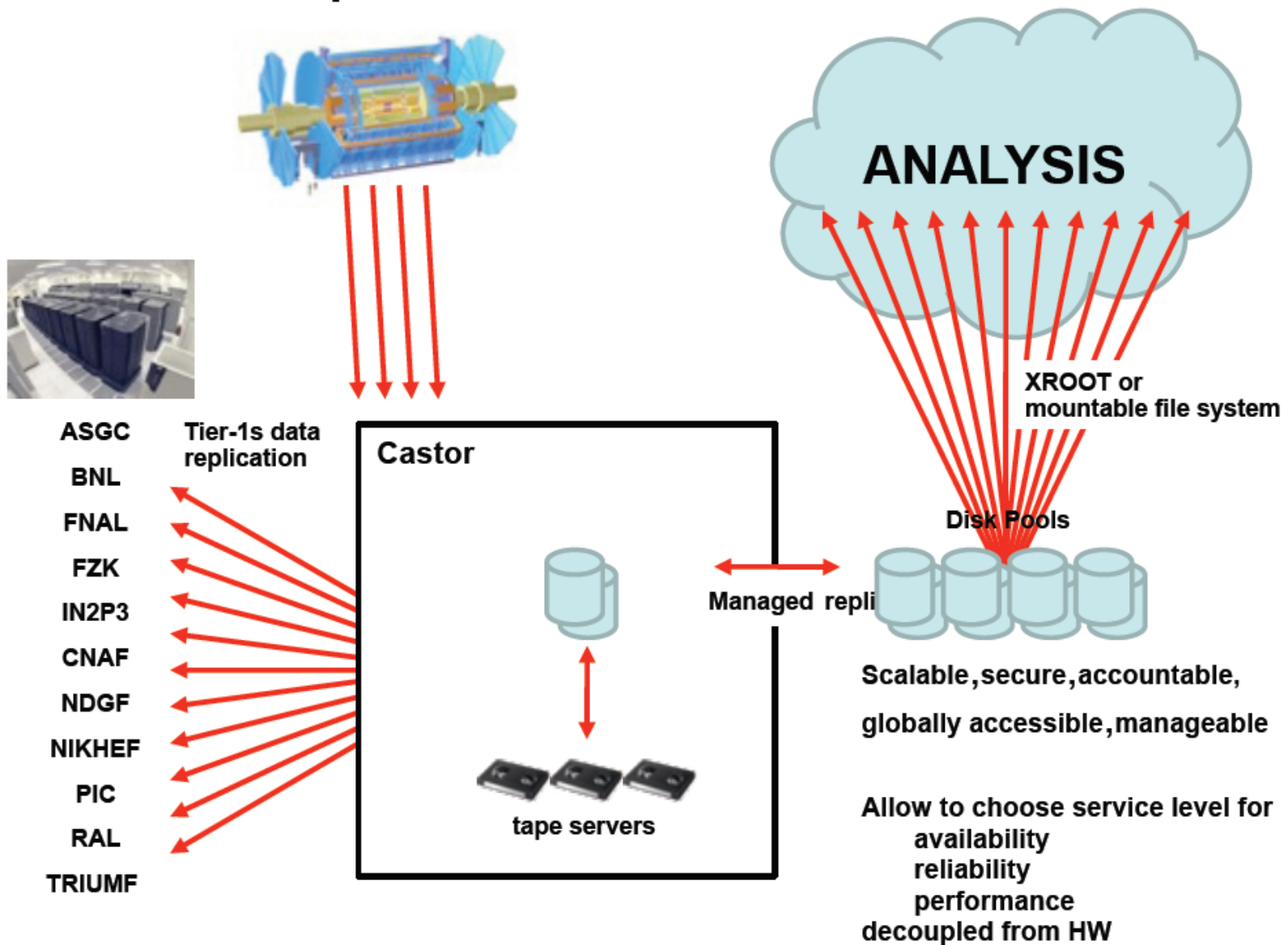
Outline

- Introduction to CERN, LHC and Experiments
- The LHC Computing Challenge
- Preparation
- Reality
- **Future Outlook**
 - Data Access
 - Virtualisation
- Summary/Conclusion



