

# The ATLAS Experiment on the Grid

MWGS08 - September 18, 2008 - Chicago

Marco Mambelli - University of Chicago

[marco@hep.uchicago.edu](mailto:marco@hep.uchicago.edu)

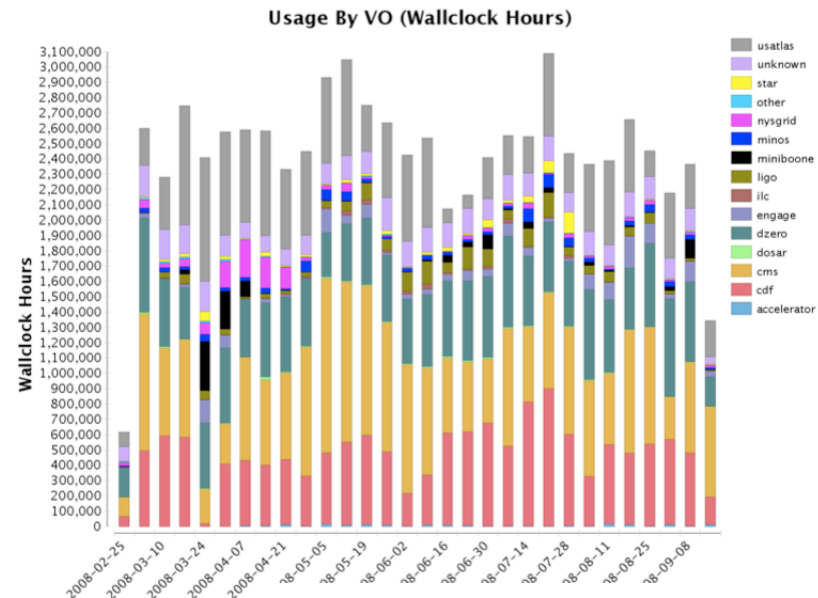
# Outline

---

- ▶ The ATLAS VO
- ▶ ATLAS and the LHC
- ▶ Computing model
- ▶ Panda
- ▶ Panda monitor and ARDA Dashboard
- ▶ Distributed Data Management
- ▶ Operation Shifts
- ▶ Production and Analysis overview

# 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](http://atlas.web.cern.ch/Atlas/ATLASreg_form.pdf)



## ATLAS REGISTRATION FORM



Please complete all sections and return this form to the  
ATLAS SECRETARIAT, Bldg 40-4-D01, fax+41227678350

\*SURNAME:

\*FIRST NAME:

\*DATE OF BIRTH (DAY/MONTH/YEAR):

\*SEX: M/F

\*MARITAL STATUS: Single / Married / Divorced

\*NATIONALITY:

\*CERN IDENTIFICATION NUMBER:

\*CERN PHONE NUMBER:

\*CERN OFFICE:

\*PRIVATE HOME ADDRESS AND PHONE NUMBER:

\*HOME ADDRESS WHEN AT CERN (e.g. CERN HOSTEL MEYRIN or ST GENIS-POUILLY):

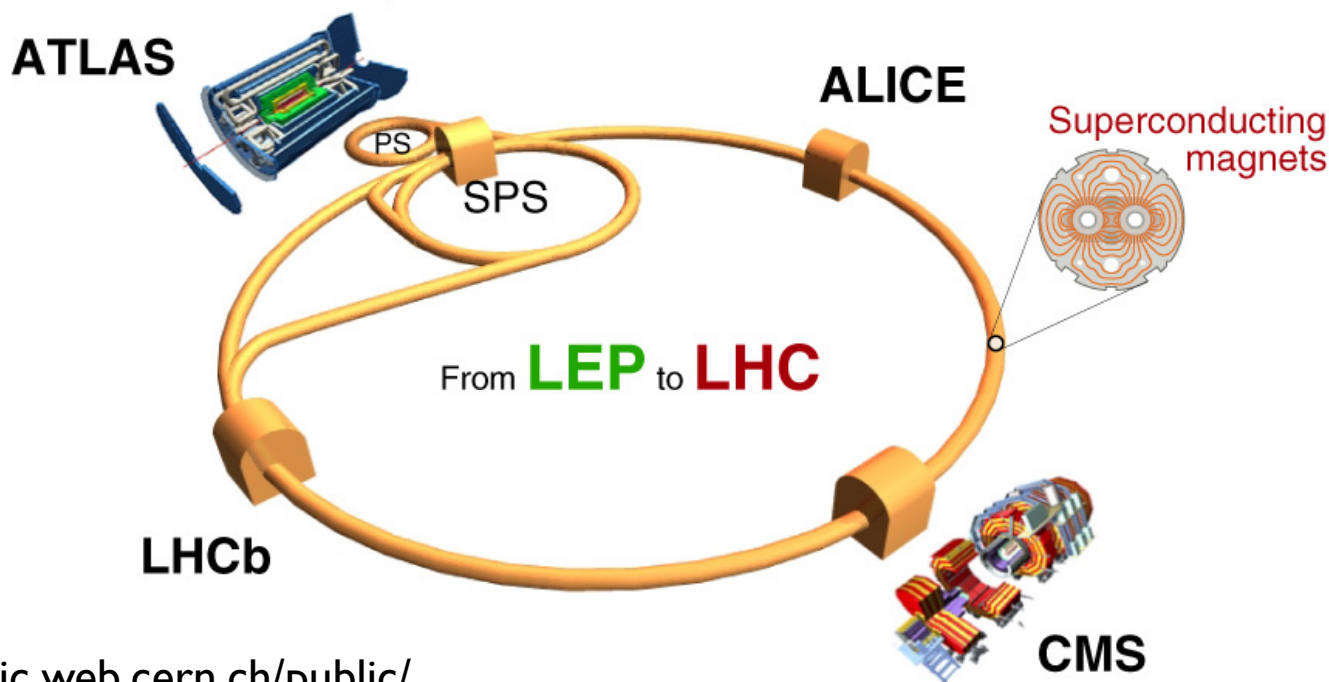
\*HOME INSTITUTE NAME AND ADDRESS:

\*HOME INSTITUTE OFFICE TELEPHONE AND FAX NO. (DIRECT):

\*EMAIL ADDRESS:

\*SUBDETECTOR OR DOMAIN OF ACTIVITY: (you may cross several)

