Using SLA to guide database transition to NoSQL on the cloud: a systematic mapping study

Fabio Leal and Martin A. Musicante
Computer Science Department, Federal University of Rio Grande do Norte, Natal, Brazil
Email: sousaleal.fabio@gmail.com, mam@dimap.ufrn.br

***Abstract*—Cloud computing became a reality, and many companies are now moving their data-centers to the cloud. A concept that is often linked with cloud computing is Infrastructure as a Service (IaaS): the computational infrastructure of a company can now be seen as a monthly cost instead of a number of different factors. Recently, a large number of organizations started to replace their relational databases with hybrid solutions (NoSQL DBs, Search Engines, ORDBs). These changes are motivated by (*i*) performance improvements on the overall performance of the applications and (*ii*) inability to a RDBMS to provide the same performance of a hybrid solution given a fixed monthly infrastructure cost. However, not always the companies can exactly measure beforehand the future impact on the performance on their services by making this sort of technological changes (replace RDBMS by another solution). The goal of this systematic mapping study is to investigate the use of Service-Level-Agreements (SLAs) on database-transitioning scenarios and to verify how SLAs can be used in this processes.**

***Keywords*—Transition to Cloud, NoSQL, Systematic Mapping.**

I. Introduction

The adoption of cloud solutions is growing fast among organizations [1]. Centralized (mostly mainframe) technology is being replaced by distributed and more flexible forms of data storage and processing. This change of paradigm is motivated by the necessity to improve the use of resources, as well as by the increasing velocity in which data is produced.

In this scenario, transitions must take into account the quality of the service delivered by the new solutions. The notion of Service-Level-Agreement (SLA) [2] may be used as a parameter in this context. SLAs are widely used to to provide service thresholds between clients and providers and are present in almost every service contract over the internet.

SLAs or OLAs - Operational Level agreements, are particularly useful to guide the process of choosing the most convenient service from a pool of service providers.

In this paper, we survey the use of SLA on database-transitioning scenarios, trying to investigate how they might be used to help the execution of this process. Our study is performed using the systematic mapping [3] technique: A set of papers is retrieved from the most popular bibliography repositories; this set is then filtered according to predefined parameters and finally, the analysis of the remaining papers is guided by a small number of research questions.

This work was partly funded by CNPq (Brazil, PDE-201118/2014-9, PQ-305619/2012-8, INCT-INES-573964/2008-4), CAPES/CNRS (Brazil/France, SticAmSud 052/14) and CAPES/ANII (Brazil/Uruguay, Capes-UdelaR 021/10).

This paper is organized as follows: Section II presents some of the concepts that are related to the transition from a traditional setting to a cloud-aware one. Section III presents our research questions and each step of our survey. The outcomes of our Systematic Mapping study can be seen on Section IV.

II. The Technological Shift

On the early 90's it was commonplace for every Information Technology (IT) company to have its own Data Center with huge servers and mainframes. IT costs were high, and high-performance computing was available only for big companies, as data centers required a large physical infrastructure and have high costs for maintenance [4].

The regular way of building a web application was to use a client-server approach, where the server was a powerful (and expensive) machine. At the same time, new players, such as Google or Yahoo, were rising with bigger missions: *"to organize the world's information and make it universally accessible and useful"* [5]. The popularization of the internet use new ways of commerce exchange, yielding an explosion in the amount of data produced and exchanged. It was *just* impossible to store the petabytes of daily-generated data in a single server.

From this point on, the community realized the economical convenience of building and maintaining several low-performance servers, instead of a single high-performance one, even if this this requires a change of culture in the administration of the new datacentres. The new approach is also incompatible with the traditional way of building applications, that usually were designed to work on a single server and database.

Several research initiatives were conducted in this area and a common solution was rising: to distribute data storage and processing. Google, Yahoo and other big IT players helped to build open source tools to make this approach possible, like Hadoop [6].

This revolution brought to life the notion of *Cloud Computing*, together with new concepts, such as Infrastructure as a Service *(IAAS)*, Platform as a Service *(PAAS)* and Software as a Service *(SAAS)* [7]. According to [7], *Cloud computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services.*

Data Integration & Polyglot Persistence: On the last years, the number of Data Base (DB) Engines grew like never before [8]. Along with the NoSQL (Not only SQL) movement and expansion of Social Networks, new concepts for Database Models appeared, like Document Store, Search Engines, Key-Value store, Wide Column Store, Multi-Model and Graph DBMS. In [8] a ranking of the most popular DB engines is presented.

Today, instead of having a single Relational Database Management System (DBMS) for the whole application, it is efficient and cost-effective to have several Data Base Engines, one for each type of data that the application handles. This concept is called *Polyglot Persistence* [9].

As [10] illustrates, polyglot persistence is very useful in the context of e-commerce applications that deal with a catalog, user access logs, financial information, shopping carts and purchase transactions, for example. The notion of polyglot persistence is built upon the observation that the *nature* of each data type is significantly different (i.e: user logs imply high volume of writes on multiple nodes, shopping carts need high availability and user sessions require rapid access for reads and writes).

As computing services started to decentralize, developers started to build applications that depended of several data-sources. By this time the use of Web Services and Service Oriented Architecture (SOA) became more popular [4].

Service Level Agreements (SLAs): According to *ITILv3's* official glossary [11], a Service Level Agreement (SLA) is "*an agreement between an IT service provider and a customer. A service level agreement describes the IT service, documents service level targets, and specifies the responsibilities of the IT service provider and the customer.*"

The agreement consists on a set of measurable constraints that a service provider must guarantee to its customers. In practical terms, it is a document that a service provider delivers to its consumers with minimum quality of service (QoS) metrics. If the service is delivered with a lower QoS than is promised on the SLA, consumers may be refunded or earn benefits that were accorded beforehand.

