A Comparison of Smart City Development and Big Data Analytics Adoption Approaches

Zohreh Pourzolfaghar[1], Christian Bremser[2], Markus Helfert[1] and Gunther Piller[2]
[1]*School of Computing, Dublin City University, Dublin, Ireland*
[2]*Business Information Technology, University of Applied Science Mainz, Mainz, Germany*

Keywords: Smart City, Information Communication Technology, Big Data Analytics, Technology Adoption.

Abstract: This paper intends to elucidate the similarities between smart city development and big data analytics adoption. Both concepts promise new opportunities: smart cities to improve citizens' life quality and big data analytics to drive companies towards the competitive edge. Consequently, the number of organisational big data initiatives and efforts to implement smart city concepts are increasing. In the context of big data analytics adoption, it could be shown that there are two distinct approaches companies follow. They either focus on the search for potential use cases or on the development of a technology infrastructure. Based on a comparison of various smart city and big data analytics use cases, this paper discloses that both of these approaches either concentrate on developing new service development or providing the required infrastructures for future services.

1 INTRODUCTION

A smart city is an innovative city that uses ICT to improve citizens' quality of life and efficiency of the urban services (Booch 2010; ITU-T FG-SSC 2014; Anthopoulos and Janssen 2016). To achieve this goal, the produced information in various city systems are combined to provide effective services. This information is produced by variety of sources, e.g. by sensors and internet of things (IoT) devices installed in the buildings, streets, vehicles and so on. Based on the definition by Pourzolfaghar and Helfert (2017a), smart services should have a goal in line with the ultimate goal for smart cities, to facilitate daily activities of citizens and improving their quality of life. To achieve this goal, governments are working in a higher level to provide the infrastructures. Also as Ryazanova and Pétercsák (2016) stated, private companies use the produced data by the devices in the infrastructure level to develop the services in smart cities. As these researchers stated, cooperation between public and private sectors in smart cities leads to develop smart zones in the cities. However, there are some difficulties to take fully advantage from this type of collaborations. By increasing number of the installed IoT devices and sensors, smart cities are facing difficulties in terms of heterogeneity of the stored data from various sources. For instance, Pourzolfaghar and Helfert (2017b) have introduced this issue as one of the barriers to take advantage from integration of information from different sources in buildings, to provide useful services by facility management companies.

Big data analytics is experiencing a situation similar to the smart city development. In the era of advancing digitisation of whole industries, the potential benefits and challenges associated with big data are important topics for companies. Big data promises new data-driven services to improve processes and enable innovative products and business models (Sivarajah, 2017). Therefore, growing number of enterprises focus their investments on the adoption of big data. Against this backdrop, Bremser et al. (2017) investigated the process how companies examine the possibilities of big data. Based on a multiple case study, they found two generic approaches that are pursued by organisations. These approaches focus either on the search for potential business opportunities or on the need to develop technology infrastructure. This paper examines, whether similar approaches can be identified in smart cities too.

The paper is structured as follows: In the next section, we introduce big data analytics and big data adoption approaches. This will be followed by

Pourzolfaghar, Z., Bremser, C., Helfert, M. and Piller, G.
A Comparison of Smart City Development and Big Data Analytics Adoption Approaches.
DOI: 10.5220/0006674501570164
In *Proceedings of the 7th International Conference on Smart Cities and Green ICT Systems (SMARTGREENS 2018)*, pages 157-164
ISBN: 978-989-758-292-9

examples how organisations use these approaches to introduce big data analytics. In the following sections smart cities use cases are introduced to describe how smart cities are developed. In the discussion section we will compare the use cases from both domains to identify the similarities between the adoption of big data analytics and the development of smart cities.

2 BIG DATA ANALYTICS IN ENTERPRISES

The TechAmerica Foundation defines big data as (TechAmerica 2012) “a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information.” In accordance with many other definitions, they use volume, velocity and variety to describe the challenges in data management. However, the literature shows that there are ambiguities about limits on the three Vs (Mikalef et al., 2017). These depend on size, industry and location of a company and result in a “three-V tipping point”. Beyond this tipping point traditional data management and analysis technologies become inadequate for deriving intelligence within a sufficient period of time.

