1. Using all.csv as the input data implement transformations that should happen during mapping and reducing stages to implement each of the following operations:
    a. Filter rows of a table based on the specified criteria "SYMBOL equals GEOMETRIC"
    b. Select specific columns from those available (eg. SELECT SYMBOL, OPEN, HIGH, LOW, CLOSE)
    c. Generate a count of the number of rows in the table

OUTPUT 1:

	SYMBOL	SERIES	OPEN	HIGH	LOW	CLOSE	LAST	PREVCLOS	TOTTRDQT	TOTTRDVA	TIMESTAMP
522	GEOMETR	EQ	125	128.5	118.1	122.05	122	126.15	261878	32241048	17-DEC-2014
621	GEOMETR	EQ	98.8	100.5	98.8	99.1	99	99.55	49782	4956717	26-APR-2013
2461	GEOMETR	EQ	71.1	72.1	68.6	69.7	70	71.35	51757	3661284	17-FEB-2011
3139	GEOMETR	EQ	71.45	74.85	71.1	73.9	73.9	71.05	333378	24609182	27-JUN-2012
4298	GEOMETR	EQ	72.5	74.9	71.1	71.45	71.5	69.8	573476	41891076	17-FEB-2012
6168	GEOMETR	EQ	91.95	91.95	89	89.6	89.3	90.35	32616	2933904	17-FEB-2014
8013	GEOMETR	EQ	75.4	77.35	73.65	76.4	76.7	75.4	132718	10079417	17-JAN-2011
9841	GEOMETR	EQ	59.9	62.25	58.35	58.95	58.4	59.8	161343	9764387	17-JAN-2012
11740	GEOMETR	EQ	114.1	114.4	111.8	112.5	112.1	113	370594	41921818	17-JAN-2013
13654	GEOMETR	EQ	100.2	101.45	97.75	98.3	98.05	101.5	181534	17991643	17-JAN-2014
15540	GEOMETR	EQ	71.3	72.45	70.1	70.6	70.8	71.05	70902	5051659	17-JUL-2012
17383	GEOMETR	EQ	103.75	109.85	103.75	105.45	105.5	104.35	356137	38410815	17-JUL-2013
18062	GEOMETR	EQ	105.6	108.25	104.1	105.2	105.2	105.2	233804	24846517	02-MAY-2014
19157	GEOMETR	EQ	142	148.5	141.4	145.45	146	140.6	707939	1.03E+08	17-JUL-2014
21121	GEOMETR	EQ	119.3	121.35	117.6	118.6	118.8	121.35	410052	48870356	17-JUL-2015
21569	GEOMETR	EQ	103.5	105.8	101.8	103.15	104	103.15	36192	3750414	27-JUN-2013
23054	GEOMETR	EQ	57.9	58.15	56.1	56.35	56.4	57.4	55082	3125131	17-JUN-2011
24865	GEOMETR	EQ	101.35	105.25	101.35	103.05	104	101.35	146150	15151746	17-JUN-2013
26671	GEOMETR	EQ	138.4	140.2	136	137.9	137.65	137.85	166832	23096655	17-JUN-2014
28631	GEOMETR	EQ	106.4	109.65	105.55	107.5	107.75	105.3	298135	32159551	17-JUN-2015
30534	GEOMETR	EQ	61.3	61.75	60.5	60.8	60.8	61.4	20346	1240667	17-MAR-2011
32371	GEOMETR	EQ	172.9	182.3	171.5	177.55	177.4	171.7	1721883	3.06E+08	17-MAR-2015
34288	GEOMETR	EQ	59.95	60.2	58.15	58.45	58.5	59.9	22705	1343170	17-MAY-2011
36112	GEOMETR	EQ	72.5	73.5	70	70.45	70.3	71.9	316512	22740078	17-MAY-2012
37168	GEOMETR	EQ	67.6	67.6	66.7	67.2	67.25	66.95	24275	1630639	02-NOV-2010
37963	GEOMETR	EQ	102.8	105.3	102.05	104.2	105.15	103.3	198177	20522729	17-MAY-2013

```
,SYMBOL,OPEN,HIGH,LOW,CLOSE
0,SYMBOL,OPEN,HIGH,LOW,CLOSE
1,20MICRONS,37.75,37.75,36.35,37.45
2,PDUMJEPULP,13.7,13.7,13.45,13.65
3,3IINFOTECH,43.75,45.3,43.75,44.9
4,PEACOCKIND,4.0,4.0,3.8,3.95
5,ABCIL,129.55,130.8,128.35,130.0
6,PFC,192.0,192.4,186.35,188.35
7,ASHCONIUL,2.55,2.85,2.55,2.85
8,RCOM,59.95,60.0,58.0,58.25
9,GUJRATGAS,273.0,276.5,272.5,274.35
10,PERIATEA,167.95,167.95,167.95,167.95
11,GEMINI,2.6,2.6,2.55,2.6
12,MMFL,64.2,67.05,62.1,62.1
13,BHUSANSTL,455.4,464.8,455.4,461.8
14,GLORY,1.35,1.35,1.35,1.35
```

