ATLAS grid experience

Marc Goulette (University of Geneva)

Introduction

ATLAS has already recorded ~ 3.46 pb^{-1} of data (31/08/10).

Therefore, in order to extract physics from the data:

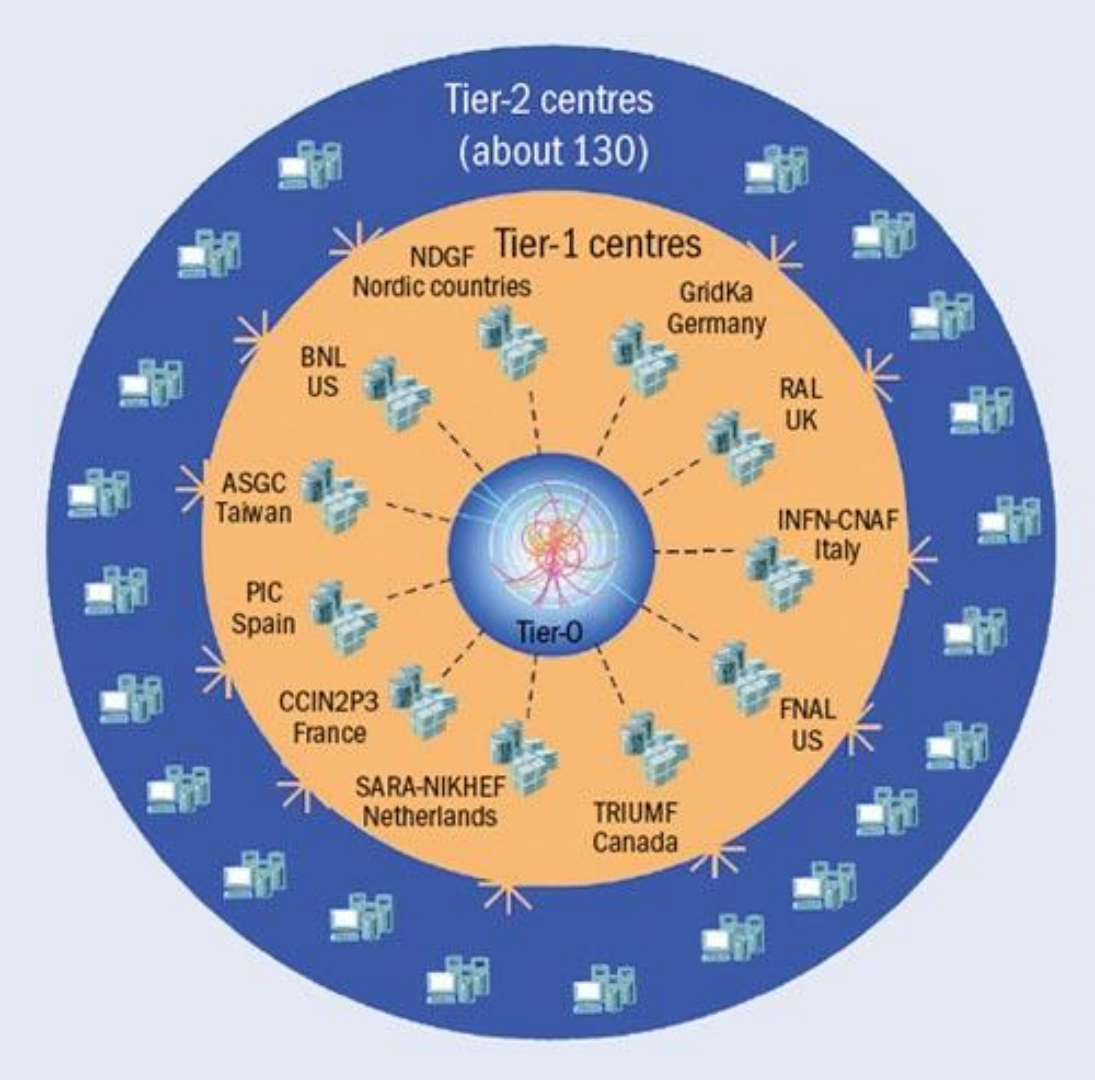
ATLAS needs a well defined and efficient computing model, which is strongly based on the LHC Computing Grid.

