HL-LHC Analysis Blueprint Meeting

Alexander Held (University of Wisconsin–Madison)
Oksana Shadura (University Nebraska–Lincoln)

In collaboration with CMS CAT & ATLAS AMG working groups and the HSF Data Analysis Working Group

ATLAS / CMS LHCC questions session

Feb 3, 2025 https://indico.cern.ch/event/1501541/

This work was supported by the U.S. National Science Foundation (NSF) cooperative agreements OAC-1836650 and PHY-2323298 (IRIS-HEP).

Motivation & context

- **LHCC Analysis Infrastructure (AI) questions 3c) and d)**
  - *Describe the plans to **develop specific use cases that can be used to benchmark different building blocks** of the Analysis Infrastructure so that a comparison can be made between different implementations.*
  - *Comment if you think that support for analysis workflows in Run-4 will need **specialized infrastructure** different from the Grid. If so, please describe what features that Analysis Infrastructure will need to provide to expand the one in the Grid*

- **Motivation:** Ensure **R&D is aligned** with future needs, clarify **AI requirements**
  - ATLAS/CMS do HL-LHC *physics* extrapolations, should extend also to *computing*!

- We focus on the **"end-user" physics analysis**: the steps after centrally organized production
  - This is generally **not well understood** nor **prescribed**

Path forward: survey, meeting, document

- **Proposed steps towards a set of HL-LHC analysis examples**
  - 1) **survey** within ATLAS & CMS experiments
  - 2) **blueprint meeting** discussing selected analysis examples
  - 3) **document with benchmark** analysis examples

- A variety of related but **differently scoped surveys** has been done previously
  - Limited in target audience (e.g. only US, site admin perspective, specific physics groups)

- Not aware of data comprehensively capturing **“end-user” analysis compute requirements**
  - Including scenarios for **extrapolation to the future** with respect to computing needs

Not duplicating efforts:
Expecting that a new survey can yield very valuable information.

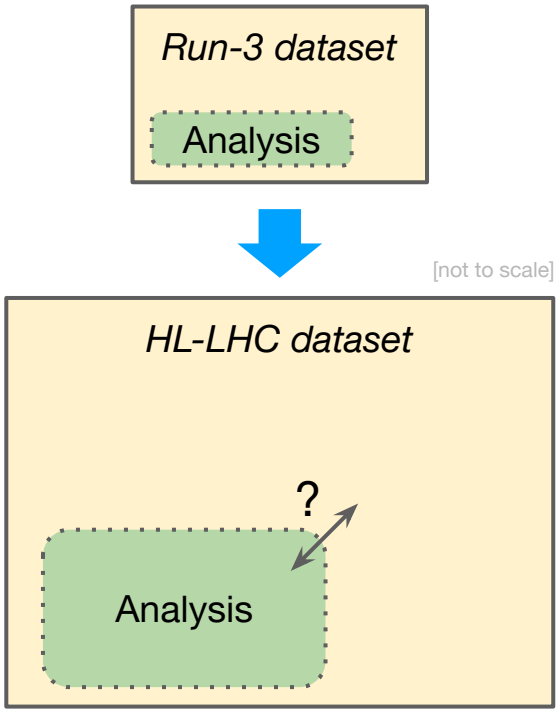
Aspects we want to capture

- **Dataset size**
- Compute **cost per event** / event rate
- **Workflow** structure (complexity, number of steps, intermediate products, optimization loops, …)
  - Including how work is **distributed across the team**
- **Service requirements** (ML inference, external databases, …)
- **Physics** target (precision measurement, model-specific search, anomaly detection, …)
- **Reproducibility** aspects
- Currently used **computing setup** (which hardware resources are used)

workflow example [https://agc.readthedocs.io/]

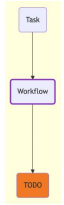
Extrapolating to HL-LHC

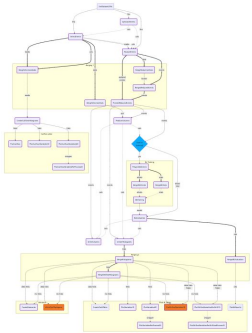
Dataset size



[New analysis approach / selection?]
[Linear scaling with integrated luminosity?]

Complexity

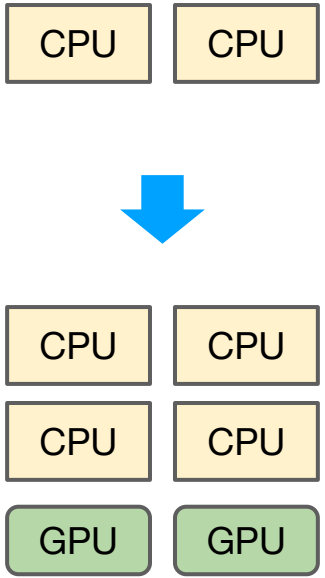




columnflow examples [https://github.com/columnflow/columnflow/wiki]

[External services?]
[Intermediate data products?]

Compute cost



[Heavy ML training / inference?]
[Special requirements, e.g. high memory?]

The blueprint meeting

- **Hybrid / virtual meeting** meeting in May
  - Targeting **two afternoons** during CERN time
- Invite speakers from **selected set of example analyses** based on survey results
- Identify set of **benchmark analyses** and **extrapolation scenarios** to HL-LHC dataset
- Capture outcome of meeting in a **document**
  - This need not be static: can **revisit assumptions** if our understanding evolves

- More information to appear at https://iris-hep.org/blueprints/HL_LHC_analysis.html
  - Official announcement of event to follow

Next steps

- **Survey design** currently ongoing
  - Will iterate with **ATLAS AMG** & **CMS CAT** for expert feedback
- Survey to be **sent out in early March**, target a 4 week circulation period
- **Analyze responses** in April, invite speakers
- Blueprint **meeting in May**

- **Feel free to get in touch** with us directly for any feedback
  - oksana.shadura@cern.ch & alexander.held@cern.ch