

Managing data recovery for Long Term Data Preservation

Tuesday, July 10, 2018 4:45 PM (15 minutes)

In the latest years, CNAF worked at a project of Long Term Data Preservation (LTDP) for the CDF experiment, that ran at Fermilab after 1985. A part of this project has the goal of archiving data produced during Run I into recent and reliable storage devices, in order to preserve their availability for further access through not obsolete technologies. In this paper, we report and explain the work done to manage the process of retrieving the aforementioned data, which were stored into about four thousands 2.5/5GB 8mm tape cartridges of different producers, which were widely popular in the nineties. The hardware setup for tape reading is briefly detailed. Particular focus is on describing in-house software tools and backend database that have been set up to drive and orchestrate the tape readers and to deal with the high number of possible problems arising during the process of reading data from hardly reliable media. The outcome of each operation is accounted into the database, making possible to monitor the overall progress and to retry unsuccessful read attempts at a later stage. The implemented solution has proved effective at reading a first 20% of the total amount. The process is currently ongoing. Eventhought a few aspects of this work are strictly dependant on how the CDF experiment organized its datasets, we believe that several decisions taken and the overall organization still make sense on a variety of use cases, where a relevant amount of data has to be retrieved from obsolete media.

Primary authors: Dr DAL PRA, Stefano (INFN); Dr FATTIBENE, Enrico (INFN - National Institute for Nuclear Physics); RICCI, Pier Paolo (INFN CNAF); FALABELLA, Antonio (Universita e INFN, Bologna (IT))

Presenter: Dr DAL PRA, Stefano (INFN)

Session Classification: Posters

Track Classification: Track 4 - Data Handling