

Building Infrastructure for Data Management in the Cloud

Keywords: Cloud DBMS, Cloud Caching, Scientific Data Management

Problem: Data management in the cloud has received a lot of attention recently. This is due to the common idea that cloud computing is able to satisfy data service needs at lower operational costs. Additionally, with the advent of public archiving large volumes of scientific data from various disciplines (such as astronomy or biology), data is massively queried by large groups of scientists. In this context the cloud is “the perfect fit” infrastructure, capable to minimize operational costs while exploiting efficiently shared data and work across query requests.

Project: This project aims in implementing and deploying the software infrastructure for data management in the cloud. The hardware architecture of the cloud consists of 10 Solaris machines, each of them equipped with 8 virtual boxes, interconnected through a high speed network connection and coordinated by a controller. The controller interconnects with a Science Server, to access the scientific data from the backend databases, and with the Internet, to allow users interacting with the cloud. To execute a query, the user contacts the controller, which will further distribute it to the appropriate cloud node or to the backend database.

Plan:

1. Build the software infrastructure for the cloud computing environment capable to offer data services to users.
2. Implement the pricing model of the cloud DBMS proposed in [1] and deploy in it on top of the infrastructure. Evaluate and perform real experimentation. Moreover, study which are the real challenges in running queries in the cloud.
3. Evaluate performance of the cloud DBMS using scientific data gathered from the MICS [2] project.

Supervisor: Prof. Anastasia Ailamaki, anastasia.ailamaki@epfl.ch

References:

1. D. Dash, V. Kantere, A. Ailamaki, An Economic Model for Self-Tuned Cloud Caching, ICDE, 2009
2. www.mics.org