Analysis use cases of the LHC experiments and corresponding Grid requirements at T2 centres

Kilian Schwarz
GSI Darmstadt
June 14, 2006
T2 Workshop, CERN

Table of contents

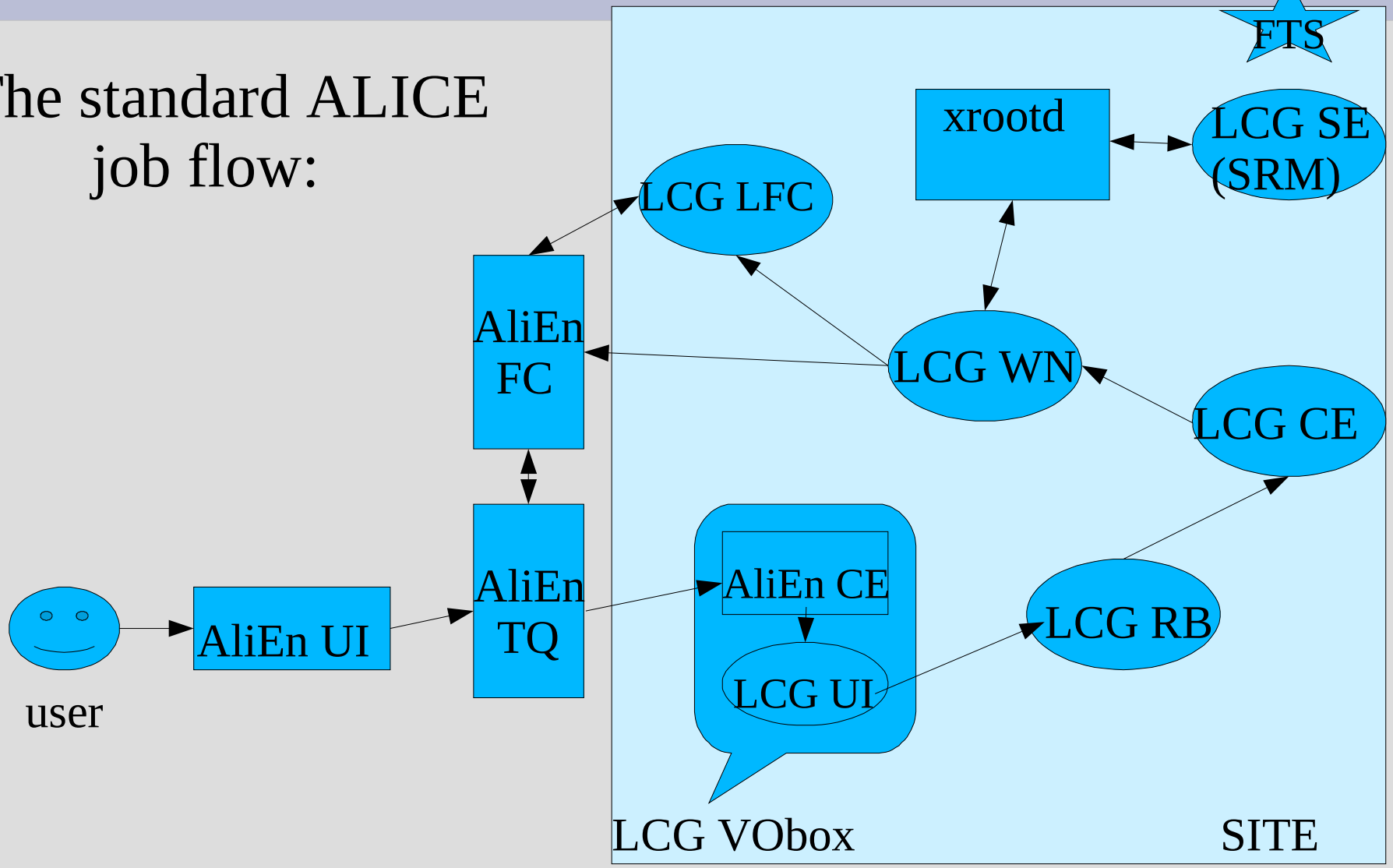
- analysis use cases of the experiments and requested Grid services at T2 centres
(partly repetition from yesterday)
- installation experiences of Grid services at T2 and partly T1 (ALICE)
- readiness for interactive analysis on the Grid

ALICE use cases at T2 centres

- (detector) simulation (MC production)
[- scheduled analysis
  - all data of a given type, data filtering for 'Sub-Analysis', output ESD/AOD]
- chaotic analysis (user analysis)
  - of local data
  - of remote data
  - single physics task, based on filtered data, output histogram files

