

# Monitoring Your Grid Jobs

Andrew Washbrook

University of Edinburgh

ATLAS Software & Computing Tutorials

14<sup>th</sup> January 2015 PUC, Chile

# Introduction and Aims

---

- This talk will show you how to monitor your jobs running on the Grid
- The focus of this talk will be getting to know the **BigPanda** monitoring tool
  - (Many) Other monitoring tools are available
- I will demonstrate this by walking through the monitoring of two simple example jobs
  - A "Hello World" python script
  - An athena based job (DPD maker)
- The job payload is not important – here we are more interested in the monitoring
- You will be doing all this yourself in the hands on session - so pay attention!
  - Don't worry - help will be available
  - All links included in this talk are persistent so you can access my example pages for comparison this afternoon

# Basic Job Submission

- The following commands submit a “hello world” job to the grid from your console
- A **Task** is created containing just one **Job**
- We will use the **jediTaskID** to track the progress of the job in the task
- We are using the **prun** tool for job submission, but can also use the **Ganga** tool for the same purpose (covered in the hands-on tutorial)

```
> setupATLAS
> localSetupPandaClient --noAthenaCheck
> prun --outDS user.wash.pruntest --exec HelloWorld.py
INFO : Need to generate a grid proxy
Your identity: /C=UK/O=eScience/OU=Edinburgh/L=NeSC/CN=andrew washbrook
Enter GRID pass phrase for this identity:
..
INFO : gathering files under /phys/linux/awashbro/tutorial/prunTest
INFO : upload source files
INFO : submit
INFO : succeeded. new jediTaskID=4617520
```

## Job Submission using prun

```
> setupATLAS
> localSetupGanga
> ganga athena --jedi --options "HelloWorld.py > out.txt" --athena_exe EXE
--outputdata 'out.txt'
```

## Job Submission using Ganga

# Basic Job Submission

- The following commands submit a “hello world” job to the grid from your console
- A **Task** is created containing just one **Job**
- We will use the **jediTaskID** to track the progress of the job in the task
- We are using the **prun** tool for job submission, but can also use the **Ganga** tool for the same purpose (covered in the hands-on tutorial)

```
> setupATLAS
> localSetupPandaClient --noAthenaCheck
> prun --outDS user.wash.pruntest --exec HelloWorld.py
INFO : Need to generate a grid proxy
Your identity: /C=UK/O=eScience/OU=Edinburgh/L=NeSC/CN=andrew washbrook
Enter GRID pass phrase for this identity:
...
INFO : gathering files under /phys/linux/awashbro/tutorial/prunTest
INFO : upload source files
INFO : submit
INFO : succeeded. new jediTaskID=4617520
```

Executable name

Proxy generation only  
needed once per session

## Job Submission using prun

Output Dataset

```
#!/usr/bin/python
print "Hello world!"
```

```
> setupATLAS
> localSetupGanga
> ganga athena --jedi --options "HelloWorld.py > out.txt" --athena_exe EXE
--outputdata 'out.txt'
```

## Job Submission using Ganga

# BigPanda Monitoring

- Next generation monitoring tool developed within the BigPanDA project
- Drop down menu with access to numerous views of job monitoring data
- Tool heavily used by ADC experts and shifters to track production activities
- Here we are more interested in monitoring our own jobs

ATLAS PanDA    Dash ▾    Tasks ▾    Jobs ▾    Errors ▾    **Users ▾**    Sites ▾    Incidents ▾    Search

ATLAS PanDA monitor home

Global concurrent running job core counts, all sites, all job types, by cloud, last 1, 7, 30 days

Slots of Running Jobs  
24 Hours from 2015-01-11 17:00 to 2015-01-12 17:39 UTC

Slots of Running Jobs  
7 Days from 2015-01-05 to 2015-01-12

Slots of Running Jobs  
30 Days from 2014-12-17 to 2015-01-12

Global concurrent running job core counts, all sites, all job types, by activity, last 1, 7, 30 days

Slots of Running Jobs  
24 Hours from 2015-01-11 17:00 to 2015-01-12 17:39 UTC

Slots of Running Jobs  
7 Days from 2015-01-05 to 2015-01-12

Slots of Running Jobs  
30 Days from 2014-12-17 to 2015-01-12

Prodsys ▾    Services ▾    VO ▾    Help ▾

aipanda054 @ 13:43, cache 3'

➤ <http://bigpanda.cern.ch>

All jobs by cloud (e.g. ES, UK, IT) running in last 1,7,30 days

All jobs by type (e.g. simulation, reconstruction, analysis) running in last 1,7,30 days

# BigPanda Search Options

Search	
PanDA job ID or name	<input type="text"/> Submit
Batch ID	<input type="text"/> Submit
Task ID	<input type="text"/> Submit
Task name	<input type="text"/> Submit
Request ID	<input type="text"/> Submit
Tasks for Request IDs	from <input type="text"/> to <input type="text"/> Submit
Jobs for Request IDs	from <input type="text"/> to <input type="text"/> Submit
Tasks for campaign	<input type="text"/> Submit
Tasks for project	<input type="text"/> Submit
Tasks for tag	<input type="text"/> Submit
Tasks for stream	<input type="text"/> Submit
Site	<input type="text"/> Submit
User	<input type="text"/> Submit
Dataset	<input type="text"/> Submit
File	<input type="text"/> Submit

- Search boxes available on BigPanda front page to get specific information on tasks, jobs and user activities
- The range of displayed information can be changed by modifying or adding the URL parameters

Enter our example  
job Task ID here

The load on the Panda  
DB will increase with  
high N

```
&limit=N  
&display_limit=N  
&hours=N  
&days=N  
&earlierthan=Nhours  
&earlierthandays=Ndays  
&date_from=2014-10-01  
&date_to=2014-10-03
```

## Example URL Modifiers

- Advanced usage: monitoring data can retrieved in *json* format to be enable results to be processed programmatically
- <http://bigpanda.cern.ch/help/>

# User View

- The user view displays your recent activities
- Useful starting point to check your jobs
- Links on this page lead to more information on an individual tasks and associated jobs

ATLAS PanDA   Dash   Tasks   Jobs   Errors   Users   Sites   Incidents   Search

PanDA user **andrew washbrook**, last 72 hours, limit 5000 per table. Params: limit=5000 user=washbrook

User page for **andrew washbrook**   Please read (and keep an eye on) the [help section below](#) to familiarize yourself with using the monitor in JEDI based analysis.

