

# From Silos to Sync: Designing a Scalable Cloud Solution for Alberta's Medical Clinics



**Prepared by:**

Nabamita De



Southern Alberta Institute of Technology

August 07, 2025

## Table of Contents

1. Abstract .....	3
2. Project Overview .....	3
2.1. Mission, Objectives, and Purpose .....	3
2.2. Why Alberta Clinics Need This Integration .....	3
2.3. Target Outcomes for Patients and Providers .....	3
3. Proposed Cloud Architecture Strategy .....	4
3.1. Data Sources .....	4
3.2. Data Sinks .....	5
3.3. High-level Azure-Based Vision Diagram .....	5
3.4. Cloud Architecture Diagram .....	6
3.5. Key Azure services used .....	6
4. Data Ingestion Strategy .....	7
4.1. Parent-Child Ingestion Strategy .....	7
4.2. Data Streaming (Real-Time via Event Hub / IoT Hub) .....	7
4.3. Data Batch Processing (via Azure Data Factory) .....	8
5. Lakehouse Architecture .....	8
5.1. Bronze Layer (Raw Zone) .....	8
5.2 Silver Layer (Curated Zone) .....	9
5.3 Gold Layer (Aggregated Zone) .....	9
5.4. Consumer Layer - Data Consumption & Access .....	9
6. Pipeline Engineering .....	10
6.1 Pipeline Design Strategy .....	10
6.2 Pipeline Failure Strategy and Monitoring .....	11
7. Conclusions and Future Scope .....	11
Conclusion .....	11
Future Scope .....	11
8. References .....	12

## 1. Abstract

This project outlines the design and implementation strategy of a modern, Azure-based cloud data solution to unify healthcare and administrative data for Alberta clinics. Using both batch and real-time ingestion strategies and Delta Lake architecture, the goal is to centralize, clean, and deliver actionable insights for clinics, staff, and patients through secure, scalable, and reliable pipelines.

## 2. Project Overview

### 2.1. Mission, Objectives, and Purpose

**Mission:**

To design and implement a scalable Azure-based cloud data platform that unifies clinical and operational data for Alberta clinics, enabling secure, real-time insights through a modern Lakehouse architecture.

**Objectives:**

- Centralize clinic data using Azure Lakehouse.
- Enable real-time and batch data processing.
- Deliver clean data for analytics and reporting.

### 2.2. Why Alberta Clinics Need This Integration

Many Alberta clinics still rely on Excel sheets and siloed digital tools, making it hard to manage patient flow, track health outcomes, and report effectively. A unified data platform bridges the gap between traditional processes and modern data-driven care.

### 2.3. Target Outcomes for Patients and Providers

- Timely access to diagnostics and feedback data
- Unified patient and staff management
- Improved reporting for compliance and funding
- Streamlined patient experiences through real-time updates

### 3. Proposed Cloud Architecture Strategy

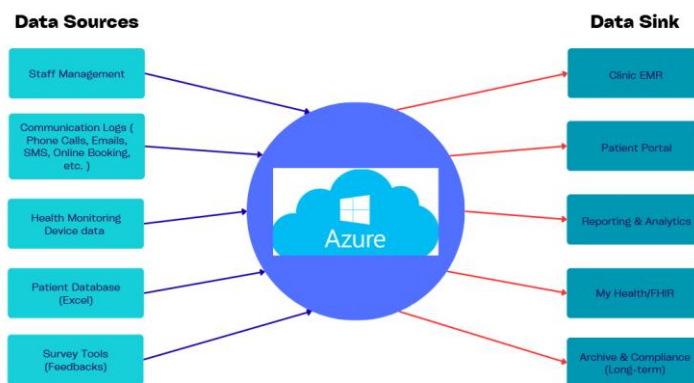
#### 3.1. Data Sources

Data Source	Nature	Format	Usage
Staff Management	Structured	SQL, JSON	Schedules, logins, HR records
Patient booking	Semi-structured	JSON, CSV	Calls, SMS, emails, Online bookings
Patient Database (CSV/XLSX)	Structured	CSV, XLSX	Patient demographics, history
Health Monitoring Device Data	Semi-Structured	JSON, Parquet	To gauge the condition of health devices.
Survey Tools (Feedbacks)	Semi-structured	JSON, CSV	Feedback, satisfaction forms