Systematic Mappings: According to [3], "*A software engineering systematic map is a defined method to build a classification scheme and structure a software engineering field of interest.*" Systematic Mapping studies provide a global view of a given research field and identify the quantity, results, and the kinds of researches in this field.

A Systematic map is composed by a number of steps (Figure 1).

On the first step, "Definition of Research question", the questions that must be answered on the survey are defined. On the "Review Scope" step, researchers target the papers/journal sources that will be taken into consideration on the systematic map. After that, the "Search" step is done using a set of predefined search engines and a body of papers ("All papers") is retrieved. After an initial "Screening of the papers", the "Relevant papers" are chosen according to inclusion and exclusion criteria defined by the research team. At this point, the papers that will participate of the study are selected. The selection is based on the title, abstracts and keywords of each paper ("Keywording using Abstracts"). After that, a "Classification Scheme" is built, defining different points-of-view (or facets) from which the body of papers will be classified. After matching each paper with the classification schema ("Data Extraction and Mapping Process"), the systematic mapping is performed. In this phase the relationships between the collected data (in the light of the classification scheme) are used to answer the research questions.

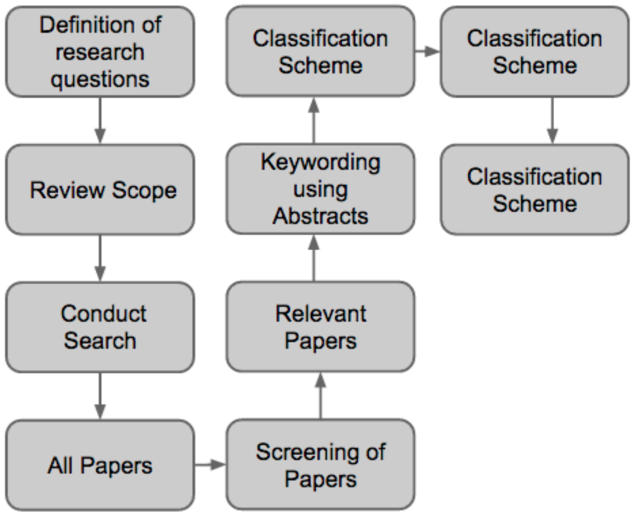

Fig. 1. Systematic Mapping Steps [3].

In the next section we build a systematic mapping to evaluate the use of SLA in database transition scenarios, from relational database management systems to NoSQL.

III. A Systematic Mapping

In this section we develop our systematic mapping, in the way described in section II.

We start by defining our research scope and by pointing out the questions to be answered by the end of the survey (Section III-A), followed by the selection of keywords that composed the search string (Section III-B). The other steps of the systematic mapping procedure follow.

After that, we defined a classification schema and classified the relevant papers. These steps (screening of papers to classification scheme) were not linear as shown on Figure 1. We had to iterate on the screened papers a few times to find the ideal classification schema to our systematic map.

A. Definition of the research questions

As we wanted to investigate the way in which SLAs has been used in database transitioning processes, we proposed three main research questions and two associated questions, as follows:

RQ1) What are the reasons to change from RDBMSs to NoSQL solutions? This question is about *why* technological changes are needed on Information Systems and what is the *motivation* behind them.

AQ1.1) What are the pros and cons to migrate from RDBMSs to NoSQL solutions? This question is particularly important, as we also want to *point out the down sides* of DB migrations.

AQ1.2) How can we measure the overall improvements promised by this change? After a migration is performed, it is also important to *measure the benefits* of this migration, and this question evidences this point.

RQ2) How can SLAs be used to guide database transitioning processes from RDBMSs to NoSQL databases in cloud-based apps? This is one of the main questions of our study, as we want to evidence the state-of-art on the use of SLAs in migration scenarios.

RQ3) Is there a standard representation of SLAs in cloud services? We also try to search for standards on the representation of SLAs to verify if future works can be done under this field.

These are relevant questions for our survey as our focus is to determine how SLAs can be used in Database migrations scenarios. Our questions are intended to cover *why* it is good to migrate from RDBMSs to a NoSQL solution and *how* we can do it. In the end of our survey we also reveal what are the most popular technologies related to migrations.

B. Definition of keywords

Our research scope can be splitted in three main categories:

1) Databases: As we wanted to investigate RDBMS to NoSQL transitions, we defined that a representative query string for this category would be

NoSQL AND (rdbms OR relational)

2) Migration: A number of terms terms were oftenly linked with database migration and transition scenarios. Most of them, however, are consistently represented by the radicals "migr" and "transit". Thus the representative search string for this category is defined by:

migr OR transit**

3) Service level agreements: Service Level Agreements can also be represented by its acronym (SLA). A search string for this category is defined as

"Service Level Agreement" OR SLA

As we defined the three main categories of interest, what we wanted to survey is their conjunction. In this way, the final search string can be represented as:

(migr OR transit*) AND (NoSQL AND (rdbms OR relational)) AND (SLA OR "Service Level agreement")*

C. Conduct Search for Primary Studies (All Papers)

The next step of our systematic map was to identify relevant publishers. We surveyed academia experts and initially chose five main sources: **Springer, ACM, Sciencedirect, IEEE and Elsevier**. All of these publishers have relevant publications about databases and Service Level Agreements, as they index publications from reputable international conferences & journals.

