MAPSIDE JOIN:

```
hive> create database hive_advance;
OK
Time taken: 0.108 seconds
hive> use hive_advance;
OK
Time taken: 0.092 seconds
hive> create external table orders(order_id int,
    > order_date string,
    > order_customer_id int,
    > order_status string)
    > row format delimited fields terminated by ','
    > stored as textfile
    > location '/user/hduser/orders';
OK
Time taken: 0.398 seconds
hive> show tables;
OK
orders
Time taken: 0.929 seconds, Fetched: 1 row(s)
hive> load data local inpath '/home/hduser/orders.csv' into table orders;
Loading data to table hive_advance.orders
Table hive_advance.orders stats: [numFiles=1, totalSize=184]
OK
Time taken: 2.946 seconds
```

orders.csv

```
70001,2012-10-05,3005,delivered
70009,2012-09-10,3001,shipped
70002,2012-10-05,3002,shipped
70004,2012-08-17,3009,packed
70007,2012-09-10,3005,delivered
70005,2012-07-27,3007,packed
```

Customers.csv

```
3002,Nick,Rimando,NewYork,5001
3007,Brad,Davis,NewYork,5001
3005,Graham,Zusi,California,5002
3008,Julian,Green,London,5002
3004,Fabian,Johnson,Paris,5006
3009,Geoff,Cameron,Berlin,5003
```

```
hive> create external table customers(
    > customer_id int,
    > customer_fname string,
    > customer_lname string,
    > customer_city string,
    > customer_zipcode string)
    > row format delimited fields terminated by ','
    > stored as textfile
    > location '/user/hduser/customers'
    > ;
OK
Time taken: 0.4 seconds
hive> load data local inpath '/home/hduser/customers.csv' into table customers;
Loading data to table hive_advance.customers
Table hive_advance.customers stats: [numFiles=1, totalSize=185]
OK
Time taken: 1.231 seconds
```

```
hive> set hive.auto.convert.join=false;
hive> select c.customer_id, c.customer_fname, c.customer_lname ,
    > o.order_id, o.order_date from orders o JOIN customers c ON
    > (o.order_customer_id=c.customer_id);
```

```
Total MapReduce CPU Time Spent: 16 seconds 930 msec
OK
3002    Nick    Rimando 70002   2012-10-05
3005    Graham  Zusi    70007   2012-09-10
3005    Graham  Zusi    70001   2012-10-05
3007    Brad    Davis   70005   2012-07-27
3009    Geoff   Cameron 70004   2012-08-17
Time taken: 178.546 seconds, Fetched: 5 row(s)
hive>
```

```
hive> set hive.auto.convert.join=true;
hive> select c.customer_id, c.customer_fname, c.customer_lname ,
    > o.order_id, o.order_date from orders o JOIN customers c ON
    > (o.order_customer_id=c.customer_id);
```

```
Total MapReduce CPU Time Spent: 3 seconds 100 msec
OK
3002    Nick    Rimando 70002   2012-10-05
3007    Brad    Davis   70005   2012-07-27
3005    Graham  Zusi    70001   2012-10-05
3005    Graham  Zusi    70007   2012-09-10
3009    Geoff   Cameron 70004   2012-08-17
Time taken: 95.677 seconds, Fetched: 5 row(s)
hive>
```

specify using some hints:

→Execute inner join as map side join using hints indicating left table is small.

```
hive> set hive.auto.convert.join=false;
hive> set hive.ignore.mapjoin.hint;
hive.ignore.mapjoin.hint=true
hive> set hive.ignore.mapjoin.hint=false;
hive> select /*+ MAPJOIN(o) */ c.customer_id, c.customer_fname, c.customer_lname ,
    > o.order_id, o.order_date from orders o JOIN customers c ON
    > (o.order_customer_id=c.customer_id);
```

```
Total MapReduce CPU Time Spent: 4 seconds 130 msec
OK
3002    Nick    Rimando 70002   2012-10-05
3007    Brad    Davis   70005   2012-07-27
3005    Graham  Zusi    70001   2012-10-05
3005    Graham  Zusi    70007   2012-09-10
3009    Geoff   Cameron 70004   2012-08-17
Time taken: 96.631 seconds, Fetched: 5 row(s)
```

