DP activities @ CDF

- October 16, 2013-

Silvia Amerio (U.Padova) and Bo Jayatilaka (Fnal)

for CDF DP task force

Goal: Level-4 preservation at both CDF and D0 (full analysis capability, including MC generation), ideally until 2020 (end of SL6 support).

Seek common solutions between experiments where possible

Division of effort between experiments and Computing Sector
- Experiments primarily responsible for software and documentation
- CS responsible for infrastructure (job submission, databases)

Bit preservation

- ~ 10 PB of data to be preserved
- Migration to new tape technology when available
- Currently migrating data from LT03/LT04 to T10K tapes
- ~ 9k/12k of tapes already migrated
- Data integrity checks
  - After each copy (Adler 32 checksum)
  - Periodic reads from each tape

Metadata stored in Databases

- CDF DBs are Oracle based; we will stay with Oracle in the long term future (moving to an open source is not feasible with the available manpower)
- Upgraded to Oracle 11. Another upgrade will probably be necessary before 2020.
- MySql DBs:
  - CDF notes moving to InSPIRE
  - People DB: freeze and archive.

Data access preservation: SAMWeb

CDF/D0 use SAM (Sequential Access via Metadata) to access data.
Current system worked us well to this point but it is difficult to maintain and it lacks security features → support for the existing SAM infrastructure will end in 2015.
A new SAM version (SAMWeb) has been developed for the Intensity Frontier experiments → long term future support

The CDF legacy software release will support SAMWeb (by the end of the year).

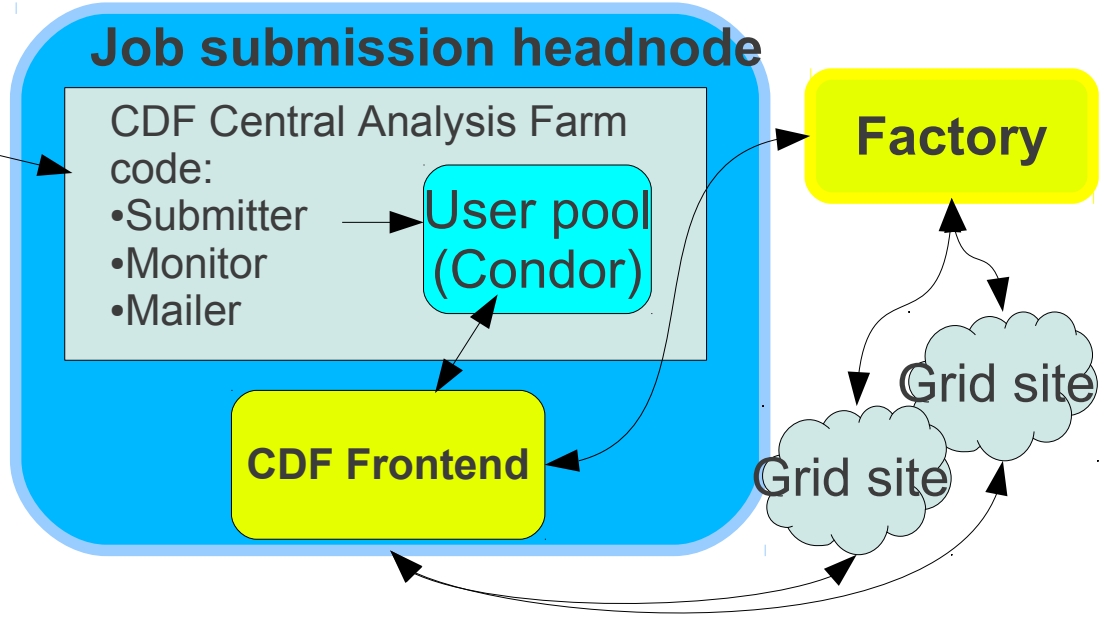
Current stable release runs on SL5 but includes components built on SL3/4 → it requires compatibility libraries

The offline team is currenlty testing a legacy release built entirely on SL5 (no compatibility libraries) and on SL6.

