The new (remote) Tier 0

What is it, and how will it be used?

Ian Bird, CERN

WLCG Workshop
19th May 2012

Accelerating Science and Innovation

Agenda

CERN IT Department

- **Why a new computer centre?**
- **Initiatives for CC extensions**
- **Call for tender**
- **Comments on the selected site**
- **How will the remote CC be used?**

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

History

- **In ~2005 we foresaw a problem with the limited power available in the CERN Computer Centre**
  - Limited to 2.5 MW
  - No additional electrical power available on Meyrin site
  - Predictions for evolving needs anticipated ~10 MW by 2020
- **In addition we uncovered some other concerns**
  - Overload of existing UPS systems
  - Significant lack of critical power (with diesel backup)
  - No real facility for business continuity if there was a major catastrophic failure in the CC
- **In ~2006 we proposed to build a new CC at CERN on the Prevessin site**
  - The idea was to have a modular design that could expand with need

History - 2

CERN IT Department

- **Studies for a new CC on Prévessin site**
  - Four conceptual designs (2008/2009)
  - Lack of on site experience
  - Expensive!
  - Uncertainty in the requirements
- **Bought 100 kW of critical power in a hosting facility in Geneva**
  - Addressed some concerns of critical power and business continuity
  - In use since mid-2010
- **Planned to consolidate existing CC**
  - Minor upgrades brought power available from 2.5 to 2.9 MW
  - Upgrade UPS and critical power facilities – additional 600 kW of critical power, bringing total IT power available to 3.5 MW
- **Interest from Norway to provide a remote hosting facility**
  - Initial proposal not deemed suitable
  - Formal offer not convincing
  - Interest from other member states

New CC history - 1

CERN IT Department

- **Call for interest at FC June 2010**
  - How much computing capacity for 4MCHF/year?
  - Is such an approach technically feasible?
  - Is such an approach financially interesting?
  - Deadline end of November 2010
- **Response**
  - Surprising level of interest – 23+ proposals
  - Wide variation of solutions and capacity offered
  - Many offering > 2MW (one even > 5MW)
  - Assumptions and offers not always clearly understood
  - Wide variation in electricity tariffs (factor of 8!)
- **Many visits and discussions in 2010/2011**
- **Official decision to go ahead taken in spring 2011 and all potential bidders informed**
- **Several new consortia expressed interest**

New CC History – 2

CERN IT Department

- **Call for tender**
  - Sent out on 12th Sept
  - Specification with as few constraints as possible
  - Draft SLA included
  - A number of questions for clarification were received and answered (did people actually read the documents?)
  - Replies were due by 7th Nov

Tender Specification - I

CERN IT Department

- **Contract length 3+1+1+1+1**
- **Reliable hosting of CERN equipment in a separated area with controlled access**
  - Including all infrastructure support and maintenance
- **Provision of full configured racks including intelligent PDUs**
- **Essentially all services which cannot be done remotely**
  - Reception, unpacking and physical installation of servers
  - All network cabling according to CERN specification
  - Smart ‘hands and eyes’
  - Repair operations and stock management
  - Retirement operations

Evaluation Process

- **The financial offers were reviewed and in some cases corrected**
- **The technical compliance of a number of offers were reviewed (those which were within a similar price range)**
- **Meetings were held with 5 consortia to ensure that**
  - we understood correctly what was being offered
  - they had correctly understood what we were asking for
  - errors were discovered in their understanding
- **Site selected and approved at FC 14th March**

Wigner Data Centre, Budapest

CERN IT Department

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

Technical Overview

CERN IT Department

- **New facility due to be ready at the end of 2012**
- **1100m² (725m²) in an existing building but new infrastructure**
- **2 independent HV lines to site**
- **Full UPS and diesel coverage for all IT load (and cooling)**
  - 3 UPS systems per block (A+B for IT and one for cooling and infrastructure systems)
- **Maximum 2.7MW**
  - In-row cooling units with N+1 redundancy per row (N+2 per block)
  - N+1 chillers with free cooling technology (under 18°C*)
- **Well defined team structure for support**
- **Fire detection and suppression for IT and electrical rooms**
- **Multi-level access control; site, DC area, building, room**
- **Estimated PUE of 1.5**

