[Stack Overflow](https://stackoverflow.com/)

1. [About](https://stackoverflow.com/company)
2. [Products](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)
3. [For Teams](https://stackoverflow.com/teams)

Top of Form



Bottom of Form

1. [Log in](https://stackoverflow.com/users/login?ssrc=head&returnurl=https%3a%2f%2fstackoverflow.com%2fquestions%2f29704333%2fspark-load-csv-file-as-dataframe) [Sign up](https://stackoverflow.com/users/signup?ssrc=head&returnurl=%2fusers%2fstory%2fcurrent)

[Home](https://stackoverflow.com/)

* 1. PUBLIC
  2. [**Stack Overflow**](https://stackoverflow.com/questions)

[Tags](https://stackoverflow.com/tags)

[Users](https://stackoverflow.com/users)

* 1. FIND A JOB

[Jobs](https://stackoverflow.com/jobs?so_medium=StackOverflow&so_source=SiteNav)

[Companies](https://stackoverflow.com/jobs/companies?so_medium=StackOverflow&so_source=SiteNav)

* 1. TEAMS

[What’s this?](javascript:void(0))

[Create a Team](https://stackoverflow.com/teams/create" \o "Stack Overflow for Teams is a private, secure spot for your organization's questions and answers.)

[Spark - load CSV file as DataFrame?](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

[Ask Question](https://stackoverflow.com/questions/ask)

Asked 5 years, 11 months ago

Active [7 months ago](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe?lastactivity)

Viewed 432k times

152

62

I would like to read a CSV in spark and convert it as DataFrame and store it in HDFS with df.registerTempTable("table\_name")

I have tried:

scala> val df = sqlContext.load("hdfs:///csv/file/dir/file.csv")

Error which I got:

java.lang.RuntimeException: hdfs:///csv/file/dir/file.csv is not a Parquet file. expected magic number at tail [80, 65, 82, 49] but found [49, 59, 54, 10]

at parquet.hadoop.ParquetFileReader.readFooter(ParquetFileReader.java:418)

at org.apache.spark.sql.parquet.ParquetRelation2$MetadataCache$$anonfun$refresh$6.apply(newParquet.scala:277)

at org.apache.spark.sql.parquet.ParquetRelation2$MetadataCache$$anonfun$refresh$6.apply(newParquet.scala:276)

at scala.collection.parallel.mutable.ParArray$Map.leaf(ParArray.scala:658)

at scala.collection.parallel.Task$$anonfun$tryLeaf$1.apply$mcV$sp(Tasks.scala:54)

at scala.collection.parallel.Task$$anonfun$tryLeaf$1.apply(Tasks.scala:53)

at scala.collection.parallel.Task$$anonfun$tryLeaf$1.apply(Tasks.scala:53)

at scala.collection.parallel.Task$class.tryLeaf(Tasks.scala:56)

at scala.collection.parallel.mutable.ParArray$Map.tryLeaf(ParArray.scala:650)

at scala.collection.parallel.AdaptiveWorkStealingTasks$WrappedTask$class.compute(Tasks.scala:165)

at scala.collection.parallel.AdaptiveWorkStealingForkJoinTasks$WrappedTask.compute(Tasks.scala:514)

at scala.concurrent.forkjoin.RecursiveAction.exec(RecursiveAction.java:160)

at scala.concurrent.forkjoin.ForkJoinTask.doExec(ForkJoinTask.java:260)

at scala.concurrent.forkjoin.ForkJoinPool$WorkQueue.runTask(ForkJoinPool.java:1339)

at scala.concurrent.forkjoin.ForkJoinPool.runWorker(ForkJoinPool.java:1979)

at scala.concurrent.forkjoin.ForkJoinWorkerThread.run(ForkJoinWorkerThread.java:107)

What is the right command to load CSV file as DataFrame in Apache Spark?

[scala](https://stackoverflow.com/questions/tagged/scala)[apache-spark](https://stackoverflow.com/questions/tagged/apache-spark)[hadoop](https://stackoverflow.com/questions/tagged/hadoop)[apache-spark-sql](https://stackoverflow.com/questions/tagged/apache-spark-sql)[hdfs](https://stackoverflow.com/questions/tagged/hdfs)

[Share](https://stackoverflow.com/q/29704333)

[Improve this question](https://stackoverflow.com/posts/29704333/edit)

Follow

[edited Jul 24 '18 at 14:00](https://stackoverflow.com/posts/29704333/revisions)

[[](https://stackoverflow.com/users/1592191/mrsrinivas)](https://stackoverflow.com/users/1592191/mrsrinivas)

[mrsrinivas](https://stackoverflow.com/users/1592191/mrsrinivas)

**27k**1010 gold badges103103 silver badges117117 bronze badges

asked Apr 17 '15 at 16:10

[[](https://stackoverflow.com/users/2411173/donbeo)](https://stackoverflow.com/users/2411173/donbeo)

[Donbeo](https://stackoverflow.com/users/2411173/donbeo)

**13.8k**2828 gold badges9292 silver badges159159 bronze badges

* check this [link for doing it in Spark 2.0](http://stackoverflow.com/questions/29704333/how-to-read-csv-file-as-dataframe/39533431#39533431) – [mrsrinivas](https://stackoverflow.com/users/1592191/mrsrinivas) [Oct 23 '16 at 5:25](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe#comment67665052_29704333)

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

14 Answers

[Active](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe?answertab=active#tab-top)[Oldest](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe?answertab=oldest#tab-top)[Votes](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe?answertab=votes#tab-top)

206

spark-csv is part of core Spark functionality and doesn't require a separate library. So you could just do for example

df = spark.read.format("csv").option("header", "true").load("csvfile.csv")

In scala,(this works for any format-in delimiter mention "," for csv, "\t" for tsv etc)

val df = sqlContext.read.format("com.databricks.spark.csv") .option("delimiter", ",") .load("csvfile.csv")

[Share](https://stackoverflow.com/a/29705881)

[Improve this answer](https://stackoverflow.com/posts/29705881/edit)

Follow

[edited Jan 4 '20 at 9:27](https://stackoverflow.com/posts/29705881/revisions)

