SMART BACKUP SOLUTION

A PROJECT REPORT

Submitted by

SHAH SHRIT TEJAS (18BECE30556)
SONI YASHVIBEN PRITESHKUMAR (18BECE30562)

In fulfillment for the award of the degree
of
BACHELOR OF ENGINEERING

in

COMPUTER ENGINEERING

LDRP Institute of Technology and Research, Gandhinagar

Kadi Sarva Vishwa Vidyalaya

APRIL, 2022

LDRP INSTITUTE OF TECHNOLOGY AND RESEARCH GANDHINAGAR

CE Department

CERTIFICATE

This is to certify that the Project Work entitled **"SMART BACKUP SOLUTION"** has been carried out by **Shah Shrit Tejas (18BECE30556)** under my guidance in fulfilment of the degree of Bachelor of Engineering in Computer Engineering Semester-8 of Kadi Sarva Vishwa Vidyalaya University during the academic year 2021-22.

Prof. Ashish Patel   **Dr. Shivangi Surati**

Internal Guide,   **Head of the Department,**

LDRP - ITR.   **LDRP - ITR.**

LDRP INSTITUTE OF TECHNOLOGY AND RESEARCH GANDHINAGAR

CE Department

CERTIFICATE

This is to certify that the Project Work entitled **"SMART BACKUP SOLUTION"** has been carried out by **Soni Yashviben Priteshkumar (18BECE30562)** under my guidance in fulfilment of the degree of Bachelor of Engineering in Computer Engineering Semester-8 of Kadi Sarva Vishwa Vidyalaya University during the academic year 2021-22.

Prof. Ashish Patel   **Dr. Shivangi Surati**

Internal Guide,   **Head of the Department,**

LDRP – ITR.   **LDRP – ITR.**

ACKNOWLEDGEMENT

I would like to express my profound gratitude to Prof. Ashish Patel as well as our HOD Dr. Shivangi Surati who gave us the golden opportunity and the guidance to work on this wonderful project, which also helped us in doing a lot of research work and we gained new skills which improved our technical knowledge. We are really thankful to them.

Shah Shrit Tejas (18BECE30556)

Soni Yashviben Priteshkumar (18BECE30562)

ABSTRACT

The IT industry has taken an immense boost in the corporate world and our personal lives, so digital data has become very valuable, making backing up data a widespread practice. Non-technical people often use the traditional backup method, i.e., Copy-Paste data in another storage drive, which has to be done at equal time intervals. This method has its cons like redundancy of data, time-taking etc.

To solve these problems, we've made a project entitled as 'SMART BACKUP SOLUTION', which works on the principle of Network Attached Storage (NAS). This server can read/write the file operations over the internet and automatically backups the files that are saved on to the local machine. Now, configuring multiple workstations as NAS server (workstation that provides services of NAS) manually is time consuming and error prone. So instead, client can run single script on his machine, which will remotely configure n numbers of workstations as NAS servers. After that client also gets facility to create cloud backup via this script, that will sync up data in local file system located at fairly distant locations.

LIST OF FIGURES

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Introduction

Our project's focus area is automating repetitive tasks of creating NAS server, establishing connection between server and client and automatic cloud backups. Setting up a NAS server follows a few steps both on server and client side.

This project utilizes various functionalities of Bash shell scripting. Client first has to download the script and execute it in local machine. And since it uses bash program as an interpreter, the script will execute in server workstation without any compatibility error of interpreter used, as bash is the default shell used in most Linux and Unix systems.

This script is highly modular and automated and it also solves any dependency related error and any other errors that may occur in the process of configuring NAS server and cloud backups.

This script also provides client, the ability to sync up multiple NAS devices across locations, by setting up a cloud storage and periodically setting up two-ways syncs between NAS and the cloud. This brings functionality of high availability, consistency and reliability to local file system. This script will not only configure server but also provides options to make necessary configuration to client as well in order to make connection with sever.

1.1.1 Storage devices

Storage devices are also referred to as digital storage, or storage media. A storage device is any hardware capable of holding information either temporarily or permanently. There are two types of storage devices used by computer: a primary storage device, such as RAM, and a secondary storage device, like a hard drive. Secondary storage can be removable, internal, or external storage. Without a storage device, your computer would not be able to save any settings or information and would be considered a dumb terminal.

1.1.2 Examples of storage devices

Magnetic storage devices:

It is one of the most common types of storage used today with computers.

- Floppy disk and zip disk

- Hard drive and Super drive

- Tape cassette

Optical storage devices:

It utilizes lasers and lights as its method of reading and writing.

- Blu-ray disc and CD-ROM disc

- CD_R and CD-W disc

Flash memory devices:

It is a non-volatile storage and it is cheaper, reliable and more efficient than magnetic media.

- Memory card and Memory stick

- SSD

Online and cloud:

It is becoming popular as people need to access their data from more than one device.

- Network media

- Dropbox

- One drive

1.2 Aims and Objectives

This project aims to make the process of setting up servers (NAS server) easy, more manageable and faster without any risk of system errors and potential security threat. Since we are automating steps of setting up server, it now becomes less prone to human errors. Also, people can utilize their own personal computers (equipped with any of the several variants of Linux OS) as a source of centralized file sharing system. The person just has to run this script, specify a few details asked and a centralized local backup file system is created within seconds.

If you want to create a local file system to serve as a centralized storage unit, either for personal use or small-scale start-up, then NAS is prefect choice for you. Why? Because NAS gives more storage on fast, local networks; better protection than hard drives; and ability to easily sync.

And to set up NAS, you can utilize power of the script created inthis project to create centralized storage unit without any knowledge of NAS as such and that too in just seconds. The final objective is to –

- Save time and be more productive
- Eliminate repetitive tasks
- Fewer mistakes & errors

1.3 Brief Literature Review

This chapter contains the theories and the background to justify the usage of NAS-based solutions. When implemented properly, it lowers the risk for a data loss as well as greatly improves the mobility of the work thus allowing employees to work abroad. The topic contains the basic knowledge of the NAS-based solutions as well as different ways to implement them.

Data storage in information technology is a technology which allows of retaining information. There are different data storage types in the industry, but the concept is very similar. Primarily, there are two types of storage. Type-I is that is used as an HDD with the computer or digital device itself. Type-II is the one that is used for data recovery or data backup.

Network-attached storage (NAS) is also hard disk storage but it's rather an external server, with its own hardware and network connectivity. Usually, NAS is configured to act only as a file server which contains tools to manage the network shares, users and user rights.

Even though there are various different OS's available, they all act more or less similarly as   they utilize and manage the hardware (memory, devices etc.). The options are either Microsoft Windows or Linux, which both offers practically same abilities when implementing a NAS. In literature the findings are mostly based on the author's own views but they do support the fact that Linux should be considered   a very stable OS in this kind of use.

1.4 Problem Definition

People generally manage their own data by storing it into their own hard drives. This generates a threat of a data loss in multiple ways: user may accidentally remove his/her own data and while there are no constant backups made, the data may not be recoverable.

