Procedia Computer Science

Volume 53, 2015, Pages 457–467

2015 INNS Conference on Big Data

Railway assets: A potential domain for big data analytics

Adithya Thaduri[1], Diego Galar[2], and Uday Kumar[3]*

[1] Lulea University of Technology, Lulea, Sweden
adithya.thaduri@ltu.se
[2] diego.galar@ltu.se
[3] uday.kumar@ltu.se

Abstract

Two concepts currently at the leading edge of todays information technology revolution are Analytics and Big Data. The public transportation industry has been at the forefront in utilizing and implementing Analytics and Big Data, from ridership forecasting to transit operations Rail transit systems have been especially involved with these IT concepts, and tend to be especially amenable to the advantages of Analytics and Big Data because they are generally closed systems that involve sophisticated processing of large volumes of data. The more that public transportation professionals and decision makers understand the role of Analytics and Big Data in their industry in perspective, the more effectively they will be able to utilize its promise. This paper gives an overview of Big Data technologies in context of transportation with specific to Railways. This paper also gives an insight on how the existing data modules from the transport authority combines Big Data and how can be incorporated in providing maintenance decision making.

Keywords: Big Data, Railways, Maintenance, Transportation

1 Introduction

Recent technological advances have sparked what amounts to a revolution in the application of Big Data cognitive and informational tools. There is enormous amount of data that has been generated by the systems using sensors by data acquisition techniques. These type of data varies primarily by 5 Vs; volume, velocity, variety, value and veracity. The types of data can be of different characteristics structured, semi-structured and unstructured. Due to the advancements in sensor technology, data handling and storage and capabilities of faster computing algorithms, the usage of Big Data is quite predominant in several sectors. This huge amount of data that are collecting from the various sources has invoked several possibilities and insights for the future technology that will embrace a new wave of application areas.

Big data is defined by as data that includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable

*Masterminded EasyChair and created versions 3.0–3.4 of the class style

Selection and peer-review under responsibility of the Scientific Programme Committee of INNS-BigData2015

doi:10.1016/j.procs.2015.07.323

elapsed time [1]. It is quite difficult to define quantitatively the Big Data by numbers because the definition comes to the question, How big is big data? We are talking the data in terms of 100 Terra bytes, 10 GB/s from 100 sources and this figures will change because of increasing installation of Internet of Things (IoT), Wireless Sensor Networks (WSN), cloud based services and condition monitoring of heavy machines. Hence there is a need of Data mining capabilities to handle this massive amount of data for efficient storage and accessible information. What is the use of this massive data? That is where the importance of Big Data Analytics is emerged. IBM [2] defined the big data analytics as the use of advanced analytic techniques against very large, diverse data sets that include different types such as structured/unstructured and streaming/batch, and different sizes from terabytes to zettabytes. By analyzing the Big Data, the data analysts, researchers, technicians, operators and business users can make better understanding of the system and provide faster and better decisions that were inaccessible before. There are tools that are still applying in the research arena that can also be utilized for big data analytics are machine learning, natural language processing, predictive analytics, data mining, statistics and strategic management methods.

Big Data has been applying in several areas like computing, telecommunications, web and mobile services, manufacturing, process industries, transportation, scientific simulations, etc. The most prominent applications of Big Data as per McKinsey Global Institute analysis are healthcare, retail, transportation, manufacturing, public services and personal data [3]. It is also worth to mention that also, from the more operational point of view, the latest trends in data visualization focus on two different lines:Augmented Reality [4] and Infographics tools [5].

2 Transport Industry

In virtually any city, one is likely to see the results of analytics in the operation of transit buses and trains that are essential for maintaining the mobility of the metro area. Furthermore, it is useful for both the public and the industry to realize how significantly public transportation has been a leading pioneer in the rich and extensive historic development of these tools. There are some published studies over usage of Big data in area of transportation. It applied for Istanbuls automated fare collection system and pricing for BRT-Bus Rapid Transit line planning with visualization metrics to obtain better recommendations for consumers [6]. Using Markov-chain approach, a multi-modal transport network in London was developed with better information clusters for efficiency of transport [7]. City Intelligent Energy Network used statistical data including city economy, construction, population, and different energy parameters to develop the comprehensive model for low-carbon emissions [8].

Using Big data, several key indicators for transport was demonstrated to measure performance of cities [9] and for sustainability [10]. Big data processing and data mining has been used to develop architecture for traffic cloud data mining and optimization of strategies (TCD-MOS) and related data processing and network optimization methods [11]. There are other case studies such as big data analytics for safety management [12], an assistant decision-supporting method for urban transportation planning using GPS (Global Positioning System) data [13], crowdsourcing for intelligent transport system[14], crowd sourcing geo-social network [15], intelligent transport system for predicting drivers behavior [16], realtime monitoring from traffic for TomTom [17] and in Netherlands [18]. Several opportunities in public transport and processing techniques and their challengers [19] were described in transport domain [20].

The are endless possibilities of Big Data analytics in the transportation and one of the main areas is to look into maintenance aspects for maximum customer experience. It can identify behaviour of bottlenecks, maximum loads, variation in traffic, unplanned delay timings,