Journal Of Harmonized Research (JOHR)

Journal Of Harmonized Research in Engineering
2(1), 2014, 62-68

ISSN 2347 – 7393

Original Research Article

CREATION OF A PRIVATE CLOUD FOR COLLEGE USING COMMODITY HARDWARE

Karan Singh Hundal*, Puneet Madhavan,Darshan Oswal.

MES College of Engineering, Pune-411007.

Abstract- This paper discusses the research project which we aim to implement for our college. It provides an introduction to cloud, clustering technologies along with the elements required to build these technologies, their advantages and the reasons behind implementing these technologies together. We discuss how existing hardware in the college can be used more effectively and the various systems that resemble our idea and have been implemented at the college level. We also shed light on the challenges the system presents and the factors upon which its effectiveness depends upon along with the steps we plan to undertake to implement it. The paper concludes with the advantages that this system presents for our college and the future scope of the same.

Key terms- Cloud computing, cluster computing.

Introduction to Cloud Computing and Its Advantages

Cloud Computing:

Could computing distributes computation task on the resource Pool, that consists of large computing devices, so the application systems can obtain the computation power, the storage space and software service according to its demand. Such a resource pool is called "cloud". The Clouds are the virtual computation resources usually large-scale server clusters. Cloud computing can put all the computing resources and manage them automatically through the software without intervene.

Cloud Computing provides resource sharing in terms of scalable hardware, middleware services and platforms for development of various applications ,and value-added business applications.

Cloud Computing Characteristics:

The resources which are provided to the users in form of service through the Internet. Examples:Amazon EC2,Google app engine. The resources provided to the users can be dynamically configured and extended by the users based on the requirements.These resources are presented as a single entity although in reality they might be distributed

For Correspondence:
karan_hundal2005ATyahoo.com
Received on: February 2014
Accepted after revision: February 2014

or shared. Examples:IBM has 8 data centres to serve its research institutes around the world.

Cloud computing Service forms:

SaaS (Software As a Service):

SaaS provider dispose the applied software on their server from where the user can subscribe this software service from the manufacturer through Internet and are charged according to the quantity of software and using time Example: Google Doc, Google Apps and Zoho Office.

PaaS (Platform As a Service):

It is a kind of distribution platform server, the manufacturers supply service to the users, such as a develop environment, server platform and hardware resources, and the users customize and develop their own application and transfer to other customers through their server and Internet. Example: Google App Engine

IaaS(Infrastructure As a Service):

IaaS takes infrastructure which is made of many servers as a measurement service to the customers. It integrates memory and I/O devices, storage and computing ability into a virtual resources pool, and provides storage resources and virtualization service for the whole industry. Example: Google Cloud.

Advantages of Cloud Computing

-Cost Effective

Cloud computing is probably the most cost efficient method to use. It is also efficient in terms of maintenance and upgradation. Cloud's Pay-as-you-go alongwith other scalable options makes it very reasonable for the company in question.

-High Storage

Storing information in the cloud gives us very high storage capacity. Hence, we no more need to worry about running out of storage space or increasing your current storage space availability.

-Backup and Recovery

Since all our data is stored in the cloud, restoring and backing it up is relatively much easier than storing the same on a physical device. Most cloud service providers are usually competent enough to handle recovery of information and thus the entire process of backup and recovery is much easy and efficient than other traditional methods of data storage.

-Easy Access to Information

Once we register ourselves in the cloud, we can access the information from anywhere, wherever there is an existing Internet connection. This convenient feature lets us free to move beyond time zone and geographic location contraints

Introduction to Cluster Computing

Computer cluster refers to a group of computers, linked together closely such that it works like a single big computer system. It provides the benefits of improved performance and/or availability compared to a single powerful computer while being relatively more cost-effective than single computers of comparable speed or availability therefore makes a better return on existing investments in computing technology by an organization. It also establishes more control over computing workloads and depending on the method of implementation tries to push important jobs to the head of the line for reliable processing and therefore improving overall computing reliability and serviceability.