Another threat is that any of the used laptops may get stolen with the data, which may

even generate a risk of misuse of the information which is located in the hard drive. One more threat is the hard drive failure which may lead to a complete data

loss of that specific hard drive content.
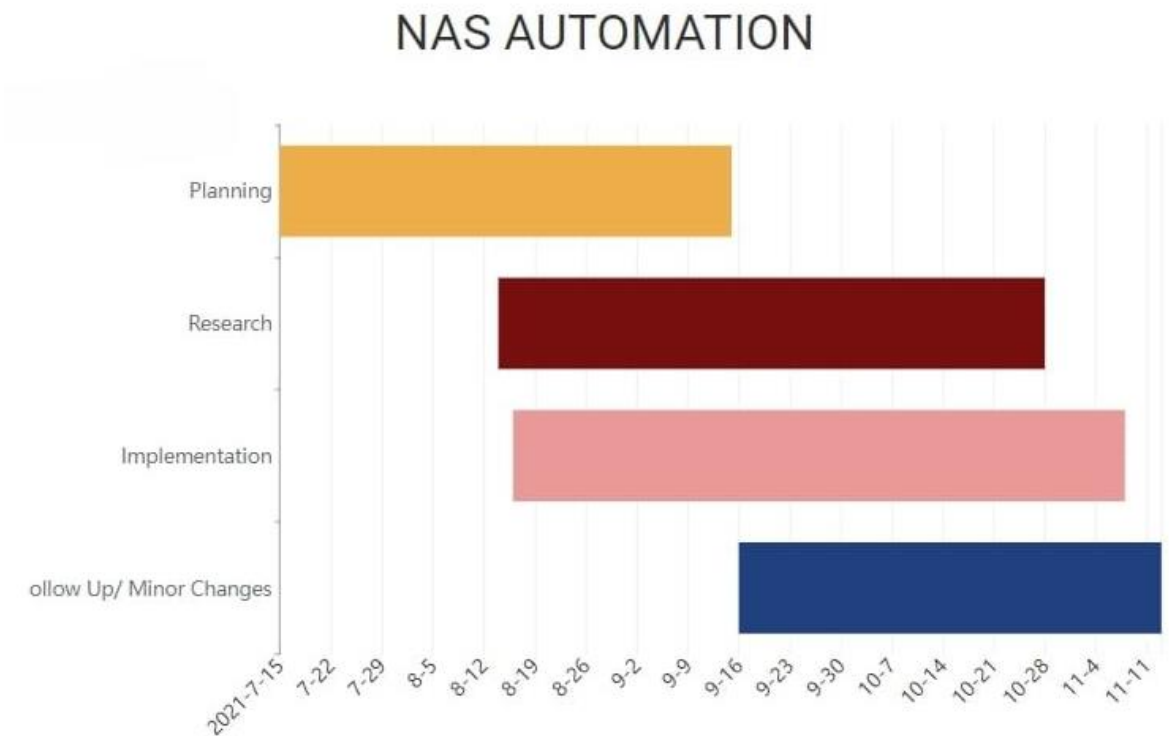
1.5 Plan of Work

Figure 1: Gantt chart

2. Technology and Literature Review

2.1 Technology

This project involves use of technologies like - NAS, NFS protocol, Shell scripting, bash interpreter, Linux operating system, AWS cloud storage for cloud backup and network communication protocol i.e., SSH protocol. In this section of report, we will discuss about all the above-mentioned tools and technologies usedin this project.

2.1.1 Automation

IT automation is the process of creating software and systems to replace repeatable processes and reduce manual intervention. With IT automation, software is used to set up and repeat instructions, processes, or policies that save time and free up IT staff for more strategic work. With the rise of virtualized networks and cloud services that require rapid, complex provisioning, automation is an indispensable strategy for helping IT teams deliver services with improved speed, consistency, and security.

Recent IT automation trends include the use of artificial intelligence and machine learning to create smarter processes that deal with more unpredictable situations. IT automation is useful to replace time- consuming tasks and allow IT staff to keep up with the increasing scale and complexity ofIT operations and cloud infrastructure.

The potential applications of automation are nearly infinite, but some of the most common ones include:

- Cloud automation
- Resource provisioning
- Configuration
- Network management

Advantages of IT automation:

- Cost savings

- Time savings

- Faster operations

- Reduced errors

- Better security

2.1.2 What is NAS? And how does it work?

Network-attached storage is a type of dedicated file storage that's typically connected to a computer network and provides Local Area Network nodes with file-based shared storage through a standard Ethernet connection. It holds data that's typically shared among several heterogeneous types of clients via network. It is often manufactured as a computer appliance – a purpose-built specialized computer. Or you can configure your personal computer (as done in this project) to serve as a NAS server and provide services of storage. NAS devices are configured and managed with a browser-based utility program. Each NAS has its own IP address in network.

Other than being just primitive storage system, it does offer features such as addition of other-network devices, media streaming, remote access, cloud tier, etc. along with regular backup and archiving functions. An important benefit of NAS is its ability to provide multiple clients on the network with access to the same files. Prior to NAS, enterprises typically had hundreds or even thousands of discrete file servers that had to be separately configured and maintained. Today, when more storage capacity is required, NAS appliances can simply be outfitted with larger disks or clustered together to provide both vertical scalability and horizontal scalability. Many NAS vendors partner with cloud storage providers to provide customers with an extra layer of redundancy for backing up files. In the home, NASes are often used for storing and serving multimedia files and for automated backups. In the enterprise, a NAS array can be used as a backup target for archiving and disaster recovery. If a NAS device has a server mode, it can also function as an email, multimedia, database or print server for a small business.
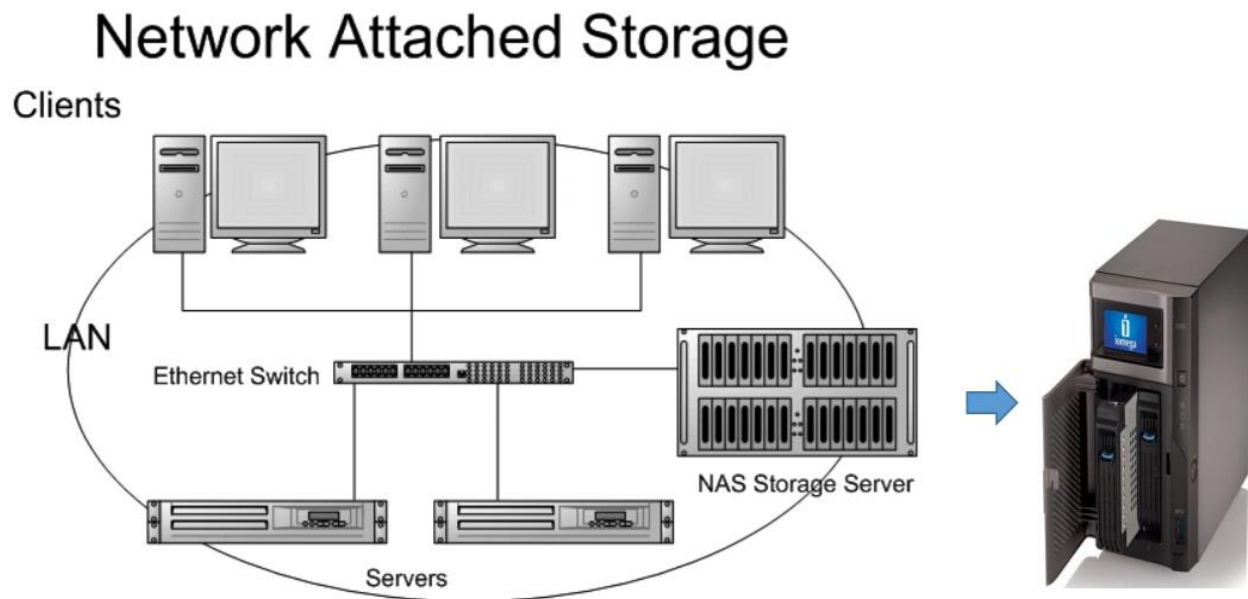
Figure 2: NAS block diagram

NAS servers provide access to files using network file sharing protocols such as NFS (popular on UNIX systems), SMB (Server Message Block – used with MS Windows systems), or AFP (used with Apple Macintosh computers).

