



Trusted Data Integration in Service Environments: A Systematic Mapping

Senda Romdhani, Nadia Bennani, Chirine Ghedira-Guegan, Genoveva Vargas-Solar

► To cite this version:

Senda Romdhani, Nadia Bennani, Chirine Ghedira-Guegan, Genoveva Vargas-Solar. Trusted Data Integration in Service Environments: A Systematic Mapping. Proceedings of ICSOC 2019, Oct 2019, Toulouse, France. hal-02315151

HAL Id: hal-02315151

<https://hal.archives-ouvertes.fr/hal-02315151>

Submitted on 14 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Trusted Data Integration in Service Environments: A Systematic Mapping ^{*}

Senda Romdhani¹, Nadia Bennani², Chirine Ghedira-Guegan¹, and Genoveva Vargas-Solar⁴

¹ University of Lyon, CNRS, University of Lyon 3, LIRIS, Lyon, France
(senda.romdhani,chirine.ghedira-guegan)@univ-lyon3.fr

² LIRIS, INSA-Lyon, Villeurbanne, France
nadia.bennani@insa-lyon.fr

³ University Grenoble Alpes, CNRS, Grenoble INP, LIG-LAFMIA, 38000 Grenoble, France genoveva.vargas@imag.fr

Abstract. This paper identifies and discusses trends and open issues regarding the use of trust in data-provisioning service environments, especially cloud environments. Applying a systematic review method [2], we propose a classification scheme used to provide a quantitative view of current trust solutions insisting in open issues. Finally, using analysis results, we give the general lines of an approach for improving data provisioning in multi-cloud using Service Level Agreement (SLA) and proposing the notion of multi-level trust.

Keywords: Trust · Data-provisioning Services · Service Environments · Multi-Cloud · Service Level Agreement

1 Introduction

In recent years, data-provisioning service environments ranging from Service Oriented Architectures (SOA) and cloud architectures have become widely used provisioning environments [1]. The cloud eases the provisioning by providing dynamically scalable and virtualized resources as services [3] under a pay as you go model. The process of making data available in trustworthy conditions using the cloud is significantly challenging. Particularly because services are deployed under multi-tenant and multi-layer configurations.

In order to illustrate these challenges, let us consider the following e-health scenario. Assume that, for preparing a surgery, a doctor needs to have information about her patient including laboratory analysis, blood pressure etc. This information is produced by several actors participating in patients medical control (e.g., chemist, cardiologist, smart devices etc.). Consider that these actors use different clouds for storing and giving access to their data. A data integration tool can provide a global vision of these data to the doctor guided by quality requirements specified in SLA. The challenge lies in consuming and composing

^{*} Funded by SUMMIT (<http://summit.imag.fr>) of the Auvergne Rhone Alpes region.

data from cloud services with different SLA properties (e.g., QoS, reliability) and that provide data under different conditions and quality properties (e.g., timeliness, security). These properties depend on the way data services were developed and on those added by the clouds where they are deployed (e.g. availability, data replication, security measures etc). Besides, the offered QoS specified in SLAs generally fluctuates due to the uncertainty and dynamics of the cloud and this explains the need for a quality warranty.

Guiding the data integration process by adding a trust management dimension as quality warranty should overpass the QoS uncertainty and improve the integration's result. Thus, by evaluating the trust level of data provisioning services and their composition it is possible to perform a reliable integration process.

The objective of our research being to identify the trends and open issues regarding trusted data integration, we deem necessary first conducting a comprehensive analysis on trust in data-provisioning service environments especially in the cloud and on the role of SLA in providing trust. In this context, the main contribution of this paper is a classification scheme that results from applying a systematic review method [2] which consists of 5 inter-dependant steps including: (i) Setting a research scope, (ii) retrieving candidate research papers from online databases, (iii) selecting relevant papers to answer research questions, (iv) defining a classification scheme, and (v) Performing statistical analysis. Based on our systematic review study we propose the general lines of aspects to be considered for providing a trusted data integration solution by composing trustworthy provisioning services on multi-cloud settings. The remainder of this paper is as follows: Section 2 describes our mapping study. Section 3 gives a quantitative analysis, identifies open issues and present our solution. Section 4 concludes the paper and discusses future work.

