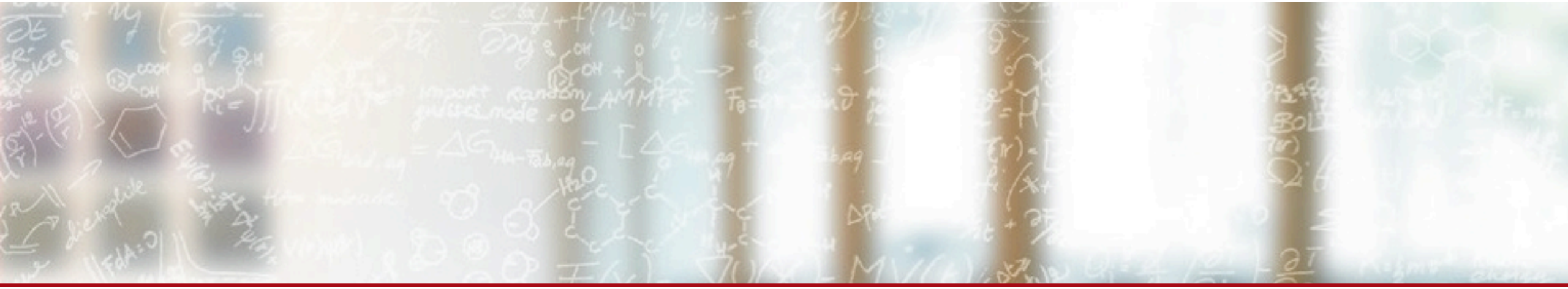
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich



