ISSN XXXX XXXX © 2018 IJESC

Research Article Volume 8 Issue No.4

Big Data and CPS (Cyber Physical System) used in Pharmacy to Alert on Expiration of Medicine

S.Malathi[1], M.Priadarsini[2], M.Dharshana[3], T.Agathiya[4]
Student[1, 3, 4], Assistant Professor[2]
Department of computer Science and Engineering
Government College of Engineering, Dharmapuri, India

Abstract:
Big data is wide process that enables organization to create, manipulate and manage very large dataset and the storage environments. This analytic technology is used for managing and analyzing medication based data. The pharmacy information focuses on application technology for pharmacists to increase patient safety. The pharmacy information is growing rapidly results in big data and it enhances the development of patient-centered pharmacy service. In this paper the expiration date of medicine is alerted by sending the notification to the pharmacy before it gets expired. The pharmacy information is collected with CPS system integrated with internet in which physical and software components are intertwined. This can be done by processing the pharmacy informatics with Apache hive and map reduce algorithm.

Keywords: Big data, CPS (Cyber Physical System), Hadoop, HDFS, Hive and Map reduce Algorithm.

1. INTRODUCTION

Big data is really critical to our life and it emerging as one of the most important technologies in modern world and it is relatively new, the act of gathering and storing large amounts of information for analysis. The amount of data that's being created and stored on a global level is almost inconceivable, and it just keeps growing. Big data is data that exceeds the processing capacity of conventional database systems that inundates a business on a day-to-day basis. Hadoop is a savior of the big data world, java-based programming framework that supports the processing and storage of large data sets in a distributed computing environment. It consists of computer cluster built from commodity hardware. Basically, Hadoop is a way of storing enormous data sets across distributed clusters of server and then running distributed analysis application in each cluster. It is designed to be robust, in that your big data application will continue to run even when individual servers or cluster fail. Hadoop is also designed to be efficient, because it doesn't require application to shuttle huge volumes of data across network. Hadoop consists of two components: **HDFS**- that allows to store data of various formats and **YARN**-processing unit. The healthcare industry is exploiting with data including pharmaceutical manufactures, research centers, chemist etc… are all collecting more drug-related data. In fact more data has been created since 2003 than in all of previous recorded history. The difficulties that are faced by pharmacists are to identify the data or to check expired medicine regularly to avoid get rid of expired medicine. Because, the usage of expired medicine causes lot of side effects, some may leads to death. Here, CPS system is used to scan the bar code or QR code and to convert them into data directly to avoid manual entry and human error.

2. MAPREDUCE ALGORITHM:

Map reduce is a processing technique and a programming model that implemented for processing and generating big data sets then computing is done based on java. The map reduce programming framework use map and reduce task in functional programming. A Map reduce job usually splits the input data-set into independent chunks and this task is done by mapper function. The framework sorts the output of the map function and give it as an input to the reduce function where input and output will store in a file-system. Map reduce is a core component and considered as the heart of Apache Hadoop software framework. It is the original framework for writing application that process large amount of structured and unstructured data stored in HDFS. All data elements in map reduce are immutable, changing an input pair does not change the input files.

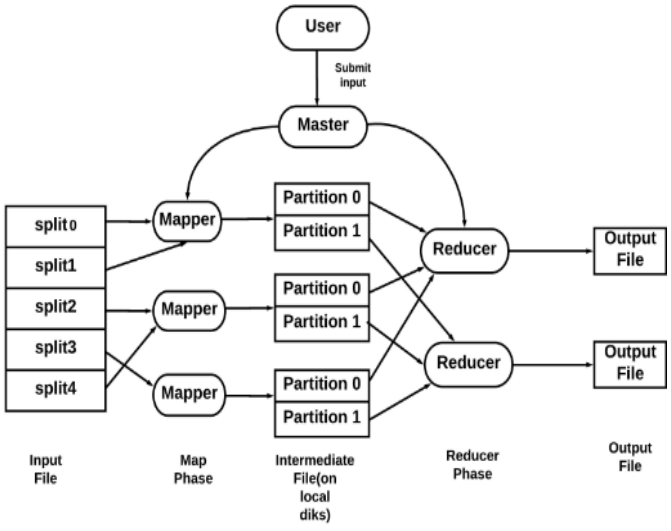

Figure.1.Mapreduce works.

Mapreduce programs transforms lists of input data elements into lists of output data elements, once for the Map and other once for the Reduce. The first phase of any Map reduce job is Map. In the Mapping phase, raw data is transformed into a set of<key, value>pairs. The reduce part of Map reduce merge values from the Map phase reduce is optional, sometimes all the work is done in the Mapper.

3.LITERATURE SURVEY:

In pharmacy the data are collected with CPS system [8] which is integrated with internet and it helps by scanning the QR code or bar code to fetch data. Therefore the pharmacy informatics is collected and stored in HDFS. [15] To process the data from various data node, map reduce algorithm should be used. [14] This method can be applied in pharmacy to identify the medicine that contains 3 to 5 days for expire [16] and then send notification to the pharmacists in detail.

