Master PySpark: From Zero to Big Data Hero!!

Joins Part 4

Coding Question:

1. Write a PySpark query to create a DataFrame that lists each employee along with their manager's name. Display columns employee and manager.
2. Modify the code to find and display only the employee(s) who do not have a manager (CEO-level employees). Display columns employee and manager.
3. Extend the code to find all employees who directly report to "Manager A." Display columns empid, ename, and mrgid.
4. Write a query to determine the hierarchy level of each employee, where the CEO is level 1, direct reports to the CEO are level 2, and so on. Display columns empid, ename, mrgid, and level.

```
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, expr

# Create a Spark session
spark =
SparkSession.builder.appName("EmployeeHierarchy").getOrCreate()

# Sample data
data = [
    (1, None, "CEO"),
    (2, 1, "Manager A"),
    (3, 1, "Manager B"),
    (4, 2, "Employee X"),
    (5, 3, "Employee Y"),
]
columns = ["empid", "mrgid", "ename"]
employee_df = spark.createDataFrame(data, columns)
# Display the result
print("emp_data:")
employee_df.show()
```

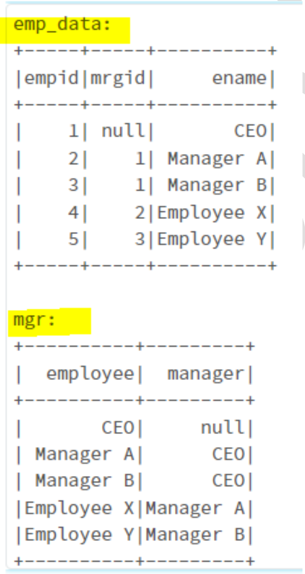
Follow me on LinkedIn – Shivakiran kotur

```
# Self-join to find the manager and CEO
manager_df = employee_df.alias("e") \
    .join(employee_df.alias("m"), col("e.mrgid") == col("m.empid"),
"left") \
    .select(
        col("e.ename").alias("employee"),
        col("m.ename").alias("manager")
    )

# Display the result
print("mgr:")
manager_df.show()

#  filter for employees without a manager (CEO)
manager_df2 = employee_df.alias("e1") \
    .join(employee_df.alias("m1"), col("e1.mrgid") ==
col("m1.empid"), "left") \
    .select(
        col("e1.ename").alias("employee"),
        col("m1.ename").alias("manager")
    ) \
    .filter(col("manager").isNull())

# Display the result
manager_df2.show()
```



```
emp_data:
+-----+-----+----------+
|empid|mrgid|     ename|
+-----+-----+----------+
|    1| null|       CEO|
|    2|    1| Manager A|
|    3|    1| Manager B|
|    4|    2|Employee X|
|    5|    3|Employee Y|
+-----+-----+----------+

mgr:
+----------+---------+
|  employee|  manager|
+----------+---------+
|       CEO|     null|
| Manager A|      CEO|
| Manager B|      CEO|
|Employee X|Manager A|
|Employee Y|Manager B|
+----------+---------+
```

Follow me on LinkedIn – Shivakiran kotur