



Testing and evaluating storage technology to build a distributed Tier1 for SuperB in Italy

^aSilvio Pardi on behalf of the SuperB Collaboration

^aINFN-Napoli -Campus di M.S.Angelo Via Cinthia– 80126, Napoli, Italy

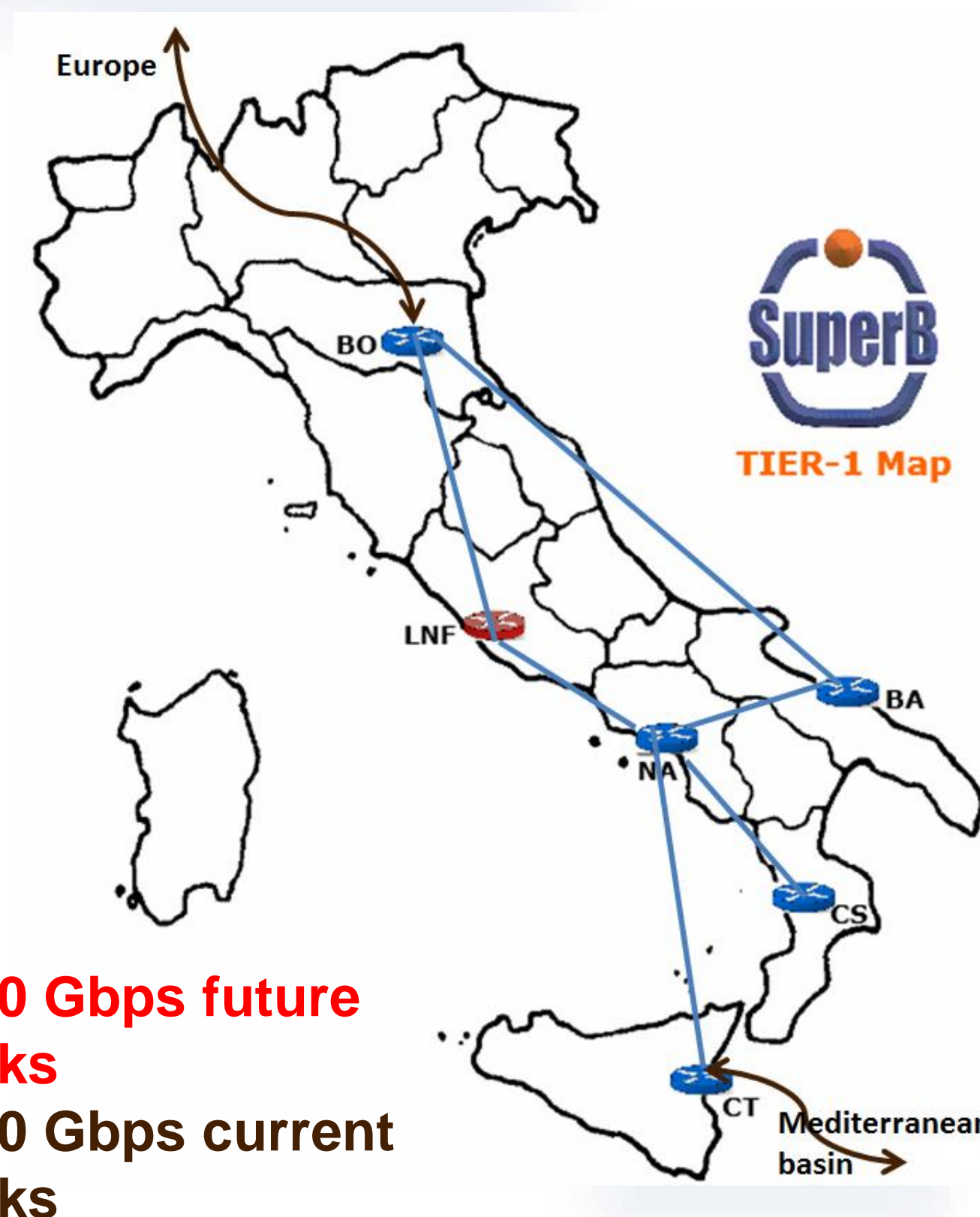
INTRODUCTION

The data analysis of the SuperB collider^[a] will be done in several centers of the International Collaboration.