# LHC experiment at CERN

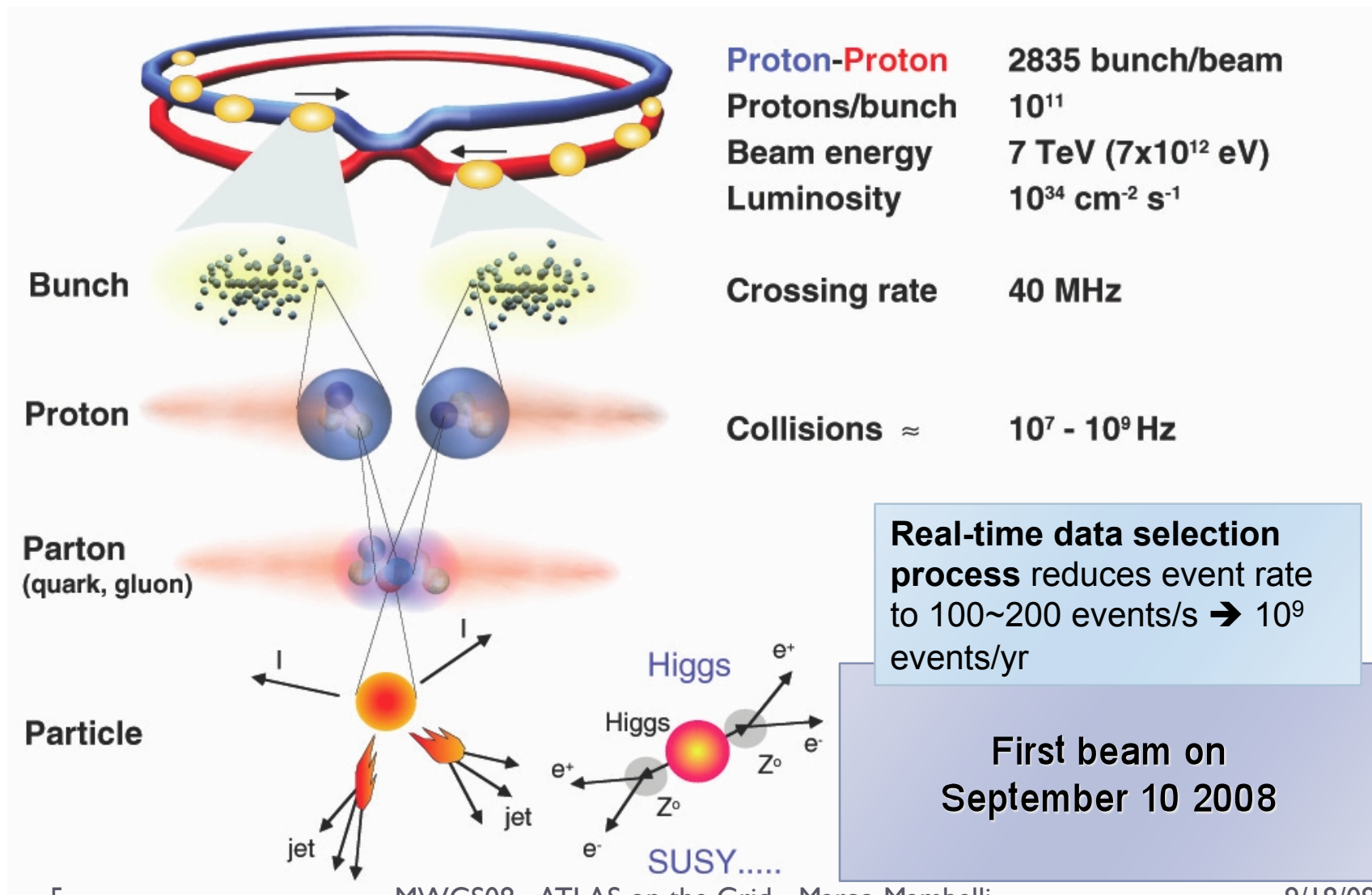


<http://public.web.cern.ch/public/>

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

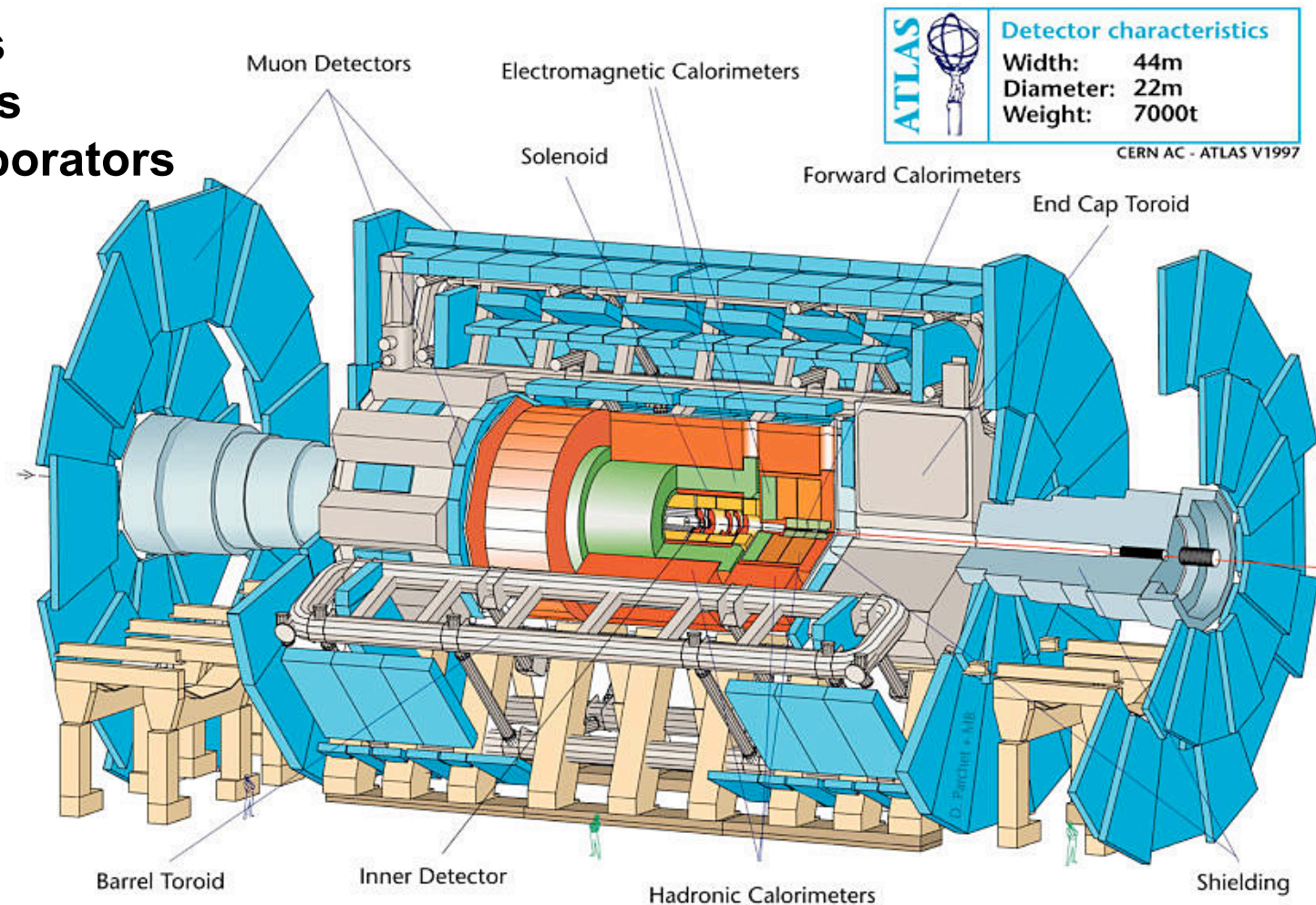
<http://www.youtube.com/watch?v=j50ZssEojtM>

# proton-proton collision



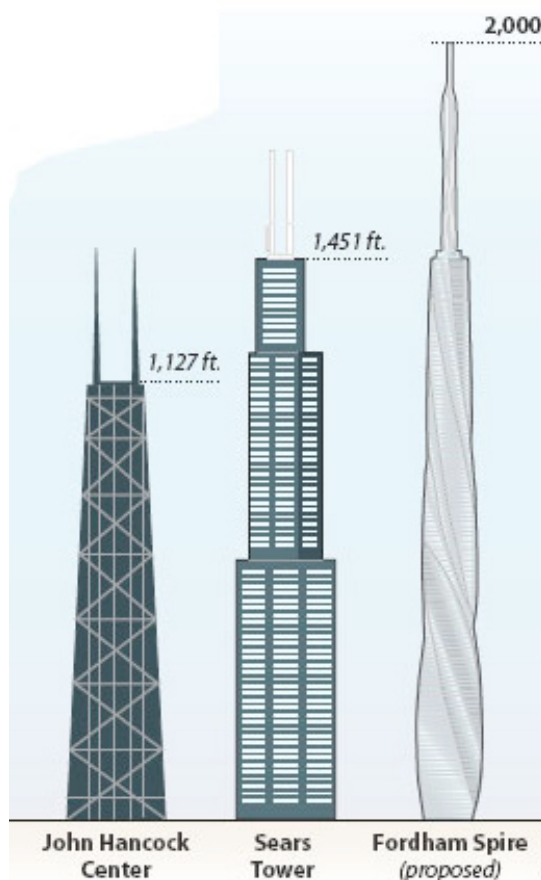
# The ATLAS experiment

**37 Countries**  
**167 Institutes**  
**~2000 Collaborators**





# Data produced



- ▶  $2 \times 10^9$  events/yr and 1 event  $\sim 1.6$  MB
- ▶ ATLAS will record about 3.2 Petabytes of data per year (3.2 million GB)
- ▶ plus 2-3 times as much simulated data
- ▶ invites comparisons like “if we wrote one year’s data on DVDs it would make a stack taller than the Sears Tower (1451 feet = 442.3m)”

