

# Create hudi table

```
Terminal Shell Edit View Window Help
demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash -- 245x64
--fs:/onetable_257/demo - docker-compose - start_demo.sh
.../onetable_257/demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash
.../onetable_257/demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash
+

docker exec
Last login: Sat Nov 25 10:53:04 on ttys021
sagar@demo:~$ docker exec -it demo-spark-master-1/bin/bash
root@96969e9b24db:/home/onetable# ls
utilities
root@96969e9b24db:/home/onetable# pyspark \
--packages org.apache.hudi:hudi-spark3.2-bundle_2.12:0.14.0 \
--conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" \
--conf "spark.sql.catalog.spark_catalog=org.apache.spark.sql.hudi.catalog.HoodieCatalog" \
--conf "spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
Python 3.10.13 (main, Nov 15 2023, 16:37:09) [GCC 10.2.1 20210118] on linux
Type "help", "copyright", "credits" or "license()" for more information.
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/opt/bitnami/spark/jars/spark-unsafe_2.12-3.2.4.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
:: loading settings :: url = jar:file:/opt/bitnami/spark/jars/ivy-2.6.1-jar!/org/apache/ivy/core/settings/ivysettings.xml
Ivy Default Cache set to: /root/.ivy2/cache
The jars for the packages stored in: /root/.ivy2/jars
org.apache.hudi:hudi-spark3.2-bundle_2.12 address as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent-408f4154-8a8d-44a8-a3e7-121c2da62f8d:1.0
conf: [default]
Found org.apache.hudi:hudi-spark3.2-bundle_2.12:0.14.0 in central
downloading https://repo1.maven.org/maven2/org/apache/hudi/hudi-spark3.2-bundle_2.12/0.14.0/hudi-spark3.2-bundle_2.12-0.14.0.jar ...
(SUCCESSFUL) org.apache.hudi:hudi-spark3.2-bundle_2.12:0.14.0:hudi-spark3.2-bundle_2.12.jar (1694ms)
:: resolution report :: resolve 1022ms :: artifacts dl 1694ms
:: modules in use:
org.apache.hudi:hudi-spark3.2-bundle_2.12:0.14.0 from central in [default]
-----
| conf | number | modules | artifacts | | | |
|---|---|---|---|---|---|---|
| default | 1 | 1 | 1 | 0 | 1 | 1 |
-----
:: retrieving :: org.apache.spark#spark-submit-parent-408f4154-8a8d-44a8-a3e7-121c2da62f8d
conf: [default]
1 artifacts copied, 0 already retrieved (1025240/74ms)
23/11/25 19:12:46 WARN NativeCodeLoader: Unable to load native-adoop library for your platform... using builtin-java classes where applicable
Using Spark's default local profile: org.apache.spark/ivy/ivy-2.6.1-jar!/org/apache/ivy/core/settings/ivysettings.xml
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to

      ____
     / ___/
    / __/
   /___/
  version 3.2.4

Using Python version 3.10.13 (main, Nov 15 2023 16:37:09)
Spark context Web UI available at https://96969e9b24db:4040
Spark context available as 'sc' (master = local[1], app id = local-1700939567564).
SparkSession available as 'spark'.
>>> from pyspark.sql.types import *
>>>
>>> # initialize the bucket
>>> table_name = "people"
>>> local_base_path = "/tmp/hudi-dataset"
>>>
>>> records = [
... (1, 'John', 25, 'NYC', '2023-09-28 00:00:00'),
... (2, 'Emily', 30, 'SFO', '2023-09-28 00:00:00'),
... (3, 'Michael', 35, 'ORD', '2023-09-28 00:00:00'),
... (4, 'Andrew', 40, 'MCO', '2023-10-28 00:00:00'),
... (5, 'Bob', 28, 'SEA', '2023-09-23 00:00:00'),
... ]
>>>
```

```
Terminal Shell Edit View Window Help
demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash -- 245x64
--fs:/onetable_257/demo - docker-compose - start_demo.sh
.../onetable_257/demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash
.../onetable_257/demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash
+

docker exec
>>>
>>> schema = StructType([
... StructField("id", IntegerType(), True),
... StructField("name", StringType(), True),
... StructField("age", IntegerType(), True),
... StructField("city", StringType(), True),
... StructField("create_ts", StringType(), True)
... ])
>>> df = spark.createDataFrame(records, schema)
>>>
>>> hudi_options = {
... 'hoodie.table.name': table_name,
... 'hoodie.datasource.write.partitionpath.field': 'city',
... 'hoodie.datasource.write.hive_style_partitioning': 'true'
... }
>>>
>>> df.write
... .format("hudi")
... .options(**hudi_options)
... .save(**(local_base_path)/(table_name))
... )
23/11/25 19:12:51 WARN DFSPropertiesConfiguration: Cannot find HUDI_CONF_DIR, please set it as the dir of hudi-defaults.conf
23/11/25 19:12:51 WARN DFSPropertiesConfiguration: Properties file file:/etc/hudi/conf/hudi-defaults.conf not found. Ignoring to load props file
23/11/25 19:12:51 WARN HoodieSparkSqlWriters: Choosing BULK_INSERT as the operation type since auto record key generation is applicable
23/11/25 19:12:51 WARN HiveConf: HiveConf of name hive.stats.jdbc.timeout does not exist
23/11/25 19:12:51 WARN HiveConf: HiveConf of name hive.stats.retries.wait does not exist
23/11/25 19:12:51 WARN ObjectStore: Version information not found in metastore. hive.metastore.schema.verification is not enabled so recording the schema version 2.3.0
