DUNE Workflow Management and Distributed Data Management mini-Workshop

Introduction

- *Workflow Management*: System that manages all aspects of running production and analysis jobs.
- *Distributed Data Management*: System that stores, catalogues, and moves data.
- LHC experiments rely on GRID resources for blue of their computing.
    - A significant fraction of the resources they use are opportunistic.
    - Access to HPCs.
    - protoDUNE and DUNE are global experiments. We should be able to utilize any resources available to us.
- IF experiments at Fermilab predominately use Fermilab resources.
- DUNE needs to come up with strategy to process protoDUNE data.
    - DUNE Computing Model.

Goal

- Understand resources and services (FTEs, software, hardware) available to DUNE, specifically at Fermilab (e.g. via FIFE)
- Survey existing systems and experiences.
    - ATLAS (PANDA), CMS, MicroBooNE, Nova
    - FIFE: plans…
- Establish the DUNE requirements.
- DUNE will have to choose a system/strategy.
    - Ideally, FIFE should evolve to meet DUNE's requirements.

Workshop Format

- Thursday Morning: Talks from other experiments/FIFE.
- Thursday Afternoon/Friday Morning:
    - First pass DUNE requirements
    - Discussion of each element/component of required systems
        - Task, Job, Event Book-keeping and submission. Job splitter. WorkFlow Mangement. Resource Provisioning. Pilot. VO Management. Resource/Job/Site Monitoring. Accounting Data Management. Catalogues. …
        - Will ask FIFE experts to provide a few slides on such topics.
- Friday Afternoon: DUNE requirements

Nomenclature

- From Tom's slides…
- *Campaign:* A large computing project
    - Examples: MCC 6.0. Reprocessing one pass of ProtoDUNE SP data.
- *Task:* A piece of a campaign that maps to a cluster of jobs.
    - Examples: SNB neutrino sample in the FD sample of MCC 6.0.
- *Stage:* Many tasks require several steps to process each event
    - Examples: MC: gen+G4+detsim+reco+ana. Data: split+reco+ana
- *Cluster:* Condor's term – a set of jobs
- *job:* single unit of batch execution – runs one one worker node for a specified amount of time. May run more than one executable or use more than one core.

Requirements

- DUNE Inherited requirements from LBNE-doc-8035-v6 (Brett and Maxim), in DUNE-doc-914-v2, "Computing Model for the DUNE Experiment".

- We'll start with these this afternoon.

- Grid and Cloud Capability Requirements

    - A **widely distributed computing infrastructure**, featuring a network of federated resources (including Grid- and Cloud-based resources) shall be implemented.

    - This widely distributed computing infrastructure shall be put in place by the **DUNE S&C Organization** in close cooperation with participating computing sites, institutions, and agencies (cf. the Open Science Grid etc).

    - The distributed resources shall include **processing facilities, storage, network hubs** and their combinations (e.g. Grid sites with large available storage capacity).

    - Necessary **tools and procedures** shall be provided, for **streamlined incorporation of new facilities** as they become available, efficient use of opportunistic resources

    - The effort and expertise across all of the Collaboration shall be leveraged in order to provide adequate operational support with a **minimum of manpower**.

    - Details of site capabilities, interfaces, configuration and other crucial information will be gathered and preserved as **documents and database records**, and made available through appropriate information systems (often termed Grid Information System).

WMS Requirements

- DUNE shall implement a **Workload Management System (WMS)** for resource management and brokerage functionality which will govern distribution of most types of computation workload in DUNE (e.g. production jobs, group analysis, etc) across a variety of resources available to the Collaboration.

- The DUNE WMS shall be capable of keeping **precise record of the software configuration** used for each and every job deployed on the Grid, including, among other things, the DUNE Offline Software Release information.

- The DUNE WMS shall be capable of **quickly suspending participating sites** due to outages, network congestion or potential security issues.

- The DUNE WMS shall be augmented with a **Workflow Management layer**, which will help **create and manage large groups of Grid task**s supporting scientific workflows.

- A DUNE **WMS Monitoring System** shall be implemented to allow efficient operation of the WMS, by helping ascertain correct execution of Grid jobs, accounting of resource utilization, identification and debugging of failure modes etc.

- The DUNE WMS Monitoring System shall keep status records for individual jobs and their groups, information related to data I/O and transmission, and include the **"big picture" performance monitoring data** for entire Grid sites utilized by DUNE.

- The DUNE WMS Monitoring System shall have interfaces conducive to integration with both **Web UI for users and operators**, **and** with **automated systems** needing the WMS data.

	ATLAS	CMS	FIFE	OSG VOs
Experiment Software	Athena (AthenaMP, AthenaMT)	CMSSW	Various (Art + LarSoft, Gaudi)	Custom/Various
Task Book Keeping	DEFT	WMStats (P) CRAB Client (A)	POMS (?) Custom	
Job Book Keeping	Panda Server (JEDI)	WMStats (P) CRAB Client (A)	Jobsub Client	HTCondor Pegasus Custom
Event Book Keeping	Panda Server (JDEI/Event Service)	WMStats CRAB Client		
Task Submission	Panda Client DEFT	ReqMgr API CRAB Client	POMS (?) Jobsub Client Custom	
Workflow Manager	Panda Server (JEDI)	WMAgent CRAB Server	(????) Custom Custom based on SAM	HTCondor Pegasus Custom
Jobs Splitter (Task ⇨ n*Jobs)	Panda Server (JEDI)	WMAgent CRAB Server	Jobsub Server Custom	HTCondor Pegasus Custom
Job Submission (Results in jobs in queue)	Implicit in Panda Server	WMAgent CRAB Server	Jobsub Server	HTCondor Custom
Resource Provisioning	Panda (Server+ APF, Crontabs, HPC - Direct ssh - BOSCO	GlideinWMS (GlideinWMS Factory + Frontend)	GlideinWMS (GlideinWMS Factory + Frontend)	GlideinWMS (GlideinWMS Factory + Frontend)

	like)			
VO Resource Manager (VO Pool/Matchmaking)	Panda Server	HTCondor	HTCondor	HTCondor
Resource Monitoring	Panda Monitoring	WLCG Dashboard, CMS Global Info from HTCondor	FIFEMon (Landscape)	
Site Availability Monitoring	SAM/ETF HammerCloud WLCG Site readiness	SAM/ETF HammerCloud WLCG Site readiness	FIFE team manual	RSV
Job Monitoring	Panda Monitoring	WLCG Dashboard WMStats (P) CRAB Client (A)	FIFEMon (Landscape) Jobsub Client	HTCondor
Information Systems	AGIS	SiteDB, GlideinWMS Factory Config, WLCG Dashboard, SiteConfig in CVMFS	GlideinWMS Factory Config	OIM, GlideinWMS Factory Config
Accounting	EGI Accounting Portal	EGI Accounting Portal	Gratia	Gratia
Data Cataloging/Management	Rucio	PhEDEx + DBS	SAM	Custom
Data Movement	Rucio + WLCG-FTS3	PhEDEx + WLCG-FTS3	Ifdh + SAM	Custom