[Volume 1 issue 11 Dec 2013]

Page No.303-305 ISSN :2320-7167

INTERNATIONAL JOURNAL OF MATHEMATICS AND COMPUTER RESEARCH

A Survey on Cloud Computing

Hemamalini. B. H.[1], Dr. Suresh L.[2], Dr. Radhika K.R.[3]

1. *Associate Professor, Dept. of CSE, BMSIT, VTU, Bangalore-India,*
2. Principal, CITech, VTU, Bangalore-India,
3. Professor, Dept. of ISE, BMSCE, Bangalore-India
bhhemaraj@gmail.com, suriakls@gmail.com, rkr@bmsce.ac.in

Abstract

Cloud is a good concept of using the existing resources efficiently, but still there is lot of room for improving the efficiency. The present work will be focused on Amazon web service namely EMR service. Today the user has to pay for 1 hour even he/ she uses the service for just 20 minutes. Similarly, if the user uses the service for one hour and ten minutes, he pays for two hours. A company who does extensive data mining, running 1000 instances of data nodes, running couple of thousand jobs for a week, will be paying a huge sum of money. The proposed tool or framework will help to reuse the existing clusters, without shutting down; saving bootstrap time & making sure the jobs are queued properly for execution.

Keywords:
Cloud computing, IaaS, PaaS, SaaS, query service, big data.

1. Introduction

Cloud computing is computing on the Internet. Like traditional computing, the user need not buy the required software, upgrade, install, maintain etc. If the user has a computer with Internet connection, it is enough. The user can access any software which is available on cloud. Cloud computing has many advantages like reduced upfront investment, infinite scalability, expected performance, high availability, and excellent fault tolerance capacity [1]. Because of these features, the big giants in the industry like Google, Amazon, Salesforce and Microsoft have adopted cloud computing. The user can rent the required service for a specific period.

2. Classification:

Cloud computing can be classified into 3 types based on the deployment:

2.1 Public cloud:

In public cloud, the service provider makes resources, such as applications and storage, available to the general public over the Internet. It may be free or offered on a pay-per-usage model.

The advantages of using a public cloud service are:

- Easy and inexpensive set-up because hardware, application and bandwidth costs are covered by the provider.
- Scalability to meet the needs of the user.
- Resources are not wasted because you pay for what you use.

2.2 Private Cloud

- Private cloud (also called internal cloud or corporate cloud) is a marketing term for a proprietary computing architecture which provides hosted services to a limited number of people behind a firewall.
- The corporate network and data center administrators have become service providers due to the advances in virtualization and distributed computing that meet the needs of their "customers" within the corporation.

2.3 Hybrid cloud

A hybrid cloud is a cloud computing environment in which an organization provides and manages some resources in-house and has others provided externally. For example, an organization might use a public cloud service, such as Amazon Simple Storage Service (Amazon S3) for archived data but continue to maintain in-house storage for operational customer data.

The hybrid approach allows a business to make use of the scalability and cost-effectiveness that a public cloud computing environment offers without exposing mission-critical applications and data to third-party vulnerabilities. This type of hybrid cloud is also referred to as hybrid IT.

To be effective, a management strategy for hybrid cloud deployment should address configuration management, security, change control, fault management and budgeting. Because a hybrid cloud combines public cloud and private data center principles, it's possible to plan a hybrid cloud deployment from either of these starting points. Picking the better starting point, however, will make it easier to address business goals.

3.Services

The cloud computing can be classified into three main categories based on the service that it provides.

3.1 IaaS (Infrastructure as a Service)

In IaaS, the vendor provides the infrastructure itself as a service. The infrastructure can be provided in the form of technology, data centre or IT service, such as offering CPU time on a per hour basis, assessing for storage usage, as well as assessing for data transfers per gigabyte, often with differing rates for uploads versus downloads. Amazon's Elastic Computing Cloud (EC2) is a good example. It is similar to traditional "outsourcing" at less cost and effort. The developer can use the infrastructure provided by the vendor for paying for it for a specified period, thus saving the enormous investment on setting up the infrastructure, and also saving maintenance cost. Providers: Amazon's S3, Sun's cloud service [2].
NIST provides on-demand self-service, resource pooling and rapid elasticity. On-demand means that the service is available to turn on or off as needed. Resource pooling means that multiple users share a bank of servers (including storage devices and other computing resources) over the Internet, as an alternative to using dedicated servers. Rapid elasticity means the cloud offering can be dramatically scaled up and down as needed. With as-a-service, you only pay for what you use, and you can use as much as you want.

3.2 PaaS (Platform as a Service)

In this, the development platform itself is provided as a service. The developer uses the platform provided by the cloud. The platform will be hosted on the cloud. The platform is accessed by the user using a browser with internet connection and the user creates and runs his application on the platform. Providers: Microsoft's Azure, Google App Engine [2].

3.3 SaaS (Software as a Service)

In traditional computing, users have to buy the licensed version of the software and install it for use. In cloud computing, the software is provided as a service on the basis of pay-per-use model. It provides multi-tenant means the at the backend, the same infrastructure is shared amongst multiple users while in the front end, each user feels that the software is dedicated to a single user. IT supports running multiple instances of the software too. Providers: Google Docs, Zoho [2].

4 Proposed Work:

The proposed system will provide efficient service mechanisms which are evaluated for performance.

