
Panda, a Pilot-based workflow manager



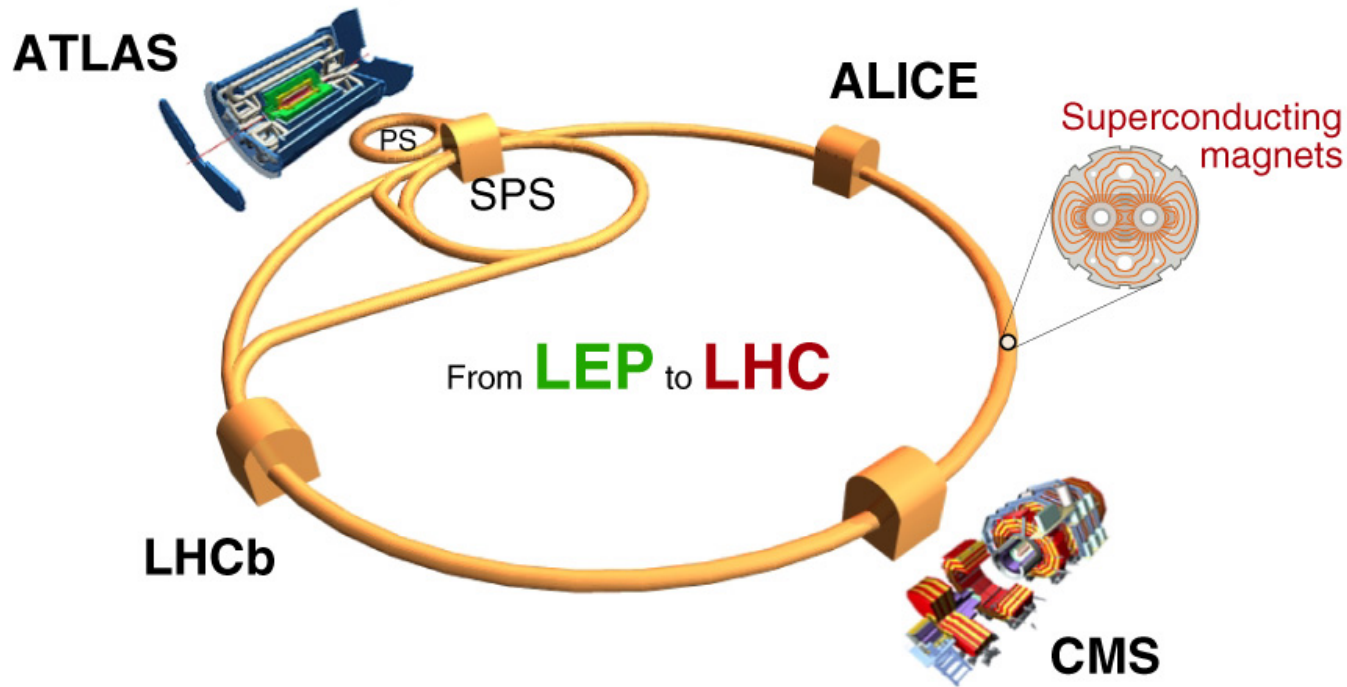
Open Science Grid

New Mexico Grid School – April 8, 2009

Marco Mambelli – University of Chicago

