User's Friendly Interface to the CDF Data Handling System

F.Ratnikov[1] for the CDF Data Handling Group
[1]Rutgers, The State University of NJ, USA

Abstract

The CDF collaboration at the Fermilab Tevatron analyses proton-antiproton interactions at a center-of-mass energy of 2 TeV. During the the collider run starting this year the experiment expects to record 1 Petabyte of data and associated data samples. The Data Handling (DH) system has online and offline components.

The DH offline component provides access to the stored data, to stored reconstruction output, to stored Monte-Carlo data samples, and user owned data samples. It serves more than 450 physicists of the collaboration. The extra requirements to the offline component of the Data Handling system are simplicity and convenience for users. More than 50 million events of the CDF Run II data have been already processed using this system.

Keywords: CDF, Data handling, Data access, User interfaces

1 Introduction

The CDF Run II Data Handling (DH) [1] system consists of the following components:

- Tape library, permanently storing all the data
- Disk Inventory Manager (DIM) [2] responsible for staging of necessary data from tape to the local disk pool, putting produced output data on tape and manipulating the local disk pool
- Data File Catalog (DFC) [3] -- Oracle database keeping track of all files, filesets and datasets, information about data quality and luminosity of different data units, data processing conditions and other relevant information

The CDF analysis framework [4] has a modular structure, including analysis modules processing the event, and input and output modules responsible for data IO. Data Handling Input and Output modules interface between the DH system and the data processing framework, retrieving necessary data from the DH system, delivering them to the analysis modules and sending produced data back to the DH system.

CDF Run II data are organized hierarchically. The minimal unit is a runsection that combines data taken during about 30 sec of continuous detector running. Several runsections are combined into files, files are combined into filesets and filesets are combined into datasets [1]. One file is the unit for DH Input and Output operation, the DIM manipulates whole filesets, and users are usually interested in particular dataset.

CDF Run II data are written in the ROOT format [5], all data being contained in one or several branches of the ROOT tree.

2 Data Input

Input data may be requested by any combination of datasets, filesets or files. Additional restrictions on runs or runsections may be applied. The DH Input communicates first with DFC building the complete list of files to be processed and filesets containing these files. After the data processing is started, the list of required filesets is delivered to the DIM which controls availability of the required data to the processing job. DH Input requests data by filesets, the order of filesets being determined by DIM for minimizing the time while process waits for the next data are staged from tape to disk. Read ahead is used to make data delivery even more effective: staging for couple of next filesets is started while previous fileset is processed. Once the DIM delivers fileset on a disk, it reports the actual data path to DH Input that selects the necessary files and reads them one by one. The user can speed up the processing of small data samples by restricting the data sample to those data, that are available in the DIM disk pool already, and therefore not requiring time consuming staging from tape.

CDF data are stored in ROOT format files, which are essentially direct access files. To use this opportunity for the direct event pick up, the event attributes like run and event numbers should be associated with index of the event in the file. DH Input builds such a catalog on the fly using either event information itself (time consuming procedure) or using extra dedicated Event Information branch of the primary ROOT tree (fast procedure). DH Input then is able to access every event directly. The DH Input also triggers "Begin of run" condition (BOR) and delivers corresponding BOR record to the analysis system before any event of new run is processed. The BOR record contains various run specific information necessary for processing events of the run.

3 Data output

To store output data in the Data Handling system, an output file should be produced, a new record should be made in the FILE record of the DFC and the file should be released to the DIM for archiving. DH Output handles streams of output events taking care that produced data are consistent. When a current output file becomes large enough, DH Output closes it and opens a new one. The name of the file describes the data contained and is uniquely defined by the output dataset and by run and runsections contained in the file. The CDF detector luminosity is determined per runsection, so all events of each runsection for a given dataset should be contained in one file to keep track of the integrated luminosity corresponding to the file. DH Output takes care that an output file is closed only when events signal the change of runsection.

Due to the varying latency of the CDF Level 3 processing farm, events are put into primary RAW data streams out of the original order. To keep events of the same runsections compact in the file, events in the output stream should be grouped by runsections. The processing of the events in the original order is also natural for many event-by-event analyses. DH Input accesses individual events directly so it can be requested to process events of each input file in natural order. Assuming that runsections are compact in input files, they thus keep compact in the output file.

The DH interface provides the possibility for analysis jobs to iterate through huge datasets. A job could crash on the way due to many different reasons and need to be restarted. The DH interface provides an efficient way to do this. Although the current output file is considered corrupted, previous files are good and have been registered in the DFC already. Then DH Output checks events of the output stream against DFC records about data of the output dataset that were processed already, avoiding double counting. Together with an ability of DH Input to exclude the already processed part of data from the input data list, this method keeps output data sample complete and consistent with minimal overhead of data processing.

4 Operating out of DH system

Although the primary goal of the DH interfaces is delivering data from the DH system to the analysis job and putting produced output back to the DH system, both DH Input and Output can also operate with any data files on disk. It provides a uniform interface to the data that is identical on both fully equipped central analysis computing system and the stand-alone desktop, having no access to DIM, and/or DFC. The ability to run out of the DH system is also important for working with relatively small data samples and for debugging purposes.

References

[1] S. Lammel et al., "First Operational Experience with the CDF Run II Data Handling System" this conference.

[2] Paul Hubbard, S. Lammel, "The CDF Run II Disk Inventory Manager" this conference.

[3] D. Litvintsev et al., "CDF Run II Data File Catalog" this conference.

[4] L. Sexton-Kennedy et al., "The Physical Design of the CDF Simulation and Reconstruction Software", CHEP2000 proceedings.

[5] R. Kennedy et al., "The CDF Run II Event Data Model", CHEP2000 proceedings.