

Mod13_Sesion2_SQL_Spark

September 23, 2023

```
[1]: spark
```

```
[1]: <pyspark.sql.session.SparkSession at 0x7f09d9b21960>
```

```
[2]: df = spark.read.format("jdbc")\
      .option("url","jdbc:mysql://bd.arcelia.net/datosabiertos")\
      .option("driver","com.mysql.cj.jdbc.Driver")\
      .option("query", "select ORIGEN, SECTOR, ENTIDAD_UM, SEXO, EDAD, \
      ↪YEAR(FECHA_INGRESO) ANIO, MONTH(FECHA_INGRESO) MES FROM COVID19MEXICO")\
      .option("user","usabierto01")\
      .option("password","datos21%")\
      .load()
```

```
[3]: df.printSchema()
```

```
root
 |-- ORIGEN: integer (nullable = true)
 |-- SECTOR: integer (nullable = true)
 |-- ENTIDAD_UM: string (nullable = true)
 |-- SEXO: integer (nullable = true)
 |-- EDAD: integer (nullable = true)
 |-- ANIO: long (nullable = true)
 |-- MES: long (nullable = true)
```

```
[4]: df.write.partitionBy("EDAD").save("/tmp/dcd/particion/covid/edad")
```

```
23/09/23 01:14:57 WARN YarnAllocator: Container from a bad node:
container_1695421740761_0005_01_000002 on host: cluster-sir-dcd05-w-1.us-
central1-a.c.carbide-trees-398419.internal. Exit status: 143. Diagnostics:
[2023-09-23 01:14:56.377]Container killed on request. Exit code is 143
[2023-09-23 01:14:56.377]Container exited with a non-zero exit code 143.
[2023-09-23 01:14:56.378]Killed by external signal
```

```
.
23/09/23 01:14:57 WARN YarnSchedulerBackend$YarnSchedulerEndpoint: Requesting
driver to remove executor 2 for reason Container from a bad node:
container_1695421740761_0005_01_000002 on host: cluster-sir-dcd05-w-1.us-
central1-a.c.carbide-trees-398419.internal. Exit status: 143. Diagnostics:
[2023-09-23 01:14:56.377]Container killed on request. Exit code is 143
```

[2023-09-23 01:14:56.377]Container exited with a non-zero exit code 143.
 [2023-09-23 01:14:56.378]Killed by external signal

.

23/09/23 01:14:57 ERROR YarnScheduler: Lost executor 2 on cluster-sir-dcd05-w-1.us-central1-a.c.carbide-trees-398419.internal: Container from a bad node: container_1695421740761_0005_01_000002 on host: cluster-sir-dcd05-w-1.us-central1-a.c.carbide-trees-398419.internal. Exit status: 143. Diagnostics: [2023-09-23 01:14:56.377]Container killed on request. Exit code is 143
 [2023-09-23 01:14:56.377]Container exited with a non-zero exit code 143.
 [2023-09-23 01:14:56.378]Killed by external signal

.

23/09/23 01:14:57 WARN TaskSetManager: Lost task 0.0 in stage 0.0 (TID 0) (cluster-sir-dcd05-w-1.us-central1-a.c.carbide-trees-398419.internal executor 2): ExecutorLostFailure (executor 2 exited caused by one of the running tasks) Reason: Container from a bad node: container_1695421740761_0005_01_000002 on host: cluster-sir-dcd05-w-1.us-central1-a.c.carbide-trees-398419.internal. Exit status: 143. Diagnostics: [2023-09-23 01:14:56.377]Container killed on request. Exit code is 143
 [2023-09-23 01:14:56.377]Container exited with a non-zero exit code 143.
 [2023-09-23 01:14:56.378]Killed by external signal

.

23/09/23 01:16:01 WARN YarnAllocator: Container from a bad node: container_1695421740761_0005_01_000003 on host: cluster-sir-dcd05-w-0.us-central1-a.c.carbide-trees-398419.internal. Exit status: 143. Diagnostics: [2023-09-23 01:15:59.461]Container killed on request. Exit code is 143
 [2023-09-23 01:15:59.462]Container exited with a non-zero exit code 143.
 [2023-09-23 01:15:59.462]Killed by external signal

.

23/09/23 01:16:01 WARN YarnSchedulerBackend\$YarnSchedulerEndpoint: Requesting driver to remove executor 3 for reason Container from a bad node: container_1695421740761_0005_01_000003 on host: cluster-sir-dcd05-w-0.us-central1-a.c.carbide-trees-398419.internal. Exit status: 143. Diagnostics: [2023-09-23 01:15:59.461]Container killed on request. Exit code is 143
 [2023-09-23 01:15:59.462]Container exited with a non-zero exit code 143.
 [2023-09-23 01:15:59.462]Killed by external signal

.

23/09/23 01:16:01 ERROR YarnScheduler: Lost executor 3 on cluster-sir-dcd05-w-0.us-central1-a.c.carbide-trees-398419.internal: Container from a bad node: container_1695421740761_0005_01_000003 on host: cluster-sir-dcd05-w-0.us-central1-a.c.carbide-trees-398419.internal. Exit status: 143. Diagnostics: [2023-09-23 01:15:59.461]Container killed on request. Exit code is 143
 [2023-09-23 01:15:59.462]Container exited with a non-zero exit code 143.
 [2023-09-23 01:15:59.462]Killed by external signal

.

