## Title: ITMD526\_Assignment\_05

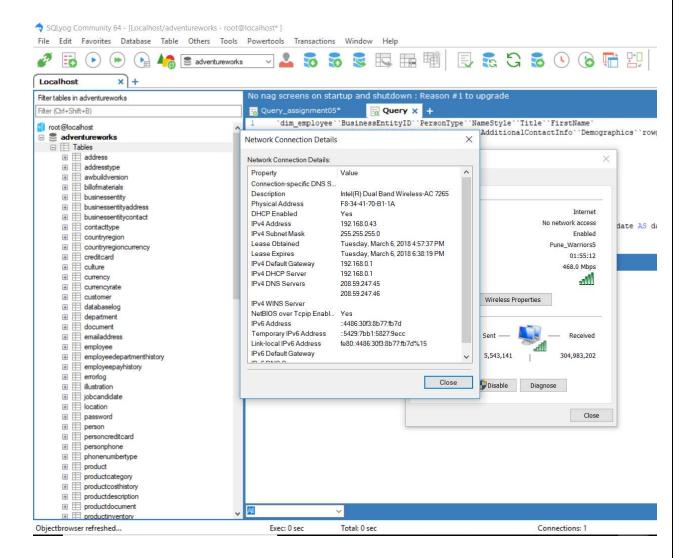
First Name	Last Name	CWID
Arpit	Khandekar	A20409171

### Table of Contents

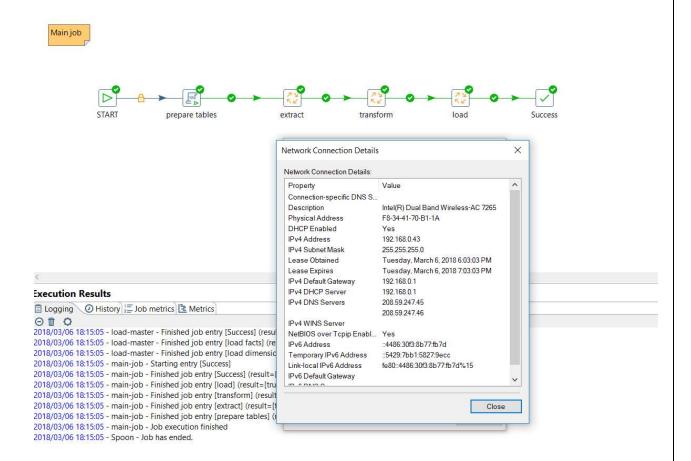
1. Normalization part of Kettle implementation	2
2. Extraction of four data tables(person, employee, employeedepartmenthistory and departmenthistory	rtment ) for
creation of dim_employee	3
3. To load database in dim employee table into salesmart db database	7

#### 1. Normalization part of Kettle implementation.

- 1. The goal is to learn how normalization works with the help of ETL transformation.
- 2. Before running the transformation, we need to import adventureworks database in our local MYSQL database, we are able to see the database name 'adventureworks' in our database.

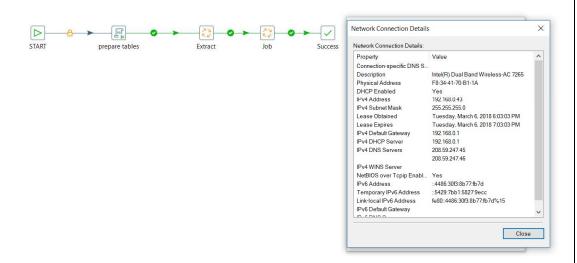


3. Job has run successfully.



- 2. Extraction of four data tables(person, employee, employeedepartmenthistory and department) for creation of dim\_employee.
  - 1. Now, Create a new main job which will contain script file 'prepare tables', and two jobs 'Extract' and 'Job'.

