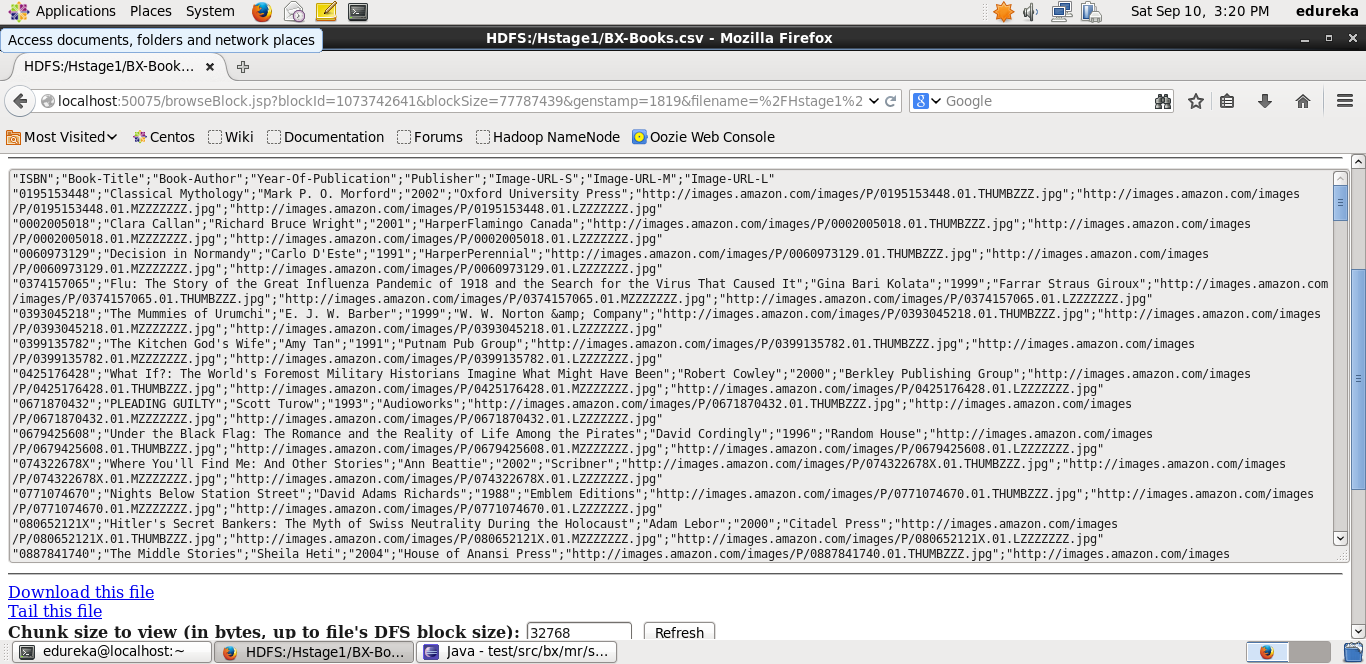
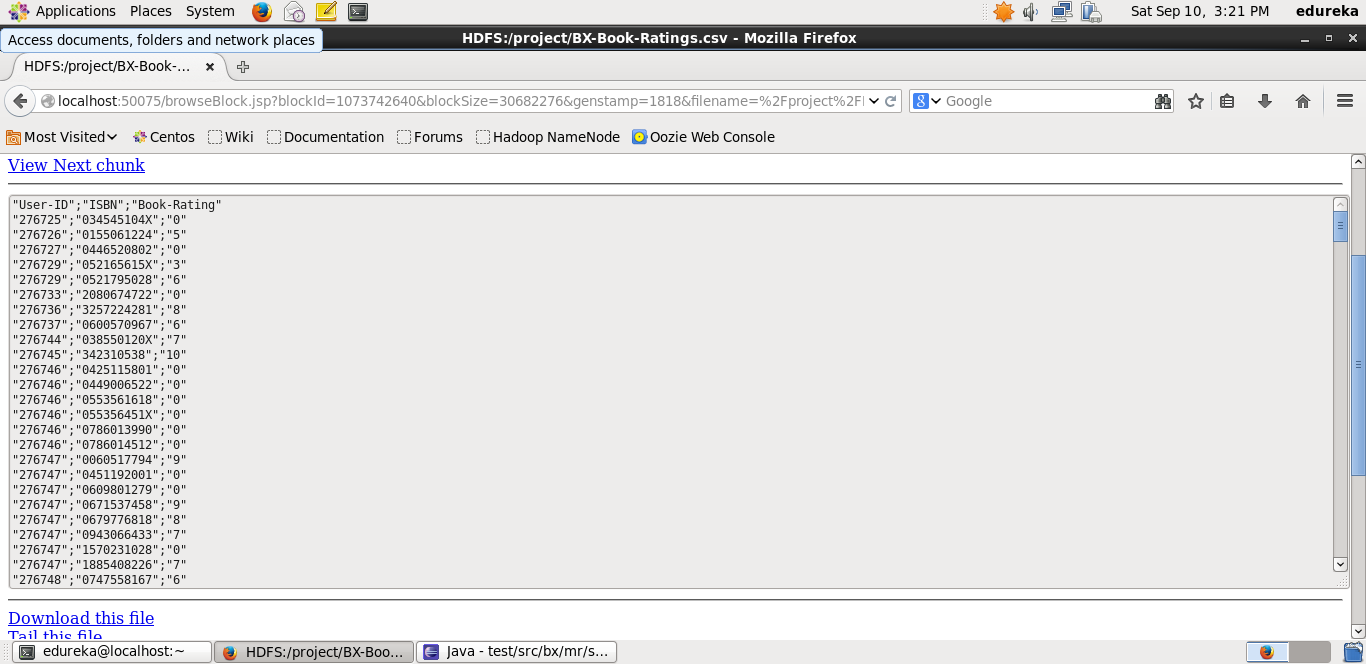
**Book Project**

