INFN-CNAF CDF long term data preservation project

- November 29, 2013 -

S. Amerio

(University of Padova, INFN)

Tevatron data preservation

INFN-CNAF CDF LTDP project is developed in collaboration with Tevatron RUN II data preservation project

At Fermilab, data preservation project funded by DOE involving CDF and D0 experiments and and the Computing Sector.

Goal:
- Maintain full analysis capability
- Seek common solutions between experiments where possible
- Until 2020 (SL6 support) *at minimum*

Work ongoing on dfferent areas
- Bit preservation: migration to new tapes and new data access system
- Software preservation: CDF legacy software release (SL6) in preparation
- Job submission: opportunistic usage of Fermilab resources using virtual machines
- Documentation: new web-page, documents archived in Inspire

CDF data preservation in Italy: motivations

Goal: preserve a complete copy of CDF data and MC samples at CNAF + services (access, data analysis capabilities)

INFN involvement in long term CDF data preservation is important for different reasons:

1) INFN strongly contributed to the success of CDF; we need to ensure INFN maintains access to data many years from now, beyond CDF collaboration → A mirror archival in Europe is a necessary safety measure.

2) Direct participation with a real case to the problem of data preservation, which is of great interest → CDF preservation system at CNAF can serve as a prototype for future experiments now supported by INFN.

3) Opportunity for CNAF to take a significant role in the long term preservation of data.

INFN-CNAF LTDP project main areas

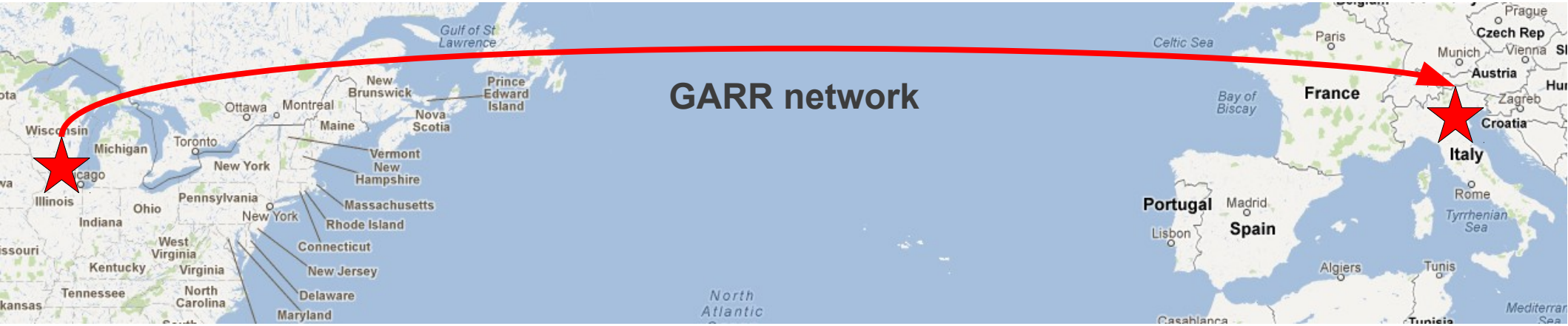
Bit preservation

- Copy CDF data and MC samples at CNAF (4 PB)
- Regular checks of data access and data integrity

Analysis capabilities preservation

- Preserve data access
- Preserve CDF reconstruction and analysis software
- Give users resources to run CDF analysis (authentication, disk space, CPU)
- Documentation

Bit preservation: implementation



The copy will be done via a dedicated link on the GARR network (5 Gb/s)

It will be splitted in two years
- end 2013 - early 2014 → All data and MC user level ntuples (2.1 PB)
- mid 2014 → All raw data (1.9 PB) + DBs

Data integrity checks:

- During the copy: compare CRC after copy with value stored in the DB
- After the copy: validation jobs run on random files

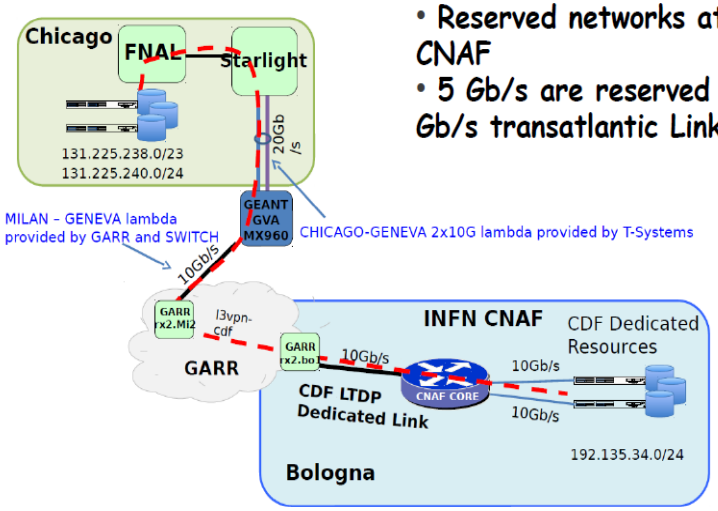
Resources needed for the copy

- **Tape**: 4PB (data raw and ntuples, MC ntuples only)
  - Tape to be procured (by the end of the year?)
- **Disk** to be used as cache for the copy: 100 TB
  - Already available, we will use 2013 CDF resources
- **2 T10K drives** dedicated to the copy*
  - Already procured
- **1 10 Gb/s gridftp and 2 tsm-hsm servers**
  - Installed and tested
- **One server** to store CDF DB
- Oracle licence
  - To be procured

