**Microsoft Data Engineer Track - Graduation Project Framework**

**1. Introduction**

* **Objective**: Develop a functional data engineering solution using Microsoft Azure data services, focusing on key aspects of data storage, transformation, and analysis.
* **Track Relevance**: Relate the project to the Microsoft Data Engineer track, emphasizing the use of Azure data storage, transformation, and analytics services.
* **Expected Outcomes**: A functional data engineering solution deployed on Azure, demonstrating core data engineering principles such as ETL (Extract, Transform, Load) processes, data warehousing, and real-time processing.

**2. Project Proposal**

* **Title**: Provide a relevant project title.
* **Team Members**: List all team members and their roles.
* **Problem Statement**: Clearly define the problem your project will solve using data engineering techniques.
* **Proposed Solution**: Describe your approach, detailing the specific Azure services and data engineering tools you'll use.
* **Scope and Limitations**: Define the scope of the project and any anticipated constraints.

**3. Project Plan**

* **Milestones**: Identify key phases of the project such as data collection, ETL process, real-time data processing, and deployment.
* **Task Breakdown**: Assign tasks to each team member and set deadlines for completion.
* **Resources Needed**: List the necessary software, hardware, datasets, or Azure services.
* **Risk Management**: Outline potential risks and propose strategies to mitigate them.

**4. Data Preparation and Analysis**

* **Data Collection**: Gather and explore datasets using Python and Azure tools like Azure Data Factory.
* **Data Preprocessing**: Apply data preprocessing techniques such as handling missing data, feature scaling, and normalization.
* **Exploratory Data Analysis (EDA)**: Use visualization tools (e.g., Matplotlib, Seaborn) to explore data, identify patterns, and prepare features for modeling.

**5. Data Engineering Implementation**

* **SQL Data Warehousing**: Set up and configure a SQL data warehouse in Azure, defining the schema and data management strategies.
* **ETL Process**: Design and implement ETL workflows using Azure Data Factory to move and transform data.
* **Real-Time Data Processing**: Implement real-time data processing solutions using Azure Stream Analytics, developing Stream Analytics jobs to handle data streams.
* **Data Integration**: Integrate data from various sources into Azure Synapse Analytics for advanced transformations and analytics.

**6. Azure Integration and MLOps**

* **Azure Synapse Analytics**: Use Synapse to integrate data and perform advanced analytics.
* **MLOps Practices**: Implement MLOps practices, including version control, model monitoring, and logging using Azure DevOps and MLflow.

**7. Model Deployment**

* **Deployment**: Deploy your data models and data pipelines on Azure, making them accessible through APIs.
* **Optimization**: Optimize the deployed models and pipelines for performance and reliability.

**8. Documentation and Final Presentation**

* **Technical Documentation**: Provide detailed documentation of the data engineering processes, Azure integration, and custom implementations.
* **Final Report**: Compile a comprehensive report detailing the project’s objectives, methodology, results, and future considerations.
* **Final Presentation**: Prepare and deliver a presentation highlighting the key aspects of the project, including a demo of the solution.

**9. Deliverables**

* **Functional Data Engineering Solution**: A fully developed data engineering solution deployed on Azure.
* **Source Code**: All scripts and code used for data processing, integration, and deployment.
* **Documentation**: Comprehensive technical documentation and a final project report.
* **Presentation Materials**: A slide deck for the final presentation, including a demonstration of the solution.

**10. Assessment Criteria**

* **Data Engineering Solution** (30%): Quality and efficiency of the data engineering solution.
* **Azure Integration** (30%): Effective use of Azure services and MLOps practices.
* **Data Preparation and Analysis** (20%): Thoroughness of data preprocessing and analysis.
* **Presentation and Documentation** (20%): Clarity and completeness of the presentation and documentation.