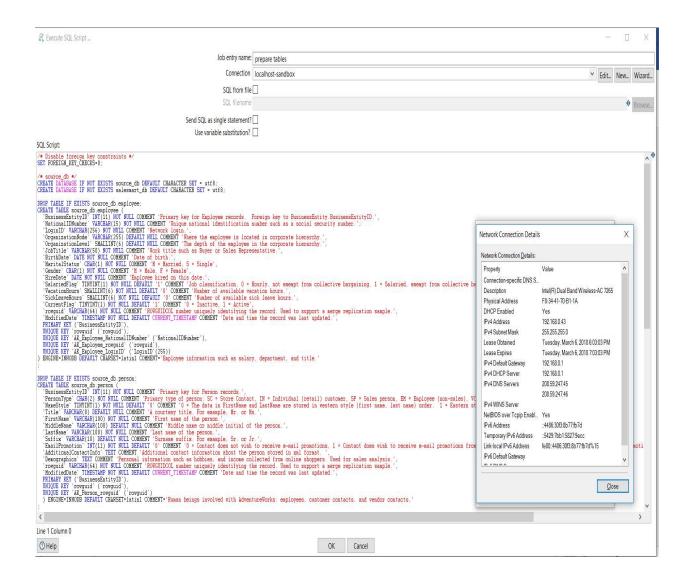
Assignment 05\_Arpit Khandekar



2. In SQL script we need to create a new Database source\_db and salesmart\_db, also we need to write SQL to create a database tables in both database.

SQL Script contains following:

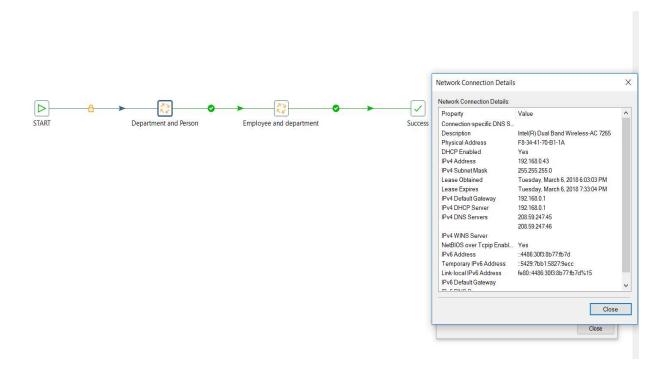
- a) Setting foreign\_key\_checks = zero;
- b) Creating two databases source\_db and salesmart\_db
- c) Creating tables employee, person, employeedepartmenthistory and department into source\_db database.
- d) Creating dim\_employee into database salesmart\_db database.
- e) Setting Foreign key checks = 1 at the end.



3. Now we will link another job to 'Extract job' of main where we will create two Jobs 'Department and Person and Employee and Department' which will contain another job which is required transformations, In those transformation we will table input connected to localhost-adventureworks database and connecting it to table output having connection localhost-source\_db.

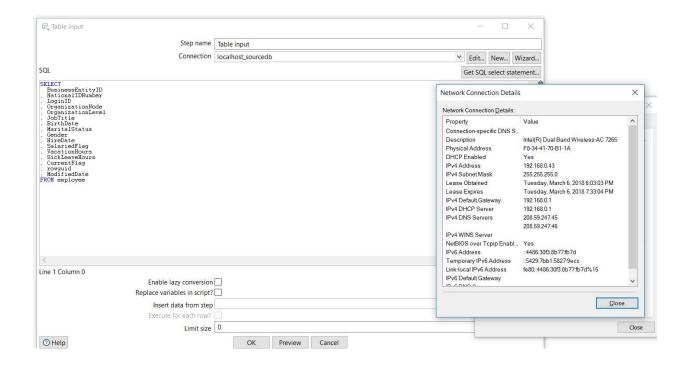


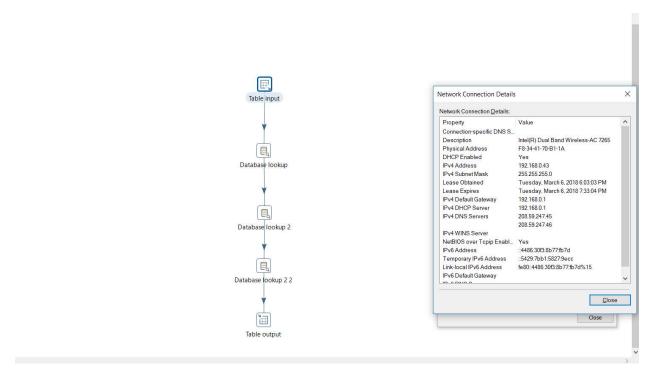
4. Now merge those two jobs to one job extract\_jobinfomain and run those jobs.