Simplified instructions for MC production (new instructions page, *unified MC submission script).*

Job submission

• CDF *Central Analysis Farm code (CAF)* provides the users with a uniform interface to resources on different Grid sites



• Three *portals* to access computing resources:
• CDFGrid → FNAL
• NamGrid → OSG
• Eurogrid → Tier1 @ CNAF and LCG

Based on *glideinWMS* workload management system (batch system = *Condor*)

In the long term future eventually move to opportunistic usage of Intensity Frontier resources for CDFGrid/NamGrid and of Tier1 @ CNAF resources for EuroGrid.

Recreate the necessary HW environment using virtualization.
Condor/grid infrastructure already in use will require little change.

Transparent for users.

CDF website

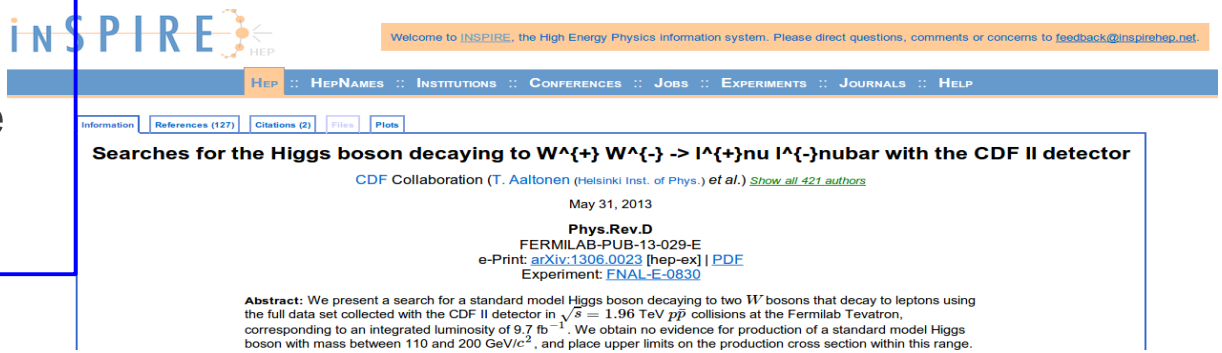
- We are checking and re-organizing all the information currently stored in the CDF webpage.

CDF Fast Navigator

	Computing				
HELP	Problem Reporting				
Infrastructure and Services	CAF / Batch CAF Home CDFGrid Monitor NamGrid Monitor EuroGrid Monitor	Code Management Code Browser CVS Browser cdfsoft2 Releases	Database Database Browser SAM DB Browser Luminosity Browser	Infrastructure Account Request ILP Cluster Disk Space Fermilab email lists	Data Handling DH Home CDF dCache CDF Enstore SAM @ CDF
Essential Processes	Production Datasets Production Periods Operations Tiki	Data Preservation SCD DP Project	Ntuples Stntuple TopEventModule	Monte Carlo Production MC Uploading cdfSim TrigSim	

CDF notes on InSPIRE

- Migration of CDF notes to Inspire ongoing



INFN long term data preservation project

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

Copy of the data

The copy will be splitted in two years

- End 2013 – begin 2014 → All data and MC user level ntuples (2.1 PB)
- Mid 2014 → All raw data (1.9 PB) + DBs

Long term future analysis framework

Based as much as possible on the current resources already available at CNAF (Eurogrid, SAM station, ...)