Therefore, this tipping point poses a threshold beyond which firms start dealing with big data and examine the value of new technologies (e.g. complex event processing, in-memory data processing, NoSQL databases) compared with their present implementations (Gandomi and Haider 2015). However, big data is meaningless without analysis. Its potential value can only be leveraged, when companies extract meaningful insights from big data. Therefore, new processes, tools and methods are required that capture, analyse, and visualise the underlying patterns in the data (Mikalef et al., 2017). Hence, big data analytics refers to multiple techniques (e.g. text analytics, social media analytics) that are used to acquire intelligence from big data (Gandomi and Haider 2015). Companies want to take full advantage of these new opportunities and explore the possibilities of big data analytics. In this regard Bremser et al. (2017) has shown that there are two distinct approaches companies follow to examine the potentials of big data:

In the approach Business First, enterprises explore big data potentials entirely from a business perspective. They search for use cases with high expected business value. These use cases span from possible improvements of existing processes to entirely new business services or business models. Typically, these use cases are developed in cloud environments as prototypes. In order to evaluate the assumptions regarding business case, the prototypes are tested in market segments with company-friendly customers.

Organisations following a Platform Building approach initially focus on an identification of key activities for the development of a future-oriented big data platform and not on the search for particular application scenarios. Their goal is to provide a technological starting point and integrate all existing and company relevant data sources. Specific application scenarios do not yet exist, but are expected to come up eventually.

In the following subsections, we will introduce four examples, how companies approached big data analytics. The examples have been adapted from the multiple case study of Bremser et al. (2017). The first two cases illuminate successfully identified business opportunities and emphasise how companies create new business services through the application of big data analytics. These examples will be followed by cases where companies introduce new technologies in order to enable themselves for future big data analytics use cases.

2.1 Big Data Analytics - Business First

2.1.1 Predictive Analytics in Utilities

A deregulated market and the energy transition cause uncertainties in the business of utilities companies. Driven by this development, the top management of one of the companies of the multiple case study of Bremser et al. (2017) set up a digital IT unit to focus on innovative and data-driven topics. One of the identified big data use cases in this company is predictive maintenance. Predictive maintenance promises to prevent cost-intensive machine failures. For this purpose, the investigated company starts to use sensor data from production plants, machines and other operating equipment. These sensors are producing data about the conditions of the machines, for example, voltage, temperature, rotation and vibration. Through the continuous generation of machine data, the processing of these large volumes and fast-moving data is required. In order to carry out first analyses, data is temporarily collected within a cloud storage. They are then analysed by data scientists for usage and fault patterns. Typically, machine learning methodologies are used for these studies (Susto et al., 2015).

Observed patterns help to identify low quality components and to monitor wear and condition of machines in real time. Third party data, such as weather data and data on other possible environmental influence factors are also integrated into the analysis to improve the prediction of possible breakdowns. In order to validate possible predictions, the company captures the uptime of various machines. If the application of an algorithm leads to a decreasing downtime, the algorithm is rolled out to other machines and production sites.

After a successful validation, also the IT infrastructure needs to be adapted to carry out the new analyses in a stable and efficient manner. For this purpose the company in question plans to integrate a lambda architectures into their existing IT landscape. This consists out of a speed layer for real time data processing and a batch layer for handling large data volumes (Kiran, 2015).

2.1.2 Demand Forecasting in Retail Trade

Right product in right places at right time is a key challenge every retail trader is confronted with. The triangulation of data from different sources, like enterprise resource planning (ERP), supply chain and web, allows retailers to predict customer trends and needs. In the past, historical sales data have only been used to determine future consumption. Current research shows that anonymised online search data is an accurate proxy for customer activities (LaRivier, 2016).

In addition, product characteristics, relationships with other products and external information like competitor prices or weather changes gain importance in the identification of demand anomalies. In order to analyse data from those different sources, analyses of time series, multiple linear regressions or machine learning methods like support vector machines or artificial neural networks are used (Carbonneau et al., 2007).The company from the case study set up a data lab in order to explore and evaluate the potential of demand forecasting. There, data scientists combine data from potentially useful data sources and develop algorithm which are tested in single stores. If these tests succeed, the developed algorithm will be integrated in the overall stock replenishment system.

