

# Introduction to PanDA Client Tools – pathena, prun and others

**Nurcan Ozturk**

**University of Texas at Arlington**

**First ATLAS-South Caucasus Software / Computing**

**Workshop & Tutorial**

**25-29 October 2010, Tbilisi, Georgia**

# PanDA Client Tools



- PanDA = Production and Distributed Analysis System for ATLAS
- PanDA client consists of five tools to submit or manage analysis jobs on PanDA
- DA on PanDA page: <https://twiki.cern.ch/twiki/bin/viewauth/Atlas/DAonPanda>
  - [pathena](#)
    - How to submit Athena jobs
  - [prun](#)
    - How to submit general jobs (ROOT, python, sh, exe, ...)
  - [psequencer](#)
    - How to perform sequential jobs/operations (e.g. submit job + download output)
  - [pbook](#)
    - Bookkeeping (browsing, retry, kill) of analysis jobs
  - [puserinfo](#)
    - Access control on PanDA analysis queues

# What is pathena?



- To submit Athena jobs to PanDA
- A simple command line tool, but contains advanced capabilities for more complex needs
- A consistent user interface to Athena
- When you run Athena with:

```
$ athena jobOptions.py
```

all you need to do is:

```
$ pathena jobOptions.py --inDS inputDatasetName --outDS outputDatasetName
```



a dataset which contains  
the input files



a dataset which will contain  
the output files

# Launching a pathena job



```
$ pathena jobOptions.pythia.py --outDS user.nurcan.pythiaEventGeneration
```

```
INFO : extracting run configuration
INFO : ConfigExtractor > No Input
INFO : ConfigExtractor > Output=STREAM1 pythia.pool.root
INFO : ConfigExtractor > RndmStream PYTHIA
INFO : ConfigExtractor > RndmStream PYTHIA_INIT
INFO : archiving source files
INFO : archiving InstallArea
INFO : checking symbolic links
INFO : uploading source/jobO files
INFO : trying to get the latest version number for DBRelease=LATEST
INFO : use ddo.000001.Atlas.Ideal.DBRelease.v120901:DBRelease-12.9.1.tar.gz
INFO : query files in ddo.000001.Atlas.Ideal.DBRelease.v120901
INFO : submit to ANALY_SLAC
```

```
=====
```

```
JobsetID : 14390
```

```
JobID : 14391
```

```
Status : 0
```

```
> build
```