The Grid is a distributed system. Its role is to build and maintain a data storage and analysis infrastructure for the entire HEP community, including the LHC and ATLAS.

The computing and the analysis models are used together to extract physics.

ATLAS computing model

LHC Computing Grid



Tier-0
(CERN)

- Data recording
- First-pass reconstruction
- Data distribution

Tier-1
(11 centres)

- Calibration
- Re-processing
- Data selection
- Data distribution

Tier-2
(>100 centres)

- Simulation (MC)
- Data selection
- Data distribution
- Analysis

Tier-3
(Universities & Institutes)

- Data selection
- Analysis

ATLAS resources in 2010

	CPU (MSi2k)	Disk (PB)	Tape (PB)
Tier-0	6.1	0.5	11.4
CERN (CAF)	4.6	2.8	1.0
Sum of Tier-1s	50	40	28.7
Sum of Tier-2s	51.5	22.1	no
Total	112.2	65.4	41.1

ATLAS Swiss activities in the computing model

- No Tier-1 in Switzerland, but associated with the FZK T1 in Karlsruhe (in the DE cloud).
- One Tier-2: CSCS (Manno).
- Two Tier-3: Bern, Geneva.

(From Aug. 2010)

	CPU cores (kSi2k)	Disk (TB)
CSCS T2 (Atlas)	~290 cores	300
CSCS T2 (Total)	768 cores (2.5 kSI2k)	800
T3 Bern (12 users)	200+300 shared (~600 kSI2k)	100
T3 Geneva (~☐60 users)	268 cores (462 kSI2k)	177

- 220 TB free for Atlas at CSCS (31/08/10).

- Disk for CSCS T2 should be 364 by the end of 2010

- 5 M SpecInt2000 is equivalent to 5000 PC at 3 GHz.)

- The Nominal Wide Area Network (WAN) in Mbits/sec is 10000.

ATLAS analysis Model

For the data available on the different sites of the grid, we then have:

Different streams (based on the Trigger used):

