Project Title

Unified Hydropower Data Platform in Microsoft Fabric

Deliverables

Data Sources**:** Fingrid, Zenodo, Hydropower Database
- Integrate three main data sources — Fingrid's real-time hydropower generation in Finland, Zenodo's historical modeled capacity factors, and a European hydropower plant metadata set containing capacity and typology details.
- Using Spark notebooks, clean and harmonize these datasets to compare actual production against installed capacity and long-term climatic potential.
- The solution would use CI/CD pipelines for automation and Power BI for visualization, enabling insights into seasonal efficiency trends and how climate conditions affect hydropower performance in Finland.

Plan of Action / Project Flow

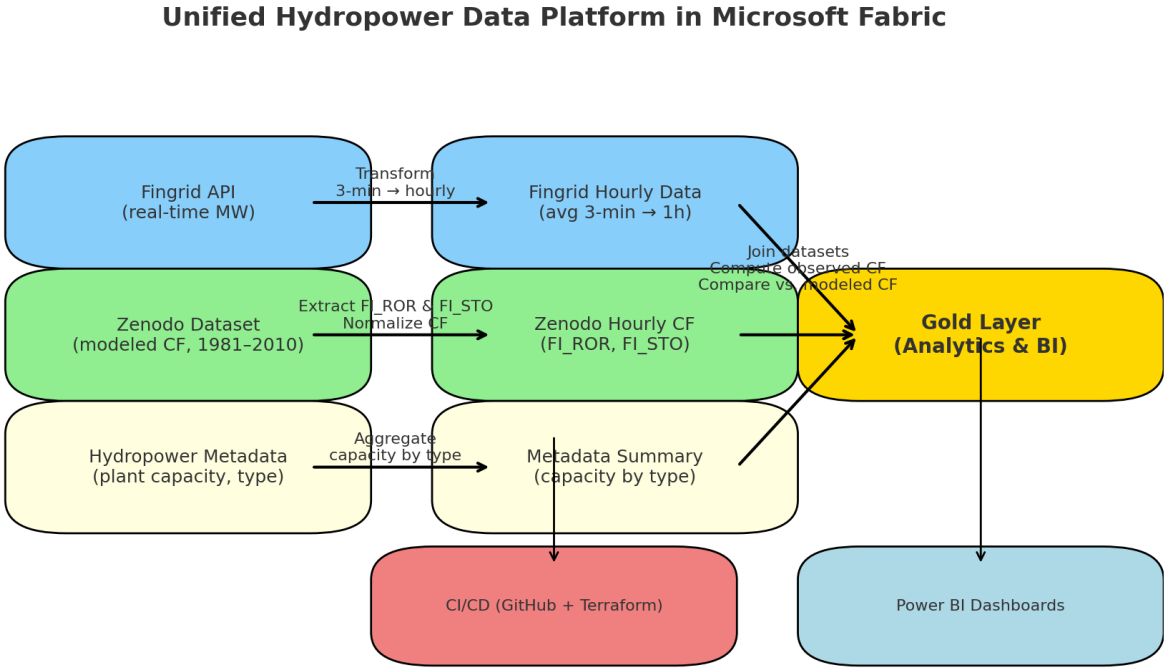
Example in Fabric Project Flow

1. **Bronze Layer**
    a. Ingest hydropower metadata CSV/API from the GitHub database (id, capacity, type, etc.)
2. **Silver Layer**
    a. Clean metadata (e.g. unify country codes, types)
    b. Compute "total capacity in Finland" by summing all plants in Finland
    c. Possibly normalize types (e.g. run-of-river vs reservoir)
3. **Gold / Analytics Layer**
    a. From Fingrid's observed output → compute **observed capacity factor**
    b. Compare observed capacity factor to modeled capacity factor (Zenodo) for Finland.
    c. Drill down by type: e.g. "What is the capacity factor of storage plants vs run-of-river in observed data vs model?"
4. **Visual & Insights**
    a. Show charts: time series of observed generation, capacity factor, model vs actual
    b. Use metadata to create charts by plant type or region

Description

Here's the **architecture diagram** showing how the three datasets - **Fingrid (real-time)**, **Zenodo (historical modeled)**, and **Hydropower Metadata (capacity/type)** . Following is the flow through Microsoft Fabric:

- **Bronze layer:** raw ingestion from APIs and CSVs.
- **Silver layer:** cleaned, aligned data (hourly averages, extracted Finland columns, capacity summaries).
- **Gold layer:** unified analytics comparing observed vs. modeled capacity factors.
- **CI/CD** automates deployment, and **Power BI** dashboards visualize performance and climate insights.



Layer	Data Source(s)	Tasks	Output Example
Bronze (Raw)	Fingrid API (real-time MW), Zenodo CSV (modeled CF 1981–2010), Hydropower Metadata CSV (plant-level info)	Ingest raw data from all sources into Fabric Lakehouse. Store each as-is in separate tables or folders.	`bronze_fingrid_raw`, `bronze_zenodo_raw`, `bronze_hydro_metadata_raw`

Silver (Cleaned & Unified)	All three	- Clean timestamps and normalize time zones. - Convert Fingrid's 3-min data → hourly averages. - Extract Finland columns (`FI_ROR`, `FI_STO`) from Zenodo. - Clean metadata (country, plant type, capacity). - Aggregate total installed capacity by type (run-of-river, storage).	`` `silver_fingrid_hourly` ``, `` `silver_zenodo_fi_cf` ``, `` `silver_capacity_summary` ``
Gold (Analytics & Modeling)	Joined dataset (Fingrid + Zenodo + Metadata)	- Join observed generation with installed capacity to compute **observed capacity factor**. - Compare observed vs. modeled (Zenodo) capacity factors. - Analyze efficiency by plant type and season. - Compute correlation between inflow (Zenodo proxy) and generation (Fingrid).	`` `gold_hydro_fi_summary` `` (capacity factors, efficiency trends, deviations)
Visualization (BI)	Power BI (connected to Fabric Lakehouse)	- Create interactive dashboards showing: • Real vs. modeled capacity factors • Efficiency trends by month/season • Breakdown by plant type • Anomalies or deviations from model	`` `hydro_power_efficiency.pbix` ``(Fabric or Power BI Service dashboard)
Streaming (Real-time layer)	Fingrid API (3-min intervals)	- Ingest real-time data into Eventstream. - Compare live generation vs. historical baseline from Zenodo. (Zenodo as long term baseline data, using scenario analysis) - Trigger alerts for underperformance or anomalies.	Real-time Fabric dashboard (Power BI streaming tiles)