

Task - Show account wise net balance

In the given dataset , **ACC001** related 3 entries are there. Two are Debit transaction types , and one is Credit transaction type.Hence, the net balance for ACC001 is -100-50+300 = +150. Similarly, check the net balance for each account number considering the transaction type.

Input					Output	
Sr_No	Date	acc_no	transaction_type	amount	acc_no	net_balance
1	2024-01-01	ACC001	Debit	100	ACC001	150
2	2024-01-02	ACC001	Debit	50	ACC002	-100
3	2024-01-03	ACC001	Credit	300		
4	2024-01-01	ACC002	Credit	100		
5	2024-01-04	ACC002	Debit	200		

Import Libraries

```
In [6]: from pyspark.sql import SparkSession
from pyspark.sql.functions import *
```

Create Spark session

```
In [2]: spark = SparkSession.builder.appName("Net-balance-app").getOrCreate()
```

Create dataframe

```
In [7]: data = [
    (1, '2024-01-01', 'ACC001', 'Debit', 100),
    (2, '2024-01-02', 'ACC001', 'Debit', 50),
    (3, '2024-01-03', 'ACC001', 'Credit', 300),
    (4, '2024-01-01', 'ACC002', 'Credit', 100),
    (5, '2024-01-04', 'ACC002', 'Debit', 200)

]
_schema = "Sr_No integer, Date string , acc_no string , transaction_type string , amount integer"
df = spark.createDataFrame(data , _schema)
df.show()
```

Sr_No	Date	acc_no	transaction_type	amount
1	2024-01-01	ACC001	Debit	100
2	2024-01-02	ACC001	Debit	50
3	2024-01-03	ACC001	Credit	300
4	2024-01-01	ACC002	Credit	100
5	2024-01-04	ACC002	Debit	200

Create a new column 'net_balance' which shows account wise net balance considering both Credit and Debit transaction types

Using Spark Dataframe API

```
In [8]: df_net_balance= df.groupby("acc_no") \
    .agg(
        sum(when(df.transaction_type =='Credit' , df.amount).otherwise(0)).alias("Credit"),
        sum(when(df.transaction_type =='Debit' , df.amount).otherwise(0)).alias("Debit")
    ) \
    .withColumn("net_balance" , col("Credit") - col("Debit")) \
    .select("acc_no" , "net_balance")

df_net_balance.show()
```

acc_no	net_balance
ACC001	150
ACC002	-100

Using PySpark SQL

Create a temporary view

```
In [9]: df.createOrReplaceTempView("account_details")
```

Write Spark SQL to get account wise net balance

```
In [15]: df_net_balance= spark.sql ('''
        SELECT acc_no,
        (SUM( CASE WHEN transaction_type == 'Credit' THEN amount END ) -
        SUM( CASE WHEN transaction_type == 'Debit' THEN amount END ))
        AS net_balance
        FROM account_details
        GROUP BY acc_no
        ''')
```

```
In [16]: df_net_balance.show()
```

acc_no	net_balance
ACC001	150
ACC002	-100

Thank You !

<https://www.linkedin.com/in/lipsa-biswas/>