Each of these publishers had their own search engine, however some inconsistent results were obtained when querying some of them directly. For instance, on one occasion IEEE search engine returned 0 results to the query string "*changes AND database AND nosql AND sla*". When we searched the same query on Google Scholar filtering *only IEEE publications* we found **90** results. We supposed that this erratic behavior is due to network traffic conditions. In order to mitigate the risks of having different results for the same query on different search engines, we used Google Scholar as a meta-search engine. It is important to mention that the use of Google Scholar made it possible to enlarge our search space. Our scope is not limited to the list of publishers identified above; these publishers represent only the minimal coverage of the literature in this systematic map.

During this step, we also noticed that 2009 would be the starting year of our survey, since this was the year when the term NoSQL was reintroduced to the community, by Johan Oskarsson (Last.fm) [12].

D. Search Strategy

Google Scholar returned 74 results for our search string. We have also manually included other particularly relevant publications that were picked from selected sources, such as [13], a master thesis about Relational Database Schema Migration to NoSQL.

These publications were either not published by the selected publishers or were not indexed by Google Scholar.

Our search strategy was composed by the following steps:

Step 1 - Basic Search: Search publications from 2009 to 2015 that matched the query string and add other relevant publications manually. This step was performed from 01 April 2015 to 11 April 2015 and returned 80 publications.

Step 2 - Dump the retrieved results on a spreadsheet. This spreadsheet is publicly available on [14].

Step 3 - Screening for keywords in of title and abstract: On the spreadsheet each publication has title, abstract, year and referenced URL. Title and abstract were considered on the initial screening. On this step we discarded publications that are notably not related to the topic of our study.

Step 4 - Apply inclusion/exclusion criteria: The inclusion/exclusion criteria are shown below. We first applied the inclusion criteria over the selected works, and then exclusion criteria removed the out of scope publications. Only papers clearly not relevant for the purposes of the study were removed. The publications that were considered on this study are marked on the spreadsheet with a green background.

1) Inclusion Criteria:
   - The publication is about a migration from RDBMS to a NoSQL technology;
   - The main focus of the publication is on RDBMs or NoSQL systems;
   - The publication makes use or references a SLA.

2) Exclusion Criteria
   - Non-English written publications;
   - Access-restriction to the original publication (could not access the source of the publication);
   - Non-technical publications;
   - The work is about migrations within the same database DBMS;
   - The work is not related to databases or SLAs.
   - The work is related to databases but no comparison is made between two technologies.

A total of 35 publications were selected after this phase.

Step 5 - Fast Reading of the papers: We've performed a fast read of all publications that were not excluded on the previous step. This helped us to classify each paper accurately. This step is extensively detailed on the next section.

E. Classification of the Papers

We classified each selected publication in three facets: **Contribution Type**, **Technologies used** and **SLA representation**.

1) **Contribution type**: On the initial screening of papers we have defined 5 main types of contribution to our study. Other systematic mapping studies, such as [15] and [16] use a similar classification schema for contribution type.
   - Benchmark
   - Migration Experience Report
   - Bibliographical Review
   - Tool
   - Framework / Process

   When a migration of databases is described in a publication, we also classify specific data to answer the questions of this study: Source and Target technology of the migration, motivation to migrate and "Uses SLA or other artifact to validate the migration?";
2) **Technologies used**: To classify and rank the technologies mentioned on each paper we have developed a *"Batch-PDF-Tokens-Matcher"*. This tool matches a list of strings against a list of PDF files. The tool was made open-source and is publicly available on [17]. In our case, we wanted to find the most cited technologies on relevant publications. As it was not possible to include all database types on this category, we chose to match only the 250 most-popular databases, ranked and made publicly available by [8].
3) **SLA representation**: One of the questions that our study wanted to answer was RQ3 (Is there a standard representation of SLAs in cloud services?). There are several ways to represent an SLA, and we clusterized some of these ways. It is also possible that no information is given about SLA representation on the publication, so we added two subcategories to this facet: "No SLA is used" and "Missing information".
   - Tool/Code/Database records
   - Language
   - Missing Information
   - No SLA is used

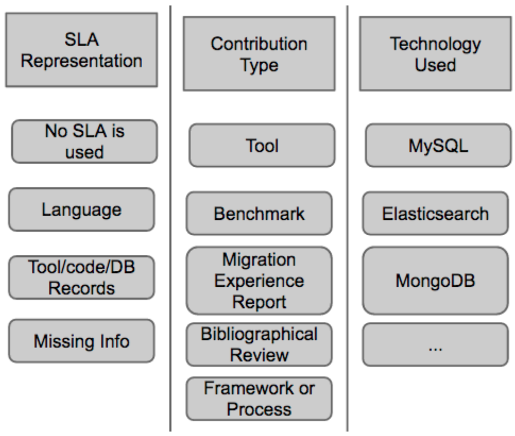

Fig. 2. Classification scheme for selected publications.

A graphical representation of our classification schema is detailed on Figure 2. We have listed only a subset of the several technology types that were found on the review.

IV. Outcomes

As mentioned in Section III-D, 74 publications were found matching our query string on IEEE, Elsevier, ACM, Springer and Sciencedirect using Google Scholar as a meta-search engine. We have manually added other 5 publications to our study. These publications were either not published by the selected publishers or were not indexed by Google Scholar.

35 out of the total of 79 publications were selected, after the screening of abstract and analysis of the inclusion/exclusion criteria[1].