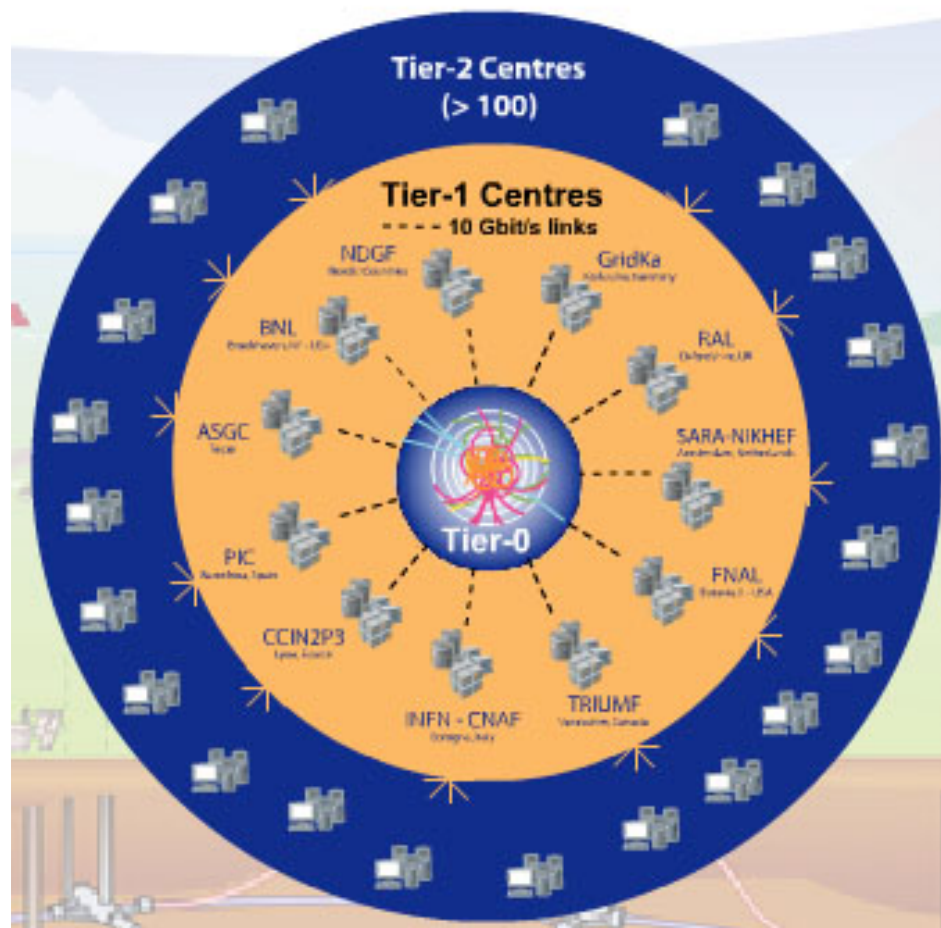
***DVD thickness: 1.2 mm***

***DVD capacity: 8.5 GB (1-side, 2-layer)***

***3.2 PB/8.5 GB = 376470 discs = 452 m***

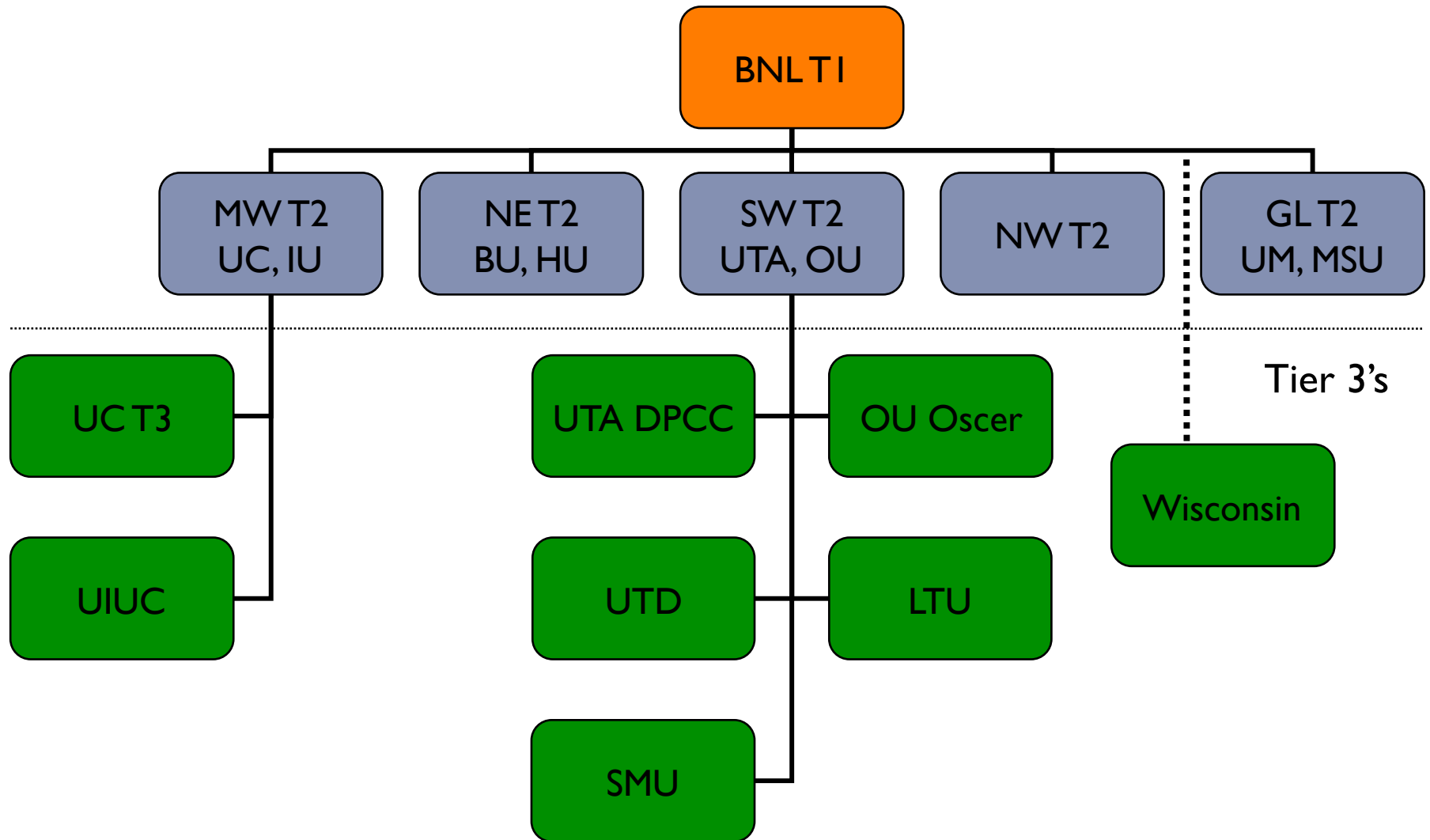
# LHC Computing Grid

- ▶ 3 major Grid Environments
  - ▶ WLCG/EGEE (Enabling Grids for E-science)
  - ▶ OSG (Open Science Grid)
  - ▶ NG (Nordugrid)
- ▶ Grids have differences in
  - ▶ Middle-ware
  - ▶ Replica catalogs to store data
  - ▶ Software tools to submit jobs





# Computing hierarchy: the US Cloud

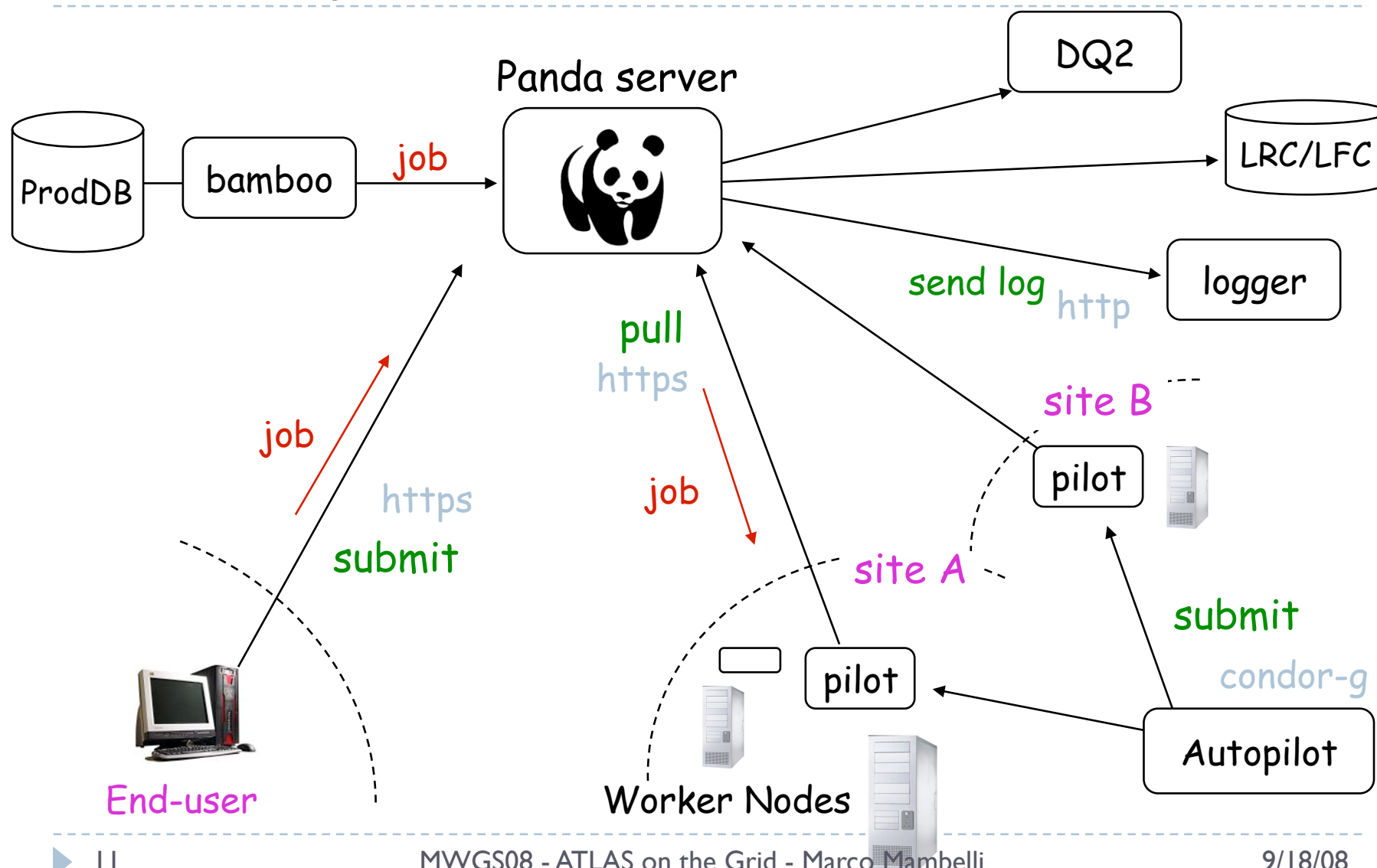


# PANDA

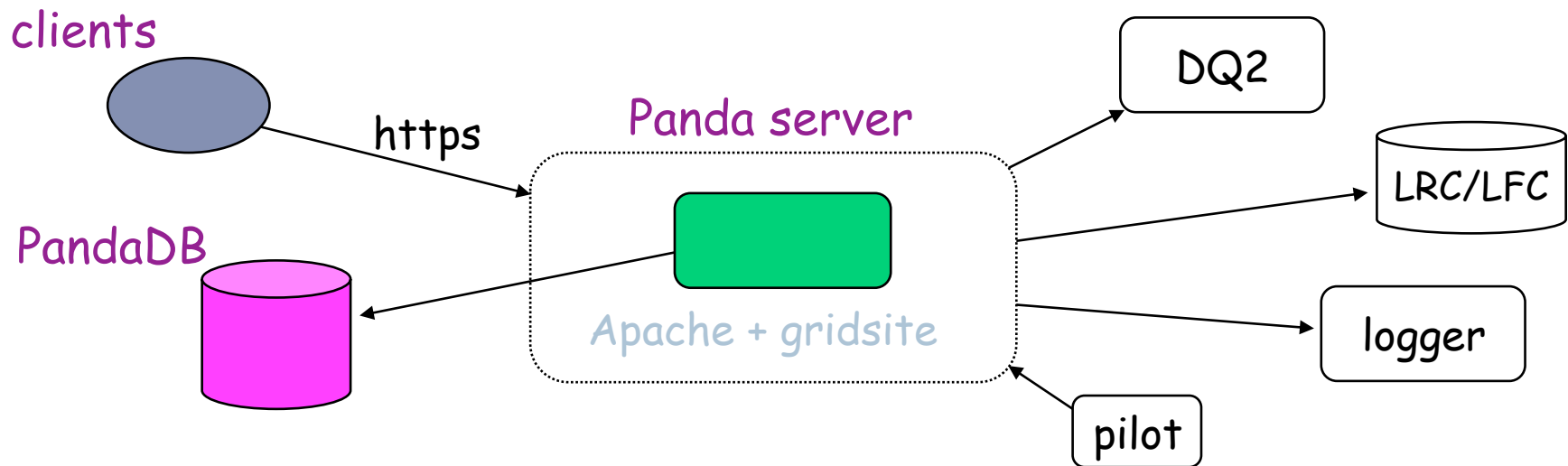
---

- ▶ **PANDA = Production AND Distributed Analysis system**
  - ▶ Designed for analysis as well as production
  - ▶ Project started Aug 2005, prototype Sep 2005, production Dec 2005
  - ▶ 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 CHARMM**

