

CIS4400

Homework 1

Tatyana Titova
CIS 4400
March 31, 2025
Professor Jefferson Bien-Aime

Title: Hourly Electric Grid Monitor

Data Source: <https://www.eia.gov/electricity/gridmonitor/dashboard/custom/pending>

Business Requirements:

The goal for this project is to create a tool that can improve grid reliability and planning as well as allow room for analysis of the effectiveness of renewable resources.

The requirements are the following:

1. Monitor Real Time Regional Electricity Demand to Support Grid Reliability and Planning
 - a. This tool must help utility companies and grid operators in their daily day to day and also in cases of unusual activities.
 - b. The tool should be able to track how much energy people are using hour by hour by region and by energy source.
 - c. Comparison analysis should be easily done to see how well distributed the energy is across all regions.
 - d. Analysis must be done for spikes in demand and record the patterns before those spikes happen.
 - i. Is it seasonal?
 - ii. What time of day do the spikes happen?
 - iii. What day of the week do the spikes happen?
 - iv. Do some regions have more spikes than others?
 - e. How can we use this data to better prepare for spikes?
 - f. This tool must be able to use historical analysis of demand trends and generation trends for the purposes of forecasting usage and being better prepared to deliver consistent service.
2. Analyze the Impact of Renewable Energy Sources vs Non-renewable Energy Sources
 - a. This tool must be able to track hourly contributions of renewable resources as well as non-renewable resources.

- i. That includes solar, wind, hydro, nuclear, coal, fossil fuel
- b. This tool should be able to track the energy source and its impact on energy generation.
 - i. Which source is generating the most energy?
 - ii. Which source generation meets the demands the most frequent?
- c. Are renewable sources reliable?
 - i. How much are renewable sources fluctuating during the day? Week? Month? Year?
 - ii. How much are non-renewable sources fluctuating during the day? Week? Month? Year?
- d. This tool should help in analysis of reliability of renewable and non-renewable sources by region.

Functional Requirements:

1. Reliability and Planning
 - a. The system must be able to retrieve hourly electricity demand and generation via EIA API.
 - b. The system must be able to have filters and visualizations of usage demands by regions.
 - c. The system must be able to have filters and visualizations of generation of energy by regions.
 - d. The system must have the ability to generate forecasts that show which regions need more energy using patterns discovered in historical data.
2. Renewable Energy Analysis
 - a. The system must be able to retrieve hourly electricity demand by region via EIA API as well for this business requirement
 - b. The system must also be able to extract the hourly generation of electricity by energy source (renewable vs nonrenewable, so wind, solar, coal, fossil fuel, etc)
 - c. The system should be able to do comparison calculations.
 - i. How much energy was generated by renewable sources vs non-renewable sources out of the total.

Data Requirements:

Data Sourcing

This system is designed to monitor and forecast electricity demand with the purpose to analyze grid reliability and renewable energy performance. It can also be used as a tool to predict

electricity usage for the purposes of better planning, especially for times of spikes in usage, as previous patterns can be observed. I will be using a Web API for this project.

Information Architecture:

The Business Processes and Functions Supported:

- Monitoring real-time electric demand across US regions
- Monitoring real-time electric supply (generation) across US regions
- Monitoring real-time electric demand by energy source
- Monitoring real-time electric supply (generation) by energy source
- Forecast electricity demand based on historical data
- Forecast electricity spikes based on historical data

The Types of Analytics Needed:

- Tracking/Monitoring Analytics: tracking real-time demands and generations per region and per energy source
- Predictive Analytics: To maintain the grid effectively, we need to track historical data and predict electricity demand
- Comparative Analytics: The goal is to evaluate performance of renewable energy across regions vs traditional non-renewable sources

Types of Decisions Affected:

- This will effect possibly the day to day operations to ensure an effectively maintained grid and also for managing spikes in energy use
- Budgeting for energy production and distribution across the different regions
- This will also affect infrastructure investment decisions depending on the analysis of the renewable vs non-renewable sources

The People Who Have Access:

- The Data Engineers: The professionals who manage the pipelines and do all of the maintenance would need to have full access.
- Energy Data Analysts: These professionals would be able to use the data to analyze the historical trends and usage. They could generate reports/dashboards in regards to the effectiveness on the grid and also the performance of the different energy sources. This could help with research in renewable energy.
- Grid operators: They would be able to monitor real time demand data and monitor the grid stability. In addition, they will have access to reports, dashboards, alerts, live data in order to better understand the electric usage, especially during events like spikes and blackouts.

- Policy Makers: Having access to this data will help the people in charge make more informed choices on infrastructure decisions and investments with forecasts that use historical data.

