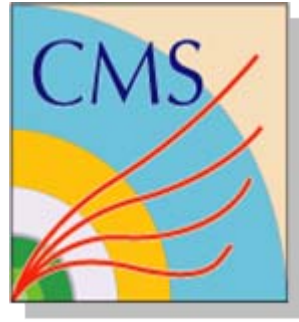


Report from the CHIPP Computing Board



Christoph Grab

Appenberg, Aug 25, 2009

Status of the Swiss Tier-2 Regional Centre

Swiss Tier-2: Facts and Figures (1)

- **The Swiss Tier-2** is operated by a collaboration of **CHIPP** and **CSCS**
(Swiss Centre of Scientific Computing of ETHZ), located in Manno (TI).
- **Properties :**
 - ➔ One centre for all three expts. **CMS, ATLAS + LHCb**; provides:
 - **Simulation** for experiment's community (*supply CH WLCG pledges*)
 - **End-user analysis** for Swiss community
 - **support** (operation and data supply) for Swiss **Tier-3 centres**
 - ➔ Standard **LINUX** compute cluster "**PHOENIX**"
- **Hardware setup increased incrementally** in phases (A,B,C) :
 - ➔ Total aim: storage : ~ 800 TB disks ; CPU : 2500 kSI2k (~1800 cores)
 - ➔ Technology choice so far: SUN blade centres + quad AMD Opterons.
 - ➔ Final size to be reached by ~ Q1/2010; now at ~ 55 % size
 - ➔ NO tapes

• **Manpower for Tier-2:**

- Operation at CSCS: sums to ~2.5 FTEs (IT experts, about 5 persons)
- support of experiment specifics by scientists of experiments; one contact person per experiment → in total ~2 FTE.
→ to be supported in 2009 through new CHIPP PostDocs

• **Financing (HW and service, no manpower) :**

- Financing of hardware mainly through SNF/FORCE (~90%), with some contributions by Universities + ETH + PSI;
- Operations and infrastructure provided by CSCS i.e. ETHZ

• **Support at CSCS: new director Prof.T.Schulthess**

- strategic direction for CSCS is HPCN for CH.
→ no further GRID research group;
- CSCS is committed to continue support for CHIPP and CH-Tier-2 (as agreed in MoU with CHIPP of March 2007; valid through 31.12.2012)

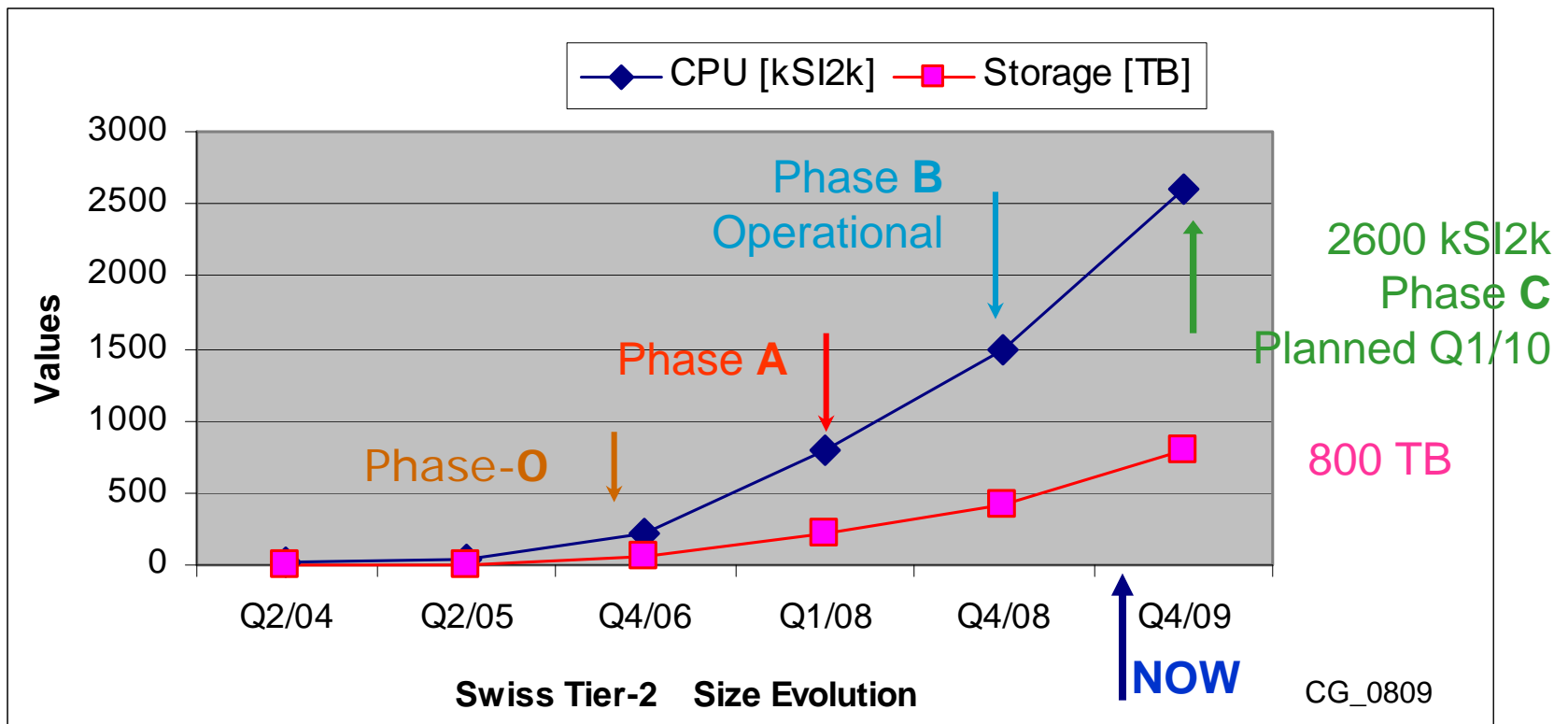
Financing : Hardware and service, no manpower till now:

- Total financial contributions (2004-2008) for incremental setup of Tier-2 cluster hardware only, manpower funded by Institutes:
 - by Universities + ETH + EPFL + PSI → ~200 kCHF
 - by Federal funding (FORCE/SNF) → ~3.9 MCHF including 2009, for last phase C
- Planned investments (FORCE/SNF):
 - >2010 onwards: rolling replacements ~ 600-700 kCHF/year;
 - and ONE FTE person to support Tier-2
- ➔ Total investment of ~ 4.1 MCHF;
annual recurring ~700 kCHF + one FTE

Swiss Tier-2 : Cluster Evolution

Growth corresponds to Swiss commitment in terms of compute resources supplied to the expt's according to the signed MoU with WLCG.

- In operation now: 960 cores → 1600 kSI2k; total 510 TB storage
- Last phase planned for Q4/09-Q1/10: → 2500 kSI2k; ~ 800 TB storage





System Phase B operational (since Q4/2008)



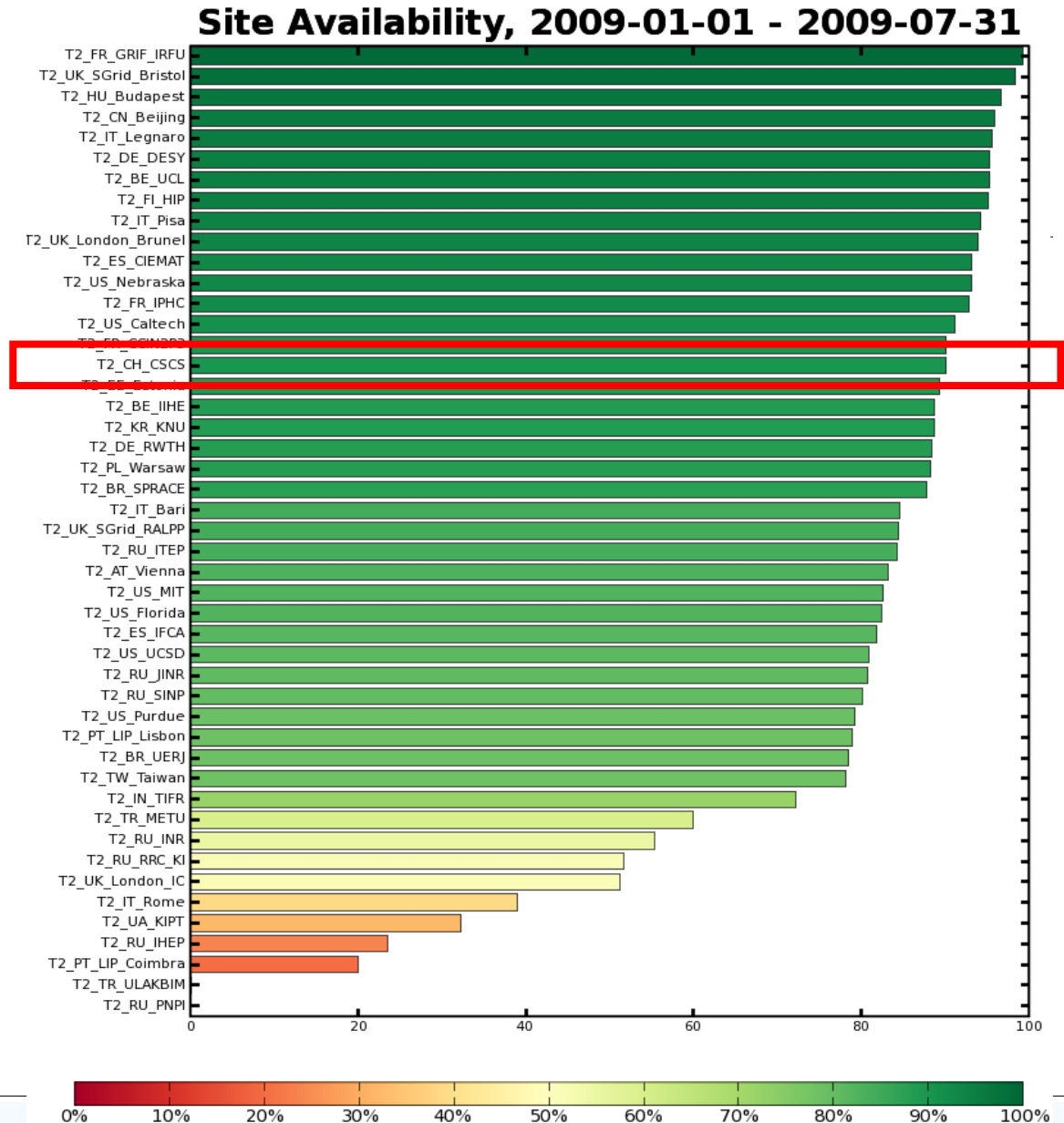
- ❖ **CPUs:** total of 1600 SI2K
SUN SB8000P blade centres;
AMD Opterons 2.6 GHz CPU (quad)