Since this project is implemented on Linux based operating system, and both server and client are assumed to have Linux environment, we use NFS - Network File System protocols on both server and client side.

2.1.3 A brief idea about NFS protocol

In simple words, NFS is a protocol that allows sharing file systems over the network. It is an internet standard, client/server protocol developed by Sun Microsystems. NFS enables a client to view, store, and update files on a remote computer as if they were locally stored. On the back end, NFS client software translates POSIX file access commands issued by applications into NFS server requests that respond with metadata, data, and status. NFS requires a client in the kernel that supports NFS mounting.

17

At present NFS allows multiple clients to access same files, thus data remains up-to-date, consistent and reliable. Also mounting file system is transparent to all users, support for heterogeneous environment, data is centralized and fewer removable disks and drives reduce security concerns.

2.1.4 Bash and Shell scripting

This project is created using bash scripting language. We created a fully automated shell script that uses Bash as a command processor/interpreter. Since this project was designed on top of Linux OS, shell is universally available on anything like Unix. Bash is a default shell on most Linux/Unix systems. Also, with shell script we can directly execute system command where special environment settings, command option, or post-processing apply automatically, but in a way that allows the new scripts to still act as a fully normal Unix command.

Because our project has only a command line interface as frontend, therefore using shell scripting is only advisable. It reduces overall code complexity since we can write directly executable commands.

2.1.5 NFS-utils package

In this project we have used nfs-utils package that provides a daemon for the kernel NFS (Network File System) server and related tools, which provides better performance than traditional Linux NFS server. This package also contains the mount.nfs, umount.nfs and showmount programs.

This script allows automatic resolution of any dependency issues if any. Like if client system is not equipped with SSH software, then script will first install OpenSSH software.

Every file system being exported to remote users with NFS, as well as the access level for those file systems, are listed in the /etc/exports file. When the nfs service starts, the /usr/sbin/exportfs command launches and reads this file, passes control to rpc.mountd (if NFSv2 or NFSv3) for the actual mounting process, then to rpc.nfsd where the file

systems are then available to remote users.

2.2   Literature Review

This topic contains the theories and the background to justify the usage of NAS-based solutions in business. When implemented properly, it lowers the risk for a data loss as well as greatly improves the mobility of the work thus allowing employees to work abroad.

2.2.1 The Basics of Data Storage

Data storage in information technology is a technology which allows of retaining information. There are different data storage types in the industry, but the concept is very similar. For example, practically almost every computer has at least one hard drive, where all the information is stored. Such information can be for example the operating system or the files created by user.

This information can be stored and accessed later on. Hard drives are called to be secondary storages as they're not directly accessed by the CPU of the computer. In computer the primary storage handles the information directly from the CPU, but this study doesn't cover this type of storage. NAS is also one type of a secondary storage.
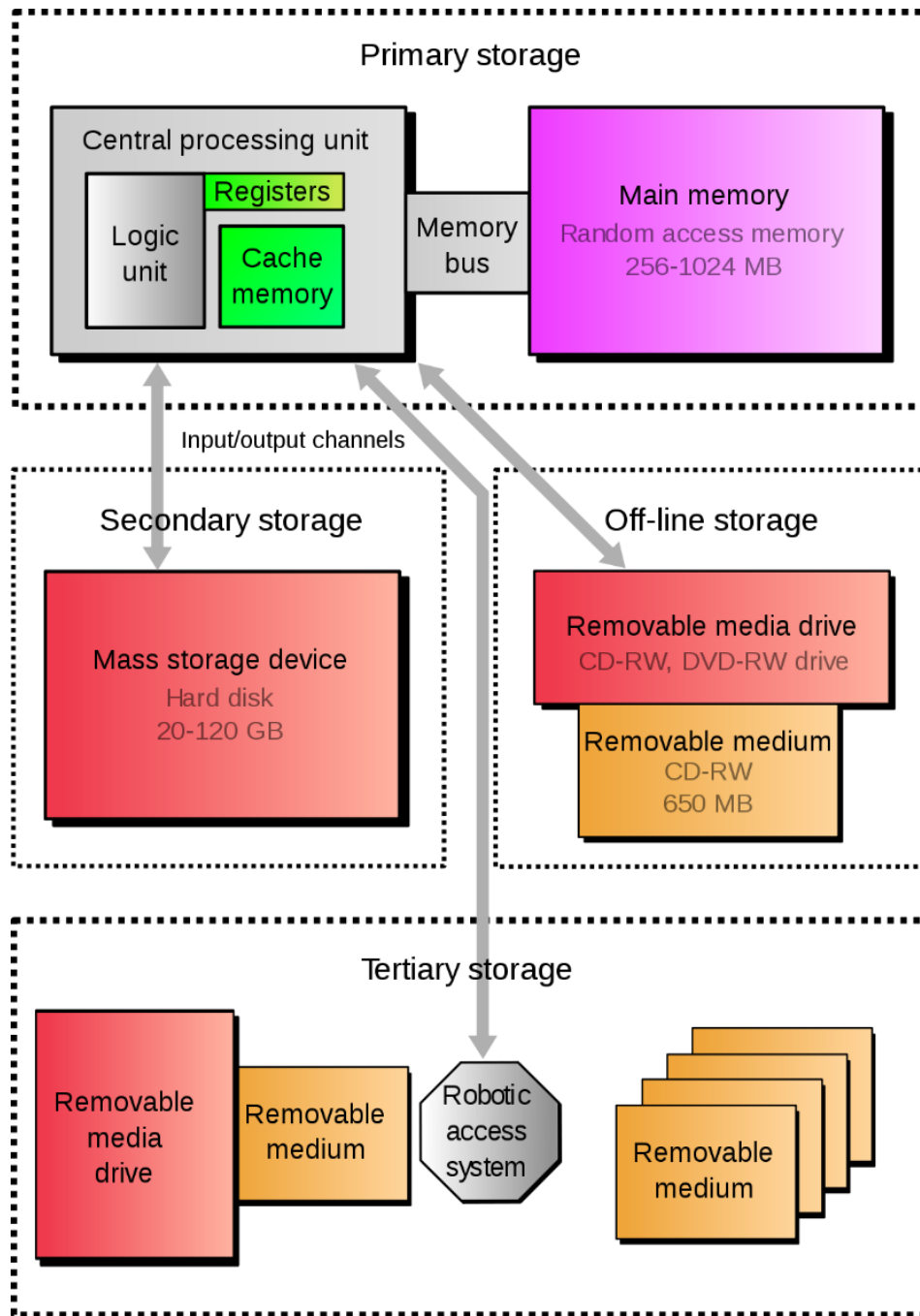
Figure 3: Various forms of Storage types

Above figure represents the different types of storages in the computer world. NAS is considered to be in the "Secondary Storage" which also includes all types of hard disks.