2.2 Big Data Analytics - Platform Building

2.2.1 Data Lake Development in Pharma

In recent years pharmaceutical companies have experienced a rapid increase of production and customer data. Now available data sources reach from anonymised electronic health records of patients to sensor data from the production sites. These data sources have different characteristics regarding growth rate, structure and privacy requirements. In order to gain valuable insights from these different data, they need to be properly integrated. Therefore, a company in the case study implemented a so called data lake. Data lakes provide a scalable and schema-less repository for raw data and provide an interface to access the stored data (O’Leary, 2014).

Typically, a data lake is a complex ecosystem of different technologies to handle the challenges of the three Vs (volume, velocity, variety) (Hai et al., 2016). To meet privacy issues and treat data appropriately, non-technical issues have also to be considered, like master data management and data governance (O’Leary, 2014). Therefore, also organisational changes are necessary, like the creation of new data related roles in IT (e.g. data steward, data engineer) or the adoption of policies guiding the responsible use of data.

This shows that the introduction of a data lake concept requires far-reaching organisational changes and can only be carried out by large organisational projects. The pharmaceutical company of the case study hopes, that the implemented data lake will serve a broad range of future big data use cases of the company. Typical examples are waste reduction in drug production or individualisation of medication.

2.2.2 Introduction of Hadoop in Banking

Thousands of movements on millions of bank accounts generate tons of data points. In order to meet regulatory requirements, these account movements need to be archived and easily accessible. In the past, the data were archived on disk robots, but the processing limits of these systems are exhausted and the extension of disk space is costly. Beside these challenges in data management, fintechs and a changing customer expectation forces banks to deal with the possibilities of big data. At the same time, regulatory pressure burdens financial resources. Against this backdrop, the investigated financial institution started to search for opportunities to replace existing IT components cost-neutrally with big data technologies. To cope with the demands in data management, the considered bank replaces existing databases with Hadoop clusters. For this purpose, traditional projects are carried out to introduce the new technology and migrate the data to the news systems. In addition, the introduction of Hadoop offers opportunities to integrate further data sources and provides a technological starting point for the adoption of future big data use cases.

3 SMART CITY DEVELOPMENT

Smart cities are innovative cities which uses information communication technologies to improve citizens' quality of life. The ultimate goal of smart cities, meaning improving quality of life, is realised through the provided services. In other word, services in smart cities are responsible to facilitate daily activities for the citizens. To achieve this goal, many devices and sensors have been developed to collect useful information to provide services. For instance, traffic cameras are collecting real time information about traffic status in different areas of the city. This status is reported through the live radio programs and assists the drivers to choose a better route to prevent congestion.

Later, statistical traffic information can be used to optimise transport services based on the real traffic information. Similarly, useful information is produced by IoT devices and sensors in smart buildings. This information can be utilised by various industries, e.g. facility management or utility companies, to provide more effective services, e.g. energy saving alarms, real time repairmen services. Apparently, a huge amount of data is available and should be considered to continuously improve smart cities.

The produced information in various smart systems (e.g. healthcare system, education systems, smart buildings and etc.), can also be utilised to develop smart services through combination. As an example, the shortest time for an emergency case is calculated by combining data from transport systems about congestion, and data on the availability of ambulances in the vicinity to an accident. Seemingly, different types of information sources are utilised to expand smartness in the cities. In the following subsections we will introduce five use cases to provide more insight into the different approaches taken by smart cities.

3.1 Smart Services in Smart Cities

3.1.1 Smart Services in Facility Management Industry

Facility management (FM) industry is responsible for providing and delivering timely, professional analysis, and consulting support services for the customers (Rondeau et al., 2012). Nowadays many buildings are equipped with smart devices, sensors and cameras to provide more comfort environment for the inhabitants. All these installed devices are producing real time information to be utilised by different industries like facility management for various purposes, e.g. for energy management, security management and etc. As an instance, Pourzolfaghar and Helfert (2017b), illustrated how the produced information by smart devices in buildings assist FM department to achieve more efficient energy management. According to their research, the plan for efficiently use of energy was related to the booked meeting room.

Based on the booking system for meeting rooms, the heating systems are turning on at the booked time. At the same time, a control system is working to look after efficient energy usage. This system works based on the received information from the motion detection sensors, cameras and any other smart devices installed in the room. In case that no movement is detected in a given time, the heating system will be switched off automatically. In this way energy consumption will be managed in an efficient way. Apparently, huge amount of the real time data is produced by sensor and smart devices in the buildings. However, the produced data is stored in heterogeneous storages. To make the buildings there is need to integrate the produced data from various sources to provide useful services to the buildings inhabitants.

3.1.2 Urban Planning in Smart Cities

The energy usage of buildings can be utilised for more effective urban planning in smart cities. In this regard Rathore et al. (2016) has provided some examples on how the collected data from various sources, e.g. smart homes, may contribute to the development of smarter cities. They believed that enabling smart cities give benefits to the government authorities and the citizens. For instance, Rathore et al. (2016) explained that the citizens may save car fuel by efficiently managing the route to reach the destination, as well as protecting themselves from environmental pollution. Likewise, Pourzolfaghar and Helfert (2016) emphasised that how energy consumption is utilised for more efficient urban planning. As an example, water consumption for all the buildings in an urban area along with weather forecast and rainfall information is essential to predict possibility of flood occurrence. Consequently, this information is useable for flood management in rainy seasons. The water consumption information in the buildings are produced by smart meters and their changes can be tracked by daily manner.

3.1.3 Smart Commerce

Smart commerce is an approach in which customer concerns is placed in the centre of the business. Yan et al. (2010) elucidated that retailers could always gain profit from having knowledge about customers' needs and willingness to buy. They also emphasised on the forecast information accuracy effect on the profit of traditional and online retailers. The smart devices with embedded sensors can provide information this type of information about the products for the related market. As such, this information is of interest of the companies which produce these devices for real time demand management. By obtaining the produced data by the embedded sensors in smart devices the companies can provide on time services to the customers. Moreover, based on the collected data about faulty devices, there will be a chance to use statistical information to improve the quality of their products. Moreover, the received information about more demand for a product in an urban area and crowd movement in that urban area are important for market analyse and the potential profits from a establishing a new service centre in the area.

3.2 Infrastructure Development in Smart Cities

3.2.1 Smart City Technologies in Dublin-Dockland

Dublin Docklands Strategic Development Zone is an extension of Dublin city centre to combine different communities. Ryazanova et al. (2016) conducted an exploratory research to specify the perception of value creation and value capture among the stakeholders. Based on the finding from their research, it has been realised that the smart technologies are similarly discussed across the public and private sectors in terms of infrastructure needs in the keeping with their responsibilities for the development of this smart zone. As Ryazanova et al. (2016) reported, public sector is mostly concentrated on hard infrastructure such as bridges and transport provision, including electric vehicles and cycling. At the same time private sectors work on wired and wireless networking infrastructure and the test-bedding of sensors and software systems. As they believe large corporations had the capacity to invest in sensing networks with potential environmental benefits such as flooding and air pollution sensors. Clearly, plenty of devices in this project are producing various datasets which are supposed to be utilised to provide smart services in Dublin Dockland smart zone. One of the potential difficulties which can be predicted is the interoperability issues.

3.2.2 Smart City Infrastructure Development in the River City

River city is a city with the intention to be smart based on the drawn vision for it. According to this decision, the authorities in this city decided to work on the architecture to develop a data pool. In this way, they will have efficient control on the provided data by various sources. As an example, Pourzolfaghar and Helfert (2017a) explored the footfall counter service to find out the benefits of this service to the smart city stakeholders. Based on their findings, the main aim of implementing the required infrastructure for the footfall counter is associated with their strategic decision to transform the city to a smart city. One of the recognised difficulties related to this project is the inability to control the produces data by the devices. To achieve the strategic goals, the authorities of the River city are working to develop an architecture to obtain the ability to control the produced data by the implemented infrastructures.

4 DISCUSSION

According to Bremser et al. (2017), the main aim of big data analytics is to provide data-driven services that improve processes and enable innovative products and business models. For this purpose, two distinct approaches are followed to achieve big data analytics goals. Companies focus either on the implementation of certain big data use cases to leverage business opportunities or on the development of a technology infrastructure. As such the main aim of smart cities is improvement of life quality for the stakeholders, by providing effective services. Referring to the explored smart city use cases, the development of smart cities has been approached in a similar way: either specific services are implemented or infrastructure components are established. To clarify these similarities, the presented use cases in both areas are compared from a use case driven and a platform perspective. For this comparison we list as dimensions: aim of activities, approach, data usage and infrastructure. A summary is presented in tables 1 and 2.

