# Business Intelligence and Big Data in the Cloud: Opportunities for Design-Science Researchers

Odette Mwilu Sangupamba [1], Nicolas Prat[2], and Isabelle Comyn-Wattiau[1,2]

[1] CEDRIC-CNAM, Paris, France
odemwilu@hotmail.com, isabelle.wattiau@cnam.fr
[2] ESSEC Business School, Cergy-Pontoise, France
{prat,wattiau}@essec.edu

**Abstract.** Cloud computing and big data offer new opportunities for business intelligence (BI) and analytics. However, traditional techniques, models, and methods must be redefined to provide decision makers with service of data analysis through the cloud and from big data. This situation creates opportunities for research and more specifically for design-science research. In this paper, we propose a typology of artifacts potentially produced by researchers in design science. Then, we analyze the state of the art through this typology. Finally, we use the typology to sketch opportunities of new research to improve BI and analytics capabilities in the cloud and from big data.

**Keywords:** Business Intelligence, Big Data Analytics, Cloud Computing, Design-Science Research, Artifact.

## 1 Introduction

Business intelligence (BI) helps managers to make informed decisions. BI technology is demonstrated an indisputable support for decision making. BI tools facilitate the presentation of more accurate reporting, improve decision making, enhance customer relationships, and increase revenue. BI must be able to deal with big volumes of data (big data analytics). According to IDC [1], the business analytics software market will grow at a 9.7% compound annual rate through 2017. The growth of the market will be driven in part by the current hype around big data.

Cloud computing also attracts many organizations, because of its potential: ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) [2, 3]. Its objective is to provide innovative services to the request of different types of users. The latter are freed from the underlying infrastructure. Beyond outsourcing, there are two concepts that are highlighted in cloud computing: virtualization and agility. Through the cloud, organizations can acquire IT services without additional intervention or human interaction. According to IDC [4], spending on public IT cloud services alone was estimated a $47.4 billion industry in 2013 and is expected to more than double by 2017. Ultimately, cloud computing enables more efficient BI tasks.

M. Indulska and S. Purao (Eds.): ER Workshops 2014, LNCS 8823, pp. 75–84, 2014.

BI and analytics raise many issues for design-science researchers. This paper focuses more specifically on the deployment of BI on the cloud and big data. The research question addressed in the paper is: what are the research topics on which design-science researchers can contribute regarding BI in the cloud and big data?

Indeed, it is useful for design-science researchers to focus on the definition of new processes and implement new models for BI to take advantage of cloud computing and big data. Given that design-science research leads to the production of artifacts, several researchers made an inventory of potential artifacts. Building on these previous papers, we propose a typology of artifacts that allows us to structure our literature review on BI in the cloud and detect open research questions. Even though business intelligence and big data in the cloud raise many new research issues for the information systems community at large (quantitative and qualitative research, IS economics, design-science research), the present papers focuses specifically on the possible contributions of design-science research.

The remainder of the paper is structured as follows: in the second section, we present our typology of design-science research artifacts. In the third section, we synthesize the current state of research for business intelligence and big data in the cloud. The fourth section describes open research issues and opportunities for design science. The last section introduces a discussion before concluding the paper.

## 2 Typology of Design-Science Research Artifacts

March and Smith [5] distinguish among four types of artifacts constituting the outputs of design-science research:

- *Construct*: a conceptualization used to describe problems within the domain and to specify their solutions.
- *Model*: a set of propositions or statements expressing relationships among constructs.
- *Method*: a set of steps (an algorithm or guideline) used to perform a task.
- *Instantiation*: the realization of an artifact in its environment.

This typology of artifacts is widely used, including in the seminal paper by Hevner et al. [6]. However, the typology is sometimes difficult to operationalize, due to the relative fuzziness of the concepts of construct, model, method, and instantiation. Therefore, it is useful to specialize the typology, specifying and defining subcategories for the four categories of artifacts.

Offermann et al. [8] specialize the typology of artifacts, proposing more specific categories. This work provides a useful basis for a classification of the various types of artifacts. However, we claim that this paper lacks some important subcategories of artifacts, some types include a large number of different concepts, and the proposed definitions may lead to confusion. Consequently, we propose our typology of the different types of artifacts (subcategories of the concepts of construct, model, method, and instantiation). This typology is shown in Table 1. For each subcategory, we propose a definition. The references in Table 1 indicate the papers from which the definitions were taken or adapted. Our typology, with precise subcategories and a definition for each subcategory, helps in the identification and characterization of design-science research artifacts. With our typology, design-science researchers can clarify the object of research and reflect on research methods appropriate to this object.

