**Code Used to Create Correlated Sample Through Join of the SRSWR and first table**

hive> create table s\_tpch1g.partsupp as select partsupp.ps\_partkey,partsupp.ps\_suppkey,partsupp.ps\_availqty, partsupp.ps\_supplycost, partsupp.ps\_comment from partsupp join s\_lineitem on partsupp.ps\_partkey =s\_lineitem.l\_partkey and partsupp.ps\_suppkey = s\_lineitem.l\_suppkey;

**Code Used to create the Part Sample Table**

hive> create table s\_tpch1g.part as select p\_partkey, p\_name, p\_mfgr, p\_brand, p\_type, p\_size, p\_container, p\_retailprice, p\_comment from part join s\_partsupp on part.p\_partkey = s\_partsupp.ps\_partkey;

**Code Used to create the Supplier Sample Table**

hive> create table s\_tpch1g.supplier as select s\_suppkey, s\_name,s\_address,s\_nationkey,s\_phone, s\_acctbal,s\_comment from supplier join s\_partsupp on supplier.s\_suppkey = s\_partsupp.ps\_suppkey;

**ORDERS TABLE IS EMPTY CANNOT CREATE THESE 4 LINKING TABLES**

**Code Used to create the Orders Table**

hive> create table s\_tpch1g.orders as select o\_orderkey, o\_custkey,o\_orderstatus,o\_totalprice,o\_orderdate,o\_orderpriority,o\_clerk,o\_shippriority,o\_comment from orders join s\_lineitem on orders.o\_orderkey = s\_lineitem.l\_orderkey;

**Code Used to create the Customer Table**

hive> create table s\_tpch1g.customer as select c\_custkey,c\_name,c\_address,c\_nationkey,c\_phone, c\_acctbal,c\_mktsegment,c\_comment from customer join s\_orders on customer.c\_custkey =s\_orders.o\_custkey;

**Code Used to create the Nation Table**

hive> create table s\_tpch1g.nation as select n\_nationkey, n\_name,n\_regionkey,n\_comment from nation join s\_customer on nation.n\_nationkey = customer.c\_nationkey;

**Code Used to create the Region Table**

hive> create table s\_tpch1g.region as select r\_regionkey, r\_name,r\_comment from region join s\_nation on region.r\_regionkey = s\_nation.n\_regionkey;

**Code Used to transfer table from Source to Sample Table**

hive> create table s\_tpch1g.lineitem as select \* from tpch1g.s\_lineitem;

**Results**

**LINEITEM PARTSUPP**

hive> select count (\*) from lineitem,partsupp where l\_partkey = ps\_partkey and l\_suppkey = ps\_suppkey;

**TPCH1G -** 14.383 seconds 6000003 count

15.162 seconds

15.010 seconds

12.499 seconds

14.697 seconds

TOTAL 71.751 seconds

AVERAGE 14.350 seconds

**S\_TPCH1G -** 07.516 seconds 64574 count

10.608 seconds

07.988 seconds

07.606 seconds

07.324 seconds

TOTAL 41.042 seconds

AVERAGE 08.208 seconds

**PARTSUPP PART**

Hive > select count (\*) from partsupp,part where ps\_partkey = p\_partkey;

**TPCH1G -** 10.590 seconds 800000 count

09.955 seconds

09.627 seconds

09.249 seconds

08.107 seconds

TOTAL 47.528 seconds

AVERAGE 09.506 seconds

**S\_TPCH1G -** 08.532 seconds 78012 count

03.745 seconds

07.435 seconds

07.195 seconds

07.353 seconds

TOTAL 34.260 seconds

AVERAGE 06.852 seconds

**PARTSUPP SUPPLIER**

Hive > select count (\*) from partsupp,supplier where ps\_suppkey = s\_suppkey

**TPCH1G -** 10.161 seconds 800000 count

07.232 seconds

09.317 seconds

06.965 seconds

07.733 seconds

TOTAL 41.408 seconds

AVERAGE 08.282 seconds

**S\_TPCH1G -** 07.497 seconds 418964 count

09.386 seconds

02.391 seconds

04.150 seconds

02.990 seconds

TOTAL 26.414 seconds

AVERAGE 05.283 seconds

**LIINEITEM ORDERS**

Hive > select count (\*) from lineitem,orders where l\_orderkey = o\_orderkey;

**TPCH1G -** 20.711 seconds 6,000,003 count

19.804 seconds

21.779 seconds

21.498 seconds

19.211 seconds

TOTAL 103.003 seconds

AVERAGE 20.601 seconds

**S\_TPCH1G -** 08.277 seconds 62274 count

02.713 seconds

07.528 seconds

07.774 seconds

08.688 seconds

TOTAL 34.980 seconds

AVERAGE 06.996 seconds

**ORDERS CUSTOMER**

Hive> select count (\*) from orders,customer where o\_custkey = c\_custkey;

**TPCH1G -** 10.242 seconds 1500000 count

10.520 seconds

08.308 seconds

07.306 seconds

03.698 seconds

TOTAL 40.074 seconds

AVERAGE 08.015 seconds

**S\_TPCH1G -** 07.775 seconds 98066 count

06.796 seconds

08.008 seconds

08.469 seconds

02.497 seconds

TOTAL 33.520 seconds

AVERAGE 06.704 seconds

**CUSTOMER NATION**

Hive> select count (\*) from customer,nation where c\_nationkey = n\_nationkey;

**TPCH1G -** 07.504 seconds 1500000 count

02.951 seconds

02.313 seconds

02.198 seconds

01.796 seconds

TOTAL 16.762 seconds

AVERAGE 03.352 seconds

**S\_TPCH1G -** 31.290 seconds 144,103,006 count

32.489 seconds

29.309 seconds

30.689 seconds

26.190 seconds

TOTAL 149.967 seconds

AVERAGE 29.993 seconds

**NATION REGION**

**TPCH1G -** 06.525 seconds 25 count

06.280 seconds

01.670 seconds

02.259 seconds

01.542 seconds

TOTAL 17.886 seconds

AVERAGE 03.577 seconds

**S\_TPCH1G -** 193.290 seconds 720,088,870 count

170.069 seconds

191.972 seconds

167.730 seconds

171.875 seconds

TOTAL 894.936 seconds

AVERAGE 178.987 seconds