Towards the Evaluation of a Big Data-as-a-Service Model: A Decision Theoretic Approach

Georgios Skourletopoulos, Constandinos X. Mavromoustakis
Mobile Systems (MoSys) Laboratory
Department of Computer Science
University of Nicosia
Nicosia, Cyprus
skourletopoulos.g@unic.ac.cy, mavromoustakis.c@unic.ac.cy

George Mastorakis, Evangelos Pallis
Department of Informatics Engineering
Technological Educational Institute of Crete
Heraklion, Crete, Greece
gmastorakis@staff.teicrete.gr, pallis@pasiphae.eu

Periklis Chatzimisios
CSSN Research Lab
Alexander Technological Educational Institute of Thessaloniki
Thessaloniki, Greece
peris@it.teithe.gr

Jordi Mongay Batalla
Warsaw University of Technology and Seidor S.A.
Nowowiejska Str. 15/19 and Eix Onze de Setembre, 41-43
Warsaw, Poland and Vic, Spain
jordim@interfree.it

***Abstract*—The rise of large data centers has created new business models, where businesses can lease storage and computing capacity and pay only for the storage they actually use, rather than making the large capital investments needed to construct and provision large-scale computer installations. In this context, investments in big-data computing are rapidly gaining ground, having extraordinary near-term and long-term benefits. The mobile cloud can be considered as a marketplace, where the storage and computing capabilities of the mobile cloud-based system architectures can be leased off. However, cloud storage is not less expensive, only that it incurs operating rather than capital expenses. This paper elaborates on a novel cost analysis model, adopting a non-linear and asymmetric approach. The proposed modelling aims to evaluate the adoption of a big data-as-a-service business model against the traditional high-performance data warehouse appliances that exist in the market in order to inform effective and strategic decision making. The lease of cloud storage is investigated, when developing the mathematical formulas, and the research approach is examined with respect to the cost that derives from the unused storage. Possible upgradation of the storage and the risk of entering into new and accumulated costs in the future are also considered in this study. A quantification tool has been also developed as a proof of concept (PoC), implementing the proposed quantitative model and intending to shed light on the adoption of big data-as-a-service business models.**

***Keywords*—big data-as-a-service; non-linear modelling; cost analysis; business model evaluation; mobile cloud computing architectures; decision theory**

I. Introduction

Advances in digital sensors, communications, computation and storage have created huge collections of data, capturing information of value to business and science through real time analytics and creating an era where data storage and computing become utilities that are ubiquitously available. In this direction, many organizations choose the most effective data sets to meet their goals and enhance decision making and problem solving by adopting mobile cloud migration approaches. Mobile and cloud computing [1] enable the development of a huge amount of applications, and therefore data, bringing innovation in data architecture. The mobile cloud computing (MCC) technology aims to utilize cloud computing techniques for storage and processing of data on mobile devices. As a result, big data needs more computing power and storage provided by cloud computing platforms [2]. In this context, servicelization is the method of offering social networking services, big data analytics and mobile internet services [3]. Everything-as-a-service is creating a big services era due to the foundational architecture (i.e., Service-Oriented Architecture) of services computing. In addition, cloud providers provide network-accessible storage priced by the gigabyte-month and computing cycles priced by the CPU-hour [4]. Although the cloud-based approach can significantly augment computation capability of mobile device users, the task of developing a reliable mobile cloud computing system is still a very challenging field. Unlike web-based big data, location data is an essential component of mobile big data, which are harnessed in order to optimize and personalize mobile services.

In this context, this research work is making progress beyond the current state-of-the-art, by contributing to a novel cost analysis model towards the evaluation of a big data-as-a-service business model against the traditional data warehouse appliances (DWH) that exist in the market. The conceptual big data warehouse architecture is shown in Fig. 1 [5]. In the mobile cloud marketplaces [6], the storage capacity and computing capabilities of the mobile cloud-based service-oriented architectures can be leased off [7]. In this paper, we prove that the investment in data warehouse appliances in order to perform statistical analysis, instead of adopting a big data-as-a-service business model, will introduce accumulated costs in the long run, which can be hardly managed, having a significant impact

