EventIndex Operations
Massive Event Picking

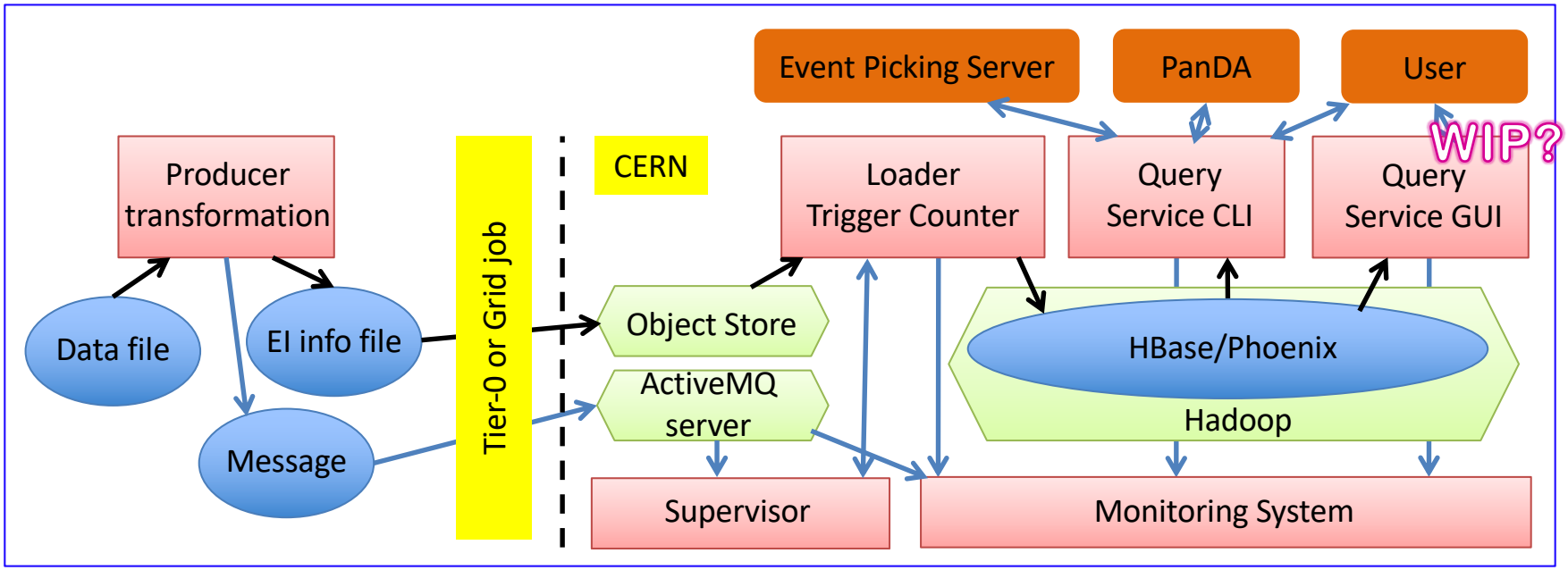
Dario Barberis

University of California, Berkeley (USA)

INFN Genova (IT)

