

# Real-Time Fraud Detection System

Mukunth Rajendran, Rashmi Khurana, Suprajah Suresh

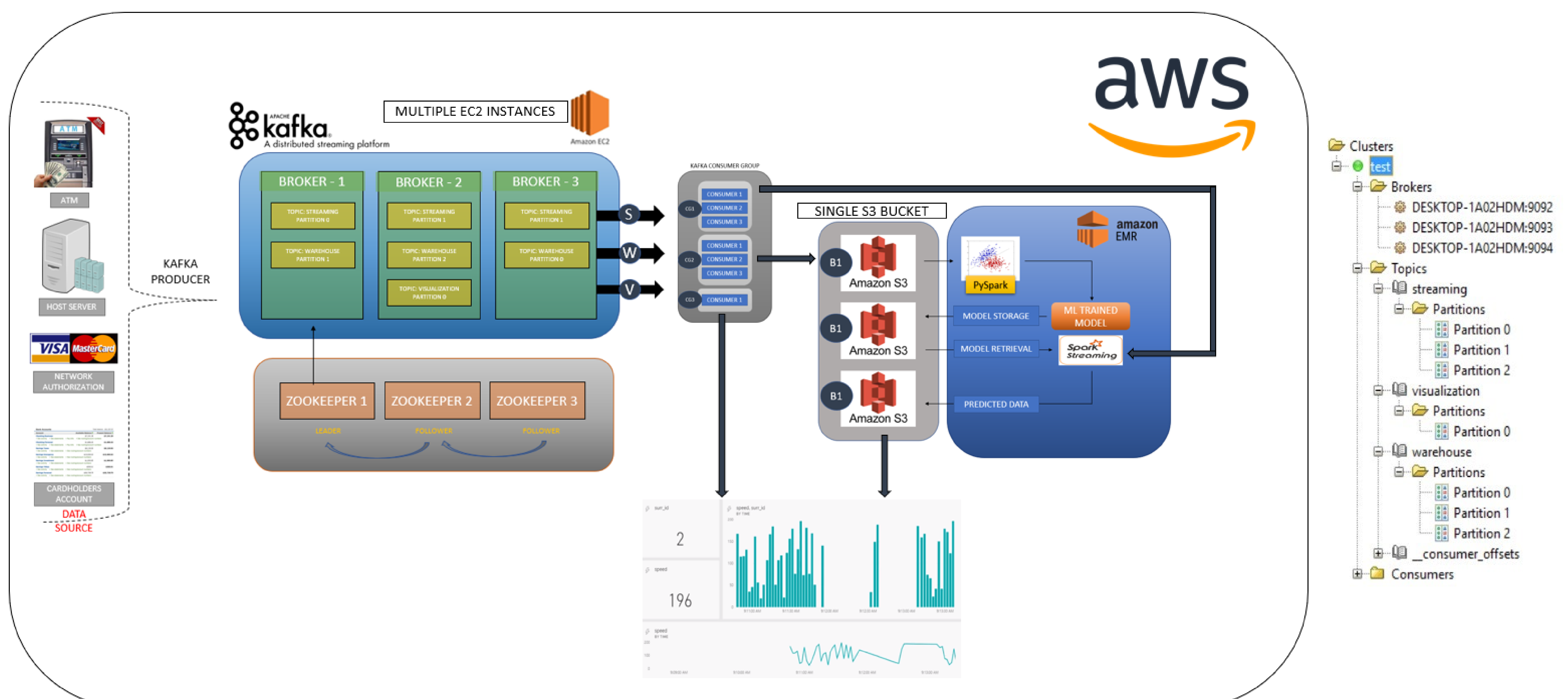
Instructor: Dr. David Belanger

## Objective:

Recognize fraudulent credit card transactions in real-time so that customers are not charged for unauthorized transactions. Instantaneous fraud detection is required to stop transaction completion. To build a scalable, multi-source stream processing system with parallel processing capabilities to predict fraudulent transactions, to provide system information and to deliver visualization of data.

## Data Pipeline:

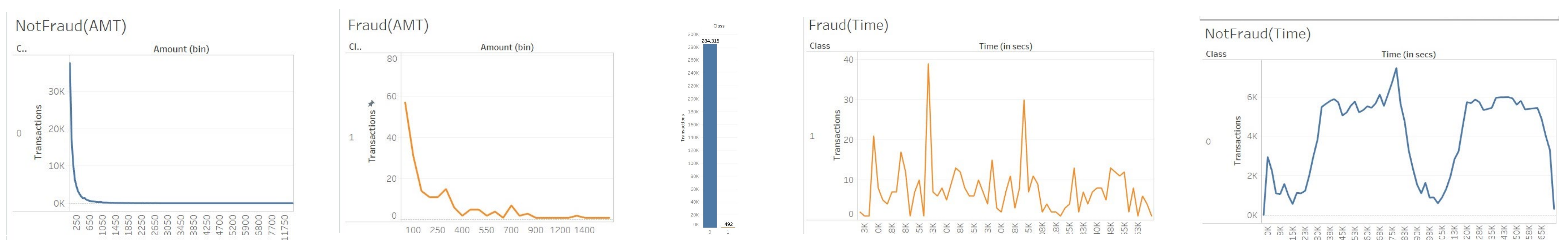
The system is built using Amazon web services to host the components on the cloud. EC2 instances are used to run a multi-node, multi-cluster Kafka structure. AWS S3 is used for storage of the streamed data, trained model and to run Spark SQL queries. Spark programs are run on Amazon EMR to predict the incoming stream. The outcome is to provide notifications for fraudulent behavior.



## Modelling & Analysis:

- ❖ Warehoused data is used for model building. Many Machine learning algorithms are used and the model with best performance is selected.
- ❖ The dataset is highly unbalanced but does not require preprocessing.
- ❖ Extracted the best performing model using cross-validation and saved the model in S3.
- ❖ Model is read by the Spark Streaming program and predictions are made in real-time.
- ❖ The Real-Time data is used for visualization and for triggering alarm.

## Visualization:



## Future Scope:

- ❖ Use other processing frameworks like Apache Flink and Apache Storm to process data in real-time and Apache Beam to provide a unified model that performs preprocessing.
- ❖ Improve performance of the system using multithreading and reactive programming for further optimization..