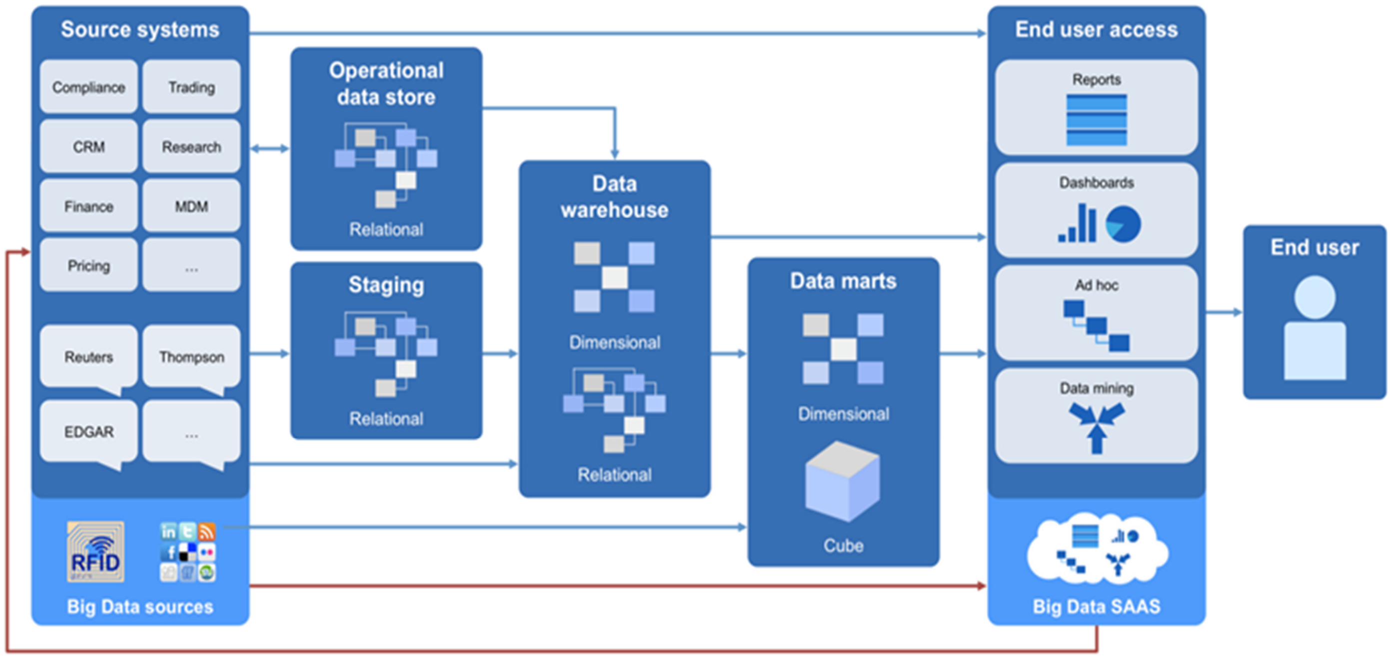

Fig. 1. A conceptual big data warehouse architecture.

on the return on investment (ROI). The model formulation is based on the lease of cloud storage capacity under the assumption that fluctuations in the demand for storage capacity occur due to the increasingly large amount of data, adopting a non-linear and asymmetric approach. Initially, the size of the additional costs is explicitly affected by the business model selection decision and the maximum cloud storage capacity, examining the need to require more capacity in the future. The evaluation results enable enterprises to make accurate predictions and examine the adoption of big data-as-a-service (BDaaS) business models to increase the ROI and gain competitive advantage by exploiting the full range of analytics capabilities provided by an outside provider.

Following this introductory section, the reminder of this paper is organized as follows: Section II presents related research efforts and the research gap that motivates the need for the evaluation of a big data-as-a-service business model. Section III proposes the novel quantitative model, while Section IV provides an evaluation analysis of two different case scenarios and Section V concludes this research paper.

II. Related Work And Research Motivation

The adoption of a big data-as-a-service business model enables the effective storage, management and assessment of huge data sets and data processing from an outside provider along with reduced costs by freeing up organizational resources. In this direction, limited research efforts are witnessed, focusing mainly on defining the big data concept along with issues and challenges [8]. The authors in [9] attempted to introduce a user experience-oriented big data-as-a-service architecture, aiming to provide visualization services from analyzing unstructured data. An overview of service-generated big data and big data-as-a-service is presented in [10] towards the proposal of an infrastructure to provide functionality for managing and analyzing different types of service-generated big data. A big data-as-a-service framework is employed to provide accessing service-generated big data and data analytics results to users in order to enhance efficiency and reduce cost. On the contrary, the authors in [11] describe the development of a cloud-supported big data mining platform, which provides rich data statistical and analytical functions. In this work, the platform's architecture is composed of the infrastructure, virtualization, data set processing and services layers, implementing the K-means algorithm. Furthermore, a big data analytics platform is proposed in [12], which manages big data and develops analytics algorithms and services through collaboration between data owners, data scientists and service developers on the web. A CCTV metadata analytics service is also implemented in the platform.