2 Trust Challenges in Service based Data-Provisioning

The aim of our systematic mapping is to (i) Categorize and quantify research contributions on trust in service environments and especially cloud computing. (ii) Categorize the key contributions of SLA-based research works. (iii) Discover open issues and limitations in existing work. Our study is guided by three research questions:

RQ1. How have published papers on trust evolved towards the cloud and other service environments? This question is devoted to identify trust solutions evolution towards the cloud, the research trends and contributions.

RQ2. What are the most and the least addressed evaluation targets and how are they combined? Trust can be associated to different entities defined as evaluation target. This question aims to determine the frequency of addressing entities implied in the data integration process in service environments and whether trust has been simultaneously considered for more than one entity

RQ3. Have SLAs been used and how was it related to trust? The question aims to determine whether SLA's have been used to evaluate trust and if so, expose contributions.

2.1 Conducting Papers' Search and Screening

This step consisted in collecting papers from three online databases: IEEE, ACM, and Science Direct. A set of keywords was chosen using taxonomies and topics from conferences considered influential in the scientific community. We used the following general query for searching papers and retrieved a total of 3351 papers.

Trust AND (multi-cloud OR cloud OR service)

As a result of the filtering process specified in the review method, only 446 papers were included. Note that this study is normally influenced by various factors like the choice of keywords used to define the query, the way they are combined into a conjunctive and disjunctive expression and the selection of databases.

2.2 Key-wording using Abstracts

This step consists in analyzing selected papers and key-wording using frequent terms derived from abstracts. First, the frequent terms are considered as facets and then each facet is organized into dimensions forming a classification scheme. Our scheme defines 5 facets for classifying trust challenges⁴:

Evaluation environment: This facet proposes 3 dimensions to classify data-provisioning service environments including, *single cloud*, *multi-cloud* (e.g. hybrid cloud, collaborative clouds etc.) and *service environment* (e.g. SOA).

Evaluation target: According to our study a (trust) evaluation target can be a *service provider*, a *service user*, a *composite service*, a *service* or *data*.

SLA: Groups dimensions describing the actions performed on SLAs w.r.t trust evaluation solutions namely works that *extend SLA* or that propose some *trust evaluation metrics* or strategies for computing and *monitoring* SLA.

Contribution: Groups the dimensions that characterize the type of contributions in papers. It classifies proposals into five dimensions, namely, models, frameworks, methods, approaches and tools.

Validation approach: This facet include 5 dimensions namely *experiments*, *comparisons*, *benchmarks*, *scenarios* and *use cases*.

2.3 Data Extraction and Mapping Process: Quantitative Analysis

RQ1. How have published papers on trust evolved towards the cloud and other service provisioning environments? Combining the facets Contribution, Evaluation Environment and Validation Approach, we observe contributions' trends on trust in the cloud (figure 1). The resulting bubble chart shows that most research papers propose trust models and that experimentation (77%) is the most used way for validating models. According to our study, few solutions address trust in multi-cloud environments (5,8%).

RQ2. What are the most and the least addressed evaluation targets in each service environment and how are they combined? The facets Evaluation, Environment and Evaluation Target (Figure 2) put the lights on the frequency of

⁴ For more information and references please use this link: https://drive.google.com/drive/folders/17SW_e8kbrROtpu0VTjNnqN9u1vQ3scdp?usp=sharing

addressing each evaluation target per service environment. We can observe that most research contributions focused on evaluating the trustworthiness of services (35% cloud services). The results also show that little attention has been given to composite services and data and that are mostly addressed for a single cloud and other service environments. It seems that trust in data integration remains an open issue when combined with multi-cloud.

Some papers addressed trust evaluation on more than one evaluation target simultaneously. Nevertheless, this multi-evaluation concerns at best two-levels.

RQ3. Were SLAs used in these publications and how was it related to trust? According to our quantitative analysis, we found that only 25 papers used SLA for trust evaluation. The facets SLA and Evaluation Environment give elements for determining which actions have been applied on SLA in each environment (Figure 2). The results shows that about 50% of proposals defined a set of SLA trust metrics. We can see that 10 papers proposed an SLA monitoring solution. These contributions are mostly deployed on single cloud. We can conclude from the results that there is merely no added new dimensions in SLA specific for trust evaluation and that papers tend to use the standard SLA form.

