Long Term Data Preservation (at INFN)

Luca dell'Agnello
INFN-CNAF
(thanks to S. Amerio, M. Pezzi and T. Boccali)

The Data Preservation problem (1)

- **Long-term Data Preservation** refers to the series of managed **activities** necessary to **ensure** continued **access** to digital materials for **as long as necessary**
  - also after the end of the life of the experiment…
- Some scientific areas, e.g. astrophysics, are well ahead in data preservation.
  - HEP experiments now focusing on this issue
  - On going at INFN for CDF (CNAF) and Aleph (Pisa)

INFN CNAF

The Data Preservation problem (2)

- **DPHEP:** Data Preservation in High Energy Physics
  Past experiments have already successful DP projects in place (e.g. Babar)
- All **LHC experiments** are devoting more efforts to data preservation

- Data preservation is one of the areas targeted in **Horizon2020**

A data preservation project can be divided into two main areas:

- *1 - Bit preservation : how preserve data*
- *2 - Analysis framework preservation : code preservation, virtualization*

INFN CNAF

Bit Preservation

- (At least) two copies of data on MSS in two different places
  - Well-defined (automatic) procedure to periodically check data integrity…
  - …and (in case) copy from the other site(s).
- Periodic migration of data from a generation of tapes to the next one…
  - Needs extra funding besides the start-up

INFN CNAF

Analysis framework preservation

- Development of the long term future analysis framework.
- Preserve data access
- Preserve reconstruction and analysis software
- Give users resources to run analysis (authentication, disk space, CPU)
- Documentation

INFN CNAF

Aleph (1)

- Experiment took data at the CERN e+e- collider LEP from 1989 to 2000.
- Data still valuable
  - still get request by Theoretical Physicists for additional checks / studies on ALEPH Data
- Preservation done at INFN-PISA
  - 150k files, avg size 200 MB (30 TB)
  - Split between real collected events and Monte Carlo simulated events
  - Processing times (analysis) on recent machines is such that 1 week is enough to process all the sample
- Current policy: any paper can use ALEPH data if among the author there is at least one former ALEPH Member

Aleph (2)

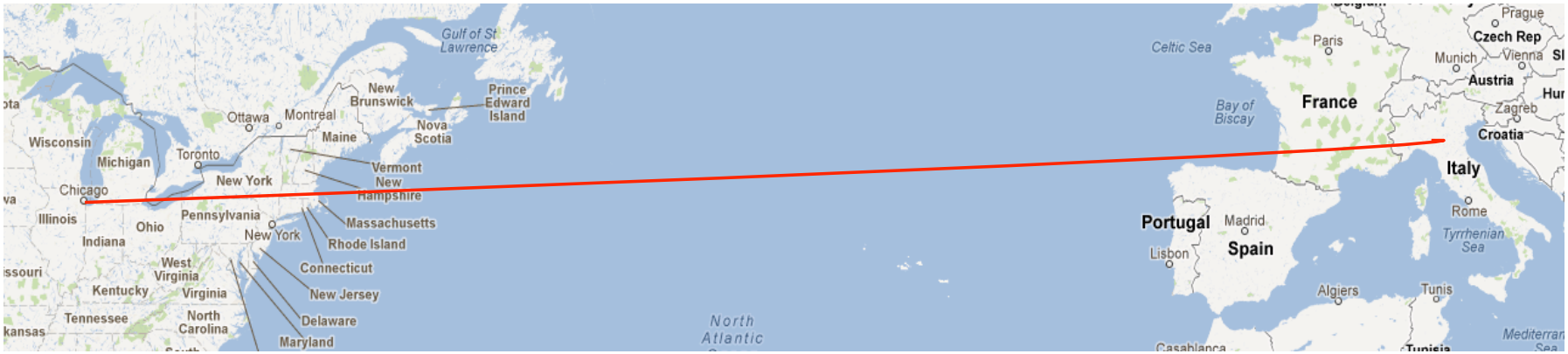
- Computing Environment via VM approach
  - Currently using uCERN-VM (SL4)
  - Provides batch ready VMs, interactove ready VMs, development ready VMs
- Data to be served via POSIX to the executables
  - Current approach (pre Eudat) was
    - Via WebDAV (Apache, Storm, …)
    - Seen by the VM as FUSE/DavFS2 mounted POSIX Filesyste
- Currently working on EUDAT

