Initial data gathering

US Census → CSV Files → Azure blob container → Azure Databricks (databricks + kafka Producer) → Azure Databricks (databricks + kafka Consumer)

Scraped websites → Azure Databricks

Azure data lake → SQL Database

Automated process

Data Factory → Azure Databricks (databricks + kafka Producer) → Azure Databricks (databricks + kafka Consumer) → SQL Database → Power BI

Machine Learning → Power BI

Data gathering

1. Download files from US census data
2. Save data as CSV
3. Upload files into Azure blob containers
4. Scrape data from internet in azure databricks
5. Import CSV files from blob containers
6. Run Kafka producers and consumers in databricks
7. Output data to azure data lake and azure SQL databases

Automation process: can be seen as direct addition to data gathering part

8. Automate data factory to run data bricks
9. Run Kafka producers and consumers in databricks
10. Update data in SQL databases
11. Import data from databases in PowerBI and apply machine learning