IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Special Issue, 4, Feb 2019 E-ISSN: 2347-2693

Real-Time Big Data Analytics: Applications and Challenges

R. Uthra[1*], A. Mahalakshmi [2]

[1,2]M.Sc.,(Computer Science) RabiammalAhamed Maideen College For Women,Thiruvarur,Tamilnadu, India

Corresponding Author: uthrarajan6@gmail.com

Available online at: www.ijcseonline.org

***Abstract*—**In recent years, time-significant processing or real-time processing andanalytics of big data have received a huge amount of attentions. There are many areas/domains where real-time processing of data and making timely decision can saves thousands of human lives, minimizing the risks of human lives and resources, enhance the quality of human lives, enhance the chance of effectiveness, capableresources management etc. This paper has describedreal-time big data analytics applications and the tools used and the technical challenges faced by the applications. In addition it presents a general overview of big data to describe a background knowledge on this extent. Some examples of these domains includesecurity, healthcare, transportation, military, and natural disaster. Several big data applications in these domains rely on fast and timely analytics based on available data to make excellence decisions.

***Keywords*—**bigdata,applications,tools,challenges

I. INTRODUCTION

Big data term is used for huge data sets having the characteristics of large in volume, variety in data, complex structure, and high motion of data. These data set are generated from various sourcessuch as, telecommunication, biological science, medical science, scientific research, financial services, military/defense intelligence, emails, audio, video, images, search queries, various online transaction, etc. These data can be used for company or organization by analyzing them properly and obtaining meaningful information from them. The Big data applications refers to the distributed applications. But if the data analytics can be done in real-time, a significant amount of benefits can be achieved. That's why, in recent time, a real-time big data application have gained a serious attention for generating a timely response. A real-time big data analytic application is an application program that process within a time frame and generate a faster response.In real-time big data analytics application, big data need to be analyzed and executed in a timely manner as accurately as possible to generate a fast responseOr making a real-time decision correctly. But successful implementationof these application is really a very challenging task, mainlybecause of its real-time computation/processing.This paper outlines some important areas where real-time big dataanalytics application is very useful and demanding.

II. BACKGROUND

A general overview about big data has been presented in this section.Mainly, big data definition and it characteristics, have been discussed in this section. The term "Big data" refers to data sets that typical database management beyond system storage, managing, and analyzing data due to their massive sizes.Big data sizes range from a few dozen terabytes to yottabytes. Big data applications generally seek new knowledge and gather intelligence from the data and convert that into business advantages in terms of enhancing operations and profitability and reducing risks and overheads. In addition, utilizing big data can provide a number of new types of advanced services that have just recently been introduced or will be introduced in the near future. These services enhance the value of life and help reduce risks and threats. Big data differs from regular data in six characteristics: volume, velocity, and variety,veracity,variability,and value.

—**Volume**: It refers to the size or quantity of data which is larger than terabytesand petabytes or even exabytes.
—**Velocity**: It refers to the high speed of data, that means, how fastdata is being produced and how fast data is being processed..
—**Variety**: In big data, data is collected from different sources and usually has three types: structured, semi-structured, and unstructured
—**Veracity:** It refers to the genuineness of the data
—**Variability :**This is factor refers to the inconsistency which can be shown by the data at times. This can hamper the process of being able to handle and manage the data effectively.
—**Value**: It refers to the quality of data for generating the intended result.

