International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 3 Issue 5, May-2014, pp: (448-456), Impact Factor: 1.252, Available online at: www.erpublications.com

Design and Implementation an Private Cloud Cluster for G-Cloud in IRAQ

Sayf Addin Mahmood N[1], Mazin Abed Mohammed[1], Rabah Nory Farhan[1]
[1]Department of Computer Science, College of computer, [1]Planning and Follow Up Department,
[1]Department of Computer Science, College of computer, University of Anbar,
[1]Anbar, Iraq

Abstract: Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (Internet).Cloud computing gets its name as a metaphor for the Internet. It can be categorized as a computing model rather than as technology in itself. In this model "customers" plug into the "cloud" to access IT resources which are priced and provided "on-demand". Delivered over an Internet connection, the "cloud" replaces the company data center or server providing the same service.Thus, Cloud Computing is simply IT services sold and delivered over the Internet.Cloud computing is everywhere. Pick up any IT focused magazine or visit any IT website or blog and talk about cloud computing will be there. However, it is also a fact that not everyone agrees on what it is and how it can be used. Ask ten different professionals what cloud computing is and you will get ten different answers. In this paper, we Design and implement a privet cloud cluster (PCC) at home for use as infrastructure as a Service (IaaS) in G-Cloud or for research in university. Use HQL to test the cluster and use the dataset as a government database.

Keywords: Cloud Computing, Privet Cloud Cluster, Hadoop, Hive, HQL, MapReduce, IaaS, G-Cloud.

Introduction

Cloud computing involves distributed computing over a network, where a program or application may run on many connected computers at the same time. It specifically refers to a computing hardware machine or group of computing hardware machines commonly referred as a server connected through a communication network such as the Internet, an intranet, a local area network (LAN) or wide area network (WAN). Any individual user who has permission to access the server can use the server's processing power to run an application, store data, or perform any other computing task.

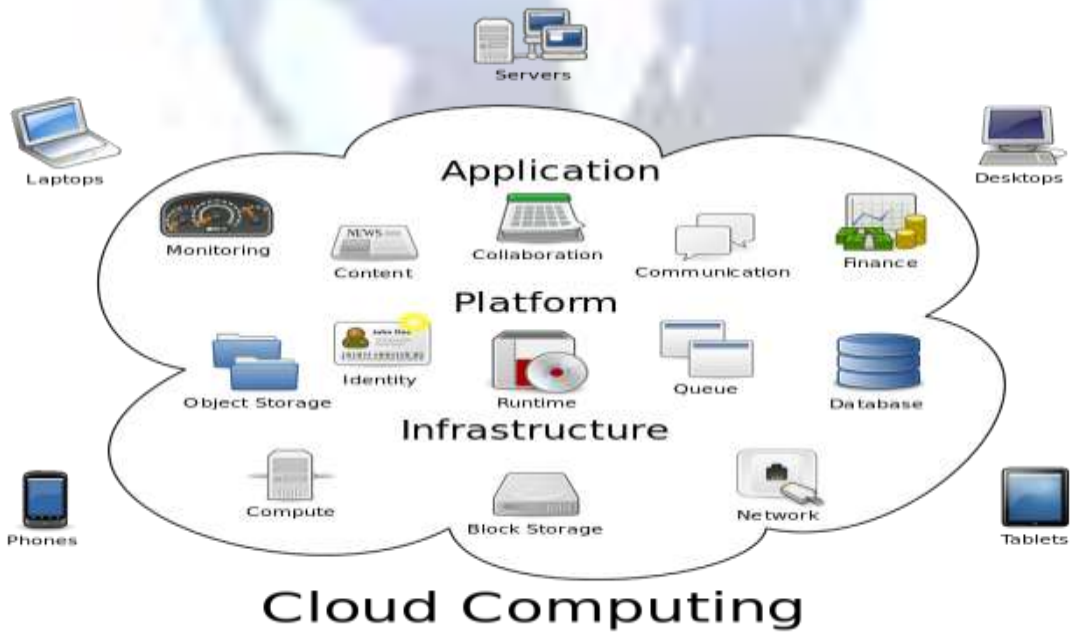


Figure 1 shows cloud computing components

Therefore, instead of using a personal computer every-time to run the application, the individual can now run the application from anywhere in the world, as the server provides the processing power to the application and the server is also connected to a network via internet or other connection platforms to be accessed from anywhere [2]. All this has become possible due to increasing computer processing power available to humankind with decrease in cost as stated in Moore's law. See figure1

