Status Report of the BFG Community

Raphaël Pesché

Rechenzentrum

Albert-Ludwigs-Universität Freiburg

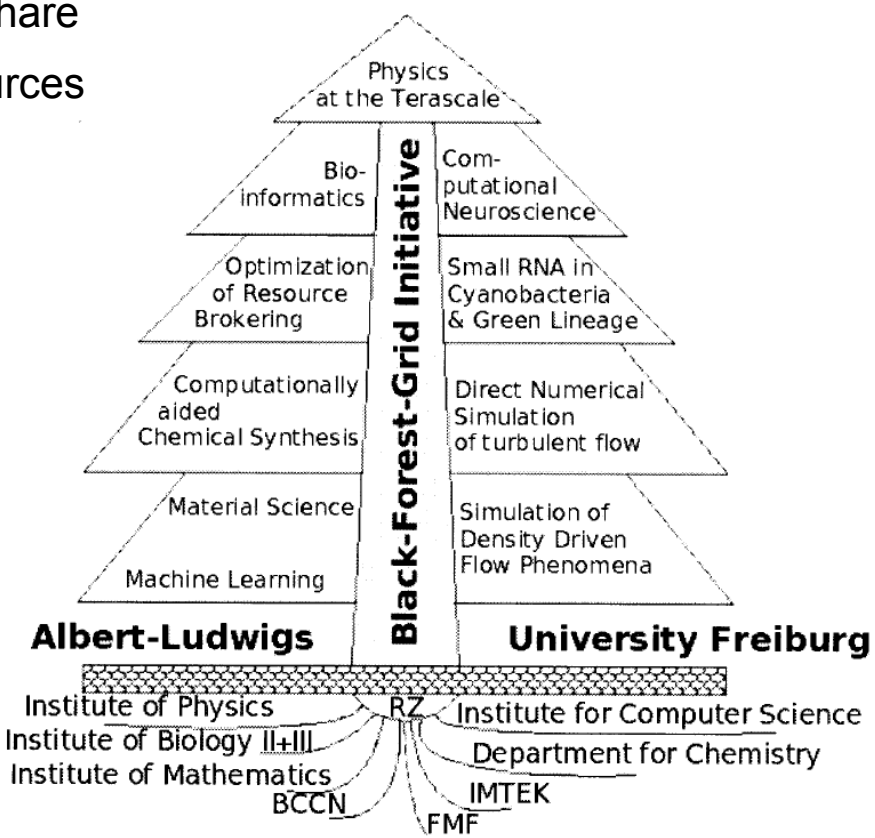
5th Black Forest Grid Workshop

23rd April 2009

Black Forest Grid

The Black Forest Grid Initiative

- **Collaborative effort of several departments** at the University of Freiburg to share available computing resources
- **Based on basic agreements** on hardware and operating systems
- HW hosted at the Computing Center



Outline

- **Projects in the BFG**
- **Status of the BFG Cluster**
  - Available computing resources
  - Monitoring and administration
  - Resource usage
- **Future improvements**
  - Software
  - Computing and storage resources

Overview of the BFG Projects

Bioinformatics
Lehrstuhl für Bioinformatik,
Institut für Informatik

Computational aided chemical syntheses
Institut für Anorganische und
Analytische Chemie

Computational Neuroscience
Bernstein Center for Computational
Neuroscience (BCCN)

Machine Learning
Computer-based New Media group
(CGNM), Institut für Informatik

Numerical Simulation of turbulent flow using Digital Lattice Boltzmann Automata
Institut für Mikrosystemtechnik

Particle Physics at the Tera Scale
Physikalisches Institut

Material Science
Service-Group Scientific Data Processing,
Freiburger Materialforschungszentrum

Simulation of Density Driven Flow Phenomena
Abteilung für Angewandte Mathematik

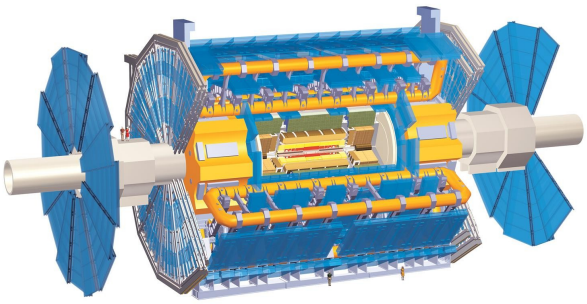
Data Mining
Lehrstuhl für maschinelles Lernen und
natürlichsprachlicher Systeme

