Logo

Description automatically generated with medium confidence

**ANL555 - Data Integration for Enterprise Automation**

**Group-based Assignment**

**January 2025 Semester**

**Group 2**

**Leader:** Wang Yiming (E2510862)

**Member 1:** Li Jian (Q2472519)

**Member 2:** Teng Han Leng(CE250886)

**Member 3:** Leow Yu Fan (B1681528)

**Submission Date:** 25 March 2025

# Background

As part of the telco company's digital transformation initiative, our data integration and automation team is tasked with addressing current challenges related to customer communication during network faults. Currently, customer and network fault information are manually extracted into CSV files from Juniper Networks and two CRM systems—an on-premises CRM (Amdocs) for post-paid customers and a cloud-based CRM (Salesforce.com) for pre-paid mobile and fibre broadband customers—and processed in Excel. This manual approach poses data privacy risks and causes delays in proactive customer notifications.

Our objectives include:

* Secure, real-time access to network service data from Juniper Networks.
* Secure integration of customer data from CRM systems (Amdocs and Salesforce.com).
* Automated proactive customer notifications via email and/or mobile app.
* Safe storage of integrated data in Snowflake Data Lake for analytics and predictive monitoring.

# Question 1:

## (a) Personally Identifiable Information (PII) Data Privacy Issues

### **Issue 1**: Insecure Manual Data Extraction and Handling

**Source Systems**: Amdocs, Salesforce.com

**Description of Issue**:

Currently, customer details and service fault information are manually extracted from Juniper Networks and different CRM systems as CSV files and processed in Excel. This manual handling significantly increases the risk of unauthorized access, accidental exposure, and improper handling of sensitive customer information.

**Integrated Safeguards:**

* **Secure API Integration**: Replace manual processes with secure API integrations, implementing strong authentication and encryption.
* **Role-Based Access Control (RBAC):** Ensure data access is restricted solely to authorized personnel.
* **Selective Data Retrieval**: Exclude unnecessary and sensitive fields, such as customer National Registration Identity Card (NRIC) numbers, to minimize unnecessary exposure.
* **Tokenization**: Apply tokenization to sensitive fields (mobile numbers, emails) during data transfer.
* **Transport Layer Security (TLS):** Ensure all data transfers are encrypted securely.

**Relevant Example:**

A notable incident underscoring the critical importance of secure data handling practices involved the e-commerce platform Carousell. Due to insufficient security on its APIs, personal data belonging to approximately 2.6 million users was inadvertently exposed and sold on a hacking forum ([Hamzah, 2024](https://www.straitstimes.com/singapore/carousell-fined-58k-for-data-breaches-including-one-where-data-of-26m-users-were-sold-on-hacking-forum)). This case highlights the necessity of robust API security measures and controlled data extraction practices.

### **Issue 2**: Error Resulted from Manual Cross-Referencing and Matching of Sensitive Data in Excel

**Source System:** Amdocs, Salesforce.com

**Description of Issue:**

The current manual cross-referencing process in Excel poses significant risks. Sensitive customer data from different sources is manually combined, increasing chances of errors, data mismatches, unintended exposure, or leakage during handling.

**Integrated Safeguards:**

* **Automated Data Matching:** Implement an automated integration and matching workflow through secure data integration platforms (e.g., IPaaS such as Workato).
* **Secure File Transfer and Encryption:** If intermediate storage is necessary, use secure file transfer protocols (SFTP) and strong encryption (e.g., AES-256) for interim storage.

**Relevant Example:**

An example underscoring similar risks is a data handling error at Health Sciences Authority Singapore, where manual processes and human error led to the accidental exposure of blood donors' personal details due to the database being placed on a server accessible by the Internet with inadequate safeguards to limit unauthorised access ([Kurohi & Choo, 2019](https://www.straitstimes.com/singapore/health/personal-information-of-over-800000-blood-donors-exposed-online-hsa)). This highlights the importance of automating sensitive data handling to minimize human error and data exposure risks.

