DUNE Host Lab Responsibilities

Stu Fuess / Scientific Computing Division

2019 ICAC

15 October 2019

DISCLAIMER

- What is presented here is a summary of internal Fermilab discussions
- Starting point for discussions with the broader collaboration
- This is still a work in progress – comments welcome

Outline

- Methodology and history
  - Snapshot of the details
- Summary of projects w/ responsibilities
- Issues?

Methodology and History

- The Fermilab half of the DUNE Computing Technical Coordinators (Kirby) organized a series of meetings to which all Fermilab service providers were invited
- All worked on a shared spreadsheet that listed existing services and projects
  - Some areas expanded to great detail
  - Example snip of the spreadsheet:

Service/Project	Project Lead	Development	Infrastructure	Operations	Support	Notes	FNAL Person(s)	FNAL Division	Key for entries
Data Storage and Movement									
Data storage devices:									
Tape	1	1	1	1	1		Robert Illingworth, Bo Jayatilaka		
Disk	1	1	1	1	1		Robert Illingworth, Bo Jayatilaka		
Data Management:									
Rucio Service Operations	1	1	1	1	1	SCD runs FNAL-based sv	Robert Illingworth		
Rucio Fermi-specific features	1	1	1	1	1		Robert Illingworth (dev), Steven Timm (deploy + Int)		
Rucio DUNE-specific features	1	2	1	1	1		External dune-funded developers + Robert (dev), Steven Timm (deploy + int)		
New Metadata catalog	1	2	1	1	1		Robert Illingworth (dev), Steven Timm (deploy + Int)		
SAM->Rucio File Catalog	1	1	n/a	2	2		Robert Illingworth (dev), Steven Timm (deploy + Int)		
Data retention policies	1	n/a	n/a	4	2		S. Timm deploy and int, hand off to DUNE experimenters		
Rucio schema and procedures	1	n/a	n/a	4	2		S. Timm deploy and int, hand off to DUNE experimenters		
Moving data from point A to B	1	n/a	n/a	4	2		DUNE experimenters		
Data Movement:									
Fermilab FTS replacement	1	1	1	2	2	Talks to Rucio and new m	Robert Illingworth (dev), Steven Timm (deploy and int) hand off to DUNE experimenters for ops		
Fermilab - FTS (legacy)	1	1	1	3	1	Need to have something a	Robert Illingworth		
CERN - FTS3	4	3	2	2	2	Assuming this means runi	N/A		
Data Protocols:									
DOMA	2	2	2	4	3	DOMA is lots of separate	Bo Jayatilaka		
DOMA-AAI						AAI structure is key techn	Bo Jayatilaka		
DOMA-3rd party copy	3	3	2	2	2	3rd party copy key techno	Bo Jayatilaka		
DOMA-caching	2	2	2	2	2	"Data lakes" are key cach	Bo Jayatilaka		
Equivalent of SAM project mgmt	1	2	2	2	2	Need to develop the equivalent machinery for running over file sets. Strong contact with workflow mgmt. The experiment needs to decide on its apprc			
						Need to note that all existing DUNE "operations" people in data management are currently FNAL employees--need to identify and train minions.			

Methodology and History

- For each task (continuing or new) listed, looked at the responsibilities for:
  - Project lead
  - Development
  - Infrastructure
  - Operations
  - Support

- For each task we marked the Host Lab responsibilities as:
  1. = Mandated to do
  2. = Want to do, but OK for others to lead
  3. = Don’t care to do, but would participate as able
  4. = Do NOT want to participate; must be done by others

Summary of Projects

- Main areas (in no particular order):
  - User Management
  - DAQ and Online Computing
  - Data Storage and Management
  - Distributed Computing
  - Databases
  - Offline Software and Tools
  - Software Development Tools
  - Collaborative Tools
  - User Facilities

Host responsibilities: User Management

- Host lab to lead, develop and operate:
  - Computer security
  - Federated identity mechanism
  - Authentication and Authorization infrastructure
  - Account and VO administration (*FERRY*)