Table 1 shows the comparison of use case driven initiatives. Here, both concepts aim on the provisioning of new data-driven services in order to improve business processes or city management services respectively. In big data analytics, companies have already implemented use case productively, while in smart cities most of the use cases are still on the level of proposals. Nevertheless, in both areas the approaches are quite similar. In big data analytics enterprises typically set up lab environments to identify and evaluate potential use cases with high business value. In smart cities, use cases are identified by public and private companies based on the available data sources and in line with the strategic goals to increase quality of life. The comparison of data usage and infrastructure shows that in both concepts a wide range of different data sources is used. These sources need to be processed in an integrated way. Therefore, in both areas required technologies are implemented, e.g. real time processing for sensor data of machines.

Table 1: Comparison of use case driven initiatives.

Use case	Aim	Approach	Data usage	Infrastructure
Utilities	Optimised asset management processes	Digital IT identifies new use cases; these are developed by data scientists in cloud environments and validated for single production machines	Sensor data from production sites & external data sources, e.g. weather data	Real time processing components
Retail trade	Improved supply chain processes	Lab environment is used to identify use cases and relevant data sources; demand forecasting is validated in single stores	ERP, supply chain data and external data sources, e.g. web search data, social media data, weather data	Real time processing components, social media screener, interfaces to other external data sources
Facility Management	Improved energy management efficiency	FM industry uses data from buildings to develop smart services	Motion sensors, cameras, IoT devices	Realtime information processing and integration capabilities to analyse the data from installed smart devices and sensors in buildings
Urban Planning	Flood prediction	Government utilise data from buildings data in an urban area to mitigate flood consequences	Smart meters in buildings, weather forecast information	Real time data processing on water usage and weather forecast
Smart Commerce	Improved demand management	Smart commerce uses data from buildings to promote their products to selected consumers	Embedded sensors in products and devices, crowd movement information	Installed smart and IoT devices and sensors in buildings

Table 2: Comparison of platform driven initiatives.

Use case	Aim	Approach	Data usage	Infrastructure
Pharma	Companywide, centralised data lake	Stepwise implementation of new technologies and integration of data sources; carried out as part of traditional projects	Data usage is expected to increase, therefore, all company relevant data source regardless their characteristics (volume, velocity, variety) are integrated	General technologies for big data management (e.g. real time and batch processing; scalable storage) and corresponding interfaces to company relevant data sources
Bank	Scalable storage for mass data	Implementation of Hadoop, migration of data and stepwise integration of further data sources; carried out within traditional projects	Account movements are expected to be used for analytical purposes and combined with other data	Hadoop File System Interfaces to existing bank systems
Dublin Dockland	Provide a networked infrastructure	Step by step integration of networks and data into the city infrastructure	Public companies to make data available to private companies for the development of smart services	Hard infrastructures e.g. transport provision, including electric vehicles and cycling
River City	Data pool	Step by step integration of diverse data into the city infrastructure	City council will have all data available to develop smart services and to transform the city to a smart city	General technologies to collect the data various sources

Table 2 compares use cases belonging to the platform approach. In both areas, the described cases aim on the provisioning of general infrastructure components. In the introduced big data analytics use cases, the infrastructure focuses on the creation of processing and storage possibilities, while in the smart city examples also data acquisition capabilities are considered. Overall, in both areas an infrastructure is being developed step-by-step and technologies and data sources are integrated. Equally, in both areas, the stepwise developed infrastructure shall serve as a technological starting point providing data and technologies for future use cases.

5 CONCLUSIONS

With the aim of comparison between big data analytics adoption in enterprises and smart city developments, this study explored and compared use cases from both areas. We found that two different approaches are followed in smart city initiatives. The first approach focuses on the development of smart services for citizens and their implementation. The second approach centres on the implementation of a basic infrastructure components, e.g. embedding sensors in the city's infrastructure or providing a central data pool. In this approach, governments develop and provide an infrastructure and expect use cases to come up in future from private companies using provisioned data and technologies. This shows that similar approaches are pursued in the development of smart cities and big data analytics adoption in enterprises Therefore, this paper concludes that the big data adoption approaches Business First and Platform Building are also present in the context of smart city developments.

