Big Data Analytical Tools for IoT in Healthcare

[1] Ayshathul Thuhara, [2]Lubna Mohammed Kunhi, [3]Rahamath, [4]Mustafa Basthikodi
[1,2,3] Student, [4]Professor, Dept. of CSE, Bearys Institute of Technology Mangalore.

Abstract - The commonness of the Internet of Things (IoT) in healthcare produces a massive amount of heterogeneous data from different types of sensors that are used by the patients in a day to day life. Therefore, to process such massive amount of heterogeneous data in an emergency situation the most challenging tasks. There are different technologies to manage emergency details such as Hadoop, Spark, Storm. This paper mainly deals with the comparison on Hadoop, Storm, Spark in medical data handling and processing.

Keywords - Hadoop, Storm, Spark, Big data, Healthcare, IoT

I.INTRODUCTION

Healthcare is one of the most important sectors in the developing world. As the world is developing, growth in the health sector should also improve. As a result the Internet of Things (IoT) on health care is introduced. People are using sensors, connected devices and so on in order to get necessary information about their health and necessary hospitality. Sensors such as medical sensors, activity sensor and coordinator sensor are implanted to human body parts. These sensors collect desperate medical data such as blood pressure, diabetes, body temperature, glucose level, sweating and other medical measurements such as sitting, walking, cycling, climbing upstairs, downstairs. Smart devices in health care such as smart pills, CYCORE, Smart CGM (continuous glucose monitoring), smart insulin pens, Smart inhalers, Smart contact lenses and so on helps in generating health factors.

Data generated from these devices are of large quantity. To handle these data and to process it in securing manner is a challenging task .Big Data analytical frameworks are used to handle these healthcare data .Each framework has its own properties and is good with some properties. Speedy and convenient data transfers are two considerations that motivate organizations of heath care to explore IoT technologies. The flood of data generated by IoT devices and gadgets used in healthcare industry could cause unseen problems if the organizations are not equipped to handle it properly and verify its quality.

This paper presents a comparison on three data frameworks Apache Hadoop, Apache Spark and Apache Storm. The comparison of the performance, reliability, security, development cost of the frameworks that will be best for health sector problems.

The paper is organized in sections. In this, Section II describes Apache Hadoop used to handle health data. Then, Section III describes Apache Storm used to handle health data. Section IV describes Apache Spark used to handle health data .Comparative Table on Apache Hadoop, Apache Storm, Apache Spark presented in Section V. The discussions of comparison are described in Section VI. Finally, Section VII concludes this paper with a summary of the main points of interest in this research.

II.APACHE HADOOP

Hadoop is an open source processing framework that supports batch processing. High speed data from the healthcare are handled by hadoop ecosystem. The system handles capturing, collecting, processing and analyzing data from number of sensors implanted on human body with body area network continuously monitors health parameter. Hadoop Processing Unit is of various master nodes and data nodes. To store data in block on multiple parallel data nodes it uses Hadoop Distributed File System. Map Reduce algorithm is implemented on each data node to generate intermediate results of decision making. Based on the output generated, the machine executes required action such as first aid, remote physician, reminding patient about doctor prescriptions and so on.

III.APACHE STORM

Apache Storm is an open source stream processing framework focuses on extremely low latency and is best suitable for workloads that need real-time processing .Storm can handle large quantities of data and can deliver data with less latency compared to other frameworks .Storm streaming process works by orchestrating DAGs in a framework called as topologies. These topologies describe various steps and transformation taken on each incoming piece of data as it enters the system.

These topologies are composed of:

- Streams: It is bounded data that arrives continuously at the system.
- Spouts: Sources of data stream at the edge of topology. These can be queues, APIs, etc. which produce data to be operated.
- Bolts: It represents a processing step that takes stream, applies operation on them, and outputs results as a stream. Bolts are connected with each spouts, and connect each other for necessary processing. Final bolt output produced at the end of topology can be used as an input to connected system.

Storm can guarantee that each piece of data can be processed at least once, duplicate can be present in some scenarios. Storm with Trident can be to produce micro-batches instead of pure stream processing.

III.APACHE SPARK

Apache Spark is hybrid processing framework and is an open source like Hadoop and Storm. Apache Spark is

used to monitor and integrate heterogeneous data from Internet of Things enabled in medical devices. Large scale classification, clustering, association rule mining and predictive modeling can be done using Spark's machine learning libraries. Spark can handle data coming from streaming sources. SparkML is used to handle the big set of data coming from static data sources. SparkML perform diagnosis based on structured and unstructured data processing. Large scale classification are done using Spark's Support vector Machine, Random Forest and K-means clustering libraries. Retrieval of information is done through Spark SQL and SPARQL.

Implementation of iterations in Spark as regular for-loops and executes them by loop unrolling, which means for each iterations a new set of task will be scheduled and executed. The result of each iteration is fed to the next iteration to execute, and will be stored in memory.

Spark also runs in standalone, Mesos or in cloud, makes improvement in data analytics in future days.