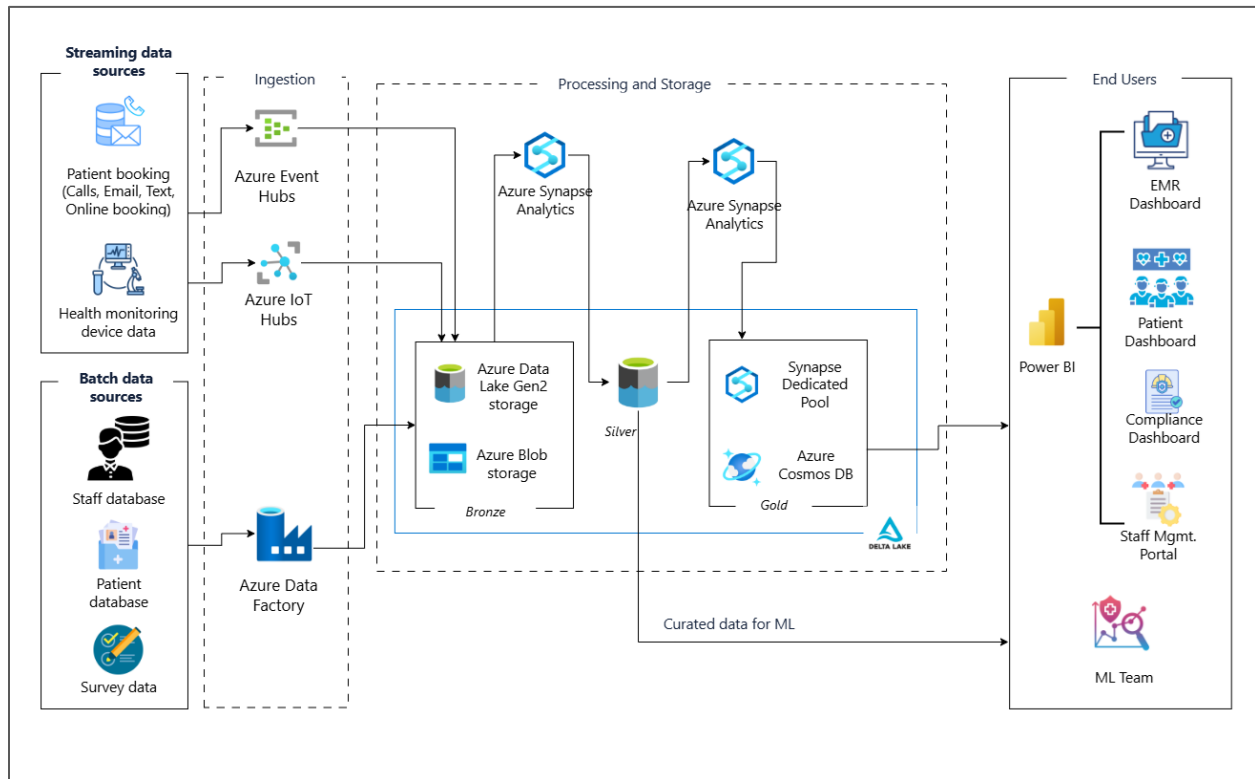
### 3.2. Data Sinks

Data Sink	Purpose	Details
EMR Dashboard	Store clinical records	Power BI
Patient Dashboard	Patient access to data	Power BI
Machine Learning Team	Predictive Analytics	SQL table, Azure Data Lake Gen 2 Storage
Staff Management Portal	Manage staff & operations	Cosmos DB, Power BI
Compliance Dashboard	Long-term retention & audit	Cosmos DB, Power BI