# Panda System

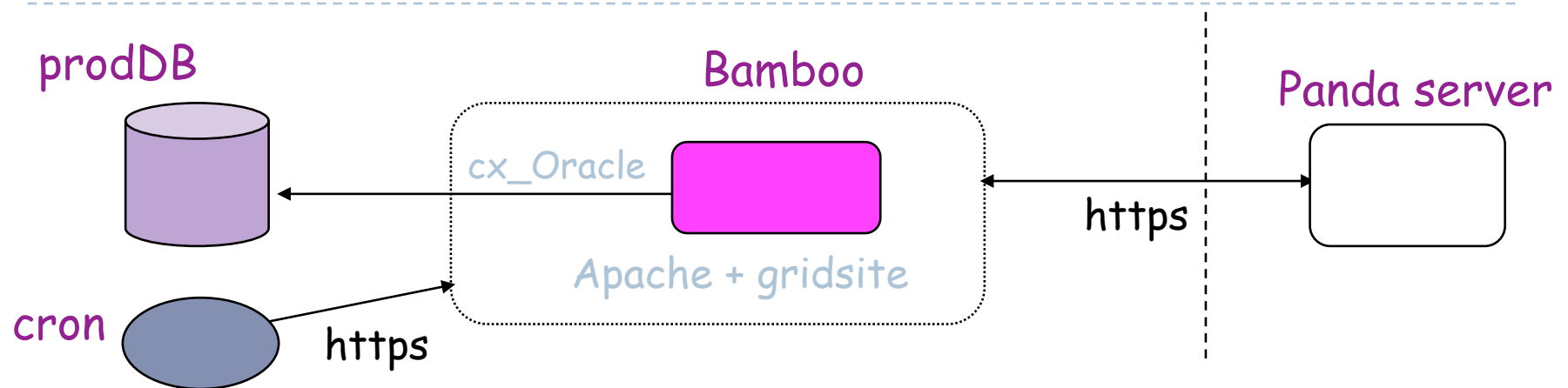


# Panda Server



- ▶ 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

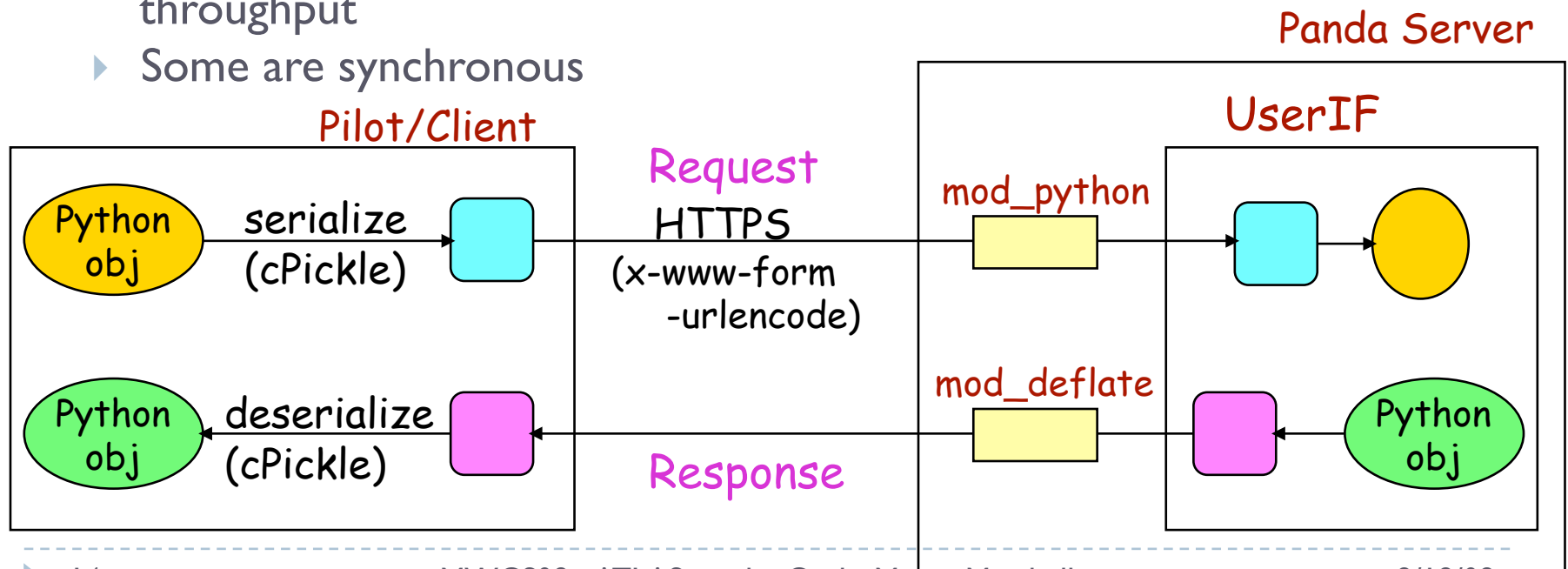
# Bamboo



- ▶ 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

# Client-Server Communication

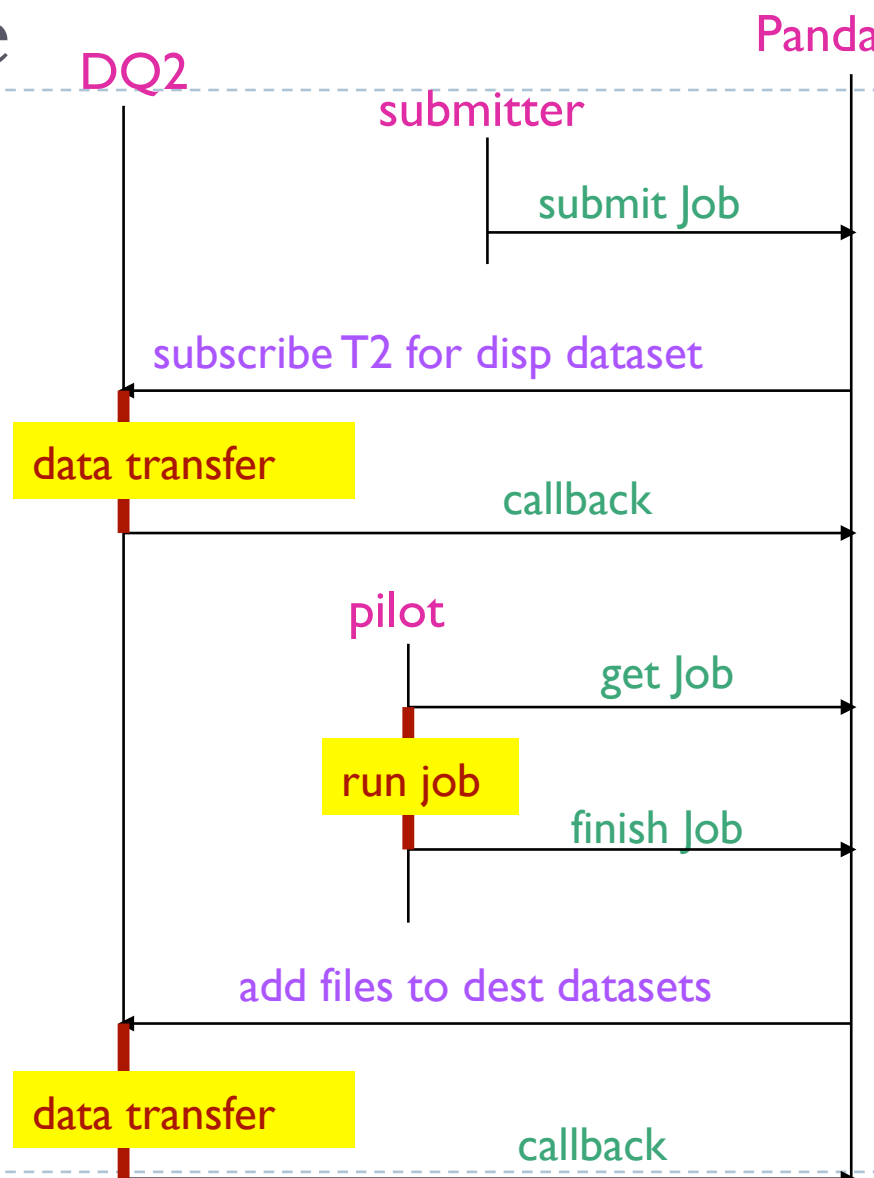
- ▶ HTTP/S-based communication (curl+grid proxy+python)
- ▶ GSI authentication via `mod_gridsite`
- ▶ Most of communications are asynchronous
  - ▶ Panda server runs python threads as soon as it receives HTTP requests, and then sends responses back immediately. Threads do heavy procedures (e.g., DB access) in background → better throughput
  - ▶ Some are synchronous





# 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 T2s and aggregate output files to T1
- ▶ 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 SE and register them to Local Replica Catalogue

# Pilot vs ATLAS Job

---

## Pilot

- ▶ Submitted by factories
  - ▶ autopilot
  - ▶ 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

## 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://pandamon.usatlas.bnl.gov: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](#)

[Bugs](#)

[CERN 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](#)  
[CTBsim Task Req](#)  
[Task list](#)  
[New Tag](#)  
[Bug Report](#)

