

```
CREATE TABLE ${hiveconf:TEMP_TABLE1} AS
SELECT d_date_sk
FROM date_dim d
WHERE d.d_date >= '${hiveconf:q08_startDate}'
AND d.d_date <= '${hiveconf:q08_endDate}';
```

```
CREATE TABLE ${hiveconf:TEMP_TABLE2} AS
SELECT DISTINCT wcs_sales_sk
FROM (
  FROM (
    SELECT wcs_user_sk ,
    (wcs_click_date_sk * 86400L + wcs_click_time_sk) AS tstamp_inSec , wcs_sales_sk ,
    wp_type
  FROM web_clickstreams
  LEFT SEMI JOIN ${hiveconf:TEMP_TABLE1} date_filter
  ON (wcs_click_date_sk = date_filter.d_date_sk and wcs_user_sk IS NOT NULL)
  JOIN web_page w ON wcs_web_page_sk = w.wp_web_page_sk
  DISTRIBUTE BY wcs_user_sk
  SORT BY wcs_user_sk ,tstamp_inSec ,wcs_sales_sk ,wp_type
  ) q08_map_output
  REDUCE wcs_user_sk ,tstamp_inSec ,wcs_sales_sk ,wp_type
  USING 'python q08_filter_sales_with_reviews_viewed_before.py
  review ${hiveconf:q08_seconds_before_purchase}'
  AS (wcs_sales_sk BIGINT)
  ) sales_which_read_reviews;
```