**Input Data in HDFS**

**BX-Books:-**



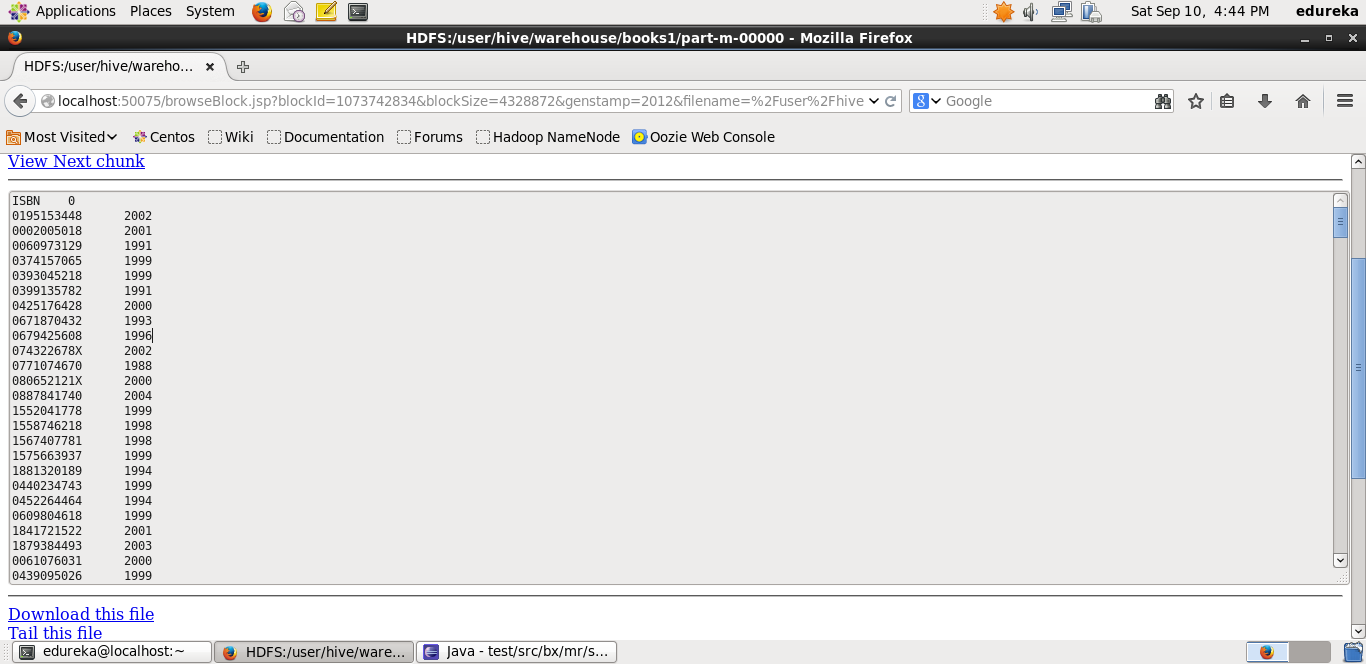
**BX-Book Ratings**



**Step1:** MapReduce program to extract ISBN and Year of Publication from BX-Books



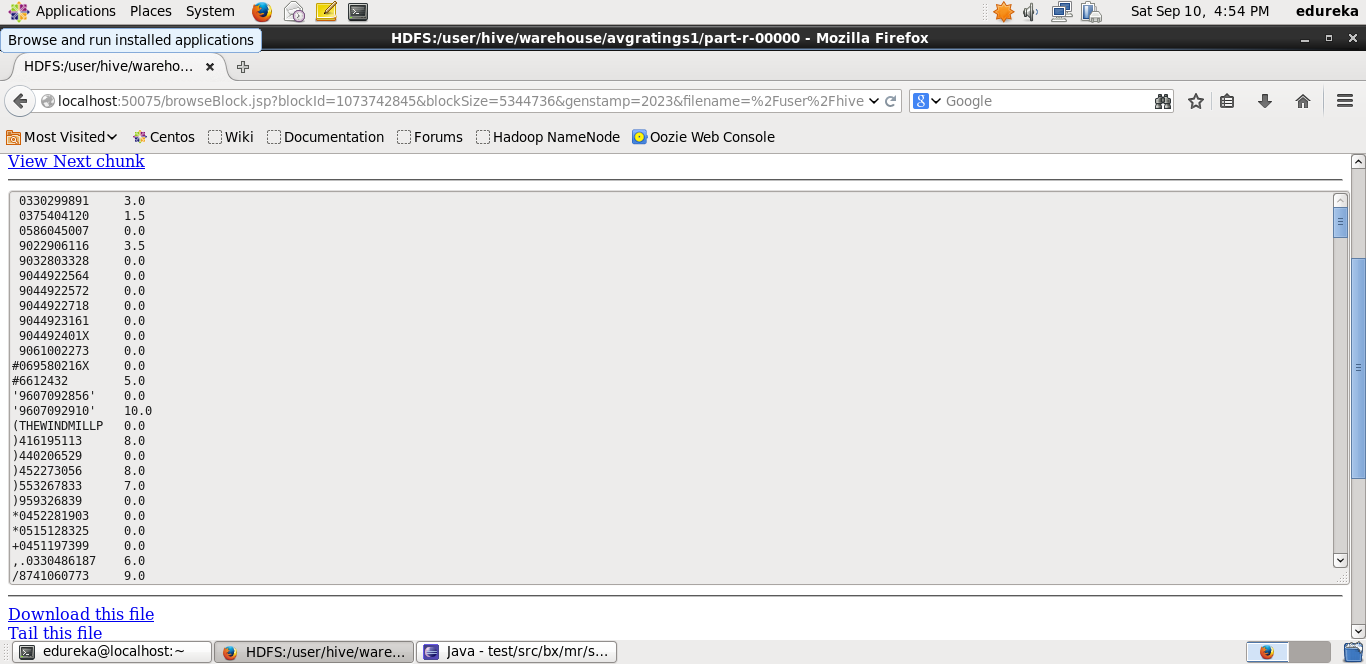
**Output:**

****

**Step2:**MapReduce program to extract ISBN and Average Rating from BX-Book Rating



**Output:**

****

**Step 3:** Pig commands to find frequency of books published each year:-

grunt > A = load ‘/projectResults/MR13/part-m-00000’ as (book:chararray, year:int);

