

Announcing Stack Overflow Documentation

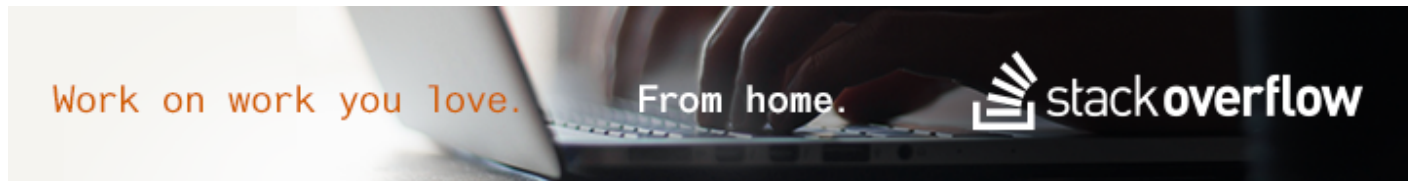
We started with Q&A. Technical documentation is next, and we need your help.

Whether you're a beginner or an experienced developer, you *can* contribute.

Sign up and start helping →

[Learn more about Documentation →](#)

saving a dataframe to JSON file on local drive in pyspark



I have a dataframe that I am trying to save as a JSON file using pyspark 1.4, but it doesn't seem to be working. When i give it the path to the directory it returns an error stating it already exists. My assumption based off the [documentation](#) was that it would save a json file in the path that you give it.

```
df.write.json("C:\Users\username")
```

Specifying a directory with a name doesn't produce any file and gives an error of "java.io.IOException: Mkdirs failed to create file:/C:/Users/username/test/_temporary/....etc. It does however create a directory of the name test which contains several sub-directories with blank crc files.

```
df.write.json("C:\Users\username\test")
```

And adding a file extension of JSON, produces the same error

```
df.write.json("C:\Users\username\test.JSON")
```

[python](#) [json](#) [apache-spark](#) [pyspark](#)

edited Jun 26 '15 at 17:48

asked Jun 26 '15 at 15:38



[TheRandomSuit](#)

962 9 28

-
- 1 I think you need to give it a complete file name, not just the directory. – [Brobin](#) Jun 26 '15 at 15:40
-
- 1 yes, i verified the permissions on that directory and used `getpass.getuser()` from python to verify that i was logged in as that user via the console. – [TheRandomSuit](#) Jun 26 '15 at 17:35
-
- 1 try an alternate approach such as `df.toJSON().saveAsTextFile(path)` – [urug](#) Jun 26 '15 at 17:41
-
- 1 I too faced such a problem when using windows.. So I changes to Linux where same code worked perfectly ... – [KcDoD](#) Jun 26 '15 at 17:53
-
- 1 Thanks for giving it a try. I figured it had something to do with Windows, ughhh.... – [TheRandomSuit](#) Jun 26 '15 at 17:54
-

2 Answers

Could you not just use

```
df.toJSON()
```

as shown [here](#)? If not, then first transform into a pandas DataFrame and then write to json.

```
pandas_df = df.toPandas()
pandas_df.to_json("C:\\Users\\username\\test.JSON")
```

answered Jun 29 '15 at 14:39



[NightHallow](#)

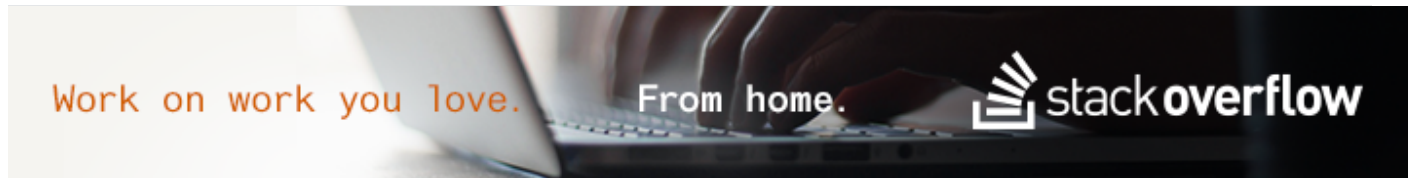
552 3 15

`df.toJSON()` doesn't seem to accept an array, but if I pass it a single line it works. i'm trying to debug this more. – [TheRandomSuit](#) Jun 29 '15 at 17:35

converting to a Pandas dataframe works perfect, I would probably just use a Pandas dataframe the entire time, unless there are memory or processing issues that would arise from a much larger data set. –

[TheRandomSuit](#) Jun 29 '15 at 17:41

Yeah, I use DataFrames as often as I can. If memory becomes a problem take a look at [Dask](#) – [NightHallow](#)
Jul 1 '15 at 14:37



I would avoid using `write.json` since its causing problems on Windows. Using Python's file writing should skip creating the temp directories that are giving you issues.

```
with open("C:\\Users\\username\\test.json", "w+") as output_file:  
    output_file.write(df.toJSON())
```

[edited Jun 29 '15 at 17:17](#)

[answered Jun 29 '15 at 14:16](#)



[Brobin](#)

1,276 1 5 20

If I use `output_file.write(df.toJSON())` it produces `TypeError: expected character buffer object`, i'm assuming it is passing it an array which then causes the failure because if I use `output_file.write(df.toJSON().first())` it will successfully create the JSON file with only one line in it. – [TheRandomSuit](#) Jun 29 '15 at 17:16

Great! I added the escape slashes to my answer. – [Brobin](#) Jun 29 '15 at 17:17