A new distributed supercomputing center, dedicated for SuperB, is under construction within the Italian Cloud. This infrastructure will provide Tier1 class services, in terms of available resources and processing activities.

The designed Infrastructure is based on 4 centers in south Italy connected by high speed network links.

The goal of the present study is to investigate models and technologies of Data-Storage in order to obtain a reliable and consistent framework to share data between the sites and support both central analysis (Skimming, MonteCarlo) and the chaotic final users' analysis.



Gluster FS Cluster in Napoli

Optical fiber Lan2 10Gbps

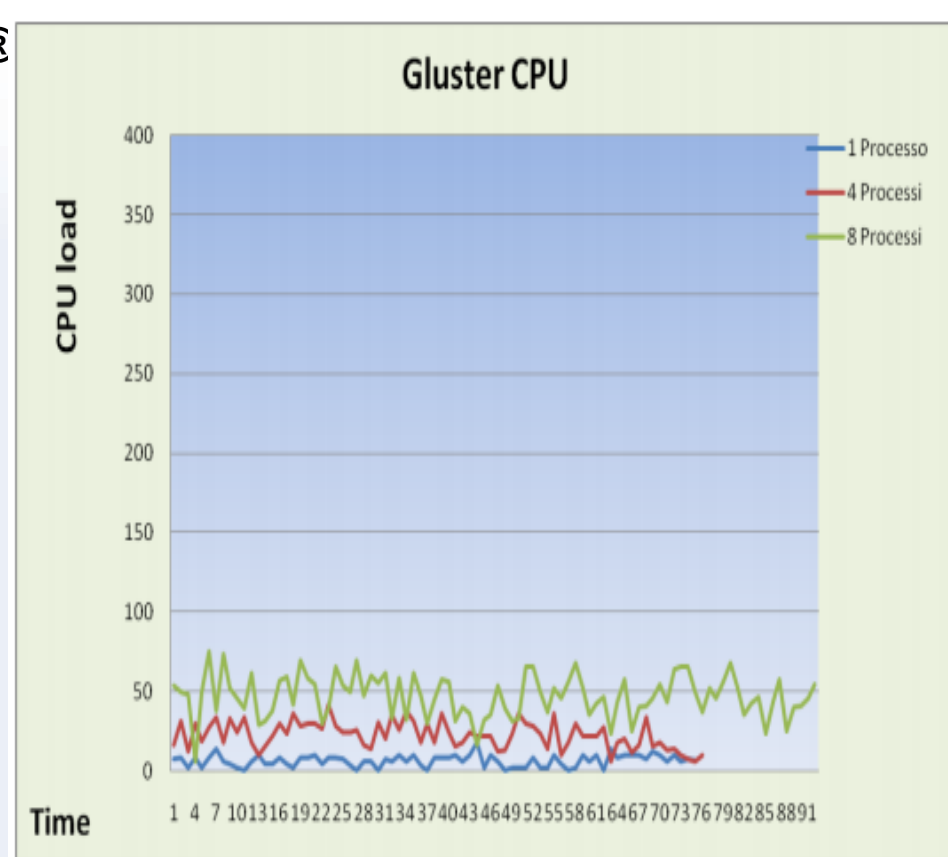
Optical fiber Lan1 10Gbps

12 Server PowerEdge R510

- 2 CPU quad-core Intel® Xeon
- 32 Gigabyte Ram 1066 Mhz
- 4 HD –Raid0 500 GB
- 7200 Rpm
- Broadcom NetXtreme I 57711
- 10GbE NIC
- O.S. Scientific Linux

