HTTP caches

Dr. Silvio Pardi

INFN-Napoli

Joint WLCG & HSF Workshop 2018

28 March 2018

INFN

Istituto Nazionale di Fisica Nucleare

SCOReS Project

Italian Acronym for: Study of a Caching system to optimize the usage of Opportunistic Resources and sites without pledged storage, for e-Science application(s) (SCOReS)

Project funded by GARR within a National call consisting in a 2Year fellowship.

- Davide Michelino - project fellowship
- Silvio Pardi – Project Tutor for INFN-Napoli
- Prof. Guido Russo

Consortium GARR

INFN
Istituto Nazionale di Fisica Nucleare

Cache Use-Cases

Goal of the activity is to setup and test an HTTP Caching system to integrate in Experiment Computing Model and for other applications- pilot experiment is Belle II.

Cache can affect performance in many scenarios:

- At site level: Cache increases Analysis performance for those job running on the same data-set.
- Cache can help all sites close to the one hosting cache
- Storage-Less Site Paradigms
- Cloud Storage
  - Multiple access to Cloud storage with limited bandwidth vs the clients
  - Limit the number of GET requests on Cloud Storage

INFN
Istituto Nazionale di Fisica Nucleare

Caching laboratory with DPM

- DPM 1.9 with Dome will allow investigation of operating WLCG storage as a cache
- Scenarios
  - Data origin a regional federation of associated sites
  - Data origin the global federation
- A volatile pool can be defined which calls out to a stager on a miss
  - Caching logic implemented in a pluggable way
  - Hybrid cache/conventional setup
- Questions to investigate
  - Cache management logic
  - Different client strategies on miss
    - blocking read, async read, redirection to origin
  - Authentication solutions
  - Workflow adaptation for locality

CHEP 2016

We are trying to answer at these questions

Concept of Volatile Pool

A Pool in DPM is a collection of File Systems managed as a single storage area.

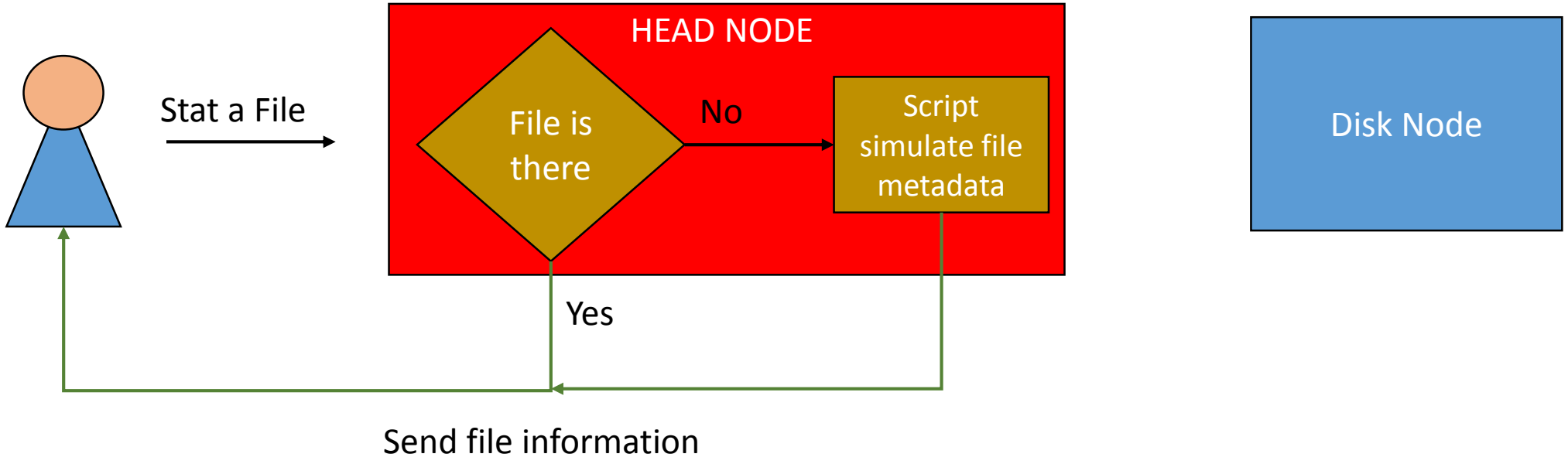
A **Volatile Pool** is a special pool that can pull files from external sources.