Express, Calibration, Minbias, L1Calo, MuonswBeam, Random, Cosmics, BPTX…

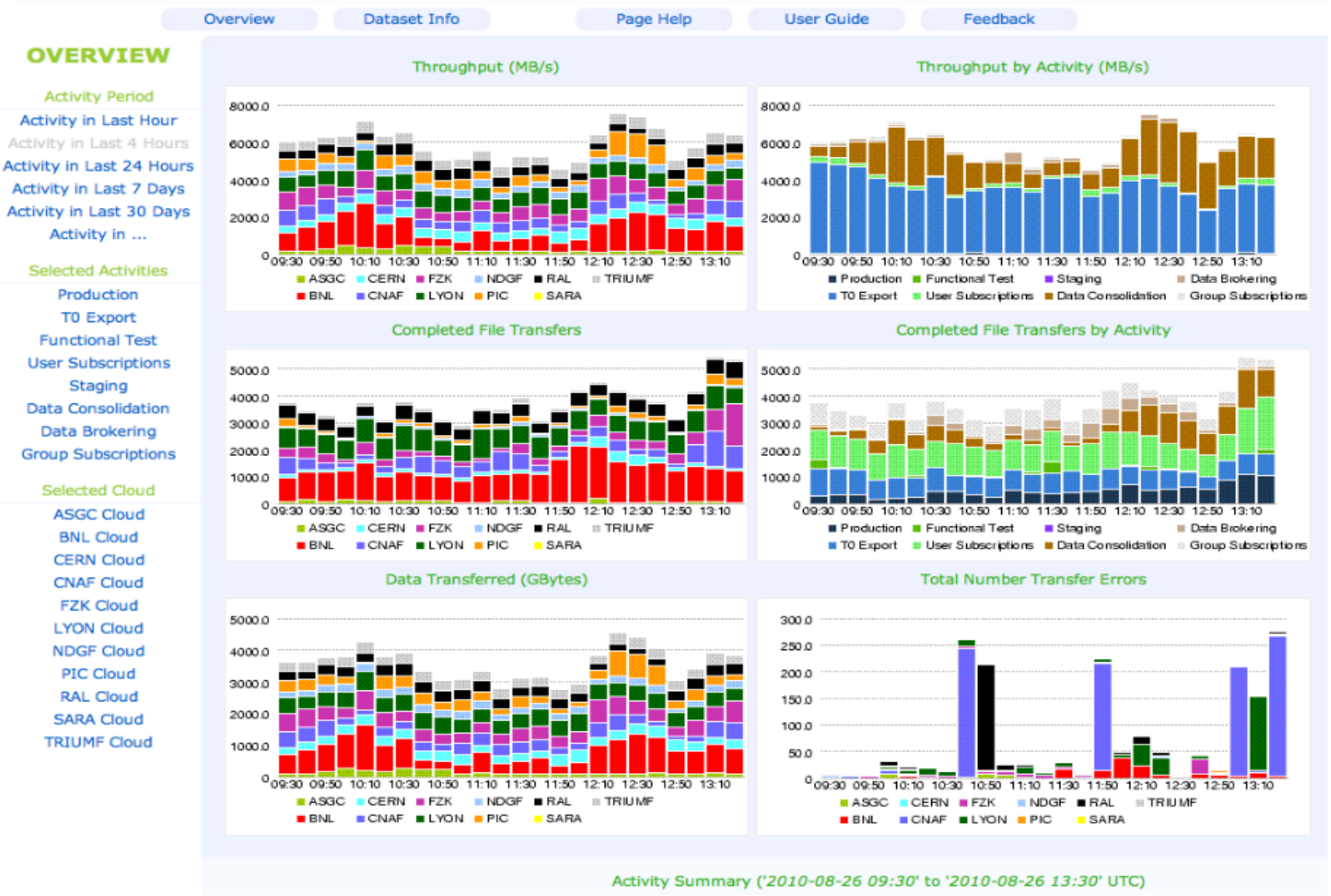
Different data formats (Event Data Model):

RAW, ESD, AOD, TAG, DPD, NTUP, …

Thus, different studies can be done, and will use the analysis model at different stages:

Most studies are using the main Framework (Athena), the EDM, some Distributed Analysis Tools (panda, ganga), DDM tools (DQ2), and grid backends for direct submission to the grid (LCG, Nordugrid). At the end, most analysis are using ROOT to produce the final plots.

GRID Monitoring



The software is centrally monitored within clouds and at sites

Overall picture using a concrete example (1)

