Stock Volatility Analysis by Map-reduce

Method and Implementation

- I used 3 mapreduce stages in order to find the top 10 stocks with minimum volatility and top 10 stocks with maximum volatility.
- In the first stage the filename which is the stock name and date fields are extracted from the file that is being processed, the stock name/month/year is set as the key and the particular line of data is stored as a value.The reducer receives the input from the mapper with the values for every stock name and every month combined.The reducer then extracts the first and the last date for every month of and calculates the Xi value and sets it back with the key.The output of the first stage is (stock name/month/year,Xi value)
- In the second stage the mapper receives the input from the first reducer,the mapper receives the key as stock name/month/year as key and Xi as value,the mapper removes the month and year from the stock name and puts stock name as the key which will result in combining of all the Xi values of a particular company for all the months.The reducer receives the input from the mapper with the key as stock name and all Xi values for that company.The reducer goes through all the Xi values and computes the stock volatility according to the formula.The output of the second reducer is(stock name,volatility). we set a counter in job 2 ‘s configuration and increment it every time when we calculate volatility of a stock and write it in the intermediate file.
- In the third stage the mapper receives input as (stock name,volatility) .As we need to display the top 10 stocks with maximum volatility and top 10 stocks with minimum volatility I sorted the data using mapreduce internal sorting the mapper always writes output sorted by the key,so the in this stage the mapper all swaps the key and value which is received in the input
- The output from the mapper will be (volatility,stock name).The reducer will receive the input in a sorted order so the top 10 stocks are the stocks with minimum volatility.we use the counter that we set during job2,which tells us the total number of stocks,using that the counter we display the top 10 stocks with maximum volatility as those would be the last 10 stocks in the intermediate file.

First stage:

The input to the mapper will be:	
Key	**Value**
AAIT/12/2014	20141231,32.95,32.95,32.95,32.95,000,32 .95
AAIT/12/2014	20141230,32.86,32.95,32.86,32.95,1200,3 2.95
AAIT/12/2014	20141229,33.18,33.22,33.18,33.20,1700,3 3.20
AAIT/12/2014	20141201,34.20,34.20,34.20,34.20,100,33 .84

The output from the mapper will be:	
Key	**Value**
AAIT/12/2014	{ (20141231,32.95,32.95,32.95,32.95,000,3 2.95),(201412-30,32.86,32.95,32.86,32.95, 1200,32.95).....(20141201,34.20,34.20,34.20,34.20,100,33.84) }

Stock Volatility Analysis by Map-reduce

The input to the reducer will be:	
Key	**Value**
AAIT/12/2014	{(20141231,32.95,32.95,32.95,32.95,000,3 2.95),(201412-30,32.86,32.95,32.86,32.95, 1200,32.95).....(20141201,34.20,34.20,34.20,34.20,100,33.84) }

The output from the reducer will be:	
Key	**Value**
AAIT/12/2014	Xi value
AAIT/11/2014	Xi value
AAIT/10/2014	Xi value
AAIT/01/2014	Xi value

Second stage:

The input to the mapper will be:	
Key	**Value**
AAIT/12/2014	Xi value
AAIT/11/2014	Xi value
AAIT/10/2014	Xi value
AAIT/01/2014	Xi value

The output from the mapper will be:	
Key	**Value**
AAIT	{Xi value,Xi value,Xi value,Xi value,Xi value}
AAL	{Xi value,Xi value,Xi value,Xi value,Xi value}

The input to the reducer will be :	
Key	**Value**
AAIT	{Xi value,Xi value,Xi value,Xi value,Xi value}
AAL	{Xi value,Xi value,Xi value,Xi value,Xi value}

The output from the reducer will be:	
Key	**Value**
AAIT	volatility
AAL	volatility
AAME	volatility

Third stage:

Stock Volatility Analysis by Map-reduce

The input to the mapper will be:	
Key	**Value**
AAIT	volatility
AAL	volatility
AAME	volatility

The output from the mapper will be:	
Key	**Value**
volatility	AAIT
volatility	AAL
volatility	AAME

The input to the reducer will be:	
Key	**Value**
volatility	AAIT
volatility	AAL
volatility	AAME
The output from the reducer will be:	
Key	**Value**
AAIT	volatility
AAL	volatility
AAME	volatility

The output at the last stage will be sorted according to the volatility.The third reduce map stage is only used for sorting.By default the map phase sorts the output according to the key.So we sort the output according to the volatility and the reducer displays the top 10 stocks with minimum and maximum volatility.