OUTPUT 3:

1893059

Exercise 2: Calculation of various statistical quantities and decision making using HIVE

1. Only lines with value "EQ" in the "Series" column should be processed. As the first stage, filter out all lines that do not fulfil this criteria.

```
>>SELECT * FROM NSEDATA WHERE SERIES != 'EQ';
>>hive -e "SELECT * FROM NSEDATA WHERE SERIES != 'EQ'" >
>>/home/hduser/hive/savefiles/stockstats/q2_1.csv
```

AICHAMP	BE	24.5	24.5	24.5	24.5	24.5	23.5	1005	24622.5	1-Apr-11
ALICON	BE	53.95	53.95	53.95	53.95	53.95	51.5	1	53.95	1-Apr-11
BHARATFORG	W2	140	140	140	140	140	133.3	100000	14000000	1-Apr-11
BIRLAERIC	BE	17.35	17.95	17.35	17.6	17.5	17.35	651	11500	1-Apr-11
BRITANNIA	N1	165.5	172	165.5	172	172	165.5	2	337.5	1-Apr-11
CEBBCO	BE	41.85	43.5	41.75	42.45	43.5	43.95	89715	3783142	1-Apr-11
DECOLIGHT	BE	10.35	10.35	9.65	10.2	9.7	9.85	4891	49282.85	1-Apr-11
FAME	BE	55.8	58.7	53.1	58.2	58.7	55.9	7657	421357.8	1-Apr-11
FTFTF12BGR	MF	9.01	9.01	9.01	9.01	9.01	9	2000	18020	1-Apr-11
GOKEX	BE	98	102.6	93.4	98.2	100	97.7	1911	188788.3	1-Apr-11
HDFC	W1	150	152	148	151.95	151.8	150.45	49950	7509890	1-Apr-11
IMFA	BE	567.65	588	566.1	588	588	586.7	284	164600.5	1-Apr-11
INDHOTEL	N4	104.94	104.95	104.94	104.95	104.95	104.41	4	419.78	1-Apr-11
ISPATIND	P1	7.55	8.4	7.55	8.05	8	8.3	266635	2179305	1-Apr-11
JAINSTUDIO	BE	10.1	10.15	9.45	9.6	9.55	9.85	6145	59193.9	1-Apr-11
JCTEL	BE	1.4	1.55	1.4	1.45	1.5	1.45	50721	75275.85	1-Apr-11
JEYPORE	BE	142.4	142.4	142.4	142.4	142.4	136.45	1	142.4	1-Apr-11
JYOTISTRUC	N1	114	114	114	114	114	114	40	4560	1-Apr-11
JYOTISTRUC	W1	10	10	9	10	10	9.75	635	6322	1-Apr-11
KALYANIFRG	BE	141	141	141	141	141	141	29	4089	1-Apr-11
KOVAI	BE	120	121.6	119	121.6	121.6	115.85	515	62061	1-Apr-11
L&TFINANCE	N1	1010	1010	1010	1010	1010	1013.6	1	1010	1-Apr-11
L&TFINANCE	N2	995	998.95	995	995	995	993.5	530	529032.6	1-Apr-11
L&TFINANCE	N3	1143	1150	1142.1	1149	1150	1142	180	205896	1-Apr-11
L&TFINANCE	N4	1032.95	1032.95	1020.2	1022.05	1022	1034.49	2872	2940003	1-Apr-11
L&TFINANCE	N6	999	999	969.8	969.8	969.8	965	52	51188.8	1-Apr-11