As a consequence of the big data and analytics growth [13]–[15], enterprises can now turn to big data-as-a-service solutions to bridge the storage and processing gap [16]. In this context, a big data provisioning service is presented in [17], aiming to incorporate hierarchical and peer-to-peer data distribution techniques to speed-up data loading into the virtual machines (VMs) used for data processing. The proposed service achieves to reduce time by at least 30 percent over current state of the art techniques, coupled with classic declarative machine configuration techniques, which simplify the deployment of big data in the cloud. On the other hand, the work in [18] elaborates on a big data-as-a-service solution based on Hadoop, which extracts data from the social networks and constructs a graph for further analysis, while the authors in [19] propose an admission control and resource scheduling algorithm, which achieves to satisfy the Quality of Service (QoS) requirements of requests according to the Service Level Agreements (SLAs) guarantees, save cost and boost profits for the Analytics-as-a-Service (AaaS) providers. The practice of Magnetic, Agile, Deep (MAD) data analysis was also presented in [20] as a method to replace traditional Enterprise Data Warehouses (EDW) and Business

Intelligence (BI) solutions, exploiting data parallel algorithms for sophisticated statistical techniques with a focus on density methods.

Although there are related work approaches in the literature regarding big data from different viewpoints, there is currently little work in the area of cost-benefit analysis to evaluate big data-as-a-service business models and, therefore, a research gap for classifying the big data-as-a-service concept is witnessed as various types of services compete with very different business models. In this research work, the proposed quantitative model contributes towards the evaluation of a big data-as-a-service business model, adopting a cost-driven and non-linear approach (i.e., fluctuations in the demand for storage capacity), against the traditional high-performance data warehouse appliances that exist in the market. The proposed quantification modelling approach is an extension of previous research works conducted by Skourletopoulos et al. in [21]–[25] from the big data-as-a-service perspective.

III. Model Formulation

The proposed quantitative cloud-inspired model aims to reveal the benefits of the adoption of a big data-as-a-service business model against the conventional high-performance data warehouse appliances that exist in the market. The lease of cloud storage is considered, when developing the mathematical formulas, and the model formulation is based on cost and benefits analyses, measuring the amount of profit not earned due to the underutilization of the storage capacity under the assumption that fluctuations in the demand for storage capacity occur (i.e., non-linear and asymmetric approach). The hypothesis is that the business model selection decision is made with respect to the predicted benefits over the period of *l*-years as the total cost for adopting a business model is taken into consideration. The cloud storage capacity to be leased off is evaluated with respect to the following assumptions and statements:

- The cloud storage is subscription-based and the billing vary over the period of *l*-years due to the fluctuations in the demand for storage capacity (i.e., gigabyte per month). A pricing overview is witnessed in [26].
- Fluctuations in the demand for storage capacity are predicted over the period of *l*-years, affecting the total cost in order to ensure the Quality of Service (QoS) and Quality of Experience (QoE).
- The scalability and elasticity provided by the cloud are taken into account when developing the model, as the cost variations, which result from the annual variation in the demand for storage capacity, are composed of data and document (unstructured) storage, maintenance services, network, on-demand I/O, operations (i.e., service requests), server and technical support costs.
- The total network cost consists of costs related to bandwidth usage, egress and data transfer between regional and multi-region locations. As the cloud-based, always-on mobile services are usually sensitive to network bandwidth and latency [27], the additional network cost is expected to satisfy the outbound network traffic demands in order to avoid delays. Furthermore, custom metadata headers are accounted for in the monthly storage and bandwidth usage.
- The additional on-demand I/O cost enables to increase the throughput [28] when the content retrieval from a bucket should be faster than the default.
- The additional server cost includes those costs that result from the additional CPU cores and the amount of memory required for processing.

In this context, the cost analysis (CA) modelling from the traditional data warehouse appliance (DWH) viewpoint would take the following form (the variable description is presented thoroughly in Table 1):

$$CA_i = 12 * \left(Cs_{/m} * S_{max}\right), i \geq 1 \text{ and } S_{curr} \leq S_{max} \quad (1)$$

where,

$$Cs_{/m} = Cs_{/m(max)} = C\alpha_{/m(max)} + C\gamma_{/m(max)} + C\eta_{/m(max)} + C\theta_{/m(max)} + C\kappa_{/m(max)} + C\lambda_{/m(max)} + C\mu_{/m(max)} + C\sigma_{/m(max)}$$

As the benefits of cloud computing and big data-as-a-service (i.e., scalability and elasticity) do not stand in data warehouse appliances, the adopted cost analysis approach does not consider the storage capacity currently used (S_{curr}) in order to adapt the usage cost. As a result, cost variations due to the fluctuations in the demand for storage capacity do not apply as long as $S_{curr} \leq S_{max}$. In conclusion, the true benefits and cost difference are always zero ($C_D = 0$) over the period of *l*-years. It is important to mention that in case of such an increase in the demand for storage capacity that $S_{curr} > S_{max}$, then incremental capacity should be added to the storage systems with overhead and downtime, resulting in accumulated costs, not to mention that there were charges for more storage capacity than the actual usage.

