ATLAS analysis workflows using the EventIndex and the Event Picking Server for massive event picking and enhanced processing

CHEP2024 – Computing in High Energy and Nuclear Physics 2024, 21-25 October 2024, Kraków, Poland

D. Barberis[1*], E. Alexandrov[2], I. Alexandrov[2], A. Dimitrievska[3], A. Iakovlev[2], S. Pagan Griso[4], S. Turchikhin[1]

[1] University of Genoa and INFN, Genoa, Italy, [2] JINR, Dubna, Russia,
[3] University of Birmingham, United Kingdom, [4]LBNL, Berkeley, USA
[*]Presenter and corresponding author

EventIndex: an event-level metadata catalogue for all ATLAS events

Use cases:

- **Event Picking.**
- **Counts or selections.**
- **Overlaps:**
  - of triggers in a dataset.
  - of events between derivations.
- **Production checks.**

Architecture:

- Designed to scale for **big data** [1,2]:
  - Able to keep **trillions** of event records.
  - Able to ingest 10k records/s.
- Data platform with open-source components.
  - Hadoop ecosystem:
    - HBase, Phoenix, YARN, Spark, Scala, HDFS.

Event record content:

Event records with immutable event information:

- Run and event number.
- Event location (GUID).
- Provenance.
- Trigger information.
- Luminosity block, Bunch crossing identifier.

Event Picking Server: an automated tool to search and retrieve large numbers of single events

Use case:

- **Automate event picking for large requests**
  (from thousands to millions of events across all ATLAS data)

Architecture:

- Three components [3,4]:
  - GUI for user requests, monitoring and results
  - Daemon to process the requests
  - Backend database (PostgreSQL) to store the requests and their status/progress

Workflow:

- The user submits a request through the GUI, supplying a list of run/event numbers, data type and (if needed) trigger and other auxiliary information
- The Daemon does the bulk of the work:
  - splits the list by run number,
  - queries the EventIndex to retrieve the GUIDs of the files with the events,
  - submits event picking jobs to the ATLAS PanDA distributed workflow management system,
  - collects the output files into datasets placed at CERN,
  - notifies the user of completion.
- The user can monitor the progress through the GUI, then retrieve the output datasets and process the events

Analysis workflows with massive event picking

Some physics analyses need direct access to lower-level detector information than is available in processed data (AODs or DAODs)

- **To apply improved calibrations with respect to those available at the time of global processing**
- **To run additional reconstruction algorithms that would be too time-consuming if run on all events**

These analyses can be split into different stages:

First analysis stage:

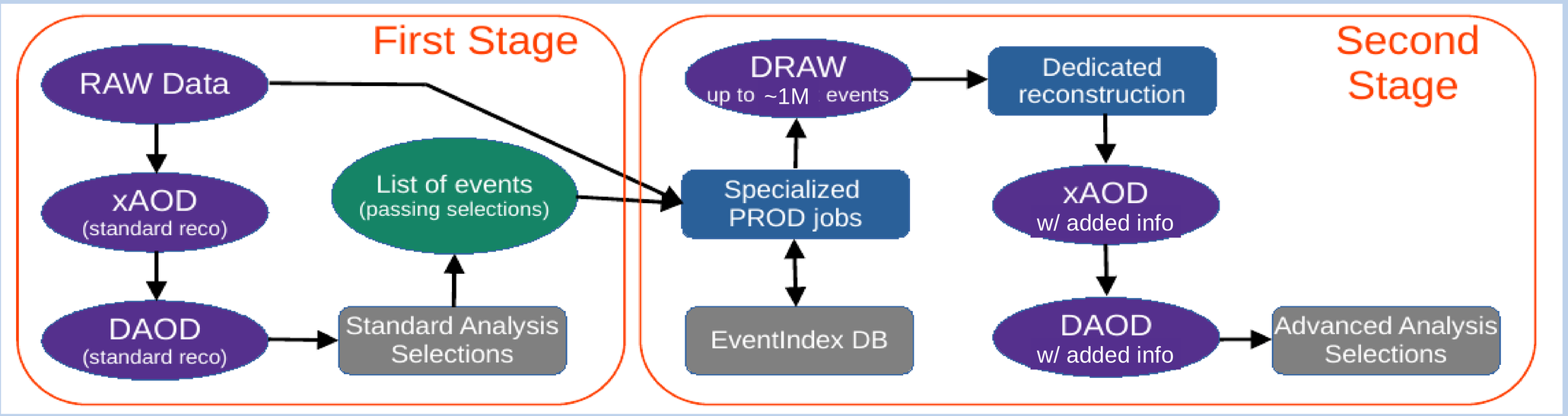
- **Run the analysis on the output of standard reconstruction, normally in DAOD format**
- **Create a list of selected events (signal, background and validation regions)**

Event Picking stage:

- **Use the Event Picking Server to extract the selected events from the RAW data files (so far ~30 billion events on tape)**
- **Collect them into datasets grouped by run number**

Second analysis stage:

- **Apply dedicated calibrations and reconstruction algorithms to the selected events, adding information to the AOD and DAOD formats**
- **Complete the analysis using the additional info**



Examples of massive event picking for physics analyses

Analysis of γγ→WW scattering

- Main background is suppressed by requiring isolation of reconstructed tracks; residual background remains from non-reconstructed low-p_T particles in underlying event
  - Reconstruction of low-p_T particles with a custom algorithm helps to reduce the background but is time-consuming
- (2019) **50k events** selected for signal and control regions using standard reconstruction were extracted with massive event picking, processed with low-p_T tracking and analysed

$B_c^{\pm *} \rightarrow B_c^{\pm}(\rightarrow J/\psi\, \mu^{\pm}\nu_{\mu})\, \gamma(\rightarrow e^{+}e^{-})$ decays

- Low-p_T tracking needed to reconstruct low-energy photon conversions ($p_T(e^{\pm}) > 50$ MeV)
  - Gives factor 100 more acceptance than standard tracking threshold ($p_T(e^{\pm}) > 500$ MeV)
- (2023) **650k events** selected using the Event Picking Server and processed with enhanced reconstruction

Search for long-lived SUSY particles

- Particles with $m > 100$ GeV and $\tau > 1$ ns would leave large ionisation energy deposits in the pixel detectors
- The collected charge decreases with radiation damage, so module and time-dependent calibrations are needed
- (2024) **Over 1M events** selected in signal and background regions to apply dE/dx recalibrations before final analysis

References

[1] Barberis D. *et al.*, "The ATLAS EventIndex: A BigData Catalogue for All ATLAS Experiment Events", *Comput. Softw. Big Sci.* 7 (2023) 1, 2
[2] Gallas E. J. *et al.,* "Deployment and Operation of the ATLAS EventIndex for LHC Run 3", *EPJ Web Conf.* 295 (2024) 01018
[3] Alexandrov E. I. *et al.*, "Development of the ATLAS Event Picking Server", *Proc. 9th Int. Conf. on Distributed Computing and Grid Technologies in Science and Education* (2021) 223-228
[4] Alexandrov E. I. *et al.*, "The ATLAS Event Picking Service and Its Evolution", *Phys. Part. Nucl.* 55 (2024) 3, 437-440