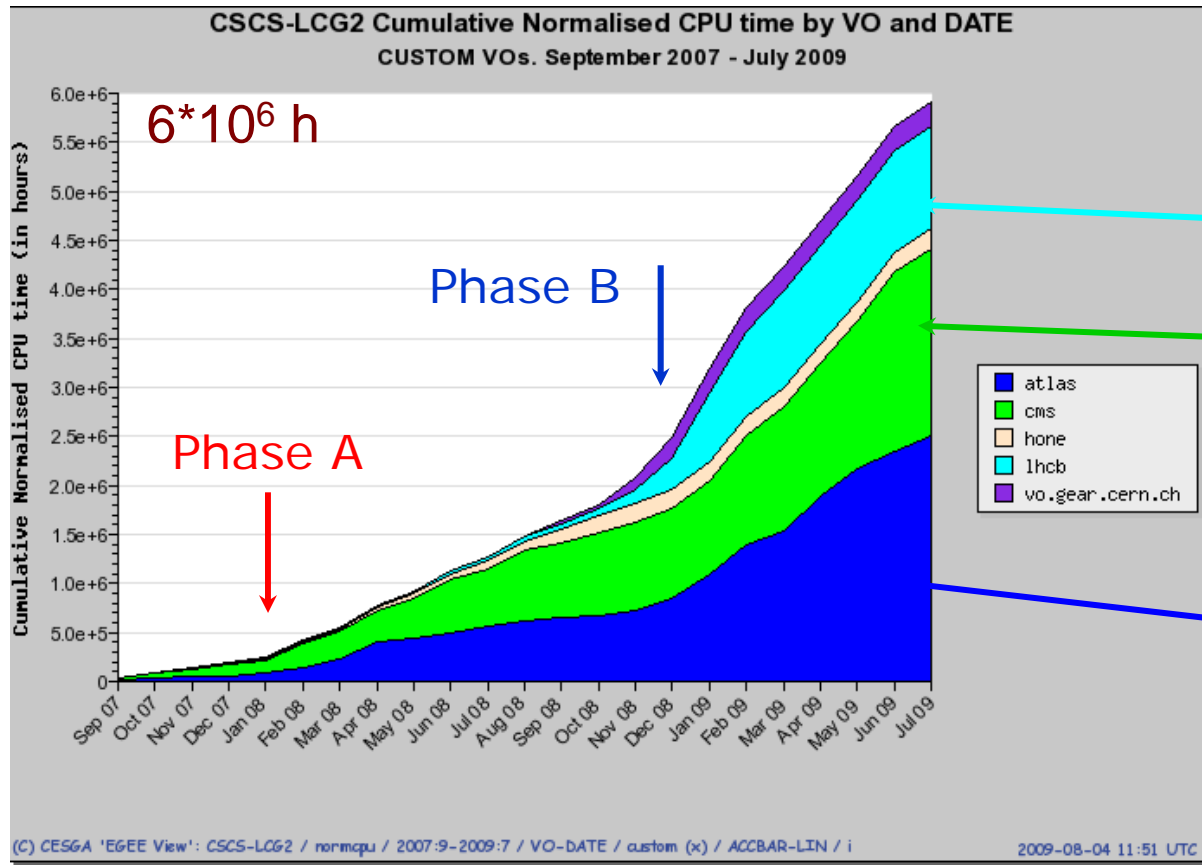
- ❖ **Storage:** 27 X4500 - systems
net capacity of 510 TB total

Performance Plots of Tier-2 Operations

Swiss Tier-2 availability (1/09-7/09)

- **Site availability**
for all CMS Tier-2
(for CMS running)
- **Our Tier-2**
T2_CH_CSCS
reached ~ 94%
availability
- **want to reach >95%**





