Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2019) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

The 10th International Conference on Ambient Systems, Networks and Technologies (ANT)
April 29 – May 2, 2019, Leuven, Belgium

Business Innovation in the Development of Big Data Toolboxes for the Management of Professional Vehicle Fleets

Athanasios Koumparos[a*], George Koutalieris[a], Christos Doulkeridis[b], Harris Georgiou[b], Akrivi Vlachou[b], Anagnostis Delkos[a], George Kaptsis[a]

[a]*Vodafone Innovus S.A. Athens-Lamia National Road, 12.5 Km, 14552 Metamorphosi, Greece*
[b]*University of Piraeus (UPRC), Greece*

Abstract

This paper introduces project Track&Know that aims at developing a new software framework to improve the efficiency of big data applications in geospatial contexts. We focus on our concept for innovating the fleet management use case and on the respective elicitation of related business requirements. Business innovation is considered with respect to the Track&Know research objectives and to their link with real business problems. Our approach foresees a continuous refinement process where research capabilities are introduced to product stakeholders and business requirements are continuously refined to solve real-world business problems. We conclude our presentation with an initial list of tacit business requirements aligned with specific Track&Know research objectives.

Keywords: Business Requirements, Big Data Analytics, Value Chain, Fleet Management, Mobile Asset Tracking

1. Introduction

The rapid progress of information and communication technologies (ICTs) have resulted major leaps in transportation data collection and management practices. In our times, the omnipresence of data feeds in this market, is posing new challenges both for big data researchers, developers and product managers wishing to exploit the advanced data generation capabilities created by connected and automated vehicle systems, and by the rapid growth of the Internet of Things (IoT).

* Corresponding author. Tel.: +306943220930;
E-mail address: athanasios.koumparos@vodafoneinnovus.com

Thanks to on-board devices, sensors, and wireless connectivity, big data have been expanding into the automotive and transportation sectors. Transport operators are increasingly capitalizing on the ability to corelate data from vehicles and people's behavior to data about the environment in which the vehicle is operating (e.g. traffic, weather, emergencies, etc.). Furthermore, the increasing ability of modern vehicles to communicate their location, velocity and other useful data, has enabled services that promise improved routing, traffic congestion reduction and management efficiency.

The enormous volume of mobility data in this new era has posed new challenges in the world of mobility data management [8] and Big Data Analytics (BDA) are now creating new capabilities to process and analyse massive amounts of complex data, reveal hidden patterns and identify correlations. Many transportation agencies see big data and its applications as an opportunity to improve the management and operation of transportation systems, increase the accuracy of prediction, enable informed decision making, and optimize transportation services.

Track&Know is a project funded by the European Commission under the Horizon 2020 Programme and is conceptualized within this framework, developing specialized applications (Toolboxes), tailor-made to the specific needs of the Transport industries. In this paper we are focusing on the Fleet Management domain which is one of the three Track&Know Demonstrators, the other two being: car insurance and healthcare telematics.

The rest of this paper is structured as follows: In Section 2, we discuss the role of big data in fleet management and in the supply chain. Section 3 outlines the related business needs for big data analytics. Then, in Section 4, we present the research challenges addressed by the Track&Know toolboxes. Section 5 presents new business requirements for effective fleet management, and Section 6 concludes the paper

2. The role of Big Data in Fleet Management and in the Supply Chain

Fleet Management systems rely on continuous real time monitoring of vehicles updates in order to maximise efficiency and provide quick responses to the fleet operator. Technologies such as mobile communications, Global Positioning Systems (GPS) and Geographical Information Systems (GIS) are combined with information systems that are storing collected data and provide fleet management applications to end users based on their needs. The pertinent Digital Transformation objectives of modern businesses, and their clear focus on Data Value generation [16] reposition modern Fleet Management Systems as key enablers of the overall digital transformation strategy of their organization, eliminating challenges and bringing down costs. For example, in temperature-controlled supply chains -i.e. Cold Chains- of the food sector, fleet generated data constitute an essential element of food safety and are required by numerous systems participating in the value chain [1].

Therefore, the role of fleet data management transcends the purpose of fleet operational improvements and is rapidly generating an essential element for the digital transformation of supply chains.

