I used Hive to get the desired result

Find out the frequency of books published each year

```
hive> create database bookcross;
OK
Time taken: 1.11 seconds
hive> show databases;
OK
book
bookcross
default
Time taken: 0.143 seconds, Fetched: 3 row(s)
hive> use bookcross;
OK
Time taken: 0.037 seconds
hive> create table books(ISBN string,BookTitle string,BookAuthor string,YrOfPub int,Publisher string)
row format delimited fields terminated by '\;' stored as textfile;
OK
Time taken: 0.6 seconds
hive> describe books;
OK
isbn            string
booktitle       string
bookauthor      string
yrofpub         int
publisher       string
Time taken: 0.793 seconds, Fetched: 5 row(s)
hive> load data local inpath '/home/edureka/Muthu/BX-Books-test.csv' into table books;
Copying data from file:/home/edureka/Muthu/BX-Books-test.csv
Copying file: file:/home/edureka/Muthu/BX-Books-test.csv
Loading data to table bookcross.books
Table bookcross.books stats: [numFiles=1, numRows=0, totalSize=73443360, rawDataSize=0]
OK
Time taken: 4.433 seconds
hive> select yrofpub,count(booktitle) from books group by yrofpub;
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
```

Starting Job = job_1431312707020_0001, Tracking URL =
http://localhost:8088/proxy/application_1431312707020_0001/
Kill Command = /usr/lib/hadoop-2.2.0/bin/hadoop job -kill job_1431312707020_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2015-05-10 23:46:45,316 Stage-1 map = 0%, reduce = 0%
2015-05-10 23:46:58,718 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.57 sec
2015-05-10 23:47:08,569 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.98 sec
MapReduce Total cumulative CPU time: 3 seconds 980 msec
Ended Job = job_1431312707020_0001
MapReduce Jobs Launched:
Job 0: Map: 1 Reduce: 1 Cumulative CPU: 3.98 sec HDFS Read: 73443592 HDFS Write: 966 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 980 msec
OK

NULL	6056
0	4589
1376	1
1378	1
1806	1
1897	1
1900	3
1901	7
1902	2
1904	1
1906	1
1908	1
1909	2
1910	1
1911	19
1914	1
1917	1
1919	1
1920	33
1921	2
1922	2
1923	11
1924	2
1925	2
1926	2
1927	1
1928	2
1929	7
1930	13
1931	3
1932	5

1933	4
1934	1
1935	3
1936	7
1937	5
1938	7
1939	9
1940	35
1941	10
1942	12
1943	8
1944	4
1945	8
1946	13
1947	14
1948	8
1949	11
1950	31
1951	40
1952	33
1953	63
1954	54
1955	69
1956	74
1957	75
1958	77
1959	102
1960	129
1961	130
1962	121
1963	129
1964	148
1965	170
1966	182
1967	170
1968	226
1969	317
1970	426
1971	492
1972	699
1973	807
1974	929
1975	1147
1976	1521

1977	1833
1978	2086
1979	2153
1980	2626
1981	3224
1982	4124
1983	4422
1984	4900
1985	5256
1986	5765
1987	6441
1988	7391
1989	7820
1990	8560
1991	9250
1992	9708
1993	10391
1994	11498
1995	13192
1996	13626
1997	14466
1998	15367
1999	16980
2000	16863
2001	16976
2002	17298
2003	14054
2004	5745
2005	45
2006	3
2010	2
2011	2
2012	1
2020	3
2021	1
2024	1
2026	1
2030	7
2037	1
2038	1
2050	2

Time taken: 43.834 seconds, Fetched: 116 row(s)

Find out in which year maximum number of books were published

```
hive> select yrofpub,count(booktitle) as maxcnt from books group by yrofpub sort by maxcnt desc limit 1 ;
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1431312707020_0002, Tracking URL =
http://localhost:8088/proxy/application_1431312707020_0002/
Kill Command = /usr/lib/hadoop-2.2.0/bin/hadoop job  -kill job_1431312707020_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2015-05-10 23:59:52,557 Stage-1 map = 0%,  reduce = 0%
2015-05-11 00:00:02,196 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.49 sec
2015-05-11 00:00:11,813 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.8 sec
MapReduce Total cumulative CPU time: 3 seconds 800 msec
Ended Job = job_1431312707020_0002
Launching Job 2 out of 3
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1431312707020_0003, Tracking URL =
http://localhost:8088/proxy/application_1431312707020_0003/
Kill Command = /usr/lib/hadoop-2.2.0/bin/hadoop job  -kill job_1431312707020_0003
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2015-05-11 00:00:28,445 Stage-2 map = 0%,  reduce = 0%
2015-05-11 00:00:36,950 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.91 sec
2015-05-11 00:00:45,600 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.03 sec
MapReduce Total cumulative CPU time: 2 seconds 30 msec
Ended Job = job_1431312707020_0003
Launching Job 3 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
```

set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
set mapreduce.job.reduces=<number>
Starting Job = job_1431312707020_0004, Tracking URL =
http://localhost:8088/proxy/application_1431312707020_0004/
Kill Command = /usr/lib/hadoop-2.2.0/bin/hadoop job -kill job_1431312707020_0004
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2015-05-11 00:01:03,554 Stage-3 map = 0%, reduce = 0%
2015-05-11 00:01:11,343 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 0.96 sec
2015-05-11 00:01:21,042 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 2.2 sec
MapReduce Total cumulative CPU time: 2 seconds 200 msec
Ended Job = job_1431312707020_0004
MapReduce Jobs Launched:
Job 0: Map: 1 Reduce: 1 Cumulative CPU: 3.8 sec HDFS Read: 73443592 HDFS Write: 2631 SUCCESS
Job 1: Map: 1 Reduce: 1 Cumulative CPU: 2.03 sec HDFS Read: 2997 HDFS Write: 119 SUCCESS
Job 2: Map: 1 Reduce: 1 Cumulative CPU: 2.2 sec HDFS Read: 485 HDFS Write: 11 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 30 msec
OK
2002 17298
Time taken: 100.376 seconds, Fetched: 1 row(s)

