Customer Transaction Management

Logo, company name

Description automatically generated

**Prepared by: Shontina Wesley**

TAble of  
Contents

sYsTEM iMPLEMENTATION 1

Customer statement of requirements 1

System requirements 2

Functional requirements specification 4

Effort estimation 6

Domain analysis 6

Interaction diagrams 8

Class diagrams and interaction specification 10

System architecture and system design 11

Algorithms and data structure 14

User interface design and implementation 14

Design of tests 14

History of work, current status, and future work 15

Glossary of terms 17

References 23

# sYsTEM iMPLEMENTATION

## Customer statement of requirements

My insurance agency has grown over the last few years, and I realize the need to take steps to streamline my workflow processes. I currently have an agency management system (AMS) and cloud-based software for database management. The data in the database is derived from legacy technology and disparate systems that are not always timely nor accurate. Despite my AMS I must utilize emails from carriers and carrier landing page links to review customers updates. I have to copy data from the emails/carrier site and then paste into an excel spreadsheet. After that the data must be manually reviewed and reconfigured for action. There is no single reliable data repository for all my customer accounts. I have lapses in policies and must rewrite and reinstate policies as a practice. I have no reports to tell me how my business is doing at any point in time. I need a robust reporting system to monitor my business’ performance. I need to automate some of the repetitive manual tasks so I can think more strategically about growing my business.

## System requirements

|  |  |  |
| --- | --- | --- |
| No. | Priority Weight | Description |
| 1.) Email Management | High | Power Automate workflow will listen for emails from specific carriers. The system will parse HTML email for embedded spreadsheet of customer updates into JSON file. |
| 2.) Retrieve direct carrier updates | High | Power Automate – Robotics Processing Automation (RPA) workflow will go to carrier site and extract customer billing notices from carrier landing page. The workflow will upload excel file into data lake house. |
| 3.) Create database -Map and Ingest data | High | Use Azure explorer/synapse to (etl) the data, create schema, and create pipeline to orchestrate the flow of the code into the SQL serverless pool. |
| 4.) Connect to reporting interface | Medium | The system will create Power Bi reports on top of data lake for data visualization of business intelligence and provide ad-hoc-query tools based on customer needs and business rules. |
|  |  |  |

## Functional requirements specification

Fully-dressed use cases

|  |  |
| --- | --- |
| Name of use case: | Extract and Parse Data from Emails |
| Actor: | Power Automate |
| Successful Completion: | 1.Power Automate listens for specified emails in designated email box.  2.Email (HTML) is parsed and converted to a JSON file that contains only extract customer transactional information.  3. A dated JSON file is downloaded to Azure data lake into the transient zone. |
| Alternative: | 1.Power Automate does not listen for specified emails in designated email box.  2.Email (HTML) is not parsed and converted to a JSON file that contains only extract customer transactional information.  3. A dated JSON file is not downloaded to Azure data lake into the transient zone. |
| Precondition | Power Automate Flow is active and error free |
| Postcondition: | JSON files are in transient zone |
| Assumption: | Emails arrives specified options |

|  |  |
| --- | --- |
| Name of use case: | Create database |
| Actor: | Data Engineer |
| Successful Completion: | 1. Database is created with the proper structure and code to yield information deemed useful to the client. |
| Alternative: | 1. Database is created with the proper structure and code to yield information deemed useful to the client. |
| Precondition | Client does not have a single reliable data repository. |
| Postcondition: | Client has a single reliable data repository. |
| Assumption: | Data engineer has all access to systems and data to create the data base. |

|  |  |
| --- | --- |
| Name of use case: | Generate Reports and Analytics |
| Actor: | Power BI |
| Successful Completion: | 1. User selects the data source from the **Get Data** window and select **Connect**. 2. Power BI Desktop makes the connection to the data source, and it presents the available data sources in the **Navigator**. |
| Alternative: | 1. User does not select the correct data source from the **Get Data** window. 2. Power BI does not make the connection to the data source, and no data presents in the **Navigator**. |
| Precondition | Accesses Power BI desktop |
| Postcondition: | User is able to successfully access the database from the Power Bi desktop. |
| Assumption: | User has the privileges to access the data |

Casual description for use cases that will be implemented for future work.

