**PROJECT PROPOSAL: Prediction Models for CPU and I/O Usage of Queries in Teradata**

Kritika Versha, Vishal Joshi

# **Problem Statement:**

As the load on databases vary from time to time, the performance of the resulting activity changes accordingly. Performance fluctuations depend on many parameters such as the change in the ratio of transactions, the time of transaction processing, the types of queries, and the complex interactions between various transactions. Identifying resource consumption on a per transaction basis is crucial. If this consumption can be predicted beforehand, it is possible to optimize the utilization of resources in the database. We find a solution to this by developing models using machine learning to identify crucial parameters that predict the usage of CPU and I/O queries.

## Proposed Approach and Milestones:

1. Use the Teradata DBQLogTable to fetch various parameters of the query logs.
2. Identify machine learning models (ex. SVM, Multinomial Naive Bayes, Linear Regression) which would work for training the load data.
3. Extract and test features to identify which of them works best for prediction models.
4. Use clustering models to categorize the queries.
5. Test the accuracy of models using ROC curves and benchmarks such as TPC-DS.
6. Create of API/CLI to give prediction output.