CNAF-CDF project

- Goal: preserve a complete copy of CDF data and MC samples at CNAF and services (access, data analysis capabilities)…
- …using "standard" tools and shared resources!
- First problem: copy all (required) data from FNAL to CNAF
- Implement bit preservation
- Implement analysis framework for LTDP
- Aleph approach (i.e. EUDAT) not viable

INFN CNAF

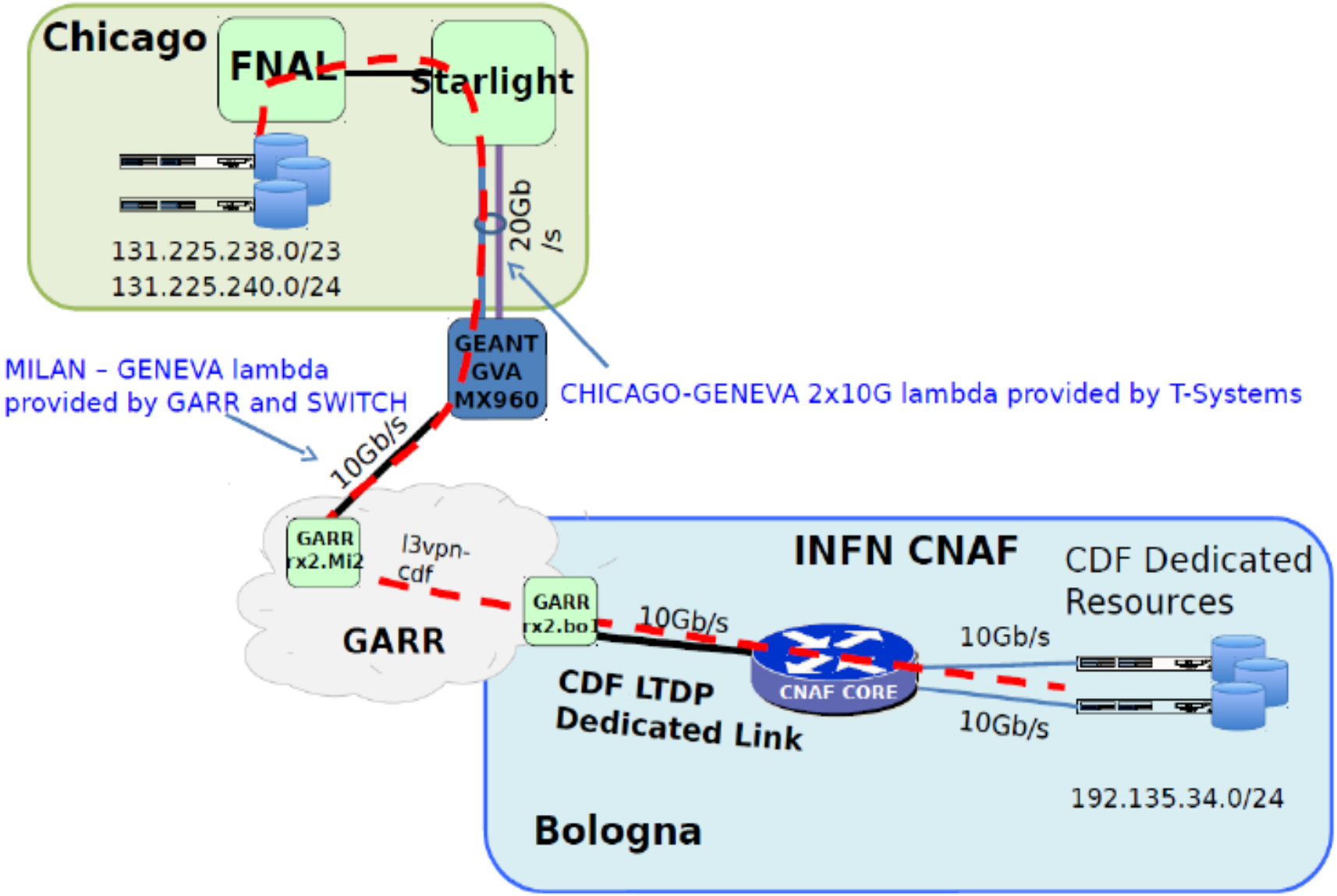
CNAF-CDF project: the copy (1)



