ScienFIST.org

INTERNATIONAL JOURNAL OF INFORMATION TECHNOLOGIES, ENGINEERING AND MANAGEMENT SCIENCE

Reframing Architecture Frameworks using Cloud Computing and Open (Big) Data

Martin LNĚNIČKA
Department of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Czech Republic
martin.lnenicka@gmail.com

Abstract

Businesses as well as governments and scientists look at the growing amounts of data as a serious problem for traditional technological and organizational frameworks. They try to find a new solution to bridge the gap between heterogeneous enterprise architectures and frameworks on the one hand, and the different data storage and analytics on the other hand. This paper develops the issue of cloud computing and open (big) data and their impact on the enterprise architecture frameworks. It offers a set of steps to reframe the existing frameworks with these emerging approaches and technologies.

Keywords: Cloud computing, big data, open data, enterprise architecture, architecture framework

Introduction

The continuous emergence of new Information and Communication Technologies (ICT) and digital channels such as cloud computing, open and big data, mobile broadband, social media is yet another phase of the ongoing revolution in ICT [2], [19], [21].

The technical specifications of any technology-based product or service has to be based on a full understanding of all of the business processes that are required to ensure that a service can be delivered. Thus, the overall enterprise architecture is critical to be compliant with the constant changing trends in ICT. Well-structured and effective business processes together with the ICT and enterprise architecture are the essential building components of a successful and profitable organization.

Materials and methods

This paper identifies and defines the new trends in ICT and offers a solution to incorporate them into the selected enterprise architecture frameworks.

These trends affect the enterprise architecture and related frameworks, which need to be updated and reframed based on the main principles of them. In this paper, cloud computing and open and big are chosen as the most important trends and the existing architecture frameworks are compared with respect to them. Finally, a set of steps is proposed and described to help reframe and improve the existing enterprise architecture frameworks and considering the practical benefits for the enterprise architecture.

Method of analysis, decomposition, aggregation and benefits comparison is used in this paper.

Literature review and theoretical background

The importance of cloud computing

As technology focus evolved from mainframes through client-server to the internet era and now to cloud computing, each enterprise has to continuously evaluate emerging architectures and frameworks, and plan the evolution of its ICT environment [16].

Cloud computing changes the way users access the ICT systems, which means resources (computers, infrastructures, data storage, and application services) and the way of managing and delivering computing technologies, services and solutions. It offers a new model of delivering computing resources in a pay-per-use approach, which allows organizations to use their applications without any software installation and access their personal files at any computer with internet access. It is a way to significantly lower cost of infrastructure while increasing scale of operations.

Cloud computing can be distinguished into four types on the basis of location where the cloud is hosted – public cloud (the customer has no visibility over the location of the infrastructure, the computing infrastructure is shared between all organizations); private cloud (ICT architecture is dedicated to the customer and is not shared with others, is more expensive, considered and secure than public cloud); hybrid cloud (organizations host some critical, secure applications in private clouds, the not so critical applications are hosted in the public cloud) and community cloud (the cloud infrastructure is shared between the organizations of the same community). In general, cloud providers fall into three categories [1], [16]:

- Infrastructure as a Service (IaaS): offering web-based access to storage and computing power. The user does not need to manage or control the underlying cloud infrastructure but has control over the operating systems, storage, and deployed applications.
- Platform as a Service (PaaS): provides a new platform giving developers the tools to build and host web applications.
- Software as a Service (SaaS): is a software licensing model in which applications are accessible from client devices through a thin client interface such as a web browser.

The emergence of cloud platforms has given rise to new paradigms for dealing with distributed data in the cloud, parallel computing using very large computing clusters as well as rapid-scale application development tools for various specialized domains. The MapReduce programming paradigm then makes large-scale analytics tasks easy to define. MapReduce implementations such as Apache Hadoop, Cloudera or Hortonworks allow massive computing clusters to be used while tolerating faults that are inevitable at such scales [5], [8], [16].

The importance of open and big data

Nowadays, organizations are acquiring more and more data from both free and subscription public sources and all of them vary in structure, quality and volume [10]. The benefits from large amounts of data, which constantly increases and contains great hidden potential, can be maximized, e.g., by using the knowledge base [17].

