Long Term Data Preservation for CDF at INFN-CNAF

2014 J. Phys.: Conf. Ser. 513 042011

(http://iopscience.iop.org/1742-6596/513/4/042011)

Journal of Physics: Conference Series **513** (2014) 042011 doi:10.1088/1742-6596/513/4/042011

Long Term Data Preservation for CDF at INFN-CNAF

S. Amerio[1], L. Chiarelli[2], L. dell'Agnello[3], D. De Girolamo[3], D. Gregori[3], M. Pezzi[3], A. Prosperini[3], P. Ricci[3], F. Rosso[3], and S. Zani[3]

[1] University of Padova, Italy
[2] GARR, Rome, Italy
[3] INFN-CNAF, Bologna, Italy

E-mail: silvia.amerio@pd.infn.it

Abstract. Long-term preservation of experimental data (intended as both raw and derived formats) is one of the emerging requirements coming from scientific collaborations. Within the High Energy Physics community the Data Preservation in High Energy Physics (DPHEP) group coordinates this effort. CNAF is not only one of the Tier-1s for the LHC experiments, it is also a computing center providing computing and storage resources to many other HEP and non-HEP scientific collaborations, including the CDF experiment. After the end of data taking in 2011, CDF is now facing the challenge to both preserve the large amount of data produced during several years of data taking and to retain the ability to access and reuse it in the future. CNAF is heavily involved in the CDF Data Preservation activities, in collaboration with the Fermilab National Laboratory (FNAL) computing sector. At the moment about 4 PB of data (raw data and analysis-level ntuples) are starting to be copied from FNAL to the CNAF tape library and the framework to subsequently access the data is being set up. In parallel to the data access system, a data analysis framework is being developed which allows to run the complete CDF analysis chain in the long term future, from raw data reprocessing to analysis-level ntuple production. In this contribution we illustrate the technical solutions we put in place to address the issues encountered as we proceeded in this activity.

1. Introduction

Interest in the long term preservation of scientific data and their availability to general public is growing. Data collected in High Energy Physics (HEP) experiments are the result of a significant human and financial effort. The preservation of HEP data beyond the lifetime of the experiment is of crucial importance to ensure the long term completion and extension of scientific programs, to allow cross collaboration analysis, analyzing data from several experiment at once, to perform new analysis with new theoretical models and techniques and for education, training and outreach.

HEP data preservation poses many technical and organizational challenges: data preservation implies migration to new storage media when available, adjusting data access methods if needed; moreover data analysis capabilities must be preserved ensuring the experiment legacy software runs on new platforms, or on old ones with no security issues; validation systems have to be set up to regularly check data access and analysis framework; all the information needed to access and analyze data has to be properly organized and archived.

In this paper we describe a project to preserve at CNAF computing center in Italy a complete copy of the data of CDF experiment at Tevatron. The project aims having at CNAF a copy of all raw data and user level ntuples (4 PB) and providing users with data access and data analysis capabilities in the long term future.

2. CDF Computing Model

The data taking phase of the CDF experiment ended in September 2011. Total collision and simulation data amount to about 10 PB, of which 4 PB are raw and ntuple-level data. These samples are currently stored on the dCache [1] Mass Storage Sytem at Fermilab (more specifically the data are saved on a T10K technology tape library) and the data handling is performed through a specific tool, SAM (Sequential data Access via Metadata) [2].

CDF reconstruction and analysis code is written in C, C++ and Python languages and it is preserved in frozen releases in CVS repositories. The latest version of CDF code runs on SL5 operating system; a SL6 legacy release is being prepared and will be ready by the end of 2013. The CDF Central Analysis Farm code (CAF) provides the users with a uniform interface to resources on different Grid sites [3]. Three portals based on glideinWMS [4] allow users to access computing resources at Fermilab, OSG, CNAF Tier 1 and other LCG sites. Authentication is based on Kerberos.

3. FNAL-CNAF data transfer and storage

To minimize usage contention for tape library access at CNAF, we plan to copy all CDF data (4 PB) before the start of LHC data taking in 2015. To meet this tight constraint we setup a dedicated system able to transfer data at 5 Gb/s sustained rate, and copy them to CNAF tape system, automatically updating the FNAL database. The copy will be driven by CNAF using the CDF SAM data handling system (see Fig. 1). A request from CNAF SAM station triggers data retrieval from the tape system to the disk buffer at FNAL; the data are then copied via gridftp to CNAF, where they are automatically archived to tape. In the following sections we describe in more detail the network and storage layouts.