The Data Source, Integration, and Consumption:

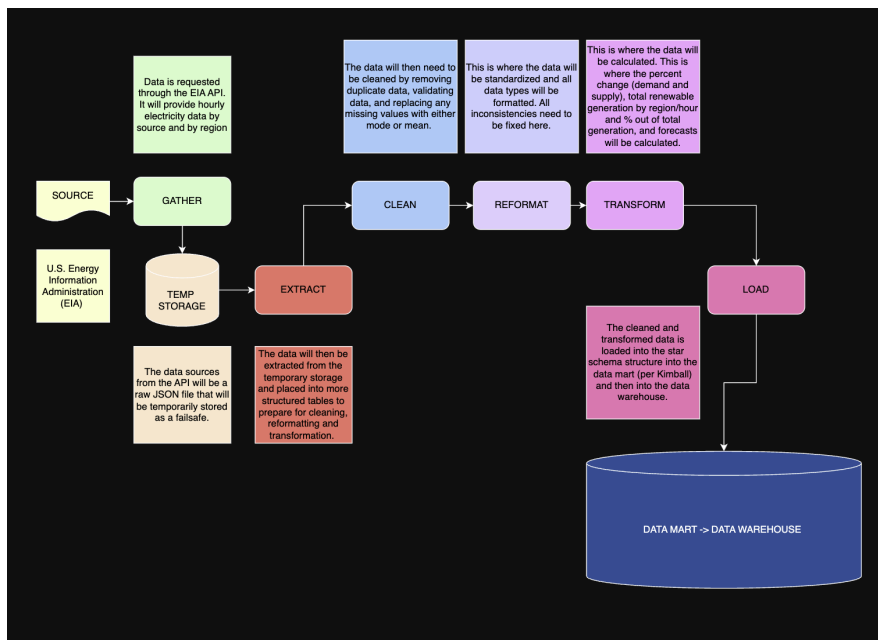
- Data Source: EIA Open Data API
- Integration: Data will be extracted using the API and moved through the data pipeline of gathering, storing, extracting, cleaning, reformatting, transforming, and loading into the data warehouse
- Consumption: The data will be presented using BI tools, PowerBI to be specific, that will support real time monitoring of data and also historical data for forecasting.

Information Flow:

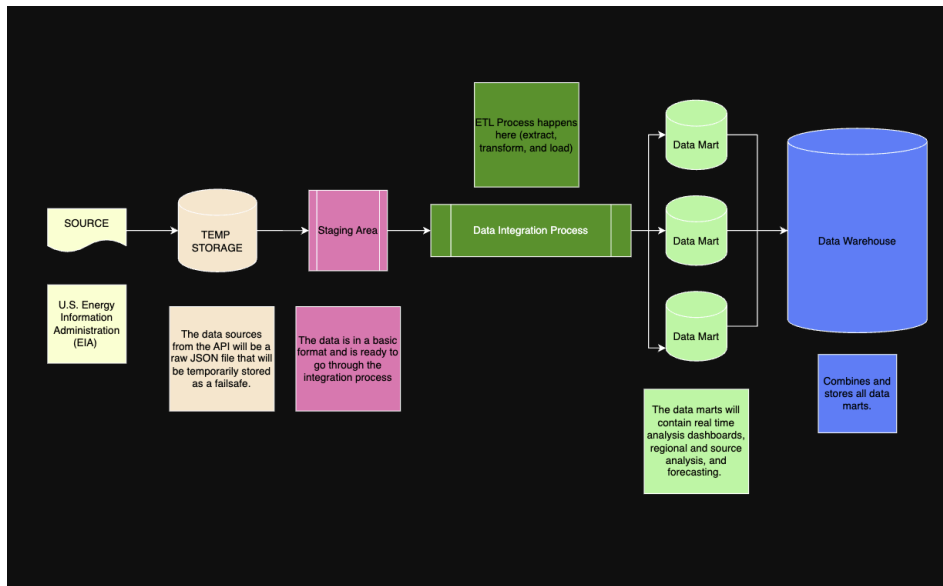
Brief Description:

The information will start at the source, that being the EIA Open Data API, from which the data will be extracted from. The data would then be stored in a temporary storage to ensure pipeline structure. Once extracted from the temporary storage, the data then will be cleaned and reformatted for proper and consistent use. Once things like missing values or misalignments are handled, then the data can be transformed. The data will then be stored in a centralized data warehouse. The users would interact with the data through dashboards, reports, and other tools that will be built on top of the warehouse for analysis and forecasting purposes. The data will be accessed by individuals mentioned earlier.

Diagram:



Data Architecture: Will use Bottom Up Approach



Data Integration Points:

- The main integration point is the EIA API, which is the source. It will retrieve the needed metrics.
 - Hourly electricity demand by region
 - Hourly electricity demand by source
 - Hourly electricity generation by region
 - Hourly electricity generation by source

Ensuring Consistency:

- The use of foreign keys like region codes, time stamps, and surrogate keys

Ensuring Integrity:

- The data will go from temporary storage into a staging area where some basic cleaning and missing values will be addressed.
- The use of surrogate keys will make sure that there is consistency and matching in the star schema.
- Relationships will be established between the main fact table and the dimension tables

Ensuring Scalability:

- Bottom-Up Data Architecture Model will allow for multiple data mart creation and you can add as many as needed. If extra information is needed like carbon emissions to do further analysis on the impact of renewable sources, it can be introduced in a data mart.

- The data warehouse will be implemented on a cloud (Azure), which would allow scalability.

Dimensional Modeling

For dimensional modeling, I will use a big (wide) fact table, two dimensions, and a reference table (not connected) to be used later with BI tools. I think it is best to keep everything in a wide fact table for easier data retrieval, have a calendar dimension, and a region dimension connected. There is only one data source here and it is the EIA Open Data API.

