Karlsruher Institut für Technologie

# GridKa Report

Joint TAB + GridPB meeting, Dresden, 2.12.2010

Steinbuch Centre for Computing

# Inhalt

- Gemessene Zuverlässigkeit
- Störungen / Probleme
- Performanz der Speichersysteme
- WLCG Site Security Challenge 4
- Personal
- Ressourcen 2011
- WLCG Diskussionen

# Site reliability

- Durchschnitt 1.5.2010 - 15.09.2010
  - http://lcg-sam.cern.ch:8080/reports/site_avail.xsql

| | CERN | KIT | BNL | IN2P3 | INFN | NDGF | NIKHEF | PIC | RAL | SARA | Taiwan | Triumf | FNAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **OPS** | **100** | **98** | **99** | **100** | **98** | **98** | **99** | **99** | **99** | **91** | **96** | **99** | **100** |
| Alice | 100 | 97 | | 98 | 99 | 100 | 99 | | 97 | 95 | | | |
| Atlas | 98 | 97 | 96 | 97 | 96 | 97 | 94 | 88 | 97 | 83 | 88 | 99 | |
| CMS | 99 | 91 | | 95 | 95 | | | 99 | 99 | | 93 | | 99 |
| LHCb | 91 | 88 | | 83 | 78 | | 0(?) | 84 | 88 | 79 | | | |

**16.09.10 bis 21.09.10:** **GridKa und einige andere Sites 'down' aufgrund 0-Day-Root-Exploit.**

# Site reliability

- Durchschnitt 22.9.2010 - 31.10.2010
  - http://lcg-sam.cern.ch:8080/reports/site_avail.xsql

| | CERN | KIT | BNL | IN2P3 | INFN | NDGF | NIKHEF | PIC | RAL | SARA | Taiwan | Triumf | FNAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **OPS** | **98** | **99** | **100** | **94** | **100** | **90** | **92** | **98** | **100** | **91** | **70** | **100** | **100** |
| Alice | 100 | 100 | | 97 | 95 | 100 | 90 | | 98 | 92 | | | |
| Atlas | 99 | 99 | 98 | 81 | 92 | 100 | 96 | 98 | 98 | 93 | 96 | 100 | |
| CMS | 100 | 99 | | 94 | 87 | | | 99 | 98 | | 99 | | 100 |
| LHCb | 95 | 96 | | 87 | 96 | | 0(?) | 91 | 77 | 90 | | | |

- Viele 'false positive' Fehler bei LHCb
- Tests bzw. 'Critical'-Klassifizierungen zum Teil veraltet.
  - Bsp.: LCG-CEs werden getestet, aber Cream-CEs werden benutzt (Alice)

# Störungen

- Kühlsystem-Ausfall  Samstag 10.7.2010, ca. 15:30 Uhr (1/2)
  - Kompletter Ausfall der SCC-Kühlanlage aufgrund von Überhitzung.
  - Anlage ist ausgelegt bis 40°C Aussentemperatur, Temp. war 37.5°C
  - Hitzestau wahrscheinlich verursacht durch Schallschutzwände.

  - Workernode-Racks haben sich abgeschaltet.
  - Viele (zentrale) Grid-Services bzw. Service-Bausteine betroffen.
  - Speichersysteme liefen weitgehend weiter (automatisch öffnende Schränke.)
  - ca. 18:30: Experten für alle wichtigen Systeme vor Ort.
  - ca. 22:30: Die meisten Services sind wieder online.
    - Workernodes bleiben über das gesamte Wochenende ausgeschaltet.

