

Big Data Business Intelligence: A quality driven approach

Supervisors:

Samira Si-said Cherfi (CEDRIC - Cnam), Professor
Faten Atigui, (CEDRIC - Cnam), Associate professor

Key words: Business Intelligence, Big Data, Data Quality, Process Quality, Analytics

With the recent advances in Big Data storage and processing, there is a real potential of data-centric systems, i.e., systems that employ analysis of large amounts of data to support runtime decisionmaking. Traditional Business Intelligence (BI) solutions [1,2] are constructed around a three layers architecture that are (1) data sources, (2) data storage and management (typically within a data warehouse) and (3) an On-Line Analytical Processing (OLAP) and reporting tools. With the emergence of big data, data analytics is gaining increasing attention in business and BI solutions become Data-Driven rather than “business” and intuition driven [3]. Indeed, the enterprise data is not always sufficient to explain observed situations. Predefined analysis models and analysis axes may not be able to provide a clear explanation on why a given phenomenon happens. The solution requires often extra data that is not included in corporate systems. Consequently, there is an obvious need to be able to cope with high technology standards to face big data specific complexity (Volume, Variety, Velocity, etc.) [3,4]. Furthermore, there is a clear and strong need to enforce the importance of business needs and expertise to enhance the “intelligence” implied by business analysis that does not always come as a by-product of data analysis. Indeed, the availability of such data with open data sources initiatives for example, do not allow immediate analysis as analysis models do not integrate this data. There is a real need to bring analytics power to data instead of integrating data into analysis systems [5]. There is a plethora of application areas which require interactive and on the fly responsive business intelligence system [9]. Reactive/Interactive systems using data from data sensors, social feeds, heterogeneous IoT are examples of applications needing such an approach.

Challenges and Research Questions

Going on a traditional vision of business intelligence solutions implies re-modelling the data warehouse, uploading, transforming and testing the new data [6]. In addition to the re-modeling problem, other issues related to data arise such as data quality, data understanding and data availability [7,8]. These problems related to data will definitely generate performance degradation going against the expected reactivity. The intended research work will follow the underlying agenda:

- Provide solutions to access data sources on the fly: this includes mechanism helping data understanding and filtering.
- Help evaluate data quality: this is a crucial step in the business intelligence solution architecture. It should be analytics needs dependent but should also be assisted by automated tools to reduce analysts' time dedicated to data cleaning task.
- Provide on the fly solutions for both analytical model redesign and analytical process configuration.
- Finally, the problem of scalability should in a Big Data context to make solutions applicable to real cases.

Candidate requirements

Applicant should meet the following requirements:

- possesses a solid background in Computer Science, Data Science, or Mathematics (demonstrated by a relevant Master);
- has a strong background information systems modeling;
- has a strong interest in data science research and real-life applications of analytics;
- has strong skills in software development to be able to realize research ideas in terms of software prototype;
- has excellent communication skills in English

References

- [1] W. H. Inmon. Building the Data Warehouse. QED Information Sciences, Inc., Wellesley, MA, USA, 1992.
- [2] Ralph Kimball. The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses. John Wiley, 1996.
- [3] Christian Bizer, Peter A. Boncz, Michael L. Brodie, Orri Erling: The meaningful use of big data: four perspectives - four challenges. SIGMOD Record 40(4): 56-60 (2011).
- [4] Hsinchun Chen, Roger H. L. Chiang, Veda C. Storey. Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly 36(4): 1165-1188 (2012).
- [5] Alfredo Cuzzocrea. Data warehousing and OLAP over Big Data: a survey of the state-of-the-art, open problems and future challenges. IJBPIM 7(4): 372-377 (2015)
- [6] Matteo Golfarelli, Stefano Rizzi. From Star Schemas to Big Data: 20+ Years of Data Warehouse Research. A Comprehensive Guide Through the Italian Database Research 2018: 93-107. 2017.
- [7] Thota S. (2017) Big Data Quality. In: Schintler L., McNeely C. (eds) Encyclopedia of Big Data. Springer, Cham, p. 1-5.
- [8] Kim, M., Zimmermann, T., DeLine, R., & Begel, A. (2017). Data Scientists in Software Teams: State of the Art and Challenges. *IEEE Transactions on Software Engineering*.
- [9] Vassiliadis, P., Marcel, P., & Rizzi, S. (2019). Beyond roll-up's and drill-down's: An intentional analytics model to reinvent OLAP. *Information Systems*.