Datasets - [search](#)  
[Dataset browser](#)  
[Aborted MC datasets](#)  
[Panda subscriptions](#)

Datasets Distribution  
[DDM Req](#)  
[Req list](#)  
[AODs](#)

**Panda Production Operations Dashboard**

Panda shift [guide](#) [calendar](#) [mailing list](#)  
ADCoS [twiki](#) [calendar](#) [mailing list](#)  
Production task support [mailing list](#)

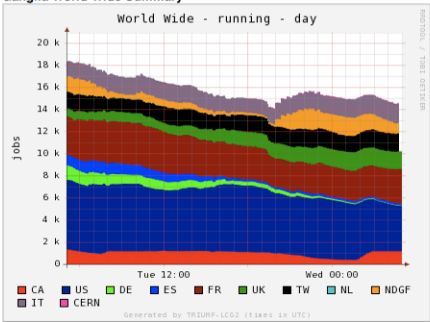
[Click for help](#)

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-Isf	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		
SLACXRD-Isf		
SWT2_CPB-pbs		
UC_ATLAS_MWT2o		
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_MC		

	IT	NL	UK	US	TW
Nodes	566	2	870	1526	468
Jobs	8372	228	8439	13514	7643
Latest	09-17 04:50	09-16 23:24	09-17 04:50	09-17 04:50	09-17 04:50
Pilots (3hrs)	255	688	804	1275	241
defined	0	0	0	0	0
assigned	15	227	0	5	10
waiting	0	0	0	0	0
activated	3417	0	2140	152	3241
sent	0	0	0	0	1
running	1776	0	1599	4113	1624
holding	50	0	46	230	29
transferring	443 / 0	0 / 0	1236 / 0	2526 / 12	267 / 0
finished	2003	0	2905	6263	2444
failed tot	668	1	513	225	27
trf other	25% 0% 25%	100% 0% 100%	15% 0% 15%	3% 0% 3%	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	69 / 0	219	0	0% 0% 0%

http://pandamon.usatlas.bnl.gov:25880/server/pandamon/query?dash=prod

19

MWGS08 - ATLAS on the Grid - Marco Mambelli

9/18/08

# Panda Monitor: Dataset browser

[Configuration](#)  
[Update](#)  
[Panda monitor](#)  
[Quick guide](#), [twiki](#)

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

[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](#)  
[Aborted MC datasets](#)

## DQ2 dataset browser for csc datasets

[Click for help](#)

Dataset lists last updated 157 min ago

Select a project:

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

Category	Count	Description
<a href="#">All</a>	147033	All datasets
<a href="#">DBrelease</a>	12	DB release datasets
<a href="#">M3</a>	386	M3 cosmics run
<a href="#">M4</a>	4210	M3 cosmics run
<a href="#">conditions</a>	35	Datasets for conditions data files
<a href="#">csc</a>	4411	Computing system commissioning production
<a href="#">ctb</a>	613	Combined testbeam production
<a href="#">dc2</a>	6	Data Challenge 2 production
<a href="#">larg</a>	53	LAr commissioning
<a href="#">mc</a>	6135	MC validation production
<a href="#">other</a>	71	Everything else
<a href="#">rome</a>	210	Rome physics workshop production
<a href="#">testpanda</a>	6439	Panda test datasets
<a href="#">tile</a>	52	Tilecal commissioning
<a href="#">user</a>	58456	User datasets
<a href="#">validation</a>	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
<a href="#">ALBERTA</a>	<a href="#">CERNCAF</a>	<a href="#">FRTIER2S</a>	<a href="#">CSCS</a>	<a href="#">CNAF</a>	<a href="#">IJST2</a>	<a href="#">IHEP</a>	<a href="#">IFAE</a>	<a href="#">ASGC</a>	<a href="#">Rutherford</a>
<a href="#">MCGILL</a>	<a href="#">CERNPROD</a>	<a href="#">LPC</a>	<a href="#">CYF</a>	<a href="#">CNAFDISK</a>	<a href="#">NDGFDISK</a>	<a href="#">ITEP</a>	<a href="#">IFIC</a>	<a href="#">ASGCDISK</a>	<a href="#">Rutherford</a>
<a href="#">MONTREAL</a>	<a href="#">TIER0DISK</a>	<a href="#">LAL</a>	<a href="#">DESY-HH</a>	<a href="#">CNAFTAPE</a>	<a href="#">NDGFT1</a>	<a href="#">JINR</a>	<a href="#">IFICDISK</a>	<a href="#">ASGCDISK V2</a>	<a href="#">Rutherford</a>
<a href="#">SFU</a>	<a href="#">TIER0TAPE</a>	<a href="#">SACLAY</a>	<a href="#">DESY-ZN</a>	<a href="#">CNAFTTEST</a>	<a href="#">NDGFT1DISK</a>	<a href="#">NIKHEF</a>	<a href="#">IFICTAPE</a>	<a href="#">ASGCTAPE</a>	<a href="#">Rutherford</a>
<a href="#">TORONTO</a>	<a href="#">LPNHE</a>	<a href="#">FZK</a>	<a href="#">LNF</a>	<a href="#">NDGFT1TAPE</a>	<a href="#">PNPI</a>	<a href="#">LIP-COIMBRA</a>	<a href="#">ASGCTAPE V2</a>	<a href="#">Rutherford</a>	<a href="#">Rutherford</a>

<http://pandamon.usatlas.bnl.gov:25880/server/pandamon/query?overview=dslist>



# Panda Monitor: error reporting

[Panda monitor](#)  
Times are in UTC

[Wiki](#) [Bugs](#)  
[Shift elog](#) [CERN 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](#)  
[CTBsim Task Req](#)  
[Task list](#)  
[New Tag](#)  
[Bug Report](#)

[Datasets - search](#)  
[Dataset browser](#)  
[Aborted MC datasets](#)  
[Panda subscriptions](#)  
  
[Datasets Distribution](#)  
[DDM Req](#)  
[Req list](#)

**Panda job error summary for last 24 hours (1.0 days)**  
All CEs and jobs. Show [production](#), [analysis](#), [test](#), all jobs/CEs

Job wall time: 317553 hrs Error losses: trans: 9971 (3.1%) panda: 8458 (2.7%) ddm: 3329 (1.0%) other: 1317 (0.4%)

Error type (type count)	Count	CPU-hrs	Latest	Code: Description
<b>All</b>	<a href="#">defined</a> :708 <a href="#">failed</a> :8358	<a href="#">assigned</a> :264 (17.1%)	<a href="#">waiting</a> :0	<a href="#">activated</a> :19020 <a href="#">sent</a> :0 <a href="#">running</a> :10202 <a href="#">holding</a> :1599 <a href="#">transferring</a> :5359 <a href="#">finished</a> :40421
brokerageErrorCode (120)	120	0.0	09-17 13:11	<a href="#">100</a> : Unknown error code
ddmErrorCode (6)	1	0.0	09-16 18:14	<a href="#">100</a> : DQ2 server error
ddmErrorCode (6)	5	14.0	09-17 13:54	<a href="#">200</a> : Could not add output files to dataset
exeErrorCode (1114)	2	2.6	09-16 13:45	<a href="#">1101</a> : LRC registration error: Connection refused
exeErrorCode (1114)	1	0.9	09-16 20:13	<a href="#">1114</a> : Put error: Failed to import LFC python module
exeErrorCode (1114)	4	30.3	09-16 18:57	<a href="#">1131</a> : Put function can not be called for staging out
exeErrorCode (1114)	31	13.6	09-17 04:50	<a href="#">1132</a> : LRC registration error (consult log file)
exeErrorCode (1114)	7	14.8	09-17 10:25	<a href="#">1133</a> : Put error: Fetching default storage URL failed
exeErrorCode (1114)	1	26.2	09-15 10:22	<a href="#">1135</a> : Could not get file size in job workdir
exeErrorCode (1114)	875	7494.9	09-16 22:32	<a href="#">1137</a> : Put error: Error in copying the file from job workdir to localSE
exeErrorCode (1114)	13	159.2	09-16 15:34	<a href="#">1154</a> : Failed to register log file
exeErrorCode (1114)	6	58.6	09-16 15:20	<a href="#">1155</a> : Failed to move output files for lost job
exeErrorCode (1114)	1	11.8	09-14 15:01	<a href="#">1176</a> : Pilot has no child process
exeErrorCode (1114)	1	22.1	09-15 07:10	<a href="#">1211</a> : Missing installation
exeErrorCode (1114)	3	51.7	09-17 13:52	<a href="#">60000</a> : segmentation violation
exeErrorCode (1114)	117	399.1	09-17 13:44	<a href="#">60010</a> : segmentation fault
exeErrorCode (1114)	5	92.2	09-17 10:47	<a href="#">61200</a> : ServiceManager Unavailable
exeErrorCode (1114)	6	107.3	09-17 13:36	<a href="#">62600</a> : AthenaCrash
exeErrorCode (1114)	30	94.8	09-17 10:27	<a href="#">64100</a> : Transform output file error
exeErrorCode (1114)	11	52.2	09-17 12:15	<a href="#">69999</a> : Unknown Transform error

OSG Errors for period 09-16-2007

Error Code	Description	Percentage
60010	segmentation fault	31.02%
60000	segmentation violation	11.92%
64100	Transform output file error	13.12%
62600	AthenaCrash	1.02%
61200	ServiceManager Unavailable	2.32%
69999	Unknown Transform error	3.62%
1137	Put error: Error in copying the file from job workdir to localSE	0.12%
1135	Could not get file size in job workdir	0.22%
1133	Put error: Fetching default storage URL failed	1.62%
1132	LRC registration error (consult log file)	0.12%
1131	Put function can not be called for staging out	0.12%
1114	Put error: Failed to import LFC python module	0.12%
1101	LRC registration error: Connection refused	0.12%
200	Could not add output files to dataset	0.12%
100	Unknown error code	0.12%

