

Weather Analytics Using Machine Learning Techniques On Openstack Cloud

¹Pratima Nerkar, ²Suresh Rathod

¹ Department of Computer Engineering, Pune University , Sinhgad Academy of Engineering
Pune, Maharashtra, India

² Department of Computer Engineering, Pune University , Sinhgad Academy of Engineering
Pune, Maharashtra, India

Abstract - Cloud computing is the latest distributed computing paradigm and it offers various opportunities to solve large-amount of scientific problems. Cloud-enabled VM for developers and scientists. Huge data analysis can be done using cloud computing and machine learning technique management for developers & scientists. We are working on the technology framework for weather analytics. Weather science is Big Data domain .Weather plays an important role in everyday life. Weather prediction has been the one of the most challenging issue around the world in last year. Various techniques are used for prediction are Statistical analysis, Data mining, Regression analysis, and neural networks. This paper represents machine learning technique for the early prediction of weather on hadoop in Openstack cloud.

Keywords - *Cloud computing, weather analytics, machine learning technique, big data, hadoop.*

1. Introduction

Cloud computing is a concept which runs in a distributed environment using virtualization technique. It is not a new technology; it is just a way of using old technologies in effective manner. It is a model which enables on demand network access to a shared pool of resources in convenient way. Cloud computing has two important concepts and they are abstraction and virtualization. Technologies such as cluster, grid, and now, cloud computing, are trying to give a single system view to distributed resources by aggregating these resources and computing them in fully virtualized environment. Buyya et al. [5] have defined Cloud as a parallel and distributed system These systems are inter-connected and virtualized computers. Vaquero et al. [15] have stated cloud as a large pool of easily usable virtualized resources (such as hardware, development platforms and/or services). These

resources are dynamically reconfigured for minimum resource utilization. These resources are used by a pay-per-use model in which guarantee is offered by the Infrastructure Provider by means of customized Service Level Agreements. Mathematics and statistics are used for Knowledge discovery from data. Weather is a big data and predictive analytics of weather is a challenging topic. This is need of a common man to get some prediction of weather in order to preplan. Various weather forecast methods are employed in weather prediction at regional and national levels. Basically, there are two methods for predicting weather. One is Empirical method and second is Dynamical method. The empirical method deals with past data of the weather and its correlation with a variety of variables. Some of the empirical methods are regression, artificial neural network; fuzzy logic and group method of data handling are used for weather prediction.

In dynamical method, predictions are done by physical models which used equations and numerical weather forecasting method [4].Regression is a method where analysis is to be done using simple data to know parameter values and standard errors. In regression response variable and explanatory variable plays important role. Regression is a statistical method where the explanatory variable and response variable are correlated depending on the model used. Machine learning technique can be widely used in business, the social and behavioral sciences, the biological sciences, weather prediction, and many other areas. supervised learning analysis includes para-metric methods such as linear and logistic regression. In this paper, weather prediction model is implemented with the use of supervised learning technique. The experimental results prove that there is a close agreement between the

predicted and actual weather. The actual development is done using OpenStack. OpenStack is open source software which the project developers and cloud computing technologist can use to setup and run the cloud. Its services can be accessed via APIs. The important components of OpenStack are Nova, Swift, Keystone and Glance, Keystone and Horizon

2. Related Work

2.1 Background

Expansion of internetworking coupled with Service Oriented Architecture gave birth to new concepts like Cloud Computing which cater exponentially increasing demand of data generation, storage, integration and communication. The key concept operating at lower level of cloud computing stack is machine. The task of seamlessly isolating VMs and managing timely allocation of resources is done by an additional software component called Hypervisor or Virtual Machine Monitor (VMM). Virtualization can be used to club together applications running on different servers, on a single host under different virtual machines (termed as server virtualization)[3].

For the success of server virtualization Hypervisor or Virtual Machine Monitor, should ensure CPU, memory, network and IO virtualization (different virtual machines get a feeling of having these resources dedicated to them but in reality they are shared with time multiplexing). This reduces infrastructure required and eventually operational and maintenance cost [3][4]. In addition to cost effectiveness virtualization brings many benefits like high availability, scalability, less power consumption etc Though Liu et al. [3] proposed a approach to minimize memory consumption and complexity, the technique still needs sufficient space and complex computation. Hadoop is an open-source distributed cluster form that includes a distributed file system, HDFS and the programming model, Map Reduce. Virtualization. Virtualization, means creating a virtually (rather than actually), such as a hardware platform, operating system, a storage device or network resources [2].

