DOMA R&D Area

Brian Bockelman

IRIS-HEP Kickoff, 31 October 2018

As a reminder…

- DOMA ~> **D**ata **O**rganization, **M**anagement, and **A**ccess.
- *Primarily*…
    - focused on *event data* for HEP experiments (as opposed to software or calibration), although *metadata* is also a key component.
    - **Organization** refers to how we serialize data in storage (e.g., ROOT file format).
    - **Management** is the policy and mechanisms for how we distribute data and manage storage.
    - And finally, **access** is how data is delivered to site storage, processing, or end users.

The DOMA Team

- Area lead: Brian Bockelman
- Participating institutions:
    - Nebraska: Brian Bockelman, Derek Weitzel (starting late Y1), Oksana Shadura (Y2).
    - Princeton: Jim Pivarski.
    - UChicago: Rob Gardner, Andrew Chien, Benedikt Reidel, Ilija Vukotic.
    - UCSC: Carlos Maltzahn, Jeff LeFevre, Reza NasiriGerdeh.
    - UCSD: Frank Wuerthwein, Edgar Fajardo.
    - UIUC: Mark Neubauer and Ben Galewski
- Note: I've listed some of the people who have been participating in the DOMA discussions, a wider set than those paid by IRIS.

Why DOMA?

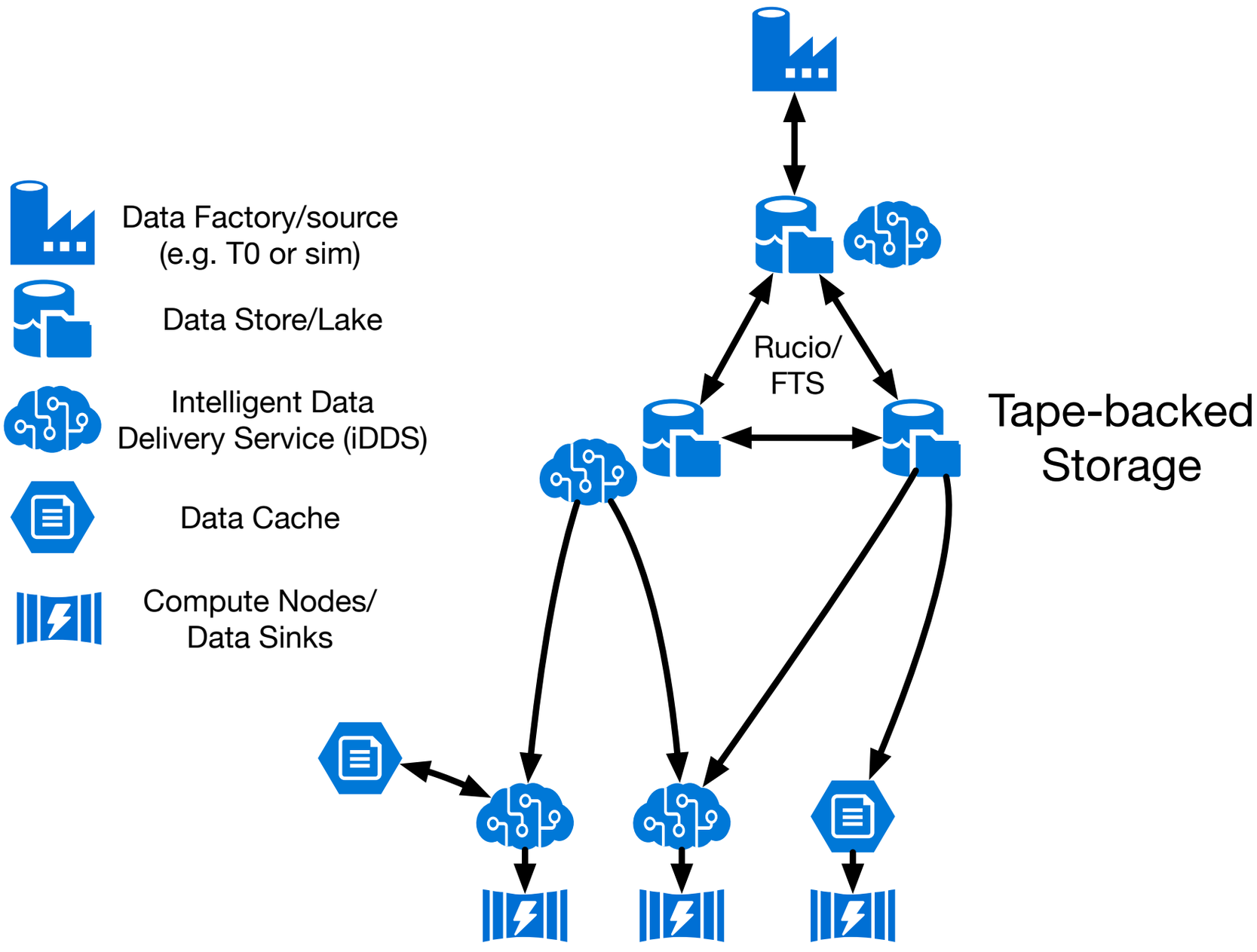
- What's the point of a DOMA R&D area?
    - We believe data management is one of the key cost drivers for the HL-LHC and our group will make key contributions to reducing this cost. **Advances in technology do not compensate** for the increase in data volumes and complexity.
        - To reduce cost, we must **reduce the data resident on disk**: reduce replicas, reduce size of the dataset, reduce the percentage of data kept on disk (vs tape).
        - We must also **reduce the barriers for user data access**. Human costs, although more difficult to quantify, loom large for the HL-LHC.
- Some of these issues are social and some are technical: we cannot tell the HL-LHC what to do, but rather work together to provide potential solution mechanisms.