23/09/23 01:16:01 WARN TaskSetManager: Lost task 0.1 in stage 0.0 (TID 1) (cluster-sir-dcd05-w-0.us-central1-a.c.carbide-trees-398419.internal executor 3): ExecutorLostFailure (executor 3 exited caused by one of the running tasks) Reason: Container from a bad node: container_1695421740761_0005_01_000003 on

```

host: cluster-sir-dcd05-w-0.us-central1-a.c.carbide-trees-398419.internal. Exit
status: 143. Diagnostics: [2023-09-23 01:15:59.461]Container killed on request.
Exit code is 143
[2023-09-23 01:15:59.462]Container exited with a non-zero exit code 143.
[2023-09-23 01:15:59.462]Killed by external signal
.
ERROR:root:KeyboardInterrupt while sending command. (0 + 1) / 1]
Traceback (most recent call last):
  File "/usr/lib/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/java_gateway.py",
line 1038, in send_command
    response = connection.send_command(command)
  File "/usr/lib/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/clientserver.py",
line 511, in send_command
    answer = smart_decode(self.stream.readline()[:-1])
  File "/opt/conda/miniconda3/lib/python3.10/socket.py", line 705, in readinto
    return self._sock.recv_into(b)
KeyboardInterrupt

```

```

-----
KeyboardInterrupt                                Traceback (most recent call last)
/tmp/ipykernel_24129/1470152337.py in <cell line: 1>()
----> 1 df.write.partitionBy("EDAD").save("/tmp/dcd/particion/covid/edad")

/usr/lib/spark/python/pyspark/sql/readwriter.py in save(self, path, format,
↳ mode, partitionBy, **options)
    966         self._jwrite.save()
    967     else:
--> 968         self._jwrite.save(path)
    969
    970     @since(1.4)

/usr/lib/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/java_gateway.py in
↳ __call__(self, *args)
    1318         proto.END_COMMAND_PART
    1319
-> 1320         answer = self.gateway_client.send_command(command)
    1321         return_value = get_return_value(
    1322             answer, self.gateway_client, self.target_id, self.name)

/usr/lib/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/java_gateway.py in
↳ send_command(self, command, retry, binary)
    1036         connection = self._get_connection()
    1037         try:
-> 1038             response = connection.send_command(command)
    1039             if binary:
    1040                 return response, self.
↳ _create_connection_guard(connection)

```

```

/usr/lib/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/clientserver.py in
    ↪send_command(self, command)
    509         try:
    510             while True:
--> 511                 answer = smart_decode(self.stream.readline()[:-1])
    512                 logger.debug("Answer received: {0}".format(answer))
    513                 # Happens when a the other end is dead. There might be
    ↪an empty

/opt/conda/miniconda3/lib/python3.10/socket.py in readinto(self, b)
    703         while True:
    704             try:
--> 705                 return self._sock.recv_into(b)
    706             except timeout:
    707                 self._timeout_occurred = True

```

KeyboardInterrupt:

[]:

```

[5]: df = spark.read.format("jdbc")\
    .option("url","jdbc:mysql://bd.arcelia.net/datosabiertos")\
    .option("driver","com.mysql.cj.jdbc.Driver")\
    .option("query", "select ORIGEN, SECTOR, ENTIDAD_UM, SEXO, EDAD,\
    ↪YEAR(FECHA_INGRESO) ANIO, MONTH(FECHA_INGRESO) MES FROM COVID19MEXICO LIMIT\
    ↪100000")\
    .option("user","usabierto01")\
    .option("password","datos21%")\
    .load()

```

[Stage 0:>

(0 + 1) / 1]

```

[6]: df.write.partitionBy("EDAD").save("/tmp/dcd/particion/covid/edad")

```

```

23/09/23 01:19:52 WARN YarnAllocator: Container from a bad node:
container_1695421740761_0005_01_000005 on host: cluster-sir-dcd05-w-0.us-
central1-a.c.carbide-trees-398419.internal. Exit status: 143. Diagnostics:
[2023-09-23 01:19:51.141]Container killed on request. Exit code is 143
[2023-09-23 01:19:51.141]Container exited with a non-zero exit code 143.
[2023-09-23 01:19:51.141]Killed by external signal
.

```

```

23/09/23 01:19:52 WARN YarnSchedulerBackend$YarnSchedulerEndpoint: Requesting
driver to remove executor 5 for reason Container from a bad node:
container_1695421740761_0005_01_000005 on host: cluster-sir-dcd05-w-0.us-
central1-a.c.carbide-trees-398419.internal. Exit status: 143. Diagnostics:
[2023-09-23 01:19:51.141]Container killed on request. Exit code is 143
[2023-09-23 01:19:51.141]Container exited with a non-zero exit code 143.