Find out how many book were published based on ranking in the year 2002

hive> create table Rating(userid int,ISBN string,bookrating int) row format delimited fields terminated by '\;' stored as textfile;
OK
Time taken: 0.083 seconds
hive> load data local inpath '/home/edureka/Muthu/BX-Book-Rating.csv' into table Rating;
Copying data from file:/home/edureka/Muthu/BX-Book-Rating.csv
Copying file: file:/home/edureka/Muthu/BX-Book-Rating.csv
Loading data to table bookcross.rating
Table bookcross.rating stats: [numFiles=1, numRows=0, totalSize=23783540, rawDataSize=0]
OK
Time taken: 0.863 seconds
hive> select r.bookrating,count(b.booktitle) from books b join rating r on (b.isbn = r.isbn) where b.yrofpub = '2002' group by r.bookrating;
Total jobs = 3
Stage-7 is selected by condition resolver.
Stage-1 is filtered out by condition resolver.
15/05/11 00:48:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
15/05/11 00:48:08 WARN conf.Configuration: file:/tmp/edureka/hive_2015-05-11_00-48-02_760_2245939365732268143-1/-local-10007/jobconf.xml:an attempt to override final parameter: mapreduce.job.end-notification.max.retry.interval; Ignoring.

15/05/11 00:48:09 WARN conf.Configuration: file:/tmp/edureka/hive_2015-05-11_00-48-02_760_2245939365732268143-1/-local-10007/jobconf.xml:an attempt to override final parameter: mapreduce.job.end-notification.max.attempts; Ignoring.
15/05/11 00:48:09 INFO Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
15/05/11 00:48:09 INFO Configuration.deprecation: mapred.min.split.size is deprecated. Instead, use mapreduce.input.fileinputformat.split.minsize
15/05/11 00:48:09 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
15/05/11 00:48:09 INFO Configuration.deprecation: mapred.min.split.size.per.node is deprecated. Instead, use mapreduce.input.fileinputformat.split.minsize.per.node
15/05/11 00:48:09 INFO Configuration.deprecation: mapred.input.dir.recursive is deprecated. Instead, use mapreduce.input.fileinputformat.input.dir.recursive
15/05/11 00:48:09 INFO Configuration.deprecation: mapred.min.split.size.per.rack is deprecated. Instead, use mapreduce.input.fileinputformat.split.minsize.per.rack
15/05/11 00:48:09 INFO Configuration.deprecation: mapred.max.split.size is deprecated. Instead, use mapreduce.input.fileinputformat.split.maxsize
15/05/11 00:48:09 INFO Configuration.deprecation: mapred.committer.job.setup.cleanup.needed is deprecated. Instead, use mapreduce.job.committer.setup.cleanup.needed
Execution log at: /tmp/edureka/edureka_20150511004848_4af7a9ae-10df-42f1-84cb-5b5d90d6a1c0.log
2015-05-11 12:48:10	Starting to launch local task to process map join;	maximum memory = 518979584
2015-05-11 12:48:14	Processing rows:	200000 Hashtable size:	199999 Memory usage:	94604064	percentage:	0.182
2015-05-11 12:48:15	Processing rows:	300000 Hashtable size:	299999 Memory usage:	142537600	percentage:	0.275
2015-05-11 12:48:16	Dump the side-table into file: file:/tmp/edureka/hive_2015-05-11_00-48-02_760_2245939365732268143-1/-local-10004/HashTable-Stage-4/MapJoin-mapfile61--.hashtable
2015-05-11 12:48:17	Uploaded 1 File to: file:/tmp/edureka/hive_2015-05-11_00-48-02_760_2245939365732268143-1/-local-10004/HashTable-Stage-4/MapJoin-mapfile61--.hashtable (15147688 bytes)
2015-05-11 12:48:17	End of local task; Time Taken: 7.041 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 2 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1431312707020_0011, Tracking URL = http://localhost:8088/proxy/application_1431312707020_0011/
Kill Command = /usr/lib/hadoop-2.2.0/bin/hadoop job -kill job_1431312707020_0011
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 0
2015-05-11 00:48:29,708 Stage-4 map = 0%, reduce = 0%
2015-05-11 00:48:43,645 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 6.15 sec
MapReduce Total cumulative CPU time: 6 seconds 150 msec

```
Ended Job = job_1431312707020_0011
Launching Job 3 out of 3
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1431312707020_0012, Tracking URL =
http://localhost:8088/proxy/application_1431312707020_0012/
Kill Command = /usr/lib/hadoop-2.2.0/bin/hadoop job  -kill job_1431312707020_0012
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2015-05-11 00:48:59,745 Stage-2 map = 0%,  reduce = 0%
2015-05-11 00:49:08,287 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.96 sec
2015-05-11 00:49:17,795 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.19 sec
MapReduce Total cumulative CPU time: 2 seconds 190 msec
Ended Job = job_1431312707020_0012
MapReduce Jobs Launched:
Job 0: Map: 1   Cumulative CPU: 6.15 sec   HDFS Read: 73443592 HDFS Write: 326 SUCCESS
Job 1: Map: 1  Reduce: 1   Cumulative CPU: 2.19 sec   HDFS Read: 692 HDFS Write: 75 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 340 msec
OK
0       53124
1       142
2       257
3       531
4       853
5       3568
6       3147
7       6569
8       9761
9       6502
10      6189
Time taken: 76.178 seconds, Fetched: 11 row(s)
hive>
```