Although there are many different sources of data, government data are important because of their scale, breadth, and status as the canonical source of information on a wide range of various subjects [20]. Consequently, they are the main source of open data available on the Web. As publicly available data can often be generated and provided in huge amounts and through multiple sources, businesses as well as public sector deal with huge quantities and varieties of data on one hand and faster expectations for analysis on the other [5], [9]. The emergence of these big data is mostly enabled by the recent hardware and software advances and it is complemented by the shift towards openness of the public sector [2].

Because of large amounts of data produced by the public sector, the open data model has evolved into the open (big) data model [19]. This intersection is as much about integrating multiple data sources, e.g. on the local, national or international level of the public sector. Marton, Avital and Jensen [13] stated that the fundamental concepts of open and big data are technical in nature as they were developed in the fields of computer science and engineering. Open data and big data are both gathered for a purpose and then normally re-purposed [6], [13]. While big data are characterized as large in volume, gathered at high-speed and may be also unstructured and come from many sources, open data are about standards on how to make data machine-readable, and hence linkable [13]. The combination of open and big data that refer to different domains and are published by different public and international organizations with other data (e.g. business's own data) could enable creating and evaluating models that were previously hard to develop [11]. The opportunity to collect and integrate these data in the enterprise data architecture offers amazing potential for organizations to increase their competitiveness.

The resulting economic gains of open (big) data can be put into three categories: resource efficiency improvements, product and process improvements and management improvements through evidence based, data-driven decision making [2]. These data also provide a variety of benefits to society such as to increase government transparency and accountability, stimulate innovation, participation, collaboration or improve business processes [6], [11]. More precisely, Kucera and Chlapek [12] presented a set of benefits that can be achieved by publishing these data and a set of risks that should be assessed when a dataset is considered for opening up. Also Cowan, Alencar and Mcgarry [6] used several practical examples in an attempt to illustrate many of the related issues and allied opportunities of open (big) data.

Demchenko, de Laat and Membrey [7] defined the basic information/semantic models, architecture components and operational models that together comprise a so called Big Data Ecosystem.

Enterprise architecture frameworks and the basic requirements for them

The goal of enterprise architecture is to create a unified ICT environment (standardized hardware and software systems) across the organization or all of the organization's business units, with tight symbiotic links to the business side of the organization and its strategy. More specifically, the goals are to promote alignment, standardization, reuse of ICT assets and the sharing of methods for software development across the organization [14]. Therefore, the function of the enterprise architecture within ICT has evolved to manage the complexities of an ever-changing technical environment [16].

Taking into account an enterprise architecture approach ensures that all technology decisions are driven from an agreed upon business context [10].

There are many different suggested frameworks to develop an enterprise architecture. However, most frameworks contain four basic domains [14]:

- business architecture: documentation that outlines the organization's most important business processes,
- information architecture: identifies where important blocks of information, such as a customer record, are kept and how one typically accesses them,
- application system architecture: a map of the relationships of software applications to one another,
- the infrastructure technology architecture: a blueprint for the gamut of hardware, storage systems, and networks.

The most widely used frameworks are Zachman Enterprise Architecture Framework, The Open Group Architecture Framework, Extended Enterprise Architecture Framework, Enterprise Architecture Planning and Federal Enterprise Architecture Framework [14].

More enterprise architecture frameworks can be found at *http://www.iso-architecture.org/ieee-1471/afs/frameworks-table.html* or *http://www.enterprise-architecture.info/*.

Cloud computing and open (big) data in the enterprise architecture framework

Although cloud computing has been broadly accepted by many organizations, research on open (big) data in the cloud remains in its early stages [15].

The paradigm has changed from traditional host or service based models to data centric architecture and operational models [7]. These approaches data structure and analytics differently than traditional information architectures. At first glance, the five "V"s define attributes and form the main challenges of open (big) data in the context of the enterprise architecture framework. These are: Volume (data coming in TB transactions, or in tables or in files), Velocity (data coming at real-time or near time or coming in streams), Variety (data coming whether in structured semi or unstructured way), Value (the insights produced during data analysis) and Veracity (refers to the trust into the data, which may be impaired by the data being uncertain or imprecise) [4], [5], [7].