The considered publications are: [18] [2] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [21] [29] [30] [31] [32] [33] [34] [35] [36] [37] [38] [39] [40] [41] [42] [43] [44] [45] [46] [47] [48] [1] [49].

The results are given in two steps: *Frequency Analysis between relevant criteria* and *Answers to research questions*.

A. Frequency Analysis between relevant criteria

Each analysis is presented below and is followed by a brief interpretation of the result.

- Publications by year: The publications by year table - Figure 3 - shows us that 2013 was the year when most of the publications were made on this research area. As we have just reached the middle of 2015 it is expected that not many publications are indexed on the search engines on this year.
- Frequency - Publication type vs Years: An interesting pattern can be found on the table Publication Type vs Years (Figure 4); "Migrations Experience Reports" were the main type of publication that our study aimed to discover, but only 3 publications of this type were found on our systematic map.

[1]A column in our spreadsheet [14] contains the justification for the rejection of each discarded publication.

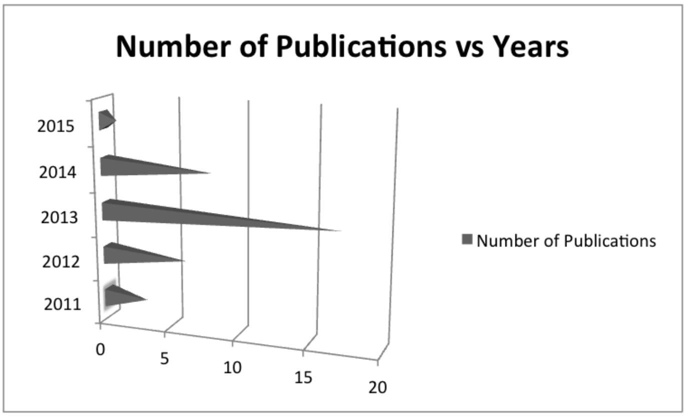

Fig. 3. Publications by year.

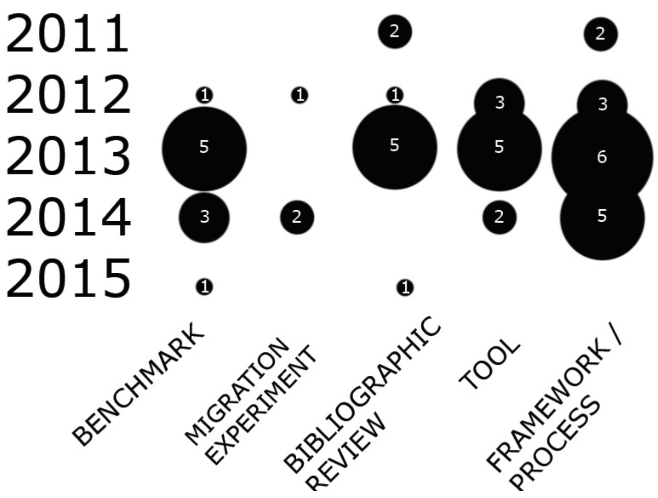

Fig. 4. Frequency - Publication type vs Years.

- Publications by country: Figure 5 shows that USA, Australia and Romania were responsible for almost 50% of the total selected publications. The other half of publications were distributed among other 16 countries, proving that the research area is not being developed by a single research group.

- Popular Technologies: Figure 6 shows that Cassandra, MongoDB and MySQL were the most-cited databases on the papers that we analyzed. This chart can be used as a starting point to propose new works in the area of database migration. The full count of each technology is available on [50].

- SLA Representation: 29% of the publications didn't make any use of SLAs. Another big part - 32% - didn't even mention how the SLA is represented internally. 3 publications propose or use DSLs (Domain Specific Languages) to represent SLAs. This evidences a clear absence of standards when defining/using SLAs. The full table is presented on Figure 7. [50].

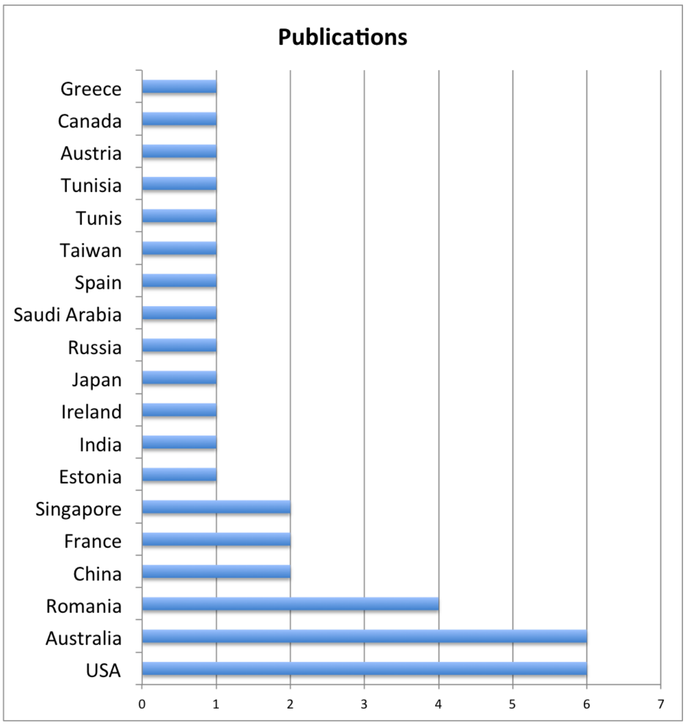

Fig. 5. Publications by country.