```

```

[2023-09-23 01:19:51.141]Killed by external signal
.
23/09/23 01:19:52 ERROR YarnScheduler: Lost executor 5 on cluster-sir-
dcd05-w-0.us-central1-a.c.carbide-trees-398419.internal: Container from a bad
node: container_1695421740761_0005_01_000005 on host: cluster-sir-dcd05-w-0.us-
central1-a.c.carbide-trees-398419.internal. Exit status: 143. Diagnostics:
[2023-09-23 01:19:51.141]Container killed on request. Exit code is 143
[2023-09-23 01:19:51.141]Container exited with a non-zero exit code 143.
[2023-09-23 01:19:51.141]Killed by external signal
.
23/09/23 01:19:52 WARN TaskSetManager: Lost task 0.3 in stage 0.0 (TID 3)
(cluster-sir-dcd05-w-0.us-central1-a.c.carbide-trees-398419.internal executor
5): ExecutorLostFailure (executor 5 exited caused by one of the running tasks)
Reason: Container from a bad node: container_1695421740761_0005_01_000005 on
host: cluster-sir-dcd05-w-0.us-central1-a.c.carbide-trees-398419.internal. Exit
status: 143. Diagnostics: [2023-09-23 01:19:51.141]Container killed on request.
Exit code is 143
[2023-09-23 01:19:51.141]Container exited with a non-zero exit code 143.
[2023-09-23 01:19:51.141]Killed by external signal
.
23/09/23 01:19:52 ERROR TaskSetManager: Task 0 in stage 0.0 failed 4 times;
aborting job
23/09/23 01:19:52 ERROR FileFormatWriter: Aborting job
7d8f3c9d-1fc2-4551-a8dd-f0d834450b7e.
org.apache.spark.SparkException: Job aborted due to stage failure: Task 0 in
stage 0.0 failed 4 times, most recent failure: Lost task 0.3 in stage 0.0 (TID
3) (cluster-sir-dcd05-w-0.us-central1-a.c.carbide-trees-398419.internal executor
5): ExecutorLostFailure (executor 5 exited caused by one of the running tasks)
Reason: Container from a bad node: container_1695421740761_0005_01_000005 on
host: cluster-sir-dcd05-w-0.us-central1-a.c.carbide-trees-398419.internal. Exit
status: 143. Diagnostics: [2023-09-23 01:19:51.141]Container killed on request.
Exit code is 143
[2023-09-23 01:19:51.141]Container exited with a non-zero exit code 143.
[2023-09-23 01:19:51.141]Killed by external signal
.
Driver stacktrace:
    at org.apache.spark.scheduler.DAGScheduler.failJobAndIndependentStages(D
AGScheduler.scala:2717) ~[spark-core_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.scheduler.DAGScheduler.$anonfun$abortStage$2(DAGSche
duler.scala:2653) ~[spark-core_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.scheduler.DAGScheduler.$anonfun$abortStage$2$adapted
(DAGScheduler.scala:2652) ~[spark-core_2.12-3.3.2.jar:3.3.2]
    at
scala.collection.mutable.ResizableArray.foreach(ResizableArray.scala:62)
~[scala-library-2.12.14.jar:?]
    at
scala.collection.mutable.ResizableArray.foreach$(ResizableArray.scala:55)
~[scala-library-2.12.14.jar:?]

```

```

        at scala.collection.mutable.ArrayBuffer.foreach(ArrayBuffer.scala:49)
~[scala-library-2.12.14.jar:?]
        at
org.apache.spark.scheduler.DAGScheduler.abortStage(DAGScheduler.scala:2652)
~[spark-core_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.scheduler.DAGScheduler.$anonfun$handleTaskSetFailed$
1(DAGScheduler.scala:1189) ~[spark-core_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.scheduler.DAGScheduler.$anonfun$handleTaskSetFailed$
1$adapted(DAGScheduler.scala:1189) ~[spark-core_2.12-3.3.2.jar:3.3.2]
        at scala.Option.foreach(Option.scala:407) ~[scala-library-2.12.14.jar:?]
        at org.apache.spark.scheduler.DAGScheduler.handleTaskSetFailed(DAGSchedu
ler.scala:1189) ~[spark-core_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.doOnReceive(D
AGScheduler.scala:2905) ~[spark-core_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.onReceive(DAG
Scheduler.scala:2847) ~[spark-core_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.onReceive(DAG
Scheduler.scala:2836) ~[spark-core_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.util.EventLoop$$anon$1.run(EventLoop.scala:49)
~[spark-core_2.12-3.3.2.jar:3.3.2]
        at
org.apache.spark.scheduler.DAGScheduler.runJob(DAGScheduler.scala:959) ~[spark-
core_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.SparkContext.runJob(SparkContext.scala:2246)
~[spark-core_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.sql.execution.datasources.FileFormatWriter$.write(Fi
leFormatWriter.scala:257) ~[spark-sql_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.sql.execution.datasources.InsertIntoHadoopFsRelation
Command.run(InsertIntoHadoopFsRelationCommand.scala:186) ~[spark-
sql_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.sql.execution.command.DataWritingCommandExec.sideEff
ectResult$lzycompute(commands.scala:113) ~[spark-sql_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.sql.execution.command.DataWritingCommandExec.sideEff
ectResult(commands.scala:111) ~[spark-sql_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.sql.execution.command.DataWritingCommandExec.execute
Collect(commands.scala:125) ~[spark-sql_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.sql.execution.QueryExecution$$anonfun$eagerlyExecute
Commands$1.$anonfun$applyOrElse$1(QueryExecution.scala:98) ~[spark-
sql_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.sql.execution.SQLExecution$. $anonfun$withNewExecutio
nId$6(SQLExecution.scala:109) ~[spark-sql_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.sql.execution.SQLExecution$.withSQLConfPropagated(SQ
LExecution.scala:169) ~[spark-sql_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.sql.execution.SQLExecution$. $anonfun$withNewExecutio
nId$1(SQLExecution.scala:95) ~[spark-sql_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.sql.SparkSession.withActive(SparkSession.scala:779)
~[spark-sql_2.12-3.3.2.jar:3.3.2]
        at org.apache.spark.sql.execution.SQLExecution$.withNewExecutionId(SQLEx

