Table of Contents

DM-14: A researcher executes an application that runs for a long period of time, scrapes data from public websites, and stores large amounts of data for analysis

A researcher executes an application that runs for a long period of time, scrapes data from public websites, and stores large amounts of data for analysis. We assume the researcher has an allocation to use the resource.

In most cases, the researcher wants to experience it as follows.

1. First, the researcher constructs a job consisting of tasks that (a) identify publicly-accessible websites to scan, (b) crawls each website to gather data, and (c) stores the data in a predefined location on the system.
2. Then, the researcher submits the job for execution on a compute resource.
3. When the job executes, subtasks running on compute nodes open connections to the public internet as needed and write the data to the predefined storage location.
4. When the job completes, the researcher either submits additional jobs to analyze the data in place (as described in use case DA-04), or transfers the data to another system for analysis (as described in use case CAN-02).

We'll accept any solution as long as the following are true.

1. In Step 2, if the researcher has more than one allocation on the resource, the researcher must be able to specify which allocation to use for this application run.
2. In Step 3, many subtasks may run in parallel.

DM-15: A researcher transfers data between a community storage system and the researcher's commercial cloud storage

A researcher transfers data between a community storage system and the researcher's commercial cloud storage, e.g., Google Drive, Box, Microsoft OneDrive, etc. We assume the researcher has an allocation to use the community storage system.

In most cases, the researcher wants to experience it as follows.

1. First, the researcher accesses the community's data transfer interface. (See use case CAN-02.)
2. Then, the researcher selects the community storage system as one end of the data transfer.
3. Then, the researcher selects the researcher's cloud storage as the other end of the data transfer.

4. Then, the researcher specifies the specific files/folders to be moved and the specific folder to which they should be copied.
5. Then, the researcher initiates the transfer.
6. When the transfer is complete, the researcher receives a notification.

We'll accept any solution as long as the following are true.

1. In Step 1, the researcher can login to the community's data transfer interface using the researcher's community identity.
2. In Step 2, the researcher can search for and identify the community storage system.
3. In Step 3, the researcher can enable the community storage system to use the researcher's cloud storage. (This likely involves specifying a specific cloud service to use and supplying credentials to access the cloud service.)
4. After Step 5, the data transfer proceeds without further involvement by the researcher or the device the researcher was using to initiate the transfer.
5. In Step 6, there is a way for the researcher to inspect the transfer results, including performance.
6. If there is a problem with the data transfer after Step 5 that cannot be automatically resolved, the researcher receives a notification with an explanation of the problem that enables the researcher to know (a) the status of the transfer, and (b) how to proceed.