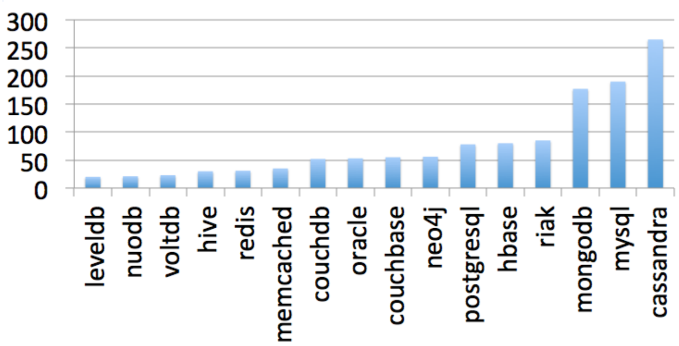

Fig. 6. Most-cited database technologies.

B. *Answers to research questions*

In this mapping study we have surveyed the state of the art in transitions from RDBMSs to NoSQL databases. We created a rigorous protocol which analyzed 79 publications to answer the research questions that we identified previously. We may consider the answer to these questions as the main outcome of this paper. The answers are summarized below:

RQ1-What are the reasons to change from RDBMSs to NoSQL solutions? and **AQ1.2**-What are the pros and cons to migrate from RDBMSs to NoSQL solutions? investigate the motivation of a database transition. We found a number of reasons to migrate from a relational database to a NoSQL/Hybrid model. As the migration is most of the time motivated by the benefits of these models, the answers to the questions RQ1 and AQ1.2 are presented together.

The publications that reference the benefits and disadvantages of NoSQL / Hybrid solutions were [46] [47] and [48].

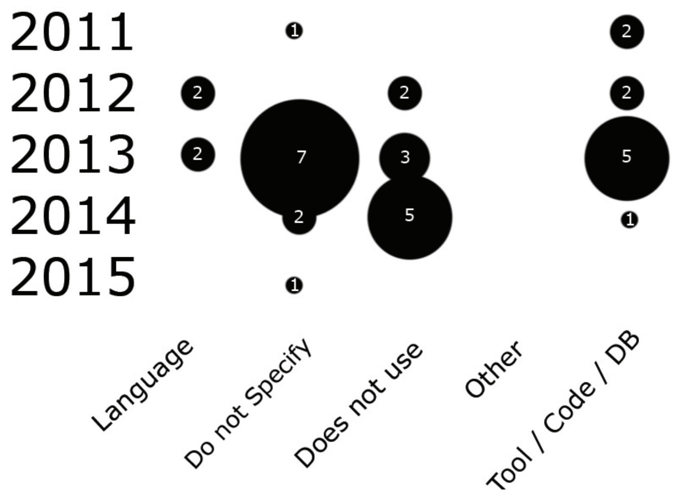

Fig. 7. SLAs representation.

Benefits::

- Segments of the data to be read and processed in parallel using a MapReduce framework, such as Hadoop;
- Schema-less data model;
- Support for large files;
- Scalability - relational databases tend to “Scale up”, which is opposed to the “Scale horizontally” strategy used by hybrid solutions;

Disadvantages::

- A big disadvantage on moving from a RDBMS to a NoSQL/Hybrid solution is the need for changes on the data models of the application. Several publications address this problem, such as [46] [51] and [52]
- Another disadvantage of NoSQL and Hybrid solutions is the fact that these concepts are relatively new. As we mentioned previously, the term NoSQL was used to reference these types of databases for the first time in 2009 [12]. The relational model has been in use for more than 30 years;As it is a new concept, it is particularly hard to find developers with a large experience in NoSQL databases.

AQ1.2-How can we measure the overall improvements promised by this change?

No publication was found addressing the problem of measuring the overall improvements after a database transition. In fact, as it is shown on IV-A, there are few publications with the “Migration Experience Report” type.

Several benchmarking frameworks, such as TPC-H, TPC-DS and YCSB were identified [21] during our survey, though. These benchmarking frameworks could be a good starting point to develop new tools and specialized frameworks to solve this problem. This seems to be a promising research theme for future works.

RQ2-How can SLAs be used to guide database transitioning processes from RDBMSs to NoSQL databases in cloud-based apps?

A number of works were found relating SLAs with Quality of Service (QoS) and Quality of Experience (QoE). Several publications, such as [22], [53] and [54] propose a SLA-centric approach to monitor and control the performance of cloud-based apps. [2], [19], [55] and [36] propose SLA-centric/User-Centric solutions to monitor the performance of web applications. All these solutions are technology-agnostic and could be used to monitor the performance improvements promised by a database transitioning process.

The question RQ2 was subject of discussion with industry experts and it was found out that there are some services, such as New Relic [56], Appsee [57] and Datadog [58] that provide SLA-monitoring tools for web apps.

RQ3-Is there a standard representation of SLAs in cloud services?

The selected publications did not present a standardized and common representation for SLAs. In fact, 32% of the selected publications did not even mention how the SLA was represented. 29% represented the SLAs as tables/documents stored on databases without any sla-oriented guidelines. This evidences a clear absence of standards when defining/using SLAs.

[1] proposes SYBL: *An Extensible Language for Controlling Elasticity in Cloud Applications. SYBL allows specifying in detail elasticity monitoring, constraints, and strategies at different levels of cloud applications, including the whole application, application component, and within application component code.* In other words, SYBL can also be seen as a language to specify SLAs in cloud environments.

