ECE 454 Assignment 3 Design Doc

# Part 1

The procedure is simple. Part 1 consists of one mapper. It finds the genes with highest expression rate, and store it in an array list. Then the array list is passed back to the main Hadoop.

# Part 2

# Part 3

# Part 4

## Part 1

First, input is read and put into the bags. Then it is passed to custom UDF, where it passes a string of genes with highest expression. Then the query is written to the output file.

## Part 2

Similar to part 4 part 1, the input is read and put into the bags. Then it is passed to custom UDF. Then the query is flattened into genes. Then, using UDF, the values are calculated and written to the output file.

## Part 3