marco@hep.uchicago.edu

LHC experiment at CERN

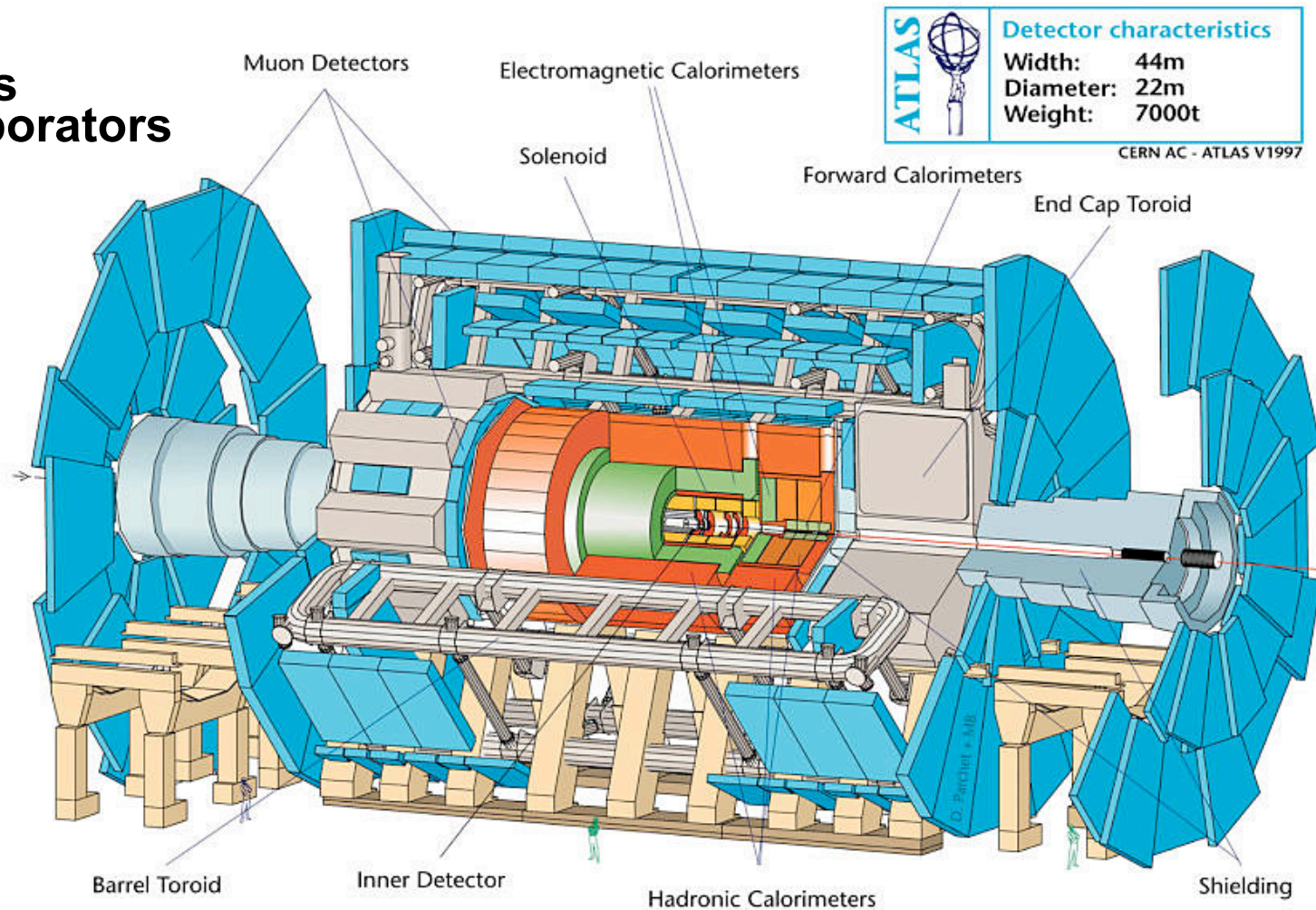


	Beams		Energy		Luminosity
LEP	e^+	e^-	200	GeV	$10^{32} \text{ cm}^{-2} \text{ s}^{-1}$
LHC	p	p	14	TeV	10^{34}

