Cloud Architecture Proposal for Canadian Blood Services

Table of Contents

Purpose

This document serves as a proposal for the migration of Canadian Blood Services' existing database system to a modern cloud-based architecture. The initiative is driven by the need to address current system limitations and to leverage the benefits of cloud computing, including scalability, disaster recovery, and enhanced security.

Objectives

The primary objective is to migrate the existing RDBMS to a cloud platform, specifically Microsoft Azure, to achieve:

- Improved scalability to handle varying loads.
- Enhanced disaster recovery capabilities.
- Strengthened security measures to protect sensitive data.

Existing Storage System

Canadian Blood Services relies on an on-premises Relational Database Management System (RDBMS) to handle a variety of critical datasets integral to their operations. The core components of the existing database system include:

Blood Donor Information	Comprehensive data on blood donors for tracking donor history, managing donation records, and maintaining donor contact information for future engagement and notifications.
Blood Inventory	Detailed inventory tracking to ensure the availability and safety of blood supplies.
Patients Database	Medical history or patient-related information to facilitate the matching of blood products to recipients.

The existing system is designed to prioritize the integrity and availability of data, which is critical for the healthcare sector. However, there are several challenges associated with this on-premises RDBMS:

Problem Statements:

Limited Data Visibility: Current systems lack real-time data visibility, hindering efficient blood monitoring and allocation across centers.

Scalability Constraints: On-premises infrastructure struggles with sudden demand changes, causing bottlenecks during increased blood product needs or donor registrations.

Disaster Vulnerability: Reliance on localized data storage poses risks of data loss during disasters, impacting operations and life-saving services.

Vision

The vision is to integrate core systems such as hospital and clinic appointment systems, patient databases, and inventory management into a unified cloud platform. This integration will be supported by advanced analytics, geo-location services, and a robust alert and notification system, all hosted on Azure.

Below diagram illustrates the vision for the project.

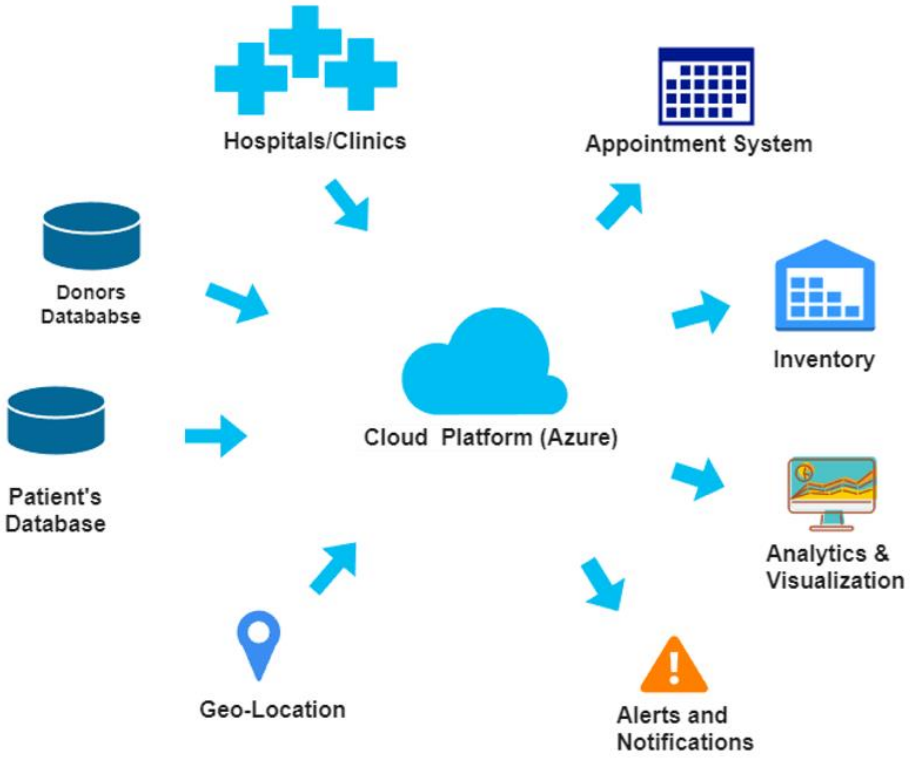

Figure 1: Vision Diagram

Proposed Cloud Architecture

Below is the proposed cloud architecture which will utilize Azure services to create a robust and scalable system:

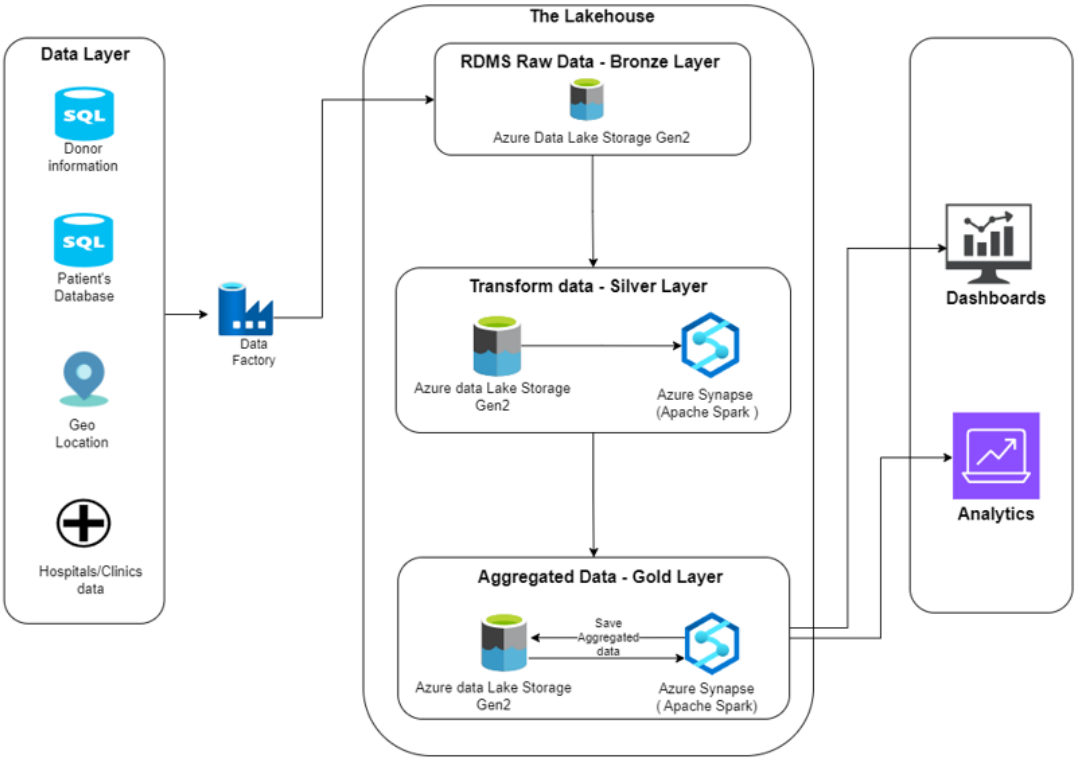

Figure 2: Proposed Cloud Architecture

The architecture will be based on a Lakehouse Data Layer with Bronze, Silver, and gold layers to manage raw, transformed, and aggregated data respectively, using Azure Data Lake Storage and Azure Synapse Analytics (Apache Spark). This layer is structured into three distinct zones:

Bronze Layer: This initial layer within Azure Data Lake Storage serves as the landing zone for raw data from sources such as patient, donor, Hospitals databases and geo-location information. The raw data will be ingested as a batch utilizing Azure Data Factory.
Silver Layer: Data will undergo transformation in this layer, Azure Synapse Analytics (Apache Spark) to refine and structure the data for further analysis.
Gold Layer: The most processed and aggregated data will be stored here, ready for consumption by end-user applications and for generating insights.

Data Pipeline Design

Below diagram illustrates the pipeline design utilizing parent-child approach:

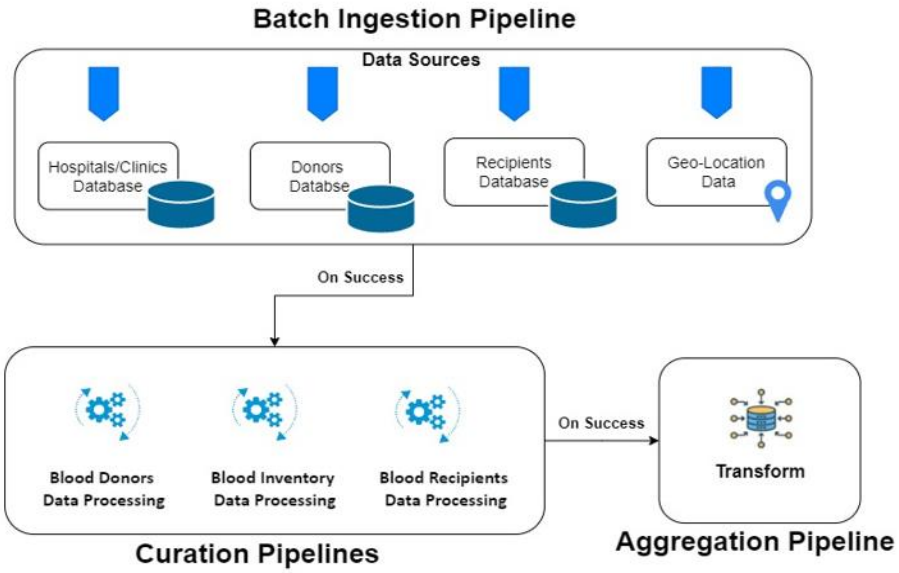

Figure 3: Pipeline design

➢ Components

The data pipeline will consist of the following components:

1. **Batch Ingestion Pipeline:** To efficiently ingest data from various sources such as hospitals, donors, and recipients.
2. **Curation Pipeline:** To clean, remove duplicates, process and refine the data. This pipeline will be triggered on successful execution of batch ingestion pipeline.
3. **Aggregation Pipeline:** To process and transform data for analytics and reporting. This pipeline will be triggered on successful execution of curation pipeline.

➢ Ingestion Strategy

Batch Processing: Chosen for its efficiency with stable data, reduced resource consumption, and simplified management.

Batch Frequency: Data is ingested every 24 hours at 12 am.

➢ Pipeline Triggers

Below are the two triggers for the pipeline to execute:

Schedule Trigger	Runs at regular intervals to process new donor registrations and inventory updates.
Event-based Trigger	Checks blood availability upon requests from hospitals.

Deployment & Data Access

End-users will access the final data through an API App, which visualizes key blood bank metrics such as blood type availability by location.

Security

The architecture will incorporate below two tools to ensure the secured data access,

1. **Role-Based Access Control (RBAC):** will manage who has access to specific data within the system.
2. **API Authentication:** Secured by Azure Active Directory (AAD) to ensure that only authorized users can access the API app.

Constraints

The constraints for this cloud architecture includes:

a. **Budget Limitations:** The need to work within a predefined budget for implementing cloud architecture.

Future Scope

The future scope of this cloud architecture includes:

- Continuous improvement and optimization of cloud services and components.
- Integration of advanced analytics and machine learning capabilities.

Acknowledgement

I would like to express my gratitude for the opportunity to develop this proposal. My appreciation also goes to Junaid Qazi for his guidance throughout this process. Thank you for considering this proposal.