Stats from the user DB (updated every few hours)

User	Jobs in last week	Latest job in the DB stats	Personal CPU-hrs - 1 day	Personal CPU-hrs - 7 days	Group CPU-hrs - 1 day	Group CPU-hrs - 7 days
andrew washbrook	87	2013-10-12 19:33	0.0	0.0	0.0	0.0

Jump to recent [tasks](#) [jobsets](#) [jobs](#)

Grid resources used

Task attribute summary, 1 tasks

corecount (1)	1 (1)
processingtype (1)	panda-client-0.5.42-jedi-run (1)
reqid (1)	3 (1)
status (1)	done (1)
taskpriority (1)	1000 (1)
tasktype (1)	analy (1)
transpath (1)	runGen-00-00-02 (1)
username (1)	andrew washbrook (1)

Task Summary

Replace with your username

➤ <http://bigpanda.cern.ch/user/?user=washbrook>

1 tasks, sorted by jeditaskid [Show in task list page](#)

ID Parent	Jobset	Task name TaskType/ProcessingType	Campaign	Group	User	Task status Logged status	Input files finish% fail% Nfinish Nfail	Modified	State changed	Priority
4617520	3	user.wash.pruntest/ anal/panda-client-0.5.42-jedi-run	andrew washbrook			done	100% 1	2015-01-09 10:15	01-09 10:15	1000

Task Details

Recent jobsets   Note that with JEDI you should pay more attention to tasks than jobsets! See the help section below.

Jobset ID	nJobs	Modification times	Priorities	defined	waiting	assigned	activated	sent	starting	running	holding	transferring	finished	failed	cancelled
1	2	2015-01-06 14:28 to 01-06 14:36	1000 to 2000	0	0	0	0	0	0	0	0	0	2	0	0

# Example Job Monitoring

- <http://bigpanda.cern.ch/task/4617520/>

Task 4617520: user.wash.pruntest/													
Task ID	Jobset	Type	WorkingGroup	User	Task status	Ninputfiles	finished	failed	Created	Modified	Cores	Priority	Parent
4617520	3	analy		andrew washbrook	done	1	1 (100%)		2015-01-06 14:22	01-09 10:15	1	1000	

View:	<a href="#">job list (access to job details and logs)</a>	<a href="#">child tasks</a>	<a href="#">prodsys task page</a>	<a href="#">brokerage logger</a>	<a href="#">JEDI action logger</a>	<a href="#">error summary</a>
-------	---	-----------------------------	-----------------------------------	----------------------------------	------------------------------------	-------------------------------

- <http://bigpanda.cern.ch/job?pandaid=2351340829>

Where the job was run

PandaID	Owner	Request Task ID	Status	Created	Time to start d:h:m:s	Duration d:h:m:s	Modified	Cloud Site	Priority
2351340829	andrew washbrook	1 4617520	finished	2015-01-06 14:23	0:0:05:05	0:0:01:13	01-06 14:36	IT ANALY_INFN-FRASCATI	1000
Job name: user.wash.pruntest/ type: panda-client-0.5.42-jedi-run transformation: runGen-00-00-02									
Datasets: Out: user.wash.pruntest.log.14345588									

<a href="#">Job information</a>	<a href="#">logfiles</a>	<a href="#">pilot job stdout, stderr, batch log</a>	<a href="#">logger records</a>	<a href="#">child jobs</a>
---------------------------------	--------------------------	---	--------------------------------	----------------------------

<a href="#">Jobset 1</a>
--------------------------

# Job Log Files

- [http://bigpanda.cern.ch/filebrowser/?guid=b2c3305e-5581-4062-9676-92205d72018a&lfn=user.wash.prntest.log.4617520.000001.log.tgz&site=ANALY\\_INFN-FRASCATI&scope=user.wash](http://bigpanda.cern.ch/filebrowser/?guid=b2c3305e-5581-4062-9676-92205d72018a&lfn=user.wash.prntest.log.4617520.000001.log.tgz&site=ANALY_INFN-FRASCATI&scope=user.wash)
- Globally Unique Identifier
- Logical File Name

## File listing

Modification time	Size [B]	File
Tue Jan 6 14:28:57 2015	0	<a href="#">athena_stderr.txt</a>
Tue Jan 6 14:28:57 2015	23820	<a href="#">athena_stdout.txt</a>
Tue Jan 6 14:28:57 2015	730	<a href="#">job_setup.sh</a>
Tue Jan 6 14:28:52 2015	188	<a href="#">matched_replicas.json</a>
Tue Jan 6 14:28:57 2015	527	<a href="#">metadata-2351340829.xml</a>
Tue Jan 6 14:28:49 2015	1391	<a href="#">pandaJobData.out</a>
Tue Jan 6 14:28:57 2015	0	<a href="#">pandatracerlog.txt</a>
Tue Jan 6 14:28:49 2015	0	<a href="#">pilot.stderr</a>
Tue Jan 6 14:28:50 2015	77	<a href="#">PILOT_INITDIR</a>
Tue Jan 6 14:28:57 2015	97	<a href="#">Pilot_VmPeak.txt</a>
Tue Jan 6 14:28:51 2015	280	<a href="#">pilotlog.out</a>
Tue Jan 6 14:29:52 2015	83922	<a href="#">pilotlog.txt</a>

- From an individual job view you can get access to log files for further diagnosis

Output and error of job payload

Pilot logs show the staging of the job at the site

# Athena Job Submission

```
> setupATLAS
> localSetupPandaClient
> pathena DPD_topOptions.py --inDS=data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975
--outDS=user.wash.tutorial.dpd
INFO : using CMTCONFIG=x86_64-slc5-gcc43-opt
..
INFO : submit
INFO : succeeded. new jediTaskID=4635475
```

Input Dataset resides on the Grid  
Output dataset

AMI reports Input  
Dataset has 29 files  
and 211,633 events

- Let us try again with an Athena based example
- The input dataset is too big to be processed in one job and does not reside on local machine
- Grid submission is more useful as input and output datasets increase in size and number
- In this example **15** Athena jobs are defined in the task

➤ <http://ami.in2p3.fr/index.php/en/ami-portal-home>

The screenshot shows the AMI Dataset Browser interface. At the top, there's a navigation bar with links for Datasets, Files, AMITags, Nomenclature, Tools, and Bookmarks. Below the navigation is a search bar with the query "1: data11\_001-real\_data". The main area is titled "Dataset Browser" and displays a table with dataset information. The table includes columns for logicalDatasetName, nFiles, totalEvents, runNumber, period, prodsysStatus, and dataType. The "logicalDatasetName" row contains "data11\_7TeV.00189090.physics\_Egamma.merge.AOD.f403\_m975" with a red link to "DO2 - Provenance - GANGA export". The "nFiles" row shows the value "29" with a red arrow pointing to it. The "totalEvents" row shows "211633". The "runNumber" row shows "189090" with a red link to "COMA Report - Periods - Run\_Summary - Run\_Query - DAQ\_Config". The "period" row shows "L2" with a red link to "more Info - COMA - All Runs - PhysCont". The "prodsysStatus" row shows "Tier 0". The "dataType" row shows "AOD". On the right side of the table, there are sections for "dataset\_extra", "dataset\_keywords", "dataset\_comment", and "files", each indicating "No records found".