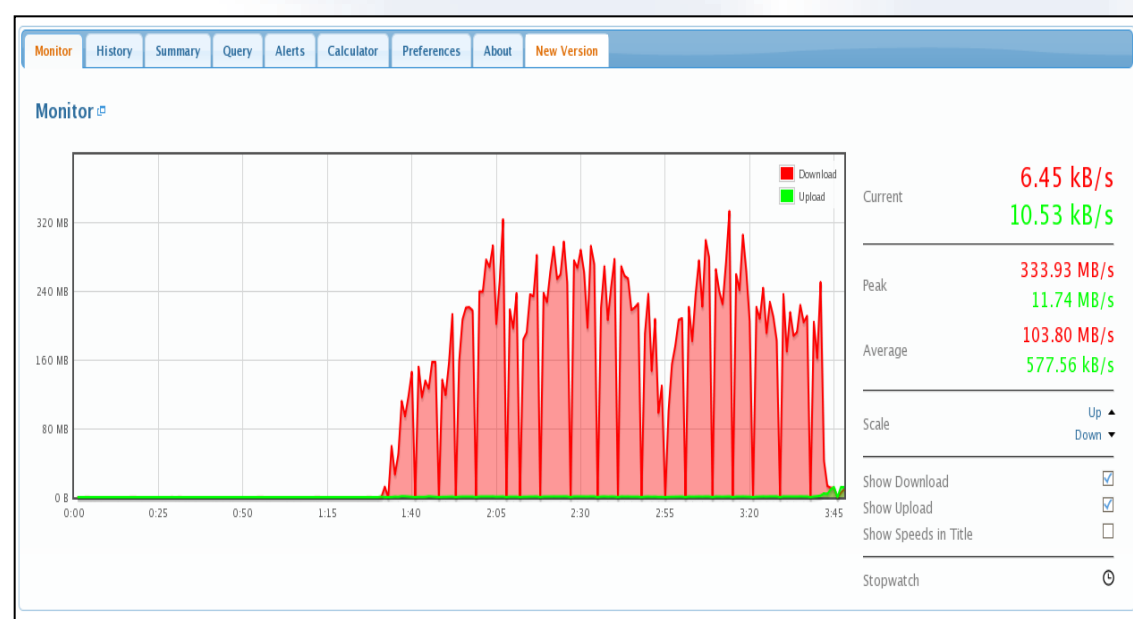
Twinax 10Gbps connection

Gigabit Ethernet connection



Network Usage 8 Jobs on a Single Node
Max rate 333MB/s

Network Usage when all the Nodes are involved
7 Jobs per Node Max rate 333MB/s



We tested the main characteristics of GlusterFS using SuperB analysis Job.

