**Progress Report and Updated Milestones**

We have split the project into 4 concrete milestones:

|  |
| --- |
| **Milestone** |
| 1. Backend Software Stack    1. Research and choose software stack **DONE**    2. Build and Configure Hadoop, Hive, Shark, Spark, and BlinkDB **DONE**    3. Load data and test BlinkDB **DONE** |
| 1. Hierarchical Clustering    1. Determine method and metric for hierarchical clustering **DONE**    2. Prototype clustering in Map Reduce **DONE**    3. Dealing with multi-variate data (different ranges, counts, categorical data) **PARTIAL** |
| 1. Visualization techniques    1. Choose appropriate visualization techniques **DONE**    2. Implement visualization interface **DONE**    3. Scale the visualization to multi-dimensional data **DONE**    4. Zoom in/out procedure **PARTIAL** |
| 1. Glue all of the parts together    1. Integrate clustering into BlinkDB **PARTIAL**    2. Build interface from BlinkDB to front end visualization    3. Final refinements to    4. Scale from 12GB data set to multi TB data set |

Additional Notes: