

Chapter 12

Distributed Streaming Big Data Analytics for Internet of Things (IoT)

Sornalakshmi Krishnan
SRM University, India

Kayalvizhi Jayavel
SRM University, India

ABSTRACT

In this chapter, a discussion on the integration of distributed streaming Big Data Analytics with the Internet of Things is presented. The chapter begins with the introduction of these two technologies by discussing their features and characteristics. Discussion on how the integration of these two technologies benefit in efficient processing of IoT device generated sensor data follows next. Such data centric processing of IoT data powered by cloud, services and other enablers will be the architecture of most of the realtime systems involving sensors and real-time monitoring and actuation. The Volume, Variety and Velocity of sensor generated data make it a Big Data scenario. In addition, the data is real time and requires decisions or actuations immediately. This chapter discusses how IoT data can be processed using distributed, scalable stream processing systems. The chapter is concluded with future directions of such real time Big Data Analytics in IoT.

INTRODUCTION

Internet of Things has been identified as an emerging technology that will transform our environment to a more connected and smarter world. Cisco predicts that over the next five years, global IP networks will support up to 10 billion new devices and connections increasing from 16.3 billion in 2015 to 26.3 billion by 2020. The projection is 3.4 devices and connections per capita by 2020—up from 2.2 per capita in 2015. And if clearly observed, every company have ventured to IoT relevant to their sector. Cisco, Juniper and other networking based companies have started talking about Edge, Mist Fog Analytics as next futuristic technologies for IoT.

DOI: 10.4018/978-1-5225-3142-5.ch012

MathWorks have acquired ThingSpeak which is a cloud based company and MathWorks have developed extensive toolbox for Internet of Things. It comprises of many open source hardware support, to name a few Raspberry Pi, Arduino and many more. IBM has come up with BluemixCloud, Google with its OS for Internet of Things, Brillo. Internet of Things allows envisaging the evolution of internet as a huge network of connected intelligent devices. These ubiquitous connected things not only sense, but also process, analyze real physical events ranging from simpler to complex and triggers actions as the need demands.

As the number of affiliated devices increases, the rate at which the data is generated and processed also increases. This requirement has led to the employment of technical advancements like Cloud, Software Oriented Architectural models, Software Defined Networks, Machine Learning, Artificial Intelligence and many more for making the things around us smarter, faster and dynamically intelligent.

In such a connected environment, the enormous amount of data generated by networked devices has to be processed in both real time as well as batch basis. The data generated by IoT devices possess the characteristics of Big Data in terms of Volume, Velocity, Variety and Value. The heterogeneous devices when connected together produce huge amounts of data from which useful inferences or decisions have to be drawn.

The powerful paradigm of MapReduce along with the implementing frameworks like Hadoop has made Big Data processing easier. The sub area of Big Data is Streaming Analytics which analyzes huge amounts of data that arrives with huge velocity and expects the actuation or decision in real time with low latency in terms of seconds. With the number of devices connecting to internet and the need for real time decision, making intelligent applications like self-driving is gaining importance. The frameworks for Streaming Analytics should possess the basic characteristics of Fault Tolerance, high availability, low Latency and Scalability. According to the requirement of application, processing can either follow the store-process-react or process-react-optional store style.

Cloud along with its different flavors and characteristics like Elasticity, Pay as per service and Scalability provides the best performance for centralized Storage, Analytics and Visualization in IoT. Cloud Services are available and provisioned without any human intervention and follow the pay-as-you-go model. Also, while utilizing cloud, the problems of over provisioning or under provisioning found in static fixed provisioning environments do not exist. Scalability and Load Balancing can help maintain the Quality of Service promised to the customers. By offering most of the components as a service, cloud environment is taking away most of the complexities handled at the user level. This enables the users to concentrate on the business processing rather than infrastructure.

Service science is an emerging companion to Streaming Analytics in cloud where there are numerous research areas such as Service Discovery, Composition and Orchestration. The service-oriented Cloud Computing is a supporting framework for cohesive set of cloud components. The big streaming data as opposed to the traditional services data is not structured or similar between services. There are a huge variety of sources in the IoT context like Wireless Sensor Network monitoring forest fire or a Weather station or Pollution monitoring or Home automation which vary exorbitantly unimaginable in the data formats. The data exchange needed before or after Data Analytics will be taken care by services.

The IoT paradigm along with Big Data, Cloud and Service Science promises revolutionary architecture which will be suitable for most of the critical IT applications ranging from smart grids to smart connected communities (Sun, Song, Jara & Bie, 2016).

34 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/distributed-streaming-big-data-analytics-for-internet-of-things-iot/198768

Related Content

An Overview of Ontology-Driven Data Integration

Agustina Buccella and Alejandra Cechich (2009). *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 471-480).

www.irma-international.org/chapter/overview-ontology-driven-data-integration/20732

Detecting Users' Dietary Preferences and Their Evolutions via Chinese Social Media

Qingqing Zhou and Chengzhi Zhang (2018). *Journal of Database Management* (pp. 89-110).

www.irma-international.org/article/detecting-users-dietary-preferences-and-their-evolutions-via-chinese-social-media/218928

The Benefits of Data Warehousing at Whirlpool

Barbara J. Haley, Hugh J. Watson and Dale L. Goodhue (2006). *Cases on Database Technologies and Applications* (pp. 135-153).

www.irma-international.org/chapter/benefits-data-warehousing-whirlpool/6209

Conceptual Data Modeling Patterns: Representation and Validation

Dinesh Batra (2005). *Journal of Database Management* (pp. 84-106).

www.irma-international.org/article/conceptual-data-modeling-patterns/3333

Model Driven Engineering for Quality of Service Management: A Research Note on the Case of Real-Time Database Management Systems

Salwa M'barek, Leila Baccouche and Henda Ben Ghezala (2016). *Journal of Database Management* (pp. 24-38).

www.irma-international.org/article/model-driven-engineering-for-quality-of-service-management/178634