**Case Study: Implementing ETL with Azure Synapse Pipeline and Azure Databricks for WNS**

**Scenario**

A global business process management company, **WNS**, wants to analyze its sales data to derive insights and optimize its operations. The company has data coming from multiple sources, including CSV files containing daily sales transactions. WNS aims to set up an ETL (Extract, Transform, Load) pipeline to process this data and store the results in a structured format in an Azure Synapse Analytics data warehouse for further analysis.

**Objectives**

1. **Extract** data from CSV files stored in Azure Data Lake Storage Gen2.
2. **Transform** the data using Azure Databricks Notebooks to clean and aggregate the data.
3. **Load** the transformed data into Azure Synapse Analytics.

**Components**

1. **Azure Data Lake Storage Gen2**: Stores raw CSV files.
2. **Azure Synapse Analytics**: Target data warehouse.
3. **Azure Databricks**: Performs data transformation.
4. **Azure Synapse Pipeline**: Orchestrates the ETL process.

**Sample Data Files**

* sales\_2024\_01.csv
* sales\_2024\_02.csv
* sales\_2024\_03.csv