Recreates the job's environment at the grid site

```
PandaID=1132989222
```

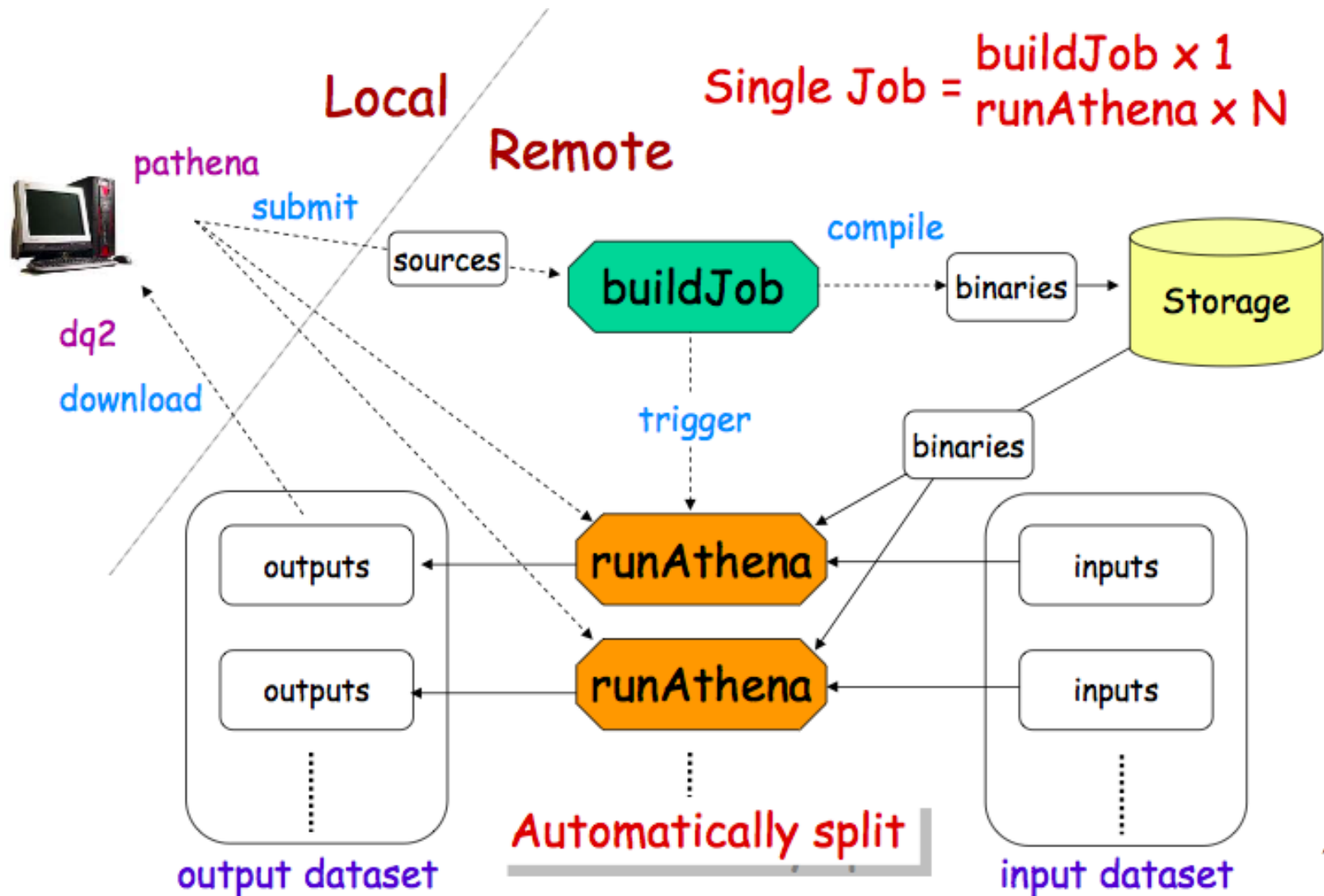
```
> run
```



Runs the job option file at the grid site

```
PandaID=1132989223
```

# Job Cycle



# What is prun?



- To submit general jobs to PanDA:
  - ROOT (ARA- AthenaRootAccess), Python, shell script, exe, ...
- ATLAS analysis has two stages
  - Run Athena on AOD/ESD to produce DPD → pathena
  - Run ROOT, Python, shell scripts, etc. to produce final plots → prun
- How to run prun:
  - Example in the twiki page:

<https://twiki.cern.ch/twiki/bin/viewauth/Atlas/SouthCaucasusComputingTutorial>

```
$ prun --outDS user.nurcan.pruntest --exec HelloWorld.py
```



output dataset name



name of the python script

# Launching a prun job



```
$ prun --outDS user.nurcan.pruntest --exec HelloWorld.py
```

```
INFO : gathering files under /afs/cern.ch/user/n/nozturk/scratch0/16.0.1/run
```

```
INFO : upload source files
```

```
INFO : submit to ANALY_SARA
```

```
=====
```

```
JobsetID : 14388
```

```
JobID : 14389
```

```
Status : 0
```

```
> build
```



Recreates the job's environment at the grid site

```
PandaID=1132981745
```

```
> run
```



Runs the job option file at the grid site

```
PandaID=1132981750
```

# What is pbook?



- Bookkeeping of PanDA jobs:
  - Browsing
  - Retry
  - Kill
- Makes a local sqlite3 repository to keep personal job information:
  - IMAP like sync-diff mechanism
  - Not scanning the global PanDA repository, thus quick response
- Dual user interface
  - Command-line
  - Graphical



# Monitoring a PanDA job (1/2)



Go to PanDA monitor at <http://panda.cern.ch> and enter your Panda jobID on the left panel

