Lab 3

**Roll No.**: J061-Anirudh VM, J069-Shubh Lilani

**Aim**: Word Count Using Map Reduce

**Objectives:**

1. To run Hive command.
2. Copy Data file from Local to HDFS.
3. Generate a Word count query.
4. Display Word count of the file

**Codes:**

//Map Reduce in HIVE hive

CREATE TABLE FILES (line STRING);

LOAD DATA INPATH 'data1.txt' OVERWRITE INTO TABLE FILES;

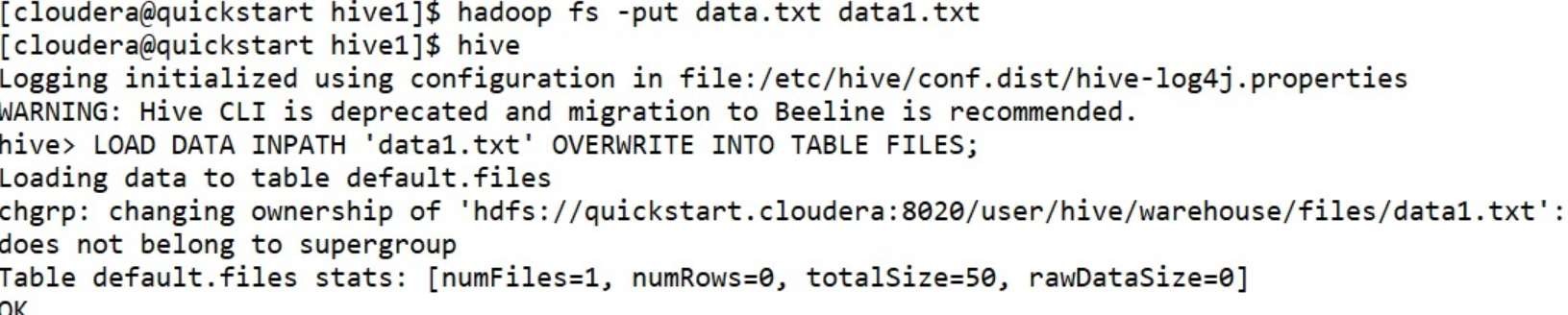
CREATE TABLE word\_count AS

SELECT w.word, count(1) AS count from

(SELECT explode(split(line, ' ')) as word from FILES) w GROUP BY w.word

ORDER BY w.word;

SELECT \* FROM word\_count ;



In order to change the average 1oad •or a reducer (in bytes ) :

Set hive.exsc.seducers.bytes.per.reducer=‹number>

In order to limit the maximum number of seducers:

set hive.exec.seducers.max:‹number›

In order to set a constant number of seducers:

set mapreduce.job.reduces=‹numben>

Starting 3ob: job 1614416l56655\_B082, Tracking URL: http://quickstart.cloudera:â0B8/proxyfapplication\_16l4416l56655\_B B2/

K:ill Command = / usn/11b/hadoop/b:in/hadoop job -k:ill job 16L4416156655 9682

I-ladoop job 1nformation for Stage -Z : number of rappers : 1i number of reducers : 1 2621—02—27 B2 : B9 : 4e, 727 Stage -2 map = 8g, reduce = a'g

2021-62-27 02 : 09 : 52, B72 Stage-2 map = 1001, reduce = B •, Curru1at1 ve CPU 1.5 sec 2621-B2-27 BZ : GB : BB, B57 Stage-2 map = 1B0g, reduce = been, Cumu1 ative CPU 5.B4 sec MapReduce Total cuoulat1ve CPU t:ime: 5 seconds 48 msec

Ended 3eb = yob\_16l44l6156655\_BB6Z

Novlng data to: hd€s: / /qul c kstart. c1oudera:ae2a/user/ hive/warehouse/word\_count

7able default.word\_count stats: [numFiles=1, numRows=7, totalSize=54, rawDatzSize=47]

NapReduce 3obs Launched:

Stage-Stage-1: Map: l Reduce: 1 Cumulative cPu: 9.74 etc HDFS Read: 7589 HDFS Write. 262 SUCCESS Stage-Stage-Z: Map: l Reduce: I Eumulative CPU: 5.04 sec HDF5 Aead: 4886 MDFS Write: 128 SUC0E5S Total MapReduce CPU Time Spent: 8 seconds 7ae msec

h1ve › CREATE TABLE word\_count A5

› **SELECT w.** xord, count (1) AS count from

» (SELECT explode( sp1It(11ne, ' ' ) ) AS word FRON FILES) u

› GROUP BY w. word

# › ORDER BY x.xord;

query ID = c louder a\_2e21e227820808\_9b7e 516d -essb-<wg -sfc6 -56dra2c78bbc

Tptal jobs = 2

Launching 3ob 1 out of 2

Number of reduce tas ks not specified. E stlmated Iron 1nput data s1ze: 1 In order to change the average 1oad -for a reducer ( in bytes ) :

set hive .exec . seducers . bytes . per . reducer=<number›

In order to limit the maximum number of seducers:

set hive .exec. seducers . max=‹number›

In order to set a constant number of seducers : set etapreduce . j ob. reduces=‹nuober›

Starting Job = job\_16L44l6l&66SS\_000l, Tracking URL: http://quickztart.cloudera:8088/proxy/application\_16144l6l56658\_ee01/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_l6l44l6156655\_ee0l Nadoop job 1nJ"-ormation -For Stage -1 : nultiber a-F rappers : 1; number of seducers : 1 2021-e2-27 ez : e9 : 83,631 Stage -1 nap = US, reduce = BE

2021-82-27 e2: e9:15, 283 Stage -1 map = 1001, reduce = 0X•, Cumu1at1ve CPU 2. e2 sec 2B21-82-27 02: B9 : 27,523 Stage -1 map = 1BBX, reduce = 18Bg, Cunu1at1ve CPU 3.74 sec I4apReduce Total cumu1at1ve CPU €1ne: 3 seconds 74ig lnsec

# Ended Job = job\_16l4416l56658\_000l

Launching 3ob 2 out of 2

Eumber o+ reduce tasks determined at compile time: 1

|  |  |  |
| --- | --- | --- |
| h1ve› | SELECT | \* FRON word\_count |
| OK |  |  |
| Th1s | 2 |  |
| a | 2 |  |
| h1ve | 1 |  |
| is | 2 |  |
| spark | 1 |  |

tutorial 1

tutorial. 1

Time taken: e.ea2 seconds, Fetched: 7 row(s)