Panda, a Pilot-based workflow manager
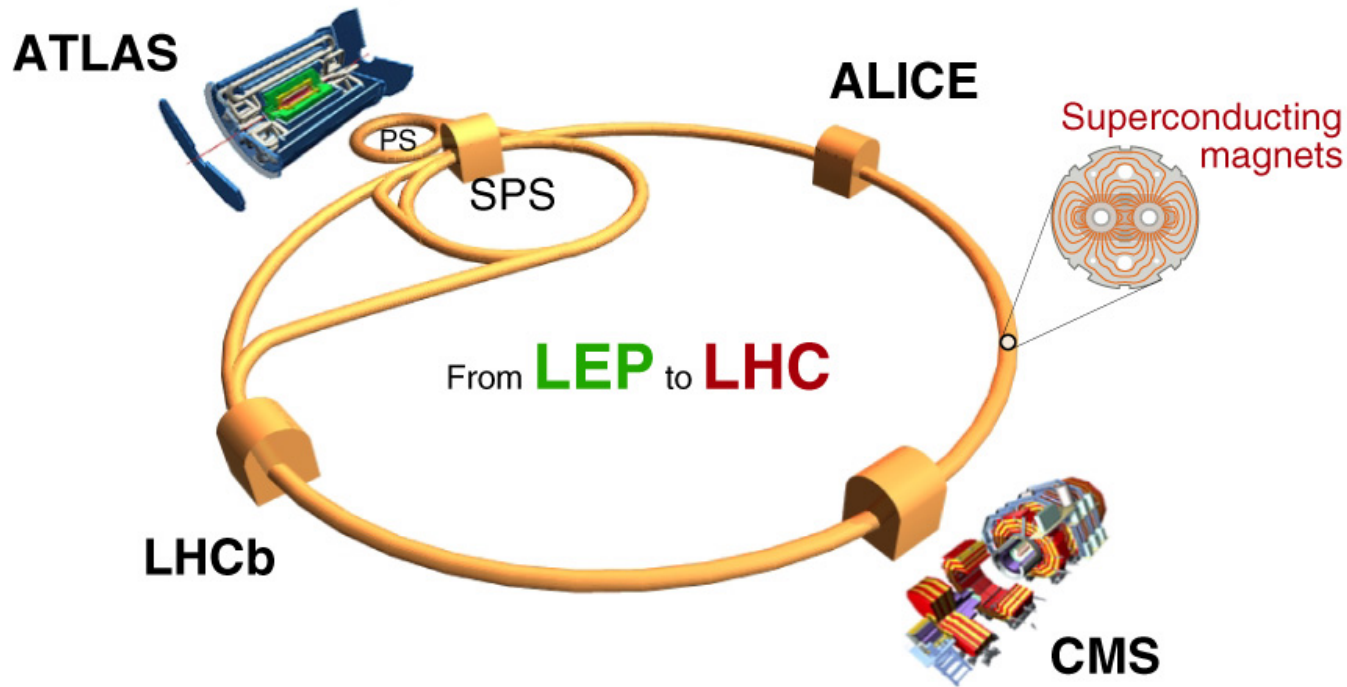
Open Science Grid

New Mexico Grid School – April 8, 2009
Marco Mambelli – University of Chicago
marco@hep.uchicago.edu