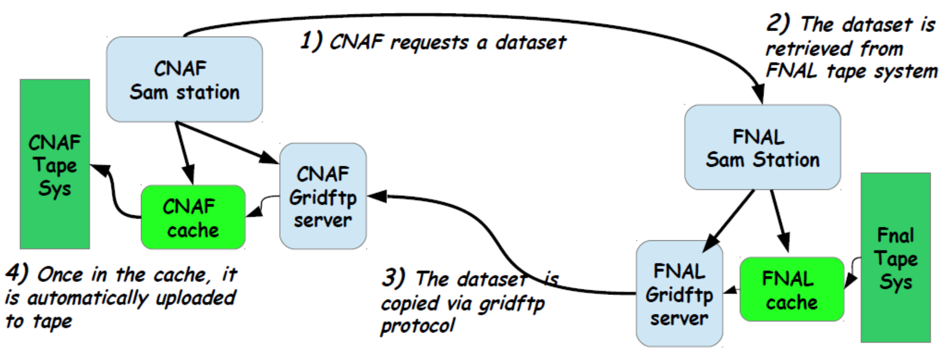

Figure 1. Mechanism to copy CDF data files from Fermilab to CNAF computing center.

3.1. Network Layout for FNAL-CNAF CDF Data Copy

As shown in Figure 2, we are using dedicated network resources for the CDF long term data preservation project. At CNAF 2 servers with 10GE network interfaces, on a dedicated VLAN, are directly connected to our core switch/router. For the geographical connectivity a dedicated 10 Gbps connection is avaiable and a dedcated class-C network is used for this purpose with a specific BGP peering with GARR network. This link is separated from the ones used by

LHCOPN/LHCONE peerings and for General Internet. The separation of CDF traffic allows us to manage and monitor CDF data movement independently from our Tier 1 network resources and to have a secure and reliable high speed channel for data transfers.

GARR, the Italian Research Network, has ulso provided dedicated network resources to this project: a L3 VPN from Bologna and Milan routers and a Lambda from Milan and Geneve with SWITCH network. Finally, from Geneva to Chicago there are 2x10G links provided by T-Systems that connect GEANT to Starlight and FNAlL: on these links a 5 Gbps dedicated channel has been assigned to this project.

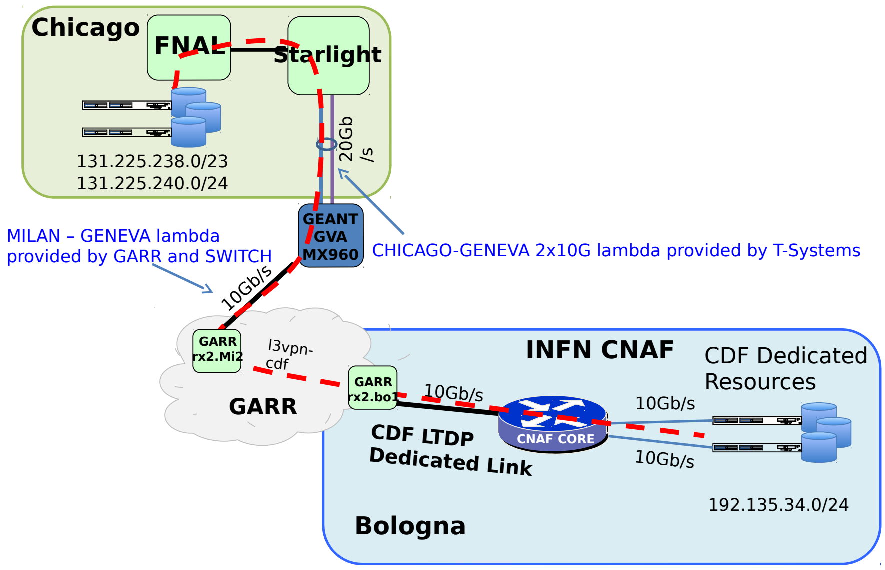

Figure 2. Layout of the FNAL-CNAF copy network.