[[](https://stackoverflow.com/users/3639306/sumitya)](https://stackoverflow.com/users/3639306/sumitya)

[sumitya](https://stackoverflow.com/users/3639306/sumitya)

**2,385**11 gold badge1414 silver badges2828 bronze badges

answered Apr 17 '15 at 17:33

[[](https://stackoverflow.com/users/332087/shyamendra-solanki)](https://stackoverflow.com/users/332087/shyamendra-solanki)

[Shyamendra Solanki](https://stackoverflow.com/users/332087/shyamendra-solanki)

**8,113**11 gold badge2727 silver badges2525 bronze badges

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

174

**Parse CSV and load as DataFrame/DataSet with Spark 2.x**

First, initialize SparkSession object **by default it will available in shells as spark**

val spark = org.apache.spark.sql.SparkSession.builder

.master("local") # Change it as per your cluster

.appName("Spark CSV Reader")

.getOrCreate;

**Use any one of the following ways to load CSV as DataFrame/DataSet**

1. Do it in a programmatic way

val df = spark.read

.format("csv")

.option("header", "true") //first line in file has headers

.option("mode", "DROPMALFORMED")

.load("hdfs:///csv/file/dir/file.csv")

Update: Adding all options [from here](https://github.com/databricks/spark-csv#features) in case the link will be broken in future

* **path**: location of files. Similar to Spark can accept standard Hadoop globbing expressions.
* **header**: when set to true the first line of files will be used to name columns and will not be included in data. All types will be assumed string. The default value is false.
* **delimiter**: by default columns are delimited using, but delimiter can be set to any character
* **quote**: by default the quote character is ", but can be set to any character. Delimiters inside quotes are ignored
* **escape**: by default, the escape character is , but can be set to any character. Escaped quote characters are ignored
* **parserLib**: by default, it is "**commons**" that can be set to "**univocity**" to use that library for CSV parsing.
* **mode**: determines the parsing mode. By default it is PERMISSIVE. Possible values are:
  + **PERMISSIVE**: tries to parse all lines: nulls are inserted for missing tokens and extra tokens are ignored.
  + **DROPMALFORMED**: drops lines which have fewer or more tokens than expected or tokens which do not match the schema
  + **FAILFAST**: aborts with a RuntimeException if encounters any malformed line charset: defaults to 'UTF-8' but can be set to other valid charset names
* **inferSchema**: automatically infers column types. It requires one extra pass over the data and is false by default comment: skip lines beginning with this character. Default is "#". Disable comments by setting this to null.
* **nullValue**: specifies a string that indicates a null value, any fields matching this string will be set as nulls in the DataFrame
* **dateFormat**: specifies a string that indicates the date format to use when reading dates or timestamps. Custom date formats follow the formats at java.text.SimpleDateFormat. This applies to both DateType and TimestampType. By default, it is null which means trying to parse times and date by java.sql.Timestamp.valueOf() and java.sql.Date.valueOf().

2. [You can do this SQL way as well](https://stackoverflow.com/a/42688281/1592191)

val df = spark.sql("SELECT \* FROM csv.`hdfs:///csv/file/dir/file.csv`")

**Dependencies**:

"org.apache.spark" % "spark-core\_2.11" % 2.0.0,

"org.apache.spark" % "spark-sql\_2.11" % 2.0.0,

**Spark version < 2.0**

val df = sqlContext.read

.format("com.databricks.spark.csv")

.option("header", "true")

.option("mode", "DROPMALFORMED")

.load("csv/file/path");

**Dependencies:**

"org.apache.spark" % "spark-sql\_2.10" % 1.6.0,

"com.databricks" % "spark-csv\_2.10" % 1.6.0,

"com.univocity" % "univocity-parsers" % LATEST,

[Share](https://stackoverflow.com/a/39533431)

[Improve this answer](https://stackoverflow.com/posts/39533431/edit)

Follow

[edited Jul 17 '20 at 5:05](https://stackoverflow.com/posts/39533431/revisions)

answered Sep 16 '16 at 14:01

[[](https://stackoverflow.com/users/1592191/mrsrinivas)](https://stackoverflow.com/users/1592191/mrsrinivas)

[mrsrinivas](https://stackoverflow.com/users/1592191/mrsrinivas)

**27k**1010 gold badges103103 silver badges117117 bronze badges

* do this session require hive? I am getting hive errors. – [Puneet](https://stackoverflow.com/users/2770775/puneet) [Nov 15 '16 at 19:28](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe#comment68470221_39533431)
* 2

No need. Only spark-core\_2.11 and spark-sql\_2.11 of 2.0.1 version is fine. If possible add the error message. – [mrsrinivas](https://stackoverflow.com/users/1592191/mrsrinivas) [Nov 16 '16 at 3:20](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe#comment68480916_39533431)

* 1

can we convert a pipe delimited file to a dataframe? – [Omkar Puttagunta](https://stackoverflow.com/users/2104303/omkar-puttagunta) [Mar 23 '17 at 14:30](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe#comment73049834_39533431)

* 3

@OmkarPuttagunta: Yes, off course! try some thing like this spark.read.format("csv").option("delimiter ", "|") ... – [mrsrinivas](https://stackoverflow.com/users/1592191/mrsrinivas) [Mar 23 '17 at 14:36](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe#comment73050111_39533431)

* 1

The other option for programmatic way is to leave off the .format("csv") and replace .load(... with .csv(.... The option method belongs to the DataFrameReader class as returned by the read method, where the load and csv methods return a dataframe so can't have options tagged on after they are called. This answer is pretty thorough but you should link to the documentation so people can see all the other CSV options available [spark.apache.org/docs/latest/api/scala/…](http://spark.apache.org/docs/latest/api/scala/index.html#org.apache.spark.sql.DataFrameReader@csv(paths:String)\*):org.apache.spark.sql.DataFrame – [Davos](https://stackoverflow.com/users/1335793/davos) [Apr 17 '18 at 13:20](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe#comment86774113_39533431)

[Show **1** more comments](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

18

It's for whose Hadoop is 2.6 and Spark is 1.6 and without "databricks" package.

