

# PySpark Count Ratio Calculation

June 2, 2023

generate a pyspark code to compute ratio by dividing current day count with historic 15 days average

Certainly! Here's an example of a PySpark code snippet that computes the ratio by dividing the count of the current day with the average count of the previous 15 days:

python

 Copy code

```
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, datediff, sum, avg, when

# Create a SparkSession
spark = SparkSession.builder.getOrCreate()

# Read the input data into a DataFrame
data = spark.read.csv("your_input_data.csv", header=True)

# Convert the date column to a date type
data = data.withColumn("date", col("date").cast("date"))

# Compute the count for the current day
current_day_count = data.filter(col("date") == "2023-06-02").groupBy().agg(

# Compute the average count for the previous 15 days
historic_average = data.filter(col("date") >= "2023-05-18" and col("date"

# Compute the ratio
ratio = when(historic_average != 0, current_day_count / historic_average)

# Display the result
ratio.show()
```

[Continue this conversation](#)