Contribution ID: **130** Type: **oral presentation**

A skimming procedure to handle large datasets at CDF

Monday 13 February 2006 16:20 (20 minutes)

The CDF experiment has a new trigger which selects events depending on the significance of the track impact parameters. With this trigger a sample of events enriched of b and c mesons has been selected and it is used for several important physics analysis like the Bs mixing. The size of the dataset is of about 20 TBytes corresponding to an integrated luminosity of 1 fb-1 collected by CDF. CDF has developed a skimming procedure to reduce the dataset by selecting events which contain only B mesons in specifics decay modes. The rejected events are almost background, and this guarantees that no signal is lost while the processing time is reduced by factor 10. This procedure is based on SAM (Sequential Access via Metadata), the CDF data handling system. Each file from the original dataset is read via SAM and processed on the CDF users farm at Fermilab. The outputs are stored and cataloged via SAM on a temporary disk location at Fermilab in order to be finally concatenated. This final step consists of copy and then store and catalog the output in Italy on disks hosted at Tier 1, permanently. These skimmed data are available in Italy for the CDF collaboration, and user can access them via the Italian CDF farm. We will describe the procedure to skim data, concatenate the output and the method used to control that each input file is processed once and only once. The tool to copy data from the users farm to temporary and permanent disk locations, developed by CDF,consists of users authentication plus a transfer layer. Users allowed to perform the copy are mapped in a gridmap file and authenticated with a Globus Security Infrastructure (GSI). Details on the tool performances and the use and the definition of a remote permanent disk location will be described in detail.

Primary author: Dr LUCCHESI, Donatella (INFN Padova)

Co-authors: Dr FELLA, Armando (INFN Pisa); Dr DELLI PAOLI, Francesco (INFN Padova); Dr CASARSA, Massimo (INFN Trieste); Dr DA RONCO, Saverio (INFN Padova)

Presenters: Dr LUCCHESI, Donatella (INFN Padova); Dr DELLI PAOLI, Francesco (INFN Padova)

Session Classification: Distributed Data Analysis

Track Classification: Distributed Data Analysis