- The subject of the tech talk you'd like to give (we'll try to allocate these in class so that we can get a good distribution of them, avoid topic conflicts, etc.)

I would like to give a tech talk about Spark. Since I am a data scientist, there are advanced topics that I can discuss in Spark that have not been addressed in the class.

- A list of any areas you'd like to see the instructors cover in class that weren't already mentioned in lecture (we may not be able to accommodate everything given our time budget, but we'll do our best)

I would like Kafka, Cassandra, and HBase covered as well as an advanced discussion on Spark Streaming

- A link to the URL for the GitHub project that will host your term project work

https://github.com/michalmonselise/scala_project

- A more fleshed out version of your project proposal that we can discuss in week 2.

My project will be a web application that maps tweet activity by keyword. The user will enter a keyword and the output will be a world map with a heat map over the map that will indicate the tweet activity. There will be multiple analytics options regarding the keyword.

My project will use Kafka for retrieving the tweets, Spark streaming for aggregating the tweets, HBase for storing the tweets. My web page will then retrieve the data from HBase and will generate a map using GeoTrellis.

I intend to divide the work into 3 parts: a data extraction pipeline, a web UI and a GeoTrellis application to plot the data.

The data extraction pipeline code will use the actor model to stream the tweets. Then the data will move to an HBase store. From there I will have a spark streaming application. The application will pick up the difference in rows and use that to perform analytics.

The web UI will get the data from Spark and display it.

Finally, I will use the geotrellis package to generate visualizations.