1) I have experienced a scenario where I need to build a data pipeline for data ingestion process. The data is loaded from source server (flat files) to Azure file storage via SMB protocol. The event is triggered for data transfer securely everyday at a scheduled time.
Azure data factory is used to design and implement ETL processes.
Azure stream analytics is used for streaming data ingestion.
Defining the data format and schema for the streaming data.
Azure active directory (Entra ID) is used for user authentication and authorization.
Azure security center is used to monitor and audit data access.
Azure key Vault is used to implement data encryption and access controls., generated tokens
Azure logic apps is used to create and run automated workflows.
The data transformations, aggregations, removing of null values, dealing with redundancy, using Azure databricks.
The data is moved from raw (bronze) layer to curated (Gold) layer
The curated data is stored in Azure Synapse.
I have used Azure autoscaling to add more processing

nodes to handle increased data volumes efficiently.

2) In a previous project (Burger king), we are using orders and feedback surveys data of stores in south-west region of the US. We get data from different sources (stores) in various formats, grouped by county and state. We use Azure data factory to run and implement the ETL process. We use Azure databricks to perform transformations, aggregations, data conversions, normalization, defines dynamic schema for streaming data of orders/surveys.
Implements optimization techniques such as partitioning the data into equal partitions, bucketing into similar clusters and caching for processing the frequently accessed data.
Maintaining data quality by data profiling, defining data rules and regulations by discussing with data stakeholders.
Develops data reports and tables with curated, sorted data, which can be used by data scientists / data reporting team to predict the most selling product and reviews to improve store value.