



Leveraging Cloud for Large Scale Business Transactions – A Survey

Neethu Mohandas¹, M Sudheep Elayidom², Sasi Gopalan³

Research Scholar, School of Engineering, Cochin University of Science and Technology, Kochi, India¹

Associate Professor, School of Engineering, Cochin University of Science and Technology, Kochi, India^{2,3}

Abstract: Leveraging cloud for the benefit of today's business world is a significant area to work on. This paper gives insight on some of the works done so far that involves migration of business applications to cloud and on the performance enhancement it brings compared to the previous on-premise business applications. Big Data analytics is a hot topic in this domain, and this paper surveys some works done so far on Big Data analytics using the cloud framework. A brief introduction about the Cloud and Big Data and their significance in today's business world is given. Our main motive is to give a brief overview of the effect of "the Cloud" and "Big Data" on today's business.

Keywords: On-premise, Migration, Cloud, Big Data analytics.

I. INTRODUCTION

As you all must have already heard, today's tech world is changing the mantras "the Cloud" and "Big Data" so frequently that even the common man, who is blissfully unaware of the technical world and its complexities, might have heard of it. According to IDC, Over the next five years spending on cloud-based Big Data and analytics (BDA) solutions will grow three times faster than spending for on-premise solutions. Hybrid on/off premise deployments will become a requirement [11]. So what exactly is this so called "Cloud" and "Big Data"? Are they just some fancy terms in the tech world or do they really revolutionize our industry? Both these terms will be explained briefly in the following sections.

Understanding our market is a crucial step for predicting its future, and that's exactly where Big Data and Cloud comes to play. An analysis on the recent history of the market, along with the present trends in it, is one way of forming a prediction about the future market. Growth in applications incorporating advanced and predictive analytics, including machine learning, will accelerate in 2015. These apps will grow 65% faster than apps without predictive functionality [11].

In the coming years, more and more data will be produced, machine generated and otherwise, that the different industries and organizations can peep into. The competitive edge will push them to analyze those data for their advantage, as much as they can. So, for the future, Big Data is the Smart Data. In other words, the future of the organizations now hang onto one task: efficiently analyzing the "Big Data" at the "right time" and applying the results in the "right market".

Google, AWS, IBM, and Microsoft are becoming the major players in the public cloud domain. The Cloud thirsty industries and organizations are becoming their customers [10]. This paper illustrates some implementations done based on cloud and Big Data by some of these major players.

II. CLOUD COMPUTING

"The Cloud" essentially refers to "Cloud Computing", "which, in simple words, is anything that delivers hosted services over the internet. The name "Cloud" is said to be inspired from the symbol that is usually used to represent internet in diagrams etc. In other words, it is a business model, where the industries can use computing resources just like another utility.

A. Public, Private and Hybrid Clouds

The Cloud could be classified as public, private and hybrid cloud. The public cloud services are provided by a third party over the internet and the customers are required to pay only for the CPU Cycles, bandwidth or storage they consume. Amazon Web Services (AWS), Microsoft Azure are some good examples.

The private cloud services, on the other hand, are provided by the company's own data center to internal users, thus focusing on security. Now, the "hybrid" cloud, as the name indicates, is a hybrid of the above mentioned public and private clouds.

The critical applications can be run on the on-premise private cloud, while those applications that should scale on demand could be run on public cloud. It brings the best out of both, and hence it is predicted to be the favored among the three.

B. PaaS, IaaS and SaaS

The cloud can provide three types of services, namely, Platform as a Service (PaaS), Infrastructure as a Service (IaaS) and Software as a Service (SaaS). PaaS such as Google App Engine hosts developmental tools which the customers can access over the internet through Application Program Interfaces (API) etc. Customers can migrate workloads to a Virtual Machine (VM), utilizing a predetermined storage capacity and available APIs in case of IaaS. A typical example for an IaaS provider is AWS. SaaS, otherwise known as Web Services, provides softwares over the internet, all the customer will require is an internet connection and a computer or mobile device.



AgileCRM (CRM for small businesses), Airwoot (Evolutionise Customer Support on Social Media), Applane (Integrated Application Suite for Business & Schools) etc are some SaaS providers in India [9].

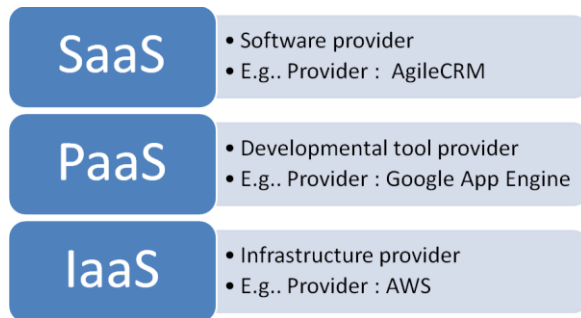


Fig 1: Types of Cloud Services

III. BIG DATA

Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making [8]. It helps organizations to store, analyze and manipulate huge amount of disparate data at the right speed at the right time.