23/11/25 19:12:53 WARN ObjectStore: SetMetaStoreSchemeVersion called but recording version is disabled: version = 2.3.0, comment = Set by MetaStore UNKNOWN@192.168.128.5
23/11/25 19:12:53 WARN ObjectStore: Failed to get database default, returning NoSubObjectException
23/11/25 19:12:53 WARN AutoClobberKeyOverwriteUtil: Preserve field is will be ignored with auto record key generation enabled
23/11/25 19:12:54 WARN MetricsConfig: Cannot locate configuration: tried hadoop-metrics2-hbase.properties, hadoop-metrics2.properties
# WARNING: Unable to get Instrumentation. Dynamic Attach failed. You may add this JAR as -javaagent manually, or supply -Djdk.attach.allowAttachSelf
# WARNING: Unable to attach Serviceability Agent. Unable to attach even with module exceptions: (org.apache.hudi.org.openjdk.jol.vm.sa.SASupportException: Sense failed., org.apache.hudi.org.openjdk.jol.vm.sa.SASupportException: Sense failed., org.apache.hudi.org.openjdk.jol.vm.sa.SASupportException: Sense failed.)
>>> exit()
root@96969e9b24db:/home/onetable# vi my_config.yaml
root@96969e9b24db:/home/onetable# java -jar utilities/target/utilities-0.1.0-SNAPSHOT-bundled.jar --datasetConfig my_config.yaml
SLF4J: No SLF4J providers were found.
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See https://www.slf4j.org/codes.html#noProviders for further details.
SLF4J: Class path contains SLF4J bindings targeting slf4j-api versions 1.7.x or earlier.
SLF4J: Ignoring binding found at [jar:file:/home/onetable/utilities/target/utilities-0.1.0-SNAPSHOT-bundled.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See https://www.slf4j.org/codes.html#ignoringBindings for an explanation.
WARNING: sun.reflect.Reflection.getCallerClass is not supported. This will impact performance.
2023-11-25 19:13:21 INFO io.onetable.utilities.RunSync:150 - Running sync for basePath file:///tmp/hudi-dataset/people for following table formats [DELTA, ICEBERG]
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/home/onetable/utilities/target/utilities-0.1.0-SNAPSHOT-bundled.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
2023-11-25 19:13:22 INFO io.onetable.client.OneTableClient:120 - No previous OneTable sync for target. Falling back to snapshot sync.
2023-11-25 19:13:22 INFO io.onetable.client.OneTableClient:120 - No previous OneTable sync for target. Falling back to snapshot sync.
# WARNING: Unable to get Instrumentation. Dynamic Attach failed. You may add this JAR as -javaagent manually, or supply -Djdk.attach.allowAttachSelf
# WARNING: Unable to attach Serviceability Agent. You can try again with escalated privileges. Two options: a) use -Djoi.trythisdundertrue to try with sudo; b) echo 0 | sudo tee /proc/sys/kernel/yama/ptrace_scope
2023-11-25 19:13:27 INFO io.onetable.client.OneTableClient:120 - OneTable Sync is successful for the following formats [DELTA, ICEBERG]
root@96969e9b24db:/home/onetable# pyspark \
--packages io.delta:delta-core_2.12:1.0 \
--conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension" \
--conf "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
Python 3.10.13 (main, Nov 15 2023, 16:37:09) [GCC 10.2.1 20210118] on linux
Type "help", "copyright", "credits" or "license()" for more information.
WARNING: An illegal reflective access operation has occurred
```

# Create delta table

```
Terminal Shell Edit View Window Help
demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash -- 245x64
--jss/onetable_257/demo - docker-compose - start_demo.sh
.../onetable_257/demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash
.../onetable_257/demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash

docker exec
root@969e9b24db:/home/onetable pyspark \
--packages io.delta:delta-core-2.12:2.1.0 \
--conf "spark.sql.extensions=org.apache.spark.extensions.DeltaSparkSessionExtension" \
--conf "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
Python 3.10.13 (main, Nov 15 2023, 16:37:09) [GCC 10.2.1 20210118] on linux
Type "help", "copyright()", "credits()" or "license()" for more information.
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/opt/bitnami/spark/jars/spark-unsafe-2.12-3.2.4.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
:: loading settings :: url = jar:file:/opt/bitnami/spark/jars/ivy-2.6.1.jar!/org/apache/ivy/core/settings/ivysettings.xml
Ivy Default Cache set to: /root/.ivy2/cache
The jars for the packages stored in: /root/.ivy2/jars
io.delta:delta-core-2.12 added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent-f97dab3e-8f3e-4021-8360-1838308bb08c;1.0
confs: [default]
Found io.delta:delta-core-2.12:2.1.0 in central
Found io.delta:delta-storage:2.1.0 in central
Found org.apache.spark:antlr4-runtime:4.8 in central
Found org.codehaus.jackson:jackson-core-asl:1.9.13 in central
Downloading https://repo1.maven.org/maven2/io/delta/delta-core-2.12/2.1.0/delta-core-2.12-2.1.0.jar ...