import org.apache.spark.sql.types.{StructType,StructField,StringType,IntegerType};

import org.apache.spark.sql.Row;

val csv = sc.textFile("/path/to/file.csv")

val rows = csv.map(line => line.split(",").map(\_.trim))

val header = rows.first

val data = rows.filter(\_(0) != header(0))

val rdd = data.map(row => Row(row(0),row(1).toInt))

val schema = new StructType()

.add(StructField("id", StringType, true))

.add(StructField("val", IntegerType, true))

val df = sqlContext.createDataFrame(rdd, schema)

[Share](https://stackoverflow.com/a/42898960)

[Improve this answer](https://stackoverflow.com/posts/42898960/edit)

Follow

[edited Mar 20 '17 at 9:07](https://stackoverflow.com/posts/42898960/revisions)

answered Mar 20 '17 at 8:35

[[](https://stackoverflow.com/users/7706918/eric-yiwei-liu)](https://stackoverflow.com/users/7706918/eric-yiwei-liu)

[Eric Yiwei Liu](https://stackoverflow.com/users/7706918/eric-yiwei-liu)

**201**22 silver badges66 bronze badges

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

14

With Spark 2.0, following is how you can read CSV

val conf = new SparkConf().setMaster("local[2]").setAppName("my app")

val sc = new SparkContext(conf)

val sparkSession = SparkSession.builder

.config(conf = conf)

.appName("spark session example")

.getOrCreate()

val path = "/Users/xxx/Downloads/usermsg.csv"

val base\_df = sparkSession.read.option("header","true").

csv(path)

[Share](https://stackoverflow.com/a/39715436)

[Improve this answer](https://stackoverflow.com/posts/39715436/edit)

Follow

[edited Oct 30 '16 at 6:34](https://stackoverflow.com/posts/39715436/revisions)

[[](https://stackoverflow.com/users/2308683/onecricketeer)](https://stackoverflow.com/users/2308683/onecricketeer)

[OneCricketeer](https://stackoverflow.com/users/2308683/onecricketeer)

**123k**1313 gold badges8989 silver badges180180 bronze badges

answered Sep 27 '16 at 3:17

[[](https://stackoverflow.com/users/6146234/penny-chan)](https://stackoverflow.com/users/6146234/penny-chan)

[penny chan](https://stackoverflow.com/users/6146234/penny-chan)

**585**88 silver badges1515 bronze badges

* 7

Is there a difference between spark.read.csv(path) and spark.read.format("csv").load(path)? – [Eric](https://stackoverflow.com/users/7906521/eric) [Jun 5 '18 at 14:49](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe#comment88415363_39715436)

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

9

In Java 1.8 This code snippet perfectly working to read CSV files

POM.xml

<dependency>

<groupId>org.apache.spark</groupId>

<artifactId>spark-core\_2.11</artifactId>

<version>2.0.0</version>

</dependency>

<!-- https://mvnrepository.com/artifact/org.apache.spark/spark-sql\_2.10 -->

<dependency>

<groupId>org.apache.spark</groupId>

<artifactId>spark-sql\_2.10</artifactId>

<version>2.0.0</version>

</dependency>

<!-- https://mvnrepository.com/artifact/org.scala-lang/scala-library -->

<dependency>

<groupId>org.scala-lang</groupId>

<artifactId>scala-library</artifactId>

<version>2.11.8</version>

</dependency>

<dependency>

<groupId>com.databricks</groupId>

<artifactId>spark-csv\_2.10</artifactId>

<version>1.4.0</version>

</dependency>

Java

SparkConf conf = new SparkConf().setAppName("JavaWordCount").setMaster("local");

// create Spark Context

SparkContext context = new SparkContext(conf);

// create spark Session

SparkSession sparkSession = new SparkSession(context);

Dataset<Row> df = sparkSession.read().format("com.databricks.spark.csv").option("header", true).option("inferSchema", true).load("hdfs://localhost:9000/usr/local/hadoop\_data/loan\_100.csv");

//("hdfs://localhost:9000/usr/local/hadoop\_data/loan\_100.csv");

System.out.println("========== Print Schema ============");

df.printSchema();

System.out.println("========== Print Data ==============");

df.show();

System.out.println("========== Print title ==============");

df.select("title").show();

[Share](https://stackoverflow.com/a/40105291)

[Improve this answer](https://stackoverflow.com/posts/40105291/edit)

Follow

[edited Oct 30 '16 at 6:32](https://stackoverflow.com/posts/40105291/revisions)

[[](https://stackoverflow.com/users/2308683/onecricketeer)](https://stackoverflow.com/users/2308683/onecricketeer)

[OneCricketeer](https://stackoverflow.com/users/2308683/onecricketeer)

**123k**1313 gold badges8989 silver badges180180 bronze badges

answered Oct 18 '16 at 10:05

[[](https://stackoverflow.com/users/2517035/rajeev-rathor)](https://stackoverflow.com/users/2517035/rajeev-rathor)

[Rajeev Rathor](https://stackoverflow.com/users/2517035/rajeev-rathor)

**1,234**1717 silver badges1414 bronze badges

* While this may be useful to someone. The question has a Scala tag. – [OneCricketeer](https://stackoverflow.com/users/2308683/onecricketeer) [Oct 30 '16 at 6:33](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe#comment67908423_40105291)

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

6

There are a lot of challenges to parsing a CSV file, it keeps adding up if the file size is bigger, if there are non-english/escape/separator/other characters in the column values, that could cause parsing errors.

The magic then is in the options that are used. The ones that worked for me and hope should cover most of the edge cases are in code below:

### Create a Spark Session

spark = SparkSession.builder.master("local").appName("Classify Urls").getOrCreate()

### Note the options that are used. You may have to tweak these in case of error

html\_df = spark.read.csv(html\_csv\_file\_path,

header=True,

multiLine=True,

ignoreLeadingWhiteSpace=True,

ignoreTrailingWhiteSpace=True,

encoding="UTF-8",

sep=',',

quote='"',

escape='"',

maxColumns=2,

inferSchema=True)

Hope that helps. For more refer: [Using PySpark 2 to read CSV having HTML source code](https://blog.codonomics.com/2018/08/using-pyspark-2-to-read-csv-having-html.html#more)

Note: The code above is from Spark 2 API, where the CSV file reading API comes bundled with built-in packages of Spark installable.