3. Business needs for Big Data Analytics in Fleet Management Applications

Big data research can address many facets of fleet data management such as driver performance, trajectory data and maintenance history. It is of significant value to use and analyse the large amounts of fleet generated data to deliver evidence-based support to fleet managers. New services can be developed in every step of the value chain and Big Data can be used to detect problems before they happen, preventing vehicles downtime and minimizing operational costs.
This specific market opportunity is the focus of Track&Know fleet management use case. During our business requirements analysis we had to align the fleet customer expectations with the specifications of the Track&Know Toolboxes. Since this pilot is based on an existing commercial Fleet Management product, the main sources for business requirements classification were the marketing and product management inputs provided by the vendor. Due to the existing product road-mapping and the established interactions of the vendor with the customer base, the identified priorities do not only constitute opportunities for the demonstration of Track&Know, but they also correspond to tacit customers' needs that can generate significant value and market acceptance.

3.1. Business opportunities

The drivers for Fleet Management improvements based on Big Data are both technical and customer related and serve the purpose of mining business value in data sources, going beyond standard Business Intelligence. The key benefits of the Track&Know platform are summarized below:

- Clean data with minimum errors,
- Insights due to correlations with the operational environment of the vehicles (e.g. weather, traffic etc.),
- Advanced management of large amounts of data on demand,
- Provide fleet managers with robust solutions for: driver behavior; fuel efficiency; and, predictive maintenance.

The overarching business objective is to provide insights of business value and user-friendly spatio-temporal analytics to fleet managers, via clean system responses, for matters related to operational fleet efficiency, on demand.

4. Track&Know Big Data Toolboxes

Track&Know is aiming at delivering big data innovation in business cases related to transport and mobility. To this end, a set of toolboxes is envisioned: Big Data Processing (BDP), Big Data Analytics (BDA), Complex Event Recognition (CER), Visual Analytics (VA). The overall positioning of the Track&Know Toolboxes against the BDVA reference model architecture is shown in Figure 1. In this paper, we consider business innovation related with toolboxes BDP and BDA because their associated research challenges have been scrutinised to support the contemporary needs of fleet management applications and business.

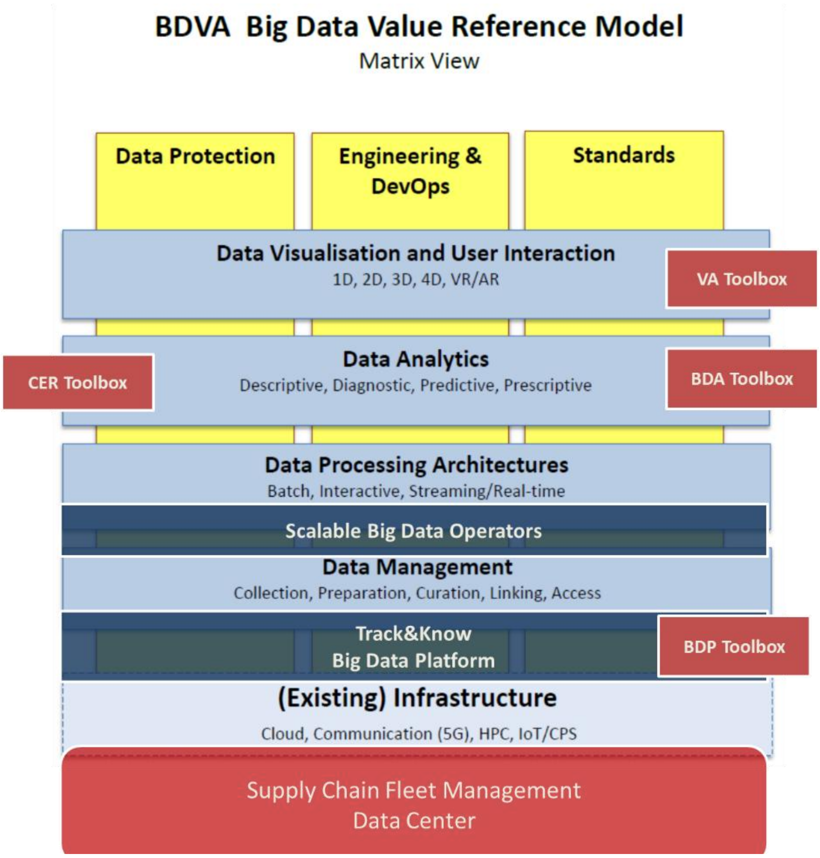

Figure 1: Positioning of the Track&Know Toolboxes against the BDVA reference model architecture

4.1. The Big Data Processing (BDP) Toolbox

The Big Data Processing (BDP) Toolbox aims at supporting novel, scalable, solutions of high throughput addressing storage, efficient access, indexing, partitioning and load balancing for Big spatio-temporal data.

