

Status of grid file transfer

Gérald Grenier

IP2I (previously IPN) Lyon

November 5th, 2020

Context

Grid file catalogue

- A kind of database that matches physical files on a grid server with logical file names.
- Historical catalogue for ILC/CALICE is based on LFC, run on DESY servers.
- LFC runs on SL6 machine that are at end of support.
- All files in the LFC catalogue should either be transferred to DIRAC File Catalogue or deleted.
- Dealing with the LFC content under `/grid/calice/SDHCAL`

The SDHCAL directory

- The `/grid/calice/SDHCAL` directory contains test beam data and various simulation/reconstruction/analysis outputs.
- Test beam data are under `/grid/calice/SDHCAL/TB`
- All TB files have been transferred to DIRAC.
- All LFC **raw data files** have also been transferred to CERN eos, file checksum check ongoing. (checking if files in CERN eos have the same checksum than files in the LFC).
- Have also transferred 15 files that were under the `/grid/calice/SDHCAL/TOMUVOL` directory.

The SDHCAL data

- Data in LFC were saved under directories with various names.
- Raw, streamouted, trivalent data were not necessarily under the same directory name.
- Data that were in LFC under `/grid/calice/SDHCAL/TB/CERN/<NAME>` have been put in DIRAC under `/calice/SDHCAL/TB/CERN/<NAME>`
- When, for a same TB period, raw and other data (streamout, trivent) were in different directories, they have been put under the RAW data name directory.
- All Raw data have been recopied to CERN eos with a different directory name.
- In most of DIRAC-FC directories, there is a file named `lfc_listing.txt` that matches original full LFC file names with full file names in Dirac

SDHCAL data directories in LFC

Directories under `/grid/calice/SDHCAL/TB/CERN` in LFC

Avril2015

SPS_AugSep2012

Mai2015

SPS_Dec2014

PS_April2012

SPS_Jun2016

PS_May2015 → Mai2015

SPS_May2012

SPS_2012 special (see below)

SPS_Nov2012

SPS_2014 → SPS_Dec2014

SPS_Oct2015

SPS_2015 → Avril2015

SPS_Oct2016

SPS_April2015 → Avril2015

SPS_Sept2017

The directory SPS_2012 was containing streamout and trivent files covering more than one test period of 2012. It has been copied to DIRAC directory `/calice/SDHCAL/TB/CERN/ArnaudSteen/SPS_2012`

Structure of Test Beam Data in DIRAC

- Directories in red in the previous slides exist in DIRAC catalogue under `/calice/SDHCAL/TB/CERN/<NAME>`
- Under each of these directories, you may find the following subdirectories :
 - RAW (always there)
 - STREAMOUT
 - STREAMOUT_GGARILLOT
 - TRIVENT
 - TRIVENT_GGARILLOT
 - ROOT
- Special case, the DESY Test Beam data have been transferred from LFC's `/grid/calice/SDHCAL/TB/DESY/` to DIRAC's `/calice/SDHCAL/TB/DESY/` as is.

SDHCAL raw data locations

run range	Directory name in		
	LFC	DIRAC	CERN eos
713775-714207	PS_April2012	PS_April2012	PS_04_2012
714228-714698	SPS_May2012	SPS_May2012	SPS_05_2012
715377-715787	SPS_AugSep2012	SPS_AugSep2012	SPS_08_2012
716200-716433	SPS_Nov2012	SPS_Nov2012	SPS_11_2012
726163-726414	SPS_Dec2014	SPS_Dec2014	SPS_12_2014
728094-728456	Avril2015	Avril2015	SPS_04_2015
728501-728682	Mai2015	Mai2015	PS_06_2015
730437-730926	SPS_Oct2015	SPS_Oct2015	SPS_10_2015
732665-732909	SPS_Jun2016	SPS_Jun2016	SPS_06_2016
733598-733759	SPS_Oct2016	SPS_Oct2016	SPS_10_2016
736443-736576	SPS_Sept2017	SPS_Sept2017	SPS_09_2017
743782-744385		SPS_Sept2018	SPS_09_2018

new directories

SDHCAL raw data locations (continued)

DIRAC

- Data under `/calice/SDHCAL/TB/CERN/<NAME in DIRAC>/RAW/`
- To retrieve data :

```
source /cvmfs/clicdp.cern.ch/DIRAC/bashrc
dirac-proxy-init -g calice_user
dirac-dms-user-lfns -b /calice/SDHCAL/TB/CERN/Avril2015/RAW
dirac-dms-get-file calice-SDHCAL-TB-CERN-Avril2015-RAW.lfns
```
- Few files in CERN eos not yet in DIRAC.

CERN eos

- Data under `/eos/project/s/sdhcal/data/<NAME in eos>/Raw`
- Access from lxplus.
- In addition DESY TB data under `/eos/project/s/sdhcal/data/DESY_01_2012/Raw`

Non Test Beam data files

Considering all the files that are under `/grid/calice/SDHCAL/` but not under `/grid/calice/SDHCAL/TB/`

SDHCAL subdirectories done

Deleted items

Mokka_Output Some old ILD simulations.

ShowerSpliter Rémi's stuff to erase according to Rémi.

rete Rémi's stuff to erase according to Rémi.

pandora_libs Shared libraries

pingault Antoine's green light to delete, only few script/configuration files kept.

rhan test files.

tarballs Shared libraries and executables.

SDHCAL subdirectories done (continued)

Items transferred to DIRAC

Marlin My data for previous studies on MST.

stages_output ROOT files produced by my training students.

generation Source code used by Robert.

kieffer ILD simulation with Mokka of single Klong and WW/ZZ to 2 and 4 jets events **with various cell sizes**.

petr TMVA weight outputs, status unclear but small amount of data, around 100 xml files.

- Files transferred from these directories are stored in my personal user space on DIRAC (under `/calice/users/g/ggrenier/LFC_transfer/`)

Pending SDHCAL subdirectories

- bo** ARBORPFA tests, will be deleted.
- steen/simu** Mostly prototype simulations. Everything can be deleted according to Arnaud and Guillaume but the ww and zz subdirectories (ww and zz have been transferred to DIRAC).
- steen/data** TB analysis output ROOT files, threshold scan results, efficiency maps. Can be deleted according to Arnaud.
- simu** Various ILD and prototype simulations done by Alexei, Arnaud, Rémi and Guillaume. Missing information on what is what. Proposal is to delete.
- haddad** Yacine's stuff. Don't know what it is. No news from Yacine on the subject.

Outlook

- Data transfer from LFC close to the end.
- Non transferred data will be lost, please check urgently if some non-transferred data are to be saved.
- LFC recursive listing in the talk material.

Correspondence userID-people in the file listing

44542 GG

44543 Ran Han

44550 Robert Kieffer

44730 Yacine Haddad

44731 Arnaud Steen

44770 Alexei Petrukhin

44802 Rémi Été

44916 Guillaume Garillot

44932 Antoine Pingault

44989 Bo Li

Backup