DOUGLAS

DOUGLAS COLLEGE

COMMERCE & BUSINESS ADMINISTRATION
COMPUTING STUDIES & INFORMATION SYSTEMS
COMPUTER AND INFORMATION SYSTEMS (PBD)

CSIS 4495-050: APPLIED RESEARCH PROJECT

Progress Report 4:
End-to-End Data Engineering Solution for HR Analytics

Student Name: Bruno do Nascimento Beserra | 300392300

Instructor: Dr. Bambang Sarif

NEW WESTMINSTER/BC
FALL/2025

1.0 Work Hours

Student Name: Bruno do Nascimento Beserra

Date	Number of Hours	Description of Work Done
10/29/2025	1	Updated bronze notebook with path lineage change
10/29/2025	1	Updated silver notebook with path lineage change
10/29/2025	1	Updated gold notebook with path lineage change
10/29/2025	0.5	Finish notebook with call sample of the data
11/01/2025	0.5	Review how implement a job in databricks
11/01/2025	1.5	Created a job inside databricks to run the pipeline automatically everyday (Inside Databricks)
11/02/2025	3	Create first sample of code for the Slowly Changing Dimension type 2 algorithm
11/06/2025	3	Meeting with my team to decide key points of the midterm video report
11/07/2025	1.5	Update Slowly Changing Dimension type 2 algorithm and upload it to github
11/07/2025	4	Second meeting with my team to decide key points of the midterm video report and tentative of recording the video
11/07/2025	1	Create Progress Report 4 Document

2.0 Description of Work Done

During Week 9, I worked on updating the notebooks created in the past. I created a better version of the sample to pull data from the gold layer, updated the layer notebooks to change the database paths of the data and split them by quality layer in order to have a better management of who have access to what data. I also worked on the creation of the job that will run the pipeline everyday to update the data automatically.

At start of week 10, I mainly worked in designing and implementing the initial version of the Slowly Changing Dimension (SCD) Type 2 algorithm, then I conducted several tests to make sure everything is working properly.

I also participated in two meetings where we discussed about the key points for our midterm video report, attempted to record the final video presentation, and prepared this progress report document.

3.0 Repo Check-in of Implementation Completed

Over this period, I checked in to github repository three times. Two under the implementation folder. The first commit was related to updates to the notebooks, the second one had the first version of the implementation of the Slowly Changing Dimension type 2 technique, and the last one was to upload this report to the Document folder. The work related to the Databricks job was done directly in Databricks, so it was not added to GitHub.