Data Futures – I

LHC Experiments



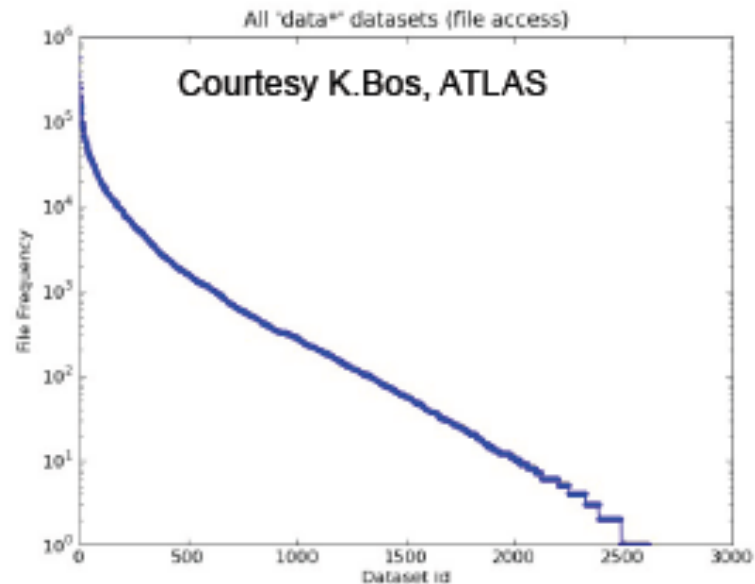


Data Futures – II

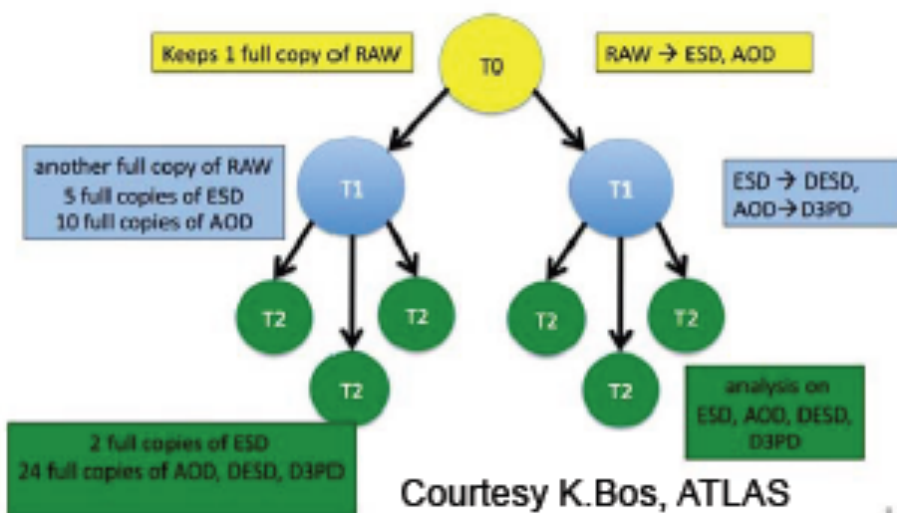
- Address hardware reliability issues in software
 - ... as is done elsewhere...
- Bring back model where storage system maintains multiple replicas of files, but drop disk mirroring
 - CERN switched from parity RAID a few years ago for I/O performance reasons.
- Growing interest in HADOOP at Tier2s.

