

"Abstract Job" = dataset + algorithms (what the user sets up).

"Concrete Job" = single invocation of abstract job at user level.

"Subjob" = part of a concrete job sent to a specific executor.

"Dataset" = a collection of data that can be processed event by event. Has to provide a means to gather the data for each event. Should have a unique identifier, ideally with option to produce a human readable identifier.

The term API here refers to a language-neutral API that may in fact be implemented in a variety of different languages. Some of these APIs may be implemented as services. The term API may be used to describe a collection of sub-APIs, therefore some of the APIs listed below may be subcomponents of other listed APIs

Abstract Job Submission API

Pass it an abstract job, it creates a concrete job.

Concrete Job Interaction API

May contain Concrete Job Control API and Concrete Job Status API

Concrete Job Control API

After you have created the concrete job, gives you ability to control the concrete job, for example, tells concrete job to pause so that I can have a look at some partial results, tells concrete job to start over again from the beginning.

Concrete Job Status API

Allows you to get full or partial results, or other status information, may involve collating results.

Concrete Job Capabilities API

Allows client to find out capabilities of a concrete job, what interfaces a concrete job supports, such as "does this concrete job allow me to get partial results? Can I upload a root script into this concrete job". Requires there to be an agreed upon list of what interfaces might be supported.

Subjob Management API

Manages the subjob lifecycle. May be very specific to the application.

Estimator API

Could be implemented by various services to tell "how expensive is the execution of the task that I want you to execute". Units of expense depends on what service is talking.

Data Mover API

May be called by various components to move data. Some user analysis tools might not use this directly. Would like asynchronous notification of completion. Might want to move some data to my laptop.

Dataset Catalog Service Query API

Let me do ls, cd, etc. to eventually get to set of datasets or logical files. Let me find metadata about a dataset. I'd like to be able to make queries based on metadata ("show me all the datasets that satisfy the following criteria"). Include quality of service data such as where the data happens to be, disk or tape, etc. Dataset Catalog API may as likely get data from a relational or object database as from an actual physical catalog. Likewise, the fact that the API may be able to present a hierarchical view of the data does not necessarily imply that the data is actually stored in such a hierarchy. May in turn use Replica Location API and perhaps the Metadata Catalog API.

Dataset Catalog Service Management API

Let me add information to the dataset catalog, set access controls, allow locks to protect datasets from deletion

Replica Location API

Given a logical file name, gives a list of physical replicas. May no longer need to be directly accessed by user application if Dataset Catalog API hides this.

Metadata Catalog API

Allows you to get metadata about datasets, logical files and physical files. Or allows you to find datasets, logical files and physical files by a metadata query.

Matchmaker API

Matches requested abstract job requirements with computing resources. Location of data is one of the important factors. Does the matchmaker consider possibilities of moving data to get a better match? In EDG, you don't talk directly to the matchmaker. The resource broker talks to the matchmaker for me.

Resource Discovery API

Where can I get resources (CPU and network) that I'm authorized to use and that I can currently reserve

Resource Reservation API

Reserve these resources for me. May require a bunch of resources to be co-scheduled.

Grid Storage API

Interface to grid storage, probably SRM

Sign On API

Authentication, probably GSI

Software Installer API

Installs and configures software on compute nodes.