Quick guide to the Panda monitor

Not logged in. [List users](#)

Find yourself here

Quick search  
Panda job ID

Quick search

Production: Panda Production Operations Dashboard. Summary of Panda production status  
Clouds: Organization and task assignment of clouds (Tier 1 + Tier 2/3s) processing Panda jobs  
DDM: Summary of DDM systems information and tools  
PandaMover: Panda DQ2 dataset mover status. Monitors Panda jobs that replicate datasets using dq2-cr.  
AutoPilot: Pilot submission system serving all of OSG and LCG  
Sites: Collection of grid-wide and site-level monitoring links  
Analysis: Information on Panda-based analysis using pathena  
Physics data: ATLAS data discovery and access info and tools for physicists  
Usage: CPU usage by user  
ProdDash: Link to the ARDA ATLAS production dashboard  
DDMDash: Link to the ARDA ATLAS DDM dashboard  
List users: On extreme right, lists Panda users and gives access to 'your' Panda page

Jobs running in Panda: Job links at left list the running, activated (ready for pickup by a pilot), waiting (waiting for input data availability), assigned (brokered and waiting for completion of input data transfer to processing site), defined (awaiting brokerage), finished, failed and cancelled jobs. Analysis jobs (as opposed to managed production jobs) can be listed separately. The 'old archive' contains all finished/failed/cancelled jobs older than 3 days.

Quick searches: Enter a Panda job name or ID, dataset name or ID, or task name or ID into the appropriate field and hit return in order to do a quick lookup.

Summaries: Enter a day count in the desired summary field and hit return to bring up a summary covering the last N days. The 'blocks' summary shows the production datablocks (datasets) currently being processed in the production system, with details on where they are being processed, job states etc. The 'errors' summary shows overall production status at all Panda sites with details of the error conditions encountered. The 'nodes' summary shows worker nodes active at all production sites with statistics on processed jobs and states.

Tasks: Task request forms are provided for entering generic, event generation and CTB tasks. The full task list gives statistics on tasks by grid and a listing of all tasks. The task browser

# Monitoring a PanDA job (2/2)



pathena run job finished.

Panda jobs

http://panda.cern.ch:25980/server/pandamon/query?job=1132989223

Configuration: [Production](#) [Clouds](#) [Incidents](#) [DDM](#) [PandaMover](#) [AutoPilot](#) [Sites](#) [Releases](#) [Analysis](#) [Stats](#) [Users](#) [Physics data](#) [ProdDash](#) [DDMDash](#)

Update

**Panda monitor**  
Times are in UTC

Not logged in. [List users](#)

**Panda jobs**

Jobs: 1132989223

[Click for help](#)

PandaID, Owner, Working group	Job	Status	Created	Time to start	Duration	Ended/Modified	Cloud/Site, Type	Priority
<a href="#">1132989223</a> <a href="#">Nurcan Ozturk</a>	jobsetID=14390 runAthena-00-00-11	finished	2010-10-21 09:37	0:12:56	0:02:15	10-21 09:52	US/ANALY_SLAC, analysis-run	1000

Out: [user.nurcan.pythiaEventGeneration/](#)

Associated build job: [1132989222](#)

Run jobs in this job set: 1132989223

**Job 1132989223 details**

[Show pilot information](#)

4 files for job 1132989223:

Filename	Type	Status	Dataset
<a href="#">DBRelease-12.9.1.tar.gz</a> guid=5ee1fcfb-498e-4a6f-aaa6-58529f60b0d1	input	ready	<a href="#">ddo.000001.Atlas.Ideal.DBRelease.v120901</a>
<a href="#">user.nurcan.1021093741.873687.lib_014390.lib.tgz</a> guid=847fc200-86e9-4b21-9244-52d50e3f2230	input	ready	<a href="#">user.nurcan.1021093741.873687.lib_014390</a>
<a href="#">user.nurcan.014390_1132989223.log.tgz</a> guid=2be59764-b45a-4f3e-81ee-08e1d40253a1 Space token SLACXRD_USERDISK	log	ready	<a href="#">user.nurcan.pythiaEventGeneration/</a> (destination block: <a href="#">user.nurcan.pythiaEventGeneration.101021113741_sub011293800</a> )
<a href="#">user.nurcan.014390.Stream1_00001.pool.root</a> guid=B4BD0AEA-F8DC-DF11-B9FF-A4BADB09A283 Space token SLACXRD_USERDISK	output	ready	<a href="#">user.nurcan.pythiaEventGeneration/</a> (destination block: <a href="#">user.nurcan.pythiaEventGeneration.101021113741_sub011293800</a> )