Note: PySpark is a Python wrapper for Spark and shares the same API as Scala/Java.

[Share](https://stackoverflow.com/a/51947156)

[Improve this answer](https://stackoverflow.com/posts/51947156/edit)

Follow

answered Aug 21 '18 at 10:57

[[](https://stackoverflow.com/users/498604/karthiks)](https://stackoverflow.com/users/498604/karthiks)

[karthiks](https://stackoverflow.com/users/498604/karthiks)

**6,107**55 gold badges4242 silver badges5555 bronze badges

* Thankyou So much, You saved my life :D – [Khubaib Raza](https://stackoverflow.com/users/8111190/khubaib-raza) [Jun 6 '20 at 13:01](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe#comment110062530_51947156)

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

4

Penny's Spark 2 example is the way to do it in spark2. There's one more trick: have that header generated for you by doing an initial scan of the data, by setting the option inferSchema to true

Here, then, assumming that spark is a spark session you have set up, is the operation to load in the CSV index file of all the Landsat images which amazon host on S3.

/\*

\* Licensed to the Apache Software Foundation (ASF) under one or more

\* contributor license agreements. See the NOTICE file distributed with

\* this work for additional information regarding copyright ownership.

\* The ASF licenses this file to You under the Apache License, Version 2.0

\* (the "License"); you may not use this file except in compliance with

\* the License. You may obtain a copy of the License at

\*

\* http://www.apache.org/licenses/LICENSE-2.0

\*

\* Unless required by applicable law or agreed to in writing, software

\* distributed under the License is distributed on an "AS IS" BASIS,

\* WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

\* See the License for the specific language governing permissions and

\* limitations under the License.

\*/

val csvdata = spark.read.options(Map(

"header" -> "true",

"ignoreLeadingWhiteSpace" -> "true",

"ignoreTrailingWhiteSpace" -> "true",

"timestampFormat" -> "yyyy-MM-dd HH:mm:ss.SSSZZZ",

"inferSchema" -> "true",

"mode" -> "FAILFAST"))

.csv("s3a://landsat-pds/scene\_list.gz")

The bad news is: this triggers a scan through the file; for something large like this 20+MB zipped CSV file, that can take 30s over a long haul connection. Bear that in mind: you are better off manually coding up the schema once you've got it coming in.

(code snippet Apache Software License 2.0 licensed to avoid all ambiguity; something I've done as a demo/integration test of S3 integration)

[Share](https://stackoverflow.com/a/40116607)

[Improve this answer](https://stackoverflow.com/posts/40116607/edit)

Follow

answered Oct 18 '16 at 19:34

[[](https://stackoverflow.com/users/2261274/stevel)](https://stackoverflow.com/users/2261274/stevel)

[stevel](https://stackoverflow.com/users/2261274/stevel)

**9,582**11 gold badge2828 silver badges4242 bronze badges

* I hadn't seen this csv method or passing a map to options. Agreed always better off providing explicit schema, inferSchema is fine for quick n dirty (aka data science) but terrible for ETL. – [Davos](https://stackoverflow.com/users/1335793/davos) [Feb 15 '18 at 7:15](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe#comment84605989_40116607)

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

2

In case you are building a jar with scala 2.11 and Apache 2.0 or higher.

There is no need to create a sqlContext or sparkContext object. Just a SparkSession object suffices the requirement for all needs.

Following is mycode which works fine:

import org.apache.spark.sql.{DataFrame, Row, SQLContext, SparkSession}

import org.apache.log4j.{Level, LogManager, Logger}

object driver {

def main(args: Array[String]) {

val log = LogManager.getRootLogger

log.info("\*\*\*\*\*\*\*\*\*\*JAR EXECUTION STARTED\*\*\*\*\*\*\*\*\*\*")

val spark = SparkSession.builder().master("local").appName("ValidationFrameWork").getOrCreate()

val df = spark.read.format("csv")

.option("header", "true")

.option("delimiter","|")

.option("inferSchema","true")

.load("d:/small\_projects/spark/test.pos")

df.show()

}

}

In case you are running in cluster just change .master("local") to .master("yarn") while defining the sparkBuilder object

The Spark Doc covers this: <https://spark.apache.org/docs/2.2.0/sql-programming-guide.html>

[Share](https://stackoverflow.com/a/53056342)

[Improve this answer](https://stackoverflow.com/posts/53056342/edit)

Follow

[edited Oct 30 '18 at 2:31](https://stackoverflow.com/posts/53056342/revisions)

answered Oct 30 '18 at 1:59

[[](https://stackoverflow.com/users/4608441/swapnil-shashank)](https://stackoverflow.com/users/4608441/swapnil-shashank)

[swapnil shashank](https://stackoverflow.com/users/4608441/swapnil-shashank)

**466**33 silver badges88 bronze badges

* This is same as existing answers – [mrsrinivas](https://stackoverflow.com/users/1592191/mrsrinivas) [Nov 17 '19 at 3:08](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe#comment104058829_53056342)

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

1

Add following Spark dependencies to POM file :

<dependency>

<groupId>org.apache.spark</groupId>

<artifactId>spark-core\_2.11</artifactId>

<version>2.2.0</version>

</dependency>

<dependency>

<groupId>org.apache.spark</groupId>

<artifactId>spark-sql\_2.11</artifactId>

<version>2.2.0</version>

</dependency>

//Spark configuration:

val spark = SparkSession.builder().master("local").appName("Sample App").getOrCreate()

//Read csv file:

val df = spark.read.option("header", "true").csv("FILE\_PATH")

// Display output

df.show()

[Share](https://stackoverflow.com/a/60761297)

[Improve this answer](https://stackoverflow.com/posts/60761297/edit)

Follow

answered Mar 19 '20 at 16:30

[[](https://stackoverflow.com/users/3227848/s-k)](https://stackoverflow.com/users/3227848/s-k)

[S\_K](https://stackoverflow.com/users/3227848/s-k)

**496**55 silver badges88 bronze badges

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

1

With Spark 2.4+, if you want to load a csv from a local directory, then you can use 2 sessions and load that into hive. The first session should be created with master() config as "local[\*]" and the second session with "yarn" and Hive enabled.