[SUCCESSFUL] io.delta:delta-core-2.12:2.1.0/delta-core-2.12.jar (143ms)
Downloading https://repo1.maven.org/maven2/io/delta/delta-storage/2.1.0/delta-storage-2.1.0.jar ...
[SUCCESSFUL] io.delta:delta-storage:2.1.0/delta-storage.jar (15ms)
Downloading https://repo1.maven.org/maven2/org/antlr/antlr4-runtime/4.8/antlr4-runtime-4.8.jar ...
[SUCCESSFUL] org.apache.spark:antlr4-runtime:4.8/antlr4-runtime.jar (22ms)
Downloading https://repo1.maven.org/maven2/org/codehaus/jackson/jackson-core-asl/1.9.13/jackson-core-asl-1.9.13.jar ...
[SUCCESSFUL] org.codehaus.jackson:jackson-core-asl:1.9.13/jackson-core-asl.jar (16ms)
:: resolution report :: resolve 1126ms :: artifacts dl 297ms
:: modules in use:
io.delta:delta-core-2.12:2.1.0 from central in [default]
io.delta:delta-storage:2.1.0 from central in [default]
org.apache.spark:antlr4-runtime:4.8 from central in [default]
org.codehaus.jackson:jackson-core-asl:1.9.13 from central in [default]
+-----+-----+
| conf | number | modules | artifacts |
|-----+-----+
| default | 4 | 4 | 4 | 0 | 4 | 4 |
+-----+-----+
:: retrieving :: org.apache.spark#spark-submit-parent-f97dab3e-8f3e-4021-8360-1838308bb08c
confs: [default]
4 artifacts copied, 0 already retrieved (3759k/5ms)
23/11/25 19:13:36 WARN NativeCodeLoader: Unable to load native-heap-optimizer for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to

      / \
     /   \
    /_____\
   /         \
  /             \
 /               \
/                 \
version 3.2.4

Using Python version 3.10.13 (main, Nov 15 2023, 16:37:09)
Spark context Web UI available at http://969e9b24db:4040
Spark context available as 'sc' (master = local[*], app id = local-17080939619344).
SparkSession available as 'spark'.
>>> from pyspark.sql.types import *
>>> # initialize the bucket
>>> table_name = 'people'
>>> local_base_path = '/tmp/delta-dataset'
```

```
Terminal Shell Edit View Window Help
demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash -- 245x64
--jss/onetable_257/demo - docker-compose - start_demo.sh
.../onetable_257/demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash
.../onetable_257/demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash

docker exec
>>>
>>> records = [
... (1, 'John', 25, 'NYC', '2023-09-28 00:00:00'),
... (2, 'Emily', 30, 'SFO', '2023-09-28 00:00:00'),
... (3, 'Michael', 35, 'LAX', '2023-09-28 00:00:00'),
... (4, 'Andrew', 40, 'NYC', '2023-10-28 00:00:00'),
... (5, 'Bob', 20, 'SEA', '2023-09-23 00:00:00'),
... (6, 'Charlie', 31, 'DFW', '2023-08-29 00:00:00')
... ]
>>>
>>> schema = StructType([
... StructField('id', IntegerType(), True),
... StructField('name', StringType(), True),
... StructField('age', IntegerType(), True),
... StructField('city', StringType(), True),
... StructField('create_ts', StringType(), True)
... ])
>>>
>>> df = spark.createDataFrame(records, schema)
>>>
>>> df.write
... .format("delta")
... .partitionBy("city")
... .save(local_base_path)
>>>
>>> exit()
root@969e9b24db:/home/onetable# rm my_config.yaml
root@969e9b24db:/home/onetable# vi my_config.yaml
root@969e9b24db:/home/onetable# java -jar utilities/target/utilities-0.1.0-SNAPSHOT-bundled.jar --datasetConfig my_config.yaml
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See https://www.slf4j.org/codes.html#NoProviders for further details.
SLF4J: Class path contains SLF4J bindings targeting slf4j-api versions 1.7.x or earlier.
SLF4J: Ignoring binding found at [jar:file:/home/onetable/utilities/target/utilities-0.1.0-SNAPSHOT-bundled.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See https://www.slf4j.org/codes.html#ignoredBindings for an explanation.
WARNING: sun.reflect.Reflection.getCallerClass is not supported. This will impact performance.
2023-11-25 19:14:19 INFO io.onetable.utilities.RunSync:150 - Running sync for basePath file:///tmp/delta-dataset/people for following table formats [HUDI, ICEBERG]
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/home/onetable/utilities/target/utilities-0.1.0-SNAPSHOT-bundled.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
2023-11-25 19:14:24 INFO io.onetable.hudi.HudiTableManager:73 - Hudi table does not exist, will be created on first sync
2023-11-25 19:14:24 INFO io.onetable.client.OnetableClient:120 - No previous Onetable sync for target. Falling back to snapshot sync.