2.2.2 Secondary Data Storage Possibilities

There are various types of secondary data storages available today. Most common is the magnetic disk, which can be found practically from every PC and server. This type of a disk uses magnetic coating which allows the data to be "written" as magnetic particles. The problem with this type of storage is that it is a mechanical 11 devices and it may stay healthy for a long period of time but there's no guarantee of that to happen. There is, however, a way to avoid data loss from a magnetic disk failure and it's called RAID.

One of the Secondary Data Storage types is the Solid-State Device. It's not an optical drive, like CD-ROM drive, nor it's not a magnetic drive either, but it's a drive which utilizes memory chips. As the drive is not mechanical and it doesn't have any moving parts, it can be considered to be less fragile than typical hard drives are. SSD's are coming to home use and in server environments as well, but since the technology is still new, it's not as trusted as magnetic drives in environments where there is no room for errors. The gain from SSD is clear; it's very fast and it consumes only a fracture of power compared to a magnetic drive. Network-attached storage (NAS) is also hard disk storage but it's rather an external server, with its own hardware and network connectivity. Usually, NAS is configured to act only as a file server which contains tools to manage the network shares, users and user rights.

2.2.3 Operating Systems

Even though there are various different OS's available, they all act more or less similarly as they utilize and manage the hardware (memory, devices etc.). Different OSs are designed for different purposes and this research is interested especially in server OSs as the NAS is a server platform. The options are either Microsoft Windows or Linux, which both offers practically same abilities when implementing a NAS. In literature the findings are mostly based on the author's own views but they do support the fact that Linux should be considered a very stable OS in this kind of use.

2.2.4 Importance of Data Security

As the NAS has to be connected to the Internet and it has to be accessible from outside also, the following aspects has to be thought and changed if necessary. An unsecure network and an unsecure NAS-server can lead for a total disaster, which is essential to understand when applying this type of service in a corporation.

Intrusion Detection and Prevention Systems (so called IDPSs) are designed to detect any unauthorized use either in the inner network or in a single computer. There are two types of IDPS –systems. First is the host-based IDPS, which is a single host version. It monitors the status of the files located in the hard drive(s). Second is the network-based IDPS which monitors the network usage and tries to find any suspicious network traffic.

A sophisticated network in a company applies different procedures to protect their LAN from unauthorized usage. Both firewall and VPN are designed to prevent this type of use. A firewall is a piece of hardware, or a subnet, which either blocks or allows connections from inside of LAN to Internet or vice versa. VPN is a non-direct network connection between multiple communication entities, which is virtualized with VPN software over the Internet. The idea behind VPN is to offer secure communications between two physically different places and extend the LAN into more than one place. This can be, for example, utilized to create a LAN between two offices, which belongs to one company.

Cryptography is an essential aspect when creating a NAS which should be accessible from the Internet. If the communication would happen without any type of encryption, the data could be compromised and it could lead into a severe damage for the company. For example, a VPN-connection can be encrypted to allow fairly secure connections to the corporation LAN.

2.2.5 Importance of Data Backup

Backing up the data is important since whenever we're dealing with essential data, we have to have a disaster recovery plan and backups are a very important side of it. There are different ways how this can be done. A full copy of a system, which holds the data, might be the best choice but it's also the solution which costs most. It depends how tolerant system should it be on how big impact it would generate if a specific data would not be accessible for a period of time.

In some cases, it is relatively meaningless if a specific data would not be accessible for couple of days. Yet, there are cases and systems where the data should be accessible no matter what happens. These aspects should be thought when building a backup plan and a system which utilizes backups.

In that case, the company the case company should use the NAS-server for storing all the documents which are made. Therefore, constant backups should be made and they should be relatively easy and fast to recover to an existing system or a new one. RAID technologies are utilized and can be thought to utilize constant backup. This 19 alone isn't enough to offer a needed security, since all the data is still located physically in the same place. Therefore, a service outside of the premises of the company should be utilized for backups. What is a very important to notice is that these backups are never to be considered as a working solution if the backups are never tested to recover the data.

3   System Requirements Study

3.1   User Characteristics

An accurate, comprehensive insight into how a product is used by its users in practice is essential for designing products that meet user expectations. A user centered approach in product development processes asks for a deeper understanding of user characteristics related to the problem's users face when interacting with particular products and services. The user must have following characteristics:

- User must have knowledge of computers and networking concepts.

- User should have scripting knowledge and should be able to troubleshoot if required. And user must be a system administrator or have similar work experience.

- In order to run this program, user must be comfortable working with Linux operating system.

- User must have basic knowledge about what is NAS and it's working principal.

3.2   Hardware and software requirements

3.2.1 Software requirements

- ➢ Operating system: Debian or RedHat based Linux

- ➢ CLI: equipped with bash program

- ➢ Oracle VM Virtual Box (optional)

- ➢ Supported file system for NFS on server machines

3.2.2   Hardware requirements

- ➢ AMD, Intel, or ARM 64-bit architectures processors

- ➢ Stable network connectivity and Minimum 100 MB RAM

- ➢ Laptop, desktop or even virtual machine can also work

3.3   Assumptions and Dependencies

3.3.1 Assumptions

User's command line should have bash program for shell. Client and server can have either RedHat based or Debian based operating system. For cloud-based configuration, a VM in any cloud having a public IP Address is supported.

3.3.2 Dependencies

- ➤ nfs-utils library is dependent on
  - rpcbind
  - nfsidmap
  - gsproxy
  - libevent
  - device-mapper