Cloud computing gets its name as a metaphor for the Internet. It can be categorized as a computing model rather than as technology in itself. In this model "customers" plug into the "cloud" to access IT resources which are priced and provided "on-demand". Delivered over an Internet connection, the "cloud" replaces the company data center or server providing the same service. (Ross &Beath, 2002).Thus, Cloud Computing is simply IT services sold and delivered over the Internet see figure 2.

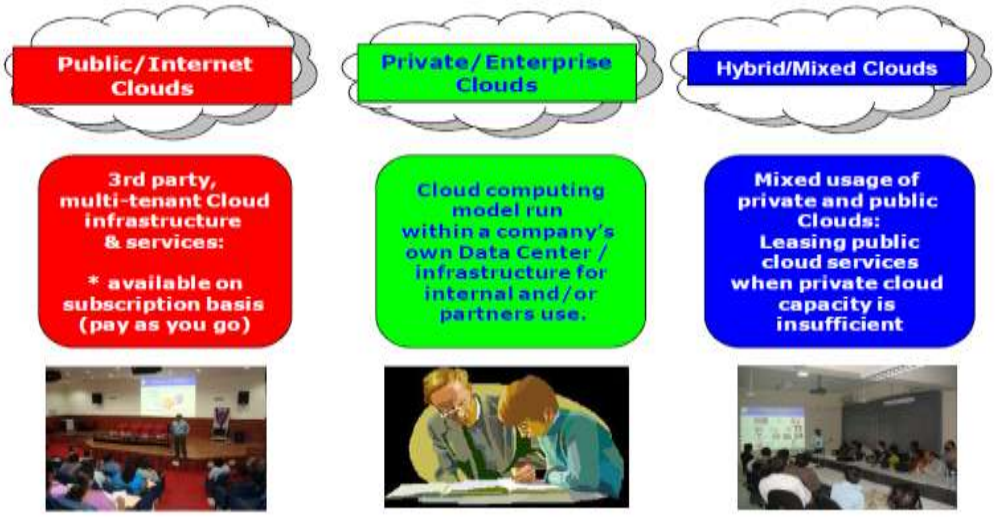


Figure 2 shows cloud computing types

In common usage the term "the cloud" is essentially a metaphor for the Internet.[1] Marketers have further popularized the phrase "in the cloud" to refer to software, platforms and infrastructure that are sold "as a service", i.e. remotely through the Internet. Typically, the seller has actual energy-consuming servers which host products and services from a remote location, so end-users don't have to; they can simply log on to the network without installing anything. The major models of cloud computing service are known as software as a service, platform as a service, and infrastructure as a service. These cloud services may be offered in a public, private or hybrid network.

Cloud computing relies on sharing of resources to achieve coherence and economies of scale, similar to a utility (like the electricity grid) over a network.[4] At the foundation of cloud computing is the broader concept of converged infrastructure and shared services.

The cloud also focuses on maximizing the effectiveness of the shared resources. Cloud resources are usually not only shared by multiple users but are also dynamically reallocated per demand. This can work for allocating resources to users. For example, a cloud computer facility that serves European users during European business hours with a specific application (e.g., email) may reallocate the same resources to serve North American users during North America's business hours with a different application (e.g., a web server). This approach should maximize the use of computing power thus reducing environmental damage as well since less power, air conditioning, rackspace, etc. are required for a variety of functions. With cloud computing, multiple users can access a single server to retrieve and update their data without purchasing licenses for different applications[7].

The term "moving to cloud" also refers to an organization moving away from a traditional CAPEX model (buy the dedicated hardware and depreciate it over a period of time) to the OPEX model (use a shared cloud infrastructure and pay as one uses it)[5].