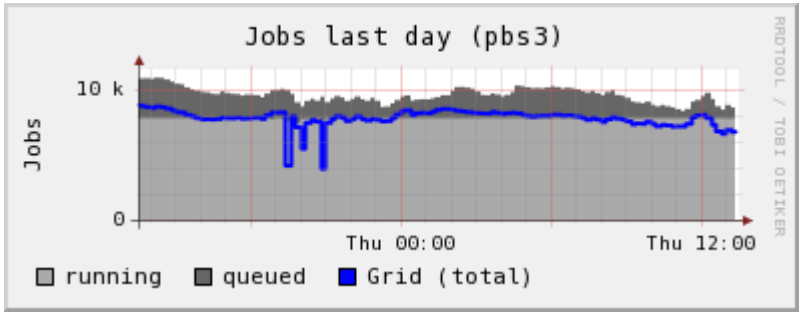
# Störungen

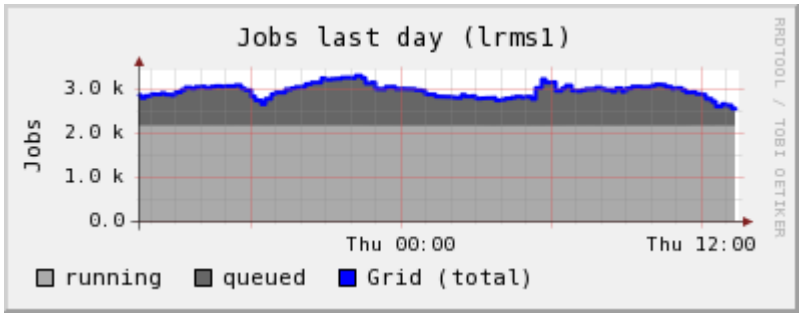
- Kühlsystem-Ausfall  Samstag 10.7.2010, ca. 15:30 Uhr (2/2)
  - Probleme beim KIT Alarm-Workflow (Keine Meldung des Kühlanlagen-Ausfalls an die GridKa-Rufbereitschaft.)
  - Defekte Hardware nach dem Kühlungsausfall:
    - Speicher-Controller(!)
    - Daten-Festplatten
    - CMS dCache Headnode Systemplatte
    - Workernode Netzteile und Systemplatten

# Probleme

- Seit Anfang 2010: Schlechte Performance des Batchsystems
  - Lange Antwortzeiten
  - Lange Scheduling-Zyklen

- => Ende Oktober: Aufteilung des Clusters in 2 getrennte Cluster
  - ca. 8000 Job-Slots + 2000 Job-Slots
  - LHC-Experimente Atlas, CMS und LHCb auf beiden Clustern
  - Nicht-LHC-Experimente und Alice auf größerem Cluster
- => Deutliche Verbesserung der Antwortzeiten und Scheduling-Zyklen
- Bisher gute Auslastung beider Cluster!



# Probleme

- Linux Kernel 'kswapd' Problem
  - kswapd Kernel-Thread verbraucht 100% CPU-Zeit
  - Workernodes nicht mehr ansprechbar.
  - Netzwerkverbindungen werden z.T. noch akzeptiert, blockieren aber.
  - Kernel gerät in diesen Zustand nach Problemen mit NFS-Shares.
  - PBS Server blockiert, wenn Workernodes in diesem Zustand sind
    - Neustart des PBS-Servers ist einzige Möglichkeit
  - Kein Bugfix Kernel/PBS in Aussicht.
  - Tool zum schnellen Auffinden solcher Workernodes mittlerweile vorhanden.
  - PBS Neustart nach Clustersplit deutlich schneller.

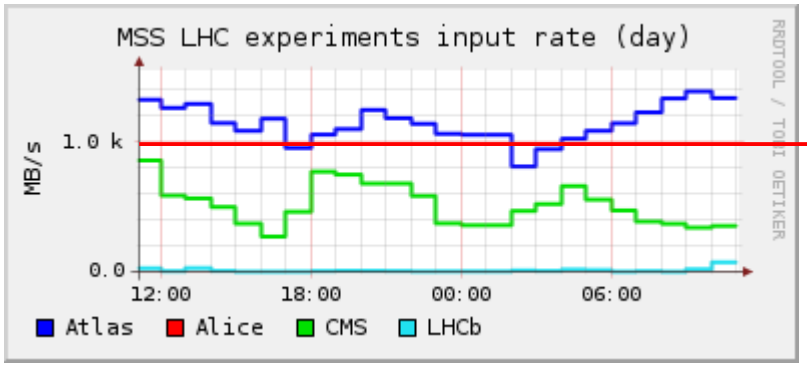
# Sun Grid Engine als alternatives Batchsystem?

