Data Management: High Level Architecture

# Introduction

# User Roles

User roles can be either system wide, or experiment specific. Examples of system wide roles are:

* Administrator
  + Manages experiments
  + Manages experiment user roles
  + Manages experiment data
* User

Examples of experiment-specific roles are:

* Beamline Staff (BLS):
  + Can add/delete experiment users
  + Can modify experiment file permissions
  + Can modify experiment policies
  + Can delete experiment files
* Principal Investigator (PI):
  + Can add/delete experiment users
  + Can modify experiment file permissions
* Experiment User:
  + Can read experiment data

# Experiment Policies

The DM System will support various policy types, such as:

* Disk quota
* Auto-delete policy
* Auto-transfer policy
* Backup policy

Some of those may have set of allowed values (e.g., “Auto-delete” policy type may mean that “experiment data will be removed from the system in N days”, where N may be 30, 60, or 90).

Template policy sets may be used for initial experiment setup.

# System Components

* Data Acquisition Service (DAQS):
  + Runs on each detector node
  + Provides REST interfaces that allow DM administrators to setup, start and stop an experiment
  + Monitors local file system for new experiment data files and transfers those to medium term storage using, for example, GridFTP client
  + Manages local storage (e.g., removes data that belong to previous experiments)
* Experiment Setup UI:
  + Installed on every detector node
  + Used by DM administrators to setup, start and stop an experiment
  + Retrieves data from scheduling system
  + DM LDAP provides user authentication
* Data Storage Service (DSS):
  + Controls data storage file system: creates/removes experiment directories, updates file/directory permissions according to information stored in the database
  + Permissions are updated for new files, and after DB modifications via DM web portal; they are also checked periodically for all files to match information stored in the DB
  + Schedules directory removal/archiving for old (closed) experiments
  + Provides REST interfaces for experiment setup and permission updates, scheduling deletion for unused directories, etc.
  + DM LDAP provides user authentication, DM DB provides user authorization
* Data Management Database:
  + Contains information about experiment directory permissions, experiment users/owners
  + Eventually (later stages of the DM project), the DB may be extended to host replica/metadata catalogs
* Data Management Web Portal:
  + Presents various views of the DM system
  + Allows experiment owners to manage their experiments
  + Allows DM administrators to manage all experiments
  + DM LDAP provides user authentication, DM DB provides user authorization
  + Eventually (later stages of the DM project), the portal may be extended to view and edit file metadata and replica catalogs
* LDAP Management Service:
  + Regularly synchronizes user accounts in LDAP and DM DB with information present in external user DB
  + Presents DM administrative REST interfaces for user account-related tasks (e.g., force account synchronization)

# Use Cases

## Experiment Setup

1. DM Administrator provides username/password to the Experiment Setup UI
2. UI retrieves data from the scheduling DB in order to present possible experiments
3. DM Administrator selects experiment and choses local storage area to be monitored for experiment data files
4. UI makes REST web service call to the Data Acquisition Service to setup chosen experiment
5. Data Acquisition Service calls Data Storage Service experiment setup REST interface, which updates the database, creates necessary directories and establishes initial file system permissions. It also logs the event and sends notifications to subscribed clients.
6. Data Acquisition Service stops file system monitoring thread for the previous experiment (if any), and starts new monitoring thread for the provided local storage directory
7. Data Acquisition Service schedules removal of old directory/data for the previously running experiment according to configured policy (e.g., “leave data on the local storage 7 days after experiment completes”).

## Detector DAQ

1. AD software closes newly written file.
2. Data Acquisition Service notices that new file has been written, and invokes file transfer (e.g., using GridFTP). Upon transfer completion, DAQS notifies DSS.
3. Data Storage Service receives notification (from the file system, or from DAQS) that a new experiment file was written, and it queues permissions change request.
4. Queued permission change request will subsequently trigger file permission modification according to settings in DM DB. This event will also be logged and subscribed clients will be notified.

## Directory Permissions Update in DM DB

1. Experiment owner logs into DM Web Portal and requests permission changes for one of the experiment data directories.
2. Web Portal modifies DM DB, and notifies Data Storage Service.
3. Data Storage Service queues request.
4. Queued permission change request will subsequently trigger permission modification for all directory files (including subdirectories) according to settings in DM DB. This event will also be logged and subscribed clients will be notified.