GOVERNMENT OF INDIA

INDIRA GANDHI CENTRE FOR ATOMIC RESEARCH

KALPAKKAM – 603102

SYSTEM	HIGH PERFORMANCE COMPUTING FACILITIES
TITLE	BACKUP PROCEDURE FOR 134 NODE NEHA CLUSTER USING NAS SERVER

No.	REVISIONS	DATE	APPROVED
A	ORIGINAL ISSUE	25/6/15	

COMPUTER DIVISION				**High Performance Computing**					
COMPUTING SYSTEMS SECTION				**IGCAR/CD/CSS/HPC/08**					
	NAMES	SIGN	DATE						
PREPARED	**M.L. JAYALAL**		25.6.2015						
	SUJA RAMACHANDRAN		25.6.2015						
CHECKED	**S. RAJESWARI**		25/6/2015	**REV**	**A**				
APPROVED	**S.A.V. SATYA MURTY**		25/6/15	Distribution: Head, CSS; Head, CD; Director, EIRSG.					

SUMMARY

A detailed description about the backup procedure for NEHA cluster using the NAS server is given in the document. An introduction to different types of backup is given. Then, the Lustre File System and NAS Storage system of NEHA are covered. The backup utility named *'rsync'* and the backup and restore procedure for NEHA cluster are covered in detail.

Backup Procedure for 134 Node NEHA Cluster Using NAS Server

1. Introduction

The objective of this document is to give details about the backup procedure for 134 Node Dual hex core Xeon based NEHA supercomputing cluster by using the "rsync" command on the Network Attached Storage (NAS) Server.A backup refers to the copying and archiving of computer data so that it may be used to restore the original data, after a data loss event. The primary purpose of backup is to recover data after its loss, be it by data deletion or corruption. In addition, backups can recover data from an earlier time, according to a user-defined data retention policy, typically configured within a backup application for how long copies of data are required. An efficient backup procedure is crucial for the High Performance Computing (HPC) Systems in view of unexpected loss of critical data from the main storage space.

2. Different types of Backup

There are a number of backup types, of which the major ones are full backup, incremental backup and differential backup.

a. Full Backup

Full backup is a method of backup where all the files and folders selected for the backup will be backed up. When subsequent backups are run, the entire list of files will be backed up again. The advantage is that restores are fast and easy as the complete list of files are stored each time. The disadvantage is that each backup run is time consuming as the entire list of files is copied again. Also, full backups take up a lot more storage space when compared to incremental or differential backups.

b. Incremental backup

Incremental backup is a backup of all changes made since the last backup. With incremental backups, one full backup is done first and subsequent backup runs are just the changes made since the last backup. The result is a much faster backup than a full backup for each backup run. Storage space used is much less than a full backup and less than with differential backups. Restores are slower than with a full backup and a differential backup.

c. Differential backup

Differential backup is a backup of all changes made since the last full backup. With differential backups, one full backup is done first and subsequent backup runs are the changes made since the last full backup. The result is a much faster backup than a full backup for each backup run.

Storage space used is much less than a full backup but more than with incremental backups. Restores are slower than with a full backup but usually faster than with incremental backups.

3. Lustre File System of NEHA

In the case of NEHA cluster, the main storage space consists of a parallel Lustre filesystem with 24TB storage space. The storage system of NEHA is configured with the open-source parallel cluster file system named Lustre. The Lustre file system of NEHA consists of scalable storage system with 24 TB of usable storage space configured with 500GB SATA HDDs in RAID configurations. It has a distributed storage across multiple storage nodes to enable parallel access by cluster nodes. The overall architecture of the Lustre storage system of NEHA is shown in Fig 1.

