
Glossary

A

Abstract Workflow See DAX

C

Concrete Workflow See Executable Workflow

Condor-G A task broker that manages jobs to run at various distributed sites, using Globus GRAM to launch jobs on the remote sites.<http://cs.wisc.edu/condor>

Clustering The process of clustering short running jobs together into a larger job. This is done to minimize the scheduling overhead for the jobs. The scheduling overhead is only incurred for the clustered job. For example if scheduling overhead is x seconds and 10 jobs are clustered into a larger job, the scheduling overhead for 10 jobs will be x instead of 10x.

D

DAGMan The workflow execution engine used by Pegasus.

Directed Acyclic Graph (DAG) A graph in which all the arcs (connections) are unidirectional, and which has no loops (cycles).

DAX The workflow input in XML format given to Pegasus in which transformations and files are represented as logical names. It is an execution-independent specification of computations

Deferred Planning Planning mode to set up Pegasus. In this mode, instead of mapping the job at submit time, the decision of mapping a job to a site is deferred till a later point, when the job is about to be run or near to run.

E

Executable Workflow A workflow automatically generated by Pegasus in which files are represented by physical filenames, and in which sites or hosts have been selected for running each task.

F

Full Ahead Planning Planning mode to set up Pegasus. In this mode, all the jobs are mapped before submitting the workflow for execution to the grid.

G

Globus The Globus Alliance is a community of organizations and individuals developing fundamental technologies behind the "Grid," which lets people share computing power, databases, instruments, and other on-line tools securely across corporate, institutional, and geographic boundaries without sacrificing local autonomy.

See Globus Toolkit

Globus Toolkit	Globus Toolkit is an open source software toolkit used for building Grid systems and applications.
GRAM	A Globus service that enable users to locate, submit, monitor and cancel remote jobs on Grid-based compute resources. It provides a single protocol for communicating with different batch/cluster job schedulers.
Grid	A collection of many compute resources , each under different administrative domains connected via a network (usually the Internet).
GridFTP	A high-performance, secure, reliable data transfer protocol optimized for high-bandwidth wide-area networks. It is based upon the Internet FTP protocol, and uses basic Grid security on both control (command) and data channels.
Grid Service	A service which uses standardized web service mechanisms to model and access stateful resources, perform lifecycle management and query resource state. The Globus Toolkit includes core grid services for execution management, data management and information management.

L

Logical File Name	The unique logical identifier for a data file. Each LFN is associated with a set of PFN's that are the physical instantiations of the file.
-------------------	---------------------------------------------------------------------------------------------------------------------------------------------

M

Metadata	Any attributes of a dataset that are explicitly represented in the workflow system. These may include provenance information (e.g., which component was used to generate the dataset), execution information (e.g., time of creation of the dataset), and properties of the dataset (e.g., density of a node type).
Monitoring and Discovery Service	A Globus service that implements a site catalog.

P

Physical File Name	The physical file name of the LFN.
Partitioner	A tool in Pegasus that slices up the DAX into smaller DAX's for deferred planning.
Pegasus	A system that maps a workflow instance into an executable workflow to run on the grid.

R

Replica Catalog	A catalog that maps logical file names on to physical file names.
Replica Location Service	A Globus service that implements a replica catalog

S

Site	A set of compute resources under a single administrative domain.
------	------------------------------------------------------------------

Site Catalog A catalog indexed by logical site identifiers that maintains information about the various grid sites. The site catalog can be populated from a static database or maybe populated dynamically by monitoring tools.

T

Transformation Any executable or code that is run as a task in the workflow.

Transformation Catalog A catalog that maps transformation names onto the physical pathnames of the transformation at a given grid site or local test machine.

W

Workflow Instance A workflow created in Wings and given to Pegasus in which workflow components and files are represented as logical names. It is an execution-independent specification of computations