On the contrary, the cost analysis (CA) and benefits/cost difference (C_D) modelling from the big data-as-a-service (BDaaS) perspective takes the following form (the variable description is shown in detail in Table 1) during the first year (i.e., equations 2 and 4) and from the second year and onwards (i.e., equations 3 and 5):

$$CA_1 = 12 * \left(Cs_{/m} * S_{curr}\right) \quad (2)$$

$$CA_i = 12 * (\Delta_{i-2} * B_{i-2}), i \geq 2 \quad (3)$$

$$C_{D_1} = 12 * \left[Cs_{/m} * (S_{max} - S_{curr})\right] \quad (4)$$

$$C_{D_i} = 12 * [\Delta_{i-2} * (S_{max} - B_{i-2})], i \geq 2 \quad (5)$$

where,

$$Cs_{/m} = Cs_{/m(curr)} = C\alpha_{/m(curr)} + C\gamma_{/m(curr)} + C\eta_{/m(curr)} + C\theta_{/m(curr)} + C\kappa_{/m(curr)} + C\lambda_{/m(curr)} + C\mu_{/m(curr)} + C\sigma_{/m(curr)}$$

$$\Delta_0 = (1 + \delta_1\%) * Cs_{/m}$$

$$\Delta_i = (1 + \delta_{i+1}\%) * \Delta_{i-1}, i \geq 1$$

$$\delta_i\% = a_i\% + \gamma_i\% + \eta_i\% + \theta_i\% + \kappa_i\% + \lambda_i\% + \mu_i\% + \sigma_i\%, i \geq 1$$

$$B_0 = (1 + \beta_1\%) * S_{curr}$$

$$B_i = (1 + \beta_{i+1}\%) * B_{i-1}, i \geq 1$$

Two possible types of benefits/cost difference results are encountered, when leasing cloud storage:

- Positive calculations, which reveal the underutilization of the storage capacity and the probability to satisfy a possible future increase in the demand.
- Negative calculations, which indicate the immediate need for upgradation. This need stimulates additional costs; however, the total amount of accumulated cost in traditional data warehouse appliances is not comparable, as the earnings by adopting a big data-as-a-service business model can be reinvested on the additional storage required, maximizing the return on investment (ROI).

IV. Performance Evaluation Analysis And Numerical Results

This paper contributes to novel cost and benefits analysis models that support the evaluation of a big data-as-a-service business model against traditional data warehouse appliances under the assumption that fluctuations in the demand for cloud storage capacity occur. From the research viewpoint, a quantification perspective is adopted when leasing cloud storage, presenting a complex model. The models are characterized by extensibility as more parameters can be added, indicating how customizable the formulas are. The level of comprehension of the model formulation depends on the expertise of the user. Furthermore, a quantification tool has been developed as a proof of concept (PoC), which implements the proposed formulas, intending to compare the big data-as-a-service model against the traditional high-performance data warehouse appliances. From the technical point of view, the cloud-supported web application is targeted to be deployed in the Google Cloud Platform supported by the Google App Engine and it was implemented using the Java programming language.

Table 1. ABBREVIATIONS AND VARIABLE DESCRIPTION.

Symbol	**Variable Description**
CA	The cost analysis calculation results represented in monetary units.
C_D	The cost difference/benefits calculations represented in monetary units.
i	The index of the year.
$Cs_{/m}$	The initial monthly cost for leasing cloud storage represented in monetary units.
S_{max}	The maximum storage capacity.
S_{curr}	The storage currently used.
Δ_0	The cost formation for leasing cloud storage regarding the second year of the period of *l*-years, once the corresponding variation in the monthly cost is applied (represented in monetary units).
δ_1	The total variation regarding the cost for leasing cloud storage for the second year of the period of *l*-years, which is represented as percentage.
Δ_i	The cost formation for leasing cloud storage from the third year until the end of the period of *l*-years, once the corresponding variation in the monthly cost is applied (represented in monetary units).
δ_i	The total variation regarding the cost for leasing cloud storage from the third year until the end of the period of *l*-years, which is represented as percentage.
B_0	The storage used during the second year of the period of *l*-years, once the corresponding variation in the demand is applied.
β_1	The variation in the demand for storage capacity regarding the second year of the period of *l*-years represented as percentage.
B_i	The storage used from the third year until the end of the period of *l*-years, once the corresponding variation in the demand is applied.
β_i	The variation in the demand for storage capacity from the third year until the end of the period of *l*-years, which is represented as percentage.
C_α	The data storage cost.
C_γ	The document storage cost.
C_η	The maintenance services cost.
C_θ	The network cost.
C_κ	The on-demand I/O cost.
C_λ	The operations cost.
C_μ	The server cost.
C_σ	The technical support cost.
$\alpha_i\%$	The variation in the monthly data storage cost represented as percentage.
$\gamma_i\%$	The variation in the monthly document storage cost represented as percentage.
$\eta_i\%$	The variation in the monthly maintenance services cost represented as percentage.
$\theta_i\%$	The variation in the monthly network cost represented as percentage.
$\kappa_i\%$	The variation in the monthly on-demand I/O cost represented as percentage.
$\lambda_i\%$	The variation in the monthly operations cost represented as percentage.
$\mu_i\%$	The variation in the monthly server cost represented as percentage.
$\sigma_i\%$	The variation in the monthly technical support cost represented as percentage.