- ~4 PB of data to be copied and preserved
  - All data and MC user level n-tuples (2.1 PB)
  - All raw data (1.9 PB)
  - Databases
- Dedicated 5 Gbps link FNAL-CNAF
  - Originally foreseen to complete the copy before Q2 2014
  - Delay in tape procurement ☹
  - Copy will be completed at the end of Q3 2014…

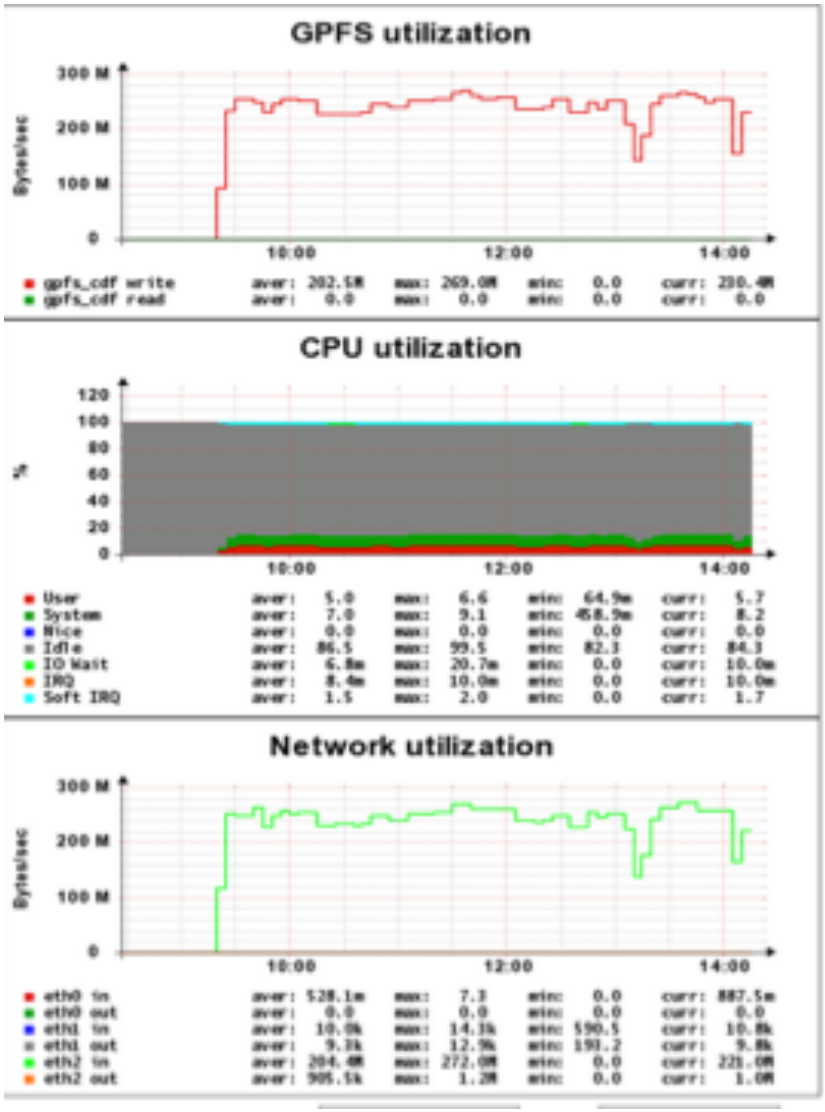
INFN CNAF

Data transfer: network layout



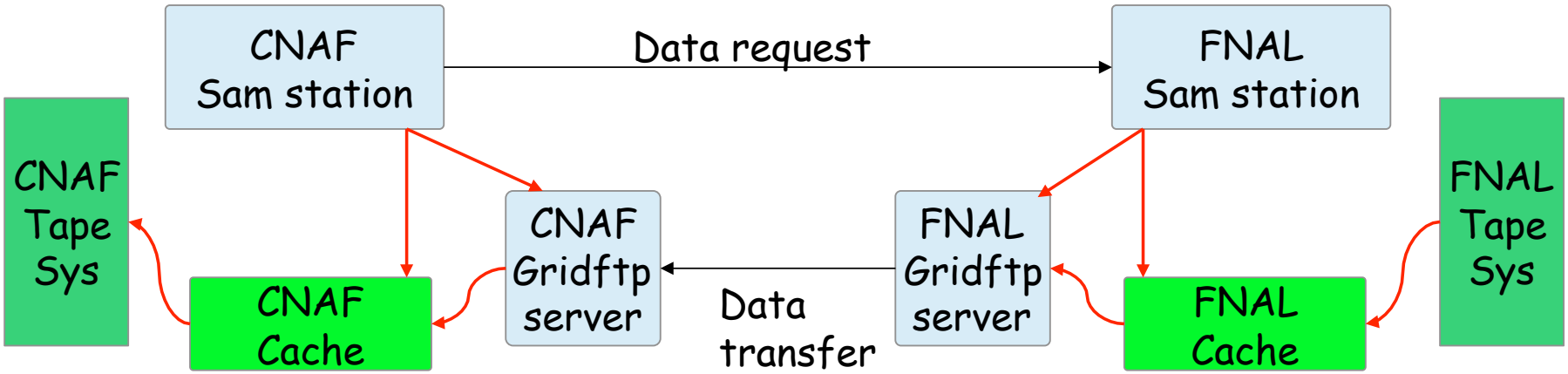
CNAF-CDF project: the copy (2)

