

## Explain

STAGE DEPENDENCIES:

Stage-1 is a root stage

Stage-2 depends on stages: Stage-1

Stage-10 depends on stages: Stage-2

Stage-4 depends on stages: Stage-10

Stage-0 depends on stages: Stage-4

Stage-11 depends on stages: Stage-0

Stage-5 depends on stages: Stage-11

STAGE PLANS:

Stage: Stage-1

Map Reduce

Map Operator Tree:

TableScan

alias: customers

Statistics: Num rows: 41898 Data size: 544674 Basic stats: COMPLETE Column stats: NONE

Filter Operator

predicate: (first\_name is not null and last\_name is not null and email is not null) (type: boolean)

Statistics: Num rows: 41898 Data size: 544674 Basic stats: COMPLETE Column stats: NONE

Select Operator

expressions: first\_name (type: string)

outputColumnNames: \_col0

Statistics: Num rows: 41898 Data size: 544674 Basic stats: COMPLETE Column stats: NONE

Union

Statistics: Num rows: 83109 Data size: 1080417 Basic stats: COMPLETE Column stats: NONE

Group By Operator

keys: \_col0 (type: string)

mode: hash

outputColumnNames: \_col0

Statistics: Num rows: 83109 Data size: 1080417 Basic stats: COMPLETE Column stats: NONE

Reduce Output Operator

key expressions: \_col0 (type: string)

sort order: +++++++

Map-reduce partition columns: \_col0 (type: string)

Statistics: Num rows: 83109 Data size: 1080417 Basic stats: COMPLETE Column stats: NONE

TableScan

alias: customers

Statistics: Num rows: 41211 Data size: 535743 Basic stats: COMPLETE Column stats: NONE

Filter Operator

predicate: (first_name is not null and last_name is not null and email is not null) (type: boolean)
Statistics: Num rows: 41211 Data size: 535743 Basic stats: COMPLETE Column stats: NONE
Select Operator
expressions: first_name (type: string)
outputColumnNames: _col0
Statistics: Num rows: 41211 Data size: 535743 Basic stats: COMPLETE Column stats: NONE
Union
Statistics: Num rows: 83109 Data size: 1080417 Basic stats: COMPLETE Column stats: NONE
Group By Operator
keys: _col0 (type: string)
mode: hash
outputColumnNames: _col0
Statistics: Num rows: 83109 Data size: 1080417 Basic stats: COMPLETE Column stats: NONE
Reduce Output Operator
key expressions: _col0 (type: string)
sort order: ++++++
Map-reduce partition columns: _col0 (type: string)
Statistics: Num rows: 83109 Data size: 1080417 Basic stats: COMPLETE Column stats: NONE
Reduce Operator Tree:
Group By Operator
keys: KEY._col0 (type: string)
mode: mergepartial
outputColumnNames: _col0
Statistics: Num rows: 41554 Data size: 540202 Basic stats: COMPLETE Column stats: NONE
File Output Operator
compressed: false
table:
input format: org.apache.hadoop.mapred.SequenceFileInputFormat
output format: org.apache.hadoop.hive ql.io.HiveSequenceFileOutputFormat
serde: org.apache.hadoop.hive.serde2.lazybinary.LazyBinarySerDe
Stage: Stage-2
Map Reduce
Map Operator Tree:
TableScan
Union
Statistics: Num rows: 58445 Data size: 759785 Basic stats: COMPLETE Column stats: NONE
Group By Operator
keys: _col0 (type: string)
mode: hash