```

```

ecution.scala:64) ~[spark-sql_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.execution.QueryExecution$$anonfun$eagerlyExecute
Commands$1.applyOrElse(QueryExecution.scala:98) ~[spark-
sql_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.execution.QueryExecution$$anonfun$eagerlyExecute
Commands$1.applyOrElse(QueryExecution.scala:94) ~[spark-
sql_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.catalyst.trees.TreeNode.$anonfun$transformDownWi
thPruning$1(TreeNode.scala:584) ~[spark-catalyst_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.catalyst.trees.CurrentOrigin$.withOrigin(TreeNod
e.scala:176) ~[spark-catalyst_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.catalyst.trees.TreeNode.transformDownWithPruning
(TreeNode.scala:584) ~[spark-catalyst_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.catalyst.plans.logical.LogicalPlan.org$apache$sp
ark$sql$catalyst$plans$logical$AnalysisHelper$$super$transformDownWithPruning(Lo
gicalPlan.scala:30) ~[spark-catalyst_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.catalyst.plans.logical.AnalysisHelper.transformD
ownWithPruning(AnalysisHelper.scala:267) ~[spark-catalyst_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.catalyst.plans.logical.AnalysisHelper.transformD
ownWithPruning$(AnalysisHelper.scala:263) ~[spark-catalyst_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.catalyst.plans.logical.LogicalPlan.transformDown
WithPruning(LogicalPlan.scala:30) ~[spark-catalyst_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.catalyst.plans.logical.LogicalPlan.transformDown
WithPruning(LogicalPlan.scala:30) ~[spark-catalyst_2.12-3.3.2.jar:3.3.2]
    at
org.apache.spark.sql.catalyst.trees.TreeNode.transformDown(TreeNode.scala:560)
~[spark-catalyst_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.execution.QueryExecution.eagerlyExecuteCommands(
QueryExecution.scala:94) ~[spark-sql_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.execution.QueryExecution.commandExecuted$lzycomp
ute(QueryExecution.scala:81) ~[spark-sql_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.execution.QueryExecution.commandExecuted(QueryEx
ecution.scala:79) ~[spark-sql_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.execution.QueryExecution.assertCommandExecuted(Q
ueryExecution.scala:116) ~[spark-sql_2.12-3.3.2.jar:3.3.2]
    at
org.apache.spark.sql.DataFrameWriter.runCommand(DataFrameWriter.scala:860)
~[spark-sql_2.12-3.3.2.jar:3.3.2]
    at
org.apache.spark.sql.DataFrameWriter.saveToV1Source(DataFrameWriter.scala:390)
~[spark-sql_2.12-3.3.2.jar:3.3.2]
    at
org.apache.spark.sql.DataFrameWriter.saveInternal(DataFrameWriter.scala:363)
~[spark-sql_2.12-3.3.2.jar:3.3.2]
    at org.apache.spark.sql.DataFrameWriter.save(DataFrameWriter.scala:239)
~[spark-sql_2.12-3.3.2.jar:3.3.2]
    at jdk.internal.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
~[?:?]