** Used full time by CDF only for a limited amount of time; then will be available for all INFN experiments supported by CNAF.*

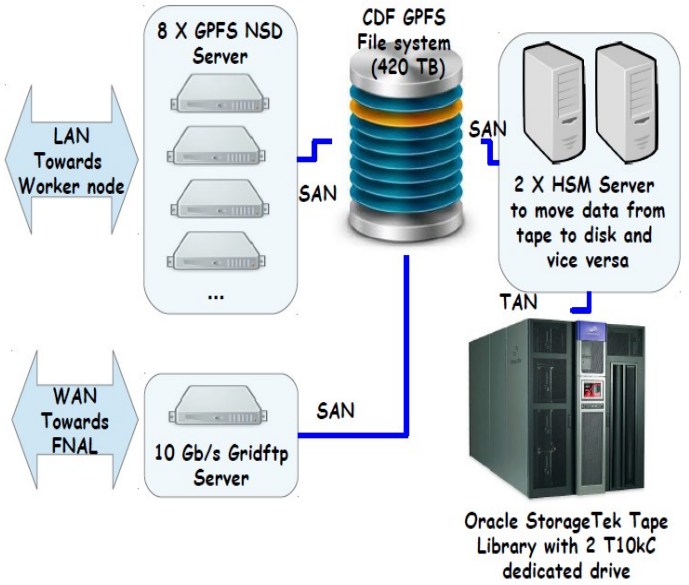
Network and data storage layouts

Network Layout for FNAL-CNAF CDF Data Copy



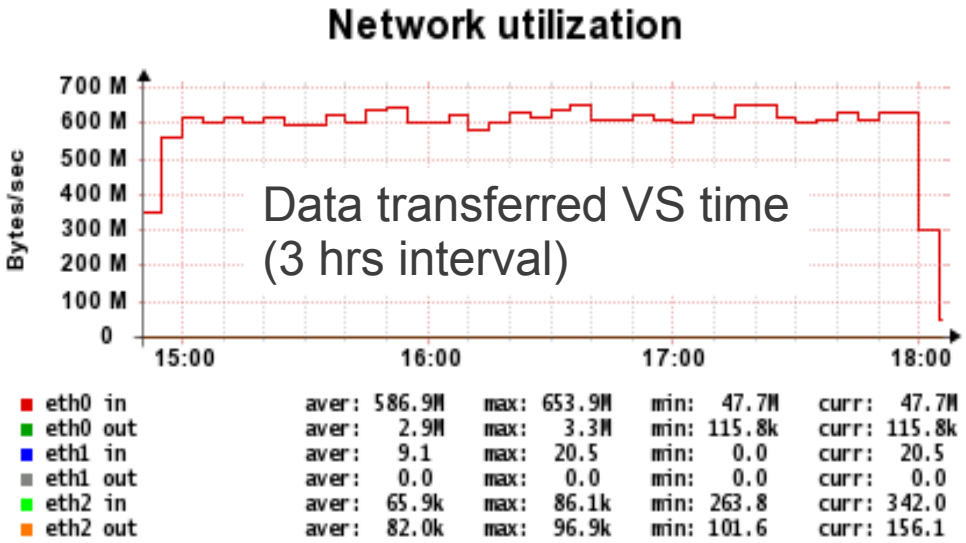
- Reserved networks at FNAL and CNAF
- 5 Gb/s are reserved to CDF on 20 Gb/s transatlantic Link

Storage Layout for FNAL-CNAF CDF Data Copy



Data are copied from FNAL via gridftp.
GEMSS HSM moves data to/from tape.
Retrieval from tape using standard CDF commands.

GEMSS is disk-tape managment system developed at CNAF used by all LHC and non-LHC experiments in production at CNAF for all storage classes



Optimization of the FNAL-CNAF network setup ✓

Optimization of the data copy scripts. ✓

Tests on real CDF datasets ✓

With ~ 50-80 parallel copy processes we exploit at the best the available bandwidth.
Data transfer rate stable over time.

Analysis preservation: implementation

Resources needed to access and analyse the data in the long term future

- Machine to **access data**
- **Disk cache** where requested data can be temporarily stored for analysis
- **Users areas** (on demand)
- **Job submission portal** to submit CDF jobs on CNAF resources (opportunistic usage of CNAF resources)
- **CDF code volume**

Plan

- All these resources are *already available at CNAF*; the long term data access and analysis framework will use as much as possible of the current system.
- To run CDF legacy code we plan to use a dynamic virtual infrastructure through the INFN-developed WNoDeS framework

- *Upgrade CDF machines to SL5/SL6 and all the services to the latest versions of the code.* 
- *Adapt the job submission portal to handle jobs on both SL5 and SL6. Ongoing...*
- *Test analysis framework on an existing analysis.*