2023-11-25 19:14:24 INFO io.onetable.client.OnetableClient:120 - No previous Onetable sync for target. Falling back to snapshot sync.
# WARNING: Unable to get Instrumentation. Dynamic Attach failed. You may add this JAR as -jvmsync manually, or supply -Djdk.attach.allowAttachSelf
# WARNING: Unable to attach ServiceabilityAgent. You can try again with escalated privileges. Two options: a) use -Djvmsync=true to try with sudo; b) echo 0 | sudo tee /proc/sys/kernel/yama/ptrace_scope
2023-11-25 19:14:26 INFO io.onetable.client.OnetableClient:123 - Onetable Sync is successful for the following formats [HUDI, ICEBERG]
root@969e9b24db:/home/onetable# pyspark \
--packages org.apache.iceberg:iceberg-spark-runtime-3.2:2.12.1.4.1 \
--conf "spark.sql.extensions=org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions" \
--conf "spark.sql.catalog.spark_catalog=org.apache.iceberg.spark.SparkSessionCatalog"
Python 3.10.13 (main, Nov 15 2023, 16:37:09) [GCC 10.2.1 20210118] on linux
Type "help", "copyright()", "credits()" or "license()" for more information.
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/opt/bitnami/spark/jars/spark-unsafe-2.12-3.2.4.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
:: loading settings :: url = jar:file:/opt/bitnami/spark/jars/ivy-2.6.1.jar!/org/apache/ivy/core/settings/ivysettings.xml
Ivy Default Cache set to: /root/.ivy2/cache
The jars for the packages stored in: /root/.ivy2/jars
org.apache.iceberg:iceberg-spark-runtime-3.2:2.12 added as a dependency
```









# Read hudi and iceberg source tables in delta

```
Terminal Shell Edit View Window Help demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash -- 245x64
--fs:/onetable_257/demo - docker-compose - start_demo.sh .../onetable_257/demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash
docker exec
  confs: [default]
  0 artifacts copied, 4 already retrieved (8kB/2ms)
23/11/25 19:12:27 WARN NativeCodeLoader: Unable to load native-adoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to
  version 3.2.4

Using Python version 3.10.13 (main, Nov 15 2023 16:37:09)
Spark context Web UI available at http://956591b24db:4040
Spark context available as 'sc' (master = local[*]), app id = local-1700939848300.
SparkSession available as 'spark'
>>> spark.read.format("delta").load("/tmp/hudi-dataset/people").show()
+-----+-----+-----+-----+-----+-----+-----+-----+
|hoodie_commit_time|hoodie_commit_seqno|hoodie_record_key|hoodie_partition_path|hoodie_file_name|id|name|age|city|create_ts|
+-----+-----+-----+-----+-----+-----+-----+-----+
|20231125191251581|20231125191251581|...|city=DW|dc557a99-b7ca-4d1...|6|Charles|31|DW|2023-08-29 00:00:00|
|20231125191251581|20231125191251581|...|city=ORD|737cef1f-4618-4c2...|3|Michael|35|ORD|2023-09-28 00:00:00|
|20231125191251581|20231125191251581|...|city=NYC|1a0a9220-3a41-48b...|4|Andrew|40|NYC|2023-10-28 00:00:00|
|20231125191251581|20231125191251581|...|city=SFO|9aff7921-8d1b-419...|2|Emily|30|SFO|2023-09-28 00:00:00|
|20231125191251581|20231125191251581|...|city=NYC|d9ee4b08-3f6a-491...|1|John|25|NYC|2023-09-28 00:00:00|
|20231125191251581|20231125191251581|...|city=SEA|3ee30e31-cb77-4e1...|5|Bob|28|SEA|2023-09-23 00:00:00|
+-----+-----+-----+-----+-----+-----+-----+-----+

>>> spark.read.format("delta").load("/tmp/iceberg-dataset/people").show()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "/opt/bitnami/spark/python/pyspark/sql/readerwriter.py", line 158, in load
    return self._df(self._reader.load(path))
  File "/opt/bitnami/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/java_gateway.py", line 1321, in _call
  File "/opt/bitnami/spark/python/pyspark/sql/utlis.py", line 111, in deco
    return f(*a, **kw)
  File "/opt/bitnami/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/protocol.py", line 326, in get_return_value
py4j.protocol.Py4JError: An error occurred while calling sql.load.
