Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 64 (2017) 369 – 374

www.elsevier.com/locate/procedia

The 9th CIRP IPSS Conference: Circular Perspectives on Product/Service-Systems

A big data analytical architecture for the Asset Management

Jaime Campos[a]*, Pankaj Sharma[b], Unai Gorostegui Gabiria[c], Erkki Jantunen[d], David Baglee[e]

[a]*Linnaeus University, Faculty of Technology, Department of Informatics,Sweden*
[b]*Department of Mechanical Engineering, , IIT Delhi, New Delhi, India*
[c]*Electronics and Computing Department Mondragon University, Mondragon, Spain*
[d] *VTT Technical Research Centre of Finland, P.O.Box 1000, FI-02044 VTT, Finland*
[e]Department of Computing, Engineering and Technology, University of Sunderland, UK

* Corresponding author. Jaime Campos. Tel.: +46-(0) 470-708829 E-mail address:jaime.campos@lnu.se

Abstract

The paper highlights the characteristics of data and big data analytics in manufacturing, more specifically for the industrial asset management. The authors highlight important aspects of the analytical system architecture for purposes of asset management. The authors cover the data and big data technology aspects of the domain of interest. This is followed by application of the big data analytics and technologies, such as machine learning and data mining for asset management. The paper also presents the aspects of visualisation of the results of data analytics. In conclusion, the architecture provides a holistic view of the aspects and requirements of a big data technology application system for purposes of asset management. The issues addressed in the paper, namely equipment health, reliability, effects of unplanned breakdown, etc., are extremely important for today's manufacturing companies. Moreover, the customer's opinion and preferences of the product/services are crucial as it gives an insight into the ways to improve in order to stay competitive in the market. Finally, a successful asset management function plays an important role in the manufacturing industry, which is dependent on the support of proper ICTs for its further success.

Peer-review under responsibility of the scientific committee of the 9th CIRP IPSS Conference: Circular Perspectives on Product/Service-Systems.

Keywords: Asset Management; Big data; Big data analytics; Data mining.

1. Introduction

When preventive maintenance is applied, a large amount of data on operations and maintenance are produced. However, companies and the asset management function are currently storing a huge amount of data and are at the same time not using it [1]. Consequently, there is almost no use of the stored and existing data produced by the equipment that could result in increased efficiency and organisational performance of an asset management function. The analysis of the data produced by a company is extremely important for improved decision making.

The area of Information and Communication Technologies (ICTs) is dynamic, and new technologies emerge rather frequently. Therefore companies need to understand their various characteristics to be able to develop, implement and use these ICTs successfully. The authors provide a big data analytical architecture at a conceptual level where the data scientist and the maintenance staff are part of the system. In addition, it highlights important aspects of a system to be used for the purpose of asset management. There are several ICTs applications and systems suggested and implemented in the industrial domain [2; 3]. However, some of these ICTs do not provide an overall picture of the system where the data

Peer-review under responsibility of the scientific committee of the 9th CIRP IPSS Conference: Circular Perspectives on Product/Service-Systems.
doi:10.1016/j.procir.2017.03.019

scientist and user are part of the system, especially when it is based on the big data approach. In the current work, the authors provide an analytical architecture, based entirely on a big data approach at a conceptual level. A data scientist requires innovative solutions in order to perform different elements of the CRISP Methodology including business and data understating, data preparation, modelling, evaluation and deployment aspects of a big data solution or project. The next part of the paper is ordered in the following way. In section 2, the paper discusses data and big data characteristics. In section 3, big data analytics, machine learning, and data mining features in connection to the domain of interest are discussed. Thereafter, in section 4, important aspects of user interface and visualisation are covered. In section 5, an Asset Management technical and analytical framework is presented and discussed. Finally, conclusions are presented in section 6.