Data Futures – III

- Only a small subset of data distributed is actually used
- Experiments don't know a priori which dataset will be popular
 - CMS has 8 orders magnitude in access between most and least popular



ATLAS Data placement model

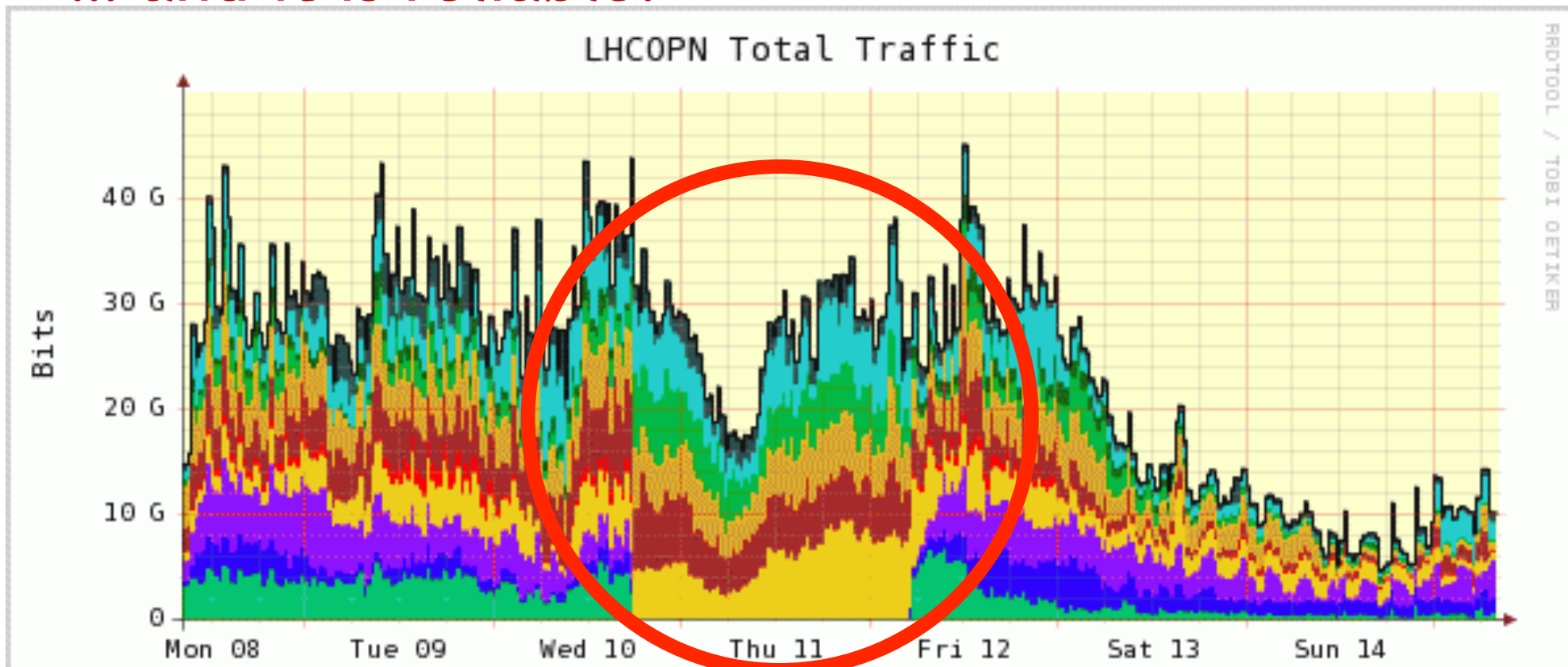


Courtesy K.Bos, ATLAS

**Dynamic data replication:
create copies of popular
datasets at multiple sites.**

Data Futures – IV

- Network capacity is readily available...
- ... and it is reliable:



Fibre cut during tests in 2009

Capacity reduced, but alternative links took over



Data Futures – IV

- Network capacity is readily available...
- ... and it is reliable.
- So why not simply copy data from another site
 - rather than recalling from tape?
 - if it not available locally?