Optimization of Resource Brokering
Service-Group Scientific Data Processing,
Freiburger Materialforschungszentrum

Seizure Prediction
Epilepsy Center, BCCN, FDM

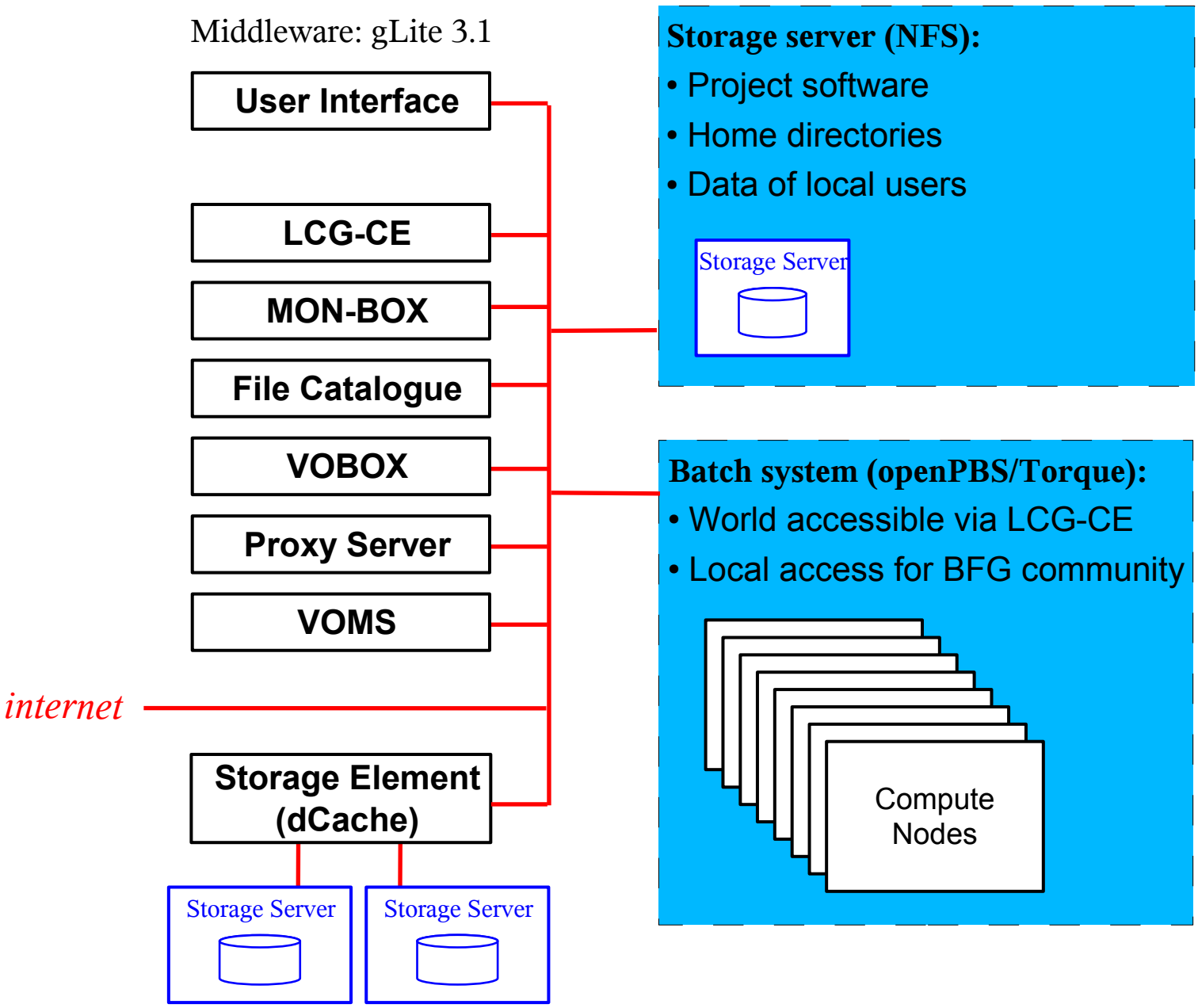
Computer Based Meaning Research
Deutsches Seminar