### 3.3. High-level Azure-Based Vision Diagram



### 3.4. Cloud Architecture Diagram



### 3.5. Key Azure services used

- **Azure Data Factory** (Batch ingestion)
- **Azure Event Hub / IoT Hub** (Streaming)
- **Azure Data Lake Gen2, Azure Blob Storage** (Bronze layer)
- **Azure Synapse Analytics** (Processing & SQL pools)
- **Azure Cosmos DB** (Query & API Access)
- **Power BI** (Visualization)

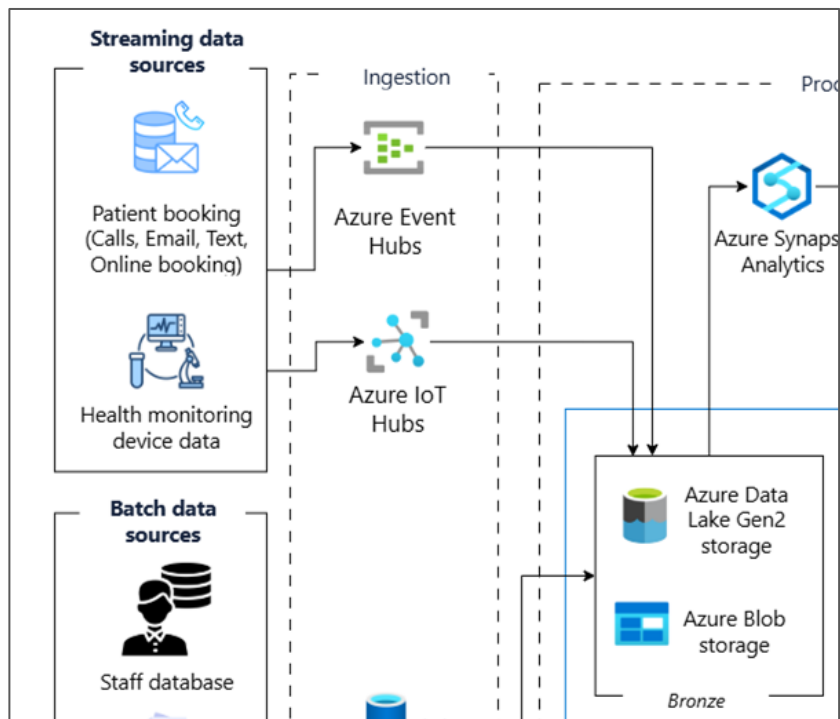
## 4. Data Ingestion Strategy

### 4.1. Parent-Child Ingestion Strategy

Data is modeled in parent-child relationships (e.g., Appointment → WebBooking, PhoneBooking). This reduces duplication and supports scalable data modeling.

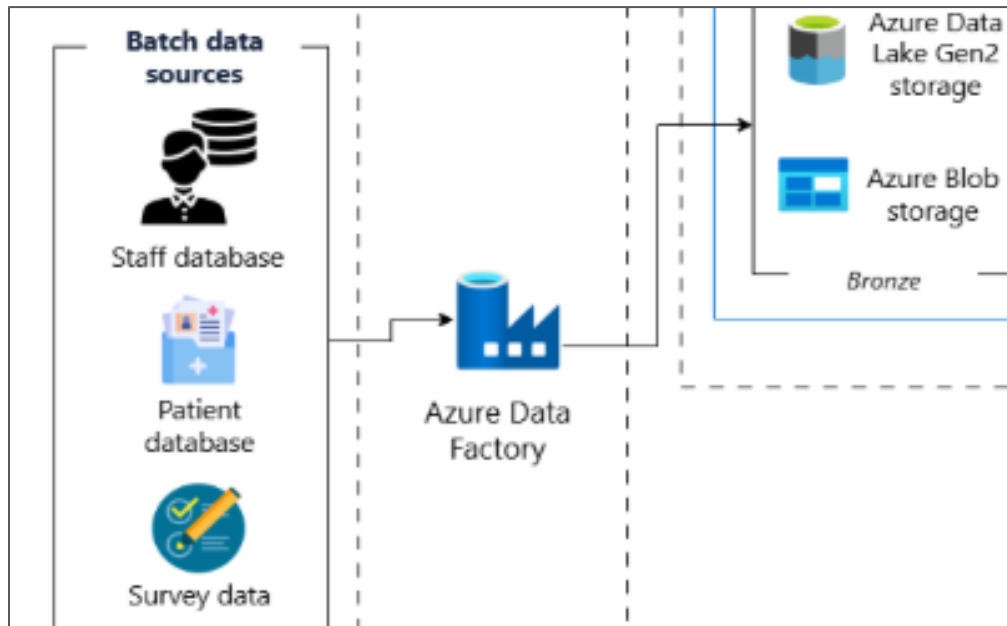
### 4.2. Data Streaming (Real-Time via Event Hub / IoT Hub)

