**Multi-Clinic Patient Monitoring & Analytics Platform**

**Problem Statement**

A healthcare chain operates **multiple clinics across different cities**. Each clinic maintains patient records (EHR), lab tests, pharmacy sales, and IoT vitals (BP, HR, glucose) independently. Data is spiled at each clinic, making it difficult to:

* Get a **consolidated patient view** across clinics.
* Track **clinic-wise performance** (visits, prescriptions, revenue, patient outcomes).
* Monitor **real-time patient vitals** for high-risk patients.
* Enable **predictive insights** (readmission risks, chronic illness management).
* Ensure **compliance** (GDPR) while sharing data across clinics.

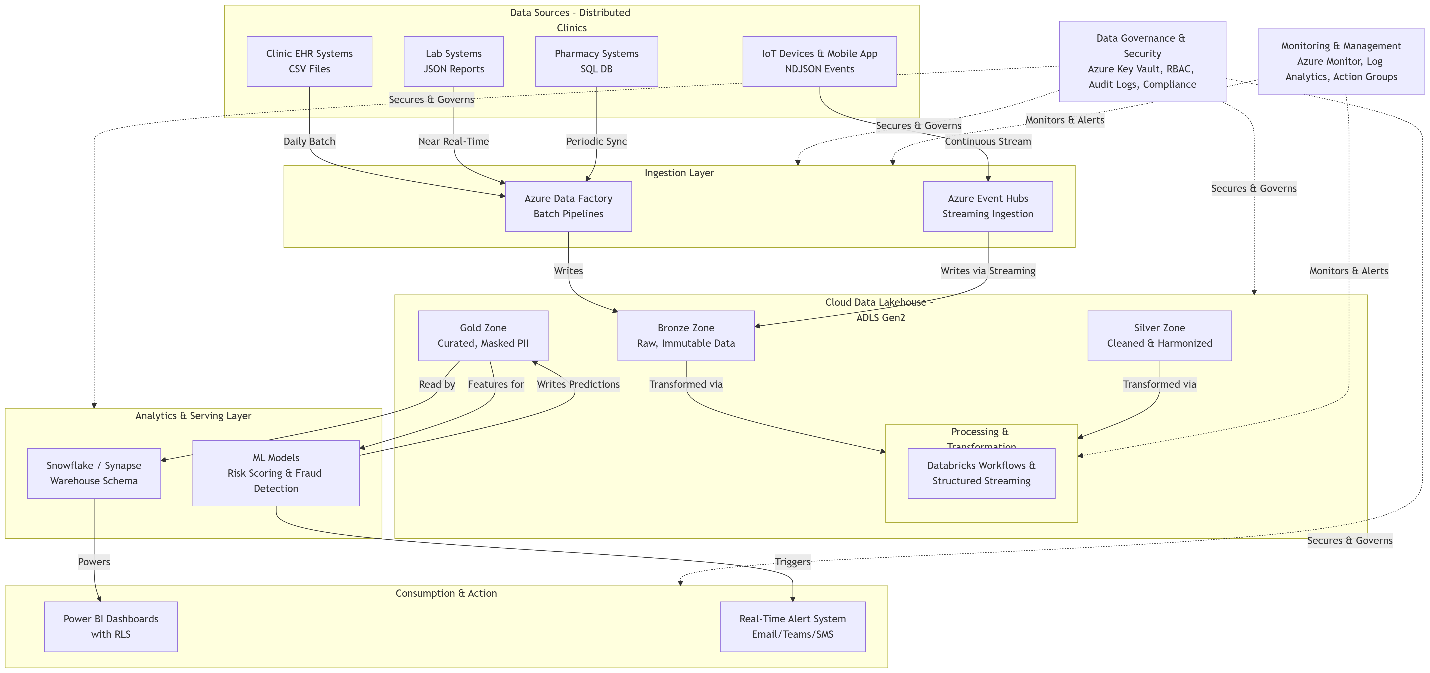
The project aims to build a **cloud-based patient monitoring & analytics platform** to unify data from all clinics, provide real-time insights, and enable better care + cost optimization.

