**Please answer the following questions and attach to your resume.**

**Generic SQL Questions:**

Sample table to be used for solutions below –

Employee ( empid integer, mgrid integer, deptid integer, salary integer)

Dept (deptid integer, deptname text)

**1.** Find employees who do not manage anybody.

**2.** Find departments that have maximum number of employees. (solution should consider scenario having more than 1 departments that have maximum number of employees). Result should only have following information for selected department - deptname, count of employees sorted by deptname.

**3.** Find top 3 employees (salary based) in every department. Result should have deptname, empid, salary sorted by deptname and then employee with high to low salary.

**4.** List all employees, their salary and the salary of the person in their department who makes the most money but less than the employee.

Sample source data (Do not assume E\_ID will always be sequential for your solution):

|  |  |  |
| --- | --- | --- |
| E\_ID | D\_ID | SALARY |
| 100 | 201 | 10,000 |
| 101 | 201 | 9,500 |
| 102 | 201 | 11,000 |
| 103 | 205 | 10,500 |
| 104 | 205 | 8,000 |
| 105 | 205 | 9,500 |

Expected Output:

|  |  |  |  |
| --- | --- | --- | --- |
| E\_ID | D\_ID | SALARY | OTHER\_SALARY |
| 100 | 201 | 10,000 | 9,500 |
| 101 | 201 | 9,500 | Null |
| 102 | 201 | 11,000 | 10,000 |
| 103 | 205 | 10,500 | 9,500 |
| 104 | 205 | 8,000 | Null |
| 105 | 205 | 9,500 | 8,000 |

**Hive Specific Questions:**

LoudAcre Mobile is a mobile phone service provider that is moving a portion of their customer analytics workload to Hadoop. Before they can use their customer data, they want you to clean it and make it consistent.

Errors were found while looking at the customer records. Unfortunately, different input methods wrote date fields in different formats. Your task is to standardize these date fields into a consistent format.

Data Description

The Hive metastore contains a database named problem1 that contains a table named customer. The customertable contains 90 million customer records (90,000,000), each with a birthday field.

Sample Data (birthday is in bold)

1904287 Christopher Rodriguez Jan 11, 2003

96391595 Thomas Stewart 6/17/1969

2236067 John Nelson 08/22/54

Output Requirements

**5.** Create a new table named solution in the problem1 database of the Hive metastore

Your solution table must have its data stored in the HDFS directory /user/cert/problem1/solution

Your solution table must have exactly the same columns as the customer table in the same order, as well as keeping the existing file format

For every row in the solution table, replace the contents of the birthday field with a date string in “MM/DD/YY” format.

MM is the zero-padded month (01-12),

DD is the zero-padded day (01-31),

YY is the zero-padded 2-digit year (00-99)

**Python / Scala or Java Algorithms**

* Describe the code to find the greatest common denominator of two numbers.