**CUSTOMER 360: The Customer’s Choice**



**Customer 360** is a modern, cloud-based retail analytics platform designed to streamline online shopping experiences. It integrates customer behavior, transaction patterns, and sales performance into a unified system. With interactive dashboards, personalized insights, it empowers businesses to make data-driven decisions and enhance customer satisfaction.



LAKKAKULA SAI AKASH

**Table of Contents**

1. [Problem Statement](https://anblicks-my.sharepoint.com/:w:/p/saiakash_lakkakula/ERcEwO-ClfZNq0N7wvBGdUsB2QaBvXGHbuAH-LSdfL_rjw?e=gWqTRa&nav=eyJoIjoiOTY4ODI4MDgxIn0%3D)
2. Software Requirements Specification
3. Architecture
4. Project Implementation Steps
5. Results
6. Outcomes of Analysis
7. Future Scope
8. References

**Problem Statement**

Customer 360 is focused on delivering a unified, data-driven understanding of customer behavior to support strategic business decisions. However, critical customer-related data—such as **customer profile details**, **support ticket information**, **transaction records**, and **web interaction logs.** This fragmentation makes it challenging to consolidate and analyze the data to gain a complete and accurate view of the customer journey.

This project aims to build a **robust and scalable data pipeline** that integrates all relevant customer data into a centralized data warehouse using **Snowflake**. By consolidating diverse datasets into a single source of truth, Customer 360 can eliminate data silos and ensure consistency across analytical outputs.

Data transformation processes will be carried out using **SQL and PySpark**, including data cleansing, normalization, and enrichment. The curated data will be visualized using **Power BI dashboards**, offering insights into customer behavior, issue resolution patterns and engagement metrics.

With this end-to-end architecture, Customer 360 will be able to deliver accurate analytics, enhance customer experiences, and enable data-driven decision-making across the organization.

**Software Requirements Specification**

**Technologies Used:**

1. **Azure Blob Storage**  
   Serves as the **initial storage location** for raw .xlsx files containing customer profile details, transaction records, support ticket data, and web interaction logs. Azure Blob Storage offers scalable, cost-effective storage with native integration into Azure services.
2. **Azure Data Lake Storage (ADLS)**  
   Functions as the **intermediate storage layer** for Delta files produced by Databricks. ADLS provides a hierarchical namespace and is optimized for large-scale analytical workloads, supporting efficient data access and management.
3. **Azure Databricks**  
   Used to read and transform .xlsx files from Blob Storage using **PySpark**. Databricks allows for scalable and distributed data processing. The transformed outputs are written in **Delta format** and stored in ADLS for downstream loading.
4. **Snowflake**  
   Serves as the **centralized cloud data warehouse**. Data is ingested from Delta files stored in ADLS using **external stages**. Snowflake manages structured data across a multi-layer architecture (e.g., Bronze, Silver, Gold) and supports high-performance querying and schema management.
5. **Power BI**  
   Provides **interactive visualizations and reporting** based on the curated data in Snowflake. Power BI dashboards enable business stakeholders to access real-time insights on customer behavior, transactions, support interactions, and engagement trends.

**Software Requirements:**

1. Microsoft Azure subscription
2. Snowflake account
3. Power BI account
4. Databricks workspace

**Architecture:**

The data pipeline architecture follows a **modular, cloud-native design** that supports scalable data processing and analytics. It consists of several interconnected layers that enable ingestion, transformation, storage, and visualization of data.

**Data sources:**

The following .xlsx files serve as input to the pipeline. Each file corresponds to a core area of customer data required to generate the unified Customer 360 view. All files are initially stored in **Azure Blob Storage** and later transformed and processed in the pipeline.

1. **Customer\_master.xlsx:**

**Description:** Contains customer personal and identifiable information

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| customer\_id | Unique identifier for each customer |
| name | Full name of the customer |
| email | Email address |
| phone | Contact number |
| signup\_date | Date the customer registered on the platform |
| location | Customer's geographical location |

1. **Customer\_support.xlsx**

**Description:** Logs support ticket details raised by customers.

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| **ticket\_id** | Unique ID of the support ticket |
| **customer\_id** | Identifier linking the ticket to a customer |
| **created\_at** | Timestamp when the ticket was raised |
| **resolved\_at** | Timestamp when the ticket was resolved |
| **issue\_type** | Category or nature of the issue reported |
| **satisfaction\_score** | Customer’s rating of support experience (e.g., 1–5) |

