Contribution ID: **237** Type: **oral presentation**

Real-time dataflow and workflow with the CMS Tracker data

Thursday 6 September 2007 17:10 (20 minutes)

The Tracker detector has been taking real data with cosmics at the
Tracker Integration Facility (TIF) at CERN.
First DAQ checks and on-line monitoring tasks are executed at the
Tracker Analysis Centre (TAC) which is a dedicated Control Room at TIF with
limited computing resources. A set of software agents were developed
to perform the real-time data conversion in a standard Event Data Model format, the
copy of RAW data to CASTOR storage system at CERN and the registration of them in the
official CMS bookkeeping systems.
According to the CMS computing and analysis model, most of the
subsequent data processing has to be done in remote Tier-1 and Tier-2 sites, so
data are automatically injected for the transfer from the TAC to the sites
interested to analyze them, currently Fermilab, Bari and Pisa.
Official reconstruction in the distributed environment is triggered in
real-time from Bari by using the ProdAgent tool, currently used with
simulated data. Data are reprocessed with the most recent (pre-)releases
of the official CMS software to provide immediate feedback to the
software developers and the users. Automatic end-user analysis of published
data is performed via CRAB tool to derive the distributions of the most
important physics variables.
A monitoring system to check all the steps of the processing chain is
also under development. An overview of the status of the tools developed
is given, together with the evaluation of the real-time performance
of the chain of tasks.

Summary

The Tracker detector has been operated with cosmic events at the
Tracker Integration Facility (TIF) at CERN. Data are checked via on-line data quality
monitoring tools running at the Tracker Analysis Centre (TAC) which is a dedicated
Control Room with limited computing resources.
Procedures are also developed and executed in real-time to make data officially
available to the CMS community so raw data are firstly converted in a standard
format, then archived on CASTOR storage system at CERN and registered in the
official CMS data bookkeeping (DBS) and data location (DLS) systems.

The local storage available at TAC computers is sufficient to cache incoming data for
about 10 days, and is clearly the best solution for fast-response analyses and DAQ
checks. On the other side, a large community is expected to analyze data taken at the
TAC, and this cannot happen at because of the limited resources.

Data are expected to flow from the TAC to the CMS Tier-1 and Tier-2 remote sites and
to be accessed using standard CMS tools. Once data are registered in the DBS and DLS
they are ready to be transferred in remote sites using the CMS official data movement

tool PhEDEx. This operation requires that data are injected in the PhEDEx transfer management database to be routed to destination sites; a set of scripts are developed to perform this operation periodically in order to send data in the sites interested to analyze them, currently Fermilab, Bari and Pisa.

Official reconstruction in the distributed environment is automatic and triggered in real-time from a Bari machine by using a set of scripts optimized to run the ProdAgent tol, just used to reconstruct Monte Carlo simulated data. Data are reprocessed with the most recent releases and prereleases of the official CMS software to provide immediate feedback to the software developers and the users. A parallel reprocessing executed at FNAL with the last package patches is used to test the performances of the track reconstruction algorithms in real-time with read data.

Reconstruction, re-reconstruction, calibration and alignment tasks running at remote sites which requires to access data in the offline database located at CERN are run by using FroNTier software to access those data remotely.

Automatic end-user analysis of published data is performed via CRAB tool to derive the distributions of the most important physics variables.
A monitoring system to check all the steps of the processing chain is also under development. An overview of the status of the tools developed is given, together with the evaluation of the real-time performance of the chain of tasks.

Authors: Dr NOEDING, Carsten (Fermi National Accelerator Laboratory, Batavia (IL), USA); Dr GIORDANO, Domenico (Università di Bari e INFN Sezione di Bari); Dr PALLA, Fabrizio (INFN - Sezione di Pisa); Dr BAGLIESI, Giuseppe (INFN - Sezione di Pisa); Dr DE FILIPPIS, Nicola (INFN Bari); Dr SARKAR, Subir (INFN - Sezione di Pisa); Dr BOCCALI, Tommaso (INFN - Sezione di Pisa); Dr CIULLI, Vitaliano (Università di Firenze e INFN Sezione di Firenze)

Co-authors: Dr MIRABITO, Laurent (CERN PH/CMT); Dr BAINBRIDGE, Robert (High Energy Physics Group Blackett-Lab - Imperial College, London)

Presenter: Dr DE FILIPPIS, Nicola (INFN Bari)

Session Classification: Distributed data analysis and information management

Track Classification: Distributed data analysis and information management