Hadoop Validation

Services

* HDFS
* MapReduce
* Yarn
* Hive
* Spark

Hive Table creation using Beeline

Command :

0: jdbc:hive2://netf-pmbigdat01.epmtelco.com.> CREATE EXTERNAL TABLE IF NOT EXISTS school(

. . . . . . . . . . . . . . . . . . . . . . .>         Id int,

. . . . . . . . . . . . . . . . . . . . . . .>         Name STRING)

. . . . . . . . . . . . . . . . . . . . . . .>     ROW FORMAT DELIMITED

. . . . . . . . . . . . . . . . . . . . . . .>     FIELDS TERMINATED BY ','

. . . . . . . . . . . . . . . . . . . . . . .>     STORED AS TEXTFILE

. . . . . . . . . . . . . . . . . . . . . . .>     location '/user/data';

INFO  : Compiling command(queryId=hive\_20170824112929\_560584f4-844c-4cd5-b03e-14c25d1d4be1): CREATE EXTERNAL TABLE IF NOT EXISTS school(

        Id int,

        Name STRING)

    ROW FORMAT DELIMITED

    FIELDS TERMINATED BY ','

    STORED AS TEXTFILE

    location '/user/data'

INFO  : Semantic Analysis Completed

INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)

INFO  : Completed compiling command(queryId=hive\_20170824112929\_560584f4-844c-4cd5-b03e-14c25d1d4be1); Time taken: 0.022 seconds

INFO  : Executing command(queryId=hive\_20170824112929\_560584f4-844c-4cd5-b03e-14c25d1d4be1): CREATE EXTERNAL TABLE IF NOT EXISTS school(

        Id int,

        Name STRING)

    ROW FORMAT DELIMITED

    FIELDS TERMINATED BY ','

    STORED AS TEXTFILE

    location '/user/data'

INFO  : Completed executing command(queryId=hive\_20170824112929\_560584f4-844c-4cd5-b03e-14c25d1d4be1); Time taken: 0.002 seconds

INFO  : OK

No rows affected (0.082 seconds)

Validating Mapreduce & Yarn using Beeline

Command:

0: jdbc:hive2://netf-pmbigdat01.epmtelco.com.> insert into school values ('3', 'sumewer');

Output:

INFO  : Compiling command(queryId=hive\_20170824115555\_56fb64b2-f183-4735-87ed-b00768cc9e5b): insert into school values ('3', 'sumewer')

INFO  : Semantic Analysis Completed

INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:\_col0, type:int, comment:null), FieldSchema(name:\_col1, type:string, comment:null)], properties:null)

INFO  : Completed compiling command(queryId=hive\_20170824115555\_56fb64b2-f183-4735-87ed-b00768cc9e5b); Time taken: 0.092 seconds

INFO  : Executing command(queryId=hive\_20170824115555\_56fb64b2-f183-4735-87ed-b00768cc9e5b): insert into school values ('3', 'sumewer')

INFO  : Query ID = hive\_20170824115555\_56fb64b2-f183-4735-87ed-b00768cc9e5b

INFO  : Total jobs = 3

INFO  : Launching Job 1 out of 3

INFO  : Starting task [Stage-1:MAPRED] in serial mode

INFO  : Number of reduce tasks is set to 0 since there's no reduce operator

INFO  : number of splits:1

INFO  : Submitting tokens for job: job\_1503581872672\_0010

INFO  : The url to track the job: http://netf-pmbigdat02.epmtelco.com.co:8088/proxy/application\_1503581872672\_0010/

INFO  : Starting Job = job\_1503581872672\_0010, Tracking URL = http://netf-pmbigdat02.epmtelco.com.co:8088/proxy/application\_1503581872672\_0010/

INFO  : Kill Command = /opt/cloudera/parcels/CDH-5.8.4-1.cdh5.8.4.p0.5/lib/hadoop/bin/hadoop job  -kill job\_1503581872672\_0010

INFO  : Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

INFO  : 2017-08-24 11:55:47,881 Stage-1 map = 0%,  reduce = 0%

INFO  : 2017-08-24 11:55:53,043 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.87 sec

INFO  : MapReduce Total cumulative CPU time: 1 seconds 870 msec

INFO  : Ended Job = job\_1503581872672\_0010

INFO  : Starting task [Stage-7:CONDITIONAL] in serial mode

INFO  : Stage-4 is selected by condition resolver.

INFO  : Stage-3 is filtered out by condition resolver.

INFO  : Stage-5 is filtered out by condition resolver.

INFO  : Starting task [Stage-4:MOVE] in serial mode

INFO  : Moving data to: hdfs://netf-pmbigdat02.epmtelco.com.co:8020/user/data/.hive-staging\_hive\_2017-08-24\_11-55-43\_212\_6617139602049464738-4/-ext-10000 from hdfs://netf-pmbigdat02.epmtelco.com.co:8020/user/data/.hive-staging\_hive\_2017-08-24\_11-55-43\_212\_6617139602049464738-4/-ext-10002

INFO  : Starting task [Stage-0:MOVE] in serial mode

INFO  : Loading data to table default.school from hdfs://netf-pmbigdat02.epmtelco.com.co:8020/user/data/.hive-staging\_hive\_2017-08-24\_11-55-43\_212\_6617139602049464738-4/-ext-10000

INFO  : Starting task [Stage-2:STATS] in serial mode

INFO  : Table default.school stats: [numFiles=4, numRows=3, totalSize=40, rawDataSize=27]

INFO  : MapReduce Jobs Launched:

INFO  : Stage-Stage-1: Map: 1   Cumulative CPU: 1.87 sec   HDFS Read: 3415 HDFS Write: 80 SUCCESS