: org.apache.spark.sql.delta.DeltaAnalysisException: '/tmp/iceberg-dataset/people' is not a Delta table.
  at org.apache.spark.sql.delta.DeltaErrorsBase.notDeltaTableException(DeltaErrors.scala:459)
  at org.apache.spark.sql.delta.DeltaErrorsBase.notDeltaTableException(DeltaErrors.scala:488)
  at org.apache.spark.sql.delta.DeltaErrorsBase.notDeltaTableException(DeltaErrors.scala:2293)
  at org.apache.spark.sql.execution.DeltaTableV2.resolveRelation(DeltaTableV2.scala:174)
  at org.apache.spark.sql.delta.sources.DeltaDataSource.$anonfun$createRelation$$deltaData$Data$Source$.scala:209)
  at org.apache.spark.sql.delta.metering.DeltaLogging.recordFrameProfile(DeltaLogging.scala:139)
  at org.apache.spark.sql.delta.metering.DeltaLogging.recordFrameProfile(DeltaLogging.scala:137)
  at org.apache.spark.sql.delta.sources.DeltaDataSource.recordFrameProfile(DeltaDataSource.scala:49)
  at org.apache.spark.sql.execution.datasources.DataSource.resolveRelation(DataSource.scala:398)
  at org.apache.spark.sql.DataFrameReader.loadV1Source(DataFrameReader.scala:274)
  at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:243)
  at scala.collection.mutable.ArrayOps$ofRef.map$mcOption$or$mcUnit$or$mcBoolean$or$mcByte$or$mcShort$or$mcInt$or$mcLong$or$mcFloat$or$mcDouble$or$mcChar$or$mcString$or$mcObject$or$mcAny$.scala:238)
  at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:210)
  at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:210)
  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 py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)
  at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.java:357)
  at py4j.Gateway.invoke(Gateway.java:282)
  at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.java:132)
  at py4j.commands.CallCommand.execute(CallCommand.java:79)

Terminal Shell Edit View Window Help demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash -- 245x64
--fs:/onetable_257/demo - docker-compose - start_demo.sh .../onetable_257/demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash
docker exec
  confs: [default]
  0 artifacts copied, 4 already retrieved (8kB/2ms)
23/11/25 19:12:27 WARN NativeCodeLoader: Unable to load native-adoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to
  version 3.2.4

Using Python version 3.10.13 (main, Nov 15 2023 16:37:09)
Spark context Web UI available at http://956591b24db:4040
Spark context available as 'sc' (master = local[*]), app id = local-1700940048179.
SparkSession available as 'spark'
>>> spark.read.format("hudi").load("/tmp/iceberg-dataset/people/data").show()
+-----+-----+-----+-----+-----+-----+-----+-----+
|id|name|age|city|create_ts|
+-----+-----+-----+-----+-----+-----+-----+-----+
|6|Charles|31|DW|2023-08-29 00:00:00|
|3|Michael|35|ORD|2023-09-28 00:00:00|
|4|Andrew|40|NYC|2023-10-28 00:00:00|
|2|Emily|30|SFO|2023-09-28 00:00:00|
|1|John|25|NYC|2023-09-28 00:00:00|
|5|Bob|28|SEA|2023-09-23 00:00:00|
+-----+-----+-----+-----+-----+-----+-----+-----+

>>> exit()
root@956591b24db:/home/onetable# pyspark --packages org.apache.hudi:hudi-spark3.2-bundle_2.12:0.14.0 --conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" --conf "spark.sql.catalog.spark_catalog=org.apache.spark.sql.hudi.catalog.HoodieCatalog" --conf "spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
Python 3.10.13 (main, Nov 15 2023, 16:37:09) [GCC 10.2.1 20210118] on Linux
Type "help", "copyright", "credits" or "license()" for more information.
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/opt/bitnami/spark/jars/spark-unsafe_2.12-3.2.4.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
:: loading settings :: url = jar:file:/opt/bitnami/spark/jars/ivy-2.5.1-jar!/org/apache/ivy/core/settings/ivysettings.xml
Ivy Default Cache set to: /root/.ivy2/cache
The jars for the packages stored in: /root/.ivy2/jars
org.apache.hudi:hudi-spark3.2-bundle_2.12 added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent-b9d97555-7784-4b65-9fa9-fe4bc72849ff;1.0
  confs: [default]
  found org.apache.hudi:hudi-spark3.2-bundle_2.12:0.14.0 in central
  :: resolution report :: resolve 57ms :: artifacts 0 ms
  :: modules in use:
  org.apache.hudi:hudi-spark3.2-bundle_2.12:0.14.0 from central in [default]
+-----+-----+-----+-----+-----+-----+-----+-----+
|conf|number|search|downloaded|evicted|number|downloaded|
+-----+-----+-----+-----+-----+-----+-----+-----+
|default|1|0|0|0|0|1|0|
+-----+-----+-----+-----+-----+-----+-----+-----+
:: retrieving :: org.apache.spark#spark-submit-parent-b9d97555-7784-4b65-9fa9-fe4bc72849ff
  confs: [default]
  0 artifacts copied, 1 already retrieved (8kB/2ms)
23/11/25 19:12:56 WARN NativeCodeLoader: Unable to load native-adoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to
  version 3.2.4

Using Python version 3.10.13 (main, Nov 15 2023 16:37:09)
Spark context Web UI available at http://956591b24db:4040
Spark context available as 'sc' (master = local[*]), app id = local-1700940048179.
