

# Data Sources

The data you will be sourcing from is contained in two sources.

1. A database ([SQLite](#)) of *customer* data
2. A [CSV file](#) of *stream* level data

You will create a simple data pipeline that:

- simplifies the data for future analysis
- performs quality assurance checks.

The process of building *the data ingestion pipeline* entails extracting data, transforming it, and loading it into an appropriate data storage technology. When constructing a pipeline it is important to keep in mind that they generally process data in batches. For example, data may be compiled during the day and the batch could be processed during the night. The data pipeline may also be optimized to execute as a streaming computation (i.e., every event is handled as it occurs).

## IMPORTANT NOTE!

If you encounter an error that the notebook cannot open a file, or cannot find a table in the database, make sure your notebook is pointing to the right file in the right directory!

If you are running the notebook in Watson Studio be sure to follow the directions at <http://ibm.biz/notebook-in-studio> to ensure that your notebooks can find the files.