→ Execute inner join as map side join using hints indicating right table is small.

```
hive> select /*+ MAPJOIN(c) */ c.customer_id, c.customer_fname, c.customer_lname ,
    > o.order_id, o.order_date from orders o JOIN customers c ON
    > (o.order_customer_id=c.customer_id);
Query ID = hduser_20220711182320_62e164e1-30a6-4391-840a-d08b0791cba4
Total jobs = 1
```

```
Total MapReduce CPU Time Spent: 3 seconds 200 msec
OK
3005    Graham  Zusi    70001   2012-10-05
3002    Nick    Rimando 70002   2012-10-05
3009    Geoff   Cameron 70004   2012-08-17
3005    Graham  Zusi    70007   2012-09-10
3007    Brad    Davis   70005   2012-07-27
Time taken: 108.14 seconds, Fetched: 5 row(s)
hive>
    > select /*+ MAPJOIN(c) */ c.customer_id, c.customer_fname, c.customer_lname ,
    > o.order_id, o.order_date from orders o LEFT OUTER JOIN customers c ON
    > (o.order_customer_id=c.customer_id);
```

can a left outer join be treated as map-side join when right table is small ?

-->> Yes.

```
Total MapReduce CPU Time Spent: 2 seconds 620 msec
OK
3005    Graham  Zusi    70001   2012-10-05
NULL    NULL    NULL    70009   2012-09-10
3002    Nick    Rimando 70002   2012-10-05
3009    Geoff   Cameron 70004   2012-08-17
3005    Graham  Zusi    70007   2012-09-10
3007    Brad    Davis   70005   2012-07-27
```

→can a right outer join be treated as map-side join when left table is small ?

→ Yes

```
hive> select /*+ MAPJOIN(o) */ c.customer_id, c.customer_fname, c.customer_lname ,
    > o.order_id, o.order_date from orders o RIGHT OUTER JOIN customers c ON
    > (o.order_customer_id=c.customer_id);
```

```
Total MapReduce CPU Time Spent: 2 seconds 310 msec
OK
3002    Nick    Rimando 70002   2012-10-05
3007    Brad    Davis   70005   2012-07-27
3005    Graham  Zusi    70001   2012-10-05
3005    Graham  Zusi    70007   2012-09-10
3008    Julian  Green   NULL    NULL
3004    Fabian  Johnson NULL    NULL
3009    Geoff   Cameron 70004   2012-08-17
Time taken: 68.432 seconds, Fetched: 7 row(s)
hive>
```

BUCKETMAP JOIN:

```
hive> create external table orders_bucketed_new(
    > order_id int,
    > order_date string,
    > order_customer_id int,
    > order_status string)
    > clustered by(order_customer_id) into 4 buckets
    > row format delimited fields terminated by ',';
OK
Time taken: 0.737 seconds
hive> insert into orders_bucketed_new select * from orders;
```

```
[hduser@localhost ~]$ hdfs dfs -ls /user/hive/warehouse/hive_advance.db/orders_bucketed_new
22/07/11 19:11:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform..
. using builtin-java classes where applicable
Found 5 items
-rw-r--r--   1 hduser supergroup         24 2022-07-11 19:10 /user/hive/warehouse/hive_advance.db/orde
rs_bucketed_new/000000_0
-rw-r--r--   1 hduser supergroup        123 2022-07-11 19:10 /user/hive/warehouse/hive_advance.db/orde
rs_bucketed_new/000001_0
-rw-r--r--   1 hduser supergroup         30 2022-07-11 19:10 /user/hive/warehouse/hive_advance.db/orde
rs_bucketed_new/000002_0
-rw-r--r--   1 hduser supergroup         29 2022-07-11 19:10 /user/hive/warehouse/hive_advance.db/orde
rs_bucketed_new/000003_0
-rw-r--r--   1 hduser supergroup        184 2022-07-11 19:04 /user/hive/warehouse/hive_advance.db/orde
rs_bucketed_new/orders.csv
[hduser@localhost ~]$
```

```
hive> create external table customers_bucketed_new(
    > customer_id int,
    > customer_fname string,
    > customer_lname string,
    > customer_city string,
    >  customer_zipcode string)
    > clustered by(customer_id) into 2 buckets
    > row format delimited fields terminated by ',';
OK
Time taken: 0.337 seconds
hive> insert into customers_bucketed_new select * from customers;
```

```
[hduser@localhost ~]$ hdfs dfs -ls /user/hive/warehouse/hive_advance.db/customers_bucketed_new
22/07/11 19:20:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform..
. using builtin-java classes where applicable
Found 3 items
-rw-r--r--   1 hduser supergroup         92 2022-07-11 19:17 /user/hive/warehouse/hive_advance.db/cust
omers_bucketed_new/000000_0
-rw-r--r--   1 hduser supergroup         93 2022-07-11 19:17 /user/hive/warehouse/hive_advance.db/cust
omers_bucketed_new/000001_0
-rw-r--r--   1 hduser supergroup        185 2022-07-11 18:57 /user/hive/warehouse/hive_advance.db/cust
omers_bucketed_new/customers.csv
```

```
hive> set hive.optimize.bucketmapjoin;
hive.optimize.bucketmapjoin=false
hive> set hive.optimize.bucketmapjoin=true;
hive> set hive.auto.convert.join=true;
hive> select c.customer_id, c.customer_fname, c.customer_lname ,
    > o.order_id, o.order_date from customers_bucketed_new c JOIN orders_bucketed_new o ON
    > (o.order_customer_id=c.customer_id);
```

