Computing Report

ATLAS Bern

Christian Haeberli

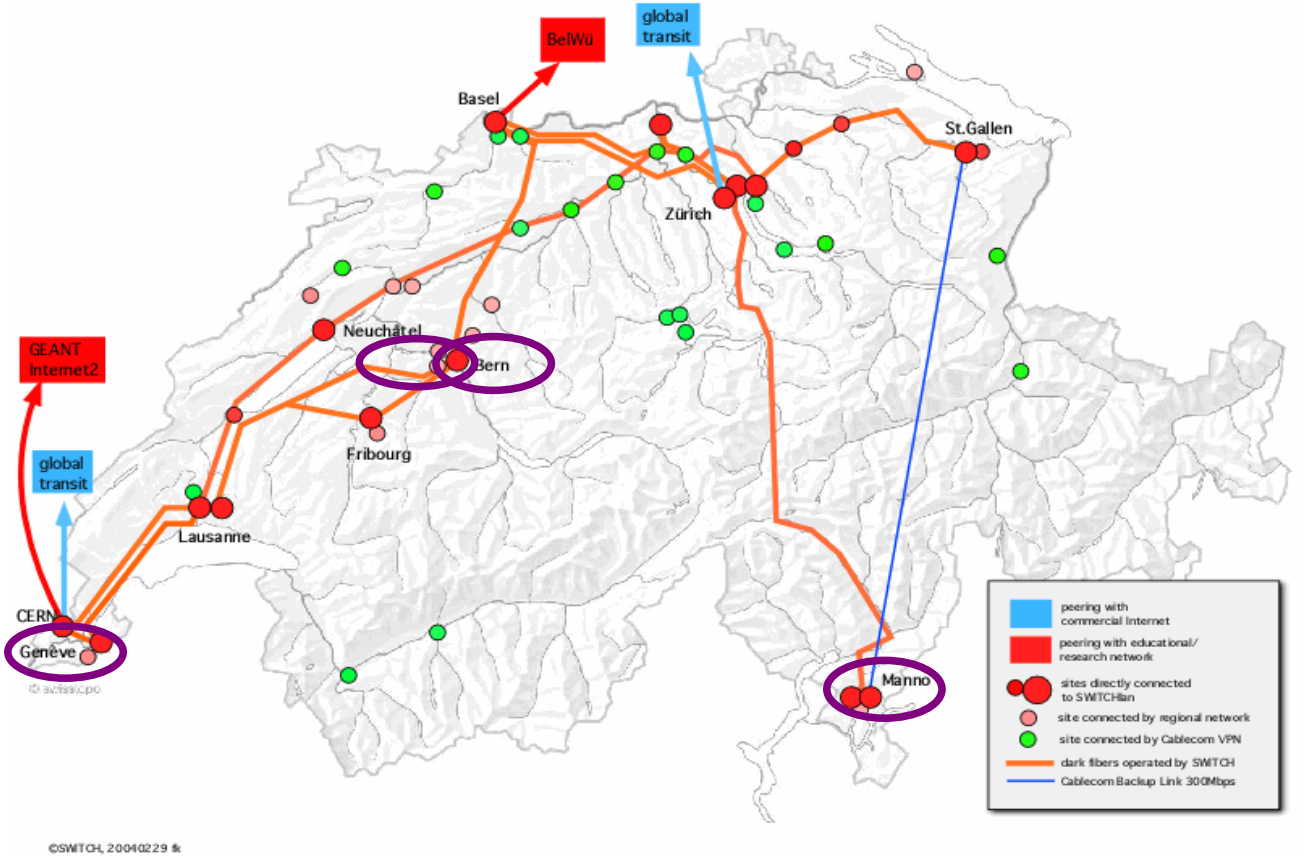
The Bern ATLAS Cluster

- In November 2003 we started to build a small Linux cluster:
  4 worker nodes (8 CPUs) and a 0.5 TB RAID storage
- The ATLAS Software was successfully deployed on the cluster
- In summer 2004 we wanted to bring in our modest resources into the ATLAS DataChallenge 2 effort. There was no other Swiss contribution at that time.
  - We needed to integrate the cluster in one of the three Grid flavours supported by the ATLAS production system: LCG, NorduGrid, Grid3
  - A short evaluation resulted in favour of NorduGrid:
    - Our cluster was to small to be integrated in LCG
    - LCG was too complex to be maintained by a small university group
    - Grid3 was not production ready at that time
  - NorduGrid was installed in July
- The Bern ATLAS Cluster is available for the ATLAS production system since. DC 2 and Rome production were successfully run during the last 13 months.

