MEE: A Hybrid Cloud Management Solution with Data Governance Method

Hao Wei, Joaquín Salvachúa Rodriguez*
Departamento de Ingeniería de Sistemas Telemáticos, Universidad Politécnica de Madrid, 28040 Madrid, Spain.

* Corresponding author. Tel.: +34 91 549 5700 ext. 3056; email: joaquin.salvachua@upm.es
Manuscript submitted February 17, 2018; accepted March 15, 2018.
doi: 10.17706/ijcce.2018.7.3.58-67

Abstract: Cloud computing represents the modern way of IT efficiency. Though cloud adoption is growing rapidly, the overall adoption rate is still not ideal. This paper discusses the concerns about cloud computing from an enterprise perspective and proposes a solution of hybrid cloud with data governance methodologies. This solution uses holistic point of view, which combines data management methods and cloud computing technologies together, to answer and ease the concerns. Various topics between private cloud and public cloud are discussed and compared in the paper, include security, customization, data transfer, data/vendor lock-in, Software as a Service (SaaS) integration, cost, and resource provision. It is important to note these topics because cloud infrastructure management is not simply about cloud technologies but also data quality management. After a brief overview of data governance, a solution with modeling, enforcement, and evaluation (MEE) is proposed to manage a hybrid cloud infrastructure. This kind of research has the possibility to influence better adoption of cloud computing and improve data quality in the enterprise area.

Key words: Cloud computing, data governance, hybrid cloud management.

1. Introduction

Cloud computing is an old concept came into reality. The concept date back to 60s, has tremendous popularity in modern software development. It provides “infinite” computing resource for software that needs fast scaling, shorten the time for people to get sufficient resource and only need to pay for what they have used [1]. It enables Infrastructure as Code, DevOps which improved the productivity of software development. Cloud computing accelerates the SaaS applications deployment and availability of getting online also facilitates the applications to reach customers.

From the access perspective, cloud computing could be classified into three types of cloud, public cloud, private cloud, and hybrid cloud. Take the well-accepted definitions from U.S. National Institute of Standards and Technology, a public cloud is the cloud for public open use, a private cloud is running on-premises in dedicated facilities for a single organization, while a hybrid cloud is a combination of private and public cloud [2].

Though cloud computing is not a young technology, from a survey [3] conducted in 2016, it shows that cloud adoption of organizations is not ideal. This survey shows that only 29% of respondents are heavily using cloud infrastructure, while in enterprises which have more than 1000 employees, this number drops

to 25%. Overall, enterprise community still has concerns about cloud adoption. According to another report from IDG Survey [4], the top three concerns in enterprise are data location, security and vendor lock-in. The survey key result is shown in Fig. 1. It is also worth to note that concern surrounding security has significantly reduced from 2015, which was top one concern, and data location has rated as the most concerned factor by 53% of respondents.

Challenges to Implementing a Cloud Strategy	2016	2015	2014
Concerns about where data is stored	53% 1	N/A	N/A
Concerns about the security of cloud computing solutions (risk of unauthorized access, data integrity and protection)	52% 2	67% 1	61% 1
Concerns about vendor lock-in	44% 3	N/A	N/A
Concerns surrounding integration (making info. available to applications outside the cloud, preserving a uniform set of access privileges)	39%	43% 2	46% 2
Lack of the right skill sets to manage and derive the maximum value from cloud investments	37%	N/A	N/A
Concerns about the ability of cloud computing solutions to meet enterprise and/or industry standards (compliance)	34%	35% 3	27%
Difficulty measuring ROI/determining accurate economic value of cloud solutions	33%	27%	30%
Concerns about information governance (eDiscovery and other information management requirements)	32%	30%	35% 3
Lack of clear strategy or help from key vendors in adapting their applications to cloud computing platforms	31%	16%	20%
Business leaders are not receptive to cloud computing solutions	25%	15%	11%
Employees are not receptive to cloud computing solutions	18%	10%	10%

Fig. 1. Challenges to implementing a cloud strategy [4].