GRID Computing and Atlas experiment



- **Atlas** is a detector for the LHC at the CERN
- A huge amount of data is produced by the ATLAS experiment (**> 100TB** per year)
- ATLAS follows decentralized grid computing approach
  - Hierarchical model of computing centers (Tiers), each Tier with a different role
  - Wordwide: 10 Tier-1s, ~50 Tier-2s
  - The BFG installation in Freiburg serves as an official ATLAS Tier-2 center

Overview of the BFG cluster



The batch system

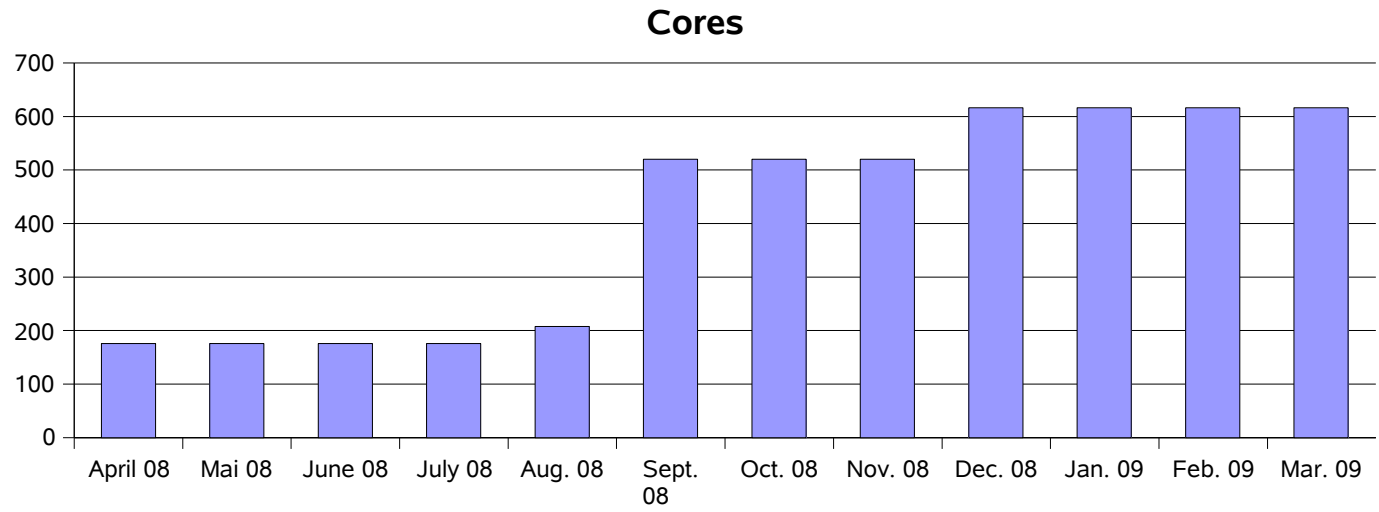
- Cluster resources interfaced by batch system (**openPBS/Torque**)
  - World accessible via LCG Computing Element
  - Virtual organizations:
    BFG, ATLAS, DECH, DTEAM, GHEP, ILC, OPS
  - Local access for BFG community
- **Shared home directory** throughout the cluster
- Additional **storage space** available via NFS
  - project software installation
  - data of local users