```

```

        at jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAcce
ssorImpl.java:62) ~[?:?]
        at jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMe
thodAccessorImpl.java:43) ~[?:?]
        at java.lang.reflect.Method.invoke(Method.java:566) ~[?:?]
        at py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)
~[py4j-0.10.9.5.jar:?]
        at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.java:357)
~[py4j-0.10.9.5.jar:?]
        at py4j.Gateway.invoke(Gateway.java:282) ~[py4j-0.10.9.5.jar:?]
        at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.java:132)
~[py4j-0.10.9.5.jar:?]
        at py4j.commands.CallCommand.execute(CallCommand.java:79)
~[py4j-0.10.9.5.jar:?]
        at
py4j.ClientServerConnection.waitForCommands(ClientServerConnection.java:182)
~[py4j-0.10.9.5.jar:?]
        at py4j.ClientServerConnection.run(ClientServerConnection.java:106)
~[py4j-0.10.9.5.jar:?]
        at java.lang.Thread.run(Thread.java:829) ~[?:?]
23/09/23 01:19:52 WARN TaskSetManager: Lost task 0.0 in stage 1.0 (TID 4)
(cluster-sir-dcd05-w-0.us-centrall1-a.c.carbide-trees-398419.internal executor
6): org.apache.spark.SparkException: Task failed while writing rows.
        at org.apache.spark.sql.errors.QueryExecutionErrors$.taskFailedWhileWrit
ingRowsError(QueryExecutionErrors.scala:654)
        at org.apache.spark.sql.execution.datasources.FileFormatWriter$.executeT
ask(FileFormatWriter.scala:379)
        at org.apache.spark.sql.execution.datasources.FileFormatWriter$.anonfun
$write$22(FileFormatWriter.scala:268)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:136)
        at
org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Executor.scala:548)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1504)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:551)
        at java.base/java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoo
lExecutor.java:1128)
        at java.base/java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPo
olExecutor.java:628)
        at java.base/java.lang.Thread.run(Thread.java:829)
Caused by: org.apache.hadoop.ipc.RemoteException(java.io.FileNotFoundException):
File does not exist: /tmp/dcd/particion/covid/edad/_temporary/0/_temporary/attempt_2023092301193022057684773912605_0001_m_000000_4/EDAD=0/part-00000-119d3e04-
dfc3-4357-a615-154912f6bec0.c000.snappy.parquet (inode 16614) Holder
DFSCClient_attempt_2023092301193022057684773912605_0001_m_000000_4_162004577_31
does not have any open files.
        at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkLease(FSName
system.java:3071)

```



```

        at org.apache.hadoop.hdfs.server.namenode.FSDirWriteFileOp.completeFileInternal(FSDirWriteFileOp.java:704)
        at org.apache.hadoop.hdfs.server.namenode.FSDirWriteFileOp.completeFile(FSDirWriteFileOp.java:690)
        at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.completeFile(FSNamesystem.java:3115)
        at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.complete(NameNodeRpcServer.java:959)
        at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolServerSideTranslatorPB.complete(ClientNamenodeProtocolServerSideTranslatorPB.java:639)
        at org.apache.hadoop.hdfs.protocol.proto.ClientNamenodeProtocolProtos$ClientNamenodeProtocol$2.callBlockingMethod(ClientNamenodeProtocolProtos.java)
        at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:604)
        at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:572)
        at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:556)
        at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:1093)
        at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:1043)
        at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:971)
        at java.base/java.security.AccessController.doPrivileged(Native Method)
        at java.base/javax.security.auth.Subject.doAs(Subject.java:423)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1878)
        at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2976)

        at org.apache.hadoop.ipc.Client.getRpcResponse(Client.java:1612)
        at org.apache.hadoop.ipc.Client.call(Client.java:1558)
        at org.apache.hadoop.ipc.Client.call(Client.java:1455)
        at org.apache.hadoop.ipc.ProtobufRpcEngine2$Invoker.invoke(ProtobufRpcEngine2.java:242)
        at org.apache.hadoop.ipc.ProtobufRpcEngine2$Invoker.invoke(ProtobufRpcEngine2.java:129)
        at com.sun.proxy.$Proxy32.complete(Unknown Source)
        at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.complete(ClientNamenodeProtocolTranslatorPB.java:570)
        at
        java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.base/java.lang.reflect.Method.invoke(Method.java:566)
        at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:422)
        at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeMethod(RetryInvocationHandler.java:165)

```

```

        at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invoke(RetryIn
vocationHandler.java:157)
        at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeOnce(Ret
ryInvocationHandler.java:95)
        at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocat
ionHandler.java:359)
        at com.sun.proxy.$Proxy33.complete(Unknown Source)
        at
org.apache.hadoop.hdfs.DFSOutputStream.completeFile(DFSOutputStream.java:952)
        at
org.apache.hadoop.hdfs.DFSOutputStream.completeFile(DFSOutputStream.java:909)
        at
org.apache.hadoop.hdfs.DFSOutputStream.closeImpl(DFSOutputStream.java:892)
        at
org.apache.hadoop.hdfs.DFSOutputStream.close(DFSOutputStream.java:847)
        at org.apache.hadoop.fs.FSDataOutputStream$PositionCache.close(FSDataOut
putStream.java:77)
        at
org.apache.hadoop.fs.FSDataOutputStream.close(FSDataOutputStream.java:106)
        at org.apache.parquet.hadoop.util.HadoopPositionOutputStream.close(Hadoo
pPositionOutputStream.java:64)
        at
org.apache.parquet.hadoop.ParquetFileWriter.end(ParquetFileWriter.java:1106)
        at org.apache.parquet.hadoop.InternalParquetRecordWriter.close(InternalP
arquetRecordWriter.java:132)
        at org.apache.parquet.hadoop.ParquetRecordWriter.close(ParquetRecordWrit
er.java:164)
        at org.apache.spark.sql.execution.datasources.parquet.ParquetOutputWrite
r.close(ParquetOutputWriter.scala:41)
        at org.apache.spark.sql.execution.datasources.FileFormatDataWriter.relea
seCurrentWriter(FileFormatDataWriter.scala:64)
        at org.apache.spark.sql.execution.datasources.BaseDynamicPartitionDataWr
iter.renewCurrentWriter(FileFormatDataWriter.scala:267)
        at org.apache.spark.sql.execution.datasources.DynamicPartitionDataSingle
Writer.write(FileFormatDataWriter.scala:365)
        at org.apache.spark.sql.execution.datasources.FileFormatDataWriter.write
WithMetrics(FileFormatDataWriter.scala:85)
        at org.apache.spark.sql.execution.datasources.FileFormatDataWriter.write
WithIterator(FileFormatDataWriter.scala:92)
        at org.apache.spark.sql.execution.datasources.FileFormatWriter$.anonfun
$executeTask$1(FileFormatWriter.scala:362)
        at org.apache.spark.util.Utils$.tryWithSafeFinallyAndFailureCallbacks(Ut
ils.scala:1538)
        at org.apache.spark.sql.execution.datasources.FileFormatWriter$.executeT
ask(FileFormatWriter.scala:369)
        ... 9 more