INFN-CNAF CDF LTPD project timeline

Maggio 2012	Prima presentazione in GR1
Sett. 2012	Approvazione progetto suddiviso in 2 anni. Fondi primo anno sub-iudice al contratto per il responsabile progetto lato CDF.
Sett 2012-Maggio 2013	Ottimizzazione setup per il trasferimento FNAL-CNAF
Giugno 2013	Sblocco fondi sub-iudice. Inizio procedura acquisto tape.
Sett. 2013	Approvazione fondi secondo anno, sub-iudice alla copia di una prima parte significativa dei dati
Sett-Nov.2013	Test di trasferimento FNAL-CNAF; ottimizzazione script di copia.
Febbraio 2014	Richiesta sblocco fondi secondo anno.
Giugno 2014	Fine copia prima parte dei dati.
Settembre 2014	Test framework analisi al CNAF
Dicembre 2014	Fine copia seconda parte dei dati. Finalizzazione framework analisi dati e test con una vera analisi.

CDF-Italy
S.Amerio (Coordinator), G.Punzi, L.Ristori

CNAF
L.dell'Agnello, D.De Girolamo, D.Gregori, M.Pezzi, A.Prosperini,P.Ricci, D.Salomoni, F.Rosso, S.Zani

GARR
L.Chiarelli

CDF-Fnal

B. Jayatilaka, C.Vellidis

Fermilab computing sector

D.Litvinsev, G.Oleynik, P.Demar

Conclusions

INFN-CNAF CDF LTPD project approved in Sept 2012; full speed since May 2013.

First INFN approved project on long term data preservation.

Important use case for INFN and CNAF: it will serve as a prototype for other experiments, inside and outside HEP.

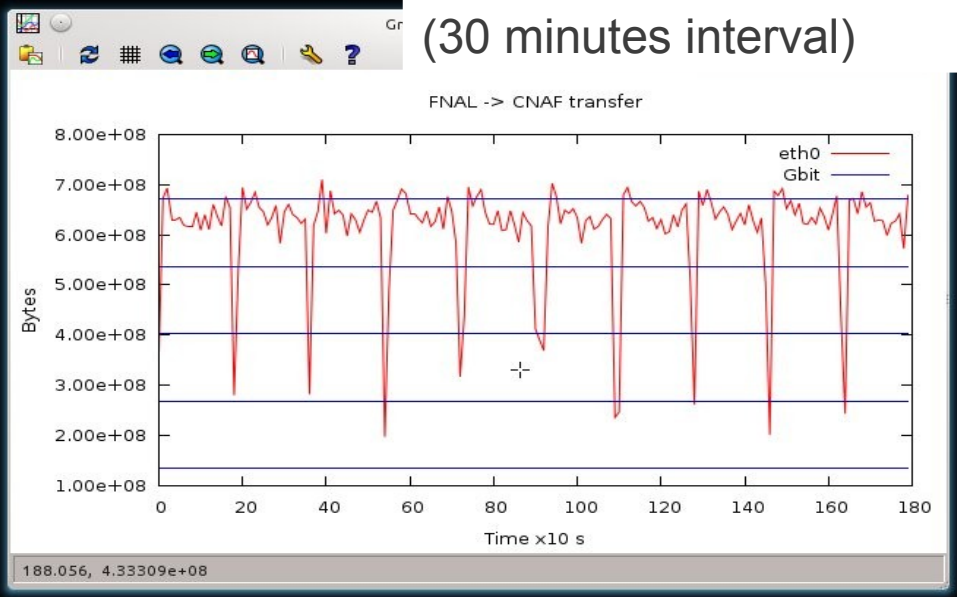
Opportunity for collaboration with other experiments, e.g. Babar and Aleph on virtualization techniques.

First step towards a common framework for long term data preservation of HEP data.