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

The ATLAS experiment

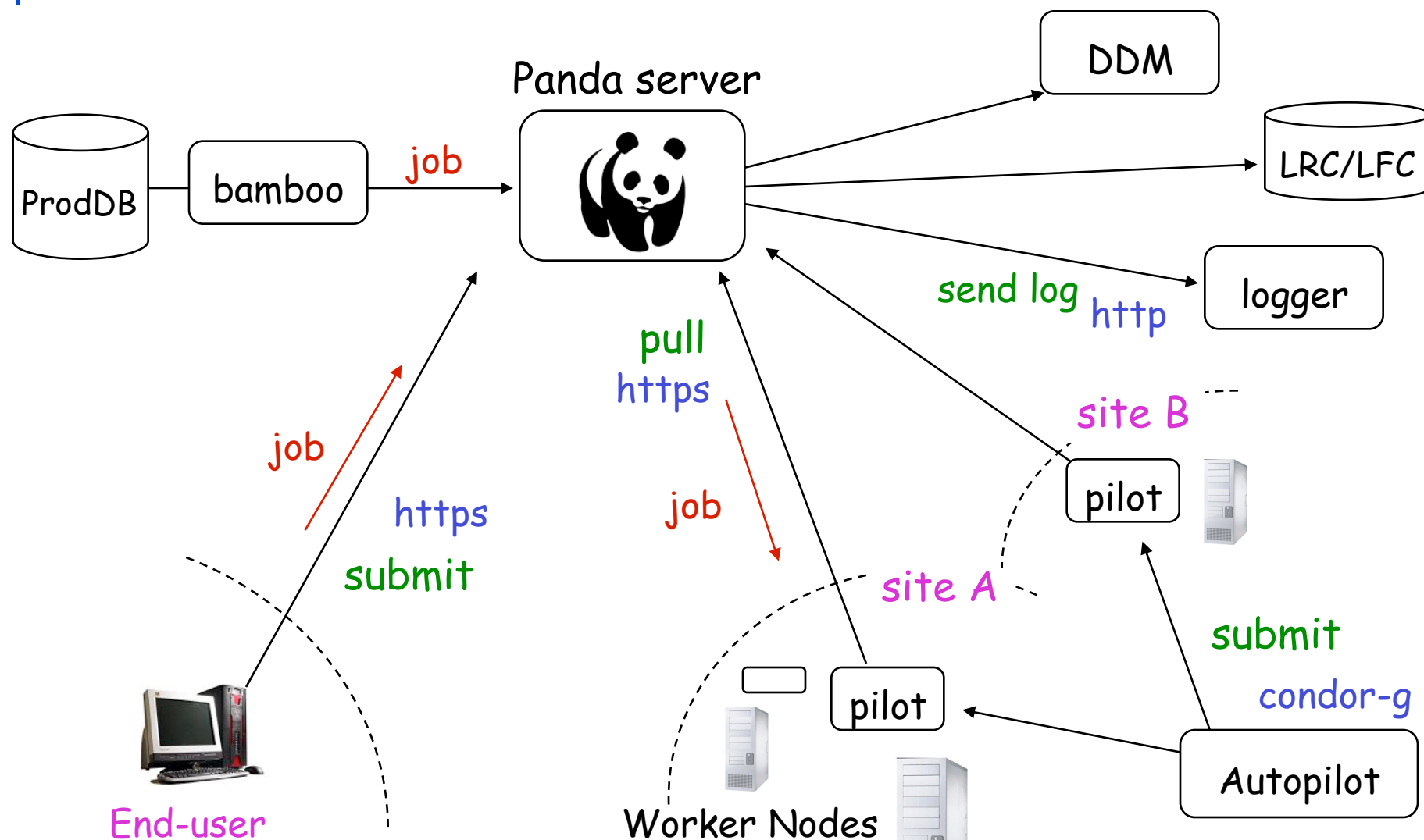
37 Countries
167 Institutes
~2000 Collaborators