## (b) Recommended Data Integration Solution

Considering the complexity of the scenario, involving multiple CRM systems, a network monitoring platform, strict security policies, and requirements for real-time notifications and analytics, the most suitable integration method is an API-led integration approach combined with an event-driven microservice architecture. This approach not only ensures real-time data flows and automation but also robustly addresses data privacy and security concerns.

### Detailed Solution Explanation:

### 1. Event-Driven Microservice for Network Monitoring (Juniper Networks):

To comply with security policies prohibiting direct internet access to Juniper Networks' APIs, we introduce a secure internal microservice as an intermediary. This microservice serves two critical functions:

* **Internal API Proxy:** Securely retrieves real-time network service status information from Juniper Networks without exposing their APIs directly to the internet.
* **Real-Time Event Triggers:** When Juniper Networks detects a service fault (e.g., a network outage in a specific service area), it immediately triggers an event notification through this microservice. This event notification is securely passed to the integration platform, enabling real-time responsiveness.

### 2. API Gateway for CRM Integration (Amdocs and Salesforce):

An API Gateway is implemented as a centralized layer between the two CRM systems Amdocs and Salesforce, and the Integration Platform as a Service (IPaaS), here we use Workato. The API Gateway provides the following capabilities critical to addressing data privacy concerns:

* **Secure API Integration:** Enables secure real-time API calls to both CRM systems, replacing manual extraction of CSV files, thus eliminating manual data handling and significantly reducing privacy risks.
* **Role-Based Access Control (RBAC):** Ensures access to sensitive CRM data is strictly limited to authorized services and personnel, adhering to the principle of least privilege.
* **Selective Data Retrieval and Filtering:** Explicitly filters out highly sensitive fields, such as customer National Registration Identity Card (NRIC) numbers, to ensure they are never transferred downstream, thereby reducing data exposure risk.
* **Tokenization:** Implements tokenization of sensitive data fields such as customer mobile numbers and email addresses at the API Gateway, ensuring downstream systems, including the IPaaS and analytics data lake, never directly handle real personally identifiable information.
* **End-to-End TLS Encryption:** Ensures all data transmitted through APIs is securely encrypted using Transport Layer Security (TLS), maintaining data integrity and confidentiality during transit.

### 3. Integration Platform as a Service (IPaaS - Workato):

Workato receives real-time event notifications from the Juniper Networks microservice and responds by:

* Retrieving the corresponding tokenized customer details from the API Gateway-connected CRM systems in real-time.
* Executing business logic and data processing to identify affected customers based on service area information.
* Preparing and packaging notifications for affected customers.

### 4. Real-Time Notification Delivery (via API Gateway):

Workato forwards the prepared notifications back through the API Gateway, which securely performs:

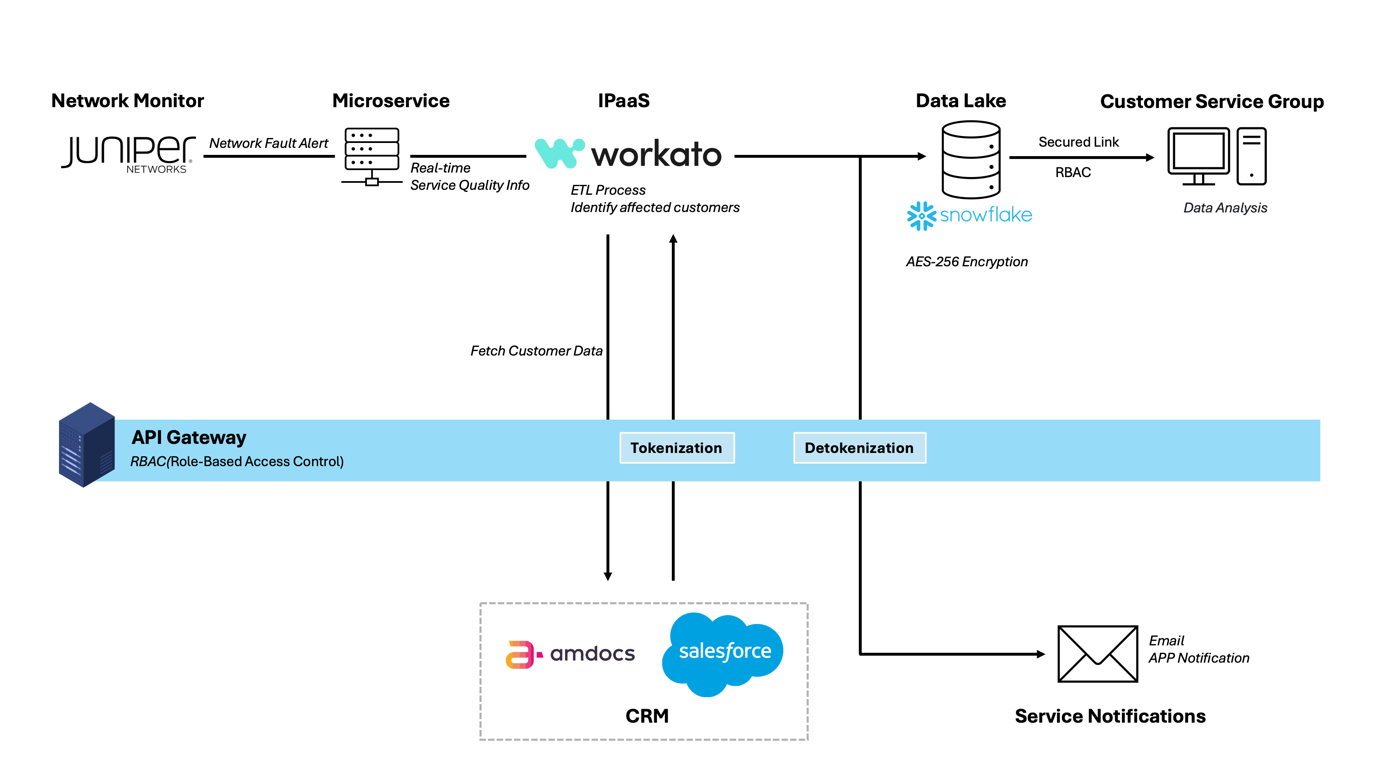
* **Detokenization:** Converts tokenized customer contact information back into actual email addresses and mobile numbers.
* **Secure Notification Dispatch:** Immediately sends proactive notifications to customers via email and/or the telco’s mobile app, providing timely and transparent communication about network service disruptions.

### 5. Data Lake Integration and Predictive Analytics (Snowflake):

Simultaneously, Workato sends the processed, tokenized network fault data to the Snowflake data lake for historical storage and analysis. The Snowflake integration ensures:

* **AES-256 Encryption at Rest:** Ensures all data stored in the Snowflake data lake remains encrypted and secure, safeguarding against unauthorized access and maintaining compliance with data privacy regulations.
* **Historical Analysis & Predictive Modelling:** Enables the customer service and analytics teams to leverage stored historical data for trend analysis, reporting, and development of predictive monitoring models, ultimately improving proactive customer service.

**Architecture Diagram for the Recommended Data Integration Solution**



# Question 2:

## (a) Scope Definition and Data Integration Solution Choice

**Scope of MVP**

The Minimal Viable Product (MVP) aims to demonstrate a streamlined, automated, and secure solution that proactively notifies customers in real-time when a network fault occurs. The MVP specifically focuses on:

* **Real-Time Network Fault Detection:** Immediate detection of simulated network fault events using an internal microservice representing Juniper Networks.
* **Automated Customer Data Retrieval (Tokenized):** Automated real-time retrieval of tokenized customer information from simulated CRM systems (representing Amdocs and Salesforce CRM), triggered directly by detected network faults.
* **Proactive Notifications to Customers:** Automated preparation and real-time dispatch of notifications via simulated channels, such as email and the telco's mobile app.
* **Secure Data Storage and Analytics:** Secure storage of processed and tokenized network fault and customer data into a simulated data lake (Snowflake), facilitating subsequent analytics and predictive monitoring.

To simplify feasibility testing, the MVP prototype will simulate a single network fault event (e.g., "Sentosa area outage") with representative customer data.