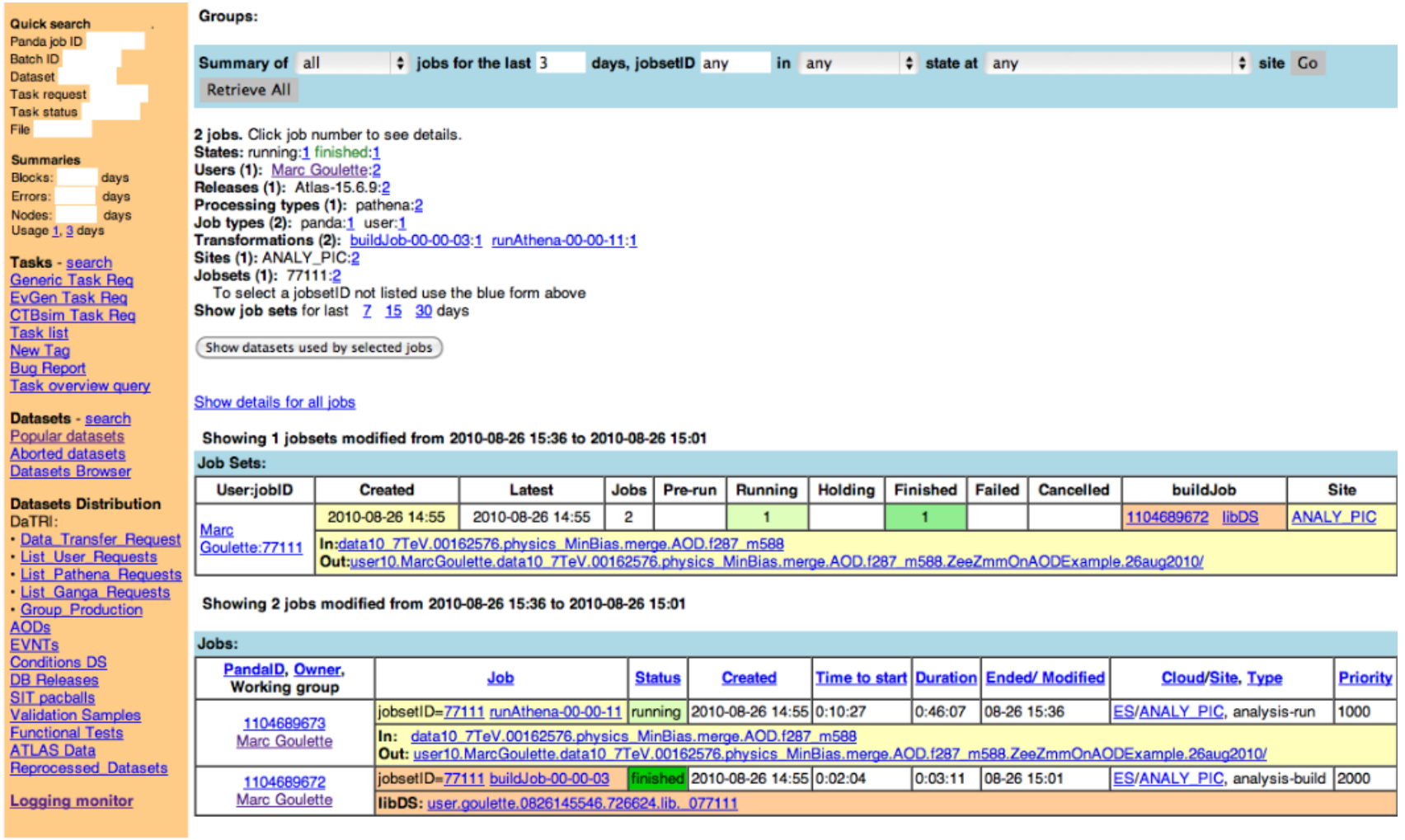
- You connect to a server at your institute or lab.
- You setup an Athena Release (software area…)
- You initiate the grid, Panda, DQ2
- You have prepared your analysis code (compiled…)
- Then you can send your job to the grid:

```
Pathena ZeeZmmOnAODExample_jobOption.py
--i□nDS data10_7TeV.00162576.physics_MinBias.merge.AOD.f287_m588
--outDS user10.MarcGoulette.data10…._m588.ZeeZmmOnAODExample.26aug2010

INFO : extracting run configuration
INFO : ConfigExtractor > Input=POOL
…
INFO : query files in data10_7TeV.00162576.physics_MinBias.merge.AOD.f287_m588
INFO : use 4 files
INFO : submit to ANALY_PIC
…
  > build
    PandaID=1104689672
  > run
    PandaID=1104689673
```

Overall picture using a concrete example (2)

- You can check the status of your job:



Quick search
Panda job ID
Batch ID
Dataset
Task request
Task status
File

Summaries
Blocks: days
Errors: days
Nodes: days
Usage 1, 3 days

Tasks - search
Generic Task Req
EvGen Task Req
CTBsim Task Req
Task list
New Tag
Bug Report
Task overview query

Datasets - search
Popular datasets
Aborted datasets
Datasets Browser

Datasets Distribution
DaTRI:
- Data_Transfer_Request
- List_User_Requests
- List_Pathena_Requests
- List_Ganga_Requests
- Group_Production

AODs
EVNTs
Conditions DS
DB Releases
SIT pacballs
Validation Samples
Functional Tests
ATLAS Data
Reprocessed_Datasets

Logging monitor

Groups:

Summary of all **jobs for the last** 3 **days, jobsetID** any **in** any **state at** any **site** Go
Retrieve All