# Task Monitoring

- The new task is visible from my user view page in a *registered* state

2 tasks, sorted by jeditaskid <a href="#">Show in task list page</a>										
ID Parent	Jobset	Task name TaskType/ProcessingType	Campaign	Group	User	Task status Nfiles	Input files finish% fail% Nfinish Nfail	Modified	State changed	Priority
4635475	6	user.wash.tutorial.dpd/ anal/panda-client-0.5.42-jedi-athena	andrew washbrook			registered 29		2015-01-09 15:47	01-09 15:47	1000
4617520	5	user.wash.pruntest/ anal/panda-client-0.5.42-jedi-run	andrew washbrook			done 1	100% 1	2015-01-09 14:44	01-09 14:44	1000

Task 4635475: user.wash.tutorial.dpd/										
Task ID	Jobset	Type	WorkingGroup	User	Task status	Detailed JEDI status	Ninputfiles   finished   failed	Created	Modified	
4635475	6	analy		andrew washbrook	submitting	scouting	29   0 (0%)	2015-01-09 15:47	01-09 15:48	

States of jobs in this task (merge jobs excluded) <a href="#">Show all jobs</a>															
defined	waiting	pending	assigned	throttled	activated	sent	starting	running	holding	transferring	finished	failed	cancelled	merging	
11															

[Jump to job parameters, task parameters](#)

View: [job list \(access to job details and logs\)](#) [child tasks](#) [prodsys task page](#) [brokerage logger](#) [JEDI action logger](#) [error summary](#)

**Input containers**  
data11\_7TeV.00189090.physics\_Egamma.merge.AOD.f403\_m975

**Output containers**  
user.wash.tutorial.dpd\_MyFirstD3PD/  
user.wash.tutorial.dpd.log/

➤ <http://bigpanda.cern.ch/user/?user=washbrook>

# Task Monitoring (2)

Job attribute summary Sort by count, alpha	
atlasrelease (1)	Atlas-17.2.7 (11)
attemptnr (1)	1 (10)
cloud (1)	DE (11)
computingsite (1)	ANALY_FMPhi-UNIBA (11)
destinationse (1)	ANALY_FMPhi-UNIBA (1)
homepackage (1)	AnalysisTransforms-AtlasPhysics_17.2.7.4.1 (11)
inputfileproject (1)	data11_7TeV (10)
inputfiletype (1)	AOD (10)
jeditaskid (1)	4635475 (11)
jobsetid (1)	6 (11)
jobstatus (2)	defined (10) running (1)
outputfiletype (1)	? (11)
priorityrange (3)	900:999 (6) 1000:1099 (4) 2000:2099 (1)
processingtype (1)	panda-client-0.5.42-jedi-athena (11)
prodsourcelabel (2)	panda (1) user (10)
produsername (1)	andrew washbrook (11)
reqid (1)	6 (11)
specialhandling (1)	ddm:rucio (11)
transformation (2)	buildJob-00-00-03 (1) runAthena-00-00-12 (10)

➤ <http://bigpanda.cern.ch/task/4635475/>

- Switching to the Task view we can now see one job is running at the computing site ANALY\_FMPhi-UNIBA
- Can follow individual progression of running jobs by clicking on Job Status

Job list Sort by Pandaid, ascending mod time, descending mod time, priority, attemptnr, duration											
PanDA ID Attempt#	Owner Group	Request Task ID	Transformation	Status	Created	Time to start d:h:m:s	Duration d:h:m:s	Mod	Cloud Site	Priority	
2354305514 Attempt 0	andrew washbrook	6 4635475	buildJob-00-00-03	running	2015-01-09 15:48	0:0:16	0:0:19	01-09 15:49	DE ANALY_FMPhi-UNIBA	2000	

# Task Monitoring (3)

States of jobs in this task (merge jobs excluded)					Show all jobs													
defined	waiting	pending	assigned	throttled	activated	sent	starting	running	holding	transferring	finished	failed	cancelled	merging				
								10			1							

Job attribute summary	
atlasrelease (1)	Atlas-17.2.7 (11)
attemptnr (1)	1 (10)
cloud (1)	DE (11)
computingsite (1)	ANALY_FMPHI-UNIBA (11)
destinationse (1)	ANALY_FMPHI-UNIBA (1)
homepackage (1)	AnalysisTransforms-AtlasPhysics_17.2.7.4.1 (11)
inputfileproject (1)	data11_7TeV (10)
inputfiletype (1)	AOD (10)
jeditaskid (1)	4635475 (11)
jobsetid (1)	6 (11)
jobstatus (2)	finished (1) running (10)
outputfiletype (1)	? (11)
priorityrange (3)	900:999 (4) 1000:1099 (6) 2000:2099 (1)
processingtype (1)	panda-client-0.5.42-jedi-athena (11)
prodsourcelabel (2)	panda (1) user (10)
produsername (1)	andrew washbrook (11)
reqid (1)	6 (11)
specialhandling (1)	ddm:rucio (11)
transformation (2)	buildJob-00-00-03 (1) runAthena-00-00-12 (10)

➤ <http://bigpanda.cern.ch/task/4635475/>

- Checking some time later there are 10 jobs in a *running* state with 1 job *finished*

# Task Monitoring (4)

Task 4635475: user.wash.tutorial.dpd/														
Task ID	Jobset	Type	WorkingGroup	User	Task status	Ninputfiles	finished	failed	Created	Modified	Cores	Priority	Parent	
4635475	6	analy		andrew washbrook	done	29	29 (100%)		2015-01-09 15:47	01-09 17:49	1	1000		

States of jobs in this task (merge jobs excluded) <a href="#">Show all jobs</a>														
defined	waiting	pending	assigned	throttled	activated	sent	starting	running	holding	transferring	finished	failed	cancelled	merging
											15			

Job attribute summary Sort by count, alpha	
atlasrelease (1)	Atlas-17.2.7 (15)
attemptnr (2)	1 (10) 2 (3)
cloud (2)	DE (11) UK (4)
computingsite (2)	ANALY_FMPHI-UNIBA (11) ANALY_SHEF_SL6 (4)
destinationse (2)	ANALY_FMPHI-UNIBA (1) ANALY_SHEF_SL6 (1)
homepackage (1)	AnalysisTransforms-AtlasPhysics_17.2.7.4.1 (15)
inputfileproject (1)	data11_7TeV (13)
inputfiletype (1)	AOD (13)
jeditaskid (1)	4635475 (15)
jobsetid (1)	6 (15)
jobstatus (1)	finished (15)
outputfiletype (1)	? (15)
priorityrange (3)	900:999 (4) 1000:1099 (9) 2000:2099 (2)
processingtype (1)	panda-client-0.5.42-jedi-athena (15)
prodsourcelabel (2)	panda (2) user (13)
produsername (1)	andrew washbrook (15)
reqid (1)	6 (15)
specialhandling (1)	ddm:rucio (15)
transformation (2)	buildJob-00-00-03 (2) runAthena-00-00-12 (13)

➤ <http://bigpanda.cern.ch/task/4635475/>

- And then later all the jobs have finished and the task status is *done*
- Note that 4 of the 15 jobs were sent to a computing site in a different cloud (UK)
- Do not expect all your jobs to run at the same computing site