SparkSession available as 'spark'
>>> spark.read.format("hudi").load("/tmp/iceberg-dataset/people/data").show()
23/11/25 19:12:56 WARN DFSPropertiesConfiguration: Cannot find HADOOP_CONF_DIR, please set it as the dir of hudi-defaults.conf
23/11/25 19:12:56 WARN DFSPropertiesConfiguration: Properties config file:/etc/hudi/conf/hudi-defaults.conf not found. Ignoring to load props file
+-----+-----+-----+-----+-----+-----+-----+-----+
|id|name|age|create_ts|city|
+-----+-----+-----+-----+-----+-----+-----+-----+
|6|Charles|31|2023-08-29 00:00:00|DW|
|3|Michael|35|2023-09-28 00:00:00|ORD|
|4|Andrew|40|2023-10-28 00:00:00|NYC|
|2|Emily|30|2023-09-28 00:00:00|SFO|
|1|John|25|2023-09-28 00:00:00|NYC|
|5|Bob|28|2023-09-23 00:00:00|SEA|
+-----+-----+-----+-----+-----+-----+-----+-----+

```

```
Terminal Shell Edit View Window Help demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash -- 245x64
--fs:/onetable_257/demo - docker-compose - start_demo.sh .../onetable_257/demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash
docker exec
  confs: [default]
  0 artifacts copied, 1 already retrieved (8kB/2ms)
23/11/25 19:12:56 WARN NativeCodeLoader: Unable to load native-adoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to
  version 3.2.4

Using Python version 3.10.13 (main, Nov 15 2023 16:37:09)
Spark context Web UI available at http://956591b24db:4040
Spark context available as 'sc' (master = local[*]), app id = local-1700940048179.
SparkSession available as 'spark'
>>> spark.read.format("hudi").load("/tmp/iceberg-dataset/people/data").show()
23/11/25 19:12:56 WARN DFSPropertiesConfiguration: Cannot find HADOOP_CONF_DIR, please set it as the dir of hudi-defaults.conf
23/11/25 19:12:56 WARN DFSPropertiesConfiguration: Properties config file:/etc/hudi/conf/hudi-defaults.conf not found. Ignoring to load props file
+-----+-----+-----+-----+-----+-----+-----+-----+
|id|name|age|create_ts|city|
+-----+-----+-----+-----+-----+-----+-----+-----+
|6|Charles|31|2023-08-29 00:00:00|DW|
|3|Michael|35|2023-09-28 00:00:00|ORD|
|4|Andrew|40|2023-10-28 00:00:00|NYC|
|2|Emily|30|2023-09-28 00:00:00|SFO|
|1|John|25|2023-09-28 00:00:00|NYC|
|5|Bob|28|2023-09-23 00:00:00|SEA|
+-----+-----+-----+-----+-----+-----+-----+-----+

```

# Reading delta and iceberg source tables in hudi

```
Terminal Shell Edit View Window Help demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash -- 245x64
~/ss/onetable_257/demo - docker-compose - start_demo.sh ...ss/onetable_257/demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash
docker exec
>>> spark.read.format("hudi").load("/tmp/iceberg-dataset/people").show()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "/opt/bitnami/spark/python/pyspark/sql/readwriter.py", line 158, in load
    return self._df(self._reader.load(path))
  File "/opt/bitnami/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/java_gateway.py", line 1321, in _call_
  File "/opt/bitnami/spark/python/pyspark/sql/utills.py", line 111, in deco
    return f(*a, **kw)
  File "/opt/bitnami/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/protocol.py", line 326, in get_return_value
py4j.protocol.Py4JJavaError: An error occurred while calling o42.load.
: org.apache.hudi.exception.TableNotFoundException: Hoodie table not found in path /tmp/iceberg-dataset/people
at org.apache.hudi.DataSourceUtils.getTablePath(DataSourceUtils.java:88)
at org.apache.hudi.DefaultSource.createRelation(DefaultSource.scala:111)
at org.apache.spark.sql.execution.datasources.DataSource.resolveRelation(DataSource.scala:358)
at org.apache.spark.sql.DataFrameReader.loadV1Source(DataFrameReader.scala:274)
at org.apache.spark.sql.DataFrameReader.$anonfun$load$3(DataFrameReader.scala:245)
at scala.Option.getOrElse(Option.scala:189)
at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:245)
at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:188)
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 py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)
at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.java:357)
at py4j.Gateway.invoke(Gateway.java:282)
at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.java:132)
at py4j.commands.CallCommand.execute(CallCommand.java:79)
at py4j.ClientServerConnection.waitForCommands(ClientServerConnection.java:182)
at py4j.ClientServerConnection.run(ClientServerConnection.java:196)
at java.base/java.lang.Thread.run(Thread.java:829)

>>> spark.read.format("hudi").load("/tmp/iceberg-dataset/people/data").show()
-----
| id | name | age | create_ts | city |
-----
| 4 | Charlie | 31 | 2023-09-28 00:00:00 | DFW |
| 3 | Michael | 35 | 2023-09-28 00:00:00 | ORD |
| 4 | Andrew | 48 | 2023-10-28 00:00:00 | NYC |
| 2 | Emily | 38 | 2023-09-28 00:00:00 | SFO |
| 1 | John | 25 | 2023-09-28 00:00:00 | NYC |
| 5 | Bob | 28 | 2023-09-23 00:00:00 | SEA |
-----

>>> hudi_options = {"hoodie.metadata.enable": "true", "hoodie.datasource.write.hive_style_partitioning": "true"}
>>> spark.read.format("hudi").options(hudi_options).load("/tmp/iceberg-dataset/people/data").show()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "/opt/bitnami/spark/python/pyspark/sql/readwriter.py", line 158, in load
    return self._df(self._reader.load(path))
  File "/opt/bitnami/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/java_gateway.py", line 1321, in _call_
  File "/opt/bitnami/spark/python/pyspark/sql/utills.py", line 111, in deco
    return f(*a, **kw)
  File "/opt/bitnami/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/protocol.py", line 326, in get_return_value
py4j.protocol.Py4JJavaError: An error occurred while calling o65.load.