SMB (Sort Merge Bucket Join)

```
hive> drop table orders_bucketed_new;
OK
Time taken: 0.522 seconds
hive> drop table customers_bucketed_new;
OK
Time taken: 0.418 seconds
hive>
```

```
[hduser@localhost ~]$ hdfs dfs -rm -R /user/hive/warehouse/hive_advance.db/orders_bucketed_new
22/07/11 19:27:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform..
. using builtin-java classes where applicable
22/07/11 19:27:15 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minu
tes, Emptier interval = 0 minutes.
Deleted /user/hive/warehouse/hive_advance.db/orders_bucketed_new
[hduser@localhost ~]$ hdfs dfs -rm -R /user/hive/warehouse/hive_advance.db/customers_bucketed_new
22/07/11 19:27:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform..
. using builtin-java classes where applicable
22/07/11 19:27:43 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minu
tes, Emptier interval = 0 minutes.
Deleted /user/hive/warehouse/hive_advance.db/customers_bucketed_new
[hduser@localhost ~]$
```

```
hive> create external table orders_bucketed_new(
    > order_id int,
    > order_date string,
    > order_customer_id int,
    > order_status string)
    > clustered by(order_customer_id) sorted by(order_customer_id) into 3 buckets
    > row format delimited fields terminated by ',';
OK
Time taken: 0.68 seconds
```

```
hive> create external table customers_bucketed_new(
    > customer_id int,
    > customer_fname string,
    > customer_lname string,
    > customer_city string,
    > customer_zipcode string)
    > clustered by(customer_id) sorted by(customer_id) into 3 buckets
    > row format delimited fields terminated by ',';
OK
Time taken: 0.622 seconds
hive> insert into customers_bucketed_new select * from customers;
```

```
hive> insert into orders_bucketed_new select * from orders;
```

```
[hduser@localhost ~]$ hdfs dfs -cat /user/hive/warehouse/hive_advance.db/orders_bucketed_new/* | wc -l
22/07/11 19:47:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
8
[hduser@localhost ~]$ hdfs dfs -ls /user/hive/warehouse/hive_advance.db/orders_bucketed_new
22/07/11 19:48:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
Found 3 items
-rw-r--r--   1 hduser supergroup         53 2022-07-11 19:43 /user/hive/warehouse/hive_advance.db/orders_bucketed_new/00000
0_0
-rw-r--r--   1 hduser supergroup         59 2022-07-11 19:43 /user/hive/warehouse/hive_advance.db/orders_bucketed_new/00000
1_0
-rw-r--r--   1 hduser supergroup         94 2022-07-11 19:44 /user/hive/warehouse/hive_advance.db/orders_bucketed_new/00000
2_0
```

```
hive> select c.customer_id, c.customer_fname, c.customer_lname ,
    > o.order_id, o.order_date from customers_bucketed_new c JOIN orders_bucketed_new o ON
    > (o.order_customer_id=c.customer_id);
```

```
Total MapReduce CPU Time Spent: 11 seconds 520 msec
OK
3009    Geoff   Cameron 70004   2012-08-17
3007    Brad    Davis   70005   2012-07-27
3002    Nick    Rimando 70002   2012-10-05
3005    Graham  Zusi    70007   2012-09-10
3005    Graham  Zusi    70001   2012-10-05
Time taken: 74.015 seconds, Fetched: 5 row(s)
hive>
```