Proponents claim that cloud computing allows companies to avoid upfront infrastructure costs, and focus on projects that differentiate their businesses instead of infrastructure.[5] Proponents also claim that cloud computing allows enterprises to get their applications up and running faster, with improved manageability and less maintenance, and enables IT to more rapidly adjust resources to meet fluctuating and unpredictable business demand.[5][6][7] Cloud providers typically use a "pay as you go" model. This can lead to unexpectedly high charges if administrators do not adapt to the cloud pricing model.

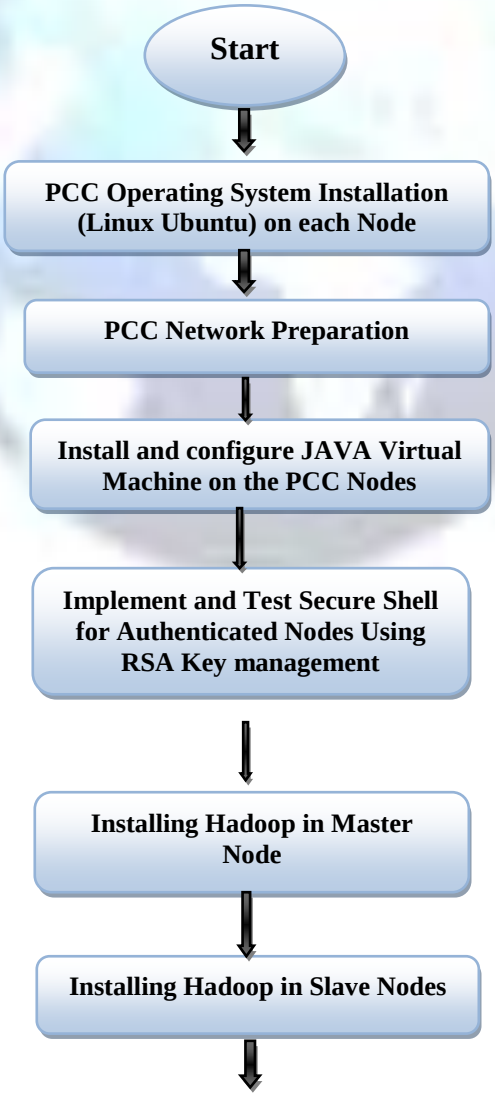
In this paper, we show the characteristics and successive design and implementation of our Private Cloud Cluster (PCC). Since we adopting to use very efficient Cloud platform (Hadoop) which must be installed on Linux Operating system, extensive work procedure using bash shell commands to configure the Hadoop must be undertaken. Under Linux, there are some powerful tools that, for all practical purposes, are unavailable under Windows. One of these tools is something called "shell programming". This means writing code that a command shell executes. There is something like this under Windows but, as usual, the Windows version is a weak imitation.
The most common Linux shell is named "Bash". There are many other shells available. Unless there is a compelling reason not to, we recommend Bash shell for PCC configuration, because this increases the chance that our scripts will be portable between machines, distributions, and even operating systems. The PCC architecture will be shown through multiple stages. The PCC will be initiated with Map Reduce jobs and invoke the HIVE platform for testing HQL queries that is essential for the G-Cloud

Hardware Requirements for Private Cloud Cluster

The PCC essentially consists of determined number of computers connected via LAN. of One Master Server which plays the role of cloud manager since it controls the execution of the PCC. The remaining computers will be configured as slave nodes. Since the installed cluster must not alter the operating system of these computers, VMware workstation was installed on these computers and initiated as virtual machines.

PCC Architecture

Our proposed cluster, as shown in the cluster description, consists of various nodes installed on VMware workstations configured for Local Area Network connections. This cluster represents the backbone of our PCC. PCC architecture design consists of multiple stages built in successive manner due to the fact that each stage depends on previous stages. These stages are outlined in the figure (1).