**Skill Tower to Develop the Project**

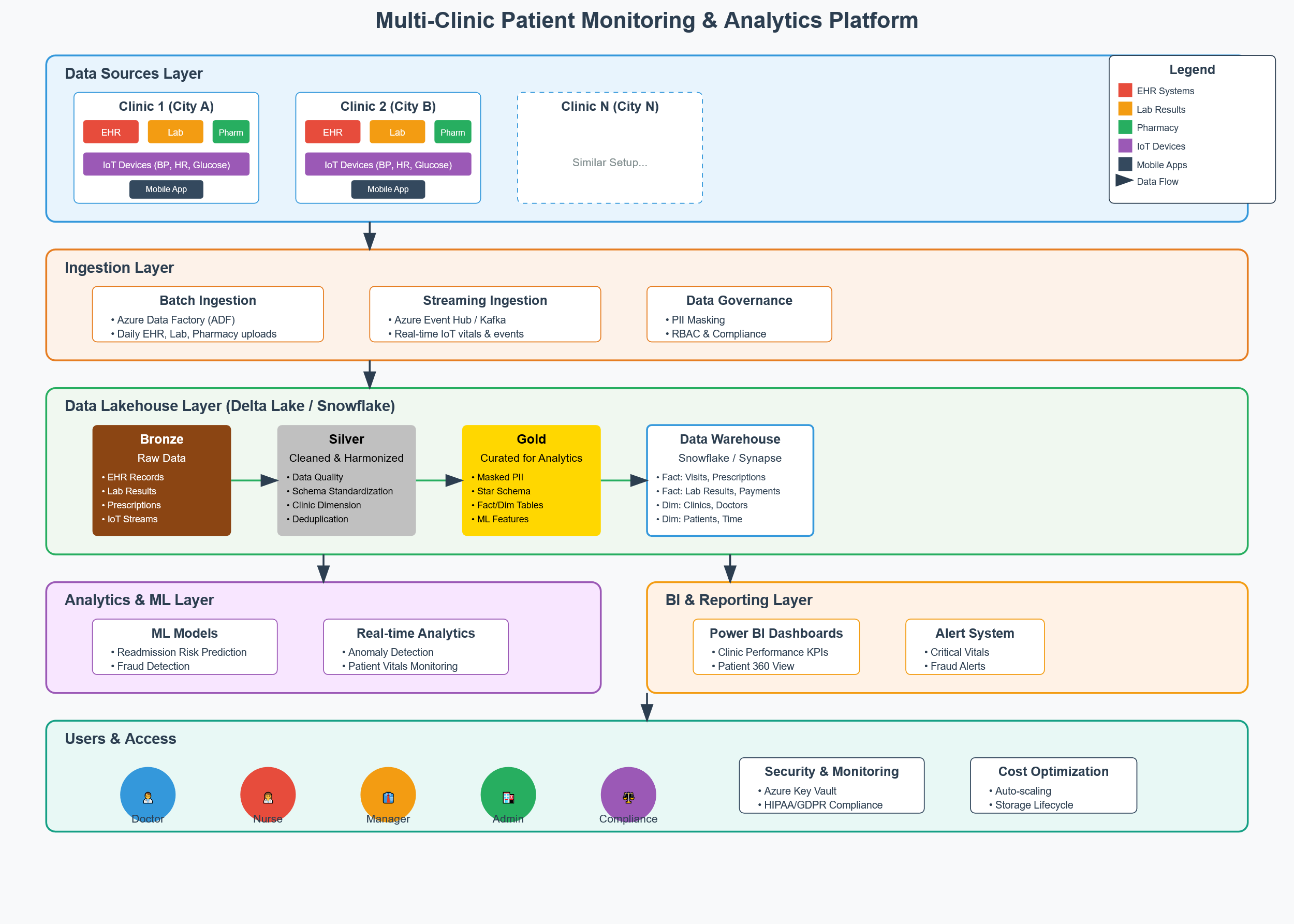
* **Data Engineering**: Batch + streaming pipelines from each clinic.
* **Cloud Data Lakehouse**: Delta Lake/Snowflake/Synapse to unify clinic data.
* **IoT & Streaming**: Event ingestion from connected medical devices.
* **Data Governance & Security**: PII masking, RBAC, compliance checks.
* **Analytics & ML**: Patient risk scoring, fraud detection (fake claims, overbilling).
* **BI**: Clinic-level dashboards (patient load, treatments, revenue, outcomes).

**Use Case / Architecture Diagram**

**Flow Diagram: -**



**Architecture diagram:**



**Data Sources**

* **Clinic EHR Systems**: Patient visits, diagnosis, treatment.
* **Lab/Pharmacy Systems**: Lab reports, prescriptions, medicine sales.
* **IoT Devices**: Glucose monitors, ECG patches, BP cuffs.
* **Mobile App**: Patient-reported symptoms.

**Ingestion Layer**

* **Batch (ADF/Glue)**: Daily uploads from clinics.
* **Streaming (Event Hub/Kafka)**: IoT vitals + appointment events.

**Lakehouse (Delta Lake / Snowflake)**

* Bronze: Raw data from all clinics.
* Silver: Cleaned + harmonized with clinic dimension.
* Gold: Curated for BI/ML, with masked PII.

**Analytics & ML**

* Predictive models for readmission risk, chronic disease alerts.
* Fraud detection (e.g., unusually high billing at one clinic).

**BI & Reporting**

* Clinic performance dashboards.
* Patient 360 view across all clinics.
* Real-time patient alert system.

**User Stories**

1. As a **doctor**, I need a single patient record even if the patient visits different clinics.
2. As a **nurse**, I want real-time alerts when IoT vitals cross thresholds.
3. As a **clinic manager**, I want dashboards to compare patient visits, revenue, and outcomes across clinics.
4. As a **hospital admin**, I want to detect abnormal billing/claims from clinics.
5. As a **compliance officer**, I need PII masked before data enters reporting.

