# First Example

We need to specify hint in query and set below two properties

hive> set hive.optimize.bucketmapjoin = true;

hive> set hive.ignore.mapjoin.hint=false;

* Logical Optimization   
    
  In the logical optimization, mapjoin optimization was followed by bucketmapjoin optimization, bucketmapjoin optimization take the MapJoin operator tree as input and convert it into BucketMapJoin, so a hinted query would be first transformed into a mapjoin and then a bucketmapjoin. Therefore, hint disabled logical optimisation would do nothing join optimisation on the join tree.
* Physical Optimization   
    
  In the physical optimisation, the hive.auto.convert.join was tested and MapJoinResolver was used and just convert a reduce join into a MapJoin. No further BucketMapJoin Optimization rules in this phase. That's why I just get Mapjoin in my question.

Here postimportBucketed table is bucketed on username with 30 buckets

|  |
| --- |
| hive> explain select /\*+ MAPJOIN(a) \*/ a.\* from postimportBucketed a join postimportBucketed b on a.username = b.username;  OK  STAGE DEPENDENCIES:  Stage-3 is a root stage  Stage-1 depends on stages: Stage-3  Stage-0 depends on stages: Stage-1  STAGE PLANS:  Stage: Stage-3  Map Reduce Local Work  Alias -> Map Local Tables:  a  Fetch Operator  limit: -1  Alias -> Map Local Operator Tree:  a  TableScan  alias: a  Statistics: Num rows: 42286538 Data size: 19438765021 Basic stats: COMPLETE Column stats: NONE  Filter Operator  predicate: username is not null (type: boolean)  Statistics: Num rows: 21143269 Data size: 9719382510 Basic stats: COMPLETE Column stats: NONE  HashTable Sink Operator  keys:  0 username (type: string)  1 username (type: string)  Stage: Stage-1  Map Reduce  **Map Operator Tree:**  TableScan  alias: b  Statistics: Num rows: 42286538 Data size: 19438765021 Basic stats: COMPLETE Column stats: NONE  Filter Operator  predicate: username is not null (type: boolean)  Statistics: Num rows: 21143269 Data size: 9719382510 Basic stats: COMPLETE Column stats: NONE  Map Join Operator  condition map:  Inner Join 0 to 1  keys:  0 username (type: string)  1 username (type: string)  outputColumnNames: \_col0, \_col1, \_col2, \_col3, \_col4, \_col5, \_col6, \_col7, \_col8, \_col9, \_col10, \_col11, \_col12, \_col13, \_col14, \_col15  Statistics: Num rows: 23257596 Data size: 10691320992 Basic stats: COMPLETE Column stats: NONE  File Output Operator  compressed: false  Statistics: Num rows: 23257596 Data size: 10691320992 Basic stats: COMPLETE Column stats: NONE  table:  input format: org.apache.hadoop.mapred.TextInputFormat  output format: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat  serde: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe  Local Work:  Map Reduce Local Work  Stage: Stage-0  Fetch Operator  limit: -1  Processor Tree:  ListSink  Time taken: 0.413 seconds, Fetched: 58 row(s) |

Just make sure that no **reducer operator tee** is there in explain plan

If we want to do simple(shuffle) join we can remove the hint

|  |
| --- |
| explain select a.\* from postimportBucketed a join postimportBucketed b on a.username = b.username;  OK  STAGE DEPENDENCIES:  Stage-5 is a root stage , consists of Stage-1  Stage-1  Stage-0 depends on stages: Stage-1  STAGE PLANS:  Stage: Stage-5  Conditional Operator  Stage: Stage-1  Map Reduce  **Map Operator Tree:**  TableScan  alias: a  Statistics: Num rows: 42286538 Data size: 19438765021 Basic stats: COMPLETE Column stats: NONE  Filter Operator  predicate: username is not null (type: boolean)  Statistics: Num rows: 21143269 Data size: 9719382510 Basic stats: COMPLETE Column stats: NONE  Reduce Output Operator  key expressions: username (type: string)  sort order: +  Map-reduce partition columns: username (type: string)  Statistics: Num rows: 21143269 Data size: 9719382510 Basic stats: COMPLETE Column stats: NONE  value expressions: machinename (type: string), machineid (type: string), ipaddress (type: string), domainname (type: string), pcmlogeventid (type: string), modulename (type: string), filename (type: string), productname (type: string), companyname (type: string), description (type: string), text (type: string), activetime (type: double), datetime (type: timestamp), timeslot (type: timestamp), date (type: date)  TableScan  alias: b  Statistics: Num rows: 42286538 Data size: 19438765021 Basic stats: COMPLETE Column stats: NONE  Filter Operator  predicate: username is not null (type: boolean)  Statistics: Num rows: 21143269 Data size: 9719382510 Basic stats: COMPLETE Column stats: NONE  Reduce Output Operator  key expressions: username (type: string)  sort order: +  Map-reduce partition columns: username (type: string)  Statistics: Num rows: 21143269 Data size: 9719382510 Basic stats: COMPLETE Column stats: NONE  **Reduce Operator Tree:**  Join Operator  condition map:  Inner Join 0 to 1  keys:  0 username (type: string)  1 username (type: string)  outputColumnNames: \_col0, \_col1, \_col2, \_col3, \_col4, \_col5, \_col6, \_col7, \_col8, \_col9, \_col10, \_col11, \_col12, \_col13, \_col14, \_col15  Statistics: Num rows: 23257596 Data size: 10691320992 Basic stats: COMPLETE Column stats: NONE  File Output Operator  compressed: false  Statistics: Num rows: 23257596 Data size: 10691320992 Basic stats: COMPLETE Column stats: NONE  table:  input format: org.apache.hadoop.mapred.TextInputFormat  output format: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat  serde: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe  Stage: Stage-0  Fetch Operator  limit: -1  Processor Tree:  ListSink  Time taken: 0.184 seconds, Fetched: 58 row(s) |

