International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

A Study on Difference between Hadoop and Micro Services

Abhishek Matal
Master of Computer Application
Bharti Vidyapeeth Institute of Management and Information Technology.
University of Mumbai

Prof. Niddhi Poonia
Dept.: Master of Computer Application
Bharti Vidyapeeth Institute of Management and Information Technology.
University of Mumbai

Abstract-

The cloud is an emerging paradigm which leads the way for different approaches and standards. The architectural styles are evolving based on the requirements of the cloud as well. In recent years microservices is seen as the architecture style for scalable, fast evolving cloud applications. As part of this paper, a systematic mapping study was carried out around microservices.
Keywords: Cloud ,SOA ,Web services ,Microservices ,Systematic mapping.

Big data has earned a place of great importance and is becoming the choice for new researches. To find the useful information from massive amount of data to organizations, we need to analyze the data. Mastery of data analysis is required to get the information from unstructured data on the web in the form of texts, images, videos or social media posts. This paper presents an overview on big data, advantages and its scope for the future research. Big data present opportunities as well as challenges to the researchers.

Keywords— big data; Hadoop; Map Reduce; HDFS; data mining

INTRODUCTION-

Service-oriented architecture (SOA) has emerged as a means of developing distributed systems where the components are stand-alone services . Services are basic units which are developed independently and made accessible over the Internet. Standard internet protocols are used for service communication among different computers. SOA provides many advantages to develop easy and economic distributed software systems and, therefore, it is the leading technology for interoperability on today's internet world. Service-oriented software engineering defines evolution of existing software engineering approaches to develop dependable and reusable services considering the requirements and characteristics of this technology. Service-oriented computing(SOC) is the paradigm that utilizes services as the fundamental elements for developing applications. Therefore, service-oriented software engineering aims at designing and developing service-based applications consonant with SOC paradigm and SOA principles using software engineering methodologies

Big data is a vague topic and there is no exact definition which is followed by everyone. Data that has extra-large volume, comes from variety of sources, variety of formats and comes at us with a great velocity is normally refer to as big data. Big data can be structured, unstructured or semi-structured, which is not processed bythe conventional data management methods. Data can be generated on web in various forms like texts, images or videos or social media posts. In order to process these large amount of data in an inexpensive and efficient way ,parallelism is used .there are four characteristics for big data. They are volume, velocity, variety and veracity

LITRETURE REVIEW:

This part of the paper presents selected literature discussing microservices and comparing its performance to other architecture such as monolithic and service-oriented architectures. Other researchers' work is also included containing fundamental concepts and ideas about microservices architecture and microservices architecture .

In a research that was done by Singh and Peddoju, the performance of a monolithic application is compared to a microservices application that they developed and compared the performance of the two applications, their tests consisted of 2000 threads. Their results exhibited that microservices architecture has a better performance in terms of throughput when it is used for a large number of requests [1].

Microservices are often compared to Service Oriented Architecture as Mark Richards did in his book. The comparison of microservices' architecture with Service Oriented architecture in term of service and architecture characteristics as well as architecture capabilities are presented in [2].

Villamizar et al. compared the costs of using the cloud to run web applications with different architectures such as microservices, monolithic and Amazon Web Services Lambda architectures. In addition to cost comparison, they also compared response time of each application that they created, which showed that response time increased when microservices architecture was utilized compared to the monolithic architecture because each request must go through the gateway to every microservice in the system [3].

In another paper, the performance of microservices in container-based and virtual machine (VM)-based environment was compared. Amazon cloud environment was applied to conduct their experiments, and they compared the performance of these environments regarding throughput, response time, and CPU consumption. This paper concludes that VM-based environments on Amazon cloud services outperformed container-based environments on Amazon cloud environment, especially concerning response time where VM-based environment showed a better performance of 125% over container-based environment [4].

Docker is a widely used container-based virtualization software that is utilized in microservices architecture and uses Linux containers for the operating system virtualization [5].

The main reason behind using Docker for microservices is the minimal impact of its imposes on processing, memory, and network [6].

Service discovery is an essential component of any microservices application because the location of a microservice is not assigned at the design stage. Also, it may be deployed in a cloud-based environment which means services could relocate and replicate at production systems [7]

HADOOP FRAMEWORK ARCHITECTURE:

Hadoop is open source software used to process the Big Data. It is very popular used by organizations/researchers to analyze the Big Data. Hadoop is influenced by Google's architecture, Google File System and MapReduce. Hadoop processes the large data sets in a distributed computing environment. An Apache Hadoop ecosystem consists of the Hadoop Kernel, MapReduce, HDFS and other components like Apache Hive, Base and Zookeeper

A. Hadoop consists of two main components:

1) Storage: The Hadoop Distributed File System (HDFS): It is a distributed file system which provides fault tolerance and designed to run on commodity hardware. HDFS provides high throughput access

to application data and is suitable for applications that have large data sets. HDFS can store data across thousands of servers. HDFS has master/slave architecture .Files added to HDFS are split into fixed-size blocks. Block size is configurable, but defaults to 64 megabytes.

