

PUBS

David Caratelli, Eric Church, David Kaleko, Michael Kirby,
Zarko Pavlovic, YunTse Tsai, Kazuhiro Terao, Katherine Woodruff

November 20, 2019

Abstract

This document describes PUBS projects used for online data management. It is meant for mostly current and future experts who may want to have an overview about the online data management strategy.

1 Introduction

The online data management uses modulated PUBS projects to stream data files from DAQ to ENSTORE (permanent storage). Each PUBS “project” is an action triggered by a unique (run,subrun) ID. Such projects include a transfer of binary file, metadata extraction, automated detection of troubles and execution of pre-registered solution. PUBS employs a daemon to periodically execute each project to continuously drain data files created by DAQ. It is essential for PUBS to handle incoming data stream to avoid a pile up of data files in the online computing farm.

1.1 Software Dependency

PUBS core component depends on Postgres SQL (9.1 and above) and standard Python 2.7 packages with one additional extension PsyCpg2 for Postgres database access API. However the online PUBS projects do have external dependencies:

- uboonedaq_datatypes v6_15_03
- uboonencode v4.26.00
- ifdhc v1.8.6
- sam_web_client v1_9
- json
- numpy
- matplotlib
- pyqtgraph

1.2 Software Repository

PUBS is maintained at github repository: <https://github.com/drinkingkazu/pubs>

2 EVB Project description

New Run Registration : This project searches for new files produced by the DAQ and parses the file name to search for the (run, subrun) associated with this file. If found, the (run, subrun) information is registered to the SMC database.

- name : register_new_run
- read : ubdaq file
- write : SMC Postgres Server

Binary Checksum : This project calculates the Adler32 checksum on the binary file written by the DAQ. Ultimately the checksum is used for the SAM metadata field.

- name : prod_binary_checksum_evb
- read : ubdaq file
- write : SMC Postgres Server
- activated w/ parent status : kSTATUS_DONE
- successful output status : kSTATUS_DONE

Metadata Registration : This project registers the extracted metadata (json file) to the SAM File Database. This project depends on samweb_cli Python package.

- name : prod_register_binary_evb
- parent : prod_binary_metadata_near1
- read : json file
- write : SAM File Database
- activated w/ parent status : kSTATUS_DONE
- successful output status : kSTATUS_DONE::kSTATUS_VALIDATE_DATA

Binary Data Deletion on evb : This project deletes the .ubdaq binary data file from the evb machine, once the successful transfer to the dropbox has been checked. The project also removes a metadata json file as no longer needed.

- name : clean_binary_evb
- parent : compare_binary_checksum_evb::mv_binary_near1::swizzle_data_near1
- remove : .ubdaq and .ubdaq.json file
- activated w/ parent status : kSTATUS_REMOVE_DATA::kSTATUS_DONE::kSTATUS_DONE
- successful output status : kSTATUS_DONE

3 Near1 Project description

Binary Metadata : This project extracts the metadata from binary file written by the DAQ. It uses uboonaDAQ_datatypes ups product, in particular a binary executable dumpEventHeader. Extracted metadata is stored in a json file along with the original file (same directory with “.json” file extension).

- name : prod_binary_metadata_near1
- parent : prod_binary_checksum_evb
- read : ubdaq file
- create : json file
- activated w/ parent status : kSTATUS_DONE
- successful output status : kSTATUS_DONE

Binary Transfer evb to dCache This project transfers the binary file from evb to the dCache Scratch Dropbox area. It is executed at near1 to take an advantage of dedicated near1 to dCache network as well as private network on which the source file system is mounted. Both network speed is rated as 10 GBit/sec.

- name : prod_transfer_binary_evb2dropbox_near1
- parent : prod_register_binary_evb
- move : ubdaq file
- activated w/ parent status : kSTATUS_DONE
- successful output when file is transferred to Dropbox : kSTATUS_VALIDATE_DATA
- successful output when file cannot be transferred to Dropbox : kSTATUS_TRANSFER_BINARY_NEAR1

Binary Transfer Validation on evb : This project validates that the transfer to the Dropbox has occurred successfully by comparing the binary checksum stored in the SAM database to that stored in the Postgres database.

- name : prod_verify_binary_evb2dropbox_near1
- parent : prod_transfer_binary_evb2dropbox_near1
- read : dCache Scratch Dropbox, SMC Postgres Server
- activated w/ parent status : kSTATUS_VALIDATE_DATA
- DISABLEDD w/ status : kSTATUS_TRANSFER_BINARY_NEAR1
- successful output status for sampling : kSTATUS_SWIZZLE_DATA
- successful output status for removal : kSTATUS_REMOVE_DATA

Binary File Drain to Near1 : This project moves the binary and json file from evb to near1. The project is activated in case the regular path to the dCache space is not available.

- name : prod_transfer_binary_evb2near1_near1
- parent : transfer_binary_dropbox_evb
- move : .ubdaq and .json files
- activated w/ parent status : kSTATUS_TRANSFER_BINARY_NEAR1
- disabled w/ status : kSTATUS_VALIDATE_DATA
- successful output status : kSTATUS_DONE

Binary Transfer on near1 : This project moves the binary file from near1 to the dropbox, and is only relevant when there are data files drained from evb to near1 due to a problem that prohibits direct transfer of files from evb to dropbox.