We can see from explain plan that there are Map and Reduce phase and map join operator is not available instead join Operator is available

|  |
| --- |
| hive> select /\*+ MAPJOIN(a) \*/ a.username from postimportBucketed a join postimportBucketed b on a.username = b.username;  Query ID = root\_20170404060101\_e8ded26d-4c6e-45e1-99b9-49a352f6e137  Total jobs = 1  Execution log at: /tmp/root/root\_20170404060101\_e8ded26d-4c6e-45e1-99b9-49a352f6e137.log  2017-04-04 06:01:29 Starting to launch local task to process map join; maximum memory = 1908932608  2017-04-04 06:01:34 Dump the side-table for tag: 0 with group count: 136 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000000\_0.hashtable  2017-04-04 06:01:34 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000000\_0.hashtable (4726908 bytes)  2017-04-04 06:01:38 Dump the side-table for tag: 0 with group count: 114 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000001\_0.hashtable  2017-04-04 06:01:38 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000001\_0.hashtable (4400919 bytes)  2017-04-04 06:01:44 Dump the side-table for tag: 0 with group count: 145 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000002\_0.hashtable  2017-04-04 06:01:45 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000002\_0.hashtable (5105430 bytes)  2017-04-04 06:02:00 Dump the side-table for tag: 0 with group count: 120 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000003\_0.hashtable  2017-04-04 06:02:00 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000003\_0.hashtable (4914776 bytes)  2017-04-04 06:02:32 Dump the side-table for tag: 0 with group count: 139 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000004\_0.hashtable  2017-04-04 06:02:32 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000004\_0.hashtable (6190440 bytes)  2017-04-04 06:03:03 Dump the side-table for tag: 0 with group count: 131 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000005\_0.hashtable  2017-04-04 06:03:03 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000005\_0.hashtable (5124036 bytes)  2017-04-04 06:03:10 Dump the side-table for tag: 0 with group count: 129 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000006\_0.hashtable  2017-04-04 06:03:10 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000006\_0.hashtable (6573412 bytes)  2017-04-04 06:03:15 Dump the side-table for tag: 0 with group count: 129 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000007\_0.hashtable  2017-04-04 06:03:15 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000007\_0.hashtable (5048218 bytes)  2017-04-04 06:03:42 Dump the side-table for tag: 0 with group count: 135 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000008\_0.hashtable  2017-04-04 06:03:42 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000008\_0.hashtable (5264436 bytes)  2017-04-04 06:03:58 Dump the side-table for tag: 0 with group count: 128 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000009\_0.hashtable  2017-04-04 06:03:58 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000009\_0.hashtable (5271813 bytes)  2017-04-04 06:04:49 Dump the side-table for tag: 0 with group count: 132 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000010\_0.hashtable  2017-04-04 06:04:49 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000010\_0.hashtable (6177005 bytes)  2017-04-04 06:05:06 Dump the side-table for tag: 0 with group count: 145 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000011\_0.hashtable  2017-04-04 06:05:06 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000011\_0.hashtable (5501312 bytes)  2017-04-04 06:05:35 Dump the side-table for tag: 0 with group count: 127 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000012\_0.hashtable  2017-04-04 06:05:35 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000012\_0.hashtable (5998053 bytes)  2017-04-04 06:06:04 Dump the side-table for tag: 0 with group count: 133 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000013\_0.hashtable  2017-04-04 06:06:04 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000013\_0.hashtable (5622444 bytes)  2017-04-04 06:07:01 Dump the side-table for tag: 0 with group count: 110 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000014\_0.hashtable  2017-04-04 06:07:02 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000014\_0.hashtable (5591608 bytes)  2017-04-04 06:07:19 Dump the side-table for tag: 0 with group count: 130 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000015\_0.hashtable  2017-04-04 06:07:19 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000015\_0.hashtable (5656639 bytes)  2017-04-04 06:07:25 Dump the side-table for tag: 0 with group count: 135 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000016\_0.hashtable  2017-04-04 06:07:25 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000016\_0.hashtable (6443900 bytes)  2017-04-04 06:07:31 Dump the side-table for tag: 0 with group count: 125 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000017\_0.hashtable  2017-04-04 06:07:31 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000017\_0.hashtable (5823179 bytes)  2017-04-04 06:07:39 Dump the side-table for tag: 0 with group count: 148 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000018\_0.hashtable  2017-04-04 06:07:39 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000018\_0.hashtable (7683129 bytes)  2017-04-04 06:07:45 Dump the side-table for tag: 0 with group count: 126 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000019\_0.hashtable  2017-04-04 06:07:45 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000019\_0.hashtable (5960519 bytes)  2017-04-04 06:08:24 Dump the side-table for tag: 0 with group count: 136 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000020\_0.hashtable  2017-04-04 06:08:24 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000020\_0.hashtable (6032486 bytes)  2017-04-04 06:08:47 Dump the side-table for tag: 0 with group count: 133 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000021\_0.hashtable  2017-04-04 06:08:47 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000021\_0.hashtable (6361470 bytes)  2017-04-04 06:09:11 Dump the side-table for tag: 0 with group count: 130 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000022\_0.hashtable  2017-04-04 06:09:11 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000022\_0.hashtable (5526829 bytes)  2017-04-04 06:09:38 Dump the side-table for tag: 0 with group count: 106 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000023\_0.hashtable  2017-04-04 06:09:38 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000023\_0.hashtable (4962102 bytes)  2017-04-04 06:09:56 Dump the side-table for tag: 0 with group count: 110 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000024\_0.hashtable  2017-04-04 06:09:56 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000024\_0.hashtable (7182266 bytes)  2017-04-04 06:10:27 Dump the side-table for tag: 0 with group count: 118 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000025\_0.hashtable  2017-04-04 06:10:27 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000025\_0.hashtable (6154275 bytes)  2017-04-04 06:10:42 Dump the side-table for tag: 0 with group count: 120 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000026\_0.hashtable  2017-04-04 06:10:42 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000026\_0.hashtable (4663182 bytes)  2017-04-04 06:11:07 Dump the side-table for tag: 0 with group count: 132 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000027\_0.hashtable  2017-04-04 06:11:08 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000027\_0.hashtable (5309035 bytes)  2017-04-04 06:11:13 Dump the side-table for tag: 0 with group count: 135 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000028\_0.hashtable  2017-04-04 06:11:13 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000028\_0.hashtable (5691551 bytes)  2017-04-04 06:11:47 Dump the side-table for tag: 0 with group count: 119 into file: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000029\_0.hashtable  2017-04-04 06:11:47 Uploaded 1 File to: file:/tmp/root/a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec/hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000029\_0.hashtable (5198256 bytes)  2017-04-04 06:11:47 End of local task; Time Taken: 617.225 sec.  Execution completed successfully  MapredLocal task succeeded  Launching Job 1 out of 1  Number of reduce tasks is set to 0 since there's no reduce operator  Starting Job = job\_1490094830773\_0095, Tracking URL = http://mac55:8088/proxy/application\_1490094830773\_0095/  Kill Command = /opt/cloudera/parcels/CDH-5.8.2-1.cdh5.8.2.p0.3/lib/hadoop/bin/hadoop job -kill job\_1490094830773\_0095  Hadoop job information for Stage-1: number of mappers: 30; number of reducers: 0  2017-04-04 06:12:06,352 Stage-1 map = 0%, reduce = 0% |