The ATLAS VO

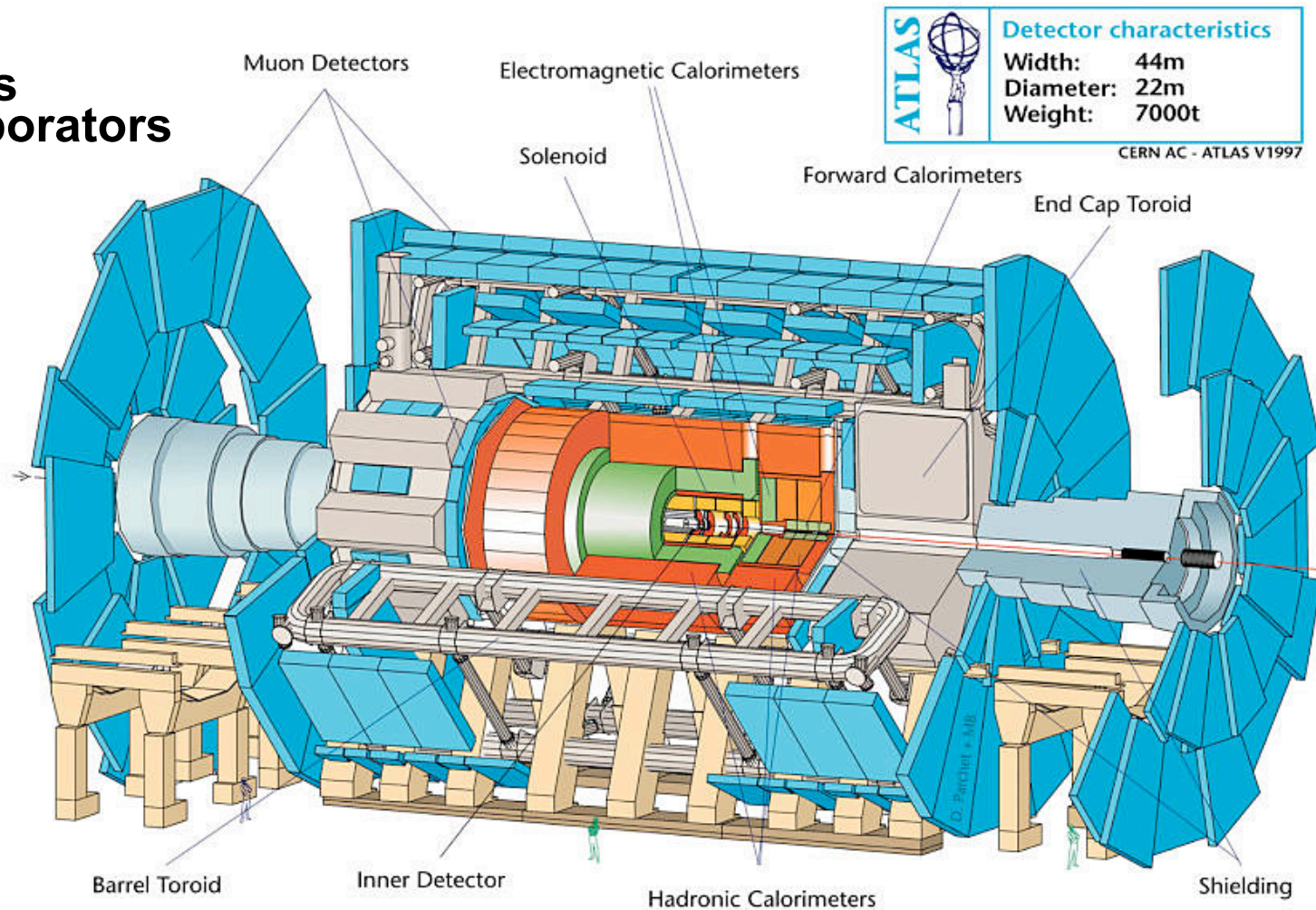▸ Virtual Organization in OSG (and other Grids)
  ▸ In OSG since the beginning
  ▸ https://twiki.grid.iu.edu/bin/view/VO/ATLAS
  ▸ https://lcg-voms.cern.ch:8443/vo/atlas/vomrs

▸ Collaboration for the ATLAS experiment in the LHC at CERN
  ▸ http://atlas.ch/
  ▸ http://atlas.web.cern.ch/Atlas/ATLASreg_form.pdf

Open Science Grid

LHC experiment at CERN

	Beams		Energy		Luminosity
LEP	e^+	e^-	200	GeV	$10^{32}\,cm^{-2}s^{-1}$
LHC	p	p	14	TeV	10^{34}

http://www.youtube.com/watch?v=j50ZssEojtM

Panda, Pilot-based WFM - Marco Mambelli

3

The ATLAS experiment

37 Countries
167 Institutes
~2000 Collaborators

Open Science Grid

PANDA

- PANDA = Production ANd Distributed Analysis system
  - Designed for analysis as well as production for High Energy Physics
  - Works both with OSG and EGEE middleware