Basically it is a technique that divides a physical computer into several isolated machines known as virtual machines (VM). Multiple virtual machines can run on a host computer, each possessing its own operating system and applications. This gives an illusion to the processes running on virtual machines as if they are using dedicated hardware resources, but in reality they are sharing the

physical hardware of the host

2.2 Private Cloud

Sonali Yadav[11] introduced importance of private cloud with that its deployment is explained. Various open source software's such as Eucalyptus, Openstack, OpenNebula have been studied. Bram et. Al [12] introduces a test method to determine the value of deployment in cloud. He focused on the essential characteristics of cloud. Rashmie et al.[13]focused on dynamic scaling, high availability, multitenancy and effective resource allocation.

2.3 Weather Prediction

Ozlem Terzi [3] proposed a model for weather prediction using data mining technique. The weather data is collected per month from three different places. The error is detected of minimum value. Z. Ismail [4] introduced a forecasting model in order to predict gold price. He concluded that the multiple linear regression gives better performance than the native prediction method. Wint Thida Zaw, et al. [5] presented a technique, which described non-linear regression. He introduced relationship between input and output to predict weather. Giang H.Nguyen et al. [9] proposed a model which is used to predict the sales of factory using past data of sales and finance indicator as prediction variable.

3. Supervised Learning Model

Regression is a method where the mapping of output from input is learned. It is classified as supervised learning problem. For simplicity weather data set is to be taken as input attribute and a linear model is to be used. The machine learn-ing program optimizes the parameters in order to minimize the error.

$$Y = w_1X + w_0 \quad (1)$$

Here, Y=prediction variable
X=weather data attribute
 w_0, w_1 =regression parameters

Surveying past weather data, we can collect a training data that is weather data and machine learning has a function to learn Y as a function of X.

(2) Represents the multiple linear regression

$$g(X) = w_1X_1 + w_2X_2 + \dots + w_mX_m + w_0 \quad (2)$$

4. Weather Analytics as a service

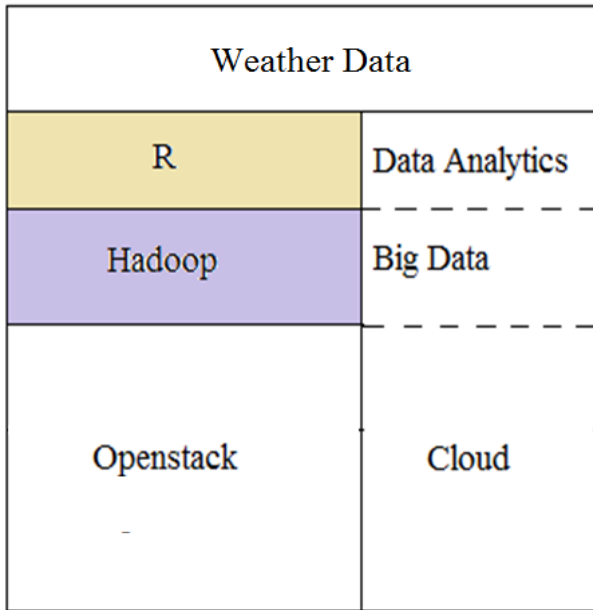


Fig 1: Weather Analytics as a service Stack

Fig 1, shows a weather analytics as a service stack which consist of cloud at base, then on it hadoop and then on hadoop R and on R we can analyze the weather data. Openstack is an opensource software which can be used for building a private cloud by the project developers and the cloud computing technology users. They use openstack in order to setup i.e install and run cloud. The openstack services can be accessed via API's. Openstack has following important components and they are Nova , Glance, Swift, Keystone and Horizon also .Sahara is openstack API's for hadoop and R integration. Hadoop is an opensource framework. It is built using java in order to process and query a very large amount of data on a very large number of clusters of commodity hardware Apache Hadoop has two main basic features. And they are HDFS (Hadoop Distributed File System) MapReduce

R for revolution analytics .It is having more than 3000 packages, It is increasing day-by-day. Now it has been really impossible for a book to even try to explain all these packages .To perform statistical analysis R is used. R facilitates with a very wide variety of the statistical, machine learning (linear and nonlinear modeling, classic statistical tests, time-series analysis, classification, clustering) and graphical techniques, and it is really very extensible.

