The Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP) was established in 2018 to close the gap between LHC computing needs and expected resources and to optimize the interfaces used by physicists for data analysis and FPGAs to accommodate the full HL-LHC dataset (Figure 1).

In 2026, the High Luminosity LHC (HL-LHC) will begin running and will aim to collect ~10 times as much data as all the standard LHC runs in 2010-2023.

IRIS-HEP is funded by the National Science Foundation and is distributed across several US universities. It incorporates six different focus areas: Innovative Algorithms; Data Organization, Management, and Access (DOMA); Analysis Systems; the Scalable Systems Laboratory (SSL); the Open Science Grid (OSG); and the Software Sustainability Core (SSC).

Innovative Algorithms
- Improve use of vectorization and parallelization techniques in HEP software
- Research algorithms that can be run on non-CPU processors such as GPUs and FPGAs
- Investigate use of machine learning techniques for trigger and reconstruction software

DOMA
- Assess needs in terms of atomic data sizes and shapes, compression, and caching models
- Prototype consolidated, high-capacity “data lakes” to adapt to demand
- Study possible use of other data distribution techniques like content delivery networks

Analysis Systems
- Develop analysis interfaces focusing on abstract declarative programming rather than imperative methods
- Investigate new query-based interfaces for retrieving data
- Improve ease of analysis reproducibility and data reinterpretation
- Provide IRIS-HEP and HL-LHC experiments with access to scalable platforms for development and testing
- Research into accelerated hardware that is or will potentially be available
- Facilitate integration with Open Science Grid infrastructure

SSL
- Operate and maintain infrastructure providing US computing services for the LHC experiments
- Develop software that can reduce network traffic and latency
- Monitor and support network performance and security

OSG
- Provide computing education and outreach for the HEP community
- Run workshops and training sessions at all skill levels
- Identify and promote software best practices

IRIS-HEP: A new software institute to prepare for the data from the High Luminosity Large Hadron Collider in the exabyte era

Mason Proffitt, Gordon Watts, Emma Torró (University of Washington, Seattle)

Introduction

- The Large Hadron Collider (LHC) is the most powerful particle collider and a vital tool in the search for new physics, as demonstrated by its discovery of the Higgs boson in 2012
- Since starting in 2008, the LHC has presented enormous computing challenges in the collection, transfer, storage, and processing of hundreds of petabytes (PB) of collision data
- In 2026, the High Luminosity LHC (HL-LHC) will begin running and will aim to collect ~10 times as much data as all the standard LHC runs in 2010-2023
- Expected increases in disk space and performance of the currently used storage and CPU technologies will not be sufficient to accommodate the full HL-LHC dataset (Figure 1)
- The Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP) was established in 2018 to close the gap between LHC computing needs and expected resources and to optimize the interfaces used by physicists for data analysis

Figure 1: Projected ATLAS computing demands (blue) versus available resources with a flat annual budget (red)

Figure 2: Schematic of prototype analysis interface

Analysis Systems Example
- At the University of Washington, we have begun work on a proof-of-concept project demonstrating a query-based analysis interface (Figure 2)
- GitHub repository:
  - https://github.com/gordonwatts/BDTTrainingAnalysisLanguage/

Table: Resources for ATLAS computing

<table>
<thead>
<tr>
<th>Year</th>
<th>PB</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>0</td>
</tr>
<tr>
<td>2018</td>
<td>0</td>
</tr>
<tr>
<td>2019</td>
<td>0</td>
</tr>
<tr>
<td>2020</td>
<td>0</td>
</tr>
<tr>
<td>2021</td>
<td>0</td>
</tr>
<tr>
<td>2022</td>
<td>0</td>
</tr>
<tr>
<td>2023</td>
<td>0</td>
</tr>
<tr>
<td>2024</td>
<td>0</td>
</tr>
<tr>
<td>2025</td>
<td>0</td>
</tr>
<tr>
<td>2026</td>
<td>0</td>
</tr>
<tr>
<td>2027</td>
<td>0</td>
</tr>
<tr>
<td>2028</td>
<td>0</td>
</tr>
</tbody>
</table>

Summary

- A new institution called IRIS-HEP has been established to coordinate HL-LHC computing services and resources and design the tools necessary for data analysis
- The University of Washington team has started work on a high-level, query-based analysis interface that is capable of running on several different data file formats
- Proof-of-concept code has been demonstrated and tested at a basic level
- Project improvement and development driven by real physics analysis needs

References


Acknowledgements

This material is based upon work supported by the National Science Foundation under Cooperative Agreement OAC-1836650. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.