When an **User stat a file** in a Volatile pool, the **DPM head node runs a script** and then reply with file information or with a standard output if the file is not already there.

When an **User get a file** from the Volatile pool, the **Disk Node** providing the file system of the pool, send the file to the client if ready, otherwise a script is locally run in order to retrieve the file from some external source.

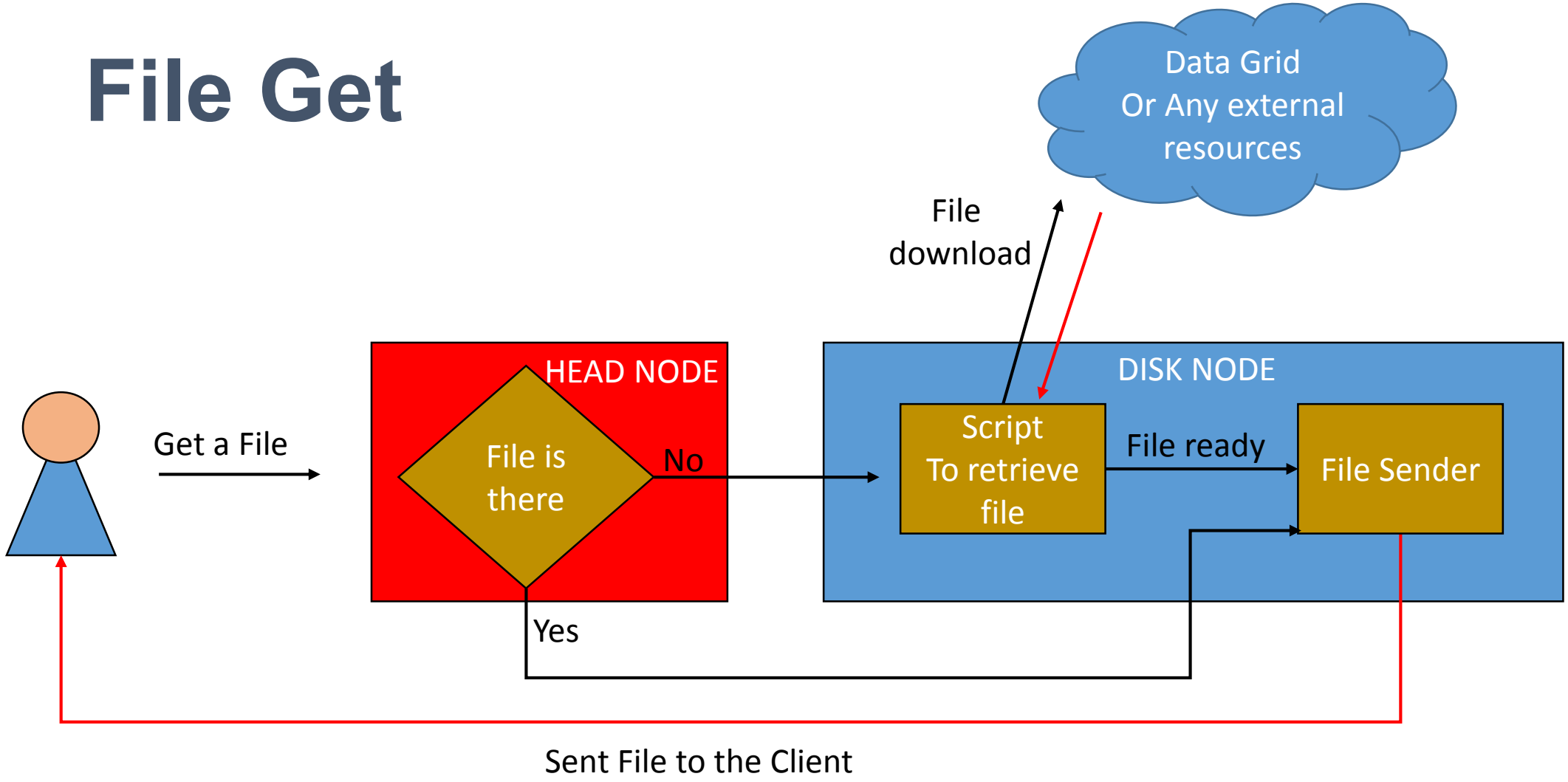
INFN
Istituto Nazionale di Fisica Nucleare

