

Draft proposal

One query-multiple databases : An interface to mutually access Hadoop/Oracle databases

Arjun Datt Sharma(asharm24), Arnab Saha(asaha), Nishtha Garg(ngarg)

Department of Computer Science, North Carolina State University

Abstract

In today's digital ecosystem, where a number of technologies to maintain and analyze the huge amounts of data have evolved, querying and analyzing data from multiple data repositories with different data mechanics is challenging; esp. when one is RDBM system(like Oracle) and the other is NoSQL system(like Hadoop). For instance, if a company (which uses Hadoop) expands its product into a new location, by acquiring a similar product by another company (which uses Oracle), then processing the entire data using a single interface is an intricate task.

In this project, we intend to solve the aforementioned problem in two steps. As the databases pertain to the same product, they have similar information but not necessarily similar structure. So we aim at providing a mapping for all the similar attributes in both datasets. Secondly we aim at creating a common "SQL-like" API to access data from both datasets simultaneously. The ultimate goal is to provide a cost effective alternative to the approach of migrating the entire dataset from one technology to another.