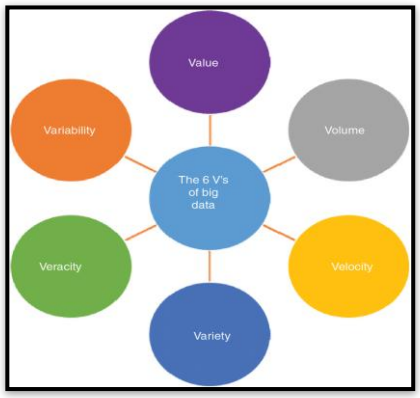

Figure 1- 6 V's characteristics of big data

III. RELATED WORKS

In paper [3], Data type and amount in human society is growing in amazing speed which is caused by emerging new services as cloudcomputing, internet of things and location-based services, the era of big data has arrived. As data has been fundamental resource, how to manage and utilize big data better has attracted much attention. Especially, with the development of internet of things, how to processing large amount real-time data has become a great challenge in research and applications. Recently, cloud computing technology has attracted much attention with high-performance, but how to use cloud computing technology for large-scale real-time data processing has not been studied. This paper studied the challenges of big data firstly and concludes all these challenges into six issues. In order to improve the performance of real-time processing of large data, this paper builds a kind of real-time big data processing (RTDP) architecture based on the cloud computing technology and then proposed the four layers of the architecture, and hierarchical computing model. This paper proposed a multi-level storage model and the LMA-based application deployment method to meet the real-time and heterogeneity requirements of RTDP system.

In paper [4], With Increase in use of technology the data generated from different source also increasing from past years. The rapid growth of data creates an opportunity to analyze and take decision or action. It is important to make use of it, which in terms gives many advantages in different areas like intelligent transport system, Emergency response, Military decision making, Stock market, Cyber security, Internet of things etc. Since all the listed applications areas are highly dynamic it is very important to take decision in real time using historical data and real time data. In overall collecting real-time data to take decision in real-time or near to real time is challenging. The use of emerged technologies and tools helps in the processing and analyzing of data in timely manner, which helps in getting significant benefits from huge data.

IV. REALTIME PROCESSING FRAMEWORK

In addition to powerful computing ability, real-time big data processing system must have strong timeliness which means it must quickly respond tothe request from systemterminals in a very short time delay. This paper, according to the demands on the computing ability of real-time big data processing system and the timeliness, divides the RTDP (Real-Time Data Processing) framework into four layers–Data, Analytics, Integration and Decision from a functionallevel.

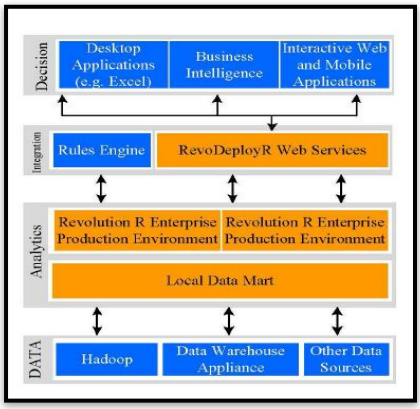

Figure 2- Architecture of Realtime processing

1. Data - This layer mainly used for data collection andstorage, but also including data cleaning and some simple data analysis, preparing data for Analytics.
2. Analytics - This layer is the corelayer of real time data processing system and the critical layer to determine the performance of real time data processing system.
3. Integration - This layer plays a connecting role in RTDP system. In this layer, many common data processing algorithm packages are combined.
4. Decision - This layer makes decisions with the results of dataanalysis - the highest layer of data processingsystem as well as the ultimate goal of data analysis process.

V. APPLICATIONS

A. TRANSPORTATION

One of the most important applications of real-time big data analytics is enabling intelligent transportation systems. These days, different sensing technologies are available to monitor traffic conditions in big and crowded streets and cities. These sensing technologies are divided into two categories: road sensors and vehicle sensors. Examples of road sensors are road monitoring cameras, vehicle inductive loop sensors, and capacitance mats. Vehicle sensors include on-board cameras, GPS systems, proximity sensors, and speedometers. The

roads and vehicular sensors can generate big data that can be used to provide advanced intelligent transportation services in big cities. Current GPS systems offer drivers best route information; however, many cannot respond quickly enough to sudden changes in traffic conditions such as accidents or roadblocks. Another advanced service that can be provided for goods delivery drivers is a real-time and dynamic solution of the traveling salesman problem for defining the order and routes of the delivery points such that delivery time is minimized.

