

# Project: Flight Data Analysis

Group Members: Nilesh Samnani

Mushir Shaikh

## Algorithm:

### i. Delay in flight:

- In mapper class, we have considered arrival delay with threshold 5 min.
- We have bifurcated the data into 2 parts based on threshold. We have mapped the entries with delay time more than threshold to value 0 and entries having delay time less than threshold to value 1.
- The output of mapper would be having 2 keys (0,1) and the entries with their respective category.
- In reducer, we have sum of all values under same key airline. And find the probability of on being scheduled or not.
- Then, in cleanup function, we've sort them base on probability and commit highest 3 and lowest 3 values for on being scheduled.

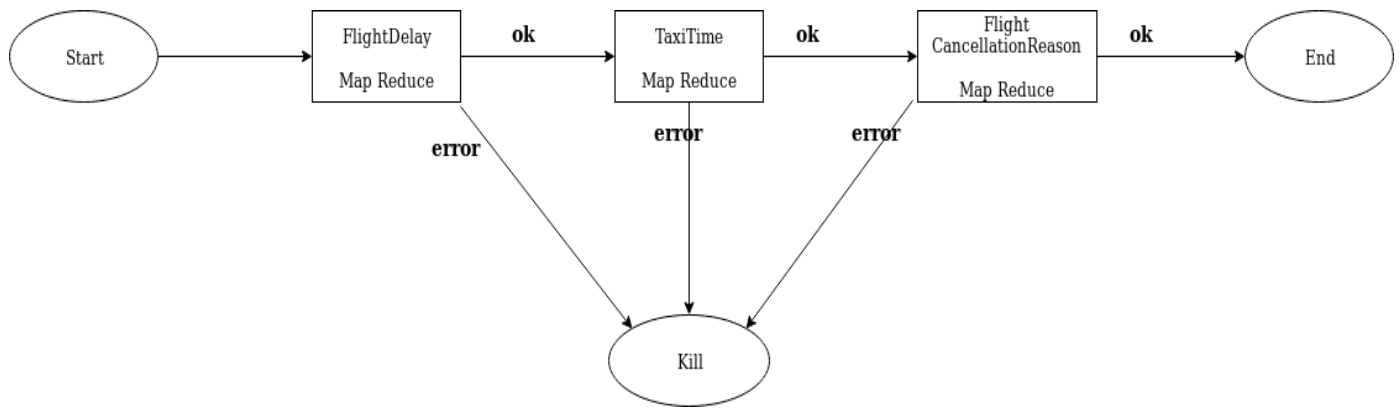
### ii. Taxi Time for flight:

- We have used one mapper and one reducer for this.
- In Mapper class we have first filter the entries having value in 4 columns(Origin, Dest, Taxi In, Taxi Out) and commit the entries in the following 2 ways.
- 1 (Origin, Taxi Out)
- 2 (Dest, Taxi In)
- In reducer, we have averaged taxi time by counting total taxi time and divide it with total count value.
- In cleanup function, we've sort them based on avg taxi time and then commit longest 3 and shortest 3 values of avg taxi time with airport.

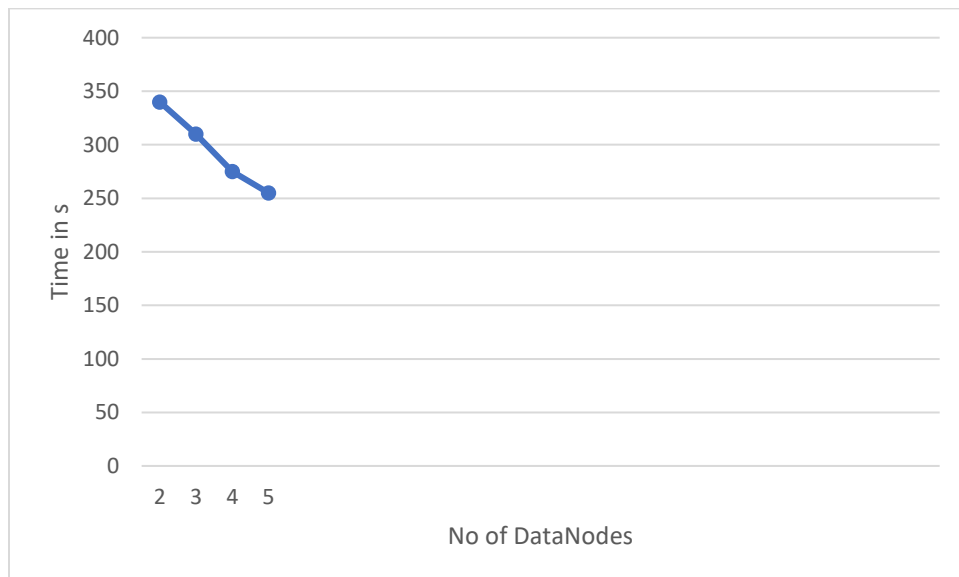
### iii. Reason for flight cancellation:

- We've used one mapper and one reducer for this.
- In mapper class, we've first read the file and filter entries having cancelled value and code. And then we've committed that entry with value (cancellation\_code, 1).
- In reducer class, we've just calculated the sum for each cancellation code and committed only code with maximum sum.

## Oozie Workflow

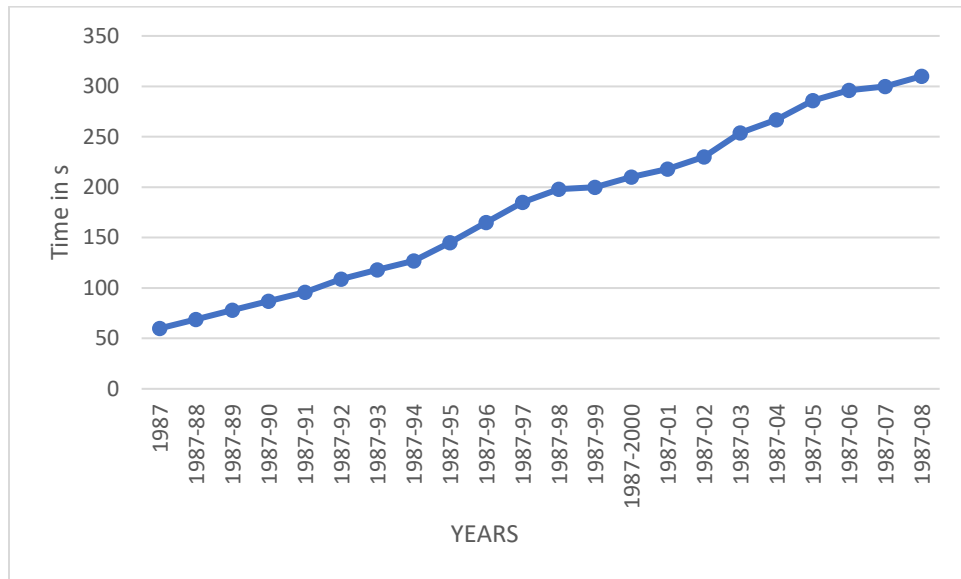


### Performance by increasing the instances



We have examined that as the number of instances of data node is increased the execution time decreases.

### Performance by increasing data



In our project we have observed that as the size of data of increasing the time required for execution is also increases.