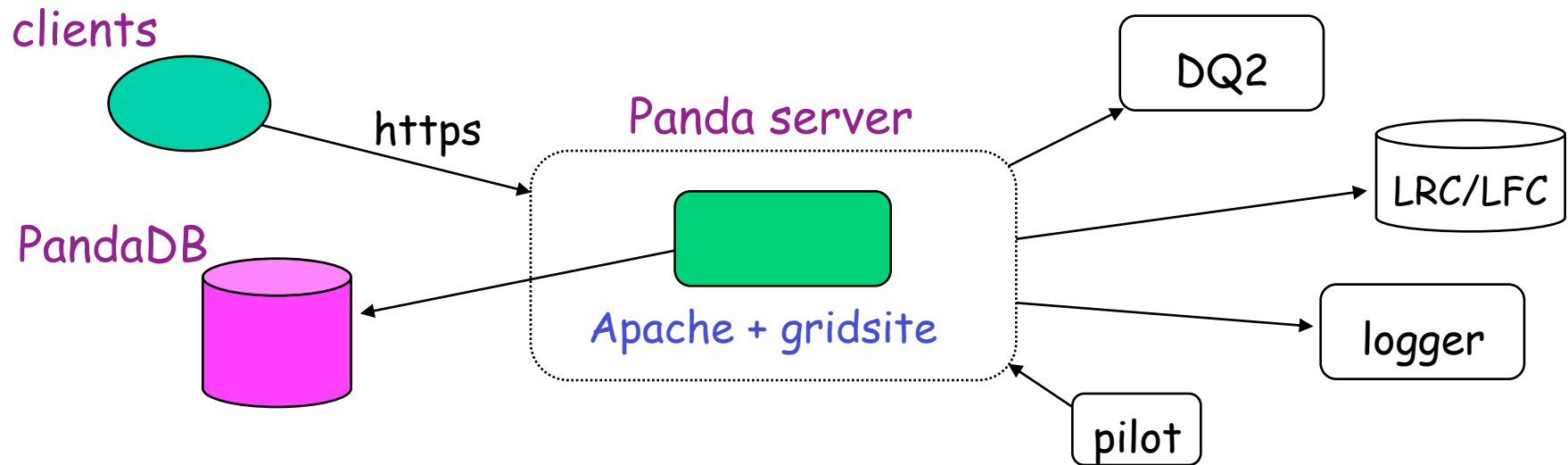
PANDA

- ▶ 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

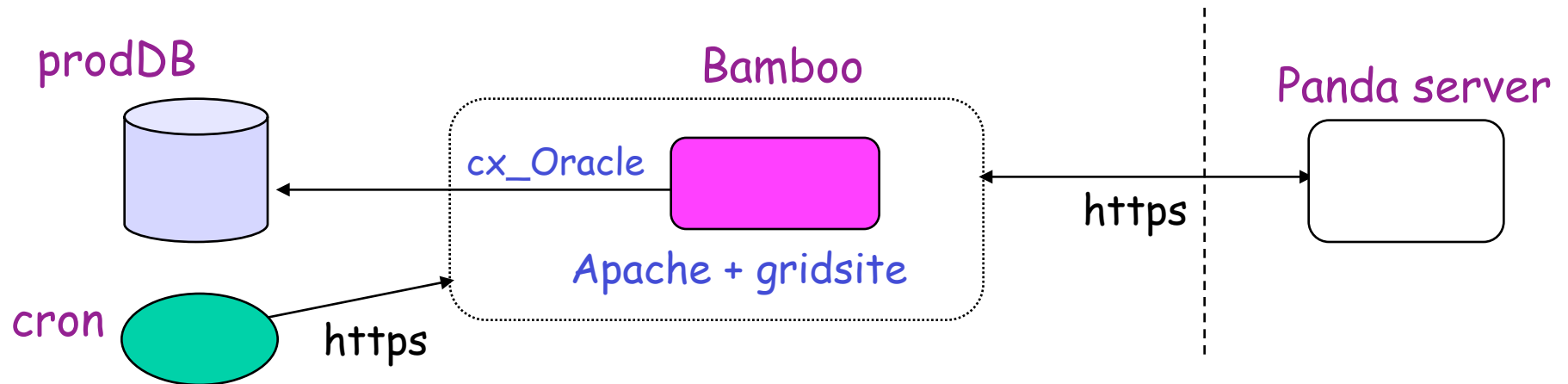


Panda Server



- ▶ 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

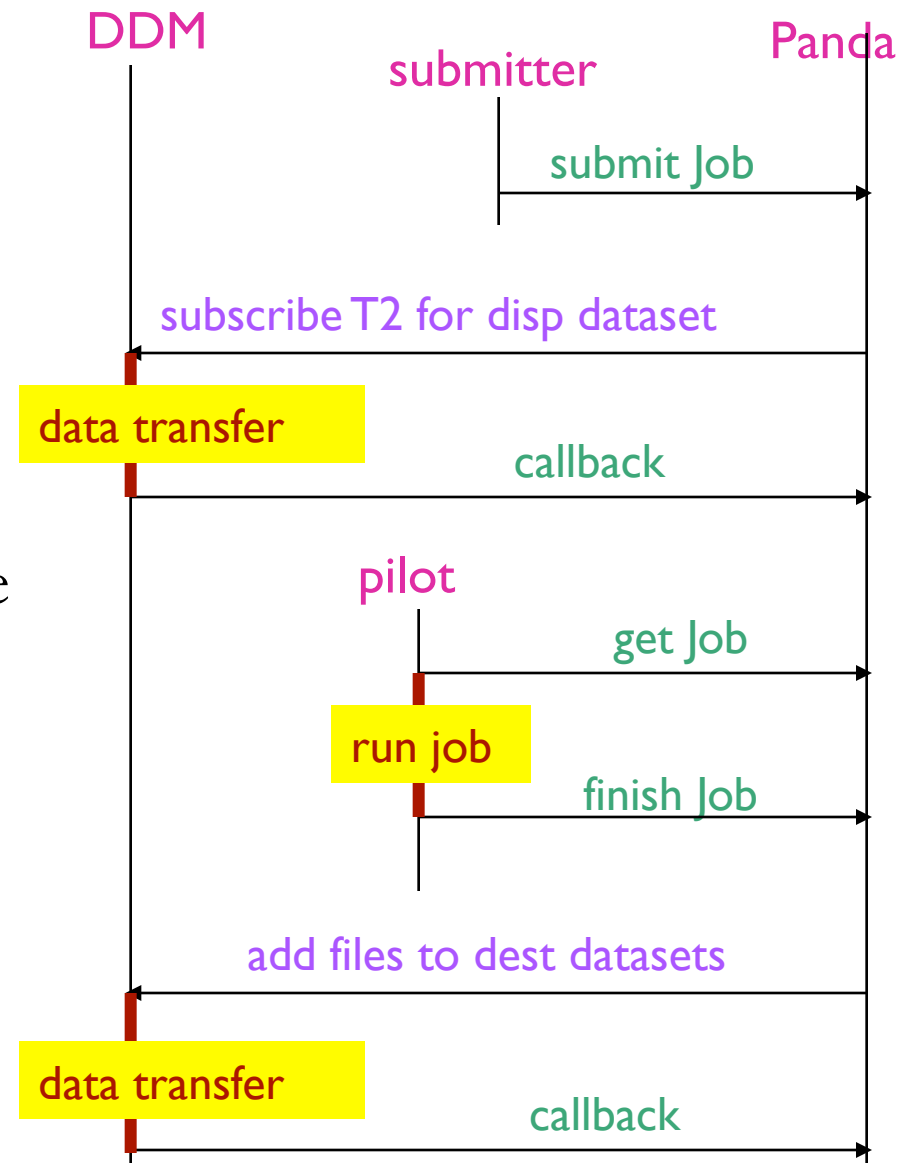
Bamboo



- ▶ 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

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



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

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

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

Panda monitor
Times are in UTC

Wiki
Shift elog

Jobs - search
Recent running,
activated, waiting,
assigned, defined,
finished, failed jobs
Select analysis, prod,
install, test jobs

Quick search
Job
Dataset
Task request
Task status
File

Summaries
Blocks: days
Errors: days
Nodes: days
Daily usage

Tasks - search
Generic Task Req
EvGen Task Req
CIBsim Task Req
Task list
New Tag
Bug Report

Datasets - search
Dataset browser
Aborted MC datasets
Panda subscriptions

Datasets Distribution
DDM Req
Req list
ADCs

Panda Production Operations Dashboard

Panda shift guide calendar mailing list
ADCoS twiki calendar mailing list
Production task support mailing list

Click for help

Servers: BNL:OK BNLdev:OK CERN:OK Logger:OK Bamboo:OK