The below one worked for me.

import org.apache.log4j.{Level, Logger}

import org.apache.spark.\_

import org.apache.spark.rdd.\_

import org.apache.spark.sql.\_

object testCSV {

def main(args: Array[String]) {

Logger.getLogger("org").setLevel(Level.ERROR)

val spark\_local = SparkSession.builder().appName("CSV local files reader").master("local[\*]").getOrCreate()

import spark\_local.implicits.\_

spark\_local.sql("SET").show(100,false)

val local\_path="/tmp/data/spend\_diversity.csv" // Local file

val df\_local = spark\_local.read.format("csv").option("inferSchema","true").load("file://"+local\_path) // "file://" is mandatory

df\_local.show(false)

val spark = SparkSession.builder().appName("CSV HDFS").config("spark.sql.warehouse.dir", "/apps/hive/warehouse").enableHiveSupport().getOrCreate()

import spark.implicits.\_

spark.sql("SET").show(100,false)

val df = df\_local

df.createOrReplaceTempView("lcsv")

spark.sql(" drop table if exists work.local\_csv ")

spark.sql(" create table work.local\_csv as select \* from lcsv ")

}

When ran with spark2-submit --master "yarn" --conf spark.ui.enabled=false testCSV.jar it went fine and created the table in hive.

[Share](https://stackoverflow.com/a/61770726)

[Improve this answer](https://stackoverflow.com/posts/61770726/edit)

Follow

answered May 13 '20 at 9:17

[[](https://stackoverflow.com/users/6867048/stack0114106)](https://stackoverflow.com/users/6867048/stack0114106)

[stack0114106](https://stackoverflow.com/users/6867048/stack0114106)

**7,576**22 gold badges1010 silver badges2929 bronze badges

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

0

To read from relative path on the system use System.getProperty method to get current directory and further uses to load the file using relative path.

scala> val path = System.getProperty("user.dir").concat("/../2015-summary.csv")

scala> val csvDf = spark.read.option("inferSchema","true").option("header", "true").csv(path)

scala> csvDf.take(3)

spark:2.4.4 scala:2.11.12

[Share](https://stackoverflow.com/a/58869396)

[Improve this answer](https://stackoverflow.com/posts/58869396/edit)

Follow

[edited Apr 1 '20 at 7:48](https://stackoverflow.com/posts/58869396/revisions)

answered Nov 15 '19 at 2:14

[[](https://stackoverflow.com/users/1504589/venkat-kotra)](https://stackoverflow.com/users/1504589/venkat-kotra)

[Venkat Kotra](https://stackoverflow.com/users/1504589/venkat-kotra)

**9,104**33 gold badges3939 silver badges4747 bronze badges

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

-1

Default file format is Parquet with spark.read.. and file reading csv that why you are getting the exception. Specify csv format with api you are trying to use

[Share](https://stackoverflow.com/a/51626693)

[Improve this answer](https://stackoverflow.com/posts/51626693/edit)

Follow

answered Aug 1 '18 at 6:30

[[](https://stackoverflow.com/users/6266343/tazak)](https://stackoverflow.com/users/6266343/tazak)

[tazak](https://stackoverflow.com/users/6266343/tazak)

**1**11 bronze badge

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

-1

Try this if using spark 2.0+

For non-hdfs file:

df = spark.read.csv("file:///csvfile.csv")

For hdfs file:

df = spark.read.csv("hdfs:///csvfile.csv")

For hdfs file (with different delimiter than comma:

df = spark.read.option("delimiter","|")csv("hdfs:///csvfile.csv")

Note:- this work for any delimited file. Just use option(“delimiter”,) to change value.

Hope this is helpful.

[Share](https://stackoverflow.com/a/56811655)

[Improve this answer](https://stackoverflow.com/posts/56811655/edit)

Follow

[edited Jun 28 '19 at 18:28](https://stackoverflow.com/posts/56811655/revisions)

[[](https://stackoverflow.com/users/2560327/ajay-ahuja)](https://stackoverflow.com/users/2560327/ajay-ahuja)

[Ajay Ahuja](https://stackoverflow.com/users/2560327/ajay-ahuja)

**849**99 silver badges2525 bronze badges

answered Jun 28 '19 at 18:25

user11714872

* This is same as existing answers – [mrsrinivas](https://stackoverflow.com/users/1592191/mrsrinivas) [Nov 17 '19 at 3:07](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe#comment104058825_56811655)

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

-1

With in-built Spark csv, you can get it done easily with new SparkSession object for Spark > 2.0.

val df = spark.

read.

option("inferSchema", "false").

option("header","true").

option("mode","DROPMALFORMED").

option("delimiter", ";").

schema(dataSchema).

csv("/csv/file/dir/file.csv")

df.show()

df.printSchema()

There are various options you can set.

* header: whether your file includes header line at the top
* inferSchema: whether you want to infer schema automatically or not. Default is true. I always prefer to provide schema to ensure proper datatypes.
* mode: parsing mode, PERMISSIVE, DROPMALFORMED or FAILFAST
* delimiter: to specify delimiter, default is comma(',')

[Share](https://stackoverflow.com/a/58610421)

[Improve this answer](https://stackoverflow.com/posts/58610421/edit)

Follow

answered Oct 29 '19 at 15:20

[[](https://stackoverflow.com/users/7338171/piyush-patel)](https://stackoverflow.com/users/7338171/piyush-patel)

[Piyush Patel](https://stackoverflow.com/users/7338171/piyush-patel)

**1,284**11 gold badge1111 silver badges2020 bronze badges

[Add a comment](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)

[**Highly active question**](https://stackoverflow.com/help/privileges/protect-questions). Earn 10 reputation in order to answer this question. The reputation requirement helps protect this question from spam and non-answer activity.