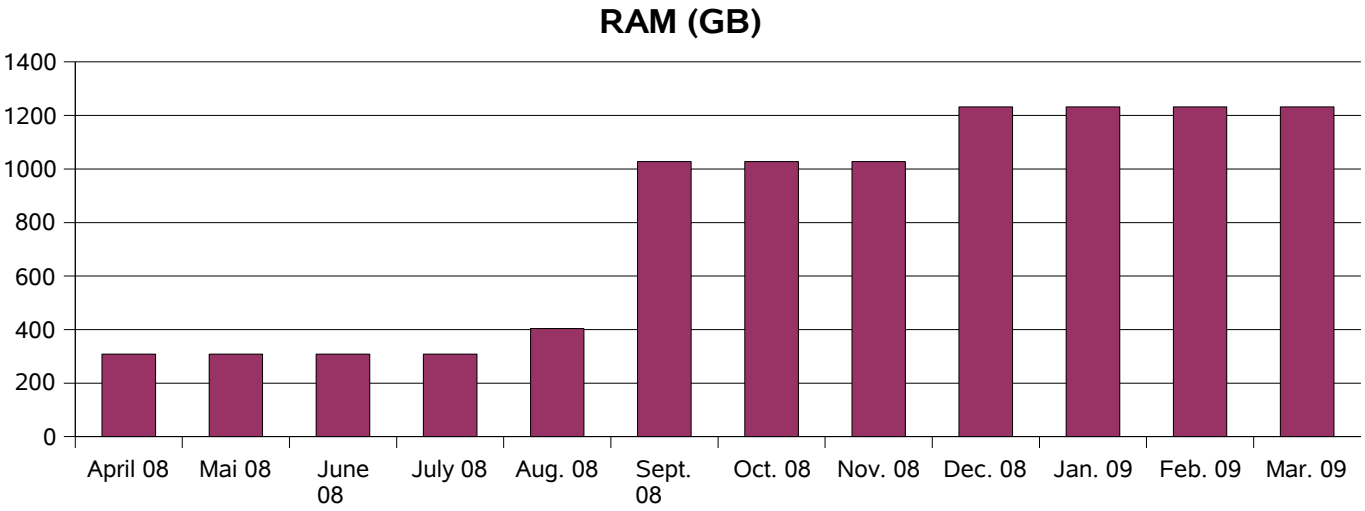
Compute nodes

- **SUN x4100** (31) + **SUN v20z** (7)
  - 2x AMD Opteron 248, 2.2 GHz
  - 4GB RAM
- **SUN x4600** (1)
  - 16x AMD Opteron 8220, 2.8 GHz
  - 32GB
- **DELL PE1950** (19)
  - 4x Intel Xeon 5160, 3.0 GHz
  - 8GB RAM
- **HP BL460c** (40)
  - 8x Intel Xeon E5345, 2.33 GHz
  - 16GB RAM
- **HP BL220c** (16)
  - 8x Intel Xeon E5440, 2.83 GHz
  - 16GB RAM

→ **616 Cores / 1.2TB RAM**

Worker Nodes: Rise of the Machines





Cluster management with quattor

- Toolkit for automated installation, configuration and management of clusters
- Started in the scope of **EDG**, development coordinated by **CERN**
- Configuration management infrastructure:
  - CVS-based configuration database (CDB)
  - Configuration expressed in a dedicated language (PAN)
  - PAN templates are to compiled to XML
  - One XML file per node containing a full description of its configuration