- PBSPro kann nicht mit 'hängenden' Wokernodes umgehen. ('kswapd')
- Skalierungsprobleme bei PBSPro
  - Vorübergehend beseitigt durch Split in 2 Instanzen
  - Altair kündigte Performance-Verbesserungen im nächsten Release an.

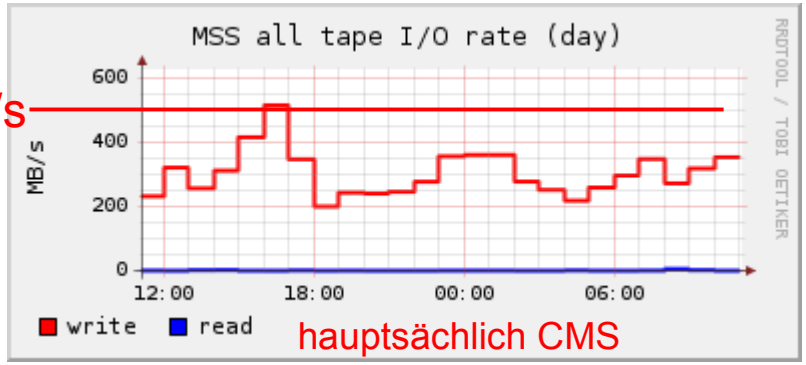
- Erste Tests mit Sun Grid Engine (SGE) sind vielversprechend.
- Unklare Zukunkt von SGE nach Übernahme von SUN durch Oracle
  - Kommerzieller Zweig + Open-Source-Zweig?
  - Kosten der kommerziellen Version?
  - Open-Source-Release: Weiterentwicklung? Bugfixes (Security!)?

# Netzwerk / Speichersysteme: Performance

- Snapshot vom 10./11.11.
  - Atlas ESD/AOD Verteilung nach Reprozessierung
  - Pb-Pb Daten-Import