The tool emphasizes on shedding light on the adoption of big data-as-a-service business models, examining this challenging research problem from the storage capacity perspective. The calculations are significant to understand the progress of different case scenarios and prove that the investment in data warehouse appliances instead of adopting a big data-as-a-service business model, will introduce accumulated costs in the long run, which can be hardly managed, having a significant impact on the return on investment (ROI). An indicative and illustrative example of the evaluation that was performed, emphasizes on the need to consolidate data from different sources, as a consequence of the increase in semi-structured and unstructured data gathered from online interactions, and make those data available for analysis and management reporting (e.g., sales, marketing, operations and finance). In this direction, cost analysis and cost difference/benefits comparisons are performed between a banking data warehouse (BDW) model and a big data-as-a-service business model under the necessity of managing the data-intense workloads of advanced data analytics at the storage level (business intelligence and predictive analytics are also considered in this example, as part of the business continuity planning). A 5-year period of time ($l = 5$) is examined prior to adoption of either a traditional data warehouse or a big data-as-a-service business model, enabling to do a what-if analysis on two different case scenarios under the assumption that fluctuations in the demand for storage occur. The adopted variations in the demand for storage regarding the two case scenarios are presented in Table 2.

Table 2. VARIATIONS IN THE DEMAND FOR STORAGE FOR THE TWO CASE SCENARIOS.

Term	Case Scenario 1	Case Scenario 2
Year 1 to 2	$\beta_1\% = 5\%$	$\beta_1\% = 10\%$
Year 2 to 3	$\beta_2\% = 15\%$	$\beta_2\% = 22\%$
Year 3 to 4	$\beta_3\% = 20\%$	$\beta_3\% = 35\%$
Year 4 to 5	$\beta_4\% = 23\%$	$\beta_4\% = 40\%$

Towards a better understanding of the cost and benefits analysis quantification, the calculations for the first two years take the following form:

$$CA_1 = 12 * \left(Cs_{/m} * S_{curr}\right)$$

$$C_{D_1} = 12 * [Cs_{/m} * (S_{max} - S_{curr})]$$

$$CA_2 = 12 * [(1 + \delta_1\%) * Cs_{/m} * (1 + \beta_1\%) * S_{curr}]$$

$$C_{D_2} = 12 * \{(1 + \delta_1\%) * Cs_{/m} * [S_{max} - (1 + \beta_1\%) * S_{curr}]\}$$

Having explained the quantification rules, the values presented in Tables 3 and 4 are applied to the formulas (1), (2), (3), (4) and (5) accordingly. The choice of the specific values and case scenarios enables to obtain accurate and comparable results towards the evaluation of the two different business models. Table 4 includes the cost variations for leasing additional cloud storage for the two case scenarios, which are dependent on the variations in the demand for storage. The obtained evaluation results are shown thoroughly in Tables 5 and 6, while the cost analysis flows and the cost difference/ benefits comparisons over the 5-year period are witnessed in Fig. 2, 3, 4 and 5, respectively. In this direction, the first case scenario points out that adopting a big data-as-a-service model is more cost-effective than a traditional data warehouse one, as the cost analysis results for the big data-as-a-service model reveal always the least positive values over the 5-year period, despite the increase in the demand for storage. In addition, the benefits calculations are always positive in big data-as-a-service (the decline in the results is due to the increase in the demand for storage capacity), while the cost difference results are always zero in traditional data warehouse business models.

Table 3. VALUES TO BE APPLIED TO FORMULAS (1) TO (5).

