Big Data Project Report

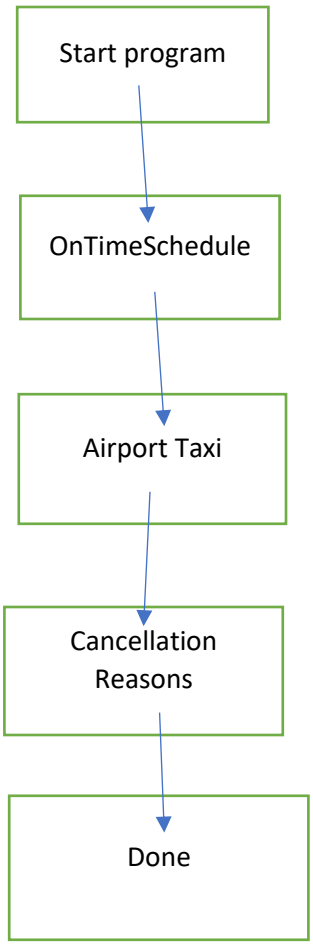
CS644

Professor: Chase Wu

Student: Zibin Guan

Minjuan Zhang

1. Structure of Ozzie Workflow



2. Algorithm

   On time schedule airline:

   Mapper:

   Reading the data line by line inside the mapper function and less than 20 minutes delay, we will consider that flight is on time.

   Reducer:

   Read the data from mapper, and then calculate the probability for each airline to find out which airline delay the most or less.

   Probability function: delay/total flight

   The cleanup function is doing the sorting and find out the three highest and lowest probability.

   And then we can find out what are the airlines.

Taxi time:

Mapper:

Read the data line by line, get the data from 20 and 21 columns for taxi in and taxi out.

Reducer:

Like on time schedule function, the function for taxi is normal on time/ average taxitime of each airport

Cleanup function will sort the list and output the 3 longest and shortest taxi time of the airport

Cancellation:

Mapper:

Read the data, get the data from column 23 which about the cancellation reasons.

Reducer:

Sum up all the reason by the keys which totally have 4.

Clean up function helps us to sort the data and then output the most common reason

3.