# Job Monitoring with pbook

---

```
> localSetupPandaClient
> pbook
INFO : Synchronizing local repository ...
..
INFO : Done
>>> show(4635475)
INFO : Getting status for TaskID=4635475 ...
INFO : Updated TaskID=4635475 ...
=====
jediTaskID : 4635475
taskStatus : scouting
JobsetID : 6
type : panda-client-0.5.42-jedi-athena
release : Atlas-17.2.7
cache : AtlasPhysics-17.2.7.4.1
inDS : data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975
outDS : user.wash.tutorial.dpd_MyFirstD3PD/,user.wash.tutorial.dpd.log/
creationTime : 2015-01-09 15:47:01
lastUpdate : 2015-01-09 16:27:04
params : pathena ../share/DPD_topOptions.py --inDS=data11_7TeV.
00189090.physics_Egamma.merge.AOD.f403_m975 --outDS=user.wash.tutorial.dpd
inputStatus :
activated : 1
running : 7
tobedone : 19
holding : 2
```

- You can also monitor the status of your jobs from the console using **pbook**
- pbook can also be used to *kill*, *retry*, *reassign* and *reproduce* your jobs
- Use the help commands in pbook for more information - e.g. `help()`, `help(show)`

# Task Logging

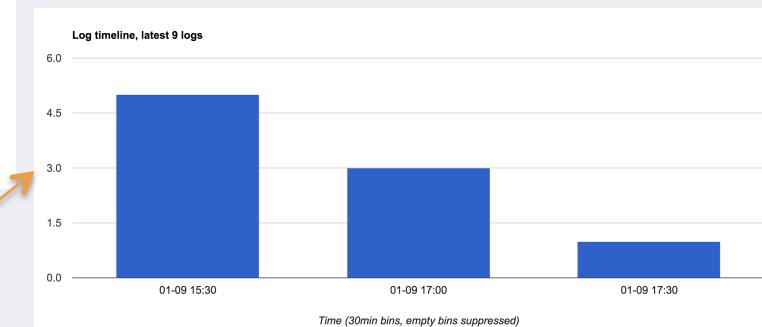
- A timeline of your task processing is shown on the Task logging view
- This page is more useful for larger task definitions (i.e. thousands of jobs)

View:	worker nodes	jobs, job errors	brokerage actions	pilots	Site status board	elogs	DDM source destination
-------	--------------	------------------	-------------------	--------	-------------------	-------	------------------------

➤ <http://bigpanda.cern.ch/task/4635475/>

Latest 9 log records				
Logged at	Category	Type	Level	Log message
2015-01-09 17:49	panda.mon.jedi	postprocessor	info	jediTaskID=4635475 set task.status=done
2015-01-09 17:05	panda.mon.jedi	jobgenerator	info	jediTaskID=4635475 set task.status=running
2015-01-09 17:05	panda.mon.jedi	jobgenerator	info	jediTaskID=4635475 successfully submitted 4/4
2015-01-09 17:05	panda.mon.jedi	jobgenerator	info	jediTaskID=4635475 start to generate jobs
2015-01-09 15:48	panda.mon.jedi	jobgenerator	info	jediTaskID=4635475 set task.status=scouting
2015-01-09 15:48	panda.mon.jedi	jobgenerator	info	jediTaskID=4635475 successfully submitted 11/11
2015-01-09 15:47	panda.mon.jedi	jobgenerator	info	jediTaskID=4635475 start to generate jobs
2015-01-09 15:47	panda.mon.jedi	contentsfeeder	info	jediTaskID=4635475 set task.status=ready
2015-01-09 15:47	panda.mon.jedi	taskrefiner	info	jediTaskID=4635475 set task.status=defined

Jedi logs generated during lifetime of Athena task



<http://bigpanda.cern.ch/logger/?category=panda.mon.jedi&jeditaskid=4635475>

# Task Output

Input containers
data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975

➤ <http://bigpanda.cern.ch/task/4635475/>

Output containers
user.wash.tutorial.dpd_MyFirstD3PD/

Datasets										
Name	ID	Type	Stream	Task	State	Status	Nfiles	Created	Modified	
user.wash.tutorial.dpd.log.14603407	14603407	log	LOG0	4635475	closed	done	10	2015-01-09 15:48	01-09 17:49	
user.wash.tutorial.dpd.log.14608187	14608187	log	LOG0	4635475	closed	done	3	2015-01-09 17:05	01-09 17:49	

➤ <http://bigpanda.cern.ch/datasetList/?containername=user.wash.tutorial.dpd.log/>

Dataset user.wash.tutorial.dpd.log.14603407									
ID	Type	Scope	Stream	Task	State	Status	Nfiles	Created	Modified
14603407	log		LOG0	4635475	closed	done	10	2015-01-09 15:48	01-09 17:49

Show files in dataset

➤ <http://bigpanda.cern.ch/datasetInfo/?datasetname=user.wash.tutorial.dpd.log.14603407>

user.wash.tutorial.dpd.log.4635475.000007.log.tgz	user.wash	log	0.11	4635475	2354327047	1	finished	2015-01-09 15:48	e73dd6f0-4bad-41e5-8724-	786323587
user.wash.tutorial.dpd.log.4635475.000004.log.tgz	user.wash	log	0.11	4635475	2354305518	0	finished	2015-01-09 15:48	bd39a3be-7af8-4ff7-877e-a22febdcb9da	786323568
user.wash.tutorial.dpd.log.4635475.000008.log.tgz	user.wash	log	0.11	4635475	2354305522	0	finished	2015-01-09 15:48	33d42d3-0714-479b-be9f-09f14cba3eeef	786323592

File Size (MB)

Job ID

➤ <http://bigpanda.cern.ch/fileList/?datasetname=user.wash.tutorial.dpd.log.14603407>

# Task Output (2)

- Return to the job view to show the files processed and generated by a single job