```

```
[10]: ! hdfs dfs -ls /tmp/dcd/particion/covid/edad
```

Found 107 items

```
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=0
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=1
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=10
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=100
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=101
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=102
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=103
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=105
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=108
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=11
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=12
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=13
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=14
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=15
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=16
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=17
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=18
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=19
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=2
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=20
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=21
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=22
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
```

```

/tmp/dcd/particion/covid/edad/EDAD=23
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=24
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=25
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=26
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=27
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=28
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=29
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=3
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=30
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=31
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=32
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=33
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=34
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=35
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=36
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=37
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=38
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=39
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=4
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=40
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=41
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=42
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=43
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=44
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19

```

```

/tmp/dcd/particion/covid/edad/EDAD=45
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=46
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=47
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=48
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=49
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=5
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=50
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=51
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=52
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=53
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=54
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=55
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=56
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=57
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=58
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=59
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=6
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=60
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=61
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=62
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=63
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=64
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=65
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=66
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19

```

```

/tmp/dcd/particion/covid/edad/EDAD=67
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=68
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=69
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=7
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=70
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=71
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=72
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=73
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=74
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=75
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=76
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=77
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=78
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=79
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=8
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=80
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=81
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=82
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=83
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=84
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=85
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=86
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=87
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=88
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19

```

```

/tmp/dcd/particion/covid/edad/EDAD=89
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=9
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=90
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=91
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=92
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=93
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=94
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=95
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=96
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=97
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=98
drwxr-xr-x  - root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/EDAD=99
-rw-r--r--  2 root hadoop          0 2023-09-23 01:19
/tmp/dcd/particion/covid/edad/_SUCCESS

```

[]:

```
[11]: df.write.partitionBy("EDAD").save("gs://dcd05-sir-bucket/dcd/particion/covid/
      ↪edad")
```

```
[12]: df.write.partitionBy("ANIO", "MES").save("hdfs:///tmp/dcd/particion/covid/
      ↪anio_mes")
```

```
[13]: df.write.partitionBy("ANIO", "MES").save("gs://dcd05-sir-bucket/dcd/particion/
      ↪covid/anio_mes")
```

```
[14]: ! hdfs dfs -ls -R /tmp/dcd/particion/covid/anio_mes
```

```

drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2020
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2020/MES=1
-rw-r--r--  2 root hadoop        23181 2023-09-23 01:35 /tmp/dcd/particion/covid/

```

```

anio_mes/ANIO=2020/MES=1/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2020/MES=10
-rw-r--r--   2 root hadoop        1752 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2020/MES=10/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2020/MES=11
-rw-r--r--   2 root hadoop        1611 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2020/MES=11/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2020/MES=12
-rw-r--r--   2 root hadoop        1709 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2020/MES=12/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2020/MES=2
-rw-r--r--   2 root hadoop       23508 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2020/MES=2/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2020/MES=3
-rw-r--r--   2 root hadoop       62946 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2020/MES=3/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2020/MES=4
-rw-r--r--   2 root hadoop      105391 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2020/MES=4/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2020/MES=5
-rw-r--r--   2 root hadoop        1906 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2020/MES=5/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2020/MES=6
-rw-r--r--   2 root hadoop        1944 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2020/MES=6/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2020/MES=7
-rw-r--r--   2 root hadoop        1823 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2020/MES=7/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35

```



```

/tmp/dcd/particion/covid/anio_mes/ANIO=2020/MES=8
-rw-r--r--  2 root hadoop      1792 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2020/MES=8/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2020/MES=9
-rw-r--r--  2 root hadoop      1669 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2020/MES=9/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2021
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2021/MES=1
-rw-r--r--  2 root hadoop      1704 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2021/MES=1/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2021/MES=10
-rw-r--r--  2 root hadoop      1440 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2021/MES=10/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2021/MES=2
-rw-r--r--  2 root hadoop      1617 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2021/MES=2/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2021/MES=3
-rw-r--r--  2 root hadoop      1425 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2021/MES=3/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2021/MES=4
-rw-r--r--  2 root hadoop      1440 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2021/MES=4/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
drwxr-xr-x  - root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/ANIO=2021/MES=8
-rw-r--r--  2 root hadoop      1499 2023-09-23 01:35 /tmp/dcd/particion/covid/
anio_mes/ANIO=2021/MES=8/part-00000-da024029-3402-4b87-ba62-
a5aae6b2b8f9.c000.snappy.parquet
-rw-r--r--  2 root hadoop          0 2023-09-23 01:35
/tmp/dcd/particion/covid/anio_mes/_SUCCESS

