Enterprise Business Intelligence Approach With Cloud-Based Analytics

1st Jelena Plašić
Department of Information Technologies
Faculty of Technical Sciences,
University of Kragujevac
Čačak, Serbia
jelena.plasic@ftn.edu.rs

2nd Nenad Stefanović
Department of Information Technologies
Faculty of Technical Sciences,
University of Kragujevac
Čačak, Serbia
nenad@ftn.edu.rs

3rd Andrijana Gaborović
Department of Information Technologies
Faculty of Technical Sciences,
University of Kragujevac
Čačak, Serbia
andrijana.gaborovic@ftn.edu.rs

***Abstract* — In today's rapidly changing e-business environment enormous amount of heterogeneous and high-velocity data is being generated. A critical component for success of the modern enterprise is its ability to take advantage of all available information. In this paper, business intelligence (BI) system for advanced analytics based on modern cloud-based data warehouse architecture and machine learning models is presented. Its main purpose is to support insights-driven business decision making by integrating relevant data form various sources, and transforming it into usable and accessible information and knowledge. In order to demonstrate the effectiveness and usefulness of the proposed BI system, an experiment with the real-world dataset from the retail industry has been carried out. Data has been integrated into the single multidimensional data warehouse. For advanced analytics several machine learning (ML) models have been created. These ML models complement data warehouse reporting with advanced knowledge that can be used for effective decision making and proactive actions.**

Keywords — business intelligence, data science, data warehousing, machine learning, cloud analytics

I. INTRODUCTION

The modern business environment, characterized by high competition growth and great dynamism, puts companies in a situation to struggle to maintain market share and achieve better results, and thus to increasingly take into account the problem of managing large amounts of data.

Business intelligence (BI), data science and big data technologies encompass the collection of platforms, services, and tools, as well as data modelling and analytical methods for efficient data integration, storage, processing and reporting.

BI is used by retailers to understand customer needs, optimize price alignment with current trends and determine upcoming trends [1]. It is useful for addressing supply chain risks in categories of demand, supply, process and environmental risk in the retail industry, which is crucial for retail companies [2].

Reliance on adequate and relevant data allows businesses to make effective decisions and ultimately achieve a competitive advantage [3]. According to [4] the need for continual analysis and innovation to remain competitive through new opportunities is what sets a lot of successful businesses apart from failing ones.

II. BUSSINESS INTELIGENCE MODEL

The purpose of Business Intelligence is to make it possible to combine data from multiple sources, analyze and systemize information, and then disseminate the information to relevant stakeholders. That way, companies are able to see the big picture and make smarter business decisions.

Business intelligence includes processes like data preparation, in which raw data from different source systems should be integrated, consolidated and cleansed, storing this data in a data warehouse, analytical query of this data and distribution of the obtained information to help influence and drive business decisions.

Due to the growing diffusion of big data in the business context, enterprises increasingly need to redefine their knowledge management systems in order to make them capable of managing the different types of complex data available in a dynamic and transparent way [5].

Traditionally, enterprises have been gathering data in data warehouse which was an on-premises operation and the larger it grew, the more IT infrastructure and resources were required to support it. The traditional enterprise data warehouse has never been designed to support near real-time transactions or event processing, it hasn't been able to keep up with variety of new data sources and explosive volumes of data and has been slowing down the ability to do advanced analytics.

The traditional data warehouse was designed specifically to be a central repository for all data in a company. Data was extracted, transformed, and loaded (ETL) into the warehouse within an overall relational schema, and with increasing data volumes, new sources and types of data, the schema model was expanding in order to keep up.

The traditional data warehouse was built on symmetric multi-processing (SMP) technology. A symmetric multi-processing architecture means multiple processors for extra power, but these processors share a single operating system, memory pool and they share access to the disks. Enterprises can use different design strategies to design data warehouses for improving the performance and managing the volume efficiently, but oftentimes this means adding more resources and spending more money for addressing the requirement. Because of the need for bigger, more powerful hardware and ever-larger forklift

migrations to support volume growth, they are looking for an alternative that won't break their budget.

Besides increasing data volume, there are few more reasons why enterprises should consider moving to another kind of architecture, like the need for real-time data and new sources and types of data.

The traditional data warehouse was not architected to support the capacity, loading and complexity of the real-time transactions, so performance of the query might not be at the expected level. Traditional data warehouse also requires well-structured, sanitized, relational formatted data, but there is a variety of new data types, from mobile and social channels, devices, growing amount of cloud-born relational business data and other sources outside the business, that have the potential to enhance business operations, but could not easily fit the business schema mode.