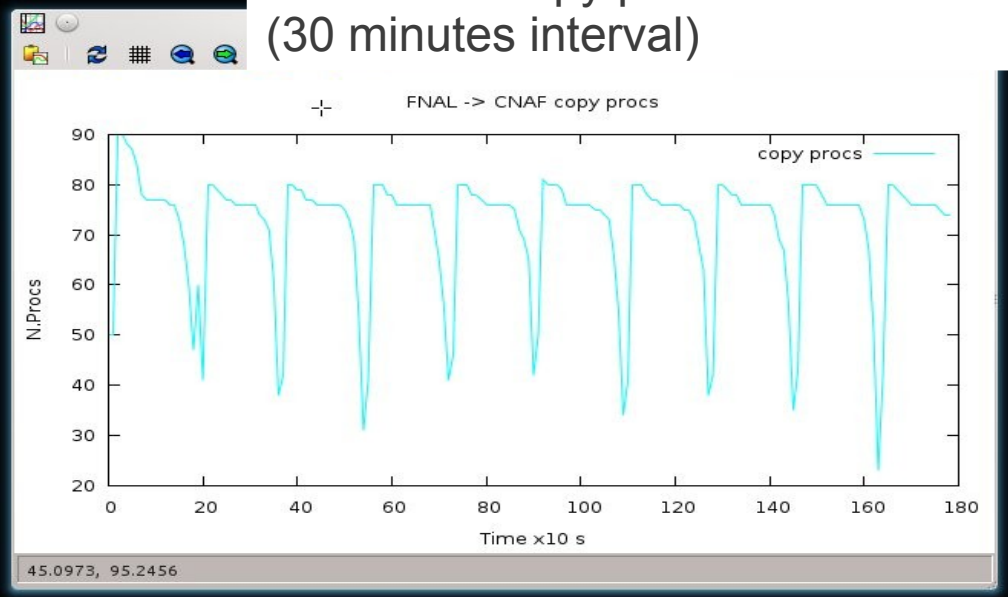
- Backup -

Network setup

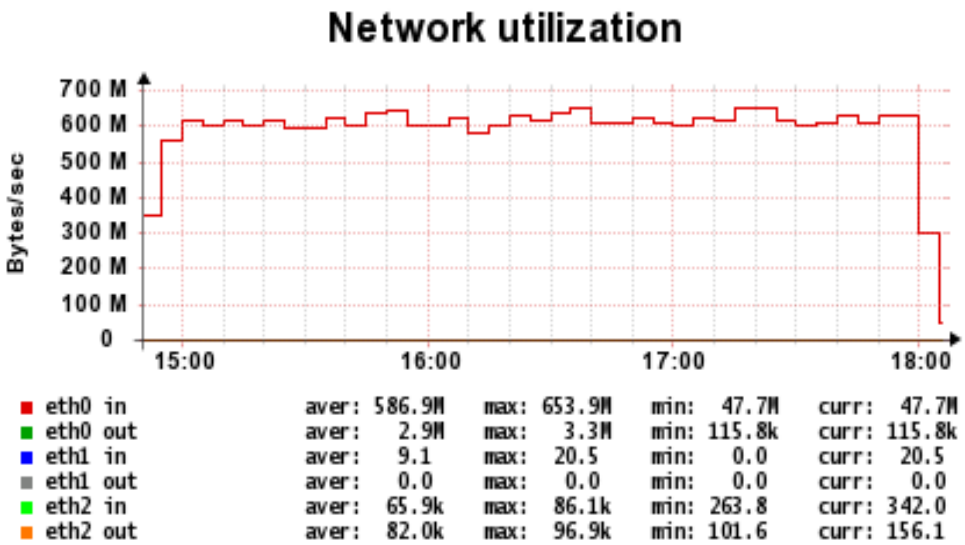
Data transferred VS time
(30 minutes interval)



Number of copy processes vs time
(30 minutes interval)



Data transferred VS time
(3 hrs interval)