3 Open Issues and Outlook

Our systematic review shows that trust is an important property considered by proposals dealing with data provision, services, and the (multi)-cloud. Still, there are open issues regarding trusted data integration as it remains unexplored in multi-cloud environments. As explained in our scenario, data integration combines trust issues from data (data providers) and from the integration process itself which uses composite services and the cloud. Thereby we conclude that trusted data integration on multi-cloud is important and must consider 3 trust levels: data, service and cloud. Yet, current trust solutions do not cover the 3 trust levels simultaneously (see section 2). Our work will propose a trusted data service composition algorithm based on SLA to compute query results considering all the chain trustworthiness. We also identify as promising research area the need of enhancing SLA beyond cloud resources quality. To do so, it is important to identify the set of data integration requirements and the missing information in SLA that can lead to a three-dimensional trust solution.

4 Conclusion and Future Work

A multi-cloud is a collaborative environment where service providers can increase their access to multiple cloud resources and tune their conditions. This collaboration generates the proliferation of data-provisioning services offering to end users heterogeneous SLAs. This facility may generate doubts for users who delegated data management to the cloud and they may want some trust warranties (e.g., completeness of data, resource availability etc.).

This paper aimed at presenting a systematic mapping study about trust in data provisioning environments and especially the cloud. It identified trends

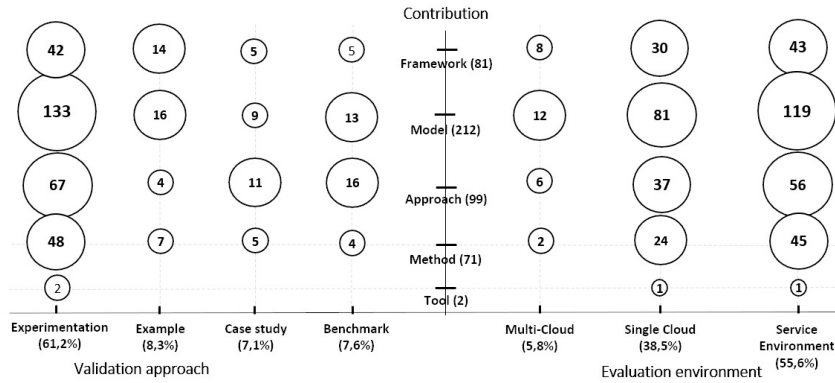


Fig. 1. Contribution trends on trust in the cloud

and open issues and presented the general lines of a multi-level trust-based data integration solution. In our ongoing work, we intend to develop a solution by focusing on trust in the cloud and also study SLAs in more depth to adapt them and use them to guide trusted data integration.

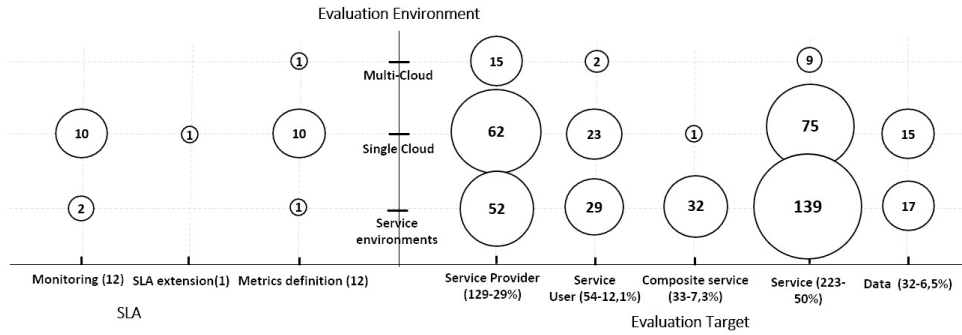


Fig. 2. Dealing with trust on different evaluation targets using SLA

References

1. Carvalho D.A.S., Neto P.A.S., Vargas-Solar G., Bennani N., Ghedira C.: Can Data Integration Quality Be Enhanced on Multi-cloud Using SLA?. In: Database and Expert Systems Applications. LNCS, vol 9262. Springer, Cham (2015).
2. Petersen, K., Feldt, R., Mujtaba, S., Mattsson, M.: Systematic mapping studies in software engineering. In: Ease, Vol. 8, pp. 68–77. ACM, Italy (2008)
3. Zhang, Q., Cheng, L., Boutaba, R.: Cloud computing: state-of-the-art and research challenges. Journal of internet services and applications, 1(1), 7–18. (2010)