Supporting Multi Data Stores Applications in Cloud Environments

The production of huge amount of data and the emergence of cloud computing have introduced new requirements for data management. Many applications need to interact with several heterogeneous data stores depending on the type of data they have to manage: traditional data types, documents, graph data from social networks, simple key-value data, etc. Interacting with heterogeneous data models via different APIs, and multiple data store applications imposes challenging tasks to their developers. Indeed, programmers have to be familiar with different APIs. In addition, the execution of complex queries over heterogeneous data models cannot, currently, be achieved in a declarative way as it is used to be with mono-data store application, and therefore requires extra implementation efforts. Moreover, developers need to master and deal with the complex processes of cloud discovery, and application deployment and execution.

In this paper we propose an integrated set of models, algorithms and tools aiming at alleviating developers task for developing, deploying and migrating multiple data stores applications in cloud environments. Our approach focuses mainly on three points. First, we provide a unifying data model used by applications developers to interact with heterogeneous relational and NoSQL data stores. Based on that, they express queries using OPEN-PaaS-Database API (ODBAPI), unique REST API allowing programmers to write their applications code independently of the target data stores. Second, we propose virtual data stores, which act as a mediator and interact with integrated data stores wrapped by ODBAPI. This run-time component supports the execution of single and complex queries over heterogeneous data stores. Finally, we present a declarative approach that enables to lighten the burden of the tedious and non-standard tasks of (1) discovering relevant cloud environment and (2) deploying applications on them while letting developers to simply focus on specifying their storage and computing requirements. A prototype of the proposed solution has been developed and is currently used to implement use cases from the Open PaaS project.