The provided code analyzes New York Stock Exchange (NYSE) values for the year of 2008. The analysis is done using the Hadoop tech stack. MapReduce is used to split the data into maps and reduce code snippets. Apache Hive is used to create the query. Hive uses an SQL-like language called Hive-QL.

The algorithm is built to first create a table that holds all of the provided NYSE data. The table holds information such as the stock’s symbol, date, opening price, high and low for the day, and the closing price. A select statement is then created to select all stock symbols in 2008 and perform the covariance calculation. The covariance calculation is performed by first finding the average of the product for two stock prices. This product is then divided by the product of the averages for the same two stocks.

The end goal of this code is to find the covariance between stocks and display the relationship between them. The result from the analysis shows that stocks QRR and QTM have a positive covariance and therefore are more likely to mirror each other in their movement. Another example is the stocks QRR and QXM. These stocks have a negative covariance and are more likely to move in opposite directions of each other. These conclusions would be useful to a stockbroker for trading stocks on the market.

The provided solution uses the Hadoop Distributed File System (HDFS). This system allows for large amounts of data to be stored and processed on file systems that are not colocated in one location. It offers the ability to scale the compute power by adding more nodes for helping with the processing tasks.

Many tools have a method for finding covariance among large amounts of numbers. Tools such as Microsoft Excel can perform the calculation for us. This method falls apart when the data pool gets large enough. This is when big data processing technologies, such as Hadoop, must be used instead.