Specifically related to SLAs on Cloud Services, [49] and [59] propose CSLA (Cloud Service Level Agreement): a language for Cloud Elasticity Management. CSLA features concepts related to SLAs on Cloud service, as QoS/functionality degradation and penalty models. These features are very useful, as it allows providers to express contracts with fines and penalties.

Some other initiatives that tried to define the mechanisms by which Webservice SLAs are established are WS-Agreement [60] and WSLA [61]. Also, [62] presents a framework for managing the mappings of the Low-level resource Metrics to High-level SLAs (LoM2HiS framework).

It is worth to notice that future works on standardizing the representations of SLAs are needed.

V. Conclusions

By analyzing the charts and the answers to the research questions we can conclude that the research area of this mapping study is still not mature.

There is a lack of consensus and standards on this research area, as no official guidelines for migrating from a relational database to a NoSQL/Hybrid database were found. It was also not found a standard way to represent and create SLAs. These two points seems to be promising research topics and will be covered on future works.

References

[1] G. Copil, D. Moldovan, H.-L. Truong, and S. Dustdar, "Sybl: An extensible language for controlling elasticity in cloud applications," in *Cluster, Cloud and Grid Computing (CCGrid), 2013 13th IEEE/ACM International Symposium on*, May 2013, pp. 112–119.

[2] S. Sakr and A. Liu, "Sla-based and consumer-centric dynamic provisioning for cloud databases," in *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*, June 2012, pp. 360–367.

[3] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering*, ser. EASE'08. Swinton, UK, UK: British Computer Society, 2008, pp. 68–77. [Online]. Available: http://dl.acm.org/citation.cfm?id=2227115.2227123

[4] A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and I. Stoica, "Above the clouds: A berkeley view of cloud computing," Tech. Rep., 2009.

[5] A. Spector, P. Norvig, and S. Petrov, "Google's hybrid approach to research," *Commun. ACM*, vol. 55, no. 7, pp. 34–37, Jul. 2012. [Online]. Available: http://doi.acm.org/10.1145/2209249.2209262

[6] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in *Mass Storage Systems and Technologies (MSST), 2010 IEEE 26th Symposium on*, May 2010, pp. 1–10.

[7] A. e. a. CONNOLY, D. FOX, "A view of cloud computing - http://delivery.acm.org/10.1145/1730000/1721672/p50-armbrust.pdf," April 2010, [Online; posted 01-April-2010].

[8] "Db engines ranking," http://db-engines.com/en/ranking, Accessed: 2014-05-01.

[9] P. J. Sadalage and M. Fowler, *NoSQL distilled: a brief guide to the emerging world of polyglot persistence*. Pearson Education, 2012.

[10] G. V. Solar, "Addressing data management on the cloud: tackling the big data challenges," May 2014. [Online]. Available: http://db-engines.com/en/ranking

[11] Axelos, "Best management practice portfolio: common glossary of terms and definitions," Oct. 2012. [Online]. Available: http://www.axelos.com/gempdf/Axelos_Common_Glossary_2013.pdf

[12] E. Evans, "Eric evans blog - nosql," http://blog.sym-link.com/2009/05/12/nosql_2009.html, Accessed: 2014-05-01.

[13] R. Lamllari, "Extending a methodology for migration of the database layer to the cloud considering relational database schema migration to nosql," Master's thesis, University of Stuttgart, 2013.

[14] F. Leal, "Systematic mapping spreadsheet," https://docs.google.com/spreadsheets/d/1N3DboEqthdiKG3VDMKlqyjFdHC3pxrN1XtXtNoNrhX8, Accessed: 2015-05-01.

[15] A. Tahir and S. Macdonell, "A systematic mapping study on dynamic metrics and software quality," in *Software Maintenance (ICSM), 2012 28th IEEE International Conference on*, Sept 2012, pp. 326–335.

[16] D. Ameller, X. Burgus, O. Collell, D. Costal, X. Franch, and M. P. Papazoglou, "Development of service-oriented architectures using model-driven development: A mapping study," *Information and Software Technology*, vol. 62, no. 0, pp. 42 – 66, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950584915000361

[17] F. Leal, "Fabio leal - pdf tokens matcher," https://github.com/fabiosl/pdf-tokens-finder, Accessed: 2015-05-01.

[18] C.-W. Huang, W.-H. Hu, C.-C. Shih, B.-T. Lin, and C.-W. Cheng, "The improvement of auto-scaling mechanism for distributed database - a case study for mongodb," in *Network Operations and Management Symposium (APNOMS), 2013 15th Asia-Pacific*, Sept 2013, pp. 1–3.

[19] L. Zhao, S. Sakr, and A. Liu, "A framework for consumer-centric sla management of cloud-hosted databases," *Services Computing, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2013.

[20] Z. Zheng, J. Zhu, and M. Lyu, "Service-generated big data and big data-as-a-service: An overview," in *Big Data (BigData Congress), 2013 IEEE International Congress on*, June 2013, pp. 403–410.

[21] R. Moussa, "Benchmarking data warehouse systems in the cloud," in *Computer Systems and Applications (AICCSA), 2013 ACS International Conference on*, May 2013, pp. 1–8.

[22] P. Xiong, "Dynamic management of resources and workloads for rdbms in cloud: A control-theoretic approach," in *Proceedings of the on SIGMOD/PODS 2012 PhD Symposium*, ser. PhD '12. New York, NY, USA: ACM, 2012, pp. 63–68. [Online]. Available: http://doi.acm.org/10.1145/2213598.2213614

[23] E. Alomari, A. Barnawi, and S. Sakr, "Cdport: A framework of data portability in cloud platforms," in *Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services*, ser. iiWAS '14. New York, NY, USA: ACM, 2014, pp. 126–133. [Online]. Available: http://doi.acm.org/10.1145/2684200.2684324

[24] C. Pahl and H. Xiong, "Migration to paas clouds - migration process and architectural concerns," in *Maintenance and Evolution of Service-Oriented and Cloud-Based Systems (MESOCA), 2013 IEEE 7th International Symposium on the*, Sept 2013, pp. 86–91.