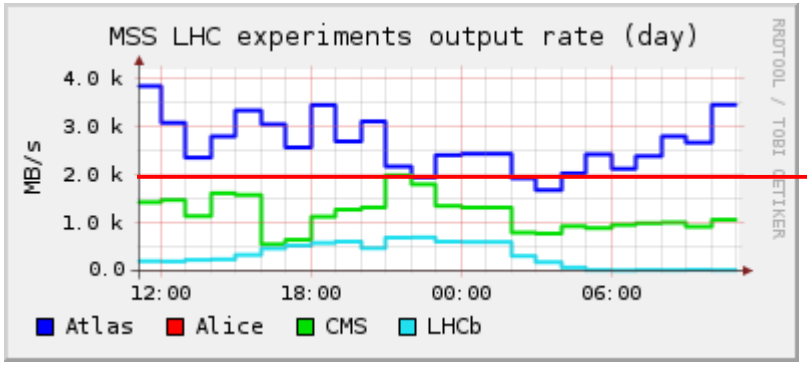



1GB/s




500 MB/s

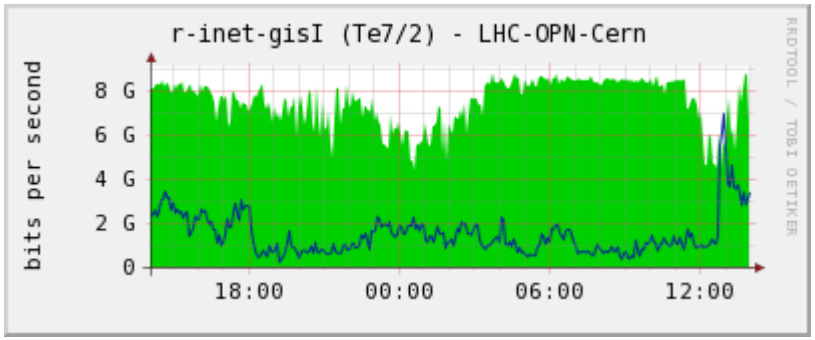
hauptsächlich CMS



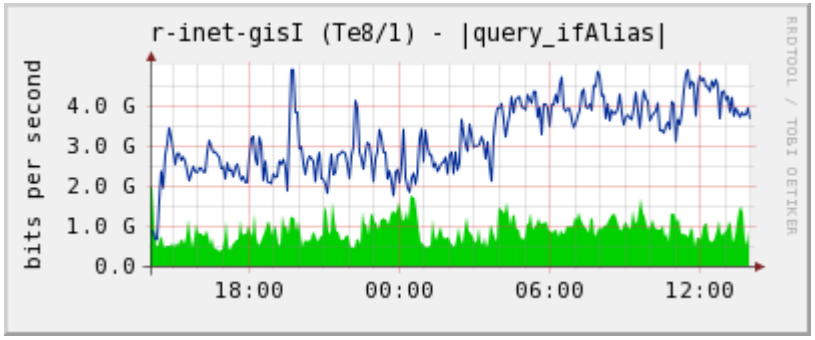

2GB/s

**Sehr gute Performance der GridKa Speichersysteme!**

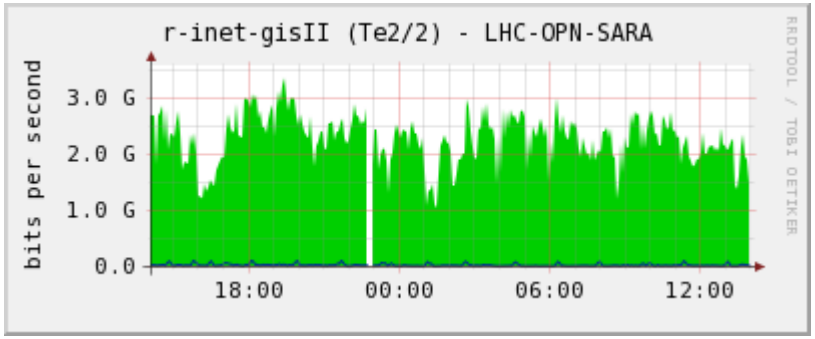
# Netzwerk / Speichersysteme: Performance