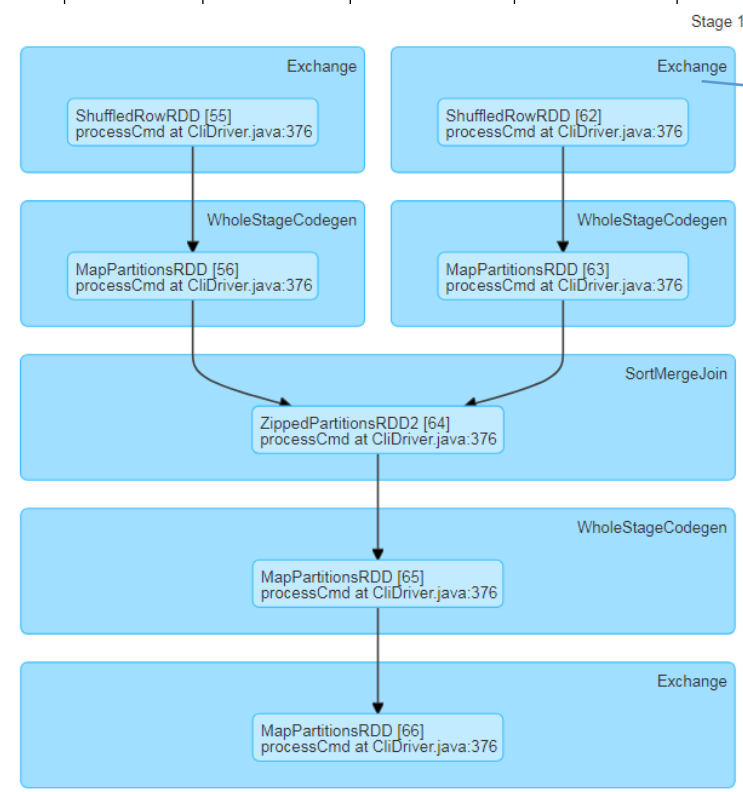
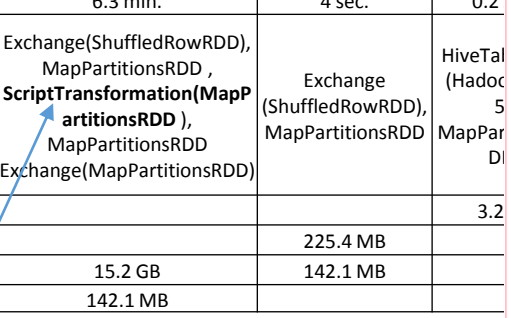
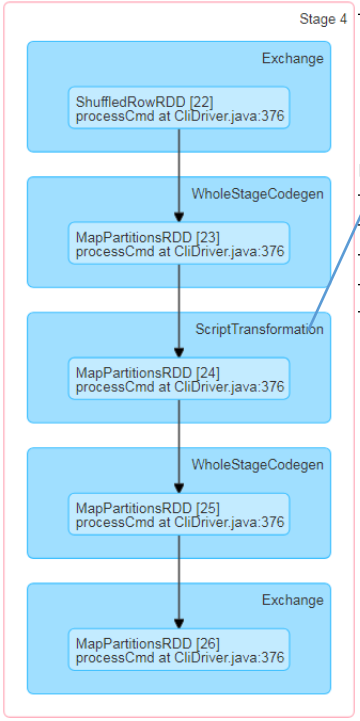
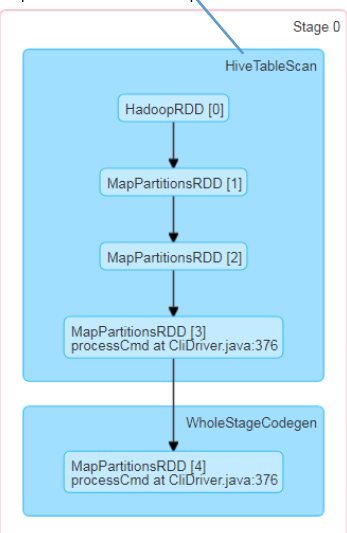
```
CREATE TABLE IF NOT EXISTS
${hiveconf:TEMP_TABLE3} AS
SELECT ws_net_paid , ws_order_number
FROM web_sales ws
JOIN ${hiveconf:TEMP_TABLE1} d ON (
ws.ws_sold_date_sk = d.d_date_sk);
```

```
DROP TABLE IF EXISTS ${hiveconf:RESULT_TABLE};
CREATE TABLE ${hiveconf:RESULT_TABLE} (
  q08_review_sales_amount BIGINT ,
  no_q08_review_sales_amount BIGINT)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' , LINES TERMINATED BY '\n'
STORED AS ${env:BIG_BENCH_hive_default_fileformat_result_table}
LOCATION '${hiveconf:RESULT_DIR}';

INSERT INTO TABLE ${hiveconf:RESULT_TABLE}
SELECT q08_review_sales.amount AS q08_review_sales_amount ,
q08_all_sales.amount - q08_review_sales.amount AS
no_q08_review_sales_amount
FROM (
  SELECT 1 AS id, SUM(ws_net_paid) as amount
  FROM ${hiveconf:TEMP_TABLE3} allSalesInYear
  LEFT SEMI JOIN ${hiveconf:TEMP_TABLE2} salesWithViewedReviews
  ON allSalesInYear.ws_order_number = salesWithViewedReviews.wcs_sales_sk
  ) q08_review_sales JOIN (
  SELECT 1 AS id, SUM(ws_net_paid) as amount
  FROM ${hiveconf:TEMP_TABLE3} allSalesInYear
  ) q08_all_sales ON q08_review_sales.id = q08_all_sales.id;
```

Q08 ORC No Compression

Source Code	lines 12-16	lines 19-42	lines 45-48	lines 51-75
Stages	0	1	2	3
Tasks	1	2	1	1
Duration	3 sec.	4 sec.	6.3 min.	4 sec.
Operations	HiveTableScan (HadoopRDD, 4x MapPartitionsRDD)	HiveTableScan (HadoopRDD, 4x MapPartitionsRDD)	Exchange(ShuffledRowRDD), MapPartitionsRDD, ScriptTransformation(MapPartitionsRDD), MapPartitionsRDD	Exchange(ShuffledRowRDD), MapPartitionsRDD
Input				
Output				
Shuffle Re:				
Shuffle Wr				



Stage 12	12	13
0	200	1
ec.	9 sec.	0.2 sec.
leScan pRDD, rtitions D)	2xExchange (ShuffledRowRDD), 2xMapPartitionsRDD, SortMergeJoin (ZippedPartitionsRDD 2), 2xMapPartitionsRDD	Exchange (ShuffledRow RDD), MapPartitions RDD
GB		
	2.1 GB	25.0 B
8 MB	11.5 KB	11.5 KB