

## 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 →](#)

## Getting specific field from chosen Row in Pyspark DataFrame

Work on work you love. From home.



I have a Spark DataFrame built through *pyspark* from a JSON file as

```
sc = SparkContext()
sqlc = SQLContext(sc)

users_df = sqlc.read.json('users.json')
```

Now, I want to access a *chosen\_user* data, where this is its `_id` field. I can do

```
print users_df[users_df._id == chosen_user].show()
```

and this gives me the full Row of the user. But suppose I just want one specific field in the Row, say the user gender, how would I obtain it?

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

edited Mar 1 at 11:04



zero323

69k 16 82 148

asked Mar 1 at 10:23



martina

633 12 25

## 1 Answer

Just filter and select:

```
result = users_df.where(users_df._id == chosen_user).select("gender")
```

or with `col`

```
from pyspark.sql.functions import col
```

```
result = users_df.where(col("_id") == chosen_user).select(col("gender"))
```

Finally PySpark `Row` is just a `tuple` with some extensions so you can for example

`flatMap` :

```
result.rdd.flatMap(list).first()
```

or `map` with something like this:

```
result.rdd.map(lambda x: x.gender).first()
```

edited Mar 1 at 12:28

answered Mar 1 at 10:29



zero323

69k 16 82 148

Can you also comment on the way to extract the actual value and put into a variable maybe? With your solution I get `[Row(gender=u'male')]` when running an action 'collect' at the end. I'd like to really put 'male' into a variable then. – [martina](#) Mar 1 at 12:01

Sure. Please check the edit. – [zero323](#) Mar 1 at 12:28

Perfect, thanks very much. – [martina](#) Mar 1 at 13:18

