

# Master PySpark: From Zero to Big Data Hero!!

## Windows Function in PySpark Part 4

highest sal in dept

```
from pyspark.sql import functions as F
from pyspark.sql.window import Window
# Updated sample data for employees
emp_data = [(1, "Alice", 1, 6300),
            (2, "Bob", 1, 6200),
            (3, "Charlie", 2, 7000),
            (4, "David", 2, 7200),
            (5, "Eve", 1, 6300),
            (6, "Frank", 2, 7100)]

# Sample data for departments
dept_data = [(1, "HR"),
            (2, "Finance")]

# Create DataFrames for employees and departments
emp_df = spark.createDataFrame(emp_data, ["EmpId", "EmpName",
"DeptId", "Salary"])
dept_df = spark.createDataFrame(dept_data, ["DeptId", "DeptName"])
```

```
+-----+-----+-----+-----+
| EmpId | EmpName | DeptId | Salary |
+-----+-----+-----+-----+
| 1 | Alice | 1 | 6300 |
| 2 | Bob | 1 | 6200 |
| 3 | Charlie | 2 | 7000 |
| 4 | David | 2 | 7200 |
| 5 | Eve | 1 | 6300 |
| 6 | Frank | 2 | 7100 |
+-----+-----+-----+-----+
```

```
+-----+-----+
| DeptId | DeptName |
+-----+-----+
| 1 | HR |
| 2 | Finance |
+-----+-----+
```

```
# Window specification for ranking salaries within each department
window_spec =
Window.partitionBy("DeptId").orderBy(F.desc("Salary"))

# Add a rank column based on the highest salary within each
department
ranked_salary_df = emp_df.withColumn("Rank",
F.rank().over(window_spec))

# Filter to get only the top rank (highest salary) for each
department
result_df = ranked_salary_df.filter(F.col("Rank") == 1)
print("result_df")
result_df.show()
```

```
+-----+-----+-----+-----+
| EmpId | EmpName | DeptId | Salary | Rank |
+-----+-----+-----+-----+
|    1  |  Alice  |    1   |  6300  |    1  |
|    5  |   Eve   |    1   |  6300  |    1  |
|    4  |  David  |    2   |  7200  |    1  |
+-----+-----+-----+-----+
```

```
# Join the department names to get department names
result_df = result_df.join(dept_df, ["DeptId"], "left")

# Show the employees with the highest salary in each department
result_df.select("EmpName", "DeptName", "Salary").show()
```

```
+-----+-----+-----+
| EmpName | DeptName | Salary |
+-----+-----+-----+
|  Alice  |    HR   |  6300  |
|   Eve   |    HR   |  6300  |
|  David  | Finance |  7200  |
+-----+-----+-----+
```