
Technical challenge for the data engineer role at BayWa r.e Data Services GmbH

As you already know, at BayWa r.e we're at the forefront of renewable energy and always eager to get our hands on new datasets in the field. We recently found out about the Energy-charts open datasets and would like to provide our data scientist with data they can crunch.

You are tasked with building a data pipeline and storage layer with the following requirements:

1. **Build the following ingestion data pipeline**

Source data: [Energy-Charts API - Swagger UI](#)

- Public power data 15minutes interval: daily ingestion
- Price data: daily ingestion
- Installed power: monthly ingestion

We should be able to change at least:

- Ingestion frequency
- Start and end time for backfilling

Restrict the data volume as you think fits a local execution of the pipeline.

2. **Create and document a data model that allows the following BI/ML use cases.**

- Trend of daily public net electricity production in Germany for each production type.
- Prediction of underperformance of public net electricity on 30min intervals.
- Analysis of daily price against the net power for offshore and onshore wind

Store the data ingested into relevant tables fitting the use cases. Data should be able to flow from the source to the data model tables each time we run the pipeline.

Create SQL queries that would be the basis for each report.

Preferred Stack

- Storage type: Delta Table
- Data Processing: pySpark

3. **Monitor the pipeline run and data quality**

Some basic pipelines run monitoring is required as well as basic data quality measure can be checked.

Any kind of method or framework is accepted for local use.

Deliverable

- The code should be hosted on a git repository and shared with us.
- The solution can be run locally by anyone, code should have basic comments.
- A read README explaining your approach, any assumptions and decision made, and how to set up and test the solution.
- Document a cloud production grade data ingestion, data warehouse and monitoring implementation of it. Diagrams of basic cloud components and data flow are enough.
- Prepare for discussing the chosen solution at the next interview stage.