**Choice of Data Integration Solution**

Consistent with the recommendation from Question 1, the chosen integration method is an API-led, event-driven architecture, incorporating the following components:

* **Internal Microservice**: Provides a secure communication proxy for Juniper Networks’ APIs, adhering to the organization's security guidelines.
* **API Gateway**: Centrally manages secure communication with CRM systems, ensuring robust privacy through data filtering, tokenization, role-based access control (RBAC), and encryption.
* **Integration Platform as a Service (Workato)**: Core platform to automate real-time event processing, customer notifications, and integration with the data lake.
* **Simulated CRM Systems and Snowflake Data Lake**: Represented by databases and cloud storage platforms with mock or tokenized customer data, secured using appropriate encryption methods.

# (b) MVP Prototype Implementation

**MVP Prototype Overview**

The MVP prototype was built using **Workato**, a no-code integration and automation platform, to demonstrate how real-time customer notifications can be triggered automatically when network faults are detected. The goal of this MVP is to replace the current manual process with a secure, automated, and scalable solution that integrates customer data from two CRMs (Salesforce and Amdocs), detects service faults from Juniper Networks, and proactively notifies affected customers.

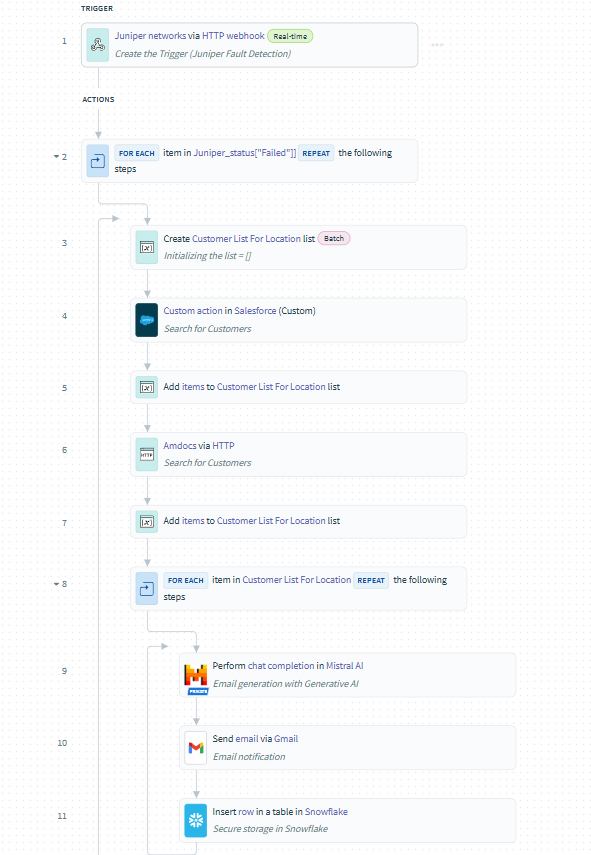
This prototype aligns with the telco’s objective of enhancing customer service quality and reducing complaint response times by delivering real-time fault alerts through email.

**Architecture Summary**

The solution simulates the following data sources and tools:

* **Juniper Networks** – triggers a webhook with a JSON payload listing affected locations under "Failed" status.
* **Salesforce CRM** – holds pre-paid and broadband customer information.
* **Amdocs CRM** – holds post-paid customer information, accessed via HTTP.
* **Mistral AI** – generates polite and location-specific message templates.
* **Gmail** – sends real-time email alerts to customers.
* **Snowflake** – stores a record of each notification for analytics.

