

SERVER-LESS REAL TIME DATA PROCESSING APPLICATION

Abstract: - Big data is key factor for success for many business and service domains, as it provide timely analytics over large set of data or big data. Examples of these domains are service, transport, entertainment, energy etc. Many big data applications works in these domains to provide fast and timely analysis of real time data produces. But it is of no worth if data is only stored or analyzed one at a time. For producing more efficient outputs for real time data in these domains, the storing and analysis of big data should be done in parallel so that it can produce more precise output and saves time for storing and processing of these data. This project is designed to demonstrate this concept by making an application which will store and analyze the real time data with the help of AWS.

Keywords: Big Data, AWS, Ride sharing, Real-time data processing.

I. INTRODUCTION

In today's modern data world, data is main commodity which is required to store and process for the understanding the needs for customers in market and for business organization to grow. The data which is beyond the storage capacity and the data which is beyond the processing power are called **Big Data**. **Big Data** refers to data collected from various sources like sensors, social sites applications, or any other mediums in huge amount [1]. Big data can't be stored in traditional databases

due to its massive size. Now-a day's millions of individual produces data which will contribute in increasing the size of data which leads to form big data. Companies stores these data to understand the recent trends in market, likes and dislikes of customers in live market and make applications accordingly. With the help of big data the performance in domains likes health care, entertainment, luxury, security etc. can be enhance and provide better services in domains.

The application should be capable enough to handle such massive data so that it can make fast decisions and analysis. Delay in decision making leads to affect the performance of these applications in analyzing big data and output produce will of no use.

With the help of cloud computing, the processing of big data can be done easily and saves time for big data engineer to process large set of data. Cloud computing provides many services for processing of big data and also it can store data present in petabytes or may be larger.

II BACKGROUND STUDY

The Big data refers to data in large size which traditional database can't handle or store [5]. Big data have characteristics like:-

Velocity states that how fast data can be generated, produced, created, or refreshed. It defines the speed at which data comes from different applications or sources.

Volume defines the amount of data in big data. Basically it gives idea about the size of data produced. The data may be in petabytes to exabytes. Till 2020, 40000 ExaBytes of data will be produced.

Variety defines the nature of data present in big data. The data can be in structured, unstructured or semi-structured in nature from different sources. Data may have different formats, but will store at a common place.

Cloud Computing :- It is the process of using remote servers on the internet to store, manage and process data rather than a local server or a P.C [6].

Amazon Web Service is secure cloud computing platform which offers functionalities like compute power, data storage, content delivery, and many other services which help business to scale and grow and pay only for what they use and how long they use services of AWS. It provides on-demand cloud computing services to individuals, companies or government, and charges for services which they use.

Service models of cloud computing [7]:-

- 1) IaaS (Infrastructure as a Service)
- 2) SaaS (Software as a Service)
- 3) PaaS (Platform as a Service)

Advantages of Cloud Computing:

- 1) No upfront cost.
- 2) Better security.
- 3) Maintenance free.
- 4) Access Anywhere

5) Fast Implementation

6) Instant Scalability

Real Time Data Streaming :- It is big data technology which is used to query the continuous data that is generated from different sources and detects pattern from the data within small time period from the time of receiving the data [8]. Real-time data analytics, streaming analytics, Complex Event Processing, real-time data streaming analytics, and event processing are some other names for real time data streaming. Real time processing helps in many fields and knows much information about the behavior and activities of people in real or live market.

Real time data helps to create productive and profitable projects. Some advantages of Real time data processing are:-

- It saves money in developing many applications in different fields.
- Increases accuracy of application
- It improves decision-making and crisis management in application.
- It provides immediate resolution to problems.

Real Time data Analytics is the analysis of data as soon as data enter in database from different sources. Basically it is use of or capability of process the data which continuously generated and store in database.

It helps to analyze the pattern in data and make decisions accordingly.

Real time data is information that is obtained or delivers after it is collected and stored. Real time data is used in many fields like navigation system, live tracking, location sharing, comments on social media, likes dislikes etc.

Cloud Services plays an important role in making an application which will store and

process real time data. The cloud computing services also provides functionality to make it server less so that, there is no need to maintain the server from developer side as cloud services will take care of it.

III. METHODOLOGY

The problem with earlier system is that the real time data is store and analyze by different tools and technology, thus organization have make a relation between these technologies in order to process big data. They store data in databases or in cloud and sent this data to hadoop clusters or big data tools like hive, pig, spark etc. to process it. But this problem can solve, adopting an idea by which both the process can be parallel.

This idea can be demonstrate by an application which will store and process real time data with the help of AWS Cloud computing like kinesis, Athena, kinesis data firehose etc. So in this section Architecture and methodology of this application will discuss.

