International Journal of Research (IJR)
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Vol. 8 Issue 6
June 2021

Requirement of Cloud Environment for Large Information Handling

Bharat Batham
Research Scholar, Assistant Professor, Department of Computer Science and Application, Atal Bihari Vajpayee Hindi University, Bhopal, Madhya Pradesh, India

Dr. Shailja Sharma
Associate Professor, Department computer Science & Engineering, Rabindranath Tagore University, Bhopal, Madhya Pradesh, India

Abstract:
The modern day advancement is increasingly digitizing our lives which have led to a rapid growth of data. Such multidimensional datasets are precious due to the potential of unearthing new knowledge and developing decision-making insights from them. Analyzing this huge amount of data from multiple sources can help organizations to plan for the future and anticipate changing market trends and customer requirements. While the Hadoop framework is a popular platform for processing larger datasets, there are a number of other computing infrastructures, available to use in various application domains. The primary focus of the study is how to classify major big data resource management systems in the context of cloud computing environment. We identify some key features which characterize big data frameworks as well as their associated challenges and issues. We use various evaluation metrics from different aspects to identify usage scenarios of these platforms. The study came up with some interesting findings which are in contradiction with the available literature on the Internet.

Keyword: multidimensional datasets, data, Hadoop framework, computing infrastructures, Cloud computing.

Introduction

You've likely heard the terms "Big Data" and "Cloud Computing" before. If you're involved with cloud application development, you may even have experience with them. The two go hand-in-hand, with many public cloud services performing big data analytics.

With Software as a Service (SaaS) becoming increasingly popular, keeping up-to-date with cloud infrastructure best practices and the types of data that can be stored in large quantities is crucial.

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Vol. 8 Issue 6
June 2021

We'll take a look at the differences between cloud computing and big data, the relationship between them, and why the two are a perfect match, bringing us lots of new, innovative technologies, such as artificial intelligence.[1]

The Difference between Big Data & Cloud Computing

Before discussing how the two go together, it's important to form a clear distinction between "Big Data" and "Cloud Computing". Although they are technically different terms, they're often seen together in literature because they interact synergistically with one another.[2]

Image 1:

What is big data?

Gartner defines big data as high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.[3]

Whoa, that's a mouthful.

Building on Gartner's definition, the concept of big data and what it encompasses can be better understood with four Vs:

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Vol. 8 Issue 6
June 2021

- *Volume*. The amount of data accumulated by private companies, public agencies, and other organizations on a daily basis is extremely large. This makes volume the defining characteristic for big data.
- *Velocity*. It's a given that data can and will pile up really fast. But what matters is the speed with which you can process and examine this data so that it becomes useful information.
- *Variety*. The types of data that get collected can be very diverse. Structured data contained in databases, and unstructured data such as tweets, emails, images, videos, and more, need to be consumed and processed all the same.
- *Veracity*. Because of its scale and diversity, big data can contain a lot of noise. Veracity thus refers to the the certainty of the data and how your big data tools and analysis strategies can separate the poor quality data from those that really matter to your business.

Technology leaders also name a fifth V – value. But this one isn't inherent within the huge amounts of raw data. Instead, the true value of big data can only be realized when the right information is captured and analyzed to gain actionable insights.

To get a better idea of how big big data is, let's review some statistics:

- Over 1 billion Google searches are made and 294 billion emails are sent everyday
- Every minute, 65,972 Instagram photos are posted, 448,800 tweets are composed, and 500 hours worth of YouTube videos are uploaded.
- By 2020, the number of smartphone users could reach 6.1 billion. And taking Internet of Things (IoT) into account, there could be 26 billion connected devices by then.

For sure, big data is really big.

Why big data in the cloud makes perfect sense

The benefits of moving to the cloud are well documented. But these benefits take on a bigger role when we talk of big data analytics.

Big data involves manipulating petabytes (and perhaps soon, exabytes and zettabytes) of data, and the cloud's scalable environment makes it possible to deploy data-intensive applications that power business analytics. The cloud also simplifies connectivity and collaboration within an

organization, which gives more employees access to relevant analytics and streamlines data sharing.[4]

While it's easy for IT leaders to recognize the advantages of putting big data in the cloud, it may not be as simple to get C-suite executives and other primary stakeholders on board. But there's a business case to be made for the big data + cloud pairing because it gives executives a better view of the business and boosts data-driven decision making.