- Some modifications needed for:
  - CDF copy system to use third party gridftp transfers
  - the SAM station to use our MSS
- Optimization of network configuration
  - Saturation of available network (5 Gb/s)
    - Limited to allow repack
  - Gridftp: 80 simultaneous processes each of which divided into 20 parallel streams
- Complete (re)use of tools and standard Tier1 infrastructure
  - Minimal FTE overhead
    - excepting the start-up!



CNAF-CDF project: the copy (3)

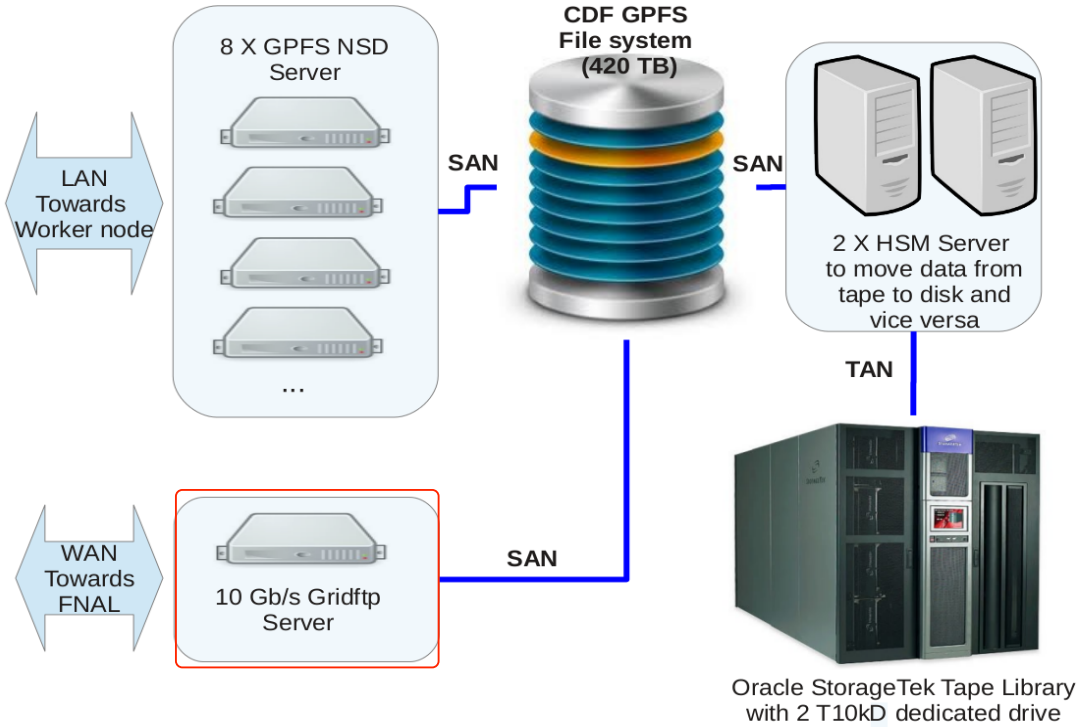
- CNAF requests data from FNAL that are staged at FNAL cache
  - The pre-staging done in parallel with the data transfer
- Data are copied via gridftp protocol (SAM performs checksum control)
- Once the data are in the CNAF cache, they are automatically migrated to tape



Data transfer archiving

1 - Data are copied from FNAL via gridftp.
2 - GEMSS moves data to tape
3 - Retrieval from tape using standard CDF commands (SAM station).