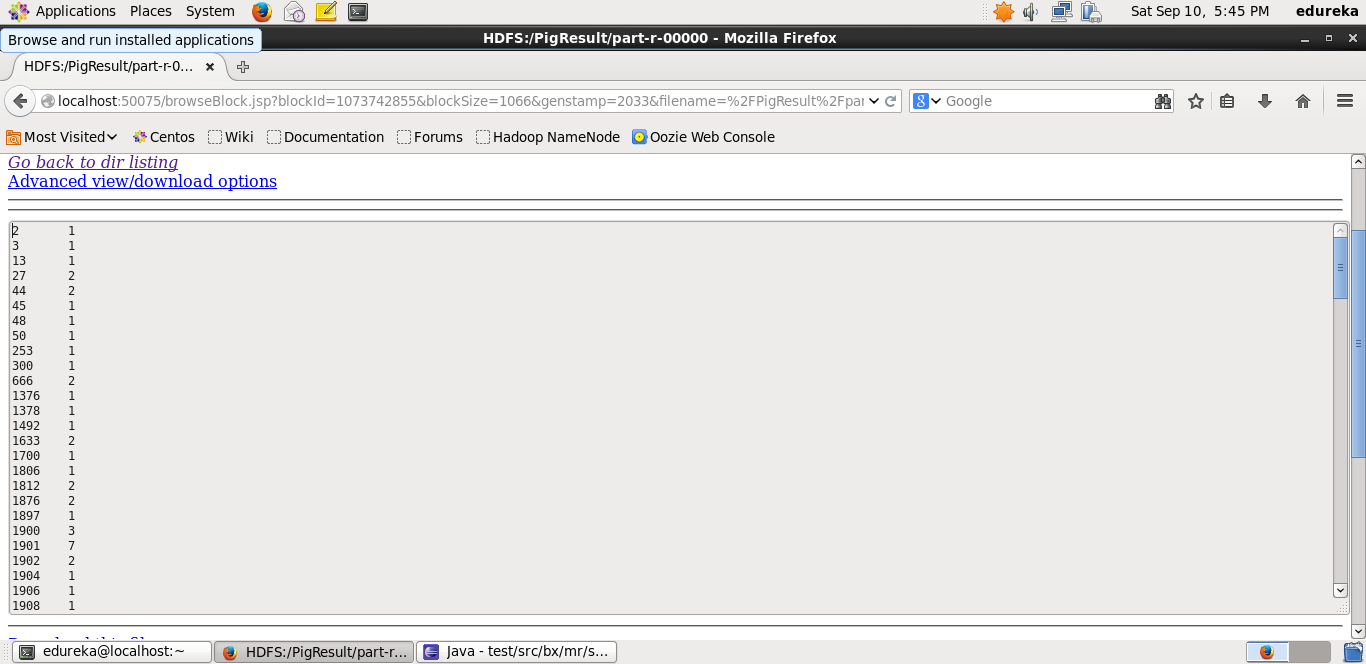
grunt > B= filter A by (not year==0);

grunt > C= group B by year;

grunt > D= foreach C generate group,COUNT(B.book);

grunt > store D into “/PigResult’;

**Output:**

****

**Step 4:**Pig commands to find year in which maximum number of books were published

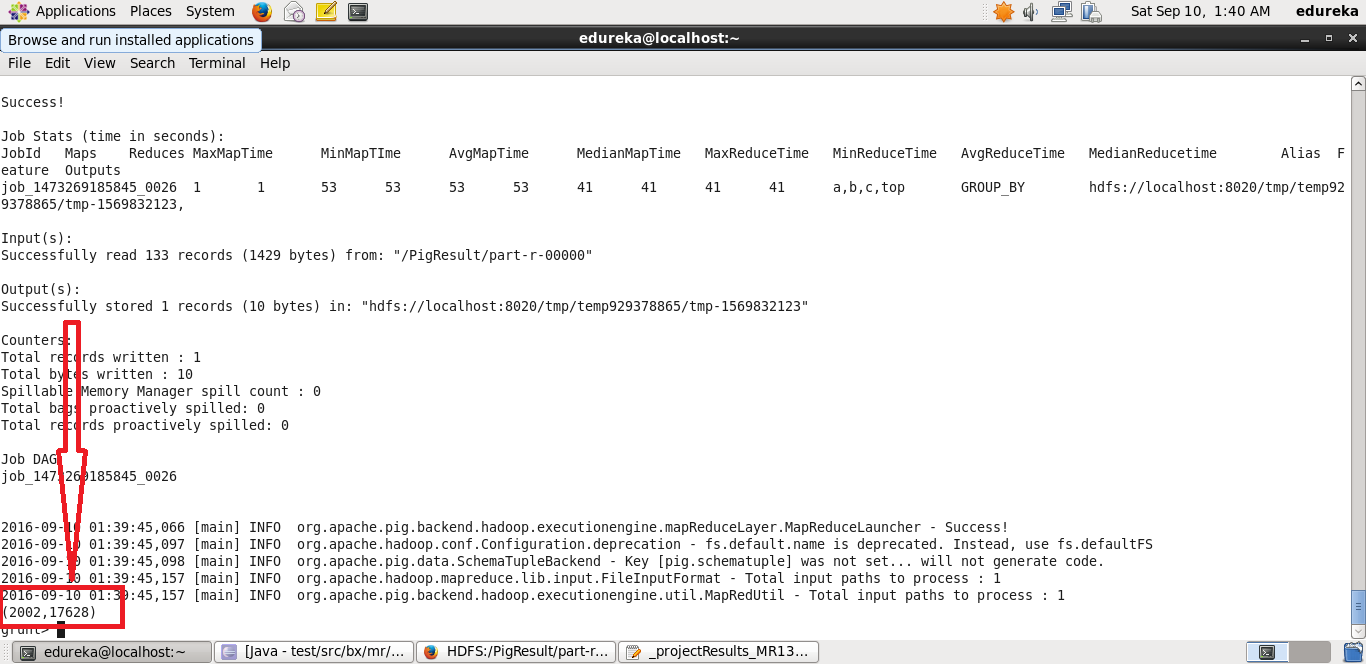
grunt> A= load ‘/PigResult/part-r-00000’ as (year:int, no\_books:int);