- A single task queue and pilots
  - Apache-based Central Server
  - Pilots retrieve jobs from the server as soon as CPU is available → late scheduling

- Highly automated, has an integrated monitoring system

- Integrated with ATLAS Distributed Data Management (DDM) system

- Not exclusively ATLAS: has its first OSG user in CHARMM (Chemistry at HARvard Molecular Mechanics)

Panda System

Panda server

DDM

LRC/LFC

ProdDB — bamboo — *job* → Panda server

send log *http* → logger

pull
https

job

job

https

submit

site B

pilot

site A

End-user

Worker Nodes

pilot

submit

condor-g

Autopilot

Open Science Grid

Panda Server

clients

PandaDB

https

Panda server

Apache + gridsite

DQ2

LRC/LFC

logger

pilot
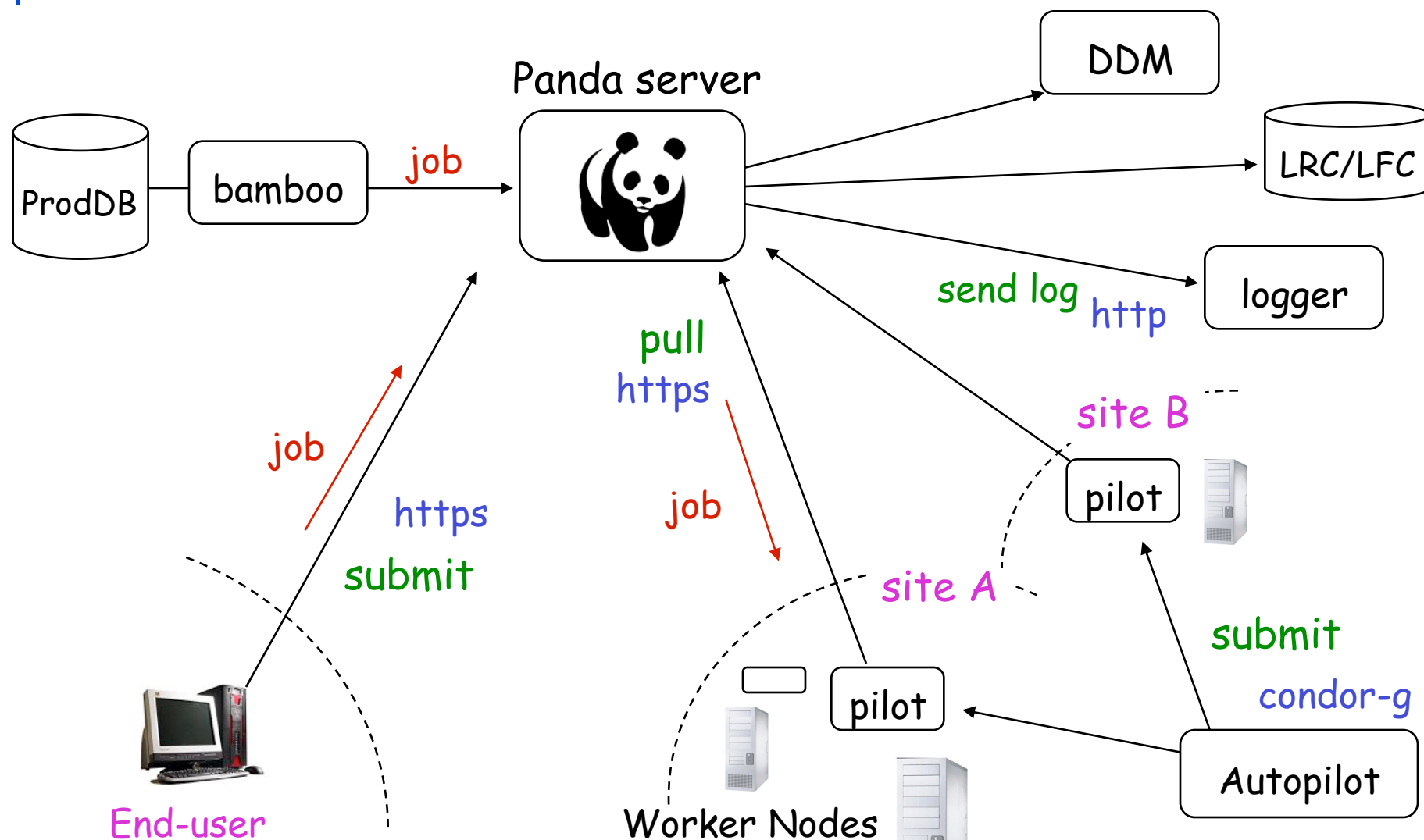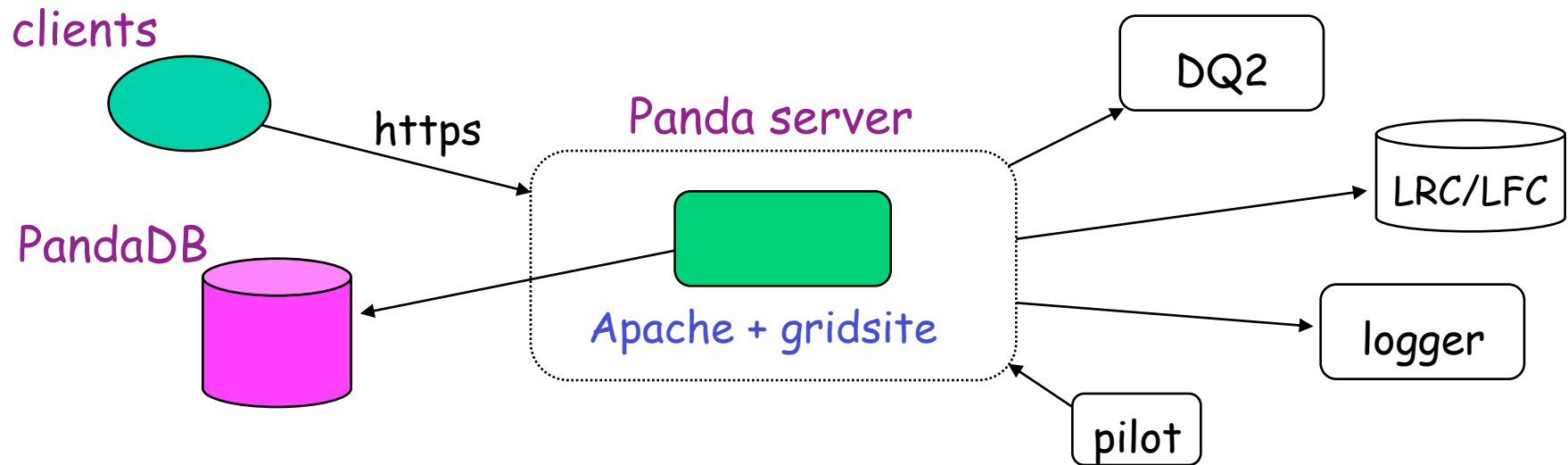
▸ Central queue for all kinds of jobs

▸ Assign jobs to sites (brokerage)

▸ Setup input/output datasets

　　▸ Create them when jobs are submitted
　　▸ Add files to output datasets when jobs are finished