The concerns of the enterprise community are hindering the adoption rate of cloud and the productivity of organization. This paper proposes a hybrid cloud management solution with data governance approach to solve the problems or reduce the concerns in enterprise, in order to encourage the wider adoption and release more potential of cloud computing. Our solution takes into account of special requirement and regulation of large organization. To achieve this, the data governance method with proven result is introduced into the proposed solution.

The paper makes following contributions. We compared and discussed seven aspects of private and public cloud in the context of enterprise reality, and our analysis suggests that hybrid cloud is a suitable choice for enterprise. We introduced data governance method into hybrid cloud research and proposed a three steps hybrid cloud management solution with data governance method to improve data quality in cloud.

The remaining sections of this paper are organized as follows. The analysis and comparison of the private and public cloud are discussed in section II. In section III, we argue that hybrid cloud is an ideal solution for enterprise. Section IV presents a three steps hybrid cloud management solution with data governance. The analysis of the proposed solution is discussed in section V. Section VI concludes the paper and outlines the future work.

2. Comparison of Private and Public Cloud

2.1. Security

In security, private cloud has more advantages than public cloud. The security topic in cloud computing has been discussed in several publications [5], [6]. Major public cloud providers have heavily invested in it; significant improvement has been made. In private cloud, the customer could have full knowledge on every security details and full control on all implementations. Furthermore, customer's security strategy or customized requirement can be implemented and applied immediately, while security practices of public cloud providers are still lacking in transparency. Customers of public cloud cannot conduct the self-risk evaluation; neither can find accountability from the public cloud provider.

2.2. Customization

Public cloud providers normally target general customer requirement on the market, have slow or no reaction to a specific requirement. That brings another topic, the public cloud would be slow or difficult to adopt specific requirement from customer, while private cloud could be highly customized to meet such requirement. This ability of customization is critical for the business unit, also influence the cloud adoption decision.

2.3. Data Transfer

Data transfer is a bottleneck for cloud adoption. When bringing the infrastructure to the cloud, every company faces legacy data. Transport data to cloud normally is expensive and requires significant effort. The methods include data transfer via Internet and by logistic shipping. This gives advantage to private cloud, because private cloud is likely built in near place of legacy data storage, or it even reuses of existing storage system. This situation facilitates the data movement to the cloud. One exception may exist is that new relatively independent business which has no legacy data.

2.4. Data/Vendor Lock-in

For enterprise, not be able to gain full control of the data is one of the worst scenarios. Data lock-in is already a well-discussed obstacle in cloud computing. Similarly, vendor lock-in is also a key concern in cloud adoption. According to the report, 78 percent of IT decision makers have concerns about vendor lock-in that prevent their organizations maximize the benefit of cloud [7]. All the providers have various services to help customers transfer data onto the cloud platform, while it is not often to see the service of withdrawing data. Besides this, public cloud providers now offer an extensive range of SaaS and tend to differentiate their service in the market. On one hand, the SaaS help customers build application more easily and bring the advanced R&D experience and technologies from cloud providers to the customers. The other hand, once the customers built the dependency on cloud provider's SaaS, it would be difficult to switch to another provider. This vendor lock-in is one of the most important factors in the evaluation of cloud adoption.

2.5. SaaS Integration

The SaaS integration experience could be very divergent in the context of public cloud and private cloud. Give the lack of critical attention paid to SaaS integration, it is worth to note that there is a certain number of specialized SaaS vendors on the market, which provide professional SaaS with low cost or even free tier. These SaaS could contribute and improve company's R&D and operation efficiency. This saves the cost and time effort to bring the product to market and customer. However, most of the such SaaS customer scenarios are focused on public cloud users only. For private cloud user, there are fewer choices with a much higher price. Therefore, SaaS integration is more expensive and more limited for private cloud user.

2.6. Cost

In recent years various approaches [8], [9] have been proposed to further reduce the cloud storage cost. Additionally, prior research has suggested [1], as the scale of data center grows, the cost of utilities, operation and material would plunge with the factors of 5 to 7.

Normally public cloud has large or extra-large data center while private cloud is usually built upon small to medium size data center. Thus, we believe the public cloud price would be lower than Private cloud.

