Intel Data Center

INTRODUCTION: Intel, the semiconductor manufacturing powerhouse, is planning on building a new data center. Energy availability and usage are some of the key considerations in deciding on a location of the data center. For example, which regions produce a surplus of energy, and are therefore more likely to provide energy at cheaper prices? Which regions rely more on renewable energy sources?

In this project, co-designed with Intel's Sustainability Team, you'll write SQL queries that will power your analysis and create visualizations that will help the Intel team select the best location for the new data center.

— Data Sets

`intel.energy_data`: Contains information about daily energy production and consumption for different regions in the United States.

- `balancing_authority` - A Balancing Authority is responsible for maintaining the electricity balance within its region. This is a company that makes sure electricity is being exchanged between electric providers and regions so that no region runs out of electricity due to high demand.
- `date` - The date the energy was produced.
- `region` - The electric service area within a geographic area of the USA. e.g. California, Midwest, etc.

- `time_at_end_of_hour` - The time and date after energy was generated, .e.g., energy generated between 1pm-2pm will show up as 2pm in this field.
- `demand` - The energy demand in megawatts (MW) on the grid (what the houses/business are using).
- `net_generation` - The energy produced in MW in the region by all sources e.g., wind, coal, nuclear, etc.
- `all_petroleum_products` - The energy produced in MW by petroleum products.
- `coal` - The energy produced in MW by all coal products
- `hydropower_and_pumped_storage` - The energy produced in MW by water power and pumped heat sources.
- `natural_gas` - The energy produced in MW by natural gas sources
- `nuclear` - The energy produced in MW from nuclear fuel sources
- `solar` - The energy produced in MW by solar panels and other solar energy capturing methods.
- `wind` - The energy produced in MW from wind turbines and other wind sources.

`intel.power_plants`: Contains general information about power plants in the United States.

- `plant_name` - The name of the power plant.
- `plant_code` - The unique identifier of the plant.
- `region` - The region in the US where the power plant is located. Matches the regions in the intel.energy_data
- `state` - The state where the power plant is located.
- `primary_technology` - The primary technology used to generate electricity at the power plant.

`intel.energy_by_plant`: Contains total energy production information at the plant for the year 2022.

- `plant_name` - The name of the power plant.
- `plant_code` - The unique identifier of the plant.
- `energy_type` - The kind of energy generated by the power plant. Either renewable energy or fossil fuel.
- `energy_generated_mw` - The total energy generated, in MegaWatts, at the plant for the year 2022.

— **Task 1:** Energy Generation

Let's first identify regions that are net energy producers. Not all regions generate enough energy to meet the local demand. Some regions purchase power from other regions, while others sell their surplus to regions in need.

A. A query using the `intel.energy_data` table that calculates the sum total of energy produced, grouped by each region. Sorted the output by highest total energy.

```
SELECT
  region,
  SUM(net_generation - demand) AS total_gen
FROM intel.energy_data
GROUP BY region
ORDER by total_gen DESC
```

Which region has the highest positive total energy?

> The region with the highest positive total energy is Mid-Atlantic with 31693087.

B. Intel is interested in regions that generate a large amount of energy from renewable sources. Renewable energy is defined as any energy generated from `hydropower_and_pumped_storage`, `wind`, and `solar` sources.

A query that calculates the sum total of renewable energy by region. Sorted the output by the region with the highest renewable energy.

```
SELECT
```

```
  region,
  SUM(hydropower_and_pumped_storage + solar + wind) AS
sum_solar
FROM intel.energy_data
GROUP BY region
ORDER by sum_solar DESC
```

What are the top two regions for total renewable energy production?

