

Schema of customers and orders table:

```
hive> use demo_db;
OK
Time taken: 0.065 seconds
hive> desc orders;
OK
id                bigint
product_id        string
customer_id        int
quantity           int
amount             double
Time taken: 0.346 seconds, Fetched: 5 row(s)
hive> desc customers;
OK
id                bigint
name              string
address           string
Time taken: 0.318 seconds, Fetched: 3 row(s)
```

Setting property:

```
hive> set hive.auto.convert.join=false;
hive> set hive.auto.convert.join;
hive.auto.convert.join=false
hive>
```

join

```
hive> select c.id, c.name, c.address ,
> o.id, o.product_id from orders o JOIN customers c ON
> (o.customer_id=c.id);
Query ID = hduser_20220714193143_525835b6-6857-49d8-b4da-1cb81f5d0c80
Total jobs = 1
```

Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1

```
OK
1      Deepti  Hyderabad      1457      Desktop1
1      Deepti  Hyderabad      1232      Phone1
1      Deepti  Hyderabad      1457      Desktop
1      Deepti  Hyderabad      1232      Phone
2      Alekhya Nizamabad  1234      Watch1
2      Alekhya Nizamabad  1234      Watch
3      Likhith Pune      3412      Camera1
3      Likhith Pune      3412      Camera
4      Keerthana Vizag      3468      Laptop1
4      Keerthana Vizag      3468      Laptop
Time taken: 120.014 seconds, Fetched: 10 row(s)
```

Mapside join

```
hive> set hive.auto.convert.join=true;
hive> set hive.auto.convert.join;
hive.auto.convert.join=true
hive> select c.id, c.name, c.address ,
> o.id, o.product_id from orders o JOIN customers c ON
> (o.customer_id=c.id);
Query ID = hduser_20220714193906_0d34b660-0212-495c-a51e-a0602a3a983f
Total jobs = 1
```

Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0

```

1      Deepti Hyderabad      1232      Phone
3      Likhith Pune          3412      Camera
4      Keerthana Vizag       3468      Laptop
2      Alekhya Nizamabad     1234      Watch
1      Deepti Hyderabad     1457      Desktop
1      Deepti Hyderabad     1232      Phone1
3      Likhith Pune          3412      Camera1
4      Keerthana Vizag       3468      Laptop1
2      Alekhya Nizamabad     1234      Watch1
1      Deepti Hyderabad     1457      Desktop1
Time taken: 73.188 seconds, Fetched: 10 row(s)

```

```

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;

```

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

```

hive> select /*+ MAPJOIN(o) */ c.id, c.name,
> o.id, o.product_id from orders o JOIN customers c ON
> (o.customer_id=c.id);
Query ID = hduser_20220714194426_c0471e28-10ae-49df-b004-5a799b802dec
Total jobs = 1

```

```

1      Deepti 1232      Phone
1      Deepti 1457      Desktop
1      Deepti 1232      Phone1
1      Deepti 1457      Desktop1
2      Alekhya 1234      Watch
2      Alekhya 1234      Watch1
3      Likhith 3412      Camera
3      Likhith 3412      Camera1
4      Keerthana 3468      Laptop
4      Keerthana 3468      Laptop1
Time taken: 62.369 seconds, Fetched: 10 row(s)

```

Query for indicating right table is small

```

hive> select /*+ MAPJOIN(c) */ c.id, c.name,
> o.id, o.product_id from orders o JOIN customers c ON
> (o.customer_id=c.id);
Query ID = hduser_20220714194645_0eb0c1b8-9003-49fe-88b2-82a08e39c323
Total jobs = 1

```

```

1      Deepti 1232      Phone
3      Likhith 3412      Camera
4      Keerthana 3468      Laptop
2      Alekhya 1234      Watch
1      Deepti 1457      Desktop
1      Deepti 1232      Phone1
3      Likhith 3412      Camera1
4      Keerthana 3468      Laptop1
2      Alekhya 1234      Watch1
1      Deepti 1457      Desktop1
Time taken: 64.846 seconds, Fetched: 10 row(s)

```

LEFT OUTER JOIN

```
hive> select /*+ MAPJOIN(c) */ c.id, c.name,  
  > o.id, o.product_id from orders o LEFT OUTER JOIN customers c ON  
  > (o.customer_id=c.id);  
Query ID = hduser_20220714194855_6b51bdb8-87c1-4b31-a207-b71bdb323cf8  
Total jobs = 1
```

```
OK  
1      Deepti  1232    Phone  
3      Likhith 3412    Camera  
4      Keerthana      3468    Laptop  
2      Alekhya 1234    Watch  
1      Deepti  1457    Desktop  
1      Deepti  1232    Phone1  
3      Likhith 3412    Camera1  
4      Keerthana      3468    Laptop1  
2      Alekhya 1234    Watch1  
1      Deepti  1457    Desktop1  
Time taken: 65.398 seconds, Fetched: 10 row(s)  
hive>
```

RIGHT OUTER JOIN

```
hive> select /*+ MAPJOIN(o) */ c.id, c.name,  
  > o.id, o.product_id from orders o RIGHT OUTER JOIN customers c ON  
  > (o.customer_id=c.id);  
Query ID = hduser_20220714195145_7796fb99-08cc-4230-9b33-7a814cda6468  
Total jobs = 1
```

```
OK  
1      Deepti  1232    Phone  
1      Deepti  1457    Desktop  
1      Deepti  1232    Phone1  
1      Deepti  1457    Desktop1  
2      Alekhya 1234    Watch  
2      Alekhya 1234    Watch1  
3      Likhith 3412    Camera  
3      Likhith 3412    Camera1  
4      Keerthana      3468    Laptop  
4      Keerthana      3468    Laptop1  
Time taken: 59.812 seconds, Fetched: 10 row(s)
```

BUCKET JOIN

```
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 4 buckets  
  > row format delimited fields terminated by ',';  
OK  
Time taken: 4.237 seconds
```

```
hive> load data local inpath "/home/hduser/Customers.txt" into table Customers_bucketed_new;  
Loading data to table demo_db.customers_bucketed_new  
Table demo_db.customers_bucketed_new stats: [numFiles=1, totalSize=251]  
OK  
Time taken: 1.82 seconds
```

```
hive> create external table orders_bucketed_new(
  > order_id int,
  > order_amount float,
  > order_date string,
  > order_customer_id int)
  > clustered by(order_customer_id) into 8 buckets
  > row format delimited fields terminated by ',';
```

```
OK
Time taken: 0.376 seconds
```

```
hive> load data local inpath "/home/hduser/orders.txt" into table orders_bucketed_new;
Loading data to table demo.db.orders_bucketed_new
Table demo.db.orders_bucketed_new stats: [numFiles=1, totalSize=352]
OK
Time taken: 1.454 seconds
```

```
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) limit 10;
Query ID = hduser_20220714210713_6d27b874-1765-4f5d-a4d9-3aca1dc1394d
Total jobs = 1
```

```
OK
3005   Graham   Zusi      70001   2012-10-05
3001   Brad     Guzan     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
3002   Nick     Rimando   70008   2012-09-10
3004   Fabian    Johnson   70010   2012-10-10
3009   Geoff    Cameron   70003   2012-10-10
3008   Julian   Green     70012   2012-06-27
Time taken: 90.46 seconds, Fetched: 10 row(s)
```

Sort Merge Bucket Join

```
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) limit 10;
Query ID = hduser_20220714210713_6d27b874-1765-4f5d-a4d9-3aca1dc1394d
Total jobs = 1
```