Clusters of computers are now filling the processing capacity that was once occupied by powerful stand-alone machines. Commodity clusters are becoming so popular mainly because these systems provide a way to reuse seemingly obsolete systems while at the same time high performance clusters can easily compete with the best supercomputers IBM or SGI have to offer.

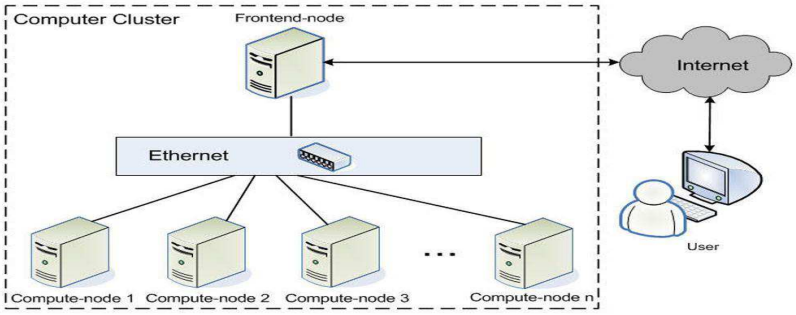

Fig 1. A Cluster Of Computers [IMAGE]

Advantages of Cluster Computing:

-Easy to deploy: The cluster computing system is very simple to deploy. In this system software is installed and also configured automatically. Using the web interface, the cluster nodes can be easily added, removed, managed and thus, reduces efforts and saves time.

-Flexible: Cluster computing being an open system, it is very flexible. It supports real-world topologies and synchronizes the cluster files without re-installation.

-Expandable: It is very east to add new, future hardware models and cluster node at any time as per the need.

-Optimized: Optimization is an important advantage of cluster computing system. The system is optimized not only for performance but also for its simplicity.

Motivation for creating a private cloud platform using Commodity hardware

General implementations of cloud include the clients hosting their services on a third party data centre and just using their client devices with web access to use them. In our implementation of the cloud, we make use of the commodity hardware that is present in our college as the required infrastructure(which mainly consists of a cluster of old unused PC's). Several processes that run on a cloud are implemented on a cluster of computers using parallel programming. The cluster also acts as a cloud server. Hence the advantages of clusters namely-faster processing, availability and load balancing are put into use as these are the necessary provisions required in a cloud infrastructure

A **computer cluster** consists of a grouped set of loosely or tightly connected computers that work together so that in many respects they can be viewed as a single system which is exactly what is required in a cloud implementation.The components of a cluster are usually connected to each other through fast local area networks("LAN") but not always, with each *node*(computer used as a server) running its own instance of an operating system (OS) . Clusters have emerged as a result of convergence of a number of computing trends including the availability of cost efficient microprocessors, software for high performance distributed computing and high speed networks. Clusters are deployed in order to improve availability and performance over that of a single computer, while typically being much more cost-effective than single computers of comparable speed or availability and hence the existing resources in our college can be used more effectively and this is also cost effective.

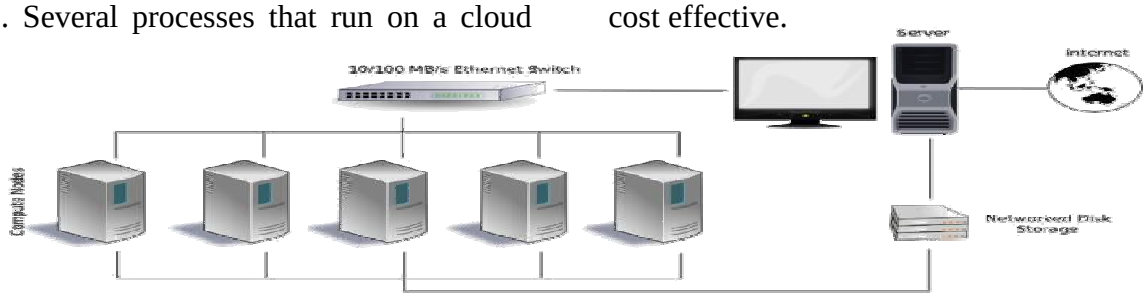

Fig 2. Cluster of inter-connected computers.