A. Characteristics of Big Data

The three main characteristics of Big Data are:

- Volume
- Velocity
- Variety

Big Data is, in simple words, the data which requires exceedingly high processing capacity that the traditional database systems are unable to offer. What is so unique about Big Data? The structure and size of the data makes it unique. It represents both the information as well as the way the data can be analyzed. Nowadays huge amount of machine generated data makes it difficult for industries to analyze them.

I) Volume:

Let's say, for example, there is a huge amount of information generated by the smart phones all over the world. It will be a complicated task to analyze this data to understand the customer preferences for a product. This represents the volume of Big Data. The ever increasing source of data along with the larger size of data produces increasingly huge volume of data.

II) Velocity:

The second characteristic, Velocity, deals with the pace at which this information flows in. The massive amount of data flows in continuously from various sources. The industries are aiming at analyzing the real time data to spot the current market trends in order to predict the future market. This data can be helpful for decision making in this competitive world, helping in ROI.

III) Variety:

The Variety implies the disparity in the type of data that flows in. It could be structured or unstructured. The data

could be in the form of pictures, videos, text etc in different formats. This creates a problem in storage, mining as well as analyzing the data.

Interpreting Big Data to find out hidden trends is not easy, and is almost next to impossible using traditional methods. Apache Hadoop is an example framework which can process Big Data. It is an open source framework, and consequently we will get different versions of Hadoop online depending on which vendor you downloaded it from, namely, Cloudera, Amazon etc.

A deep need exists for the structure to parse the data to separate out the useless information and find the useful threads to uncover opportunities. Even more potential has opened for those who can orchestrate this feat.

IV. BIG DATA IN CLOUD

Here we briefly mention some of the works done in the Big Data domain, utilizing the cloud framework.

A. Product Range Effect in Purchase Data

In today's market, the customers are driven by two utility functions [1]:

- Generic utility function
- Personal Utility function

A customer is driven both by a generic utility function (cost minimization) and by a personal utility function (fulfillment of unique desires). Using an analytical framework which can scale higher, an analysis on customer purchase data was done which brought out some interesting results.

From the analysis, a "Range Effect" of products was discovered. It was found out from the purchase data that the more sophisticated a product is, the more cost the customers were willing to pay in terms of the distance to travel, more than in terms of the price of the product itself. As an application of the analysis, a framework to accurately predict how long a customer will travel to buy a product was developed, based on the product's sophistication.

To be considered "sophisticated" a product needs to satisfy two constraints:

- It has to be sold to few customers
- Customers buying it have to buy all products that are less sophisticated than it.

The logic is that each product satisfies a need and a customer buys a product if and only if she already satisfied all more basic needs. Based on this concept, a suitable sophistication index was developed which will help predict the "Product Range" of the concerned product, or simply put, the distance the customer will be willing to travel to buy the product.

B. Recommending similar items in online Markets

A recommendation system for similar items for big online markets was developed based on Hadoop Map-Reduce framework [2]. This system was implemented in eBay which resulted in increased user interactions. Similarity



recommendation is a frequently used service in most of the online markets. But the requirements for such a system varies, and becomes challenging for dynamic marketplaces such as eBay where the product listings are huge and the products get sold out within a matter of 1-2 weeks.

The main differences incorporated in this system:

- Capturing user preferences for similarity using clusters learned from user queries
- Striking a balance between quality and similarity

The developed recommendation system was found beneficial to the multi-billion company in terms of profit. The system was developed based on an offline process and an online system.

The offline process groups item listings which are short-lived, and produces a long-term cluster definition based on it. The offline process is implemented using a Hadoop framework.

The clusters are formed based on user queries. The online system then checks the important similarity dimensions and then focuses on the trade-off between similarity and quality factors. Utilizing the user queries in building the system is said to have reaped profits in the company.

C. Minimizing Big Data Problems using Cloud Computing Based on Hadoop

To reduce the Big Data related problems, the industries' infrastructure is utilized and based on the requirements, merging of infrastructures based on Hadoop in the form of different public and private clouds are introduced as a solution. It improves the processing time and capacity of the big data requests.

An analysis shows improvement in performance and capacity for incoming requests upon increase in the number of clouds. IT organizations are advised to setup big data infrastructure according to the proposed cloud-based approach to overcome their big data problems in less time and less cost.

D. Big Data Analytics in Retail

This work was done by Intel, by teaming up with Living Naturally (a trusted name in retail technology) investigates in detail the following factors in order to improve the productivity and marketing capabilities of retailers and suppliers:

I) Product Pipeline Tracking:

When inventory levels are not in sync with the demand, an application was developed for make recommendations to avoid such a situation in future and to maximize ROI.

II) Market Basket Analysis:

Lets the retailers know about related products that benefits from a sales increase. An application which makes suggestions about the related products is developed.