Swiss ATLAS Grid

- After having seen how smoothly NorduGrid was running for DC 2 production, we (together with Geneva) wanted to use NorduGrid for our own purposes.
- Build a NorduGrid based Swiss ATLAS Computing Grid, consisting of:
  - PHOENIX at CSCS
  - ATLAS Cluster at the DPNC Geneva
  - ATLAS Cluster at the LHEP Bern
  - UBELIX Cluster at the University of Bern
    UBELIX is a common Linux Cluster to the whole university with ~100 CPUs (about to triple its size this autumn)
- The Swiss ATLAS Grid unifies ~140 CPUs to a country-wide batch facility

Swiss ATLAS Grid



Swiss ATLAS Grid

- This country-wide batch facility is successfully used for:
  - ATLAS SUSY simulation
  - ATLAS e-gamma trigger reconstruction
  - Reconstruction of CTB data
  - the Bern cluster is still available for central ATLAS production, but *Swiss* jobs have priority
  - five happy users so far...

Why NorduGrid?

- For us NorduGrid was the choice, because
  - LCG can only be used on dedicated cluster and not on shared clusters like UBELIX, due to requirements on OS, Scheduler, Cluster Management...
  - NorduGrid is designed to be plugged on already existing clusters
  - NorduGrid can in contrary to LCG be sensibly deployed on small clusters
  - NorduGrid is easy to install and configure
  - Everybody can install the NorduGrid user interface on his laptop (even on his AFS area)
  - → We could not have built the Swiss ATLAS Grid with LCG

Use case SUSY

SUSY Study by Eric Thomas (4 points SUx)

Numbers:

- Event Generation: 360’000 events
- G4 Simulation: 136’500 events
- Digitization: 136’500 events
- Reconstruction: 155’850 events
- Data Volume: 750 GB
- Number of successful jobs: 5428
- Success rate: 75%

Frequent Failures

- Oracle DB cluster at CERN, which is holding the ATLAS geometry DB, was overloaded and refusing connections
- Cluster internal problems (e.g. NFS)
- ATLAS software: 5% of the reconstruction jobs crashed
- ATLAS software distribution: inconsistencies of the distribution kits
- Problems with file handling (e.g. stage-in from CASTOR at CERN)
- User mistakes (expired grid proxies, invalid destinations for output files, invalid job descriptions)

User Support

- A new production *always* starts with a failure rate of 100%, mostly due to user mistakes
- Many things can go wrong: the experiment software is complex and not always stable, the jobs are running remote, files are stored remote, users are not familiar with the Grid tools…
- Therefore user support is crucial during a production period, especially in the early phase:
  - Help the user with the job description
  - Get error reports and try to trace the problem
  - Monitor the production (7/7) to see if something goes/will go wrong
  - Patch software releases
  - Setup backup services (e.g. DB replica) in case CERN services are overloaded
  - Replicate data from high to low latency storage (from castor to a RAID array)
- If there is not enough support or if the support is not coming fast enough, people will move back to lxplus/lxshare