3.2. Storage Layout for FNAL-CNAF CDF Data Copy

The adopted solution at CNAF for the so-called 'bit preservation' of CDF data is GEMSS [5], a well established system, in production since several years at CNAF. GEMSS consists of a pool of disks managed by GPFS, a tape library infrastructure for the archive back-end managed by TSM and an integration system to transfer data from disk to tape and vice versa. On top of this system, a srm service, StoRM, provides access control for the LHC experiments over the WAN; this is not used from CDF.

The storage layout system is composed by several elements. SAM data handling tools are installed on a dedicated machine, SAM station in the following. The SAM station is the core of the transfer method and it is basically a Virtual Machine (VM) controlling the data transfer and updating the database that contains the locations of all CDF files. The GridFtp servers are the machines that effectively transfer the files from FNAL to the CNAF storage system. We have two dedicated GridFTP servers with a 10GE connection to the CNAF Tier 1 10GE network backbone and connected via the Storage Area Network (SAN) directly to the CDF GPFS file system. This allows a plain method for transferring data to FNAL from CNAF through a single point. The transferred data are stored on a dedicated GPFS cluster file system. At present we have 8 GPFS servers to access a single GPFS file system of 420 TB. On this file system two different directories were created. The first directory (cache) is the actual SAM disk cache that is used during network transfer. When a dataset is copied to the disk cache and the databases are updated, it is moved to a second directory (durable) from which files are regularly and automatically migrated to tape. Two dedicated HSM servers are used for moving data from tape to disk and vice versa using 2 Oracle StorageTek T10000C tape drives (with a capacity of 5 TB each tape cartridge) through the Tape Area Network (TAN) The whole Storage system layout is represented in Figure 3.

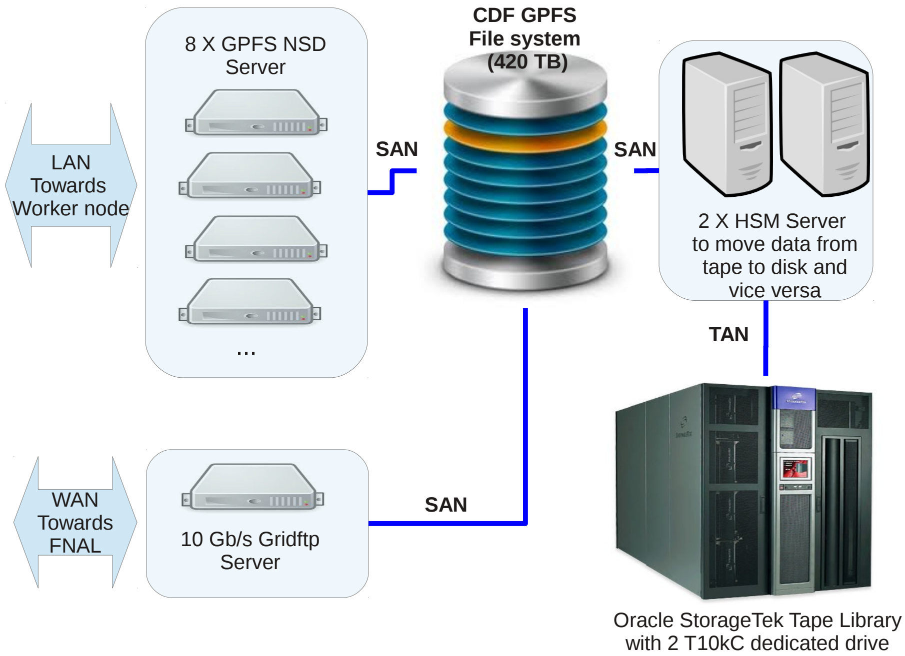

Figure 3. Storage Layout for FNAL-CNAF CDF data transfer

We performed multiple data transfer test to tune the GridFtp server and SAM station parameters. Results for the best configuration are presented in Figure 4. The required bandwidth of 5Gbit/s can be exploited wihen $\sim$ 50-80 files are being copying in parallel. The data transfer rate has shown to be stable over time.

4. CDF data analysis in the long term future

CNAF already offers a set of services to analyse CDF data. Data can be accessed via SAM and stored on a dedicated cache. Users can submit their analysis jobs to LCG via a dedicated portal, Eurogrid [6]. CDF analysis code is accessible via AFS. All these services are replicas of CDF services at Fermilab, installed as virtual machines on SL5 and SL6 operating systems. In the long term future these virtual machines will have to be migrated to archival mode. Running CDF legacy code requires addressing several issues, like availability of suitable hardware resources, software maintenance and handling of computer and network security. CNAF proposes that services used to access CDF data be eventually migrated to a dynamic virtual infrastructure.