1. **Transaction.xlsx:**

**Description**: Contains details of transactions made by customers.

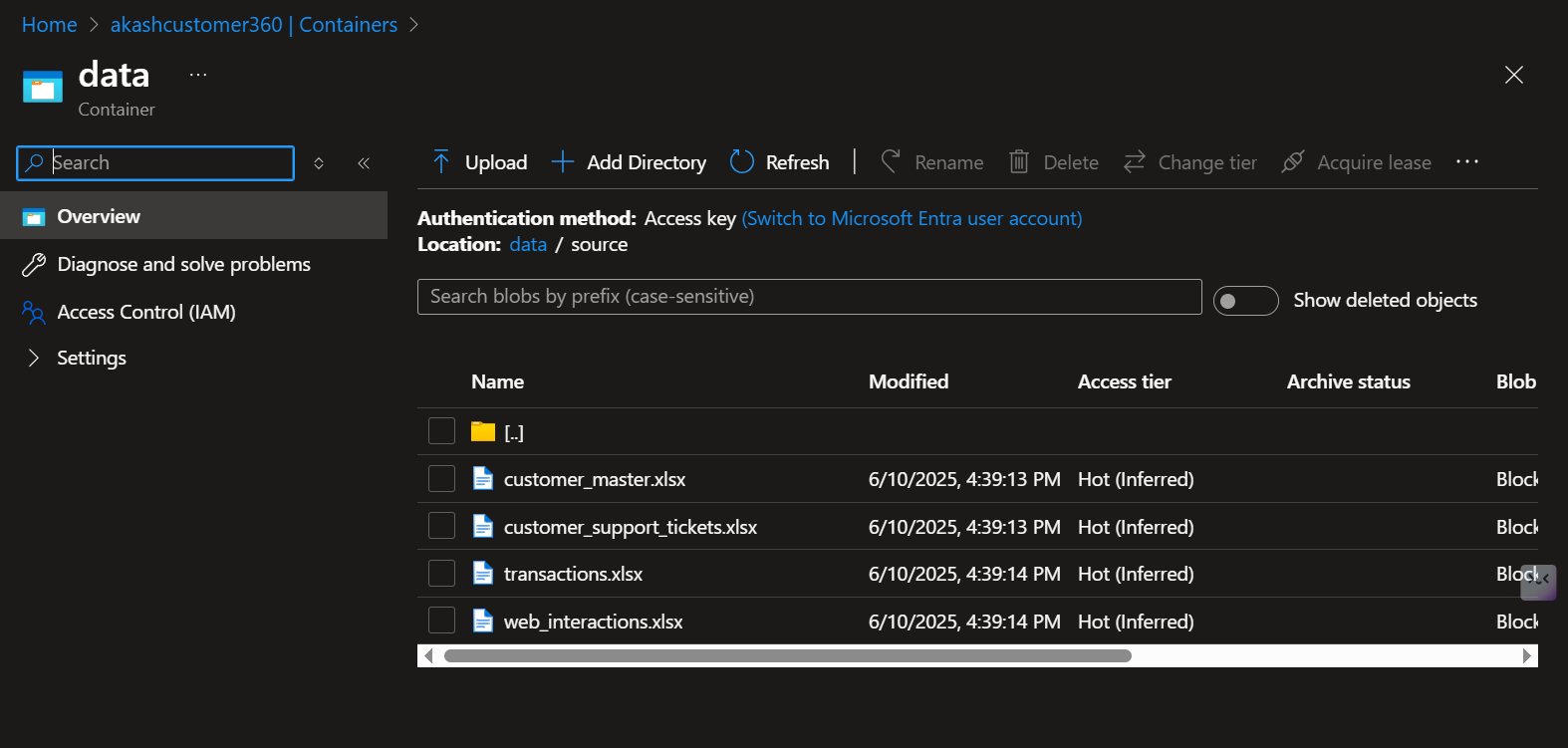
|  |  |
| --- | --- |
| **Field Name** | **Description** |
| **transaction\_id** | Unique transaction identifier |
| **customer\_id** | Customer who performed the transaction |
| **transaction\_date** | Date the transaction occurred |
| **amount** | Value of the transaction |
| **payment\_method** | Mode of payment (e.g., credit card, UPI) |
| **store\_location** | Store or service location where transaction happened |

1. **Web\_Interactions.xlsx:**

**Description:** Captures customer browsing behavior and session data

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| **session\_id** | Unique session identifier |
| **customer\_id** | Customer associated with the session |
| **page\_views** | Total number of pages viewed in the session |
| **session\_duration** | Total time spent in session (in seconds/minutes) |
| **device** | Device used for the session (e.g., mobile, desktop) |
| **interaction\_date** | Date the interaction occurred |

As the data is present in the blob storage initially like this :



**Key Features:**

The solution delivers a scalable, cloud-native data pipeline designed to unify customer data and enable business intelligence. Below are the key features of the implementation:

### **1. Unified Customer View**

* Consolidates customer profile, support, transaction, and interaction data into a single analytical model.
* Enables 360-degree visibility into customer behavior and engagement.