The BDP Toolbox specifically considers the needs of researchers, scientists, practitioners, and developers that work with big mobility data [2]. A common problem encountered in the Big Data era is the lack of abstractions and easy-to-use interfaces for providing transparent access to scalable storage (NoSQL stores). This is intensified by the many different types of NoSQL stores [3], most notably: key-value stores, document stores, wide-column stores, and graph databases. To address this shortcoming, we develop a set of big data operators that are tailored to mobility data, thus supporting spatial, spatio-temporal, and trajectory-based retrieval of big mobility data. Typical examples of operators include range queries, aggregate range queries, and k-nearest neighbor queries. Moreover, in the context of Track&Know, implementations of the big data operators are provided for various NoSQL stores, thus hiding from the developer the peculiarities of each individual NoSQL store. By analogy with the relational DB landscape, the big data operators provided by BDP can be considered as the JDBC/ODBC equivalent for NoSQL stores.

In terms of research contributions, the functionality offered by BDP can be classified in two categories: (a) scalable storage and indexing, and (b) efficient query processing. In the former category, the focus is on efficient indexing by considering effective one-dimensional mapping of spatio-temporal data to integer values that serve as keys in NoSQL stores, thus facilitating efficient spatio-temporal filtering over keys [4,5]. Furthermore, distributed spatio-temporal partitioning and indexing techniques are also of interest. In the latter category, the research objective is efficient processing of complex query operators (such as spatio-temporal joins), by parallel data processing techniques that scale over vast quantities of mobility data [6,7].

All afore-described research targets are applicable to the fleet management use-case in Track&Know and particular attention has been put on describing the potential of these advances to business requirements related to trajectories of moving objects.

4.2. The Big Data Analysis (BDA) Toolbox

The Big Data Analysis (BDA) Toolbox will deliver scalable trajectory data mining techniques for voluminous data and real-time techniques to incrementally capture recurring or rapidly evolving phenomena.

New challenges of mobility data management are considered in terms of storing, querying, analyzing and extracting knowledge out of them in an efficient way [8]. One of these challenges is cluster analysis. The typical approach is to either transform trajectories to vector data, in order for well-known clustering algorithms to be applicable, or to define appropriate trajectory similarity functions, which is the basic building block of every clustering approach. Analyzing spatio-temporal data has the potential to discover hidden patterns or result in non-trivial insights, especially when its immense volume is considered. In this context, a useful data analysis task is Hot spot analysis, which is the process of identifying statistically significant clusters. However, there is practically very little research work on hot spot analysis for Big trajectory data [9].

Another critical functionality is matching raw or resampled GPS trajectories to the unrelying road network or "map". This can be implemented either as a single-point process, usually when the goal is to quickly discard invalid GPS data, or using a complete segment of a trajectory, when a specific sequence of road segments is considered as the most probable path in the maximum-likelihood sense (e.g. HMM-based) [17,18]. Both cases can be applied for batch or online processing, as an autonomous module for arbitrary use or inside the BDA (or BDP) main pipeline.

In many cases, trajectories have to be resampled for acquiring fixed-rate data rates or in sections where there are missing data. The interpolation module includes several processing steps that eventually produce properly resampled trajectory data at arbitrary fixed rates. This includes statistical analysis of the temporal dimension (reporting intervals), fitting appropriate statistical models to its distribution and estimating the corresponding limits for predefined probability bounds (e.g. 95% messages included & 5% outliers discarded). Subsequently, trajectory interpolation can be applied either with high-quality algorithms in batch mode, e.g. Hermite cubic spline using location & speed reference points, or in online mode, e.g. linear or quadratic using only few location reference points as they arrive. The result of this component is a new fixed-rate resampled trajectory that can be used for subsequent processing.