outputColumnNames: _col0
Statistics: Num rows: 58445 Data size: 759785 Basic stats: COMPLETE Column stats: NONE
Reduce Output Operator
key expressions: _col0 (type: string)
sort order: +++++++
Map-reduce partition columns: _col0 (type: string)
Statistics: Num rows: 58445 Data size: 759785 Basic stats: COMPLETE Column stats: NONE
TableScan
alias: customers
Statistics: Num rows: 16891 Data size: 219583 Basic stats: COMPLETE Column stats: NONE
Filter Operator
predicate: (first_name is not null and last_name is not null and email is not null) (type: boolean)
Statistics: Num rows: 16891 Data size: 219583 Basic stats: COMPLETE Column stats: NONE
Select Operator
expressions: first_name (type: string)
outputColumnNames: _col0
Statistics: Num rows: 16891 Data size: 219583 Basic stats: COMPLETE Column stats: NONE
Union
Statistics: Num rows: 58445 Data size: 759785 Basic stats: COMPLETE Column stats: NONE
Group By Operator
keys: _col0 (type: string)
mode: hash
outputColumnNames: _col0
Statistics: Num rows: 58445 Data size: 759785 Basic stats: COMPLETE Column stats: NONE
Reduce Output Operator
key expressions: _col0 (type: string)
sort order: +++++++
Map-reduce partition columns: _col0 (type: string)
Statistics: Num rows: 58445 Data size: 759785 Basic stats: COMPLETE Column stats: NONE
Reduce Operator Tree:
Group By Operator
keys: KEY._col0 (type: string)
mode: mergepartial
outputColumnNames: _col0
Statistics: Num rows: 29222 Data size: 379886 Basic stats: COMPLETE Column stats: NONE
File Output Operator
compressed: false
table:

input format: org.apache.hadoop.mapred.SequenceFileInputFormat
output format: org.apache.hadoop.hive ql.io.HiveSequenceFileOutputFormat
serde: org.apache.hadoop.hive.serde2.lazybinary.LazyBinarySerDe
Stage: Stage-10
Map Reduce Local Work
Alias -> Map Local Tables:
\$hdt\$_0:\$hdt\$_1:p
Fetch Operator
limit: -1
Alias -> Map Local Operator Tree:
\$hdt\$_0:\$hdt\$_1:p
TableScan
alias: p
Statistics: Num rows: 100000 Data size: 1000000 Basic stats: COMPLETE Column stats: NONE
Filter Operator
predicate: (first_name is not null and last_name is not null and email is not null) (type: boolean)
Statistics: Num rows: 100000 Data size: 1000000 Basic stats: COMPLETE Column stats: NONE
Select Operator
expressions: first_name (type: string)
outputColumnNames: _col0
Statistics: Num rows: 100000 Data size: 1000000 Basic stats: COMPLETE Column stats: NONE
HashTable Sink Operator
keys:
0 _col0 (type: string)
1 _col0 (type: string)
Stage: Stage-4
Map Reduce
Map Operator Tree:
TableScan
Map Join Operator
condition map:
Inner Join 0 to 1
keys:
0 _col0 (type: string)
1 _col0 (type: string)
outputColumnNames: _col3
Statistics: Num rows: 110000 Data size: 1100000 Basic stats: COMPLETE Column stats: NONE
Select Operator
expressions: _col6 (type: int)
outputColumnNames: _col0
Statistics: Num rows: 110000 Data size: 1100000 Basic stats: COMPLETE Column stats: NONE

Group By Operator
aggregations: count(_col3)
keys: _col0 (type: int)
mode: hash
outputColumnNames: _col0
Statistics: Num rows: 110000 Data size: 1100000 Basic stats: COMPLETE Column stats: NONE
Reduce Output Operator
key expressions: _col0 (type: int)
sort order: +++
Map-reduce partition columns: _col0 (type: int)
Statistics: Num rows: 110000 Data size: 1100000 Basic stats: COMPLETE Column stats: NONE
value expressions: _col3 (type: bigint)
Local Work:
Map Reduce Local Work
Reduce Operator Tree:
Group By Operator
aggregations: count(VALUE._col0)
keys: KEY._col0 (type: int)
mode: mergepartial
outputColumnNames: _col0
Statistics: Num rows: 55000 Data size: 550000 Basic stats: COMPLETE Column stats: NONE
Select Operator
expressions: _col1 (type: string)
outputColumnNames: _col0
Statistics: Num rows: 55000 Data size: 550000 Basic stats: COMPLETE Column stats: NONE
File Output Operator
compressed: false
Statistics: Num rows: 55000 Data size: 550000 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
name: default.subscribe_statistics
Stage: Stage-0
Move Operator
files:
hdfs directory: true
destination: hdfs://namenode:8020/user/hive/warehouse/subscribe_statistics
Stage: Stage-11
Create Table Operator:
Create Table
columns: company string

input format: org.apache.hadoop.mapred.TextInputFormat
output format: org.apache.hadoop.hive.ql.io.IgnoreKeyTextOutputFormat
serde name: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
name: default.subscribe_statistics
Stage: Stage-5
Stats-Aggr Operator