Grid Tools: PanDA Production and Distributed Analysis

Nurcan Ozturk

University of Texas at Arlington

PanDA Client Tools

- **PanDA** is the main ATLAS **Production** and **Distributed Analysis** system
- **PanDA Client** consists of several tools for job execution on the grid and bookkeeping
  - **pathena** – for submitting athena jobs
  - **prun** – for general jobs (e.g. ROOT and Python scripts)
  - **pbook** is a bookkeeping tool for all analysis jobs (used e.g. for job retries and kills)

https://twiki.cern.ch/twiki/bin/view/PanDA/PandaTools#Documentation_on_individual_tool

What is pathena?

- Client tool for PanDA used to submit user-defined jobs from the command line
- Works on the athena runtime environment
- A consistent user-interface to athena

When you run athena with

$ athena jobOptions.py

all you need to do to submit a job to the grid is

$ pathena jobOptions.py [--inDS inputDataset] --outDS outputDataset

where **inputDataset** is a dataset which contains input files (optional), and **outputDataset** is a dataset which will contain output files (required)

- Simple to use, has advanced capabilities for complex needs (see pathena -h)

What is prun?

- **prun** (PanDA run) is a tool for submitting general jobs to PanDA, e.g. ROOT and Python scripts
- ATLAS analysis has two stages:
  - Run athena on DAODs (Derived Analysis Object Data) to produce smaller DAODs (pathena)
  - Run ROOT, Python, or shell scripts to produce final plots (prun)
- PanDA part of the hands-on tutorial will start with sending a simple "Hello world" job to the grid and learning how to monitor the job and find the output!

PanDA Analysis Workflow

JEDI: Job Execution and Definition Interface
https://twiki.cern.ch/twiki/bin/view/PanDA/PandaJediAnalysis
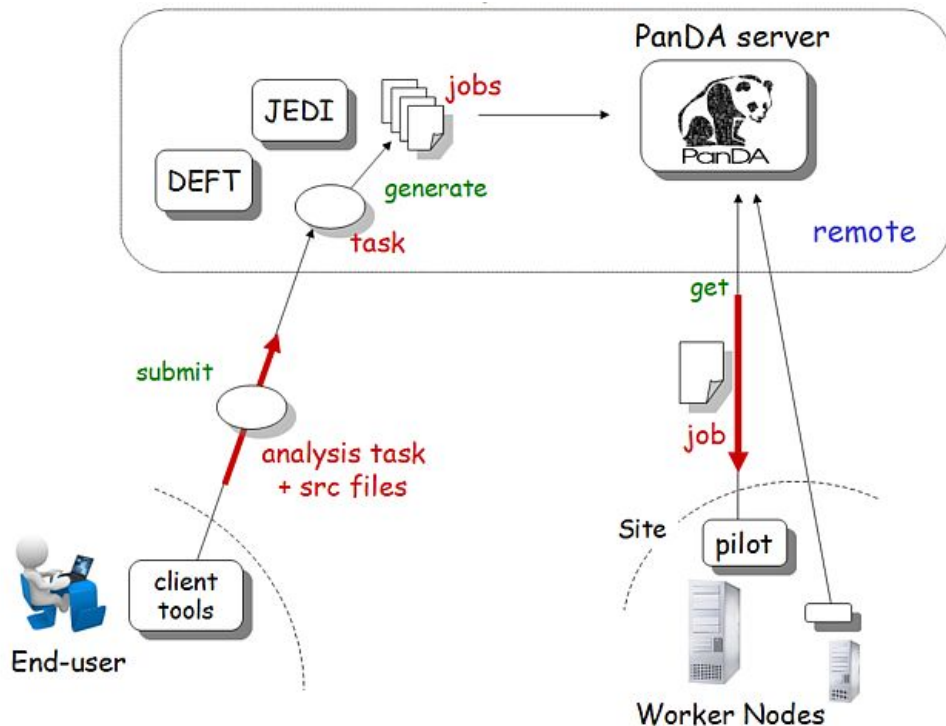
What happens when a job is submitted to the grid?