### **2. Scalable Cloud Architecture**

* Built on Azure and Snowflake, supporting large volumes of structured and semi-structured data.
* Easily extensible to incorporate additional data sources or subject areas.

### **3. Automated Data Ingestion and Transformation**

* Supports seamless ingestion of .xlsx files from Azure Blob Storage.
* Data transformation and enrichment are performed using **PySpark in Azure Databricks**.
* Transformed data is stored in Delta format for optimized performance.

### **4. Integration via External Stages**

* Uses **Snowflake external stages** to load Delta files from ADLS efficiently without duplication or unnecessary movement of data.

### **5. Layered Data Warehouse Design**

* Follows the **Bronze-Silver-Gold** schema approach in Snowflake:
  + **Bronze**: Raw ingested data.
  + **Silver**: Cleaned and validated data.
  + **Gold**: Business-ready analytical views.

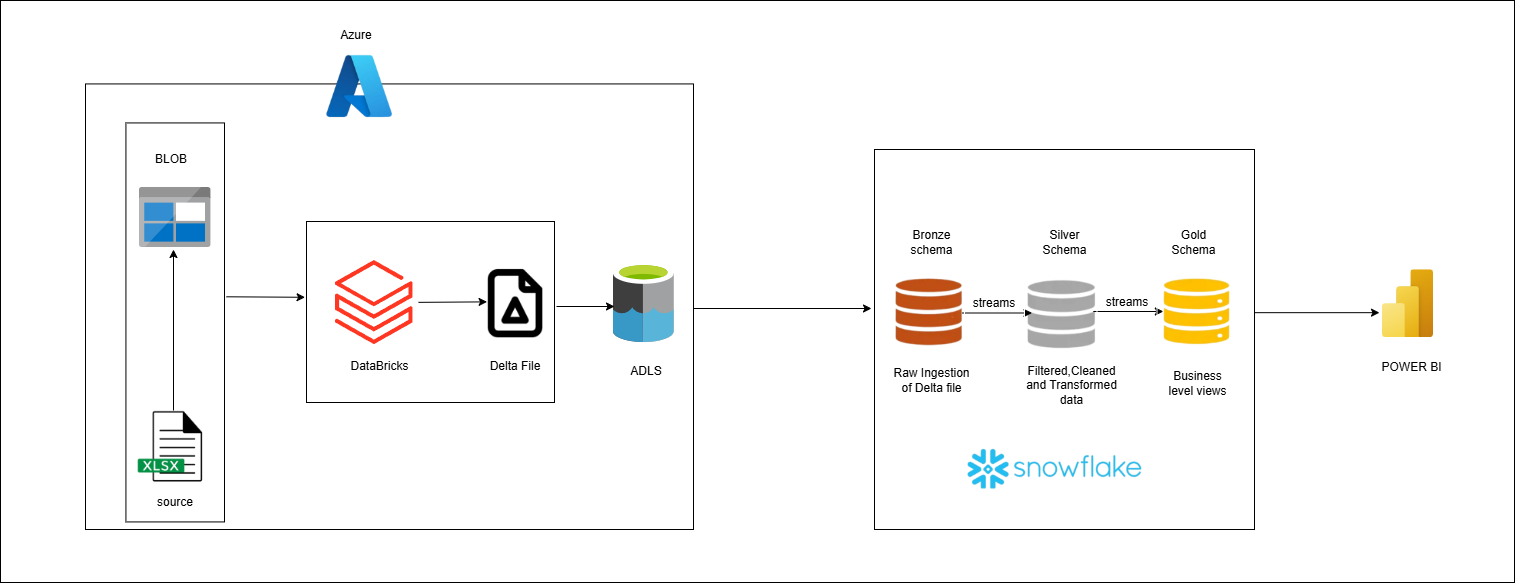
### **6. Real-Time Interactive Dashboards**

* Connects Power BI directly to Snowflake's Gold layer.
* Provides rich visualizations and filters for customer insights, support performance, and transaction trends.

### **7. Modular and Maintainable**

* Each layer of the pipeline (ingestion, transformation, storage, visualization) is logically separated and independently manageable.
* Facilitates easy debugging, scaling, and auditing.

