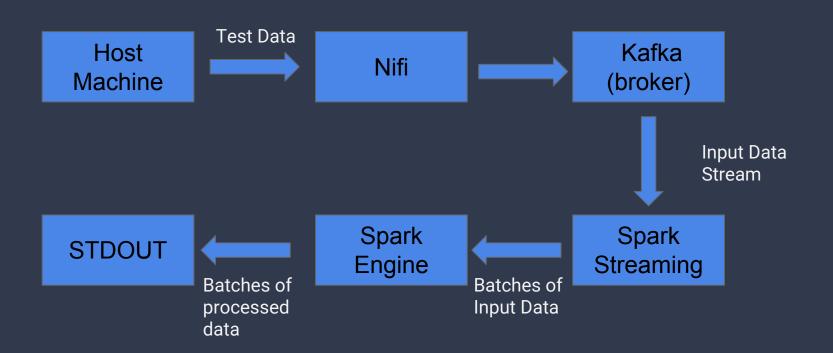# E-COMMERCE DELIVERY

Team: KnightFighters
Members:
1) Anshul Kumar   ME15B082
2) Ashutosh Raj   ME15B086
3) Asit Tarsode   ME15B087

PROCESS FLOWCHART

Host Machine → Test Data → Nifi → Kafka (broker)

Kafka (broker) → Input Data Stream → Spark Streaming

Spark Streaming → Batches of Input Data → Spark Engine

Spark Engine → Batches of processed data → STDOUT

Pre-processing:

<u>Methods used:</u>

- Handling of null data points
- Handling columns with date- time format
- One-hot encoding of categorical data

Inferences:

Random Forests gave the best accuracy

Conclusion:

An area under ROC curve of 70% shows that it is possible to get an approximate prediction of when the order is not going to be delivered on time. Thus it becomes easy to modify the expected delivery date for the customer or take measures to expedite failing deliveries.

THANK YOU!