4   System Diagrams

4.1   Use Case Diagram

4.1.1   Use case diagram of CLI interface
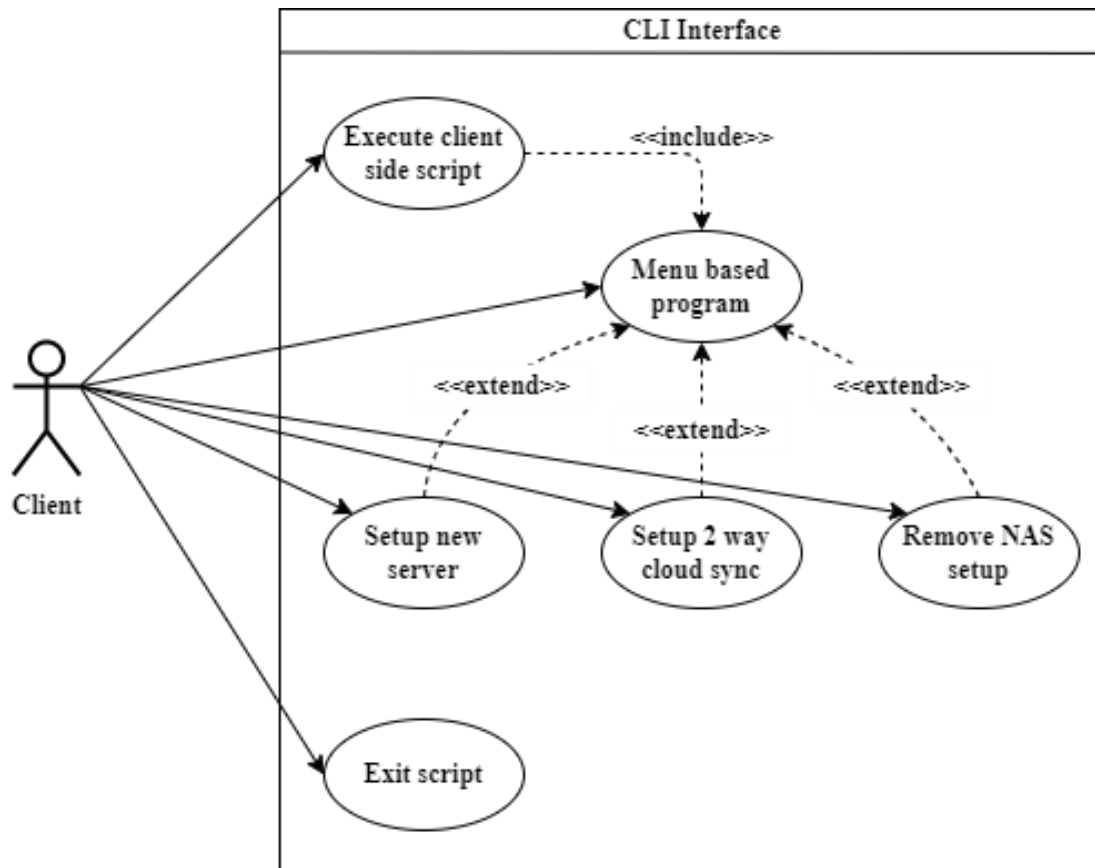
Figure 4: Use case diagram of CLI interface

Scenario for CLI interface

Introduction: the first thing client encounters when executing script is main menu options to choose from

Actor: Client

Precondition: Client must be well versed with CLI interface and bash

Postcondition: Client gets to choose from various modes of configuration and also an option to exit from program.

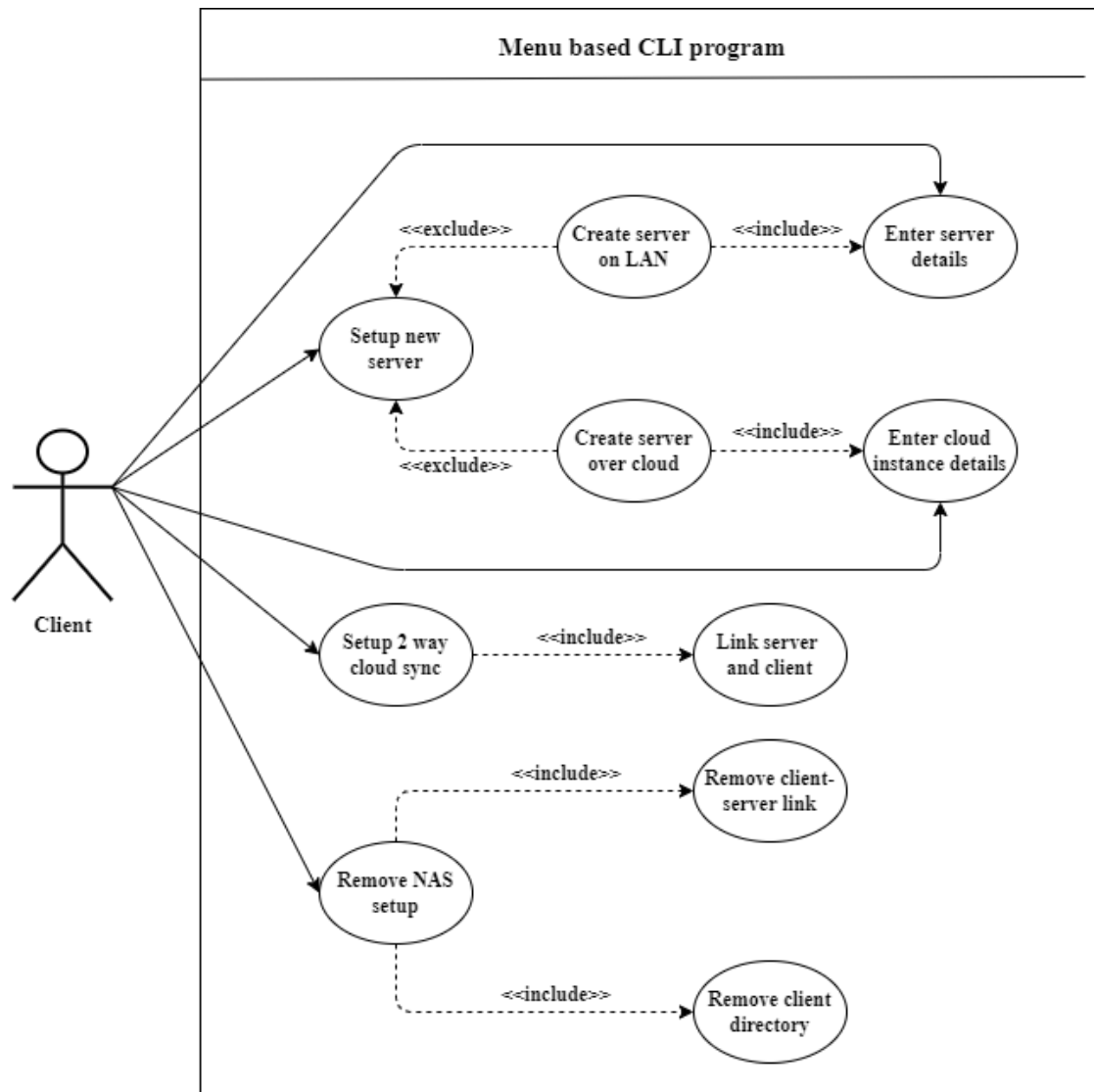### 4.1.2 Use case diagram of Menu program

Figure 5: Use case diagram of menu program

Scenario for main menu

Introduction: Client gets options to choose mode of configuration

Actor: Client

Precondition: Client must be well versed with CLI interface and bash program

Postcondition: Client will then be asked to enter few basic details of server machine and client information will be automatically collected.

Basic Flow: -   Client will select local mode of configuration

           -    Program will ask for details like IP address of server, username and password etc.