ALICE: Grid services required at T2 centres

The standard ALICE job flow:



ALICE: elementaryGrid service requirements at T2 centres

based on the standard job flow (see last trancparency) the most elementary Grid services needed at T2s are:

- AliEn site services (on Vobox)
- LCG Vobox services
- LCG CE
- LCG WN

to make a site complete additionally the following additional services should be installed:

- LCG LFC (in principle either another site's LFC or the central FC can be used)
- xrootd or LCG SE (classic or SRM) --- in principle the produced data can also be stored at CERN or at the local T1 centre, and even according to the current computing model the data produced at T2s should be permanently stored at the corresponding T1s. So T2s need only tactical storage

ALICE: higher level services and services needed for analysis use cases

for performance reasons and to bring more into the game:

- LCG RB (in principle any RB supporting the ALICE VO can be used, though

Analysis use cases

a) analysis of local data:
- in principle no additional services needed.
- but for transparent and smooth data access for local users using ROOT based applications (AliRoot) xrootd on file servers is of advantage.

b) analysis of remote data:
- LCG SE (SRM) and FTS client necessary

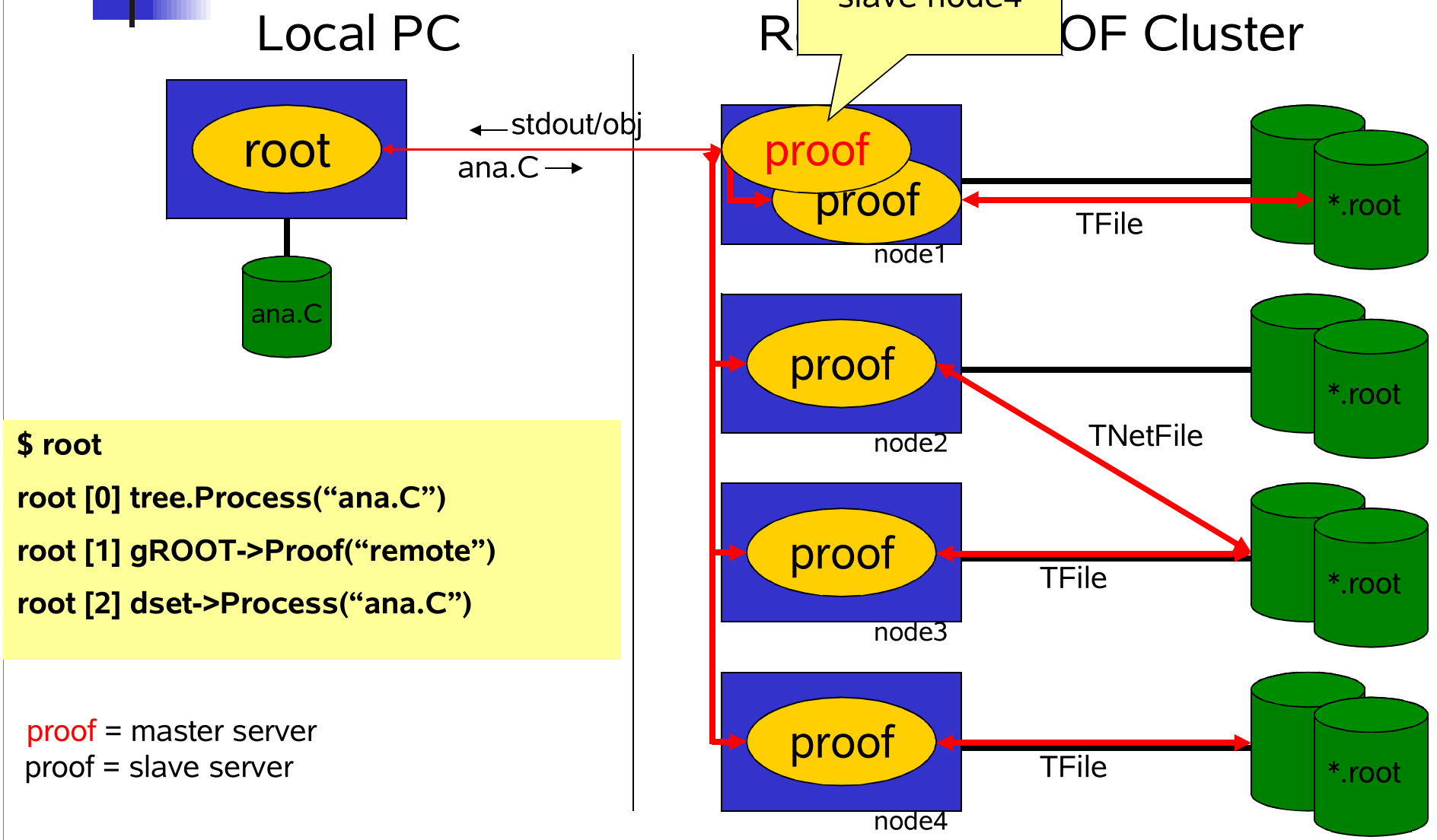
ALICE: Grid services needed for various analysis use cases

c) batch like analysis
- in combination with a) or
  b) no additional services needed.

