CMS Status Report
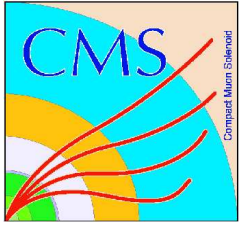
PCP for CMS DC04 at FZK
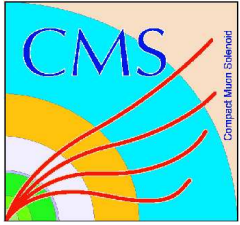
Beware of acronyms!

The German CMS Production Team:

Klaus Rabbertz, Jens Rehn, Serge Sushkov, Joanna Weng

Data Challenge 04

* Start: January 2004

* 5% test of expected computing requirements for CMS running in 2007

  1) Data distribution and access

  2) Data calibration

  3) Distributed user analysis jobs

     ➡ Jobs on data on multiple Tiers

Pre-Challenge Production
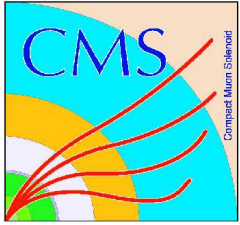
- Simulation of events necessary for DC04

- Anticipated: 50 M events until December

- Requested from physics working groups:
  69 M events

- German participation: 10 %

    - 5 M events for GridKa

- Current total rate: 12 M / month

- But: No pile-up yet, only „old" Geant3 simulation
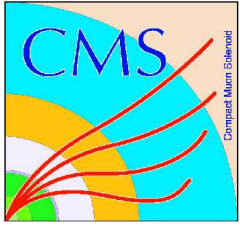
Assignment Scheme

* **RefDB: Central database at CERN**

  * Distribution and tracking of MC production (via e-mail)

* **CMKIN, CMSIM, OSCAR:**

  * CMS simulation software (n-tuples, Geant3, Geant4)

* **McRunJob:**

  * Local tracking and job submission (Python scripts)

* **SRB: Storage Resource Broker:**

  * Data and software distribution

Last three months ...

- ⁕ 23.06. GridKa validated for CMKIN production

- ⁕ 01.07. Start of Pre-Challenge Production

- ⁕ 02.07. First CMKIN assignment

- ⁕ 18.07. GridKa validated for CMSIM production

- ⁕ 18.07. First CMSIM assignment

- ⁕ Today: 26th assignment running

  - ➡ 2.6 M events generated (CMKIN), 3.8 % over all Tiers
  - ➡ 3.2 M events simulated (CMSIM), 9.6 % over all Tiers
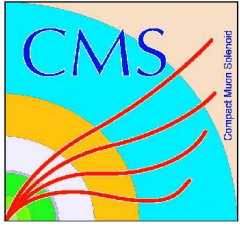
Problems with ... fileservers

* CMS: fileservers cms1 – cms3

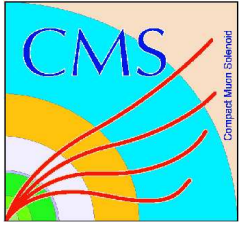   1) cms1: 1.8 TB, used for start of production, ok

   2) cms2: 1.5 TB, very problematic, last incident: 1. October

   → used for checks only

   3) cms3: 6.5 TB, available since 13.07., after small start-up problems runs very stable

   → Migrated production to cms3 on 28.08.

* Home fileserver sometimes not working

Problems with ... mails

* **CMS uses mails from compute nodes for production surveillance**

    1) Sending mails from within NAT ok, but

        * Rejected as SPAM at CERN

    2) Set-up change at GridKa requested: June 2003

        * Completion announced: 08.09.2003
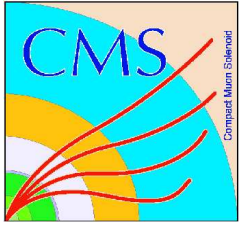   * Really finished: 26.09.2003

Problems with ... firewall

* CMS employs SRB for file import/export

  1) Need one primary port: 5543

  2) Need range for multithreading: 20000 - 20200

     → Blocked by firewall despite our request because ...