File Stat



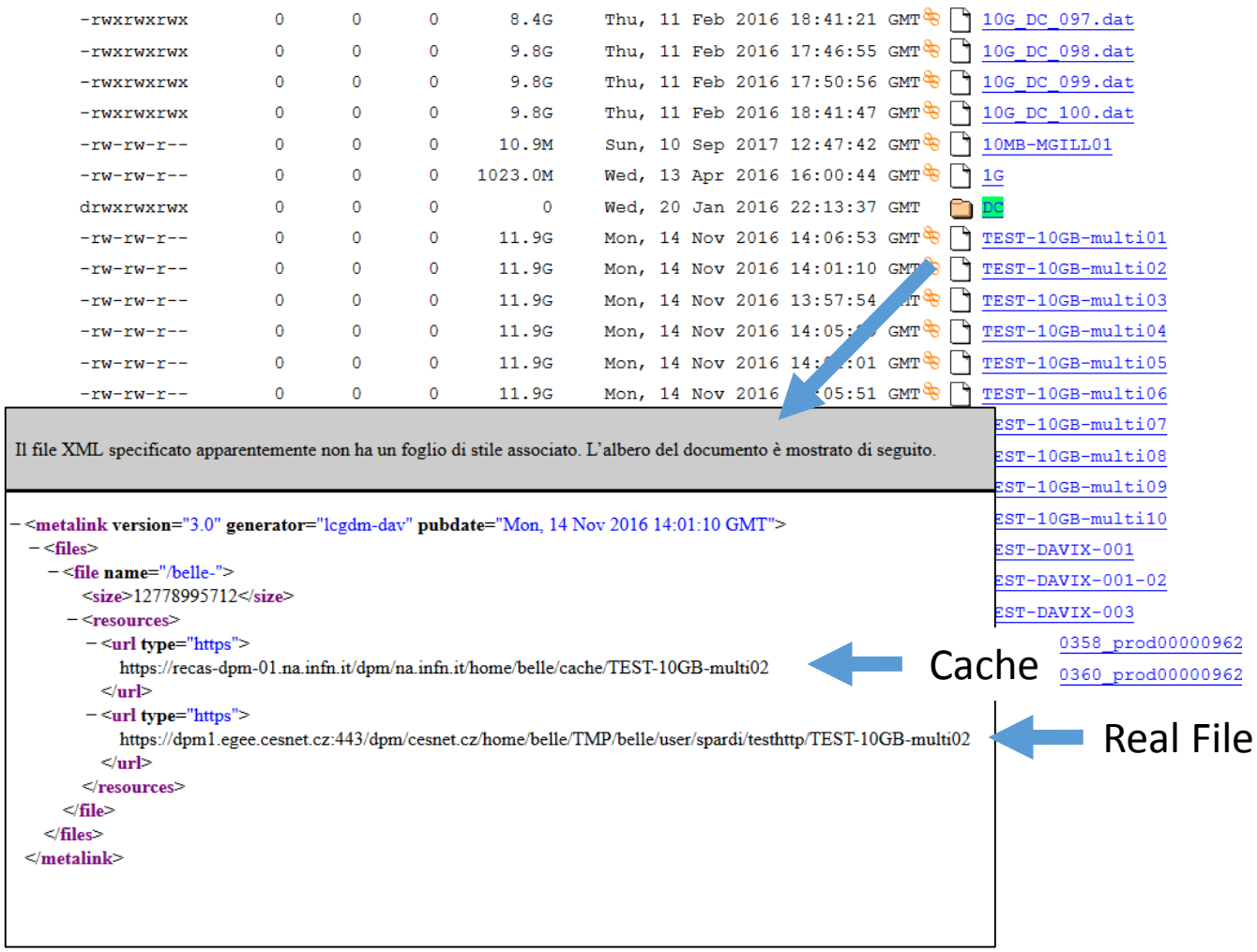
INFN
Istituto Nazionale di Fisica Nucleare

File Get



INFN
Istituto Nazionale di Fisica Nucleare

Dynafed + Volatile Pool



What happen if we aggregate a Webdav endpoint with a DPM Volatile Pool?

