



Contribution ID: 40

Type: Poster

The Long Term Data Preservation (LTDP) project at INFN CNAF: CDF user case.

Tuesday, 2 September 2014 08:00 (1 hour)

In the last years the problem of digital preservation of valuable scientific data has significantly become one of the most important points to consider in scientific collaborations. In particular the long term preservation of almost all experimental data, raw and all related derived formats including calibration information, is one of the emerging requirements within the High Energy Physics (HEP) community for experiments that have already concluded the data taking phase. The DPHEP group (Data Preservation in HEP) coordinates the local teams within the whole collaboration and the different Tiers (computing centers). The INFN CNAF Tier1 is one of the reference sites for data storage and computing in the LHC community but it also offers resources to many other HEP and non-HEP collaborations. In particular the CDF experiment has used the INFN CNAF Tier1 resources for many years and after the end of data taking in 2011, it is now facing the challenge to both preserve the large amount of data produced during several years and to retain the ability to access and reuse the whole amount of it in the future. According to this task the CDF Italian collaboration, together with the INFN CNAF computing center, has developed and is now implementing a long term future data preservation project in collaboration with the FNAL computing sector. The project comprises the copy of all CDF raw data and user level ntuples (about 4 PB) at the INFN CNAF site and the setup of a framework which will allow to access and analyze the data in the long term future. Therefore a big portion of the 4 PB of data (raw data and analysis-level ntuples) are currently being copied from FNAL to the INFN CNAF tape library backend and the system, which subsequently allows the data access, is being set up. In addition to this data access system, a data analysis framework is being developed in order to run the complete CDF analysis chain in the long term future, from raw data reprocessing to analysis-level ntuples production. In this contribution we first illustrate the difficulties and the technical solutions adopted to copy, store and maintain CDF data at the INFN CNAF Tier1 computing center. In addition we describe how we are exploiting virtualization techniques for the purpose of building the long term future analysis framework, and we also briefly illustrate the validation tests and techniques under development in order to check data integrity and software operation efficiency over time.

Primary authors: GREGORI, Daniele (Istituto Nazionale di Fisica Nucleare (INFN)); DELL'AGNELLO, Luca (INFN-CNAF); RICCI, Pier Paolo (INFN CNAF); Ms AMERIO, Silvia (University of Padova & INFN); PEZZI, Michele (Infn-cnaf)

Presenter: RICCI, Pier Paolo (INFN CNAF)

Session Classification: Poster session

Track Classification: Computing Technology for Physics Research