

Table of Contents

Processing procedures.....	1
ReDigi and ReReco requests.....	1
Prestaging.....	1
Assigning requests.....	1
Questions and Comments.....	2
Announcing requests.....	2

Processing procedures

Information about how to deal with processing requests.

ReDigi and ReReco requests

The three main steps for handling ReDigi and ReReco requests are (1) prestaging, (2) assignment, and (3) announcing. Before running any of the scripts please see https://twiki.cern.ch/twiki/bin/view/CMSPublic/CompOpsWorkflowOperationsWMAgentScripts#Loading_WMAgent for details about how to setup the WMAgent environment.

Prestaging

There are two ways to get lists of input datasets which need to be subscribed to disk at a Tier-1. The first way is to run `assignWorkflowsAuto.py`. This will produce a list of workflows which can't be assigned because the input datasets are not on disk. From this list of workflows a list of input datasets can be produced. Alternatively, the script <https://github.com/CMSCompOps/WmAgentScripts/blob/master/listWorkflows.py> run without any arguments lists all ReDigi and ReReco workflows in assignment-approved as well as the input datasets.

The second way is to run the script <https://github.com/alahiff/WmAgentScripts/blob/AndrewFixes/assignInputDatasetsToSites.py>. This will produce a suggested list of PhEDEx subscriptions, and if run with the `-e` option will automatically make these requests. The script assumes that all input datasets should be subscribed to disk at the custodial Tier-1. Datasets with CERN as the custodial site are subscribed to FNAL. Note that no attempt is made to distribute work evenly amongst the Tier-1s. Therefore, I generally manually make subscriptions using the PhEDEx page <https://cmsweb.cern.ch/phedex/prod/Request::Create?type=xfer> after checking http://www.gridpp.rl.ac.uk/cms/reprocessingcampaigns_totals.html to see how much work there is at each site.

Assigning requests

The script <https://github.com/alahiff/WmAgentScripts/blob/AndrewFixes/assignWorkflowsAuto.py> is used to assign ReDigi and ReReco workflows. There are 3 different ways in which it can be used, depending on the options specified:

- **-w** specify a single workflow, e.g. `assignWorkflowsAuto.py -w workflow_name`
- **-f** specify a file containing a list of workflows (one per line), e.g. `assignWorkflowsAuto.py -f file_name`
- with neither `-w` nor `-f`, a list of workflows in assignment-approved will be obtained automatically

By default workflows will not actually be assigned. In order to assign workflows the option `-e` must be specified.

Workflows can be assigned to a specific site using the `-s` option. For example, to assign a workflow to FNAL, use `-s T1_US_FNAL`. If no site is specified, the script will assign the workflow to the Tier-1 which has the input dataset (almost) fully on disk. In order to run a workflow at a site which doesn't have the input dataset on disk, you must specify the `-o` option.

The script will refuse to assign workflows under any the following conditions:

- The input dataset is not VALID or PRODUCTION

- The input dataset is not fully on disk at the site the workflow was assigned to, except if the `-o` option is used
- The pileup dataset (if any) is not fully on disk at the site the workflow was assigned to

Use the `-h` option to get a fully list of possible arguments.

Generally it is enough just to run `assignWorkflowsAuto.py` followed by `assignWorkflowsAuto.py -e` in order to assign workflows. A list of workflows not assigned is also give, as well as the reason why they were not assigned.

Questions and Comments

- the transfer request goes with normal priority, should it be high priority replication ?
- the transfer is not auto approved, what is the latency ?

Announcing requests

I use the script

<https://github.com/alahiff/WmAgentScripts/blob/AndrewFixes/announceReDigiWorkflows.py> to generate lists of workflows which can be announced. It assumes that the following conditions are true for a workflow to be announceable:

- The workflow is in the closed-out state
- The campaign is one of a hardwired list of campaigns
- The output datasets are not already VALID
- The output datasets are $\geq 95\%$ and $\leq 100\%$ complete
- The dataset name doesn't contain `test`, `TEST` or `None`

Clearly there are some missing checks, for example, checking that the number of jobs created was above the appropriate minimum threshold for run-dependent MC. For these you need to check that the number of jobs created was above the appropriate threshold (500 for PU_RD1 or 2000 for PU_RD2).

Example usage:

```
$ python announceReDigiWorkflows.py
Considering: dmason_HCA-Spring14dr-Backfill-00002_00187_v1__141010_072615_4131
Considering: dmason_HCA-Spring14dr-Backfill-00002_00187_v1__141022_060213_3297
Considering: dmason_HCA-Spring14dr-Backfill-00002_00187_v1__141115_031745_5294
Considering: dmason_HCA-Spring14dr-Backfill-00002_00187_v1__141115_031805_9215
Considering: pdmvserv_HIN-HiWinter13DR53X-00071_00013_v1_nomix_150317_175305_5059
Announcing: pdmvserv_HIN-HiWinter13DR53X-00071_00013_v1_nomix_150317_175305_5059
DATA: /Pythia_D0tokaonpion_Pt0_TuneZ2_2760GeV/HiWinter13DR53X-nomix_STARTHI53_V28-v2/DQM
DATA: /Pythia_D0tokaonpion_Pt0_TuneZ2_2760GeV/HiWinter13DR53X-nomix_STARTHI53_V28-v2/GEN-SIM-RECO
```

If run with the `-e` option, then for the announceable workflows:

- Will be changed to the announced state
- Output datasets will be set to VALID
- The script `assignDatasetToSite.py` is run for any AOD, AODSIM, or MINIAODSIM output datasets

I usually run the script above and keep the output:

```
python announceReDigiWorkflows.py > announce
```

To get a list of workflows which can be announced:

CompOpsProcessing < CMSPublic < TWiki

```
cat announce | grep Announc | awk '{print $2}'
```

and to get the corresponding list of datasets

```
cat announce | grep DATA | awk '{print $2}'
```

Once the above has been done, a message should then be sent to the Datasets Announcements HyperNews listing the appropriate datasets (one email per campaign).

This topic: CMSPublic > CompOpsProcessing

Topic revision: r5 - 2015-04-19 - unknown



Copyright &© 2008-2021 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](#)