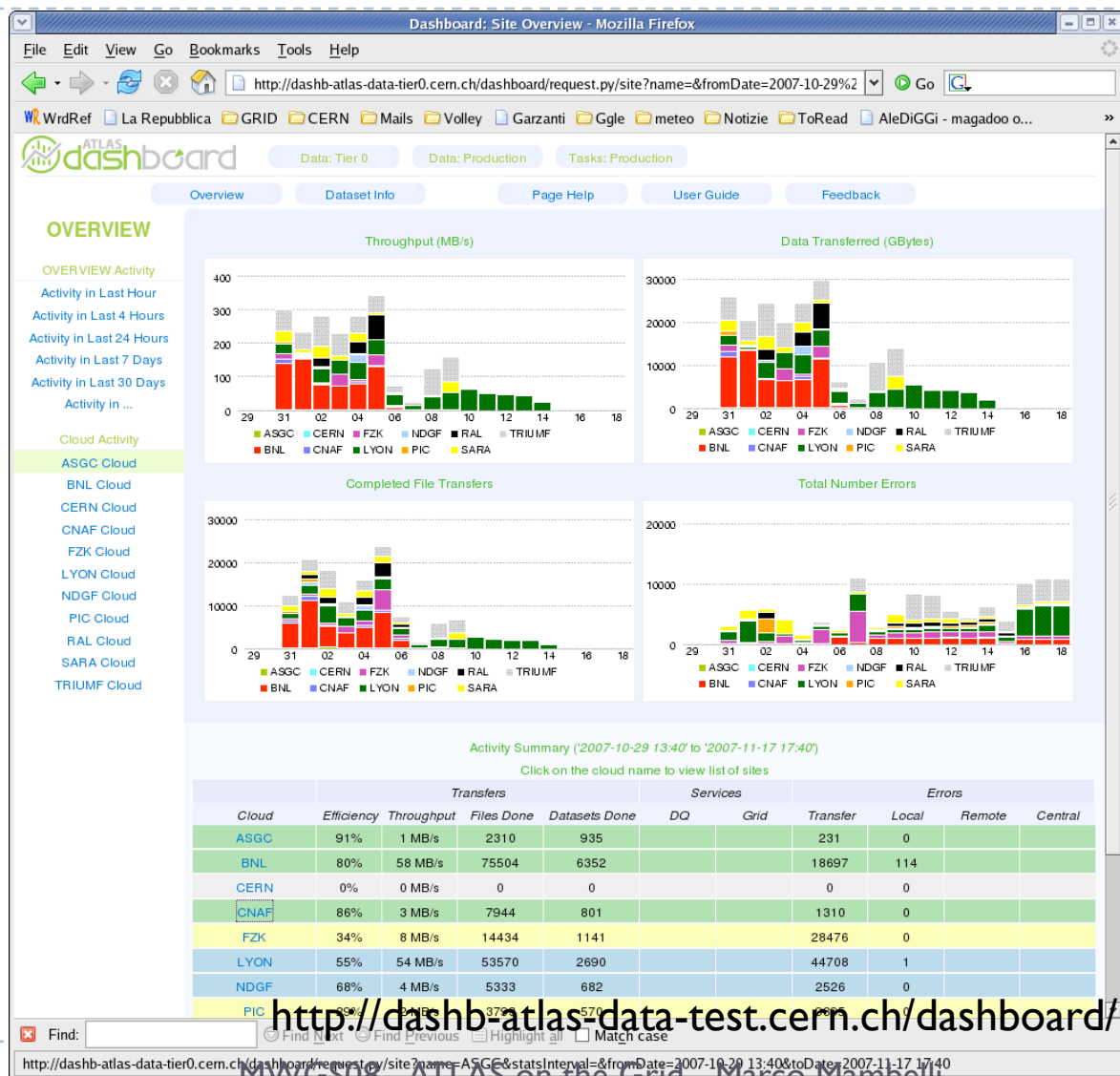
<http://pandamon.usatlas.bnl.gov:25880/server/pandamon/query?days=1&overview=errorlist>

21

MWGS08 - ATLAS on the Grid - Marco Mambelli

9/18/08

# DDM Dashboard: overview



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

# DDM Dashboard: Dataset detail

Dashboard: Datasets @ CNAFDISK - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://dashb-atlas-data-tier0.cern.ch/dashboard/request.py/dataset?site=CNAFDISK&orderField=

WrdRef La Repubblica GRID CERN Mails Volley Garzanti Ggle meteo Notizie ToRead AleDiGGI - magadoo o...

ATLAS dashboard

Data: Tier 0 Data: Production Tasks: Production

Overview Dataset Info Page Help User Guide Feedback

**CNAFDISK**

Dataset List  
Most Recent Events  
Most Recent Errors

All Cloud Sites  
CNAFDISK  
CNAFTAPE  
CNAFTTEST  
LNF  
MILANO  
NAPOLI  
ROMA1

Expected Date Format: 'YYYY-MM-DD HH:MM:SS'

☒ QUEUED (36) ☒ INCOMPLETE (11) From: 2007-10-29 13:40:00 To: 2007-11-17 17:40:00  
☒ COMPLETE (643) ☒ CANCELED (0) Limit: 100 Reset Refresh  
☒ BROKEN (0)

Click on the 'Last Update' cell of each row to see dataset details

100 to 200 of 690

Last Update	Dataset Name	State
2007-11-14 07:18:31	M5.0029563.DefaultLL1TT-b00000100.ESD.v13003013	INCOMPLETE
2007-11-13 20:10:48	M5.0029549.DefaultLL1TT-b00000001.CBNT.v13003013	COMPLETE
2007-11-09 02:52:16	M5.0029147.DefaultLL1TT-b00000001.ESD.v13003013	COMPLETE
2007-11-09 00:16:00	M5.0029117.DefaultLL1TT-b00001000.CBNT.v13003013	COMPLETE
2007-11-08 23:40:03	M5.0029549.DefaultLL1TT-b00000010.CBNT.v13003013	COMPLETE
2007-11-08 23:26:11	M5.0029113.DefaultLL1TT-b00000001.ESD.v13003013	COMPLETE
2007-11-08 23:10:30	M5.0029104.DefaultLL1TT-b11110001.CBNT.v13003013	COMPLETE
2007-11-08 22:57:51	M5.0029576.DefaultLL1TT-b00000100.CBNT.v13003013	COMPLETE
2007-11-08 22:32:55	M5.0029414.DefaultLL1TT-b00000001.CBNT.v13003013	COMPLETE
2007-11-08 22:29:50	M5.0029425.DefaultLL1TT-b00000001.CBNT.v13003013	COMPLETE
2007-11-08 22:29:06	M5.0029502.physics.cosmics.CBNT.v13003013	COMPLETE
2007-11-08 22:14:38	M5.0029251.DefaultLL1TT-b00000100.CBNT.v13003013	COMPLETE
2007-11-08 22:01:52	M5.0029104.DefaultLL1TT-b11101111.CBNT.v13003013	COMPLETE
2007-11-08 21:57:44	M5.0029563.DefaultLL1TT-b11100010.CBNT.v13003013	COMPLETE
2007-11-08 21:56:34	M5.0029108.DefaultLL1TT-b00001000.ESD.v13003013	COMPLETE
2007-11-08 21:54:15	M5.0029545.DefaultLL1TT-b11100011.CBNT.v13003013	COMPLETE
2007-11-08 21:51:53	M5.0029502.physics.cosmics.ESD.v13003013	COMPLETE
2007-11-08 21:47:35	M5.0029549.DefaultLL1TT-b11100000.ESD.v13003013	COMPLETE
2007-11-08 21:44:12	M5.0029147.DefaultLL1TT-b00000001.CBNT.v13003013	COMPLETE
2007-11-08 21:36:13	M5.0029563.DefaultLL1TT-b00000001.CBNT.v13003013	COMPLETE

Find: Find Next Find Previous Highlight all Match case

Done

http://dashb-atlas-data-test.cern.ch/dashboard/request.py/dataset?site=AGLT2\_DATADISK

