The Physics Analysis Tools project for the ATLAS experiment

Bruno Lenzi

CEA - Saclay

30/09/2009

ICCMSE 2009
Rhodes, Greece

Physics Analysis Tools project

Motivation

- The analysis of ATLAS data (≈ 1 PB per year) poses a great challenge to the ATLAS computing
- Computing Model and Event Data Model defined

Mission

- Provide tools to help users to perform physics analysis, “hiding” details of the ATLAS framework “ATHENA”

Physics Analysis Tools (PAT) group

- 2 convenors
- About 20 “permanent” contributors (0.1 - 0.4 FTE)
- More than 2000 clients $\longleftrightarrow$ contributors

Outline: the Physics Analysis Tools actions

- 1 The ATHENA framework and the ATLAS Event Data Model
- 2 Filtering and selecting the ATLAS data
- 3 Reading and analysing the ATLAS data
- 4 Common analysis tools
- 5 Conclusions

The ATHENA framework

ATHENA is the standard framework for simulation, reconstruction and physics analysis in ATLAS

- Handles the configuration and execution of various C++ packages
- Configuration defined via python scripts (jobOptions):
 - What to run, in which order
 - How many times, input/output
- Components:
 - **Algorithm:** application that run once per event and read / produce data
 - **Tool:** piece of code shared between algorithms
 - **Data object:** physics (muon, electron) or detector objects (cells)
 - **Service:** globally available software
 - Eg: data access (StoreGate), message printing
 - **Other components:** Auditor, sequence, filter, converter, property

The ATHENA framework

The ATLAS Event Data Model

Event Summary Data (ESD)

- Output of reconstruction
 - Tracks and hits
 - Calorimeter cells
 - Combined reconstr. objects
- POOL/ROOT format

Analysis Object Data (AOD)

- Summary of reconstruction
 - “Physics” objects
 - e , μ , τ , jets
 - Few detector obj.
- POOL/ROOT format

Filtering and selecting the ATLAS data

PAT group responsible for providing tools for data selection

Derived Physics Data (DPD)

- Selection of events, objects, object information from Analysis Object Data
 - Can be flat ROOT ntuple in the latest stage
- Specific for each physics and performance (sub-)group
- Input for physics analysis, together with AOD

TAG: Event-level metadata

- Database (or ROOT files) used to quickly select events in AOD and/or ESD
- Generated when Analysis Object Data are produced / merged
- About 200 variables for each event
 - Event identification, trigger decisions, number of physics objects, ...

Filtering and selecting the data: TAGs

- TAG files can speed up the analysis by skipping uninteresting events
 - Used in the production of Derived Physics Data
 - Exercised during detector commissioning
 - Helped spotting first events in ATLAS

Filtering and selecting the data: TAGs

- TAG files can speed up the analysis by skipping uninteresting events
 - Used in the production of Derived Physics Data
 - Exercised during detector commissioning
 - Helped spotting first events in ATLAS
- TAG database replicated
 - Web server can output tables, histograms and ROOT files

The diagram illustrates the ELSSI (Event Level Selection Service) web interface and its outputs. The interface is divided into several sections:

- Banner:** Welcomes the user to ELSSI and provides contact information.
- Browser:** Allows users to search for data using various criteria like Temporal cut, Kinematics, Data Quality, Trigger, and Physics attributes.
- Session:** Provides a way to save and manage search sessions.
- Selection Criteria:** A section where users can define their search criteria, including a range of values for specific variables.
- Results:** Displays the results of the search, including a table of data points and histograms.

The **Results** section shows a table of data points for two variables: **LOOSEELECTRONETAI** and **LOOSEELECTRONPHI**. The table displays the first 100 of 3602 rows. The data points are as follows:

LOOSEELECTRONETAI	LOOSEELECTRONPHI
-191533968	-3.0434804
-312644333	-353945196

Below the table, two histograms are shown, one for **LOOSEELECTRONETAI** and one for **LOOSEELECTRONPHI**, displaying the distribution of the data points.