According to [6] a cloud-based environment eliminates many common issues with data warehousing and also offers many new opportunities. Since there is desperate need for such elastic capabilities as data warehousing and big data, data warehouse must modernize in order to keep up. The modern data warehouse starts with the ability to handle both relational and non-relational data sources and to rapidly provision and release resources to match what a workload requires.

Massively parallel processing (MPP) environment has capabilities to handle these complexities and improve the overall performance. MPP is the coordinated processing of a single task by multiple processors, each processor using its own operating system and memory and communicating with each other using some form of messaging interface. Key differences between SMP and MPP environments are illustrated in Figure 1.

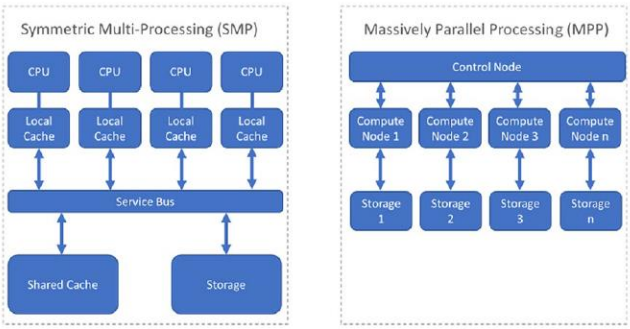

Fig. 1. Key differences between SMP and MPP environments

With MPP environment adding additional storage after the initial implementation might not be as expensive as SMP. MPP helps to process unstructured data much efficient than SMP - since data can be distributed with multiple nodes, performance on data retrieval is fast and real-time queries are efficiently handled.

Modern data warehouse is designed so that can handle both relational and non-relational data, provides a way to easily interface with all types of data through one query model, and can handle "big data" while providing very fast queries. Its layers are illustrated in Figure 2.

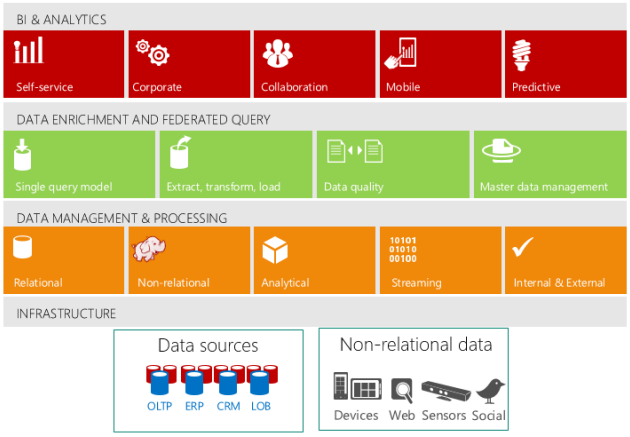

Fig. 2. Modern data warehouse

Recent studies have begun to empirically demonstrate the value that big data and business analytics have on organizational-level outcomes, such as agility [7], innovation [8] and competitive performance [9][10].

According to [11] increasing competition has further contributed to the complexity and made achieving business growth and sustainability very challenging, therefore involvement in work, consistency and mission further enable an organization's members to understand the need for, and the usefulness of, the system and to make efforts to use it effectively.

Those who adopted approaches in business intelligence, besides competitive advantages, have experienced improvement in decision making capabilities, better customer service and increased revenue, as claimed by [12].

III. Results and Discussion

Machine learning uses well-researched statistical principles to discover patterns in large sets of data. Typically, these patterns cannot be discovered by traditional data exploration because the relationships are too complex or because there is too much data. Applying the machine learning algorithms makes it possible to forecast trends, identify patterns, create rules and recommendations, analyze the sequence of events in complex data sets, and gain new insights.

Dataset that has been used is sourced from multinational manufacturer and seller of bicycles and accessories. The products were classified into five categories: bicycles (mountain, road and tourist), accessories (such as helmets and water bottles), clothing, components and services. There was a huge amount of information of this company in the transaction system, so queries over the transaction database could reveal a lot about the dynamics of business and the nature of data, but in order to obtain better data analysis, a data warehouse and business intelligence were needed.

Figure 3 illustrates proposed BI solution.