Variable Description	Values
Maximum storage capacity (in terabytes)	$S_{max} = 4$
Storage currently used (in terabytes)	$S_{curr} = 2$
Initial monthly cost for leasing cloud storage (in USD)	$Cs_{/m} = 390$

Table 4. TOTAL COST VARIATIONS FOR LEASING ADDITIONAL CLOUD STORAGE FOR THE TWO CASE SCENARIOS.

Variable Description	Case Scenario 1	Case Scenario 2
Cost variation for leasing additional storage	$\delta_1\% = 2\%$ $\delta_2\% = 5\%$ $\delta_3\% = 18\%$ $\delta_4\% = 20\%$	$\delta_1\% = 5\%$ $\delta_2\% = 10\%$ $\delta_3\% = 25\%$ $\delta_4\% = 18\%$

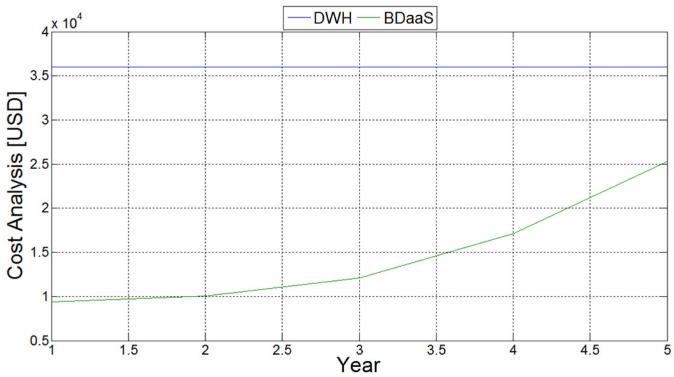

Fig. 2. Case Scenario 1: Cost analysis flow.

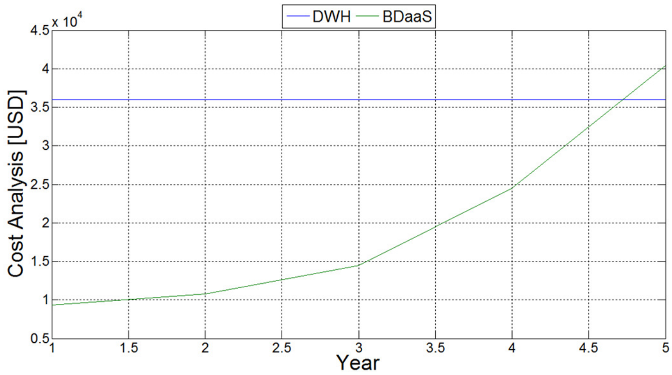

Fig. 3. Case Scenario 2: Cost analysis flow.

It is worthy to mention that the cost analysis and cost difference/benefits results regarding the traditional data warehouse appliances remain the same throughout the period, because there are charges for the full storage capacity and not the actual one used, as it is provided by cloud computing. On the contrary, the second case scenario demonstrates the cost-effectiveness and the benefits gained by adopting the big data-as-a-service model during the first four years. However, the cost difference/benefits results become negative during the fifth year, which reveal the need for immediate upgradation that will be motivated in order to meet the demand requirements. The necessity for upgradation can be also observed at the increased costs compared to those in the conventional data warehouse approach. In this case, the earnings gained throughout the period, due to the selection of the dig data-as-a-service business model, will be reinvested on the additional storage required, maximizing the return on investment.

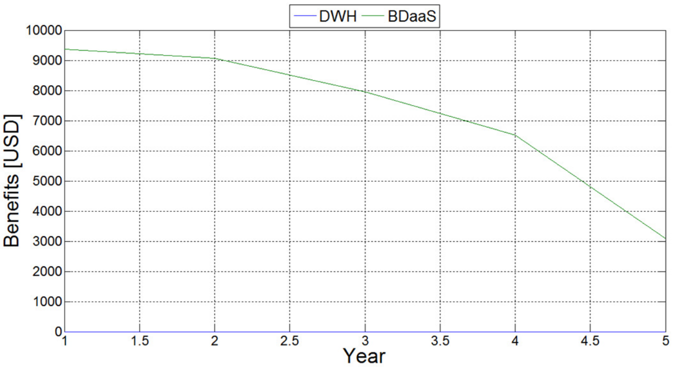

Fig. 4. Case Scenario 1: Cost difference/Benefits comparison.

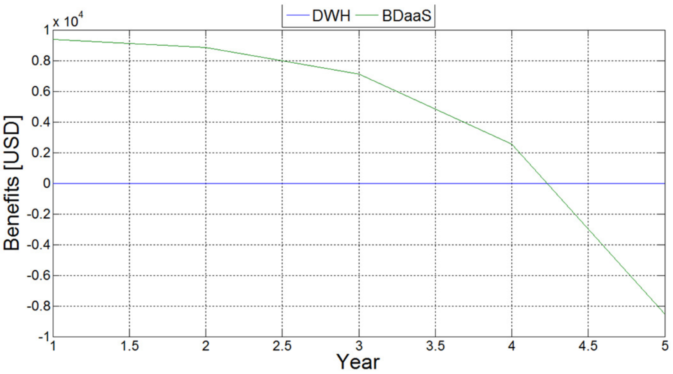

Fig. 5. Case Scenario 2: Cost difference/Benefits comparison.