Make a relocatable tar file using cmake which contains binary and other runtime files

Upload the tar file to remote storage
Executed first on the grid

Executed in parallel when build job finishes (automatically split into sub jobs)

email received when jobs completed
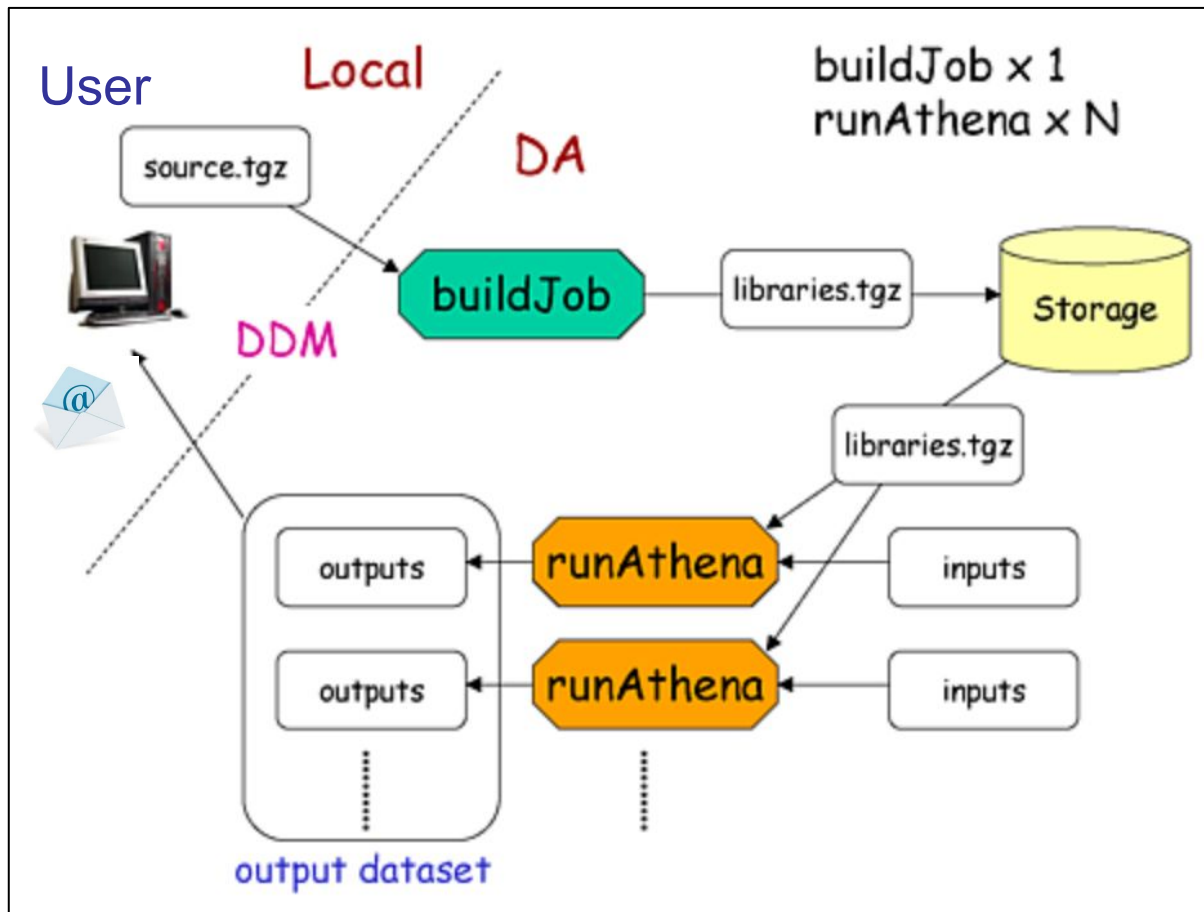
Launching a prun job

- Setup the environment, athena, PanDA client tools as described at:
  - https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/SoftwareTutorialUsingTheGrid

- Make a python file (executable), HelloWorld.py, with lines:
  - #!/usr/bin/python
  - print "Hello world!"