Another important fact is that cloud computing is about pay as you go model, customers can use computing resources without a contract or advance commitment. Whereas, to build a private cloud, organizations need to invest a solid budget and undertake the risk.

2.7. Resource Provision

Several studies, for instance [10]-[12], have been carried out on resource allocation and tasks scheduling in cloud. Those effort has been made with the purpose of improving data center resource utilization efficiency and performance.

However, we still see the difference between public cloud and private cloud on resource provision. As noted earlier, due to the size of the data center of public cloud, the resource from public cloud could be seen as "infinite" for cloud user. We believe that compared with public cloud private cloud would be easier to hit the upper limit of resource capacity.

First of all, the scale of private cloud is smaller than public cloud, because of different target users. Private cloud is targeting users in a single organization and public cloud is for public customers. In most cases, the resource of private cloud is less than public cloud. As time goes and also the usage grows up, the ability of resource provision of private cloud would hit on the limit due to the initial design of the data center.

Secondly, the motivations are different. Public cloud providers have strong commercial motivation to offer as much as possible resource, therefore resource limit would rarely be a problem. While private cloud is a part of internal IT organization, the motivation to expand the current data center would rely on business decision. It would have a certain delay to respond the requirement of internal customers.

3. Hybrid Cloud for Enterprise

We believe that the hybrid cloud solution is a suitable cloud solution for enterprise. Because base on existing fact that for a private organization the most important data is already stored and managed by themselves, the IT infrastructure is already invested and existed. To give up the existing infrastructure and move to public-cloud-only infrastructure is not investment efficient neither time efficient. In such circumstance, hybrid cloud is a low-cost solution. It can transform into private cloud simply based on existing infrastructure, also take the advantage of nearby data storage location.

As a result, hybrid cloud, that can connect the existing infrastructure to public cloud and bring two types of cloud together, can deliver the best benefit of cloud. On private cloud part, an organization can apply their effort and implement the chosen solution to enhance security. The highly customized cloud can be done immediately after the organization's decision. Normally the facilities of private cloud are owned by enterprise or by dedicated provider and co-located, so the data location is easy to be inspected. As mentioned above, private cloud can also reuse legacy resource. This can easily bring existed data onto the cloud and have little or no data transfer cost. Since in hybrid cloud, the private cloud part is fully under control, the risk of not being able to switch public cloud vendors is low. With the method that distributes and regulates data in hybrid cloud, the effort on switching public cloud vendors also decreases. This method will be extended in the following solution section.

On public cloud part of hybrid cloud, it takes benefits of public cloud, like the low cost of the resource, unlimited resource provision, broad choices of SaaS suppliers. However, in hybrid cloud, this is compromised degree compared with public cloud, it is still obvious improvement than private cloud.

Admittedly, hybrid cloud will introduce extra complexity into IT infrastructure management. Hybrid cloud also adds more challenges to an enterprise which is lack of experience and expertise to manage hybrid cloud infrastructure. Besides, both in industry and academia, there only are few solutions or research that address the challenge of hybrid cloud in enterprise.

4. Hybrid Cloud with Data Governance

4.1. Data Governance Literature Review

Data governance is a young but emerging research topic. It is a framework of decision rights and accountabilities of data compliance [13]. Data governance is a series of programs that make sure the data

asset's quality throughout a company.

In industry, there already are several organizations which conducted practices in such area. IBM created IBM Data Governance Council Maturity Model to measure organization's data governance competencies [14]. The Data Governance Institute (DGI) publishes the DGI Data Governance Framework to help the implementation of data governance in organization [15]. The Data Management Association (DAMA) developed DAMA Guide to the Data Management Body of Knowledge (DAMA DMBK) which collects processes and knowledge areas as the best practice of data management [16]. An open sourced delivery framework Method for an Integrated Knowledge Environment (MIKE 2.0) provides comprehensive methodology and solutions for information management projects [17]. In the scope of this paper, we compare the frameworks in four dimensions. Fig. 2 shows the comparison of the frameworks.

Previous academic research is more about building models, proposing frameworks of data governance [13], [18], [19]. As to technique and solution research, we find there is still a gap.