5. Experimental Setup

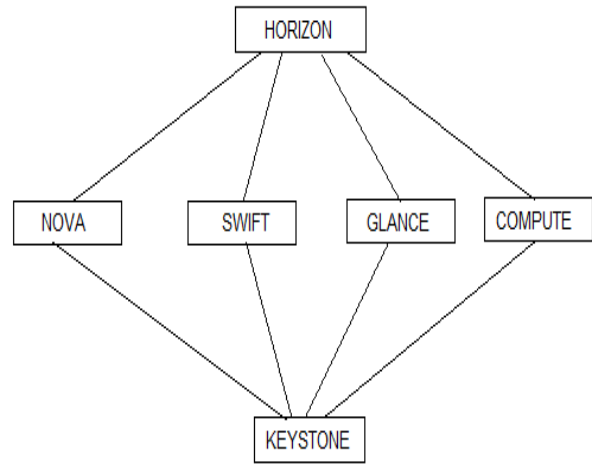


Fig: 2 Openstack Architecture

To set-up our experimental scenario, we started by setting up a test bed for hosting a Cloud platform. We decided to use the OpenStack which is solution that allow us to control a very large number of resources throughout a datacenter on the test bed [14]. To set-up our experimental scenario, we started by setting up a test bed for hosting a Cloud platform. The OpenStack as shown in Fig. 3. currently a very hot topic in development technology. Openstack has a potential and this potential is due to its architecture as well as due to large community and the support of its partners. Openstack is supported by various companies in the world .All the thing is based on the code used by NASA and Rackspace Cloud. It uses number of hypervisors such as Xen, KVM, HyperV, and Qemu..On openstack we will be installing hadoop and the integration of hadoop with openstack is done using sahara. After that R will be installed on system and integration of R and hadoop is done using various packages like Rhadoop, Rhive. Thus, We can achieve a stack for weather analytics.

6. Conclusions

Weather has a great impact on agriculture, economy not only in India but across the whole world In this paper we have proposed a method for weather analytics prediction. This is the only prediction regarding weather but not accurate because of weather factors. Thus we are going to meet the Big Data Challenges of Weather Science through Cloud-Enabled Weather Analytics-as-a-Service

