

Table of Contents

Publish to DBS(DAS) Yourself.....	1
Contacts.....	1
Introduction.....	1
Retrieving / Producing fjr files.....	1
Making a JSON File of Your Dataset.....	2
Publish to DBS.....	3
Elevation to global DBS.....	3
How to change the dataset path.....	3

Publish to DBS(DAS) Yourself

Contacts

- Chris Silkworth
- Lukas Vaneldereren
- Eric Vaandering

Introduction

This twiki explains how to publish something to DBS (DAS), w/o the use of crab. The procedure encompasses the following steps:

- bookkeep file properties in Framework Job Report (fjr) files,
- create your own JSON file,
- install WMAgent on a virtual SLC5-x86_64 machine,
- publish your dataset to the analysis DBS,
- optional: elevation to global DBS
- how to change the datasetpath

The procedure has been tested for only few use cases. So, in case of problems, please contact the authors.

It is assumed that your datasets are stored on some storage element that is accessible from the grid. Your files should be located in a subdirectory of the CMS store directory of the particular storage element. This store directory is usually of the form

```
/<SOMETHING>/cms/<SOMETHING ELSE>/store/
```

Notice:

- When publishing files to the analysis DBS, dataset files are not altered and not copied. In other words, publishing to analysis DBS is nothing more than a bookkeeping procedure.
- Elevating a dataset from the analysis DBS to global DBS encompasses an official copy of your dataset (this is done for you). In global DBS, the copies are publishes and not the original files. For technical reasons, events that were in one single original file may get scattered over several global DBS copies or events from several original files might end up in one single DBS copy.

Disclaimer:

We, the authors of this page, are not DBS experts (far from). We're just sharing our experiences. Let us know if our prescription does not work for you.

Retrieving / Producing fjr files

Framework Job Report (fjr) files summarize the properties of a CMSSW file. In case you have managed your production with crab, fjr files are available in the crab directory. After retrieving the output with 'crab -getoutput -c ', you may list the fjr files as follows:

```
ls <YOUR_CRAB_DIR>/results/crab_fjr_*.xml
```

In case you produced your data sets outside crab, you may produce the fjr files yourself with the script mkfjr.py as follows

```
cd CMSSW_A_B_C/src
```

```
cmsenv
python mkfjr <YOUR_CMSSW_FILE>.root <OUTPUT_FJR_FILE>.xml
```

e.g. for DESY

```
python dcap://dcache-cms-dcap.desy.de:22125//pnfs/desy.de/cms/tier2/store/user/lveldere/pMSSMInte
```

This script will automatically retrieve the lumisections in the CMSSW file, required for publication. For large files this might take a while. Therefore you might consider setting the lumisections by hand using the third optional argument as follows

```
python mkfjr <YOUR_CMSSW_FILE>.root <OUTPUT_FJR_FILE>.xml '{"runId1":[lumiId1,lumiId2,...],"runId
```

mind the quotation marks " versus ', the parser of this option is very sensitive to the syntax.

Making a JSON File of Your Dataset

The information from the fjr is summarized in a json file with the script mkjson.py. First, put all fjr files in a single directory, then run the script as follows:

```
cd CMSSW_A_B_C/src
cmsenv
python mkjson.py <fjr files directory> \
    <dataset path> \
    <version number> \
    <global tag> \
    <application family> \
    <CMSSW version> \
    <storage location> \
    <acquisition era> \
    <output json filename>
```

With

- **dataset path** the desired path for the dataset,
the format of the dataset path is
<PRIMARY DATASET>/<PROCESSED DATASET>/<TIER>
e.g.
/SMS-T1tttt_Mgluino-350to2000_mLSP-0to1650_8TeV-Pythia6Z/Spring12-PU_START52_V9_FastSim-v1/
- **version number** the version number for the dataset,
e.g. 1
- **global tag** the global tag used for the production of the data set
e.g. START42_V11::All
- **application family**
e.g. FastSim
- **CMSSW version**
e.g. CMSSW_5_2_4_patch1
- **storage location**
e.g. cmsrm.fnal.gov (for fermilab), dcache-se-cms.desy.de (for DESY)
- **acquisition era**
e.g. Spring12
- **output json filename**
e.g. T2tt.json

Usual format for datasetpath

```
/<short description of the considered process>/<acquisition era>-<pile up scenario>_<globaltag w/
```

IMPORTANT:

If you plan to elevate your dataset to global DBS, the dataset path should be of the form

```
/<A NAME>/<YOUR HYPERNEWS NAME>--<A NAME>/<A NAME>
```

Elevation can only work when YOUR HYPERNEWS NAME is known to hypernews.

NOTE:

The script is not tested on the fjr files produced by crab. Probably it will not work on these files, but minor changes should make it do its job. Please let us know if you have tested the script of fjr files produced by crab.

EXAMPLE JSON:

example.json

Publish to DBS

Publish to DAS with the script publish2.py. Run it as follows:

```
cd CMSSW_A_B_C/src
cmsenv
python publish.py <file.json> 100
```

where 100 is the block size (this value is supposed to be appropriate).

Note The script will check which of the files listed in the json files are already published. files that are already published are skipped. So, it is easy to publish the files in several goes.

Elevation to global DBS

After publishing to the analysis dbs, a dataset can be elevated to global dbs. This requires approval from the PAG or POG convenors. After approval, follow the procedure outlined in [WorkBookGroupActivities#The_StoreResults_Service](#)

Important Once a dataset is published in global dbs, it is impossible to add further files

How to change the dataset path

A file can only be published once to a given dbs instance and can only be associated to one datasetpath. Thus if you want to change the datasetpath for a certain dataset, you have to rename all files

-- ChristopherSilkworth - 19-Apr-2012

This topic: LCG > PublishToDASYourself

Topic revision: r12 - 2012-10-22 - LukasVanelderren



Copyright &© 2008-2019 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

Ideas, requests, problems regarding TWiki? Send feedback