**Table 1.** Typology of artifacts

| | |
|---|---|
| **Construct** | |
| Language | *A set of concepts, or more generally symbols, rules for combining them (syntax), and rules for interpreting combinations of symbols (semantics) [7].* |
| Metamodel | *A set of concepts represented in graphical notation, with rules for combining the concepts.* |
| Concept | *A new concept added to an extant language or metamodel.* |
| **Model** | |
| System design | *A structure or behavior-related description of a system, commonly using some graphical notation and possibly text [8].* |
| Ontology | *An explicit formal specification of a shared conceptualization [9].* |
| Taxonomy | *A classification of objects in a domain of interest, based on common characteristics [10].* |
| Framework | *A logical structure for organizing complex information [11].* |
| Architecture | *A blueprint representing the fundamental organization of a system embodied in its components, their relationships to each other, and to the environment [12, 13].* |
| Requirement | *A condition or capability that must be met or possessed by a system [12].* |
| **Method** | |
| Methodology | *A predefined set of steps and guidelines, with associated techniques and tools. It is aimed at, or used by, individuals who work in a discipline [12] [14].* |
| Guideline | *A suggestion regarding behaviour in a particular situation [8]. Examples: design principles (broad guidelines), heuristics, rules (detailed guidelines) [15].* |
| Algorithm | *An executable sequence of operations for performing a specific task [8] [12].* |
| Method fragment | *A method component that can be treated as a separate unit and reused in different contexts [16]. Example: design patterns.* |
| Metric | *A function that assigns a number or symbol to an entity in order to characterize an attribute or a group of attributes. The value of the metric is called a measure [17].* |
| **Instantiation** | |
| Implemented system | *An implemented software or hardware system. Example: a prototype or finalized tool.* |
| Example | *Any other concrete materialization of an abstract artifact (construct, model, or method). Examples: the application of a query language to an illustrative scenario, the illustration of a design-theory framework with concrete examples of design theories, the application of a project methodology to a real project.* |

In the next sections, we use our typology of design-science research artifacts to synthetize research on BI and big data in the cloud, and identify research opportunities.

## 3 Current State of Research for Business Intelligence and Big Data in the Cloud

In this section, we summarize the state of the art of BI and big data on the cloud. Then, we analyze the papers through our typology of artifacts. We address this analysis in three themes:

- Data management: includes all artifacts related to representation and manipulation of data in the cloud.
- Service management: describes the potential of cloud services and artifacts relating thereto.
- Security management: describes the issues related to security, privacy, trust, and availability in the cloud.

We propose this categorization because at the heart of BI in the cloud, there is the issue of managing massive amounts of data (big data) and all other services provided by the cloud. Furthermore, security is a major challenge facing the cloud.

### 3.1 Data Management

Data management has a prominent place in BI. First, it will be necessary to develop new database management systems (DBMS) specifically architectured for BI in the cloud [18]. In addition, big data require developing models and tools capable of analyzing these masses of heterogeneous data that accumulate at high speed. On the one hand, traditional data warehouses can migrate to the cloud warehouse where they will integrate all data structures: documents, spreadsheets, e-mails, images, text and social media content. It is also necessary to think about the integration of big data [19]. Transactional data keep an important role in BI. Big data must be added to the internal data organization for best results of analysis [20]. Also, internal data organizations are structured whereas big data are generally not structured. Hence, the appropriate techniques must be defined.

Several researchers have striven to develop tools from MapReduce and its primitives as Hadoop. Thus, Herodotou et al. [21] developed an *architecture* and a tool called Starfish. Starfish fills a vacuum by allowing different users and Hadoop applications to automatically obtain good performance throughout the life cycle of data in analysis, without the need to understand and manipulate the numerous nodes available. Pedersen et al. [22] introduced the new *concept* of cloud warehouse. In addition, they defined a query *language* called SQLxm. Abadi [18] provided properties that users would appreciate finding in analysis tools. D'Orazio and Bimonte [23] proposed a data *architecture* in the cloud to optimize storage costs and an *algorithm* to convert these structures into Pig data. Chaudhuri et al. [24] presented a BI *architecture* including analysis tools for big data. Analysis tools include many *algorithms* for demand and implementation services as well as resources security. *Methodologies* to parameterize these tools were also implemented [25].