Not the answer you're looking for? Browse other questions tagged [scala](https://stackoverflow.com/questions/tagged/scala) [apache-spark](https://stackoverflow.com/questions/tagged/apache-spark) [hadoop](https://stackoverflow.com/questions/tagged/hadoop) [apache-spark-sql](https://stackoverflow.com/questions/tagged/apache-spark-sql) [hdfs](https://stackoverflow.com/questions/tagged/hdfs) or [ask your own question](https://stackoverflow.com/questions/ask).

**The Overflow Blog**

* [State of the Stack: a new quarterly update on community and product](https://stackoverflow.blog/2021/03/10/state-of-the-stack-community-product-q1-2021/?cb=1)
* [Podcast 320: Covid vaccine websites are frustrating. This developer built a…](https://stackoverflow.blog/2021/03/12/podcast-320-covid-vaccine-websites-are-frustrating-this-developer-built-a-better-one/?cb=1)

**Featured on Meta**

* [State of the Stack Q1 2021 Blog Post](https://meta.stackexchange.com/questions/361934/state-of-the-stack-q1-2021-blog-post?cb=1)

[**Looking** for a job?](https://stackoverflow.com/jobs?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange)

**[Software Engineering](https://stackoverflow.com/jobs/517494/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)**

**[JPMorgan Chase Bank, N.A.](https://stackoverflow.com/jobs/517494/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)***[Hyderabad, India](https://stackoverflow.com/jobs/517494/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)*

[apache-sparkjava](https://stackoverflow.com/jobs/517494/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)

[Be one of the first applicants](https://stackoverflow.com/jobs/517494/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)

**[Software Engineering](https://stackoverflow.com/jobs/517232/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)**

**[JPMorgan Chase Bank, N.A.](https://stackoverflow.com/jobs/517232/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)***[Hyderabad, India](https://stackoverflow.com/jobs/517232/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)*

[apache-sparkjava](https://stackoverflow.com/jobs/517232/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)

[Be one of the first applicants](https://stackoverflow.com/jobs/517232/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)

**[Software Engineering](https://stackoverflow.com/jobs/514701/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)**

**[JPMorgan Chase Bank, N.A.](https://stackoverflow.com/jobs/514701/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)***[Hyderabad, India](https://stackoverflow.com/jobs/514701/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)*

[pythonapache-spark](https://stackoverflow.com/jobs/514701/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)

**[Software Engineering](https://stackoverflow.com/jobs/514953/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)**

**[JPMorgan Chase Bank, N.A.](https://stackoverflow.com/jobs/514953/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)***[Hyderabad, India](https://stackoverflow.com/jobs/514953/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)*

[apache-sparkjava](https://stackoverflow.com/jobs/514953/software-engineering-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)

* 

**[DevOps Engineer](https://stackoverflow.com/jobs/514373/devops-engineer-apple?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "DevOps Engineer at Apple. Click to learn more." \t "_blank)**

**[Apple](https://stackoverflow.com/jobs/514373/devops-engineer-apple?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "DevOps Engineer at Apple. Click to learn more." \t "_blank)***[Hyderabad, India](https://stackoverflow.com/jobs/514373/devops-engineer-apple?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "DevOps Engineer at Apple. Click to learn more." \t "_blank)*

[hadoopapache-spark](https://stackoverflow.com/jobs/514373/devops-engineer-apple?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "DevOps Engineer at Apple. Click to learn more." \t "_blank)

**[Software Engineering - Big Data ETL developer](https://stackoverflow.com/jobs/512145/software-engineering-big-data-etl-developer-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering - Big Data ETL developer at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)**

**[JPMorgan Chase Bank, N.A.](https://stackoverflow.com/jobs/512145/software-engineering-big-data-etl-developer-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering - Big Data ETL developer at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)***[Hyderabad, India](https://stackoverflow.com/jobs/512145/software-engineering-big-data-etl-developer-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering - Big Data ETL developer at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)*

[hadoopsql](https://stackoverflow.com/jobs/512145/software-engineering-big-data-etl-developer-jpmorgan-chase-bank-na?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Software Engineering - Big Data ETL developer at JPMorgan Chase Bank, N.A.. Click to learn more." \t "_blank)

* 

**[Lead NLP Data Scientist](https://stackoverflow.com/jobs/475362/lead-nlp-data-scientist-tessian?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Lead NLP Data Scientist at Tessian. Click to learn more." \t "_blank)**

**[Tessian](https://stackoverflow.com/jobs/475362/lead-nlp-data-scientist-tessian?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Lead NLP Data Scientist at Tessian. Click to learn more." \t "_blank)***[London, UK](https://stackoverflow.com/jobs/475362/lead-nlp-data-scientist-tessian?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Lead NLP Data Scientist at Tessian. Click to learn more." \t "_blank)*

**[£85K - £95KREMOTERELOCATION](https://stackoverflow.com/jobs/475362/lead-nlp-data-scientist-tessian?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Lead NLP Data Scientist at Tessian. Click to learn more." \t "_blank)**

[pythonapache-spark](https://stackoverflow.com/jobs/475362/lead-nlp-data-scientist-tessian?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Lead NLP Data Scientist at Tessian. Click to learn more." \t "_blank)

* 

**[Senior Data Scientist, Commerce Analytics for Apple Media Products](https://stackoverflow.com/jobs/514415/senior-data-scientist-commerce-analytics-for-apple?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Senior Data Scientist, Commerce Analytics for Apple Media Products at Apple. Click to learn more." \t "_blank)**

**[Apple](https://stackoverflow.com/jobs/514415/senior-data-scientist-commerce-analytics-for-apple?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Senior Data Scientist, Commerce Analytics for Apple Media Products at Apple. Click to learn more." \t "_blank)***[Hyderabad, India](https://stackoverflow.com/jobs/514415/senior-data-scientist-commerce-analytics-for-apple?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Senior Data Scientist, Commerce Analytics for Apple Media Products at Apple. Click to learn more." \t "_blank)*

[machine-learningsql](https://stackoverflow.com/jobs/514415/senior-data-scientist-commerce-analytics-for-apple?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \o "Senior Data Scientist, Commerce Analytics for Apple Media Products at Apple. Click to learn more." \t "_blank)