[25] L. Zhao, S. Sakr, and A. Liu, "Application-managed replication controller for cloud-hosted databases," in *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*, June 2012, pp. 922–929.

[26] S. Sakr, "Cloud-hosted databases: technologies, challenges and opportunities," *Cluster Computing*, vol. 17, no. 2, pp. 487–502, 2014. [Online]. Available: http://dx.doi.org/10.1007/s10586-013-0290-7

[27] S. Sakr and A. Liu, "Is your cloud-hosted database truly elastic?" in *Services (SERVICES), 2013 IEEE Ninth World Congress on*, June 2013, pp. 444–447.

[28] V. Andrikopoulos, T. Binz, F. Leymann, and S. Strauch, "How to adapt applications for the cloud environment," *Computing*, vol. 95, no. 6, pp. 493–535, 2013. [Online]. Available: http://dx.doi.org/10.1007/s00607-012-0248-2

[29] E. Boytsov, "Designing and development of an imitation model of a multitenant database cluster," *Automatic Control and Computer Sciences*, vol. 48, no. 7, pp. 437–444, 2014. [Online]. Available: http://dx.doi.org/10.3103/S0146411614070049

[30] B. Sodhi and T. Prabhakar, "Assessing suitability of cloud oriented platforms for application development," in *Software Architecture (WICSA), 2011 9th Working IEEE/IFIP Conference on*, June 2011, pp. 328–335.

[31] D. Petcu, G. Macariu, S. Panica, and C. Crciun, "Portable cloud applicationsfrom theory to practice," *Future Generation Computer Systems*, vol. 29, no. 6, pp. 1417 – 1430, 2013, including Special sections: High Performance Computing in the Cloud & Resource Discovery Mechanisms for {P2P} Systems. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X12000210

[32] G. Giannikis, D. Makreshanski, G. Alonso, and D. Kossmann, "Workload optimization using shareddb," in *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '13. New York, NY, USA: ACM, 2013, pp. 1045–1048. [Online]. Available: http://doi.acm.org/10.1145/2463676.2463678

[33] K. Grolinger, W. Higashino, A. Tiwari, and M. Capretz, "Data management in cloud environments: Nosql and newsql data stores," *Journal of Cloud Computing: Advances, Systems and Applications*, vol. 2, no. 1, p. 22, 2013. [Online]. Available: http://www.journalofcloudcomputing.com/content/2/1/22

[34] A. Copie, T.-F. Fortis, V. Munteanu, and V. Negru, "Service datastores in cloud governance," in *Parallel and Distributed Processing with Applications (ISPA), 2012 IEEE 10th International Symposium on*, July 2012, pp. 473–478.

[35] A. Copie, T.-F. Fortis, and V. Munteanu, "Determining the performance of the databases in the context of cloud governance," in *P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2013 Eighth International Conference on*, Oct 2013, pp. 227–234.

[36] P. Xiong, Y. Chi, S. Zhu, J. Tatemura, C. Pu, and H. HacigümüŞ, "Activesla: A profit-oriented admission control framework for database-as-a-service providers," in *Proceedings of the 2Nd ACM Symposium on Cloud Computing*, ser. SOCC '11. New York, NY, USA: ACM, 2011, pp. 15:1–15:14. [Online]. Available: http://doi.acm.org/10.1145/2038916.2038931

[37] S. Srirama and A. Ostovar, "Optimal resource provisioning for scaling enterprise applications on the cloud," in *Cloud Computing Technology*

and Science (CloudCom), 2014 IEEE 6th International Conference on, Dec 2014, pp. 262–271.

[38] M. Dayarathna and T. Suzumura, "Towards scalable distributed graph database engine for hybrid clouds," in *Data-Intensive Computing in the Clouds (DataCloud), 2014 5th International Workshop on*, Nov 2014, pp. 1–8.

[39] L. Qiao, K. Surlaker, S. Das, T. Quiggle, B. Schulman, B. Ghosh, A. Curtis, O. Seeliger, Z. Zhang, A. Auradar, C. Beaver, G. Brandt, M. Gandhi, K. Gopalakrishna, W. Ip, S. Jgadish, S. Lu, A. Pachev, A. Ramesh, A. Sebastian, R. Shanbhag, S. Subramaniam, Y. Sun, S. Topiwala, C. Tran, J. Westerman, and D. Zhang, "On brewing fresh espresso: Linkedin's distributed data serving platform," in *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '13. New York, NY, USA: ACM, 2013, pp. 1135–1146. [Online]. Available: http://doi.acm.org/10.1145/2463676.2465298

[40] S. Sakr, A. Liu, D. Batista, and M. Alomari, "A survey of large scale data management approaches in cloud environments," *Communications Surveys Tutorials, IEEE*, vol. 13, no. 3, pp. 311–336, Third 2011.