**Architecture:**



**Project Implementation Steps**

### **Step 1: Upload Source Files to Azure Blob Storage**

* Upload the following .xlsx files to Azure Blob Storage:
  + Customer\_master.xlsx
  + Customer\_support.xlsx
  + Transaction.xlsx
  + Web\_Interactions.xlsx
* These files contain critical data required to build the Customer 360 view.

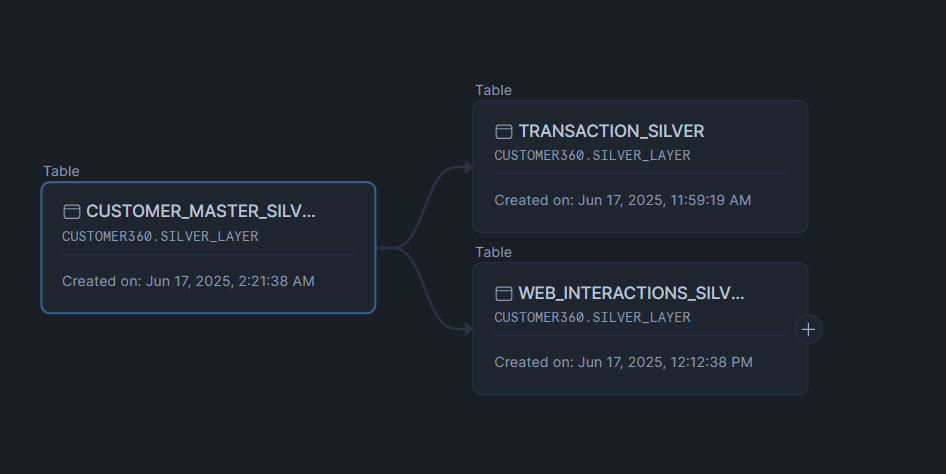
### **Step 2: Convert .xlsx Files to Delta Format using Azure Databricks**

* Launch an **Azure Databricks notebook** to:
  + Read .xlsx files directly from Azure Blob Storage using PySpark.
  + Perform basic schema inference and data type handling (no transformations).
  + Write the datasets into **Delta format**.
* Store the resulting Delta files in **Azure Data Lake Storage (ADLS)** in organized folders, one per dataset.

### **Step 3: Load Delta Files into Snowflake via External Stages**

* In **Snowflake**, configure:
  + **External Stages** pointing to the Delta file locations in ADLS.
  + **File formats** for reading Delta (or converted Parquet files) using Snowflake’s external table support.
* Use COPY INTO commands to ingest the raw data into **Bronze tables** in Snowflake.

### **Step 4: Create Streams on Bronze Tables**

* Define **streams** on each Bronze table to track newly inserted records.
* These streams will feed into the transformation logic for Silver tables.
* 

### **Step 5: Implement Stored Procedures for Bronze to Silver Transformation**

* Create **stored procedures** to:
  + Read from Bronze table streams.
  + Perform all **data transformations**, including:
    - Null handling
    - Data cleaning
    - Field formatting
    - Business rule application
  + Load transformed data into **Silver tables**.
  + Schedule these procedures using **Snowflake Tasks** or invoke them manually.

### **Step 6: Configure Streams on Silver Tables**

* Define streams on **Silver tables** to capture newly inserted clean records.
* These will trigger updates to the **Gold schema**.

### **Step 7: Automate Gold Layer Population**

* Develop additional **stored procedures or tasks** to populate **Gold tables and views** using Silver stream data.
* This includes:
  + Aggregations
  + Derived KPIs
  + Business metrics

### **Step 8: Define Views in Gold Schema**

* Create business-level **views** for reporting and analytics:
  + vw\_customer\_profile
  + vw\_transaction\_summary
  + vw\_support\_kpis

### **Step 9: Build Dashboards in Power BI**

* Connect **Power BI** to the **Snowflake Gold schema** using a direct connector.
* Create dashboards to visualize:
  + Customer engagement
  + Support efficiency
  + Transaction metrics
  + Session activity