Outline

- Introduction to CERN, LHC and Experiments
- The LHC Computing Challenge
- Preparation
- Reality
- **Future Outlook**
 - Data Access
 - **Virtualisation**
- Summary/Conclusion



Batch Virtualisation

Platform & Engineering Services

CERN IT Department

Integration of virtual machines in the batch system at CERN

S. Goasguen, M. Guijarro, B. Moreira, E. Roche, U. Schwickerath, Ricardo Silva, R. Wartel

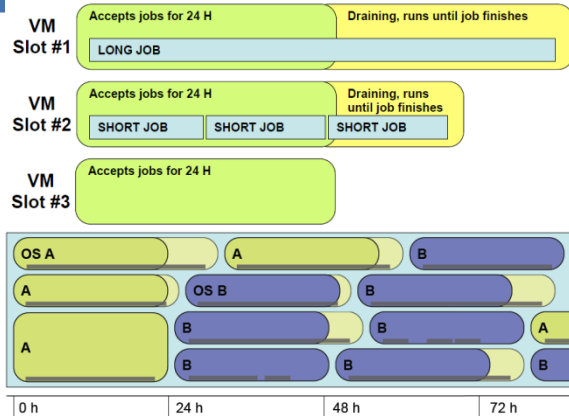
CHEP 2010
October 19, 2010

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it



Virtual Batch Worker Nodes

CERN IT Department



Integration of virtual machines in the batch system at CERN

5

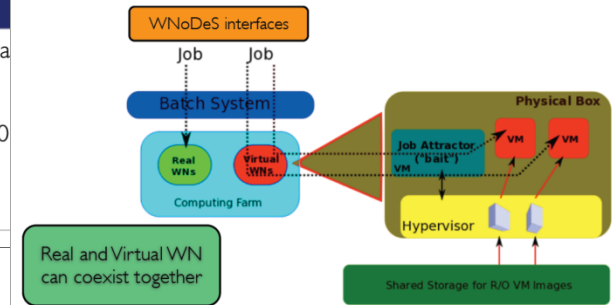
CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