Comparison Table of Data Governance Frameworks

Frameworks	Data definition & rules	Enforcement	Auditing	Has implemented technique/practice
IBM Data Governance Council Maturity Model	Yes	No	Yes	No
DAMA-DMBOK2	Yes	Yes	Yes	No
MIKE2.0	No	Yes	No	Yes
DGI Data Governance Framework	Yes	No	No	No

Fig. 2. Data governance frameworks comparison.

4.2. Proposed Hybrid Cloud Solution with Data Governance

In the field of data governance research, the researchers believe that accountability is a key factor of corporate-wide data quality management [13]. Introducing responsibility of data decision domains has significant impact on improving data quality. Vijay Khatri, et al. [19] in a well-grounded article proposed five decision domains for data governance. They are data principles, data quality, metadata, data access, and data lifecycle. Data principles define the owner of the data; data quality measures data in four dimensions, accuracy, timeliness, completeness, and credibility; metadata describes the precise concept of what is data; data access is the access requirement of data; data lifecycle includes data generation, retention, and termination.

Inspired by the promising research result of data governance, we propose a three steps hybrid cloud management solution with data governance method. The solution is called MEE, the abbreviation of three steps of the solution, modeling, enforcement, and evaluation. Measured by four data dimensions from data governance, the modeling step divides the hybrid cloud into different infrastructure domains and generates data policy for each domain. The enforcement step uses policy-based application deployment to manage application lifecycle. The accountability is implemented and traced in the evaluation step by application topology discovery and visualization, and exceptional log data analysis. The three steps go iteratively, keep updating and evolving to continuously achieve the best of data quality.

4.2.1. Modeling

In modeling, the corporate data needs to be reviewed and classified into several data labels, which will be used to divide the whole cloud infrastructure into different domains. The corresponding result of this step will shape the whole cloud infrastructure.

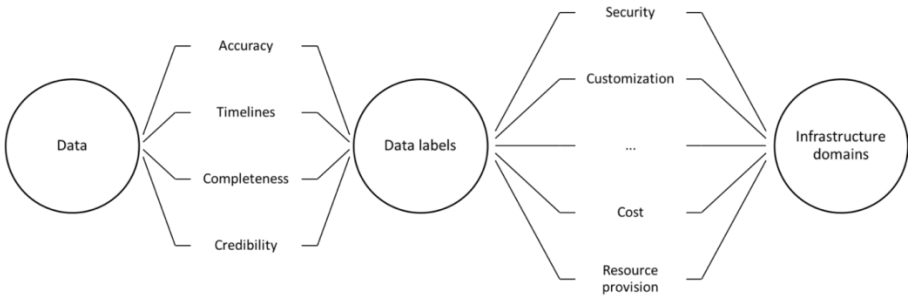

Fig. 3. Procedure of modeling.

The procedure of modeling is shown in Fig. 3. Firstly, we use the four data dimensions include accuracy, timeliness, completeness, and credibility that mentioned in [19] to measure the corporate data and group the data with similar characters. One group of data is called data label. For example, in an E-commerce system, the precise price information and quick transaction handling is extremely important. Therefore, the data in the system has very high requirement of accuracy and timeliness. In an Enterprise Resource Planning (ERP) system, people care more about the trustworthiness of the data. The credibility has higher priority than other dimensions. The data in two systems would fall into two different data labels.

Secondly, the data labels divide the cloud infrastructure into several domains. In section II, we discussed seven aspects of cloud infrastructure, security, customization, data transfer, data/vendor lock-in, SaaS integration, cost, and resource provision. Here these seven aspects together with the data labels will help to form the infrastructure domains, which host the different types of applications. Continued with the above example, the data label of E-commerce system data would benefit significantly from fast and low-costed resource provision. The infrastructure domains on public cloud suit the label very well. Obviously, the data label of ERP system data fits better in the infrastructure domains in private cloud.

In the end of modeling step, the data labels and infrastructure domains are expected deliverables. Data label describes the data characters and infrastructure domain to reside.

4.2.2. Enforcement

The enforcement step takes input from modeling step into execution and implementation. In this step, policy-based application deployment is adopted to manage the data lifecycle. In both research and industry, policy-based deployment has drawn some attentions. [20] discussed the security policy-based deployment model. Reference [21] presented an implementation of policy-based deployment across multiple computing infrastructures.