Table 5. THE COST AND BENEFITS ANALYSIS RESULTS REGARDING BIG DATA-AS-A-SERVICE FOR CASE SCENARIO 1.

	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**
CA	9360	10024.56	12104.66	17140.19	25298.93
C_D	9360	9069.84	7944.46	6517.77	3090.63

Table 6. THE COST AND BENEFITS ANALYSIS RESULTS REGARDING BIG DATA-AS-A-SERVICE FOR CASE SCENARIO 2.

	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**
CA	9360	10810.8	14508.09	24482.41	40444.94
C_D	9360	8845.2	7113.51	2544.59	-8553.08

V. CONCLUDING REMARKS

In this paper, a novel cloud-inspired cost and benefits analysis model is proposed towards the evaluation of a big data-as-a-service (BDaaS) approach against conventional data warehouses. The lease of cloud storage is investigated under the assumption that predicted fluctuations in the demand occur, adopting a non-linear and asymmetric approach. The evaluation analysis indicates that possible upgradation of the storage in high-performance data warehouse appliances will introduce accumulated costs, while the adoption of a big data-as-a-service model will bring earnings to organizations that can be reinvested on the additional storage required, maximizing the return on investment.

REFERENCES

[1] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, “A survey of mobile cloud computing: architecture, applications, and approaches,” *Wirel. Commun. Mob. Comput.*, vol. 13, no. 18, pp. 1587–1611, Dec. 2013.

[2] D. Agrawal, S. Das, and A. El Abbadi, “Big data and cloud computing: current state and future opportunities,” in *Proceedings of the 14th International Conference on Extending Database Technology (EDBT/ICDT ’11)*, 2011, pp. 530–533.

[3] L.-J. Zhang, “Editorial: Big services era: Global trends of cloud computing and big data,” *IEEE Trans. Serv. Comput.*, vol. 5, no. 4, pp. 467–468, 2012.

[4] G. Mastorakis, E. Markakis, E. Pallis, C. X. Mavromoustakis, and G. Skourletopoulos, “Virtual network functions exploitation through a prototype resource management framework,” in *2014 IEEE 6th International Conference on Telecommunications and Multimedia (TEMU)*, Heraklion, Crete, Greece, 2014, pp. 24–28.

[5] Informatica Corporation, “Making Sense of Big Data,” 2015. [Online]. Available: https://marketplace.informatica.com/solutions/making_sense_of_big_data. [Accessed: 27-Nov-2015].

[6] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, “Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility,” *Future Gener. Comput. Syst.*, vol. 25, no. 6, pp. 599–616, Jun. 2009.

[7] V. Nallur and R. Bahsoon, “A decentralized self-adaptation mechanism for service-based applications in the cloud,” *Softw. Eng. IEEE Trans. On*, vol. 39, no. 5, pp. 591–612, 2013.

[8] P. Neves and J. Bernardino, “Big Data Issues,” in *Proceedings of the 19th International Database Engineering & Applications Symposium*, 2015, pp. 200–201.

[9] E. Xinhua, J. Han, Y. Wang, and L. Liu, “Big Data-as-a-Service: Definition and architecture,” in *Communication Technology (ICCT), 2013 15th IEEE International Conference on*, Guilin, 2013, pp. 738–742.

[10] Z. Zheng, J. Zhu, and M. R. Lyu, “Service-generated big data and big data-as-a-service: an overview,” in *Big Data (BigData Congress), 2013 IEEE International Congress on*, Santa Clara, CA, 2013, pp. 403–410.

[11] F. Ye, Z. Wang, F. Zhou, Y. Wang, and Y. Zhou, “Cloud-Based Big Data Mining & Analyzing Services Platform Integrating R,” in *Advanced Cloud and Big Data (CBD), 2013 International Conference on*, Nanjing, 2013, pp. 147–151.

[12] K. Park, M. C. Nguyen, and H. Won, “Web-based collaborative big data analytics on big data as a service platform,” in *Advanced Communication Technology (ICACT), 2015 17th International Conference on*, Seoul, 2015, pp. 564–567.

[13] M. Chen, S. Mao, and Y. Liu, “Big data: A survey,” *Mob. Netw. Appl.*, vol. 19, no. 2, pp. 171–209, 2014.