Query Service: In big data world searching or running query on a terabyte of data is a huge problem; Today Google does a search on peta bytes of data. In case of Google they need to store data permanently & they are not worried about the cost, as they already have a good operations team to maintain the infrastructure. This is not the case with most of the companies. Companies are getting into cloud, because they have less overhead of operations & it is cost efficient.

If the data is not stored permanently & if the data growth is unpredicted, the existing query services like Hive, won't scale. It takes about days to query & get the results from a 100 terabyte data sitting in Apache HDFS (Hard Disk File System).

When it comes to cloud, keeping a 100 terabyte is very costly. People like data scientists bring the data they want to analyze only when required & query on them. The tools & solutions provided will be used by Data Scientists. They query the data & extracts the information from it & say these data points indicate something, which help the companies to improve their business.

Cloud is a good concept of using the existing resources efficiently, but still there is lot of room for improving the efficiency. The present work will be focused on Amazon web service namely EMR service. Today the user has to pay for 1 hour even he/ she uses the service for just 20 minutes. Similarly, if the user uses the service for one hour and ten minutes, he pays for two hours. A company who does extensive data mining, running 1000 instances of data nodes, running couple of thousand jobs for a week, will be paying a huge sum of money.

The proposed tool or framework will help to reuse the existing clusters, without shutting down, saving bootstrap time & making sure the jobs are queued properly for execution.

For this a model has to be developed which can say this job will run for certain time. Depending on that it will choose which cluster to run the job.

5 Conclusion

Cloud Computing and Big data is a buzzword these days. More and more companies start to step into Cloud and provide services for public use, because of the nature of Cloud Systems, i.e. reduced upfront cost, high availability, infinite scalability, expected performance, tremendous fault tolerance capability and so on. The present work will be focused on Amazon web service namely EMR service. A company who does extensive data mining, running 1000 instances of data nodes, running couple of thousand jobs for a week, will be paying a huge sum of money.

6 Acknowledgement:

We would like to acknowledge our honourable Chief Mentor and Professor of Eminence Dr. S. C. Sharma, our Principal Dr. S. Venkateswaran, our Head of the department Dr. Thippeswamy. G and my collegue Dr. Keshav Prasanna who has supported in bringing out this paper.

7 References:

1. Minqi Zhou, Rong Zhang, DadanZeng, WeiningQian, "Services in the Cloud Computing Era: A Survey", Software Engineering Institute, East China Normal University, Shanghai 200062, China. National Institute of Information and Communications Technology, Kyoto 619-0289, Japan. IUCS 2010, 978-1-4244-7820-0/10/$26.00 ©2010 IEEE
2. Mamoun Hirzalla,"Realizing Business Agility Requirements through SOA and Cloud Computing", DePaul University, College of Computing and Digital Media andIBM, Chicago, IL USA, 2010 18th IEEE International Requirements Engineering Conference,

1090-705X/10 $26.00 © 2010 IEEE, DOI 10.1109/RE.2010.70, pp379-380
3. Hailong Sun, Xu Wang, Chao Zhou, Zicheng Huang, Xudong Liu, "Early Experience of Building a Cloud Platform for Service Oriented Software Development", School of Computer Science and Engineering, Beihang University, Beijing, China, 978-1-4244-8396-9/10/$26.00 ©2010 IEEE.
4. Michael Mattess, Christian Vecchiola, and RajkumarBuyya, "Managing Peak Loads by Leasing Cloud Infrastructure Services from a Spot Market", Cloud Computing and Distributed Computing (CLOUDS) Laboratory, Department of Computer Science and Software Engineering, The University of Melbourne, Australia, 12th IEEE International Conference on High Performance Computing and Communications, 2010, 978-0-7695-4214-0/10 $26.00 © 2010 IEEE, DOI 10.1109/HPCC.2010.77, pp180-188
5. Li Guo, YikeGuo_ and XiangchuanTian, "IC Cloud: A Design Space for Composable Cloud Computing", Department of Computing, Imperial College London, UK, IEEE 3rd International Conference on Cloud Computing, 2010, 978-0-7695-4130-3/10 $26.00 © 2010 IEEE, DOI 10.1109/CLOUD.2010.18, pp 394-401
6. SameeraAbdulrahmanAlmulla, Chan YeobYeun, "*Cloud Computing Security Management",* Khalifa University of Science, Technology and Research (KUSTAR), Shrjah Campus, P.O. Box 573, Sharjah, United Arab Emirates,
7. Paul Marshall, Kate Keahey1,2 and Tim Freeman1, **"Elastic Site:** Using Clouds to Elastically Extend Site Resources", Department of Computer Science, University of Colorado at Boulder Boulder, CO USA, 1Computation Institute, University of Chicago, 2Argonne National Laboratory, Chicago, IL USA, 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, 2010m 978-0-7695-4039-9/10 $26.00 © 2010 IEEE, DOI 10.1109/CCGRID.2010.80, pp 43-52
8. WesamDawoud, Ibrahim Takouna, ChristophMeinel. "Infrastructure as a service, Security: Challenges Solutions", HassoPlattner Institute Potsdam, Germany, Ministry of Education & Higher Education, Palestine.
9. Ammar Khalid, "Cloud Computing: Applying Issues in Small", Department of Computer Science & IT, The Islamia University of Bahawalpur, Bahawalpur, Pakistan. International Conference on Signal Acquisition and Processing, 2010, 978-0-7695-3960-7/10 $26.00 © 2010 IEEE, DOI 10.1109/ICSAP.2010.78, pp 278-281