Optimization of Sustainable Energy Systems through Data Engineering

A Technical Report submitted to the Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

McKayla Thomas
Spring, 2022
Technical Project Team Members
McKayla Thomas

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Briana Morrison, Department of Computer Science

ABSTRACT
Arbo, a company specializing in transforming energy regulatory data into business intelligence, needed a more convenient way to access up-to-date information about energy generation in the U.S. As a data and software engineer intern, I developed a website that uses a custom API to access a singular dataset. The skills I used mainly consisted of React (JavaScript), Django(Python), Docker, and MySQL. The final goal of the project is to consolidate EIA data so it can be accessed via API and interacted with through an additional platform. The project design allows for adaptability so there would not be much difficulty to make edits. The next phase is to alter the Frontend to fit accessibility standards and test on existing customers.

1. INTRODUCTION

During the summer of 2022, I interned as a data and software engineer at Arbo, a company that specializes in transforming energy regulatory data into business intelligence and insights driving infrastructure permitting and commerce decisions. I was given several tasks related to the status of sustainable energy generation around the US. Arbo utilizes government data to assist customers in energy related projects such pipelines, buying and selling generators, and new sustainable energy projects.

Unfortunately, government documents are usually scattered and vary in jargon. This makes it difficult to collect and utilize the data for any type of project. For instance, the EIA has yearly filings (860A/860B) of the status of the energy generators throughout the US. The keywords and codes may have different meanings based on the year of the filing. This is an inconvenience to any customer looking for up to date information about the energy generation.

The accessibility and organization of the data allows customers to visualize the possible usage of the dataset. Developing a website that uses a custom API to access a singular dataset provides the accessibility and organization needed.

2. RELATED WORKS

The EIA website (2022) provided the filings needed for downloading to my website. Each downloaded file contained a zip folder. In order to obtain the zip, I needed to create a function that used the requests module.

I consulted Django.com (2022) for information about the framework, which I used to create the backend support for the models outlined in the 860 filings. For example the generator model fields represent the corresponding sections listed in the document. To protect the local machine and the existing code base, I used Docker containers.

3. PROCESS DESIGN

With most projects, the engineer cannot just begin creating the project. The first step is to plan. My plan began with familiarizing myself with the existing architecture of the data warehouse. At the time I did not know much about docker so I began researching and reading the documentation. This included using Docker preliminary applications. I did tutorials centered around creating docker containers and finding out how they are customized. Once I got comfortable with docker, I created a container to allow me to create Django models without having to worry about alternating the functional aspects of the warehouse.

Next, I needed to understand the structure of the EIA filings I would be storing. This included hours of research into what many of the codes and jargon meant. Following this research, I focused my attention on creating a graphical representation of how each model would interact with other models.

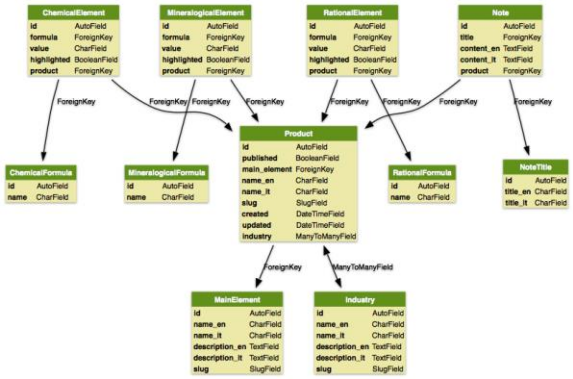

Figure 1: Model Graph

My graph was similar to Figure 1 but far more complicated. Some of the models contained foreign keys, which are used for many-to-one relationships. For example, if one has a reporter model and an article model and every article has an associated report, it is possible to reference the reporter model via foreign key. In my case, the reporter model would be the generator model. The generator model was associated with many of the other models.

Once I understood how the models were related, I had to create a data structure to house all the preexisting codes defined by the EIA. The data structure that associated the codes with their definitions then were set as fields for the related model. This prevents any issues regarding codes not being interpreted correctly.

The next step I needed to complete was the handler extractor and transformer. The responsibility of the handler was to call the extractor and transformer based on a schedule. For instance, the EIA filings were updated monthly and yearly so the handler needed to tell the extractor when to get that information from the EIA website and process and store it in the data warehouse. The extractor purpose was to extract the data from the zip files provided by the EIA. Once extracted, the data would then be passed to the transformer, which was responsible for transforming the data by parsing it and associating the models with the proper data. After these steps were completed, the API was able to pull this information to the frontend application so users would be able to interact with it as needed.

4. RESULTS

Every step of the development of this project was carefully planned, which led to highly successful results. The dataset was implemented on a separate platform dedicated to assisting customers with finding data on any active or inactive energy generations. This includes ownership information and capabilities for a specific generator. Before this application, companies would have to manually search through endless EIA filings.

This application also provides companies with access to accurate status reports on the generator of their choice. For example, if a company is interested in creating a power grid that uses a given generator, this information is now easily supplied to them.

5. CONCLUSION

My work for Arbo allowed me to use a variety of self-taught and university-taught techniques to concisely design a data structure and frontend integration. The project will assist in helping companies transition to sustainable energy production.

6. FUTURE WORK

The next phase will focus on accessibility standards and further usage. The accessibility aspects for this project will be making the frontend more visibility flexible. In addition, complete implementation through the other Saas projects. The design is adaptable, and can be used for many of the products offered by Arbo.

7. UVA EVALUATION

The UVA computer science program taught me critical methodologies like Agile. Also, I learned both Django and python which was the foundation of all my work at Arbo.

REFERENCES

Eia.gov. 2022. Form EIA-860 detailed data with previous form data (EIA-860A/860B). [online] Available at: <https://www.eia.gov/electricity/data/eia860/> [Accessed 23 September 2022].

Djangoproject.com. 2022. The web framework for perfectionists with deadlines | Django. [online] Available at: <https://www.djangoproject.com/> [Accessed 23 September 2022].

Docker documentation (2022) Docker Documentation. Available at: https://docs.docker.com/ (Accessed: November 24, 2022).