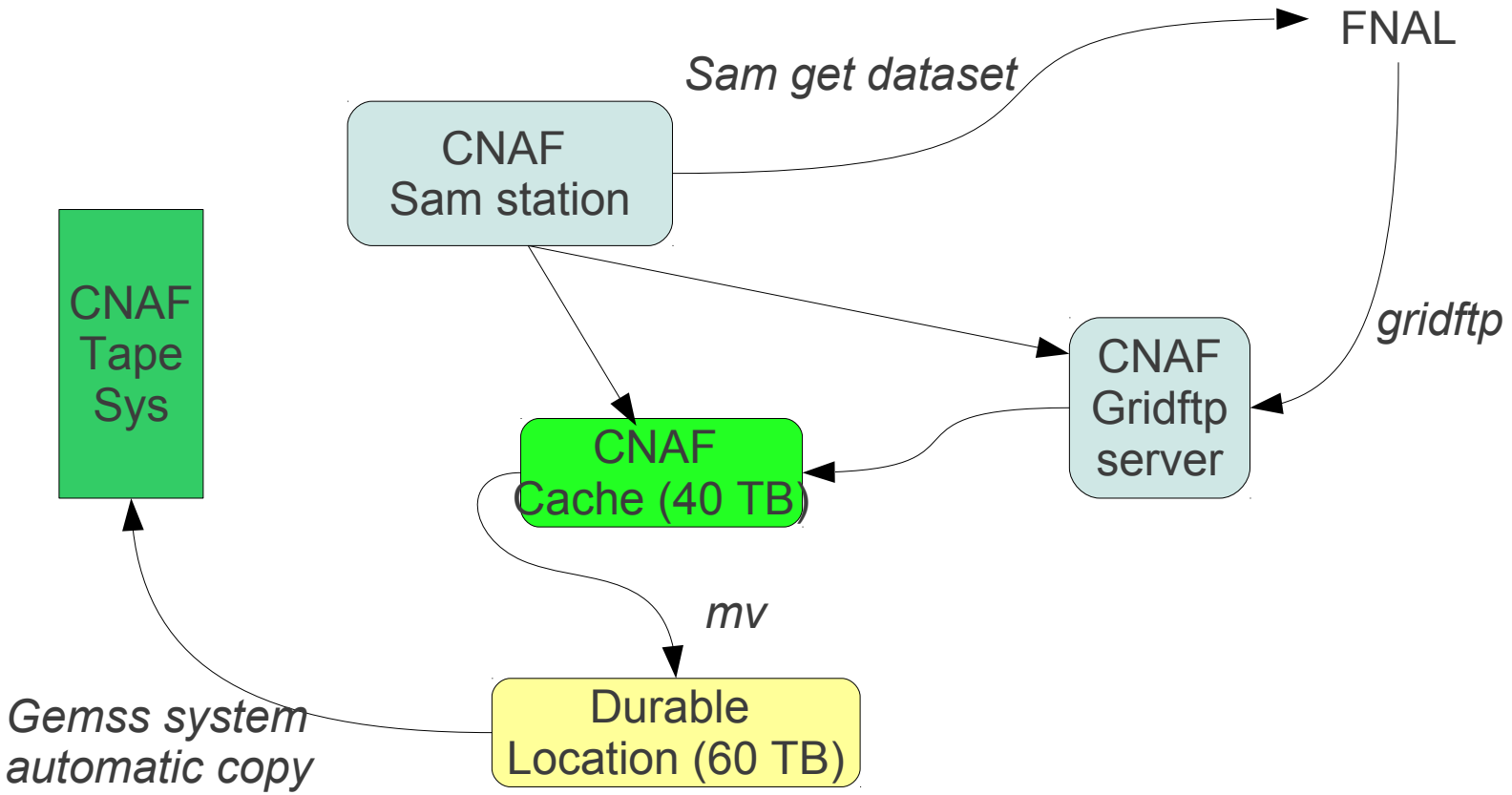


With ~ 50-80 parallel copy processes we exploit at the best the available bandwidth.

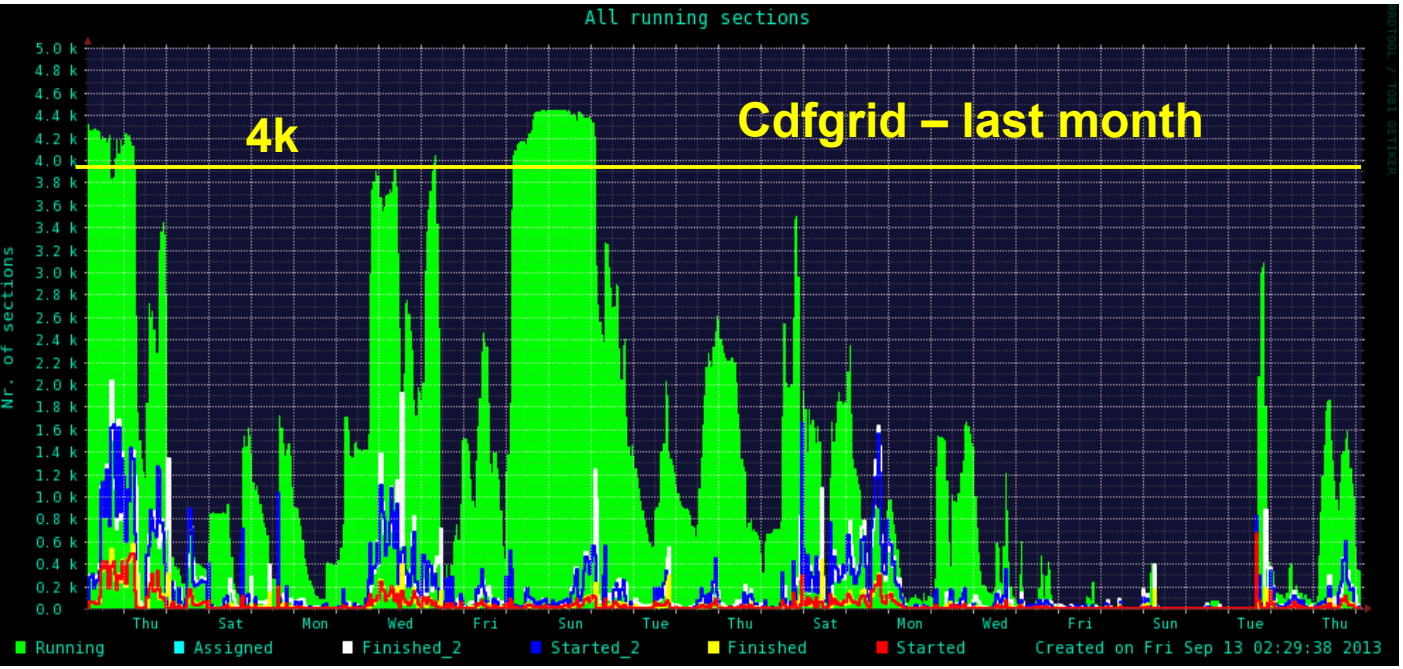
Data transfer rate stable over time.

