Would create a Data PIPELINE using for example Apache NIFI,
where all three sources of data is ingested into a Database

Kafka node will be consumed by a NIFI module, where the data should have already been transformed to correct mapping to the destination database

FILES assuming xlsx or csv files will be read in a specific interval from an STPF folder, where the data will be ingested into the same database but a different table if data points are different from the realtime kafka ingestion.

The database tables should be set with a Primary Key, so data can be upsert and IGNORE duplicates.

API calls will need to be configured in a specified interval and pass the necessary params to extract the required data. This can be computed in a python/java code, or even scheduled by NIFI to read the API and transform the data to the format required by the database before ingestion.

For more parallel processing Apache Spark can be used to run processes in parallel over multiple computing nodes.

The data above is assumed structured, if it is not structured, mongo db or ElasticSearch can be used to store and index the data with the right flattened schema.

Mongo or Elastic can be configured in a cluster mode to enable scalability.

For Dashboarding:

Depending on the use case,

We can use Power BI,Tableau or even Kibana, to build an interactive Dashboard.

For some huge data aggregations, we could simply the dashboard load by pre-aggregating these values using scheduled stored procedure to update specific tables that will be used for Dashboarding purpose.

Dashboard can show a trend chart over events of frauds, where users can visualize the trends during specific events. We could also show the Monetary benefits of catching Fraud in the dashboard for management transparency to see value.

Trends in event of the fraud could provide insights into future trends.

Using ML, we could either used supervised machine learning on pre-tagged data of fraud cases

or for post event , use anomaly detection to see behavior in customers account that is out of his norm. for unsupervised model using algorithm such as Isolation Forest could provide anomaly behavior detected, that differs form the customers normal pattern.

There are many approaches to this project, the above is just a simplified overview of an approach, which I have configured in the past.