## Team 20 Code Sample doc

## Environment setup

1. Spark installation: MAC:

<https://github.com/charlie-ph/BigDataAnalytics/blob/master/Installations-HowTos/How-To-Install-Spark-On-MACOS.md>

Windows:

<https://github.com/charlie-ph/BigDataAnalytics/blob/master/Installations-HowTos/How-To-Install-Spark-On-Windows.md>

1. PostgreSQL installation:

<https://www.postgresql.org/docs/current/tutorial-install.html>

## How to run the code

Part 1: PostgreSQL

1. run code/Postgre/create\_file.sql to create Data-warehouse in Postgre
2. import all csv files in /data into your database
3. run code/Postgre/SQL\_basic\_search.sql to get SQL results

Part2: Spark

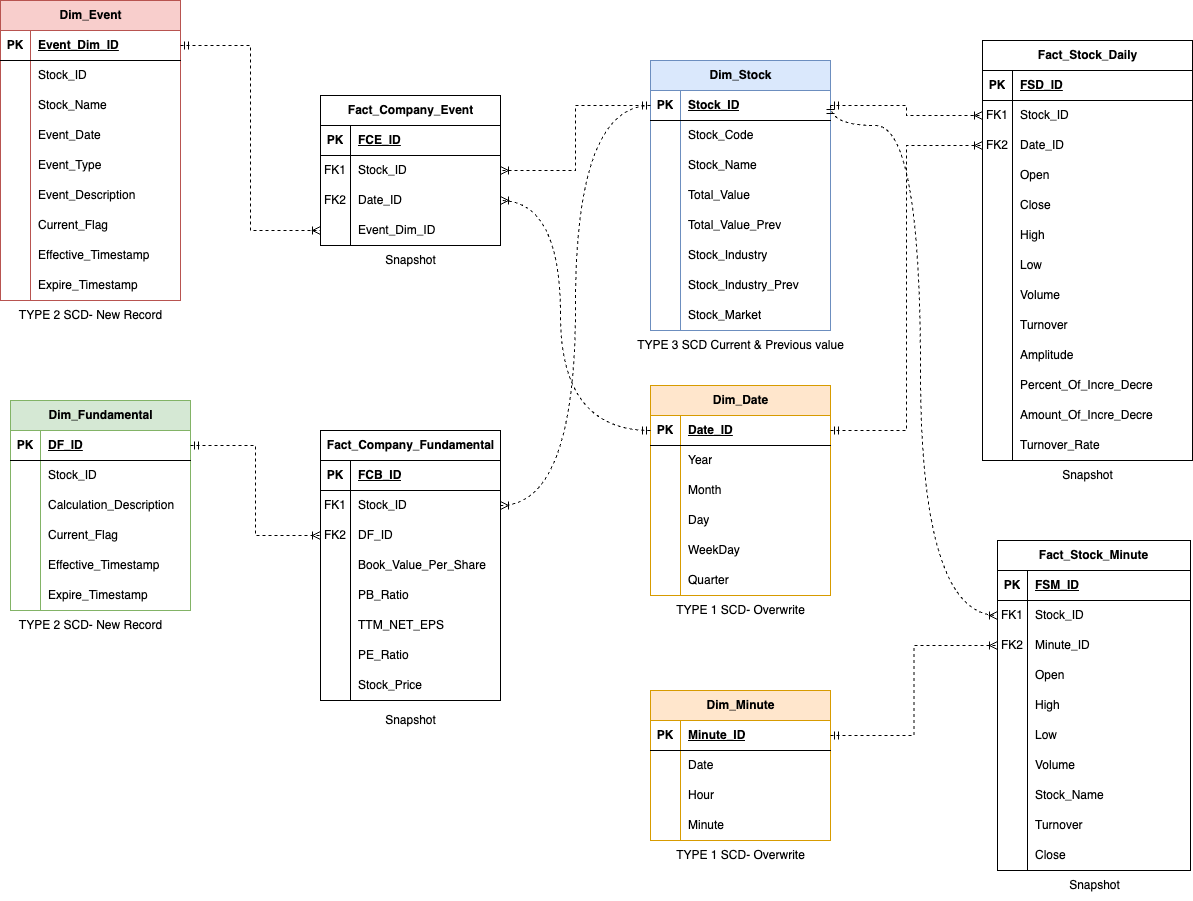
1. run code/Spark/Example.ipynb in your notebook.
2. make sure you have correct address to access all csv files.

## Dataset Explanation

I designed this database according to some basic requirements of my own trading system.



Database ERD:



For each table, we provide one notebook to finish the ETL process(code/Database/\*.ipynb). All the data is from this API: akshare(https://akshare.akfamily.xyz/)

**Results of running the code with data**