**Workflow Overview**

1. **Ingestion (Bronze Layer)**
   * Pull Fingrid hydro power API data (JSON/CSV) using Spark notebook.
   * Store directly in Fabric Lakehouse (bronze tables).
2. **Cleaning & Transformation (Silver Layer)**
   * Normalize timestamps, units, missing values.
   * Join with **hydro plant metadata** (capacity, type).
3. **Analytics (Gold Layer)**
   * Aggregations: hourly/daily averages, regional summaries.
   * Calculate load factor = actual output / capacity.
   * Store as clean fact tables for dashboards.
4. **Streaming Simulation**
   * Use Fabric Eventstream (or Python generator) to push real-time power readings.
   * Spark Structured Streaming job consumes → stores into Lakehouse.
5. **CI/CD**
   * GitHub repo with all notebooks & SQL scripts.
   * GitHub Actions workflow runs unit tests (Spark job validation, SQL checks).
   * Deploy updated pipelines automatically into Fabric workspace.
6. **Visualization (Power BI)**
   * Simple dashboard with:
     + Hydro production trends (hourly, daily)
     + Regional breakdowns
     + Anomaly detection (e.g., sudden dips in output)

| **Dataset** | **What It Tells** | **Time Dimension** | **Granularity** | **Purpose** |
| --- | --- | --- | --- | --- |
| **Fingrid API** | Actual real-time generation (MW) | Minutes | National total | Observed performance |
| **Zenodo** | Modeled long-term capacity factors | Hourly (1981–2010) | Country/bidding zone | Historical baseline |
| **Metadata** | Installed plant capacity, type, location | Static | Plant-level | Reference and scaling factor |

| **Layer** | **Dataset** | **Granularity** | **What It Brings** | **How It Connects** |
| --- | --- | --- | --- | --- |
| **Silver** | silver\_fingrid\_hourly | Hourly | Real generation (MW) | Used as numerator in Observed CF |
| **Silver** | silver\_meta\_capacity\_fi | Static | Installed capacity (MW) by type | Denominator for Observed CF |
| **Silver** | silver\_zenodo\_fi\_cf | Hourly (1981–2010) | Historical capacity factors by hour/month | Baseline to compare observed CF |
| **Gold** | — | Hourly (aligned) | Joins all three → Observed CF, Historical CF, and Deviation | Final KPI output |

A black text on a white background

AI-generated content may be incorrect.

┌────────────────────────┐

│ Fingrid (3-min API) │

│ → Hourly Averages │

└────────────┬──────────┘

│

▼

silver\_fingrid\_hourly

│

┌─────────────────┼──────────────────┐

▼ ▼

┌──────────────────┐ ┌────────────────────┐

│ JRC Metadata (CSV) │ │ Zenodo JingHydro.csv │

│ Installed MW by type│ │ 1981–2010 CF by hour │

└──────────┬──────┘ └───────────┬───────┘

│ │

▼ ▼

silver\_meta\_capacity\_fi silver\_zenodo\_fi\_cf

└────────────┬──────────────────────────┘

▼

┌────────────────────┐

│ GOLD LAYER │

│ Observed CF = │

│ Fingrid / JRC │

│ Historical CF = │

│ Zenodo avg\_cf │

│ Deviation = Δ │

└────────────────────┘