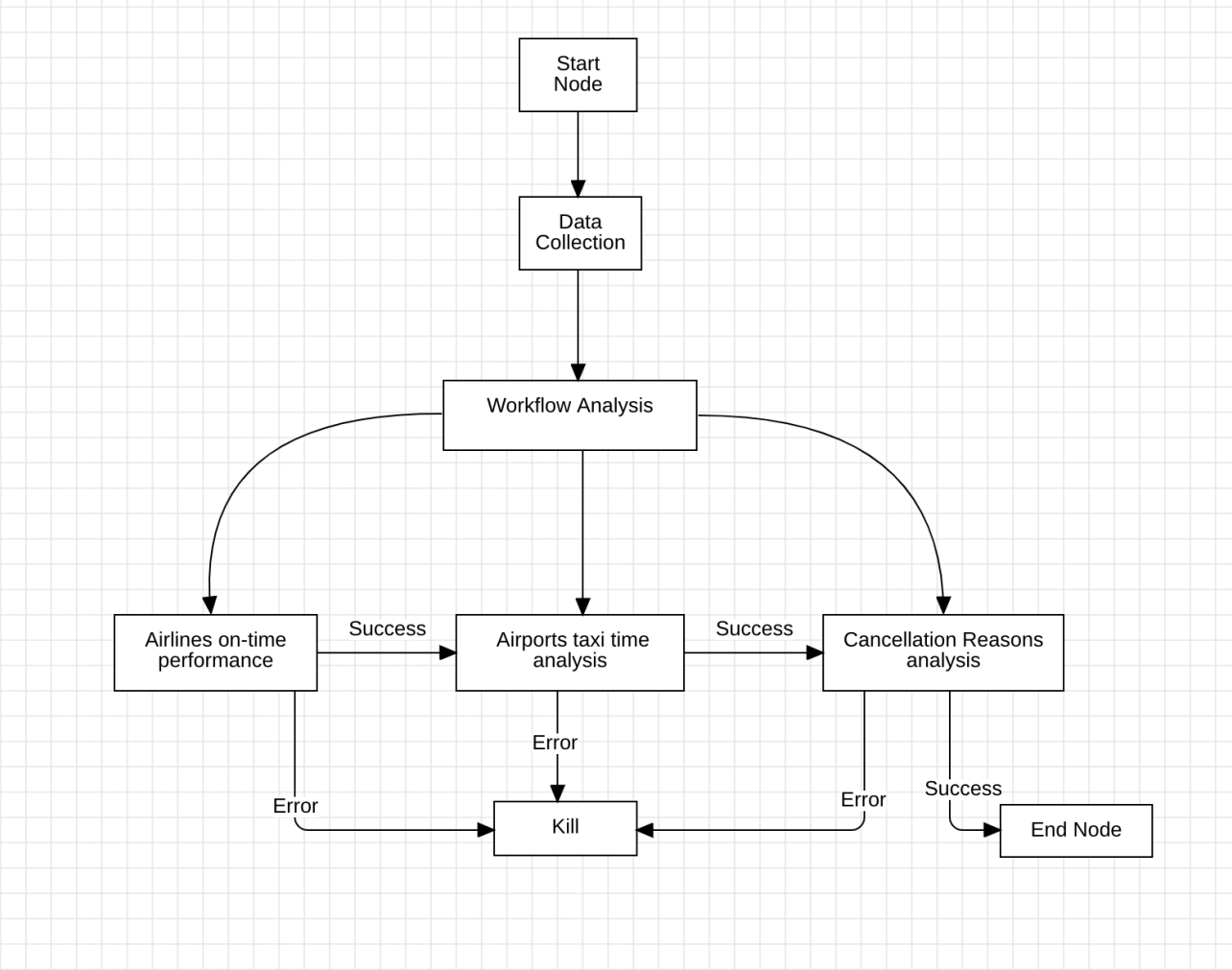
Flight Data Analysis - Ozzie Workflow

Milestone 1-DS644-004

Team
Udeda.Suchithra - Su252
Sanjana Akula - Sa2844
Alagani Sai Krupesh Goud - Sa2834

1.A diagram that shows the structure of Oozie workflow



2.A detailed description of the algorithm you designed to solve each of the problems

a) Airline on-time performance

1.Mapper
- Key-value pairs, with the airline code as the key and the on-time indicator as a tuple (1 for on-time, 0 for delayed), should be sent for each flight record.

2.Reducer
- Compute each airline's total number of flights and total number of on-time flights.

- Calculate the on-time percentage (total on-time flights / total flights) for each airline.

3.Sorting
- The airlines are sorted according to their on-time %; the highest likelihood airlines are found by sorting them in ascending order, while the lowest probability airlines are found by sorting them in descending order.

b)Airports taxi time analysis

1.Mapper
- Emit key-value pairs for every flight record, where the airport code is the key and the value is a tuple with the taxi-in and taxi-out times.

2.Combining
- To maximise efficiency, combine intermediate records.

3.Reducer
- Add up all of the taxi arrival and departure times for every airport.
- Determine how many flights there are in total for each airport.
- Calculate the average taxi time (total taxi time / total flights) for each airport.

4.Sorting
- To determine which airports have the longest average taxi times per flight, sort the airports based on their average taxi times in descending order; to find the airports with the lowest average taxi times per flight, sort them in ascending order.

c)Cancellation reasons analysis

1.Mapper
- Emit key-value pairs, where the key is the cause for the cancellation and the value is 1, for each flight record that has been marked as cancelled.

2.Combining
- To maximise efficiency, combine intermediate records.

3.Reducer
- Add up all of the cancellation reasons.
- Find the cancellation reason that has the largest number.