Without a doubt, cloud computing is playing an essential role for many use cases as a data source, providing real-time streams, analytical services or as a device transaction hub [10]. Many SaaS platforms have been developed specifically for processing open (big) data such as sequence analysis, alignment and mapping [8]. Cloud computing technologies simplify the building of related infrastructure and provision it on-demand [7].

The main difficulties related to open (big) data include capture, storage, search, sharing, analytics, and visualizing. It indicates that efficient platforms and technologies together with suitable methods are needed to analyze and process these data [18]. This part is closely related to cloud computing, which may solve the issue of storage, sharing and analytics.

Scalability is at the core of the expected new technologies to meet the challenges coming along with open (big) data. The maturing cloud computing technologies deliver the most promising platforms to meet the requirements on infrastructure for open (big) data, e.g. cost efficiency, elasticity, smooth upgrading and downgrading and parallelism capacities [4], [15].

Consequently, the scalability is the main factor in the new enterprise architect framework. It affects the infrastructure as IaaS, storage capacities for open (big) data and the necessity of the platforms as PaaS.

The other challenges and opportunities are [5], [7], [8], [15]:

- availability of the data stored in the cloud,
- elasticity, because resources are provisioned dynamically according to workload change,
- data integrity, because data can be modified only by authorized parties or the data owner to prevent misuse,
- data representation (i.e. heterogeneity in structure, type, organization, granularity and accessibility),
- data transformation and related tools,
- redundancy reduction and data compression,
- data life cycle management, which means easier maintenance and upgrade,
- energy management and expendability.

In the context of the architecture's current state, each element, step, task, and/or phase of the related framework should be assessed, compared, prioritized, modified and integrated according to the principles of cloud computing and open (big) data.

However, the enterprise architecture may often suffer from a lack of semantics which is reflected in issues with communication between people, between systems or even directly between people and systems [3]. Therefore, better knowledge management within enterprise architecture could help resolve the problem mentioned above [3].

Results and discussion

The basic components identified in the previous sections, which influence and direct the development of related enterprise architecture frameworks, are: infrastructure and resources, data analytics, data structures and models, data life cycle management and data security.

Based on the literature review and comparison of the selected enterprise architecture frameworks in the context of cloud computing and open (big) data, these findings are found and they should be used in the new enterprise architecture framework:

- assess the existing hardware and software components and resources (e.g. cost-benefit analysis) in the context of IaaS and PaaS,
- investigate their compatibility with the new technologies and then integrate them,
- restructure the data network together with address structures and formats for running them on different operating systems and with the use of different formatting products,
- redefine the data transaction flow based on different technology characteristics of cloud computing,
- redefine the control structures based on the different technological constraints,
- integrate business data with freely available open data to create new analytical models and gain a competitive advantage,
- use different data storage technologies such as cloud storages to backup business data in the public or private cloud,
- use different database management systems products such as various NoSQL databases to support real-time operations,
- cloud computing and open (big) data also require the change of security principles and practices due to external services providers.

The limitation of the new framework may be its ability to explain dynamic and interrelated processes in the context of cloud computing and open (big) data maturity. For this reason, the organizations should apply different maturity models to determine the current state of these technologies in the organization and identify specific capabilities that are lacking or lagging. Several maturity models are freely available for each topic, more can be found in [1], [5], [10] or [18]. In general, these models and their stages are defined through the perspectives such as functions characteristics, incorporating and integrating of data sources, architecture, infrastructure, organization and people, state of governance, etc.

Conclusion

New ways of collecting, storing and analyzing information allow organizations to gather data at low costs and with minimum effort. It resulted into the recent research topic focusing on the importance of cloud computing and open (big) data in the enterprise architecture frameworks. But, the physical separation of data centers, distinct security policies, ownership of data and data quality processes, in addition to the impact of each of the five "V"s requires architecture decisions.

Therefore, this paper reviewed these two topics in the context of enterprise architecture frameworks and tried to show that there is a need to reframe the existing frameworks. The main priorities should be focused on the infrastructure and data management. The choice of the concrete platform should be based on minimizing the total cost of ownership including transition costs to any new technology.

Future research will be focused on the concrete platforms such as Apache Hadoop, Apache Storm, Cloudera or MapR and their role in the architecture framework and its data life cycle management. These approaches are covered e.g. in Švalec et al. [17], where authors introduced some possible approaches towards filling the knowledge base from unstructured data sources using text and data mining methods and also processing of extensive sets of data with aid of decomposition and parallelism.