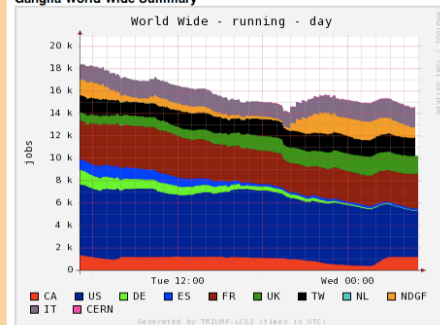
Active tasks: CA:14 CERN:1 DE:15 ES:4 FR:22 IT:12 NL:1 TW:15 UK:6 US:18

Bamboo task brokerage, job submissions, status over last 12 hours

Jobs updated >12 hrs ago: activated:5950 running:none

Jobs updated >36 hrs ago: transferring:12

Ganglia World Wide Summary



Space available at sites:

Site GB As of (UTC)

US

AGLT2-condor 6921 09-17 02:19

ANALY_AGLT2-condor 6933 09-17 01:10

ANALY_MWT2-pbs 66815 09-17 04:27

ANALY_NET2-pbs 41342 09-17 04:21

ANALY_SLAC-lsf 16401 09-17 04:50

ANALY_SWT2_CPB-pbs 144839 09-17 01:08

BU_ATLAS_Tier2 41342 09-17 03:39

BU_ATLAS_Tier2o 41342 09-17 04:49

HU_ATLAS_Tier2 41342 09-17 02:40

IU_OSG-pbs 76176 09-17 04:49

MWT2_IU-pbs 76176 09-17 04:41

MWT2_UC-pbs

OU_OCHEP_SWT2

SLACKPD-lsf

SWT2_CPB-pbs

UC_ATLAS_MWT2

Other SEs reporting

ALBERTA-LCG2 PF

FZK-LCG2 MCDISK

IN2P3-CC MCDISK

INFN-T1 MCDISK

NDGF-T1 MCDISK

PIC MCDISK

RAL-LCG2 MCDISK

SARA-MATRIX MCI

IT	566	8372	09-17 04:50	255	0	15	0	3417	0	1776	50	443 / 0	2003	668	25%	0%	25%
NL	2	228	09-16 23:24	688	0	227	0	0	0	0	0	0 / 0	0	1	100%	0%	100%
UK	870	8439	09-17 04:50	804	0	0	0	2140	0	1599	46	1236 / 0	2905	513	15%	0%	15%
US	1526	13514	09-17 04:50	1275	0	5	0	152	0	4113	230	2526 / 12	6263	225	3%	0%	3%
TW	468	7643	09-17 04:50	241	0	10	0	3241	1	1624	29	267 / 0	2444	27	1%	0%	1%

