Agents And Daemons

Automating Data Quality Monitoring Operations

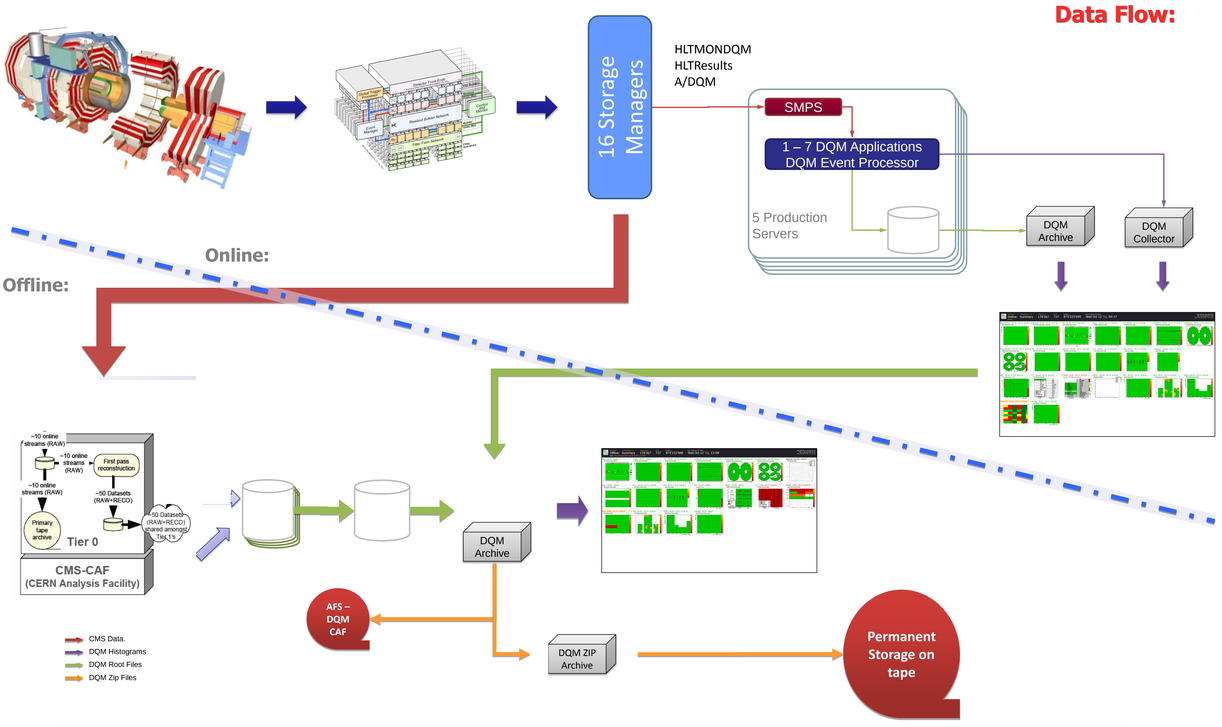
Introduction:

Since 2009 when the LHC came back to active service, the Data Quality Monitoring (DQM) team was faced with the need to homogenize and automate operations across all the different environments within which DQM is used for data certification.

The main goal of automation is to reduce operator intervention at the minimum possible level, especially in the area of DQM files management, where long-term archival presented the greatest challenges. Manually operated procedures cannot cope with the constant increase in luminosity, datasets and time of operation of the CMS detector. Therefore a solid and reliable set of agents has been designed since the beginning to manage all DQM-data related work-flows. This allows to fully exploit all available resources in every condition, maximizing the performance and reducing the latency in making data available for validation and certification. The agents can be easily fine-tuned to adapt to current and future hardware constraints and they proved to be flexible enough to include unforeseen features, like an ad-hoc quota management and a real time sound alarm system.

