**Course:**

E63 (Big Data Analytics) Harvard Extension School - Spring 2016

**Professor:**

Prof. Zoran B. Djordjević

**Topic:**

Spark: Event Stream Processing and Spark mllib

**Purpose of the project:**

Read the live highway information feed provided by Massachusetts Department of Transportation. Send the live information from the feed to Kakfa cluster. Consume the processed feed with a windowed streaming Kafka consumer so that end user can see a graph of the average speed on Massachusetts highways for last N minutes (N should be configurable).

Also, predict the average speed on Massachusetts highways for last N/10 minutes based on the speed data of last N minutes. Show the predicted to the user on the same graph.

**Can the running instance of project be accessed on internet:**

Yes. Project is deployed on AWS and can be accessed at: **https://s3.amazonaws.com/e63-course-final-project/html/final\_project\_e63\_visualization.html**

**Data set:**

Massachusetts DoT live highway travel time feed ( <https://www.massdot.state.ma.us/feeds/traveltimes/RTTM_feed.aspx>) which is updated almost every minute.

More information about this feed is provided by MassDOT here: <http://www.massdot.state.ma.us/DevelopersData.aspx>

Challenges:

Keeping the application running continuously all the time.

**Real world use of the project:**

The tool that comes out of this project can be used by law enforcement officers, town planners, Department of Transport personnel in any situation/decision that needs live speed data on the highways.

**Potential future enhancements to the project:**

The time period for which average speed data is shown to the user, can be made configurable on the UI so that end user can change the time period in turn making the tool fully customizable.

More live information about the highways (like accidents, warnings, construction work) can also be shown on the graph.