

The results are in! See what nearly 90,000 developers picked as their most loved, dreaded, and desired coding languages and more in the 2019 Developer Survey.

Spark sql group by and sum changing column name?

[Ask Question](#)

1

In this data frame I am finding total salary from each group. In Oracle I'd use this code



```
select job_id,sum(salary) as "Total" from hr.employees group by job_id;
```



1

In Spark SQL tried the same, I am facing two issues

```
empData.groupBy($"job_id").sum("salary").alias("Total").show()
```

1. The alias total is not displaying instead it is showing "sum(salary)" column
2. I could not use \$ (I think Scala SQL syntax). Getting compilation issue

```
empData.groupBy($"job_id").sum($"salary").alias("Total").show()
```

Any idea?

[scala](#)[apache-spark](#)

edited Oct 11 '18 at 10:43



[user10465355](#)

2,139 2 5 21

asked Oct 11 '18 at 10:07



[Learn Hadoop](#)

449 4 17

Possible duplicate of [Column alias after groupBy in pyspark](#) – [user10465355](#) Oct 11 '18 at 10:28

1 Answer



1

Use Aggregate function `.agg()` if you want to provide alias name. This accepts scala syntax (`$" "`)



```
empData.groupBy($"job_id").agg(sum($"s
```

By using our site, you acknowledge that you have read and understand our [Cookie Policy](#), [Privacy Policy](#), and our [Terms of Service](#).

```
empData.groupBy($"job_id").sum("salary",  
$"sum(salary)".alias("Total")).show()
```

edited Oct 11 '18 at 10:24

answered Oct 11 '18 at 10:14



vdep

2,084 1 18 33

perfect.. thanks a lot – [Learn Hadoop](#) Oct 11 '18 at 10:27