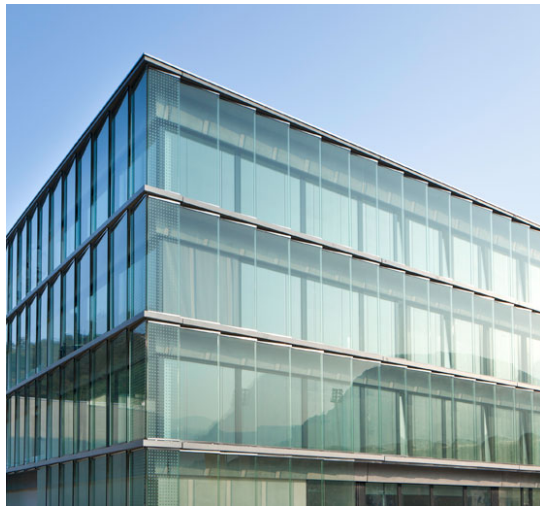
CHIPP - CSCS F2F meeting

CSCS, Lugano
August 21, 2015

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

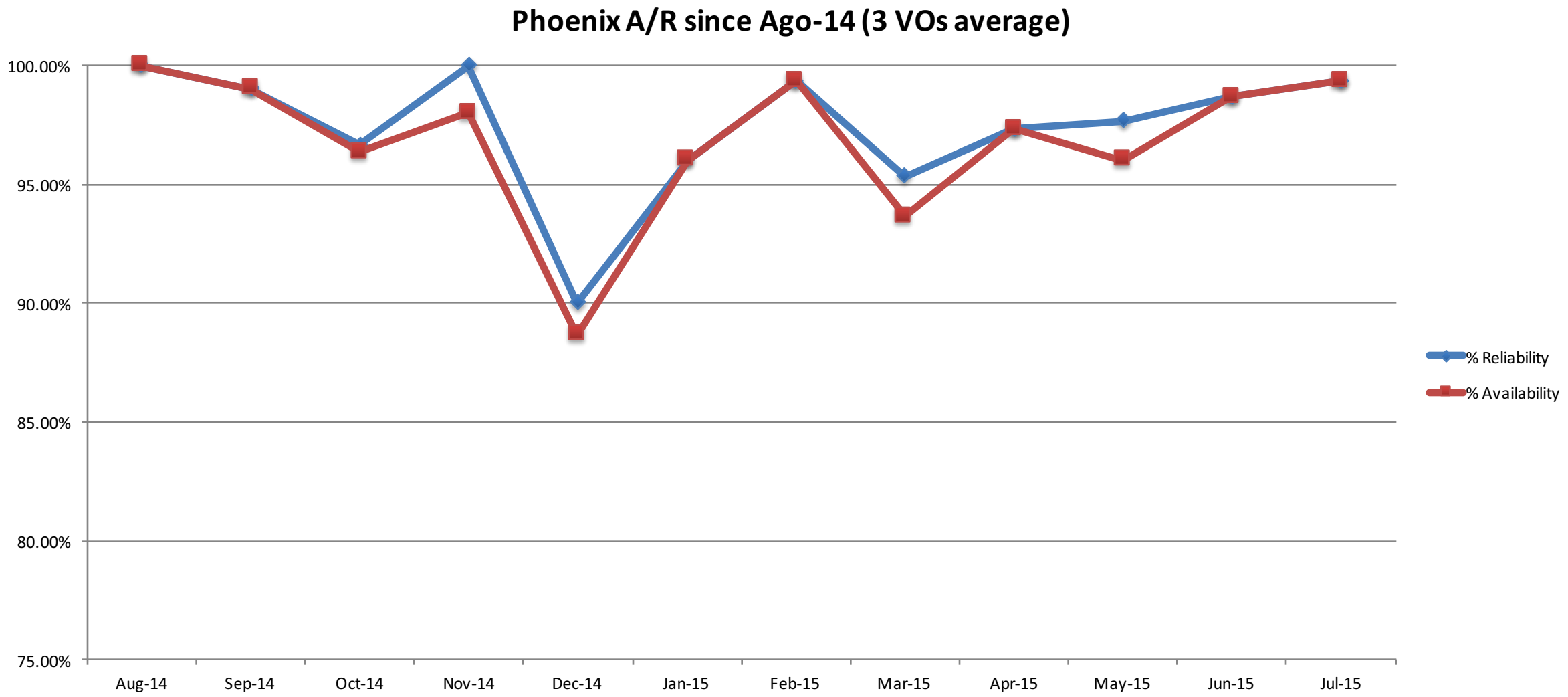
Tier 2 status and plans
CSCS



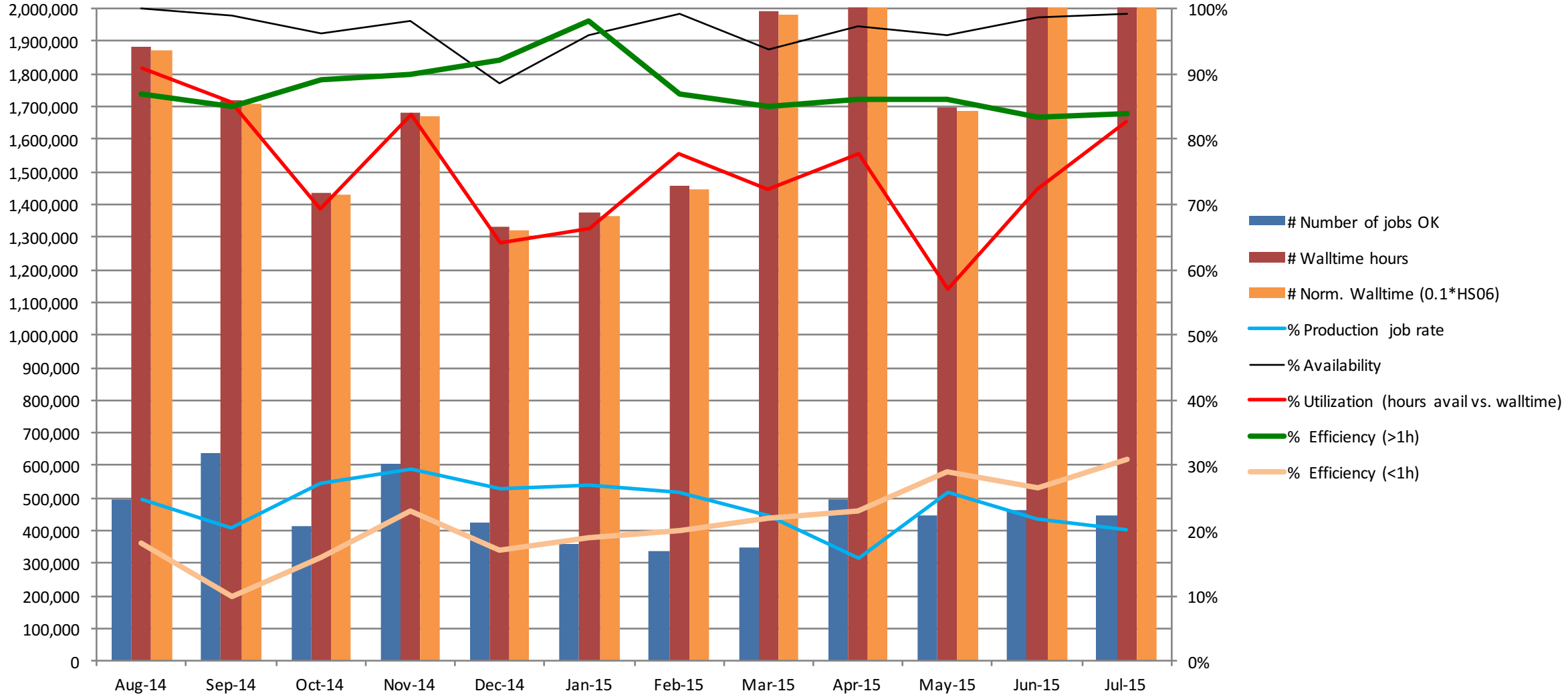
1. **Statistics**
   a. Availability/Reliability
   b. CPU usage
   c. Storage usage
2. **Operations**
   a. Systems
   b. Storage
3. **Plans**
   a. Phases of Phoenix
   b. Pledges
   c. Purchases
4. **Open discussion: network**
   a. Current Network Layout
   b. Network considerations