4.2 Activity diagram
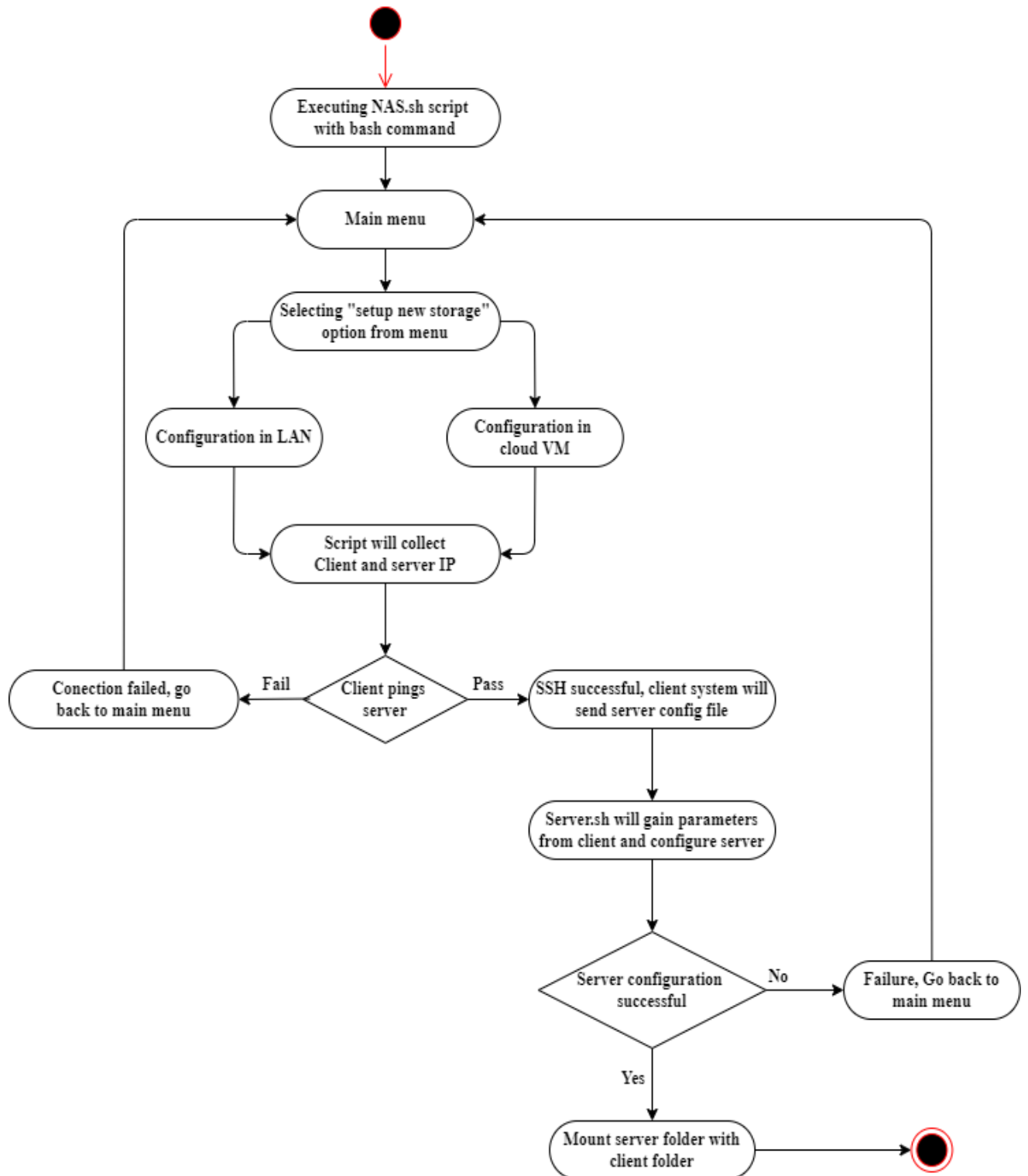
4.2.1 Activity diagram for configuration of NAS setup

Figure 6: Activity diagram for configuration of NAS setup

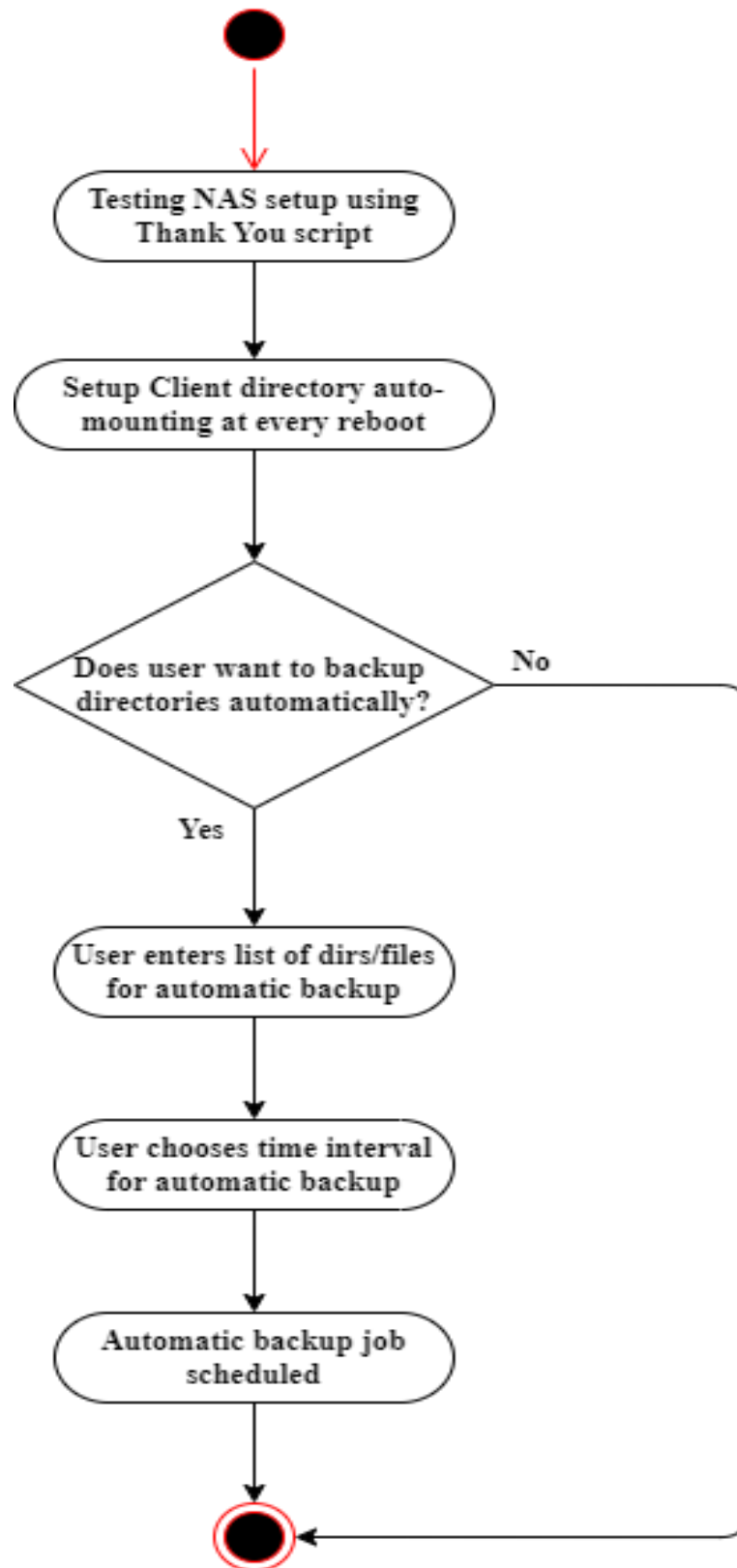### 4.2.2   Activity diagram for scheduling Backup Automation

