

Table of Contents

Workflow Dictionary.....	1
--------------------------	---

Workflow Dictionary

The purpose of this list is to define some commonly used words, in order to understand emails, e-logs and some other stuff. Is intended to use as a reference, specially for new team members:

- ACDC: A partial retry of a workflow, it retries only failed jobs. Nobody knows what the acronym stands for.
- Black-Hole-Node (or just Black-Node): When something in a node is misconfigured, and makes every job fail in that node. Since jobs fail, the node keep getting jobs assigned so it will quickly sucks a lot of jobs that will become failed. Black-Nodes are usually dealt with by marking the node as "draining"
- CMSWeb:
- CMSSW: The CMS experiment's SoftWare package.
- Component: WMAgent is made of components that interact to each other through the data.
- Condor: An open-source software framework used to distribute and run processes around a Grid. Does all the dirty work. WMAgent creates condor jobs, then they are send to GlideIn front-end, then to the sites when they really run.
- Cool-off: When a job has failed but is going to be resubmitted.
- Couch: Abbreviation for Couch DB, a JSON, web oriented database made by Apache. We use it for managing request and workflow status. There is a Central Couch and each WMAgent has also a local couch. We intend to change it to BigCouch.
- Dashboard: <http://dashb-cms-job.cern.ch/dashboard/templates/web-job2/> the main monitoring tool for CMS. It shows statistics about running jobs, failed, successful, etc.
- DAS: <https://cmsweb.cern.ch/das> Data Aggregation System, is a web application that provides a single point for requesting information about datasets (location, size, files, blocks, etc.), by querying other systems such as PhEDEx, DBS3 and McM.
- DBS3: An index of the datasets and metadata about blocks, files and lumis.
- Dataset: A logical unit of information around the experiment. It is composed of blocks, which are composed of files. A single workflow may have an input dataset, may use a pile-up dataset, and always produce one or more output datasets.
- Drain: When some error is detected on one site, or is scheduled to be down due to some maintainance, it is set to "drain": to finish its assigned jobs while not getting assigned any new ones.
- Dynamic Data Management (DDM): System of deciding where to store data.
- E-groups: <https://e-groups.cern.ch> , where access and permissions groups are handled (at least part of).
- E-Log: the main communication tool for CMS computing, you should write everything here.
- GGus: <https://ggus.eu> Some ticket managing system, this replaces the previously used Savannah.
- GlideIn Factory: A system that creates GlideIn Pilots inside the sites. Those pilots will eventually will pull condor jobs from the GlideIn Front End
- GlideIn Front-End: Another system that gets jobs from the WMAgents and sends them to sites.
- Hypernews: <https://hypernews.cern.ch/>, is the mail-history system at CERN. Used mainly for massive-send messages to groups and teams that will be stored. Each email and discussion has it's own URL so it can be referenced. If you're not getting some emails, probably you are not subscribed to the right groups.
- Job: A process that is run in a single site. Usually Workflows make several tasks, and tasks are executed by running several jobs.
- Late-binding: When the processing site of a Job is decided by Condor when the job is about to run,
- MC: Abbreviation for Monte Carlo
- Merge: When some jobs from the same workflow are finished, their output is usually merged so it occupies enough space in order to store it in tape in an efficient way.
- Monte Carlo: (Simulation), some simulation used to produce sample data for different purposes.
- Morgue: A site which has been in the waiting room for a long time gets moved to the morgue. For sites in the morgue, no one is actively trying to bring it back online.

- Overflow: Workflows which run on a site that is not included on the white list, because all of the sites on the white list are busy or not functional.
- PANDA:
- PhEDEx: Physics Experiment Data Export. A system that we use to track the location of each dataset and request data transfers from one site to another.
- ReDigi:
- Reprocessing:
- Request: A petition made by the physics team, that will become a workflow inside the WMAgents.
- Request Manager: An application that receives and handles all request from PDD (i.e. the "pyhichisists").
- ReReco:
- Resubmission: When a workflow is re-injected to WMAgent for some purpose (retry jobs, data had errors, retry from the beginning, etc)
- Robust Merge:
- Savannah: A system peviously used to handle tickets, it was replaced by GGus
- Site: A logical abstraction of an organization (University, Government Institute, etc) which is providing computational power and storage space for our experiment.
- Site whitelist: The list of sites in which a workflow can run.
- Slots: The number of jobs that a site can run simultaneously, this usually relates to the number of CPU cores available in some site.
- SSB: Site Status Board, shows status from it site among several other statistics.
- Stuck Workflow/ Stuck Job: Sometimes workflows don't seem to move, i.e. they won't get completed and won't start any new jobs. Usually happens when some other component (Couch DB, database, Phedex) crashed, so somehow it's status is lost.
- T0 (Tier-0):
- T1 (Tier-1):
- T2 (Tier-2):
- T3 (Tier-3):
- Waiting room: If a site is in the waiting room, it is having issues. No new workflows will be sent to the site.
- WMAgent: Workflow Management Agent, the main tool that we use for creating and handling workflows. It is a system that recieves a request and split it into condor jobs, and send it to the GlideIn Front-end so they can run.
- WMStats: Is a part of the WMAgent, mainly a web application useful for tracking progress of workflows and verifying health status of the WMAgents.
- Workflow: What follows from a job request send by a phycisist/phycicist team. It usually comprises several tasks that create jobs, which are run on several sites.
- WF: Abbreviation for Work Flow
- xrootd: A file access protocol-system based on srm, it is used for reading files from a remote site.

This topic: CMSPublic > WorkflowDictionary

Topic revision: r11 - 2016-03-29 - AllisonReinsvoldHall



Copyright &© 2008-2022 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

or Ideas, requests, problems regarding TWiki? use Discourse or Send feedback