Statistics

Statistics – Availability & Reliability



Statistics – CPU usage

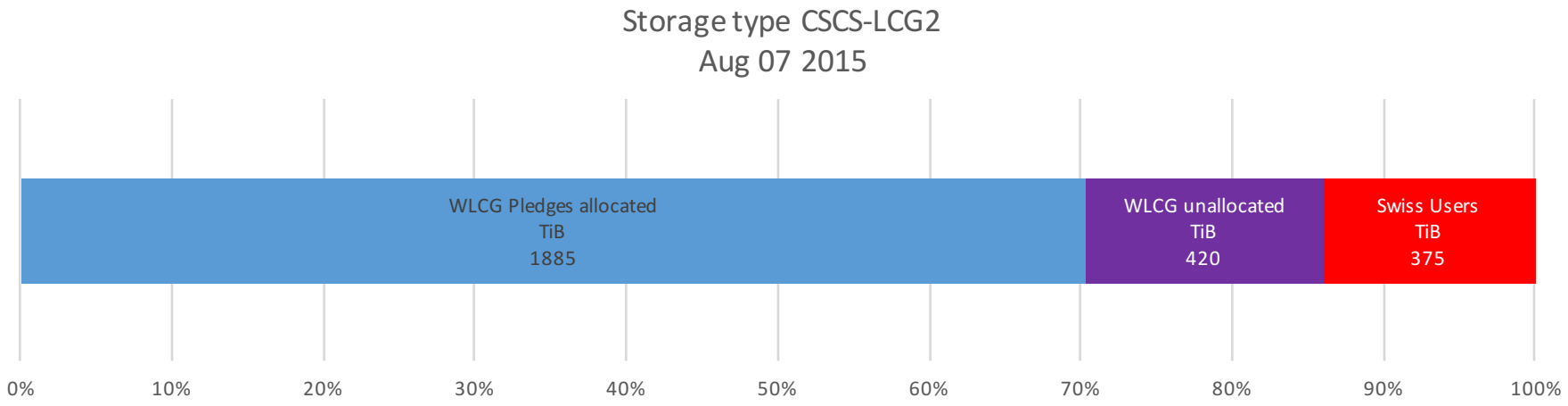


Statistics – Storage usage (1/2)

- Two ‘types’ of storage in the same dCache SE:
  - WLCG Pledges: 2305 TiB
  - Swiss Users: 375 TiB

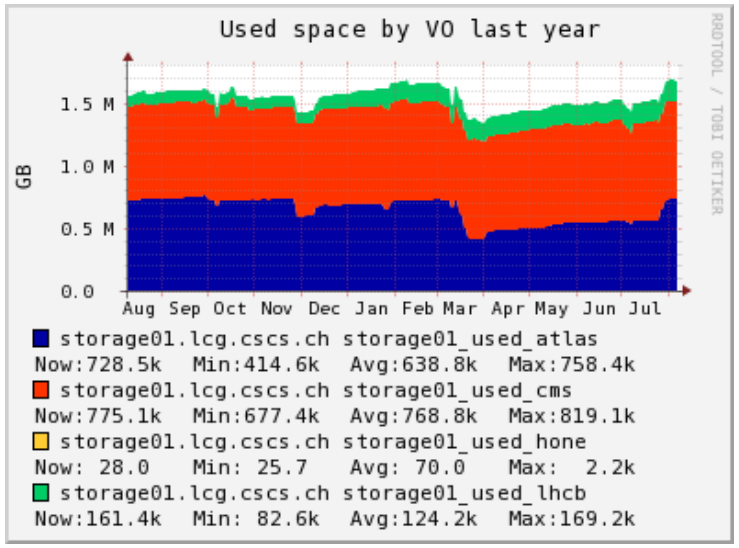
- Total capacity exceeding current Pledges (2305TiB vs. 2300 TiB), not including Swiss users storage. Real capacity well above pledges (~2600TiB)

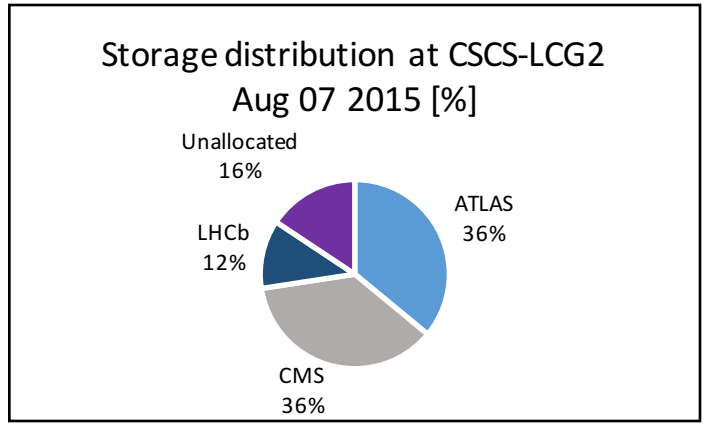
- Need to distribute 420TiB to meet current pledges (Phase J), now covered with Swiss users storage.