References

- [1] Quyet Thang NGUYEN, Nguyen QUANG-HUNG, Nguyen HUYNH TUONG, Van Hoai TRAN, Nam THOAI, "Virtual Machine Allocation in Cloud Computing for Minimizing Total Execution Time on Each Machine," 978-1-4673-2088-7, IEEE, 2013.
- [2] Han Chen, Minkyong Kim, Zhe Zhang, Hui Lei, "Empirical Study of Application Runtime Performance using On-demand Streaming Virtual Disks in the Cloud".
- [3] "Cloud computing," <http://www.ibm.com/cloud-computing/us/en/whatis-cloud-computing.html>.
- [4] Zhiming Shen, Zhe Zhang, Andrzej Kochut, Alexei Karve, Han Chen, Minkyong Kim Hui Lei, Nicholas Fuller, "VMAR: Optimizing I/O Performance and Resource Utilization in the Cloud".
- [5] K. H. Kim, W. Y. Lee, J. Kim, R. Buyya. "SLA-Based Scheduling of Bag-of- Tasks Applications on Power-Aware Cluster Systems," IEICE Transactions on Information and Systems, Issue 12, pp. 3194-3201, 2010.
- [6] P. Padala, "Understanding live migration of virtual machines." Available: <http://tinyurl.com/24bdaza>, Jun. 2010.
- [7] D. Breitgand, G. Kutiel, and D. Raz, "Cost-aware live migration of services in the cloud," in 2011 USENIX Workshop on Hot Topics in Management of Internet, Cloud, and Enterprise Networks and Services.
- [8] S.Hacking and B. Hudzia, "Improving the live migration process of large enterprise applications," in Proceedings 2009 International Workshop on Virtualization Technologies in Distributed Computing.
- [9] A. Beloglazov, R. Buyya. "Energy Efficient Allocation of Virtual Machines in Cloud Data Centers," 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, pp. 577-578, 2010.
- [10] Girish L S, Dr. H S Guruprasad, "Building Private Cloud using OpenStack ," , *International Journal of Emerging Trends & Technology in Computer Science* Vol. 3, Issue 3, May – June 2014 , ISSN 2278-6856
- [11] SonaliYadav, "Comparative Study on Open Source Software for Cloud Computing Platform: Eucalyptus, Openstack and Opennebula", *International Journal Of Engineering And Science*, Vol.3, Issue 10 (October 2013), pp 51-54, ISSN (e): 2278-4721, ISSN (p):2319-6483.
- [12] Bram Rongen, "Making the case for migration of information systems to the cloud", 16th Student Conference on IT, Enschede, The Netherlands, Jan 27 2012, Copyright 2011, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.
- [13] Rashmi, Dr. Shabana, Mehruz, Dr. G. Sahoo, "A fivephased approach for the cloud migration", *International Journal of Emerging Technology and Advanced Engineering*, April 2012, ISSN 2250-2459.
- [14] A. Beloglazov, J. Abawajy and R. Buyya, "Energy-aware resource allocation heuristics for e_cient management of data centers for Cloud computing," *Future Generation Computer Systems*, vol. 28, no. 5, pp. 755-768, 2012. DOI:10.1016/j.future.2011.04.017.
- [15] Rajkumar Buyya , Andrzej Goscinski , " CLOUD COMPUTING -Principles and Paradigms "
- [16] Carlo Mastroianni, Michela Meo, and Giuseppe Papuzzo, " Probabilistic Consolidation of Virtual Machines in Self-Organizing Cloud Data Centers," *IEEE TRANSACTIONS ON CLOUD COMPUTING*, VOL. 1, NO. 2, JULY-DECEMBER 2013
- [17] Mladen A. Vouk, " Cloud Computing – Issues, Research and Implementations," *Journal of Computing and Information Technology - CIT* 16, 2008, 4, 235–246
- [18] Worachat Chawarut, Lilakiatsakun Woraphon, " Energy-Aware and Real-time Service Management In Cloud Computing," *IEEE*, 2013
- [19] T. Swathi, K. Srikanth, S. Raghunath Reddy, " VIRTUALIZATION IN CLOUD COMPUTING," *IJCSMC*, Vol. 3, Issue. 5, May 2014, pp. 540 – 546
- [20] Neha Khandelwal et.al " Climatic Assessment Of Rajasthan's Region For Drought With Concern Of Data Mining Techniques" in *International Journal Of Engineering Research and Applications (IJERA)* ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 5, September- October 2012, pp. 1695-1697.
- [21] Pratima D. Nerkar, Prof. Suresh B. Rathod, "A Survey on Performance and Energy Management in cloud computing," *International Journal of Science and Research* , , Vol. 3, Issue. 11, 2319-7064 , Nov. 2014
- [22] Pratima D. Nerkar, Prof. Suresh B. Rathod, "A survey on virtualization technology in cloud computing ," *International journal of Advance Research in Science and Engineering* , , Vol. 3, Issue. 11, 2319-8354 , Nov. 2014.

Authors Profile

Pratima Devidas Nerkar completed B.E. in Computer Engineering in 2011 from K.K. Wagh C.O.E., Nashik.

Suresh Baliram Rathod completed B.E. in Information Technology in 2007 from STBCE; Tuljapur. He has also completed his M.E. from SCOE, Pune.

Can we use machine learning as a game changer in this domain? Using features like the latest announcements about an organization, their quarterly revenue results, etc., machine learning techniques have the potential to unearth patterns and insights we didn't see before, and these can be used to make unerringly accurate predictions. In this article, we will work with historical data about the stock prices of a publicly listed company. We will implement a mix of machine learning algorithms to predict the future stock price of this company, starting with simple algorithms like averaging and linea