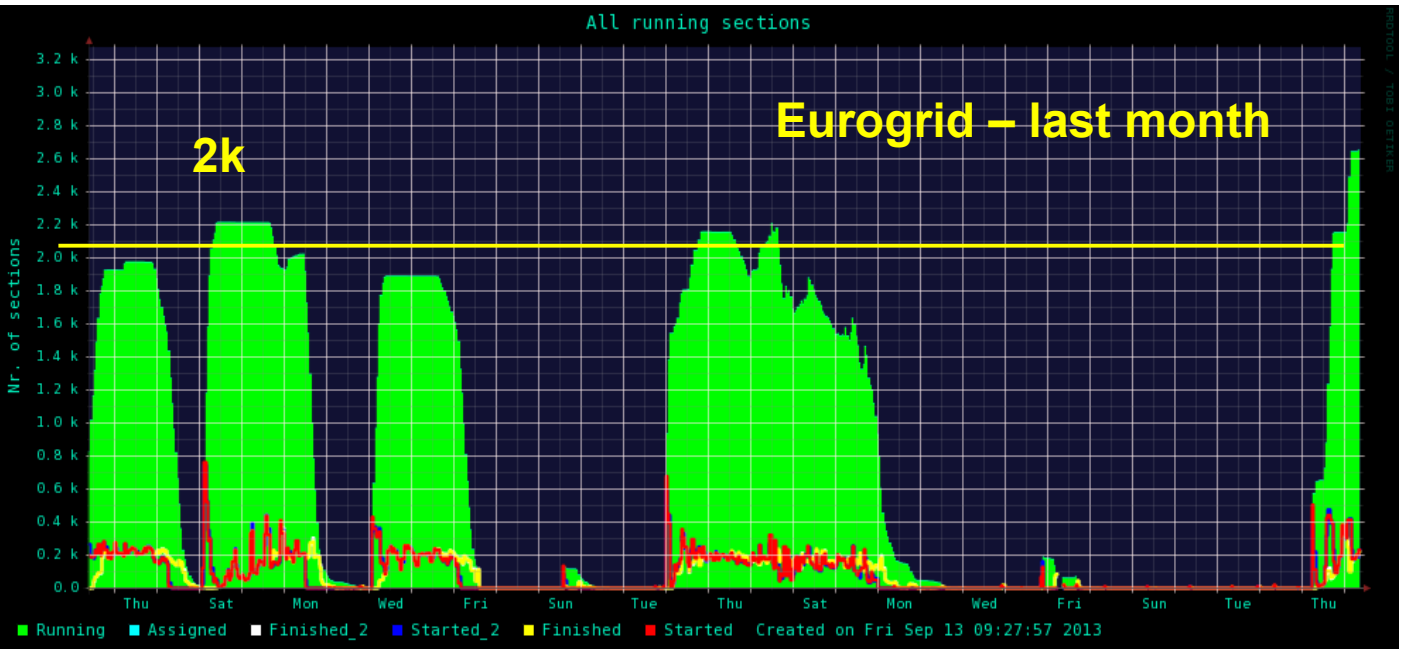
SAM station code optimization

- Sam code has been modified to do the transfer via the CNAF gridftp server (third-party transfer)
- A set of scripts has been optimized to drive the copy and transfer the files to tape.
- A small dataset has been transferred via the dedicated link using the Sam commands, uploaded to tape and retrieved from tape to disk.



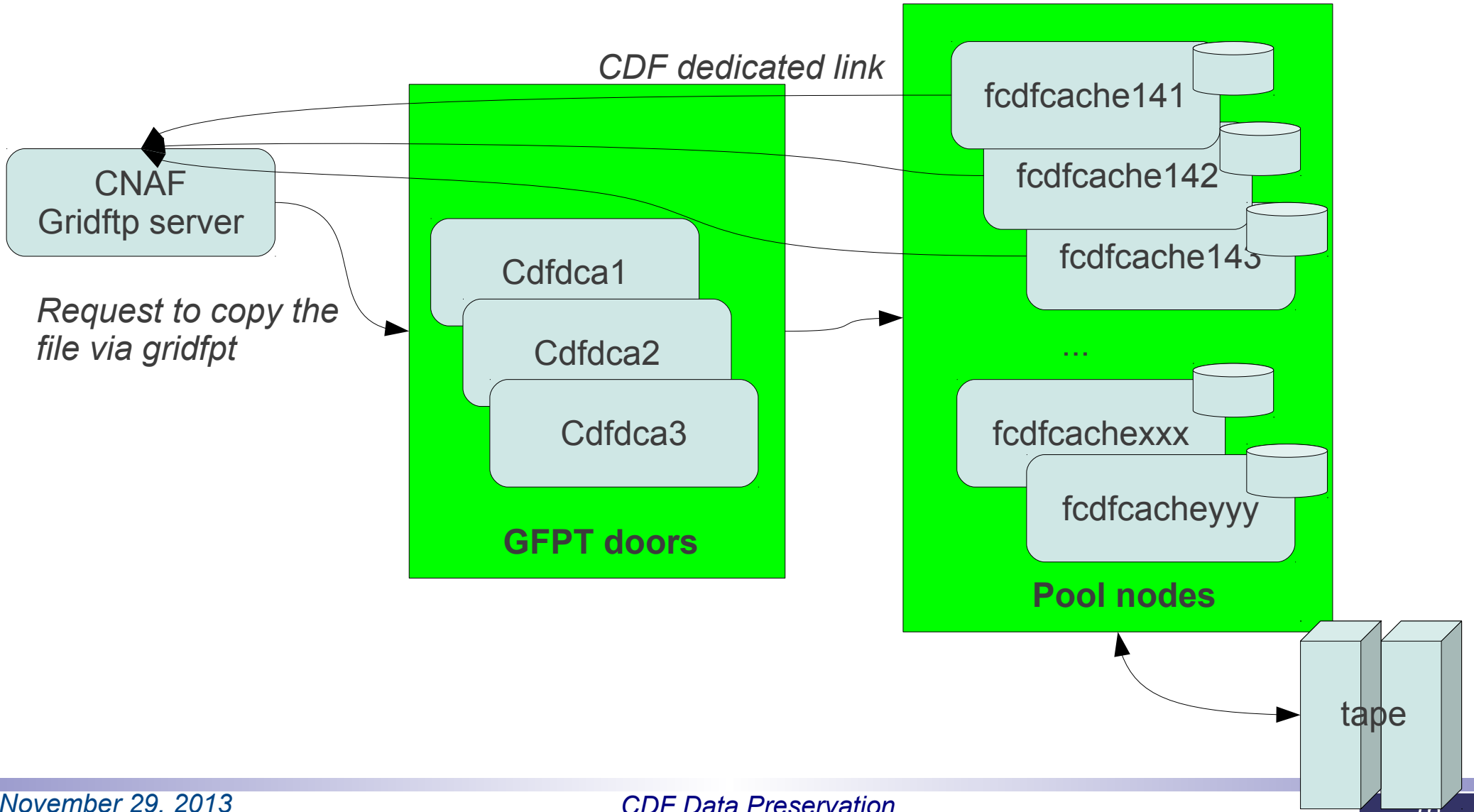
CDF computing resources usage: current status