EventIndex in Run 3

Functional schema and data flow

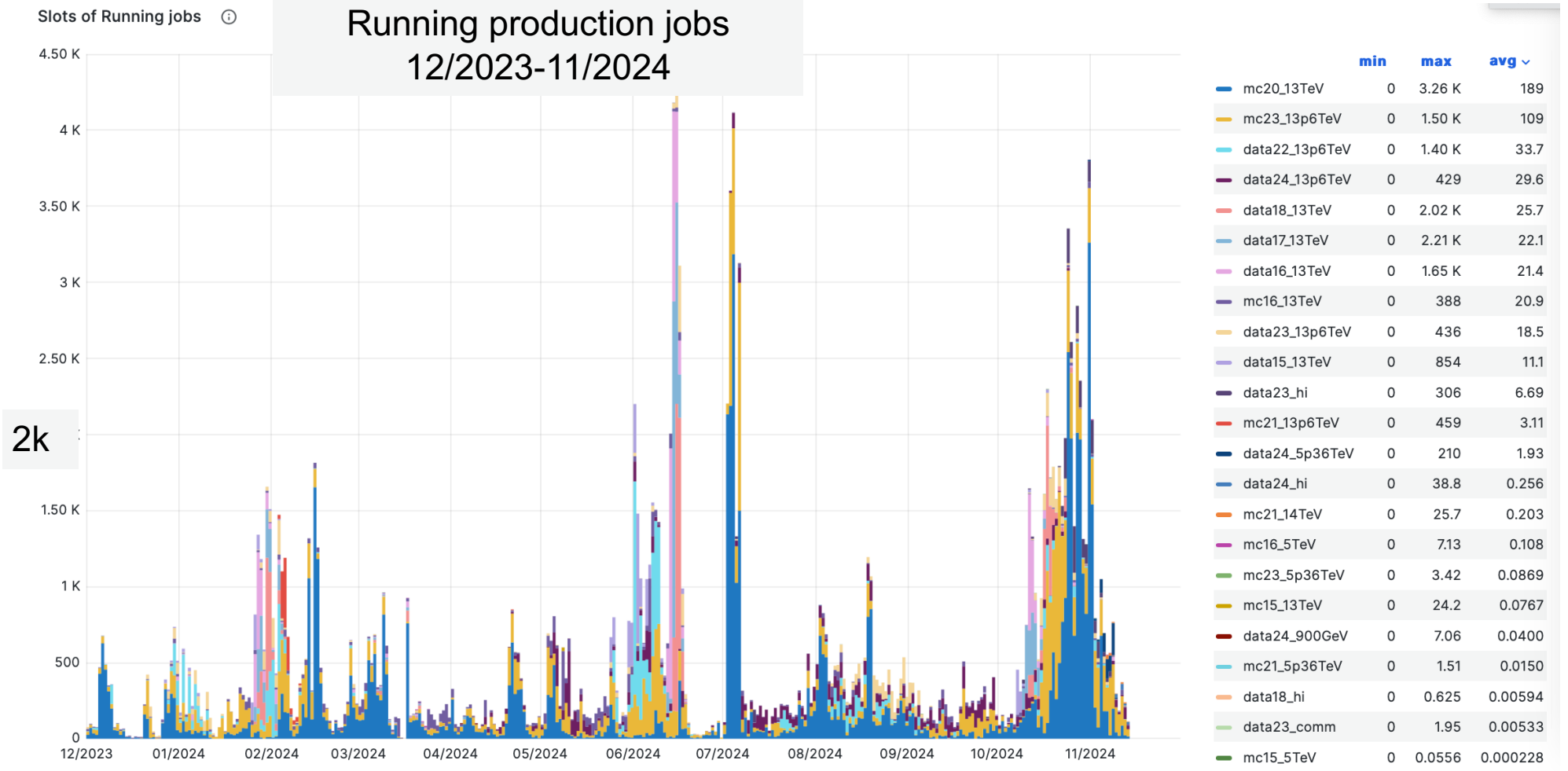


Most components updated or replaced during LS2

All major functionalities in stable operation

A few loose ends remaining

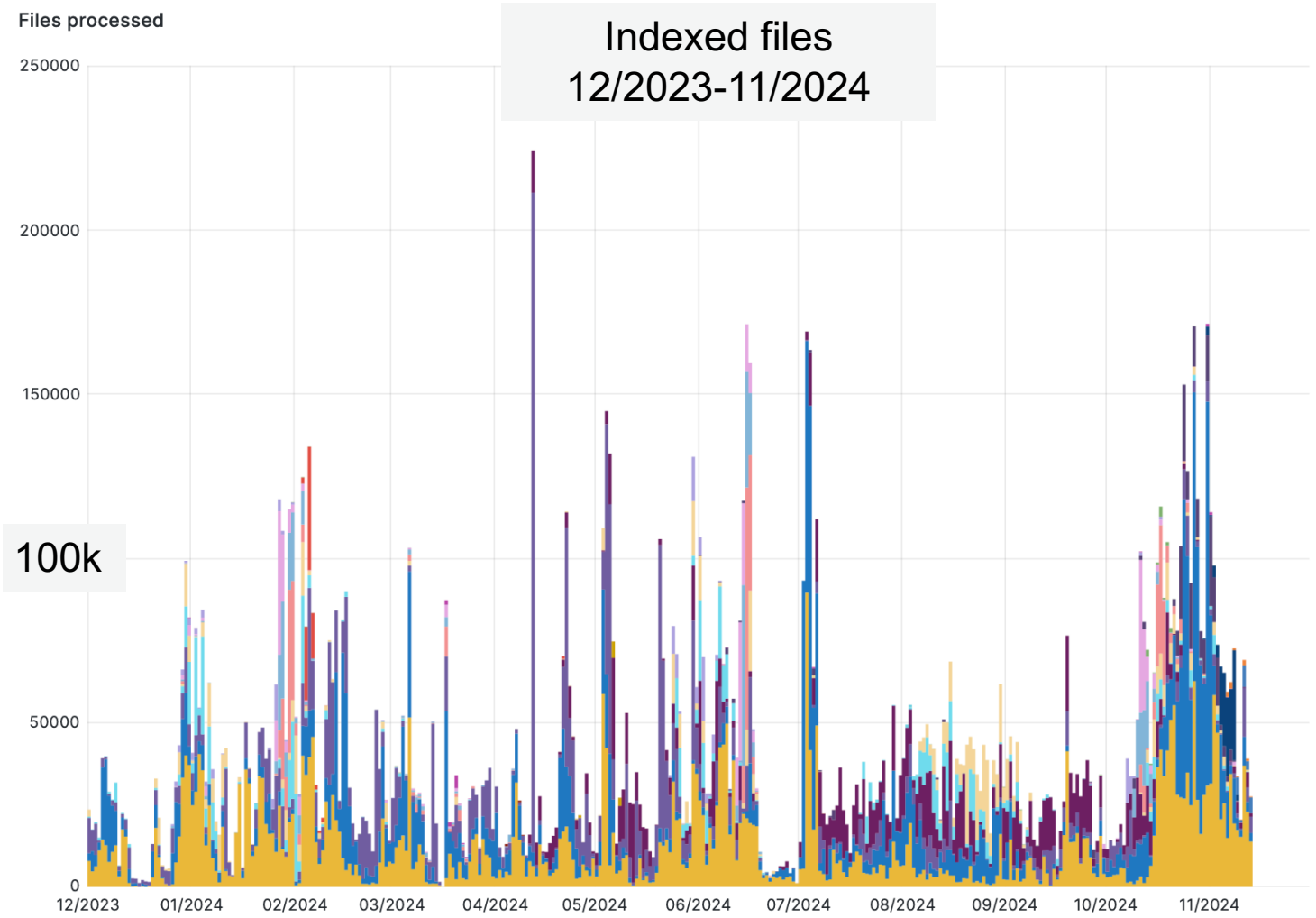
2024 Operations (1)



Average: 500

https://monit-grafana.cern.ch/d/000000696/job-accounting-historical-data?orgId=17&var-bin=1d&var-groupby=inputfileproject&var-processingtype=eventIndex&var-processingtype=t0_eventindex&var-es_division_factor=24&var-pledges=CPU&from=1701388800000&to=173301

