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

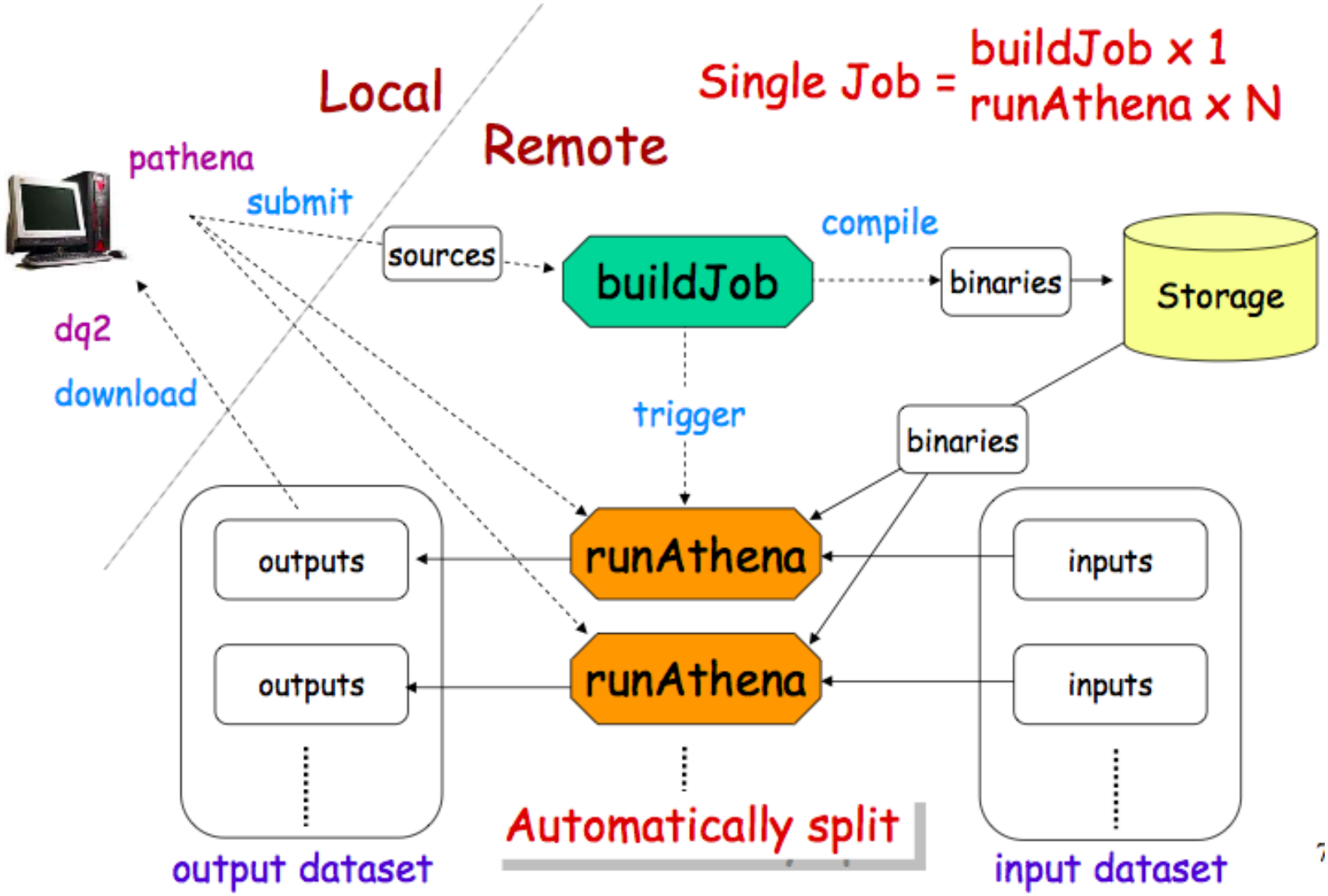
Launching a pathena job


```
$ 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
        PandaID=1132989222
       > run
       PandaID=1132989223
```