Reading and analysing the ATLAS data

Standard way: ATHENA framework

- Services, Tools, Algorithms in C++
- All functionalities with the benefits and expenses

Alternatives provided within the scope of Physics Analysis Tools project

- Like a ROOT Tree using AthenaROOTAccess
- Still in ATHENA but with PyAthena
- Using ntuples (like D3PDs) directly in ROOT
 - Need to be produced inside ATHENA
- Other frameworks

Reading the ATLAS data: AthenaROOTAccess

Allows accessing POOL data (ESD / AOD / DPD) directly from ROOT

- Sets up a transient TTree, with branches corresponding to StoreGate keys
 - Familiar environment: TTree::Draw / TBranch::GetEntry
 - First plots within seconds!
- Invokes automatically POOL converters (when a specific entry is requested)
 - Relies on the ATHENA infrastructure (classes / converters)
- Complete analysis can be done in C++ and/or python

Reading the ATLAS data: AthenaROOTAccess

Allows accessing POOL data (ESD / AOD / DPD) directly from ROOT

- Sets up a transient TTree, with branches corresponding to StoreGate keys
 - Familiar environment: TTree::Draw / TBranch::GetEntry
 - First plots within seconds!
- Invokes automatically POOL converters (when a specific entry is requested)
 - Relies on the ATHENA infrastructure (classes / converters)
- Complete analysis can be done in C++ and/or python

Limitations

- Needs ATHENA installation
- Databases, detector description, services, some tools, ... are not available
- Objects cannot be read if they require the above (eg: calorimeter cells)

Reading the ATLAS data: PyAthena

Allows the creation / retrieval of ATHENA components in python

- C++ components just need a dictionary to be loaded
- No compilation, good option for interactive use and debugging

Development cycle in ATHENA (C++)

Reading the ATLAS data: PyAthena

Allows the creation / retrieval of ATHENA components in python

- C++ components just need a dictionary to be loaded
- No compilation, good option for interactive use and debugging

Development cycle in ATHENA (C++)

Reading the ATLAS data: PyAthena

Allows the creation / retrieval of ATHENA components in python

- C++ components just need a dictionary to be loaded
- No compilation, good option for interactive use and debugging

Development cycle in ATHENA (C++)

Reading the ATLAS data: PyAthena

Allows the creation / retrieval of ATHENA components in python

- C++ components just need a dictionary to be loaded
- No compilation, good option for interactive use and debugging

Development cycle in ATHENA (C++)

Development cycle in PyAthena

Reading the ATLAS data: PyAthena

Allows the creation / retrieval of ATHENA components in python

- C++ components just need a dictionary to be loaded
- No compilation, good option for interactive use and debugging

Development cycle in ATHENA (C++)

Development cycle in PyAthena

Reading the ATLAS data: other frameworks

Several analysis frameworks exist on top of ATHENA and/or AthenaROOTAccess

Examples

- EventView
- Every Where Physics Analysis
- Atlas Modular Analysis
- ATLAS Watchman

Main applications

- Configure packages
- Convert data
- Dump ROOT ntuples
- Add user data

Common analysis tools

Physics Analysis Tools group collects and maintains code to perform physics analysis with ATHENA and/or other frameworks

Some examples:

- Data Handling tools (grid)
- Tools for Monte Carlo corrections
- Detector performance measurement and bookkeeping tools
- Tools for matching trigger, monte carlo and reconstructed objects
- **Isolation tools for different types of objects**

Common analysis tools: Isolation tools as an example

- Isolation is related to the activity around a given particle
 - Can help identifying the process that originated it
 - One of the main discriminants in Standard Model analyses, Higgs searches and beyond
- In practice we talk about two types:
 - Track isolation
 - Calorimeter isolation

Common analysis tools: Isolation tools

Example: calorimeter isolation for muons

- Part of standard muon reconstruction
- Interesting quantity: $\sum E_T$
 - Energy deposited around the particle for a given cone
 - Attached to every *muon* object
- Muons leave energy in the calorimeter
 - **Core energy** → has to be subtracted

Integration

- Interface for muons and tracks
- Core energy follows the calorimeter segmentation
 - Configurable for each particle / analysis

Common analysis tools: Isolation tools

- Isolation tools ran in the end of reconstruction
 - Attach variables to physical objects (“dressing”)
- Information can be recalculated at analysis level
 - Input can be ESD, AOD or DPD
- Different implementations exist for:
 - Tracks
 - Muons
 - Electrons / photons
- Interfaces were unified
 - Specific issues handled in each implementation

Conclusions

■ Analysing the ATLAS data is a major challenge

- Framework, computing model and data types defined
- More than 2000 clients $\longleftrightarrow$ developers

■ Physics Analysis Tools project ongoing

- Providing ways to filter and select data
 - Managing tools and lightweight data formats (TAG)
- Supporting alternatives to ATHENA framework
 - From ROOT standards to complete frameworks
- Maintaining general tools for analysis
 - Integration between different communities