We plan to implement this infrastructure so that CDF services can be instantiated on-demand on pre- packaged virtual machines (VMs). These VMs run in a controlled environment, where in- and out-bound access to these services and connection to storage data is administratively controlled.

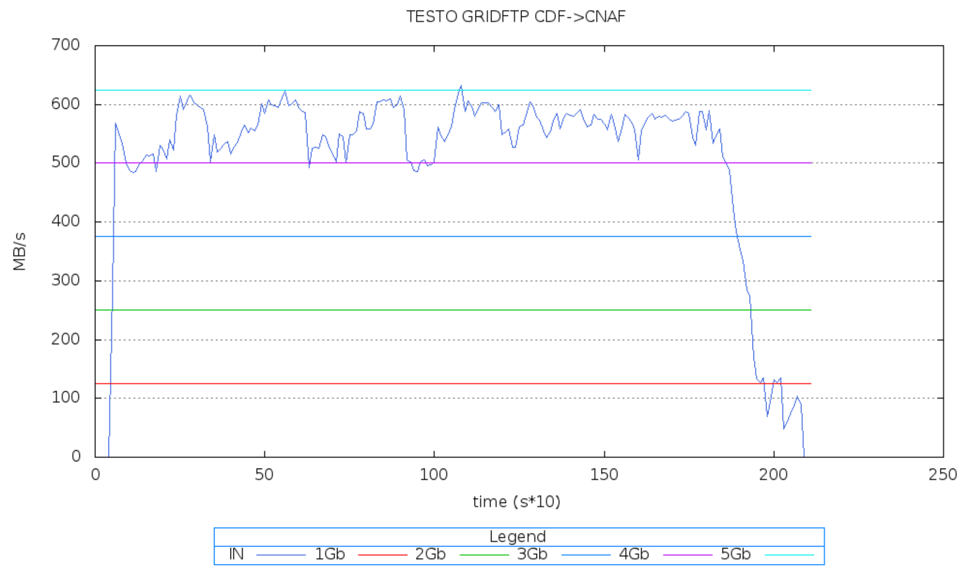

Figure 4. Results of data transfer test between FNAL and CNAF. With ~ 50-80 parallel copy processes the available bandwidth is fully exploited.

The set-up will be such that, when authorized access to CDF data is requested, instantiation of the virtual services will happen automatically and the VMs will be placed into a suitably isolated network infrastructure. We propose to realize this dynamic virtual infrastructure through OpenStack.

5. Conclusions

A project for the long term future preservation of CDF data is being implemented at INFN-CNAF computing center. The goal is to copy all CDF raw and user-level data files (4 PB) from FNAL to CNAF storage system, and provide users with tools to access and analyse the data. A copy mechanism able to copy the data at 5 Gb/s rate has been setup and successfully tested. For the analysis, an infrastructure based on virtualization is being developed, allowing the CDF analysis services to be instantiated on de-mand in a controlled environment.

5.1. Acknowledgments

We thank CDF Data Preservation task force and Fermilab Computing Sector for the support in the development of the project. In particular many thanks to Gene Oleynik, Dmitry Litvinse, Phil DeMar and Andrey Bobyshev.

References

[1] dCache group URL http://www.dcache.org/ 2
[2] Stonjek S, Baranovski A, Kreymer A, Lueking L, Ratnikov F *et al.* 2005 1052–1054 2
[3] Lucchesi D (CDF Collaboration) 2010 *J.Phys.Conf.Ser.* **219** 062017 2
[4] Sfiligoi I, Wurthwein F, Andrews W, Dost J, MacNeill I *et al.* 2011 *J.Phys.Conf.Ser.* **331** 072031 2
[5] Ricci P P, Bonacorsi D, Cavalli A, Dell'Agnello L, Gregori D *et al.* 2012 *J.Phys.Conf.Ser.* **396** 042051 3
[6] Amerio S, Benjamin D, Dost J, Compostella G, Lucchesi D *et al.* 2012 *J.Phys.Conf.Ser.* **396** 032001 4