Migration of relational databases to NoSQL - Methods of Analysis

Fábio Oliveira*
Email: fabio_oliveira@iscte-iul.pt
Abilío Oliveira*
Email: abilio.oliveira@iscte-iul.pt
Bráulio Alturas*
Email: braulio.alturas@iscte-iul.pt

* Instituto Universitário de Lisboa (ISCTE-IUL), and Information Sciences, Technologies and Architecture Research Centre (ISTAR-IUL), Lisboa, Portugal.

ABSTRACT
The amount of data to store, organize and manage in any organization, is very high and increases every day, fact well-known by companies as Facebook, Google or SAS. With this current growth rate, technologies must adapt to the amount of disposable data, and as a result, each technology is required. Big Data techniques in isolation are not focused, and this is a reason for a greater spread of NoSQL database models. The purpose of this article is to validate the existing (and already used) migration methods and to adapt them, to understand the most efficient method to migrate a relational database to a NoSQL database. We will show the methodology used and what were the steps followed for the implementation, as well as the configuration of the environment used during the tests. Results show that in this migration process, the most efficient method is what referred to as automatic offline migration. However, it requires a window of unavailability greater than the window of online migration, which in turn requires more resources from the operating system to migrate. Therefore, the most efficient method to migrate a database will depend on the application availability, and the computational resources available for it. We hope to make an important contribution in helping to choose a migration method to use, and the metrics that can be collected to better evaluate the performance of a migration.

METHOD
We have a case study of a qualitative nature, in which researchers define the starting point according to their own experience, or situations related to their practical life. Although in a case study different data collection techniques can be used, in this research we privileged the technique of observation, since we submit the data to tests and observe the results from them.
We carried out an extensive qualitative study with a sample of 9 cases of migrations, where we started with three approaches to scenarios to migrate using three different migration methods. In all 9 cases metrics were collected and the windows and migration were defined and which processes would be running in each window, so with that division we could measure and analyze the proposed metrics.

RESULTS AND DISCUSSION
In summary, we have verified the performance of three possible migration methods between relational and NoSQL databases, using a set of metrics that provide us with a detailed view of resource consumption as well as details about the migration process. From the theoretical point of view, the main contribution of this study is to show how methods of migrating relational databases to NoSQL and their associated metrics can be used, since these methods can be applied in several DBMS systems, not being linked only to this work, or to any software used here.

We identify the requirements and the various phases of a migration; After the state of the art survey, several points were collected during a migration from relational database to NoSQL, based on other studies, as well as on processes that suppliers or software maintainers indicate as the way forward - being able to consult the reference in the listing below. The following points were gathered to be validated as requirements for a migration, coming from some research done, and others added by us.

- Number of records / Initial situation (Rodrigues, 2009; Yaqub, 2012)
- Mapping the data types (Davenport & Dych, 2013; Gomes, 2011; Pereira, 2014)
- Restrictions and triggering (Antaño et al., 2014)
- Character encoding (Antaño et al., 2014; Neta, Neta, Juniose, & Oliveira, 2013)
- Tests (added by the researcher)
- Implementation (added by the researcher)
- Final Monitoring / Validation (added by the researcher)
- Staging area / Non staging area (added by the researcher)
- Failure of migrating database (added by the researcher)
- Data modelling (Gomes, 2011)

We can divide the migration method in phases, as there is four phases for the online method and two for the offline methods.

CONCLUSIONS
We have verified the performance of three possible migration methods between relational and NoSQL databases, using a set of metrics that provide us with a detailed view of resource consumption as well as details about the migration process. From the theoretical point of view, the major contribution of this study is to show how methods of migrating relational databases to NoSQL, and their associated metrics can be used, since these methods can be applied in several systems.
We can thus respond, after applying and analyzing metrics in migration cases and migration methods proposed here, that the most efficient migration method is the automatic offline migration method. However, this requires a large window of unavailability, and if it is not possible to stop the application or even the database during the migration process, the most efficient method will be the one indicated. With this context of information processing, it is always necessary to validate resources and unavailability as requirements of the project and not the migration itself, because it can happen without or with unavailability, however, with direct reflection in migration times.

* This work has been supported by the projects from ISTAR – IUL, and by the research project ‘DBMSIL’ funded by the Portuguese Foundation for Science and Technology (FCT) with reference number PTDC/STP/5781/2014.

In this metric, we can clearly see the amount of CPU consumed by the application, which is identified by the "CPU user" chart. We can also see the amount of CPU consumed by the operating system, the amount of CPU responsible for the I/O used.
In figure 8 we have the disk activity, that is, how much writing activity and how much reading activity we have in this phase of the migration, and here we can notice the reading behavior.

Therefore, in order to respond to this objective, we analyzed the metrics collected in each of the phases, a small sample of these metrics will be visualized here, and however, all the metrics can be analyzed in the work done in (Oliveira, 2017).
In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.

In figure 7 we can see the CPU consumption during the first phase of the OLTP scenario migration through the online migration method.