Activities

- The DOMA group will be moving from the initial Google Groups to doma@iris-hep.org shortly.
- 2 weeks ago we had an informal meet-up in Chicago to fill in a strawman technical plan.
  - *The intent of the strawman is to beat it up* - please don't take the following slides as anything set in stone.
  - We have a draft WBS for those who care to look at it.
- The next few slides walk through the top-level activity areas and the motivations.
  - I've purposely tried to highlight a few areas to generate discussions.
  - WBS activities try to follow pattern of (a) establishing metrics, (b) developing simulation / prototype and (c) documenting results.

Organizing “Road Map”

- Carefully avoided the words “architecture diagram”



Notes:

- We treat computation and data storage as more separate in this model
- There will always be multiple independent data stores that need to be managed.
- Some data stores may look like today’s storage elements (file-based) while others may have an internally optimized format.

3.1. Modeling

- We believe that modeling and simulation is one key area we've underinvested in for the past 20 years.

- Both ATLAS and CMS have remarkable datasets outlining the data demands and the computation needs.

- Would like to start a simulation activity of how these data demands - and corresponding scheduling policies - can be affected by different hardware and storage investment strategies.

- Goal is to build a functional model that can demonstrate the characteristics of the current system and help us predict the impact of different strategies.

3.2 Data Organization

- For data organization, we have two drivers / activities:
    - Investigate alternate data organization methods for reducing disk size.
    - Investigate alternate organization methods to accelerate delivery of events to users. “Accelerate” may cover speed (MB/s) or simpler semantics presented to the analysis user.
- The two are not necessarily conflicting: e.g., object store methods may tackle both goals.
    - First WBS item is to deliver a planning document outlining potential gains for each approach.

3.3. Data Access

- The main goal here is to augment the traditional semantics of files.
    - Allow users to interact with storage systems in terms datasets / events /objects, not filenames and byte offsets.
        - This has potential to mesh better with object-store-based layouts than files anyway!
- Goal is to provide interfaces more natural to “physics users” -> will require significant coordination with Blueprint and AS groups.

3.3. Intelligent Data Delivery Service (iDDS)

- We think of a **cache** as a service that can deliver specific bytes to the user. Responsible for finding the right data (either locally or remotely) and sending it.

- We see a **data delivery service** as something that can deliver the requested data - possibly transforming or pre-conditioning things as needed.

- A cache responds to “deliver bytes 1024 - 2048 of file /foo/bar/baz.root”.

    - A data delivery service responds to “deliver the electrons and muons from events 1 - 10,000 of dataset foo_baz_AOD”.

- The iDDS will be responsible for handling user queries, determining the appropriate data source, transforming the data (potentially working with another iDDS at the source side), deciding whether to store it locally, and finally delivering to the user.

3.4. Data Management

- Some closer-in deliverables:
  - Improve and integrate inter-site transfer protocol.
  - Ensure US LHC facilities are prepared for non-X509-based authorization models.
- More R&D topics:
  - Work with the WLCG QoS sub-group to help storage system and data management develop a common protocol on storage costs.
  - Express policies in “better semantics”, i.e., time-dimension or costs.
  - Develop and deliver report on cache usage within the WLCG.
- Joint with Blueprint, participate / organize workshop on open data management issues in the community.

DOMA and other areas

- DOMA has some clear ties with other parts of the institute:
    - **SSL**: We'll need to use this infrastructure for data transfer and for researching data accelerators.
    - **OSG-LHC**: OSG will be essential for maintaining parts
    - **AS**: We envision the iDDS to be a key component in delivering data to future analysis systems. Hoping to rely on the close ties, especially at UIUC.
    - **IA**: Less clear than the others — potentially related to delivering data to resources (e.g., HPC) with alternate architectures