4.PROPOSED SYSTEM:

The main idea of this paper is to help the pharmacy to identify the expired medicine without any manual check. A notification will be received by the pharmacists before the expired date of drugs with detailed information about drug. This can be done by using mapreduce algorithm to process and analyze the data stored in multinode. The data that are collected with CPS System is structured data and they are stored in Apache Hive for mapreduce process to identify the expired medicine. Finally the processed data is send as a notification to the pharmacists.

5. IMPLEMENTATION:

This implementation consists of following modules:

a**). DATA COLLECTION LAYER:**

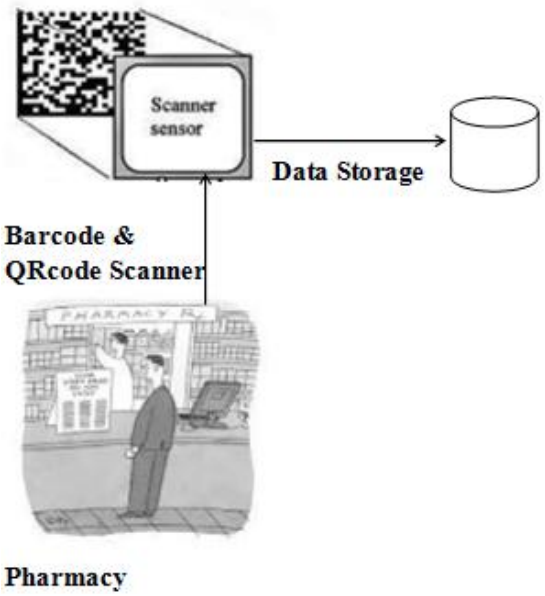

Figure.2. Data Collection Layer.

In data collection layer, medicine data that are collected with CPS system and then stored in HDFS (Hadoop Distributed File System) for further process. The types of data contained in a data layer can be numerous and varied. The data that are collected are distributed to each data node for parallel processing.

b).DATA MANAGEMENT LAYER:

Once data is collected, it should be managed in an appropriate manner. Data management act as a layer between the objects and device generating the data. It includes Data access and manipulation logic and storage design. Data management layer is designed using four-step process:

- Selecting the format of the storage.
- Mapping the problem domain classes to the selected format.
- Optimizing the storage to perform efficiently.
- Designing the necessary data access and manipulation classes.

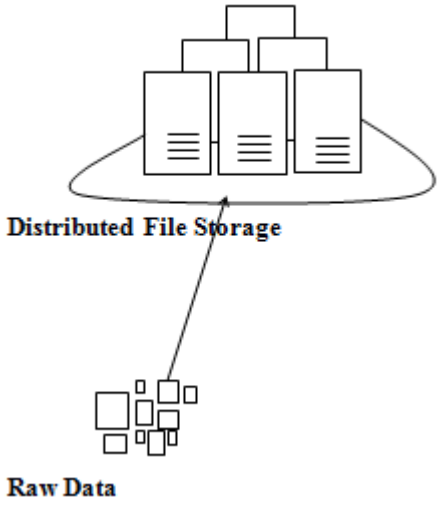

Figure. 3. Data Management Layer.

Here the data stored in HDFS need to be manage. This can be done in different ways such as by using HDFS commands line interface (CLI- also referred as FS command) or by writing mapreduce job, hive or pig queries and even by web browser to browse the files.

c).DISTRIBUTED PARALLEL COMPUTING:

The data stored in HDFS are computed by using multiple processors in parallel computing environment. The data stored in HDFS are distributed across different data nodes and operations are done in parallel. Each map task has a circular memory that is basically a buffer which is used for writing out output of each map task and then spilling output to disk will takes parallel. The thread divides the data into partition before it spilling into disk corresponding to reducers. The sorted and compact map output will be produced that needs to be transfer to reduce. The output of map task is compressed to save disk space and amount of data to be transfer to reduce. The map partitioned output will be copied directly to reduce JVM memory if it is small in numbers or else they are copied to disk as soon as by thread in parallel. Then reduce function receives map output in sorted format then it will performs it operation and write their output to HDFS. Finally. The output of the reduce function will be send as a notification to the pharmacists.

6. CONCLUSION:

The pharmacist receives notification about the medicine before the date of expiration that contains manufacture, delegated and expiration details. By using HIVE and the map reduce algorithm the data can be effectively stored and processed effectively and efficiently. This helps pharmacists to detect human error and to avoid usage of expired medicine as well as it save time. In big data, the data are maintained and parallel computation is done therefore the result will be received efficiently. But there is no use of sending notification to pharmacists if the medicine is delegated.

7. FUTURE WORK:

In future, we can enhance this idea by sending notification to the patient who had bought the medicine. And the notification must give full details about the drugs and pharmacy detail that they bought from. This can helps them to know the requirement of medicine and prevent them from the effect of

expired medicine. If drug is not yet sold and remind in pharmacy then the notification should be send to the pharmacy.