Prediction of the future location of trajectories is a key objective of Track&Know. Related challenges involve Future Location Prediction (FLP) and Trajectory Prediction (TP) tasks. The FLP problem finds two broad categories of application scenarios. The first scenario involves cases where the moving entities are traced in real-time to produce analytics and compute short-term predictions, which are time-critical and need immediate response. Short-term FLP can be extremely important in domains where safety, adaptiveness and responsiveness out outmost importance and a decision-making process. The second scenario involves cases where long-term FLP is important to identify cases which exceed regular mobility patterns, detect anomalies, and determine a position or a sequence of positions of special interest at a given time interval in the future. In this case, although response time may not be a critical factor per se, it is still crucial in order to identify correlations between historical mobility patterns and patterns that are expected to appear, e.g. approach to a restricted area. Similarly to FLP, the TP task addresses the prediction problem in the sense of complete trajectories. That is, given the recent history of S previous trajectories of one or more moving objects, i.e., each consisting of its time-stamped data points recorded in the past, predict the anticipated future trajectory of the same or "similar" objects, based on some common reference initialization (e.g. starting point, time frame, region of interest, etc). The main factors for any TP algorithm are size of the history (N) and how it is exploited by a predictive model. In principle, the TP problem can be approached as a generalization of the FLP problem [10, 11, 12], but this approach results in exponentially increasing global errors due to accumulating local errors as the look-ahead time window increases.

A more recent approach for addressing some aspects of these challenges comes from the area of Predictive Queries (PQ) [13, 14], which is one of the most exciting research topics in spatio-temporal data management, especially for the Big Data context. In many location-based services, including traffic management, ride sharing, targeted advertising, etc, there is a specific need to detect and track mobile entities within specific areas and within specific time frames. In Range Queries (RQ), the task is focused on identifying POIs and mobility patterns related to the current locations of moving objects. Instead, Predictive Range Queries (PRQ) address the same task but for future time frames.
By exploiting these research capabilities and interacting with the Fleet Management product roadmap, we plan to extend the state-of-the-art of the Fleet Management applications, aiming in the same time at extending the applicability in other European industries.

5. Innovating the Business Requirements

Business requirements innovation requires collaborative requirements elicitation techniques to allow for a wide-range of topics to be discussed among stakeholders [15]. In Track&Know the research objectives toolboxes were introduced to the product managers of the fleet management solution in order to elicit unsolved customer needs. These needs provide the innovation opportunity for further market take-up of big data in transport and mobility applications. Table 1 summarises our key findings for new fleet management business requirements, mapped against the BDP and BDA toolboxes capabilities.

Track&Know Capabilities	**Related Fleet Management Business Requirements**
BDP: Reliable data collection modes	- Increase the number of external sources integrated in by the Big Data Platform (e.g. weather, holidays and geographic points identification, etc.)
BDP: Track&Know Big Data Operators	- Reduce invalid coordinates introduced due to errors in the fleet monitoring system - Reduce cases of invalid speed calculations due to errors in the fleet monitoring system
BDA: Support for computing intensive, analytic processing, and machine learning techniques	- Identify driving behavior excess per driver - Provide recommendations for fuel consumption reduction based on driver behavior - Identify patterns leading to improved fleet maintenance costs - Support preventive maintenance recommendations based on tracked parameters (service downtime, tire life, etc.)
BDA: Future Location Prediction (FLP)	- Proactive identification of traffic hot spots per day - Alternative routes per identified hot spot
BDA: Trajectory	- Provide recommendations for fuel consumption reduction based on the overall fleet performance

Prediction (TP)	optimization - Provide accurate estimations of future travel distances - Increase the recommendations for alternative routes based on fuel economy and road conditions

Table 1: Mapping fleet pilot business needs to Big Data platform capabilities

As shown in Table 1, there are different needs aligned with different capabilities of the Toolboxes. This is an iterative process and further concepts are expected to be developed as the research results will feed the product management process.

6. Discussion and Conclusion

In this paper, we introduced the main business challenges related with the ability to use big data analytics to deliver data-driven business value for Fleet Management applications. Business innovation is elicited via collaborative experimentation with software toolboxes research achievements. Additional adjustments and service specification refinements are expected to take place according to the system's evolving needs and pilot evaluation results.

As part of our future work, we are implementing data connectors for a wide variety of data sources, addressing challenges such as online filtering and compression, information extraction, cleaning and noise elimination as part of the Big Data Processing toolbox. Also, we are developing customized data analysis methods and tools over Big Mobility Data, including cluster analysis and motion pattern detection, by exploiting enriched and integrated fleet management data sources, in the context of the Big Data Analytics toolbox.

