Industrial Training Report

Name: Saamira Yasmin

Roll: 19CH10039

Position: Data Engineer Intern, Data and Services, **Mastercard, Pune**

Duration: May 16, 2022 to July 15, 2022

The Onboarding Experience

My day 1 of internship journey as a Data Engineer intern at Mastercard commenced on 16th May, 2022. I was extremely excited and looking forward to an offline internship experience and understanding the corporate environment. I was part of the Data and Services Team of Mastercard, Pune.

The onboarding experience was seamless with lots to learn and begin with. My co-interns were from IIT Kharagpur, IIT Kanpur and NIT Warangal. We started off with training sessions where the basic workflow and concepts were briefed to us. These concept sessions were very informative and helped me understand how different teams work together at Mastercard to create a priceless experience for its customers.

About the Company, Mastercard

Mastercard Inc. is the second-largest payment-processing corporation worldwide. It offers a range of financial services. Its headquarters are in Purchase, New York.

Mastercard works to connect and power an inclusive digital economy that benefits everyone, everywhere by making transactions safe, simple, smart and accessible. Using secure data and networks, partnerships and passion, the company’s innovations and solutions help individuals, financial institutions, governments and businesses realize their greatest potential.

Training Project

Once the training sessions were completed, I was onboarded to the training project. This was a dummy project (not actual client project) to help me understand their ‘Test and Learn’ platform and learn the various steps involved in making data ready for the client. The training project was an intense 1 week of learnings and meeting the set targets. My buddy, mentor and manager were very supportive and entertained my doubts and queries. My manager was very happy with my pace of learning and I was the first one to complete the training project from the Data Engineer Team (interns).

Client Project

Right after the training project, I was staffed onto an ongoing client project, KEB Hana Card. KEB Hana Card is a credit card company based in South Korea. It sends daily and weekly transaction data to Mastercard for detailed analysis of their business operations. Through this project, I could apply the knowledge and skills that I had gained during the training project on a live scenario with the scope of a business profit for the company.

In the last week of internship, I was staffed onto a new client project, LATAM Airlines (Chile) for which I had to inspect the raw data files, prepare the data dictionary and formulate the questions to be asked to the client team regarding the deliverables and expectations.

Problem Statement

KEB Hana Card uses a set of metrics or quantities on which they want to have analysis on. They require the final data populated on the front-end for easy access and visualization. The existing import built by Mastercard for this purpose had bugs and required improvements for better performance and accurate results. My task was to follow the best practices and modify this import to create a one-click ready setup. The given import tasks were not automated as the code review was pending. Due to this, every day the files sent by the client team had to be processed manually requiring extra time and effort. I was supposed to automate the import, get the code review completed and make the entire process seamless. I had to implement different logics and write SQL scripts from scratch based on the client requirements.

Value Addition

* Created view to show last processed file date for all metrics and attributes
* Created one-click ready import
* Updated metric based on logic sent by client
* Added validations for typed table to master table and master table to staging table
* Removed duplicates
* Created metrics and financial tables from Data Setup
* Filed code review

Meetings

I used to have two meetings per week with the consultant team of KEB Hana Card where all client deliverables, to-do tasks and any bugs/errors were discussed. These meetings were online over MS Teams.

Apart from this, I had regular offline meetings with my buddy, project mentor and manager regarding project specifications, my performance feedback and discussions related to my internship experience as a whole.

Technologies/Softwares Used

SQL – Sequential Query Language (SQL) is a programming language

Splunk – Software to find resource space

Git – Collaborative space to share and exchange files

Test and Learn – Mastercard’s inbuilt platform for running all its major functionalities

qwbitbucket – Software for acquiring access to instance

JIRA – File code review tickets or other to-do task tickets

Confluence – Open-source platform for sharing resources

Virtual Environments (Remote Server Connection)

I got the opportunity to work on both Identified environment and the general (AVO Asia) environment. For my second client project, LATAM Airlines I used the identified environment for user privacy reasons. LATAM Airlines sent its transaction data to Mastercard with identifiable figures that could lead to concerns for an individual. Hence, it was important to use identified environment to maintain confidentiality of user’s data. There were many restrictions on this environment, for example I could not even open my mail box or download any new software. For my 1st client project KEB Hana Card, I used the general (AVO Asia) environment to connect to the virtual server and perform tasks.

Challenges Faced

Some of the challenges that I faced in my client projects are as listed below:

1. Mismatch between historical data and new (incremental) data for opt into 1Q pay metric:

Validation failed when inserting data from master table to staging table.

Resolved the mismatch by updating the typed scripts and maintaining similar format of data in all tables.

1. Identified low memory space in Fin database using Splunk.

Solved this problem by creating the new production tables in Fin2 database instead.

1. Getting access to instance (server and database) was a time-consuming procedure
2. Dealing with Korean names and working on a Korean client project was challenging as well as fun. At times, it was difficult to understand the deliverables and interpret the raw data sent by client.

Results

* Updated metrics with the best practices
* Data populated on front-end for easy visualisation
* Completed code review and automated the import
* Created one-click ready process

Learnings

Technical:

1. SQL functionalities like CTE to remove duplicates
2. Loading historical data from staging to production archive
3. Creating financial tables for metrics
4. Using conditional-run in data import tasks
5. Filing code review

Personal:

1. Interacting with consultants from different nationalities
2. Meeting deadlines and prioritizing tasks
3. Working in a team and learning from my peers
4. Simplifying tasks
5. Ownership of work

…………………………………………………………………………………………