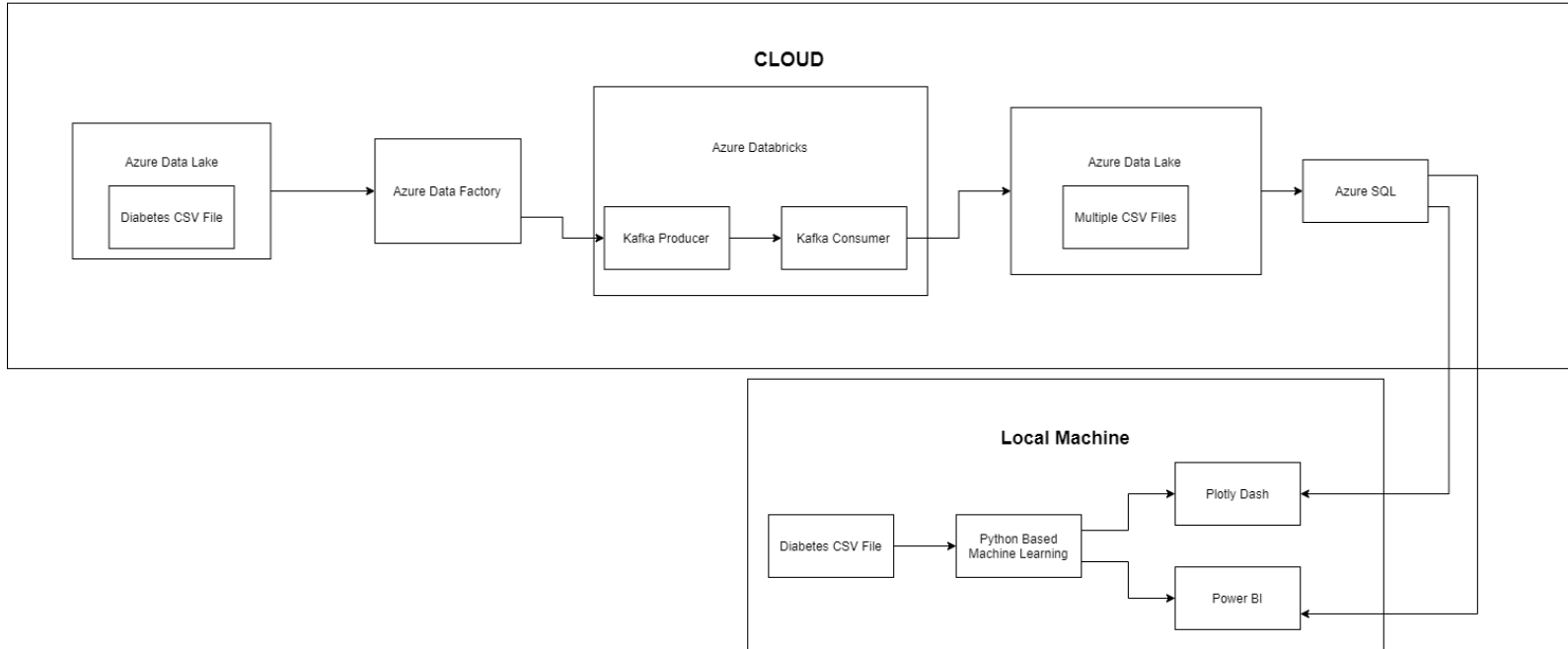


Data Platform

Outline of Data Platform



STEPS

Establishing Connection in the Cloud:

1. Store Diabetes CSV file in Data Lake.
2. Pull data from Data Lake containing CSV file into data factory. This is also known as ingesting data from source.

Simulation of real time streaming of Data:

3. Establish connection between data factory and Producer/Consumer codes.
4. Run Producer and Consumer code in Azure Databricks.

Storing Data:

5. Store Consumer output as multiple small csv files in our Data Lake.
6. Load output CSV files into Azure SQL.

Analysis of Data:

7. Load visualizations and tables made in SQL into either Power BI or Plotly Dash.

Building Machine Learning Model on Local Machine:

8. Open Diabetes CSV file.
9. Build machine learning model using python-based machine learning.
10. Load model visualizations into either Power BI or Plotly Dash