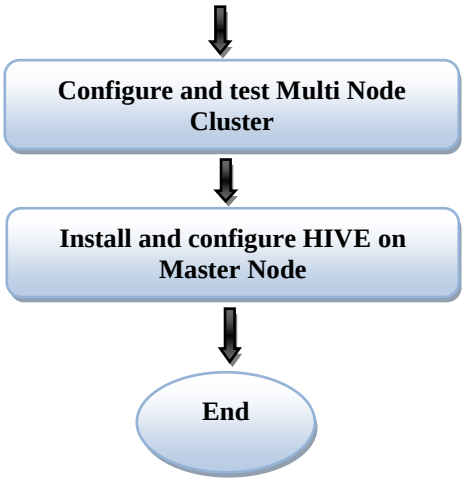


Figure 3: PCC Architecture

A. **PCC Operating System Installation**

Ubuntu is the world's favorite free operating system, with more than 20 million people preferring it to commercial alternatives. It's open source and free. There are also special editions for servers. All editions share common infrastructure and software, making Ubuntu a unique single platform that scales from consumer electronics to the desktop and up into the cloud for enterprise computing.We can Install Ubuntu by download it from Ubuntu download site and install it using VMware as virtual machine.

B. **PCC Network Preparation**

Ubuntu is shipped with a number of graphical utilities to configure our network devices. This document is geared toward server administrators and will focus on managing wired network on the command line and GUI tools. The following steps show sample of our Ubuntu network configurations:

1. Open Network Connections to set up network settings in Ubuntu as shown in Figure 2. Go to "System", "Preferences" and chose "Network Connections.".

Figure 4: System", "Preferences" and chose "Network Connections

2. Under the "Wired" tab in Figure 3 click on "Auto eth0" and select "Edit."

Figure 5: Auto eth0" and select "Edit

3. Click on IPV4 Settings tab shown in Figure 4. Chose the Manual option from the Method drop-down list and assign the IP address, net mask and gateway manually. You'll also have to assign the DNS Server address For internet if you want to update our system and Click Apply.

Figure 6: the Manual option from the Method drop-down list and assign the IP address

4. For 3 Nodes Cluster we assign
   IP for server (Master node) =192.168.0.20
   IP for node1 (Slave1)=192.168.0.1
   IP for node2 (Slave2)=192.168.0.2
   For 4 Nodes Cluster we assign
   IP for server (Master node) =192.168.0.20
   IP for node1 (Slave1)=192.168.0.1
   IP for node2 (Slave2)=192.168.0.2
   IP for node3 (Slave3)=192.168.0.3
   For 5 Nodes Cluster we assign
   IP for server (Master node) =192.168.0.20
   IP for node1 (Slave1)=192.168.0.1
   IP for node2 (Slave2)=192.168.0.2
   IP for node3 (Slave3)=192.168.0.3
   IP for node4 (Slave4)=192.168.0.4

Install and configure JAVA Virtual Machine on the PCC Nodes

You must add store support Andrew@Webupd8 for Ubuntu repository so that it could download Java 7, in terminal windows:

sudo add-apt-repository ppa:webupd8team/java

We must update all store types before installing

sudo apt-get update

Installing java JDK

sudo apt-get install oracle-java7-installer

Note that we must connect the node to the internet for download and update its programs or application and it take some time.

Implement Secure-- Shell for Cluster

Hadoop uses SSH access to manage its nodes. That mean Hadoop remote machines and our local machine must have SSH on it.In terminal windows use a command that generate an ssh key

ssh-keygen –t rsa –p ""

To enable SSH access to our local machine with this newly created key use

cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys

We can test the SSH setup by connecting to our local machine and use

sshlocalhost

The step is also needed to save our local machine's host key fingerprint to the user's known hosts file and the output for last command like :

The authenticity of host' localhost (192.168.0.20)' can't be established.
RSA key fingerprint is 3b:21:b3:c0:21:5c:7c:54:2f:1e:2d:96:79:eb:7f:95.
Are you sure you want to cntinue connecting (yes/no)? yes
Warning: Permanently added ' localhostl' (RSA) to the list of known hosts.
Linux ubuntu 2.6.20-16-386 #2 Thu Jun 7 20:16:13 UTC 2013 i686

Configuring of Hadoop for Multi Node

First copy a last single node to another VMware. We can copy N node to configure it as Slave 1, Slave 2... Slave n, and keep the first node as Master node, now follow the general step to configure the master node and slave1, slave N node, For Multi Node cluster.Do not forget the process of Copy the public key file from master to the slave1, slave2,slave n and we can test our system to connect with SSH or not using ssh master command:

ssh master

and we can see the connection to master node and add it to the list of known hosts the output like:

The authenticity of host 'master (192.168.0.20)' can't be established.