> The top two regions for total renewable energy production are the Northwest: 199266574 and Texas: 131367234

C. Modified query slightly so that it calculates the **percentage** of renewable energy by region.

```
SELECT
  region,
  SUM(hydropower_and_pumped_storage + solar + wind) AS
sum_solar,
  CONCAT(ROUND(SUM(hydropower_and_pumped_storage + solar +
wind) / SUM(net_generation) * 100, 2), '%') AS
renew_percentage
FROM intel.energy_data
GROUP BY region
ORDER BY sum_solar DESC;
```

D. Which regions change from the top 3 when looking at total renewable energy vs percentage of renewable energy?

> In percentage renewable energy, the top three are: Northwest, Texas, and Central.

In total renewable energy, the top three are also: Northwest, Texas, and Central.

— **Task 2:** Generating New Data by Energy Type

Intel would like to know how renewable energy and fossil fuels trend over time. In order to do this, you will first need to generate a new table using your SQL knowledge and the `intel.energy_data` table before visualizing trends in Tableau Cloud.

A. A query that calculates the renewable energy generated for each row. Returning only the `date`, `region`, and `energy_generated_mw` columns.

```
SELECT
  date,
  region,
   SUM(hydropower_and_pumped_storage + solar + wind) AS
energy_generated_mw
FROM intel.energy_data
GROUP BY date, region;
```

After showing the result of the query to your manager, she tells you that she wants it to be clear that the `energy_generated_mw` column is referring to renewable energy types. She asks you to create a new column called `energy_type` that has the value 'renewable energy' for each row.

A colleague teaches you a simple method to do this. When writing your query, add an additional column after your select statement. Here is an example:

```
SELECT
```

```
  *, -- any relevant fields to the query
  'renewable energy' AS energy_type
FROM intel.energy_data
```

B. Modified query from Part **A.** to include the `energy_type` column.

```
SELECT
  date,
  region,
  'renewable_energy' AS energy_type,
  SUM(hydropower_and_pumped_storage + solar + wind) AS
energy_generated_mw
FROM intel.energy_data
GROUP BY date, region;
```

C. A **new** query that calculates the fossil fuel energy generated for each row. As in Part **A.**, returning only the `date`, `region`, and `energy_generated_mw` columns, where `energy_generated_mw` is now the alias for `all_petroleum_products + coal + natural_gas + nuclear + other_fuel_sources`.

```
SELECT
  date,
  region,
  SUM(all_petroleum_products + coal + natural_gas + nuclear +
other_fuel_sources) AS energy_generated_mw
FROM intel.energy_data
GROUP BY date, region;
```

D. Modified the query in Part **C.** to include the `energy_type` column. This column now has the value 'fossil fuel' for each row.

```
SELECT
```

```
  date,
  region,
  'fossil_fuel' AS energy_type,
  SUM(all_petroleum_products + coal + natural_gas + nuclear +
other_fuel_sources) AS energy_generated_mw
FROM intel.energy_data
GROUP BY date, region;
```

E. The queries from Parts **B.** and **C.** should both have the columns `date`, `region`, `energy_generated`, and `energy_type`. Here is a final query that `UNION`s these two together.

```
SELECT
  date,
  region,
  'renewable_energy' AS energy_type,
  SUM(hydropower_and_pumped_storage + solar + wind) AS
energy_generated_mw
FROM
  intel.energy_data
GROUP BY
  date,
  region
UNION
SELECT
  date,
  region,
  'fossil_fuel' AS energy_type,
  SUM(
    all_petroleum_products + coal + natural_gas + nuclear +
other_fuel_sources
  ) AS energy_generated_mw
FROM
  intel.energy_data
GROUP BY
  date,
```

```
  region;
```

Task 3: Aggregating Power Plant Data

Intel has provided you with additional data in order to reach the best conclusion about the location of its next data center. In this task you will be working with two tables `intel.power_plants` and `intel.energy_by_power_plant`. You will need to join these tables before you can aggregate them to help the Intel team with their analysis.

A. Joined the `intel.power_plants` and `intel.energy_by_power_plant` data on the `plant_code`. This joined table will form the basis for the rest of the task.