B. HEALTH CARE

Clinical research in real-time big data refers making correct prediction in real-time so that physicians can provides better treatment and fast accurate decisions to their patients by analyzing patients data in a timely and reliable manner. The amount of data produced within medical area or in clinic has grown to be huge in volume, where analysis of of those data and generating time response can improve the quality of health care for the patients.However, there are a number of challenges that arise when dealing with these huge quantities of data such as, how to analyze this data in a reliable manner and generating real-time result to offer right treatment to the patient in real-time. By making it possible, the risk of human life can be minimized. The main purpose of health care base application is to provide real-time health care to the end users (i.e. patients) by collecting the real world medical data from all levels of human existence and analyzing them in a reliable manner in real-time.

C. NATURAL DISASTERS

There are significant number of natural disasters the world faced so far, which costs huge number of human lives, health, economies and various resources. Example of such natural disaster includes earthquakes, floods, tsunami, cyclones, volcanoes, etc. Early predicting and warnings of such natural disasters can save the lives of thousands of people and resources. The early warnings can includeinformation about shelter, to-do, required action, emergency support, etc. But this early warning systems for natural disaster involve real-time processing of huge amount of distributed data that also collected in real-time. For this type of data collection various sensors and GPS or Satellite technology can be used. However the main challenging tasks is to analyze those data in timely fashion to provide early warnings.

VI. TOOLS

As technology has many advanced tools and techniques , the selection of right tool will give the good advantage in different kind of applications.

SPLUNK

Splunk is aanalytics tool. It creates an index of the data as if the data was a book or a block of text. though databases build indices, Splunk's approach resembles more to a text search process. This indexing makes more flexible. Splunk tool is already tuned to a particular application, making it easier to make out the log files. The index helps to correlate the data in these and otherseveral common server-sideapplications.

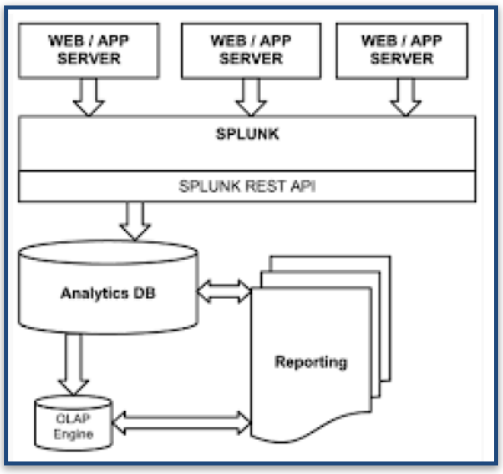

Figure 3: Splunk architecture

Splunk will take text strings and search approximately in the index. Splunk finds the URLs one wish to find and packages them into a timeline built around the time stamps, it discovers in the data. In the diagram , the data is fetched from the web servers and carried to the Splunk tool. This process data then transferred to the analytics database. The analyzed data is then transported to OLAP engine.

PRESTO

Presto is an open source software developed by Facebook to query and analyze large amount of data collected every day. Presto tool is mainly designed to querying large amount of data with distributed queries. If data is terabyte or Exabyte, it is more like to use presto with Hadoop and HDFS. It is a distributed SQL query engine particularly designed to run interactive analytical queries of all size against different sources of data. Since Presto has been used by Facebook from years, it becomes more successful in Hadoop field. For beginners Presto is mainly developed to gratify the response time which range from milliseconds to minutes. The are many platform increasing which includes Postgres, MySQL, Cassandra and Kafka which means this allow query to search where the data is present. A single Presto query will combine multiple datasource and allows analytics across entire organization. Presto is best for Interactive queries, to explore data promptly and suitable for joins with many small dimension tables and large Fact tables.

