

```
jlyad@Ubuntu-20: ~/mongo-spark
21/05/31 18:50:08 WARN SparkSessionBuilder: Using an existing SparkSession; some spark core configurations may not take effect.
21/05/31 18:50:08 WARN SparkSessionBuilder: Using an existing SparkSession; some spark core configurations may not take effect.
21/05/31 18:50:08 WARN SparkSessionBuilder: Using an existing SparkSession; some spark core configurations may not take effect.
21/05/31 18:50:09 WARN SparkSessionBuilder: Using an existing SparkSession; some spark core configurations may not take effect.
21/05/31 18:50:09 WARN SparkSessionBuilder: Using an existing SparkSession; some spark core configurations may not take effect.
21/05/31 18:50:09 INFO Mongorelation: requiredColumns: counter, filters:
21/05/31 18:50:09 INFO AbstractConnector: Stopped Spark@41fd5f69(HTTP/1.1,[http/1.1])[127.0.0.1:4040]
[info] Test run started
[info] Test com.mongodb.spark.MongoSparkTest.shouldBeAbleToCreateDataFrameUsingJavaBean started
[info] Test com.mongodb.spark.MongoSparkTest.shouldBeAbleToHandleNoneExistentCollections started
[info] Test com.mongodb.spark.MongoSparkTest.shouldThrowWhenCreatingADatasetWithInvalidData started
[info] Test com.mongodb.spark.MongoSparkTest.useDefaultValuesWhenCreatingADatasetWithMissingData started
[info] Test com.mongodb.spark.MongoSparkTest.shouldBeAbleToCreateDataFrameByInferringTheSchemaUsingSparkSession started
[info] Test com.mongodb.spark.MongoSparkTest.shouldBeAbleToCreateDataFrameByInferringTheSchema started
[info] Test com.mongodb.spark.MongoSparkTest.shouldBeAbleToQueryViaAPipeline started
[info] Test com.mongodb.spark.MongoSparkTest.shouldBeAbleToCreateADatasetUsingJavaBeanWithSparkSession started
[info] Test com.mongodb.spark.MongoSparkTest.useCustomPartitioner started
[info] Test com.mongodb.spark.MongoSparkTest.shouldBeCreatableFromTheSparkContextWithAlternativeReadAndWriteConfigs started
[info] Test com.mongodb.spark.MongoSparkTest.shouldBeCreatableFromTheSparkContext started
[info] Test com.mongodb.spark.MongoSparkTest.shouldBeAbleToHandleDifferentCollectionTypes started
[info] Test com.mongodb.spark.MongoSparkTest.shouldBeAbleToCreateADatasetUsingJavaBean started
[info] Test run finished: 0 failed, 0 ignored, 13 total, 3.52s
[info] ReadPreferenceConfigSpec
[info] ReadPreferenceConfig
[info] - should have the expected defaults
[info] - should have the expected readPreference
[info] - should be creatable from SparkConfig
[info] - should roundtrip options
[info] - should validate values
[info] ScalaTest
[info] Run completed in 1 minute, 36 seconds.
[info] Total number of tests run: 196
[info] Suites: completed 28, aborted 0
[info] Tests: succeeded 196, failed 0, canceled 7, ignored 0, pending 0
[info] All tests passed.
[info] Passed: Total 234, Failed 0, Errors 0, Passed 234, Canceled 7
21/05/31 18:50:11 INFO MongoClientCache: Closing MongoClient: [localhost:27017]
[success] Total time: 3002 s, completed May 31, 2021 6:50:11 PM
[info] Aggregating coverage from subprojects...
[info] Found 0 subproject report files []
[info] No subproject data to aggregate, skipping reports
[success] Total time: 0 s, completed May 31, 2021 6:50:11 PM
[info] Waiting for measurement data to sync...
[info] Reading coverage instrumentation [/home/jlyad/mongo-spark/target/scala-2.12/scovage-data/scovage.coverage.xml]
[info] Reading coverage measurements...
[info] Generating coverage reports...
[info] Written Cobertura report [/home/jlyad/mongo-spark/target/scala-2.12/coverage-report/cobertura.xml]
[info] Written XML coverage report [/home/jlyad/mongo-spark/target/scala-2.12/scovage-report/scovage.xml]
[info] Written HTML coverage report [/home/jlyad/mongo-spark/target/scala-2.12/scovage-report/index.html]
21/05/31 18:50:14 INFO MongoClientCache: Closing MongoClient: [localhost:27017]
[info] Statement coverage.: 80.02%
[info] Branch coverage....: 76.42%
[info] Coverage reports completed
[info] All done. Coverage was [80.02%]
[success] Total time: 3 s, completed May 31, 2021 6:50:14 PM
jlyad@Ubuntu-20:~/mongo-spark$
```

