ATLAS Analysis Workshop Summary

Matthew Feickert[1]

[1]Southern Methodist University

March 29th, 2016
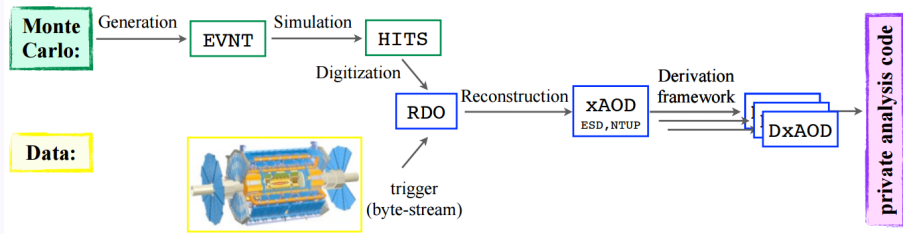
- ATLAS analysis tutorial organized jointly by The University of Chicago and the Argonne National Laboratory Analysis Support Center (ANL-ASC)
- March 7th, 2016 through March 8th, 2016

ATLAS data and simulation workflow

- MC Generator $\rightarrow$ Event Level $\rightarrow$ Detector Simulation $\rightarrow$ Raw Data Object
- Data $\rightarrow$ Raw Data Object
- RDO $\rightarrow$ xAOD $\rightarrow$ DxAOD $\rightarrow$ analysis group/user code $\rightarrow$ physics

Run 2 ATLAS Analysis Model Summary

- xAOD
  - ▸ The reconstruction output is the same as analysis input
  - ▸ ROOT and Athena readable
- Derivation Framework
  - ▸ DxAOD: production of small user xAOD
- Analysis Release
  - ▸ Recommended system for finding and collecting analysis software tools
- CP Tools
  - ▸ Common interface for all analysis tools

How Can I View an xAOD?

- Good news—can quickly browse xAOD with TBrowser

```
$ root -l PATHTOXAOD
root[0] new TBrowser
```

- Can use either interactive ROOT or PyROOT to explore and play with the xAOD to learn more about it and its containers
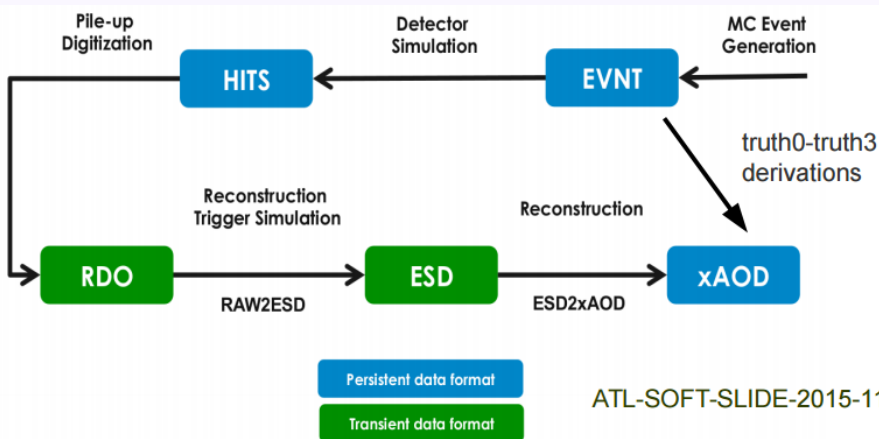
How does ATLAS make Monte Carlo?

- Event generators are used to generate events
  - ► Users control generation and physics parameters through Python job files
  - ► Provides a uniform interface, and allows for dynamic parameter settings

MC Simulation Flow

ATL-SOFT-SLIDE-2015-117

based on CLHEP2015 poster
J.Chapman et al.

MC Simulation Flow

- Most interesting is reconstruction phase
  - Simulation is reconstructed in the same manner as data would be
  - Trigger is fully simulated
  - From the main reconstruction (RAW) Event Summary Data (ESD) is derived
  - From ESD Analysis Object Data (xAOD) is derived via fast slimming process
  - xAOD is then used by both Athena and ROOT

MC TRUTH Derivations

- Can lean how to generate EVNT files and convert to derivations with Sergei's tutorial

Run Query: Everything about real data

- ATLAS website (https://atlas-runquery.cern.ch/) that allows users to query all available run information

Example Query

```
find run 2012.g1 and events 100000+ / show streams phy*
```

- Useful for detector and operation experts and for analysis
- Web tool and command line tool with identical functionality

What would I do with Run Query?

Examples

- access trigger menus for a particular run
- get integrated luminosity information for runs
- access specific machine conditions or bunch intensity profiles for particular run
- · · ·

Can start to get information with nothing more then just the run number.

Conditions (and Configuration) Metadata for ATLAS (COMA)

Metadata for all the run data

- Period Menu: access information on groups of runs (periods)
- Report Mneu: access metadata of specific conditions for runs

ATLAS Metadata Interface (AMI)

"AMI is a generic framework for metadata catalogues."

- Allows for finding names of valid datasets
- Detailed information on datasets
  - ► size
  - ► file counts, event counts
  - ► software config parameters
  - ► MC parameters
  - ► lumi blocks

Rucio

What does rucio do?

- Find datasets and where they are located in the world
- List the files of a dataset
- List the parent dataset for a file
- Retrieve datasets
- Create datasets
- Upload datasets

- Get data and MC with Rucio
- Replacing DQ2 from Run I
- Both a command line and web interface

Federated ATLAS Xrootd (FAX)

Allows direct access to most of ATLAS data from anywhere

- Read only
- Can get data from 64 sites (US, Germany, UK, France, CERN, . . .)
- Corresponds to ~98% of all ATLAS data

FAX allows data to be asked for from an endpoint and will deliver the closest (physically) located version

Why use FAX?

- Increases redundancy
  - If your job requires a file that is inaccessible, FAX will find another copy and your job still runs
- FAX finds files for you
  - Your file names are independent of run site
- Can merge data from multiple physical locations to a central output location
- Remote skimming and slimming of datasets

BigPanDA

- PanDA is the system that lets production jobs and user jobs run on the GRID
- Compared to running on a Tier3 (ManeFrame)
  - Allows for more computing resources
  - Can run on data stored offsite
  - Can take longer for jobs in the queue to run (Global vs. local)
  - Percentage of jobs that die is higher
- When is running on the GRID the best? When size matters
  - Production of ntuples from large DxAODs
  - MC production
  - When xAOD/RAW is required

BigPanDA

After setup

PanDA: HelloWorld

```
$ mkdir PanDA-helloworld
$ cd PanDA-helloworld
$ lsetup panda
$ echo 'print "hello world"' > helloworld.py
$ prun --outDS user.YOURNAME.testing.helloworld_v1 --exec 'python helloworld.py'

INFO : succeeded. new jediTaskID=1234567
```

Then can monitor tasks on the BigPanDA web interface

BigPanDA Tutorials/Resources

- ATLAS Computing Software Tutorial Using the Grid
- prun Examples
- pathena Examples

Backup