- Expect collaboration to provide:
  - Computer security coordinator: contact person within the collaboration to chase down issues and contact offenders

Host responsibilities: DAQ and Online

- Host lab to lead, develop and support:
  - artDAQ (DAQ framework)
  - Network provider interface (to ESNET) for SURF
  - Remote control room (e.g. ROC West facility)

- A number of areas still TBD on responsibilities, but aiming for collaboration to take lead roles with Fermilab consultation:
  - DAQ and Online system architecture
  - DAQ and Online system administration
  - DAQ and Online software management
  - Provisioning of DAQ (and other hardware) test stands

Host responsibilities: Data Storage and Management

- Host lab to lead, develop and support:
  - "Tier-0" physical infrastructure and operations: tape, disk
  - Data management (e.g. *Rucio*) tools
    - Responsible for Fermilab-specific features
    - Responsible for core service operations
- Expect collaboration to contribute to development and operations:
  - Metadata catalog
  - DUNE-specific *Rucio* features
  - DOMA functions (AAI, 3rd party copies, caching)
  - *Fermi-FTS* to *FTS3* transition
- Expect collaboration to manage day-to-day operations:
  - Data Movement

Host responsibilities: Distributed Computing

- Host lab to lead, develop and operate:
  - Local resources (Institutional Cluster – e.g. FermiGrid plus other local resources)
  - Portal to external resources (*HEPCloud*)
  - Job submission mechanism (*jobsub*)
  - Accounting & Monitoring
    - Infrastructure
    - Services (including *fifemon*, *elasticsearch* / *grafana* / *kibana*, LHC tools)
- Expect collaboration to contribute to development:
  - Production operations and support
  - Workflow management
- Expect collaboration to provide user and day-to-day support:
  - All the above

Host responsibilities: Databases

- Host lab to lead, develop and operate:
  - Beam information DB
  - Metadata backend DB
  - Collaboration membership DB
  - Infrastructure for some of the following…
- Expect collaboration to lead, develop, and operate:
  - Hardware DB
  - Slow controls DB
  - Configuration / conditions DB

Host responsibilities: Offline Software and Tools

- Host lab to lead, develop and support:
  - Container management
    - Infrastructure
    - Build mechanisms
  - POMS (workflow and production system)
  - Project management (e.g. *sam* or successor)
  - Job data handling utilities (e.g. *ifdh*)

- Want strong participation, but do not necessarily need to lead:
  - Production and/or analysis framework (e.g. *art* or successor)
  - LArSoft (reconstruction)
  - Software package management (e.g. *spack*)

Host responsibilities: Software Development Tools

- Host lab to lead, develop and operate:
  - *cvmfs* repositories
  - Software build systems
  - Infrastructure for CI (e.g. *Jenkins*)
  - Infrastructure for service development and integration
- Expect collaboration to lead, develop, and operate:
  - Continuous integration
- Software repository solution is likely to be github

Host responsibilities: Collaborative Tools

- Host lab to lead, develop (if needed) and operate:
  - Web site, wiki
  - Document databases (*DocDB*, *EDMS*)
  - Video conference tools
  - Project management tools
  - Meeting management tools (*Indico*)
  - Mail lists
  - Communication tools (e.g. *slack*)
- Expect collaboration to lead, develop, and operate:
  - Electronic logbook(s)

Host responsibilities: User Facilities

- Host lab to provide:
  - Office facilities for on-site visitors
  - User support to navigate bureaucracy (embedded funded person?)

- Host lab to lead, develop (if needed) and operate:
  - Interactive computing
  - LPC-like analysis computing facility

- Expect collaboration to lead, develop, and operate:
  - Education (e.g. Data Analysis School)

Conclusion: Issues?

- What is missing that should be a Host Lab responsibility?
- Should more things be left to the collaboration / consortium?
  - Are leadership assignments useful in procuring funding?
- Does the lab have the effort / funding to provide the items listed?
- What opportunities for global collaboration (e.g. WLCG, HSF, …) should be pursued in addition to those mentioned (e.g. DOMA, *Rucio*)?
  - Would any of these reduce the Host Lab task list?