IV.COMPARISON ON APACHE HADOOP, APACHE STORM AND APACHE SPARK

Criteria	Apache Hadoop	Apache Storm	Apache spark
Performance	Relatively Slow	Relatively fast	Faster by 10 to 1000 than others.
Data Store	Stores data on disk	Process data stream as it enters and sends data to other systems.	Stores data in memory
Processing	Batch processing	Stream processing	Hybrid processing systems :both batch and stream processing
Fault Tolerance	Slightly more fault tolerant	If process fails, the supervisor process restart that process automatically as it is managed by Zookeeper.	Less fault tolerant compared to hadoop
Security	Less secure	More secure than hadoop	More secure
Implementation	Implemented in Java. Complex and used with any programming language	Implemented in Clojure. Simple and used with any programming language	Using APIs for Java,Scala and Python. Lesser language support.
latency	High	Relatively less	Less
cost	Less cost for software and more infrastructure	More development cost on software.	Less cost for software and more on RAM

V.DISCUSSION

Storm is option when an application needs second latency without data loss. Storm is usually an honest alternative once interval directly affects user expertise, for instance once feedback from the process is fed directly back to a visitor's page on an internet site. Storm can be used where every data has to be processed. But Storm is incredibly complicated for developers to develop applications with limited resources. Hadoop can be used in place where the data required is not that emergency. If the health data concentrates only on security than Spark can best option. Incase if the health data are to be stored and are not bothered about other features such as security, latency, so on then hadoop is best choice.

VI.CONCLUSION

Healthcare is the important sector where mass of data are generated. These data can be analyzed by big data analytical frameworks such as Spark, Storm and Hadoop. All of these framework are easy to setup and operate. These framework provide support flexibility, real-time processing, and ad-hoc or other queries on large data sets. As I understood that the frameworks are competitive among each other comparative must be based on cases oriented view.

VI.REFERENCES

[1] Wenzhi Liu, Qi Li, Yunpeng Cai, Ye Li, "A Prototype of Healthcare Big Data Processing System based on Spark", *Eighth International Conference of Bio Medical Engineering and Informatics,*2015.

[2] V. Jagadeeswari, V.Subramiyaswamy ,R. Logesh and V. Vijayakumar, "A study on medical Internet of Things and Big Data in personalized healthcare system",2018.

[3] Renu Dwivedi, Prof. Satpal Singh, Prof. Sumit Nema, "An Exclusive Survey on Big Data Analytics, Applications and Tools", *International Journal of Innovative Research in Computer and Communication Engineering*, 2017, Issue 9, DOI: 10. 15680/IJIRCCE.2017.0509085.

[4] R K Jena,"Big Data Computing Framework: A compact Review", *International Journal of Engineering Development and Research*, 2017, vol. 5, Issue 2, ISSN: 2321-9939.

[5] Jagreet Kaur and Dr. Kulwinder Singh Mann," AI based HealthCare Platform for Real Time, Predictive and Perspective Analytics using Reactive Programming", *Journal of Physics: Conference Series*,2018.

[6] Md. Reazual Karim, Ratnesh Sahy and Dietrich Rebholz- Schuhmann, "A Scalable, Secure and Realtime Healthcare Analytics Framework with Apache Spark", *www.researchgate.net*,2015, DOI:10.13140/RG.2.1.3331.7208.

[7] Pauwels, Elenore and Denton, Sarah W, "The Internet of Bodies: Life and Death in the Age of AI", *California Western Law Review*, 2018, vol. 55, No. 1.

[8] Subhajit Chatterjee, Shreya Chatterjeee, Soumyadeep Choudhary, Sayan Basak, Srijin Dey, Suparna Sain, Kali Shreyo Ghosal, Niket Dalmia, sachet Sircar, "Internet of Things and Body Area Network- An Integrated Future", *IEEE*,2017.

[9] Stephanie Baker, Wei Xiang, "Internet of Things for smart Healthcare: Technologies, Challenges, and Oppurtunities", *IEEE*, 2017, DOI: 10.1109/ACCESS.

[10] Sreekanth Rallapalli, Gondkar RR, Uma Pavan Kumar Ketavarapu," Impact of Processing and Analyzing Healthcare Big Data on Cloud Computing Environment by Implementing Hadoop Cluster", *International Conference on Computational Modeling and Security (CMS2016)*, 2016, DOI: 10.1016/j.procs.2016.05.171.

[11] M. Mazhar Rathore, Awais Ahmad, Anand Paul, "The Internet of Things based Medical Emergency Management using Hadoop Ecosystem*",IEEE,* 2015.

[12] Ovidiu- Cristian Marcu, Alexandru Costan, Gabriel Antoniu, Marıa S. Perez-Hernandez," Spark versus Flink: Understanding Performance in Big Data Analytics Frameworks*", IEEE*, 2016, DOI 10.1109/CLUSTER.2016.22.

[13] Hongyong Yu, Deshuai Wang," Research and Implementation of Massive Health Care Data Management and Analysis Based on Hadoop", *Fourth International Conference on Computational and Information Sciences*, 2012, DOI 10.1109/ICCIS.2012.225.

[14] Wei Liu, and Dedao Gu, "Research on construction of smart medical system based on the social security card", *International Conference on Electronic and Mechanical Engineering and Information Technology (EMEIT)*, pp. 4697–4700, 2011.

[15] Burghard C," Big Data and Analytics Key to Accountable Care Success", *IDC Health Insights*, 2012

[16] Raghupathi W,"Data Mining in Health Care. In Healthcare Informatics: Improving Efficiency and Productivity", *Edited by Kudyba S. Taylor & Francis*; 2010, 211–223.

[17]"Apache Spark." http://spark.apache.org.

[18]"What is Hadoop?." https://www.talend.com/resources/what-is-hadoop/.

[19] "Apache Storm." http://storm.apache.org.

[20]Sasippriya Saminathan,K.Geetha,"A survey on health monitoring system using IoT", *International Journal of Pure and Applied Mathematics*, 2017, vol. 117, no. 17, 249-254.