d) interactive and parallel analysis of large data sets (PROOF)
- for data access xrootd is necessary !!!
  (see next transparency)

- Installation of a PROOF cluster with access to Grid enabled SE
- First installation at CERN CAF (after successful evaluation T2AFs)

Parallel Analysis ... Data



$ root

root [0] tree.Process("ana.C")

root [1] gROOT->Proof("remote")

root [2] dset->Process("ana.C")

proof = master server
proof = slave server

ATLAS: use cases at T2 centres

- 20 MB/s of reprocessed AOD data are transferred to T2s coming from T1s and T0. Large T2s should store a full set (200 TB/year of data taking) and more sets are distributed between smaller T2s. 1/3 of the AOD sample is stored at T2s. Also selected RAW and ESD data.

- generation of simulated data takes place mainly at T2s. They are transferred to T1s for permanent storage (CPU resources at T2s have to be matched to storage capacity at T1)

- Calibration

- Analysis, mainly chaotic user analysis

ATLAS: DDM (Don Quijote) – close binding between T1 and T2s.

locate data and bring the jobs to the site.
==> Distributed Data Management (DDM) consisting of FTS and LFC, responsible for data registration and replication.

since FTS and LFC need administration, it has been decided that these services are hosted by T1s. Each T1 provides services for a dedicated group of T2s.
==> centres should be geographically close and network link must be fast and reliable.

ATLAS: 3 Grids

- LCG/EGEE (DDM: only registration component)
  - gLite RB (better, but not yet as matured as LCG RB)
  - or CondorG (fast and production proof)
  - GANGA (job submission framework) for distributed analysis
- OSG (DDM fully in production, site services also at T2s)
  - PANDA as WMS (pilot jobs ..., production and analysis combined)
  - pathena for distributed analysis
- Nordugrid (DDM: only registration, FC not yet clear)
  - 0.5 to 1 s/job

DDM is shared among Grids

WMS is Grid specific, also different submission tools

common ATLAS UI planned

ATLAS: Grid services required at T2 centres

- SRM based SE (minimum 20 MB/s * 24 h = 2 TB)
- FTS channels should be set up from and to corresponding T1
- gLite 3.0
- Analysis and production needs to be done in parallel
- no part of the DDM has to be installed at T2s, but being connected to DDM services on T1s is necessary
- sufficient network connection

Data Analysis tests in August and autumn

CMS: use cases at T2 centres

- user analysis: T2 centres provide
  the largest resources for this task

  - but still a small fraction of these resources is used.
  - efficiency of using them needs to increase significantly.

CMS – data flow

- data are served and archived by T1 centres (any of them, no dedicated T1-T2 link)
- at the beginning frequent raw data access is necessary. Depending on size, and frequency of selections a data flow of 2.5 Gb/s – 10 Gb/s to T2s is reached. Assuming 200 TB of local disk space this would last 2 days to 1 week.

CMS – Central and local Grid Services

Central Experiment Services:
- Dataset Bookkeeping Service (DBS)
  - several files are registered as a file block
  - a file list from DBS allows job splitting
- Dataset Location Service (DLS)

Site Services (including T2s):
- Data Placement Service (PhEDEx) – can use FTS, srmcp, gridFTP
  - layer above individual transfers, deals with replication of datasets
- various agents, e.g. Production_Agent ...
- LFC
- CRAB (see next slide)

at T2s additionally finer grained authorisation for roles and groups through VOMS is needed.

