**Technical challenge: Building a FHIR-Based Pipeline (File-Based)**

**Goals and expectations**

So you want to join the DWH team for the Patient Finder Solution? Great! Before we get started, let’s have you write down some lines of code in order to have a technical conversation and exchange ideas on something concrete, something that you created and that you can show us around and about.

This coding challenge is intended to explore your profile along the following traits:

* Problem solving
* Coding style and proficiency
* Solutions design
* Product thinking

We’d like to receive from you a submission which includes:

* The full codebase: any Git[Hub|Lab|\*\*\*] link works, but also a zip file of the sources if you prefer
* A working solution to the challenge: we can locally run your solution
* Documentation: at the minimum, provide a README that includes instructions on how to run the project

Your time is valuable, and we understand that fitting this exercise into your schedule can be a challenge. That’s why we’re aiming for a flexible but focused approach: we expect to receive your solution in around two days, though if you need more time, just let us know, we’re happy to accommodate. Once we receive your submission, we’ll review it and determine whether to invite you for a technical interview. Your code will serve as a starting point for the conversation, so be ready to discuss the decisions you made during the development process.

**Objective**

The goal of this assignment is to design and implement a robust pipeline that processes healthcare data formatted according to the FHIR (Fast Healthcare Interoperability Resources) standard. Instead of connecting to a live FHIR server, you will work with pre-exported NDJSON files that simulate real-world clinical data exports.

The dataset includes information about patients and their documented allergies or intolerances, represented using the AllergyIntolerance resource. This resource captures clinical assessments related to a patient's adverse reactions to substances, including the allergen, type of reaction, severity, and the clinical context in which the information was recorded.

Your pipeline should be capable of ingesting, transforming, and modeling this data to support downstream analytical use cases such as patient-level reporting or clinical trend analysis.

**Assignment Tasks**

1. Data Ingestion

* Work with provided NDJSON files containing FHIR resources.
* Load the data into your processing pipeline from local files.

2. Data Processing

* Parse the NDJSON files and extract relevant fields.
* Normalize and clean the data.
* Join related resources.
* Implement graceful handling of missing or inconsistent data.

3. Data Modeling

* Design a simplified, star-schema data model consisting of:
  + Dimension Table (i.e. patient, allergen type, reaction type, severity)
  + Fact Table (i.e. the observation of an allergy event)
* Store the final structured data in a local relational database.

**Requirements**

* Apply best practices in data engineering and software development in general.

Good luck!