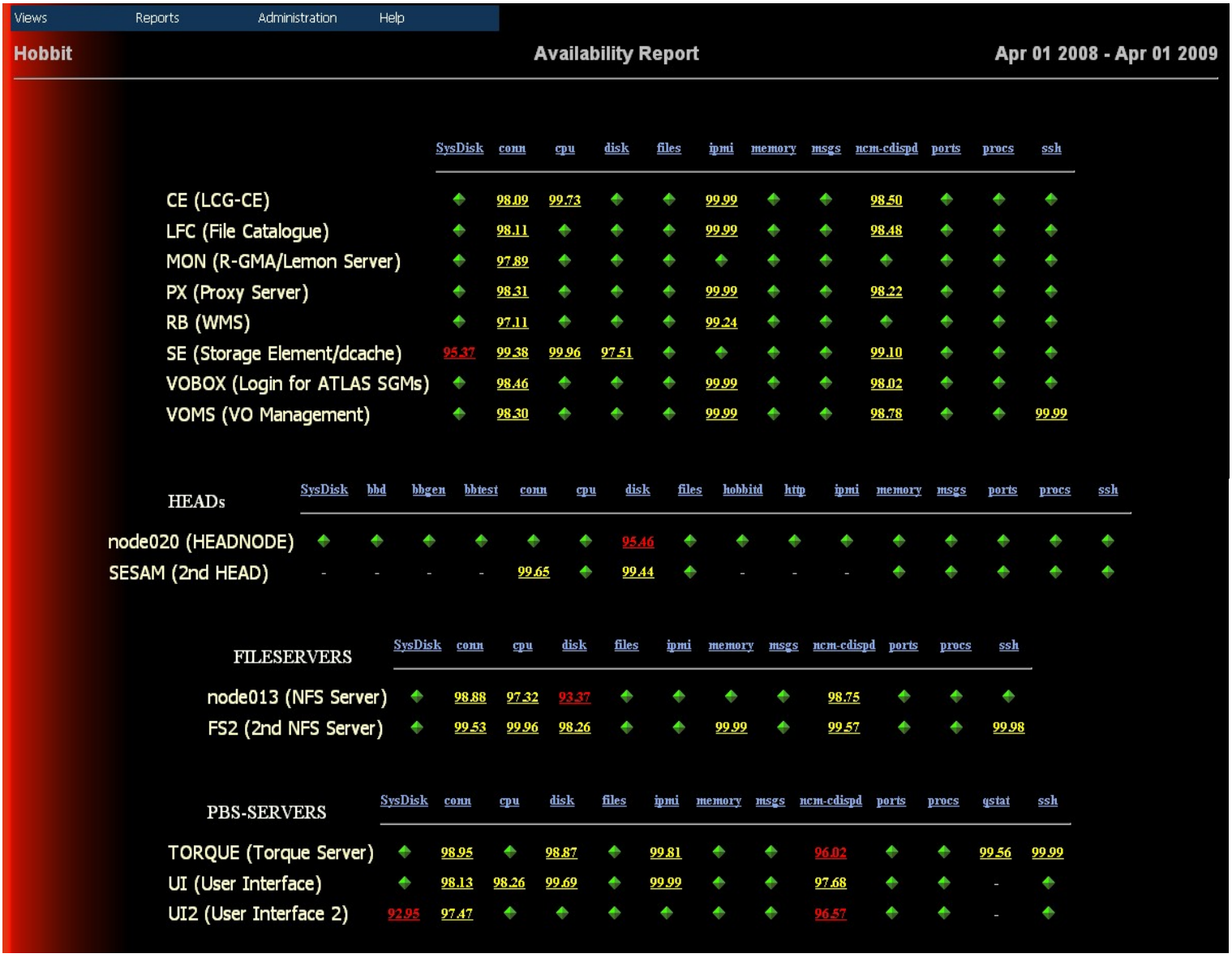
Monitoring of the cluster

- Automated monitoring of cluster resources with web-based front-ends ***Hobbit*** and ***Lemon***
- ***Hobbit***: monitoring servers and networks
  → monitors availability of network connectivity, service, and applications running on cluster servers and compute nodes
- ***Lemon*** (LHC Era Monitoring): server/client based monitoring system
  - Sensors running on the monitored systems measure information about CPU and network utilization, disk I/O, memory consumptions etc.
  - Information is locally cached and forwarded to the lemon server

Monitoring of the cluster:

availability report of the cluster servers

(April 2008 – April 2009)



Views Reports Administration Help

Hobbit — **Availability Report** — **Apr 01 2008 - Apr 01 2009**

	SysDisk	conn	cpu	disk	files	ipmi	memory	msgs	ncm-cdispd	ports	procs	ssh
CE (LCG-CE)		98.09	99.73			99.99			98.50			
LFC (File Catalogue)		98.11				99.99			98.48			
MON (R-GMA/Lemon Server)		97.89										
PX (Proxy Server)		98.31				99.99			98.22			
RB (WMS)		97.11				99.24						
SE (Storage Element/dcache)	95.37	99.38	99.96	97.51					99.10			
VOBOX (Login for ATLAS SGMs)		98.46				99.99			98.02			
VOMS (VO Management)		98.30				99.99			98.78			99.99

HEADs	SysDisk	bbd	bbgen	bbtest	conn	cpu	disk	files	hobbitd	http	ipmi	memory	msgs	ports	procs	ssh
node020 (HEADNODE)							95.46									
SESAM (2nd HEAD)	-	-	-	-	99.65		99.44		-	-	-					