Incremental CPU-usage over last 2 years; reach 6*10⁶ hours

LHCb

CMS

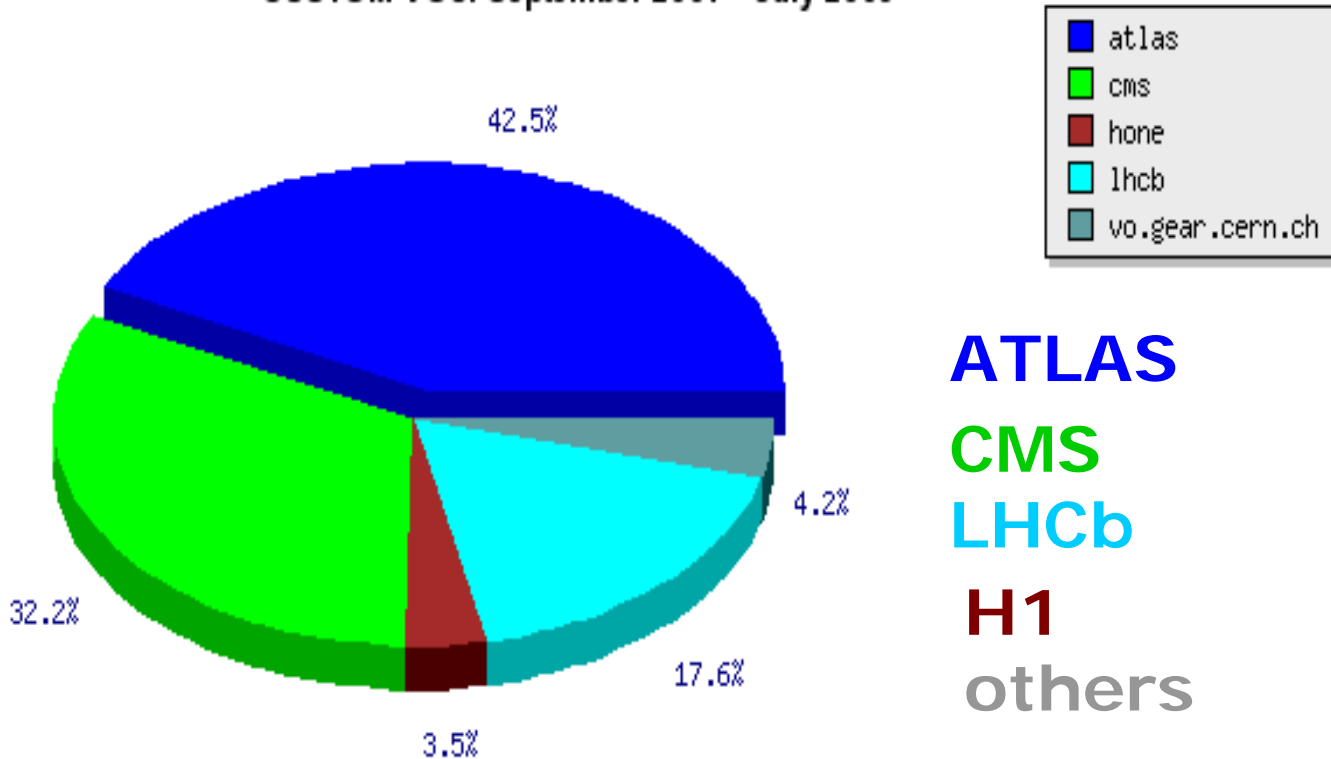
ATLAS

- **Tier-2** is up and has been in stable operation ~4 years ! continuous contributions of resources to experiments.
- Spare cycles given to other VO (eg. H1, theory (CERN) ...)

CSCS-LCG2 Normalised CPU time per VO (Excluded dteam and ops VOs)

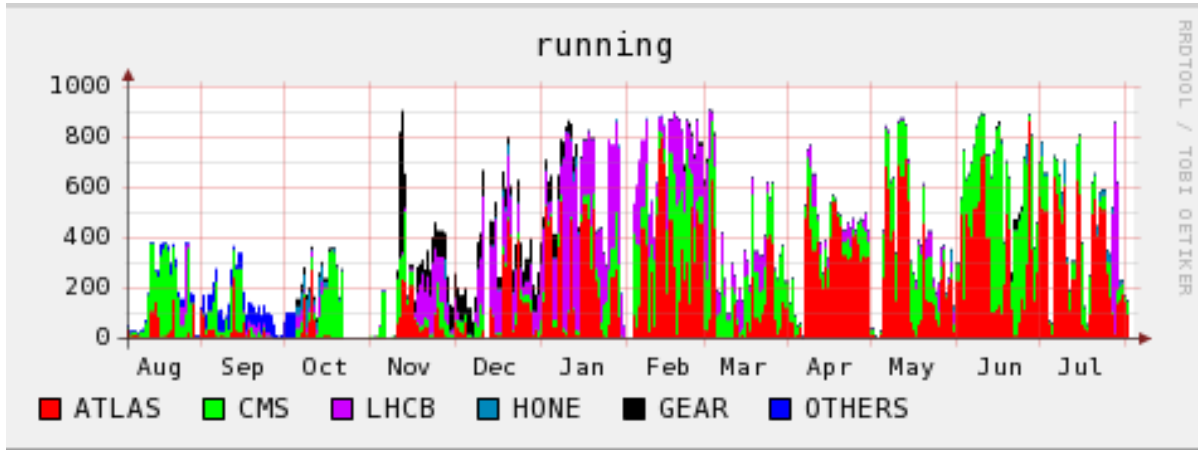
Swiss Tier-2

CUSTOM VOs. September 2007 - July 2009



- Shares between VOs overall reasonably balanced.

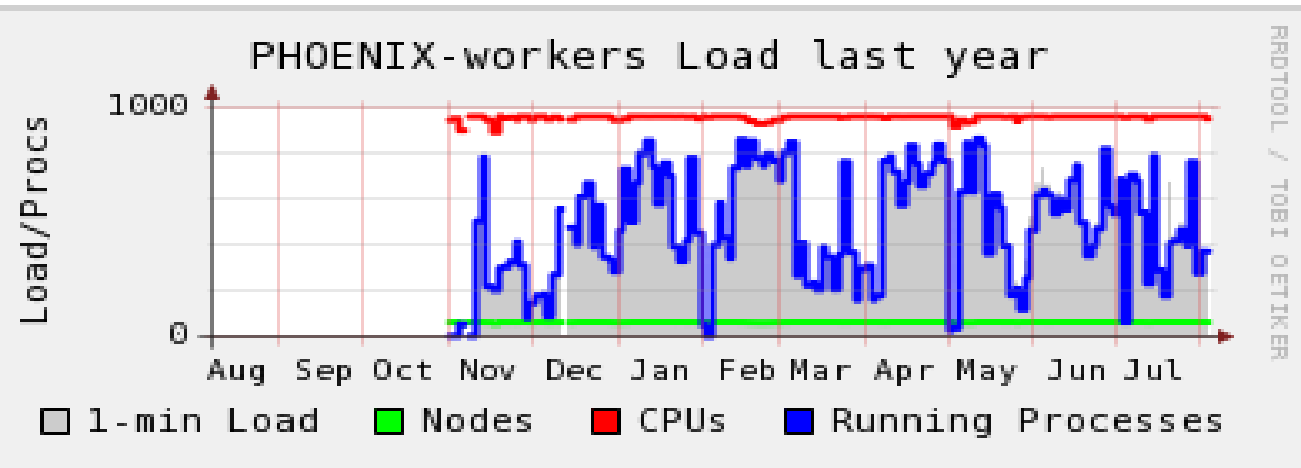
- Number of running Jobs



- CPU usage: running processes

- Load :
Up to the 960 running jobs !

→ cluster is pretty well utilised



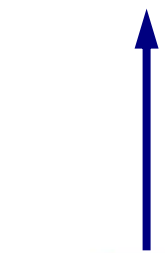
Status - Phase C

This is the last phase of ramp-up

- Evaluation + negotiations with CSCS and SUN about different options in progress (Jul/Aug/Sep 09)
- Plan to follow a **staged scenario**:
 - ➔ add an **higher bandwidth** Infiniband **backbone**
 - ➔ **replace local disk scratch** system by global system (Lustre?)
 - ➔ **upgrade/add** storage elements (SE with ZFS on disk)
 - ➔ **upgrade/add** worker nodes (CPU)
 - ➔ low power, high-density (question of availability, Intel ?)
- **Schedule**:
 - ➔ Finalize the architecture and staging process in ~Sep '09
 - ➔ **goal: maintain Phoenix-B operation in parallel to upgrade !**
NEVER interrupt production running of cluster after Nov.1
 - ➔ SE + WN upgrade/additions in parallel in Q1/10 .
 - ➔ need to comply with power restrictions at CSCS
(time with decommissioning of IBM Power-5 system).