In [2]: !which python3

/usr/bin/python3

```
In [3]: from pyspark.sql import SparkSession
from pyspark.sql import functions as F
import os
os.environ['PYSPARK_PYTHON']='usr/bin/python3'
```

```
In [7]: Jiyad_Spark_Context = SparkSession\
        .builder \
        .appName("JiyadKhan") \
        .config("spark.mongodb.input.uri", "mongodb://127.0.0.1:27017/JiyadKhan.i19-1771") \
        .config("spark.mongodb.output.uri", "mongodb://127.0.0.1:27017/JiyadKhan.i19-1771") \
        .config('spark.jars.packages', 'org.mongodb.spark:mongo-spark-connector_2.12:3.0.0')\
        .getOrCreate()
```

In [9]: Jiyad_Spark_Context

Out[9]: SparkSession - in-memory
SparkContext

[Spark UI](#)

Version

v3.1.1

Master

local[*]

AppName

JiyadKhan

```
In [16]: df.write.format("mongo").mode("append").save()
```

```
In [17]: new_df=Jiyad_Spark_Context.read.format("mongo").load()
```

```
In [18]: new_df.head(5)
```

```
Out[18]: [Row(_id=Row(oid='60b54f397a0f7d19d2e27c93'), fname='Musa', lname='DC', name='Musa, DC'),
Row(_id=Row(oid='60b54f397a0f7d19d2e27c94'), fname='Ahmed', lname='Wadood', name='Ahmed, Wadood'),
Row(_id=Row(oid='60b54f397a0f7d19d2e27c95'), fname='Ali', lname='Kamal', name='Ali, Kamal'),
Row(_id=Row(oid='60b54f397a0f7d19d2e27c96'), fname='Shayan', lname='Mansoor', name='Shayan, Mansoor'),
Row(_id=Row(oid='60b54f397a0f7d19d2e27c97'), fname='Mahreen', lname='Athar', name='Mahreen, Athar')]
```

```
In [19]: new_df.show()
```

_id	fname	lname	name
{60b54f397a0f7d19...}	Musa	DC	Musa, DC
{60b54f397a0f7d19...}	Ahmed	Wadood	Ahmed, Wadood
{60b54f397a0f7d19...}	Ali	Kamal	Ali, Kamal
{60b54f397a0f7d19...}	Shayan	Mansoor	Shayan, Mansoor
{60b54f397a0f7d19...}	Mahreen	Athar	Mahreen, Athar
{60b54f397a0f7d19...}	Saifullah	Rahool	Saifullah, Rahool
{60b54f397a0f7d19...}	Arsal	Ahmad	Arsal, Ahmad

```
In [21]: Jiyad_Spark_Context.stop()
```

```
In [ ]:
```

```
jlyad@Ubuntu-20:~$ mongo
MongoDB shell version v4.4.5
connecting to: mongodb://127.0.0.1:27017/?compressors=disabled&gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("380d5892-0ce5-4a11-a81e-cb85a8226142") }
MongoDB server version: 4.4.5
...
The server generated these startup warnings when booting:
  2021-05-31T16:15:44.709+00:00: Using the XFS filesystem is strongly recommended with the WiredTiger storage engine. See http://dochub.mongodb.org/core/prodnotes-filesystem
  2021-05-31T16:15:46.279+00:00: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted
...
...
  Enable MongoDB's free cloud-based monitoring service, which will then receive and display
  metrics about your deployment (disk utilization, CPU, operation statistics, etc).

  The monitoring data will be available on a MongoDB website with a unique URL accessible to you
  and anyone you share the URL with. MongoDB may use this information to make product
  improvements and to suggest MongoDB products and deployment options to you.

  To enable free monitoring, run the following command: db.enableFreeMonitoring()
  To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
...
> show dbs
jlyadkhan  0.000GB
admin      0.000GB
config     0.000GB
local      0.000GB
> use jlyadkhan
switched to db jlyadkhan
> db["119-1771"].find()
{ "_id" : ObjectId("60b54f397a0f7d19d2e27c93"), "lname" : "DC", "fname" : "Musa", "name" : "Musa, DC" }
{ "_id" : ObjectId("60b54f397a0f7d19d2e27c94"), "lname" : "Wadood", "fname" : "Ahmed", "name" : "Ahmed, Wadood" }
{ "_id" : ObjectId("60b54f397a0f7d19d2e27c95"), "lname" : "Kamal", "fname" : "Ali", "name" : "Ali, Kamal" }
{ "_id" : ObjectId("60b54f397a0f7d19d2e27c96"), "lname" : "Mansoor", "fname" : "Shayan", "name" : "Shayan, Mansoor" }
{ "_id" : ObjectId("60b54f397a0f7d19d2e27c97"), "lname" : "Athar", "fname" : "Mahreen", "name" : "Mahreen, Athar" }
{ "_id" : ObjectId("60b54f397a0f7d19d2e27c98"), "lname" : "Rahool", "fname" : "Saifullah", "name" : "Saifullah, Rahool" }
{ "_id" : ObjectId("60b54f397a0f7d19d2e27c99"), "lname" : "Ahmad", "fname" : "Arsal", "name" : "Arsal, Ahmad" }
> db.dropDatabase()
{ "dropped" : "jlyadkhan", "ok" : 1 }
>
```