Discussion of an Architecture Model for a Business and Big Data Analytics Platform

Masterarbeit

zur Erlangung des akademischen Grades „Master of Science (M. Sc.)“ im Studiengang Wirtschaftswissenschaft der Wirtschaftswissenschaftliche Fakultät
der Leibniz Universität Hannover

vorgelegt von:

Name: Paßlick [redacted]

Vorname: Jens Hendrik [redacted]

Prüfer:
Prof. Dr. Michael H. Breitner

Hannover, den 29.02.2016

Table of Contents

1 Introduction

For many companies, the importance of Data and Information is growing (Demirkan & Delen 2013). The purpose of Business Intelligence (BI) is to turn data into information in order to make decisions based on this obtained information. The development of the last years shows the effective use of a Business Intelligence platform is not a competitive advantage anymore but a requirement to be able to keep up with the global competition.
Architectural models can help to get an overview of possible BI elements of companies. They can guide organizations and help them with the implementation of new architecture elements. Moreover, architectural models are helpful for identifying research trends and gaps. (Phillips-Wren et al. 2015, p. 453) Based on a definition by Ponniah (2001, p.127), a Business Intelligence Architecture can be described as *"the structure"* that connects all the BI components.

A study by Russom (2014, p. 11) confirms the need for an architecture model by pointing out that 79% of respondents think that the architecture is extremely important for the success of a Data Warehouse and its related platforms. There is also a movement within BI architectures. 76% of respondents said that their Data Warehouse environment is evolving (Russom 2014, p. 8). According to the study (Russom 2014), especially advanced analytics, big data aspects and real-time analysis are reasons for the modification of an existing architecture.

Besides the need for improved BI through advanced analytics, there is also a demand for Big Data and real-time analysis. Big Data is currently one of the most discussed topics in IT. It arises from trends such as the Internet of Things, Industry 4.0 or even new business models (Dittmar 2016). Many of the aforementioned use cases cannot be processed with conventional techniques. If this is the case, it is per definition a Big Data event. The realization of many Big Data use cases need new technologies. For many companies the question arising is how to change the existing BI architecture in order to meet these new demands and how to include these new technologies into their existing architectures.

After conducting a structured literature review based on Webster & Watson (2002), several aspects have been identified to be possibly relevant for an advanced BI / Big Data architecture. Various architectural models already exist. They all have a different focus. But at the moment, there is no architecture model which includes all identified aspects at once. Pospiech & Felden (2013) and Essaidi (2010) developed reference architectures for service-oriented BI. Watson (2014) focuses on the differences between top-down and bottom-up BI. The architecture provided by Parenteau et al. (2015) shows how a realization of Self-Service BI could be possible. Consequently there are some architecture models but they nearly all emphasize different aspects.

Scientific research is needed to combine the benefits of existing architectures and to find a consensus on how an optimal BI / Big Data architecture is built.
Resulting from the study by Russom (2014) there is a demand for an efficient BI architecture model for companies, which leads to the first research question:

> RQ1: How is an advanced Business Intelligence or Big Data architecture constructed?

This research question also addresses knowledge management strategies, practices and strategies for sharing BI knowledge. Phillips-Wren et al. (2015) propose this research direction. Based on the literature review a new model is developed. Afterwards, experts discuss the model in 18 semi-structured expert interviews and help to improve it.

Another question arises when companies want to renew components of their BI architecture. Due to the high expenses caused by restructuring, companies have to set priorities in which order they have to change their architecture. This leads to the next research question:

> RQ2: Which elements and which characteristics have priority in a Business Intelligence or Big Data architecture?

Statements from the expert interviews and a quantitative survey, which is conducted after the expert interviews, should generate findings for this research question.

The structure of this thesis is Design Science oriented. More specifically, the DSRM Process Model by Peffers et al. (2007) is used. The focus of this thesis is to develop a new artifact which is according to the first research question a BI / Big Data architecture model. In the next chapter, a definition for the terms Business Intelligence, Data Warehouse and Big Data is given. After that the research design is described. The problem is identified with a literature review based on Webster & Watson (2002). This chapter is similar to the second stage in the DSRM Process Model by Peffers et al. (2007) which is called “Define Objectives of a Solution”. The fifth chapter deals with the initial model which was developed based on the literature review. Then the refinement with expert interviews is described in the next chapter. This structure was chosen in a way that the changes can be traced. Then, the final model or “artifact” is pictured in detail. The results are discussed followed by implications and recommendations resulting from this research. Afterwards, the limitations of this study are presented. The final chapter concludes the thesis and addresses further research.

11 Conclusion and Outlook

This thesis presented a new BI and Big Data architecture model. The process from raw data to the presentation of information is described. There is also a discussion about the different user groups of a BI / Big Data architecture. The model represents a significant progress to the proposal by Phillips-Wren et al. (2015). Many more ideas have been added. The ideas result from both practical and academic literature as well as from interviews with experts. In addition, coarse statements about the importance of several characteristics and elements of the BI architecture are possible.

Especially self-service BI seems to be an important aspect of a BI architecture. The experts agree that it is needed. But not all of them are of the opinion that self-service BI is feasible everywhere. It is also unclear how a realization would be possible in detail. The developed architecture tries especially with the Knowledge Database and with the Collaboration rooms to support Self-Service BI.
Another discussion point is the definition of different user groups of a BI architecture. Although many opinions differ, most literature sources and experts proposed a separation into three groups. For some companies, e.g. it can also be quite useful to divide the user group of the "Business Analysts".

The proposed reference architecture model gives the possibility to evaluate a current BI company architecture. Obviously each missing element has to be evaluated for the company because some functions are not necessary relevant for every company.
Further research can be done when asking how the architecture must be adjusted according to the size of the company. As mentioned by the experts, for example, the effectiveness of the knowledge database is very dependent on the amount of input data. It is assumed that small to medium-sized enterprises cannot provide enough data to obtain meaningful results from the database, if they do not want to give data to the cloud.

The researched weights of the elements and characteristics of the architecture could be the starting points for a maturity model. The "must have" elements could represent the aspects of architectures with a low maturity level. The "nice to have" items could suit architectures with a high level of maturity.
Another result could also be a phase model. This phase model could give an idea of a useful implementation order of architecture elements. This could support companies with the implementation or renewing of their BI / Big Data architecture.

Another aspect for further research could be the different model customization for different sectors. It is assumed that the individual quantitative values of the architectural elements and characteristics are different depending on the company sector. Especially for those elements which are characterized as “nice to have” the value could be dependent on the branch of the company.

In association with that it would be helpful to get an idea of the cost and value of the “nice to have” elements. The value is very hard to quantify. But estimating the cost of an implementation might be possible.

The differences between the various user groups and their implications for the architecture could be another research aspect. If companies can assign all of their BI users into clearly defined groups, perhaps the differences in the weighting of the BI architecture elements is more distinct.

More research is needed on the question of the extent to which certain kinds of analysis can be carried out directly on source systems and without the use of a data warehouse. This question was posed by the expert interviews. The majority of experts believe that a component for data integration will still be needed. But to what extent can data integration be performed directly on source systems in future analysis?

Further research is necessary for the implementation of self-service BI. Collaboration rooms and the knowledge database can support self-service BI but what are critical factors of success? Also, it remains unclear to what extent the actual decision may be involved in a BI architecture.