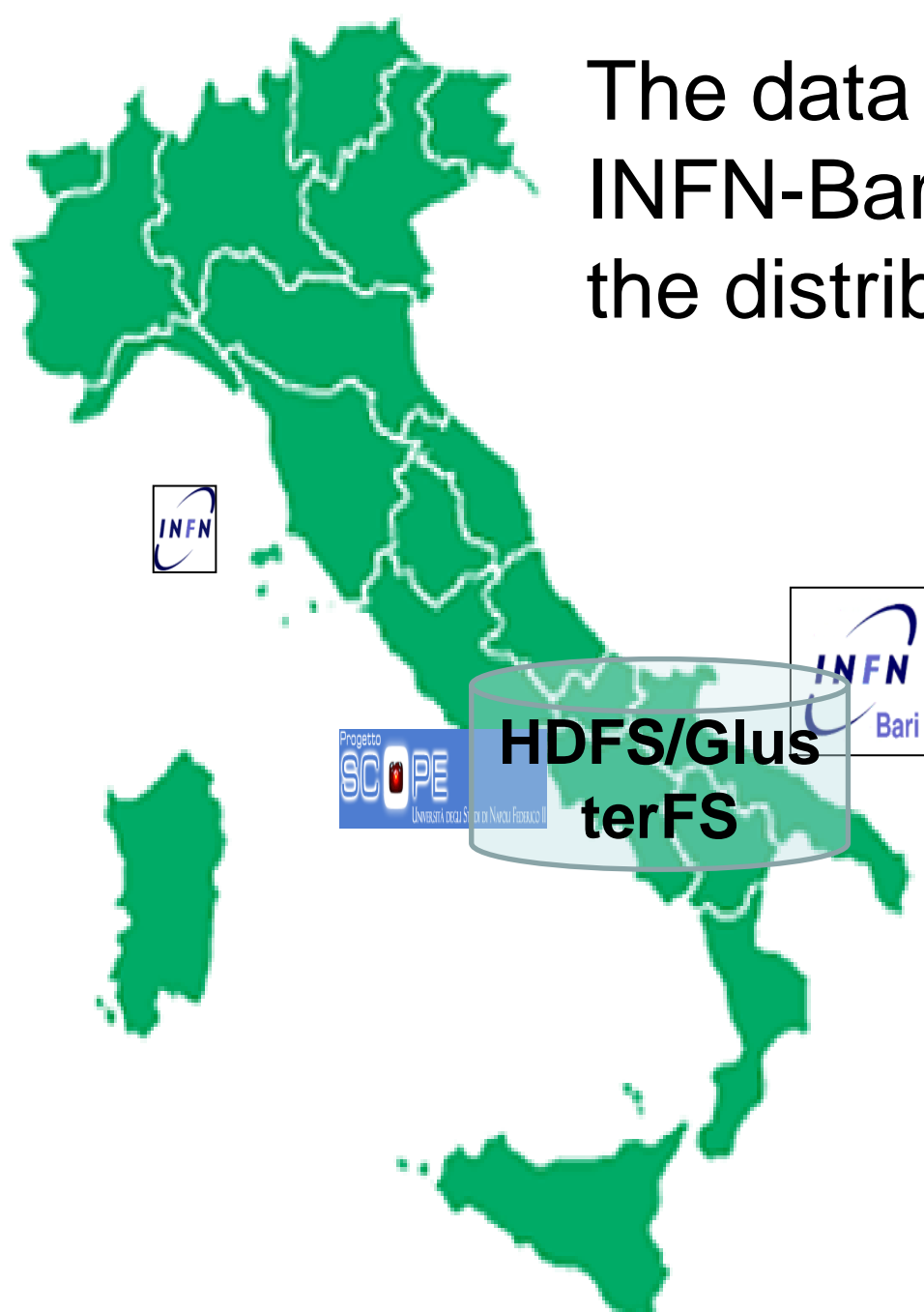
Test on the network usage shows that GlusterFS can scale over a 10Gbit/s in a small/medium cluster of 12-24 nodes. Analysis of CPU of the Gluster Daemon suggests to dedicate a core for node in order to don't affect drastically the performances in terms of jobs/s rate.

The table below shows the time spent for the computation of 1, 2, 4 to 24 jobs on the same machine.

File System	1 Proc.	2 Proc.	4 Proc.	8 Proc.	16 Proc.	24 Proc.
GlusterFS	115	116	114	135	249	421

500GB of SuperB simulate data are replicated from Pisa to INFN-Naples and INFN-Bari in order to test analysis job with HADOOP,