	US Sites	Nodes	Jobs	Latest	Pilots (3hrs)	defined	assigned	waiting	activated	sent	running	holding	transferring	finished	failed tot	trf	other
	Site Name	1526	13514	09-17 04:50	1275	0	5	0	152	0	4113	230	2526 / 12	6263	225	3%	0% 3%
	AGLT2 🚩	114	2009	09-17 04:50	197	0	0	0	0	0	529	48	935 / 0	397	100	20%	0% 20%
	BNL ATLAS 1 🚩	318	2580	09-17 04:50	184	0	0	0	1	0	914	46	0 / 0	1605	14	1%	0% 1%
	BNL ATLAS 2 🚩	0	0		0	0	0	0	0	0	0	0	0 / 0	0	0		
	BU ATLAS_Tier2 🚩	12	44	09-17 04:46	0	0	0	0	13	0	22	6	2 / 0	0	1	100%	0% 100%
	BU ATLAS_Tier2o 🚩	85	845	09-17 04:50	85	0	0	0	0	0	299	5	88 / 0	429	24	5%	5% 0%
	GLOW-ATLAS 🚩	47	160	09-17 04:40	0	0	0	0	73	0	1	7	76 / 0	2	1	33%	0% 33%
	HU ATLAS_Tier2 🚩	24	74	09-17 04:44	83	0	5	0	0	0	0	0	0 / 0	0	69	100%	0% 100%
	IU_OSG 🚩	17	232	09-17 04:50	19	0	0	0	6	0	124	0	41 / 0	60	1	2%	0% 2%
	LTU CCT 🚩	0	0		0	0	0	0	0	0	0	0	0 / 0	0	0		
	MWT2_IU 🚩	118	1569	09-17 04:50	283	0	0	0	0	0	608	32	166 / 0	762	1	0%	0% 0%
	MWT2_UC 🚩	184	2916	09-17 04:50	14	0	0	0	21	0	761	20	567 / 0	1545	2	0%	0% 0%
	OU OCHEP SWT2 🚩	60	455	09-17 04:50	102	0	0	0	0	0	161	6	89 / 0	219	0	0%	0% 0%

Panda Monitor: Dataset browser

[Configuration](#)[Production](#)[Clouds](#)[DDM](#)[PandaMover](#)[AutoPilot](#)[Sites & Grids](#)[Analysis](#)[Physics data](#)[Usage & Quotas](#)[Plots](#)[ProdDash](#)[DDMDash](#)

[Update](#)

[Panda monitor](#)**DQ2 dataset browser for csc datasets**[Click for help](#)

Dataset lists last updated 157 min ago

[Quick guide, twiki](#)

Jobs - [search](#)
Recent [running](#),
[activated](#), [waiting](#),
[assigned](#), [defined](#),
[finished](#), [failed](#) jobs
Select [analysis](#),
[production](#), [test](#) jobs

Quick search
Job
Dataset
Task
File

Summaries
Blocks: days
Errors: days
Nodes: days
[Daily usage](#)

Tasks - [search](#)
[Generic Task Req](#)
[EvGen Task Req](#)
[CTBsim Task Req](#)
[Task list](#)
[Task browser](#)

Datasets - [search](#)
[Dataset browser](#)
[New datasets](#)
[Abolished MC datasets](#)

Select a project:

Or (the old way) select a dataset category *Counts are totals, exclusive of selections*