[14] H. Chen, R. H. Chiang, and V. C. Storey, “Business Intelligence and Analytics: From Big Data to Big Impact.,” *MIS Q.*, vol. 36, no. 4, pp. 1165–1188, Dec. 2012.

[15] X. Sun, B. Gao, L. Fan, and W. An, “A cost-effective approach to delivering analytics as a service,” in *Web Services (ICWS), 2012 IEEE 19th International Conference on*, Honolulu, HI, 2012, pp. 512–519.

[16] F. Zulkernine, P. Martin, Y. Zou, M. Bauer, F. Gwadry-Sridhar, and A. Aboulnaga, “Towards Cloud-Based Analytics-as-a-Service (CLAaaS) for big data analytics in the cloud,” in *Big Data (BigData Congress), 2013 IEEE International Congress on*, Santa Clara, CA, 2013, pp. 62–69.

[17] L. Vaquero, A. Celorio, F. Cuadrado, and R. Cuevas, "Deploying Large-Scale Data Sets on Demand in the Cloud: Treats and Tricks on Data Distribution.," *Cloud Comput. IEEE Trans. On*, vol. 3, no. 2, pp. 132 – 144, Jun. 2015.

[18] S. Yousfi and D. Chiadmi, "Big Data-as-a-service solution for building graph social networks," in *Cloud Technologies and Applications (CloudTech), 2015 International Conference on*, Marrakech, 2015, pp. 1–6.

[19] Y. Zhao, R. N. Calheiros, G. Gange, K. Ramamohanarao, and R. Buyya, "SLA-Based Resource Scheduling for Big Data Analytics as a Service in Cloud Computing Environments," in *Parallel Processing (ICPP), 2015 44th International Conference on*, Beijing, China, 2015, pp. 510–519.

[20] J. Cohen, B. Dolan, M. Dunlap, J. M. Hellerstein, and C. Welton, "MAD skills: new analysis practices for big data," *Proc. VLDB Endow.*, vol. 2, no. 2, pp. 1481–1492, 2009.

[21] G. Skourletopoulos, R. Bahsoon, C. X. Mavromoustakis, G. Mastorakis, and E. Pallis, "Predicting and quantifying the technical debt in cloud software engineering," in *2014 IEEE 19th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, Athens, Greece, 2014, pp. 36–40.

[22] G. Skourletopoulos, C. X. Mavromoustakis, G. Mastorakis, J. J. P. C. Rodrigues, P. Chatzimisios, and J. M. Batalla, "A fluctuation-based modelling approach to quantification of the technical debt on mobile cloud-based service level," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM 2015) - The Fourth IEEE International Workshop on Cloud Computing Systems, Networks, and Applications (CCSNA)*, San Diego, California, USA, 2015.

[23] G. Skourletopoulos, R. Bahsoon, C. X. Mavromoustakis, and G. Mastorakis, "The technical debt in cloud software engineering: a prediction-based and quantification approach," in *Resource Management of Mobile Cloud Computing Networks and Environments*, 1st ed., Hershey, PA: IGI Global, 2015, pp. 24–42.

[24] G. Skourletopoulos, C. X. Mavromoustakis, G. Mastorakis, E. Pallis, J. M. Batalla, and G. Kormentzas, "Quantifying and evaluating the technical debt on mobile cloud-based service level," in *Proceedings of the 2016 IEEE International Conference on Communications (ICC 2016) - Communications QoS, Reliability and Modelling Symposium (CQRM)*, Kuala Lumpur, Malaysia, 2016.

[25] G. Skourletopoulos, C. X. Mavromoustakis, G. Mastorakis, J. M. Batalla, and J. N. Sahalos, "An evaluation of cloud-based mobile services with limited capacity: a linear approach," *Soft Comput. J.*, 2016.

[26] Google Cloud Platform, "Google Cloud Storage Pricing," Dec-2015. [Online]. Available: https://cloud.google.com/storage/pricing. [Accessed: 27-Nov-2015].

[27] Y. Kryftis, C. X. Mavromoustakis, J. M. Batalla, G. Mastorakis, E. Pallis, and G. Skourletopoulos, "Resource usage prediction for optimal and balanced provision of multimedia services," in *2014 IEEE 19th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, Athens, Greece, 2014, pp. 255–259.

[28] J. M. Batalla, M. Kantor, C. X. Mavromoustakis, G. Skourletopoulos, and G. Mastorakis, "A novel methodology for efficient throughput evaluation in virtualized routers," in *Proceedings of the 2015 IEEE International Conference on Communications (ICC 2015) - Communications Software, Services and Multimedia Applications Symposium (CSSMA)*, London, UK, 2015, pp. 6899–6905.