Network

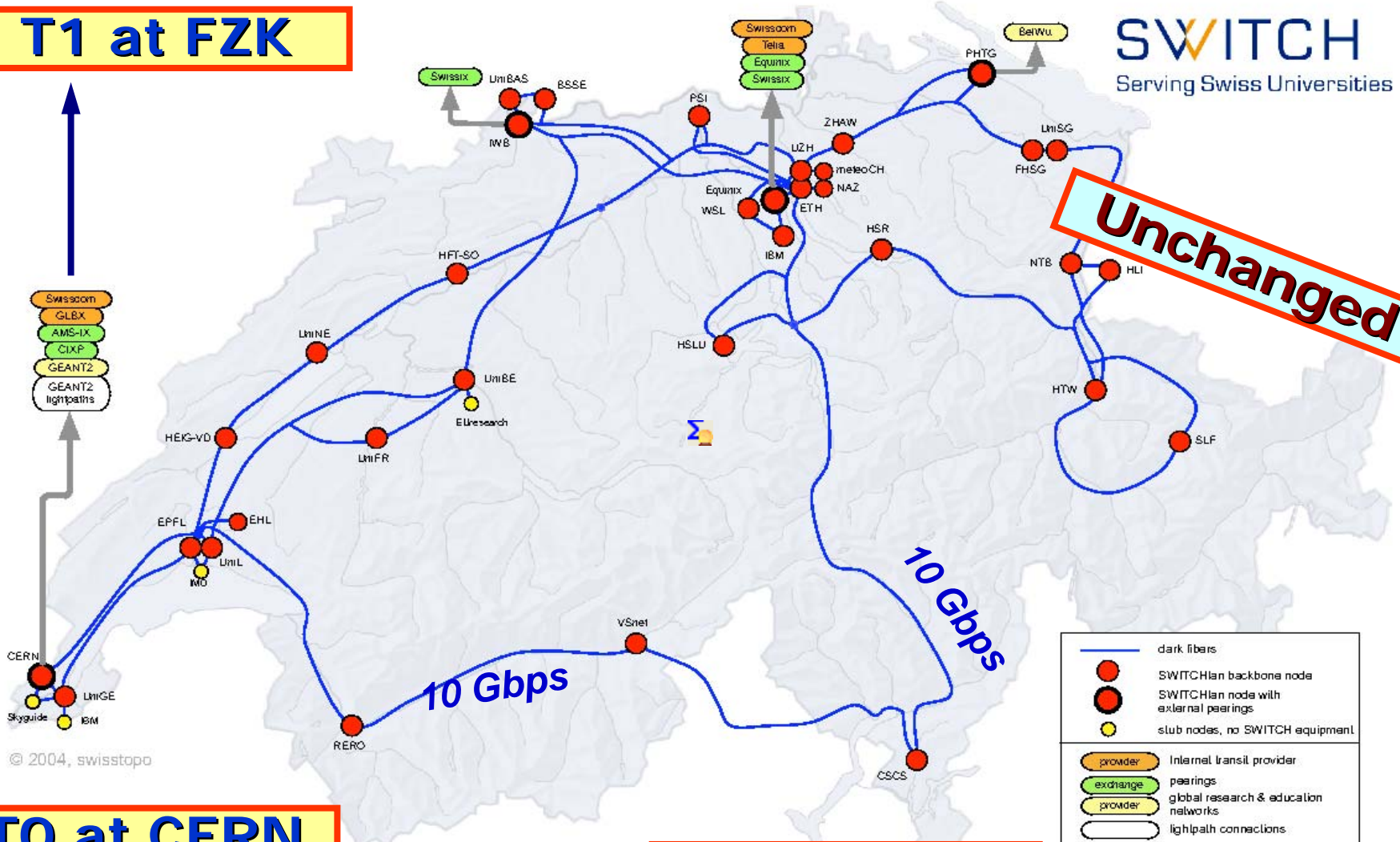
T1 at FZK



- Swisscom
- GLBX
- AMS-IX
- CIXP
- GEANT2
- lightpaths

SWITCH
Serving Swiss Universities

Unchanged



T0 at CERN

T2 at CSCS

SWITCHlan Backbone: Dark Fiber Topology May 2009

Status of the Swiss Tier-3 Centres

Swiss Tier-3 Efforts

- **Large progress** seen over last year for all 3 experiments. upgrades in progress nearly everywhere !



- **Close national collaboration between Tiers:**

- Tier-3 contacts are ALSO experiment's site contacts for CH Tier-2.
- close contacts to **Tier-1 at FZK** .



- **ATLAS** : operates the **Swiss ATLAS Grid** → federation of clusters at

- **Bern** uses local HEP + shares university resources
- **Geneva** operates local cluster