2024 Operations (2)

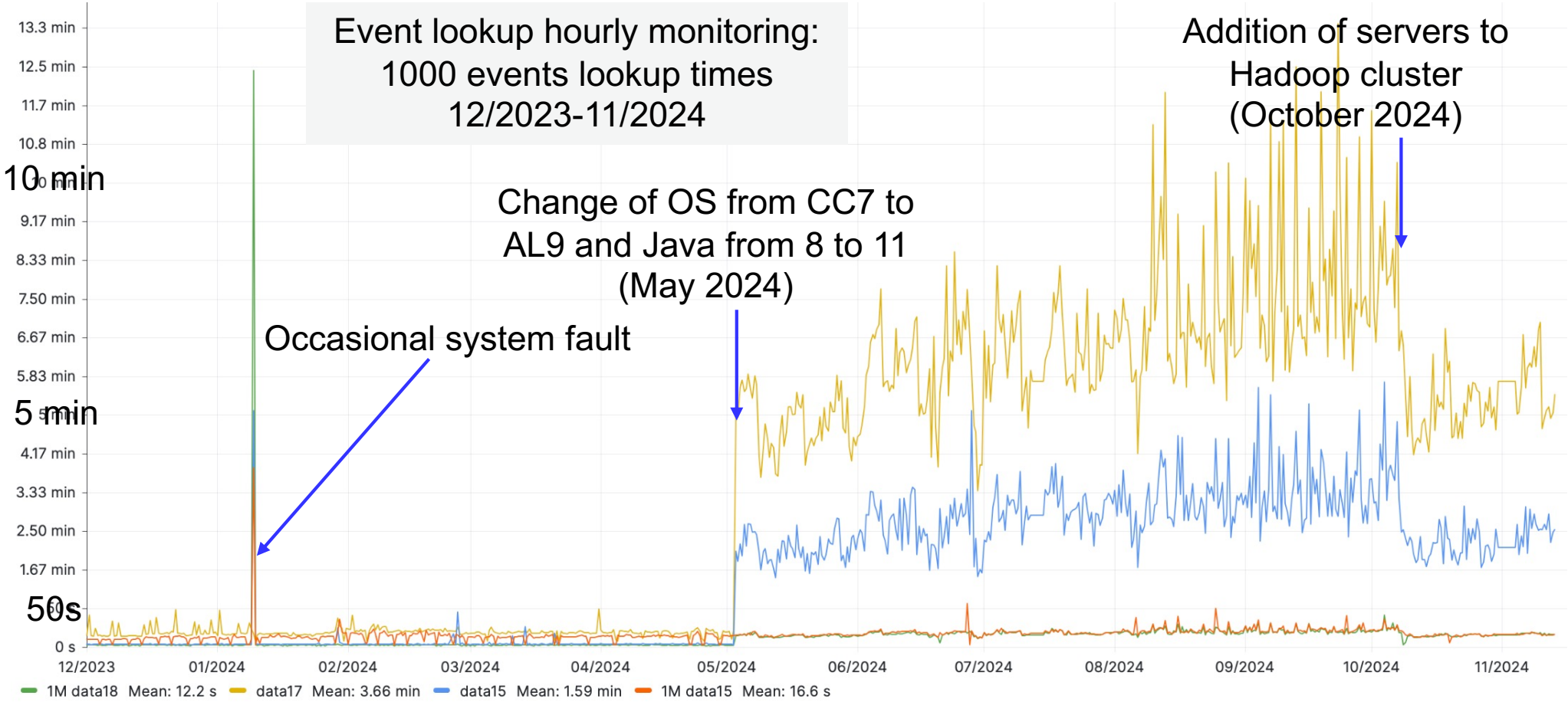


	min	max	avg	total
mc23_13p6TeV	0	89573	11911	4359269
mc20_13TeV	0	116895	9163	3353683
mc16_13TeV	0	200827	6718	2458929
data24_13p6TeV	0	30905	4354	1593589
data22_13p6TeV	0	29846	1964	718801
data23_13p6TeV	0	30245	1873	685587
data18_13TeV	0	73765	1193	436675
data17_13TeV	0	54989	1007	368500
data16_13TeV	0	46476	881	322609
data23_hi	0	27323	631	230812
data15_13TeV	0	19395	374	136946
data24_5p36TeV	0	33528	352	128714
mc21_13p6TeV	0	37514	219	80260
data24_hi	0	6604	54.9	20101
mc23_5p36TeV	0	3099	31.4	11486
mc16_5TeV	0	3924	29.2	10699
mc15_13TeV	0	4999	23.1	8447
mc21_14TeV	0	2034	20.4	7456
mc21_5p36TeV	0	875	5.62	2056
data24_900GeV	0	611	3.64	1332
data18_hi	0	95	0.825	302
data23_comm	0	271	0.740	271
mc15_5TeV	0	80	0.301	110

Total: 15M

https://monit-grafana.cern.ch/d/000000696/job-accounting-historical-data?orgId=17&var-bin=1d&var-groupby=inputfileproject&var-processingtype=eventIndex&var-processingtype=t0_eventindex&var-es_division_factor=24&var-pledges=CPU&from=1701388800000&to=173301

2024 Operations (3)



https://monit-grafana.cern.ch/d/oGCeepxnz/atlas-ei-primary-dashboard?from=1701385200000&orgId=17&to=1733007599000&viewPanel=33

Investigations pending but lacking dedicated personpower

Work done since Nov 2023

- Infrastructure:
  - Upgraded the operating system of all servers to AlmaLinux9
- Athena:
  - Moved job configuration from jobOptions to Component Accumulator
- Data Production and Supervisor:
  - Moved to Python 3
- Event Picking Server:
  - (More) Robust error recovery
  - Tests of massive event picking in view of R2R4 Milestone DB4.1
    - Demonstrate the ability to pick 1M events in 3 weeks by the end of 2024

Work needed

- Supervisor:
  - Integration of Data Production scripts 
  - Check and cleanup of obsolete datasets
  - Duplicated event counts by dataset
  - Trigger Counter and Query interface
- COMA interface:
  - Needed for production completion checks and web interface to event lookup