2. Data& big data in the domain of interest

Big data is a term that makes reference to a great quantity of data, which exceeds conventional software's capacity to handle it. The processes that are used to find patterns in them are the so-called predictive analytics and user behaviour analytics. Data gathering is the first step towards an exhaustive analysis for the asset management. Data analysis can use data collected from different sources, which depends on the objective of the analysis. One of the domains where data can be obtained is the web and social media. Search logs clicked web content, or streams are collected from various web services [4]. The information is then used to understand the customer's needs and improve the experience for the user. In the industrial sector, for instance, there is a need to monitor different assets, usually with the help of transducers to keep the equipment in optimal condition and reduce the downtime caused by maintenance. These transducers are typically named sensors. They transform a physical property or phenomena into an electrical signal. The signal then is manipulated by means of filtering to reduce the electrical noise and, if necessary, is converted to a digital signal. After some signal conditioning and manipulation, it is sent to a database where meaningful information can be attained. The data can be obtained in several formats such as structured, unstructured as well as semi-structured [5]. The structured data can be defined as the data that has a well-defined format such as date, numbers or character strings and allows storage and generation of information. Databases or spreadsheets are examples of structured data. Unstructured data refer to data in the same format as were collected, such as emails, PDFs, or documents [6]. The semi-structured data are not limited to determined fields, but contains separators to divide the data. Usually, they cannot be managed in a standardised way, but they contain their metadata that describes the objects and their relationships, i.e. XML or HTML [7].

The storage systems can be classified as DAS (Direct Attached Storage), NAS (Network Attached Storage) and SAN (Storage Area Network) [8]. However, these storage systems have limitations when creating a large-scale distributed storage system [9]. NoSQL (Not-Only SQL) databases provide a more flexible and concurrent storage systems and allow a much faster manipulation and search of information than relational databases. Four different types of NoSQL databases are distinguished: Key-Value, Documental, Graph and Column Oriented [8]. Different data storage architectures have to be tested to find the most appropriate one. Features such as Data Model, Data Storage, Concurrency Control or Consistency have to be taken into account while deciding which architecture suits best the system [8].

The data processing for big data can be done in various ways, depending on the application. Real-time applications such as navigation, finance, Internet of Things (IoT) or intelligent transportation rely heavily on timeliness [9]. The processing power needed for these real-time applications is very high, and the cloud computing and distributed technologies are useful tools to be able to analyse data at such speeds. Also, it is important to understand the characteristics of big data in connection with the 3V's, i.e. volume, velocity, and variety in the domain of interest. Therefore, the maintenance data can be, further on, divided into traditional thinking maintenance data and data connected to the 3Vs for a better understanding of its characteristics in the domain [5]. Traditional data when it comes to volume involves condition monitoring, maintenance plan, work orders, etc. The non-traditional data part of this volume is the data that is indirectly related data, such as purchase contract, production, scheduling, asset depreciation value, etc. The traditional data part of the variety includes semistructured data, such as spreadsheets and data stored in relational databases. Variety data being a part of the big data is semi-structured like emails, XML files and log files with some specific formats. Also, the unstructured data part of the variety consists of pictures, audios, videos, web pages and other documents, such as Word and PDF. Finally, part of the velocity and traditional maintenance data are transaction data and multidimensional data. Whereas, the data considered outside the traditional maintenance data and part of the velocity is the real-time condition monitoring data collected by sensors and instruments. In conclusion, the division of the data is processed in such a way that the data outside the limits of the traditional maintenance data are considered to be part of the big data approach.

3. Big data Analytics & Data Mining

With the emergence of the Big data approach and its technologies, there is a need to use ICTs, such as data mining software algorithms and statistics, for more fruitful information and knowledge as well as possibilities to find hidden patterns from the data. However, to analyse data to discover relationships and non-obvious patterns is not a new concept. Data mining, knowledge extraction, information discovery, information harvesting, data archaeology and data pattern processing are some of the names to the techniques that use data analysis for decision making. For instance, Knowledge Discovery in Databases (KDD) is a concept that involves data mining techniques. It maps low-level data that is