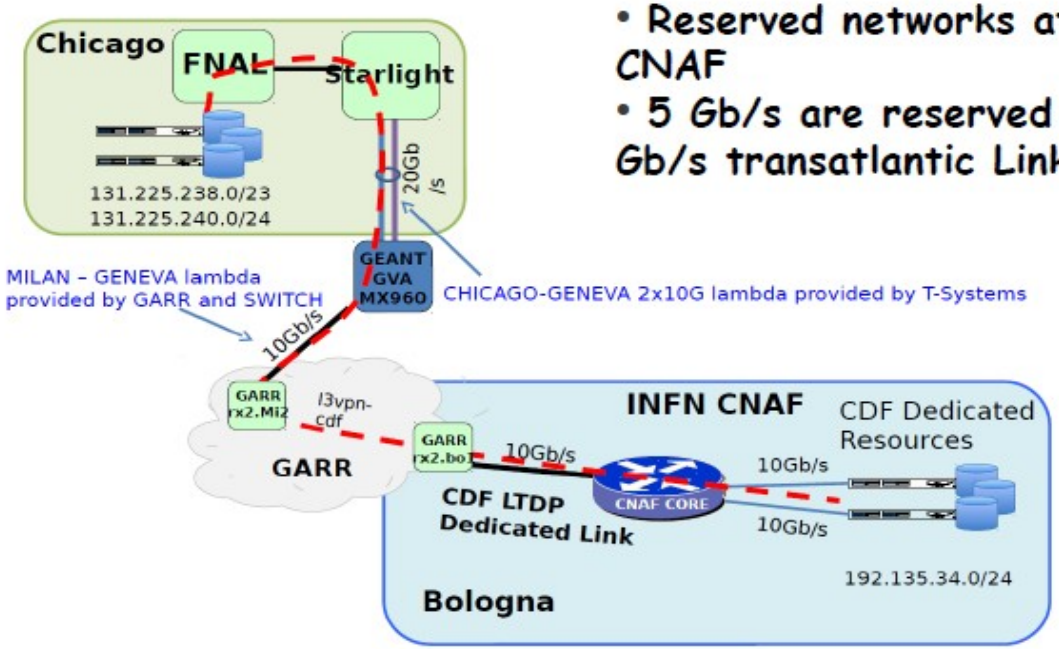
To run CDF legacy code we plan to use a dynamic virtual infrastructure through the INFN-developed WNoDeS framework.

We expect to move to archival mode in 2015.

The project has been developed and is now being implemented with the support of CDF DP task force and Fermilab CS.

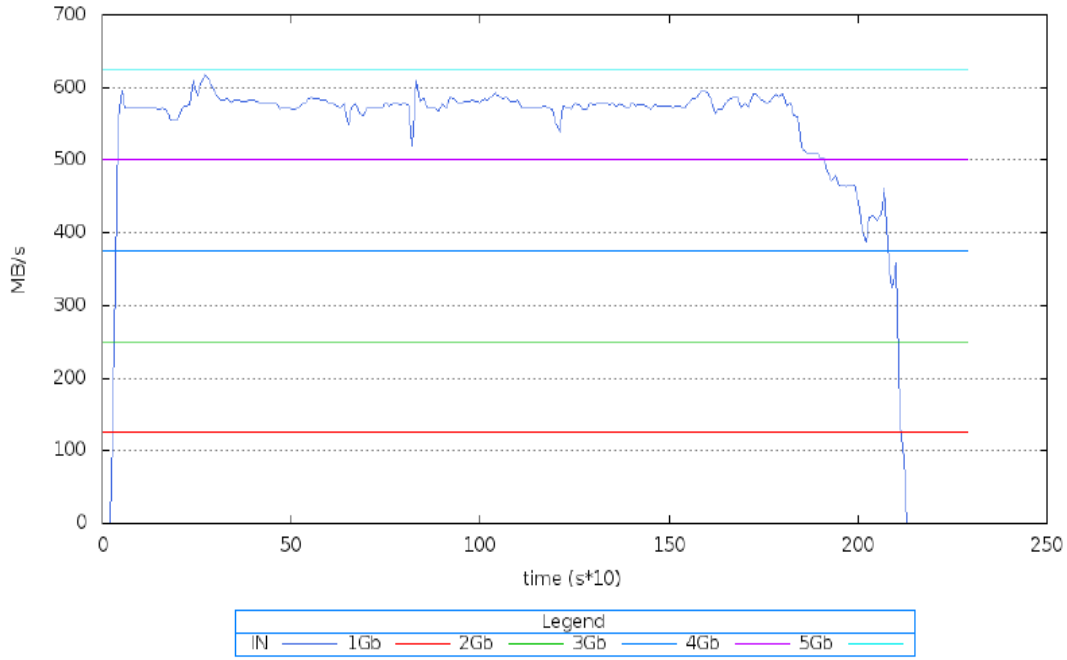
Good opportunity to share experience on DP issues between CNAF and FNAL and develop common solutions when possible.

Network layout



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

Data transfer rate FNAL-CNAF



Data transfer rate stable over time

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