Assignment 10 - Azure Data Factory

This assignment aims to assess your understanding and practical skills in using Azure Data Factory (ADF) for advanced ETL processes, including integration with Azure Data Warehouses (DWH), performing data transfor-mation using Data Flows, handling incremental loads, real-time data migration between on-premises and cloud systems, and implementing advanced transformations such as filtering, sorting, aggregation, merge operations, lookup, select. The assignment also tests your ability to use foreach loops, and expressions in ADF for efficient data pipeline management and automation.

1. Design an ADF pipeline to copy data from an on-premise Azure SQL database to Azure Cosmos DB, ensuring data consistency and performance optimization. Pick correct options of partitioning for better performance.
2. Create Pipeline using Azure Data Flow in Azure Data Factory to apply Filter and Sort transformations on datasets.
3. Design an ADF pipeline to implement aggregate operations, such as sum, average, max, min and count, within an Azure Data Flow.
4. Create best approach to bulk copy data from multiple homogenous sources into Azure SQL Database using ADF pipelines. Show usage of Lookup, For Each Loop and Expressions in Azure Data Factory.
5. Implement incremental load Pipeline in Azure Data Factory for handling datasets, ensuring efficient insert/upsert/updates to the target storage without re-inserting the entire dataset?
6. What are the key steps to connect Azure Databricks to Cosmos DB for real-time analytics and data transformation using spark and Databricks.