The data can be shared with INFN-Bari hadoop Rack through the distributed HDFS



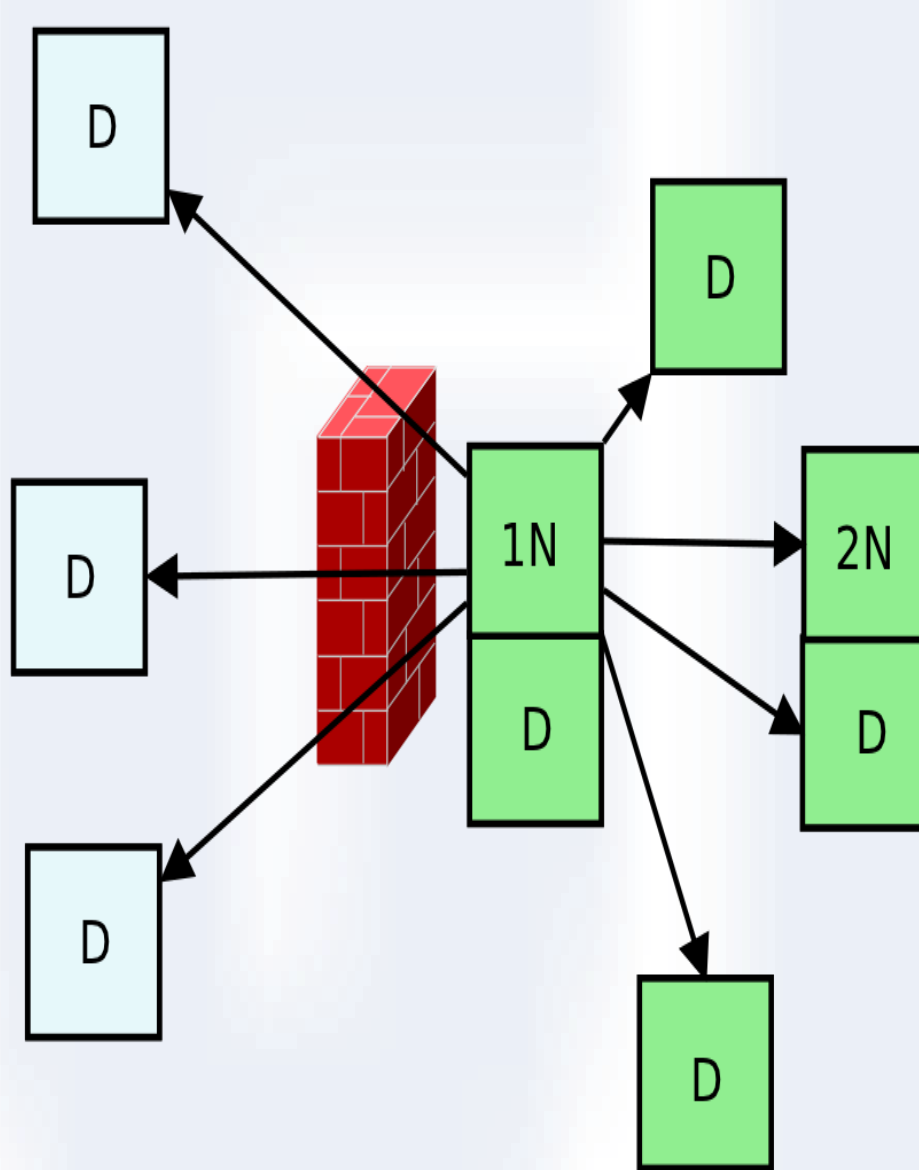
Technology Under Investigation

Exploiting Dynamic Data Placement and automatic data distribution system through the native feature of distributed file system.

- HDFS – The Hadoop File System the fuse module.
- GlusterFS configured as a distributed FS with replica policy.

The interest is to explore those technologies both in a Local Area Network farm and on a Wide Area Network test bed