16 job files					
Filename (Type)	Scope	Size (MB)	Status	Attempt (max)	Dataset
data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975_lb0091_0001.1 (input) Find creator job	data11_7TeV	1577	finished		data11_7TeV: data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975
data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975_lb0092_0001.1 (input) Find creator job	data11_7TeV	1617	finished		data11_7TeV: data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975
data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975_lb0093_0001.1 (input) Find creator job	data11_7TeV	1605	finished		data11_7TeV: data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975
data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975_lb0094_0001.1 (input) Find creator job	data11_7TeV	1621	finished		data11_7TeV: data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975
data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975_lb0095_lb0096_0001.1 (input) Find creator job	data11_7TeV	2089	finished		data11_7TeV: data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975
data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975_lb0091_0001.1 (input) Find creator job	data11_7TeV	1577	ready		data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975
data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975_lb0092_0001.1 (input) Find creator job	data11_7TeV	1617	ready		data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975
data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975_lb0093_0001.1 (input) Find creator job	data11_7TeV	1605	ready		data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975
data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975_lb0094_0001.1 (input) Find creator job	data11_7TeV	1621	ready		data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975
panda.0109170546.915527.lib_.4635475.787346101.lib.tgz (input) Find creator job	panda	0	ready		panda.0109170546.915527.lib_.4635475 (dispatch block: panda.0109170546.915527.lib_.4635475 pinkpanda rucio)
DBRelease-27.5.1.tar.gz (input) Find creator job	ddo	1373	ready		ddo:ddo.000001.Atlas.Ideal.DBRelease.v270501
data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975_lb0095_lb0096_0001.1 (input) Find creator job	data11_7TeV	2089	ready		data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975
user.wash.tutorial.dpd.log.4635475.000013.log.tgz (log)	user.wash	0	finished		user.wash.tutorial.dpd.log.14608187
user.wash.tutorial.dpd.log.4635475.000013.log.tgz (log)	user.wash	0	ready		user.wash.tutorial.dpd.log/ (destination block: sub0201463524 pinkpanda rucio)
user.wash.4635475.MyFirstD3PD_.000013.root (output)	user.wash	1	finished		user.wash.tutorial.dpd._MyFirstD3PD_.14608188
user.wash.4635475.MyFirstD3PD_.000013.root (output)	user.wash	1	ready		user.wash.tutorial.dpd._MyFirstD3PD/ (destination block: sub0201463523 pinkpanda rucio)

➤ [http://bigpanda.cern.ch/  
job?pandaid=2354357830](http://bigpanda.cern.ch/job?pandaid=2354357830)

Job Input Files

File user.wash.tutorial.dpd.log.4635475.000013.log.tgz					
Scope	# events	Size (MB)	Created	GUID	ID
user.wash	None	0.13	2015-01-09 17:05	8f4e5a62-acc4-42de-ae78-b45de9ede0fc	787346107

➤ [http://bigpanda.cern.ch/fileInfo/?  
Ifn=user.wash.tutorial.dpd.log.  
4635475.000013.log.tgz&scope=user.wash&panda  
Id=2354357830](http://bigpanda.cern.ch/fileInfo/?Ifn=user.wash.tutorial.dpd.log.4635475.000013.log.tgz&scope=user.wash&pandaId=2354357830)

Occurrences of the file in processing (most recent jobs listed first)							
Dataset	Task	Type	Job	First event	End event	Status	Attempt (max)
user.wash.tutorial.dpd.log.14608187	4635475	log	2354357830	None	None	finished	

# Accessing Task Data

- The job output is now available on the Grid for retrieval *and* for further use in subsequent jobs
- You can retrieve log files and job output from the Grid for local processing using DQ2 tools - all you need is the Dataset name

```
> localSetupDQ2Client
> dq2-ls -f user.wash.tutorial.dpd.log/
> dq2-get -f user.wash.tutorial.dpd.log.4635475.000013.log.tgz user.wash.tutorial.dpd.log/
Using ROAMING profile
Querying DQ2 central catalogues to resolve datasetname user.wash.tutorial.dpd.log/
Datasets found: 2
Number of datasets requested: 2
Total number of files in datasets: 16
Number of files or lfn patterns specifically requested by user: 1
Number of file download attempts by dq2-get (not including retrials): 1
Number of successful file download attempts: 1
..
Finished
```

-bash-4.1\$ dq2-ls -f user.wash.tutorial.dpd.log/

File Path	File ID	Ad ID	Last Modified
[ ] user.wash.tutorial.dpd.log.4635475.000001.log.tgz	102979A9-7933-4B95-BABA-489FD2C47BB1	ad:503587b4	109167
[ ] user.wash.tutorial.dpd.log.4635475.000005.log.tgz	89FD020B-825D-40D8-8093-E12876C7AEA9	ad:eb12b787	9649
[ ] user.wash.tutorial.dpd.log.4635475.000008.log.tgz	338D42D3-0714-479B-BE9F-09F14CBA3EEF	ad:28370c91	108905
[ ] user.wash.tutorial.dpd.log.4635475.000010.log.tgz	2B410481-5955-4ADE-87F6-ECA3D01A3DF7	ad:2e896ce4	108699
[ ] user.wash.tutorial.dpd.log.4635475.000013.log.tgz	8F4E5A62-ACC4-42DE-AE78-B45DE9EDE0FC	ad:345593e6	128449
[ ] user.wash.tutorial.dpd.log.4635475.000004.log.tgz	BD39A3BE-7AF8-4FF7-877E-A22FEBCDC89DA	ad:48a9dc4f	112259
[ ] user.wash.tutorial.dpd.log.4635475.000006.log.tgz	F000CF14-207D-4C11-8DDC-41C6047FD69C	ad:6fb97e56	108444
[ ] user.wash.tutorial.dpd.log.4635475.000002.log.tgz	CB8B4245-3AB1-47C8-85A0-A98422E75D10	ad:29686124	108256
[ ] user.wash.tutorial.dpd.log.4635475.000011.log.tgz	5A85C6CB-1EE7-4E84-B95F-EDF616C388D6	ad:90e1ea9	140593
[ ] user.wash.tutorial.dpd.log.4635475.000003.log.tgz.2	10130438-B538-4A2B-AF07-FD3298F51804	ad:0d73fcf2	108139
[ ] user.wash.tutorial.dpd.log.4635475.000007.log.tgz.2	E73DD6F0-4BAD-41E5-8724-07D76223091E	ad:5a65ef95	111144
[ ] user.wash.tutorial.dpd.log.4635475.000007.log.tgz	1B4C82B8-90B9-41C3-BDDD-7CF72154BE13	ad:4848bf46	10056
[ ] user.wash.tutorial.dpd.log.4635475.000005.log.tgz.2	CBCDCE27-B516-4558-BC1D-0F9929B47071	ad:d1fa19b2	107306
[ ] user.wash.tutorial.dpd.log.4635475.000003.log.tgz	FF8A1121-52B4-46F2-97B7-2682F4A85D3Z	ad:c63dd5a3	9684
[ ] user.wash.tutorial.dpd.log.4635475.000012.log.tgz	172B2B87-33BE-4B56-8D4A-7F704B711C24	ad:1ac9eb07	141969
[ ] user.wash.tutorial.dpd.log.4635475.000009.log.tgz	F536FDD4-21EF-4748-B7B2-410420F40A5Z	ad:0c65c47d	109587

total files: 16  
local files: 0  
total size: 1532306  
date: 2015-01-09 15:48:12

# Error Reporting and Diagnosis

- So that was easy! All the jobs ran and completed without problems and are available for downloading with just a few commands. Isn't the Grid wonderful?
- **The brutal reality:** No matter how hard you try your jobs will still fail
  - Don't worry - it is sometimes not your fault!
- The ATLAS Distributed Computing system compensates for occasional failures very well

Job attribute summary Sort by count, alpha	
atlasrelease (1)	Atlas-17.2.7 (15)
attemptnr (2)	1 (10) 2 (3)
cloud (2)	DE (11) UK (4)
computingsite (2)	ANALY_FMPHI-UNIBA (11) ANALY_SHEF_SL6 (4)
destinationse (2)	ANALY_FMPHI-UNIBA (1) ANALY_SHEF_SL6 (1)
homepackage (1)	AnalysisTransforms-AtlasPhysics_17.2.7.4.1 (15)
inputfileproject (1)	data11_7TeV (13)
inputfiletype (1)	AOD (13)
jeditaskid (1)	4635475 (15)
jobsetid (1)	6 (15)
jobstatus (1)	finished (15)
outputfiletype (1)	? (15)
priorityrange (3)	900:999 (4) 1000:1099 (9) 2000:2099 (2)
processingtype (1)	panda-client-0.5.42-jedi-athena (15)
prodsourcelabel (2)	panda (2) user (13)
produsername (1)	andrew washbrook (15)
reqid (1)	6 (15)
specialhandling (1)	ddm:rucio (15)
transformation (2)	buildJob-00-00-03 (2) runAthena-00-00-12 (13)

➤ <http://bigpanda.cern.ch/task/4635475/>

3 jobs needed 2 attempts  
to complete

- In fact - although the last example completed not all of the jobs worked first time
- So what happened?

# Task Errors

- First check for errors associated with the task

Overall error summary			
Category:code	Job list	Nerrors	Sample error description
pilot:1195	jobs	3	cmtsite command was timed out: 28175, (timed out)
taskbuffer:107	jobs	3	retrying at the same site. new PandaID=2354326308

Site error summary			
ANALY_FMPHI-UNIBA	6	Total errors in 3 jobs.	Finished: Failed: % failed: Holding: Cancelled:
	3	pilot:1195	cmtsite command was timed out: 28175, (timed out)
	3	taskbuffer:107	retrying at the same site. new PandaID=2354326308

User error summary			
andrew washbrook	6	Total errors	
	3	pilot:1195	cmtsite command was timed out: 28175, (timed out)
	3	taskbuffer:107	retrying at the same site. new PandaID=2354326308

Error string is “cmtsite command was timed out”

Task error summary			
jeditaskid	4635475	user.wash.tutorial.dpd/	
6	Total errors in 3 jobs.	Finished: 15 Failed: 3 % failed: 16 Holding: 0 Cancelled: 0	
3	pilot:1195	cmtsite command was timed out: 28175, (timed out)	
3	taskbuffer:107	retrying at the same site. new PandaID=2354326308	

➤ <http://bigpanda.cern.ch/errors/?jeditaskid=4635475>

Error code

All 3 jobs failed at same site

## Error occurrences by job attribute

cloud (1)	DE (3)
computingsite (1)	ANALY_FMPHI-UNIBA (3)
jeditaskid (1)	4635475 (3)
jobstatus (1)	failed (3)
processingtype (1)	panda-client-0.5.42-jedi-athena (3)
prodsourcelabel (1)	user (3)
produsername (1)	andrew washbrook (3)
specialhandling (1)	ddm:rucio (6)
transformation (1)	runAthena-00-00-12 (3)

# Job Errors

- We can also look at an individual failed jobs for more information

3 jobs were dropped from this listing because they were retried. Where there were retries, the latest retry job (only) is listed.  
Click to show/hide dropped jobs

Dropped (retry):

2354305521 (2354327047) 2354305519 (2354326307) 2354305517 (2354326308)

➤ <http://bigpanda.cern.ch/jobs/?jeditaskid=4635475&jobstatus=failed&piloterrorcode=1195>

Job details for PanDA job 2354305521

aipanda0:

PandaID	Owner	Request Task ID	Status	Created	Time to start d:h:m:s	Duration d:h:m:s	Modified	Cloud Site
2354305521	andrew washbrook	6 4635475	failed	2015-01-09 15:48	0:0:14:59	0:0:19:40	01-09 17:05	DE ANALY_FMPhi-UNIBA

Job name: user.wash.tutorial.dpd/ type: panda-client-0.5.42-jedi-athena transformation: runAthena-00-00-12

Datasets: In: data11\_7TeV:data11\_7TeV.00189090.physics\_Egamma.merge.AOD.f403\_m975  
Out: user.wash.tutorial.dpd\_MyFirstD3PD.14603408

Error description	pilot: cmssite command was timed out: 28175, (timed out) taskbuffer: retrying at the same site. new PandaID=2354327047
Pilot error code	1195
Pilot error message	cmssite command was timed out: 28175, (timed out)
TaskBuffer error code	107
TaskBuffer error message	retrying at the same site. new PandaID=2354327047

This job has descendants downstream, with relationship shown

2354327047:retry

Show script to re-create job for offline debugging: 2354305521

Job information	logfiles	pilot job stdout, stderr, batch log	logger records	child jobs
-----------------	----------	-------------------------------------	----------------	------------

Jobset 6 build job 2354357827 16 run jobs

➤ <http://bigpanda.cern.ch/job/2354305521/>

# Site Status

- Is this a problem with my job?
- Check the site status to determine if this is job related error or a site problem

PanDA resource ANALY\_FMPHI-UNIBA

ANALY_FMPHI-UNIBA information	
GOC name	FMPHI-UNIBA
Status	online
Comment	HC.Blacklist.set.online
Cloud	DE
Tier	T2
DDM endpoint	FMPHI-UNIBA_SCRATCHDISK,FMPHI-UNIBA_LOCALGROUPDISK
Maximum memory	4.1 GB
Maximum time	96.0 hours
Space	4 TB as of 01-10 15:56
Last modified	2013-09-16 22:06

View: worker nodes | jobs, job errors | brokerage actions | pilots | Site status board | elogs | DDM source destination

Further information  
on site status

No incidents reported  
during job execution

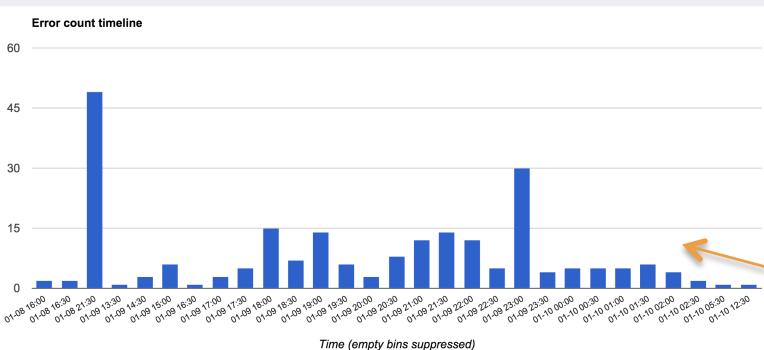
➤ [http://bigpanda.cern.ch/site/  
ANALY\\_FMPHI-UNIBA/](http://bigpanda.cern.ch/site/ANALY_FMPHI-UNIBA/)