It will create 30 hashtable files and upload them to local file system

In case of Bucket map join,

1) hive runs a local map reduce join to create HashTable files upload it to local file system, 2) it compress and archive the file and load to distributed cache (i- Large hash table file will slow down the propagation of Distributed Cache. ii- Mappers are waiting for the hash tables file from Distributed Cache.)

3) Loads to the mappers of the map Join task

|  |
| --- |
| [root@mac55 root]#pwd  /tmp/root/  [root@mac55 root]# ls -l  total 756  drwx------ 3 root root 63 Apr 4 06:01 a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec  -rw-r--r-- 1 root root 0 Apr 4 06:00 a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec2602694806286850027.pipeout  -rw-r--r-- 1 root root 0 Apr 4 06:00 a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec66650766381595699.pipeout  -rw-r--r-- 1 root root 86051 Apr 4 06:01 hive.log  -rw-r--r-- 1 root root 137798 Mar 22 07:44 hive.log.2017-03-22  -rw-r--r-- 1 root root 135353 Mar 30 07:43 hive.log.2017-03-30  -rw-r--r-- 1 root root 196360 Mar 31 05:41 hive.log.2017-03-31  -rw-r--r-- 1 root root 116943 Apr 3 04:30 hive.log.2017-04-03  -rw-r--r-- 1 root root 5100 Apr 3 03:29 root\_20170403032929\_0580d116-9a68-42b9-8cea-ab6dd64999ee.log  -rw-r--r-- 1 root root 11263 Apr 3 04:32 root\_20170403043030\_7b2ecbf7-d7fa-4e59-97fa-51fc1f7b4835.log  -rw-r--r-- 1 root root 59376 Apr 4 06:10 root\_20170404060101\_e8ded26d-4c6e-45e1-99b9-49a352f6e137.log  [root@mac55 root]# cd a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec  [root@mac55 a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec]# ls -l  total 0  drwx------ 5 root root 63 Apr 4 06:01 hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1  [root@mac55 a6f9dc95-ef64-48b9-b3d2-5a08f71dd0ec]# cd hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1/  [root@mac55 hive\_2017-04-04\_06-01-27\_122\_2306327856718048607-1]# ls -l  total 0  drwxr-xr-x 3 root root 30 Apr 4 06:01 -local-10003  drwxr-xr-x 2 root root 41 Apr 4 06:01 -local-10005  drwxr-xr-x 2 root root 47 Apr 4 06:01 -local-10006 |

# Second Example with bucket map join

|  |
| --- |
| CREATE TABLE esplus\_project\_tbl\_bucketed1(  username string,  location string,  department string,  designation string,  grade string,  project string,  proj\_realease\_dt date)  CLUSTERED BY (username,project) INTO 10 BUCKETS; |

|  |
| --- |
| CREATE TABLE esplus\_project\_tbl\_bucketed2(  username string,  location string,  department string,  designation string,  grade string,  project string,  proj\_realease\_dt date)  CLUSTERED BY (username,project) INTO 20 BUCKETS; |

|  |
| --- |
| set hive.enforce.bucketing = true; |
| insert overwrite table esplus\_project\_tbl\_bucketed1 select \* from esplus\_project\_tbl; |

