Good [morning/afternoon], everyone. I’m Manjo, a Data Engineer, and I’m excited to present my progress and achievements over the past month. This presentation outlines the work I’ve done in database management, stored procedure development, and code refactoring, showcasing how my technical expertise has contributed to our project goals. My primary tools include Microsoft SQL Server, Azure Databricks, and PySpark, and I’ve leveraged them extensively to ensure efficient data handling and workflow optimization. Under the guidance of my manager Ishaan, I’ve addressed various challenges while maintaining focus on delivering high-quality results.

To start, let me summarize my key responsibilities. My role as a Data Engineer involves designing and managing data pipelines, creating database schemas, and developing code solutions to optimize data workflows. Specifically, I’ve worked with Databricks for big data transformations, PySpark for efficient parallel processing, and stored procedures for robust database operations. These tools have been essential in achieving the project objectives and ensuring data integrity across all stages of processing.

Over the past month, I’ve completed several critical tasks. First, I designed and created all the necessary database tables based on the provided schema design. This foundational step ensured that our database structure aligned with the project’s requirements. Next, I collaborated with Ishaan to generate dummy data using Documentum and set up this data in an Azure Databricks notebook, laying the groundwork for further development and testing. I also refactored a shared notebook to improve its modularity and readability, enabling smoother integration into our data workflows.

Another significant accomplishment was the development of stored procedures, or SPs, which streamline database operations. I created two JSON-driven SPs: one for inserting and parsing data into the user upload table and another for fetching data with dynamic filters. These SPs are highly efficient, flexible, and reusable, and I demonstrated their functionality in the development database. Following feedback, I refined the JobTrigger SP, separating its functionality for standalone use to enhance maintainability. Additionally, I developed SPs for managing the submission and file details tables, ensuring conditional checks and error handling were in place for data integrity.

In terms of testing, I conducted end-to-end validations of all SPs, covering both positive and negative scenarios. For example, the SPs for the file details and binary tables underwent rigorous testing to ensure they processed JSON data accurately and aligned with Ishaan’s specifications. My testing approach guaranteed that the workflows functioned as intended, with no disruptions or data inconsistencies.

Code refactoring was another focus area. I modularized the notebook provided by Ishaan, making it more readable and maintainable. This involved optimizing existing logic and restructuring the code to improve performance. While this task was challenging due to the complexity of the original code, the final outcome was a significant improvement in both functionality and efficiency.

Looking back at my achievements, I’m proud to say that I completed all assigned tasks within the project timeline. My work on dynamic SPs has enhanced the project’s functionality, while my code refactoring efforts have improved overall quality. These contributions align with our team’s goal of delivering scalable and efficient data solutions.

Of course, there were challenges along the way. Developing complex SPs required meticulous planning to ensure functionality across different scenarios. Handling JSON data added another layer of complexity, as it demanded precise parsing and processing. Refactoring existing code was also a demanding task, requiring a deep understanding of its logic and flow. However, by collaborating closely with Ishaan and leveraging my technical skills, I was able to overcome these challenges effectively.

Looking ahead, I plan to continue refining our SPs for better performance and scalability. I’m also eager to explore advanced features in Databricks and PySpark that can further enhance our data workflows. My ultimate goal is to contribute to the project’s long-term success by driving innovation and efficiency in data engineering.

To conclude, I hope this presentation has provided a clear overview of my contributions and achievements. I’ve made significant progress in database management, SP development, and code optimization, and I look forward to building on this foundation in the months ahead. Thank you for your attention, and I’m happy to address any questions you may have.