For instance, optimization of the supply chain and efficient tracking of defects – both principal concerns of a COO of a physical product company – is made easier with material data on hand. Data is also key for the CMO looking to increase customer engagement and loyalty, and for the CFO seeking new opportunities for cost reduction, revenue growth, and strategic investments.

And all of these insights can be easily presented to the CEO to inform fast, strategic decision making.

Whatever perspective you may have, big data complemented with an agile cloud platform can affect significant change in the way your organization does business and achieves your objectives.[5]

Many enterprises are already making the move. A Forrester Research survey in 2017 revealed that big data solutions via cloud subscriptions will increase about 7.5 times faster than on premise options.

Cloud Computing: This refers to the processing of anything, including Big Data Analytics, on the "cloud". The "cloud" is just a set of high-powered servers from one of many providers. They can often view and query large data sets much more quickly than a standard computer could.

Essentially, "Big Data" refers to the large sets of data collected, while "Cloud Computing" refers to the mechanism that remotely takes this data in and performs any operations specified on that data.

Relationship between Big Data & Cloud Computing

Cloud Computing providers often utilize a "software as a service" model to allow customers to easily process data. Typically, a console that can take in specialized commands and parameters is available, but everything can also be done from the site's user interface. Some products that are

usually part of this package include database management systems, cloud-based virtual machines and containers, identity management systems, machine learning capabilities, and more.[6]

In turn, Big Data is often generated by large, network-based systems. It can be in either a standard or non-standard format. If the data is in a non-standard format, artificial intelligence from the Cloud Computing provider may be used in addition to machine learning to standardize the data.[7]

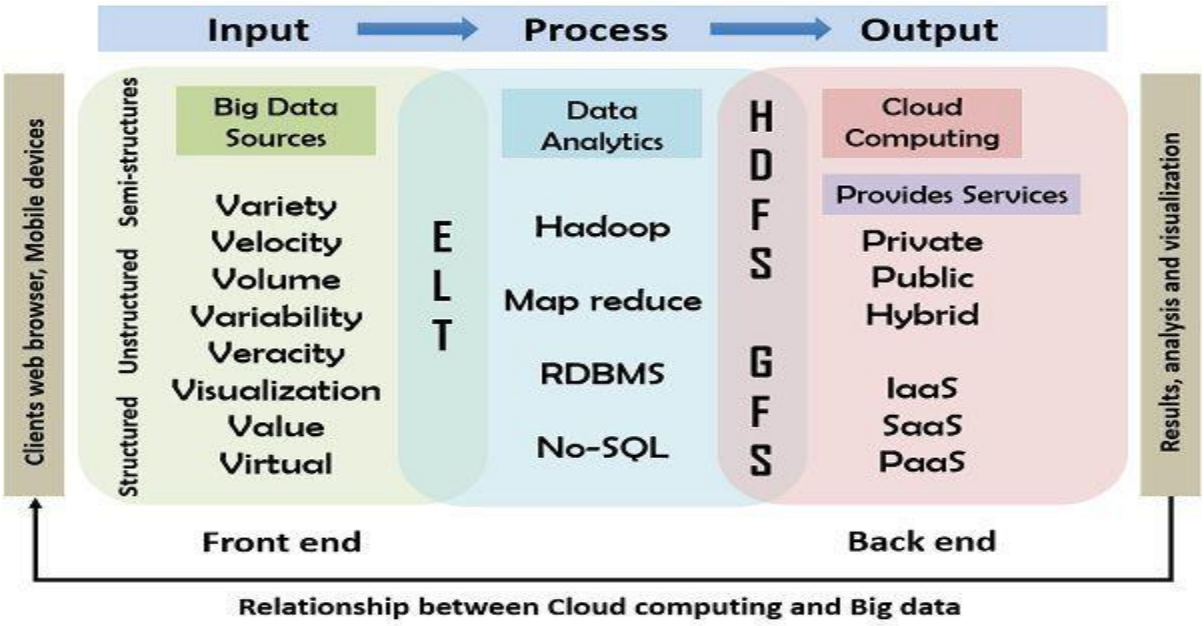

Image 2:

From there, the data can be harnessed through the Cloud Computing platform and utilized in a variety of ways. For example, it can be searched, edited, and used for future insights.