PRIVATE CLOUD provides the advantage of total access to our college. It can create a vast array of applications using our system and have full rights on its technology without any billing costs involved. They can provide facilities specifically for our college or even other colleges and would therefore could really start a revolution in our current education system . Cloud computing is one of the emerging paradigm which has the capability of taking the concept of computing to another level by introducing a platform for resource as well as data sharing over the internet.

Problem Description

In our creation of **The Private Cloud** we have to take care of the following points[5]:

- Proper dynamic provisioning of virtual machines.
- Load failure handling capabilities.
- High availability of resources when required.
- Proper file management system
- Proper division of work among the resources
- Multiple query handling capacity
- Scalable system,
- Easier access
- Ability to host services.

Operating Environment

The operating environment for our project is selected in accordance with the software tools that are going to be used in the cloud. These tools are compatible with certain OS. Therefore our operating environment is decided on the grounds of compatibility with our software tool. The main foundation software tool for our architecture is OpenStack, which is compatible with popular OS like Ubuntu, Fedora, and CentOS. So we have selected Fedora as the OS for all our nodes and server.

RDO is a community of people using and deploying OpenStack on Red Hat Enterprise Linux, Fedora. OpenStack relies on the underlying OS and hypervisor.

The Hypervisors that easily integrate with OpenStack are XEN and KVM. Considering the latest survey on Internet it is found that KVM is the most popular and widely used and readily integrates with the Nova (Compute) component of the Open Stack.

For the GUI, OpenStack itself provides a dashboard for monitoring all the functions and processes with the system. The dashboard is just one way to interact with Open Stack resources. It also provides am ability to automate access or build tools to manage their resources using the native Open Stack API or the EC2 compatibility API.

For the backend of our system, that is to manage the large data and information processing, Hadoop File Distribution System is used. Hadoop is an open distributed software framework that enables programmers to run an enormous number of nodes handling terabytes of data. One of its most significant abilities is allowing a system to continue to operate even if a significant number of nodes fail.

Since Hadoop is an independent tool and needs to work and integrate with the other tools in our system there is a need for a tool which collaborates the functionalities between Open Stack and Hadoop. This task is performed by the Savanna tool which is also an open source tool released in the market which efficiently handles the data communication between Open Stack and Hadoop. Savanna is closely associated with Hadoop.

Savanna project was started by Mirantis with contributions from Red Hat and is now available with their latest stable version 0.2 released for use. The Savanna project is designed to provide users with a simple means to provision a Hadoop cluster on Open Stack by specifying just a few parameters, such as the Hadoop version, the cluster topology, and node hardware details and thus we get a working Hadoop cluster to use with Open Stack as its infrastructure.

Implementation Plan

Install server Operating System

Fedora OS is chosen over other Linux OS as it provides ready-to-run cloud images for each supported Fedora release. The images are in compressed raw image format and for

immediate use with Open Stack. These are available in Amazon EC2 for direct download.

These images are for use with the local private cloud and allow us to modify and upload to a public cloud. The qcow2 images can be used directly by Open Stack

The latest release of Fedora is Fedora 20.

Adopt a machine virtualization strategy

Fedora uses the libvirt family of tools as its virtualization solution. By default libvirt on Fedora will use Qemu (Quick EMUlator) to run guest instances. Qemu can emulate a host machine in software, or given a CPU with hardware support can use KVM to provide a fast full virtualization.

However Open Stack also has the ability to provide virtualization solution. Open Stack consists of number of services for running IaaS clouds. They are the Object Store (Swift), Compute (Nova) and Image (Glance) services.

