Spark Read/Write Cheat Sheet

Read CSV

```
>>> df=spark.read.format("csv").option("header","true").load(filePath)
```

Infer Schema

```
>>> df=spark.read.format("csv").option("inferSchema","true").load(filePath)
```

Custom Schema

```
>>> csvSchema = StructType([StructField("id",IntegerType(),False)])
>>> df=spark.read.format("csv").schema(csvSchema).load(filePath)
```

Write CSV

```
>>> df.write.format("csv").mode("overwrite).save(outputPath/file.csv)
```

Read JSON

```
>>> df=spark.read.format("json").option("inferSchema","true").load(filePath)
```

Write JSON

```
>>> df.write.format("json").mode("overwrite).save(outputPath/file.json)
```

Read Parquet

```
>>> df=spark.read.format("parquet).load(parquetDirectory)
```

OR

```
>>> df=spark.read.parquet(parquetDirectory)
```

Write Parquet

```
>>> df.write.format("parquet").mode("overwrite").save("outputPath")
```

Write Parquet Partition By

```
>>> df.write.format("parquet").partitionBy("keyColumn").save("outputPath")
```

Read Delta

Spark SQL

```
>>> SELECT * FROM delta. `/path/to/delta_directory`
```

Spark SQL Unmanaged Table

```
>>> spark.sql(""" DROP TABLE IF EXISTS delta_table_name""")
>>> spark.sql(""" CREATE TABLE delta_table_name USING DELTA LOCATION '{}'
""".format(pathToDelta))
```

Write Delta

```
>>>someDataFrame.write.format("delta").partitionBy("someColumn").save(path)
```