ACKNOWLEDGEMENTS

This work was supported by the Science Foundation

Ireland grant “13/RC/2094” and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero - the Irish Software Research Centre (www.lero.ie).

REFERENCES

Anthopoulos, L., Janssen, M., Weerakkody, V., 2016. Smart Service Portfolios: Do the Cities Follow Standards? In Proceedings of the 25th International Conference Companion on World Wide Web.

Booch, G., 2010. Enterprise architecture and technical architecture. IEEE software vol. 27, no. 2.

Bremser, C., Piller, G., Rothlauf, F., 2017. Strategies and Influencing Factors for Big Data Exploration, Proceeding of 23rd Americas Conference on Information Systems, AMCIS 2017, Boston.

Carbonneau, R., Laframboise, K., Vahidov, R., 2008. Application of machine learning techniques for supply chain demand forecasting, European Journal of Operational Research vol. 184, no. 3.

Gandomi, A., Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management vol. 35, no. 2.

Hai, R., Geisler, S., Quix, C., 2016. Constance: An intelligent data lake system, In Proceedings of the 2016 International Conference on Management of Data, ACM.

ITU-T FG-SSC, 2014. Smart sustainable cities: An analysis of definitions,” Technical report, ITU-T.

Kiran, M., Murphy, P., Monga, I., Dugan, J., Baveja, S. S., 2015. Lambda architecture for cost-effective batch and speed big data processing, In Big Data IEEE International Conference.

LaRiviere, J., McAfee, P., Rao, J., Narayanan, V. K., Sun, W., 2016. Where Predictive Analytics Is Having the Biggest Impact, Harvard Business Review Digital Articles.

Mikalef, P., Pappas, I. O., Krogstie, J., 2017. Big Data Analytics Capabilities: A Systematic Literature Review and Research Agenda, Inf. Syst. E-bus. Manag.,.

O'Leary, D. E., 2014. Embedding AI and crowdsourcing in the big data lake, IEEE Intelligent Systems, vol. 29, no. 5.

Pourzolfaghar, Z., Helfert, M., 2016. Investigating HCI Challenges for Designing Smart Environments, In International Conference on HCI in Business, Government and Organizations. Springer International Publishing.

Pourzolfaghar, Z., Helfert, M., 2017a. Taxonomy of Smart Elements for Designing Effective Services, In Proceeding of 23rd Americas Conference on Information Systems, Boston.

Pourzolfaghar, Z., Helfert, M., 2017b. Utilising Latent Data in Smart Buildings: A Process Model to Collect, Analyse and Make Building Data Accessible for Smart Industries, In Proceeding of the Federated Conference on Computer Science and Information Systems, vol. 13.

Rathore, M. M., Ahmad, A., Paul, A., Rho, S., 2016. Urban planning and building smart cities based on the internet of things using big data analytics, Computer Networks, vol. 101.

Rondeau, E. P., Brown, R. K., Lapides, P. D., 2012. Facility management, John Wiley & Sons.

Ryazanova, O., Pétercsák, R., Heaphy, L., Connolly, N., Donnellan, B., 2016. Perception of Value in Public-Private Ecosystems: transforming the Dublin Docklands through smart technologies, Thirty Seventh International Conference on Information Systems.

Sivarajah, U., Kamal, M. M., Irani, Z., Weerakkody, V., 2017. Critical Analysis of Big Data Challenges and Analytical Methods, J. Bus. Res, vol. 70.

Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., Beghi, A., 2015. Machine learning for predictive maintenance: A multiple classifier approach, IEEE Transactions on Industrial Informatics, vol. 11, no. 3.

TechAmerica Foundation: Demystifying Big Data: A Practical Guide to Transforming the Business of Government, https://www.attain.com/sites/default/files/take-aways-pdf/Solutions_Demystifying Big Data - A Practical Guide To Transforming The Business Of GovernmGov.pdf.

Yan, R., Ghose, S., 2010. Forecast information and traditional retailer performance in a dual-channel competitive market, Journal of Business Research, vol. 63.