```

```
[15]: df.write.partitionBy("MES").save("hdfs:///tmp/dcd/particion/covid/mes")
```

```
[16]: df.write.partitionBy("MES").save("gs://dcd05-sir-bucket/dcd/particion/covid/
↪mes")
```

```
[17]: ! hdfs dfs -ls -R /tmp/dcd/particion/covid/mes
```

```
drwxr-xr-x   - root hadoop           0 2023-09-23 01:49
/tmp/dcd/particion/covid/mes/MES=1
-rw-r--r--   2 root hadoop       23600 2023-09-23 01:49 /tmp/dcd/particion/covid/
mes/MES=1/part-00000-a0e82301-1324-4d0f-88a2-95dc36d52732.c000.snappy.parquet
drwxr-xr-x   - root hadoop           0 2023-09-23 01:49
/tmp/dcd/particion/covid/mes/MES=10
-rw-r--r--   2 root hadoop        2036 2023-09-23 01:49 /tmp/dcd/particion/covid/
mes/MES=10/part-00000-a0e82301-1324-4d0f-88a2-95dc36d52732.c000.snappy.parquet
drwxr-xr-x   - root hadoop           0 2023-09-23 01:49
/tmp/dcd/particion/covid/mes/MES=11
-rw-r--r--   2 root hadoop       1877 2023-09-23 01:49 /tmp/dcd/particion/covid/
mes/MES=11/part-00000-a0e82301-1324-4d0f-88a2-95dc36d52732.c000.snappy.parquet
drwxr-xr-x   - root hadoop           0 2023-09-23 01:49
/tmp/dcd/particion/covid/mes/MES=12
-rw-r--r--   2 root hadoop       1975 2023-09-23 01:49 /tmp/dcd/particion/covid/
mes/MES=12/part-00000-a0e82301-1324-4d0f-88a2-95dc36d52732.c000.snappy.parquet
drwxr-xr-x   - root hadoop           0 2023-09-23 01:49
/tmp/dcd/particion/covid/mes/MES=2
-rw-r--r--   2 root hadoop      23810 2023-09-23 01:49 /tmp/dcd/particion/covid/
mes/MES=2/part-00000-a0e82301-1324-4d0f-88a2-95dc36d52732.c000.snappy.parquet
drwxr-xr-x   - root hadoop           0 2023-09-23 01:49
/tmp/dcd/particion/covid/mes/MES=3
-rw-r--r--   2 root hadoop     63303 2023-09-23 01:49 /tmp/dcd/particion/covid/
mes/MES=3/part-00000-a0e82301-1324-4d0f-88a2-95dc36d52732.c000.snappy.parquet
drwxr-xr-x   - root hadoop           0 2023-09-23 01:49
/tmp/dcd/particion/covid/mes/MES=4
-rw-r--r--   2 root hadoop    105840 2023-09-23 01:49 /tmp/dcd/particion/covid/
mes/MES=4/part-00000-a0e82301-1324-4d0f-88a2-95dc36d52732.c000.snappy.parquet
drwxr-xr-x   - root hadoop           0 2023-09-23 01:49
/tmp/dcd/particion/covid/mes/MES=5
-rw-r--r--   2 root hadoop       2172 2023-09-23 01:49 /tmp/dcd/particion/covid/
mes/MES=5/part-00000-a0e82301-1324-4d0f-88a2-95dc36d52732.c000.snappy.parquet
drwxr-xr-x   - root hadoop           0 2023-09-23 01:49
/tmp/dcd/particion/covid/mes/MES=6
-rw-r--r--   2 root hadoop       2210 2023-09-23 01:49 /tmp/dcd/particion/covid/
mes/MES=6/part-00000-a0e82301-1324-4d0f-88a2-95dc36d52732.c000.snappy.parquet
drwxr-xr-x   - root hadoop           0 2023-09-23 01:49
/tmp/dcd/particion/covid/mes/MES=7
-rw-r--r--   2 root hadoop       2089 2023-09-23 01:49 /tmp/dcd/particion/covid/
mes/MES=7/part-00000-a0e82301-1324-4d0f-88a2-95dc36d52732.c000.snappy.parquet
drwxr-xr-x   - root hadoop           0 2023-09-23 01:49
```

```

/tmp/dcd/particion/covid/mes/MES=8
-rw-r--r--    2 root hadoop      2081 2023-09-23 01:49 /tmp/dcd/particion/covid/
mes/MES=8/part-00000-a0e82301-1324-4d0f-88a2-95dc36d52732.c000.snappy.parquet
drwxr-xr-x    - root hadoop          0 2023-09-23 01:49
/tmp/dcd/particion/covid/mes/MES=9
-rw-r--r--    2 root hadoop      1935 2023-09-23 01:49 /tmp/dcd/particion/covid/
mes/MES=9/part-00000-a0e82301-1324-4d0f-88a2-95dc36d52732.c000.snappy.parquet
-rw-r--r--    2 root hadoop          0 2023-09-23 01:49
/tmp/dcd/particion/covid/mes/_SUCCESS