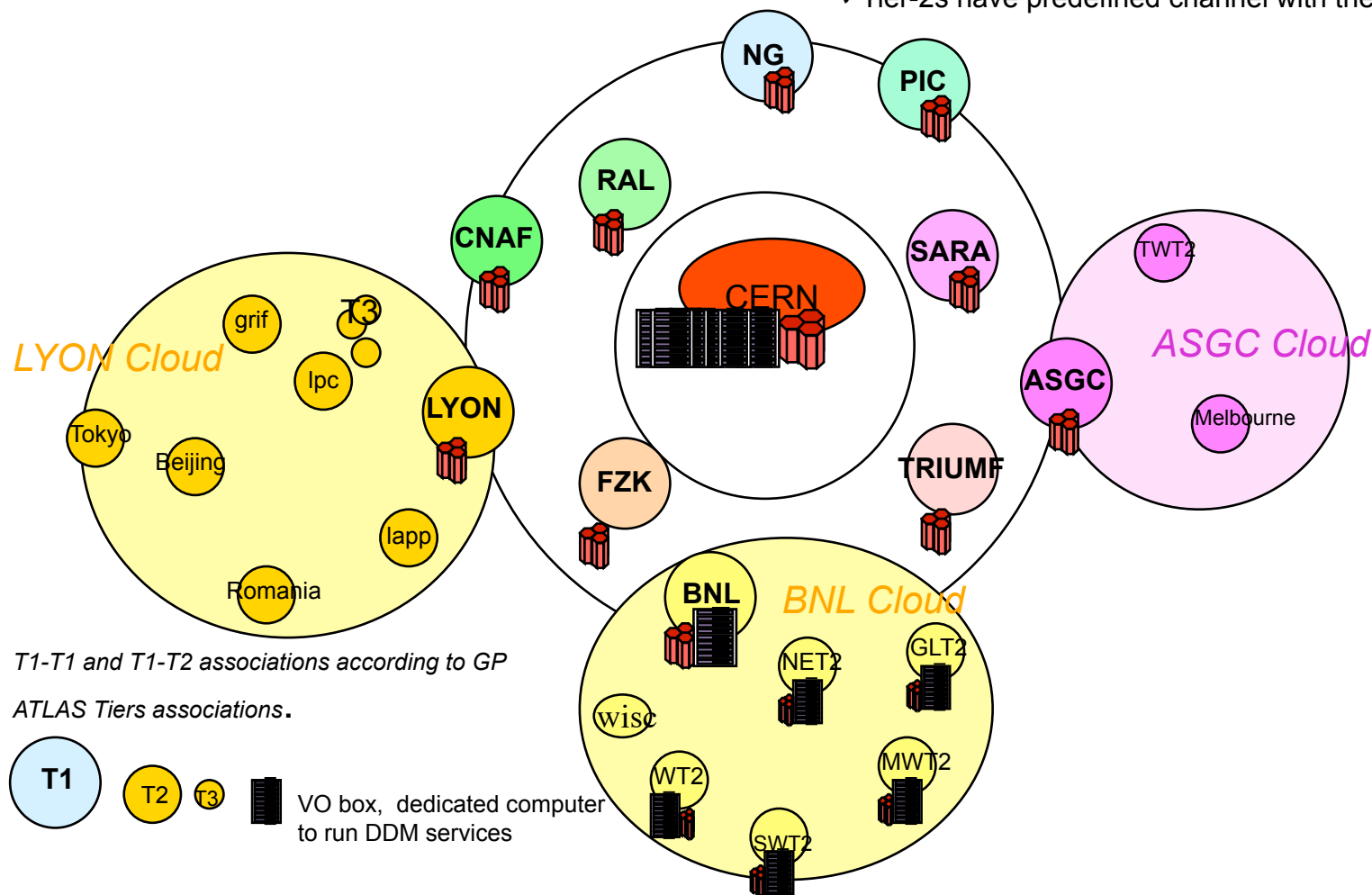
# ATLAS Data Management Software: Don Quijote

---

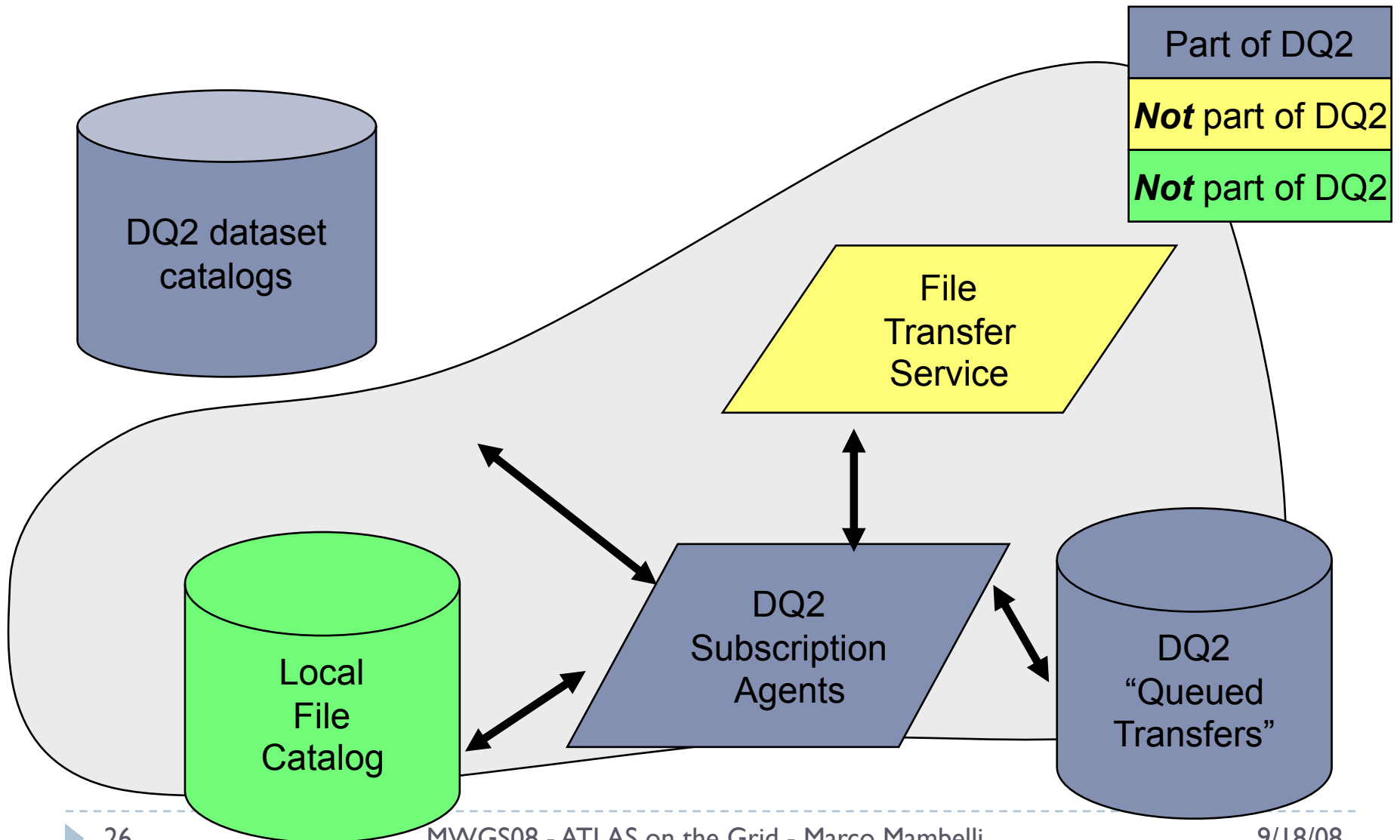
- ▶ DQ2 - Don Quijote (second release)
- ▶ DQ2 is built on top of Grid data transfer tools
  - ▶ Dataset based approach
    - ▶ Datasets : an aggregation of files plus associated DDM metadata
      - Open: latest version is open, so new files may be added or existing files may be removed
      - Closed: latest version is closed, no changes can be done, but a new version may be created
      - Frozen: latest version is closed, No versions may be added
    - ▶ Datasets is a unit of storage and replication
    - ▶ Automatic data transfer mechanisms using distributed site services
      - Subscription system
      - Notification system
  - ▶ Hierarchical structure
    - ▶ Sites subdivided in 'clouds'
    - ▶ Topology in TiersOfATLASCache.py

# Cloud Structure

- ✓ All Tier-1s have predefined (software) channel with CERN and with each other.
- ✓ Tier-2s are associated with one Tier-1 and form the cloud
- ✓ Tier-2s have predefined channel with the parent Tier-1 only.



# Simplified DDM Schema and interactions





# Subscriptions

---

- ▶ **Subscription**
  - ▶ Request for the full replication of a dataset (or dataset version) at a given site
- ▶ Requests are collected by the centralized subscription catalog
- ▶ And are then served by a site of agents – the *site services*
- ▶ **Subscription on a dataset version**
  - ▶ One time only replication
- ▶ **Subscription on a dataset**
  - ▶ Replication triggered on every new version detected
  - ▶ Subscription closed when dataset is frozen

# Site Services

---

- ▶ Agent based framework
- ▶ **Goal:** Satisfy subscriptions
- ▶ Each agent serves a specific part of a request
  - ▶ *Fetcher*: fetches up new subscription from the subscription catalog
  - ▶ *Subscription Resolver*: checks if subscription is still active, new dataset versions, new files to transfer, ...
  - ▶ *Splitter*: Create smaller chunks from the initial requests, identifies files requiring transfer
  - ▶ *Replica Resolver*: Selects a valid replica to use as source
  - ▶ *Partitioner*: Creates chunks of files to be submitted as a single request to the FTS
  - ▶ *Submitter/PendingHandler*: Submit/manage the FTS requests
  - ▶ *Verifier*: Check validity of file at destination
  - ▶ *Replica Register*: Registers new replica in the local replica catalog
  - ▶ ...

# Interaction with the grid middleware

---

- ▶ **File Transfer Services (FTS)**
  - ▶ One deployed per Tier0 / TierI (matches typical site services deployment)
  - ▶ Triggers the third party transfer by contacting the SRM or Gridftp servers, needs to be constantly monitored
- ▶ **LCG File Catalog (LFC)**
  - ▶ One deployed per Tier0 / TierI (matches typical site services deployment)
  - ▶ Keeps track of local file replicas at a site
  - ▶ Main source of replica information by the site services
  - ▶ Currently is deployed as LRC (alternative version)
- ▶ **Storage Resource Manager (SRM)**
  - ▶ Extra level of abstraction on top of file transfers (e.g. gridftp)
  - ▶ Allows operations like pinning and space reservation

# ADC Operations Shifts

---

- ▶ **ATLAS Distributed Computing Operations Shifts (ADCoS)**
  - ▶ World-wide shifts
  - ▶ 24h coverage (Asia, EU, USA)
  - ▶ To monitor all ATLAS distributed computing resources
  - ▶ To provide Quality of Service (QoS) for all data processing
  - ▶ Shifters receive official ATLAS service credit (OTSMoU)

# Typical Shift Plan

---

- ▶ Browse recent shift history
- ▶ Check performance of all sites
  - ▶ File tickets for new issues
  - ▶ Continue interactions about old issues
- ▶ Check status of current tasks
  - ▶ Check all central processing tasks
  - ▶ Monitor analysis flow (not individual tasks)
  - ▶ Overall data movement
- ▶ File software (validation) bug reports
- ▶ Check Panda, DDM health
- ▶ Maintain elog of shift activities