▸ Dispatch jobs

Open Science Grid

Bamboo

prodDB   Bamboo   Panda server

cx_Oracle

Apache + gridsite

cron   https   https
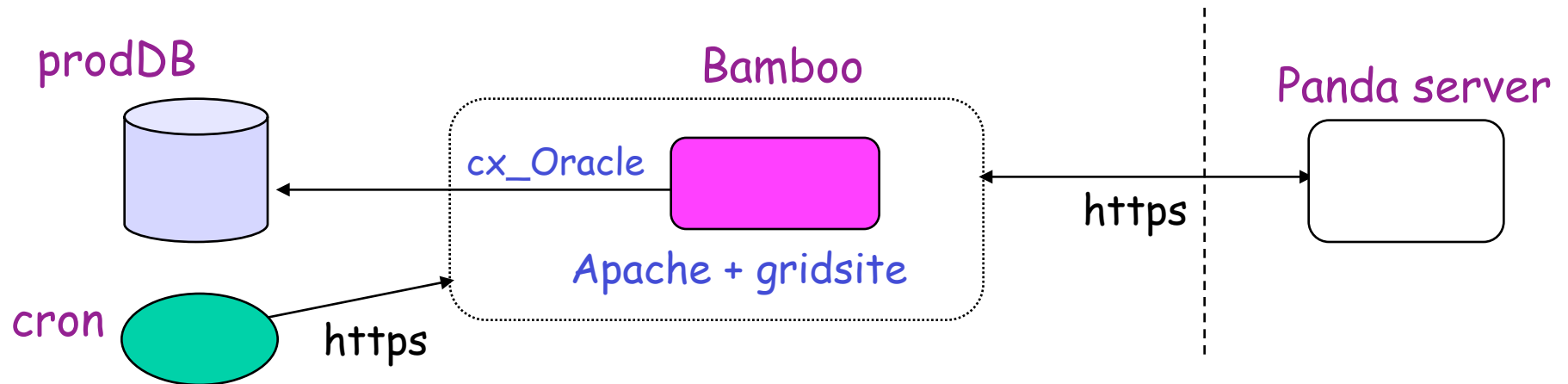
▸ Get jobs from prodDB to submit them to Panda
▸ Update job status in prodDB
▸ Assign tasks to clouds dynamically
▸ Kill TOBEABORTED jobs

▸ A cron triggers the above procedures every 10 min

Open Science Grid

Panda Job Timeline

▸ Rely on ATLAS DDM
  - ▸ Panda sends requests to DDM
  - ▸ DDM moves files and sends notifications back to Panda
  - ▸ Panda and DDM work asynchronously

▸ Dispatch input files to execution sites and aggregate output files to destination

▸ Jobs get 'activated' when all input files are copied, and pilots pick them up
  - ▸ Pilots don't have to transfer data (asynchronous)
  - ▸ Data-transfers and Job-executions can run in parallel