- Submit the prun job:

  $prun --outDS user.nozturk.pruntest.Oct19 --exec HelloWorld.py
  - INFO : gathering files under /eos/home-n/nozturk/PandaTutorial/prunTest
  - INFO : upload source files
  - INFO : submit user.nozturk.pruntest.Oct19/
  - INFO : succeeded. new jediTaskID=19539553 ⟵

Monitoring PanDA jobs (1/5)

Go to: https://bigpanda.cern.ch and enter your jediTaskID

Monitoring PanDA jobs (2/5)

Task is "registered" status first then in "running".

Monitoring PanDA jobs (3/5)

Task is done.

Monitoring PanDA jobs (4/5)

Task details, list of jobs. Click on "PanDA ID" to go to the job details, job log files.

Job list. Sort by **PandaID, time since last state change, ascending mod time, descending mod time, priority, attemptnr, ascending duration, descending duration**

PanDA ID Attempt# of maxAttempts#	Owner Group	Request Task ID	Transformation	Status	Created	Time to start d:h:m:s	Duration d:h:m:s	Mod	Cloud Site	Priority	N input events (N input files)	Maximum PSS	Job info
4521594654 Attempt 1 of 3	Nurcan Ozturk	15181 19539553	runGen-00-00-02	finished	2019-10-23 10:35:38	0:0:10:26	0:0:02:29	2019-10-23 10:54:34	IT ANALY_INFN-COSENZA-RECAS online no active blacklisting rules defined	1000	0 (1)		

click →

run job: runs the job at the grid site

Job name: user.nozturk.pruntest.Oct19/.4521594654#1

Datasets: **Out:** user.nozturk.pruntest.Oct19.log.276582212

| 4521594651 Attempt 0 of 2 | Nurcan Ozturk | 15181 19539553 | buildGen-00-00-01 | finished | 2019-10-23 10:35:38 | 0:0:00:44 | 0:0:02:31 | 2019-10-23 10:44:19 | IT ANALY_INFN-COSENZA-RECAS online no active blacklisting rules defined | 2000 | 0 (0) | | |

Job name: user.nozturk.pruntest.Oct19/#0

build job: recreates the athena environment at the grid site

Datasets: **Out:** panda.1023103535.158040.lib._19539553

Monitoring PanDA jobs (5/5)

How to see the log files in case of failures.

Email notifications from PanDA

atlas-adc-panda-no-reply@cern.ch
JEDI notification for TaskID:19539553 (1/1 All Succeeded)
To: Nurcan Ozturk

Summary of TaskID:19539553

Created : 2019-10-23 10:35:14 (UTC)
Ended   : 2019-10-23 10:54:43.429977 (UTC)

Final Status : done

Total Number of Inputs : 1
        Succeeded : 1
        Failed   : 0
        Cancelled : 0

Error Dialog : None

Log : user.nozturk.pruntest.Oct19.log/

Parameters : prun --outDS user.nozturk.pruntest.Oct19 --exec HelloWorld.py

PandaMonURL : http://bigpanda.cern.ch/task/19539553/

Report Panda problems of any sort to

 the eGroup for help request
   hn-atlas-dist-analysis-help@cern.ch

 the JIRA portal for software bug
   https://its.cern.ch/jira/browse/ATLASPANDA

More information

- Where to get help:
    - hn-atlas-dist-analysis-help@cern.ch (Distributed Analysis Support Team - DAST)

- Tutorial:
    - https://twiki.cern.ch/twiki/bin/view/AtlasComputing/SoftwareTutorialUsingTheGrid

- Frequently Asked Questions (FAQs):
  - https://twiki.cern.ch/twiki/bin/view/PanDA/PandaTools
  - pathena examples: how to run production transformations, multiple transformations in a single job, on good run lists, event picking, etc
  - prun examples: how to run CINT macro, C++ ROOT, python job, AthenaRootAccess job, to skim RAW/AOD/ESD data, merge ROOT files, use containerImage etc

- Please provide feedback on the usage of the tools and the bigpanda monitor to DAST list