A black text on a white background

Description automatically generated

**Work Diary and Check-in Log**

**Name of the apprentice: Shreya VR**

**Title: Apprentice-Tech**

**Leader’s name: Shwetha M**

**Curriculum name: Data Engineering**

**Govt. Registration No: A022534830**

**Date of joining: 19-02-2025**

**Instructions:**

Apprentices are required to update their work diaries on a bi-weekly basis, after discussing with their leader/peer buddy. The check-in log should be updated at the end of every 3 months (from the start date of the apprenticeship) – this has to be strictly done in partnership with the leader.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sl. No.** | **From (date)** | **To (date)** | **Topics covered** | **Trainings attended (if any)** | **Key Learnings** |
| 1. | 19-02-2025 | 21-02-2025 | Ethics, Cybersecurity, Professional etiquette, Outlook and collaboration tools and Excel training. | Excel training | 1. To ensure our work aligns with ethical standards it is important to remain mindful of the potential impact on ourselves and Target as a brand. 2. Adopting safe practices is essential to maintain cybersecurity which includes avoiding tailgating, utilizing a VPN when accessing network from home. 3. Learned to use excel, outlook, slack. |
| 2. | 24-02-2025 | 28-02-2025 | Data Warehousing, Data pipeline, Hadoop and its core components, Hadoop uses, basics of Object Oriented programming. Java Basics, Functions, OOPS Data Types |  | 1. Data Pipeline and its concepts, Hadoop, OOPS, Java Basics. |
| 3. | 03-03-2025 | 07-03-2025 | Introduction to Apache Spark, Architecture and its components, Spark Data Frames |  | 1. Apache Spark Overview, Data Frames and its APIs |
| 4. | 10-03-2025 | 14-03-2025 | Ingestion of Data , Kafka architecture overview and Shepherd |  | 1. Data Ingestion and processing steps. 2. Overview and working of Shepherd. |
| 5. | 17-03-2025 | 21-03-2025 | Task-1 Work Space Reservation pipeline replication |  | 1. Cloning the given pipeline repository and understanding the codebase. 2. Cloned the already existing ingestion pipeline , understood the parameters required for the DIP setup and learnt how configuration file works along with the architecture of KAFKA for storage of data and KELSA to land the data in ephemeral storage of HDFS. |
| 6. | 24-03-2025 | 28-03-2025 | Task-1 Work Space Reservation pipeline replication  Task –2 Timeliness updation for few Property Management Workflows |  | 1. Understanding the data processing pipeline with the transformations made for the data in the given pipeline and setup of it in shepherd. 2. Familiarized myself with Shepherd, the working of VELA and the entire flow of how the data is processed and landed on HDFS after processing. 3. Changing the asset-id of few given property management workflows and updating their timeliness and lineage on data-portal. |
| 7. | 01-04-2025 | 11-04-2025 | Task 1- Timeliness updation for few IS&P pipelines and updating their lineage on data-portal.  Task 2- History redemption for in-store positioning data for the year 2022, 2023 and 2024 |  | 1. Changing the asset-id of given IS&P workflows which were scheduled to oozie but had to be updated with to shepherd job and updating their lineage on data-portal. 2. Understood what timeliness is and the necessary of updating the lineages for the workflows. 3. Monitoring the given pipelines for history redemption and making sure data of every date is loaded and reloading if the pipeline fails with necessary changes in the configuration file. |
| 8. | 14-04-2025 | 17-04-2025 | Task 1- History redemption of in\_store\_positioning data for the year 2025. Task 2- Worked on retrieving the partition count of in\_store\_positioning data table for stage and prod of the years 2022, 2023 and 2024.  Task 3- Checking duplicate values for the refrigeration\_register\_history table |  | 1. Made necessary changes in the 2022 workflow after its completion, to load the 2025 data also for history redemption of TrIPS data. 2. Learnt how to run HIVE query on edge-node and ran query the query to retrieve the partition count for both stage and prod table of in\_store\_positioning data table for the year 2022,2023 and 2024. 3. Also tried how to retrieve the count in alternative method by writing a shell script for the same query by looping over the date, uploaded it on HDFS and created a workflow to get the count for the all the data loaded for the year 2025. |