* Ability for staff to login to reporting tool(requires training on PowerBI and need Log-ins)
* Ability for staff to access database to review all transactional updates (new business, pending cancels, cancels, renewals) to customer accounts
* Ability for staff to download generated reports

Traceability Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Req- ID | UI Requirement | Priority Weight | Use Case ID | Use Case Description | Status |
| 3 | Create database | High | UC02 | Staff can access and query database | Complete |
| 4 | Generate Analytics and Reporting | Medium | UC04 | Power Bi can present visual information on queried database | Complete |
| 1 | Email extraction | High | UC06 | Power Automate Extracts emails and converts to JSON files  JSON Files are download to data lake | Complete |
| 2 | Excel Extraction | High | UC07 | Power Automate will download excel file from carrier site and download to database | Complete |

## Effort estimation

|  |  |
| --- | --- |
| **Use case** | **Least Amount of Clicks** |
| Excel Extraction | 0(automated) |
| Email extraction | 0(automated) |
| Create database | 0(client does not access directly) |
| Generate Analytics and Reporting | 5 |

## Domain analysis

Physical data model

Diagram

Description automatically generated

Table structures and data types

Agency number varchar(5)

insured varchar(255)

company varchar(50)

date CHAR(8)

policy number varchar(25)

status varchar(15)

*System Operation Contracts (for user access)*

|  |  |
| --- | --- |
| Contract Req-id 1: Log in |  |
| Operation | Log in |
| Cross References | Use Case: Log-in |
| Preconditions | Accesses link to Power BI app |
| Postconditions | User’s login is processed successfully and user can access Power Bi |

|  |  |
| --- | --- |
| Contract Re-id 2: Access database |  |
| Operation | Access database |
| Cross References | Use Case: Access database |
| Preconditions | Accesses Power BI desktop |
| Postconditions | User is able to successfully access the database from the Power Bi desktop. |

## Interaction diagrams

This diagram was created to reflect the interactions of the different objects in the system. The ETL process creates files as part of a bronze, silver, gold architecture. These layers enable agility for the service data integrations and data products it contains.

Diagram

Description automatically generated

Diagram, schematic

Description automatically generated

## Class diagrams and interaction specification

This system does not have classes. This is a star schema data model appropriate for this relational database system. This model has 2 dimension tables and one fact table. The factless fact table stores the relationships and observations of the data.

Chart, line chart

Description automatically generated

## System architecture and system design

This web-based SAAS has a 3 tier Client-Server Architecture which segregates the system into two applications, where the client makes requests to the server(through Power BI). In this case, the server is a database(Azure Data Lakehouse Gen2) with application logic represented as stored procedures. The tiers in detail are as follows:

Presentation Tier – Power BI is the user interface where the end user interacts with this desktop application.

Application Tier – This tier is the heart of the design, known as the middle tier where the information is collected in the presentation tier is processed using Python(Spark) and Dax.

Data Tier- Sometimes called the database tier or back-end where the information processed is stored and managed. This solution uses Microsoft Azure for real-time lakehouse processing through Azure SQL database.

Data Lakehouse – Is a new open data management architecture that combines the flexibility, cost-efficiency, and scale of data lakes with the data management and ACID transactions of data warehouses, enabling BI and ML on all data.

*Persistent Data Storage-* There is not persistent data storage with Azure BLOB in which Azure Data Lake Storage Gen 2 is built on. Azure Blob storage is an object storage solution where each object is stored in a flat name space. Whereas, Azure Data Lake Gen2 allows a hierarchical namespace similar to a HADFS. Azure does have a persistent store for long-term storage of file. In this case, this system uses an Azure Blob storage tier classified as a “hot tier” which is optimal for frequent reading and writing access to stored data. The data will be stored in a parquet file format and then stored in folders. The objects stored in the folders will be JSON files, excel files, and folders created during the transformation from unstructured to structed data within the landing zones.

*Global Control Flow-*

* Execution order: Azure Functions is a serverless compute option. It uses an event-driven model, where a piece of code (a function) is invoked by a trigger. In this architecture, when events arrive at Event Hubs, they trigger a function that processes the events and writes the results to storage.
* Time dependency: Is it a near real time system.
* Concurrency: Does your system use multiple threads? NO
* Hardware Requirements
  + What system resources your system depend on?
  + Internet Connection
  + Active accounts for Microsoft 365, Azure, and Power BI.
  + A device that runs Windows 10 Home, Windows 10 Pro, Windows 10 Enterprise, Windows 11 Home, Windows 11 Pro, Windows 11 Enterprise, Windows Server 2016, Windows Server 2019 or Windows Server 2022. ARM devices aren't supported. Minimum hardware- Processor: 1.00 GHz or faster with two or more cores. For unattended mode, four or more cores are needed. Storage: 1 GB,RAM: 2 GB

## Algorithms and data structure

There are no algoritms.

*Data Structure*

The data lake utilizes Delta file format made up parquet and json data structures. These structures are built to work on high data volumes and best fitted for analytic workloads. Delta cache uses local storage so the amount of storage dedicated can be configured. This reduces execution time and there is consistent performance of specific queries. Another reason is the ability to time travel. You can query your data as it was of a specific timestamp, query change history, and even restore a table as it was in the past. The Delta format also has the ability to merge two datasets to benefit from parallel processing. Optimization, file reduction, and schema evolution which can enrich data are also reasons that this structure was chosen for this workload.