Business Continuity

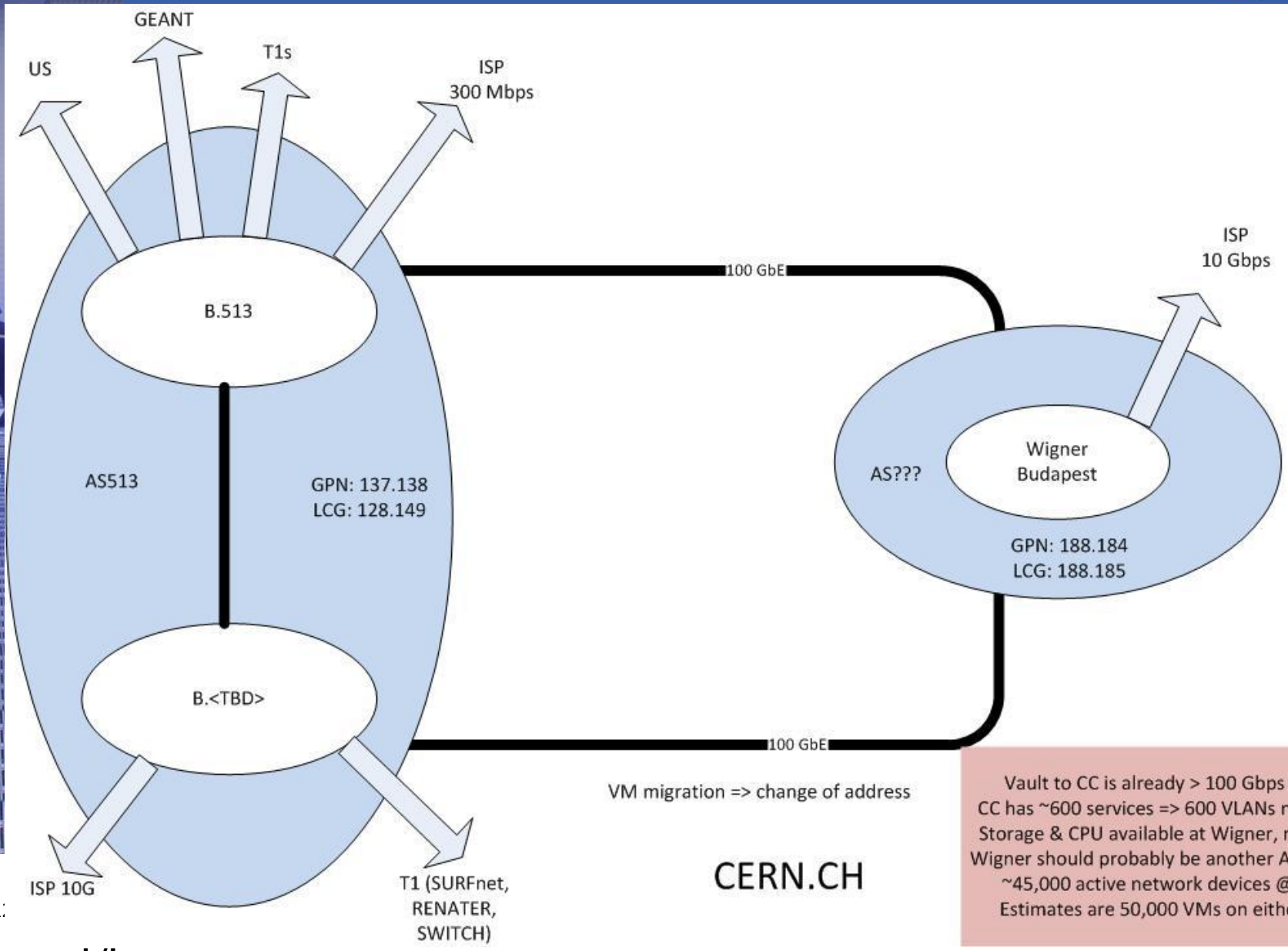
CERN IT Department

- **With a 2nd DC it makes sense to implement a comprehensive BC approach**
- **First classification of services against three BC options:**
  1. Backup
  2. Load balancing
  3. Re-installation
- **An internal study has been conducted to see what would be required to implement BC from a networking point of view**
- **Requires a network hub to be setup independently of the CC**
  - Requires an air-conditioned room of about 40-50 sqm for ~ 10 racks with 50-60kW of UPS power and good external fibre connectivity
  - Several locations are being studied
- **All IT server deliveries are now being managed as if at a remote location**

Use of Remote Hosting

- **Logical extension of physics data processing**
  - Batch and disk storage split across the two sites
  - The aim is to make the use as invisible as possible to the user
- **Business continuity**
  - Benefit from the remote hosting site to implement a more complete business continuity strategy for IT services

Possible Network Topology



Vault to CC is already > 100 Gbps today
CC has ~600 services => 600 VLANs necessary
Storage & CPU available at Wigner, not tapes
Wigner should probably be another AS number
~45,000 active network devices @CERN
Estimates are 50,000 VMs on either side