- **CMS** : **ETHZ + PSI+ UZH** run a **combined Tier-3**

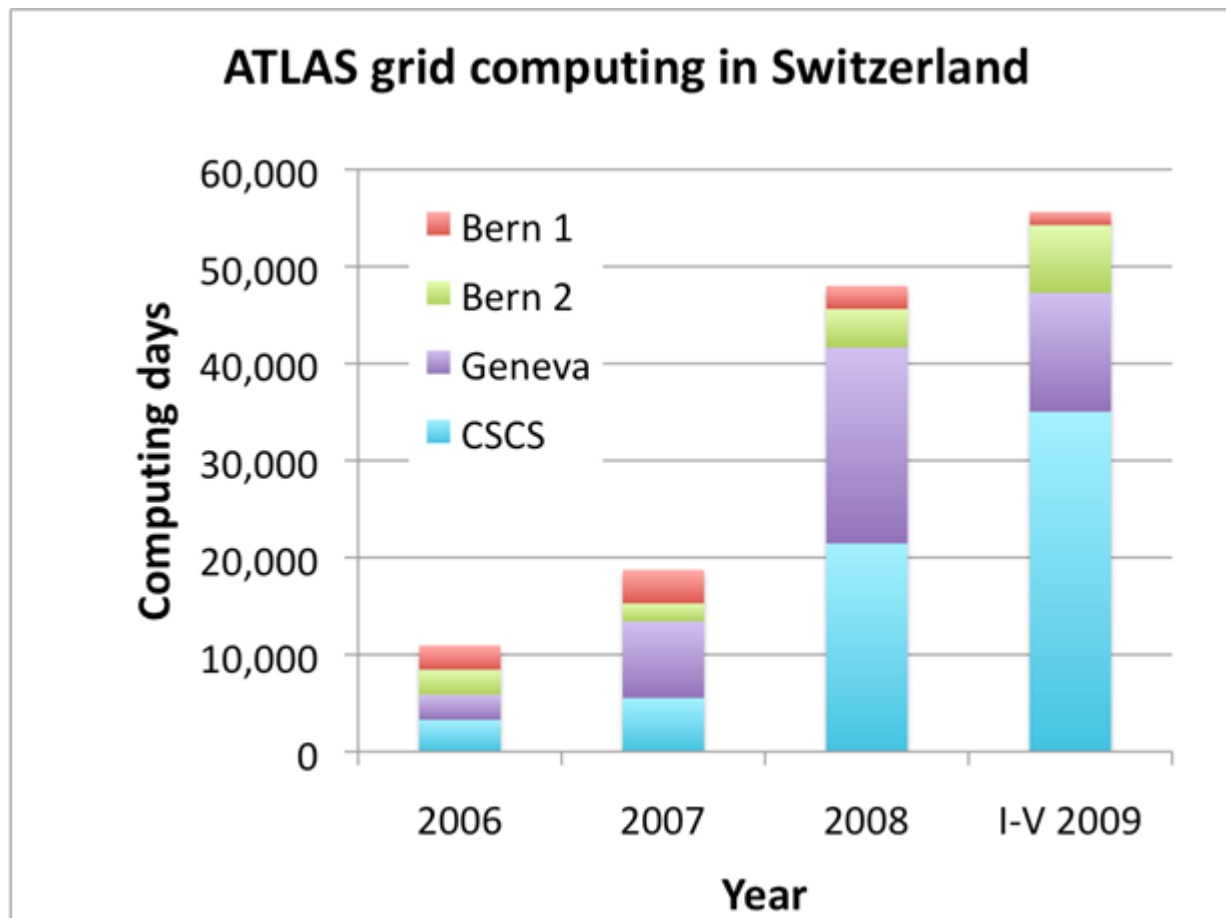
- located at and operated by PSI IT



- **LHCb** :

- **EPFL** : operates large local cluster → DIRAC pilot site
- **UZH** uses local HEP + shares university resources (upcoming)





~1.4%
of ATLAS
computing
in 2008

Summary: Swiss Tier-3 Efforts (Q3/09)

Site (#users)	Nr cores	CPU (kSI2k)	Storage (TB)	Comments
ATLAS BE (10) GE (~55)	30+300sh 268	~600 462	33 180	BE: standing Atlas production; GE: identical SW-environment to CERN; direct line to CERN.
CMS ETHZ, PSI, UZH (21)	72	~250	105	GRID SE + UI :direct GRID access. → triple size upgrade in Q4/09
LHCb EPFL (15) UZH (4)	464 shared	~800 125	36 15	EPFL is DIRAC site; identical machines as in pit UZH:MC production; shared → large upgrade in Q4/09
Total Tier-3		~2250	366	cf: Tier-2: 1600 kSI2k, 510 TB

- Tier-3 capacities : similar size as Tier-2
- Substantial investment of resources for MC production+local analysis.
- Note: CPU numbers are estimates; upgrades in progress ...



- Up to now the overall “operation, R&D etc.” for the GRID was done by EGEE (Enabling Grid for E-scienceE) [2004-2009]

- In 2010, EGEE-III will be succeeded by EGI, the “European Grid Initiative”,
 - “EGI will provide a more sustainable way to centrally coordinate, evolve and operate the current European grid infrastructure”
 - EGI depends on and is supported by the **NGI (National Grid Initiatives)**, set up per country to manage the national grids
 - Status: MoU provides a framework for setting up the EGI – in progress (to be able to participate in upcoming FP7 call of the EU in Nov 09)



- Switzerland has NO official NGI yet:
 - SWING (Swiss National Grid Initiative)** is destined to assume this role.
 - ad interim (2a), **SWITCH** agreed to act as **lead institution** and take on this role;
 - ➔ **SWITCH has officially signed the MoU as of last Friday!**



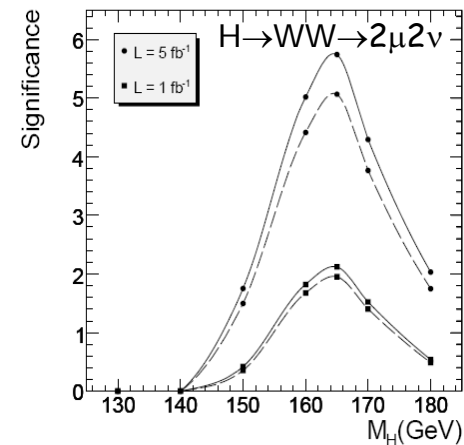
- Swiss NGI will have to provide services:
 - CHIPP/CSCS will have to provide essentially the same services**, as provided now to WLCG (through EGEE): support, operations etc.
 - further resources expected from SER for sustainable operation as NGI.