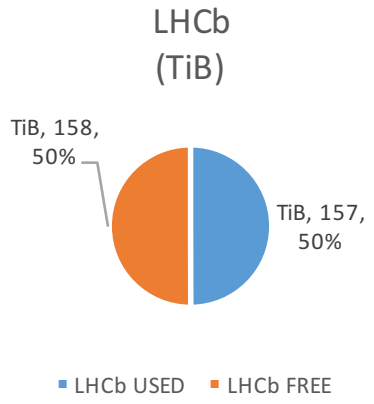


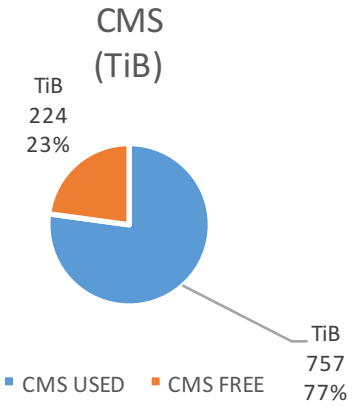
Statistics – Storage usage (2/2)

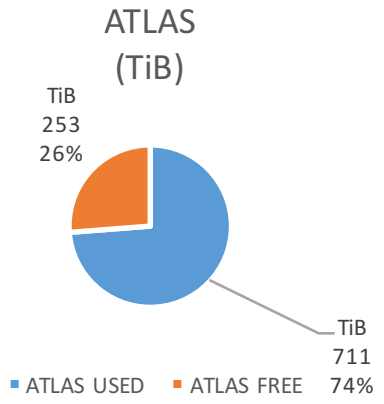
- Swiss storage distributed across the VOs as part of their pledged space.
- However, 420TiB need to be allocated as this pledged space is being covered with Swiss users storage (temporary measure).

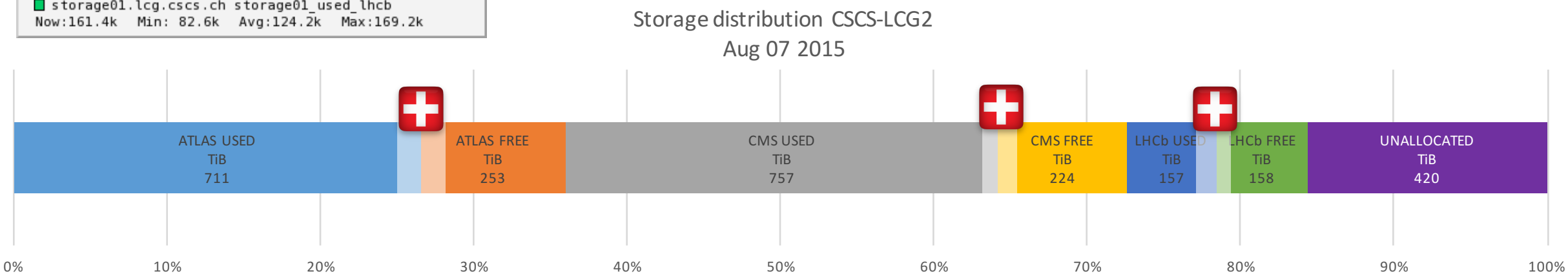












Statistics Summary

- Compute
  - Relatively unstable operation
  - Improved reliability & availability in Q3’15
- Storage
  - Need to allocate ~420TiB
  - Distribute to VOs according to needs with agreement of everyone

Operations

Operations - Systems

- Since Jan 2015
  - `arc01` re-installed and configured via Foreman+Puppet
  - `cms02` re-installed: working on finalizing Puppet configuration
  - A few production and all pre-production services moved from KVM to the VMware infrastructure
  - CFE to Puppet migration on-going (BDII ready and tested, to be deployed in production soon)

- Future
  - CMS VO boxes: `cms01` re-installed as soon as `cms02` has been completed
  - CREAM CEs:
    - Considering to reduce the number of deployed instances (from 4 to 2)
    - Investigating their possible decommission in the middle term should all VOs run successfully on ARC CEs only
  - ARC CEs reinstallation and future
    - `arc02` to be re-installed soon
    - Investigating how to let LHCb and CMS run on ARC
  - VMs migration to VMware shared infrastructure
    - Migration to be completed by the Q4 2015/ Q1 2016
    - KVM servers can be re-purposed at that time