1. ARCHITECTURE

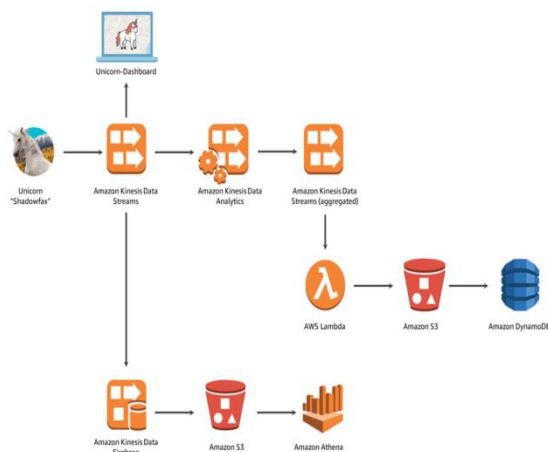


Fig. 1 Architecture diagram for real time data processing application.

In this architecture diagram, the services used are Athena, Kinesis, kinesis data firehose, Lambda function, Dynamodb as database, and Amazon S3 bucket service for temporal storage of data in batched form.

Services used for storing of data are Amazon kinesis data stream, lambda function, dynamodb, Amazon Kinesis analytics application. And services used for processing of real time data are Amazon kinesis data firehose, Amazon S3, Amazon kinesis stream and Amazon Athena.

In this project, the data of ride sharing company is use. The company provides rides on their unicorns equipped with IOT sensors which will send data about unicorn health, status and other information to headquarters of company. This application is totally server less and applications used in this project are also serverless. As Server less application doesn't require to provision, scale, and manage server by developer or organization. The cloud will automatically and timely manage servers and scale up and down according to traffic comes on application. Server less application provides benefits and work more efficiently as compare to server dependent applications.

2. MODULES

The project is divided in to 4 modules:-

a) Build a kinesis stream :- In this module, data stream is created in kinesis for unicorns on the live map. Also, for granting permission to access the stream on live map, Amazon Cognito identity pool is created.

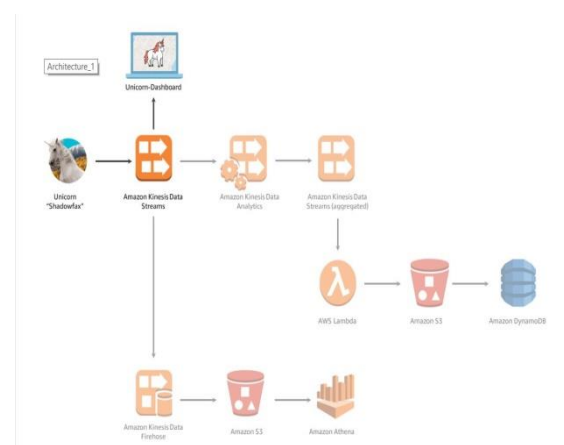


Fig. 2 Architecture for first module in project

b) Aggregation of data :- In this module, Kinesis data analytics application is developed

which will read data from stream and aggregate information like unicorn health, distance travelled in each minutes.

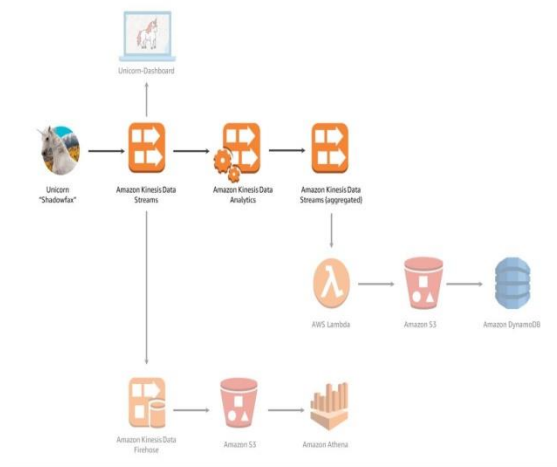


Fig.3 Architecture consists of kinesis data analytics application, source- destination kinesis steams, producer and consumer Command line clients.

c) Processing of Data Streams:- In this module Lambda function is used to process data from kinesis stream. It is created and configures to read data from stream and write the processed data into dynamodb database.