2. For every stock, for every year, calculate the following statistical parameters and store the generated information in properly designed tables:
    - Min, Max, Mean, Standard Deviation

SYMBOL	MIN	MAX	AVG.	STDDEV	YEAR
20MICRON	27.3	45.7	34.21424	3.807995	2015
20MICRON	30.1	37.55	31.96111	1.887647	2014
20MICRON	29.85	160.8	41.64174	32.16465	2013
20MICRON	61.7	163.9	100.753	26.01453	2012
20MICRON	35.85	69.95	49.14251	8.142063	2011
20MICRON	36.85	60.9	53.64202	5.724678	2010
3IINFOTEC	2.4	9.75	5.952848	2.079894	2015
3IINFOTEC	6.05	13.1	8.42428	1.28447	2014
3IINFOTEC	3.7	9.2	5.814314	1.204313	2013
3IINFOTEC	7.1	20.65	11.7	3.697611	2012
3IINFOTEC	11.65	62.2	37.35304	12.29286	2011
3IINFOTEC	52.35	70.2	61.3649	4.025086	2010
3MINDIA	6294.9	11641.4	8061.372	943.428	2015
3MINDIA	3527.55	6648.65	4866.371	1123.928	2014
3MINDIA	2974.5	4196.9	3487.371	281.7409	2013
3MINDIA	3447.45	4449.65	3859.688	186.9571	2012
3MINDIA	3127.95	4798.15	3921.363	412.142	2011
3MINDIA	3312.6	4224.35	3709.075	188.8821	2010
3RDROCK	151.95	205.5	181.1622	16.17491	2015
3RDROCK	79	158.9	132.4781	26.53929	2014
8KMILES	617.8	929.15	745.8544	83.02142	2015
8KMILES	98.2	775.05	292.8946	212.8573	2014
A2ZINFRA	12.45	31.4	18.61867	4.827566	2015
A2ZINFRA	17.15	17.15	17.15	0	2014
A2ZMES	9.75	35.1	17.91818	6.921921	2014
A2ZMES	7.6	61.35	18.58246	13.1155	2013
A2ZMES	56.1	141.25	90.66174	23.86299	2012

3. Select any year for which data is available. In your report clearly mention the year selected.
    - For that year, create a table that contains data only for those stocks that have an average total traded quantity of 3 Lakhs or more per day. Print out the first 25 entries of the Table and submit.

