Introduction

Big data is a term used to describe the availability and exponential growth of data. It can be defined as “High volume, high velocity, high value, high veracity and high variety of information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization”. Decision support systems are interactive, computer-based systems that help users in judgment and choice activities. They provide data retrieval and storage but enhance the traditional information access and retrieval functions with support for model building and model-based reasoning. They support framing, modeling, and problem solving.

YCSB

- The Yahoo! Cloud Serving Benchmark is a Big Data benchmark. The main goal that the YCSB developers make is to increase the variety of cloud serving systems. This benchmark wants to build tiers to evaluate performance, scalability, elasticity, availability, and replication.
- YCSB runs from the command line and it can create an arbitrary number of threads that will query the system under test. It will measure throughput in operations per second and record the latency in performing these operations.

TPC-H

- TPC-H is a benchmark that simulates a Decision Support System database environment. The transactions in such environments are characterized by business intelligence intensive complex data mining queries and concurrent data modifications.
- The benchmark is composed of 22 read-only queries and 2 update queries. These queries are performed on considerably large amounts of data, have a high degree of complexity, and were chosen to give answers to critical business questions.

Conclusion

Both Big Data and DSS systems have gained a lot of popularity in the last years and have become successful in many different ways. BigData is a hot topic. It not only deals with large amounts of data sets and the procedures and tools used to analyze and manipulate them, but also to a computational turn in research and thought. At the same time, Decision Support applications are related to Big Data as they need to deal with large datasets.

Figure 1 – Data Storage Elasticity

Figure 2 – TPC-H Database Schema

Pedro Furtado e-mail: pnf@dei.uc.pt