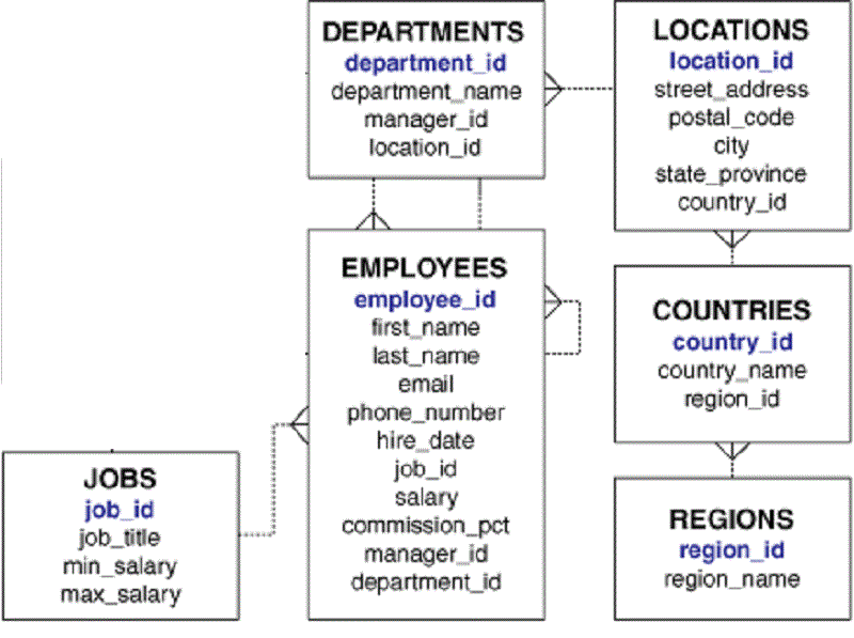
Consider the following employee-related tables in MySQL.

**Tables:**

* employees\_table (contains employee details such as id, name, email, phone, salary, department etc.)
* departments\_table (contains department details such as department id, name, location etc.)
* locations\_table (contains location information such as street name, city, pin, state, country)
* jobs\_table (contains job metadata such as job id, job name etc.)
* regions\_table (contains region details such as region id, region name)
* countries\_table (contains county details such as country id, country name)



As a Big Data Engineer, you have to perform the following tasks:

* Import all these MySQL tables to Hive tables. Hive DDLs can be inferred from table schemas in MySQL.
* Create a new table (i.e. employee\_scd) which implements a slowly changing dimension (type 2) for the employees table.
* SCD table should have additional columns, start\_date, end\_date, and active\_flag.
* Start\_date and end\_date behave as per the concept of SCD type 2. The active flag can be 0 or 1 depending on which row in the table contains the latest information of the employee.
* There is a delta file present at the HDFS path '**hdfs:///bigdatapgp/common\_folder/assignment8/emp\_details/Employeedelta.txt**' that contains the changes in employee information till the current date.
* Based on changes present in the delta file (on columns such as email, phone\_number, job\_id, department\_id, manager\_id, salary, commission\_pct) maintain the change history using start\_date, end\_date and active flag columns with old information being closed with end\_date = (current\_date -1).
* Create an updated employee table with the latest information on all the employees.
* Upload the updated employee details to a MySQL table.

**SCD Type 2 Example:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **id** | **Name** | **dept\_id** | **manager\_id** | **start\_date** | **end\_date** | **active\_flag** |
| 1 | Abc | 10 | 111 | 2019-01-01 | 9999-12-31 | 1 |
| 2 | Def | 10 | 111 | 2019-01-01 | 2019-12-08 | 0 |
| 2 | Def | 20 | 111 | 2019-12-09 | 9999-12-31 | 1 |

**For employee 2, the department got changed on 2019-12-09 from 10 to 20, hence in SCD type 2, there will be 2 records for employee id 2.**

**The schema of the output table is mentioned below. The SCD table needs to be present in the Hive whereas the updated employee table needs to be present in MySQL.**

**Output the result as per the schema mentioned in the below tables.**

1. **Employee\_scd Output Table in Hive:**

CREATE TABLE edureka\_dw.employees\_scd

   ( employee\_id int

   , first\_name VARCHAR(20)

   , last\_name VARCHAR(25)

   , email VARCHAR(25)

   , phone\_number VARCHAR(20)

   , hire\_date varchar(12)

   , job\_id VARCHAR(10)

   , salary decimal

   , commission\_pct decimal(2,2)

   , manager\_id int

   , department\_id int

   , start\_date string

   , end\_date string

   , active\_flag char(1)

   , row\_insertion\_dttm string

   ) ;

1. **Updated Employee Detail Table in MySQL:**

CREATE table edureka\_dw.employee\_details\_latest

   (employee\_id int,

   job\_id varchar(10),

   manager\_id int,

   department\_id int,

   location\_id int,

   country\_id char(2),

   first\_name varchar(20),

   last\_name varchar(25),

   salary decimal,

   commission\_pct decimal(2,2),

   department\_name varchar(30),

   job\_title varchar(35),

   city varchar(30),

   state\_province varchar(25),

   country\_name varchar(40),

   region\_name varchar(25)

   );

**Output Format:**You need to paste the commands and the code used to solve this problem.

[HTML EditorKeyboard Shortcuts](https://claas-canvas.corestack.io/courses/5/quizzes/119/take)

**ANSWER TO THE ABOVE QUESTION**

****