- **Swiss Tier-2 reliably operates and delivers the Swiss pledges to the LHC** experiments in terms of computing resources since Q2/2005
 - ➔ Phase-B operates with high availability, is well utilised
 - ➔ Last ramp-up Phase-C in progress - done by Q1/2010.
- **Tier-3 centres** are growing in size and importance, and strongly complement Tier-2 :

Running and READY for PHYSICS !

<https://twiki.cscs.ch//bin/view/LCGTier2/WebHome>



Thanks – CCB + T2/3 Personnel



C.Grab (ETHZ) [chair CCB]
D.Feichtinger (PSI)
Z.Chen (ETHZ)



S.Gadomski, A.Clark (UNI Ge)
S.Haug, H.P. Beck (UNI Bern)



R.Bernet (UNIZH)
P.Szczyпка, J. Van Hunen (EPFL)



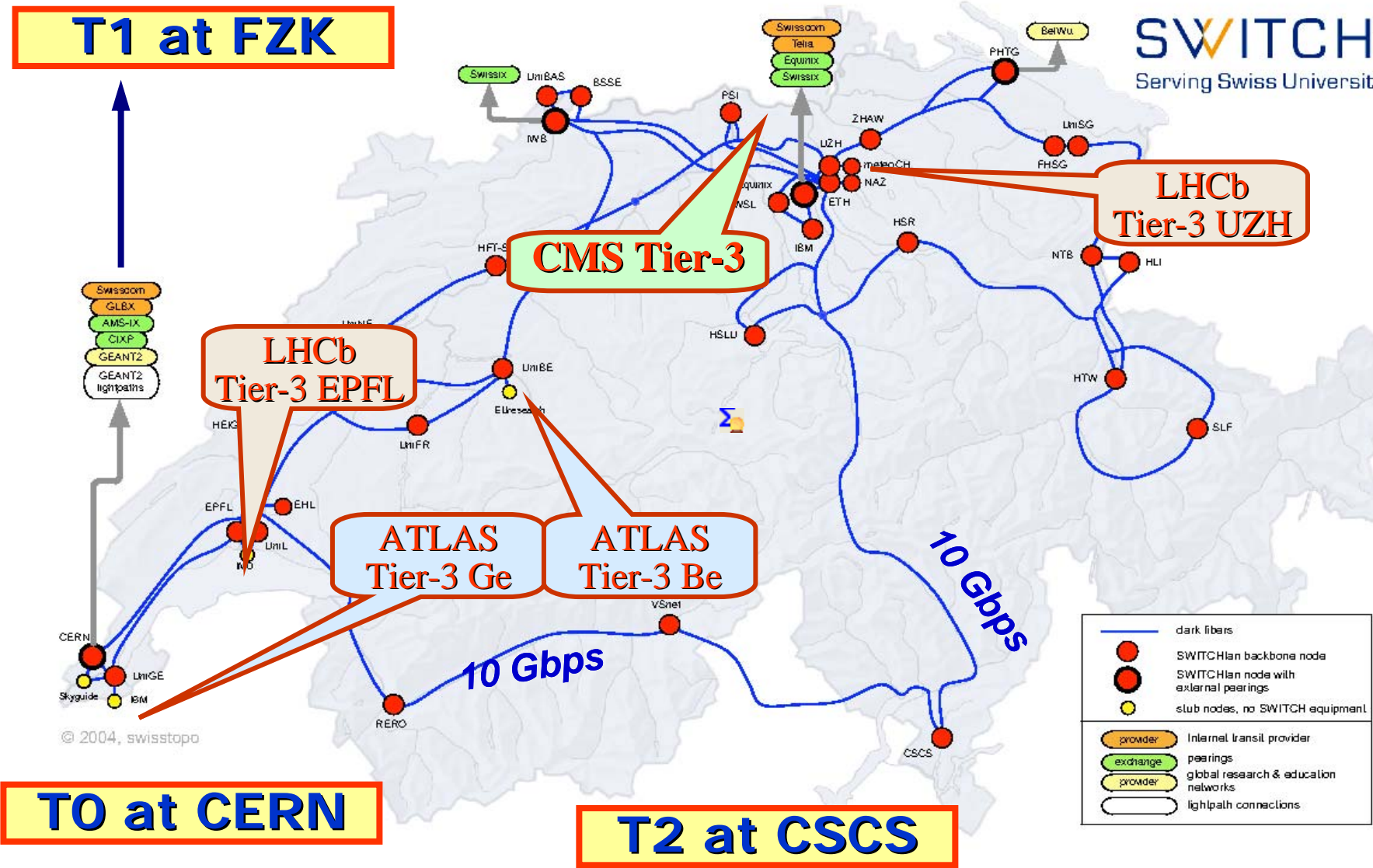
F. Georgatos, J.Temple, R.Murri, D.Ulmer [CSCS]
[+ P.Kunszt, S.Maffioletti (left)]

and many more ...

Optional slides

Status

Swiss Tier-3 Centres in Detail



SWITCHlan Backbone: Dark Fiber Topology May 2009

Two clusters with NorduGrid front ends in production since 2005. UBELIX is shared and operated by “Informatik Dienste UniBE”. LHEP is dedicated ATLAS. For local physics analysis and simulation. Fills up with ATLAS central production jobs when not used by locals.



Current T3 Hardware (UBELIX+LHEP)

- ~ 300 cores for ATLAS (~500 in 2010).
- ~ 33 TB disk (~ 50 TB in 2010).