|  |
| --- |
| [root@mac55 ~]# hdfs dfs -ls /user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1  Found 10 items  -rwxrwxrwt 3 root hive 29513 2017-04-03 03:09 /user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000000\_0  -rwxrwxrwt 3 root hive 27494 2017-04-03 03:09 /user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000001\_0  -rwxrwxrwt 3 root hive 29113 2017-04-03 03:09 /user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000002\_0  -rwxrwxrwt 3 root hive 27947 2017-04-03 03:09 /user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000003\_0  -rwxrwxrwt 3 root hive 26583 2017-04-03 03:09 /user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000004\_0  -rwxrwxrwt 3 root hive 28102 2017-04-03 03:09 /user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000005\_0  -rwxrwxrwt 3 root hive 29031 2017-04-03 03:09 /user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000006\_0  -rwxrwxrwt 3 root hive 29574 2017-04-03 03:09 /user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000007\_0  -rwxrwxrwt 3 root hive 28687 2017-04-03 03:09 /user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000008\_0  -rwxrwxrwt 3 root hive 26211 2017-04-03 03:09 /user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000009\_0 |

|  |
| --- |
| set hive.enforce.bucketing = true;  insert overwrite table esplus\_project\_tbl\_bucketed2 select \* from esplus\_project\_tbl; |

|  |
| --- |
| hive> set hive.optimize.bucketmapjoin = true;  hive> set hive.ignore.mapjoin.hint=false; |

|  |
| --- |
| hive> explain select /\*+ MAPJOIN(a) \*/ a.username,a.project from esplus\_project\_tbl\_bucketed1 a join esplus\_project\_tbl\_bucketed2 b on a.username =b.username and a.project=b.project;  OK  STAGE DEPENDENCIES:  Stage-3 is a root stage  Stage-1 depends on stages: Stage-3  Stage-0 depends on stages: Stage-1  STAGE PLANS:  Stage: Stage-3  Map Reduce Local Work  Alias -> Map Local Tables:  a  Fetch Operator  limit: -1  Alias -> Map Local Operator Tree:  a  TableScan  alias: a  Statistics: Num rows: 3847 Data size: 278408 Basic stats: COMPLETE Column stats: NONE  Filter Operator  predicate: (username is not null and project is not null) (type: boolean)  Statistics: Num rows: 962 Data size: 69620 Basic stats: COMPLETE Column stats: NONE  HashTable Sink Operator  keys:  0 username (type: string), project (type: string)  1 username (type: string), project (type: string)  Stage: Stage-1  Map Reduce  Map Operator Tree:  TableScan  alias: b  Statistics: Num rows: 3847 Data size: 278408 Basic stats: COMPLETE Column stats: NONE  Filter Operator  predicate: (username is not null and project is not null) (type: boolean)  Statistics: Num rows: 962 Data size: 69620 Basic stats: COMPLETE Column stats: NONE  Map Join Operator  condition map:  Inner Join 0 to 1  keys:  0 username (type: string), project (type: string)  1 username (type: string), project (type: string)  outputColumnNames: \_col0, \_col5  Statistics: Num rows: 1058 Data size: 76582 Basic stats: COMPLETE Column stats: NONE  Select Operator  expressions: \_col0 (type: string), \_col5 (type: string)  outputColumnNames: \_col0, \_col1  Statistics: Num rows: 1058 Data size: 76582 Basic stats: COMPLETE Column stats: NONE  File Output Operator  compressed: false  Statistics: Num rows: 1058 Data size: 76582 Basic stats: COMPLETE Column stats: NONE  table:  input format: org.apache.hadoop.mapred.TextInputFormat  output format: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat  serde: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe  Local Work:  Map Reduce Local Work  Stage: Stage-0  Fetch Operator  limit: -1  Processor Tree:  ListSink  Time taken: 0.59 seconds, Fetched: 62 row(s) |