## User interface design and implementation

The user interface is POWER BI desktop.

2 graphs have been created for reporting purposes.

A simple tabular format with a chiclet view for the carrier filter.

A simple bar chart with month year information.

## Design of tests

Title: Customer Transaction Management

**Description**: Automation- An incoming email containing customer transactional information will be parsed and uploaded as a file to the database using Power Automate. Power Automate will upload that file to Azure Data Lake. The file will be reformatted, filtered, and queried.

**Precondition**: Flow has previously been executed successfully.

**Assumption**: Access to Power Automate, Azure Data Lake(storage account), Azure Synapse

Test Steps:

1.Send email with specific to, from, and subject information.

Expected Result:

A file with data formatted according to the schema data model.

Integration Test: After the email is sent I will wait no more than 60 seconds to:

1. Log in to Power Automate.

2. Go to flow history and see if the flow just sent was executed successfully.

2. Click on the flow.

3.Click on last step of the flow.

4.Click on the raw output of the flow.

5. Log-in to Azure Storage Account (Data Lake).

6.Naviate to the gold folder.

7.Open the gold folder to view the formatted file.

## History of work, current status, and future work

*Project size estimation*

This is a small project lasting no more than 2 months and costing no more than $150 on a monthly basis, which are subscription and/or transaction-usage fees for the SAAS.

WBS

|  |  |  |
| --- | --- | --- |
| 1 | Task | task duration |
| 2 | Flow for email management | 1 day |
| 3 | Flow for data extraction from web page | 1 day |
| 4 | Configure Microsoft 365 E5 version | 4 hours |
| 5 | Configure Power BI | 4 hours |
| 6 | Configure Azure | 4 hours |
| 7 | Define dimensions tables | 4 days |
| 8 | Define fact tables | 4 days |
| 9 | Define Measures | 4 days |
| 10 | Meetings w/ stakeholder | 14 hours minimum |
| 11 | Testing | 3 days |
| 12 | Upgrades based on feedback | 5 days |
| 13 | Write test cases | 3 days |
| 14 | Research | 3 days |

Cost

Microsoft 365 business E5 version - $23 a month

Power BI premium per user - $20 month

Azure Account – free for 30 days ($100 per month thereafter)

Upgrade Power Automate - $30

*History of Work – Strikethru indicates a revised due date*

|  |  |  |
| --- | --- | --- |
| Sprints | Description | End Date |
| Week 1,2 | Environment Set-up | October 16, 2022 |
| Week 2 | Develop Triggers for email | October 16, 2022 |
| Week 3,4 | Develop RPA for extraction form carrier site | ~~October 16, 2022~~  November 22, 2022 |
| Week 5 | Bronze Zone | October 30, 2022 |
| Week 6 | Silver Zone | November 13, 2022 |
| Week 7 | Gold Zone | November 13, 2022  November 27,2022 |
| Week 5 | Develop Data Model  Record Demo for mid-term | ~~October 30, 2022~~  October 31, 2022 |
| Week 8 | Testing automation and Connect to Power BI and develop reporting | ~~November 13, 2022~~  December 3, 2022 |
| Week 8 | Record demo for final presentation | ~~November 27, 2022~~  December 7, 2022 |
| Future Work | System Upgrades based on feedback from stakeholders |  |

*Key Accomplishments*

* Successful runs of automated flows
* Completion of ETL process for data in each zone
* Successful orchestration of pipeline to connect the notebooks.
* Completion of reports in Power BI

## Glossary of terms

Glossary of terms:

Power Automate – Creates flows that automate repetitive work tasks.

A screenshot of a computer

Description automatically generated

automation- the technique of making an apparatus, a process, or a system operate automatically.

Application

Description automatically generated

SQL serverless Pool- Query service over the data in your data lake.

Diagram

Description automatically generated

Data Ingestion- Process of obtaining and importing data for immediate use or storage in a database.

Chart, funnel chart

Description automatically generated

Map data- Data mapping is the process of matching fields from one database to another. It's the first step to facilitate data migration, data integration, and other data management tasks.

Diagram

Description automatically generated

RPA- RPA is a form of business process automation that allows anyone to define a set of instructions for a robot or ‘bot’ to perform.

Graphical user interface

Description automatically generated

Robust reporting – a report where output and forecasts are consistently accurate even if one or more of the input variables or assumptions are drastically changed due to unforeseen circumstances.

A screenshot of a computer

Description automatically generated with medium confidence

parse-the process of analyzing a string of symbols, either in natural language, computer languages or data structures, conforming to the rules of a formal grammar.