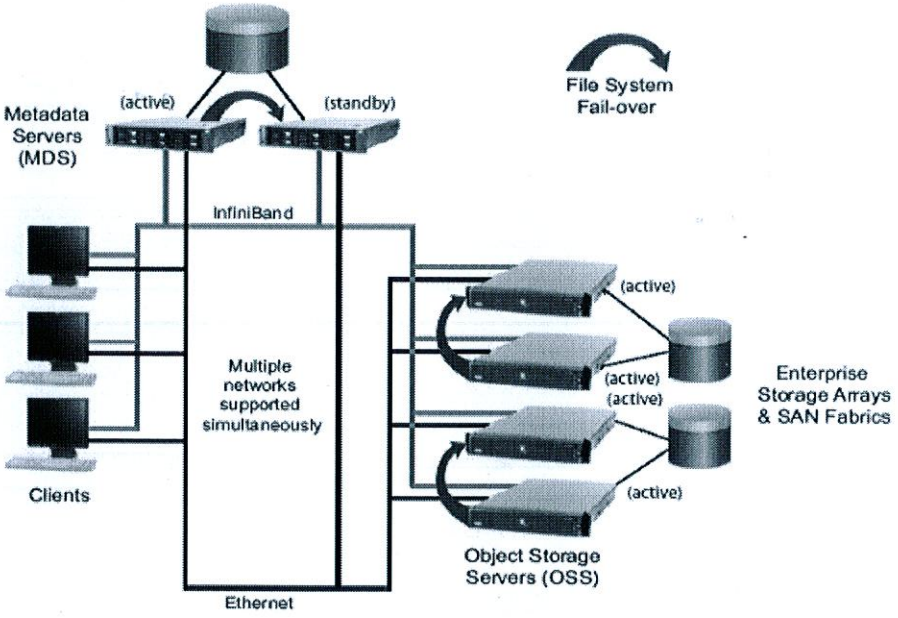

Fig. 1. Overall architecture of the Lustre File System of NEHA

The storage servers consists of two Metadata servers (MDS) and four Object Storage Servers (OSS). The metadata storage connected to each MDS provides a single combined logical volume of 2.5 TB. Similarly, the object storage targets connected to each of the OSS provides eight logical volumes having 3TB of storage. The aggregate volume of 24 TB is used to support the following user groups in NEHA – MSG, MMG, RSEG, NSEG, REG, ROMG, RPG and CG.

4. NAS Storage System Connected to NEHA

The NAS server connected to NEHA support 96 TB of raw storage capacity and 80 TB (approx.) of usable storage capacity. The NAS unit consists of a basic storage unit (QNAPTSEC 1279U SAS-RP) and one expansion unit (REXP-1200U-RP). The main unit has twelve numbers of

4TB hard disk drive and the expansion unit also has the same numbers of hard disks. The 12 disks of each unit is divided in to two RAID groups (RAID level 5), with sixth disk as hot spare. The total available storage space is divided in to four storage pools, each with 14.52 TB capacity. That implies, the basic unit contains storage pool 1 and 2; the expansion unit contains storage pool 3 and 4. Each among the storage pools can further divided in to different logical volume according to the user requirement.

5. The Backup Utility: *rsync*

The backup utility used for the backup of NEHA cluster is *'rsync'*, which is a fast and extraordinarily versatile file copying tool available in Linux platforms. It can copy locally, to/from another host over any remote shell, or to/from a remote *'rsync'* daemon. It offers a large number of options that control every aspect of its behaviour and permit very flexible specification of the set of files to be copied. It is famous for its delta-transfer algorithm, which reduces the amount of data sent over the network by sending only the differences between the source files and the existing files in the destination. *'rsync'* is widely used for backups and mirroring and as an improved copy command for Linux based systems.

The *'rsync'* utility is configured for differential backup of the group wise data in periodic intervals. Differential backup is a backup of all changes made since the last full backup. With differential backups, one full backup is done first and the subsequent backup runs are based on the changes made since the last full backup.

6. The Backup Procedure

In all backups, regardless of the backup type that is used, the data has to be stored on some data storage medium. There are wide varieties of storage media available, such as magnetic tape, hard disks, optical storage and solid state storage. An organization has to design a backup policy with the suitable backup type and data storage media.