- 23 papers submitted/accepted in 2013 to date
- Still a lot of activity in our computing farms

Network setup



Spese 2013 e richieste 2014

2013

FONDI ASSEGNATI: 89 KEURO

RIPARTIZIONE:
- utilizzo dei 2 tape drive 2013-2014: ***15 keuro***
- servers hsm : ***8 keuro***
- Tape: ***66 keuro***
  - Al costo di 52 euro/TB, potremo acquistare 1.3 PB

2014

FONDI RICHIESTI: 99 KEURO

RIPARTIZIONE
- Server per il DB: ***4 keuro***
- tape: ***95 keuro***
  - Al costo di 32 euro/TB (i tape da 5 TB possono essere riscritti con 8 TB di dati), potremo acquistare 2.9 PB

Stato acquisti 2013

I due tape drive per CDF sono già stati acquistati dal CNAF.

Per la copia si utilizzano 1 server 10 Gb/s per il trasferimento da FNAL (già presente al CNAF) e due server per il trasferimento su tape (acquistati per CDF).

Per il tape è stata fatta una richiesta d'ordine, presentata alla GE INFN il 13/09.

Data to be copied at CNAF

Raw data only + all ntuples (NO MC raw) → **4.0 PB**

The copy will be splitted in two years
- 2013 → All data and MC user level ntuples (2.1 PB)
- 2014 → All raw data (1.9 PB)

Data group	Volume (TB)
MC (raw data)	1163
MC (ntuples)	624
Data (raw)	1857
Data (production)	3834
Data (ntuples)	1492
TOTAL	8970

+ data catalogue and run conditions DB (Oracle, 250 GB)

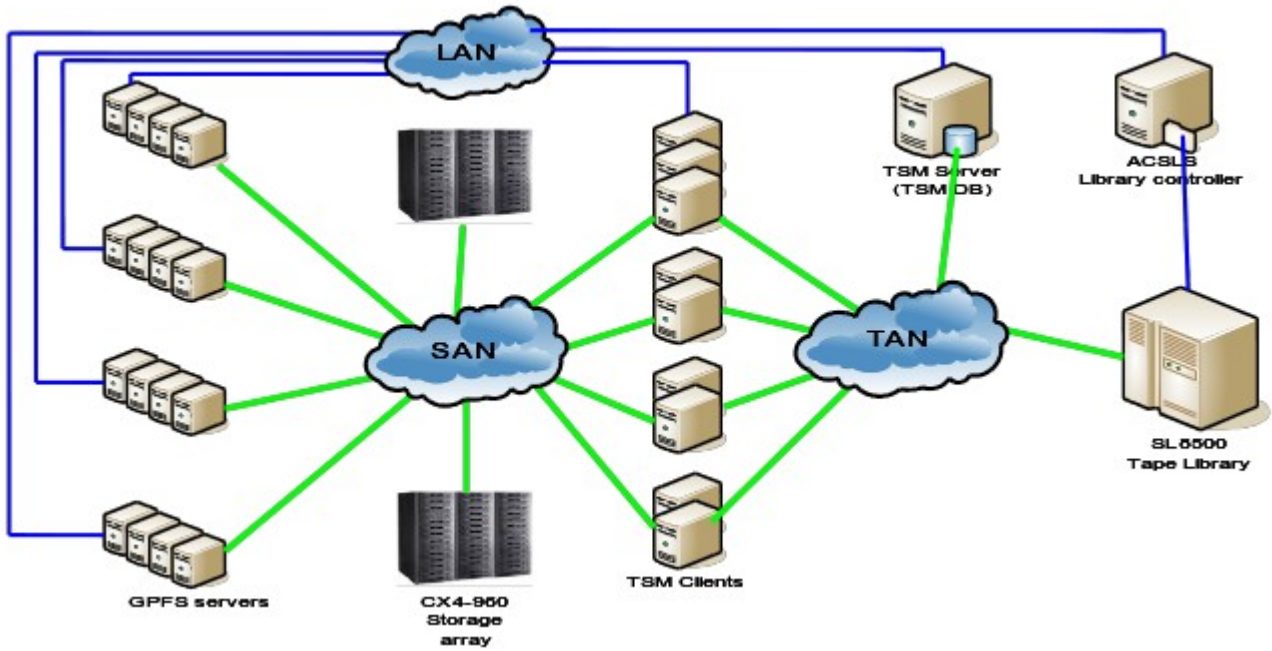
CNAF storage system in a nutshell

CNAF storage resources are organized into the GEMSS system
(Grid Enabled Mass Storage System)