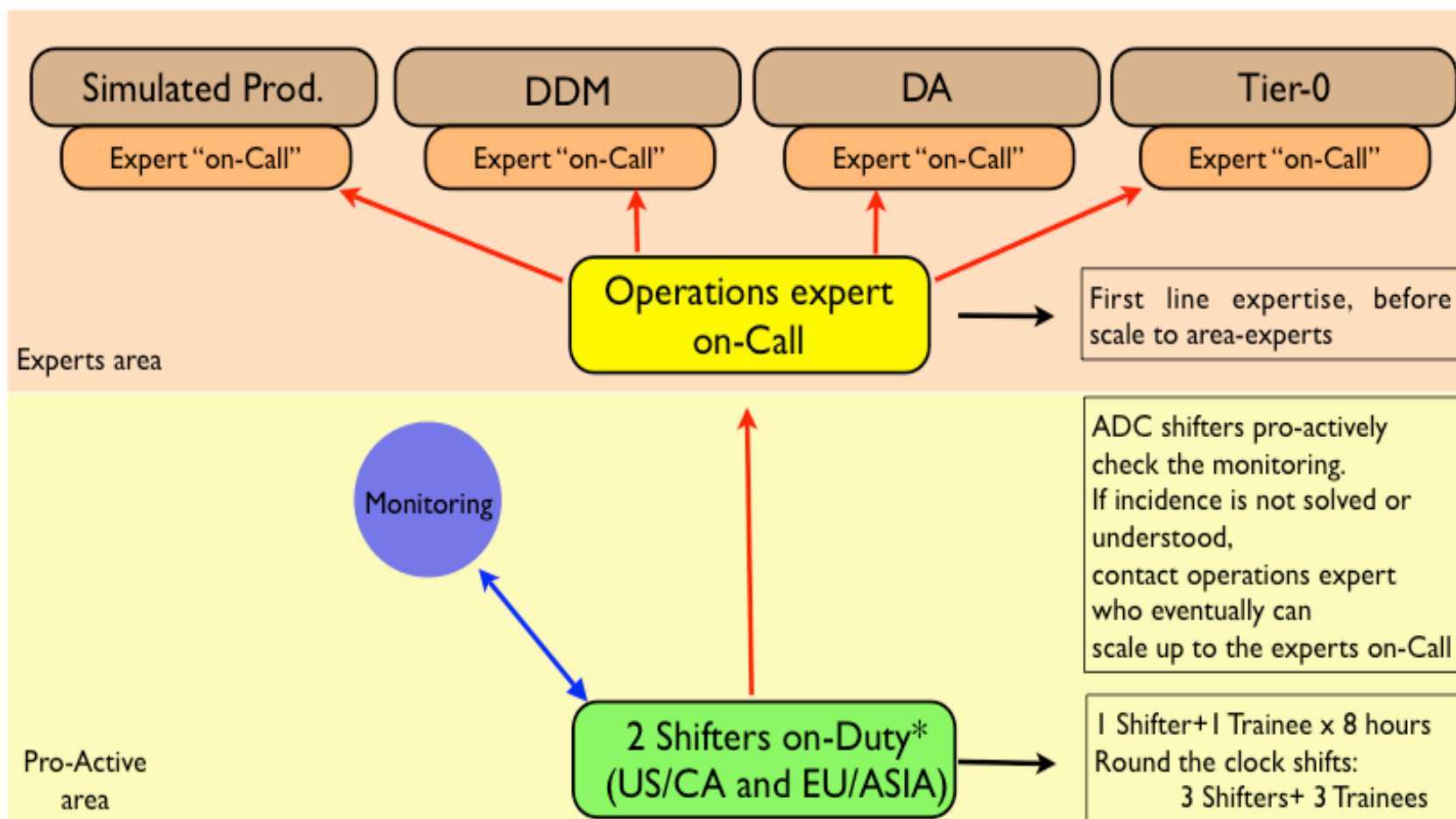
# Shift Structure

---

- ▶ **Shifter on call**
  - ▶ Two consecutive days
  - ▶ Monitor – escalate – follow up
  - ▶ Basic manual interventions (site – on/off)
- ▶ **Expert on call**
  - ▶ One week duration
  - ▶ Global monitoring
  - ▶ Advice shifter on call
  - ▶ Major interventions (service - on/off)
  - ▶ Interact with other ADC operations teams
  - ▶ Provide feed-back to ADC development teams
- ▶ **Tier I expert on call**
  - ▶ Very important



# Shift Structure Schematic



# Production and Analysis jobs

---

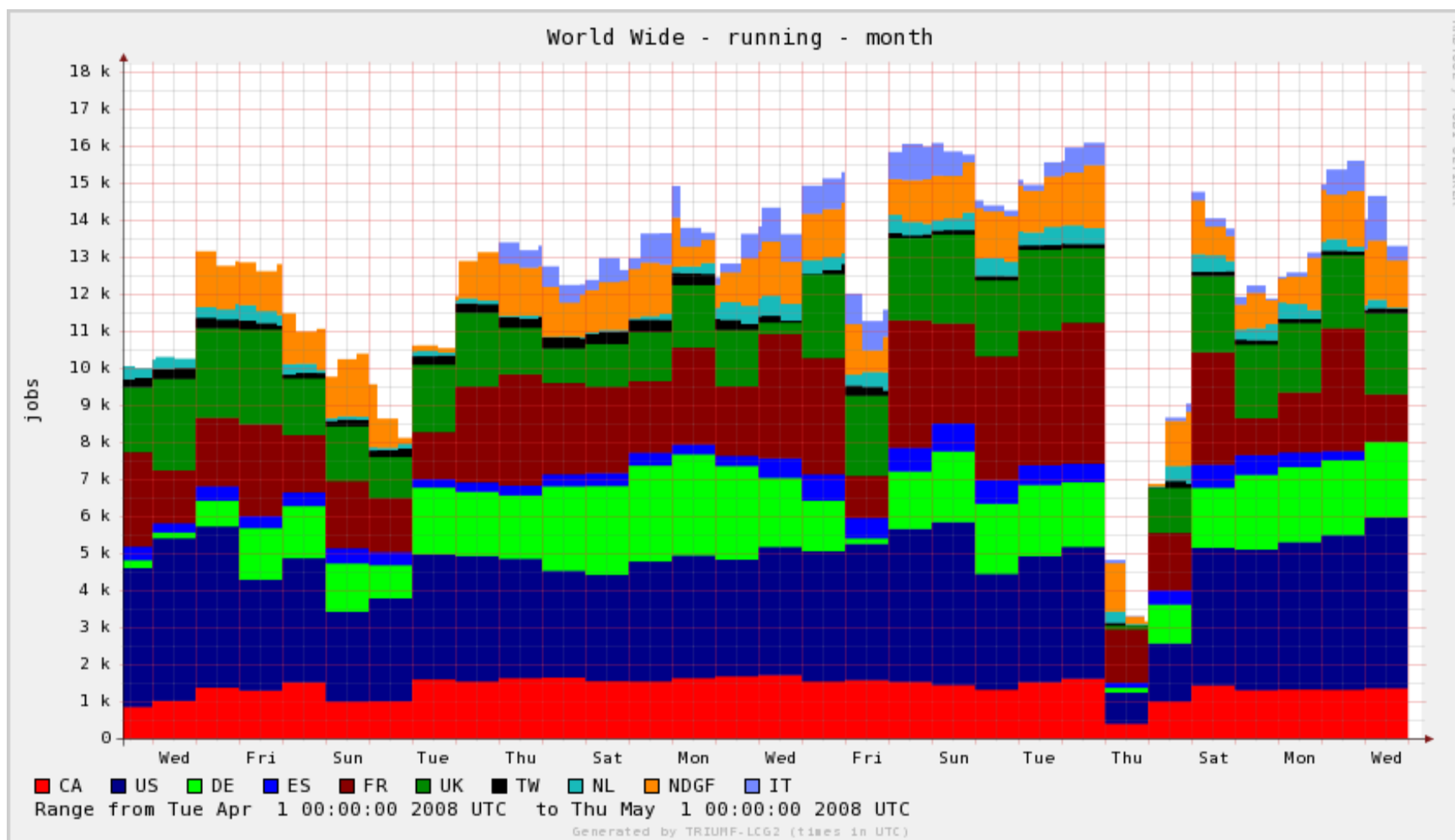
## [Managed] Production

- ▶ Panda
- ▶ Production queues
- ▶ MC simulation, FDR, Detector data processing
- ▶ Centrally managed
- ▶ Bamboo
- ▶ Triggers data transfer (DDM, FTS channels)

## Analysis

- ▶ Panda
- ▶ Analysis queues
- ▶ Varies, Data Analysis, Plot generation
- ▶ User driven
- ▶ Pathena
- ▶ Does not trigger data transfer (active staging)

# Monthly production overview



<http://gridinfo.triumf.ca/panglia/>

# For more information

---

- ▶ Panda twiki:  
<https://twiki.cern.ch/twiki/bin/view/Atlas/PanDA>
- ▶ Panda development portal:  
<https://savannah.cern.ch/projects/panda/>
- ▶ Panda software repositories:
  - ▶ <http://atlas-sw.cern.ch/cgi-bin/viewcvs-atlas.cgi/offline/Production/panda/>
  - ▶ <http://www.usatlas.bnl.gov/svn/panda/>
- ▶ DDM (DQ2):  
<https://twiki.cern.ch/twiki/bin/view/Atlas/DistributedDataManagement>
- ▶ ATLAS:  
<http://atlas.ch/>

# Acknowledgement

---

- ▶ The content presented is based on material provided by the Panda development team or the ATLAS shift team
- ▶ Specially I'd like to thank:
  - ▶ Kaushik De
  - ▶ Tadashi Maeno
  - ▶ Xavier Espinal