### 3.2 Service Management

There are six scenarii that illustrate the different service organizations existing in the cloud [26]:

- *Add-on services scenario:* Some components (e.g. components for web search) are selected from the cloud to BI infrastructure.
- *Tool replacement scenario*: The cloud makes available a complete tool, for instance a data mart or OLAP tool. This is SaaS (Software as a Service).
- *Solution provision scenario:* The cloud supports a software and hardware remote solution.
- *Business network scenario*: A solution provider acts within a corporate network. This can be a B2B market or supply chain, for example. The cloud aspect resides in the abstraction of the physical infrastructure that has become virtual.
- *Best-of-breed scenario:* The replacement of the tool is pushed to a higher level to the point where all components of the BI infrastructure are provided by an external supplier.
- *BI mashup scenario*: The BI solution is freely composed from a global market space over the internet.

Some researchers have implemented artifacts related to cloud services. Thus, Fernandez et al. [27] proposed an *architecture* that includes all services that organizations can receive in the cloud to perform their BI tasks. Hoberg et al. [28] established a *framework* according to the four dimensions following: cloud computing characteristics, adoption determinants, governance mechanisms, and business impact. Demirkan and Delen [29] implemented *requirements* and a conceptual *architecture* of service-oriented DSS. Baars and Kemper [26] presented a *framework* of BI in the cloud that can help with identification, combination and finally evaluation of potential BI services.

### 3.3 Security Management

Resource sharing in the cloud requires measuring more strongly the security level, since security is often considered as the main obstacle to the adoption of cloud computing services. Cloud computing must address the challenges of security, privacy and trust. Data have their physical existence in a given country and are governed by local regulation [29]. These regulations differ from one country to another and may be to the benefit or detriment of cloud customers. In addition, data is managed by an unknown host and customers do not control the use of their data in the cloud. Hence, design-science researchers must define the appropriate techniques.

Thus, Abadi [18] reminded the general principles of security. Many other researchers raise the issue of the cloud challenges. They call for encryption *algorithms* and focus on strengthening security policies for individual users and cloud providers [18] [29] [30].

### 3.4 Synthesis

Table 2 below synthetizes the state of the art on BI and big data on the cloud, by research theme and type of artifact.

**Table 2.** Overview of artifacts by subject

| | | Data management [18][21][22][23] [24][25] | Service management [25][26][27][28][29] | Security man-agement [18][25][29][30] |
|---|---|---|---|---|
| Construct | Language | | | |
| | Metamodel | | | |
| | Concept | | | |
| Model | System design | | | |
| | Ontology | | | |
| | Taxonomy | | | |
| | Framework | | | |
| | Architecture | | | |
| | Requirement | | | |
| Method | Methodology | | | |
| | Guideline | | | |
| | Algorithm | | | |
| | Method fragment | | | |
| | Metric | | | |
| Instantiation | Implemented system | | | |
| | Example | | | |

not covered partly covered covered

# 4 Business Intelligence in the Cloud: Open Issues and Opportunities for Design-Science Research

Table 2 shows that there are many opportunities for design-science research on the topic of BI and big data in the cloud. The cells of Table 2 that are white or grey indicate research opportunities. In Table 3, we present these opportunities. We distinguish between research opportunities identified in the literature and research opportunities identified by us, based on our typology of design-science research artifacts.

## 4.1 Data Management

In terms of data management, design-science researchers may extend the current research by producing some artifacts, for example:

- A multidimensional *metamodel* helping to design the cloud warehouse could be developed to enable the instantiation of classical multidimensional models in the specific context of the cloud or to bring new constructs enriching these models.
- Data from various sources and of different formats should all be finally consolidated. It is therefore necessary to develop conceptual and logical data models (*system design*).
- Clearer *guidelines* on how to use cloud data will help cloud customers to use the cloud data for the specific purposes of analysis and according to their needs.

**Table 3.** Overview of opportunities for design-science research

| | | Data management | Service management | Security management | Capacity management |
|---|---|---|---|---|---|
| Construct | Language | | | ▒ | |
| | Metamodel | ■ | | ■ | |
| | Concept | | | ■ | |
| Model | System design | ▒ | ■ | | |
| | Ontology | ■ | ■ | ■ | |
| | Taxonomy | | ■ | | |
| | Framework | | | | |
| | Architecture | | | ■ | ■ |
| | Requirement | | | | ■ |
| Method | Methodology | | | | |
| | Guideline | ■ | ■ | ■ | |
| | Algorithm | | | ▒ | |
| | Method fragment | | | | |
| | Metric | | | | |
| Instantiation | Implemented system | ▒ | | | |
| | Example | | | | |

■ our proposal ▒ proposal from the literature

- *Ontologies* could facilitate the integration of big data. Indeed, big data must be added to transactional data for best results of the analysis. Thus, providers and customers must have the same language. The ontology also facilitates the automation of big data integration.