Hadoop Cluster in Bari

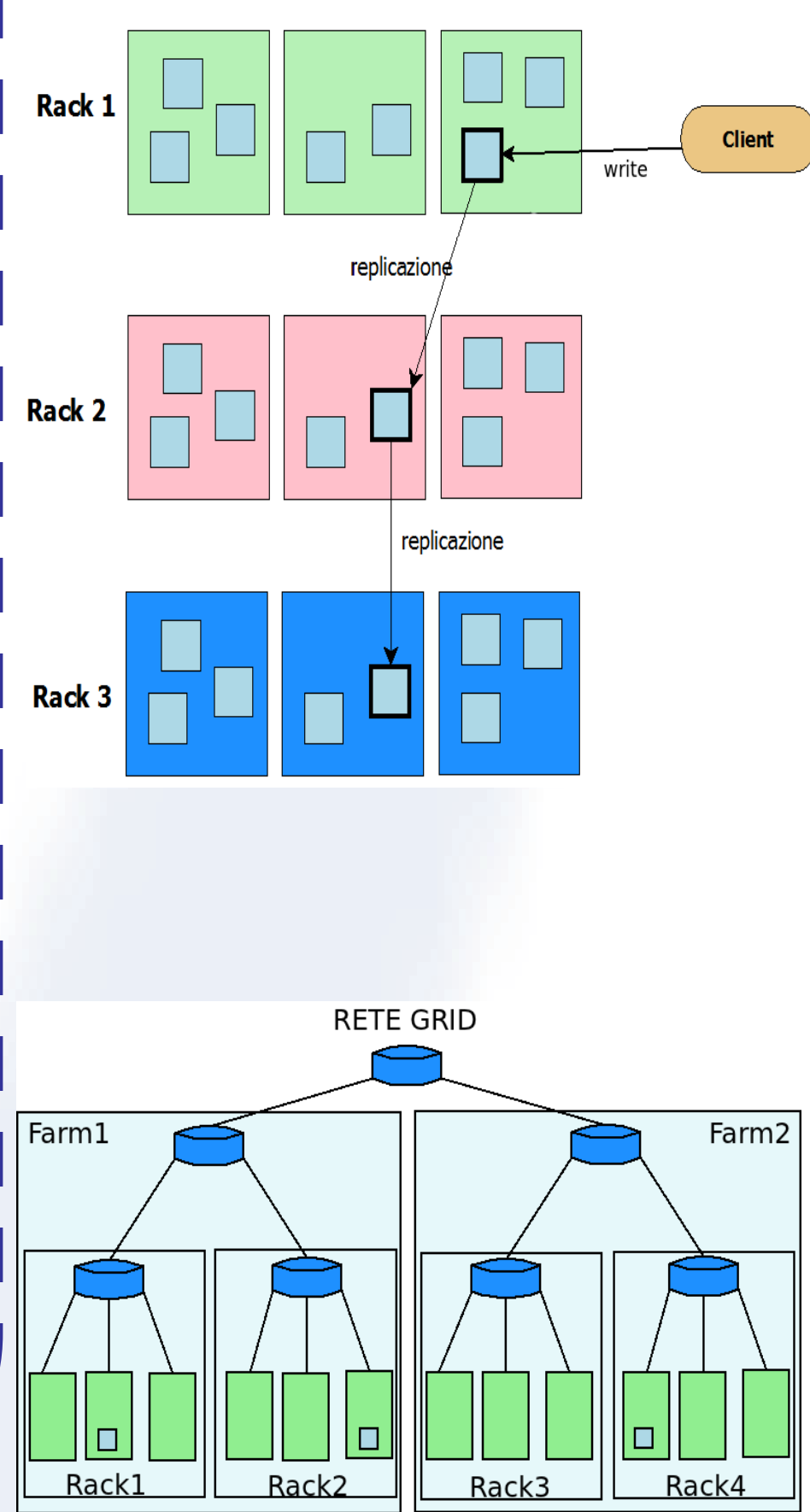


The test on HADOOP carried on at INFN-Bari is focused to the understanding of the capabilities of Hadoop file-system to cope with any kind of hardware and software failures. We have verified that each node (both name node and data nodes) could fail with small or not service interruption

At INFN-Bari, we are modifying the HADOOP replication policy in order to have a broader data distribution among different racks within a farm, and in order to provide the capability of understanding the rack topology.

This added feature could provide the capability of understanding the rack distribution among different farm or in general among different failures domain.

At the end of the work it is foreseen that this policy will allow the file-system to resist to a complete farm failure.



Conclusions and Next Steps

First Test done with GlusterFS and Hadoop showed interesting features for the SuperB purposes, in terms of functionality, scalability and reliability. Both the file systems provide a multicluster configuration extensible over WAN. A first geographic implementation of HDFS was implemented between Napoli and Bari. The next step is to setup the GlusterFS over WAN and make stress tests in order to compare the two technologies.