About 115 x 8,5 TB tapes (T10KD) written => ~ 1PByte



2 T10kD drive (130MB/s reading from the CDF GPFS Filesystem)

Actual DISK => TAPE (2 drives) bandwidth can reach 260MB/s

DISK => TAPE bandwidth will be improved with additional drives (2nd half 2014)

1 Single Point of Failure (single gridftp)

Cold spare machine avaliable (can be configured in a couple of hours)
Data transfer can "afford" a suspension in case of hardware failure

It's the same storage configuration that is used for other experiments

INFN CNAF

Present status and future development

Now we have already replicated at CNAF ≈ 1 PB of CDF data on tape

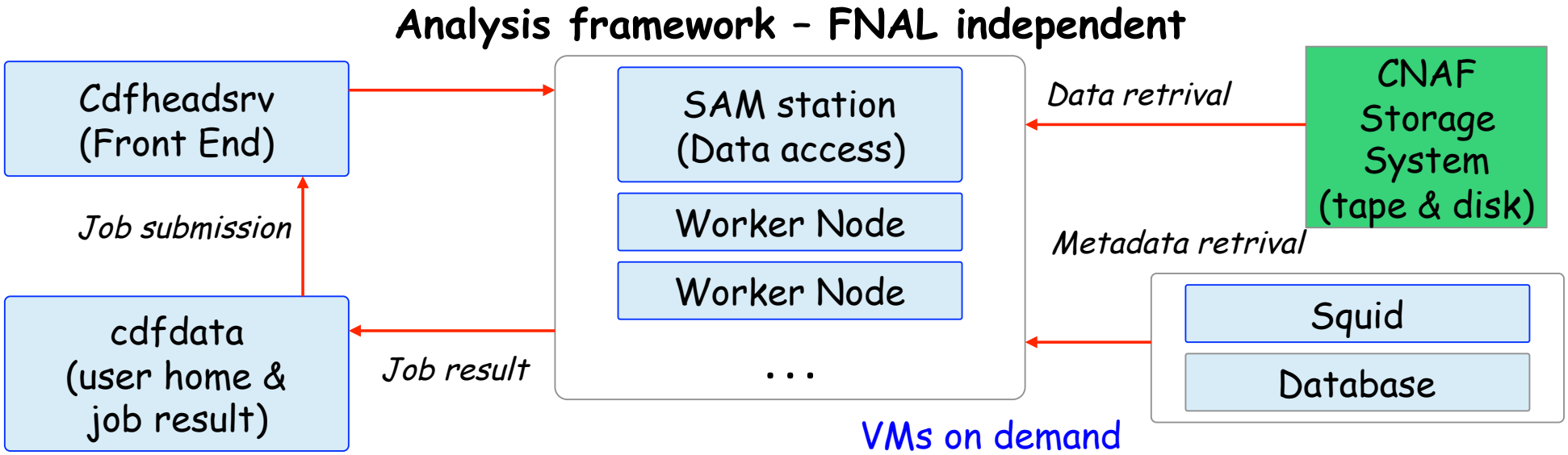
CDF data analysis in the long term future

To analyze CDF data stored at CNAF we need to implement:

- SAM station to retrieve data from tape: new version of SAM code in preparation at FNAL.
- CDF code volume, accessible via CVMFS
  Access to Batch facility
- Access to Oracle Databases (with local replicas at the INFN CNAF Tier1)
- Code preservation: CDF legacy software release (SL6) under test

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

Data analysis : future



This framework assumes limited use of CDF data in the long term future. Problems under discussion:

- Replication of Database

- Data access : QUESTION: How many years IBM will support GPFS on SL6?
  Possible solution could be NFS which provides greater compatibility with earlier version.

- Authentication : In the long future, access to the GRID resources will not be necessary.
  Possible solution could be job submission restricted to the local CNAF nodes.

INFN CNAF

Summary

- A data preservation project can be divided into two main areas:
  - Bit preservation : how preserve data
    - For CDF data transfer still on-going data (expected to be completed by the end of 2014)
  - Analysis framework preservation : code preservation, virtualization …
    - (obvious) Strategy: ause virtualization (Openstack?)
- Some issues still open (e.g. common policy for data access, "standard" way of preserving accessibility etc..)
  - Fundamental to use "standard" tools and frameworks
  - Probably some activity in H2020
- First experiences in HEP world (BABAR, Aleph, CDF) in the framework of DPHEP are the prototype for WLCG and other experiments