- **Uses Azure Event Hub** and **Azure IoT Hub** for ingesting live data.
- **Sources:** Device Readings, Online Bookings, Communication logs.
- Real-time data flows into **Bronze Layer** for raw storage.



### 4.3. Data Batch Processing (via Azure Data Factory)

- Managed by **Azure Data Factory (ADF)**.
- Sources: Patient Data, Staff Management Data, Survey tools.
- Triggered pipelines handle scheduled uploads to Data Lake.



## 5. Lakehouse Architecture

### 5.1. Bronze Layer (Raw Zone)

- Landing zone for raw, unprocessed data from all data sources.
- **Storage:** Azure Blob Storage & Azure Data Lake Gen 2 Storage
- **Format:** Original format (CSV, JSON, Parquet)
- **Purpose:** Immutable raw data storage
- **Partitioned by:** Date, Source, Entity



## 5.2 Silver Layer (Curated Zone)

- Cleaned, deduplicated, and normalized data.
- **Tool:** Azure Synapse Analytics
- **Parent-child structure:**
  - Merge child source tables into clean parent models
  - Example: Online Booking, PhoneBooking → Appointments
- Stored in **Silver Layer (Curated Zone)**
- Provides curated data for Machine Learning purpose.

## 5.3 Gold Layer (Aggregated Zone)

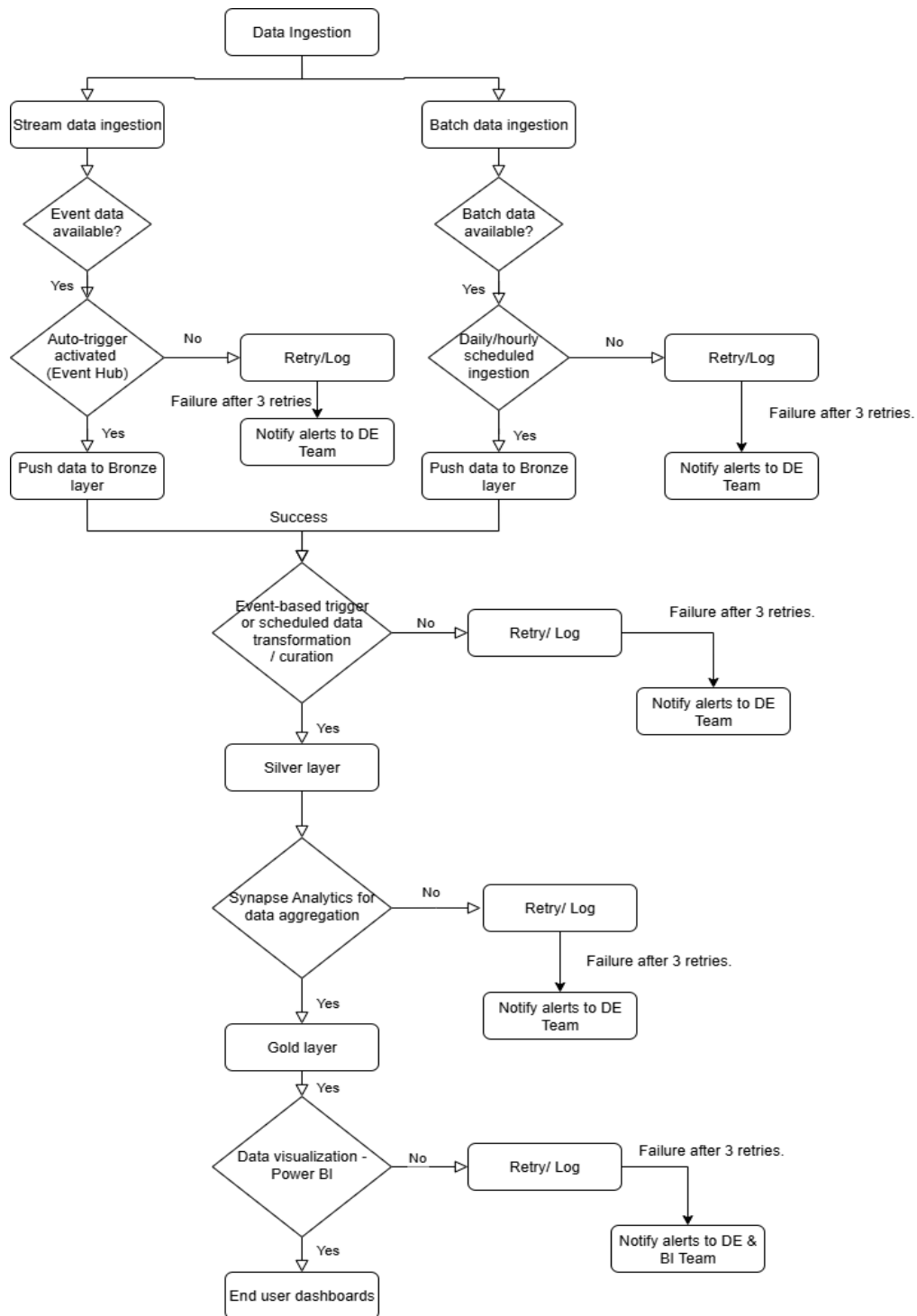
- **Tool:** Synapse Pipelines
- Aggregated KPIs, Dashboard-ready datasets
- Stored in **Gold Layer (Aggregated Zone)**