grunt> B= order A by no\_books DESC;

grunt> C= limit B 1;

grunt> dump C;

**Output:**

****

**Step 5:** Hive commands to find the number of books according to ranking published in the year 2002

hive > create table books1 (isbn string, year int) row format delimited fields terminated by ‘\t’;

>load data inpath ‘/hiveOut/books/part-m-00000’ into table books1;

-----------------------------------------------------------------------------------------------------------------

hive> create table avgRatings1 (isbn string, rating float) row format delimited fields terminated by”\t’;

>load data inpath ‘/hiveOut/ratings/part-r-00000’ into table avgRatings1;

-----------------------------------------------------------------------------------------------------------------------

hive> create table result (isbn string, rating float, year int)

>row format delimited fields terminated by ‘\t’;

>insert overwrite table result

>select a.isbn,a.rating,b.year fromavgRatings1 a join books1 b on a.isbn=b.isbn;

-----------------------------------------------------------------------------------------------------------------

hive> create table result1 (rating float, bookcount int)

> row format delimited fields terminated by ‘\t’;

>from result r

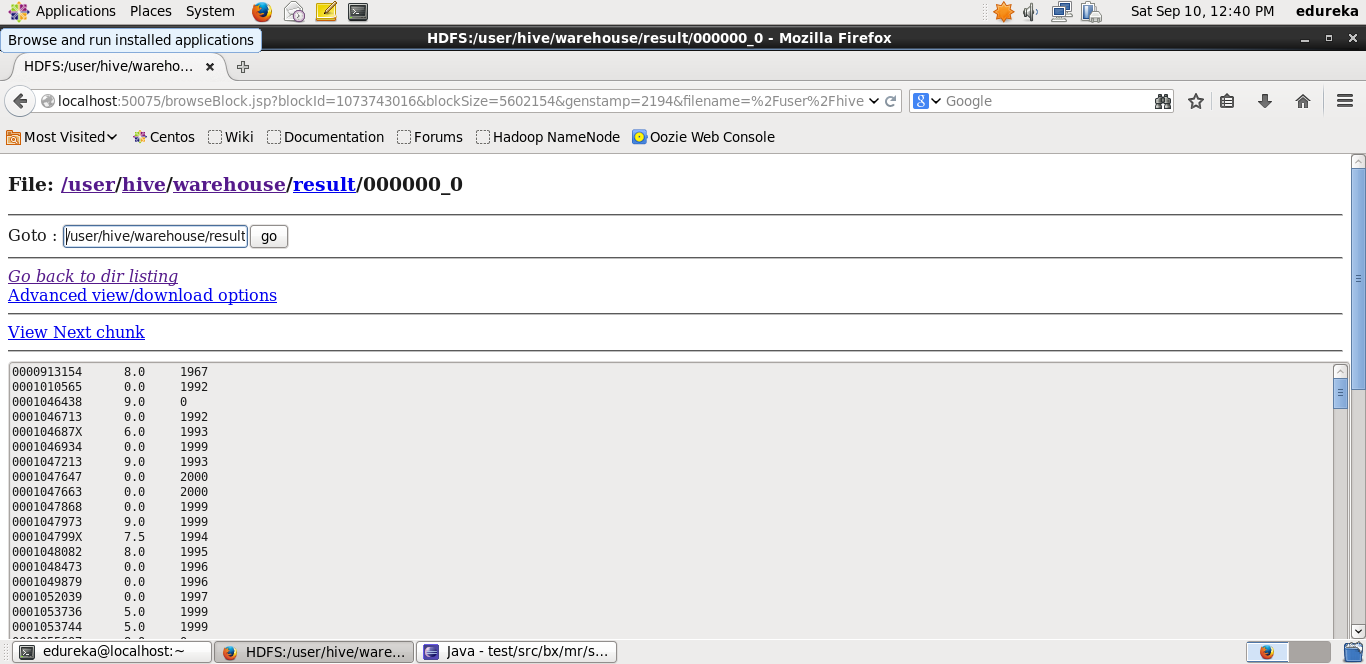
>insert overwrite table result1

>select r.rating,COUNT(r.isbn) where year= 2002

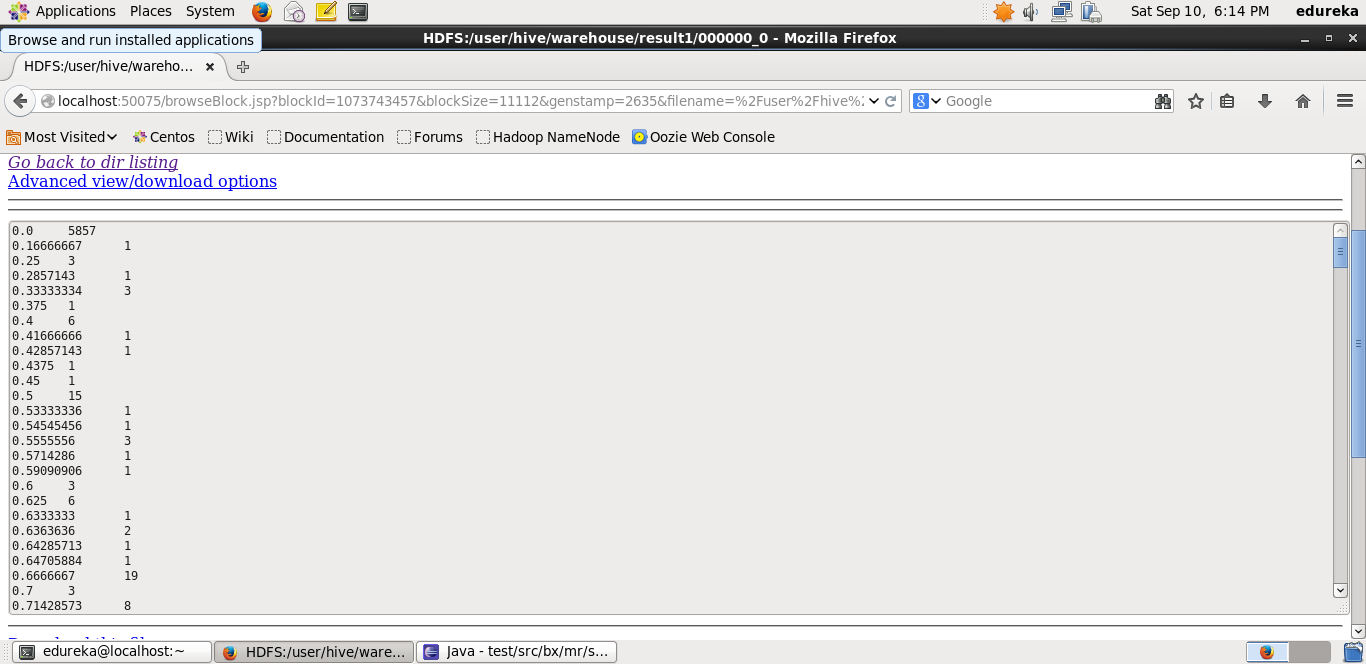
>group by r.rating;

**Output:**

**Join operation:**

****

**Query result:**

****