WNoDeS, a tool for integrated Grid/Cloud and computing farm virtualization

Alessandro Italiano, Davide Sa
INFN-CNAF
CHEP 2010, Taipei
18 - 22 October, 2010

Monday, October 18, 2010

WNoDeS: how it works



Alessandro Italiano - INFN CNAF

10

CHEP 2010

Monday, October 18, 2010

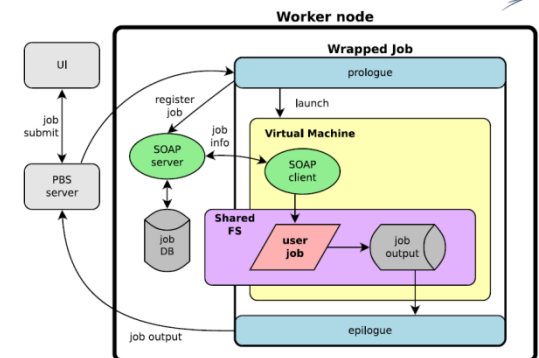
PIC
port d'informació
científica

Virtualization of PBS Jobs

Presenter: Pau Tallada Crespí

Authors: Marc Rodríguez, Pau Tallada,
Christian Neissner, Manuel Delfino

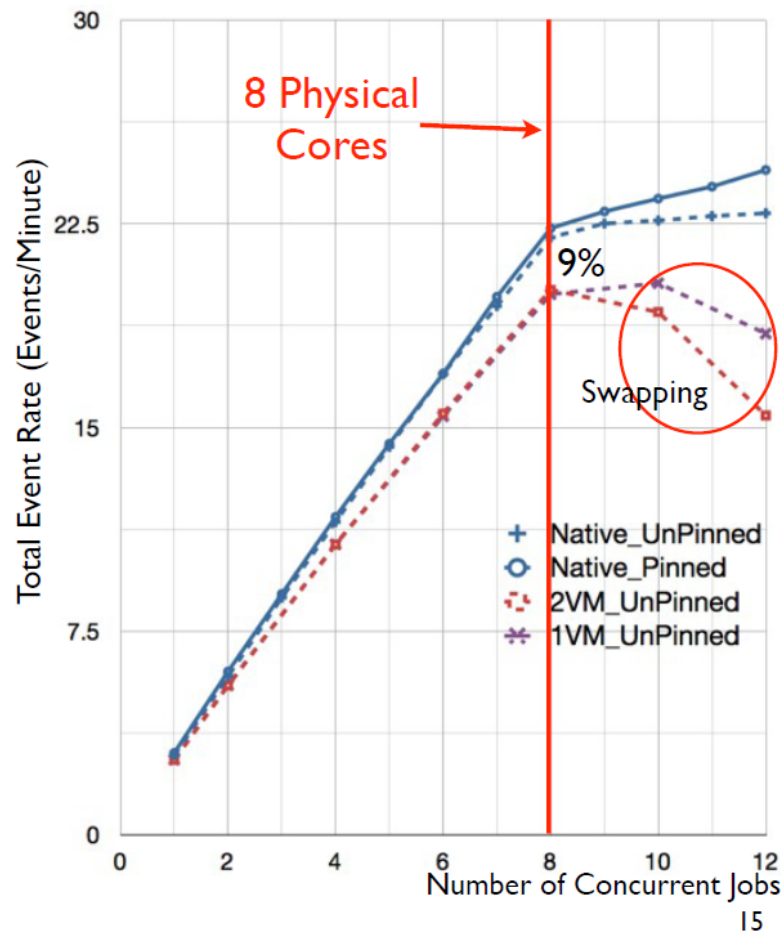
Working schema



Batch Virtualisation

- Virtualisation has a cost for users...

