**Project Management** —

**Goal(s)** — The ultimate goal of the project is to create a common SQL interface to access different types of databases. We have divided the project into two phases:

1. **Single Query Interface:** This will have the look and feel similar to any SQL interface where you can run a Select/Update query but it is capable of fetching and updating data simultaneously from disparate databases like HBase and Oracle.
2. **Schema unification:** The main challenge in using different databases to understand what an attribute in say database1 correspond to in say database2. So we are planning to develop the Schema Unification Engine which will automatically map the attributes from one database to another. For this step we are assuming that the data in both the databases are similar and therefore consist of similar attributes.

**Milestones/Tasks** —

1. **Learning and setup phase:** This is the fundamental phase which will pave the path for the development of the project. This phase consists of learning about the architecture and SQLs of the different DBs and testing it on their native environment. *[Duration: 1 week, Depends on None]*
2. **Sample data creation:** Simple testing data needs to be created on different DBs before development which will have similar schemas. *[Duration: 1 week, Depends on i]*
3. **Manual mapping:** Though the final step in the development will be to achieve automation in schema unification, we will initially provide the attribute mapping manually and store it in the meta database. [Duration: 3 weeks, Depends on ii]
4. **Query interface creation:** This step marks the end of Phase I. At the end of this phase we will be able to fetch/update data from multiple databases with the same query. The data though spread across various databases will behave like sharded data on single database. *[Duration: 3 weeks parallel with iii, Depends on ii]*
5. **Automated mapping:** The principal part of the project is to develop Schema unification Engine which will automate the mapping of the attributes across diverse databases. *[Duration: 1 month 1 week, Depends on iv]*
6. **Testing and efficiency improvement:** Our intention is to do these two steps in parallel to improve our efficiency. While our primary focus is to improve the efficiency of the schema unification engine, we plan to reduce the bugs in the latest version. *[Duration: 1 month 1 week parallel with v, Depends on v]*

**Verification** —

The first phase of the project is to develop Single Query Interface using manual mapping. We will test it with five test data ranging from simple schemas to complex schemas (We are yet to decide on the structure of the schemas). Once it runs successfully on the five test data we will move on to the second part of the project.

The second phase of the project deals with automated mapping across different databases. As domain profiling and mapping is not foolproof, we are not certain about the efficiency we might be able to achieve. Like most real world problems, absolute optimality in automated mapping might not be achieved thereby we are applying Pareto optimality for the requirements.