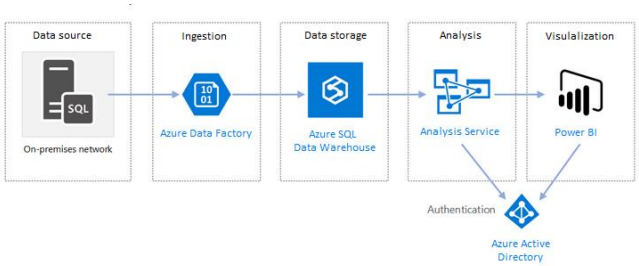

Fig. 3. Proposed BI solution

Successful implementation of a business intelligence depends on the appropriate integration of many tasks and components.

Business intelligence process includes data preparation, in which raw data from different source systems should be integrated, consolidated and cleansed. Since Azure Data Factory is data integration service, it is suitable for this process.

Based on this data Azure SQL data warehouse can be created. By the ETL process, the data from several sources are stored in tables of dimensions and facts, suitable for online analytical data processing.

For advanced analytics, such as predictions, classifications, and associations, several machine learning (ML) models have been created. These ML models complement data warehouse reporting and could be the basis for making effective decisions.

Besides advanced analytics and knowledge extraction, it is very important to deliver information to decision-makers in a timely manner and in adequate format. Reporting is realized through cloud-based BI service which enable rich visualizations, self-service analysis and collaborative decision-making.

Microsoft Azure provides an integrated platform for predictive analytics that encompasses data cleansing and preparation, machine learning, and reporting. Therefore this environment and its machine learning models were used. Results are presented in Figure 4.

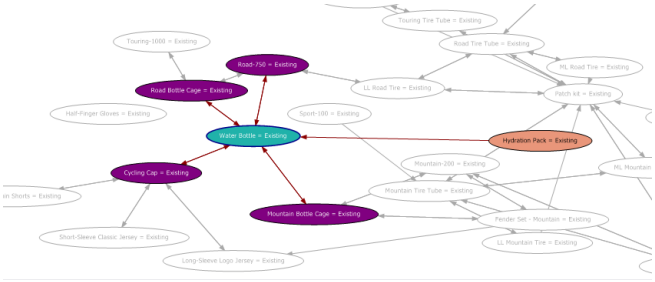

Fig. 4. Dependency network

Once the model is created and processed, the generated results can be viewed. This figure shows the result of the applying ML model with Association Rules algorithm in the form of a dependency network, where interactions of different items in the model can be explored. Each node in the view represents an item, while the lines between them represent rules. Prediction connections can be observed by selecting a node. In some cases, there is a two-way connection between the items, which means that they often appear in the same transaction. The line connecting the two items indicates that these items are likely to appear in the transaction together which means that customers usually purchase these items together.

The graph shows all nodes by default, but when complex this can be less clear, so it is possible to zoom in further if necessary to see the details. By filtering, it is possible to see which connections are stronger and which are weaker.

Predictions of an association model can be very useful because they can recommend items to a customer, based on prior or related purchases, find related events and identify relationships in or across sets of transactions.

In order to generate more insights and discover hidden trends and patterns several more ML models and algorithms were created. Results of another ML models that were used are presented with Power BI, cloud-based BI service, in order to provide better interactive visualizations and gain some insights.

Category outliers ML model automatically extract hidden knowledge by highlighting cases where one or two categories have much larger values than other categories. For example, the Figure 5 shows that those are accessories and bikes according to sales orders.

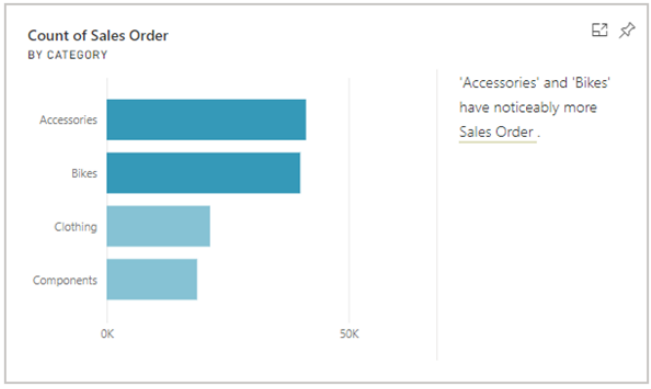

Fig. 5. The result of using the category outliers

If we use scoped insights, ML models can give us some addition insights - there is correlation between sales order and quantities, there was the most accessories in 2020 years (Figure 6).