When Dynafed stat files inside the real webdav endpoint, it receive always a reply even from the Volatile Pool.

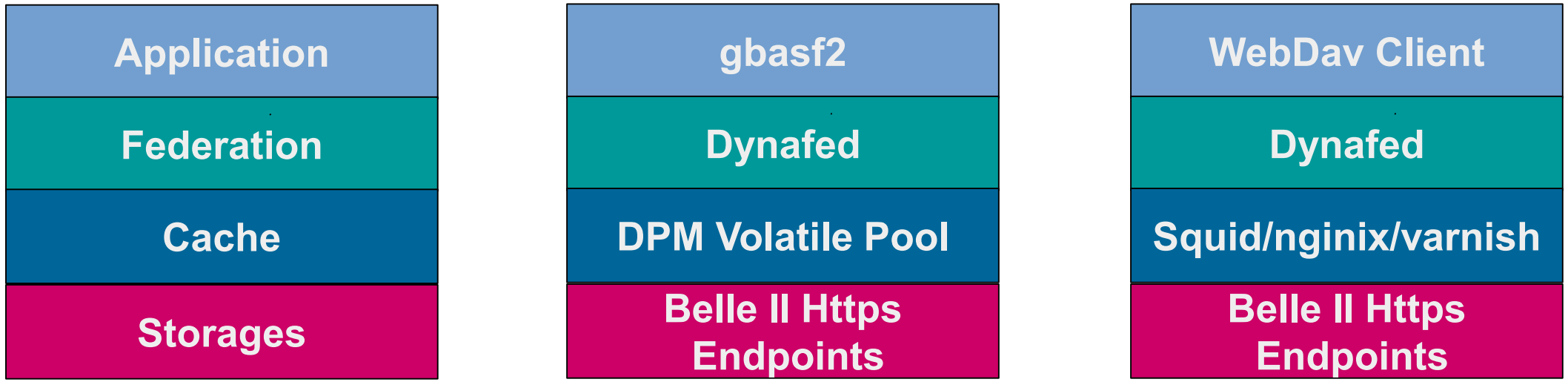
So that the metalink representing a file in Dynafed, included always the real URL and the corresponding virtual copy in the cache (even if the latter does not exist yet)

Moreover thanks to the GeoPlugin, Dynafed prioritize the cache copy if the Volatile Pool is local to the Client or close to it.