INFO  : Total MapReduce CPU Time Spent: 1 seconds 870 msec

INFO  : Completed executing command(queryId=hive\_20170824115555\_56fb64b2-f183-4735-87ed-b00768cc9e5b); Time taken: 11.043 seconds

INFO  : OK

No rows affected (11.155 seconds)

Spark Validation

Command :

cd /opt/cloudera/parcels/CDH-5.8.1.cdh5.8.4.p0.5/lib/spark/examples/

spark-submit --class org.apache.spark.examples.SparkPi --master yarn-cluster --num-executors 1 --driver-memory 512m --executor-memory 512m --executor-cores 1 lib/spark-examples-1.6.0-cdh5.8.4-hadoop2.6.0-cdh5.8.4.jar 10

Output:

17/08/24 11:46:30 INFO client.RMProxy: Connecting to ResourceManager at netf-pmbigdat02.epmtelco.com.co/10.100.74.42:8032

17/08/24 11:46:30 INFO yarn.Client: Requesting a new application from cluster with 5 NodeManagers

17/08/24 11:46:30 INFO yarn.Client: Verifying our application has not requested more than the maximum memory capability of the cluster (79159 MB per container)

17/08/24 11:46:30 INFO yarn.Client: Will allocate AM container, with 896 MB memory including 384 MB overhead

17/08/24 11:46:30 INFO yarn.Client: Setting up container launch context for our AM

17/08/24 11:46:30 INFO yarn.Client: Setting up the launch environment for our AM container

17/08/24 11:46:30 INFO yarn.Client: Preparing resources for our AM container

17/08/24 11:46:31 INFO yarn.Client: Uploading resource file:/opt/cloudera/parcels/CDH-5.8.4-1.cdh5.8.4.p0.5/lib/spark/examples/lib/spark-examples-1.6.0-cdh5.8.4-hadoop2.6.0-cdh5.8.4.jar -> hdfs://netf-pmbigdat02.epmtelco.com.co:8020/user/hdfs/.sparkStaging/application\_1503581872672\_0008/spark-examples-1.6.0-cdh5.8.4-hadoop2.6.0-cdh5.8.4.jar

17/08/24 11:46:31 INFO yarn.Client: Uploading resource file:/tmp/spark-016bebc3-cdc9-4066-a0ee-ca03710580cd/\_\_spark\_conf\_\_8310748046160292180.zip -> hdfs://netf-pmbigdat02.epmtelco.com.co:8020/user/hdfs/.sparkStaging/application\_1503581872672\_0008/\_\_spark\_conf\_\_8310748046160292180.zip

17/08/24 11:46:31 INFO spark.SecurityManager: Changing view acls to: hdfs

17/08/24 11:46:31 INFO spark.SecurityManager: Changing modify acls to: hdfs

17/08/24 11:46:31 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(hdfs); users with modify permissions: Set(hdfs)

17/08/24 11:46:31 INFO yarn.Client: Submitting application 8 to ResourceManager

17/08/24 11:46:31 INFO impl.YarnClientImpl: Submitted application application\_1503581872672\_0008

17/08/24 11:46:32 INFO yarn.Client: Application report for application\_1503581872672\_0008 (state: ACCEPTED)

17/08/24 11:46:32 INFO yarn.Client:

client token: N/A

diagnostics: N/A

ApplicationMaster host: N/A

ApplicationMaster RPC port: -1

queue: root.users.hdfs

start time: 1503593191725

final status: UNDEFINED

tracking URL: http://netf-pmbigdat02.epmtelco.com.co:8088/proxy/application\_1503581872672\_0008/

user: hdfs

17/08/24 11:46:33 INFO yarn.Client: Application report for application\_1503581872672\_0008 (state: ACCEPTED)

17/08/24 11:46:34 INFO yarn.Client: Application report for application\_1503581872672\_0008 (state: ACCEPTED)

17/08/24 11:46:35 INFO yarn.Client: Application report for application\_1503581872672\_0008 (state: RUNNING)

17/08/24 11:46:35 INFO yarn.Client:

client token: N/A

diagnostics: N/A

ApplicationMaster host: 10.100.74.38

ApplicationMaster RPC port: 0

queue: root.users.hdfs

start time: 1503593191725

final status: UNDEFINED

tracking URL: http://netf-pmbigdat02.epmtelco.com.co:8088/proxy/application\_1503581872672\_0008/

user: hdfs

17/08/24 11:46:36 INFO yarn.Client: Application report for application\_1503581872672\_0008 (state: RUNNING)

17/08/24 11:46:37 INFO yarn.Client: Application report for application\_1503581872672\_0008 (state: RUNNING)

17/08/24 11:46:38 INFO yarn.Client: Application report for application\_1503581872672\_0008 (state: RUNNING)

17/08/24 11:46:39 INFO yarn.Client: Application report for application\_1503581872672\_0008 (state: RUNNING)

17/08/24 11:46:40 INFO yarn.Client: Application report for application\_1503581872672\_0008 (state: FINISHED)

17/08/24 11:46:40 INFO yarn.Client:

client token: N/A

diagnostics: N/A

ApplicationMaster host: 10.100.74.38

ApplicationMaster RPC port: 0

queue: root.users.hdfs

start time: 1503593191725

final status: SUCCEEDED

tracking URL: http://netf-pmbigdat02.epmtelco.com.co:8088/proxy/application\_1503581872672\_0008/history/application\_1503581872672\_0008/1

user: hdfs

17/08/24 11:46:40 INFO util.ShutdownHookManager: Shutdown hook called

17/08/24 11:46:40 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-016bebc3-cdc9-4066-a0ee-ca03710580cd

SUCCESS