- name : prod_transfer_binary_near12dropbox_near1
- parent : prod_register_binary_evb
- move : .ubdaq file
- activated w/ parent status : kSTATUS_TRANSFER_BINARY_NEAR1
- disabled w/ status : kSTATUS_VALIDATE_DATA
- successful output status : kSTATUS_DONE

Binary Transfer Validation on near1 : This project validates that the transfer to the Dropbox has occurred successfully by comparing the binary checksum stored in the SAM database to that stored in the Postgres database.

- name : compare_binary_checksum_near1
- parent : transfer_binary_dropbox_near1
- read : dCache Scratch Dropbox, SMC Postgres Server
- activated w/ parent status : kSTATUS_DONE
- disabled w/ status : kSTATUS_VALIDATE_DATA
- successful output status : kSTATUS_DONE

Binary Data Deletion on near1 : This project deletes the .ubdaq binary data file from the near1 machine, once the successful transfer to the dropbox has been checked.

- name : clean_binary_near1
- parent : prod_verify_binary_ebv2dropbox_near1
- remove : .ubdaq and .ubdaq.json file
- activated w/ parent status : kSTATUS_DONE
- disabled w/ status : kSTATUS_VALIDATE_DATA
- successful output status : kSTATUS_DONE

Online Swizzling on near1 : DISABLED! DISABLED! DISABLED! DISABLED! DISABLED! This project swizzles binary files on evb.

- name : prod_swizzle_data_near1
- parent : prod_verify_binary_ebv2dropbox_near1
- read : .ubdaq file
- write : .root file
- activated w/ parent status : kSTATUS_SWIZZLE_DATA
- disabled w/ status : kSTATUS_SKIP_SWIZZLE;kSTATUS_TRANSFER_BINARY_NEAR1
- successful output status : kSTATUS_DONE

ROOT Checksum Calculation : DISABLED! DISABLED! DISABLED!
DISABLED! DISABLED! Calculate the checksum of the ROOT file created by
the online swizzling

- name : prod_root_checksum_near1
- parent : prod_swizzle_data_near1
- read : .root file
- activated w/ parent status : kSTATUS_DONE
- disabled w/ status : kSTATUS_SKIP_SWIZZLE:kSTATUS_TRANSFER_BINARY_NEAR1
- successful output status : kSTATUS_DONE

ROOT Metadata Extraction : DISABLED! DISABLED! DISABLED! DIS-
ABLED! DISABLED! Extract metadata from the ROOT file created

- name : prod_root_metadata_near1
- parent : prod_root_checksum_near1
- read : .root file
- activated w/ parent status : kSTATUS_DONE
- disabled w/ status : kSTATUS_SKIP_SWIZZLE:kSTATUS_TRANSFER_BINARY_NEAR1
- successful output status : kSTATUS_DONE

ROOT Metadata Registration : DISABLED! DISABLED! DISABLED!
DISABLED! DISABLED! Register metadata to the SAM database

- name : prod_register_root_near1
- parent : prod_root_checksum_near1
- read : .root file
- write : SAM File Database
- activated w/ parent status : kSTATUS_DONE
- disabled w/ status : kSTATUS_SKIP_SWIZZLE:kSTATUS_TRANSFER_BINARY_NEAR1
- successful output status : kSTATUS_DONE

ROOT File Transfer : DISABLED! DISABLED! DISABLED! DISABLED!
DISABLED! Transfer ROOT file to Dropbox

- name : prod_transfer_root_near12dropbox_near1
- parent : prod_register_root_near1
- move : .root file
- activated w/ parent status : kSTATUS_DONE
- disabled w/ status : kSTATUS_SKIP_SWIZZLE:kSTATUS_TRANSFER_BINARY_NEAR1
- successful output status : kSTATUS_DONE

ROOT File Transfer Validation : DISABLEDD! Verify the successful transfer of the ROOT file to the Dropbox

- name : prod_verify_root_near12dropbox_near1
- parent : prod_transfer_root_near12dropbox_near1
- read : .root file on near1 and on Dropbox
- activated w/ parent status : kSTATUS_DONE
- disabled w/ status : kSTATUS_SKIP_SWIZZLE:kSTATUS_TRANSFER_BINARY_NEAR1
- successful output status : kSTATUS_DONE

ROOT Data File Removal : DISABLED! DISABLED! DISABLED! DISABLED! DISABLED! Remove the .ROOT file from near1

- name : prod_clean_root_near1
- parent : prod_verify_root_near12dropbox_near1
- remove : .root file on near1
- activated w/ parent status : kSTATUS_DONE
- disabled w/ status : kSTATUS_SKIP_SWIZZLE:kSTATUS_TRANSFER_BINARY_NEAR1
- successful output status : kSTATUS_DONE

ROOT Metadata File Removal : DISABLED! DISABLED! DISABLED! DISABLED! DISABLED! Remove the ROOT .json file from near1

- name : prod_clean_root_json_near1
- parent : prod_verify_root_near12dropbox_near1
- remove : .json file on near1
- activated w/ parent status : kSTATUS_DONE
- disabled w/ status : kSTATUS_SKIP_SWIZZLE:kSTATUS_TRANSFER_BINARY_NEAR1
- successful output status : kSTATUS_DONE