BBI

PLATFORM MODERNIZATION WITH AZURE

https://www.bbinsight.com/

PROJECT SUMMARY

A health data leader in a niche space required a custom data platform capable of generating accurate insights into patient care through data science and analytics

BBI designed a modern, cloud-native solution that supported product development and can easily scale for future initiatives.

THE CHALLENGE

Our client handles massive amounts of confidential patient data. The new platform needed to ingest, process, and store this data to support product development. The existing system had a number of challenges, including:

- **Inefficient Ingestion Process**: Various customer data was ingested through a single job, causing delays when new data arrived before the previous customer's ingestion completed. The job was hosted on Virtual Machines that required maintenance.
- **Orchestration**: Custom .NET core scripts handled orchestration, but lacked features like function chaining and dependency management. Orchestration involved SQL tasks on a Snowflake database, making it hard to maintain.
- **Minimal Data Discovery & Understanding**: There was no data catalog or data glossary, and data quality metrics were not reported for incoming data.
- **Lack of Access Control**: Role-based access control (RBAC) was not well defined, posing security risks. Data access was not audited, potentially leading to unauthorized access.
- **Missing Monitoring and Reporting**: System had no pipeline reports, proactive failure notifications to data consumers. monitoring of data pipelines, or SLA tracking.

To address these challenges, we needed to ingest and catalog data from several source types into both a data lake and data warehouse, choose the right ETL tool to support the future data core platform, migrate the legacy processes to a cloud-native solution, and implement a cloud-native operational database system to service real-time web application operations.

Tooling Highlights

- *Snowflake*

- *Snowpark*

- *Azure Suite: SQL, Data Factory, Storage*

- *PySpark*

- *dbt*

THE SOLUTION

BBI designed a new platform that optimized data processing, security, and management while reducing operational overhead through the use of Azure's serverless and automation capabilities.

Our approach:

- Implement ingestion Azure Functions to store and catalog data in an Azure Storage Data Lake.
- Transform data into a Snowflake EDW using Snowpark and Data Factory for scheduling.
- Build operational database and servicing API using Azure SQL and Azure Functions.
- Migrate legacy SQL Server data to Data Lake using PySpark.
- Automate deployments using Azure DevOps pipelines.

New Capabilities

- *Automated data ingestion.*

- *Process flows in Azure ecosystem.*

- *Operationalized the data ingestion and data processing.*

- *Reusable data migration process.*

- *Implemented best practices for Snowpark development*

- *Automated database updates, code deployments, and infrastructure deployments.*
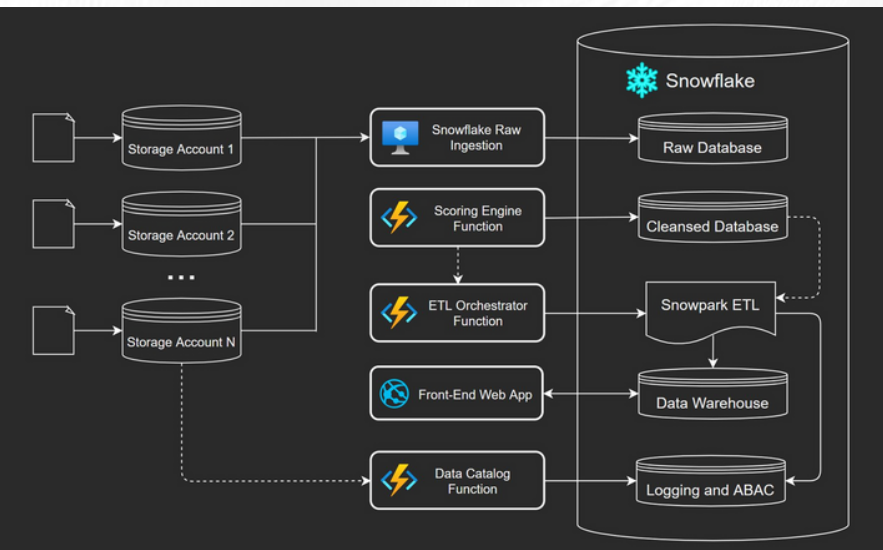
We enhanced the platform with our pre-built solution accelerators:
- ARM Template: Collection of ARM templates used to quickly standup ingestion Azure Functions and Storage Accounts
- Data Validation: PySpark scripts used to validate the integrity and equality of data from different systems.
- dbt Execution: Python scripts used to configure dbt against a target system, execute models, and return results.

BBI   https://www.bbinsight.com/