RSA key fingerprint is 3b:21:b3:c0:21:5c:7c:54:2f:1e:2d:96:79:eb:7f:95.

Are you sure you want to continue connecting (yes/no)? yes

Warning: Permanently added 'master' (RSA) to the list of known hosts.

Linux master 2.6.20-16-386 #2 Thu Jun 7 20:16:13 UTC 2013 i686

Installing And Configuring Hive

On MASTER node download the most recent stable release of Hive as a tarball from the apache site and unpack the tarball in the directory of your choice and Set the environment variable HIVE_HOME to point to the installation directory using the following command

$ tar -xzvf hive-0.9.0.tar.gz
export HIVE_HOME='/Users/Work/hive-0.9.0'
export PATH=$HADOOP_HOME/bin:$HIVE_HOME/bin:$PATH

Why Apache for Hadoop?

- Open source platform for reliable, scalable, distributed computing.
- Provides MapReduce framework for distributed processing.
- Provides a distributed file system for high-throughput data access (HDFS).

Map Reduce

By providing a data-parallel programming model, MapReduce can control job execution in useful ways:

- Automatic division of job into tasks
- Automatic placement of computation near data
- Automatic load balancing
- Recovery from failures & stragglers

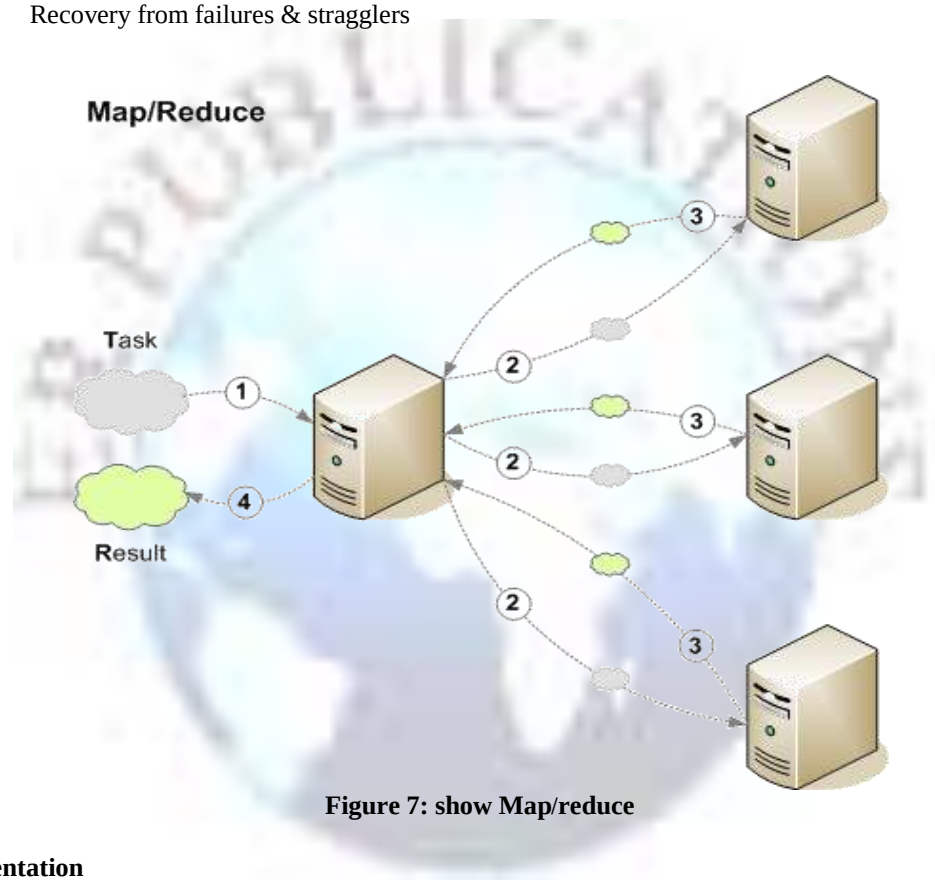


Figure 7: show Map/reduce

HQL Implementation

Create Table Query (CTQ)

