

Chapter 16

Hadoop Map Only Job for Enciphering Patient– Generated Health Data

Arushi Jain

Ambedkar Institute of Advanced Communication Technologies and Research, India

Vishal Bhatnagar

Ambedkar Institute of Advanced Communication Technologies and Research, India

ABSTRACT

Today, Big Data is being leveraged in many industries from criminal justice to health care to real estate with powerful outcomes. Organizations are using Big Data to predict the future in turn making them smarter and efficient. All the health care data such as discharge and transfer patient data maintained in Computer based Patient Records (CPR), Personal Health Information (PHI), and Electronic Health Records (EHR). The use of Big Data analytics is becoming increasingly popular at health care centres, in clinical research, and consumer based medical product development. The biggest challenge with implementation of big data is that the nature of information of public health sector is of very sensitive nature and needs to be protected from unauthorized access and release of contents. Therefore, to provide solution to the deidentifying personal health big data in this paper we author make use of only mapper job framework for data encryption.

INTRODUCTION

In recent years, there is a proliferation in the amount of data generated. Voluminous complex data is generated from various sources that cannot be processed by traditional tools. Big Data is a current and hot topic of IT industry. Harnessing such huge amount of data is a tough job and thus it requires business intelligence (BI) or analytics. Business Intelligence is required to explore new knowledge, relationships and patterns among different data elements for example the banks have to make perpetual customer centric decisions and make sure they are well informed about all the risks and stakes. One of the best ways to reach out to the customers is to constantly communicate with them. This is obviously difficult as the

DOI: 10.4018/978-1-5225-6198-9.ch016

banks have thousands of branches and more than a million customers in countries which are populous and have a huge number of customers. Therefore, the use of Big Data is really important here. By asking the customers for feedback about products and services, banks can come to know about the customer's needs and preferences. In this paper, we focused on analyzing all the complaints from the customers that can help banks make well informed decisions in optimizing the services and business process along with region or time of the year specific schemes and offers which may attract more customers.

Big Data is a collection of Data sets so large and complex that traditional database management systems cannot handle efficiently. Big Data in an umbrella term used for the humongous amount of data generated from myriad of sources such the Web, mobile devices, sensors, enterprise applications and digital repositories. The data can be structured as well as unstructured. The data ranges from terabytes to exabytes of data. The relational database management systems (RDBMS) have proven to inefficient to handle such huge volumes of data. Another important factor which renders the conventional database systems unsuitable is that the majority of data being generated is unstructured; the RDBMS systems are only adept to handle structured data. Hence new tools and schemes for data analysis and management were in order. Big Data can be characterized by the 5 V's:

1. **Volume:** This refers to the amount of data. The sheer volume of data generated these days by real time applications and other data sources such as twitter feeds, photos, videos on social media, click streams of web, sensor-enabled equipment is so mammoth it runs to petabytes and exabytes of data. Big Data technology enables us to store this amount of data on distributed systems;
2. **Velocity:** It refers to the speed at which the data is generated. For example, the social media portals generate process able data at a very high rate, the amount of credit card transactions that happen in a second. All these require data to be analyzed at a very high rate;
3. **Variety:** Refers to the myriad of sources from which data is accumulated. Data is in different formats structured, semi-structured, unstructured. A full 90% world data has been generated over the past two years and majority of it is unstructured;
4. **Veracity:** Refers to the trustworthiness of data. There are uncertainty surroundings the data being generated these days with data being incomplete and inconsistent. The Big Data analytics empowers us to work with such kind of data;
5. **Value:** Refers to techniques of deriving value from data. There is an intrinsic value that the data may possess and must discover for analysis. This makes 'value' the most important 'V' of Big Data. Modern technologies have made it possible to find the value from data.

Today, Big Data is being leveraged in many industries from criminal justice to health care to real estate with powerful outcomes. Organizations are using Big Data to predict the future in turn making them smarter and efficient. Applications from Big Data are innumerable, from retail industry where Big Data helps retailers gain insights into the customer to needs and habits, to Banking, HealthCare & Hospitality. Government agencies are increasingly incorporating Big Data analytics to curb crime and maintain law and order through social media traffic analysis and other means. Therefore, to get actionable data and perform analytics requires specialized tools. There are thousands of Big Data tools available in the market. There are open source tools like Hadoop, a term which has become synonymous to Big Data, Cloudera.

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/hadoop-map-only-job-for-enciphering-patient-generated-health-data/207063

Related Content

Research on multi-view clustering algorithm on epileptic EEG signal

(2022). *International Journal of Health Systems and Translational Medicine* (pp. 0-0).

www.irma-international.org/article//282705

Revolutionizing Diabetic Retinopathy Diagnostics and Therapy Through Artificial Intelligence: A Smart Vision Initiative

R. S. M. Lakshmi Patibandla, B. Tarakeswara Rao and M. Ramakrishna Murty (2024). *Transformative Approaches to Patient Literacy and Healthcare Innovation* (pp. 136-155).

www.irma-international.org/chapter/revolutionizing-diabetic-retinopathy-diagnostics-and-therapy-through-artificial-intelligence/342824

Assessing the Risks and Success Factors of Telehealth Development Projects in an Academic Setting

Mehmet Serdar Kilinc (2022). *International Journal of Health Systems and Translational Medicine* (pp. 1-16).

www.irma-international.org/article/assessing-the-risks-and-success-factors-of-telehealth-development-projects-in-an-academic-setting/291981

Modeling Stochastic Gene Regulatory Networks Using Direct Solutions of Chemical Master Equation and Rare Event Sampling

Youfang Cao, Anna Terebus and Jie Liang (2016). *Emerging Research in the Analysis and Modeling of Gene Regulatory Networks* (pp. 80-122).

www.irma-international.org/chapter/modeling-stochastic-gene-regulatory-networks-using-direct-solutions-of-chemical-master-equation-and-rare-event-sampling/155027

Topical Use of Plant Extract-Based Oil Blend in Relieving the Symptoms of Primary Dysmenorrhea: An Independent Clinical Study

Amul S. Bahl (2021). *International Journal of Health Systems and Translational Medicine* (pp. 47-61).

www.irma-international.org/article/topical-use-of-plant-extract-based-oil-blend-in-relieving-the-symptoms-of-primary-dysmenorrhea/270953