

# An Application of Time-Series Database in Stock Market

Chen Jiabei 2020141410356

## 1. Background

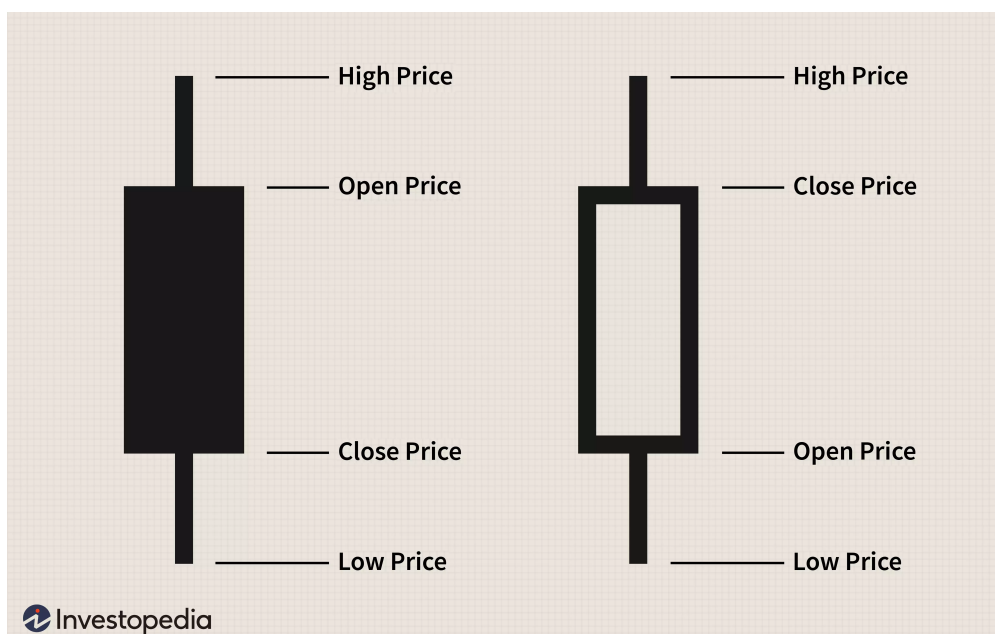
A **time series database (TSDB)** is a software system that is optimized for storing and serving time series through associated pairs of times and values. One of the most common uses for time-series data is to collect intraday securities information. Now, we will talk about how to apply time-series database into stock market data.

## 2. Prerequisites

- Python
- Pandas
- TimeScaleDB

## 3. Schema Design

As we all know, candlestick charts (also called K-line) are widely used in stock market, which shows the market's open, high, low, and close price for the day and helps traders to determine possible price movement based on past patterns. Here is what the candlestick chart looks like.



The data fields we are using are as follows:

Field	Description
time	starting time of the minute
symbol	ticker symbol
price_open	opening price of the stock
price_close	closing price of the stock
price_low	lowest price in the minute
price_high	highest price in the minute
trading_volume	trading volume in the minute

## 4. Get the Stock Data

We use Alpha Vantage API to get top 100 US ticker symbol's intraday stock data in the last three month. We can have a glimpse of the Apple's stock data fetched below.

AAPL							
		time	open	high	low	close	volume
0	2022-07-07	20:00:00	145.9200	145.9300	145.90	145.90	3436
1	2022-07-07	19:57:00	145.9100	145.9200	145.91	145.92	810
2	2022-07-07	19:55:00	145.9600	145.9600	145.96	145.96	381
3	2022-07-07	19:54:00	145.9001	145.9001	145.90	145.90	1883
4	2022-07-07	19:52:00	145.9500	145.9600	145.95	145.96	604
...	...	...	...	...	...	...	...
16113	2022-06-08	04:08:00	148.1800	148.1800	148.15	148.15	410
16114	2022-06-08	04:06:00	148.2300	148.2400	148.18	148.18	1818
16115	2022-06-08	04:04:00	148.3900	148.3900	148.30	148.30	716
16116	2022-06-08	04:03:00	148.4100	148.5000	148.41	148.50	1513
16117	2022-06-08	04:01:00	147.8500	148.4100	147.85	148.30	3247

## 5. Create Databases/Tables

First, we should create a regular table which consists of all the columns needed to store the stock data as shown in the schema design. The SQL to create a table is as shown below.

```
CREATE TABLE stocks_data ( "time" timestampz NOT NULL,
                             symbol text NULL,
                             price_open double precision NULL,
                             price_close double precision NULL,
                             price_low double precision NULL,
                             price_high double precision NULL,
                             trading_volume int NULL )
```

Next, the most important part is to create a **hypertable** which is the core of TimescaleDB. Hypertables enable TimescaleDB to work efficiently with time-series data. The SQL to create a hypertable is as shown below.

```
SELECT create_hypertable('stocks_data', 'time');
```

## 6. Insert the Stock Data

The SQL to insert data is as shown below.

```
SQL = "INSERT INTO stocks_data(time, symbol, price_open, price_close,
price_low, price_high, trading_volume) VALUES(%s,%s,%s,%s,%s,%s,%s);"
for stock_data in stock_datas:
    cursor.execute(SQL, stock_data)
```

Instead of using common SQL to insert data, we use **pgcopy** library to sufficiently insert data into hypertables with high performance. The code is as shown below.

```
COLUMNS = ('time', 'symbol', 'price_open', 'price_close', 'price_low',
'price_high', 'trading_volume')
mgr = CopyManager(conn, 'stocks_data', COLUMNS)
mgr.copy(stock_datas)
```

## 7. Query and Analyze

First, we can have a look at the stock price change of Tesla over last three month. The SQL query is as shown below.

```
SELECT time_bucket('1 day', time) AS bucket,
       last(price_close, time) AS last_closing_price
FROM stocks_data
WHERE symbol = 'TSLA'
GROUP BY bucket
ORDER BY bucket
```

Tesla's daily stock price over last three months



Then, we can show our interest in the 15-min candlestick chart of Google on 2022-07-05. The SQL query is as shown below.

```
SELECT time_bucket('15 min', time) AS bucket,
       FIRST(price_open, time) AS price_open,
       LAST(price_close, time) AS price_close,
       MAX(price_high) AS price_high,
       MIN(price_low) AS price_low
FROM stocks_data
WHERE symbol = 'GOOG' AND date(time) = date('2022-07-05')
GROUP BY bucket
```

15-min candlestick chart of Google, 2022-07-05



## 8. Summary

Through this practice, we get more familiar with the time-series database and the real application to our life. All the code will be published in my repository.

## 9. References

- [1] [Overview | Timescale Docs](#)
- [2] [Free Stock APIs in JSON & Excel | Alpha Vantage](#)
- [3] [Understanding a Candlestick Chart \(investopedia.com\)](#)