This combination allow to create a cache system

INFN
Istituto Nazionale di Fisica Nucleare

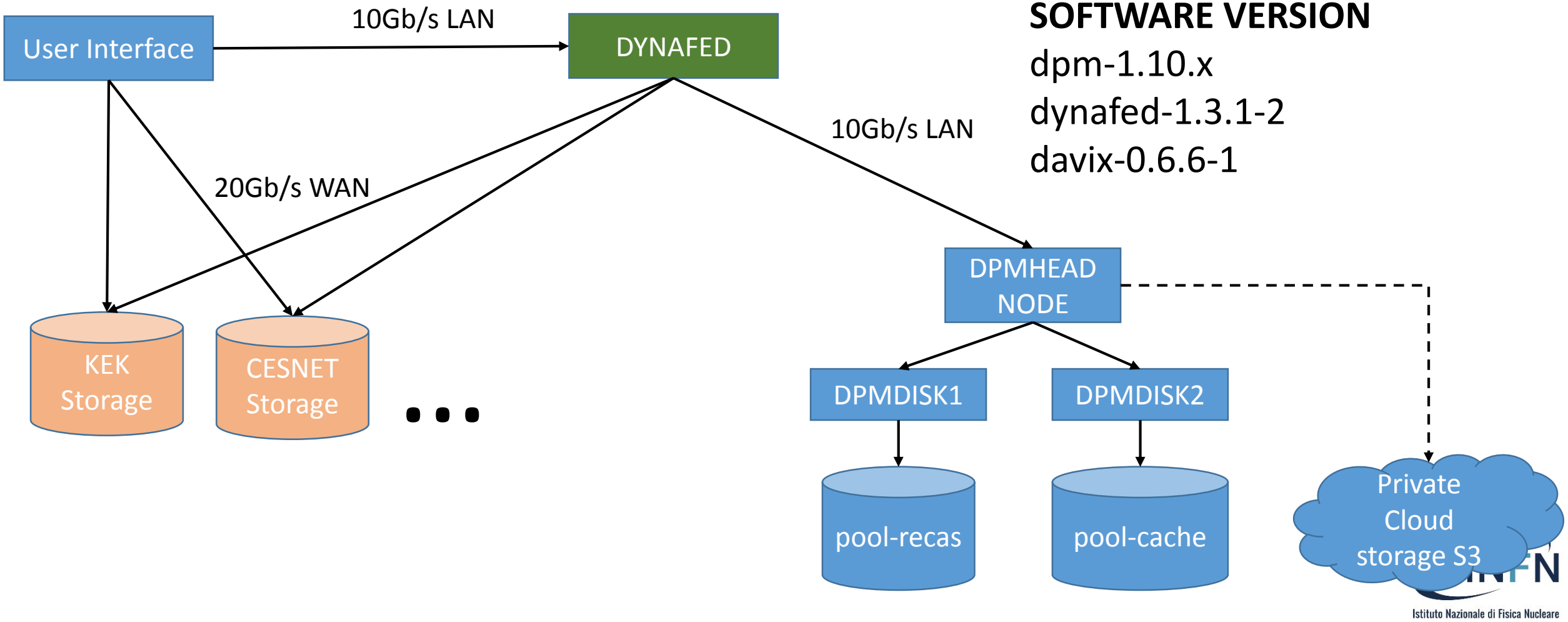
Dynafed and Cache: Model and implementation



Two challenges: User HTTP in the application workflow and implement a caching system

INFN
Istituto Nazionale di Fisica Nucleare

The testbed



SOFTWARE VERSION

dpm-1.10.x

dynafed-1.3.1-2

davix-0.6.6-1

Dynafed Server for Belle II

#	STORGE NAME	HOSTNAME	TYPE
1	DESY-DE	dcache-belle-webdav.desy.de	DCACHE
2	GRIDKA-SE	f01-075-140-e.gridka.de	DCACHE
3	NTU-SE	bgrid3.phys.ntu.edu.tw	DCACHE
4	SIGNET-SE	dcache.ijs.si	DCACHE
5	UVic-SE	charon01.westgrid.ca	DCACHE
6	BNL-SE	dcbldoor01.sdcc.bnl.gov	DCACHE
7	Adelaide-SE	coepp-dpm-01.ersa.edu.au	DPM
8	CESNET-SE	dpm1.egee.cesnet.cz	DPM
9	CYFRONNET-SE	dpm.cyf-kr.edu.pl	DPM
10	Frascati-SE	atlasse.lnf.infn.it	DPM
11	HEPHY-SE	hephyse.oeaw.ac.at	DPM
12	Melbourne-SE	b2se.mel.coepp.org.au	DPM
13	Napoli-SE	belle-dpm-01.na.infn.it	DPM
14	ULAKBIM-SE	torik1.ulakbim.gov.tr	DPM
15	IPHC-SE	sbgse1.in2p3.fr	DPM
16	CNAF-SE	ds-202-11-01.cr.cnaf.infn.it	STORM
17	ROMA3-SE	storm-01.roma3.infn.it	STORM
18	KEK-SE	Kek-se03.cc.kek.jp	STORM
19	McGill-SE	gridftp02.clumeq.mcgill.ca	STORM

Testing Dynafed server in Napoli since Feb 2016

In January 2018 we installed the new new version of Dynafed on CENTOS-7

https://dynafed-belle.na.infn.it/myfed

19 SRM production (about 75%)

Proxy generated by a robot certificate

Version on SL6 Still available
https://dynafed01.na.infn.it/myfed/

INFN
Istituto Nazionale di Fisica Nucleare

Dynafed Setup

Two views configured:

1. Aggregation of a set of Belle II storage endpoints [path /belle]

2. Aggregation of a set of Belle II storage endpoints + with the cache endpoint in Napoli. [path /belle-cache-path]

```
Example configuration for the view that include cache
….
locplugin.*.xlatepfx: /belle-cache-path/ /
….
glb.locplugin[]: /usr/lib64/ugr/libugrlocplugin_dav.so CESNET-SE 5 https://dpm1.egee.cesnet.cz:443/dpm/cesnet.cz/home/belle/TMP/belle/MC/merge1/

glb.locplugin[]: /usr/lib64/ugr/libugrlocplugin_dav.so SCOReS-CacheSE 5 https://recas-dpm-01.na.infn.it/dpm/na.infn.it/home/belle/cache/
```

Behaviour: in the example before, Dynafed creates a metalink with two endpoints, even in the file is not yet in the cache.

If the geoip plugin is activate the first endpoint for a client in Napoli will be always the local cache.

INFN
Istituto Nazionale di Fisica Nucleare

Cache Implementation via DOME

Script on the Head Node:

The implemented script recognize if the requested path is a file or a directory then reply to the client consequently.

Script on the Disk Node:

When a file is not in the cache, the disk node pulls the requested file by resolving the location via Dynafed using view that does not contain the cache in order to avoid loop.

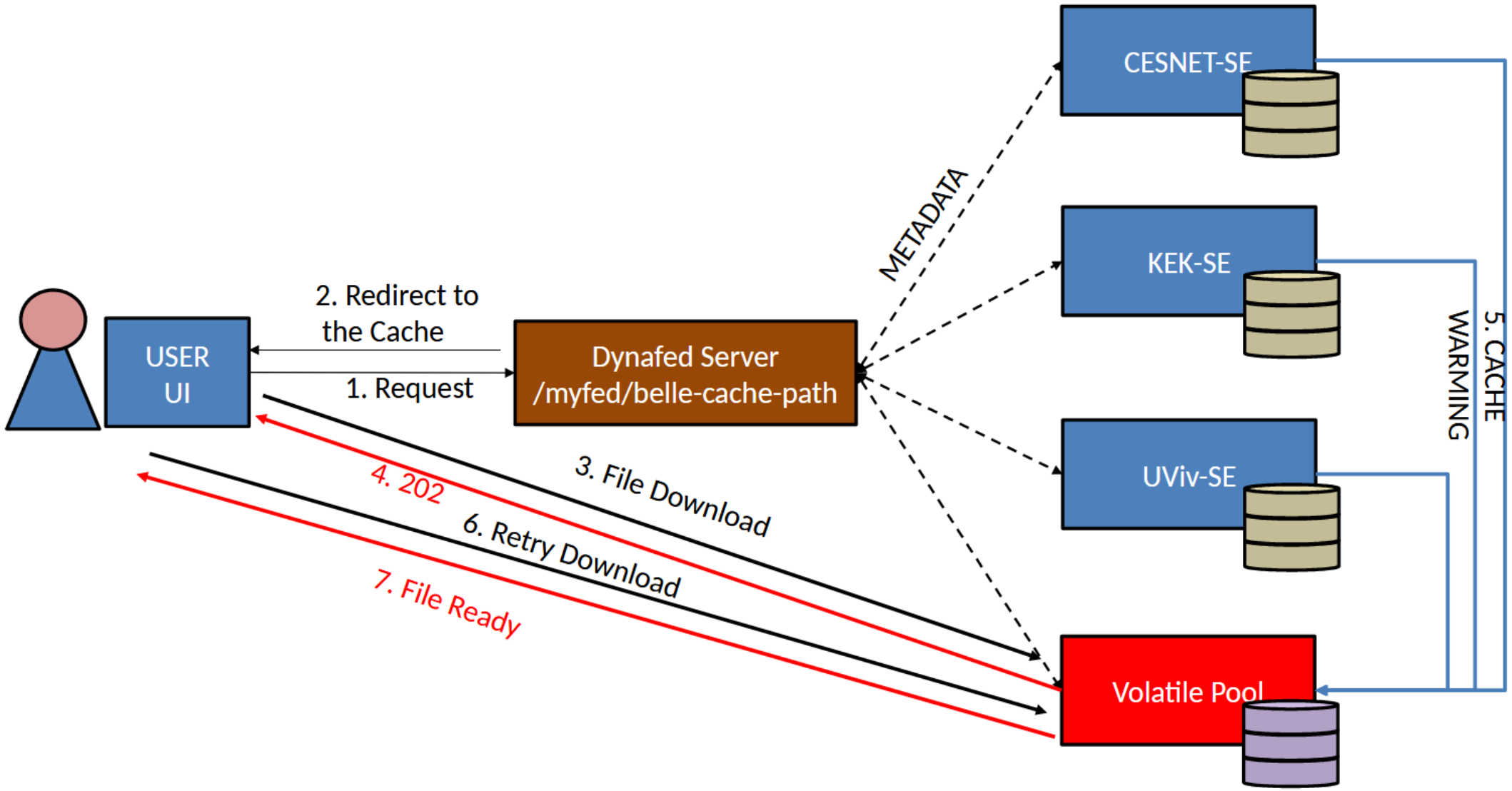
INFN
Istituto Nazionale di Fisica Nucleare