Figure 7: Activity diagram for scheduling Backup Automation
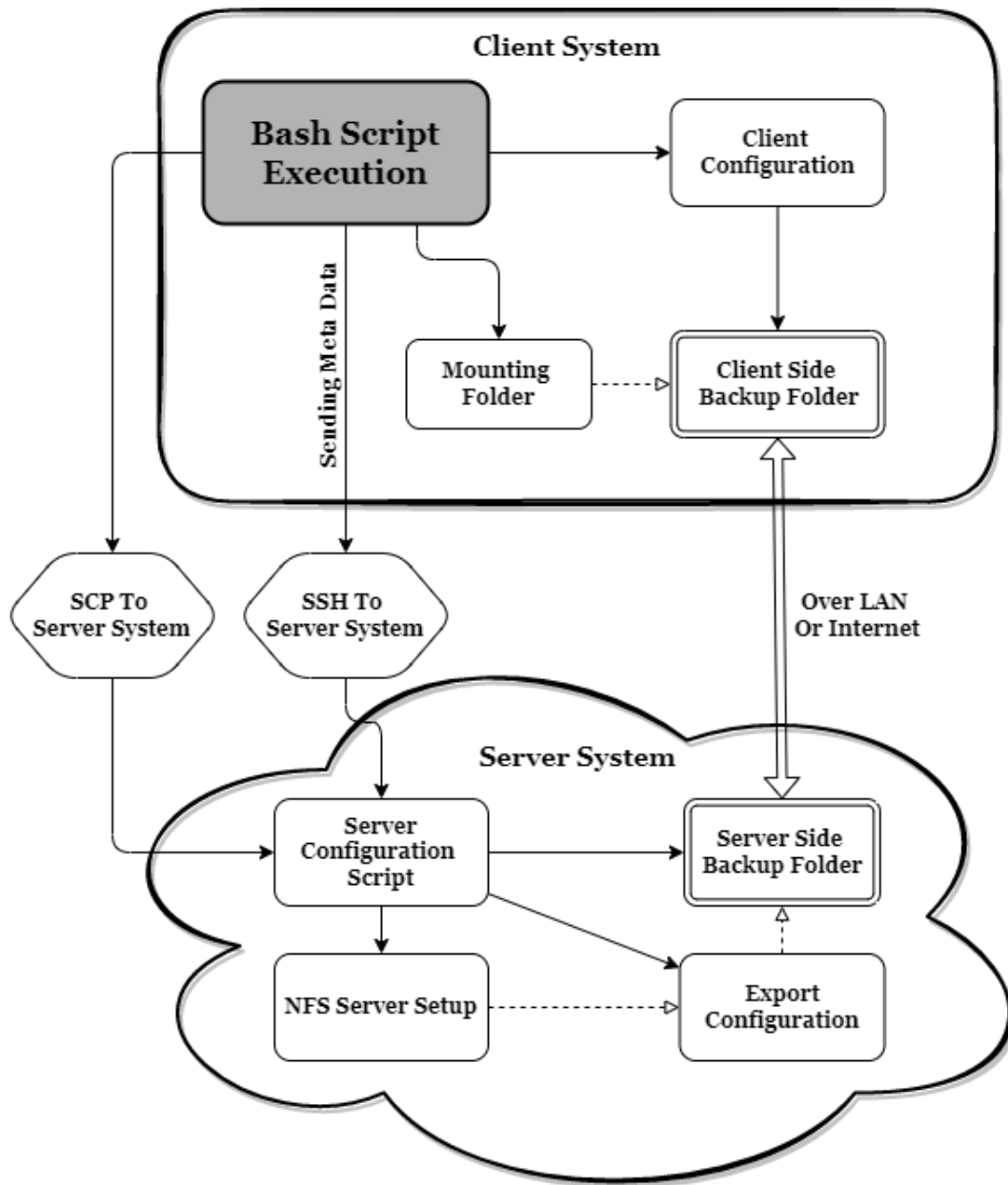
4.3 Basic system design

Figure 8: Basic system design diagram

➢ As you can see from above diagram, user/client will only interact with his machine at his workstation, and every communication and message passing from client to server is handled by Bash Scripting. SSH and SCP will make connection to server, share necessary modules to server in order to perform configuration steps directly on server.

4.4 Sequence Diagram

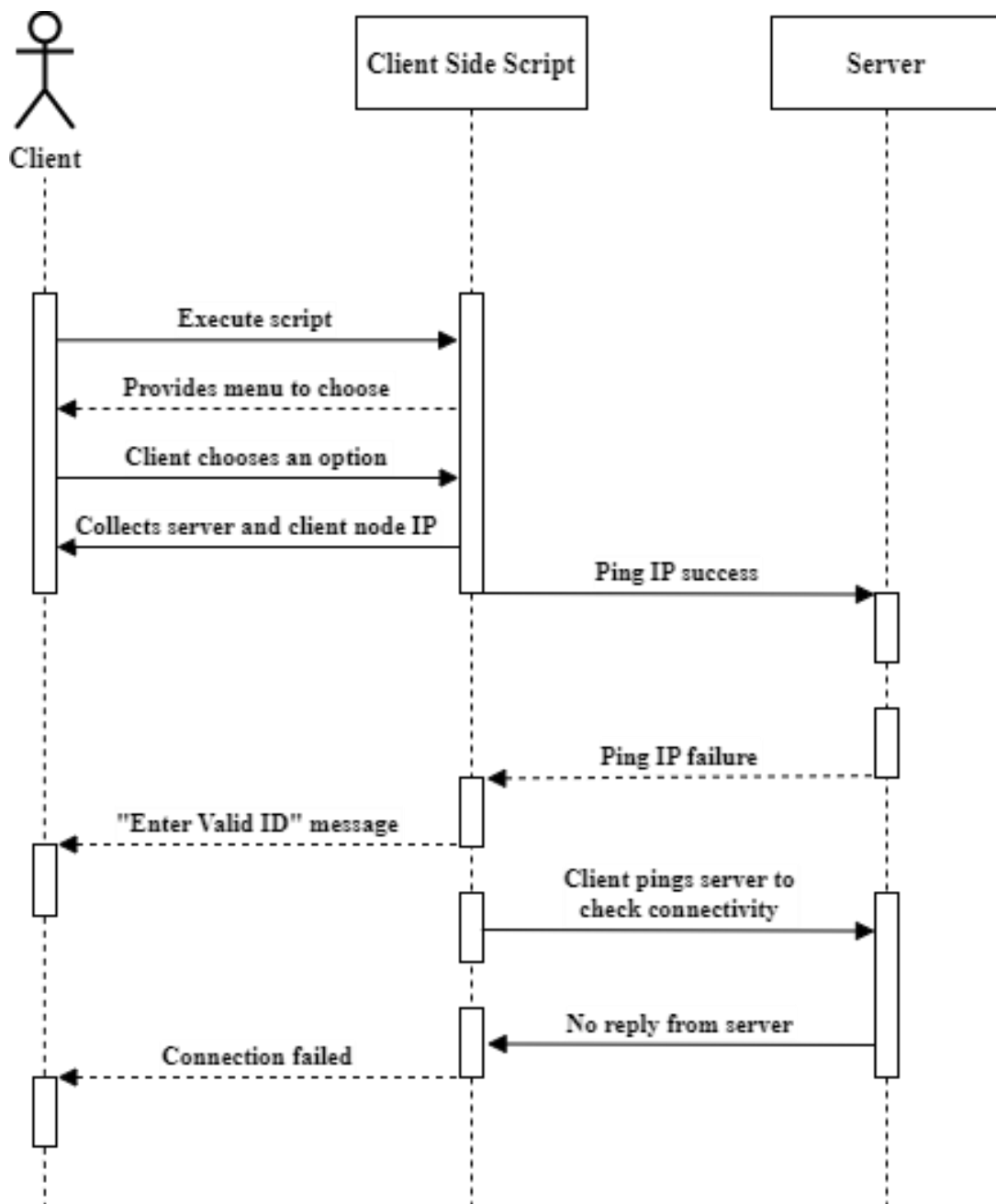### 4.4.1 Sequence Diagram of IP validation

