**Code Explanation:**

A diagram of a computer process

Description automatically generated

**1. Batch Processing Code (mainapp\_batch.py)**

* Performs a full load for batch processing via **PySpark**.
* Reads all input data from .csv files located in the specified input directory.
  + Input file contains fields such as TransactionID, TransactionTime, Amount, CustomerID, Location, etc.
* Applies business logic to classify transactions:
  + Legitimate transactions are marked as 0.
  + Fraudulent transactions are marked as 1.
* Conducts basic analysis, including:
  + Determining the time-of-day transactions occur.
  + Evaluating transaction amounts.
  + Assessing location-based risk of transactions.
* Saves the processed data in the output directory as **parquet files**.
* Processes data every 10 minutes.

**2. Masked Data Preparation (fake\_data\_preparation.py)**

* Contains code for generating masked and fake data.
* The output of this program serves as input for the batch processing program (mainapp\_batch.py).

**3. Job Scheduling (airflow\_spark\_submit.py)**

* Schedules the main batch processing job (mainapp\_batch.py) using **Airflow**.

**4. Kafka Producer (kafka\_producer.py)**

* Sends a specified number of rows to Kafka depending on input parameters.

**5. Stream Processing (mainapp\_stream.py)**

* Automatically reads messages from the Kafka topic.
* Applies business logic to classify transactions:
  + Legitimate transactions are marked as 0.
  + Fraudulent transactions are marked as 1.