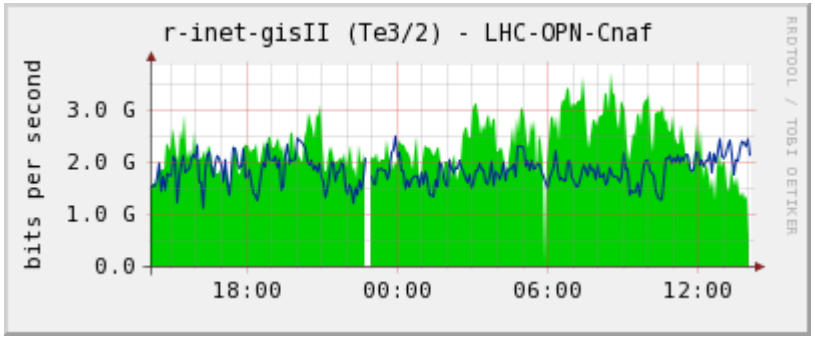

CERN

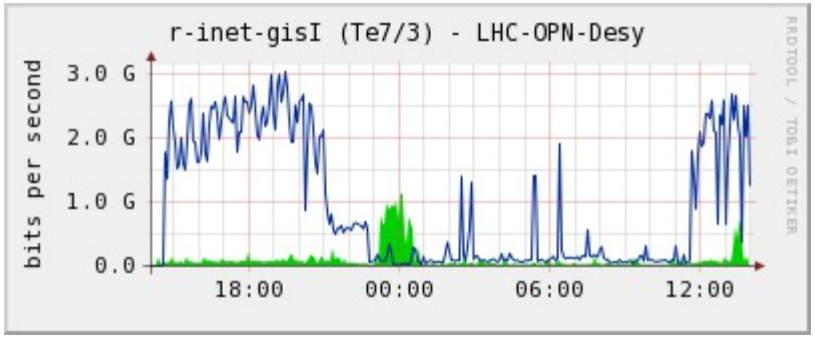



DFN XWIN




SARA und CNAF



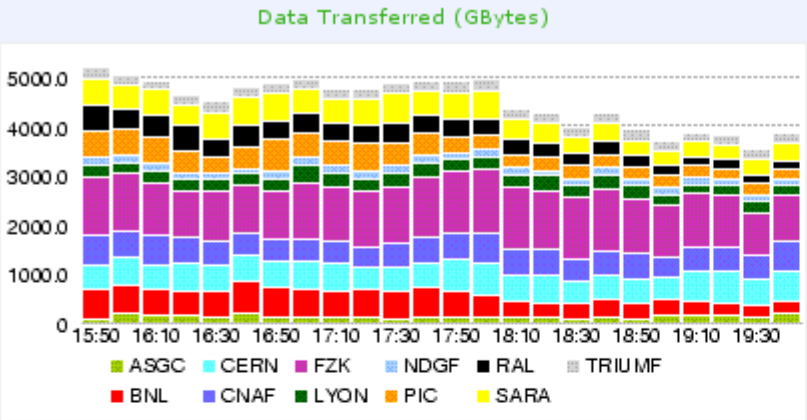

DESY

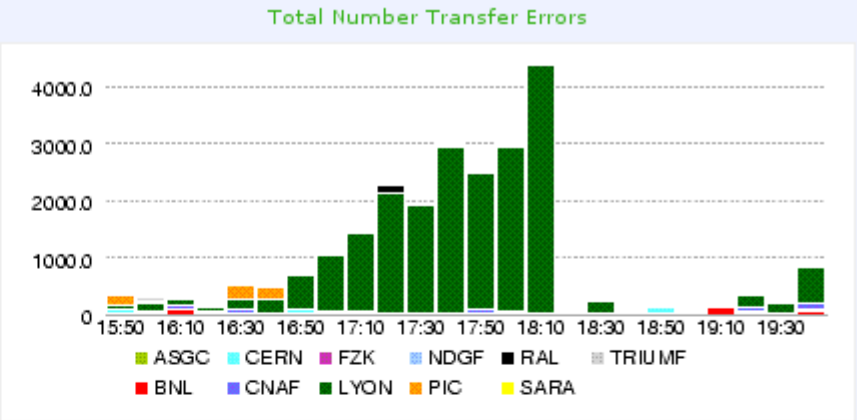
**Sehr gute Performance der GridKa Speichersysteme und des Netzwerks!**

# Netzwerk / Speichersysteme: Performance

Karlsruher Institut für Technologie

## Atlas Dashboard




Activity Summary ('2010-11-10 15:50' to '2010-11-10 19:50' UTC)