Figure 9: Sequence diagram of IP validation
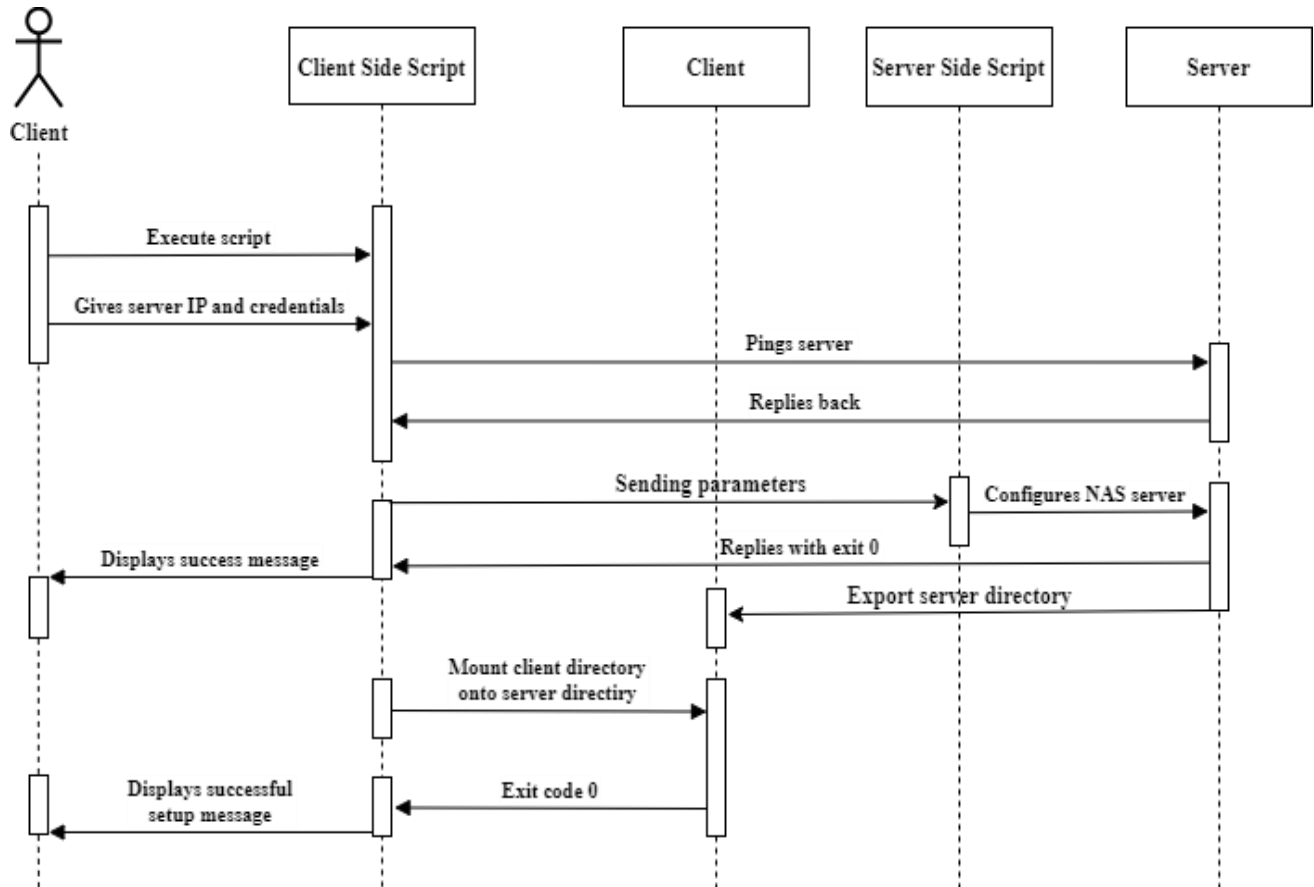
4.4.2 Sequence diagram for configuring Server

Figure 10: Sequence diagram for configuring server

5 Result, Limitation and Future enhancements

5.1 Results after project development

Upon completion of this project, we were able to create an automation script that would allow a admins or someone with a very beginner's knowledge about Linux and other technologies used in this project, to create a Network Attached Storage server on a local machine as well as VM on AWS cloud (as per user's requirement). And this script will automatically mount client folder on NAS server with a minimum interaction of user. This will create a system for automatic backups. And user will be able to schedule cronjobs for automatic backups at user specified time intervals.

With this project we got experience of working with scripting language such as shell script. Our project comprises of 99.9% of Shell and 0.01% of awk.

Also, we were able to implement some networking concepts like SSH and SCP and how to configure server over cloud remotely.

5.2 Limitations

Currently this script is developed and tested in Red Hat Enterprise Linux 8, Fedora 35, Ubuntu 20.04 LTS desktop operating systems along with Raspbian OS (Raspberry Pi) and Amazon Linux 2 VM and is compatible only in Debian and RedHat based operating systems.

Other Linux distributions like Arch based Linux or Gentoo Linux are not supported.

This project is more like Linux utility program so the frontend is a Command Line Interface based menu created using shell script. So, user must be comfortable with shell environment.

6 Conclusion

This script work by configuring remote machines as a server via SSH and once configured as NAS server, folder mounted on the server machine from the client machine acts as an automated backup. Anything and everything that is uploaded to this client folder on client machine will automatically be available on server machine, which can be either local or on cloud.

7 Bibliography

- What is NFS server? - https://bit.ly/3otlvdM

- For error solutions - https://stackoverflow.com/

- NAS collab with cloud - https://bit.ly/3kAVLLn

- AWS EC2 instance guide - https://go.aws/3HlYe6n

- Linux Shell explanation - https://explainshell.com/

- Script commands manual - https://cht.sh/

- Cron schedule expression editor - https://crontab.guru/