- GEMSS is a full HSM (Hierarchical Storage Management) integration of GPFS, TSM and STORM developed at CNAF
- Combined GPFS and TSM specific features with StoRM to provide a transparent Grid-friendly HSM solution
- An Interface between GPFS and TSM has been implemented to minize mechanical operations on tape robotics (mount/dismount, search/rewind)
- StoRM has been extended to include the SRM methods required to manage the tapes.

GEMSS is used by all LHC and non-LHC experiments in production at CNAF for all storage classes

Site wide GEMSS Setup

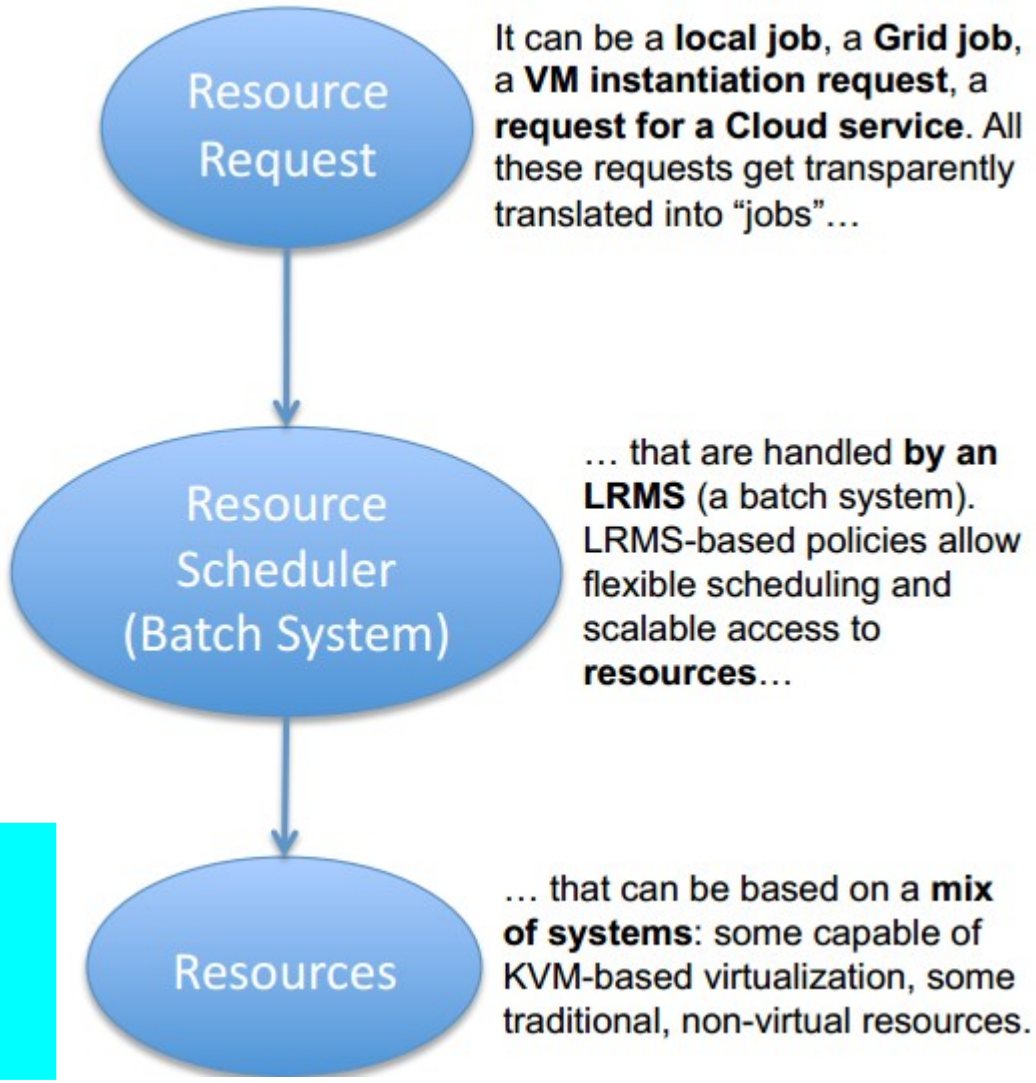


WNoDeS → Worker Nodes on Demand Service

In **production at several Italian centers, including the INFN Tier-1 since November 2009** (Currently managing about 2000 on demand Virtual machines there)

Dynamic virtual networks, new feature under development: dynamic instantiation of private VLANs and address assignement for VM isolation.

In the long term future: CDF services and analysis computing resources can be instantiated on demand on pre-packaged VMs in a controlled environment.



CDF data handling system @ FNAL