[View more jobs on](https://stackoverflow.com/jobs?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \t "_blank)**[Stack Overflow](https://stackoverflow.com/jobs?so_medium=Ad&so_source=JobAd&so_campaign=GenericOrange" \t "_blank)**

Linked

[2](https://stackoverflow.com/q/50293714?lq=1" \o "Vote score (upvotes - downvotes))

[Spark: Convert RDD[String] to DataFrame](https://stackoverflow.com/questions/50293714/spark-convert-rddstring-to-dataframe?noredirect=1&lq=1)

[-2](https://stackoverflow.com/q/50293837?lq=1" \o "Vote score (upvotes - downvotes))

[How to read data from hdfs using scala language](https://stackoverflow.com/questions/50293837/how-to-read-data-from-hdfs-using-scala-language?noredirect=1&lq=1)

[1](https://stackoverflow.com/q/53265804?lq=1" \o "Vote score (upvotes - downvotes))

[Converting Rdd[Array[String]] to DataFrame/DataSet in scala](https://stackoverflow.com/questions/53265804/converting-rddarraystring-to-dataframe-dataset-in-scala?noredirect=1&lq=1)

[0](https://stackoverflow.com/q/45268346?lq=1" \o "Vote score (upvotes - downvotes))

[Using Scala and SparkSql and importing CSV file with header](https://stackoverflow.com/questions/45268346/using-scala-and-sparksql-and-importing-csv-file-with-header?noredirect=1&lq=1)

[22](https://stackoverflow.com/q/41254011?lq=1" \o "Vote score (upvotes - downvotes))

[SparkSQL - Read parquet file directly](https://stackoverflow.com/questions/41254011/sparksql-read-parquet-file-directly?noredirect=1&lq=1)

[19](https://stackoverflow.com/q/25362942?lq=1" \o "Vote score (upvotes - downvotes))

[Parse CSV as DataFrame/DataSet with Apache Spark and Java](https://stackoverflow.com/questions/25362942/parse-csv-as-dataframe-dataset-with-apache-spark-and-java?noredirect=1&lq=1)

[2](https://stackoverflow.com/q/47966322?lq=1" \o "Vote score (upvotes - downvotes))

[Unable to create dataframe using SQLContext object in spark2.2](https://stackoverflow.com/questions/47966322/unable-to-create-dataframe-using-sqlcontext-object-in-spark2-2?noredirect=1&lq=1)

[4](https://stackoverflow.com/q/49303286?lq=1" \o "Vote score (upvotes - downvotes))

[I am getting error while loading my csv in spark using SQlcontext](https://stackoverflow.com/questions/49303286/i-am-getting-error-while-loading-my-csv-in-spark-using-sqlcontext?noredirect=1&lq=1)

[5](https://stackoverflow.com/q/47604140?lq=1" \o "Vote score (upvotes - downvotes))

[Spark DataFrame - Read pipe delimited file using SQL?](https://stackoverflow.com/questions/47604140/spark-dataframe-read-pipe-delimited-file-using-sql?noredirect=1&lq=1)

[3](https://stackoverflow.com/q/31397845?lq=1" \o "Vote score (upvotes - downvotes))

[How to create Data frame from csv in Spark(using scala) when the first line is the schema?](https://stackoverflow.com/questions/31397845/how-to-create-data-frame-from-csv-in-sparkusing-scala-when-the-first-line-is-t?noredirect=1&lq=1)

[See more linked questions](https://stackoverflow.com/questions/linked/29704333?lq=1)

Related

[0](https://stackoverflow.com/q/27614411?rq=1" \o "Vote score (upvotes - downvotes))

[Submit Spark Job to Google Cloud Platform](https://stackoverflow.com/questions/27614411/submit-spark-job-to-google-cloud-platform?rq=1)

[6](https://stackoverflow.com/q/28493991?rq=1" \o "Vote score (upvotes - downvotes))

[Spark give Null pointer exception during InputSplit for Hbase](https://stackoverflow.com/questions/28493991/spark-give-null-pointer-exception-during-inputsplit-for-hbase?rq=1)

[7](https://stackoverflow.com/q/31525965?rq=1" \o "Vote score (upvotes - downvotes))

[How to debug Scala Future?](https://stackoverflow.com/questions/31525965/how-to-debug-scala-future?rq=1)

[2](https://stackoverflow.com/q/31693363?rq=1" \o "Vote score (upvotes - downvotes))

[Error in twitter streaming](https://stackoverflow.com/questions/31693363/error-in-twitter-streaming?rq=1)

[1](https://stackoverflow.com/q/32905535?rq=1" \o "Vote score (upvotes - downvotes))

[Upgrade ReactiveMongo Play 2.3 app from 2.6 to 3.0](https://stackoverflow.com/questions/32905535/upgrade-reactivemongo-play-2-3-app-from-2-6-to-3-0?rq=1)

[2](https://stackoverflow.com/q/39737886?rq=1" \o "Vote score (upvotes - downvotes))

[StandardTitanGraph java.lang.Throwable: Hook creation trace error](https://stackoverflow.com/questions/39737886/standardtitangraph-java-lang-throwable-hook-creation-trace-error?rq=1)

[0](https://stackoverflow.com/q/45462524?rq=1" \o "Vote score (upvotes - downvotes))

[apache spark streaming kafka integration error JAVA](https://stackoverflow.com/questions/45462524/apache-spark-streaming-kafka-integration-error-java?rq=1)

[2](https://stackoverflow.com/q/49509784?rq=1" \o "Vote score (upvotes - downvotes))

[Scala Reflection exception during creation of DataSet in Spark](https://stackoverflow.com/questions/49509784/scala-reflection-exception-during-creation-of-dataset-in-spark?rq=1)

[Hot Network Questions](https://stackexchange.com/questions?tab=hot)

