

Stock Anomaly Detection

Distributed Information Systems Workshop

Spring Semester 2025

Soham Roy Scott Barcomb

Introduction

- We used Apache Spark to perform real time data analysis on streaming data
- Data used - **TSLA** stock data via yfinance, ingested as live streaming data
- We used 3 key metrics to analyze and predict trends before suggesting an action
- Web App deployed on AWS
- Apache Spark application run on a three-node cluster of EC2 instances

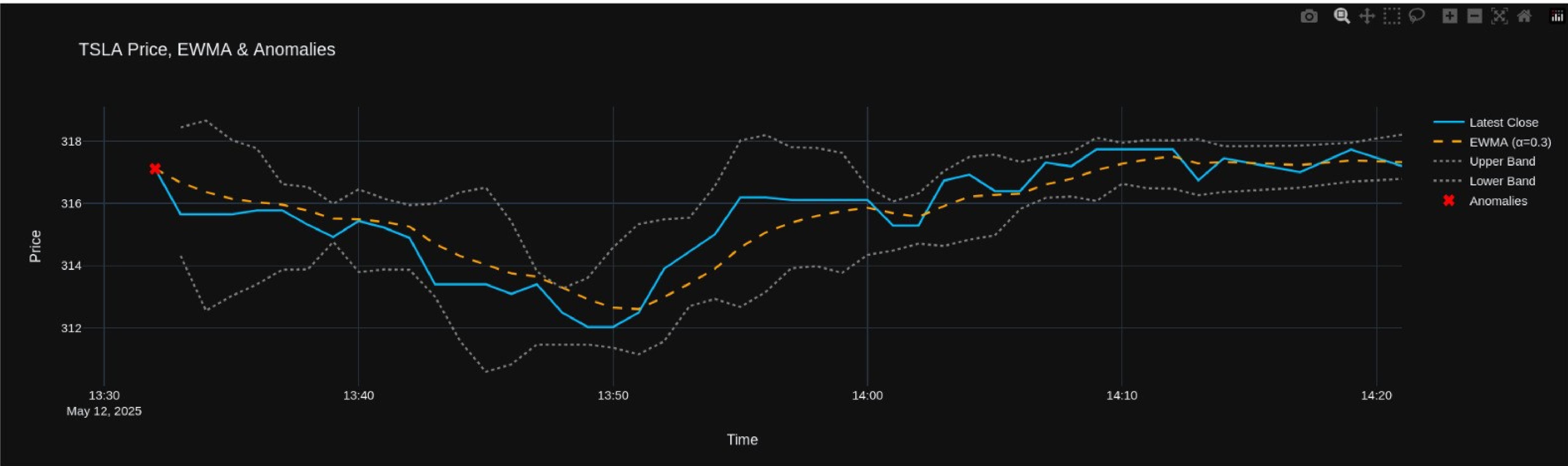
Analysis Metrics

- **EWMA** (Exponentially Weighted Moving Average) - *Detect price drift and capture short-to-mid-term trends.*
- **Z-Score** (Statistical Anomaly Detection) - *Identify outliers based on statistical deviation. $|Z| > 2$ indicates strong Anomaly*
- **Bollinger Bands** - *Assess Price Volatility - Price outside bands → unusual volatility*

Suggested Actions

- **Bullish** - A belief or signal that the stock **price will rise**. Optimism - investors may **buy** or **hold** the stock.
 - **Price > EWMA** (i.e., upward momentum)
 - Possibly approaching or exceeding the **upper Bollinger Band**
- **Bearish** - A belief or signal that the stock price will fall. Pessimism — investors may sell, avoid, or even short the stock.
 - **Price < EWMA** (downward trend)
 - **Z-score < -2** or nearing **lower Bollinger Band**

TSLA Anomaly Detection Dashboard



Suggested Action

Bearish: Price below EWMA. Consider selling, shorting, or avoiding the stock

Errors Callbacks | v3.0.4 | Server |

Dashboard displaying live analysis and suggesting action on the stock

Implementation

- Local → Cluster → Local pipeline
- The first script takes the locally stored dataset and feeds it to HDFS to mimic live data updates
- The second script is the Spark application run on the cluster
- The third script continuously reads the Spark application output on HDFS, displaying the data

Setup

- Three EC2 instances – one NameNode and two DataNodes
- HDFS – distributed file system
- Apache Hadoop YARN – resource negotiator
- Apache Spark – structured streamer