



**10\_Employee Bonus**

Write a solution to report the name and bonus amount of each employee with a bonus less than 1000.

Return the result table in any order

**Difficult Level :** EASY

**DataFrame:**

**# Define the schema for the "Employee"**

**employee\_schema = StructType([**

**StructField("empId", IntegerType(), True),**

**StructField("name", StringType(), True),**

**StructField("supervisor", IntegerType(), True),**

**StructField("salary", IntegerType(), True)**

**])**

**# Define data for the "Employee"**

**employee\_data = [**

**(3, 'Brad', None, 4000),**

**(1, 'John', 3, 1000),**

**(2, 'Dan', 3, 2000),**

**(4, 'Thomas', 3, 4000)**

**]**

**# Define the schema for the "Bonus"**

**bonus\_schema = StructType([**

**StructField("empId", IntegerType(), True),**

**StructField("bonus", IntegerType(), True)**

**])**

**# Define data for the "Bonus"**

**bonus\_data = [**

**(2, 500),**

**(4, 2000)**

**]**



**INPUT**

| **INPUT-1 EMPLOYEE** | | | |
| --- | --- | --- | --- |
| **EMPID** | **NAME** | **SUPERVISOR** | **SALARY** |
| **3** | **Brad** |  | **4,000** |
| **1** | **John** | **3** | **1,000** |
| **2** | **Dan** | **3** | **2,000** |
| **4** | **Thomas** | **3** | **4,000** |

| **INPUT-2 BONUS** | |
| --- | --- |
| **EMPID** | **BONUS** |
| **2** | **500** |
| **4** | **2,000** |

**OUTPUT**

| **OUTPUT** | |
| --- | --- |
| **NAME** | **BONUS** |
| **Brad** |  |
| **John** |  |
| **Dan** | **500** |



**# Creating Spark Session**

**from pyspark.sql import SparkSession**

**from pyspark.sql.types import StructType,StructField,IntegerType,StringType**

**#creating spark session**

**spark = SparkSession. \**

**builder. \**

**config('spark.shuffle.useOldFetchProtocol', 'true'). \**

**config('spark.ui.port','0'). \**

**config("spark.sql.warehouse.dir", "/user/itv008042/warehouse"). \**

**enableHiveSupport(). \**

**master('yarn'). \**

**getOrCreate()**

**# Define the schema for the "Employee"**

**employee\_schema = StructType([**

**StructField("empId", IntegerType(), True),**

**StructField("name", StringType(), True),**

**StructField("supervisor", IntegerType(), True),**

**StructField("salary", IntegerType(), True)**

**])**

**# Define data for the "Employee"**

**employee\_data = [**

**(3, 'Brad', None, 4000),**

**(1, 'John', 3, 1000),**

**(2, 'Dan', 3, 2000),**

**(4, 'Thomas', 3, 4000)**

**]**

**# Define the schema for the "Bonus"**

**bonus\_schema = StructType([**

**StructField("empId", IntegerType(), True),**

**StructField("bonus", IntegerType(), True)**

**])**

**# Define data for the "Bonus"**

**bonus\_data = [**

**(2, 500),**

**(4, 2000)**

**]**

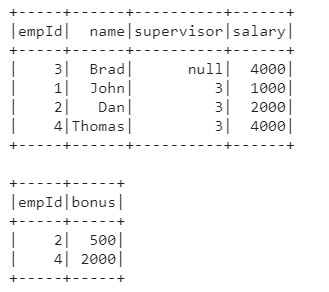
**# Create a PySpark DataFrame**

**emp\_df = spark.createDataFrame(employee\_data,employee\_schema)**

**bonus\_df = spark.createDataFrame(bonus\_data,bonus\_schema)**

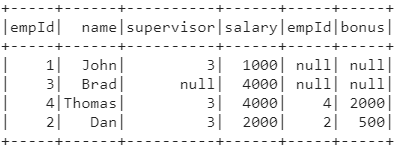
**emp\_df.show()**

**bonus\_df.show()**

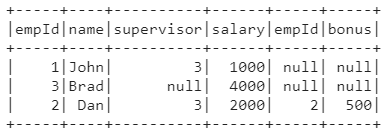
****

**join\_df=emp\_df.join(bonus\_df,emp\_df.empId==bonus\_df.empId,"left")**

**join\_df.show()**

****

**join\_df.filter( (join\_df.bonus < 1000) | col("bonus").isNull() ).show()**

****