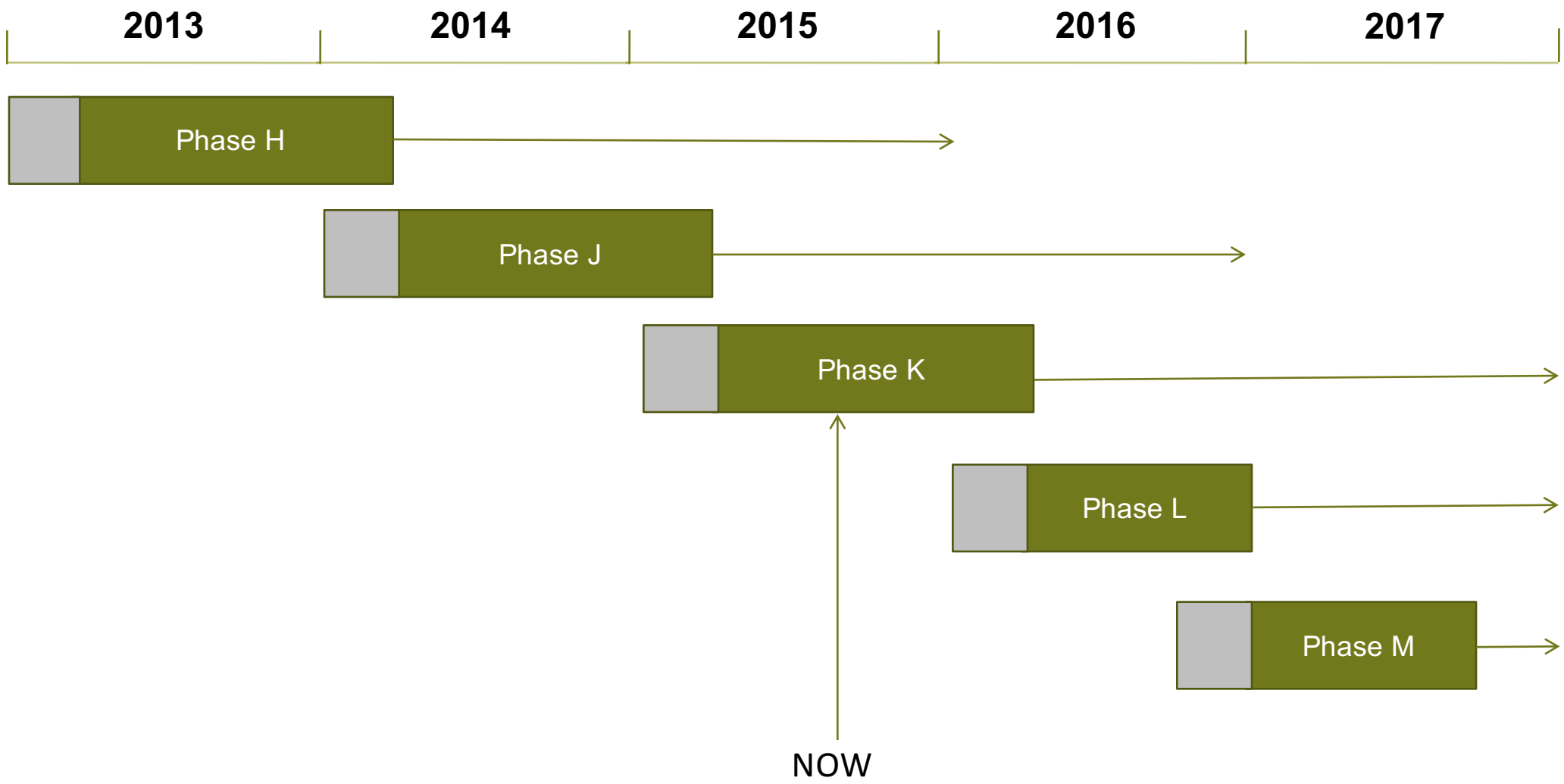
Operations - Storage

- Since Jan 2015
  - Updated dCache to latest 2.6 release
  - Kept busy with hardware issues (IBM, NetApp, etc.) and GPFS/dCache bugs!