     → CMS installation recipe was wrong:

        • Primary port 5544

        • Range 20000 - 20200 is default ... it's not!
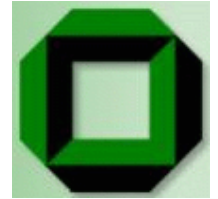
Problems with ... /tmp

* Maximal Walltime: 96 h

* /tmp clean-up after: 240 h

* Left-over files:

  1) Fill up /tmp (e.g. CDF: 15 GB from crashed job)

     → Such a „free" node may empty the queues

  2) Confuse jobs (e.g. CMS: diff. jobs read from same directory /tmp/job.pid by accident)

     → Wrong output copied back

     → CMS clean-up will be improved
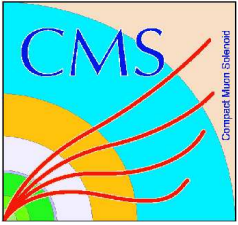
Problems with ... I/O

* Very bad Walltime to CPUtime ratio!

* Currently only jobs w/o pile-up for CMS

* I/O will be a major issue for user analyses

~1h/50h
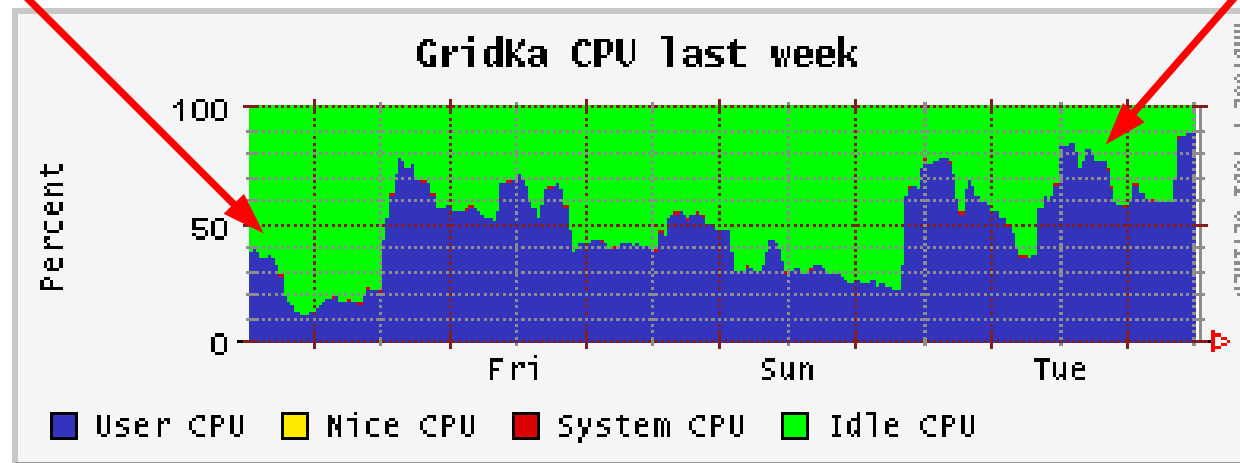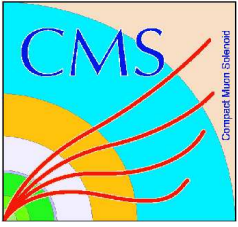
Problems with ... I/O

⁕ ## Too many I/O intensive jobs block each other

➡ occupy slots in queues,

➡ spoil GridKa performance and

➡ prevent the other experiments to run their jobs
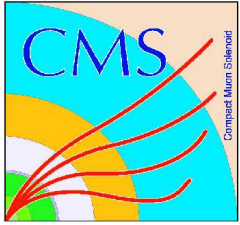
Impact of I/O int. jobs from BaBar

