**END TO END AZURE DATA ENGINEERING PROJECT**

**Flow Chart:**

Power BI

Aggregates

Databricks

Azure Storage

SQL SERVER

Azure Data Factory

Pyspark Spark SQL

**Business Requirement:**

* We have data available in SQL server and we need to move into cloud storage
* We have to build one pipeline which will transfer data and run daily
* Connect the Databricks to Storage account
* Build one aggregate table for the reporting with all the business related KPI
* Build dashboard using Power BI

**Steps:**

* Go to SQL server and upload csv file in one of the tables
* Create Azure Storage account
* Create Azure Data Factory
* Azure setup integration run time – Azure Data Factory
* Azure load data from SQL to Storage account – Data pipeline
* Connect Databricks to Storage account – Pyspark
* Databrick aggregates table building – spark SQL
* Power BI Dashboard

**KPI:**

* Total Pizza sold
* Total orders, total revenue, avg pizza per orders
* Sales vs orders month over month
* Daily orders, monthly, hourly orders trend
* Sales by pizza category
* Sales by pizza size
* Top 5 orders pizza