Or we can use explain extended to check if bucket map join is going to perform

|  |
| --- |
| hive> explain extended select /\*+ MAPJOIN(a) \*/ a.username,a.project from esplus\_project\_tbl\_bucketed1 a join esplus\_project\_tbl\_bucketed2 b on a.username =b.username and a.project=b.project;  OK  ABSTRACT SYNTAX TREE:  TOK\_QUERY  TOK\_FROM  TOK\_JOIN  TOK\_TABREF  TOK\_TABNAME  esplus\_project\_tbl\_bucketed1  a  TOK\_TABREF  TOK\_TABNAME  esplus\_project\_tbl\_bucketed2  b  and  =  .  TOK\_TABLE\_OR\_COL  a  username  .  TOK\_TABLE\_OR\_COL  b  username  =  .  TOK\_TABLE\_OR\_COL  a  project  .  TOK\_TABLE\_OR\_COL  b  project  TOK\_INSERT  TOK\_DESTINATION  TOK\_DIR  TOK\_TMP\_FILE  TOK\_SELECT  TOK\_HINTLIST  TOK\_HINT  TOK\_MAPJOIN  TOK\_HINTARGLIST  a  TOK\_SELEXPR  .  TOK\_TABLE\_OR\_COL  a  username  TOK\_SELEXPR  .  TOK\_TABLE\_OR\_COL  a  project  STAGE DEPENDENCIES:  Stage-3 is a root stage  Stage-1 depends on stages: Stage-3  Stage-0 depends on stages: Stage-1  STAGE PLANS:  Stage: Stage-3  Map Reduce Local Work  Alias -> Map Local Tables:  a  Fetch Operator  limit: -1  Alias -> Map Local Operator Tree:  a  TableScan  alias: a  Statistics: Num rows: 3847 Data size: 278408 Basic stats: COMPLETE Column stats: NONE  GatherStats: false  Filter Operator  isSamplingPred: false  predicate: (username is not null and project is not null) (type: boolean)  Statistics: Num rows: 962 Data size: 69620 Basic stats: COMPLETE Column stats: NONE  HashTable Sink Operator  keys:  0 username (type: string), project (type: string)  1 username (type: string), project (type: string)  Position of Big Table: 1  Bucket Mapjoin Context:  Alias Bucket Base File Name Mapping:  a {000000\_0=[000000\_0], 000001\_0=[000001\_0], 000002\_0=[000002\_0], 000003\_0=[000003\_0], 000004\_0=[000004\_0], 000005\_0=[000005\_0], 000006\_0=[000006\_0], 000007\_0=[000007\_0], 000008\_0=[000008\_0], 000009\_0=[000009\_0], 000010\_0=[000000\_0], 000011\_0=[000001\_0], 000012\_0=[000002\_0], 000013\_0=[000003\_0], 000014\_0=[000004\_0], 000015\_0=[000005\_0], 000016\_0=[000006\_0], 000017\_0=[000007\_0], 000018\_0=[000008\_0], 000019\_0=[000009\_0]}  Alias Bucket File Name Mapping:  a {hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000000\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000000\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000001\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000001\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000002\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000002\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000003\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000003\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000004\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000004\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000005\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000005\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000006\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000006\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000007\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000007\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000008\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000008\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000009\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000009\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000010\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000000\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000011\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000001\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000012\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000002\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000013\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000003\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000014\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000004\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000015\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000005\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000016\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000006\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000017\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000007\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000018\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000008\_0], hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000019\_0=[hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed1/000009\_0]}  Alias Bucket Output File Name Mapping:  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000000\_0 0  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000001\_0 1  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000002\_0 2  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000003\_0 3  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000004\_0 4  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000005\_0 5  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000006\_0 6  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000007\_0 7  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000008\_0 8  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000009\_0 9  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000010\_0 10  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000011\_0 11  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000012\_0 12  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000013\_0 13  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000014\_0 14  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000015\_0 15  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000016\_0 16  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000017\_0 17  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000018\_0 18  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2/000019\_0 19  Stage: Stage-1  Map Reduce  Map Operator Tree:  TableScan  alias: b  Statistics: Num rows: 3847 Data size: 278408 Basic stats: COMPLETE Column stats: NONE  GatherStats: false  Filter Operator  isSamplingPred: false  predicate: (username is not null and project is not null) (type: boolean)  Statistics: Num rows: 962 Data size: 69620 Basic stats: COMPLETE Column stats: NONE  Map Join Operator  condition map:  Inner Join 0 to 1  keys:  0 username (type: string), project (type: string)  1 username (type: string), project (type: string)  outputColumnNames: \_col0, \_col5  Position of Big Table: 1  Statistics: Num rows: 1058 Data size: 76582 Basic stats: COMPLETE Column stats: NONE  BucketMapJoin: true  Select Operator  expressions: \_col0 (type: string), \_col5 (type: string)  outputColumnNames: \_col0, \_col1  Statistics: Num rows: 1058 Data size: 76582 Basic stats: COMPLETE Column stats: NONE  File Output Operator  compressed: false  GlobalTableId: 0  directory: hdfs://mac55:8020/tmp/hive/root/ab14d5e9-b08b-4b00-ad1c-bd5aef4819f3/hive\_2017-04-12\_02-38-47\_716\_4288285061459494963-1/-mr-10000/.hive-staging\_hive\_2017-04-12\_02-38-47\_716\_4288285061459494963-1/-ext-10001  NumFilesPerFileSink: 1  Statistics: Num rows: 1058 Data size: 76582 Basic stats: COMPLETE Column stats: NONE  Stats Publishing Key Prefix: hdfs://mac55:8020/tmp/hive/root/ab14d5e9-b08b-4b00-ad1c-bd5aef4819f3/hive\_2017-04-12\_02-38-47\_716\_4288285061459494963-1/-mr-10000/.hive-staging\_hive\_2017-04-12\_02-38-47\_716\_4288285061459494963-1/-ext-10001/  table:  input format: org.apache.hadoop.mapred.TextInputFormat  output format: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat  properties:  columns \_col0,\_col1  columns.types string:string  escape.delim \  hive.serialization.extend.additional.nesting.levels true  serialization.format 1  serialization.lib org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe  serde: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe  TotalFiles: 1  GatherStats: false  MultiFileSpray: false  Local Work:  Map Reduce Local Work  Path -> Alias:  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2 [b]  Path -> Partition:  hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2  Partition  base file name: esplus\_project\_tbl\_bucketed2  input format: org.apache.hadoop.mapred.TextInputFormat  output format: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat  properties:  COLUMN\_STATS\_ACCURATE true  bucket\_count 20  bucket\_field\_name username  columns username,location,department,designation,grade,project,proj\_realease\_dt  columns.comments  columns.types string:string:string:string:string:string:date  file.inputformat org.apache.hadoop.mapred.TextInputFormat  file.outputformat org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat  location hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2  name esplus.esplus\_project\_tbl\_bucketed2  numFiles 20  numRows 3847  rawDataSize 278408  serialization.ddl struct esplus\_project\_tbl\_bucketed2 { string username, string location, string department, string designation, string grade, string project, date proj\_realease\_dt}  serialization.format 1  serialization.lib org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe  totalSize 282255  transient\_lastDdlTime 1491979084  serde: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe  input format: org.apache.hadoop.mapred.TextInputFormat  output format: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat  properties:  COLUMN\_STATS\_ACCURATE true  bucket\_count 20  bucket\_field\_name username  columns username,location,department,designation,grade,project,proj\_realease\_dt  columns.comments  columns.types string:string:string:string:string:string:date  file.inputformat org.apache.hadoop.mapred.TextInputFormat  file.outputformat org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat  location hdfs://mac55:8020/user/hive/warehouse/esplus.db/esplus\_project\_tbl\_bucketed2  name esplus.esplus\_project\_tbl\_bucketed2  numFiles 20  numRows 3847  rawDataSize 278408  serialization.ddl struct esplus\_project\_tbl\_bucketed2 { string username, string location, string department, string designation, string grade, string project, date proj\_realease\_dt}  serialization.format 1  serialization.lib org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe  totalSize 282255  transient\_lastDdlTime 1491979084  serde: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe  name: esplus.esplus\_project\_tbl\_bucketed2  name: esplus.esplus\_project\_tbl\_bucketed2  Truncated Path -> Alias:  /esplus.db/esplus\_project\_tbl\_bucketed2 [b]  Stage: Stage-0  Fetch Operator  limit: -1  Processor Tree:  ListSink  Time taken: 0.06 seconds, Fetched: 218 row(s) |