: java.io.FileNotFoundException: File /tmp/delta-dataset/people/data does not exist
at org.apache.hadoop.fs.RawLocalFileSystem.deprecatedGetFileStatus(RawLocalFileSystem.java:779)
at org.apache.hadoop.fs.RawLocalFileSystem.getFileLinkStatusInternal(RawLocalFileSystem.java:1180)
at org.apache.hadoop.fs.RawLocalFileSystem.getFileStatus(RawLocalFileSystem.java:769)
at org.apache.hadoop.fs.FilterFileSystem.getFileStatus(FilterFileSystem.java:462)
at org.apache.hudi.common.util.TablePathUtils.getTablePath(TablePathUtils.java:58)
at org.apache.hudi.DataSourceUtils.getTablePath(DataSourceUtils.java:79)
at org.apache.hudi.DefaultSource.createRelation(DefaultSource.scala:111)
at org.apache.spark.sql.execution.datasources.DataSource.resolveRelation(DataSource.scala:358)
at org.apache.spark.sql.DataFrameReader.loadV1Source(DataFrameReader.scala:274)
at org.apache.spark.sql.DataFrameReader.$anonfun$load$3(DataFrameReader.scala:245)
at scala.Option.getOrElse(Option.scala:189)
at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:245)
at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:188)
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 py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)
at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.java:357)
at py4j.Gateway.invoke(Gateway.java:282)
at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.java:132)
at py4j.commands.CallCommand.execute(CallCommand.java:79)
at py4j.ClientServerConnection.waitForCommands(ClientServerConnection.java:182)
at py4j.ClientServerConnection.run(ClientServerConnection.java:196)
at java.base/java.lang.Thread.run(Thread.java:829)

>>> spark.read.format("hudi").load("/tmp/delta-dataset/people/data").show()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "/opt/bitnami/spark/python/pyspark/sql/readwriter.py", line 158, in load
    return self._df(self._reader.load(path))
  File "/opt/bitnami/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/java_gateway.py", line 1321, in _call_
  File "/opt/bitnami/spark/python/pyspark/sql/utills.py", line 111, in deco
    return f(*a, **kw)
  File "/opt/bitnami/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/protocol.py", line 326, in get_return_value
py4j.protocol.Py4JJavaError: An error occurred while calling o65.load.
: java.io.FileNotFoundException: File /tmp/delta-dataset/people/data does not exist
at org.apache.hadoop.fs.RawLocalFileSystem.deprecatedGetFileStatus(RawLocalFileSystem.java:779)
at org.apache.hadoop.fs.RawLocalFileSystem.getFileLinkStatusInternal(RawLocalFileSystem.java:1180)
at org.apache.hadoop.fs.RawLocalFileSystem.getFileStatus(RawLocalFileSystem.java:769)
at org.apache.hadoop.fs.FilterFileSystem.getFileStatus(FilterFileSystem.java:462)
at org.apache.hudi.common.util.TablePathUtils.getTablePath(TablePathUtils.java:58)
at org.apache.hudi.DataSourceUtils.getTablePath(DataSourceUtils.java:79)
at org.apache.hudi.DefaultSource.createRelation(DefaultSource.scala:111)
at org.apache.spark.sql.execution.datasources.DataSource.resolveRelation(DataSource.scala:358)
at org.apache.spark.sql.DataFrameReader.loadV1Source(DataFrameReader.scala:274)
at org.apache.spark.sql.DataFrameReader.$anonfun$load$3(DataFrameReader.scala:245)
at scala.Option.getOrElse(Option.scala:189)
at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:245)
at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:188)
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 py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)
at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.java:357)
at py4j.Gateway.invoke(Gateway.java:282)
at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.java:132)
at py4j.commands.CallCommand.execute(CallCommand.java:79)
at py4j.ClientServerConnection.waitForCommands(ClientServerConnection.java:182)
at py4j.ClientServerConnection.run(ClientServerConnection.java:196)
at java.base/java.lang.Thread.run(Thread.java:829)

>>> spark.read.format("hudi").load("/tmp/delta-dataset/people").show()
-----
| id | name | age | create_ts | city |
-----
| 3 | Michael | 35 | 2023-09-28 00:00:00 | ORD |
| 4 | Andrew | 48 | 2023-10-28 00:00:00 | NYC |
