A Software Design Document: Backup/Restore Utility

Member Name	RollNo	Email
Shravan R Patel	B14225	b14225@students.iitmandi.ac.in
Umang Agarwal	B14137	b14137@students.iitmandi.ac.in
Sumant Garg	B14134	b14134@students.iitmandi.ac.in
Rohit Kumar Verma	B14125	b14125@students.iitmandi.ac.in

9th May 2017

The Backup/Restore Utility
Design Document

Revision History

Version	Date	Authors	Description
v1.0	05/01/17	Umang,Shravan,Sumant,Rohit	Initial version

Table of Contents

1 Introduction

1.1 Design Overview

Backup/Restore system is based on rsync utility which is commonly found on UNIX like systems. It helps in efficiently transferring and synchronizing files across a network. Clients would be able to backup their files/folders by assigning low/high priority to them. A global node which handles the request from different clients for backup/restore would help in keeping higher number of copies for a file with higher priority across multiple servers available as backup disks. This would guarantee high availability and reliability to clients during restore phase.

1.2 Intended Audience

This document is intended for Professors and Teaching assistants of course System Practicum.

1.3 References

[1] https://en.wikipedia.org/wiki/Rsync

[2] https://unix.stackexchange.com/questions/211595/linux-dump-which-folders-files-are-excluded-from-first-backup

[3] https://unix.stackexchange.com/questions/183504/how-to-rsync-files-between-two-remotes

2 Detailed Design

Backup/Restore utility backs up data every minute by using a configuration file which is created/maintained by the client on his system. Cron a time based scheduler would be running a job on every client node which will push the configuration file from the client node to the global node every minute. Global node would extract the path, priority, backup/restore/BackupToTheCurrent flag from the configuration file and prepare a todo file for the client. Whenever a client wants to backup a new folder/file, the global node will assign a backup disk(s) with most available space to that file/folder in the todo file. Client would be continuously looking for that file in its designated directory at global server. As soon as the client finds it, it pulls the todo file from Global node. A cron job at client node will then run the todo file commands and the commands from log file which where partially executed during the last execution. Global node would maintain a database which will contain client name, path, disk id in tabular form. Each storage device will have a separate directory for every client and in each of those client directories there would be many directories with name as path name to store data of that client (refer **fig 1)**.

2.1 Architecture

Backup/Restore Utility has a skeleton which can be more or less visualized in **fig 1**.

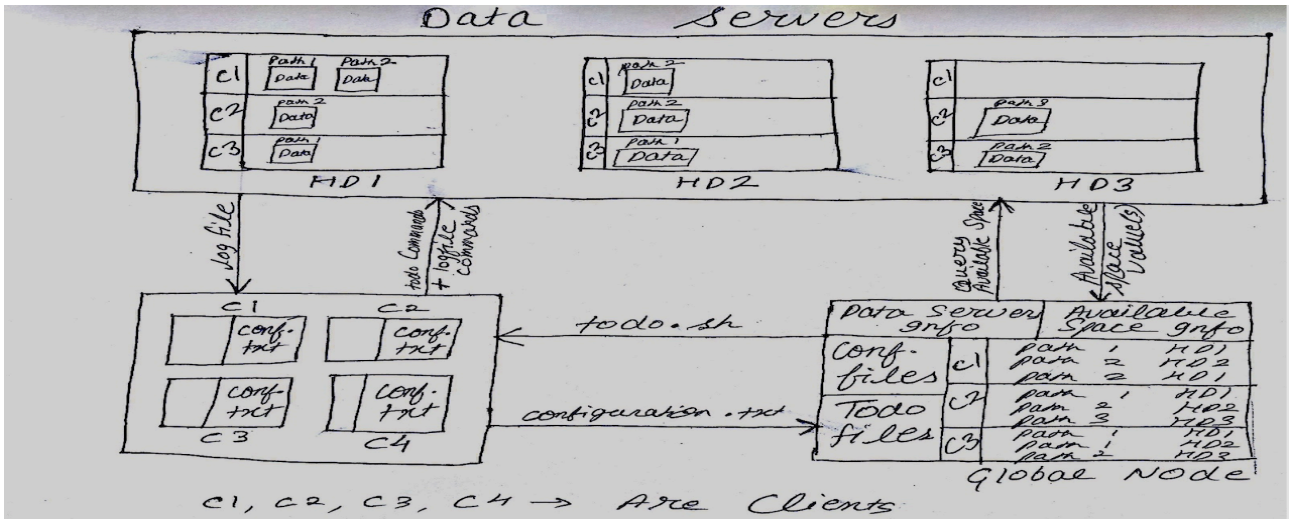

Fig. 1: Backup-Restore Utility Architecture

Components

Client Node are those nodes which uses the Utility.

Global Node help clients in Backing Up and Restoration of their files and keeps track of those files.

Storage Disks are servers which are used to backup data from client machines.

All the components have open ssh server and rsync utility installed.

Interfaces

The interface between all pairs of components is a Command Line Interface (CLI) over ssh Protocol.

2.2 Algorithms and Data Structures

Global node will check in the database whether a folder/file is already backed up or not. If it is already backed up in certain disk(s), global node will place rsync commands with source as those disk(s) in todo file. If it is a new folder/file, global node will assign disks(s) having highest available space. A log file is maintained inside each client node, which keeps commands that were partially executed (due to failures) during the last execution.

2.3 External Data

Databases

1) In Global Node
    1) A table to store client name, path, disk id.
    2) A table to Store Client User Name and Client Id.
    3) A table to store Disk id, Ip Address, Available Space.
    4) To Store configuration files of the Clients and todo files specific to a client.

Files

1) Configuration file containing Client id at the top and path of folder/file, priority and backup/restore/BackupToTheCurrent flag.
2) Todo file containing rsync commands to be executed at the client node.
3) Log file at client node which keeps track of partially executed rsync commands.

2.4 Performance

The utility will be benchmarked on different volumes of data over different clients for Backup and Restoration.

2.5 Test Scripts

Change Configuration.pl It can be allowed to run on each client machine. Script would automatically create/modify a folder on the machine and add that path along with relevant priority and backup/restore/BackupToTheCurrent flag to the configuration file.

Also we could manually switch off some data servers and then show difference in restoration of some high and low priority files.

priority files already backed up.