| 9. | 21-04-2025 | 25-04-2025 | Task 1- Worked on code changes for compaction of refrigeration\_register\_history |  | 1. Made code changes in the in\_store\_positioning\_remediation pipeline and made it abstract so that it can be run across other pipelines with the same code but different constraints. |
| 10 | 28-04-2025 | 02-05-2025 | Task 1 – started loading the data from old referigeration\_register\_history table to the new external table |  | 1. Monitored and made changes as required to load the data up to date for the refrigeration\_register\_history new table . |
| 11 | 05-05-2025 | 09-05-2025 | Task 1 –count check and data count for refrigeration\_register\_history table |  | 1. Retrieved the counts as well as the data counts for all the loaded values in the refrigeration\_register\_history table . |
| 12 | 12-05-2025 | 23-05-2025 | Task 1- Learning of project\_detail\_master workflow  Task 2 -  1) Ethics at Target  2) Core Competencies at Target |  | 1. Familiarized with the various data sources integrated into the pipeline. 2. Gained a step-by-step understanding of the data validation process. 3. Learned how the system handles corrupt records and performs schema validation using a predefined template. 4. Explored how API data is fetched by making requests to external endpoints. 5. Developed a clear understanding of the overall workflow of the data pipeline. 6. Reviewed and understood the key functions used within the codebase. 7. Traced the end-to-end data flow, from ingestion to the final write operation into the CDS table. 8. 1) Annual Ethics Training, and the employee handbook. Explains that the Code provides expectations to consistently guide ethical decision making, and do what’s right for Target, our team and guests.   2) Target's core competencies include a strong brand image, focus on social responsibility, and a commitment to optimal internal processes. |
| 13 | 26-05-2025 | 27-05-2025 | Task 1- Reverse KT on the learnings from project\_detail\_master workflow |  | 1. Documented and presented the project\_detail\_master’s end to end workflow. |
| 14. | 28-05-2025 | 30-05-2025 | Task 1 – Understanding the business use case of project\_detail\_master |  | 1. Learned about SQL views and how they are used in the workflow. 2. Analyzed the specific SQL view provided in the project and understood its business use case and relevance. |
| 15. | 02-06-2025 | 06-06-2025 | Task 1 – Enhancement for project\_detail\_master . |  | 1. Studied various SQL window functions and their applications. 2. Understood the enhancement task provided and identified the appropriate window function required for its implementation. |
| 16. | 09-06-2025 | 16-06-2025 | Task 1 – Implementation of the enhancement |  | 1. Understood the enhancement task, which involved: 2. Addition of two new fields: expiry\_date and effective\_date. 3. Understanding the logic behind adding these fields. 4. Using window functions to derive and implement these new fields. 5. Tested the changes in the staging environment. 6. Cloned the existing workflow for testing. 7. Created a new external table and loaded data into it to support the modified workflow. |
| 17. | 17-06-2025 | 20-06-2025 | Task 1 – addition of the implemented attributes as metrics in grafana |  | 1. Got an understanding as to how grafana works. 2. Learnt how to add customized metrics on grafana. |
| 18. | 23-06-2025 | 27-06-2025 | Task 1 -RFI Topics ingestion and Analysis Task 2 – History redemption for in\_store\_position\_table . |  | 1. Learnt how to setup security and add consumers for the received rfi topics on Kafka 2. Created configuration files for all the topics and ingested data and landed it on hdfs by creating a workflow on shepherd 3. Performed gap analysis of the data. 4. Monitoring the given pipelines for history redemption and making sure data of every date is loaded and reloading if the pipeline fails with necessary changes in the configuration file. |
| 19. | 30-06-2025 | 14-07-2025 | Task 1- SCA Vulnerabilities remediation |  | 1. Explored and gained a clear understanding of TAP. 2. Updated the library dependencies in the Consumer API's Gradle files . 3. Tested the changes both locally and on TAP. |
| 20. | 21-07-2025 | 04-08-2025 | Task 1- SCA Vulnerabilities remediation |  | 1. Explored and gained a clear understanding of TAP. 2. Updated the library dependencies in the Consumer API's Gradle files . 3. Tested the changes both locally and on TAP. |