### **Step 10: Schedule and Monitor**

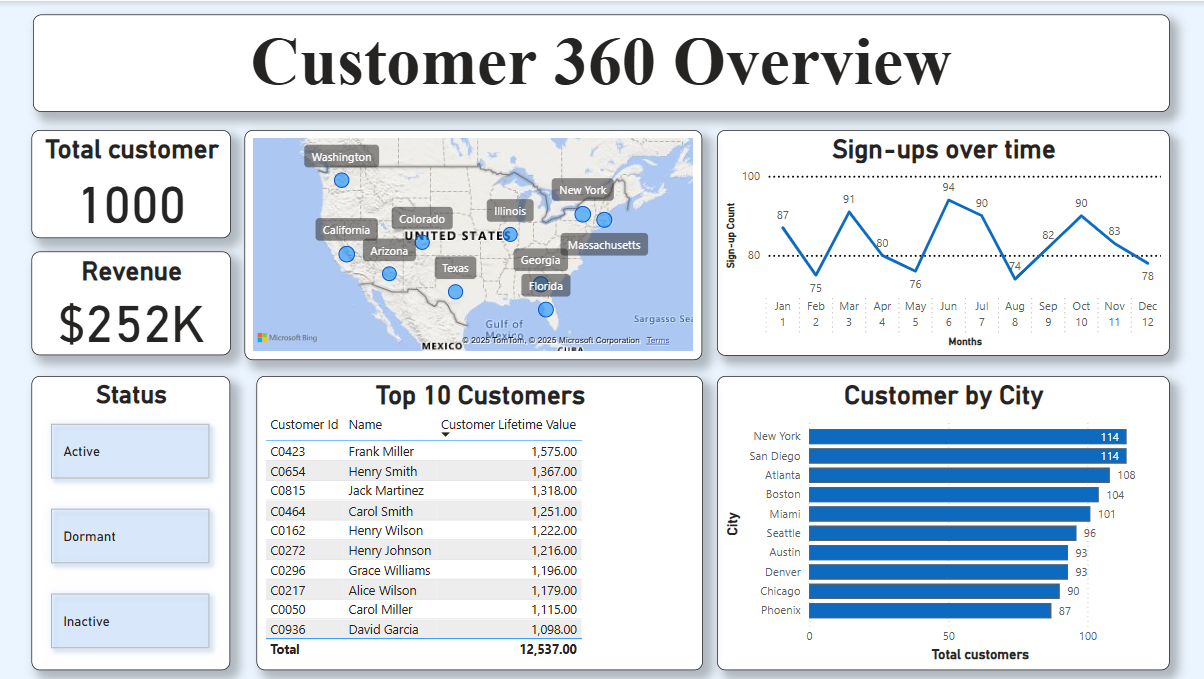
* Automate Databricks notebooks for periodic file format conversion.
* Use Snowflake **Tasks** to run stored procedures at set intervals.
* Monitor stream consumption and data load success/failures.

**Results**

The **Customer 360 Overview Dashboard** serves as the central visualization layer for stakeholders to monitor customer metrics, behavioral trends, and revenue performance in real-time. Built using **Power BI** and connected to the **Gold schema** in **Snowflake**, this dashboard integrates and visualizes data from multiple domains—customer details, transactions, support, and web interactions.

**Dashboard 1: Customer 360 Overview:**

The Customer 360 Overview Dashboard is the primary interface for visualizing unified customer data across multiple domains. Powered by Power BI and connected to the Gold layer in Snowflake, this dashboard consolidates key performance indicators and behavioral metrics into a single interactive view.



**Outcomes of the Analysis**

The analysis derived from the **Customer 360 Overview Dashboard** provides several actionable insights that can directly support strategic business initiatives, customer engagement, and revenue optimization.

### **1. Comprehensive Customer Understanding**

* Consolidation of customer data across systems (profiles, transactions, support, and interactions) provides a **single source of truth**.
* Enables accurate segmentation and personalized targeting.

### **2. Identification of High-Value Customers**

* The **Top 10 Customers** by lifetime value highlight individuals contributing significantly to total revenue.
* These customers can be prioritized for **loyalty programs**, **exclusive offers**, or **retention campaigns**.

### **3. Revenue Performance Monitoring**

* A total of **$252K in revenue** indicates the current value captured from all customer transactions.
* This enables tracking of financial performance across different periods and customer segments.

### **4. Regional Market Insights**

* Customer concentration in key cities (e.g., **New York**, **San Diego**, **Atlanta**) highlights **high-performing regions**.
* Supports **location-based marketing** and **resource allocation planning** (e.g., regional sales or support staff).

### **5. Trend Analysis of Customer Sign-ups**

* Monthly sign-up trends help in identifying **successful acquisition periods** (e.g., peaks in April and June).
* Dips in other months can indicate areas for improvement in campaigns or customer onboarding processes.

### **6. Status-Based Segmentation**

* The classification of customers into **Active**, **Dormant**, and **Inactive** supports:
  + **Re-engagement strategies** for dormant users.
  + **Churn prediction** and proactive customer retention for inactive users.