Click on the cloud name to view list of sites

| | Transfers | | | Registrations | | Errors | |
|---|---|---|---|---|---|---|---|
| Cloud | Efficiency | Throughput | Successes | Datasets | Files | Transfer | Registration |
| ASGC | 100% | 193 MB/s | 2247 | 62 | 2237 | 7 | 0 |
| BNL | 98% | 752 MB/s | 21978 | 152 | 21950 | 389 | 0 |
| CERN | 92% | 888 MB/s | 4933 | 86 | 4927 | 438 | 0 |
| CNAF | 91% | 848 MB/s | 6440 | 35 | 6445 | 599 | 0 |
| FZK | 100% | 1864 MB/s | 28597 | 220 | 28636 | 137 | 0 |
| LYON | 37% | 411 MB/s | 12476 | 138 | 12460 | 21691 | 0 |
| NDGF | 99% | 236 MB/s | 2046 | 30 | 2047 | 30 | 0 |
| PIC | 94% | 637 MB/s | 11037 | 147 | 11028 | 684 | 0 |
| RAL | 97% | 548 MB/s | 13147 | 249 | 13155 | 410 | 0 |
| SARA | 99% | 744 MB/s | 20120 | 158 | 20108 | 117 | 0 |
| TRIUMF | 79% | 354 MB/s | 3137 | 50 | 3145 | 819 | 0 |

# Grid Sicherheit - Site Security Challenge 4

## SSC4 results: FZK-LCG2 (KIT), Talk 17.09. 9:30

- Communication:
  - Heads-Up to EGI-CSIRT 15 min.
  - Heads-Up to VO-Manager 2h with info: suspicious irc-bot **and** User:CN=Sander Klous (SSC 4)
  - Notification to PJU-CA a bit late.
  - Timestamp of Update used, contained all relevant info.

- Containment:
  - All malicious jobs stopped after 30 min.
  - PJU banned after 30 min. cream-CE missed (took 4h) operational problem, solved already.
  - PJS banned/unbanned in time although PJU already identified within 2h.

**pilot job user (workload)** →

**pilot job submitter** →

- Forensics:
  - All tasks done within 4h + the only team that spotted PJU banning monitor.

Folie aus GDB-Talk von Sven Gabriel/NIKHEF

# Personal

- 3 Stellen für experiment-spezifischen Support besetzt:
  - CMS (1.2.)
  - Atlas (1.7.)
  - Alice (1.10., SCC-interner Wechsel)
- 'Experiment-Vertreter' sind im Admin-Team integriert und übernehmen auch allgemeine Aufgaben bzw. vertreten sich gegenseitig.

- SCC-interner Wechsel (Grid-Service Administration)
- KIT-interner Wechsel (Datenmanagement)
  - Stellen werden so bald wie möglich wiederbesetzt
  - Aufgaben vorübergehend übertragen auf 'Experiment-Vertreter', Storage-Team, weitere Grid-Experten (bisher D-Grid)

# Ressourcen 2011

Karlsruher Institut für Technologie

## LHC-Experimente (GridKa pledged):

| CMS | 2010 | 2011 |
|---|---|---|
| CPU | 10050 | 15000 |
| Disk | 1474 | 1950 |
| Tape | 3029 | 5700 |

| Atlas | 2010 | 2011 |
|---|---|---|
| CPU | 21600 (+4168 non-pledge) | 28250 |
| Disk | 2190 | 3125 |
| Tape | 1420 | 3750 |

| Alice | 2010 | 2011 |
|---|---|---|
| CPU | 19600 | 29250 |
| Disk | 2700 | 2130 |
| Tape | 4075 | 3250 |

| LHCb | 2010 | 2011 |
|---|---|---|
| CPU | 7480 | 11050 |
| Disk | 560 | 600 |
| Tape | 408 | 590 |

CPU: [HEPSPEC'06]
Disk, Tape: [TB]

# Ressourcen Upgrade

- 10 zusätzliche LTO4-Laufwerke in neueste Library (STK) eingebaut.
  - Wenig zuverlässige GRAU-Lib wird mittelfristig nur noch für Backup-Zwecke genutzt.