Add the Cloud solution provider on top of virtualization

The basic requirements for Open Stack on Fedora are at least 2GB RAM, processors with hardware virtualization extensions and at least one network adapter. There are options to install the components of Open Stack either separately or all-in-one go.

The Packstack installer makes installing Open Stack really easy. By using the –allinone option you can have a working self-contained RDO installation in minutes.

- Add a file management system to handle the large amounts of data

First we install Hadoop on a single node cluster and then on a multi node cluster. Hadoop can be downloaded from the Apache releases. The current stable version is 1.2.X.

Hadoop requires a working Java installation. Java SE Development Kit can be easily obtained. The next step involves configuring the Hadoop. The drawback of using Hadoop is that it is not compatible with IPv6 addresses but Fedora provides us with the option to disable IPv6 in local network.

- Add an interface tool between Cloud provider software and File Management System. The Savanna tool provides flexibility and compatibility between Open Stack and Hadoop. The latest version of Savanna tool is 0.2 and is open source. Savanna provides multiple plugins of Open Stack and Hadoop which readily integrates and is available for immediate use.
- Setting up hardware and testing the connection

This step involves setting up of the hardware architecture of the system. This includes connecting data centers to servers via a central hub and other connecting tools. Further testing of these connections is performed. After all the nodes are set up the working of the architecture will be checked by performing a dummy test.