Recreates the job's environment at the grid site

Runs the job option file at the grid site

Job Cycle


Local
Remote
Single Job = buildJob x 1
runAthena x N
pathena
submit
dq2
download
sources
buildJob
compile
binaries
Storage
trigger
binaries
outputs
outputs
runAthena
runAthena
inputs
inputs
Automatically split
output dataset
input dataset

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

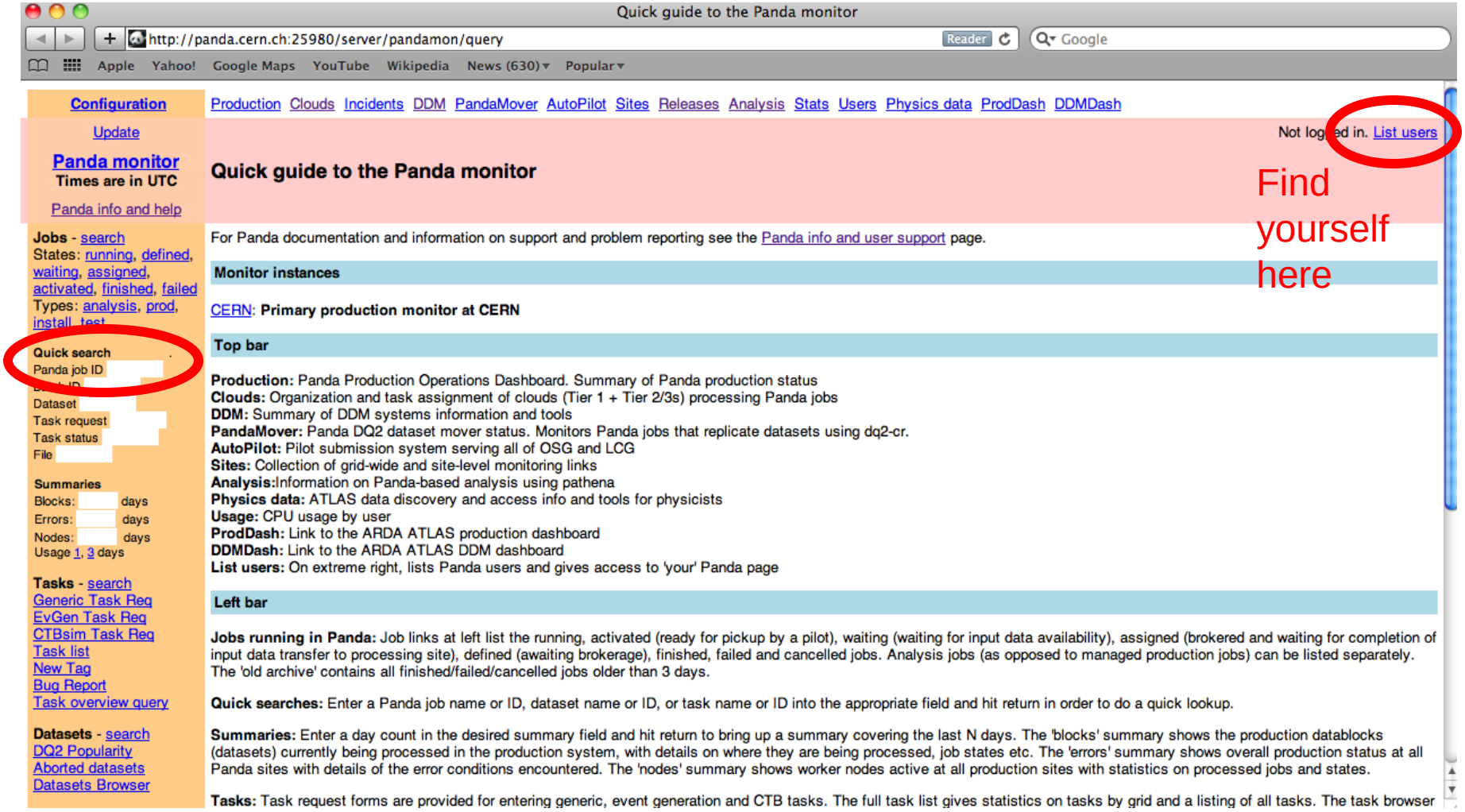
PandaID=1132981750

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

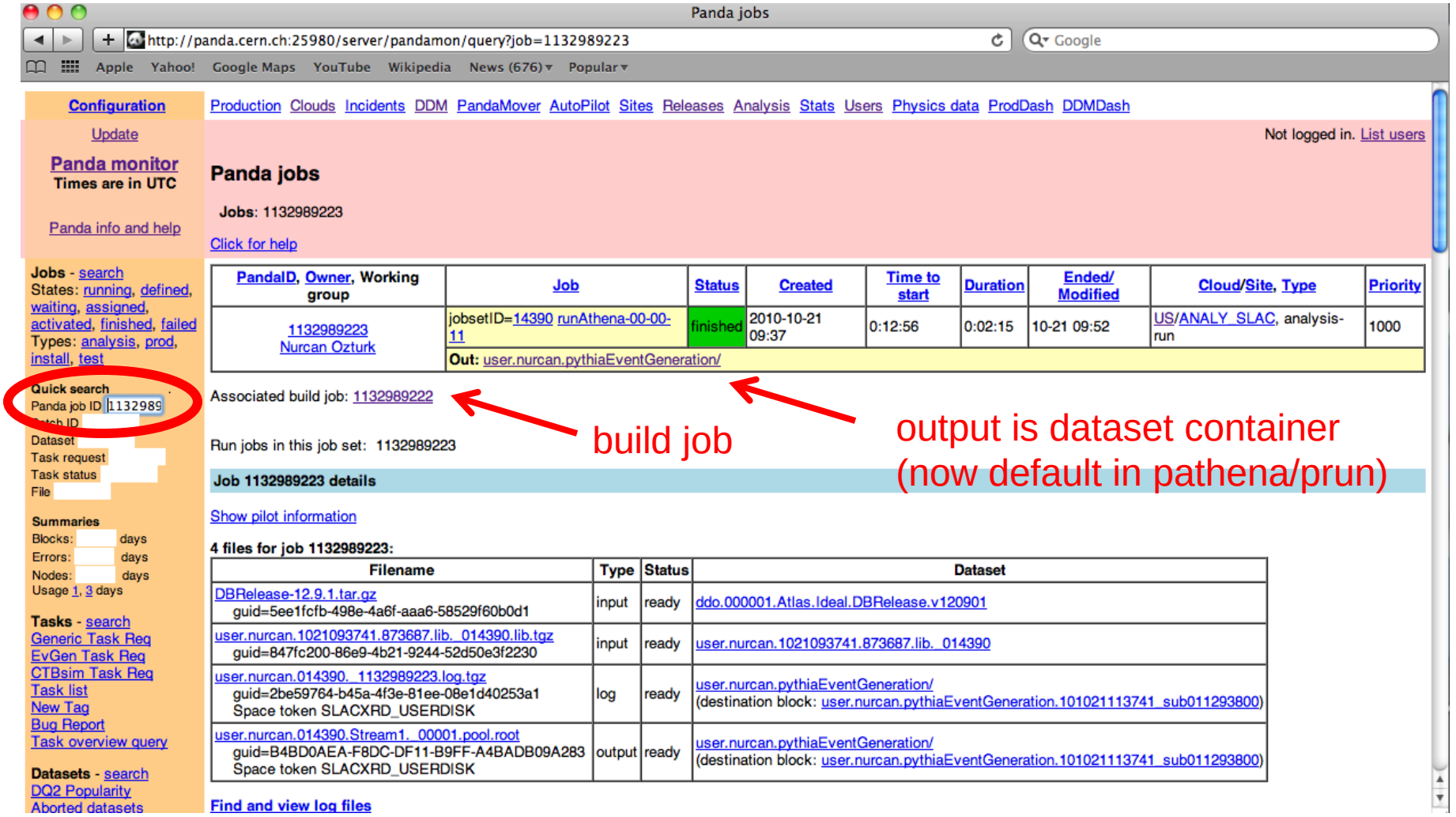
Monitoring a PanDA job (1/2)

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



Monitoring a PanDA job (2/2)

pathena run job finished.



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
  - https://groups.cern.ch/group/hn-atlas-dist-analysis-help/default.aspx