Data Intensive Computing - Project Proposal

Kai Fleischman & Martin Schuck

September 25, 2020

1 Introduction

Tackling crisis of global scale is an huge problem in today's societies. It seems that humans have a hard time at grasping phenomena which exceed their regular event horizon. This was particularly evident during the spread of the recent Corona pandemic. However, while the Corona crisis appears to be under control, an even greater thread looms in the coming years. Global warming has been one of the gravest and yet most abstract dangers to human society in recent decades. Especially now where attention has shifted to seemingly more pressing matters it is of utmost importance to remind everyone of this thread. Since graphic visualizations and daily updates proved to be most effective in getting to the public during the pandemic, we propose a real time global warming monitor as a warning. The project aims to create a world map which shows the temperature difference to the common global reference in real time.

2 Data

We plan to establish a base temperature model for each day of the year with the global climate history from the Berkeley earth dataset. Since the monitor is supposed to show real time developments, we utilize the real time weather API **XXX**.

3 Tools

The project requires the composition of a variety of tools:

- Data acquisition: Docker to achieve sufficient API requests per minute.
- Data processing: Kafka and SparkStreaming for live data processing.
- Data interpolation: MLLib to create a global interpolation model.
- Web presentation: Flask/JS for an appealing visualization.

4 Methodology

The base line temperature model will be created from averaged historical climate data samples. In order to create a truly global model from our samples, we intend to employ interpolation or regression models from MLLib. Likewise, the current temperature model is created with an 24 hour window average of the most recently streamed weather data. Our API promises to be capable of capturing more than 4000 data points around the globe on an hourly basis. By constantly updating the current temperature interpolation and calculating the deviation from the base line model around the globe, our web visualization should be able to provide a real time global warming monitor.