[Find and view log files](#)

Quick search  
Panda job ID

Summaries  
Blocks:  days  
Errors:  days  
Nodes:  days  
Usage   days

Tasks - [search](#)  
[Generic Task Req](#)  
[EvGen Task Req](#)  
[CTBSim Task Req](#)  
[Task list](#)  
[New Tag](#)  
[Bug Report](#)  
[Task overview query](#)

Datasets - [search](#)  
[DQ2 Popularity](#)  
Aborted datasets

build job

output is dataset container (now default in pathena/prun)

# More options with pathena



- `$ pathena -h`
- Usage: `pathena [options] <jobOption1.py> [<jobOption2.py> [...]]`
- 'pathena --help' prints a summary of the options
- HowTo is available at <https://twiki.cern.ch/twiki/bin/view/Atlas/PandaAthena>
- Options:
  - `-h, --help` show this help message and exit
  - `--version` Displays version
  - `--split=SPLIT` Number of sub-jobs to which a job is split
  - `--nFilesPerJob=NFILESPERJOB`  
Number of files on which each sub-job runs
  - `--nEventsPerJob=NEVENTSPERJOB`  
Number of events on which each sub-job runs
  - `--nEventsPerFile=NEVENTSPERFILE`  
Number of events per file
  - `--nGBPerJob=NGBPERJOB`  
Instantiate one sub job per NGBPERJOB GB of input files. `--nGBPerJob=MAX` sets the size to the default maximum value
  - `--site=SITE` Site name where jobs are sent (default:AUTO)



Many options available

# More options with prun



- `$ prun -h`
- Usage: `prun [options]`
- HowTo is available at <https://twiki.cern.ch/twiki/bin/view/Atlas/PandaRun>
- Options:
  - `-h, --help` show this help message and exit
  - `--version` Displays version
  - `--inDS=INDS` Name of an input dataset or dataset container
  - `--goodRunListXML=GOODRUNLISTXML`
    - Good Run List XML which will be converted to datasets by AMI
  - `--goodRunListDataType=GOODRUNDATATYPE`
    - specify data type when converting Good Run List XML to datasets, e.g, AOD (default)
  - `--goodRunListProdStep=GOODRUNPRODSTEP`
    - specify production step when converting Good Run List to datasets, e.g, merge (default)
  - `--goodRunListDS=GOODRUNLISTDS`
    - A comma-separated list of pattern strings. Datasets which are converted from Good Run List XML will be used when they match with one of the pattern strings.



Many options available

# More Information



- Documentation about PanDA tools together with analysis examples and FAQ (Frequently Asked Questions):
  - <https://twiki.cern.ch/twiki/bin/viewauth/Atlas/DAonPanda>
  - **pathena examples**: how to run production transformations, TAG selection, on good run lists, event picking, etc.
  - **prun examples**: how to run CINT macro, C++ ROOT, python job, pyROOT script, skim RAW/AOD/ESD data, merge ROOT files, etc.
- Your tutorial page:
  - <https://twiki.cern.ch/twiki/bin/viewauth/Atlas/SouthCaucasusComputingTutorial>
- Regular Offline Software Tutorial page:
  - <https://twiki.cern.ch/twiki/bin/viewauth/Atlas/RegularComputingTutorial>
- How to get support if you need help:
  - [hn-atlas-dist-analysis-help@cern.ch](mailto:hn-atlas-dist-analysis-help@cern.ch)
  - <https://groups.cern.ch/group/hn-atlas-dist-analysis-help/default.aspx>