[41] J. Montes, A. Snchez, B. Memishi, M. S. Prez, and G. Antoniu, "Gmone: A complete approach to cloud monitoring," *Future Generation Computer Systems*, vol. 29, no. 8, pp. 2026 – 2040, 2013, including Special sections: Advanced Cloud Monitoring Systems & The fourth {IEEE} International Conference on e-Science 2011 e-Science Applications and Tools & Cluster, Grid, and Cloud Computing. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X13000496

[42] A. J. Elmore, S. Das, A. Pucher, D. Agrawal, A. El Abbadi, and X. Yan, "Characterizing tenant behavior for placement and crisis mitigation in multitenant dbmss," in *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '13. New York, NY, USA: ACM, 2013, pp. 517–528. [Online]. Available: http://doi.acm.org/10.1145/2463676.2465308

[43] M. Dayarathna and T. Suzumura, "Graph database benchmarking on cloud environments with xgdbench," *Automated Software Engineering*, vol. 21, no. 4, pp. 509–533, 2014. [Online]. Available: http://dx.doi.org/10.1007/s10515-013-0138-7

[44] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: A technology tutorial," *Access, IEEE*, vol. 2, pp. 652–687, 2014.

[45] D. Shue and M. J. Freedman, "From application requests to virtual iops: Provisioned key-value storage with libra," in *Proceedings of the Ninth European Conference on Computer Systems*, ser. EuroSys '14. New York, NY, USA: ACM, 2014, pp. 17:1–17:14. [Online]. Available: http://doi.acm.org/10.1145/2592798.2592823

[46] A. Schram and K. M. Anderson, "Mysql to nosql: Data modeling challenges in supporting scalability," in *Proceedings of the 3rd Annual Conference on Systems, Programming, and Applications: Software for Humanity*, ser. SPLASH '12. New York, NY, USA: ACM, 2012, pp. 191–202. [Online]. Available: http://doi.acm.org/10.1145/2384716.2384773

[47] C. BĂZĂR, C. S. IOSIF *et al.*, "The transition from rdbms to nosql. a comparative analysis of three popular non-relational solutions: Cassandra, mongodb and couchbase," *Database Systems Journal*, vol. 5, no. 2, pp. 49–59, 2014.

[48] A. Gomez, R. Ouanouki, A. April, and A. Abran, "Building an experiment baseline in migration process from sql databases to column oriented no-sql databases," *J Inform Tech Softw Eng*, vol. 4, no. 137, p. 2, 2014.

[49] Y. Kouki, F. de Oliveira, S. Dupont, and T. Ledoux, "A language support for cloud elasticity management," in *Cluster, Cloud and Grid Computing (CCGrid), 2014 14th IEEE/ACM International Symposium on*, May 2014, pp. 206–215.

[50] "Full db json," https://gist.github.com/fabiosl/18b5e826c9daebda5165, Accessed: 2015-04-21.

[51] R. Cattell, "Scalable sql and nosql data stores," *SIGMOD Rec.*, vol. 39, no. 4, pp. 12–27, May 2011. [Online]. Available: http://doi.acm.org/10.1145/1978915.1978919

[52] C. Mohan, "History repeats itself: Sensible and nonsensql aspects of the nosql hoopla," in *Proceedings of the 16th International Conference on Extending Database Technology*, ser. EDBT '13. New York, NY, USA: ACM, 2013, pp. 11–16. [Online]. Available: http://doi.acm.org/10.1145/2452376.2452378

[53] I. Konstantinou, E. Angelou, D. Tsoumakos, C. Boumpouka, N. Koziris, and S. Sioutas, "Tiramola: Elastic nosql provisioning through a cloud management platform," in *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '12. New York, NY, USA: ACM, 2012, pp. 725–728. [Online]. Available: http://doi.acm.org/10.1145/2213836.2213943

[54] M. Klems, D. Bermbach, and R. Weinert, "A runtime quality measurement framework for cloud database service systems," in *Proceedings of the 2012 Eighth International Conference on the Quality of Information and Communications Technology*, ser. QUATIC '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 38–46. [Online]. Available: http://dx.doi.org/10.1109/QUATIC.2012.17

[55] ——, "A runtime quality measurement framework for cloud database service systems," in *Quality of Information and Communications Technology (QUATIC), 2012 Eighth International Conference on the*, Sept 2012, pp. 38–46.

[56] "New relic," http://newrelic.com, Accessed: 2014-05-01.

[57] "Appsee," https://www.appsee.com/, Accessed: 2014-05-01.

[58] "Datadog," https://www.datadoghq.com/, Accessed: 2014-05-01.

[59] Y. Kouki and T. Ledoux, "CSLA : a Language for improving Cloud SLA Management," in *International Conference on Cloud Computing and Services Science, CLOSER 2012*, Porto, Portugal, Apr. 2012, pp. 586–591. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00675077

[60] A. Andrieux, K. Czajkowski, A. Dan, K. Keahey, H. Ludwig, T. Nakata, J. Pruyne, J. Rofrano, S. Tuecke, and M. Xu, "Web Services Agreement Specification (WS-Agreement)," Global Grid Forum, Grid Resource Allocation Agreement Protocol (GRAAP) WG, Tech. Rep., Sep. 2005.

[61] S. Nepal, J. Zic, and S. Chen, "Wsla+: Web service level agreement language for collaborations," in *Services Computing, 2008. SCC '08. IEEE International Conference on*, vol. 2, July 2008, pp. 485–488.

[62] V. Emeakaroha, I. Brandic, M. Maurer, and S. Dustdar, "Low level metrics to high level slas - lom2his framework: Bridging the gap between monitored metrics and sla parameters in cloud environments," in *High Performance Computing and Simulation (HPCS), 2010 International Conference on*, June 2010, pp. 48–54.