## 4.2 Service Management

In the field of service-based computing, the researchers could focus greater attention on the following topics:

- The definition of a business process model (*system design*) could contribute to a better understanding of service requests initiated by users.
- An *ontology* of services needed by organizations will help users to make the choice of services they need from scenarios services offered by the cloud and to share a common language with the cloud provider.
- In order to regulate the demand for services, the implementation of *guidelines* is important since the customer gets services without human interaction with the provider. The *guidelines* will allow users to properly configure the tools at their disposal to access cloud resources.
- The *taxonomy* of cloud services allows users to save time by requesting a service level higher in the hierarchy that covers all their needs rather than choosing several low-level services.

### 4.3 Security Management

The security domain is not yet sufficiently explored. The cloud raises the challenge of security even higher. Hence, design-science researchers may extend research to develop the following artifacts:

- Encryption *algorithms* for ensuring data security in the cloud.
- Scrambling techniques (*algorithms*) allowing organizations to make available rich amounts of data without risk of disclosure since data are managed by an unknown host and customers do not control the use of their data in the cloud. Thus, these techniques prevent the violation of data privacy in the cloud.

### 4.4 Capacity Management

Cloud customers must have software and hardware capacity. Even if they are exempt from the details of the underlying infrastructure in the cloud services, they need a high amount of resources in terms of hardware, software and Internet connection. To the best of our knowledge, this topic is not covered by current or past research. Thus, we argue that we need the following artifacts:

- A hardware *architecture* of the cloud is necessary for customers.
- A model of *requirement* to help the cloud customer determine the hardware and software capabilities required given his/her needs.

## 5 Discussion and Conclusion

BI and big data in the cloud is a recent topic. This topic requires investigation by IS researchers using various methodologies, including quantitative, qualitative and design-science research. This paper has focused on opportunities for design-science research. We have proposed a typology of design-science research artifacts, and used this typology to identify research gaps and opportunities. More specifically, several artifacts must be developed to improve methods, models and tools dedicated to BI in the cloud and allowing users to analyze big data preserving information security. Efforts need to be conducted on logical and physical architecture design models and tools to set analysis in the cloud.

In this article, we have highlighted the services that cloud computing offers to BI to further improve its performance. With the cloud, organizations can gain large amounts of heterogeneous data for analysis (big data). This concept of big data highlights not only the volume of data but also their variety and processing speed.

As BI is growing, traditional models, processes and techniques must be rethought. This situation creates research opportunities for design-science researchers. Based on our typology of artifacts, we have elicited the gaps to be filled by further research. Indeed, our typology allowed us to identify artifacts already implemented in the domain of BI and big data in the cloud, and those to which design-science researchers should dedicate their effort.