- Plattensysteme und Fileserver sind weitgehend angeliefert.
  - gleiche Systeme wie 2010 (DDN/Dell)

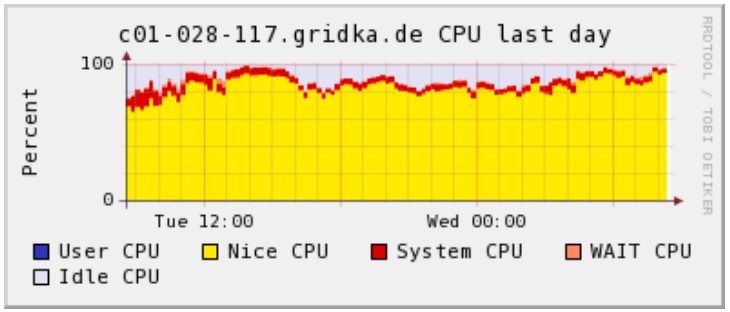
- Compute-Nodes erwartet vor Weihnachten.

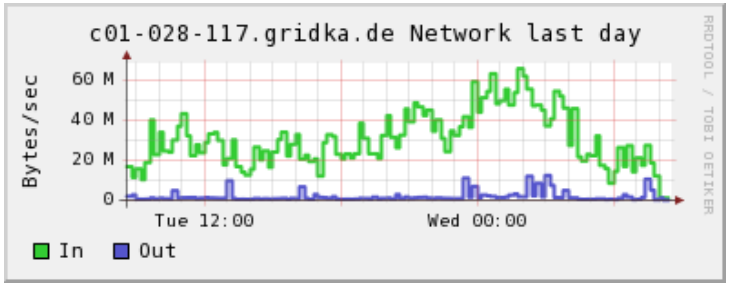
  29 Boxen, 2 HE, je 4 Nodes:
  - 2 x AMD 6168 12-core CPUs
  - 3 GB RAM per core = 72 GB per node
  - 3 x 500 GB SATA disk

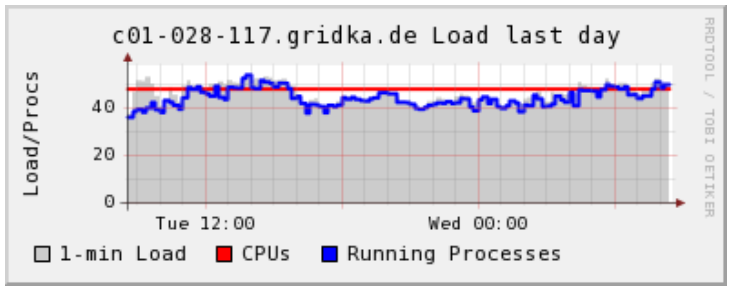
# Test-Workernode

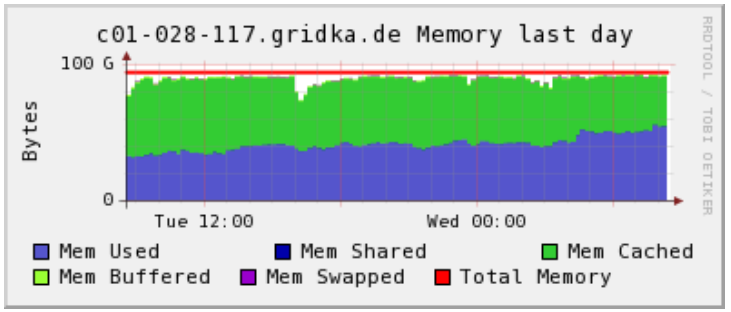
- **DELL R815 test system currently running as WN in production**
  - AMD RD890 chip set
  - **4** x AMD 6174 CPUs = 4 x 12 cores @ 2.2 GHz = **48 cores**
  - 96 GB RAM
  - 6 x 2.5” 500 GB Near Line SAS (7.2k) disks

- **No bottlenecks seen so far...**



# Aktuelle Diskussionen

- Datenmanagement-Modelle (T1 ↔ T2)
- 3D Datenbanken oder Proxy/Squid?
  - Atlas,LHCb Conditions-DBs
- Multithreaded Jobs / Reservierung ganzer Workernodes
- LHC-Pause 2012?
- Ressourcenplanung 2012