Updating the HPC platform at the Bank of Italy: Intel EE Lustre as multipurpose filesystem.

Giuseppe Bruno, Bank of Italy

Abstract

The Bank of Italy Economics and Statistics department has recently decided to update his five year old computing platform. Particular emphasis has been given to the selection of open source technologies for operating system, clustering, filesystem, Windows file sharing and platform monitoring. In order to minimize the deployment time, the whole design of the platform has been based on a cautious technology refresh, preserving the main hardware and software solutions previously adopted. The new platform has substantially improved upon the old one along the following two lines: hardware technology and architectural and software resilience aimed at remove any single point of failure. In this poster we provide some performance comparisons between the LUSTRE version 1.8.9 and 2.5.23 & 32. As an example, by running a metadata intensive bash script we achieve a substantial performance improvement. Other I/O performance indicators, derived from a benchmark suite based on standard tools such as bonnie++, iozone, dd and mdtest, show different improvements.

Introduction

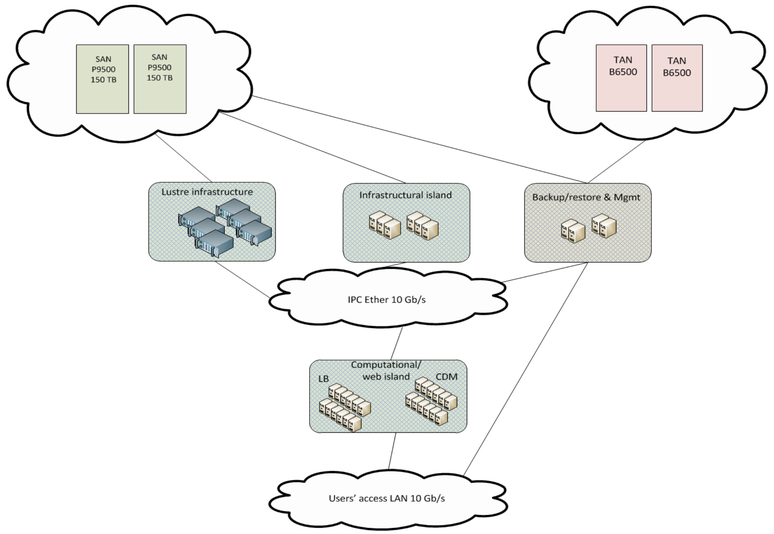
The new computing platform for the Bank's Economics and Statistics department is based on a cluster of INTEL CPU servers running RedHat Linux operating system. The platform is structured in three separated environments: development, test and production. The platform is composed of the following logical elements:

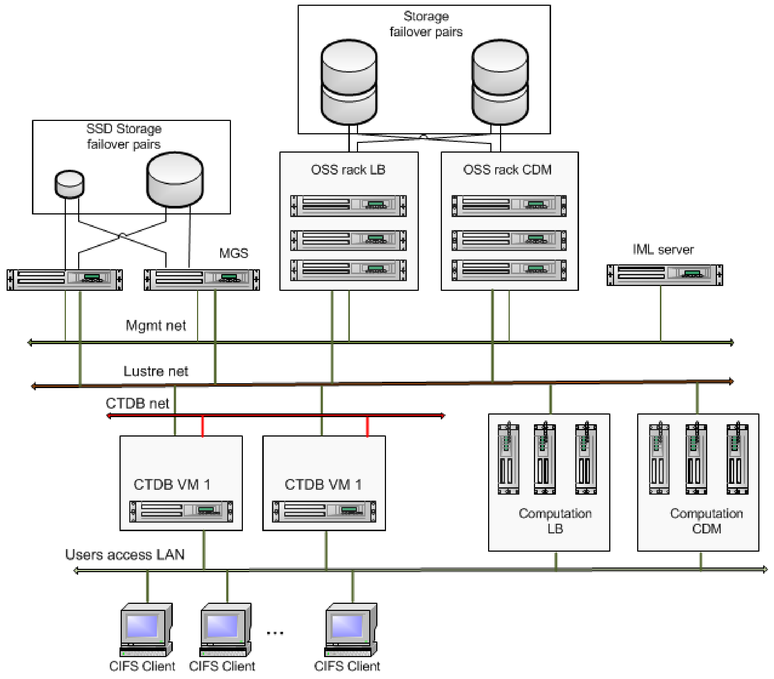
1) **an infrastructural island for DNS, LDAP, and other common services;**
2) **a computational/web island providing statistical software and http services;**
3) **a Lustre based filesystem infrastructure;**
4) **a management island for back-up/restore & platform monitoring .**

The platform employs HP blade enclosures for the computational/web island and rack-mount servers for the infrastructural island and Lustre based filesystem. The servers belong to the 380p Gen8 family.

The present filesystem is populated by 19 * 10^6 files for a disk occupation of 12 TB. The new filesystem will be hosted by a SAN which start with 50 TB with a potential for 75 TB. The median value of the size distribution is 4 KB.

Platform architecture





Monitoring and management of the Lustre infrastructure is carried out through the INTEL web tool IML.

Lustre performances

Benchmark	task			Lustre 2.5.32	
		Lustre* 1.8	Lustre* 2.5.23	Far site	Close site
dd	write	240M/s	252M/s	382M/s	413M/s
Bonnie++	read	1125M/s	946M/s	941M/s	973M/s
	write	207M/s	313M/s	398M/s	401M/s
Iozone 1M	read	1727M/s	3076M/S	3607M/s	3757M/s
	write	192M/s	794M/s	883M/s	1059M/s
mdtest	File creat	905	855	947	1640
	File read	1073	994	1009	2147
	File remov	1368	1576	1676	2778

dd if=/dev/zero of=tmp_DD bs=1M count=1k
bonnie++ -s 1000 –r 500 iozone –a –n1g –g1g mdtest -b3 –z6 –I1000 –i1
(*) worst case measures.

Conclusions

In this Poster we have presented the new computing platform for the Bank of Italy Economics and statistics department. The Lustre filesystem provides a single namespace for both computational tasks and cooperative report production. Lustre MGS/MDS employ SSD devices for its MGT/MDT needs. Disaster Recovery is pursued by distributing the platform on two sites 16 miles away from each other. CTDB provides HA functionality, cluster features and runs the samba software for CIFS clients on Windows 7.
The given performance figures show ligths as well as shadows. The small median file size might hinder further improvements.

References

1) **Lustre manual 2.x *wiki.lustre.org/images/3/35/821-2076-10.pdf***
2) **Lustre Manual Intel Manger for Lustre 2.1.1 User Guide.pdf**

Many thanks go to M. Azzarone, D. Condello for their unremitting patience in designing and testing new avenues.