Icon

Description automatically generated with medium confidence

HTML- Hypertext Markup Language, a standardized system for tagging text files to achieve font, color, graphic, and hyperlink effects on World Wide Web pages.

Text

Description automatically generated

JSON- (JavaScript Object Notation) is a lightweight data-interchange format.

Text, letter

Description automatically generated

Azure Synapse - Azure Synapse Analytics is a limitless analytics service that brings together data integration, enterprise data warehousing, and big data analytics.

Diagram

Description automatically generated

ETL- Stands for extract, transform, and load, is the process data engineers use to extract data from different sources, transform the data into a  usable and trusted resource, and load that data into the systems end-users can access and use downstream to solve business problems.

Diagram

Description automatically generated

data lake(warehouse)- A data lake is a storage repository that holds a large amount of data in its native, raw format.

Diagram

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

## References

Database design technique **-** <https://en.wikipedia.org/wiki/Don%27t_repeat_yourself>

Data base Models and UML- <https://www.microsoft.com/en-us/microsoft-365/business-insights-ideas/resources/guide-to-uml-diagramming-and-database-modeling>

ERD example using UML- <https://www.conceptdraw.com/How-To-Guide/erd-symbols-and-meanings>

The advantages of a Datalake House

<https://www.youtube.com/watch?v=g11y-kJHr3I>

This link shares the tutorial on how to use Power Automate to parse HTML table and convert to JSON

<https://www.youtube.com/watch?v=IwRKWaTnl3U&t=569s>

This link shares how to set up Azure account

<https://azure.microsoft.com/en-us/free/search/?&ef_id=Cj0KCQjwnP-ZBhDiARIsAH3FSRcyTo9Bfkv5eqUtABdJPKcHBsb8hnKzd0FbX2fn31V6OLTagkPIEJoaAgTeEALw_wcB:G:s&OCID=AIDcmm5edswduu_SEM_Cj0KCQjwnP-ZBhDiARIsAH3FSRcyTo9Bfkv5eqUtABdJPKcHBsb8hnKzd0FbX2fn31V6OLTagkPIEJoaAgTeEALw_wcB:G:s&gclid=Cj0KCQjwnP-ZBhDiARIsAH3FSRcyTo9Bfkv5eqUtABdJPKcHBsb8hnKzd0FbX2fn31V6OLTagkPIEJoaAgTeEALw_wcB>

This link share how to set up SQL serverless pool In Azure

<https://learn.microsoft.com/en-us/azure/synapse-analytics/sql/on-demand-workspace-overview>

This link shares how to create storage account to use with Azure data Lake Gen2

<https://learn.microsoft.com/en-us/azure/storage/blobs/create-data-lake-storage-account>

This link shares how to change Microsoft 365 versions

<https://learn.microsoft.com/en-us/microsoft-365/commerce/buy-or-edit-an-add-on?view=o365-worldwide#update-an-add-on>

This lihttps://signup.microsoft.com/get-started/signup?sku=a403ebcc-fae0-4ca2-8c8c-7a907fd6c235&email=&ru=https%3a%2f%2fapp.powerbi.com%3fpbi\_source%3dwebpaidmedia%26cmpid%3dpbi-free-hero-sta-signup%26redirectedFromSignup%3d1%26noSignUpCheck%3d1&products=a403ebcc-fae0-4ca2-8c8c-7a907fd6c235&ali=1nk shares how to sign up for Power BI

This link shares how to get stared with Power BI

<https://learn.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started>

This link shares information on creating the (star schema) physical data model for this project.

[Understand star schema and the importance for Power BI - Power BI | Microsoft Learn](https://learn.microsoft.com/en-us/power-bi/guidance/star-schema)

Summary of Changes

1. Upgraded customer problems statement to include a definitive statement concerning the need to automate repetitive task.
2. Updated system requirements for “retrieve direct carrier updates” as a correction to the workflow to reflect a direct upload to data lakehouse.
3. Update system requirements for “mapping and ingesting the data” to be more clear in the description.
4. Updated fully dressed use cases to reflect what was implemented at final demo. I did not have a clear understanding of what could be a use case n the beginning.
5. Updated casual description for use cases to include future work on training for staff to access reports and additional visualizations, and reporting. I was not able to test with client as the data sources did not yield all of the information the client requested. To avoid scope creep I created this as Phase 1. In Phase 2 I will look into additional data sources to expand reporting,
6. Updated sequence diagram to the proper format and flow.
7. Created new domain model to reflect final tables and relationships. With the amount of data that was derived it made sense to have less tables.
8. Added table structure and data types as a result of the change on the domain model.
9. Created proper UML as advised by professor.
10. Revised Plan of work to add task and revised date of completion.
11. Revised Star schema data model to only included 3 dimension tables.
12. Added upgrade to Power Automate for Plan of Work.