Site is in state *online*

Incidents over the last month	
2014-12-17 18:00	setonline: queue=ANALY_FMPHI-UNIBA DN=gangarbt HC.Blacklist.set.online
2014-12-17 18:00	setmanual: queue=ANALY_FMPHI-UNIBA DN=gangarbt HC.Blacklist.set.manual
2014-12-17 15:20	settest: queue=ANALY_FMPHI-UNIBA DN=adcssb02 HC.Test.Me
2014-12-17 15:20	setmanual: queue=ANALY_FMPHI-UNIBA DN=adcssb02 HC.Test.Me
2014-12-17 05:00	setoffline: queue=ANALY_FMPHI-UNIBA DN=adcssb02 set.offline.by.Switcher
2014-12-17 05:00	setmanual: queue=ANALY_FMPHI-UNIBA DN=adcssb02 set.offline.by.Switcher
2014-12-17 01:00	setbrokeroff: queue=ANALY_FMPHI-UNIBA DN=adcssb02 set.brokeroff.by.Switcher
2014-12-17 01:00	setmanual: queue=ANALY_FMPHI-UNIBA DN=adcssb02 set.brokeroff.by.Switcher

# Site Errors

- Now check all job errors at the site during the failed job execution for similar errors



➤ [http://bigpanda.cern.ch/errors/?computingsite=ANALY\\_FMPhl-UNIBA&days=2](http://bigpanda.cern.ch/errors/?computingsite=ANALY_FMPhl-UNIBA&days=2)

Job error timeline

Modify URL to select time of interest

Error occurrences by job attribute	
cloud (1)	DE (231)
computingsite (1)	ANALY_FMPhl-UNIBA (231)
jeditaskid (12)	4630856 (49) 4632375 (2) 4634601 (152) 4634611 (2) 4634621 (2) 4634625 (2) 4634629 (1) 4634886 (1) 4634897 (6) 4635371 (3) 4635475 (3) 4635515 (1)
jobstatus (1)	failed (231)
processingtype (5)	gangarobot (7) panda-client-0.5.35-jedi-athena-trf (7) panda-client-0.5.42-jedi-athena (3) panda-client-0.5.42-jedi-athena-trf (54) panda-client-0.5.42-jedi-run (160)
prodsourcelabel (1)	user (231)
produsername (8)	Fabian Spettel (5) Pieter van der Deijl pieterd@nikhef.nl (155) Robert Keyes (2) Simon Berlendis (49) adam barton (3) andrew pilkington (7) andrew washbrook (3) gangarb (7)
specialhandling (1)	ddm:rucio (448)
transformation (3)	runAthena-00-00-11 (7) runAthena-00-00-12 (64) runGen-00-00-02 (160)

Overall error summary			
Category:code	Job list	Nerrors	Sample error description
ddm:200	jobs	5	Could not add files to DDM: Details:
pilot:1098	jobs	9	Too little space left on local disk to run job: 816840704 B (need > 2147483648 B)
pilot:1100	jobs	49	Get error: (1) Replica with guid 04ae7bf5-8d90-4591-9ae3-a2e2dd991d3c not found at srm://lcgdpmse.dnp.fmph.uniba.sk/~dummy (or in the se list: ['srm://lcgdpmse.dnp.fmph.uniba.sk'])
pilot:1195	jobs	11	cmntssite command was timed out: 28175, (timed out)
taskbuffer:107	jobs	13	retrying at the same site. new PandaID=2353537682
transformation:1	jobs	151	Unspecified error, consult log file
transformation:40	jobs	6	Athena crash - consult log file

11 jobs failed with the same error

# Worker Node Errors

- This appears to be a site related problem so our investigations can stop
- For completeness we can still gather more information on the *Worker Node* that executed the job

## File lookup parameters

Scope	user.wash
LFN	<a href="http://bigpanda.cern.ch/filebrowser/?guid=1b4c82b8-90b9-41c3-bddd-7cf72154be13&amp;lfn=user.wash.tutorial.dpd.log.4635475.000007.log.tgz">user.wash.tutorial.dpd.log.4635475.000007.log.tgz</a>
Site	ANALY_FMPHI-UNIBA
GUID	1b4c82b8-90b9-41c3-bddd-7cf72154be13

## File listing

Modification time	Size [B]	File
Fri Jan 9 16:03:34 2015	194	<a href="#">job_setup.sh</a>
Fri Jan 9 16:03:02 2015	2401	<a href="#">pandaJobData.out</a>
Fri Jan 9 16:03:14 2015	0	<a href="#">pilot.stderr</a>
Fri Jan 9 16:03:34 2015	74	<a href="#">PILOT_INITDIR</a>
Fri Jan 9 16:03:43 2015	274	<a href="#">pilotlog.out</a>
Fri Jan 9 16:21:08 2015	48893	<a href="#">pilotlog.txt</a>
Fri Jan 9 16:03:43 2015	0	<a href="#">runjob.stderr</a>

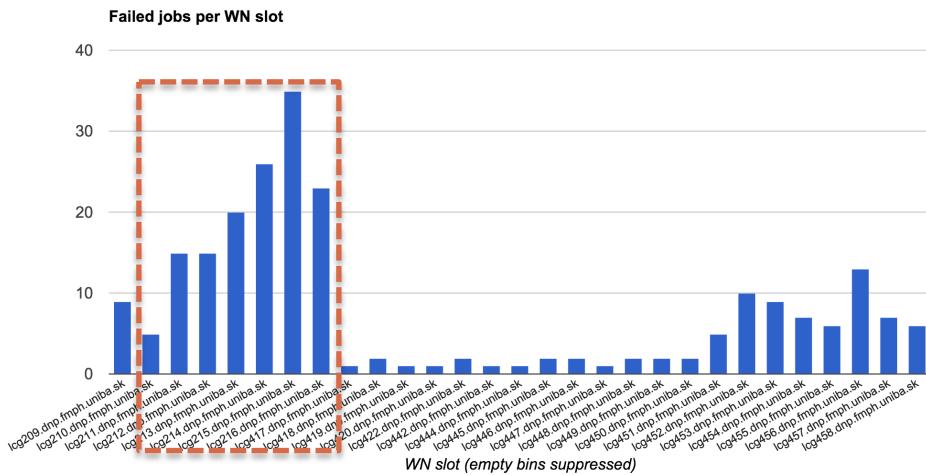
➤ [http://bigpanda.cern.ch/filebrowser/?guid=1b4c82b8-90b9-41c3-bddd-7cf72154be13&lfn=user.wash.tutorial.dpd.log.4635475.000007.log.tgz&site=ANALY\\_FMPHI-UNIBA&scope=user.wash](http://bigpanda.cern.ch/filebrowser/?guid=1b4c82b8-90b9-41c3-bddd-7cf72154be13&lfn=user.wash.tutorial.dpd.log.4635475.000007.log.tgz&site=ANALY_FMPHI-UNIBA&scope=user.wash)