The following four procedures define our solution's policy-based application deployment. The procedure of this step is shown in Fig. 4.

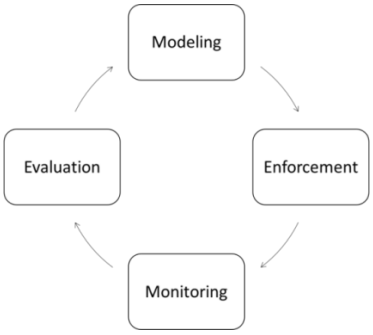

Fig. 4. Procedures of enforcement.

1. Policy definition. It takes data label as input from modeling step and generates deployment policies. The policy contains the infrastructure domains, resource types and the amount of resource.

2. Template creation. The deployment policy will be translated into deployment template by policy engine. The policy engine is responsible for interpreting the policy and acquire the deployment resource and create the deployment template. The template contains the address of deployment resource, the credential of the resource and related configurations.

The template creation algorithm is shown in Algorithm 1. It iterates through all the policies. If a policy is validated, it will be sent to the infrastructure domain broker as resource request. If it is not validated, this policy will be skipped. After the infrastructure domain broker receives a request, it returns a response. The response contains information of requested resource, e.g. IP address, access credential, and configuration information. Otherwise, the response gives error message and the policy will be skipped. The template combines all successful returned resource information. At last the algorithm returns the template.

```
Data: policies data
Result: template file with configurations
initialization;
foreach policy in policies do
    read policy;
    if policy validation is true then
        connect infrastructure domain broker;
        if request resource return true then
            read resource information;
            write resource address, credential,
              configurations into template;
        else
            skip policy;
        end
    else
        return null;
    end
end
return template;
```

Algorithm 1. Template creation.

3. Deploying. User trigger the deployment with template.

4. Policy update. When update the policy, it will go back to the beginning of the enforcement step and repeat the above procedures.

Therefore, it realizes the application lifecycle management include provision, retention, and termination.

4.2.3. Evaluation

The evaluation step detects the anomaly in hybrid cloud and traces the accountability in the corporate. We adopt the monitoring and logging technologies to enable the capabilities of this step.

The monitoring technology is application topology discovery and visualization. Previous research work [22] and patent [23] have showed the applicability of similar technology. We propose an agent-based application topology discovery and visualization technology. The agent is deployed with application in enforcement step, and it collects data of host machine and running processes. The agent also monitors the transactions among application components. With the transaction tracing, the dependencies of application components can be interpreted, and the real-time topology graph of the application can be drawn.

The logging technology is also an important part of evaluation step. The log information must contain the abnormal data behavior which could potentially decrease data quality and the application which generates the data. Thus, we could have enough information to trace the accountability of data quality.

Similar to monitoring, the logging agent is also deployed in the enforcement step. It receives the log file from application components, transforms it, and transports back to log database. Taken the large scale of

the infrastructure into the consideration, the scalability of the logging system would have some limitations. The large amount of log file can cause huge data traffic in network and also occupy too much computing resource to process log information. To optimize for better scalability, the logging format and level should be specially designed to keep the log information small and less frequent generated. A possible log format design is shown in Fig. 5. The log code shows the cross-infrastructure domain data access result, value 0 means access request failed, value 1 means access enabled/succeed, value 2 means access disabled. The log level is particular for the solution and the log with such level will only be reported on daily basis on every host to reduce traffic performance overhead.

Time	Host IP	Class	Service Name	Level	Log Code
Jun 29, 2017 11:16:20	127.0.0.1	org.xyz.SampleClass	demo-eureka-server	MEE	0
Jun 30, 2017 20:21:00	127.0.0.1	org.xyz.SampleClass2	demo-message-server	MEE	1
Aug 15, 2017 09:11:30	127.0.0.1	org.xyz.SampleClass3	demo-mysql-server	MEE	2

Fig. 5. Log format design and sample log.

In modern business, the context of data quality is rapidly changing. In this step, data model will also be updated to keep pace with business change. It is also a start point of next iteration, update requirement, apply the latest enforcement. It keeps the solution adaptive and flexible to a fast-changing environment.