DDM   submitter   Panda

submit Job

subscribe T2 for disp dataset

data transfer

callback

pilot

get Job

run job

finish Job

add files to dest datasets

data transfer

callback

Open Science Grid
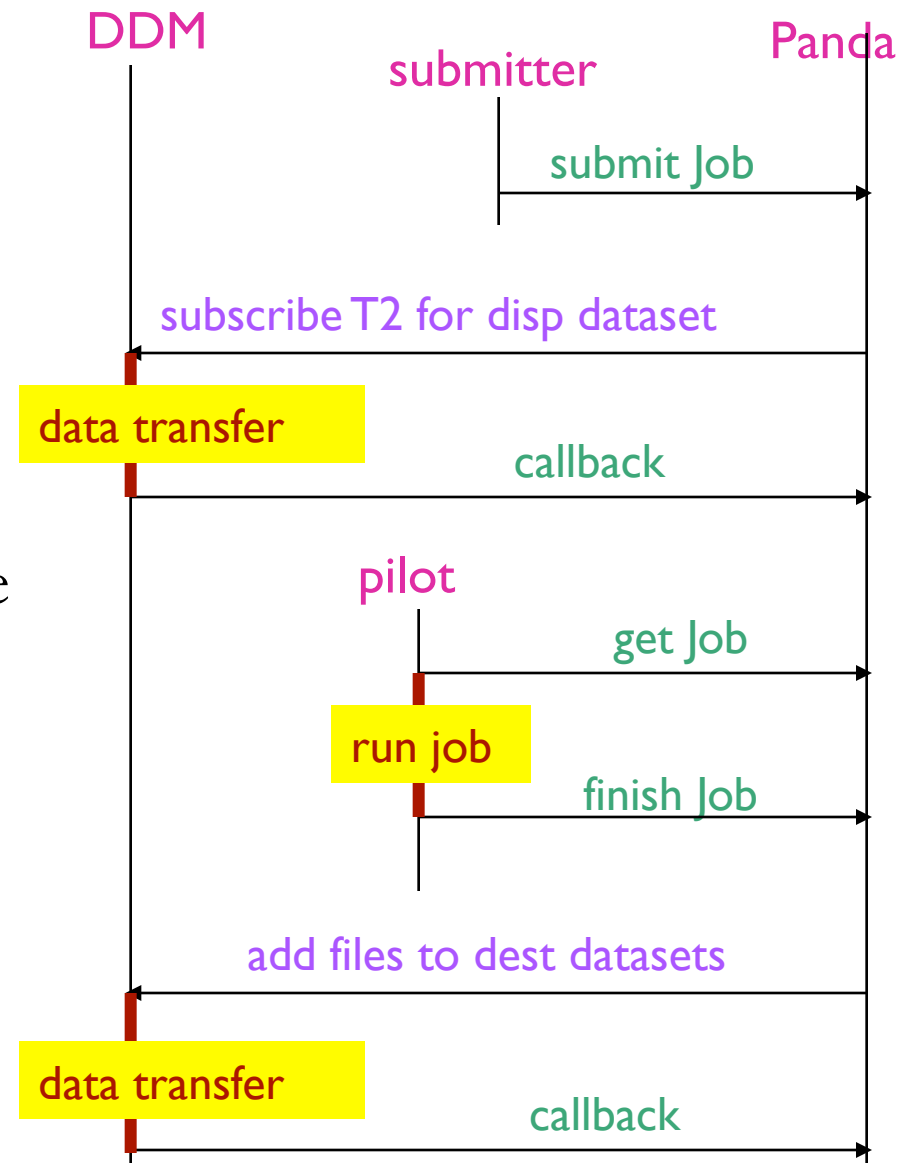
How the pilot works

▸ Sends the several parameters to Panda server for job matching (HTTP request)
  ‣ CPU speed
  ‣ Available memory size on the WN
  ‣ List of available ATLAS releases at the site

▸ Retrieves an `activated' job (HTTP response of the above request)
  ‣ activated → running

▸ Runs the job immediately because all input files should be already available at the site

▸ Sends heartbeat every 30min

▸ Copy output files to local Storage Element and register them to Local Replica Catalog

Open Science Grid

Pilot vs ATLAS Job

Pilot

- Submitted by factories
  - remote submit hosts
  - local cluster factories
- Managed by factories
- Python code to support ATLAS Job execution
- Submitted continuously
- Partially accounted
  - no big deal if some fail

ATLAS Job

- Submitted by users or production managers (Bamboo)
- Managed by Panda Server
- Runs Athena software (ATLAS libraries)
- Submitted when needed
- Fully accounted
  - error statistics are important

Open Science Grid

Some monitoring resources

▸ The following pages present some monitoring example

▸ Screenshots are just example pages, actual content varies

▸ URLs are one of the possible URLs providing a similar page
  - ▸ e.g. queries may vary the actual Site or Time interval

▸ Main URLs:
  - ▸ DDM Dashboard: http://dashb-atlas-data-test.cern.ch/dashboard/request.py/site
  - ▸ Panda Monitor: http://panda.cern.ch:25880/ or http://panda.atlascomp.org/?redirect=pandamon
    (hostname may change since there are multiple servers)

▸ Take time to navigate Panda Monitor and the Dashboard

Panda Monitor: production dashboard

http://panda.cern.ch:25880/server/pandamon/query?dash=prod

Panda Monitor: Dataset browser

http://panda.cern.ch:25880/server/pandamon/query?overview=dslist

Panda, Pilot-based WFM - Marco Mambelli

Panda Monitor: error reporting

http://panda.cern.ch:25880/server/pandamon/query?days=1&overview=errorlist

Open Science Grid

DDM Dashboard: overview

http://dashb-atlas-data-test.cern.ch/dashboard/request.py/site

Open Science Grid

?
!

