



<https://docs.mongodb.com/spark-connector/current/python-api/>

Para cargar el conector:

```
pyspark --packages org.mongodb.spark:mongo-spark-connector_2.11:2.4.0 \  
--conf "spark.mongodb.input.uri=mongodb://127.0.0.1/db.coleccion?readPreference=primaryPreferred" \  
--conf "spark.mongodb.output.uri=mongodb://127.0.0.1/db.coleccion"
```

**db** debe reemplazarse por el nombre de la base de datos

**coleccion** debe reemplazarse por el nombre de la colección

**input.uri** hace referencia a la base de datos de la que se lee la información

**output.uri** hace referencia a la base de datos en la que se escribe la información

Para leer información de la base de datos mongoDB configurada:

```
df = spark.read.format("com.mongodb.spark.sql.DefaultSource").load()
```

Para leer información de una nueva base de datos o de una nueva colección mongoDB:

```
df = spark.read.format("com.mongodb.spark.sql.DefaultSource")\  
    .option("uri", "mongodb://127.0.0.1/database.collection").load()
```

```
clientes= spark.createDataFrame([("Pedro", 50000), ("María", 10000), ("Marta", 2000),  
    ["nombre", "ingresos"])
```

Para escribir información en la base de datos mongoDB configurada:

```
clientes.write.format("com.mongodb.spark.sql.DefaultSource").mode("append").save()
```

Para escribir información en una nueva base de datos o en una nueva colección mongoDB:

```
clientes.write.format("com.mongodb.spark.sql.DefaultSource")\  
    .mode("append").option("database", "empresa")\  
    .option("collection", "clientes").save()
```