# The Intelligent Data Delivery Service

The HEP field has long been centered around a file-based data access paradigm: event data is serialized and split across several files. Whether there is batch or interactive usage, a user determines the data they are interested in accessing and "compiles" it to a list of files to access. These files are given to an application such as ROOT, which subsequently interprets their contents. Computing organizations define their data based on the file's bytestream contents as opposed to the physics contents inside the files.

We propose a higher-level interface to HEP data: the Intelligent Data Delivery Service (IDDS). By accessing the data at a higher level than byte offset ranges, we have more opportunities to optimize both data access patterns and storage:

Data Factory/source
(e.g. T0 or sim)

Data Store/Lake

Intelligent Data
Delivery Service (iDDS)

Data Cache

Compute Nodes/
Data Sinks

Rucio/
FTS

Tape-backed
Storage

- Data filtering and simple transforms can occur server-side.
- Improvements in the I/O layer can done without changes to the application.
- Storage services can continuously optimize the data layout to reduce overall data size.
- IDDS can manage data locality independently of moving full files.
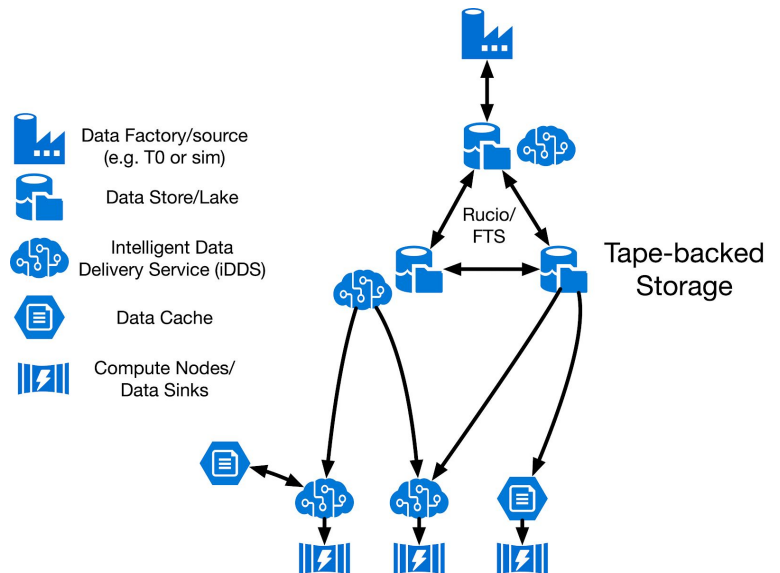
The IDDS will be tasked to deliver event- and object-level data to computing entities such as analysis processes or production jobs; it can feed individual events to jobs as in the ATLAS Event Streaming Service, supporting event-service-like computing paradigms.

By abstracting the backend storage, IRIS-HEP can work to optimize the storage of event data - using approaches such as SkyhookDM or by re-compressing data to reduce data size.

The IRIS-HEP DOMA area is looking to collect requirements and use cases from stakeholders, with a goal of building prototypes and demonstrating an IDDS viable to both analysis and production users.

Refs:

Gardner, R., Reidel, B., and Vukotic, I. Organizing, Orchestrating, and Delivering Data From Lakes. http://bit.ly/atlas-lakes

Nicolo Magini, on behalf of the ATLAS Collaboration. Towards an Event Streaming Service for ATLAS data processing. https://cds.cern.ch/record/2649518