

NoSQL: The Death of the Star

Alberto Abelló*

*Omega-125 Campus Nord UPC
Jordi Girona 1-3, 08034 Barcelona (Spain)
aabello@essi.upc.edu,
<http://www.essi.upc.edu/aabello>

1 Introduction

Nowadays, most companies externalize as many services as possible to reduce costs and be more flexible in front of fluctuations of the demand. Thus, with the cloud, the time has arrived to IT infrastructures. The National Institute of Standards and Technology (NIST) defines *cloud computing* as “a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction”.

Cloud computing, in general, is a good solution for medium to small companies that cannot afford a huge initial investment in hardware together with an IT department to manage it. With this kind of technologies, they can pay per use, instead of provisioning for peak loads. Thus, only when the company grows up (if at all), so the expenses will. The only problem is that they have to trust their data to third parties.

In Abadi (2009), we find an analysis of pros and cons of data management in the cloud. It is found completely inappropriate for transactional processing mainly due to the problems to guarantee ACID properties in such environment. However, it is adequate for analysis environments, since those properties are not needed. It also outlines the problem of having data in an untrusted environment, which would again be unacceptable in transactional processing, but can be easily solved in analytical systems by just leaving out some sensitive data or using an anonymization function. On the other hand, what cloud data management can offer to an analytical environment is elastic compute power (in the form of parallelism), replication of data (even across different regions of the world), and fault tolerance (a new machine automatically taking over from a fallen one without re-executing the whole query or process).

Data Warehouses (DW) and On-Line Analytical Processing (OLAP) tools were defined by Bill Inmon in 1992 and Edgar Codd in 1993, respectively. Thus, they are almost twenty years old and have evolved to maturity by overcoming many limitations in these years. Huge (Terabytes) relational DW exist today benefiting from techniques like materialized views, bitmap indexes, etc. Nevertheless, some challenges remain still open. Mainly, they are related to the management of ETL processes, unstructured data, and schema evolution.

Cloud computing does not mean that we cannot use a relational system. Indeed, well known alliances already exist in the market, like that between Oracle and Amazon. However,