FILESERVERS	SysDisk	conn	cpu	disk	files	ipmi	memory	msgs	ncm-cdispd	ports	procs	ssh
node013 (NFS Server)		98.88	97.32	93.37					98.75			
FS2 (2nd NFS Server)		99.53	99.96	98.26			99.99		99.57			99.98

PBS-SERVERS	SysDisk	conn	cpu	disk	files	ipmi	memory	msgs	ncm-cdispd	ports	procs	qstat	ssh
TORQUE (Torque Server)		98.95		98.87		99.81			96.02			99.56	99.99
UI (User Interface)		98.13	98.26	99.69		99.99			97.68			-	
UI2 (User Interface 2)	92.95	97.47							96.57			-	

Monitoring of the cluster: compute nodes (April 2008 - April 2009)

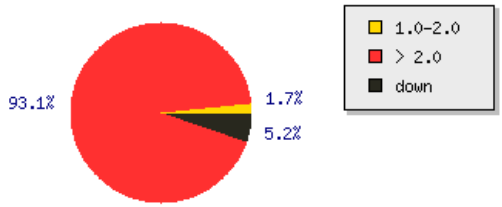
Lemon Monitoring

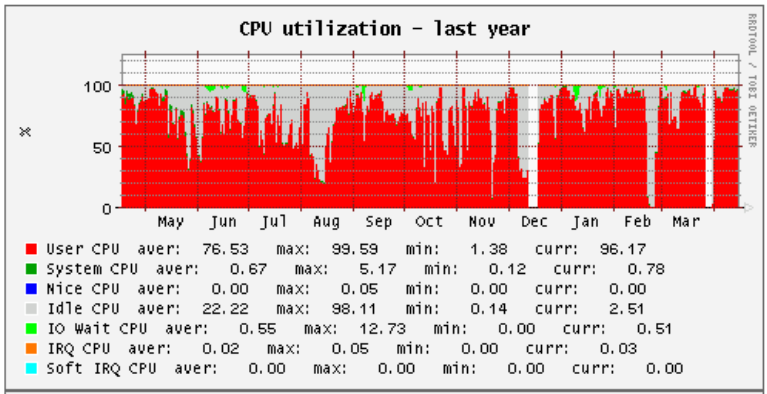
Home | Documentation | Alarms | Help

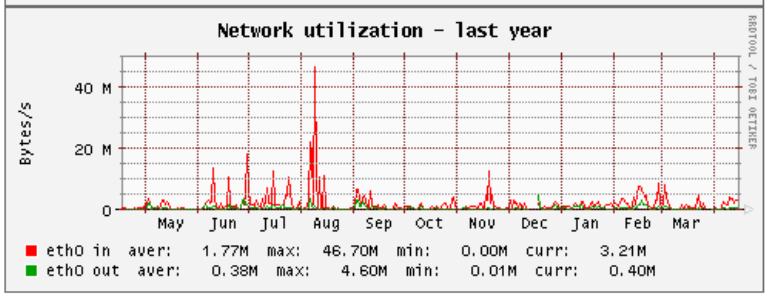
Cluster info: Worker Nodes — 16 Apr 2009 Thu 10:40:03

Cluster Information	
# of hosts (down):	58 (3)
operating system(s):	NA
# of CPUs (down):	148 (0)
average up time:	13 days, 2h:13m (boots per host)
hosts down:	node041, node073
exceptions:	NO_CONTACT
Select from hosts:	None
Metric Distributions	Correlations

Load Percentages







Last: year | Get Fresh data | Get Detailed data | Auto Update

Search host: | Virtual Clusters | Racks

Usage of cluster resources

- Major part of resources used by **BFG** and **ATLAS** VO's
- March 2008 – April 2009:

	CPU Time total [h]	Wall Time total [h]	CPU Time per job [h]	CPU/Wall per job [%]
ATLAS	754050	871235	2:00	87
BFG	2009756	1227775	14:00	163
	2763806	2099010		

Planned improvements of the BFG cluster

- **Diskless boot** of the worker nodes
- Upgrade to **Scientific Linux 64 bit**
- **Lustre** instead of NFS to mount the home directories
- **IBM Bladecenter H**: 140 x 8 Cores, 16GB each
- 20 **Dell PE 1950**
- 30 **Sun x2250**