The output is 2,504 rows.

```
SELECT
  *
FROM intel.power_plants as pp
JOIN intel.energy_by_plant as ep
ON pp.plant_code = ep.plant_code
```

B. A query that returns the total number of renewable energy power plants for each region.

```
WITH renew_energy AS (
  SELECT
    fuel_types
  FROM
    intel.power_plants
  WHERE
```

```
    fuel_types IN ('SUN', 'WAT', 'WND')
)
SELECT
  pp.region,
  'Renewable Energy' AS renewable_energy_category,
  COUNT(*) AS num_renewable_power_plants
FROM
  intel.power_plants as pp
  JOIN intel.energy_by_plant as ep ON pp.plant_code =
ep.plant_code
WHERE
  pp.fuel_types IN ('SUN', 'WTR', 'WND')
GROUP BY
  pp.region
```

Which region has the most renewable power plants?

> The Midwest with 203 renewable energy powerplants

C. A query that returns both the total number of power plants and the total energy generated, specifically from plants that use “Solar Photovoltaic” technology, grouped by each region.

```
SELECT
  pp.region,
  COUNT(*) AS total_power_plants,
  SUM(ep.energy_generated_mw) AS total_energy_generated_mw
FROM
  intel.power_plants AS pp
  JOIN intel.energy_by_plant AS ep ON pp.plant_code =
ep.plant_code
WHERE
  pp.primary_technology = 'Solar Photovoltaic'
```

```
GROUP BY
  pp.region;
```

D. Modified query from part **C** to only show regions having at least 50 power plants that use “Solar Photovoltaic” technology.

```
WITH SPV_pp AS (SELECT
  pp.region,
  COUNT(*) AS total_power_plants,
  SUM(ep.energy_generated_mw) AS total_energy_generated_mw
FROM
  intel.power_plants AS pp
  JOIN intel.energy_by_plant AS ep ON pp.plant_code =
ep.plant_code
WHERE
  pp.primary_technology = 'Solar Photovoltaic'
GROUP BY
  pp.region)

SELECT
  *
FROM SPV_pp
WHERE total_power_plants >= 50
ORDER BY total_energy_generated_mw DESC;
```

What can you infer about the efficiency (or size) of the power plants in the Midwest region relative to the other regions in your output?

Looking at the output, I can see that the Midwest ranks third for highest number of power plants, but it ranks lowest out of the 6 regions with over 50 power plants using Solar Photovoltaic for

the amount of total energy generated. This means that while it may have more plants, it isn't necessarily the most efficient region.

— **LevelUp:** Hourly Trends in Renewable Energy

Before moving on to your Tableau Visualizations, let's investigate how renewable energy generation fluctuates with the time of day.

A. A query that calculates the total **renewable** energy generated in each region for each hour of the day.

```
SELECT
  region,
  DATE_PART('hour', time_at_end_of_hour) AS hour_of_day,
  SUM(hydropower_and_pumped_storage + solar + wind) AS
total_renewable_energy
FROM
  intel.energy_data
WHERE
  hydropower_and_pumped_storage IS NOT NULL
  AND solar IS NOT NULL
  AND wind IS NOT NULL
GROUP BY
  region,
  DATE_PART('hour', time_at_end_of_hour)
ORDER BY
  region,
  hour_of_day;
```

B. Modified query to filter to the 'California' and 'Northwest' regions only.

```
SELECT
  region,
```

```
  DATE_PART('hour', time_at_end_of_hour) AS hour_of_day,
  SUM(hydropower_and_pumped_storage + solar + wind) AS
total_renewable_energy
FROM
  intel.energy_data
WHERE
  region IN ('California', 'Northwest')
  AND hydropower_and_pumped_storage IS NOT NULL
  AND solar IS NOT NULL
  AND wind IS NOT NULL
GROUP BY
  region,
  DATE_PART('hour', time_at_end_of_hour)
ORDER BY
  region,
  hour_of_day;
```