2) Processing: MapReduce : It is a programming model introduced by Google in 2004 for easily writing applications which processes large amount of data in parallel on large clusters of hardware in fault tolerant manner. This operates on huge data set, splits the problem and data sets and run it in parallel.

Two functions in MapReduce are as following:

a) Map -The Map function always runs first typically used to filter, transform, or parse the data. The output from Map becomes the input to Reduce.

b)Reduce -The Reduce function is optional normally used to summarize data from the Map function.

ADVANTAGES AND DISADVANTAGES HADOOP FRAMEWORK ARCHITECTURE :

❖ Advantages Of Hadoop

- Fast Retrieval because of the fact that the data distributed over the cluster is mapped which results in fast retrieval.
- Hadoop is easily scalable by just adding nodes into the cluster.
- Being an Open-Source, its biggest advantage is that it is cost effective compared to other database management systems.
- Hadoop is resilient to failure by replicating the data thrice by default which in case if ate is lost, it still has three copies of it.

❖ Disadvantages Of Hadoop

- When it comes to security of data, Hadoop might not be the at the top of the list since it is a really complex architecture and its security is off by default and if an amateur handle it and doesn't turn it on, then data might be at risk.
- Being written in Java, which is heavily exploited by cybercriminals, the Hadoop architecture is said to be vulnerable by nature.
- Hadoop is meant to solve problems of Big Data and it does that beautifully but when it comes to

small data, Hadoop might not be the best solution.

- Since it is open-source, it has stability issue which forces users to keep it up-to- date with the latest version.

MICROSERVICES FRAMEWORK ARCHITECTURE:

In order to compare the performance of two different architectures, first of all, there should be an application that can produce the results of microservices and monolithic applications, which can be compared and evaluated. This paper presents results created by a development platform known as JHipster utilized to generate web applications that consist of Spring Boot and Angular JS frameworks. The application that was developed for this particular paper consisted of three services.

A typical JHipster application will include three components

Two test scenarios were set up to compare the performance of the microservices application and monolithic application. Also, another test scenario was created to compare the effect of different technologies on the performance of the microservices application.. Therefore, it means dividing the total number of processed requests by the time it took to process all the requests.

The first test scenario is load testing which is used to monitor the effects of increasing the number of users on the application and how it will affect throughput and response time. It starts with 100 threads with a ramp-up of 2 minutes and holds time of another 2 minutes to analyze the concurrency of the system, then increases the number of threads until 7000 threads each time with 2 minutes for rampup and holds time.

The second test scenario is concurrency testing which is used to check how the system will hold up if all the services are used at the same time, so the test was designed to send requests to each service through their exposed APIs at the same time. It started with 100 requests for each service with no specific ramp-up time and increasing the number of requests gradually until 1000 requests. Docker was utilized to run the applications on containers, and the server that was used to run Docker environment had 16GB of memory and 2.60 GHz of CPU.

ADVANTAGES AND DISADVANTAGES OF MICROSERVCE FRAMEWORK ARCHITECTURE :

❖ Advantages Of Microservices

- Microservices are more independently deployed and allow for more team autonomy. Specific microservices can be assigned to specific development teams which allows them to focus solely on one service.
- Microservices are independently scalable, resources required can be increased for the particular services(s) rather than increasing the entire system.
- Microservices has lower downtime since it supports fault isolation. When a service fails, other services will still be working which prevents the entire system from crashing.

❖ Disadvantages Of Microservices

- Working with so many services in Microservices is a difficult task, it creates complexity for the development teams working with Microservices compared to monolithic architecture.
- Interface Control is critical because, making changes to the API can result in crashes if the services are not backward compatible.
- Microservices being a complex architecture, the total cost including skilled workers, infrastructure, security and machinery can be much higher for Microservices to work.

REFRENCE

[1] V. Singh and S. K. Peddoju, "Container-based microservice architecture for cloud applications," in 2017 International Conference on Computing, Communication and Automation (ICCCA), 2017, pp. 847–852.

[2] M. Richards, "Microservices vs. Service-Oriented Architecture," p. 57.

[3] M. Villamizar et al., "Cost comparison of running web applications in the cloud using monolithic, microservice, and AWS Lambda architectures," Serv. Oriented Comput. Appl., vol. 11, no. 2, pp. 233–247, Jun. 2017

[4] T. Salah, M. J. Zemerly, C. Y. Yeun, M. Al-Qutayri, and Y. AlHammadi, "Performance comparison between container-based and VM-based services," in 2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN), 2017, pp. 185–190.

[5] "Docker," Docker. [Online]. Available: https://www.docker.com/. [Accessed: 18-Jul-2018].

[6] N. Kratzke, "About Microservices, Containers and their Underestimated Impact on Network Performance," ArXiv171004049 Cs, Sep. 2017.

[7] F. Montesi and J. Weber, "Circuit Breakers, Discovery, and API Gateways in Microservices," ArXiv160905830 Cs, Sep. 2016