Similar for D0 this Tuesday

GridKa CPU last week

Other Problems
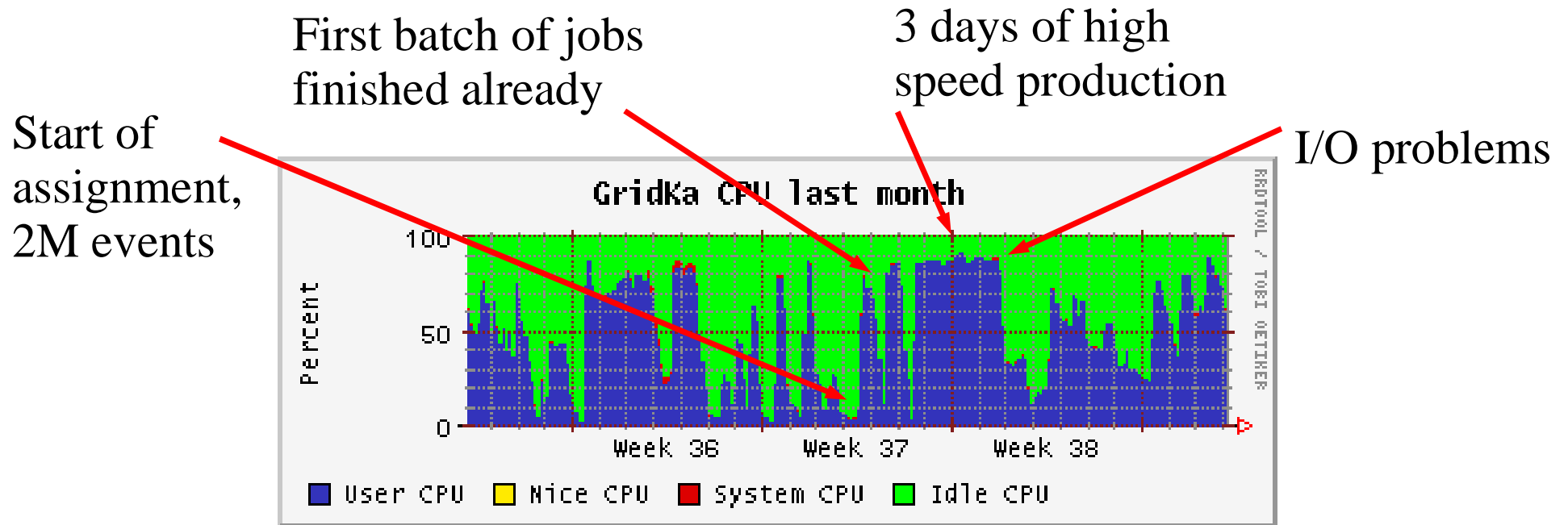
* CMKIN jobs finish too fast (~5 sec.)

  → NFS overload => increased submission delay (10 s)

* Limits for max. running jobs per queue, determined automatically?

  → Sometimes prevents jobs from running

* sshd blocks login, restart necessary

* Grid User Support, portal there, but is it in use?

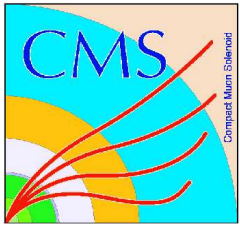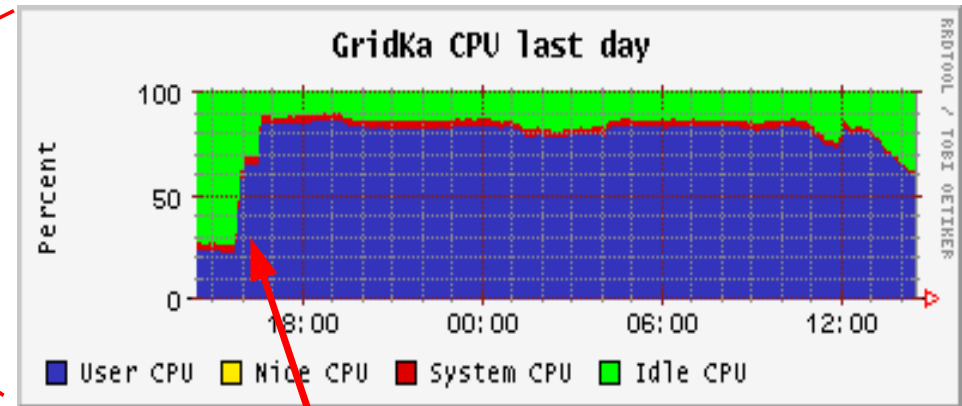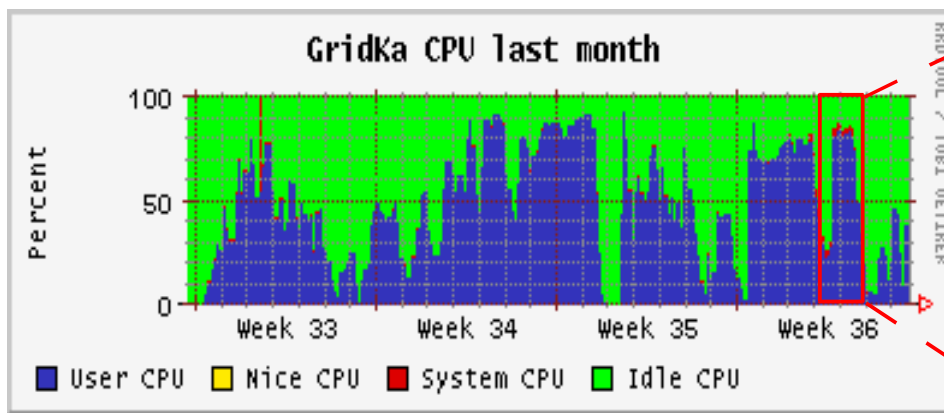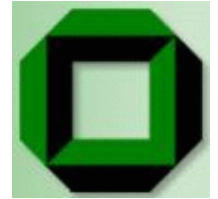 → Got automatic acknowledge for request, nothing more yet

Good News

* Very nice Ganglia interface to monitor GridKa

* It is possible to run > 390 CMSIM jobs in par.

