Instructions

We expect to receive your solution within two days of receiving the test. In those 48 hours, read this document, think about it as much as you like, and then spend **two** active hours working on it.

Evaluation criteria

We do not expect a fully working solution, but we do expect a technical design with some details at least in one of the parts. We will evaluate this solution based on the assumption you spent two hours actively working on it. If you want to spend more time on it, please do, but we will most likely notice, and it won't give you an advantage. You are welcome to use and reference any external sources (blog posts, documentation…) but we do appreciate seeing your own input and critical thinking.

We will evaluate the proposal on

- Feasibility of solution
- Clarity of the documentation provided
- Level of practical details that will make this project a success
- How the solution will meet the business needs we describe

Your task: Data and Analytics at LDC - Weather forecast management platform

At LDC, information is our strength. We celebrate this year our 170th anniversary. and we are committed to building a sustainable future, to ensure we will still be there 170 years from now. To empower our traders to make the best decisions, the Data & Analytics team is creating a world class data platform.

One of the most interesting sources of information is weather.

Design a platform that will allow our users to access the latest weather forecasts.

Our data engineers are already retrieving the data as series of flat files. Forecasts are issued four times per day and consist of grid files containing forecasted values for precipitation and temperature for each grid value over the surface of the planet.

The data arrives in our data lake straight after being emitted. The data is organised in the data lake using the time the data was received, as follows:

<source>/<year>/<date>/<hour of forecast>/

Use cases

This is how the data will be used:

1- Data scientists need to test their models based on past forecasts, following evolution of key metrics as the forecasts evolve. We need to store the full history of forecasts in a way that can be retrieved and processed by our data scientists using the platform. The data consists of 1TB of data per year.

2- Traders need to be kept up to date on the latest forecast values in the context of other indicators relevant to their work, for specific regions, in a dashboard created by our analysts and data engineer. The data needs to be updated within 15 minutes after the forecasts have been published.

3- Analysts need to retrieve historical data easily and conveniently so they can enrich them with other key indicators to create explainable prediction models and self-service analytics dashboards to share key insights with their colleagues.

All three categories of users expect the data to be reliable and need to be notified if data is missing, if there are quality issue or if it is going to arrive late.

The task

1. Propose a high-level architecture that would allow us to meet all three use cases
2. How to transform and store the data for historical access of full forecasts to test granular models?
3. How to transform and store the data for self-service BI analysis?
4. How to monitor and maintain this solution so we can proactively warn our users of any issues?
5. Share three best practices for data engineers when designing pipelines to implement your recommendations.