

Financial data & economic

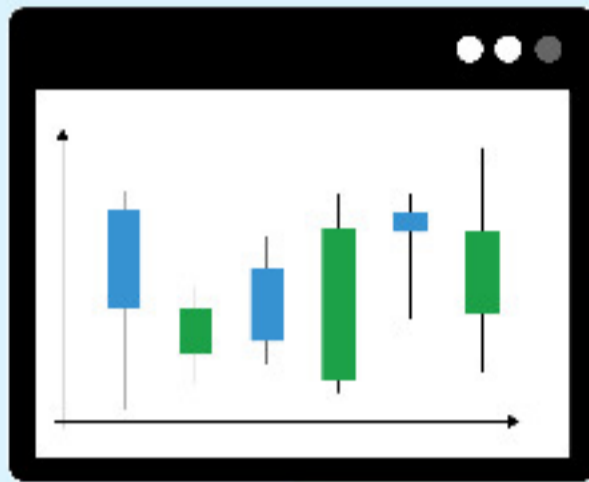
Using big data to explore the impact of economic

report analysis

news releases on Foreign Exchange trading

As the world plugs into the internet, 'big data' is becoming more mainstream. With such abundance of data it is important to adequately collect, filter, manage and store correctly in order to benefit from its unseen relationships. This is true in all industries, finance in particular, where the key asymmetrical information needed to stay ahead of competition is buried within a mountain of data. Working on behalf of a financial broker, using big data techniques, we uncovered relationships that will be used in the creation of models to reduce their risk portfolio.

Financial Data



Bid/ask price data along with the associated volume was collected live from the broker pricing feeds for each instrument. Then stored in the RAM of a kdb+ database and written to disk every 24 hours.

Data Scraping



Using a data scraper, economic news data were pulled from 'forexfactory.com'. This data included date-time, importance, predicted and actual values. Using its time-stamp it was cross referenced with the financial data.

01101010101001001
0100101000101010
11001101000101100
10000001010101001

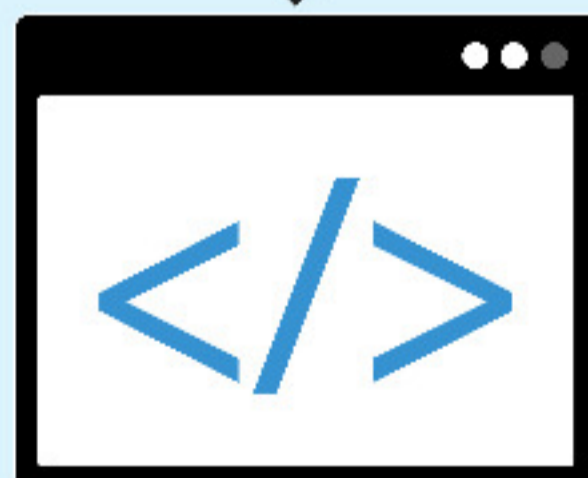
Data Cleansing



Data cleansing is an important aspect of dealing with big data. As no data sources are perfect, defects are inevitable and to draw any reliable results they must be corrected. In our case, both data sources had incorrect time-zones, invalidating all initial results. To correct this, the data scraper was altered to convert date-time stamps to UTC.

Holes were also found in the time-series financial data as a result of database software patches and price feed drop-outs. These holes were filled with an equivalent data source (Dukascopy), by back-loading data into the kdb+ database.

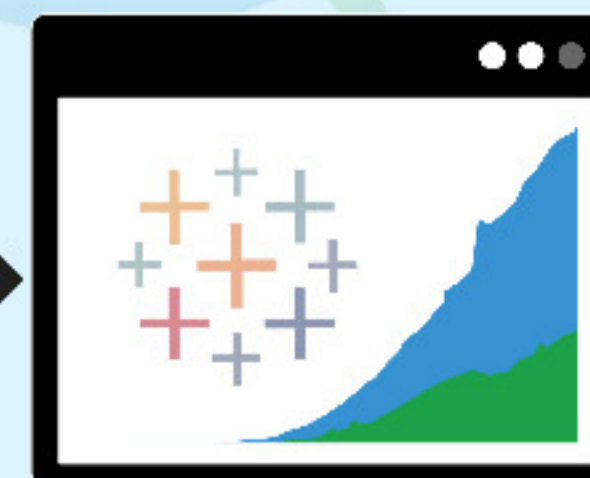
Data Restructuring



In order to read and visualise the data it needed to be restructured and relevant values to be calculated. This allowed us to use querying tools like Tableau to properly graph the data.

Due to limitations in our setup, data files were limited to around 6 Gbps, so different methods of compression and filtering were used.

Data Visualisation



Using Tableau, the data could be visualised to explore relationships. These results will be used to model price and client behaviour around the economic news releases, allowing the model to automatically hedge risk.