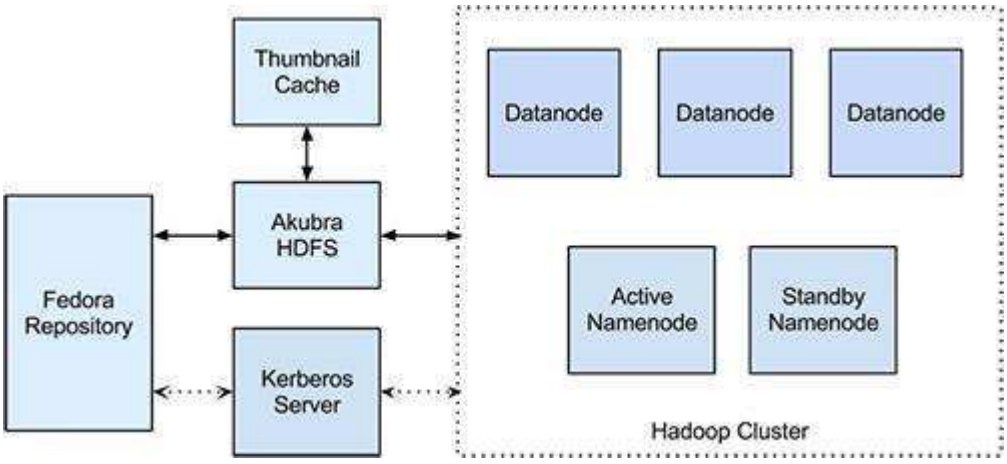

Fig 3. A typical Hadoop cluster on Fedora

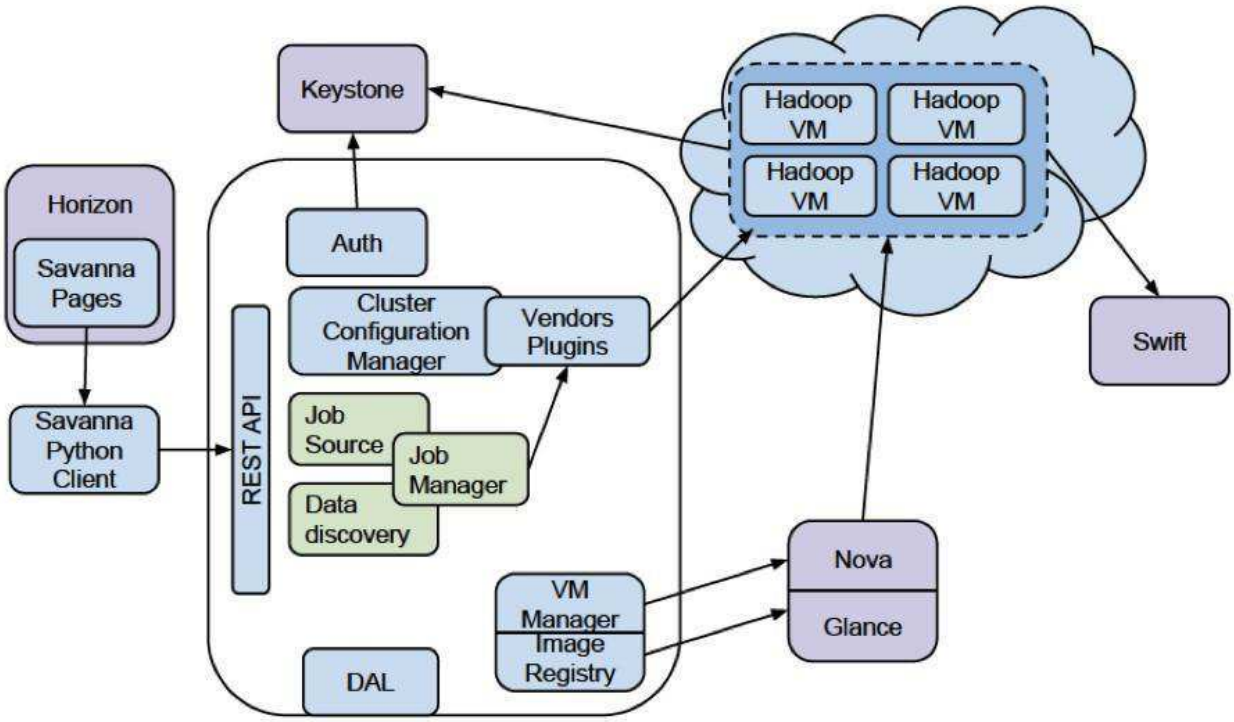

Fig 4. Havanna, openstack and Hadoop collaboration

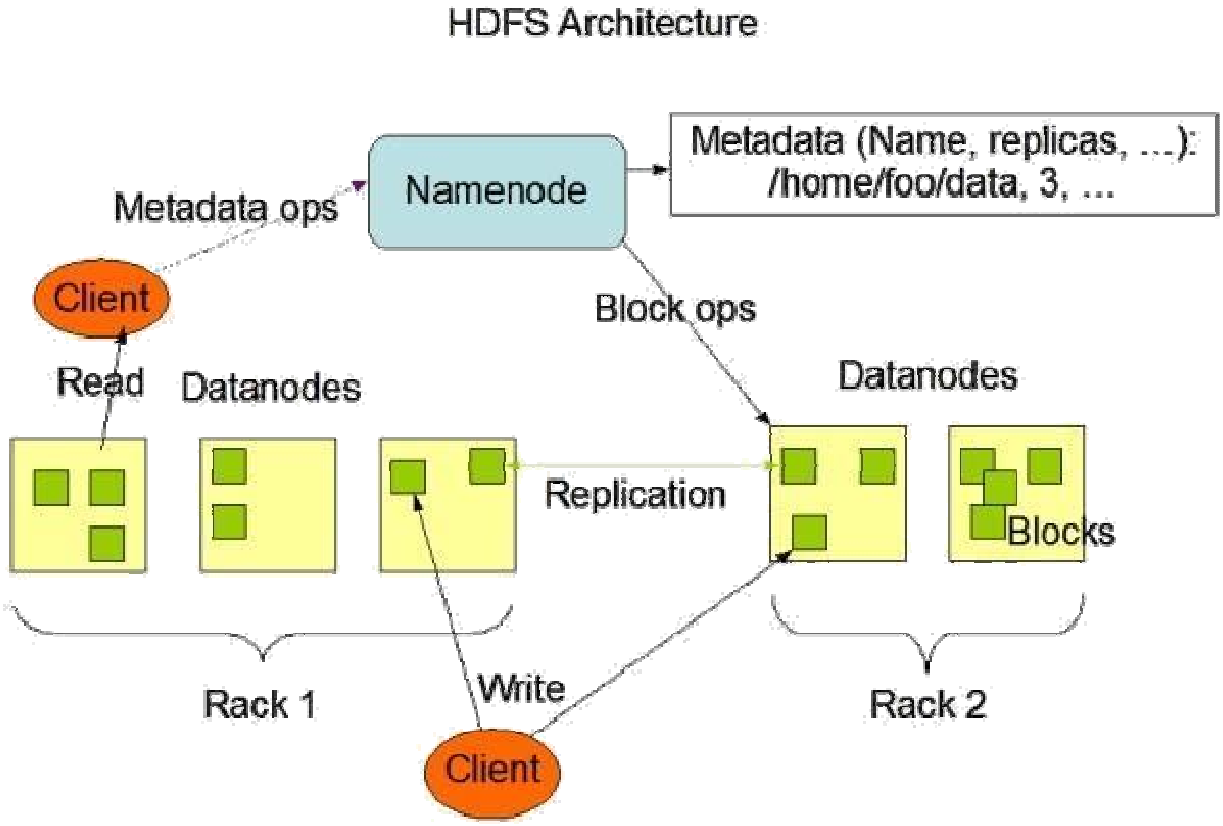

Fig 5. A Hadoop Cluster

Conclusions and Future Work

Advantages of creating this system

1. Utilization of commodity hardware that was otherwise assumed to be unessential.
2. An alternative way for private organizations to implement their private cloud over which they will have total control.
3. Acts as a research project for our educational institute and provides a better way by which students can understand the cluster and cloud computing paradigms.
4. The system is highly flexible. By adding additional nodes, a powerful high performance system can be created.
5. Our college has total control over the system and can use it to deploy useful services for the students.
6. We understand better how cloud works, parallel processing and service hosting.

Cloud computing is an upcoming computing paradigm but it is yet to realize its full capabilities in the Indian Education system. Our project is an effort in the same direction. It sheds light on computing and clustering technologies along with their collaboration to realize our system and will help us to gain more insights in these fields.