First batch of jobs
finished already

3 days of high
speed production
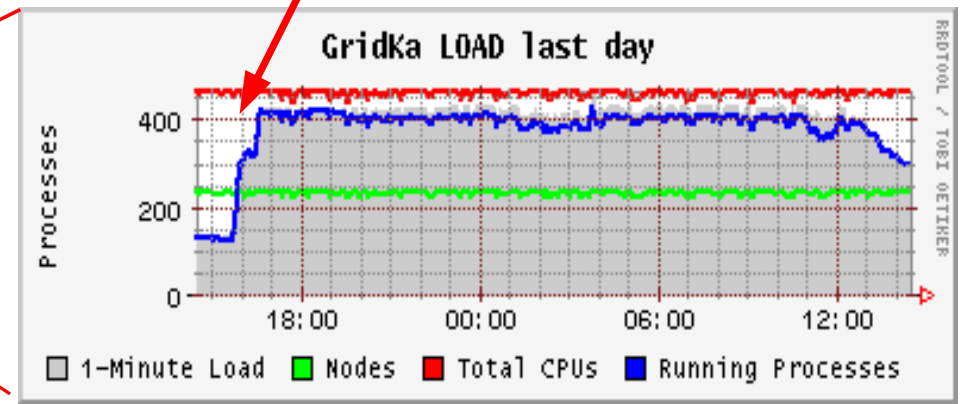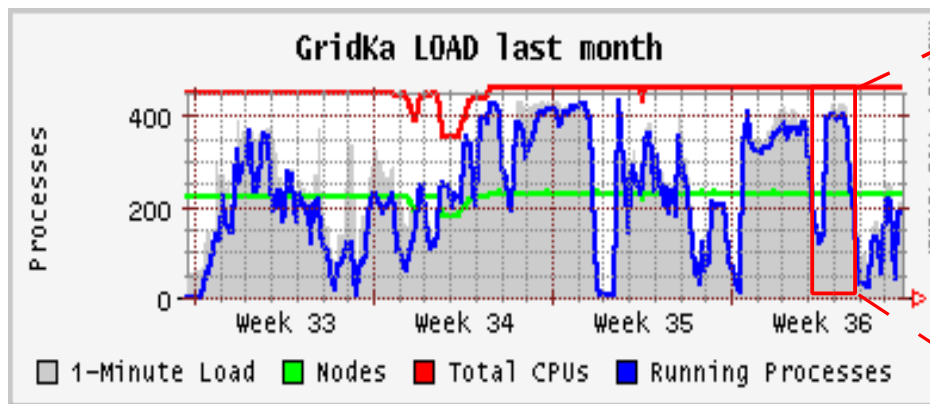
Start of
assignment,
2M events