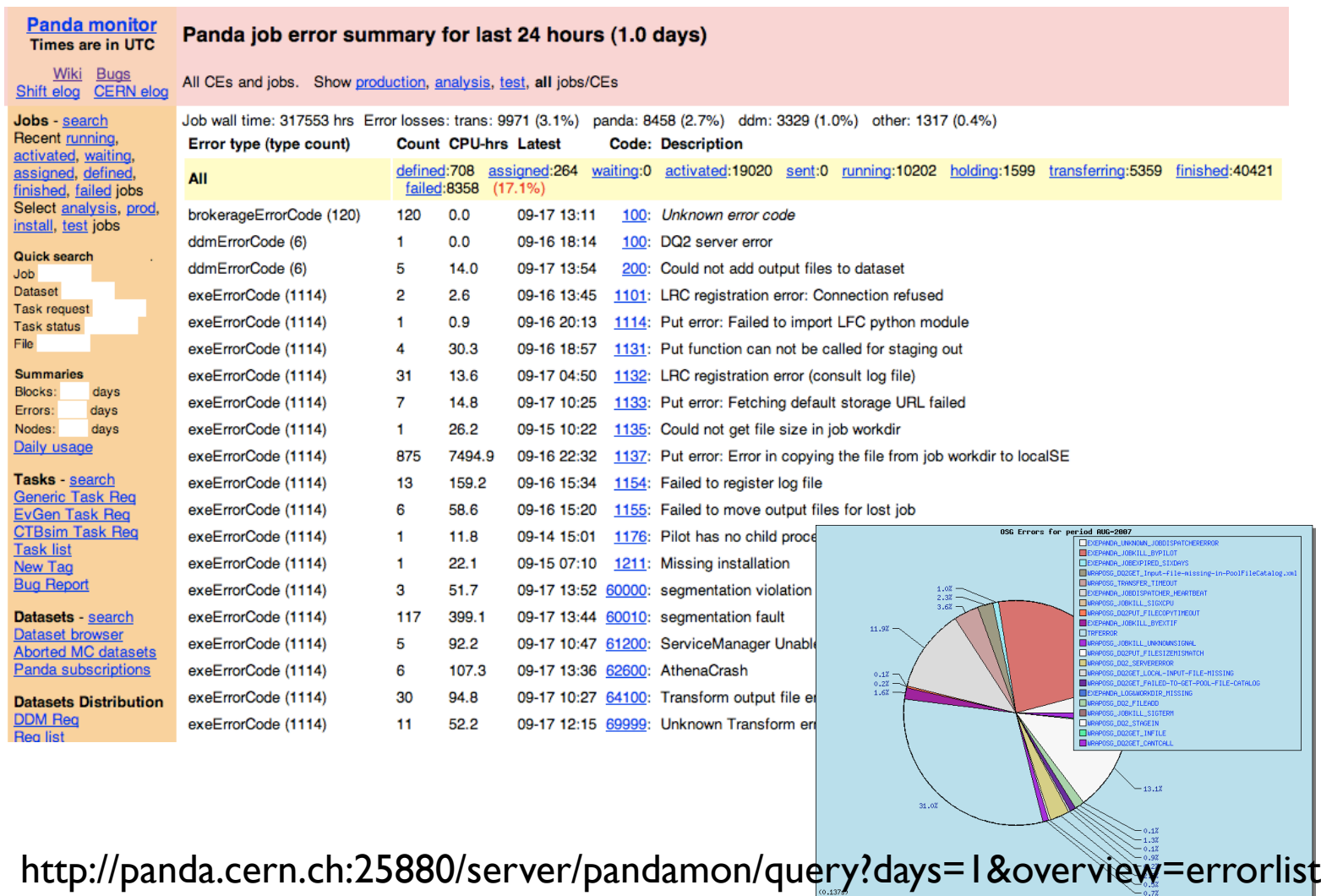
Category	Count	Description
All	147033	All datasets
DBrelease	12	DB release datasets
M3	386	M3 cosmics run
M4	4210	M3 cosmics run
conditions	35	Datasets for conditions data files
csc	4411	Computing system commissioning production
ctb	613	Combined testbeam production
dc2	6	Data Challenge 2 production
larg	53	LAr commissioning
mc	6135	MC validation production
other	71	Everything else
rome	210	Rome physics workshop production
testpanda	6439	Panda test datasets
tile	52	Tilecal commissioning
user	58456	User datasets
validation	642	Validation samples (testIdeal* etc)