- Future
  - dCache upgrade
    - In Q3/Q4 of 2015 CSCS will upgrade dCache from 2.6 to the latest supported golden release available of the 2.11 series.
    - Waited until most sites were upgraded due to high amount of bugs reported by the community.
  - GPFS upgrade
    - In Q4 2015 or Q1 2016 CSCS will upgrade GPFS 3.5 to the latest release available (likely GPFS 4).
    - This upgrade will depend on GPFS software release availability and operational issues.

Plans

Plans – Phases of Phoenix



Plans – Pledges

Phase	Compute power pledged [HS06]	Storage pledged [TB]
Phase H – April 2014	26000	1800
Phase J – April 2015	35000	2000
Phase K – April 2016	**39000**	**2600**
Phase L – April 2017	*44000*	*2900*

- Need to distribute properly 420TiB for **Phase J**
- Need to set pledges for **Phase K**
- Values for **Phase L** are a proposal!

Compute

	Phase H	Phase J	Phase K	Phase L
	HS	HS	HS	HS
ATLAS	10'400	14'000	14'500	15'600
CMS	10'400	14'000	14'500	15'600
LHCb	5'200	7'000	10'000	12'800
	26000	35000	39000	44000

Storage

	Phase H	Phase J	Phase K	Phase L
	TiB	TiB	TiB	TiB
ATLAS	792	875	955	1'040
CMS	792	875	955	1'040
LHCb	216	550	690	820
	1800	2300	2600	2900

Compute/Storage

	Phase H	Phase J	Phase K	Phase L
	HS/TB	HS/TB	HS/TB	HS/TB
ATLAS	13.13	16.00	15.18	15.00
CMS	13.13	16.00	15.18	15.00
LHCb	24.07	12.73	14.49	15.61

