

# Proposta de Mestrado

Fabio Leal

May 13, 2014

## 1 Background

The goal of this chapter is to present the technical concepts for a better understanding of our job.

### 1.1 Cloud Computing

On the early 90's it was commonplace for every Information Technology (IT) company to have its own Data Center with lots of huge servers and main-frames. IT costs were high, and high-performance computing was available only for big companies, as data centers required a lot of physical space and high costs for maintenance.

The regular way of building a web application was to use client-server approach, where the server was an extremely powerful (and expensive) machine. However, new companies, such as Google, were rising with bigger missions: *“to organize the world's information and make it universally accessible and useful”*. It was *just* impossible to store the petabytes of daily-generated data in a single server.

From this point, they also realized that it was way cheaper to build and maintain several low-performance servers than a single high-performance machine. This approach, however, was incompatible with the traditional way of building applications, as they were designed to work with a single server and database.

Lots of research 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.

This revolution brought to life new concepts, such as Infrastructure as a Service (*IAAS*), Platform as a Service (*PAAS*) and Software as a Service (*SAAS*).

[2] says that *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*.

## 1.2 Data Integration & Polyglot Persistence

On the last years, the number of Data Base (DB) Engines grew like never before. Along with the NoSQL 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.

[3] presents a ranking of the most popular DB engines.

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.

Take for instance a classic e-commerce application that deals with a catalog, user access logs, financial information, shopping carts and purchase transactions. These data have different management requirements, for instance, the catalog has a lot of reads and very few writes assuming that the catalog of a shop is more or less stable. Information about user sessions require rapid access for reads and writes but they do not need to be durable. The shopping carts need high availability across multiple locations, and can merge inconsistent writes. User activity logs imply high volume of writes on multiple nodes, while recommendations for users must provide rapid link traversals between friends, product purchases and ratings. [4]

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.

## 1.3 Service Level Agreements (SLA)

According to *ITILv3's* official glossary [1], 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.

## 1.4 Web Services & SOA

# 2 Bibliographic Review

## 2.1 SLA Guided Data integration on data environments

## References

- [1] Axelos. Best management practice portfolio: common glossary of terms and definitions, October 2012.
- [2] A. et al. CONNOLY, D. FOX. A view of cloud computing, April 2010. [Online; posted 01-April-2010].
- [3] DB Ranking. Db engines ranking, May 2014.
- [4] Genoveva Vargas Solar. Addressing data management on the cloud: tackling the big data challenges, May 2014.