2 jobs. Click job number to see details.
States: running:1 finished:1
Users (1): Marc Goulette:2
Releases (1): Atlas-15.6.9:2
Processing types (1): pathena:2
Job types (2): panda:1 user:1
Transformations (2): buildJob-00-00-03:1 runAthena-00-00-11:1
Sites (1): ANALY_PIC:2
Jobsets (1): 77111:2
To select a jobsetID not listed use the blue form above
Show job sets for last 7 15 30 days

Show datasets used by selected jobs

Show details for all jobs

Showing 1 jobsets modified from 2010-08-26 15:36 to 2010-08-26 15:01

Job Sets:

User:jobID	Created	Latest	Jobs	Pre-run	Running	Holding	Finished	Failed	Cancelled	buildJob	Site
Marc Goulette:77111	2010-08-26 14:55	2010-08-26 14:55	2		1		1			1104689672 libDS	ANALY_PIC
	In:data10_7TeV.00162576.physics_MinBias.merge.AOD.f287_m588 Out:user10.MarcGoulette.data10_7TeV.00162576.physics_MinBias.merge.AOD.f287_m588.ZeeZmmOnAODExample.26aug2010/										

Showing 2 jobs modified from 2010-08-26 15:36 to 2010-08-26 15:01

Jobs:

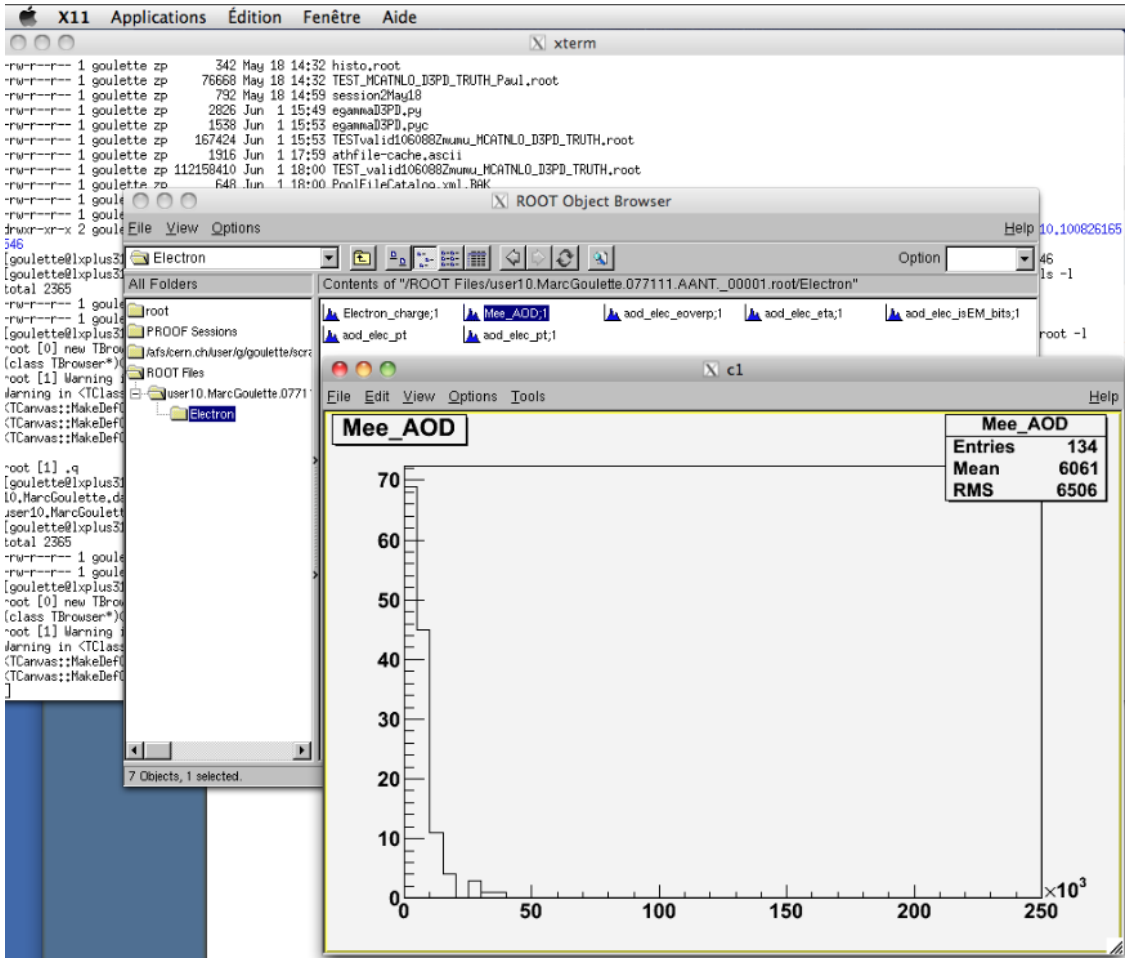
PandaID, Owner, Working group	Job	Status	Created	Time to start	Duration	Ended/ Modified	Cloud/Site, Type	Priority
1104689673 Marc Goulette	jobsetID=77111 runAthena-00-00-11	running	2010-08-26 14:55	0:10:27	0:46:07	08-26 15:36	ES/ANALY_PIC, analysis-run	1000
	In: data10_7TeV.00162576.physics_MinBias.merge.AOD.f287_m588 Out: user10.MarcGoulette.data10_7TeV.00162576.physics_MinBias.merge.AOD.f287_m588.ZeeZmmOnAODExample.26aug2010/							
1104689672 Marc Goulette	jobsetID=77111 buildJob-00-00-03	finished	2010-08-26 14:55	0:02:04	0:03:11	08-26 15:01	ES/ANALY_PIC, analysis-build	2000
	libDS: user.goulette.0826145546.726624.lib._077111							