You can see it is showing BucketMapJoin: true

|  |
| --- |
| hive> select /\*+ MAPJOIN(a) \*/ a.username,a.project from esplus\_project\_tbl\_bucketed1 a join esplus\_project\_tbl\_bucketed2 b on a.username =b.username and a.project=b.project;  Query ID = root\_20170404062222\_df20efc8-a322-4ca0-b2b9-afd659f3948c  Total jobs = 1  Execution log at: /tmp/root/root\_20170404062222\_df20efc8-a322-4ca0-b2b9-afd659f3948c.log  2017-04-04 06:22:34 Starting to launch local task to process map join; maximum memory = 1908932608  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 398 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000000\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000000\_0.hashtable (25109 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 366 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000001\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000001\_0.hashtable (23219 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 384 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000002\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000002\_0.hashtable (24616 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 374 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000003\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000003\_0.hashtable (23430 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 357 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000004\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000004\_0.hashtable (22750 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 375 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000005\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000005\_0.hashtable (23911 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 384 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000006\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000006\_0.hashtable (24519 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 393 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000007\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000007\_0.hashtable (24982 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 387 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000008\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000008\_0.hashtable (24583 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 357 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000009\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000009\_0.hashtable (22603 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 398 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000010\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000010\_0.hashtable (25109 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 366 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000011\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000011\_0.hashtable (23219 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 384 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000012\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000012\_0.hashtable (24616 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 374 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000013\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000013\_0.hashtable (23430 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 357 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000014\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000014\_0.hashtable (22750 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 375 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000015\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000015\_0.hashtable (23911 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 384 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000016\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000016\_0.hashtable (24519 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 393 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000017\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000017\_0.hashtable (24982 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 387 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000018\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000018\_0.hashtable (24583 bytes)  2017-04-04 06:22:35 Dump the side-table for tag: 0 with group count: 357 into file: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000019\_0.hashtable  2017-04-04 06:22:35 Uploaded 1 File to: file:/tmp/root/1c5c423c-319a-4581-857a-f7a438115d2d/hive\_2017-04-04\_06-22-30\_510\_6683653373174132722-1/-local-10003/HashTable-Stage-1/MapJoin-a-10-000019\_0.hashtable (22603 bytes)  2017-04-04 06:22:35 End of local task; Time Taken: 1.288 sec.  Execution completed successfully  MapredLocal task succeeded  Launching Job 1 out of 1  Number of reduce tasks is set to 0 since there's no reduce operator  Starting Job = job\_1490094830773\_0096, Tracking URL = http://mac55:8088/proxy/application\_1490094830773\_0096/  Kill Command = /opt/cloudera/parcels/CDH-5.8.2-1.cdh5.8.2.p0.3/lib/hadoop/bin/hadoop job -kill job\_1490094830773\_0096  Hadoop job information for Stage-1: number of mappers: 20; number of reducers: 0  2017-04-04 06:22:50,528 Stage-1 map = 0%, reduce = 0%  2017-04-04 06:22:58,095 Stage-1 map = 15%, reduce = 0%, Cumulative CPU 3.41 sec  2017-04-04 06:23:00,273 Stage-1 map = 70%, reduce = 0%, Cumulative CPU 18.72 sec  2017-04-04 06:23:01,300 Stage-1 map = 75%, reduce = 0%, Cumulative CPU 20.11 sec  2017-04-04 06:23:03,359 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 26.35 sec  MapReduce Total cumulative CPU time: 26 seconds 350 msec  Ended Job = job\_1490094830773\_0096  MapReduce Jobs Launched:  Stage-Stage-1: Map: 20 Cumulative CPU: 26.35 sec HDFS Read: 459715 HDFS Write: 184042 SUCCESS  Total MapReduce CPU Time Spent: 26 seconds 350 msec  OK  730061006300680069006E0062006100 axR1myLpjGPB  720061006E006A0065006500740079006100 15zmCOE3ZMSy  6D0061006E006F006800610072006B00 15zmCOE3ZMSy  73007700610070006E0069006C0073007500 Tscg4DAOuMH2  6B006500740061006E00700061006E00 qrchdU8iWjQ1  6100730068006900730068006B006100 gQy4BOK82IpQ  700061006C006C00610076006900700061007200 lcWQMY2R3Rwk  620000000000000000000000000000000000 7XHRP1L7Qtbn  7300680072007500740069006B0061006D00 PIpiAbBScQq1  6E0061007400760069006400610064006300 UJLegs3E4lpg  620068007500700065006E006400720061006300 Spk0GV590Ld5  61006E0074006F006E007000 6ItGc9hnu1xK  6200680061006E0075006200 SGizCmxJ3m6k  76006900620068006F0072007300 GrLWUc4SHXSE  760000000000000000000000 axR1myLpjGPB  7600000000000000000000000000 Bxkk2hwPCmIo  6B00650064006100720070006900 axR1myLpjGPB  6D00610079007500720069006300 PIpiAbBScQq1  760069007300680061006C006B0075006C00 GVlEV3faUQCX  700072006100760069006E006B0075006D00610072006D00 H574TTzToWiG  760069006E006100790073006900 gQy4BOK82IpQ  61006B006500640069006100 RMG  700061006C006C00610076006900730068006900 7XHRP1L7Qtbn  610073006800770069006E0069006B006F00 YcpbiiYZSvnj  6B0075006D006100720061006B00 qrchdU8iWjQ1  670065006E0065007300790073007500730065007200 H574TTzToWiG  7200610064006800650073006800790061006D007000 6E0qtRLCG44c  7600720075007300680061006C0069007000 JFOqtkGEub4x  6D006100680065006E0064007200610062006800 1FXglrY4xP6J  70006F006F006E0061006D0070006100 Tscg4DAOuMH2  73006800610069006C006500730068006D006100 KSSgwSDEUS9s  660061006900730061006C006100 kUyKIWRVf5Xv  6B0069007300680061006E006200 1FXglrY4xP6J  700061006E006A00610062006300 lcWQMY2R3Rwk  610000000000000000000000 1FXglrY4xP6J  610000000000000000000000 1FXglrY4xP6J  610000000000000000000000 1FXglrY4xP6J  610000000000000000000000000000000000 uGUhVTywmgt6  7300680069006C00700061006200680061007000 7XHRP1L7Qtbn  76006900760065006B006B0075006D0061007200 i1KKBmPP61m3  760069006B0061007300730061006700 kUyKIWRVf5Xv  6B0075006E0061006C0069006E00 KSSgwSDEUS9s  70006F006F006A0061006B006100 Nh2KS6DNkeTb  61006D006900740067006F00 RMG  700075006A0061007700 GeJstn6iWQoX  730061006E006400690070007400 kUyKIWRVf5Xv  6D006F00680061006D006D006100640073006800 Nh2KS6DNkeTb  760069006B00720061006D006D006100 6ItGc9hnu1xK  6D0069006E006E0069006100 kUyKIWRVf5Xv  6E006900740065006E006A00 7XHRP1L7Qtbn  61006E007500700061006D0061006D00 6ItGc9hnu1xK  540061006C00650065006E006400720061006400 RMG  760069006B00720061006E0074006300 UJLegs3E4lpg  630072006D00610064006D0069006E00 axR1myLpjGPB  610072006300680061006E006100730068006900 1FXglrY4xP6J  7300680065006500740061006C007000 i1KKBmPP61m3  ..  ..  Time taken: 35.115 seconds, Fetched: 4085 row(s) |

