Capstone Project- Data Pipeline for Real-time Weather Analysis Using Strom Glass API and D3 Data Visualization

Tristan Pacba

Description

The capstone project will be imitating the Southern California Edison Company and its mission to predict wildfire threats and red flag warnings[[1]](#footnote-1). By ingesting real-time weather data, including wind speed and wind gust, temperature, humidity and solar data, provided by Storm Glass API[[2]](#footnote-2), the capstone project will use data visualization to analyze trends and predict projections that can help determine helpful decisions for the company like switch power grids or place new power lines.

Design

1. Extract real-time data from Storm Glass API using Kafka
2. Transform internal analytical tables in Hive using Spark SQL
3. Load into MongoDB (possibly GCP, depending on cost)
4. Schedule data-pipeline using Airflow
5. Create data visualization using D3.js

Benefits

With increasing climate variance, even in areas where weather is generally steady, real-time weather data is crucial for accurate and relevant reporting. Private weather forecasting is a growing billion dollar industry and companies, such as AccuWeather, Earth Networks, and the Weather Co., are producing their own big data and using business-savvy analytics to solve specific real-world problems[[3]](#footnote-3). They could warn railroad company to avoid a tornado barreling toward a specific stretch of track, advise a farmer when to irrigate a particular row of crops, or determine flight scheduling to avoid severe thunderstorms.

1. <https://www.sce.com/wildfire/fire-weather> [↑](#footnote-ref-1)
2. <https://stormglass.io/> [↑](#footnote-ref-2)
3. <https://www.washingtonpost.com/business/2019/11/25/weather-is-big-business-its-veering-toward-collision-with-federal-government/> [↑](#footnote-ref-3)