CMS Analysis flow

- Since July 2005 the full Analysis work flow can be done using CRAB, the CMS Remote Analysis Builder:
  - CRAB handles data discovery (query to DBS and DLS)
  - job preparation
  - assigning jobs to files according to entries in DBS
  - submitting the application
  - submitting jobs through the LCG RB
  - submission location identified by block location in DLS

LHCb – use cases at T2 centres computing model

T0: - storage of raw data and MC samples.
- reconstruction
- stripping
- user analysis

T1s: - reconstruction
- stripping
- user analysis
- storage of: 1 copy of raw data
3 copies of MC samples

T2s: - MC production

T2s can be included in user analysis, if enough disk space is available and depending on what data are stored locally

LHCb – user analysis

- input: DSTs (from stripping) and TAG data
- typical analysis job will scan TAG data for intesesting events
- candidates will be filtered according to the complete DST information
- output: sub DST or ntuple like data sets.
- further analysis at T3 centres

LHCb – Grid usage

- user interface: GANGA
- typical user does not need to know where data are
- typical workflow:
  - code development with subset of data, prototyping
  - get full statistics via Grid

Summary of experiment use cases see talk of David Colling yesterday

- all experiments have quite similar ideas from the concept point of view Only LHCb plans no user analysis at T2s.
- the other 3 experiments do MC production and end user analysis at T2s.
- for all 3 experiments data movement following knowledge about data content and location is important
- although the concrete implementation differs greatly from experiment to experiment, all see the need to „separate the user from the complexity of WLCG“

Summary of services needed at T2s:

ALICE : - standard LCG services
- AliEn site services on Vobox
- LCG LFC
- xrootd
- ev. SRM and FTS

ATLAS: - SRM
- FTS channels have to be set up
- gLite 3.0
- good network connection

CMS :- next to LCG services
- LFC
- PhEDEx
- various agents
- CRAB

LHCb : - no user analysis at T2s

Only ATLAS has almost no additional requirements to T2 centres !!!

Grid service installation experience (ALICE point of view)

Remark: I am ALICE site admin at T2 /T1 in Germany.
So this part of the presentation will be biased :-)

LCG standard services:

- installation and configuration now relatively straight forward. Things improved considerably compared to the starting time. Although it would be helpful if more LCG services would be available as tar balls, packed accordingly, dependencies included (self sufficient tar-balls). This would simplify the installation in non SL environments (GSI Darmstadt uses Debian Linux). Still after each upgrade it is time consuming, though, until all services on different boxes work together properly and publish the correct things.
- LFC has been connected to the central Oracle DB at GSI. At that time a standard installation procedure has not been provided yet.
- at T1 centre it proved to be difficult to convince LCG site admins to provide the correct versions of LCG services (e.g. vobox) before experiment specific stuff could be started

Grid service installation experience II

various SE setups:

- SRM: a long time we struggled in setting up the SRM at GSI. Main points are here the difficulties to marry SRM with xrootd and also to find the correct backend, since the official LCG SRMs come either with DPM or dCache, but actually we wanted to connect our own home made MSS (gStore). In the end we tried almost everything and we now several prototypes running (1 based on FNAL SRM with plain disk, 1 (FNAL SRM) where we try to connect gStore, 1 SRM with dCache backend, now also with xrootd doors included.
- FTS: setup of FTS channels to and from T1 centre seems also not to be straightforward. We spent a long time configuring and trying out FTS transfers, and still the setup seems unstable. Sometimes transfers work, sometimes they don't, seems to depend on lucky times :-(

Grid service insallation experience III

various SE setups, II:

- xrootd: also the xrootd based tactical SE brings some difficulties.
  Setting up a redirector and 1 to X dataservers does not seem to be the most simple thing on earth. Especially if your redirector and the dataservers are in different networks. Various things have to be optimised in the configuration files.
  Also it is not straight forward to connect xrootd to SRM or some home made MSS (gStore) or to AliEn.
  No standard scripts seem to be available so far.
  But for the ALICE xrootd installation automatic installation scripts are on the way.

Grid Service Installation Experience IV and conclusion ...

- Experiment Specific Services on VOBox:
not much work for site admin, since they are installed automatically by the alien-installer. And then, most of the times, they work out of the box. And if they don't one has to upgrade :-)
Confusing is, that ALICE sometimes adds new configuration files or modifies the importance of existing ones, though.

Conclusion: The setup and configuration of the basic WLCG infrastructure consumes significantly more time than the setup of the additional ALICE specific services on the VOBOX, although the list of services seems to be long. Also for xrootd automatic installation is on the way.
And then ... only ATLAS does not demand additional services on T2 sites !!!

Readiness of (chaotic) multi user analysis

The the ways the experiments attack this problem seems promising. Although none of the LHC experiments really did mass production doing interactive analysis by many users at the same time.

ALICE: demos of AliEn and PROOF have been shown several times successfully (run by the CERN experts). Also tutorials are given to users. The CAF is on the way, but not yet in production.
The T2AFs are not yet existing. The transition from a production user system (aliprod) to a multi user system is not entirely smoothly. Sometimes newly registered users run into problems the traditional users don't have.

But in principle everything seems to be in place (or will be in place) and when most experiments plan to exercise Multi User analysis on the Grid in autumn this year things could actually work :-)