This cloud infrastructure allows for real-time processing of Big Data. It can take huge "blasts" of data from intensive systems and interpret it in real-time. Another common relationship between Big Data and Cloud Computing is that the power of the cloud allows Big Data analytics to occur in a fraction of the time it used to.[8]

Big Data & Cloud Computing:

As you can see, there are infinite possibilities when we combine Big Data and Cloud Computing! If we simply had Big Data alone, we would have huge data sets that have a huge amount of

potential value just sitting there. Using our computers to analyze them would be either impossible or impractical due to the amount of time it would take.

However, Cloud Computing allows us to use state-of-the-art infrastructure and only pay for the time and power that we use! Cloud application development is also fueled by Big Data.[9] Without Big Data, there would be far fewer cloud-based applications, since there wouldn't be any real necessity for them. Remember, Big Data is often collected by cloud-based applications, as well!

In short, Cloud Computing services largely exist because of Big Data. Likewise, the only reason that we collect Big Data is because we have services that are capable of taking it in and deciphering it, often in a matter of seconds. The two are a perfect match, since neither would exist without the other!

Conclusion

Finally, it's important to note that both Big Data and Cloud Computing play a huge role in our digital society. The two linked together allow people with great ideas but limited resources a chance at business success. They also allow established businesses to utilize data that they collect but previously had no way of analyzing.

More modern components of cloud infrastructure's typical "Software as a Service" model such as artificial intelligence also enable businesses to get insights based on the Big Data they've collected. With a well-planned system, businesses can take advantage of all of this for a nominal fee, leaving competitors who refuse to use these new technologies in the dust.

References:

1. "Big Data in the Cloud: Converging Technologies, (August 11)," https://www.intel.com/content/dam/www/public/emea/de/de/documents/product-briefs/big-data-cloud-technologies-brief.pdf.View at: Google Scholar
2. Q. He, J. Yan, R. Kowalczyk, H. Jin, and Y. Yang, "Lifetime service level agreement management with autonomous agents for services provision," *Information Sciences*, vol. 179, no. 15, pp. 2591–2605, 2009.View at: Publisher Site | Google Scholar

3. O. Rana, M. Warnier, T. B. Quillinan, and F. Brazier, "Monitoring and reputation mechanisms for service level agreements," in *5th International Workshop on Grid Economics and Business Models, GECON '08*, vol. 5206 of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, pp. 125–139, 2008.View at: Publisher Site | Google Scholar
4. N. Yuhanna and M. Gualtieri, "The Forrester Wave™: Big Data Hadoop Cloud Solutions, Q2 2016 Elasticity, Automation, and Pay-As-You-Go Compel Enterprise Adoption of Hadoop in the Cloud, (June 20)," https://ncmedia.azureedge.net/ncmedia/2016/05/The_Forrester_Wave__Big_D.pdf.View at: Google Scholar
5. R. Arora, "An introduction to big data, high performance computing, high-throughput computing, and Hadoop," *Conquering Big Data with High Performance Computing*, pp. 1–12, 2016.View at: Publisher Site | Google Scholar
6. F. Pop, J. Kołodziej, and B. D. Martino, *Resource Management for Big Data Platforms Algorithms, Modelling, and High- Performance Computing Techniques*, Springer, 2016.
7. M. Gribaudo, M. Iacono, and F. Palmieri, "Performance modeling of big data-oriented architectures," in *Resource Management for Big Data Platforms*, Computer Communications and Networks, pp. 3–34, Springer International Publishing, Cham, 2016.View at: Publisher Site | Google Scholar
8. G. Hesse and M. Lorenz, "Conceptual survey on data stream processing systems," in *Proceedings of the 2015 IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 797–802, Melbourne, Australia, December 2015.View at: Publisher Site | Google Scholar
9. D. Singh and C. K. Reddy, "A survey on platforms for big data analytics," *Journal of Big Data*, vol. 2, Article ID 8, 2014.View at: Publisher Site | Google Scholar