

Chapter 9

Efficient Big Data–Based Storage and Processing Model in Internet of Things for Improving Accuracy Fault Detection in Industrial Processes

Mamoon Rashid

*School of Computer Science and
Engineering, Lovely Professional
University, India*

Harjeet Singh

 <https://orcid.org/0000-0003-3575-4673>
*Department of Computer Science,
Mata Gujri College, Fatehgarh Sahib,
India*

Vishal Goyal

*Department of Computer Science,
Punjabi University, India*

Nazir Ahmad

*Department of Information Systems,
Community College, King Khalid
University, Saudi Arabia*

Neeraj Mogla

Nike Inc. Global Headquarters, USA

ABSTRACT

As the lot of data is getting generated and captured in Internet of Things (IoT)—based industrial devices which is real time and unstructured in nature. The IoT technology—based sensors are the effective solution for monitoring these industrial processes in an efficient way. However, the real—time data storage and its processing in IoT applications is still a big challenge. This chapter proposes a new big data pipeline solution for storing and processing IoT sensor data. The proposed big data processing platform uses Apache Flume for efficiently collecting and transferring

DOI: 10.4018/978-1-7998-0373-7.ch009

large amounts of IoT data from Cloud—based server into Hadoop Distributed File System for storage of IoT—based sensor data. Apache Storm is to be used for processing this real—time data. Next, the authors propose the use of hybrid prediction model of Density-based spatial clustering of applications with noise (DBSCAN) to remove sensor data outliers and provide better accuracy fault detection in IoT Industrial processes by using Support Vector Machine (SVM) machine learning classification technique.

1. CHAPTER OUTLINE

This chapter is structured around the concepts of efficient storage of sensor IoT based data and its processing in Big Data pipeline. Further the inclusion of novel prediction model will help to improve fault detection in IoT Industrial processes. In today's scenario we have to consider the big source of data generation as well as the plausible suitable platform of such huge data analysis. Therefore, the associated challenges are also included in this chapter.

1.1 Introduction to Big Data and IoT

Big Data Analytics along with Internet of Things (IoT) finds its use in areas of Smart Cities, Healthcare, Agriculture and Industrial Automation Units. The challenge of large amount of data generation in IoT devices is fulfilled by Big Data technologies in terms of its storage and processing (Chen et al., 2016). The advanced IoT devices and their applications have given rise to voluminous data in different varieties (Mavromoustakis et al., 2016). On the other side, Big Data technologies have discovered new kind of opportunities for developing IoT based systems (Rashid et al., 2013). Therefore, IoT based Systems and Big Data technologies integration will create new challenges in terms of storage and processing which needs to be addressed by the researchers (Singh et al., 2015).

1.1.1 Types of Big Data

The classification of Big Data is mostly given in terms of structure of data. The structure of data depends usually on its organization. Based on this, Big Data is classified into structured, unstructured and semi- structured data (Oussous et al., 2018). These types are explained below:

- **Structured Data:** Structured data is having fixed format and is easily stored, processed and accessed. Structured data is always following particular order as in row and column format and always results into ordered output. This

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/efficient-big-data-based-storage-and-processing-model-in-internet-of-things-for-improving-accuracy-fault-detection-in-industrial-processes/239163

Related Content

A Survey of Mobile Ticketing Services in Urban Mobility Systems

Marta Campos Ferreira, Teresa Galvão Dias and João Falcão e Cunha (2020). *International Journal of Smart Sensor Technologies and Applications* (pp. 17-35). www.irma-international.org/article/a-survey-of-mobile-ticketing-services-in-urban-mobility-systems/281601

Optimal Hop Lengths to Ensure Minimum Energy Consumption in Wireless Sensor Networks

Mekkaoui Kheireddine and Rahmoune Abdellatif (2020). *Sensor Technology: Concepts, Methodologies, Tools, and Applications* (pp. 105-124). www.irma-international.org/chapter/optimal-hop-lengths-to-ensure-minimum-energy-consumption-in-wireless-sensor-networks/249558

Blockchain Hyperledger Sawtooth Enabled Digital Forensics Chain of Custody (CoC) A Short Report

(2022). *The International Journal of Imaging and Sensing Technologies and Applications* (pp. 0-0). www.irma-international.org/article//306655

Blockchain Hyperledger Sawtooth Enabled Digital Forensics Chain of Custody (CoC) A Short Report

(2022). *The International Journal of Imaging and Sensing Technologies and Applications* (pp. 0-0). www.irma-international.org/article//306655

Wireless Video Sensor Networks: Advances in Distributed Video Coding

Abdelrahman Elamin, Varun Jeoti and Samir Belhouari (2012). *Visual Information Processing in Wireless Sensor Networks: Technology, Trends and Applications* (pp. 40-58). www.irma-international.org/chapter/wireless-video-sensor-networks/59752