References

[1] N. Antonopoulos and L. Gillam, *Cloud Computing: Principles, Systems and Applications*. London: Springer, 2010.

[2] S. Buchholtz, M. Bukowski and A. Śniegocki, *Big and open data in Europe: A growth engine or a missed opportunity*. Varšava: demosEUROPA, 2014.

[3] B. Bučko and K. Zábovská, "Knowledge in Information Systems", *International Journal of Information Technologies, Engineering and Management Science*, vol. 1, iss. 1, pp. 4-6, 2015.

[4] D. Che, M. Safran and Z. Peng, From Big Data to Big Data Mining: Challenges, Issues, and Opportunities. In: *Database Systems for Advanced Applications*, Springer, Berlin, pp. 1-15, 2013.

[5] M. Chen, S. Mao and Y. Liu, “Big Data: A Survey“, *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171-209, 2014.

[6] D. Cowan, P. Alencara and F. Mcgarry, Perspectives on Open Data: Issues and Opportunties. In: *Proceedings of the 2014 IEEE International Conference on Software Science, Technology and Engineering*, IEEE, pp. 24-33, 2014.

[7] Y. Demchenko, C. de Laat and P. Membrey, Defining Architecture Components of the Big Data Ecosystem. In: *Proceedings of the 2014 International Conference on Collaboration Technologies and Systems (CTS)*, IEEE, pp.104-112, 2014.

[8] H. Elazhary, “Cloud Computing for Big Data“, *MAGNT Research Report*, vol. 2, iss. 4, pp. 1-9, 2014.

[9] C. P. Geiger and J. Von Lucke, “Open Government and (Linked) (Open) (Government) (Data) “, *eJournal of eDemocracy & Open Government*, vol. 4, no. 2, pp. 265-278, 2012.

[10] P. Heller and D. Piziak, “An Enterprise Architect’s Guide to Big Data – Reference Architecture Overview“, Oracle, 2015, [Online]. Available: http://www.oracle.com/technetwork/topics/entarch/articles/oea-big-data-guide-1522052.pdf. [Accessed 27 08 2015].

[11] E. Kalampokis, E. Tambouris and K. Tarabanis, Linked Open Government Data Analytics. In: *Proceedings of the 12th IFIP WG 8.5 International Conference: EGOV 2013*, Springer, Berlin, pp. 99-110, 2013.

[12] J. Kucera and D. Chlapek, “Benefits and Risks of Open Government Data“, *Journal of Systems Integration*, vol. 5, no. 1, pp. 30-41, 2014.

[13] A. Marton, M. Avital and T. Blegind Jensen, Reframing Open Big Data. In: *Proceedings of the 21st European Conference on Information Systems*, Utrecht University, Netherlands, pp. 1-12, 2013.

[14] D. Minoli, *Enterprise Architecture A to Z: Frameworks, Business Process Modeling, SOA, and Infrastructure Technology*. Boca Raton: CRC Press, 2008.

[15] A. Roshdi and M. Shamsi, “REVIEW: Big Data on Cloud Computing“, *International Journal of Enhanced Research in Science, Technology & Engineering*, vol. 4, iss. 7, pp. 252-259, 2015.

[16] G. Shroff, *Enterprise Cloud Computing: Technology, Architecture, Applications*. New York: Cambridge University Press, 2010.

[17] M. Švalec et al., “Preparation of Data for Knowledge Base“, *International Journal of Information Technology and Business Management*, vol. 27, no. 1, pp. 136-139, 2014.

[18] J. M. Tien, “Big Data: Unleashing Information“, *Journal of Systems Science and Systems Engineering*, vol. 22, no. 2, pp. 127-151, 2013.

[19] United Nations, *United Nations e-Government Survey 2012: E-Government for the People*. New York: UN Publishing Section, 2012.

[20] S. Van der Waal et al., Lifting Open Data Portals to the Data Web, In: *Linked Open Data – Creating Knowledge Out of Interlinked Data*. Springer International Publishing, pp. 175-195. 2014.

[21] World Economic Forum, *The Global Information Technology Report 2014: Rewards and Risks of Big Data*. Geneva: SROKundig, 2014.