Reconstruction

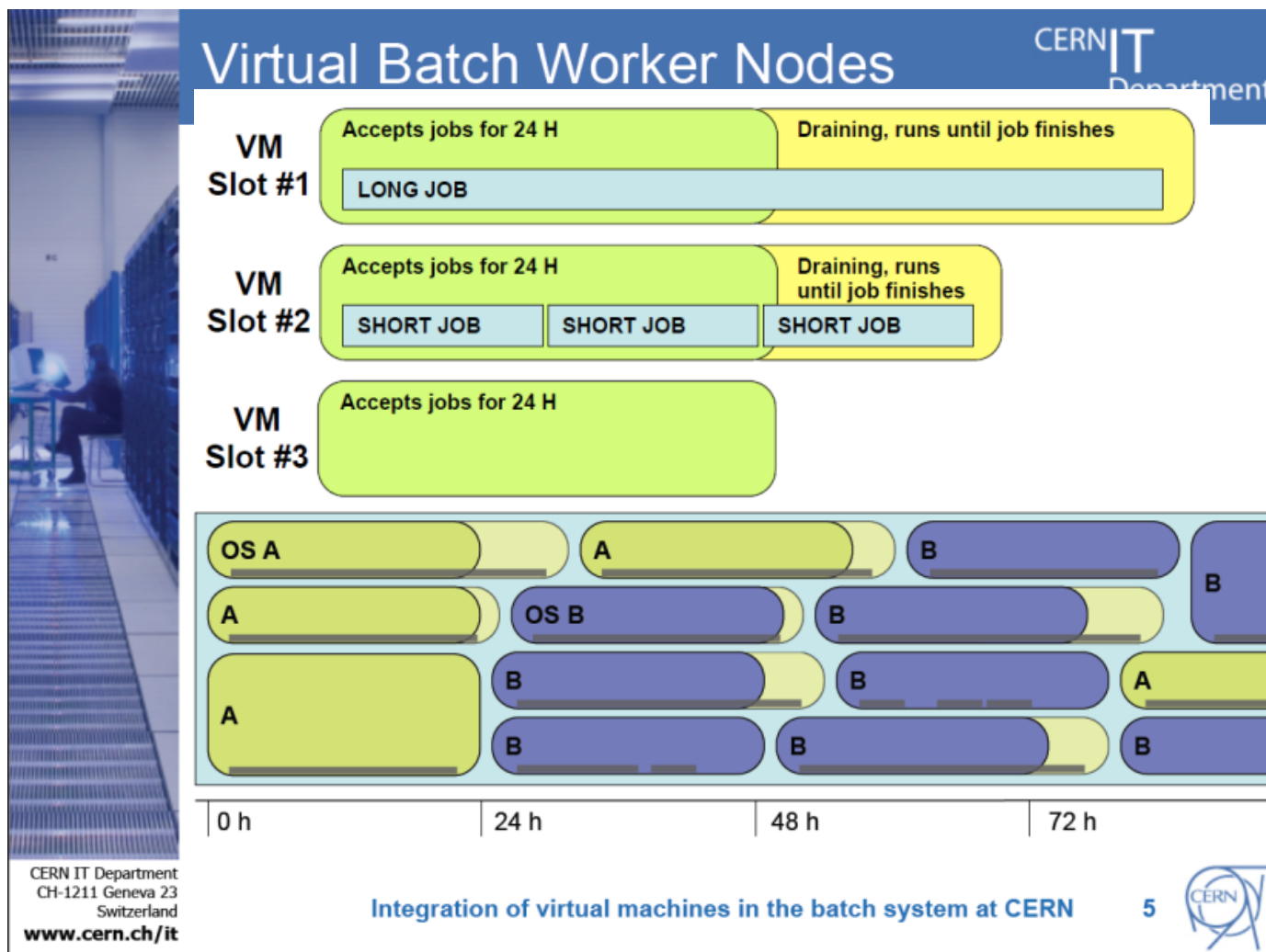


- Job uses large memory (~2GB, with 250MB shared)
- Native is ~9% better than VM.
 - Worse than Simulation, because Reconstruction jobs have more I/O activity and the memory footprint is much larger.
- VM has memory overhead, so 1VM case is swapping with high number of jobs.
- Pages can't be shared across VMs and 2VMs has more memory overhead. So 2VM case is swapping heavier.



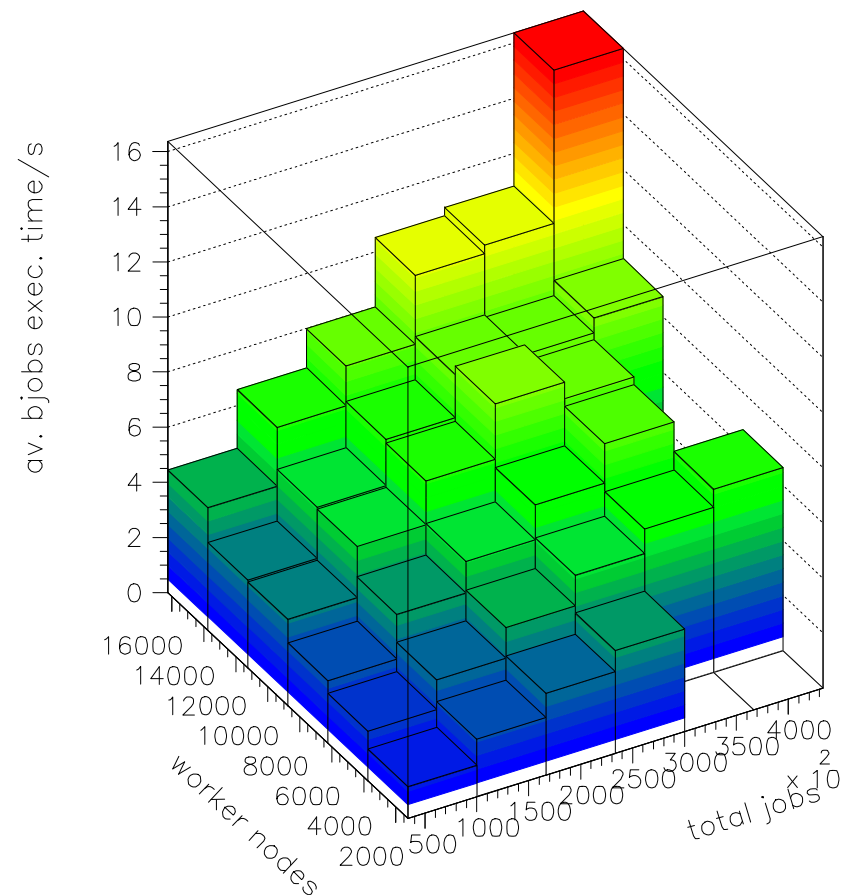
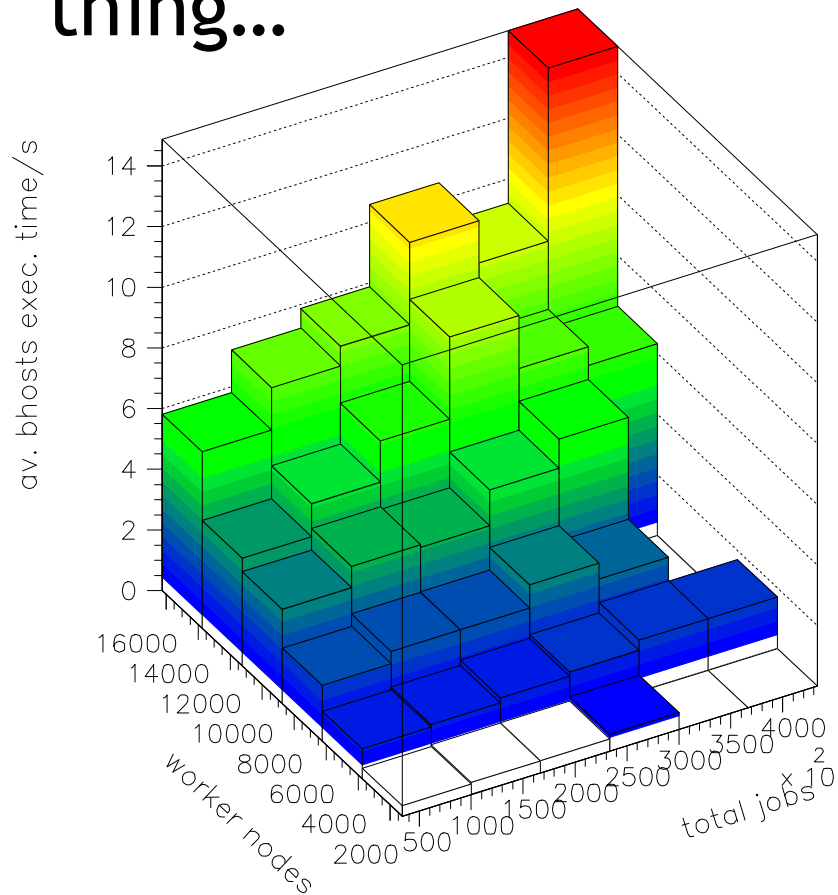
Batch Virtualisation