**Expected Deliverables**

* **Centralized Clinic Data Platform** (multi-clinic EHR, pharmacy, labs, IoT).
* **Data Lakehouse Schema** with Fact tables (Visits, Prescriptions, Lab Results) and Dim tables (Clinics, Doctors, Patients).
* **Real-time IoT Monitoring** with alerts.
* **ML Models** for patient risk + fraud detection.
* **Power BI Dashboards** (per-clinic KPIs, patient outcomes, revenue).
* **Compliance & Governance Setup** (PII masking, audit logs, RBAC).

**Milestone & Duration**

| **Milestone** | **Duration** |
| --- | --- |
| M1 – Requirements & Architecture | 1 hr |
| M2 – Batch Ingestion (EHR, lab, pharmacy) | 2 hr |
| M3 – Streaming Ingestion (IoT vitals) | 2 hr |
| M4 – Data Lakehouse Modelling (Bronze → Gold) | 2 hr |
| M5 – Advanced Analytics (ML: readmission risk, fraud detection) | 2 hr |
| M6 – BI Dashboards per Clinic | 1 hr |
| M7 – UAT, Compliance Checks & Go-Live | 1 hr |

**Implementation Notes**

* **Data Sources**:
  + Electronic Health Records (EHR) – CSV, daily batch ingestion.
  + Lab Results – JSON, near real-time ingestion.
  + Pharmacy Prescriptions – SQL DB, periodic sync.
  + IoT Vitals – NDJSON events (BP, HR, glucose), streaming ingestion.
* **Batch Example**:
  + Nightly load of patient visits, prescriptions, and billing data via ADF pipelines.
  + Incremental load for new daily patient records.
* **Streaming Example**:
  + Capture patient vitals from IoT devices every few minutes via Event Hub → Databricks Structured Streaming.
  + Feed real-time anomaly detection alerts (e.g., abnormal glucose, heart rate) monitoring dashboards.
* **Delta Lake**:
  + Bronze → Silver → Gold pipelines for ACID transactions.
  + Support schema evolution (e.g., adding new diagnostic code column without breaking pipeline).
  + Enable time travel for troubleshooting and rollback of sensitive data updates.
* **Snowflake / Synapse:**
  + Curated data warehouse layer for clinical reporting.
  + Fact tables: Visits, Prescriptions, Lab Results, Payments.
  + Dimension tables: Clinics, Doctors, Patients.
  + Use Snowpark for ML-driven patient readmission risk and fraud detection.
  + Enable secure data sharing with insurance providers and research partners.
* **Synapse / Power BI:**
  + Build interactive dashboards for patient volumes, treatment outcomes, and financial anomalies.
  + Embed RLS (Row-Level Security) for doctor/clinic-specific access.
  + Error Handling & Monitoring:
  + ADF → Integration with Azure Monitor for pipeline failure alerts.
  + Databricks → Push logs & metrics Log Analytics, enable job-level retries.
  + Action Groups → Email/Teams notifications for critical failures.
  + Build a Monitoring Dashboard for pipeline health, IoT event lag, and cluster costs.
* **Data Security:**
  + Apply masking on PII fields (patient name, contact) in curated zone.
  + Example: 9876543210 → \*\*\*\*\*\*\*3210.
  + Use Azure Key Vault for credential and connection string management.
  + Ensure RBAC & ADLS ACLs for secure data access.
  + Apply HIPAA/GDPR compliance checks for sensitive medical data.
* **Cost Estimation & Optimization:**
  + ADF – Activity & pipeline execution cost monitoring.
  + Databricks – Track cost per cluster/job, use auto-termination policies.
  + Snowflake – Estimate query credit usage, leverage caching, warehouse sizing, and query optimization.
  + Storage Lifecycle – Tier older data (e.g., >3 years) in ADLS to optimize storage costs.
  + Document cost impact of seasonal patient surge (e.g., flu season).

**Evaluation Rubrics**

| **Criteria** | **Weightage** |
| --- | --- |
| Correct Azure setup (RG, ADLS, RBAC, Lifecycle) | 10% |
| ADF pipelines for batch & incremental ingestion (EHR, lab, prescriptions) | 15% |
| Databricks transformations & Delta Lake implementation (Bronze → Gold, schema evolution) | 15% |
| Warehouse schema design (Snowflake/Synapse – Fact/Dim for clinics, patients, doctors, visits, labs, prescriptions, payments) | 15% |
| BI Dashboard (clinic dashboards, patient 360 view, anomaly/fraud alerts) | 10% |
| Error Handling & Monitoring setup (ADF + Databricks logs, alerting, Action Groups) | 10% |
| Data Security (PII masking, RBAC, compliance) | 10% |
| Cost Estimation & Optimization (ADF, Databricks, Snowflake, ADLS lifecycle) | 5% |
| Documentation (architecture diagram, troubleshooting, pipeline notes) | 5% |
| Bonus: Real-time anomaly/fraud detection or ML-driven readmission prediction | 5% |