VU DANG QUYNH GIANG

17133015

exercise1.pig

|  |
| --- |
| SET default\_parallel 5;  set job.name 'exercise1\_GROUP02';    %declare INPUT\_PATH '/home/hadoopuser/pig\_exercise/tstat-sample.txt';  %declare OUTPUT\_PATH '/home/hadoopuser/pig\_exercise/STAT/TCP-COUNT1A/';      -- Load input data from local input directory  LOGS = LOAD '$INPUT\_PATH' using PigStorage(' ') AS (ip\_c:chararray,  port\_c:int,  packets\_c:int,  rst\_c:int,  ack\_c:int,  purack\_c:int,  unique\_bytes\_c:long,  data\_pkts\_c:int,  data\_bytes\_c:long,  rexmit\_pkts\_c:int,  rexmit\_bytes\_c:long,  out\_seq\_pkts\_c:int,  syn\_c:int,  fin\_c:int,  ws\_c:int,  ts\_c:int,  window\_scale\_c:int,  sack\_req\_c:int,  sack\_sent\_c:int,  mss\_c:int,  max\_seg\_size\_c:int,  min\_seg\_size\_c:int,  win\_max\_c:int,  win\_min\_c:int,  win\_zero\_c:int,  cwin\_max\_c:long,  cwin\_min\_c:long,  initial\_cwin\_c:long,  average\_rtt\_c:double,  rtt\_min\_c:double,  rtt\_max\_c:double,  stdev\_rtt\_c:double,  rtt\_count\_c:int,  ttl\_min\_c:int,  ttl\_max\_c:int,  rtx\_RTO\_c:int,  rtx\_FR\_c:int,  reordering\_c:int,  net\_dup\_c:int,  unknown\_segments\_c:int,  flow\_control\_c:int,  unnece\_rtx\_rto\_c:int,  unnece\_rtx\_fr\_c:int,  different\_syn\_seqno\_c:int,  ip\_s:chararray,  port\_s:int,  packets\_s:int,  rst\_s:int,  ack\_s:int,  purack\_s:int,  unique\_bytes\_s:long,  data\_pkts\_s:int,  data\_bytes\_s:long,  rexmit\_pkts\_s:int,  rexmit\_bytes\_s:long,  out\_seq\_pkts\_s:int,  syn\_s:int,  fin\_s:int,  ws\_s:int,  ts\_s:int,  window\_scale\_s:int,  sack\_req\_s:int,  sack\_sent\_s:int,  mss\_s:int,  max\_seg\_size\_s:int,  min\_seg\_size\_s:int,  win\_max\_s:int,  win\_min\_s:int,  win\_zero\_s:int,  cwin\_max\_s:long,  cwin\_min\_s:long,  initial\_cwin\_s:long,  average\_rtt\_s:double,  rtt\_min\_s:double,  rtt\_max\_s:double,  stdev\_rtt\_s:double,  rtt\_count\_s:int,  ttl\_min\_s:int,  ttl\_max\_s:int,  rtx\_RTO\_s:int,  rtx\_FR\_s:int,  reordering\_s:int,  net\_dup\_s:int,  unknown\_segments\_s:int,  flow\_control\_s:int,  unnece\_rtx\_rto\_s:int,  unnece\_rtx\_fr\_s:int,  different\_syn\_seqno\_s:int,  completion\_time:double,  first\_time:double,  last\_time:double,  C\_first\_payload:double,  S\_first\_payload:double,  C\_last\_payload:double,  S\_last\_payload:double,  C\_first\_ack:double,  S\_first\_ack:double,  first\_time\_abs:double,  c\_internal:int,  s\_internal:int,  connection\_type:int,  p2p\_type:int,  p2p\_subtype:int,  ed2k\_data:int,  ed2k\_signaling:int,  ed2k\_c2s:int,  ed2k\_c2c:int,  ed2k\_chat:int,  http\_type:int,  ssl\_client\_hello:chararray,  ssl\_server\_hello:chararray,  dropbox\_id:bytearray,  fqdn:chararray);    --DUMP LOGS;    -- Group by client IP  FLOW = GROUP LOGS BY ip\_c;    -- Generate the output data  TRAFFIC = FOREACH FLOW GENERATE group, COUNT(LOGS);    -- TODO: Store output data in ./local-output/TCP\_DUMP/  STORE TRAFFIC INTO '$OUTPUT\_PATH'; |

hadoopuser@master:~/pig\_exercise$ pig -x local exercise1.pig