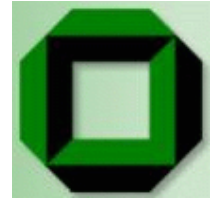I/O problems

Example of impact of CMS production on GridKa

We (CMS) could run more assignments

Impact of CMS assignment

CPU Usage Jul -> Sep

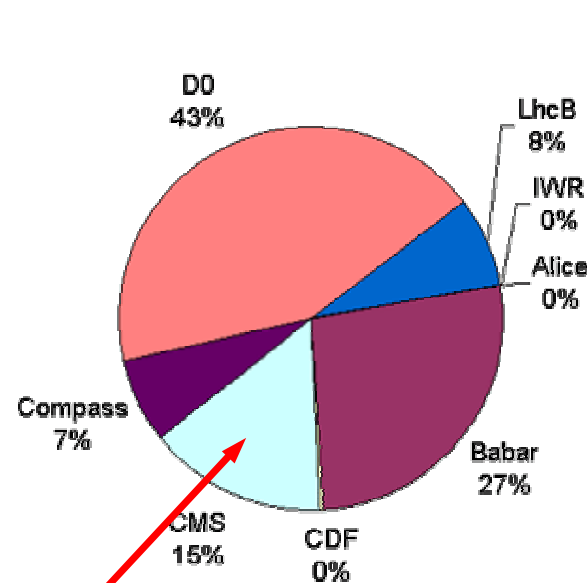### July

CPU-Time insgesamt: 163 248,31 Stunden

- LhcB 13%
- IWR 0%
- Alice 1%
- Atlas 2%
- Babar 11%
- CDF 0%
- CMS 3%
- Compass 2%
- D0 68%
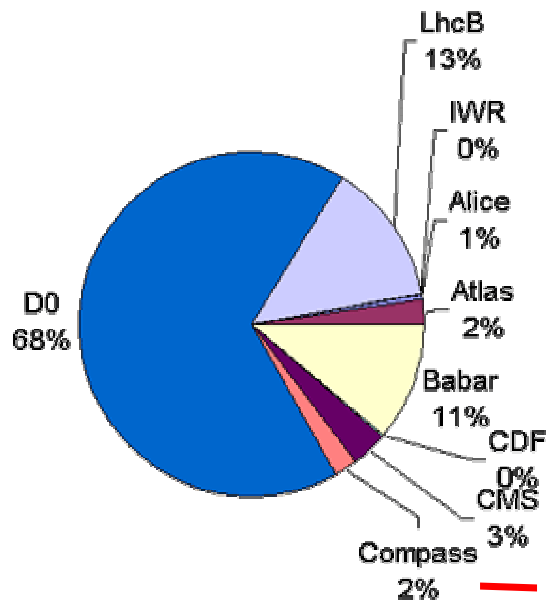
August

CPU-Time insgesamte CPU-Zeit: 124 951,23 Stunden

- D0 43%
- LhcB 8%
- IWR 0%
- Alice 0%
- Babar 27%
- CDF 0%
- CMS 15%
- Compass 7%

September

CPU-Time insgesamte CPU-Zeit: 172 431,53 Stunden

- Compass 3%
- D0 29%
- LhcB 4%
- IWR 0%
- Alice 0%
- Atlas 1%
- Babar 15%
- CDF 1%
- CMS 47%

Color coding changed!?
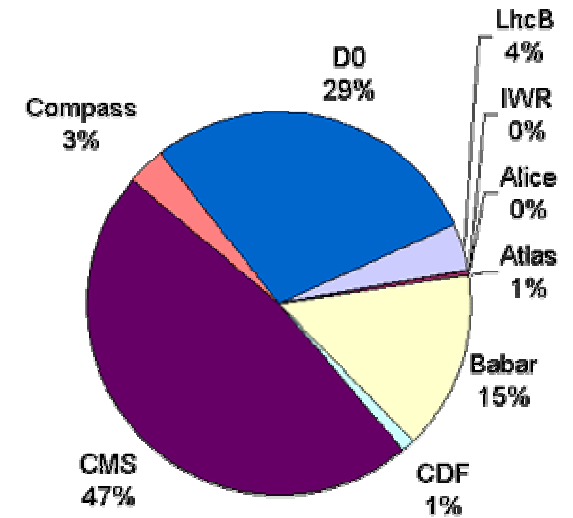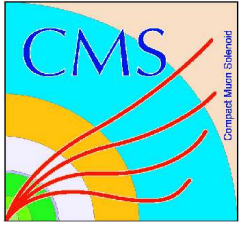
Documentation Requests

- Web archive for rdccg-user mailing list
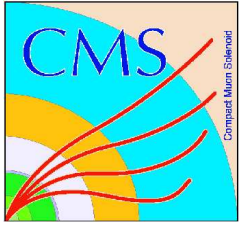 - New web pages announce GridKa-user mailing list ... which doesn't exist, yet?
- List of user accounts and corresponding e-mail addresses
  - No cat /etc/passwd plus Google search necessary
- Documentation of fileserver sharing between the experiments
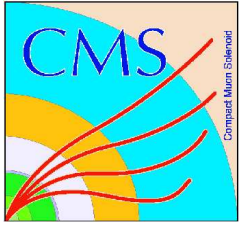- Documentation of fairshare settings/points

Other Requests

* More efficient handling of jobs with extremely bad Walltimes/CPUtime ratios necessary

* Better timing of announcements for non-emergency interventions

  → Not e.g. for tapes: Announced: 23.09. 23:23
    Scheduled: 24.09.

* One hour autologoff of ssh sessions is quite annoying from outside Karlsruhe

Summary

- Encountered many problems and resolved quite some fraction of them.

- It's a stony path to go.

- Nevertheless we're taking up speed.

Many thanks to the GridKa people, special thanks go to Manfred Alef.