- Event Picking Server:
  - Formation and correct naming of final dataset 
  - Periodic clean-up of EOS space used for output 
- Monitoring:
  - Streamlining of Grafana dashboards
    - Removal of obsolete data sources and plots
    - Dashboard re-organisation
  - Completion of data sources to include client access statistics vs time

Massive event picking for analysis

- Standard calibration and reconstruction procedures reduce this information to physics objects that can be used as input to most analyses, but not all:
  - Some very specific analyses need full information from some of the ATLAS subdetectors, or enhanced calibration and/or reconstruction algorithms.
- For these use cases, a novel workflow has been developed:
  - First select events satisfying some basic criteria
  - Then extract them in RAW data format using the EventIndex data catalogue and the Event Picking Server
  - Finally apply specialised processing algorithms
- This workflow allows us in addition:
  - to commission and use new calibration and reconstruction techniques before launching full reprocessing campaigns
  - to use algorithms and tools that are too CPU or disk intensive if run over all recorded events
  - in the future to apply AI/ML methods that start from low-level information and could profit from rapid development/use cycles

Event Picking Server

Use case:

- **Automate event picking for large requests**

  (from thousands to millions of events across all ATLAS data)

Workflow:

- The user submits a request through the GUI, supplying a list of run/event numbers, data type and (if needed) trigger stream, AMI tags and other auxiliary information
- The Daemon does the bulk of the work
- The user can monitor the progress through the GUI, then retrieve the output datasets and process the events

Architecture:

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

The Daemon:

- 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.

https://atlas-event-picking.cern.ch

Analysis workflow

First analysis stage:

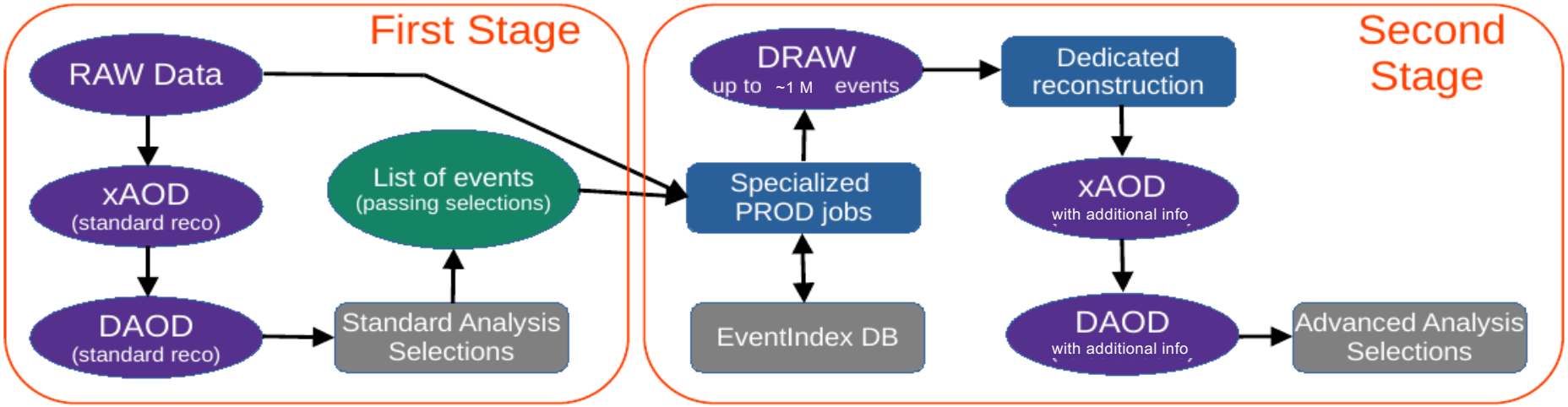
- Run the analysis on the output of standard reconstruction, normally in DAOD format
- Create a list of selected events (signal, background and validation regions)

Event Picking stage:

- Use the Event Picking Server to extract the selected events from RAW data files (~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

Analysis of the 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 reducing 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

Analysis of $B_c^{\pm *} \rightarrow B_c^{\pm}(\rightarrow J/\psi\, \mu^{\pm}\nu_{\mu})\, \gamma(\rightarrow e^+e^-)$ 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
  - Collected change is measured and digitised in each pixel through the Time-Over-Threshold (ToT)
  - The collected charge decreases with radiation damage (a factor 3 between 2015 and end 2024), 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

R2R4 Milestones and Tests

- DB4.1: Event picking of 1M events in RAW and AOD format in 3 weeks (Dec 2024)
  - Partial test with ~300k RAW events run in August 2024
    - Discovered problems with the EPS daemon getting stuck from time to time and overload of the aiatlas169.cern.ch machine where the daemon and GUI run
  - New test with ~100k events run in October to check fixes
    - Discussion on possible tape overload when staging many files; suggestion to explore using the Tape Carousel within ProdSys
  - Small scale test (18k events) early November as a system test
    - All OK, finished in <4 days
  - "50% test" (500k events) mid-November as a stepping stone towards DB4.1
    - PanDA throttled most tasks (too much data staging) – nothing really moving
      - NB: picking 1M events means staging from tape 5% of RAW data files
    - Discussion with ADC/WFMS team planned this Thursday to see how to use the Tape Carousel and have smooth operation
- DB4.2: Test the data production, collection and ingestion rates at HL-LHC rates for 3 weeks (Mar 2025)
  - Needs first implementation of data store on the Hadoop QA cluster, to be used as test setup

Outlook

- The system as it is now basically works for its major use case (event picking of a handful of events)
  - It is still incomplete relative to its implementation for Run 2
- Massive event picking works reliably till ~100k events
  - Tests of milestone DB4.1 (1M events) showed some problems that in principle can be overcome
- Next milestone DB4.2 will show us if this basic infrastructure can take the load foreseen for Run 4
  - But 2030 is rather far away…