1. Take-Home Test

Scenario

You have been asked to design and implement a data platform that pulls information from multiple source systems (for example, various EMR platforms, payroll providers, and a standalone data warehouse) into a single consolidated environment for reporting and analytics. You must also ensure the data can be surfaced in a BI tool (such as Tableau).

> **Important**: Since you do not have actual data, **feel free to create or mock up small sample datasets yourself** (or simply outline your approach conceptually if you prefer).

Tasks

1. **High-Level Architecture & Data Modeling**
   - Provide a short **architecture diagram** or description showing how data flows from multiple source systems into your data store (e.g., data warehouse, lakehouse, or another central repository) and then on to a BI tool.
   - Propose a **data model** (e.g., star schema, snowflake schema, or equivalent) to handle typical metrics, such as patient visits, billing, or payroll records, depending on your chosen scenario.
2. **ETL/ELT Pipeline Design**
   - Either outline or implement a basic pipeline that does the following:
     - **Ingests** data from at least two hypothetical source systems (e.g., "EMR_A" and "Payroll_B").
     - **Transforms**/cleans the data so it becomes consistent (for example, cleaning dates, normalizing text fields).
     - **Loads** the transformed data into your data repository.
   - Show how you would handle **incremental updates** (only new or changed records, as opposed to always doing a full refresh).
3. **Governance & Security Considerations**
   - Summarize how you would enforce **role-based access controls** so that only authorized personnel can see sensitive data (e.g., patient or employee information).
   - Note any considerations around data masking (if relevant) or encryption in transit/at rest.
4. **Queries & Reporting**
   - Write or outline **one or two example queries** that produce meaningful reporting metrics (e.g., total visits by month, total payroll cost by department, or other KPI relevant to your mock scenario).
   - Discuss how you would connect a BI tool (like Tableau) to the resulting data store—would you use live queries, extracts, or another method?

5. **Optional/Stretch**: Real-Time vs. Batch
    ○ Briefly explain how you would approach **real-time** ingestion if some data needed near-instant updates.
    ○ Contrast this with a **batch** strategy, describing how you would schedule regular loads (daily, hourly, etc.).

Deliverables

● A brief **README** describing your approach, assumptions, and any instructions on how to run or interpret your work.
● **Architecture diagram** (can be hand-drawn or a simple flowchart).
● **Data model** outline (list of tables or entities, relationships, and any columns you think are critical).
● If coding, **scripts or code snippets** to show how you would implement your ETL/ELT (this can be pseudocode if you prefer).

Time Expectation: This should be doable in **a few hours**. We're looking for clarity of approach, not an enterprise-ready solution.