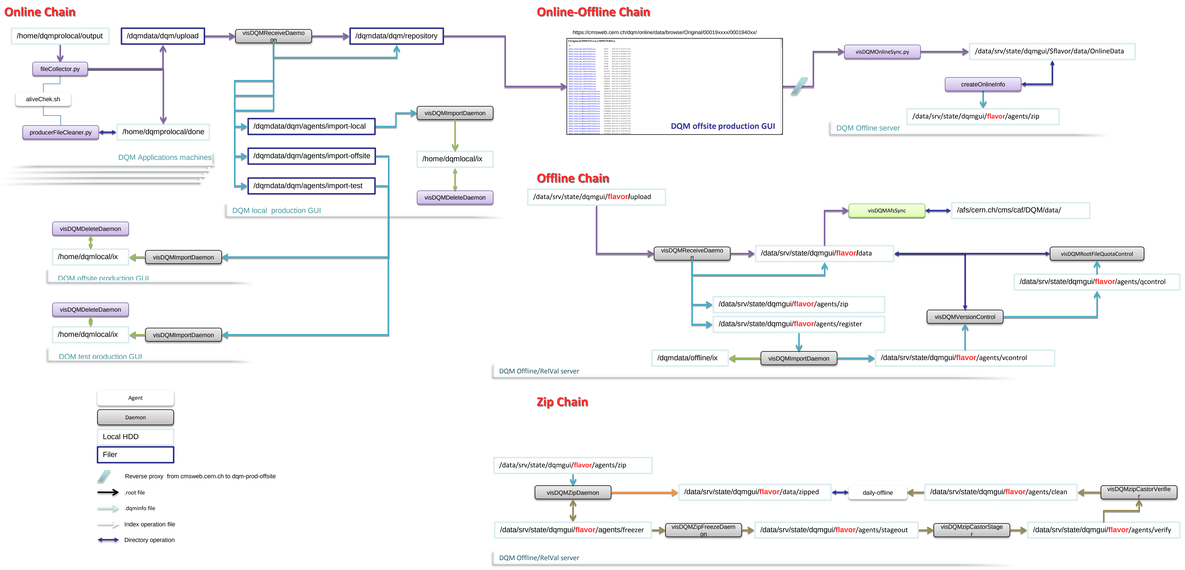
Finally an agent is a script that performs a tasks and then exits, usually invoked from a cron job, and a daemon is a script that never ends, it loops forever sleeping for small intervals as to not hug the cpu.

Data Flow:



Processing Chains:

In order to ensure that the generated DQM data is properly stored in tape and registered in the required GUIs, the daemons and agents are arranged in chains of processing. This allows to solve communication between the each agent and daemon, it also easily allows for branching to support multiple GUI indexes, which is a useful feature in case of recovery, and parallel operation since everything is file directory based, multiple instances of the same chain can be running at the same time on the same machine, without interfering with each other.



A&D Description:

A/D	Name	Dscription
A	visDQMAfsSync	Syncs partially the offline repository with a folder on afs, runned by acron
D	visDQMCreateInfoDaemon	Keeps taps on a directory and creates dqminfo files for root files that do not have them. And puts them in the zip agent drop box for archival
D	visDQMDeleteDaemon	Deletes runs/datasets based on queue files from the GUI index
D	visDQMImportDaemon	Adds root files to the index
D	visDQMIndexMergeDaemon	Not shown in the chains, is a helper daemon, It is useful to merge partial indexes with the main indexes coordinating with the import daemon
D	visDQMOnlineSyncDaemon	Downloads root files from the online GUI into the offline root repository.
D	visDQMReceiveDaemon	It picks up files from the upload directory, checks the naming conventions, and creates dqminfo files.
D	visDQMRootFileQuotaControl	Keeps root repository usage of disk space in check.
D	visDQMVerControlDaemon	Removes older version of root files, only keeping the latest available.
D	visDQMZipCastorStager	Stages zip archives in castor for permanent storage
D	visDQMZipCastorVerifier	Verifies that the archives have been copied correctly.
D	visDQMZipDaemon	Creates zip archives with the root files.
D	visDQMZipFreezeDaemon	Prevents new info being added to an archive, and marks it as ready for transfer to castror
A	aliveCheck(2).sh	Ensures that the producer machine A&D are running
D	fileCollector2.py	Collects files from producer machines in emulates an upload to the online receive daemon dropbox.
D	producerFileCleaner.py	It keeps the disk usage of a local repository of the producer machines in check
A	daily-offline	Among other things it removes the zip archives from the zipped repository that have been successfully transferred to castor.

A&D In numbers:

1. Root files processed in the last 2 Years
   - Online: 325463
   - Offline: 258977
   - Relval: 9593
   - RelvalData: 64316
   - MonteCarlo: 6519
2. Zip files processed in the last 2 Years
   - Online: 1776
   - Offline: 22180
   - Relval: 518
   - RelvalData: 4914
   - MonteCarlo: 5300
3. Amount of data transferred to castor: 52TB
4. Number of critical (unrecoverable errors): 0
5. Number of interventions per year: 2
6. Number of flavors running the A&D: 5

Luis I. Lopera, on behalf of the DQM Group
Los Andes University
Bogotá, Colombia