Future Work

We can extend the scope of this project by providing a better security to our private cloud as it is of vital importance in cloud computing since it handles the organization's sensitive data. Also an effort can be made to create better services that can be deployed on the cloud and are in turn beneficial for our college.

References

1. Research on Key Technologies of Cloud Computing, Shufen Zhang,Hongcan Yan ,Xuebin Chen
2. *Hebei United University NO.46 Xinhua West Street, Tangshan 063009, Hebei Province, China*
3. Cluster Based Cloud Computing , Ankit Bahuguna, C.L. Yogi, and Dr. Abhijit Kulshreshtha INSTITUTE OF TECHNOLOGY, NIRMA UNIVERSITY, AHMEDABAD – 382 481, 9-11 DECEMBER, 2010
4. Cluster, Grid and Cloud Computing: A Detailed Comparison , Naidila Sadashiv ,Dept. of Computer Science and Engineering,Acharya Institute of Technology, Bangalore, India and S. M Dilip Kumar
5. Dept. of Computer Science and Engineering, University Visvesvaraya College of Engineering (UVCE) Bangalore, India.
6. Cloud Computing [URL] http://en.wikipedia.org/wiki/Cloud_Computing, accessed on Jan. 2014.
7. [5].Bo Peng, Bin Cui and Xiaoming Li - *Implementation issues of a cloud computing platform*; Bulletin of the IEEE Computer Society Technical Committee on Data Engineering IEEE - 2009.