* [If women are paid less for the same work, why don't employers hire just women?](https://economics.stackexchange.com/questions/42999/if-women-are-paid-less-for-the-same-work-why-dont-employers-hire-just-women)
* [The meaning of "They know who they are"](https://english.stackexchange.com/questions/562381/the-meaning-of-they-know-who-they-are)
* [Why is Bind The Monster useful?](https://boardgames.stackexchange.com/questions/54406/why-is-bind-the-monster-useful)
* [Term to describe paradox where those with less subject matter expertise can sometimes make better teachers?](https://academia.stackexchange.com/questions/163747/term-to-describe-paradox-where-those-with-less-subject-matter-expertise-can-some)
* [How would I go about fitting a window AC into a really wide window?](https://diy.stackexchange.com/questions/218625/how-would-i-go-about-fitting-a-window-ac-into-a-really-wide-window)
* [Reduce from two pure functions to one](https://mathematica.stackexchange.com/questions/241664/reduce-from-two-pure-functions-to-one)
* [How to select outermost vertices in a shape like this?](https://blender.stackexchange.com/questions/214802/how-to-select-outermost-vertices-in-a-shape-like-this)
* [Who is the true villain of Peter Pan: Peter, or Hook?](https://movies.stackexchange.com/questions/113020/who-is-the-true-villain-of-peter-pan-peter-or-hook)
* [What could a getaway driver be charged with?](https://law.stackexchange.com/questions/62008/what-could-a-getaway-driver-be-charged-with)
* [Does a cryptographic oracle have to be a server?](https://crypto.stackexchange.com/questions/88777/does-a-cryptographic-oracle-have-to-be-a-server)
* [Just intonation on a trumpet](https://music.stackexchange.com/questions/111793/just-intonation-on-a-trumpet)
* [How do car leases work?](https://money.stackexchange.com/questions/137463/how-do-car-leases-work)
* [What is the best way to turn soup into stew without using flour?](https://cooking.stackexchange.com/questions/114744/what-is-the-best-way-to-turn-soup-into-stew-without-using-flour)
* [Can you easily make an edge loop go through a triangle?](https://blender.stackexchange.com/questions/214794/can-you-easily-make-an-edge-loop-go-through-a-triangle)
* [Is it a bad sign that a rejection email does not include an invitation to apply again in the future?](https://workplace.stackexchange.com/questions/170424/is-it-a-bad-sign-that-a-rejection-email-does-not-include-an-invitation-to-apply)
* [To-scale images useful for comparing sizes of small (SOT23 and smaller) packages typical for discrete transistors or diodes?](https://electronics.stackexchange.com/questions/552737/to-scale-images-useful-for-comparing-sizes-of-small-sot23-and-smaller-packages)
* [Can the Rats of a Hat of Vermin be valid candidates to make a Swarm of Rats from a Pipe of the Sewers?](https://rpg.stackexchange.com/questions/181836/can-the-rats-of-a-hat-of-vermin-be-valid-candidates-to-make-a-swarm-of-rats-from)
* [Should we ask ambiguous questions on an exam?](https://academia.stackexchange.com/questions/163662/should-we-ask-ambiguous-questions-on-an-exam)
* [What does "Driven low/high" mean in terms of electronics?](https://electronics.stackexchange.com/questions/552764/what-does-driven-low-high-mean-in-terms-of-electronics)
* [how to not get burned keto flour pancakes?](https://cooking.stackexchange.com/questions/114757/how-to-not-get-burned-keto-flour-pancakes)
* [Why does Wien's law not work for lower frequencies?](https://physics.stackexchange.com/questions/620666/why-does-wiens-law-not-work-for-lower-frequencies)
* [Why don't we see the Milky Way out the windows in Star Trek?](https://scifi.stackexchange.com/questions/244431/why-dont-we-see-the-milky-way-out-the-windows-in-star-trek)
* [Why might radios not be effective in a post-apocalyptic world?](https://worldbuilding.stackexchange.com/questions/197667/why-might-radios-not-be-effective-in-a-post-apocalyptic-world)
* [How do I handle players that don't care for the rules I put in place as the DM and question everything I do?](https://rpg.stackexchange.com/questions/181805/how-do-i-handle-players-that-dont-care-for-the-rules-i-put-in-place-as-the-dm-a)

[Question feed](https://stackoverflow.com/feeds/question/29704333)



[**STACK OVERFLOW**](https://stackoverflow.com/)

* [Questions](https://stackoverflow.com/questions)
* [Jobs](https://stackoverflow.com/jobs)
* [Developer Jobs Directory](https://stackoverflow.com/jobs/directory/developer-jobs)
* [Salary Calculator](https://stackoverflow.com/jobs/salary)
* [Help](https://stackoverflow.com/help)
* Mobile
* Disable Responsiveness

[**PRODUCTS**](https://stackoverflowbusiness.com/)

* [Teams](https://stackoverflow.com/teams)
* [Talent](https://stackoverflow.com/talent)
* [Advertising](https://stackoverflow.com/advertising)
* [Enterprise](https://stackoverflow.com/enterprise)

[**COMPANY**](https://stackoverflow.com/company)

* [About](https://stackoverflow.com/company)
* [Press](https://stackoverflow.com/company/press)
* [Work Here](https://stackoverflow.com/company/work-here)
* [Legal](https://stackoverflow.com/legal)
* [Privacy Policy](https://stackoverflow.com/legal/privacy-policy)
* [Terms of Service](https://stackoverflow.com/legal/terms-of-service)
* [Contact Us](https://stackoverflow.com/company/contact)
* [Cookie Policy](https://stackoverflow.com/legal/cookie-policy)

[**STACK EXCHANGE  
NETWORK**](https://stackexchange.com/)

* [Technology](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)
* [Life / Arts](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)
* [Culture / Recreation](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)
* [Science](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)
* [Other](https://stackoverflow.com/questions/29704333/spark-load-csv-file-as-dataframe)
* [Blog](https://stackoverflow.blog/?blb=1)
* [Facebook](https://www.facebook.com/officialstackoverflow/)
* [Twitter](https://twitter.com/stackoverflow)
* [LinkedIn](https://linkedin.com/company/stack-overflow)
* [Instagram](https://www.instagram.com/thestackoverflow)

site design / logo © 2021 Stack Exchange Inc; user contributions licensed under [cc by-sa](https://stackoverflow.com/help/licensing). rev 2021.3.12.38768