NEHA cluster uses parallel Lustrefilesystem and has 24TB storage space. Hardware or software failure of any Lustre component may result in a permanent data loss. The cluster is already having an automated tape library for periodic data backups. However, we have experienced that both backup and restore operations are time consuming and complex to operate. Therefore, for fast and easy access of backup data, a 96TB NAS server is procured and used to as an additional disk based backup.

In the management node of NEHA, two NAS partitions, */nasBack* and */nasBack2* are mounted, with storages sizes of 12 TB and 7 TB respectively. These partitions are used to take backup of the data. A full backup of data in the cluster is stored group-wise, using *'rsync'*.

Differential backup of the data has to be taken group-wise, in periodic intervals. In order to achieve this, backup scripts are written for each user group, which will take differential backup of that group's data using '*rsync*'.

A user account '***backupusr***' is created in the management node of NEHA, through which the operators can initiate the backup process. The stepwise description of the backup procedure is given as follows:

In management node, login as

user: backupusr

passwd: backup

In the user's home directory, the following scripts for backup are available: *backupappl.sh, backupcg.sh, backupmmg.sh, backupmsg.sh, backupnseg.sh, backupreg.sh, backupromg.sh, backuprpg.sh, backuprseg.sh, backuptest.sh*. Each of these will take backups of the respective groups namely, *appl, cg, mmg, msg, nseg, reg, romg, rpg, rseg* and *test*. A backup script will look like:

---

backupmsg.sh

```
#!/bin/bash
if [ -e /nasBack/msgdiffbackups ] ; then
        rm -rf /nasBack/msgdiffbackups
fi
rm /backupusr/msgdiffback.txt
echo "----------Starting msg Diff Backup----------"
nohuprsync -av --delete --backup --backup-dir=/nasBack/msgdiffbackups/ /home/msg/ /nasBack/msgbackup/ 2>&1 > /backupusr/msgdiffback.txt &
```

Since *'rsync'* has to be run as root to access the entire system data, the *'suid'* bit is set for all these scripts so that they can be run with root privileges. To start a backup of any one group, (eg:msg) the operator has to run:

[backupusr@mn ~]$ ***sudo ./backupmsg.sh***

This will trigger the '*rsync*' process to run and take the differential backup of /home/msg. The log messages will be written to a file corresponding to the group, such as *appldiffback.txt, cgdiffback.txt, mmgdiffback.txt, msgdiffback.txt, nsegdiffback.txt, regdiffback.txt, romgdiffback.txt, rpgdiffback.txt, rsegdiffback.txt* and *testdiffback.txt.* So once a backup job is fired, its status can be known by reading the corresponding log file. Corresponding to the example given above, the log file can be read as:

[backupusr@mn ~]$ ***tail -f msgdiffback.txt***

Once the backup is successfully completed, the log file will have a message like:

sent 861313655549 bytes received 1811090 bytes 23590897.59 bytes/sec

total size is 3649819619531 speedup is 4.24

In case the log file ends with an 'rsync error', the backup script has to be run again.

7. The Restore Procedure

The backup data will be stored in the /nasBack and /nasBack2 partitions, such as, */nasBack/msgbackup, /nasBack/regbackup, /nasBack/romgbackup, /nasBack/rpgbackup, /nasBack/rsegbackup, /nasBack/testbackup, /nasBack2/cgbackup, /nasBack2/applbackup, /nasBack2/mmgbackup* and */nasBack2/nsegbackup.* The respective files can be copied from here to the user directories for restoration.

8. Summary

A detailed description about the backup procedure for NEHA cluster using the NAS server is given in the document. An introduction to different types of backup is given first. Then, the Lustre File System and NAS Storage system of NEHA are covered. The backup utility named '*rsync*' and the backup and restore procedure for NEHA cluster are also covered in detail.