W205 Final Project Proposal

Team Sunshine: Boris Kletser, Maya Miller-Vedam, Geoff Stirling, Laura Williams

Project Framing

Domain: Renewable energy production

Problem:
One of the current barriers to widespread renewable energy adoption is that sources like wind and solar are highly variable, which presents availability and storage challenges. Utility companies, government agencies, and renewable energy research and funding organizations would all benefit from a better understanding of how much variation different sites experience and how this variation impacts energy production.

Potential clients:

- Power companies and project developers responsible for planning, building, and maintaining electricity power plants, *i.e.*, Mojave Solar Project and Green Rhino Energy
- Regional power administration agencies, responsible for regional electricity distribution, *i.e.*, Western Area Power Administration and the Bonneville Power Administration
- Public utility companies responsible for delivering local electricity directly to consumers, *i.e.*, ConEd in New York, PG&E in California
- State public utility regulatory agencies, *i.e.*, California Public Utilities Commission
- United States Dept. of Energy and its Information Administration (EIA)
- United States Environmental Protection Agency (EPA)
- Renewable energy research centers, *i.e.*, National Renewable Energy Lab in Colorado
- Environmental protection funding organizations, *i.e.*, Environmental Defense Fund

Research Question:
For the initial implementation we'll focus specifically on solar radiation and energy produced in solar panel fields that are connected to major utility companies. We will ask:

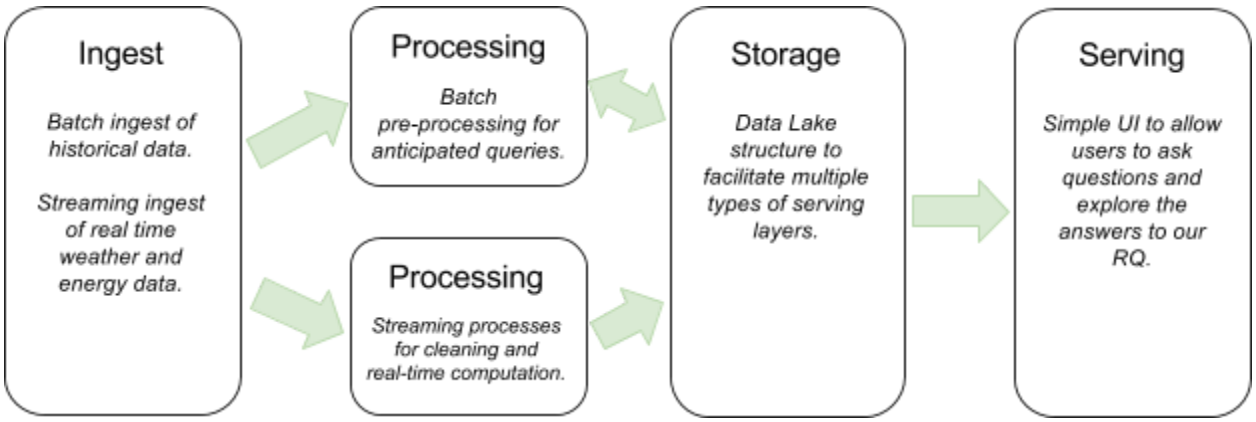
- *How does weather fluctuation impact energy production from solar fields?*
- *Given the weather patterns in a given location, would a solar field make sense as an addition to current energy production methods?*
- *TBD, if it is possible: Can we predict energy production from existing solar panel fields from real time weather data?*

Data Sources

- NOAA Quality Controlled Datasets - Monthly, daily, hourly and 5 min dumps from the US Surface Climate Reference Network including air temperature, precipitation, solar radiation, surface temperature, soil moisture and soil temperature data from weather stations around the country.
- Energy Production by Plant APIs from the Energy Information Administration - Hourly electricity operating data, including actual and forecast demand, net generation, and the power flowing between electric systems
- Solar Integration Datasets and state of the industry - National Renewable Energy Laboratory provides the energy community with modeled solar data to study the operational impacts of solar on the electric power grid.
- Solar Plants Location and Production Dataset from openei.org - overview of solar plants, locations, capacity, and facility type.

Data Architecture

As we work through this project, we will explore and choose tools to create a pipeline with ingest, processing, storage and serving layers, as outlined below:



Next Steps:

- Create table with data sources / access / limits
    - Continue to look into volume of available data. (The EIA API allows us to pull 100 series in a single request and has an updates feed which should prevent the excessive server requests. On our end, pagination can help us handle the data requested.)
- Create ER Diagram

    - Narrow down the specific data fields we will use from each of the datasets. These fields could include: Cloud cover/Visibility, Solar radiation, Precipitation, Location data to link weather with specific power plants, Renewable power plant locations and energy production)
    - Based on the data available, clarify the specific type of analysis that we can do to address our research questions. Match data to our ER Diagram.
- Build a skeleton end-to-end implementation including the four layers described above.
    - For example, as a starting point we might look at a single power plant and the weather in that specific location.
    - Decide on output layer to make analysis accessible to both technical and non-technical stakeholders

Input Sought From Arash:

- Can we start with just one variable comparison as a skeleton and plan to build it out later? (i.e.. Comparing the output of a single solar plant to that area's solar radiation)?
    - If not, do we want to try first on smaller set first?