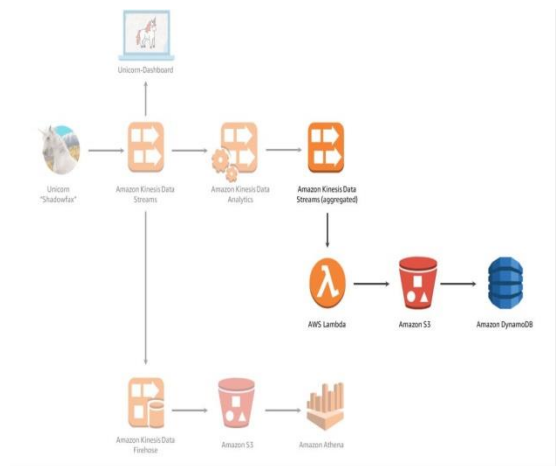


Fig. 4 Architecture for third module consist Lambda function, Dynamodb table, and stream data in aggregate form.

d) Store and Query data :- this is last module in which amazon kinesis data firehose is created to deliver the stream data to S3 bucket in batches. From S3, Athena will take data and process it in form of queries given by backend developer. This service will give the output of queries from real time data in merely 3-5 seconds.

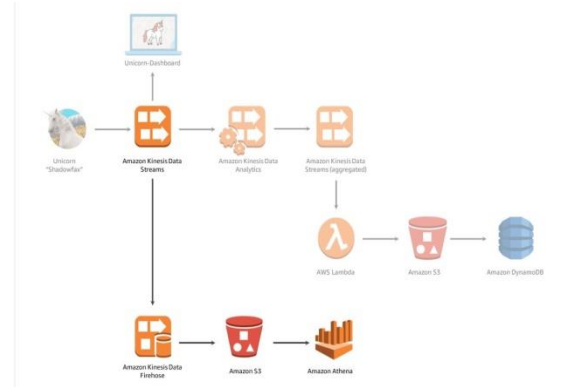


Fig. 5 Architecture consists of Athena service for query from data, S3 bucket to store data from kinesis data firehose.

IV APPLICATIONS

Some applications of real time data processing applications are:-

1) Intelligent Transportation

It is one of the most important big data applications. Various sensors are used to control and monitor traffic and analysis is done at very speed from these sensors. road monitoring cameras, road tube axle sensors, vehicle inductive loop sensors, capacitance mats, and piezoelectric axle sensors are some examples of road sensor devices [2].

2) Early warning for natural calamities or Disasters

Big data applications are very useful in predicting weather analysis and give warning in advance if some natural disaster will happen in future. By these application a lot of people's live can be save. Various sensors are installed in geographical regions where natural calamity are likely to happen [3].

3) Military Decision Making

In wars, the key to success is not power alone, but also the ability to collect information about the current scenario and make decisions quickly. If someone does this quickly, he wins [4]. In wars various items used like tanks, military vehicles, submarines, Guns, and other items. These equipment are added in resources for war and have sensors equipped on them.

By collecting enemy's location, movements, equipment data, one side can analyze it and make decisions quickly leading to attack first on enemy and defend themselves from attacks of enemy.

V CONCLUSION

Big data collection and processing is time consuming task and if proper decision making from the real time data or big data is not obtain on time then it is no use and got no benefits from that data. There are many challenges face in processing and storing real time data or big data in applications. Most of the existing applications are developed using open loop approach.

In this paper, an idea has been discussed to solve this challenge by proposing an application which will store and process real time data in minimum amount of time. This application will reduce the development and maintenance cost for big data application and due to server less in nature, there is no requirement to maintain servers from developer side as cloud will handle the servers.

VI REFERENCES

- [1]. A. McAfee and E. Brynjolfsson. "Big data: the management revolution." *Harvard business review*, Vol. 90, No. 10, pp. 60-68, 2012.
- [2]. R. M. Tyburski, "A review of road sensor technology for monitoring vehicle traffic," *ITE (Institute of Transportation Engineers) Journal*; (USA) 59, no. 8 (1988).
- [3]. A. Biem, E. Bouillet, H. Feng, A. Ranganathan, A. Riabov, O. Verscheure, H. Koutsopoulos, and C. Moran, "IBM infosphere streams for scalable, real-time, intelligent transportation services," In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pp. 1093-1104. ACM, 2010.
- [4]. D. J. Feith, "War and decision," *HarperCollins*, 2009.
- [5]. WANG Shan, WANG Hui-Ju, QIN Xiong-Pai, ZHOU Xuan. *Architecting Big Data: Challenges, Studies and Forecasts* [J]. *Chinese Journal of Computer*. 2011, 34(10): 1741-1752.
- [6]. Liu Peng. *Cloud Computing* [M]. Beijing: Electronic Industry Press, 2011.
- [7]. Koustabh Dolui, Soumya Kanti Datta, "Comparison of Edge Computing Implementations: Fog Computing, Cloudlet and Mobile Edge Computing" in *June 2017 Global Internet of Things Summit (GloTS)*.
- [8]. K. Wähner, "Real-Time Stream Processing as Game Changer in a Big Data World with Hadoop and Data Warehouse," *InfoQ*. pp. 1–9, 2014.