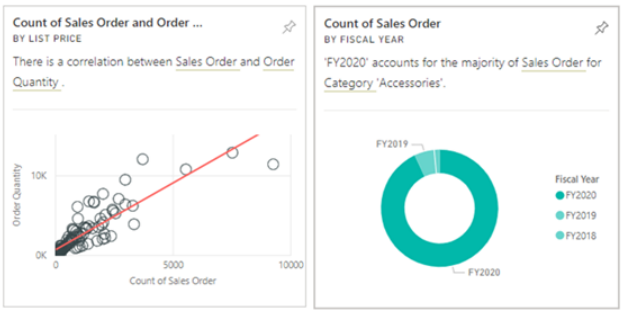

Fig. 6. Scoped insights

Figure 7 shows that the most sales order are in the United States, some regions have better sell than others and that accessories have more sales orders than other products in 2020 year.

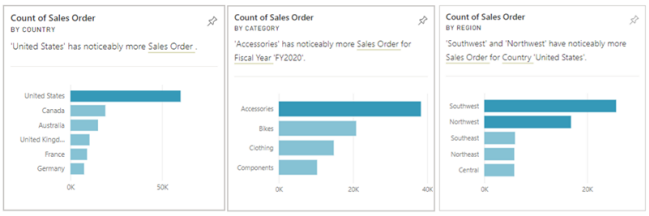

Fig. 7. Scoped insights

Applying traditional analytical methods to real-world business problems is time-consuming and challenging and requires experts in the several disciplines, but machine learning changes that, because it offers automated way to pull the most relevant information from the data.

IV. Conclusion

The world of business is constantly evolving, so the strategic importance of cloud computing is sure to become more prominent. The need for enterprises to expand on flexibility, security and speed is increasing and by incorporating cloud services, they can get all those advantages.

The study has presented business intelligence approach with cloud-based analytics based on data warehousing and machine learning models and its impact on the business performance. Systems like this are extremely important for managing today's global businesses.

Created machine learning models show that BI tools can help gain the knowledge needed to increase profitability through cross-selling, recommendations and promotions.

Enterprises should realize the benefits of using BI tools and understand their needs to successfully implement those new technologies that help them achieve successful business goal and stay ahead of the competition.

Acknowledgment

This work was supported by the Serbian Ministry of Education, Science and Technological Development (Agreement No. 451-03-9/2021-14/200132).

References

[1] Zamba, C. R., Mahlangu, G., Giyane, M., & Rebanowako, T. G. (2018). "An investigation of the potential benefits and challenges of business intelligence adoption in the retail sector in Gweru, Zimbabwe". Journal of Systems Integration, 9(2) (2018): pp. 20-25.

[2] Olexová, C. "Business intelligence adoption: a case study in the retail chain." WSEAS transactions on business and economics 11.1 (2014): pp. 95-106.

[3] Ritesh, C. & Srimannarayana, G. (2013), "Why Business Intelligence?: Significance of Business Intelligence Tools and Integrating BI Governance with Corporate Governance", International Journal of E-Entrepreneurship and Innovation (IJEEI), 4, pp. 1-14

[4] Liu, Y., Han, H. & Debello, J. (2018), "The challenges of business analytics: successes and failures".

[5] Intezari, A. and Gressel, S. (2017), "Information and reformation in KM systems: big data and strategic decision-making", Journal of Knowledge Management, Vol. 21 No. 1, pp. 71–91.

[6] How, M. "Beyond the Modern Data Warehouse - The Modern Data Warehouse in Azure". Apress, Berkeley, CA, (2020), pp. 229-274

[7] Ashrafi, A., Ravasan, A. Z., Trkman, P., & Afshari, S. (2019). "The role of business analytics capabilities in bolstering firms' agility and performance." International Journal of Information Management, 47, pp. 1-15.

[8] Lehrer, C., Wieneke, A., vom Brocke, J., Jung, R., & Seidel, S. (2018). "How big data analytics enables service innovation: materiality, affordance, and the individualization of service." Journal of Management Information Systems, 35(2), pp. 424-460.

[9] Côrte-Real, N., Ruivo, P., & Oliveira, T. (2019). "Leveraging Internet of Things and Big Data Analytics Initiatives in European and American Firms: Is data quality a way to extract business value? Information & Management."

[10] Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2019). "Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. Information & Management."

[11] Shan, S., Luo, Y., Zhou, Y., & Wei, Y. (2019). "Big data analysis adaptation and enterprises' competitive advantages: The perspective of dynamic capability and resource-based theories." Technology Analysis and Strategic Management, 31(4), pp. 406–420.

[12] Bhatiasevi, V., & Naglis, M. (2020). "Elucidating the determinants of business intelligence adoption and organizational performance." *Information Development*, *36*(1), pp. 78-96.