5. Discussion

IQ International – the International Association for Information and Data Quality (IAIDQ), conducted an industry survey on the state of information quality and data governance in 2016 [24]. From the report of the survey, 56 % enterprise participants rated good or excellent on effectiveness of data quality and data governance efforts. The survey result implies that data governance is a well-accepted data management method and is endorsed by enterprise practice. It can help organizations to achieve the goal of data quality. It is the evidence to suggest that data governance has a positive effect on data quality. Given this promising result from data governance, we believe the proposed hybrid cloud management solution would lend itself well for improving data quality and encouraging cloud adoption in enterprise.

We are already aware that one drawback for this solution is the performance overhead that is introduced by the monitoring and logging technologies in evaluation step. This performance overhead has not been thoroughly studied and may decrease the efficiency of infrastructure. However, we think that with suitable tools and right implementation, this performance overhead could be reduced to an acceptable level.

6. Conclusion and Future Work

The challenges that organization have when they adopt cloud strategy are about data. In modern data management practice, data governance is an effective method to improve data quality and reduce data security and compliance risk. Whereas, hybrid cloud management is not given enough attention from academic research. Given this orientation, we proposed the MEE solution to manage a hybrid cloud with data governance method.

This paper presents a useful solution to increase cloud adoption in enterprise. The three steps solution include modeling, enforcement, and evaluation, describing an iterative way to build a data quality focused system. The modeling step defines the data label and divides hybrid cloud infrastructure into several infrastructure domains. The enforcement step makes sure the application is deployed in the correct infrastructure domain. Furthermore, the evaluation step watches on application running status and provides insights of application running data. The three steps complete a full cycle of data management

with the proven research resulting from data governance. We believe that it can help companies to design and manage infrastructure to achieve better data quality.

The future work will involve the research of the technologies mentioned in the solution. It is worth to note that at the moment of writing the proof of concept prototype of policy-based application deployment in enforcement step is already finished. The agent-based application monitoring prototype in evaluation step is currently in progress. We will conduct a more thorough study with the implemented prototypes in real organization scenarios.

Acknowledgment

The first author wishes to thank Dr. Xiang Xu's constructive comments on earlier drafts of this paper and Dr. Kunyang Fan's valuable help.

References

[1] Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R. H., Konwinski, A., Lee, G., Patterson, D. A., Rabkin, A., Stoica, I., & Zaharia, M. (2009). Above the clouds: A berkeley view of cloud computing. *Science, 53(4)*, 50-58.

[2] Mell, T. G. P. (2011). The NIST definition of cloud computing. *Communications of the ACM, 53(6)*, 50-50.

[3] Weins, K. (2016). Cloud computing trends: 2016 state of the cloud survey. Retrieved from the website: https://www.rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2016-state-cloud-survey

[4] IDG Communications. (2016, January). 2016 IDG cloud computing survey. Retrieved from the website: https://www.idg.com/tools- for-marketers/2016-idg-enterprise-cloud-computing-survey

[5] Modi, C., Patel, D., Borisaniya, B., Patel, A., & Rajarajan, M. (2013, Feb.). A survey on security issues and solutions at different layers of cloud computing. *The Journal of Supercomputing, 63(2)*, 561-592.

[6] Islam, T., Manivannan, D., & Zeadally, S. (2016). A classification and characterization of security threats in cloud computing. *Int. J. Next-Gener. Comput*, *7(1)*.

[7] Logicworks. (2016, July). Roadblocks to cloud success report. Logicworks, 155 Avenue of the Americas Fifth Floor New York, NY 10013.

[8] Liu, J., & Shen, H. (2016, December). A popularity-aware cost-effective replication scheme for high data durability in cloud storage. *Proceedings of the 2016 IEEE International Conference on Big Data* (pp. 384-389). IEEE.

[9] Liu, J., & Shen, H. (2016, December). A low-cost multi-failure resilient replication scheme for high data availability in cloud storage. *Proceedings of the 2016 IEEE 23rd International Conference on High Performance Computing (HiPC)* (pp. 242-251). IEEE.

