NYC DoT (Open Data) Portal

Web Server

Data Ingestion

Local Spark Cluster

DoT_Traffic_Speeds_NBE.csv

Data Compression

Parquet

Compressed File Format

Local Machine (Python and Spark)

Databricks Community Edition

databricks

dbfs
Databricks File System

Data Cleaning and Preprocessing

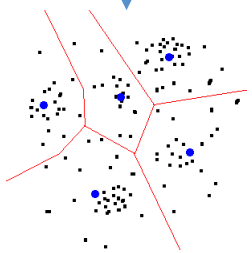
Data Visualization

Spark SQL

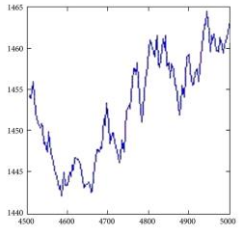
Spark MLlib

python + Spark

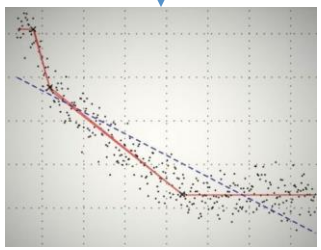
Spark + R



Clustering / PCA



Time-Series Analysis



Multivariate Adaptive Regression Splines