- Virtualisation has a cost for users...
- ... but efficiency advantages for sites.



Batch Virtualisation

- Virtualisation has a cost for users...
- ... but efficiency advantages for sites.
- Although multiplication of entities is never a good thing...





Batch Virtualisation

- Virtualisation has a cost for users...
- ... but efficiency advantages for sites.
- Although multiplication of entities is never a good thing...
- ... but maybe users will switch to requesting whole machines, not single processors.



Batch Virtualisation

- Virtualisation has a cost for users...
- ... but efficiency advantages for sites.
- Although multiplication of entities is never a good thing...
- ... but maybe users will switch to requesting whole machines, not single processors.
- Can we cut out local workload management systems and dynamically instantiate VM images that connect directly to pilot job frameworks?
 - A step to cloud computing?



Batch Virtualisation

- Virtualisation has a cost for users...
- ... but efficiency advantages for sites.
- Although multiplication of entities is never a good thing...
- ... but maybe users will switch to requesting whole machines, not single processors.
- Can we cut out local workload management systems and dynamically instantiate VM images that connect directly to pilot job frameworks?
 - A step to cloud computing?
- Sharing VM images between sites?
 - Automatic security updates for small sites?
 - Trust needed to make remote images acceptable!



Outline

- Introduction to CERN, LHC and Experiments
- The LHC Computing Challenge
- Preparation
- Reality
- Future Outlook
- **Summary/Conclusion**



Conclusions

- Preparation for LHC Computing has been
 - Long
 - Technically challenging
 - Sociologically challenging
- but
 - Successful,
 - Capable of improvements based on experience with real data
- and also
 - An exciting adventure
 - With much more detail than I have been able to give here...

