Using Docker to execute offline instances of the CMSSW workflows
Openlab Project

Alexandre Luiz Brisighello Filho

August 19, 2015
Objectives

- Provide an easy way to generate offline and system-agnostic Linux containers of CMSSW.
- Have new options to use as benchmark.
- Allow the framework to be used outside CERN or in a machine that doesn’t have network access.
Problems

Usage of two external sources:
- CMS root files (confidential data).
- Conditional Database

Adding files to the container.
- What should be added?
- How to make it smaller?

Automation and usage of extra tools.
- How to make it painless?
- What if the user wants to use extra tools with the container?
Solutions

Root Files:
- Automated identification of needed data and picking of events (reducing size).

Conditional Data:
- Usage of local squid cache with custom configuration.
- Sqlite3 modification to bypass frontier client.

Container Files:
- Use of CARE tool (http://reproducible.io/)

Automation:
- Development of a command-line tool using python that do all the work.
- Dynamic generation of execution scripts with optional parameters.

August 19, 2015
Alexandre Luiz Brisighello Filho
Using Docker to execute offline instances of the CMSSW workflows
python morpheus.py docker -l 1

#1 Run: Identifies data files and commands used

#2 Run: Packs programs needed for execution

#3 Run: Creates the container and fills up squid cache
The End

August 19, 2015

Alexandre Luiz Brisighello Filho

Using Docker to execute offline instances of the CMSSW workflows