CSA07 - primary datasets and skimming

Benedikt Hegner

DESY Hamburg

The CSA07 soup

- 130 M events for CSA07 exercise + signal samples
- http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/CMSSW/Configuration/CSA07Production/data/

Pre-CSA07 and CSA07 workflows

- Simulation with 14X and 15X
- HLT with 160
- DigiToRaw teststep
- Splitting in 11 Primary Datasets (PD)
- Raw to Digi
- RECO
- Skimming
- Shipping to T2

Requirements

- each analysis should be able to run on only 1 or two PDs

- PDs should be as small as practical

- Keeping prescaling as simple as possible

- Decision only on HLT information

"Final" list of PDs

- Tau, Muon, Electron, Photon, bJet, JetMET

- MuonCalib, bJetCalib, JetCalib, METCalib

- Failed (= both L1 and HLT have failed)

- http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/CMSSW/HLTrigger/Configuration/data/HLTrigger_PrimaryDatasets.cff

Definition

- Subset of available events and
- Subset of single event information
- Common data for one or more physics groups
- Not at single analysis level

Current status

- In total > 100 skims have been proposed
- The number is now down to 46
- Skims are tested on a basic level
- There is a more detailed talk on Thursday (by Joe Incandela)
- If we want to have skims at DESY we have to subscribe to the dataset