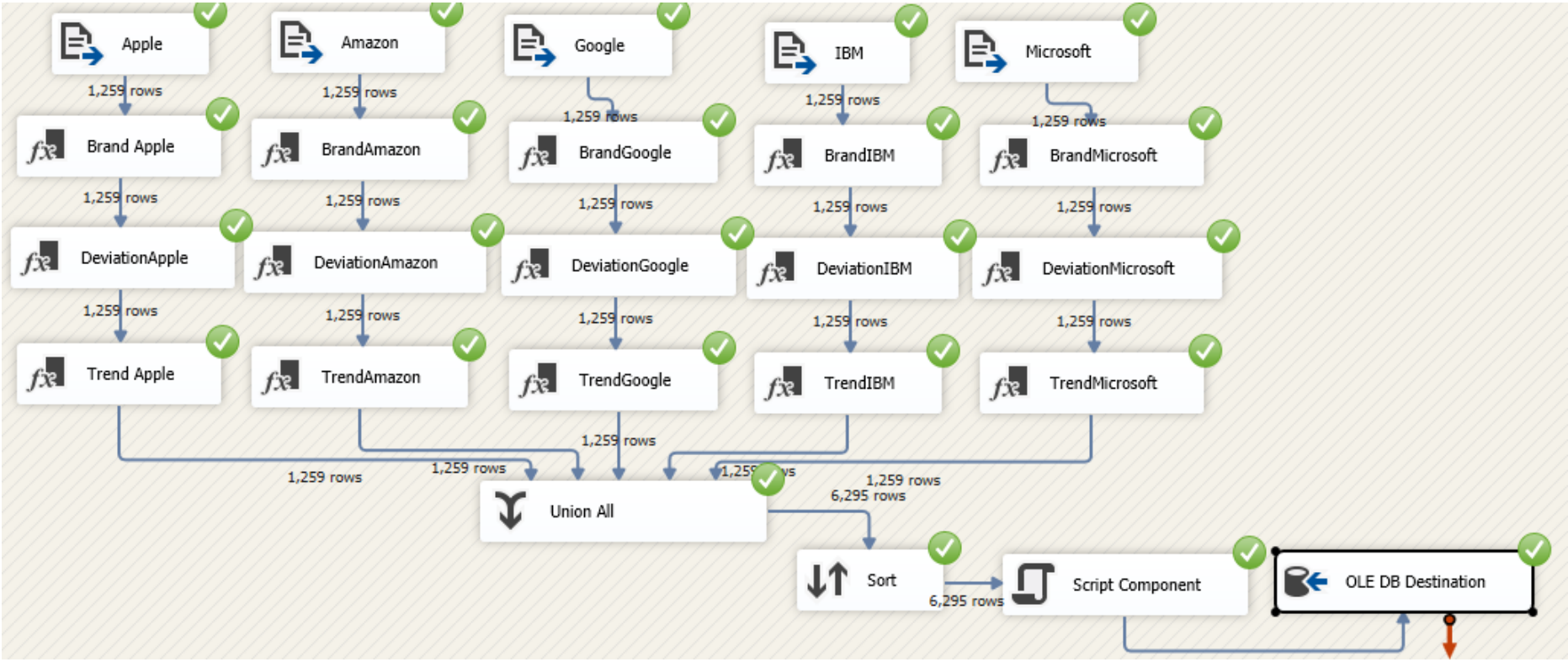
Implementation of ETL (Extract, Transform, Load) using SQL Server Integration Services (SSIS) Package



ETL creates a pipeline where data is first **extracted**, then **transformed** and finally **loaded** into a destination.

Extract: Apple, Amazon, Google, IBM and Microsoft refer to flat file (CSV files) data sources from where data is extracted.

Transform: It involves 6 main steps.

- Assigning brand name and saving it as a Brand column.
- Computation of deviation between Closing and Opening value of stocks and saving it as a Deviation Column.
- Computation of overall trend based on deviation and saving it as a Trend Column.
- Union all – combines all data sources.
- Sort – performs sorting based on DatePeriod.
- Script component – add a RowID column to given data.

Load: Saving data after necessary transformation into SQL Server database.

DatePeriod	OpenVal	HighVal	LowVal	CloseVal	Vol	Name	Brand	Deviation	Trend	RowID
2013-02-08 00:00:00.000	199.97	202.09	199.68	201.68	2893254	IBM	IBM	1.71000000000001	1	1
2013-02-08 00:00:00.000	261.4	265.25	260.555	261.95	3879078	AMZN	Amazon	0.550000000000011	1	2
2013-02-08 00:00:00.000	67.7142	68.4014	66.8928	67.8542	158168416	AAPL	Apple	0.140000000000001	1	3
2013-02-08 00:00:00.000	390.4551	393.7283	390.1698	393.0777	6031199	GOOGL	Google	2.62259999999998	1	4
2013-02-08 00:00:00.000	27.35	27.71	27.31	27.55	33318306	MSFT	Microsoft	0.199999999999999	1	5
2013-02-11 00:00:00.000	68.0714	69.2771	67.6071	68.5614	129029425	AAPL	Apple	0.490000000000009	1	6

Once the data is loaded into SQL Server database, data can be explored by executing SQL queries.

	Month	YEAR	Brand	CloseVal	Vol	Trend
1	2	2013	IBM	201.68	2893254	1
2	2	2013	Amazon	261.95	3879078	1
3	2	2013	Apple	67.8542	158168416	1
4	2	2013	Google	393.0777	6031199	1
5	2	2013	Microsoft	27.55	33318306	1
6	2	2013	Apple	68.5614	129029425	1
7	2	2013	Microsoft	27.86	32247549	1
8	2	2013	IBM	200.16	2944651	-1

Selecting DatePeriod, Brand, CloseVal, Vol, Trend columns and splitting DatePeriod using Datepart function

	Brand	HighestClosingVal	MeanClosingVal	MinimumClosingVal
1	Microsoft	95.01	51.06	27.37
2	IBM	215.8	167.26	117.85
3	Google	1187.56	682.23	383.34
4	Apple	179.26	109.07	55.7899
5	Amazon	1450.89	576.88	248.23

Implementing aggregation on closing values for each stock brand.

	Brand	MaxDeviation	MeanDeviation	MinDeviation
1	Microsoft	4.44	0.04	-3.34
2	IBM	6.28	0.03	-5.98
3	Google	50.45	-0.12	-38.22
4	Apple	8.25	0.01	-7.37
5	Amazon	81.38	0.01	-55

Implementing aggregation on deviation values for each stock brand.

	Brand	TotPctChange
1	Google	85767305.67
2	Amazon	72503296.6
3	IBM	20932393.88
4	Apple	13605597.07
5	Microsoft	6303038.04

Computing total percentage change for each stocks brand over total data period