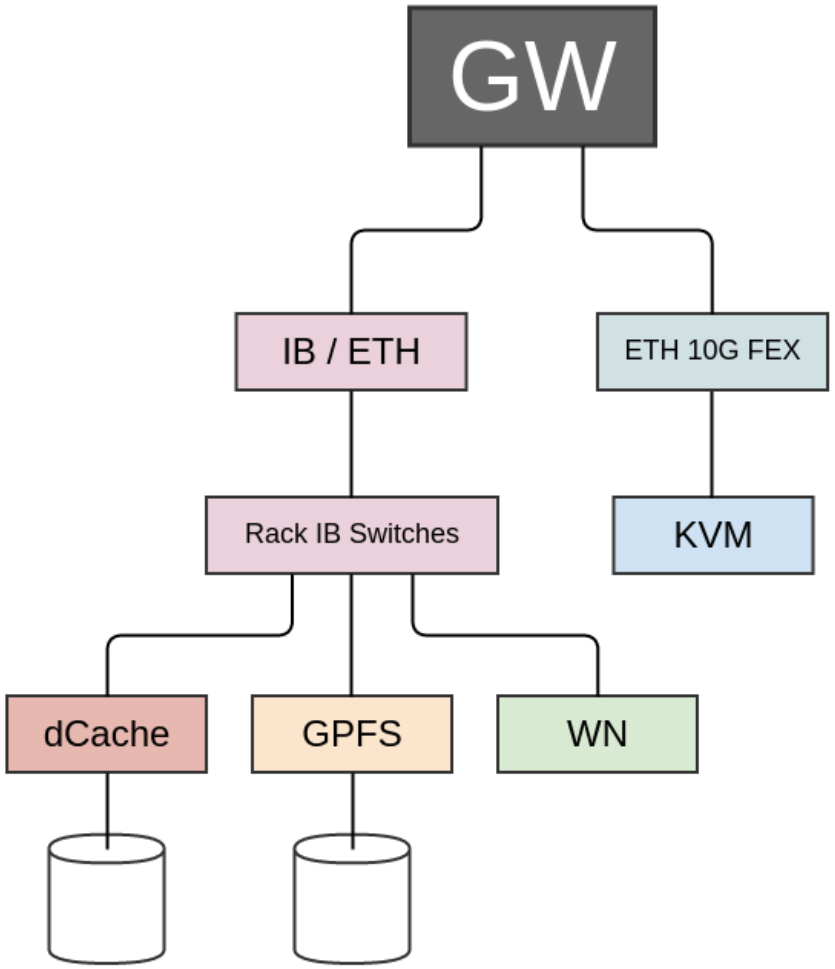
Plans – Purchases

- Storage
  - ~700TB storage, maybe 1PB if pricing allows to cover up for next year's massive decommission.
  - Integrate storage purchase in CSCS SAN.
    - Will increase availability/reliability of storage (at least 2 paths for every LUN!)
    - Potentially can allow CHIPP to benefit from CSCS massive storage purchases (big purchase = lower price per TB).

- Systems
  - 20+ WNs (likely 24 or 28 depending on final pricing). Total of 45k HS'06, well above pledges!
  - No new service nodes for now, VMware is great! ☺
  - Network infrastructure (more on this later)

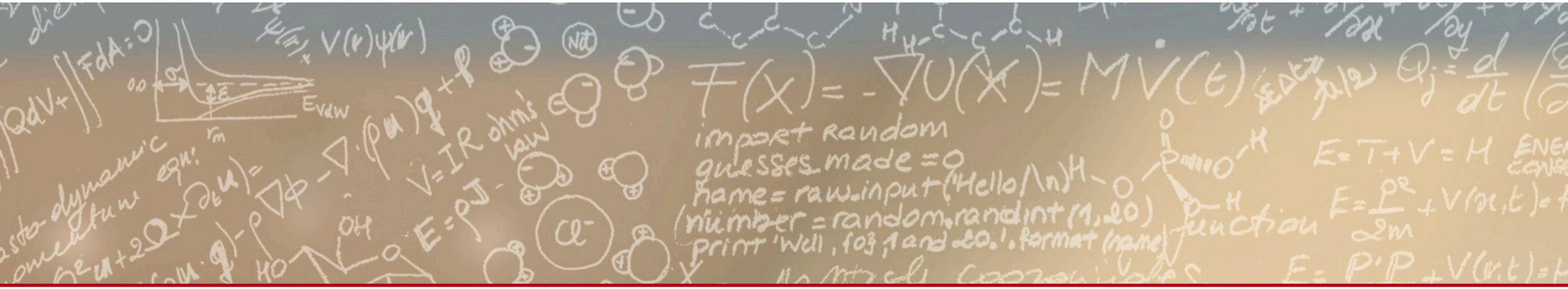
Open discussion: Network

Current Network Layout



Network Considerations

- Voltaire 4036E IB-Ethernet bridges reached the end of warranty
  - Model bought a few years ago as the only viable option
  - Requested an offer to extend the warranty for 1 or 2 years
  - Requested an offer as well for a replacement with same or equivalent model
  - In the last two years they caused several weird network issues quite hard to troubleshoot
  - Quite often we have to reboot them to restore smooth network operations
  - They cannot manage fragmentation correctly and we have been forced to configure a MTU=1500 on all our machines with all the related consequences on sustained throughput
- Evaluating alternatives
  - Current design limited to 2x10G bandwidth
  - 10G (40G?) quite affordable nowadays and already fully deployed at CSCS
  - Much easier to manage and troubleshoot compared to InfiniBand
  - InfiniBand kept for WNs access to GPFS scratch
  - Possibility to leverage the same InfiniBand card with minimal investment
  - Different network design options being discussed with our Network Team



Thank you for your attention.