Client Behaviour

- If the cache is not ready, the client receives a 202 Message that ask for waiting.
- Davix or gfal clients will retry after a n-seconds (retry_delay) up to max_retry.
- Then the file will be downloaded from the volatile pool

INFN
Istituto Nazionale di Fisica Nucleare

Implementation Detail



Preliminary Tests Details

As preliminary test, we download from a **User Interface in Napoli** a set of Belle II files, stored in CESNET, KEK and UVic . Each file set is downloaded three times as follow:

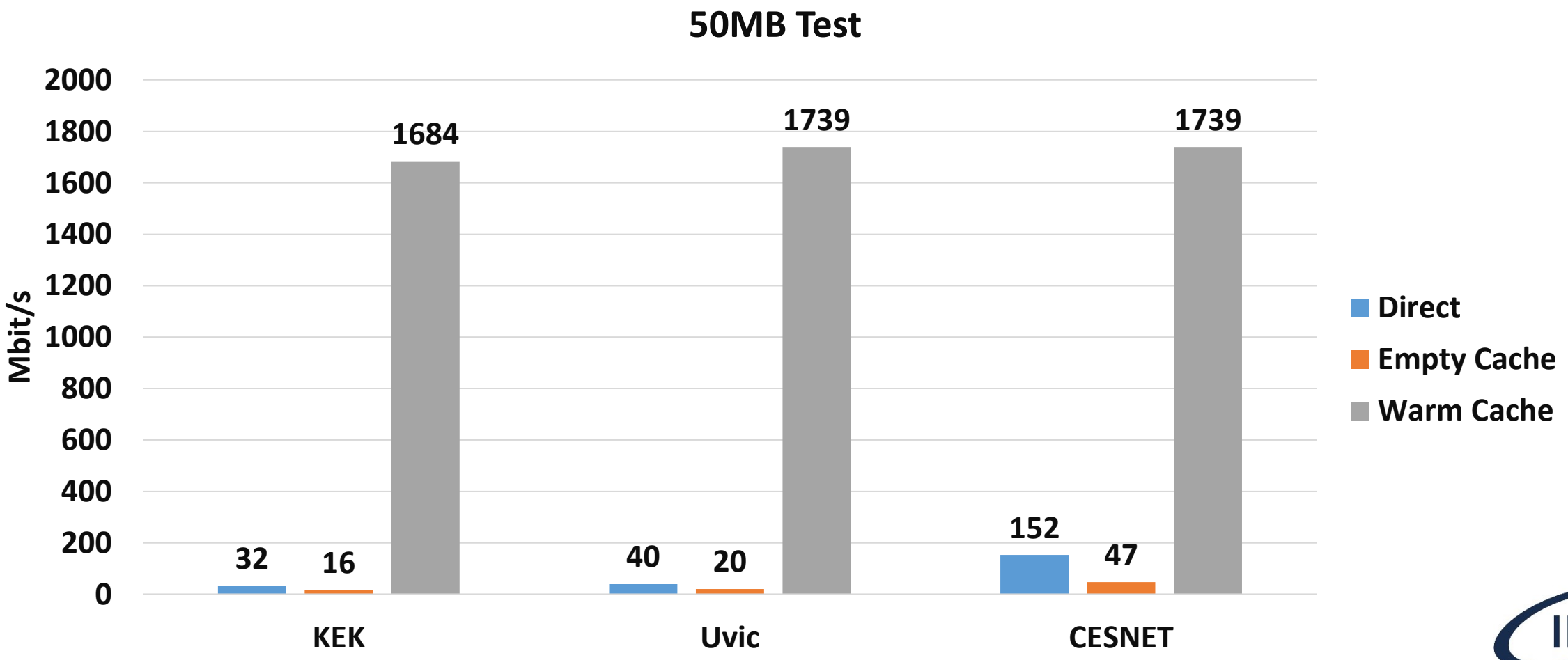
- File Download using the direct link to the remote storage
- File Download using Dynafed with Cold cache
- File Download using Dynafed with Warm cache