SELECTING YEAR 2013

3IINFOTEC	EQ	5.65	6.1	5.65	6.1	6.1	5.55	944273	5676839	1-Apr-13	1216	INE748C01020
ADANIENT	EQ	203.95	207.7	201.35	206.8	206.95	202.2	1088434	2.23E+08	1-Apr-13	14930	INE423A01024
ADANIPOR	EQ	140.8	144.2	137.35	142.45	143.25	138.7	890295	1.26E+08	1-Apr-13	10105	INE742F01042
ADANIPOV	EQ	41.7	42.7	40.55	42.2	42.2	40.65	1464569	61242356	1-Apr-13	7750	INE814H01011
AFTEK	EQ	5	5.5	4.75	5.3	5.25	4.9	337076	1764123	1-Apr-13	617	INE796A01023
ALBK	EQ	126.9	130.1	125.1	129.5	129.95	126.4	1315112	1.68E+08	1-Apr-13	17278	INE428A01015
ALOKTEXT	EQ	8.45	9	8.25	8.85	8.95	8.2	5003130	43123372	1-Apr-13	5614	INE270A01011
AMBUJACI	EQ	175	177.5	172.65	174.25	174.5	174.2	1407812	2.46E+08	1-Apr-13	14322	INE079A01024
ANANTRA	EQ	63	65.45	63	64.75	64.5	62.8	633618	40822524	1-Apr-13	5893	INE242C01024
ANDHRAB	EQ	97.65	97.65	95.1	96.3	97	94.65	700706	67294320	1-Apr-13	4646	INE434A01013
APOLLOTY	EQ	84	86.5	82.95	86.1	85.95	83.45	1225750	1.04E+08	1-Apr-13	12024	INE438A01022
ARSHIYA	EQ	18.05	19.95	18.05	18.05	18.05	19	425672	7862557	1-Apr-13	1317	INE968D01022
ARVIND	EQ	77.2	79.85	76.75	79.6	79.8	77.35	561007	44100580	1-Apr-13	5628	INE034A01011
ASHOKLEY	EQ	21.9	22.25	21.9	22.15	22.1	21.85	1420748	31341778	1-Apr-13	5424	INE208A01029
ASSAMCO	EQ	5.25	5.8	5.25	5.75	5.75	5.25	350767	1957600	1-Apr-13	609	INE442A01024
AUROPHA	EQ	158	158	146.85	151.95	151.95	145.95	3197824	4.87E+08	1-Apr-13	29097	INE406A01037
AXISBANK	EQ	1301.25	1318.5	1296	1314.9	1312	1300.7	981343	1.29E+09	1-Apr-13	42161	INE238A01026
BAJAJ-AUT	EQ	1798	1818.25	1770.05	1781.3	1772.55	1799.55	362092	6.52E+08	1-Apr-13	21567	INE917I01010
BAJAJHIND	EQ	18.95	19.8	18.7	19.5	19.6	18.95	1262862	24414454	1-Apr-13	4912	INE306A01021
BALRAMCI	EQ	43.4	45	43.4	44.8	44.8	43.6	494942	22025780	1-Apr-13	3301	INE119A01028
BANKBARO	EQ	679.85	687.2	670.55	684.7	687.2	675.4	303261	2.06E+08	1-Apr-13	7949	INE028A01013
BANKINDI	EQ	305.9	309.9	300.9	308.75	309	302.85	567220	1.73E+08	1-Apr-13	10062	INE084A01016
BEML	EQ	148.5	160	147.1	153.1	152.5	146.95	323265	50090744	1-Apr-13	8128	INE258A01016
BFUTILITIE	EQ	192.7	230.75	192.7	230.75	230.75	192.3	730494	1.58E+08	1-Apr-13	19952	INE243D01012
BHARATFO	EQ	204.05	211.55	202.5	203.85	203.3	207.1	336898	69799104	1-Apr-13	6987	INE465A01025

- From among these, select any 10 stocks from IT (eg. INFY, WIPRO, TCS, GEOMETRIC, etc) and create a Table in Hive containing their data.

0	HCLTECH	EQ	795.05	800	782.25	788.5	788.5	795	405148	3.2E+08	1-Apr-13	15156	INE860A01027
1	INFY	EQ	2910.75	2953.55	2910.75	2944.2	2939.5	2889.35	927755	2.73E+09	1-Apr-13	82593	INE009A01021
2	TCS	EQ	1565	1573.7	1551.25	1556.85	1552.1	1575.75	484406	7.56E+08	1-Apr-13	19638	INE467B01029
3	WIPRO	EQ	437.3	441.75	428.1	431	431.1	437.15	875529	3.8E+08	1-Apr-13	18537	INE075A01022
4	HCLTECH	EQ	951	951.5	926.55	930.7	930	937.55	1590949	1.49E+09	1-Aug-13	61454	INE860A01027
5	INFY	EQ	2988	2995.15	2956.75	2974.65	2969	2969.65	1086329	3.23E+09	1-Aug-13	56185	INE009A01021
6	TCS	EQ	1810.05	1834.8	1790	1815.4	1816	1815.5	1070589	1.95E+09	1-Aug-13	50550	INE467B01029
7	TECHM	EQ	1248.6	1260.8	1230.8	1251.6	1254	1246.3	706940	8.81E+08	1-Aug-13	22716	INE669C01028
8	WIPRO	EQ	438	443.8	427.5	436.4	434	438	1966728	8.6E+08	1-Aug-13	36372	INE075A01022
9	HCLTECH	EQ	689.95	696.9	682.15	687.95	688	688	719189	4.96E+08	1-Feb-13	27225	INE860A01027
10	INFY	EQ	2791.1	2801.9	2752	2770.95	2768.05	2789.5	987456	2.74E+09	1-Feb-13	42209	INE009A01021
11	TCS	EQ	1344	1352	1342.15	1348.1	1347.9	1344.15	742270	1E+09	1-Feb-13	22910	INE467B01029
12	TECHM	EQ	998.95	1015	993	1004.8	1005	998.95	403923	4.06E+08	1-Feb-13	16332	INE669C01028
13	WIPRO	EQ	410	413.45	407.05	407.6	407.15	411.55	1109376	4.55E+08	1-Feb-13	17844	INE075A01022
14	HCLTECH	EQ	624.9	627.9	619.85	622.15	622.7	618.7	305795	1.91E+08	1-Jan-13	6851	INE860A01027
15	INFY	EQ	2327.6	2329.9	2305.55	2308.85	2307.1	2318.7	369999	8.58E+08	1-Jan-13	14479	INE009A01021
16	WIPRO	EQ	396.05	399	391.15	396.9	396.5	394.5	796404	3.15E+08	1-Jan-13	10335	INE075A01022
17	HCLTECH	EQ	779.9	779.9	755.8	758.4	761	776	685642	5.25E+08	1-Jul-13	47385	INE860A01027
18	INFY	EQ	2469	2473.5	2441.1	2451	2443	2498.85	526135	1.29E+09	1-Jul-13	39596	INE009A01021
19	TCS	EQ	1512	1516.7	1483.4	1492.35	1489	1518.15	677326	1.01E+09	1-Jul-13	36062	INE467B01029
20	TECHM	EQ	1056	1069.4	1025.5	1034.45	1039	1059.8	627810	6.55E+08	1-Jul-13	33474	INE669C01028