It will create 20 hashtable files (larger bucket number) and upload them to local file system

# Second Example with map join

Since table size is small enough to fit into memory it will convert it to map side join

|  |
| --- |
| explain select a.username,a.project from esplus\_project\_tbl\_bucketed1 a join esplus\_project\_tbl\_bucketed2 b on a.username =b.username and a.project=b.project;  OK  STAGE DEPENDENCIES:  Stage-4 is a root stage  Stage-3 depends on stages: Stage-4  Stage-0 depends on stages: Stage-3  STAGE PLANS:  Stage: Stage-4  Map Reduce Local Work  Alias -> Map Local Tables:  a  Fetch Operator  limit: -1  Alias -> Map Local Operator Tree:  a  TableScan  alias: a  Statistics: Num rows: 3847 Data size: 278408 Basic stats: COMPLETE Column stats: NONE  Filter Operator  predicate: (username is not null and project is not null) (type: boolean)  Statistics: Num rows: 962 Data size: 69620 Basic stats: COMPLETE Column stats: NONE  HashTable Sink Operator  keys:  0 username (type: string), project (type: string)  1 username (type: string), project (type: string)  Stage: Stage-3  Map Reduce  Map Operator Tree:  TableScan  alias: b  Statistics: Num rows: 3847 Data size: 278408 Basic stats: COMPLETE Column stats: NONE  Filter Operator  predicate: (username is not null and project is not null) (type: boolean)  Statistics: Num rows: 962 Data size: 69620 Basic stats: COMPLETE Column stats: NONE  Map Join Operator  condition map:  Inner Join 0 to 1  keys:  0 username (type: string), project (type: string)  1 username (type: string), project (type: string)  outputColumnNames: \_col0, \_col5  Statistics: Num rows: 1058 Data size: 76582 Basic stats: COMPLETE Column stats: NONE  Select Operator  expressions: \_col0 (type: string), \_col5 (type: string)  outputColumnNames: \_col0, \_col1  Statistics: Num rows: 1058 Data size: 76582 Basic stats: COMPLETE Column stats: NONE  File Output Operator  compressed: false  Statistics: Num rows: 1058 Data size: 76582 Basic stats: COMPLETE Column stats: NONE  table:  input format: org.apache.hadoop.mapred.TextInputFormat  output format: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat  serde: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe  Local Work:  Map Reduce Local Work  Stage: Stage-0  Fetch Operator  limit: -1  Processor Tree:  ListSink  Time taken: 0.411 seconds, Fetched: 62 row(s)  hive> |