Usage/Operation

(~10 local users + central production)

- ~ 5000 Wall Time Days in 2006.
- ~ 7000 Wall Time Days in 2007.
- ~ 6500 Wall Time Days in 2008.
- ~ 15500 Wall Time Days in 2009 Jan-Jul

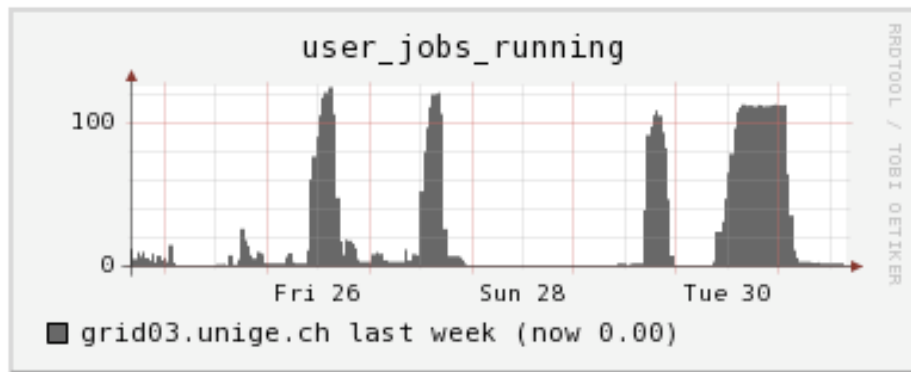
(Local users are also using Geneva T3 and CSCS T2 resources.)

S. Haug, Bern 2009-08-13

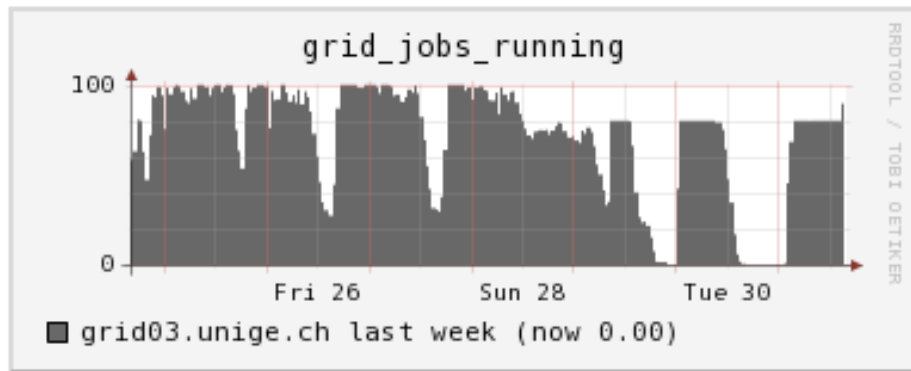


- 268 CPU cores (login + batch)
→ some total 462 kSI2k
- 180 TB for data
- the analysis facility for Geneva group
- grid batch production for ATLAS
- special features:
 - direct line to CERN at 10 Gb/s
 - latest software via CERN AFS
 - storage integrated with ATLAS data distribution systems
 - flow of data from CERN and from NDGF Tier 1
- used for Trigger development and commissioning
- complementary to CSCS Tier 2

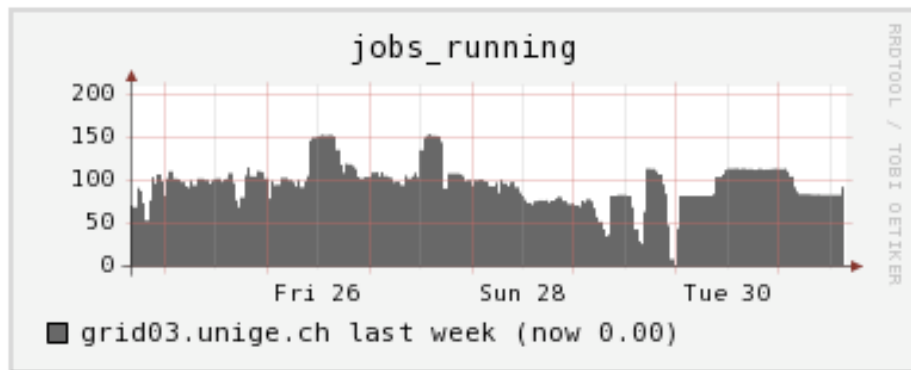
Added value by resource sharing



local jobs come in peaks



grid always has jobs



little idle time, a lot of Monte Carlo done

Common CMS Tier-3 for ETH, PSI, UZH groups in operation at PSI since Q4/2008

- 1 Gbps connection PSI → ZH
- Upgrade in 2009 by factor 3 planned.

Year	2008	2009
CPU / kSI2k	215	1000
Disk / TB	105	250

- ~ 22 users
- operates GRID storage element and user interface to enable users direct GRID access.
- Local production jobs



+ 2 more X4500

D. Feichtinger + C. Grab

Specs and details are available on the wiki

<https://twiki.cscs.ch/twiki/bin/view/CmsTier3/Tier3Overview>

Hardware and Software:

- Machines identical to those in the LHCb pit
 - 58 Worker nodes x8 cores (~840 kSI2k)
 - 36 Tb of storage
- Uses SLC4 binaries of LHCb software and DIRAC Development builds



Current Status and Operation:

- EPFL is one of the pilot DIRAC Sites
- Custom DIRAC interface for batch access
- Active development to streamline GRID usage
- Aim to run official LHCb MC production



P. Szczypka

no changes 8/09



- Zurich HEP Cluster:
 - Intel Cluster, 32-bit, openSUSE
 - DIRAC installation running SLC4 binaries
 - Local openSUSE LHCb Software Installation → local LHCb jobs
 - Hardware: CPU: 125 kSI2k, Disk: ~15 TB
 - Concerns: File Access to RAID might be bottle neck
- Zurich Schrödinger Cluster:
 - Intel Xeon Cluster, 64-bit, 4600 Cores
 - Replaces the old Matterhorn Cluster, will come in service in autumn
 - Used by other institutes, load will probably vary a lot
 - Will mainly be used for Monte Carlo production



R.Bernet (Aug.09)