Infosys Global Internship Program

Requirements Document

InStep Project Title: Data Transformation involving Azure and Snowflake using a Python HTTS endpoint

Prepared

By

**Intern Name: Gavin Aren**

With the guidanceof

**Project Mentor Name: Marella Anji Reddy**

**Co-Project Mentor Name (if any): Mahesh Paranjape**

**Unit:** DNA – Data and Analytics

|  |  |  |  |
| --- | --- | --- | --- |
| Document Name | Requirements Document | Version No. |  |
| Authorized by |  | Signature /Date |  |

Modification Log

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ver. Rev | Change Date | Author | Reviewer | Remarks |
|  |  |  |  |  |
|  |  |  |  |  |

**Table of Contents**

1 Background 4

2 Infosys need 4

3 Internship Deliverables 5

**4 Post Internship..………………………………………………………….6**

**5 Project Plan & Progress Report……………………………………….7**

# Background

In recent years, data has truly made the difference between an organisation coping with the growing competition and being beaten. Data is a valuable asset for all organisations. This is why companies like Infosys hold such a large quantity of it, collected from various sources. Analysing vast amounts of data gives them insights to help gain a competitive advantage needed to survive in today’s landscape.

Businesses want their data to do more and are moving towards an analytics driven enterprise to monetize this data. Being able to take some raw data, transform it, and then store it easily on a platform like Snowflake can be instrumental to a business’ success nowadays. Data transformation can make all the difference, improving the efficiency of analytics and business processes by striving to be more data-driven in everything they do.

The need for this project arises from the recent surge in the amount of raw data organisations seem to collect. We need to make sense of all of this data. In order for that to happen, it needs to be transformed (cleaned up, organised, flattened) and put into a data warehouse like Snowflake which can then go on to be analysed to lead business decisions with data.

Creating this endpoint enables uploading raw data easily and have it stored in a data warehouse ready to run computations on it.

# Infosys need

The need from the project simply arises from not being able to easily upload raw data in the right way to a massive data warehouse such as Snowflake, to be analysed later on. A HTTPS endpoint hosted on Microsoft Azure’s serverless services make it simple to access this function/endpoint from anywhere with no downtime and always have your data safely put away into Snowflake’s data warehouse/database. This could potentially be integrated into larger systems which collect a lot of a raw data such as IoT devices in the future, and have them configured so they send all their data to this always-available endpoint remotely for it to be stored into Snowflake directly.

The intern brings experience of working with Python, the programming language in which the HTTPS endpoint will be written in, and they will be trained on using the Snowflake tool and some of Azure’s services through various Lex courses.

**Primary Stakeholders:**

* **Developer:** The intern in this case has direct influence over the project. They are responsible for a timely delivery of the final code as well as the additional documentation and the presentation alongside it. They can provide stakeholders with advice on business idea implementation and quality estimation needed to understand the scope of work, based on their prior experience with such projects. They are also responsible for testing failed user scenarios.
* **Project mentor:** The project mentor who also happens to play the role of the customer and the project manager for the intern lays out the main requirements, project scope, the deliverables required, controls the project creation and deciding the project milestones that require incremental deliverables.

**Secondary Stakeholders:**

* **Project Supervisor**: (Toai Chowdhury) The intern’s supervisor has influence over the project and can change certain aspects of it such as the required deliverables at the end of the project.
* **Backup/Senior Mentor**: Able to make decisions with the same level of influence over the project as the primary mentor, in the case that the primary mentor is absent. Additionally, they make sure that the primary mentor is adequately contributing to, and helping out the intern with the project where possible, and that there is excellent communication between the primary mentor and the intern.
* **Project Judges:** The judges that the project will be presented to at the end will decide on the effectiveness of the project completed, and whether the final product has come out as originally desired. This will decide if this project is integrated into other larger projects down the line by Infosys.

Regular communication between all primary and some secondary stakeholders will help the intern meet the requirements and business goals of the project.

# Internship Deliverables

The internship itself will have a number of deliverables to be completed over the 8 weeks.

1. Required Documents
   * A design document talking about what the intern plans to do and a record of all the key decisions/steps they took to get to the solution
   * Final presentation that explains the project
2. Review session
   * Meeting with the mentor almost daily, update them with the progress that the intern has made. The mentor’s experience can also be leveraged by consulting with them about how to tackle problems that the intern has run into
3. Final Deliverables
   * Working HTTPS endpoint with the correct functionality and with no bugs
   * Final Presentation explaining the project

The projected project timeline with the steps of execution:

GOAL: HTTP endpoints (hosted using Azure services) + loading data into Snowflake

1. Setting up the env – 2 days
2. Creating an Azure function in local machine in the environment – 3-4 days
3. Publish code to the Azure function (host HTTPS endpoint) – 2-3 days
4. Create a snowflake table – 1 day
5. Configure the local env to write data to snowflake table – 3-4 days
6. Testing the endpoint in the local env setup – 1-2 days
7. Publishing the code to Azure function and testing HTTPS endpoint – 2-3 days
8. Creating and integrating Azure Managed Identity and KeyVault into code – 3-4 days

If the main mentor is not around, the intern’s backup mentor will be Mahesh Paranjape.

Successfully completion will involve finishing 4 working HTTP endpoints which publish alter data in a snowflake table.

**Who will carry this project forward once the intern leaves?**

**What is the implementation plan?**

**What documentation needs to be completed in order to hand the project over for the next phase?**

**4 Post Internship**

# 5 Project Plan & Progress Report

**SAMPLE PROJECT PLAN TEMPLATE**

**Deliverables expected week wise:**

|  |  |
| --- | --- |
| **Week 1 & 2** | **Week 3 & 4** |
| **Learn Azure Fundamentals through AZ900 course**  **Set Up the environment and educate myself about the function to be made (HTTPS)**  **Start working on a design document** | **Create the HTTPS endpoint**  **Publish the code to Azure and host it there**  **Creating the Snowflake table and setting up that environment, and writing data to that environment** |
| **Week 5 & 6** | **Week 7 & 8** |
| **Finish function to write data to Snowflake table, test and deploy this function**  **Integrate Azure Managed Identity and Azure KeyVault into code** | **Present the project and possibly extend the functionality** |

**Any additional comments:**

Progress report summaries for each week:

|  |  |  |
| --- | --- | --- |
| ID | Week Ending | Summary |
| 1 | 24/06/2022 | Project Outline explained and started learning about relevant tools and technologies |
| 2 | 01/07/2022 | Finished AZ-900 course and try Snowflake + Finish VM set up |
| 3 | 08/07/22 | Got connected to Snowflake locally + Start on HTTP trigger |
| 4 | 15/07/22 | Finished testing and deploying first set of HTTP triggers |
| 5 | 22/07/22 | Finished testing and deploying second set of HTTP triggers |
| 6 | 29/07/22 | Finished project and start working on documentation |
| 7 | 05/08/22 | Implemented KeyVault into code + work on final presentation |
| 8 | 12/08/22 | Presented Project |