```

[]:

```

[19]: df.write.partitionBy("ANIO", "MES").csv("hdfs:///tmp/dcd/particion/covid/
↪anio_mes_csv")

```

```

[20]: ! hdfs dfs -ls -R /tmp/dcd/particion/covid/anio_mes_csv

```

```

drwxr-xr-x    - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2020
drwxr-xr-x    - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2020/MES=1
-rw-r--r--    2 root hadoop    124874 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2020/MES=1/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x    - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2020/MES=10
-rw-r--r--    2 root hadoop      282 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2020/MES=10/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x    - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2020/MES=11
-rw-r--r--    2 root hadoop      117 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2020/MES=11/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x    - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2020/MES=12
-rw-r--r--    2 root hadoop      244 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2020/MES=12/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x    - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2020/MES=2
-rw-r--r--    2 root hadoop    129611 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2020/MES=2/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x    - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2020/MES=3
-rw-r--r--    2 root hadoop    372269 2023-09-23 01:52 /tmp/dcd/particion/covid/

```

```

anio_mes_csv/ANIO=2020/MES=3/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x  - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2020/MES=4
-rw-r--r--   2 root hadoop      605217 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2020/MES=4/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x  - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2020/MES=5
-rw-r--r--   2 root hadoop        537 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2020/MES=5/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x  - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2020/MES=6
-rw-r--r--   2 root hadoop        602 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2020/MES=6/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x  - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2020/MES=7
-rw-r--r--   2 root hadoop        410 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2020/MES=7/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x  - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2020/MES=8
-rw-r--r--   2 root hadoop        354 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2020/MES=8/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x  - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2020/MES=9
-rw-r--r--   2 root hadoop        192 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2020/MES=9/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x  - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2021
drwxr-xr-x  - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2021/MES=1
-rw-r--r--   2 root hadoop        256 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2021/MES=1/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x  - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2021/MES=10
-rw-r--r--   2 root hadoop         25 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2021/MES=10/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x  - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2021/MES=2
-rw-r--r--   2 root hadoop         99 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2021/MES=2/part-00000-74ba3594-1fb3-4386-aabe-

```

```

76bed32199f0.c000.csv
drwxr-xr-x  - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2021/MES=3
-rw-r--r--  2 root hadoop          12 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2021/MES=3/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x  - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2021/MES=4
-rw-r--r--  2 root hadoop          25 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2021/MES=4/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
drwxr-xr-x  - root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/ANIO=2021/MES=8
-rw-r--r--  2 root hadoop          26 2023-09-23 01:52 /tmp/dcd/particion/covid/
anio_mes_csv/ANIO=2021/MES=8/part-00000-74ba3594-1fb3-4386-aabe-
76bed32199f0.c000.csv
-rw-r--r--  2 root hadoop          0 2023-09-23 01:52
/tmp/dcd/particion/covid/anio_mes_csv/_SUCCESS

```

```
[21]: df.write.partitionBy("MES").csv("gs://dcd05-sir-bucket/dcd/particion/covid/mes")
```

```

-----
AnalysisException                                Traceback (most recent call last)
/tmp/ipykernel_24129/4253816011.py in <cell line: 1>()
----> 1 df.write.partitionBy("MES").csv("gs://dcd05-sir-bucket/dcd/particion/
      ↪ covid/mes")

/usr/lib/spark/python/pyspark/sql/readwriter.py in csv(self, path, mode,
      ↪ compression, sep, quote, escape, header, nullValue, escapeQuotes, quoteAll,
      ↪ dateFormat, timestampFormat, ignoreLeadingWhiteSpace,
      ↪ ignoreTrailingWhiteSpace, charToEscapeQuoteEscaping, encoding, emptyValue,
      ↪ lineSep)
    1238         lineSep=lineSep,
    1239     )
-> 1240     self._jwrite.csv(path)
    1241
    1242     def orc(

/usr/lib/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/java_gateway.py in
      ↪ __call__(self, *args)
    1319
    1320         answer = self.gateway_client.send_command(command)
-> 1321         return_value = get_return_value(
    1322             answer, self.gateway_client, self.target_id, self.name)
    1323

/usr/lib/spark/python/pyspark/sql/utils.py in deco(*a, **kw)

```

```

194                                     # Hide where the exception came from that shows a
↳non-Pythonic
195                                     # JVM exception message.
--> 196                             raise converted from None
197                             else:
198                                 raise

AnalysisException: path gs://dcd05-sir-bucket/dcd/particion/covid/mes already
↳exists.

```

```
[22]: df.write.partitionBy("ANIO").saveAsTable("covid_sql")
```

```

ivysettings.xml file not found in HIVE_HOME or
HIVE_CONF_DIR,/etc/hive/conf.dist/ivysettings.xml will be used
23/09/23 01:58:04 WARN SessionState: METASTORE_FILTER_HOOK will be ignored,
since hive.security.authorization.manager is set to instance of
HiveAuthorizerFactory.

```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```