Overall picture using a concrete example (3)

- As soon as the job is finished, you can retrieve the result using DQ2

dq2-get user10.MarcGoulette.data10…._m588.ZeeZmmOnAODExample.26aug2010

- Then, you can use ROOT do produce the final plots:



Advantages and Weak points

+ Parallel access
+ Redundancy
+ Speed
+ Efficiency
+ Reliability
+ Traceability
+ Storage features
+ Multi-task

- Complexity
- Maintenance
- Cost
- Rather new (needs maturity)
- Quickly changing (sw updates)
- Sites related issues

→ But the main point is that it helps us
to go with our complex analysis model, and thus we
should continue to learn how to use (improve) it

Summary and Conclusion

- A lot of data has been taken already, so the whole analysis chain needed to be in place beforehand.

- The analysis model is based on an important computing model.

- The computing model is strongly based on the Grid.

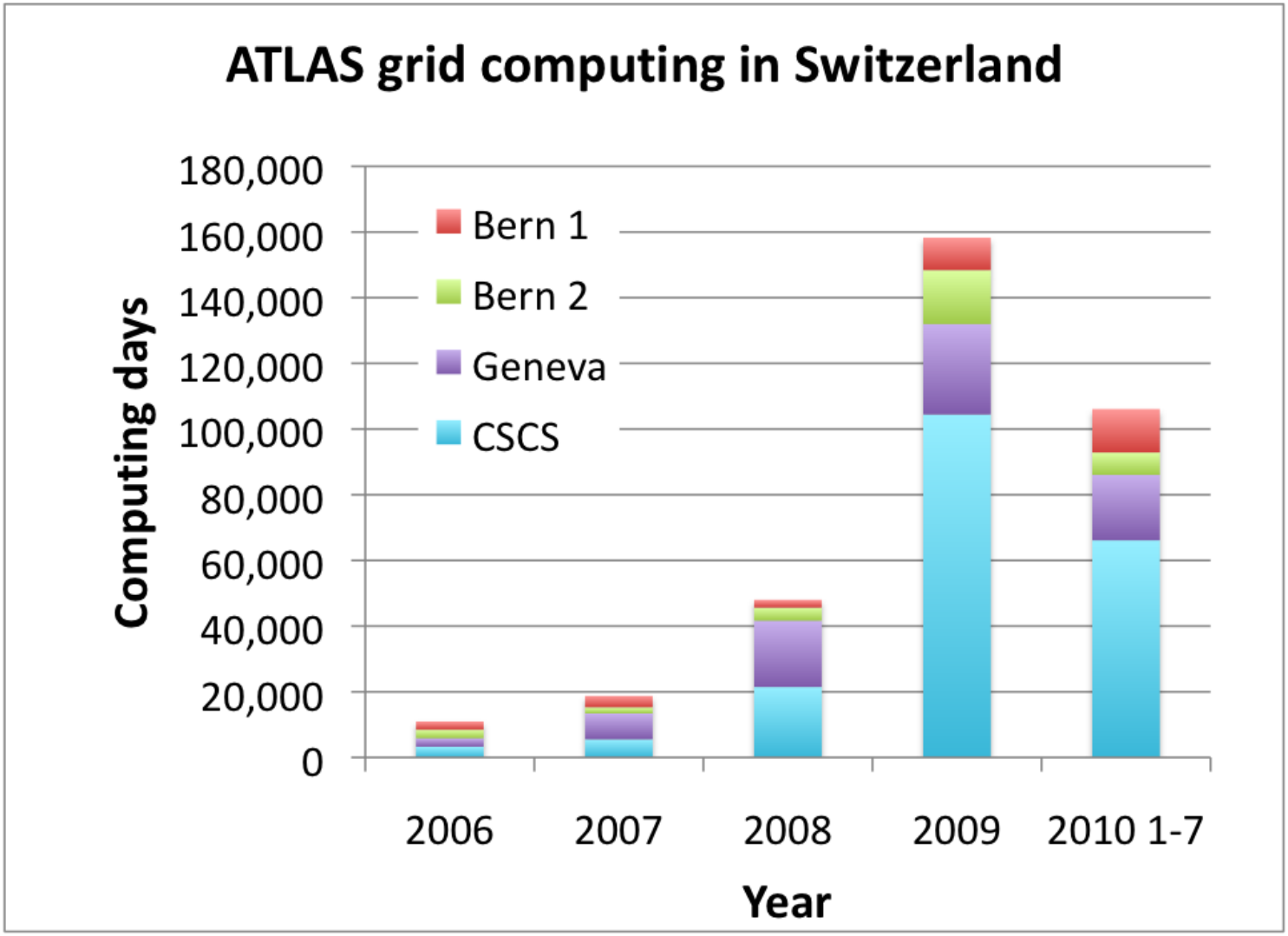
- The computing model gives many advantages that could allow the HEP community to develop and use a complex analysis model, which is needed to extract fast the early physics !

- The computing model is constantly improved and will remain an essential part of all physics analysis !

- The Swiss activities are really important in the DE cloud and for the physics groups. So, the support to the Tier-2 and the Tier-3's should follow the needs of the experiment in the coming years.

- Thanks a lot to all the people involved in the Tiers monitoring and management, in particular our colleagues from Bern, Manno and Geneva !

Significant increase of the grid usage



A few comments about how we use our T3 of Geneva

The Tier3s still play a very important role beyond the grid. For instance, here is how we contributed to the electron analysis:

1. Until recently we downloaded all the data in the AOD format to Geneva. We have than done the AOD to n-tuple step locally.

2. At some stage, as the AODs were growing, we have done the AOD to dAOD processing in Geneva and we kept only the dAOD, to free some space.

3. Now we are starting to use more the grid, like running the AOD to n-tuple step on it. So far this is no validated because we do not get the same number of events. Also the turn-around is slower. As the content of the n-tuple changes, we would need to redo this step a few times.

4. We will try to make the AOD to dAOD step on the grid and get the dAOD, containing only the electron candidates, to Geneva. We would than do (and redo) the n-tuple locally. That is the current plan, for data of periods D and E.

5. This was just an example. Every analysis will develop a different way of processing the data in some steps. Doing the last step, or the last two steps, locally rather than on the grid, should be faster and more reliable for the early data... We are limited by data volumes we can have locally. So we should be aware that the T3's would still be very busy…