## 5.4. Consumer Layer - Data Consumption & Access

- Provide end-to-end data to end users and applications.
- Access via **Power BI**.
- **EMR Dashboard:** View patient vitals, clinical history, medication, diagnostics.
- **Patient Dashboard:** View personal health records, appointment logs, test results
- **Compliance Dashboard:** Track adherence to healthcare regulations.
- **Staff Management Portal (Dashboard):** Track staffing, scheduling, certifications, and attendance.
- **ML Team:** Perform predictive analytics, model training, anomaly detection.
- Role-based access control with Azure AD.

## 6. Pipeline Engineering

### 6.1 Pipeline Design Strategy



## 6.2 Pipeline Failure Strategy and Monitoring

Component	Fail Strategy	Monitoring Tool
ADF Pipelines	Retry policy & Alert	Azure Monitor, ADF Alerts & Activity Runs
Event Hub	Capture dead-lettered events to Blob	Diagnostic Logs & Azure Log Analytics
IoT Hub	Durable functions & Queue Backups	Azure Monitor, App Insights & IoT Hub Metrics
Azure Synapse Analytics (Spark pools/SQL pools)	Try/Except blocks with logging to ADLS, timeout configs & automatic retries	Synapse Workspace Logs, monitoring dashboard, Spark Job Monitoring
Azure Data Lake Gen 2 Storage/ Blob Storage	Enable Soft Delete & Versioning	Azure Storage Metrics & Monitor logs
Cosmos DB	Throughput auto-scaling + TTL strategy	Cosmos DB Insights & Diagnostic Logs

## 7. Conclusions and Future Scope

### Conclusion

This Azure-based Lakehouse architecture brings real-time decision-making and operational clarity to Alberta clinics. It reduces data fragmentation, automates workflows, and delivers high-quality insights.

### Future Scope

- Add predictive analytics and ML pipelines
- Expand to more clinics and provinces

## 8. References

- Microsoft Azure Docs (Data Factory, Event Hub, Synapse)
- DP-203 Certification Labs and Best Practices