III) Social Media Analysis:

Some posts in social networking sites such as Twitter results in drastic change in customer demand. An

Copyright to IJIREECE

application which makes suitable recommendations before the products go viral in social media is developed.

V. LATEST TRENDS IN CLOUD

“The Cloud” has gained momentum at an alarming rate in the recent years, and most of the business organizations, small or large has already shifted to Cloud, and the rest are going to, in order for survival. It is just a matter of time. The technology trends in Cloud keeps changing, some gaining wide approval, like Hadoop. Here we will list out some current major trends in cloud computing.

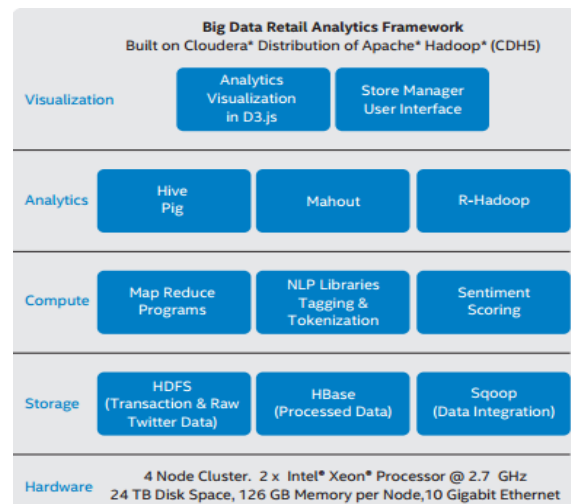


Fig 2: Big Data Retail Analytics Framework[4]

A. Increase in demand for PaaS

The companies are trying to shift the daily business to cloud [5]. The pre-requisite is that they should tailor cloud to their requirements. Hence a growing demand will occur for PaaS providers.

B. Go Programming Language

It's a programming language developed by Google for Cloud based applications. Its favourable features such as better concurrency support etc will higher demand for the language.

C. Disaster Recovery

Industries which are gradually considering cloud-based back up for disaster recovery might slowly realize they could gradually migrate to cloud altogether.

D. Hybrid Clouds

Recent surveys point out the fact that 55% of the companies prefer hybrid clouds [6].

E. Bring Your Own Device (BYOD)

End users are using their mobile devices to upload more and more data into personal cloud services for streaming, storage and syncing. Industries should find a means to integrate personal cloud services for their employees in a BYOD environment.



VI. CONCLUSION

Cloud computing is slowly revolutionizing the future of our industries. At a personal level and a professional level, access to the cloud has opened up a whole new world of possibilities, for better (photo sharing sites and virtual offices) or for worse (criminal networking and illegal downloading). Regardless of the consequences, for better or for worse, the Cloud is here to stay.

Big Data analytics play a major role in this revolution using the cloud framework. This paper has given potential insights into the Cloud and Big Data domain. Some of the Big Data analytics implementations done in Cloud are also briefly discussed along with some latest trends in cloud computing.

This survey brings out the significance of the Cloud domain in this competitive world, and the role played by Big Data analytics in predicting the future.

REFERENCES

- [1] D. Pennacchioli, Coscia, M., Rinzivillo, S., Pedreschi, D., and Giannotti, F., "Explaining the Product Range Effect in Purchase Data" in IEEE Big Data. 2013.
- [2] J. Katukuri, T. Konik, R. Mukherjee, S. Kolay, "Recommending similar items in large-scale online marketplaces" in IEEE International Conference on Big Data 2014.
- [3] M. Adnan, M. Afzal, M. Aslam, R. Jan, A. M. Martinez-Enriquez, "Minimizing big data problems using cloud computing based on Hadoop architecture" in IEEE HONET, 2014
- [4] "Getting started on Big Data Analytics in Retail", [White Paper], <http://www.intel.in/content/dam/www/public/us/en/documents/solution-briefs/retail-big-data-analytics-solution-blueprint.pdf>, Intel, 2014
- [5] <http://search.computing.techtarget.com/feature/Cloud-computing-technology-trends-in-2015>
- [6] <http://www.rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2015-state-cloud-survey#Hybrid> Cloud Remains the Preferred Strategy
- [7] <http://www.salesforce.com/uk/socialsuccess/cloud-computing/why-move-to-cloud-10-benefits-cloud-computing.jsp>.
- [8] <http://www.gartner.com/it-glossary/big-data>.
- [9] <http://www.zapstitch.com/blog/saas-companies-from-india/>.
- [10] <https://www.bbvaopenmind.com/en/cloud-computing-big-data-and-mobility-2015-tech-trends/>.
- [11] <http://www.idc.com/getdoc.jsp?containerId=prUS25329114>.
- [12] <http://www.forbes.com/sites/gilpress/2014/12/11/6-predictions-for-the-125-billion-big-data-analytics-market-in-2015/2/>.