Acknowledgements

The research described in this paper has been funded by project Track&Know that has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 780754. The authors would like to thank the fellow researchers in the Track&Know project and the anonymous reviewers for their insightful and constructive comments.

References

[1] Lucia Ramundo, Marco Taisch and Sergio Terzi, “State of the art of technology in the food sector value chain towards the IoT”, 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI) IEEE, (2016). 1 6 7740612, 10.1109/RTSI.2016.7740612 http://ieeexplore.ieee.org/document/7740612/
[2] A.Eldawy, M.F.Mokbel (2016). “The Era of Big Spatial Data: A Survey”. Foundations and Trends in Databases 6(3-4): 163-273.
[3] A.Davoudian, L.Chen, M.Liu (2018). “A Survey on NoSQL Stores”. ACM Comput. Surv. 51(2): 40:1-40:43.
[4] A.Vlachou, C.Doulkeridis, A.Glenis, G.M.Santipantakis, G.A.Vouros (2019). “Efficient spatio-temporal RDF query processing in large dynamic knowledge bases”. In Proceedings of the 34th Annual ACM Symposium on Applied Computing (SAC), 2019 (to appear).
[5] P.Nikitopoulos, A.Vlachou, C.Doulkeridis, G.A.Vouros (2018). “DiStRDF: Distributed spatio-temporal RDF queries on Spark”. In Proceedings of the Big Mobility Data Analytics (BMDA) workshop, pages 125–132, 2018.
[6] Z.Shang, G.Li, Z.Bao (2018). “DITA: Distributed In-Memory Trajectory Analytics”. SIGMOD Conference: 725-740.
[7] X.Ding, L.Chen, Y.Gao, C.S.Jensen, H.Bao (2018). “UlTraMan: A Unified Platform for Big Trajectory Data Management and Analytics”. PVLDB 11(7): 787-799.
[8] N.Pelekis, Y.Theodoridis (2014). “Mobility Data Management and Exploration”. Springer.
[9] P.Nikitopoulos, A.-I.Paraskevopoulos, C.Doulkeridis, N.Pelekis, Y.Theodoridis (2018). “Hot Spot Analysis over Big Trajectory Data”. In Proceedings of IEEE Big Data.
[10] Y.Zheng (2015). “Trajectory data mining: An overview”. Transactions on Intelligent Systems and Technology 6(3):1-41.
[11] M.Hamed, D.Gianazza, M.Serrurier, N.Durand (2013). Statistical prediction of aircraft trajectory: regression methods vs point-mass model. Proceedings of ATM.
[12] S.Theodoridis, K.Koutroumbas (2008). “Pattern Recognition”. 4th Edition, Academic Press.
[13] A.M.Hendawi, M.F. Mokbel (2012). “Predictive Spatio-Temporal Queries: A Comprehensive Survey and Future Directions”. In ACM SIGSPATIAL MobiGIS’12, November 6, 2012, CA, USA.

[14] R.Zhang, J.Qi, D.Lin, W.Wang, R.C-W.Wong (2012). "A highly optimized algorithm for continuous intersection join queries over moving objects", VLDB Journal, 21:561-586.
[15] N. Muhammad, G. Koutalieris, M. Streefkerk, N. Poot, S. Alloing and R. Wyns. "A Model to anticipate and analyse Requirements of Heritage Organisations wishing to actively participate in Europeana", In Proceedings of DIGITAL HERITAGE -2013, pages 691-694, Marsseiles/France, November 2013, IEEE Catalog Number: CFP1308W-POD, ISBN: 978-1-4799-3171-2.
[16] Curry E. (2016) The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy. Springer, Cham
[17] C.Y. Goh, J. Dauwels, N. Mitrovic, M. T. Asif, A. Oran, P. Jaillet (2012). "Online map-matching based on Hidden Markov model for real-time traffic sensing applications", Proc. 15th International IEEE Conference on Intelligent Transportation Systems, 16-19 Sept. 2012, Anchorage, AK, USA, DOI:10.1109/ITSC.2012.6338627.
[18] P. Newson, J. Krumm (2009). "Hidden Markov Map Matching Through Noise and Sparseness", Proc. ACM GIS '09 , 4-6 Nov. 2009. Seattle, WA, USA.