- Find out the Pearsons Correlation Coefficient for EVERY PAIR of stocks you have selected. Final output should be in decreasing order of the coefficient.

	A	B	C	D
1		SYMBOL1	SYMBOL2	CORR_BW_S1andS2
2	0	NIITTECH	TATAELXSI	0.99999905
3	1	MINDTREE	TECHM	0.9902777
4	2	HCLTECH	TCS	0.97837436
5	3	HCLTECH	MINDTREE	0.9722426
6	4	MINDTREE	WIPRO	0.9544632
7	5	INFY	MINDTREE	0.95037484
8	6	MINDTREE	TCS	0.94593716
9	7	HCLTECH	TECHM	0.945674
10	8	INFY	WIPRO	0.925826

4. Use the correlation information generated in step 3 in the following way:
    - Assume you have Rs 10 Lakh to invest.
    - Assume you have to invest in six stocks on the first working day of January of the **next** year.
    - By using logic / simulation / etc. Identify the stocks that you will invest in, such that at the end of the year:
        - At least your overall capital (Rs 10 Lakh) is protected.

STEPS FOLLOWED:

- SELECTED YEAR 2013 WITH 10 IT STOCKS
- CALCULATED CORRELATION BETWEEN EACH STOCKS WITH CROSS JOIN
- CALCULATED GROWTH PERCENTAGE OF EACH STOCK FROM 2011 TO 2013
- SELECTED STOCKS WITH HIGH GROWTH PERCENTAGE AND RANKED THEM FOR DECISION ON CAPITAL INVESTMENT IN EACH STOCK
- SELECTED STOCKS ARE >> ['MINDTREE', 'TCS', 'INFY', 'OFSS', 'TECHM', 'HCLTECH']
- SELECTED OFSS AS IT HAS NEGATIVE CORRELATION WITH TCS AND INFY AND ASSIGNED MAX CAPITAL ON IT AS CORRELATION IS APROX. 0.5
- AS PER RANKING TOOK DECISION ON HOW MUCH CAPITAL TO BE INVESTED AND

STARTING CAPITAL	1000000

stock	current price (2013)	priority based on %growth	capital assigned	no. of stocks to buy	total value of capital in 2014 january	price in 2014	total value of capital in 2014 december	profit/loss
MINDTREE	1549	1	250000	161	250000	1244	200775	-49225
TCS	2153	4	100000	46	100000	2692	125035	25035
INFY	3468	6	50000	14	50000	4349	62702	12702
OFSS	3274	5	200000	61	200000	3444	210385	10385
TECHM	1828	3	150000	82	150000	2653	217697	67697
HCLTECH	1258	2	250000	199	250000	1671	332075	82075
			1000000		1000000		1148668	148668

% Profit	15