Data Analysis and Green Computing: Profiling HPC Power and Tracking CO2 Emissions

Green Team Report

Supervisor: Mohsen Seyedkazemi Ardebili

April 2024

Contents

Abstract

The following report includes the observations related to our work, the goal of which is profiling the HPC (High Performance Computing) power consumption and CO2 emissions in the scope of Laboratory of Big Data course.

Contributors:
Davletiyarov Baibek
Liberatoscioli Martina
Squarzoni Lorenzo
Zavaroni Sofia

1 Introduction

introduce background info, general information about the project; the schematic representation of the computer, where and how the racks are located

2 Tools

Tools - tools that were used and what techniques/libraries were used to prepare dataset etc; 2.1. preparation of dataset: the interpolation, sorting etc; 2.2. Meta Prophet? 2.3. STL

To start, a brief introduction to the tools and methods employed for data analysis is provided.

2.1 Dataset Preparation

2.2 STL (Seasonal-Trend Decomposition)

STL (Seasonal-Trend Decomposition) is a technique used in time series analysis to break down a dataset into three main components: trend, seasonality, and remainder.
1. **Trend**: This component captures the long-term direction or progression of the data, indicating whether it is increasing, decreasing, or staying constant over time.
2. **Seasonality**: Seasonality refers to the periodic patterns or fluctuations that occur at regular intervals within the data, such as daily, weekly, or yearly cycles.
3. **Remainder**: The remainder component represents the variability in the data that cannot be explained by the trend or seasonality. It captures the random fluctuations or noise present in the data.
By decomposing the time series data into these components, STL allows to understand the underlying patterns and structures, making it easier to analyze and model the data effectively.

2.3 Prophet by Meta

Prophet[1] is a forecasting tool developed by Facebook, designed to handle time series data with seasonal patterns and uncertainty. Its key points are:
1. **Automatic Seasonality Detection**: Prophet automatically detects seasonal patterns in the data, making it suitable for datasets with irregular or changing seasonalities.
2. **Trend Flexibility**: Users can specify various components of the time series, including holidays and special events, which are incorporated into the forecasting model.
3. **Uncertainty Estimation**: Prophet provides uncertainty intervals around the forecasted values, helping users understand the range of possible outcomes.
4. **Scalability**: It's designed to be scalable and can efficiently handle large datasets.
In summary, Prophet simplifies the time series forecasting process with an intuitive interface and powerful forecasting capabilities, making it suitable for both beginners and experienced analysts.

Prophet automates seasonal modeling for forecasting, while STL offers manual control for detailed trend analysis. Choose based on data complexity and desired level of control.

3 Carbon Intensity

The distribution of avaialble data is in the range of 3 years from January 2021 to January 2024. Analysis of the data demonstrates the seasonality, which makes the data generally predictable. The available data includes the carbon intensity of LCA and Direct. elaborate LCA/Direct

4. Carbon Intensity 4.1. the outputs of raw data + plots? 4.2. trying to predict the cabon intensity? 4.2. discussion - what was observed?

3.1 3.1

3.2 Discussion

The STL decomposition as well as Prophet demonstrate the seasonality in the dataset: the decrease in Carbon Intensity occurs in the warmer monthts as well as the higher values of Carbon Intensity correspond to colder months, which is in line with the active usage of heating equipment.

4 Power and Energy Consumption

4.1 Data Visualization

graph

STL - seasonality? Meta - seasonality?

4.2 COP - Operational Carbon Footprint Calculation

4.3 Discussion

The available data spans in the range of April 2020 to October 2022 (Fig#). In terms of the power consumption, the peak can be defined as 1.3 MW and the mean approximately 0.8 MW. Based on the power consumption data, the energy consumption plot was also build. Moreover, the excess peaks and drops in the power and energy plots can be described as a testing of maximum computational power and maintenance, respectively.
The fluctuations of data does not demonstrate the pattern of seasonality visually, which is further confirmed with the STL analysis tool fig, which also did not demonstrate the seasonality pattern of energy consumption.

However, the separate regions of the graph exhibit constant trends during small periods of time, for example July-October period. This potentially can be related to the external conditions as outside temperature depending on the season of the year. Particularly, the cooling may vary depending on these external conditions, which can lead to either increased or decreased energy consumption, for example, the hotter periods require more energy for cooling. Moreover, the general increases of energy consumption can be related to the performing the computations, which leads to higher activity of the node and increased power usage, depending on the heaviness of the task.

5 Conclusion?

???

References

[1] A. Author. "Title of the Article". In: *Journal Name* Volume.Issue (Year), Page range. DOI: `DOI`. URL: `URL`.