|  |
| --- |
| select a.username,a.project from esplus\_project\_tbl\_bucketed1 a join esplus\_project\_tbl\_bucketed2 b on a.username =b.username and a.project=b.project;  Query ID = root\_20170404063333\_1b01efa7-2c94-4630-8ed0-d1d8a7867fa1  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\_1490094830773\_0098, Tracking URL = http://mac55:8088/proxy/application\_1490094830773\_0098/  Kill Command = /opt/cloudera/parcels/CDH-5.8.2-1.cdh5.8.2.p0.3/lib/hadoop/bin/hadoop job -kill job\_1490094830773\_0098  Hadoop job information for Stage-1: number of mappers: 4; number of reducers: 1  2017-04-04 06:33:37,713 Stage-1 map = 0%, reduce = 0%  2017-04-04 06:33:42,952 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 2.49 sec  2017-04-04 06:33:45,019 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.25 sec  2017-04-04 06:33:50,145 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.84 sec  MapReduce Total cumulative CPU time: 6 seconds 840 msec  Ended Job = job\_1490094830773\_0098  MapReduce Jobs Launched:  Stage-Stage-1: Map: 4 Reduce: 1 Cumulative CPU: 6.84 sec HDFS Read: 591670 HDFS Write: 184042 SUCCESS  Total MapReduce CPU Time Spent: 6 seconds 840 msecime taken: 19.25 seconds, Fetched: 4085 row(s) |

Map side join will perform better than bucket map join in this case as both tables are small

# Second Example without bucketed map join

As file size is small you we are forcefully setting **hive.auto.conver.join** to false so that it will not convert it into map side join

|  |
| --- |
| hive> set hive.auto.convert.join=false;  hive> explain select a.username,a.project from esplus\_project\_tbl\_bucketed1 a join esplus\_project\_tbl\_bucketed2 b on a.username =b.username and a.project=b.project;  OK  STAGE DEPENDENCIES:  Stage-1 is a root stage  Stage-0 depends on stages: Stage-1  STAGE PLANS:  Stage: Stage-1  Map Reduce  Map Operator Tree:  TableScan  alias: a  Statistics: Num rows: 3847 Data size: 278408 Basic stats: COMPLETE Column stats: NONE  Filter Operator  predicate: (username is not null and project is not null) (type: boolean)  Statistics: Num rows: 962 Data size: 69620 Basic stats: COMPLETE Column stats: NONE  Reduce Output Operator  key expressions: username (type: string), project (type: string)  sort order: ++  Map-reduce partition columns: username (type: string), project (type: string)  Statistics: Num rows: 962 Data size: 69620 Basic stats: COMPLETE Column stats: NONE  TableScan  alias: b  Statistics: Num rows: 3847 Data size: 278408 Basic stats: COMPLETE Column stats: NONE  Filter Operator  predicate: (username is not null and project is not null) (type: boolean)  Statistics: Num rows: 962 Data size: 69620 Basic stats: COMPLETE Column stats: NONE  Reduce Output Operator  key expressions: username (type: string), project (type: string)  sort order: ++  Map-reduce partition columns: username (type: string), project (type: string)  Statistics: Num rows: 962 Data size: 69620 Basic stats: COMPLETE Column stats: NONE  Reduce Operator Tree:  Join Operator  condition map:  Inner Join 0 to 1  keys:  0 username (type: string), project (type: string)  1 username (type: string), project (type: string)  outputColumnNames: \_col0, \_col5  Statistics: Num rows: 1058 Data size: 76582 Basic stats: COMPLETE Column stats: NONE  Select Operator  expressions: \_col0 (type: string), \_col5 (type: string)  outputColumnNames: \_col0, \_col1  Statistics: Num rows: 1058 Data size: 76582 Basic stats: COMPLETE Column stats: NONE  File Output Operator  compressed: false  Statistics: Num rows: 1058 Data size: 76582 Basic stats: COMPLETE Column stats: NONE  table:  input format: org.apache.hadoop.mapred.TextInputFormat  output format: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat  serde: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe  Stage: Stage-0  Fetch Operator  limit: -1  Processor Tree:  ListSink  Time taken: 0.072 seconds, Fetched: 57 row(s) |

|  |
| --- |
| hive> select a.username,a.project from esplus\_project\_tbl\_bucketed1 a join esplus\_project\_tbl\_bucketed2 b on a.username =b.username and a.project=b.project;  Query ID = root\_20170404063333\_1b01efa7-2c94-4630-8ed0-d1d8a7867fa1  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\_1490094830773\_0098, Tracking URL = http://mac55:8088/proxy/application\_1490094830773\_0098/  Kill Command = /opt/cloudera/parcels/CDH-5.8.2-1.cdh5.8.2.p0.3/lib/hadoop/bin/hadoop job -kill job\_1490094830773\_0098  Hadoop job information for Stage-1: number of mappers: 4; number of reducers: 1  2017-04-04 06:33:37,713 Stage-1 map = 0%, reduce = 0%  2017-04-04 06:33:42,952 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 2.49 sec  2017-04-04 06:33:45,019 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.25 sec  2017-04-04 06:33:50,145 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.84 sec  MapReduce Total cumulative CPU time: 6 seconds 840 msec  Ended Job = job\_1490094830773\_0098  MapReduce Jobs Launched:  Stage-Stage-1: Map: 4 Reduce: 1 Cumulative CPU: 6.84 sec HDFS Read: 591670 HDFS Write: 184042 SUCCESS  Total MapReduce CPU Time Spent: 6 seconds 840 msecime taken: 24.353 seconds, Fetched: 4085 row(s) |

Since files are small that is why normal join perform much faster than bucket map join as normal join doesn’t upload files to local file system