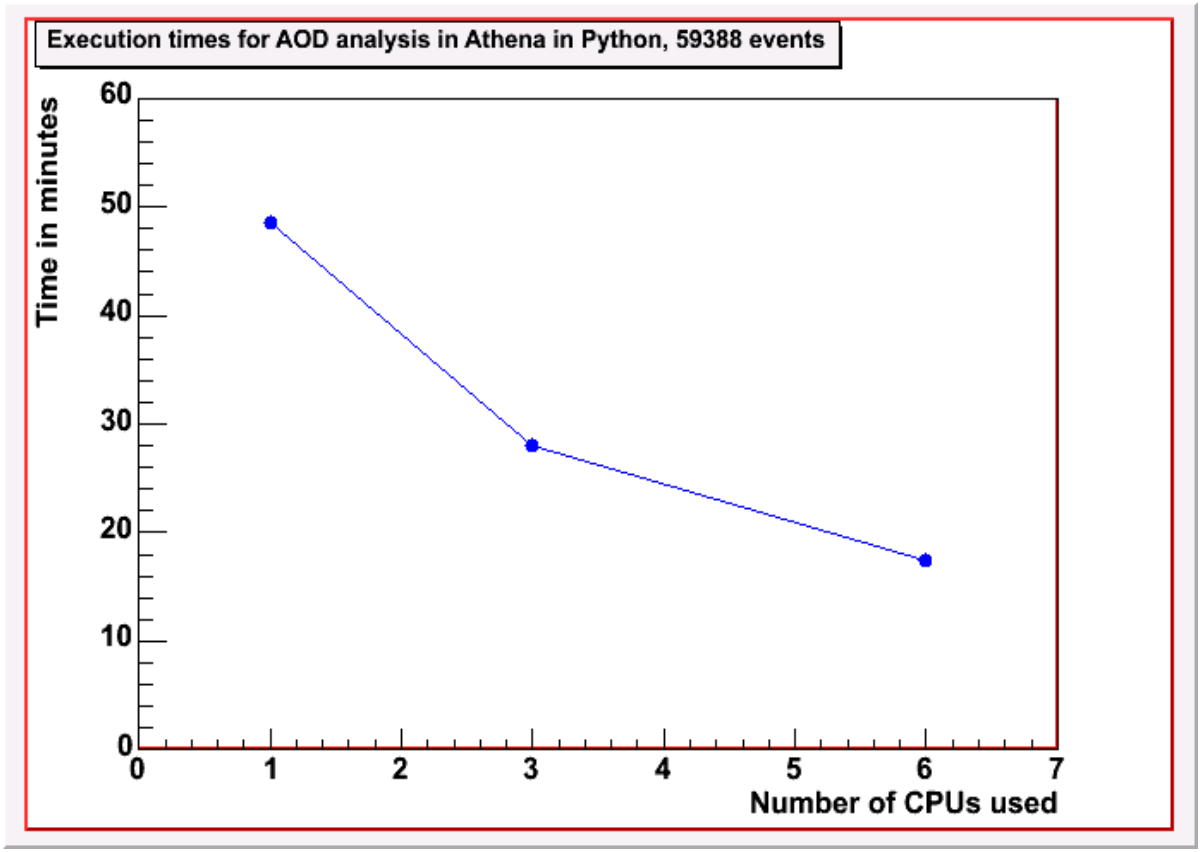
Interactive Analysis

- Infrastructure for interactive analysis in ATLAS is in bad shape. There is no coherent approach:
  - DIAL project pushed by the BNL group
    Disadvantage: Extremely complex, difficult to deploy and operate, does not address the use case of distributing jobs on a local cluster, mixing batch and interactivity
  - ATLAS Production system is more and more propagated as user tool for user analysis
    Disadvantage: Extremely complex, many components between the user and the running job, no concept of a user, no interactivity
  - DIANE project put up as back-up by some people
    Simple, successfully deployed in Bern, restricted to a local cluster, pseudo-interactive

Interactive Analysis

- Conclusion: we are on our own, nobody will provide us with a solution
- Strategy:
  - Distinguish two use cases: batch and interactive
  - Use a grid middleware for batch (today NorduGrid)
  - Restrict the interactive use case to a local cluster. This allows us to opt for a simple solution like DIANE

Run the analysis in parallel on local cluster with DIANE



- all software and all data on shared file system
- DIANE splits input by files
- distributes processing
- collects output
- merges .root files in the output (histograms are added)

Resources: Strategy in Bern

- Cover all the user needs for computing at the TIER-3 level, harvest available cycles at TIER-2 in low production seasons (TIER-2 is a facility for centrally managed ATLAS production)
- User storage needs:
  - Cover most needs at TIER-3: 2 TB/(user*year)
  - In the ATLAS Computing Model 1/3 of all AOD is stored at every TIER-2 ***in accordance with local interest***. We must enforce this policy for Manno to use our storage efficiently. If this is not possible for technical or political reasons we must invest in additional storage.
  - Profit from non-used storage at TIER-2
  - Profit from other TIER-2: there are 30 TIER-2, so there should be 10 replica of every AOD dataset available
  - Not rely on CERN storage
- Use a common university cluster for batch processing (mostly GEANT4 Simulation)
- Use a small institute owned cluster for interactive analysis

→ No requirements on TIER-2 additional to the ATLAS Computing Model

ATLAS Resources Bern

10 active physicists in ATLAS data analysis

CPU: 15 kSI2k per user (~10 modern CPUs) in 2008, assume moderate scale up by replacing old hardware with faster equipment

Storage: 2 TB per user per year

	2007	2008	2009	2010	2011	2012
CPU T3 kSI2k	66	**150**	230	300	400	460
Disk T3 TB	20	**40**	60	80	100	120

CPU Power

- The computing model at some point told us how much CPU power/user we need in 2008: 15 kSI2k
- 13.5 kSI2k of the 15 kSI2k are for Simulation, because GEANT4 simulation is very resource hungry compared to Digitization, Reconstruction and Analysis
- 15 kSI2k corresponds to ~ 10 today's CPUs
- On this 10 CPUs we can simulate and digitize ~ 1000 SUSY events/day or reconstruct 30'000 SUSY events/day
- For data analysis 1.5 kSI2k is foreseen per user. That corresponds to one powerful desktop machine. We can analyze 2 Mio AOD events/day on such a machine
- Memory budget: max 1 GB/job

Summary ATLAS Bern

- Middleware batch: NorduGrid
- Middleware interactive: DIANE
- Resources at TIER-2: no additional resources on top of the ATLAS computing model requirements
- Users: 10
- Resources at TIER-3 in 2008:
  - 150 kSI2 (100 CPUs)
    15 kSI2k/user
  - 40 TB
    2 TB/(user*year) starting in 2007

Backup

Storage

- The Computing Model foresees 1 TB / user
- To compare with other experiments:
  - CMS 3.5 TB / user
  - LHCb 1.4 TB / user
- The 1 TB / user certainly do not reflect our wish to mirror our data in Manno
- 1 TB corresponds to
  - 650’000 RAW Events
  - 2’000’000 ESD Events
  - 10’000’000 AOD Events