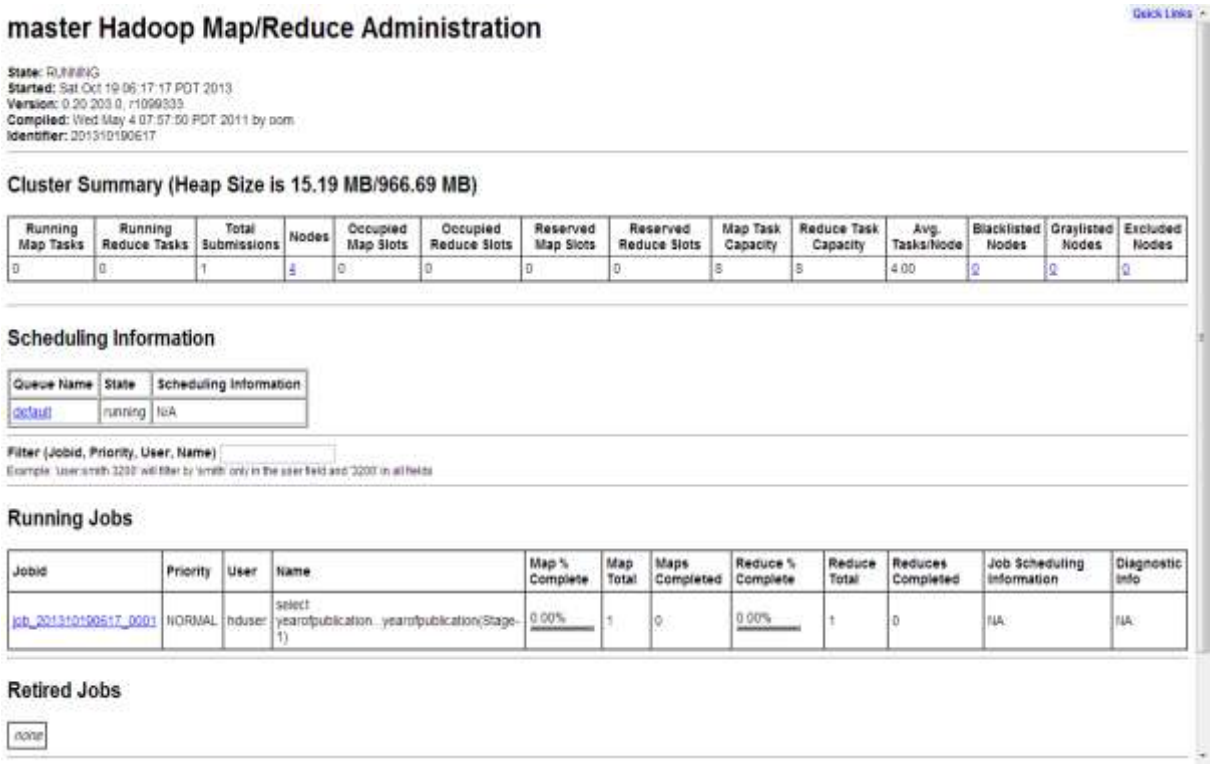
hive> CREATE TABLE IF NOT EXISTS tablename
(fields of table,)
> STORED AS TEXTFILE;

Load Data to Table Query (LDTQ)

hive> LOAD DATA INPATH 'path' OVERWRITE INTO TABLE tablename;

Select with Count Query (SCQ)

***hive> select** field1, **count**(field2) **from**tablename**group by** field1;*



master Hadoop Map/Reduce Administration

State: RUNNING
Started: Sat Oct 19 06:17:17 PDT 2013
Version: 0.20.203.0, r1099333
Compiled: Wed May 4 07:57:50 PDT 2011 by oom
Identifier: 201310190617

Cluster Summary (Heap Size is 15.19 MB/966.69 MB)

Running Map Tasks	Running Reduce Tasks	Total Submissions	Nodes	Occupied Map Slots	Occupied Reduce Slots	Reserved Map Slots	Reserved Reduce Slots	Map Task Capacity	Reduce Task Capacity	Avg. Tasks/Node	Blacklisted Nodes	Graylisted Nodes	Excluded Nodes
0	0	1	4	0	0	0	0	8	8	4.00	0	0	0

Scheduling Information

Queue Name	State	Scheduling Information
default	running	N/A

Filter (Jobid, Priority, User, Name)