Some of the advantages of using tool in real time data analytics are as follows:

- Tools helps to integrate with many other technologies.
- Tools are having great advantage of effective data visualization.
- Tools are highly capable of using structured and unstructured data generated from different sources.
- Tools helps analyst to take effective decision hastily with many advanced algorithms and cross platform computation.

VII. CHALLENGES

There are a many number of challenges facing the design, implementation and operations of real-time big data applications. Most regular big data applications are implemented using an open-loop approach.Unlike regular big data applications, real-time applications must commence fast actions that are usually bounded by specific time frames dictated by the targeted domain. Real-time big data applications are typically implemented in a closed-loop approach in which actions are usually based on the present and prior situations. To understand the challenges of designing, implementing, and operating real-time big data applications, we need to understand their common action steps which can extensively impact the action completion times. Action times in real-time big data applications are divided into five steps.

Real-time event transfer
All contemporary distributed application events should be transferred in real-time to where they can be processed. These events can be transferred from distributed sources as raw events or as filtered or aggregated events.

Real-time situation discovery
This step is considered to detect real-time business or operational situations and exceptions in the current events. Real-time situation discovery processes can define and modify the policies used to filter and aggregate event transfer. In addition, it can activate analytical or decision making processes.

Real-time analytics
This involves invocations of real-time analytical services to determine the core causes for business and operational situations and exceptions. Real-time analytical processing involves single or multiple integrated analytical services.

Real-time decision making
Based on the real-time analytics results , it is possible to select the finest option for civilizing the current business operations or effectiveness and determine the most suitable actions for a response to the business or operational environment. The most challenging part is how to define business and operational rules that makes possible , correct and timely decisions making.

Real-time responses
This involves initiating, executing and monitoring an action defined by the real-time decision making process. This process usually needs to intermingle with different distributed systems to perform its tasks. These interactions should be consistent and achieved in real-time.

VIII. CONCLUSION

Big data collection and analysis requires long period of time to complete. As a result real-time applications cannot instantlygot benefit from big data analytics. However, many situations and applications have to deal with big data and produce results in real-time. In this paper we explored different application domains that could benefit from big data applications operating in real time. Understanding the nature of these applications and the challenges in advance is the first step to create more effective and efficient applications in this domain. The next step is to identify the most appropriate solutions to these challenges and building real time big data applications that are consistent and competent of meeting the real time demands. The successful implementation of these applicationwill provides a significant amount of benefits to human,society, and the world However, to get a full advantage of these Applications, the main challenging areas have to be addressed firstlike, real-time data collection, synchronization of incoming datawith current data, extracting meaningful information from a large Amount of data, and generating timely response.

REFERENCES

[1] Nader Mohamed, Jameela Al-Jaroodi, "Real-Time Big Data Analytics: Applications and Challenges" International Conference on High-performance Computing & Simulation (HPCS), 2014
[2] Big Data Real Time Analytics: Applicationsand ToolsJayanth Kumar K1, Anisha B.S
[3] Real-Time Big Data Processing Framework: Challenges and Solutions Zhigao Zheng1,2, Ping Wang1,3,4,∗ and Jing Liu3 and Shengli Sun 1
[4] Big Data Real Time Analytics: Applications and Tools Jayanth Kumar K1, Anisha B.S.2
[5] B. Balis, T. Bartynski, M. Bubak, G. Dyk, T. Gubala, and M. Kasztelnik, "A Development and Execution Environment for Early Warning Systems for Natural Disasters," In 2013 13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), pp. 575-582. IEEE, 2013.
[6] Real-timebig data analytics-applications and challenges-Nader Mohamed Jameela Al-Jaroodi
[7] A Thommandram, JE Pugh, JM Eklund, C McGregor, andAG James. Classifying neonatal spells using real-time temporalanalysis of physiological data streams: Algorithm development.
[8] In IEEE Point-of-Care Healthcare Technologies(PHT 2013), pages 240–243, New York, USA,Bangalore, India,2013. IEEE.

