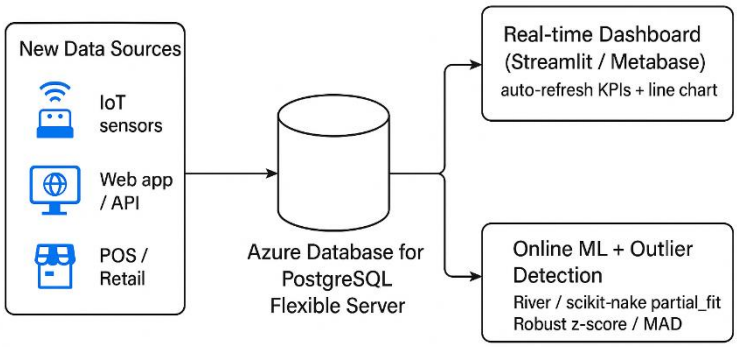
Project 1: Building a Streaming Analytics Pipeline on Azure PostgreSQL

In this team project, you will design and deploy a multi-table database hosted on Azure PostgreSQL. The database should represent a realistic business or operational setting of your choice, such as retail transactions, IoT sensor data, logistics tracking, or web traffic. You will simulate a live environment where hundreds of new observations arrive each minute and are continuously inserted into the database.



Your work goes beyond data storage. Each team will build three components on top of the streaming database:

- A dashboard that monitors key indicators in your domain. It should update in real time or near real time as new records are inserted.
- An outlier detection system that flags unusual patterns or errors in the incoming data stream.
- A machine learning model that operates on streaming or online data. It should update its predictions as new records arrive.

Timeline

1. Project idea (30 mins): Check feasibility
2. Prototype running locally (2 hours)
3. Online Deployment (1.5 hours)
4. Project wrap-up and Conclusions (0.5 hours)

Deliverables:

i. A relational schema with at least three interrelated tables deployed on Azure PostgreSQL.
ii. A data generation and ingestion pipeline that inserts hundreds of new records per minute.

iii. A working dashboard connected to the database, showing live updates.
iv. An implemented outlier detection method applied to the incoming data.
v. A machine learning component trained and evaluated on the stream, capable of processing data continuously or in micro-batches.
vi. A final report and 10-minute presentation demonstrating your system, design choices, and key results.

The goal is to simulate an end-to-end streaming analytics system, from raw data ingestion to actionable insights.

Note: See Project Ideas file to get inspiration (the team can define their own idea; this is just a reference to stimulate your creativity)