Major Challenges

CERN IT Department

- **Large scale remote hosting is new**
  - Doing more with same resources
- **Networking is complex**
- **Localization to reduce network traffic?**
- **For BC; full classification of services**
- **Latency issues?**
  - Expect 20-30ms latency to remote site
  - A priori no known issues but a number of services will need to be tested
- **Bandwidth limitations**
  - Possible use of QoS
- **Disaster recovery – backup strategy**

Status and Future Plans

CERN IT Department

- **Contract was signed on 8th May 2012**
- **During 2012**
  - Tender for network connectivity
  - Small test installation
  - Define and agree working procedures and reporting
  - Define and agree SLA
  - Integrate with CERN monitoring/ticketing system
  - Define what equipment we wish to install and how it should be operated
- **2013**
  - 1Q 2013: install initial capacity (100kW plus networking) and beginning larger scale testing
  - 4Q 2013: install further 500kW
- **Goal for 1Q 2014 to be in production as IaaS with first BC services**

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

Traffic today

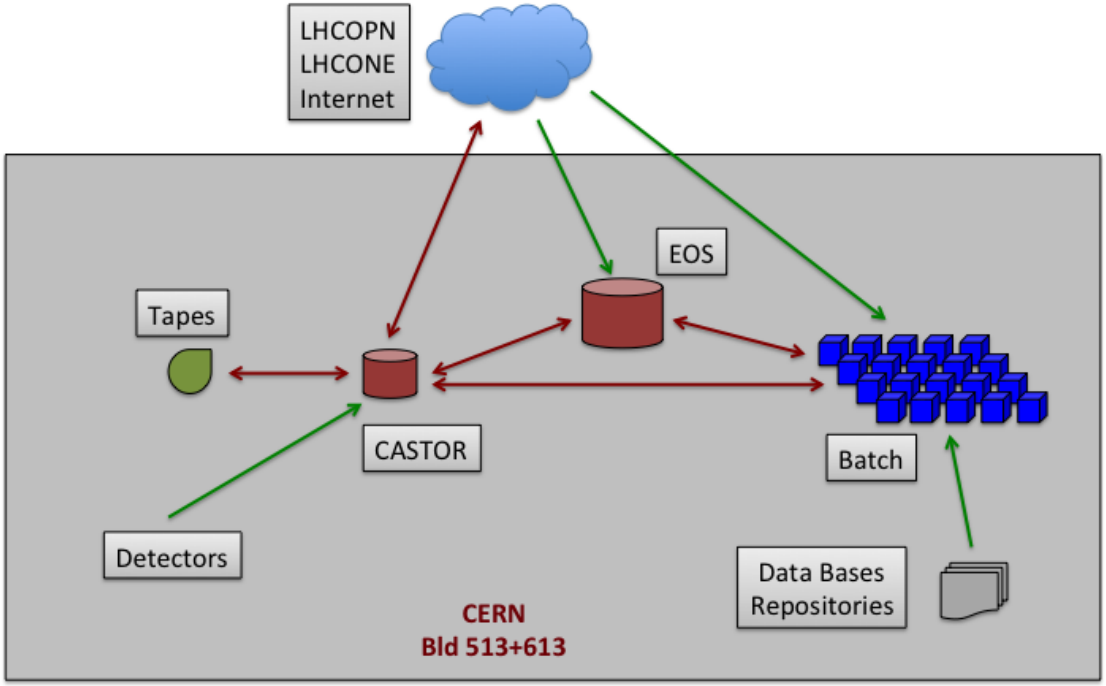
CERN IT Department

Castor IO contains:

1. Tape IO , including repack
2. LHCOPN IO
3. Batch IO
4. Part of EOS IO
5. Experiment IO (pit)

Batch IO contains:

1. Castor IO
2. EOS IO
3. Part of LHCONE IO
4. Part of Internet IO



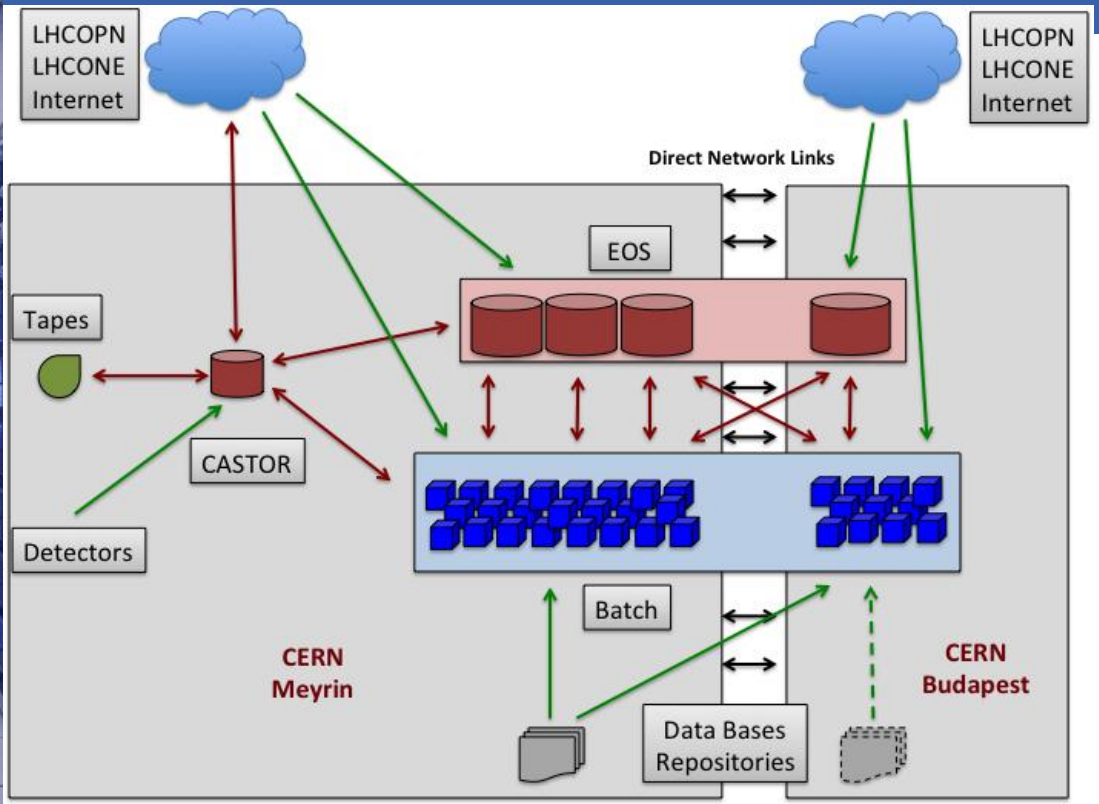
IO rates from the last 12 months

	Av. read [GBytes/s]	Av. Write [GBytes/s]	Peak read [GBytes/s]	Peak write [GBytes/s]		
CASTOR	6.3	11.0	7.0	14.2		
EOS	1.5	2.0	9.1	9.5		
Tape	0.9	0.5	3.6	2.3		
Batch	5.4	1.2	14.3	3.1		
LHCOPN	0.4	1.0	1.5	2.5		
LHCONE	0.2	0.4	0.7	0.9		
Internet	0.4	0.6	1.8	2.7		

Take max. Castor + 50% max. EOS as average reference figure
→ Average 12 GB/s Peak 19 GB/s

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

Possible usage model



Split the batch and EOS resources between the two centers, but keep them as complete logical entities

	2013	2014	2015	2016	2017
Average (peak) I/O [GB/s] CERN + ext. site	12 (19)	16 (25)	20 (32)	26 (42)	34 (54)
power ext. site [KW] ext.site size(CPU+disk)	600 17%	900 23%	1200 29%	1500 33%	1800 38%
Average (peak) I/O [GB/s] ext. site	2 (3)	4 (6)	6 (9)	9 (14)	13 (21)
# required 10 Gbit links for average (peak) I/O rates	2 (3)	4 (5)	5 (8)	8 (12)	11 (18)

Network available in 2013 → 24 GB/s

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

Possible strategies

- **If necessary to reduce network traffic**
  - Possibly introduce “locality” for physics data access
  - E.g. adapt EOS replication/allocation depending on IP address
  - Geographical mapping of batch queues and data seys
  - More sophisticated strategies later if needed

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

Latency & QoS?

- **Latency**
  - Inside CERN centre → 0.3 ms
  - CC to Safehost or experiment pits → 0,6 ms
  - CERN to Wigner (today) → 35 ms; eventual 20-30 ms
  - Possible side effects on services:
    - Response times, limit number of control statements, random I/O
  - Test now with delay box with lxbatch and EOS
- **Quality of Service**
  - Try some simple ways to implement a first order ‘QoS’ to avoid network problems in case of peak network loads; E.g. splitting TCP and UDP traffic
  - Needs to be tested

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

Summary

- **CERN CC will be expanded with a remote Centre in Budapest**
- **Intention is to implement a "transparent" usage model**
- **Network bandwidth not a problem**
  - Dependence on latency needs to be tested
  - More complex solutions involving location dependency to be tested but hopefully not necessary at least initially