8. REFERENCES:

[1]. k.Lee, T.T.Wan, and H.Kwon, ”The relatioship between healthcare information system and cost in hospital, ”Pers. Ubiquitous Comput.,vol.17,no.7, pp. 1395-1400,oct. 2013.

[2]. M.Chen, S.Mao, and Y.Liu, ”Big data: A survey, ”Mobile Netw. Appl., vol. 19, no. 2, pp. 171-209, Apr. 2014.

[3]. I. Wan, H.Yan, H.Suo, and F. Li,”Advances in cyber-physical systems research,”KSII TIIS, vol. 5, no. 11, pp. 1891-1908, 2011.

[4]. C. Lin et al,”Temporal event tracing on big healthcare data analytics,” in Proc. IEEE Int. Big data Congress, 2014, pp. 281-287.

[5]. V. Chandola, S. Sukumar, and J. Schryver,” Knlowdge discovery form massive healthcare claims data,”in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 1312-1320.

[6].O.Samuel, M. Omisore,B. Ojokoh, and E. Atajeromavwo,” Enhanced cioud based model for healthcare delivery organizations in developing countries,” Int. J. Comput. Appl., vol. 74, pp. 0975-8887, 2013.

[7]. C. Yang et al.,”Accessing medical image file with co-allocati0n HDFS in cloud,”Future Gener. Comput. Syst. , vol. 43, pp. 61-73, Feb. 2015.

[8]. Yin Zhang, Chun-Wei Tsai,Meikang Qiu.,”Health-CPS:Healthcare Cyber-Physical System Assisted by Cloud and Big Data”IEEE Trans on Big Data-Aug 2015.

[9].D.W.Bates, S.Saria, L.Ohno-Machado, A.Shah, and G.Escobar,” Big data in health care:Using analytics to identify and manage high-risk and high-cost patients, ”Health Affairs, vol. 33, pp. 1123-1131, 2014.

[10]. M.Sirota, J. T.Dudley, J. Kim, A.P. Chiang, A. A. Morgan, A. SweetCordero,J. Sage, A. J. Butte, “Discovery and preclinical validation of drug indication using compendia of public gene expression data,” Sci. Transl. Med., vol. 3, pp. 1-10, Aug. 2011.

[11]. N.S. Jahchan, J. T. Dudley, P. K. Mazur, N. Flores, D. Yang, A. Palmerton, A.-F.Zmoos, D. Vaka, K.Q.T. Tran, M. Zhou, K. Krasinska,J. W. Riess, J. W. Neal, P. Khatri, K. S. Park, A. J. Butte, and J. Saje, “A drug repositioning approach identifies try cyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors, ”Cancer Discovery, vol. 3, no. 12, pp. 1364-1377, sep. 2013.

[12]. P. Geeleher, N. J. Cox, and R. S. Hualg, “Clinical drug response can be predicted using base line gene expression levels and in vitro drug sensitivity in cell lines, ”Genome Biol., vol. 15, no. 3, pp. 1-12,2014.

[13]. M.H.Jiang and J. H. S. You, “Review of pharmacoeconomic evaluation of genotype-guided antiplatelet therapy,”Proc. Amer. Thoracic Soc., vol. 4, pp. 18-25,2007.

[14]. E. A. Mohammed, B. H. Far, and C. Naugler, “Applications of the mapreduce programming framework to clinical data analysis: current landscape and future trends,”BioData Mining, vol. 7,no. 22, pp. 1-23,2014.

[15]. Javier Andreu-Perez,Carmen C.Y.Poon, Robert D. Merrifield, Stephen T.C.Wong, and Guang-Zhong Yang, Fellow,IEEE,”Big Data for Health, ”vol. 19, no. 4, July 2015.

[16]. Ravindra Ch, G Rajesh, Annapurna G,Ch Swetha, M.Ashish Reddy, G.Goutham Krishna, “Automated Health Care Management System Using Big Data Technology, ”vol. 5,Issue 4, April 2016.

[17]. Dinda, P. A. and “Design, Implementation, and Performance of an Extensible Toolkit for Resourse Prediction in Distributed Systems,”IEEE Trans. Parallel and Distributed Systems, vol. 17, no. 2, pp. 160-173.

[18]. Massie, M.L. Chun, B.N. and Culler, D. E. “The Ganglia Distributed Monitoring System: Design, Implementation, and Experience,” Parallel Computing, vol. 30, no. 7, pp. 817-840.

[19]. Waheed et al., “An Infrastructure for Monitoring and Management in Computational Grids,” Proc. Fifth Int’l Workshop Languages, Compilers and Run-Time Systems for Scalable Computers, vol. 1915, pp. 235-245.

[20]. Wolf, F. and Mohr, B. “Hardware-Counter Based Automatic Performance Analysis of Parallel Programs,” Proc. Conf. Parallel Computing (ParCo ‘03),pp. 753-760.