## Private Cloud Setup with User Storage Management

Introduction

With the rapid growth of the internal volume of enterprise data, it causes many problems in terms of storage, such as the lack of adequate hardware storage devices, high maintenance costs, and incomplete storage environment and so on. Cloud Computing technology can provide customers with the corresponding demand services that can effectively address these issues.

Cloud Computing provides us by which we can access the applications as utilities, over the Internet. It allows us to create, configure, and customize applications online. With Cloud Computing users can access database resources via the internet from anywhere for as long as they need without worrying about any maintenance or management of actual resources.

However there are security and privacy concerns to be addressed before customers readily outsource their sensitive data. To be effective, a cloud storage service should provide, at a minimum, features and security guarantees like confidentiality, integrity, availability, reliability, efficient retrieval and data sharing Most cloud consumers want their cloud data and its usage to remain private, so along with encryption of their data objects, they also want their access patterns to be oblivious to the cloud provider.

Purpose

We will design and configure an algorithm to cloud user storage management on HADOOP (VMWARE) at operating system of Linux. Manage user directory by using HDFS (Hadoop Distributed File System) as well create remote user interface. User also managed or checked its storage capacity either available or used.

In this Report we also highlight the feature (Scalability, storage management, Network management, user I.D, multiple node,) of different platform like (Amazon, Google, iCloud, Apache Hadoop)

Cloud Storage Features

* **Scalability :** When they store data on-premises, organizations have to forecast their needs far into the future and purchase excess capacity in order to have enough space as their data continues to grow. But with cloud storage, more resources are automatically available as the organization needs them, with no need for over provisioning.
* **Security and File Encryption :** Security is a top concern when you take your business to the cloud. Your company's private data [could be compromised](https://www.businessnewsdaily.com/5215-dangers-cloud-computing.html) if the service you use lacks the right features. Make sure your service encrypts your data at least once before it ever leaves your servers. The best services also allow you to pick your own encryption key so that not even that company's employees can view your files. You should also have the ability to protect your files with a password, particularly when accessing them from a mobile device, as well as the capability to set different levels of access for different employees. Top-notch security features will let you relax when sending private information about your clients to the cloud.
* **Cloud Storage Management Complexity :**

The idea behind cloud storage management complexity is to use computer data storage software for policy-based provisioning and management of data storage independent of the underlying hardware.

Traditional data storage cannot overcome today’s challenges of scale, integration, and flexibility. If your solution for managing data growth is simply to buy more storage capacity, sooner or later you’ll be facing dramatically increased costs for both storage and management. Manually managing across heterogeneous storage systems, silos and clouds is not only error-prone but also leads to administrative overhead.

Software-defined storage addresses these challenges by separating the software that provides the intelligence for storage from the traditional hardware platform. The results include easier storage management, lower storage costs, and anywhere-anytime access to support cloud storage.

**DESIGN AND IMPLEMENTATION OF PRIVATE CLOUD STORAGE PLATFORM**

There are two ways to install Hadoop.

**Single node** and **Multi node**.

**Single Node** cluster means only one Data Node running and setting up all the Name Node, Data Node, Resource Manager and Node Manager on a single machine. This is used for studying and testing purposes. For example, let us consider a sample data set inside a healthcare industry. So, for testing whether the Oozie (Oozie is a server based Workflow Engine specialized in running workflow jobs with actions that run Hadoop Map/Reduce and Pig jobs. Oozie is a Java Web-Application that runs in a Java servlet-container) jobs have scheduled all the processes like collecting, aggregating, storing and processing the data in a proper sequence, we use single node cluster. It can easily and efficiently test the sequential workflow in a smaller environment as compared to large environments which contains terabytes of data distributed across hundreds of machines

.

**Multi node** cluster there are more than one Data Node running and each Data Node is running on different machines. The multi node cluster is practically used in organizations for analyzing Big Data. Considering the above example, in real time when we deal with peta bytes of data, it needs to be distributed across hundreds of machines to be processed. Thus, here we use multi node cluster.

**Prerequisites**

* **VIRTUAL BOX**: it is used for installing the operating system on it.
* **OPERATING SYSTEM**: we can install Hadoop on Linux based operating systems. Ubuntu and CentOS are very commonly used. In this project, we are using Linux.
* **JAVA**: we need to install the Java 8 package on your system.
* **HADOOP**: we need to install Hadoop package.

**Supported Platforms**

* GNU/Linux is supported as a development and production platform. Hadoop has been demonstrated on GNU/Linux clusters with 2000 nodes.
* Windows is also a supported platform but the followings steps are for Linux only. To set up Hadoop on Windows,

**Install Hadoop**

**Step 1:** To download the Java 8 Package. Save this file in your home directory.

**Step 2:** Extract the Java Tar File.

**Step 3:** Download the Hadoop 2.7.3 Package.

**Step 4:** Extract the Hadoop tar File.

**Step 5:** Add the Hadoop and Java paths in the bash file (bashrc).

**Step 6:** Edit the [Hadoop Configuration files](https://www.edureka.co/blog/explaining-hadoop-configuration/).

**Step 7:** Open core-site.xml and edit the property of configuration tag.

**Step 8:** Edit hdfs-site.xml and edit the property of configuration tag.

**Step 9:** Edit the mapred-site.xml file and edit the property of configuration tag.

**Step 10:** Edit *yarn-site.xml* and edit the property of configuration tag.

**Step 11:** Edit*hadoop-env.sh* and add the Java Path as mentioned in final report.

**Step 12:** Go to Hadoop home directory and format the Name Node.

**Step 13:** Once the Name Node is formatted, go to hadoop-2.7.3/sbin directory and start all the daemons.

**Step 14:** To check that all the Hadoop services are up and running.

**Step 15:** Now open the Mozilla browser and go to Local Host for check the URL.