| 2 | Emily | 38 | 2023-09-28 00:00:00 | SFO |
| 1 | John | 25 | 2023-09-28 00:00:00 | NYC |
| 5 | Bob | 28 | 2023-09-23 00:00:00 | SEA |
-----

>>> spark.read.format("hudi").options(**hudi_options).load("/tmp/delta-dataset/people").show()
-----
| id | name | age | create_ts | city |
-----
| 4 | Charlie | 31 | 2023-09-28 00:00:00 | DFW |
| 3 | Michael | 35 | 2023-09-28 00:00:00 | ORD |
| 4 | Andrew | 48 | 2023-10-28 00:00:00 | NYC |
| 2 | Emily | 38 | 2023-09-28 00:00:00 | SFO |
| 1 | John | 25 | 2023-09-28 00:00:00 | NYC |
| 5 | Bob | 28 | 2023-09-23 00:00:00 | SEA |
-----
```

```
Terminal Shell Edit View Window Help demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash -- 245x64
~/ss/onetable_257/demo - docker-compose - start_demo.sh ...ss/onetable_257/demo - com.docker.cli - docker exec -it demo-spark-master-1/bin/bash
docker exec
| 3 | Michael | 35 | 2023-09-28 00:00:00 | ORD |
| 4 | Andrew | 48 | 2023-10-28 00:00:00 | NYC |
| 2 | Emily | 38 | 2023-09-28 00:00:00 | SFO |
| 1 | John | 25 | 2023-09-28 00:00:00 | NYC |
| 5 | Bob | 28 | 2023-09-23 00:00:00 | SEA |
-----

>>> spark.read.format("hudi").load("/tmp/delta-dataset/people/data").show()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "/opt/bitnami/spark/python/pyspark/sql/readwriter.py", line 158, in load
    return self._df(self._reader.load(path))
  File "/opt/bitnami/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/java_gateway.py", line 1321, in _call_
  File "/opt/bitnami/spark/python/pyspark/sql/utills.py", line 111, in deco
    return f(*a, **kw)
  File "/opt/bitnami/spark/python/lib/py4j-0.10.9.5-src.zip/py4j/protocol.py", line 326, in get_return_value
py4j.protocol.Py4JJavaError: An error occurred while calling o65.load.
: java.io.FileNotFoundException: File /tmp/delta-dataset/people/data does not exist
at org.apache.hadoop.fs.RawLocalFileSystem.deprecatedGetFileStatus(RawLocalFileSystem.java:779)
at org.apache.hadoop.fs.RawLocalFileSystem.getFileLinkStatusInternal(RawLocalFileSystem.java:1180)
at org.apache.hadoop.fs.RawLocalFileSystem.getFileStatus(RawLocalFileSystem.java:769)
at org.apache.hadoop.fs.FilterFileSystem.getFileStatus(FilterFileSystem.java:462)
at org.apache.hudi.common.util.TablePathUtils.getTablePath(TablePathUtils.java:58)
at org.apache.hudi.DataSourceUtils.getTablePath(DataSourceUtils.java:79)
at org.apache.hudi.DefaultSource.createRelation(DefaultSource.scala:111)
at org.apache.spark.sql.execution.datasources.DataSource.resolveRelation(DataSource.scala:358)
at org.apache.spark.sql.DataFrameReader.loadV1Source(DataFrameReader.scala:274)
at org.apache.spark.sql.DataFrameReader.$anonfun$load$3(DataFrameReader.scala:245)
at scala.Option.getOrElse(Option.scala:189)
at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:245)
at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:188)
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 py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)
at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.java:357)
at py4j.Gateway.invoke(Gateway.java:282)
at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.java:132)
at py4j.commands.CallCommand.execute(CallCommand.java:79)
at py4j.ClientServerConnection.waitForCommands(ClientServerConnection.java:182)
at py4j.ClientServerConnection.run(ClientServerConnection.java:196)
at java.base/java.lang.Thread.run(Thread.java:829)

>>> spark.read.format("hudi").load("/tmp/delta-dataset/people").show()
-----
| id | name | age | create_ts | city |
-----
| 3 | Michael | 35 | 2023-09-28 00:00:00 | ORD |
| 4 | Andrew | 48 | 2023-10-28 00:00:00 | NYC |
| 2 | Emily | 38 | 2023-09-28 00:00:00 | SFO |
| 1 | John | 25 | 2023-09-28 00:00:00 | NYC |
| 5 | Bob | 28 | 2023-09-23 00:00:00 | SEA |
-----

>>> spark.read.format("hudi").options(**hudi_options).load("/tmp/delta-dataset/people").show()
-----
| id | name | age | create_ts | city |
-----
| 4 | Charlie | 31 | 2023-09-28 00:00:00 | DFW |
| 3 | Michael | 35 | 2023-09-28 00:00:00 | ORD |
| 4 | Andrew | 48 | 2023-10-28 00:00:00 | NYC |
| 2 | Emily | 38 | 2023-09-28 00:00:00 | SFO |
| 1 | John | 25 | 2023-09-28 00:00:00 | NYC |
| 5 | Bob | 28 | 2023-09-23 00:00:00 | SEA |
-----
```