**Tip:** search for this string to determine worker node job ran on

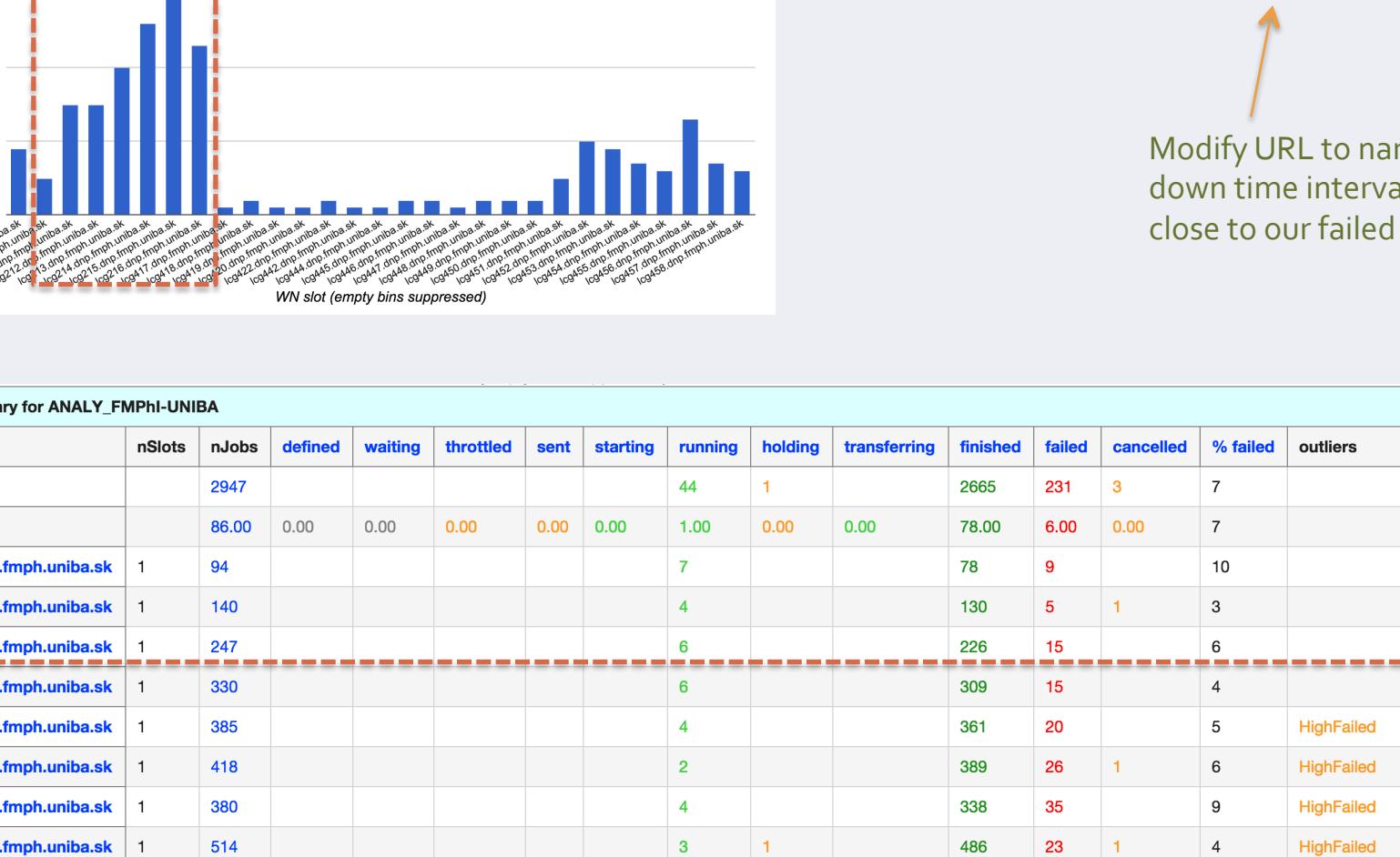
```
09 Jan 16:03:34|UpdateHandle|Connected from ('127.0.0.1', 50934)
09 Jan 16:03:34|UpdateHandle|--- TCPServer: Message received from child is : []
09 Jan 16:03:34|pUtil.py|Pilot modules have been copied to /home/pilat103/home_cream_779661867/CREAM
09 Jan 16:03:34|Monitor.py|Current time :2015-01-09T17:03:34+100
09 Jan 16:03:34|Monitor.py|The site this pilot runs on: ANALY_FMPHI-UNIBA
09 Jan 16:03:34|Monitor.py|Pilot executing on host: lcg212.dnp.fmph.uniba.sk
09 Jan 16:03:34|Monitor.py|The workdir this pilot runs on:/home/pilat103/home_cream_779661867/CREAM779
09 Jan 16:03:34|Monitor.py|The dq2url for this site is:
09 Jan 16:03:34|Monitor.py|The DQ2 SE has 999999 GB space left (NB: dCache is defaulted to 999999)
09 Jan 16:03:34|Monitor.py|New job has prodSourceLabel: user
09 Jan 16:03:34|Monitor.py|env = -l panda.0109154800.984205.lib._4635475.786323534.lib.tgz --sourceURL
['data11_7TeV.00189090.physics_Egamma.merge.AOD.f403_m975_lb0068-lb0069._0001.1']" --dbrFile=DBRelease-2
'user.wash.4635475.MyFirstD3PD._000007.root')]]}" -j "%20../share/DPD_topOptions.py"
09 Jan 16:03:34|Monitor.py| Using looping job limit: 10800 s
```

# Worker Node Errors (2)

- Finally we can check if other jobs have failed on the same worker node



➤ [http://bigpanda.cern.ch/wns/  
ANALY\\_FMPhi-UNIBA/?days=2](http://bigpanda.cern.ch/wns/ANALY_FMPhi-UNIBA/?days=2)



# Other Monitoring Tools

## Data Management



## DDM & Storage Accounting



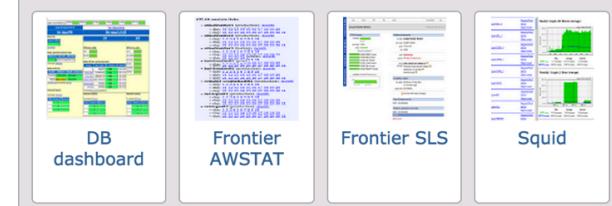
## Data Processing



## Sites and Services



## Databases



➤ <http://adc-monitoring.cern.ch/>

- Many more monitoring pages are used as part of ADC operations and by ADCOS shifters

# Conclusions

---

- Always ensure that your grid job is validated on your local machine before submission
- Clear your job of any local input or environment dependencies
- Your jobs will still fail even if you have set everything up perfectly
  - The ATLAS distributed computing system can cope with these failures
  - Jobs will be retried without any effort needed
- The BigPanda monitoring is a very useful tool to determine the progression of your Grid jobs
- Many different views of the data available - once you have mastered the navigation of the pages you should always find the information you need
- There should be no problem determining *where* jobs have run and the *how* jobs failed. The *why* is up to you to find out!