**Screenshot of the MVP Flow**

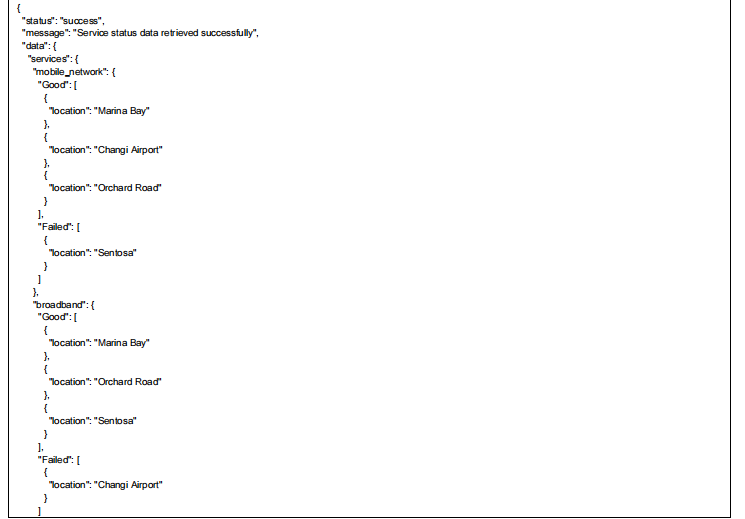


### **Workflow Logic and Explanation**

Below is the step-by-step explanation of the final Workato recipe:

#### **Step 1: Trigger – Juniper Webhook (Real-time)**

* The workflow is triggered when Juniper Networks detects a service fault and sends a webhook.
* The payload contains a ["Failed"] list of impacted service locations.

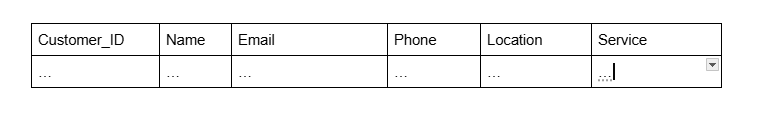


#### **Step 2: Loop Through Each Failed Location**

* Using a FOR EACH loop, we iterate through each failed location (location = “Sentosa”, location = “Changi Airport”).

#### **Step 3: Create a Customer List for That Location**

* Inside the loop, we initialize a new list variable Customer\_Location\_List for that location.
* This isolates each location’s customers and supports real-time processing per area.

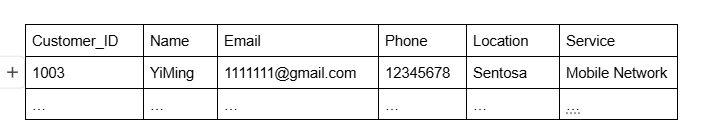


#### **Step 4: Search Salesforce CRM for Affected Customers**

* We search Salesforce records where the customer’s location matches the current fault location.
* The returned results are **tokenized** or masked to protect PII.

#### **Step 5: Add Salesforce Results to the List**

* Matching customers from Salesforce are added to the Customer\_Location\_List.



#### **Step 6: Search Amdocs CRM via HTTP**

* A GET/POST request is made to the Amdocs system to fetch customers affected in the same location.
* The returned data is in JSON format and assumed to be tokenized.

#### **Step 7: Add Amdocs Results to the List**

* The results from Amdocs are appended to the same list, ensuring both CRM sources are consolidated.