Open Science Grid
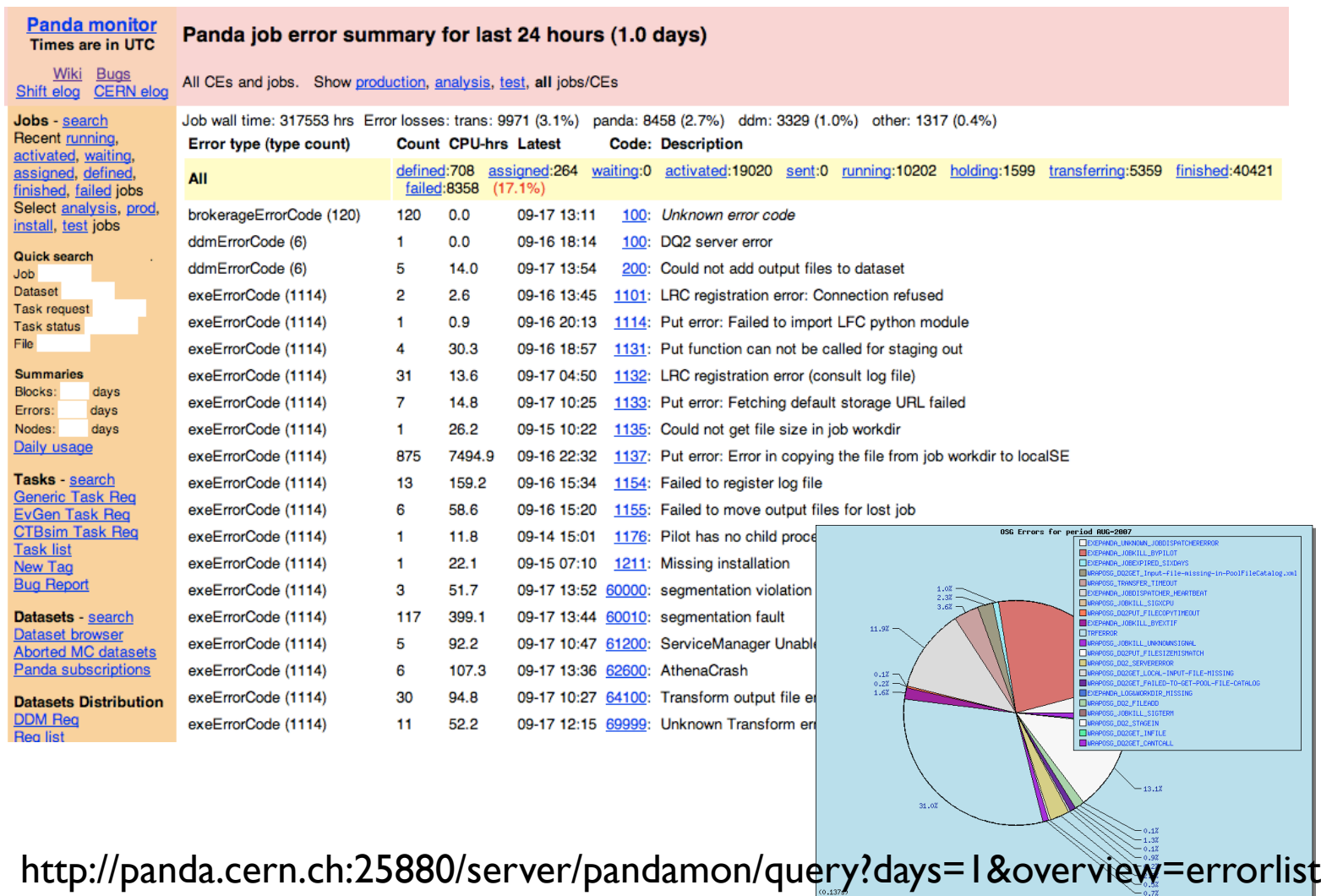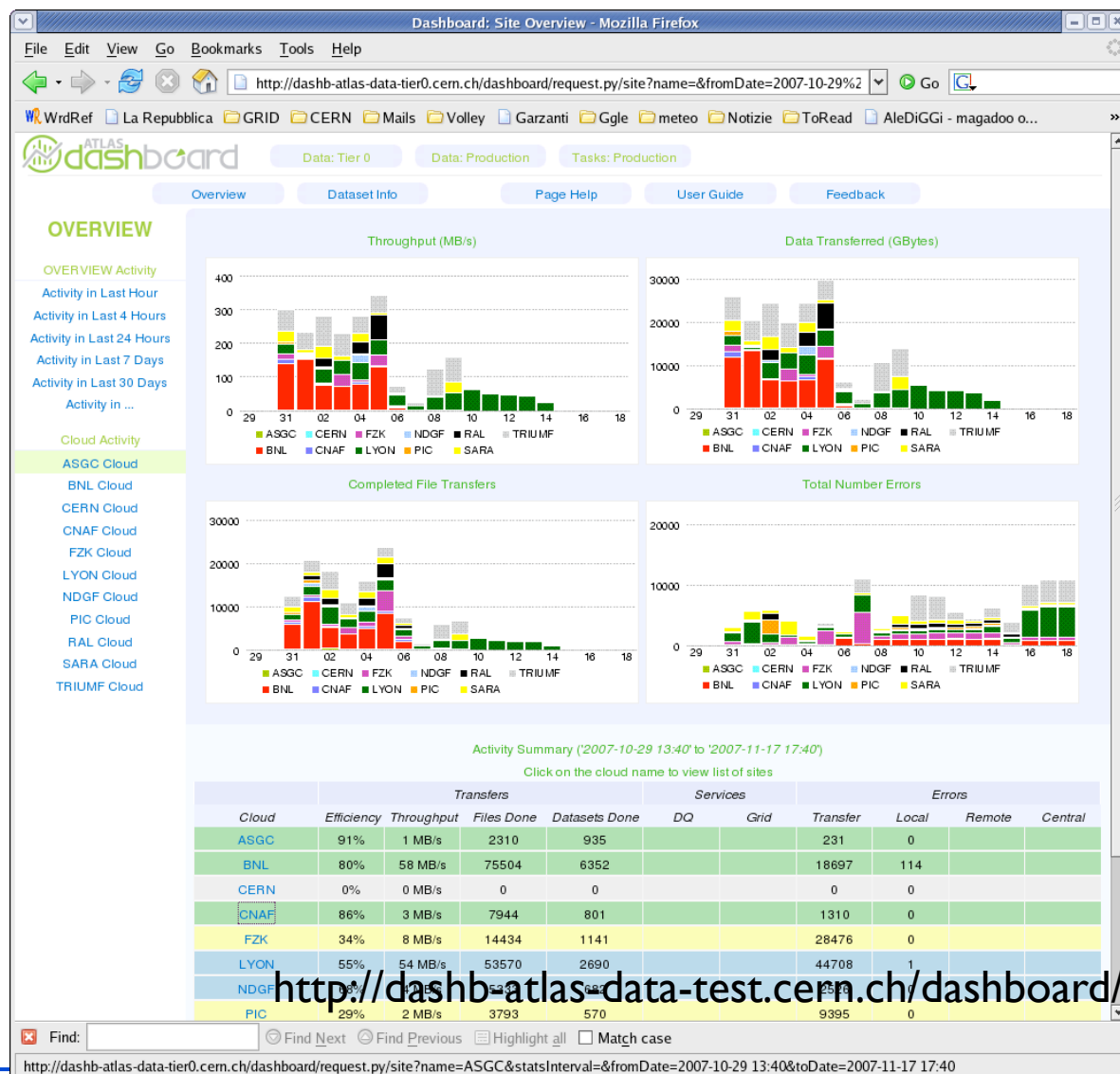
Client-Server Communication

▸ HTTP/S-based communication (curl+grid proxy+python)

▸ GSI authentication via mod_gridsite

▸ Most of communications are asynchronous

  ▸ Panda server runs python threads as soon as it receives HTTP requests, and then sends responses back immediately. Threads do heavy procedures (e.g., DB access) in background → better throughput

  ▸ Some are synchronous

Open Science Grid