[10] Liu, J., & Shen, H. (2016, December). Dependency-aware and resource-efficient scheduling for heterogeneous jobs in clouds. *Proceedings of the 2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)* (pp. 110-117). IEEE.

[11] Liu, J., Shen, H., & Chen, L. (2016, September). CORP: Cooperative opportunistic resource provisioning for short-lived jobs in cloud systems. *Proceedings of the 2016 IEEE International Conference on Cluster Computing (CLUSTER)* (pp. 90-99). IEEE.

[12] Liu, J., Shen, H., & Narman, H. S. (2016, December). CCRP: Customized cooperative resource provisioning for high resource utilization in clouds. *Proceedings of the 2016 IEEE International Conference on Big Data* (pp. 243-252). IEEE.

[13] Wende, K. (2007, December). A model for data governance - organizing accountabilities for data quality management. *Proceedings of the 18th Australasian Conference on Information Systems* (pp. 417–425).

[14] Alder, S. B. (2007, October). The IBM data governance council maturity model: Building a roadmap for effective data governance.

[15] Thomas, G. (2006). The DGI data governance framework. Retrieved from the website: http://www.datagovernance.com/the-dgi-framework

[16] International, D. (2009. March). *DAMA Guide to the Data Management Body of Knowledge - DAMA-DMBOK*. USA: Technics Publications LLC.

[17] Rindler, A., McClowry, S., Hillard, R., & Mueller, S. (2013, March). *Information Development Using MIKE 2. 0.* Motion Publishing.

[18] Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2016). A conceptual framework for designing data governance for cloud computing. *Procedia Computer Science, 94*, 160–167.

[19] Khatri, V., & Brown, C. V. (2010, January). Designing data governance. *Commun. ACM, 53(1)*, 148–152.

[20] Dulay, N., Lupu, E., Sloman, M., & Damianou, N. (2001). A policy deployment model for the ponder language. *Proceedings of the 2001 IEEE/IFIP International Symposium on Integrated Network Management* (pp. 529-543). IEEE.

[21] J. L. Hicks. (2015, June). Policy-based application deployment and continuous best - fit placement across heterogeneous computing infrastructures. U. S. Patent Application 20150169306A1.

[22] Krein, J. (2012, April). Framework for application topology discovery to enable migration of business processes to the cloud. University Stuttgart. Retrieved from the website: https://elib.uni-stuttgart.de/handle/11682/2987

[23] Berg, D. C., Blancett, B. L., Elder, M. D., Holliday, C., Kalantar, M. H., Konstantinou, A. V., Makin, N., Pouyer, T. A., & Swanke, J. E. (2014, September). Automated discovery of a topology of a distributed computing environment. U. S. Patent 8849987B2.

[24] Pierce, M. Y. A. E., & Yonke, L. (2016, June). The state of information quality and data governance - 2016 industry survey & report. Retrieved from the website: https://www.iqint.org/publication/state-information-quality-data-governance-2016-industry-survey-report/

Hao Wei is a researcher at Next Generation Internet Research Group in Universidad Politécnica de Madrid. He received his B.Sc. degree in Beijing Normal University at 2009 and M.Eng. in Beijing Jiaotong University at 2011, Beijing, China.

He has several publications in journal and conference on biomedical image processing and computer science. His main research interests are cloud computing infrastructure, cloud service, cloud management, No-SQL database, virtualization technologies, mobile computing, machine learning and artificial intelligence.

Joaquín Salvachúa Rodriguez received a M.Sc. degree in telecommunications from Universidad Politécnica de Madrid, Spain, in 1989 and Ph.D. degree from the same university in 1994. He was a lecturer at Technical University of Madrid from 1990 to 1995, when he was promoted to associate professor at the Department of Telematics Engineering.

He has participated in over 12 European research projects with several publications at international magazines, books and conferences. Current research activities include advanced multimedia applications design, human computer interaction, and multicast and P2P sharing with QoS protocol design. His main research interests are social graphs and recommendation, cloud infrastructures, P2P, DHT (Distributed HashTables), non-SQL data BASES, IaaS, PaaS, identity management and blockchain.