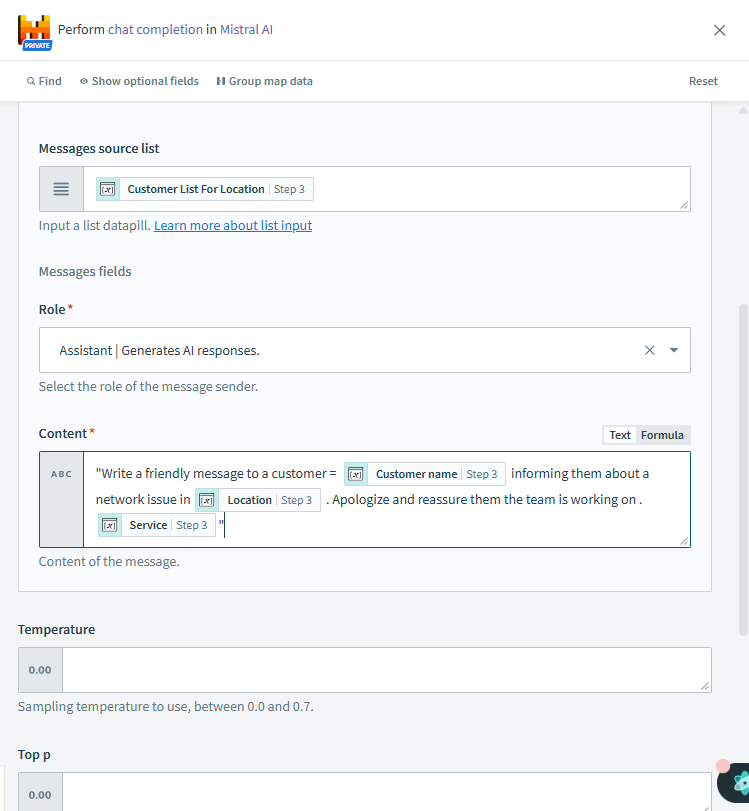


#### **Step 8: Loop Through Each Customer in the List**

* We now loop through every customer in Customer\_Location\_List to perform three key tasks per customer:

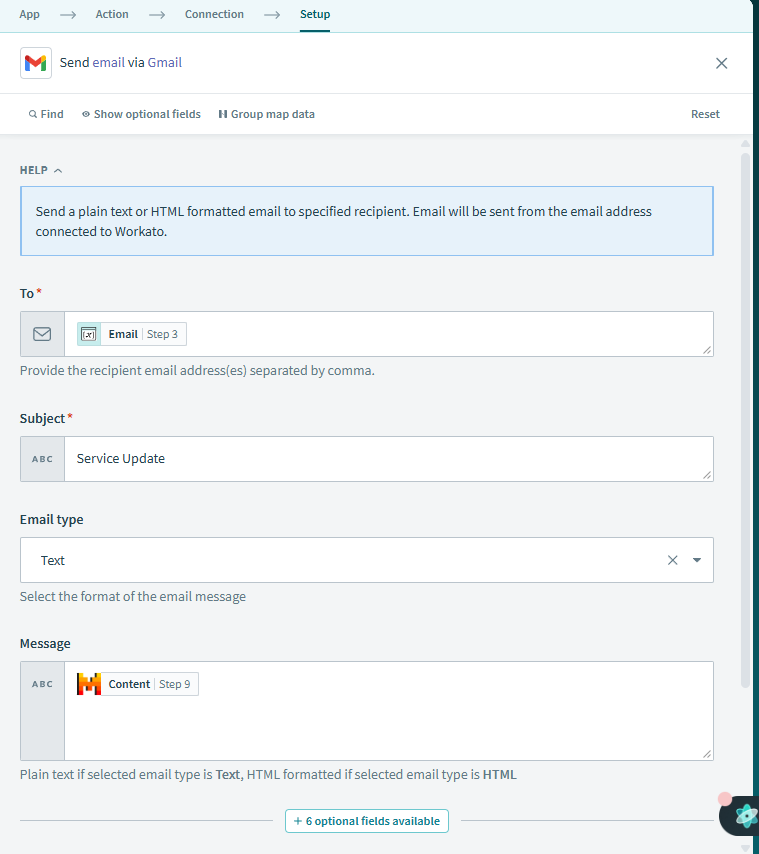
#### **Step 9: Generate a Notification Message with Mistral AI**

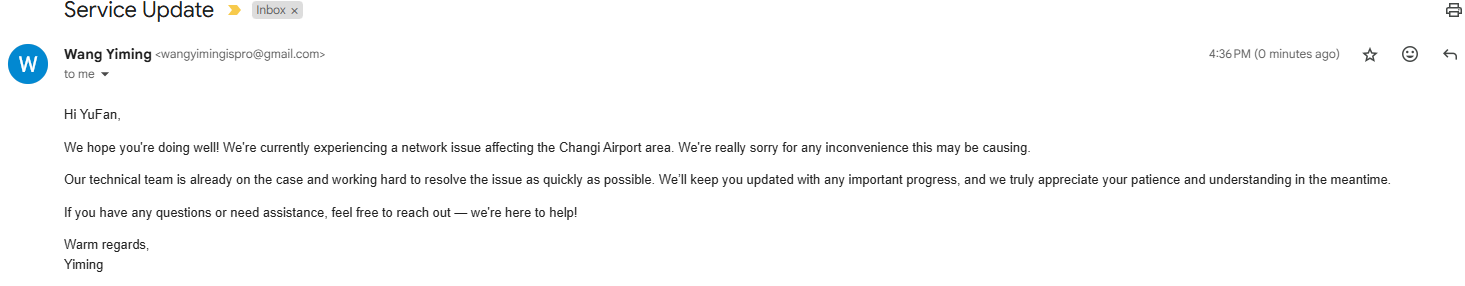
* Mistral AI is prompted with the fault location and a brief instruction to generate a message.
* Example prompt:



#### **Step 10: Send Email via Gmail**

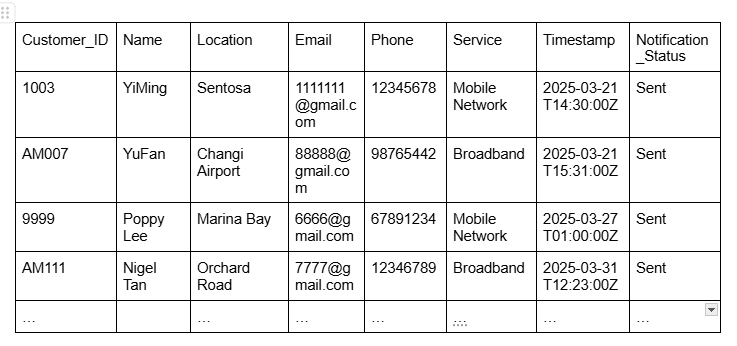
* An email is sent to the customer’s tokenized or masked address using the generated message.
* This step simulates proactive customer service in real time.





#### **Step 11: Store Notification Log in Snowflake**

* The system logs the notification event in a Snowflake table.
* Data includes:
  + location,
  + customer\_token\_id,
  + name,
  + email,
  + service,
  + phone,
  + timestamp,
  + and notification\_status



# (c) MVP Results Measurement

The MVP was evaluated based on key performance indicators: speed, accuracy, safety, and automation effectiveness. Comparisons were made against the current manual workflow used by the Customer Service Group and the Workato automated MVP solution.

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Manual Process** | **MVP Prototype** | **Improvement** |
| Average Time to Notify Customers | ~30–60 minutes | < 1 minute | 95–98% faster |
| Data Matching Accuracy | ~85–90% (manual cross-check) | >99% (automated lookup & filter) | Reduced human error |
| Proactive Notification Rate | 0% | ~100% (on trigger) | Shift from reactive to proactive |
| Data Privacy Handling (PII exposure) | High risk | Minimized (tokenization, RBAC) | Significant privacy improvement |
| Scalability | Low | High (real-time, reusable flow) | Enables future expansion e.g. customer service recovery vouchers could be sent |

The MVP will thus result in below impact:

* **Customer Satisfaction**: It is anticipated that customer annoyance during network outages will be greatly decreased with quicker and more proactive alerts. Transparency and trust can be built and enhanced by prompt communication.
* **Operational Efficiency and Accuracy**: The MVP eliminates manual tasks that are repetitive and tedious such as CSV extraction and Excel-based cross-referencing. This automation allows staff to focus on more strategic, high-value functions.
* **Data Governance Compliance**: The MVP ensures that personally identifiable information (PII) is handled securely across all integration touchpoints using tokenization, encryption (TLS, AES-256), and role-based access controls. This ensures better alignment with industry best practices and regulatory requirements.

# References

Hamzah, A. (2024, February 24). *Carousell fined $58K for data breaches, including one where data of 2.6M users was sold*. The Straits Times. <https://www.straitstimes.com/singapore/carousell-fined-58k-for-data-breaches-including-one-where-data-of-26m-users-were-sold-on-hacking-forum>

Kurohi, R., & Choo, F. (2019, March 30). *Personal information of over 800,000 blood donors was accessible online for 2 months: HSA*. The Straits Times. <https://www.straitstimes.com/singapore/health/personal-information-of-over-800000-blood-donors-exposed-online-hsa>