Choose a site if you want to restrict dataset listings to site-resident datasets

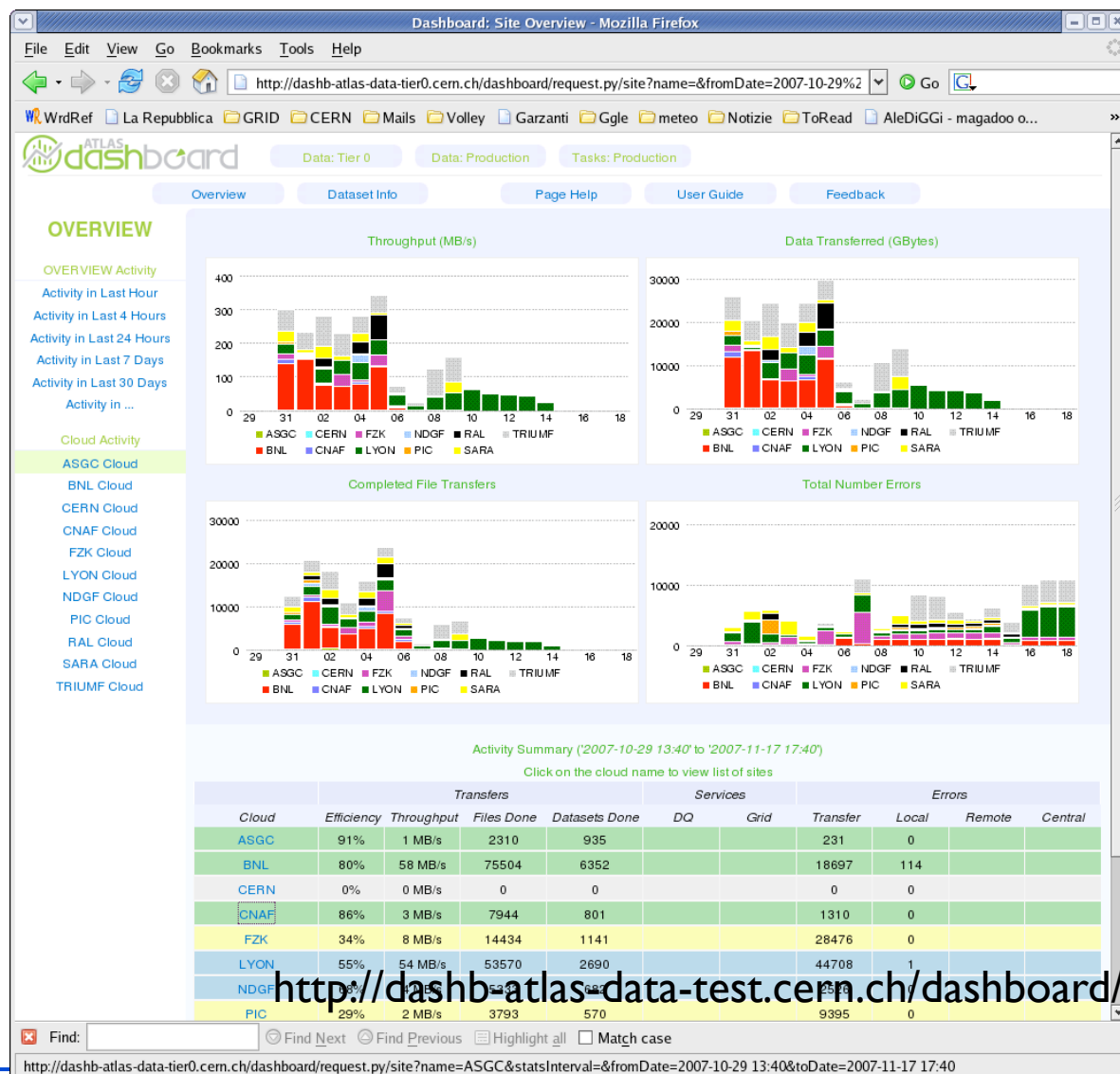
CANADA	CERN	FRANCE	GERMANY	ITALY	NDGF	NL	SPAIN	TAIWAN	UK
ALBERTA	CERNCAF	FRTIER2S	CSCS	CNAF	IJST2	IHEP	IFAE	ASGC	RAL
MCGILL	CERNPROD	LPC	CYF	CNAFDISK	NDGFDISK	ITEP	IFIC	ASGCDISK	Rutherford
MONTREAL	TIER0DISK	LAL	DESY-HH	CNAFTAPE	NDGFT1	JINR	IFICDISK	ASGCDISK_V2	Rutherford
SFU	TIER0TAPE	SACLAY	DESY-ZN	CNAFTEST	NDGFT1DISK	NIKHEF	IFICTAPE	ASGCTAPE	SLAC
TORONTO	LBNL	FZK	INFN	NDGFT1TAPE	PNPI	LIP-COIMBRA	ASGCTAPE_V2	JINR	SLAC

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

Panda Monitor: error reporting



DDM Dashboard: overview



?

!

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