Running Jobs

Jobid	Priority	User	Name	Map % Complete	Map Total	Maps Completed	Reduce % Complete	Reduce Total	Reduces Completed	Job Scheduling Information	Diagnostic Info
job_201310190617_0001	NORMAL	hduser	select yearofpublication...yearofpublication(Stage-1)	0.00%	1	0	0.00%	1	0	NA	NA

Retired Jobs

none

This cluster for $ nodes

Also, this cluster for five nodes

master Hadoop Map/Reduce Administration

State: RUNNING
Started: Fri Oct 18 10:19:49 PDT 2013
Version: 0.20.203.0, r1099333
Compiled: Wed May 4 07:57:50 PDT 2011 by oom
Identifier: 201310181019

Cluster Summary (Heap Size is 15.19 MB/966.69 MB)

Running Map Tasks	Running Reduce Tasks	Total Submissions	Nodes	Occupied Map Slots	Occupied Reduce Slots	Reserved Map Slots	Reserved Reduce Slots	Map Task Capacity	Reduce Task Capacity	Avg. Tasks/Node	Blacklisted Nodes	Graylisted Nodes	Excluded Nodes
0	0	1	5	0	0	0	0	10	10	4.00	0	0	0

Scheduling Information

Queue Name	State	Scheduling Information
default	running	N/A

Filter (Jobid, Priority, User, Name)

Running Jobs

Jobid	Priority	User	Name	Map % Complete	Map Total	Maps Completed	Reduce % Complete	Reduce Total	Reduces Completed	Job Scheduling Information	Diagnostic Info
job_201310181019_0001	NORMAL	hduser	select yearofpublication...yearofpublication(Stage-1)	100.00%	1	1	0.00%	1	0	NA	NA

Retired Jobs

none

Local Logs

Log directory, Job Tracker History

This is Apache Hadoop release 0.20.203.0

CONCLUSIONS

The focus of the paper is on analysis the investment made on adoption of cloud computing. The analysis covers both financial as well as non-financial benefits and associated costs. The same has been carried out based on a analysis framework. It also highlights the need to carry out a comprehensive risk analysis based on a risk framework taking into account various risks e.g. vendor related, technological, architectural, regulatory etc. The criticality of change management initiative for any such transition is also dealt with. This has been evaluated from the perspective of arriving at the impact of organizational change. To Accessfor thesystem with high security it's being managed by the administrators. Thedatabasesmust be under the managers to be in government servers, not in a third party.Our PCC (Infrastructure as a service) It is natural to be expensive, but safety is required in G-Cloud,Add the cost of maintenance as well as.It can be used in research centers and universities.No matter how high the costs borne by the government but in order to ensure confidentiality of data.

REFERENCES

[1]. Eric Sammer , “Hadoop Operations”, O’Reilly Media, ISBN: 978-1-449-32705-7, 2012.
[2]. Gurudatt Kulkarni, Niraj Patil, Pradip Patil, “Private Cloud Secure Computing”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012. (IVLS)
[3]. K.Mukherjee, G.Sahoo, “Cloud Computing: Future Framework for e-Governance”, International Journal of Computer Applications (0975 – 8887) Volume 7– No.7, October 2010
[4]. Marc Vael, “Cloud Computing An insight in the Governance & Security aspects”, Across Technology, 2010.
[5]. Saleh Alshomrani and Shahzad Qamar, “Cloud Based E-Government: Benefits and Challenges”, International Journal Of Multidisciplinary Sciences And Engineering, Vol. 4, No. 6, July 2013.
[6]. Sander van Vugt "Pro Ubuntu Server Administration", A press, 2009..
[7]. White Tom, "Hadoop: The Definitive Guide", Third Edition, O’Reilly Media ISBN: 978-1-449-31152-0, 2012.

AUTHOR’S PROFILE

Mr. Mazin Abed Mohammed: He received his Master of IT degree in 2011 and Bachelor in 2007. He is currently working as a Teacher Assistance in department of Planning & Follow Up in University of Anbar, Ramadi, Iraq.He is having keen interest in AI, Operating Systems, image processing and HCI. He is carrying out research work in the field of Image Processing.