## References


1. IDC: Worldwide Business Analytics Software 2013-2017 Forecast and, Vendor Shares (2012), http://idcdocserv.com/241689e_sas
2. Cuzzocrea, A., Song, I.-Y., Davis, K.C.: Analytics over Large-Scale Multidimensional Data: the Big Data Revolution! In: Proceedings of DOLAP 2011, pp. 101–104. ACM Press (2011)
3. Pring, B., Brown, R.H., Leong, L., Biscotti, F., Couture, A.W., Lheureux, B.J., Liu, V.K.: Forecast: Public Cloud Services, Worldwide and Regions, Industry Sectors. 2009-2014. Gartner Report (2010)
4. IDC: IDC Cloud, http://www.idc.com/prodserv/FourPillars/mobility/index.jsp
5. March, S., Smith, G.: Design and Natural Science Research on Information Technology. Decision Support Systems 15(4), 251–266 (1995)
6. Hevner, A., March, S., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28(1), 75–105 (2004)
7. Edwards, S., Lavagno, L., Lee, E.A., Sangiovanni-Vincentelli, A.: Design of Embedded Systems: Formal Models, Validation, and Synthesis. Proceedings of the IEEE 85(3), 366–390 (1997)
8. Offermann, P., Blom, S., Schönherr, M., Bub, U.: Artifact Types in Information Systems Design Science-a Literature Review. In: Winter, R., Zhao, J.L., Aier, S. (eds.) DESRIST 2010. LNCS, vol. 6105, pp. 77–92. Springer, Heidelberg (2010)
9. Gruber, T.R.: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition 5(2), 199–220 (1993)
10. Nickerson, R.C., Varshney, U., Muntermann, J.: A Method for Taxonomy Development and its Application in Information Systems. European Journal of Information Systems 22(3), 336–359 (2013)
11. CIO Council: Federal Enterprise Architecture Framework, version 1.1., Chief Information Officers Council, Washington D.C., USA (1999)
12. ISO/IEC, IEEE: Systems and Software Engineering – Vocabulary, standard ISO/IEC/IEEE 24765:2010(E) (2010)
13. Jarke, M., Loucopoulos, P., Lyytinen, K., Mylopoulos, J., Robinson, W.: The Brave New World of Design Requirements. Information Systems 36(7), 992–1008 (2011)
14. Nunamaker Jr., J.F., Briggs, R.O., De Vreede, G.-J., Sprague Jr., R.H.: Special Issue: Enhancing Organizations' Intellectual Bandwidth: The Quest for Fast and Effective Value Creation. Journal of Management Information Systems 17(3), 3–8 (2000)
15. Hanseth, O., Lyytinen, K.: Design Theory for Dynamic Complexity in Information Infrastructures: the Case of Building Internet. Journal of Information Technology 25(1), 1–19 (2010)
16. Kornyshova, E., Deneckère, R., Salinesi, C.: Method Chunks Selection by Multicriteria Techniques: an Extension of the Assembly Based Approach. In: Ralyté, J., Brinkkemper, S., Henderson-Sellers, B. (eds.) Situational Method Engineering: Fundamentals and Experiences. IFIP, vol. 244, pp. 64–78. Springer, Heidelberg (2007)
17. Purao, S., Vaishnavi, V.: Product Metrics for Object-Oriented Systems. ACM Computing Surveys 35(2), 191–221 (2003)
18. Abadi, D.J.: Data Management in the Cloud: Limitations and Opportunities. IEEE Data Engineering Bulletin 32(1), 3–12 (2009)
19. Bizer, C., Boncz, P., Brodie, M.L., Erling, O.: The Meaningful Use of Big Data: Four Perspectives-Four Challenges. ACM SIGMOD Record 40(4), 56–60 (2011)

20. Chen, H., Chiang, R.H., Storey, V.C.: Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly 36(4), 1165–1188 (2012)
21. Herodotou, H., Lim, H., Luo, G., Borisov, N., Dong, L., Cetin, F.B., Babu, S.: Starfish: A Self-Tuning System for Big Data Analytics. In: Proceedings of CIDR, pp. 261–272 (2011)
22. Pedersen, T.B., Pedersen, D., Riis, K.: On-Demand Multidimensional Data Integration: Toward a Semantic Foundation for Cloud Intelligence. The Journal of Super Computing 65(1), 217–257 (2013)
23. d'Orazio, L., Bimonte, S.: Multidimensional Arrays for Warehousing Data on Clouds. In: Hameurlain, A., Morvan, F., Tjoa, A.M. (eds.) Globe 2010. LNCS, vol. 6265, pp. 26–37. Springer, Heidelberg (2010)
24. Chaudhuri, S., Dayal, U., Narasayya, V.: An overview of Business Intelligence Technology. Communications of the ACM 54(8), 88–98 (2011)
25. Dean, J., Ghemawat, S.: MapReduce: Simplified Data Processing on Large Clusters. Communications of the ACM 51(1), 107–113 (2008)
26. Baars, H., Kemper, H.G.: Business Intelligence in the Cloud? In: Proceedings of PACIS 2010. Association for Information Systems, Paper 145 (2010)
27. Fernández, A., del Río, S., Herrera, F., Benítez, J.M.: An Overview on the Structure and Applications for Business Intelligence and Data Mining in Cloud Computing. In: Uden, L., Herrera, F., Bajo, J., Corchado, J.M. (eds.) 7th International Conference on KMO. AISC, vol. 172, pp. 559–570. Springer, Heidelberg (2013)
28. Hoberg, P., Wollersheim, J., Krcmar, H.: The Business Perspective on Cloud Computing - A Literature Review of Research on Cloud Computing. In: Proceedings of AMCIS 2012, Association for Information Systems, Paper 5 (2012)
29. Demirkan, H., Delen, D.: Leveraging the Capabilities of Service-Oriented Decision Support Systems: Putting Analytics and Big Data in Cloud. Decision Support Systems 55(1), 412–421 (2013)
30. Chaudhuri, S.: What Next? A Half-Dozen Data Management Research Goals for Big Data and the Cloud. In: Proceedings of PODS 2012, pp. 1–4. ACM Press, New York (2012)