Tests have been performed using files of different size: 50MB, 1GB

INFN
Istituto Nazionale di Fisica Nucleare

Test Results 50MB

Mbit/s (Higher is better)



INFN
Istituto Nazionale di Fisica Nucleare

Status of the R&D activity

A minimal set of components has been setup to create an http caching system in a federated environment.

There are still several aspects to investigate.

Massive stress tests have to be performed to validate the setup check the stability the whole solution.

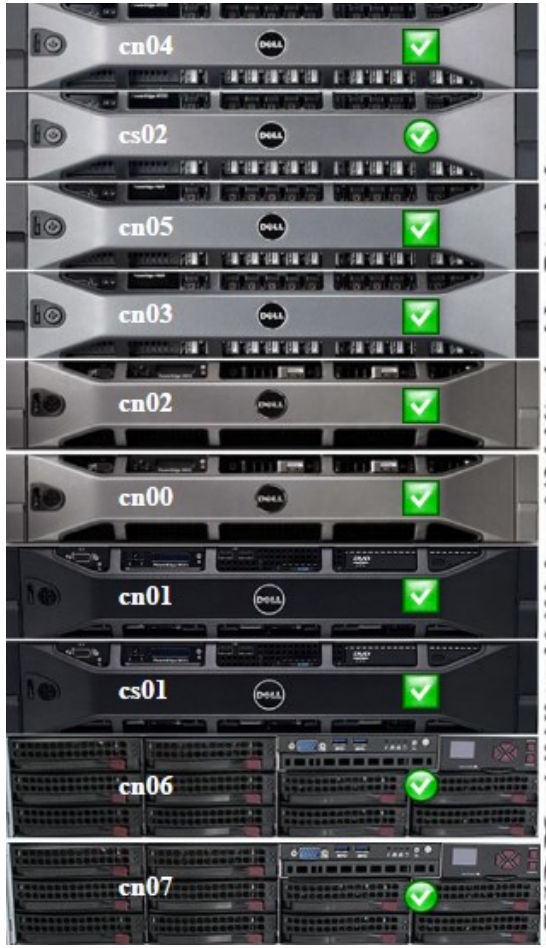
Several new ideas are emerging during the tests.

The teams of the Italian ATLAS Tier2s, Frascati, Napoli and Roma that are based on DPM, are working together to exploit volatile pools and Dynafed for additional use-cases

INFN
Istituto Nazionale di Fisica Nucleare

Thank you

Facilities



The project will integrate the caching system in the RECAS-Napoli infrastructure supporting belle II and Atlas experiment. The goal is to create a pilot system and if possible a pre-production services

For the testbed we can take advantage from a local cloud based on Openstack, with the following characteristics

- 2 Server (tot 80 Cores to store the collective service)
- 384 cores for computation
- 88TB Raw Data
- 10Gbps Network

INFN
Istituto Nazionale di Fisica Nucleare

OUR IMPLEMENTATION