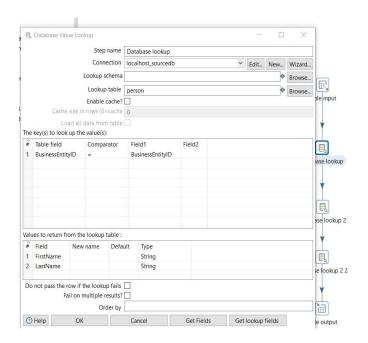
# 3. To load database in dim\_employee table into salesmart\_db database

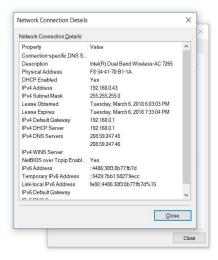
- 1. Now, we will connect a job with main job which contains two SQL script, Transformation.
- 2. Here, at first script we will set foreign key check = 0 afterwards we will set foreign key check =0 and in other script we will set foreign key check =1.
- 3. In this transformation we will select Table input and Table output where we will use three database lookup to join table person, department, employee and employeedepartmenthistory by specifying the values in lookup table for every table.



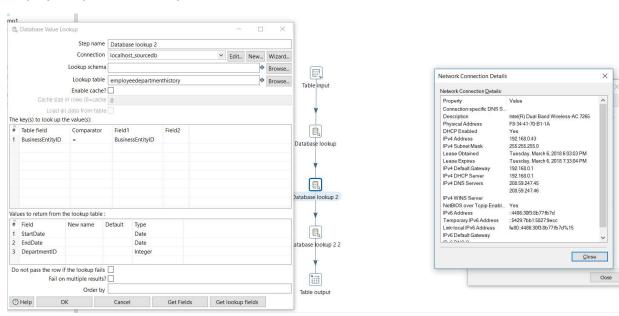


#### Person Databaselookup table:

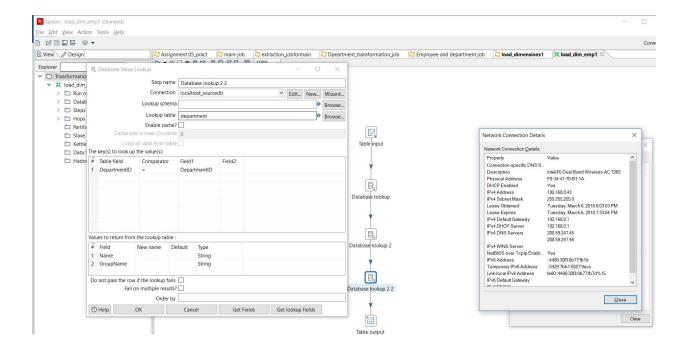




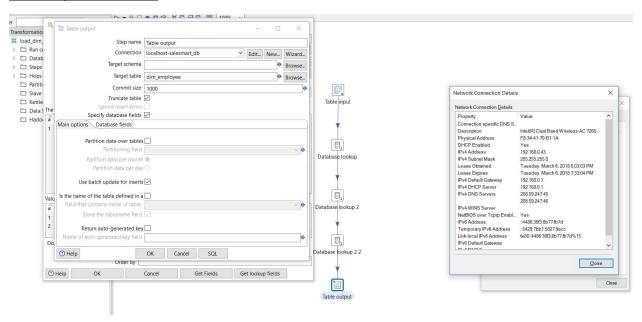
#### **Employeedepartmenthistory table:**



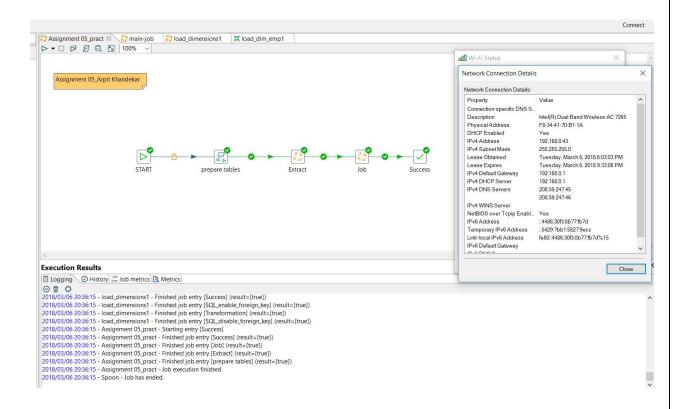
#### Department table:



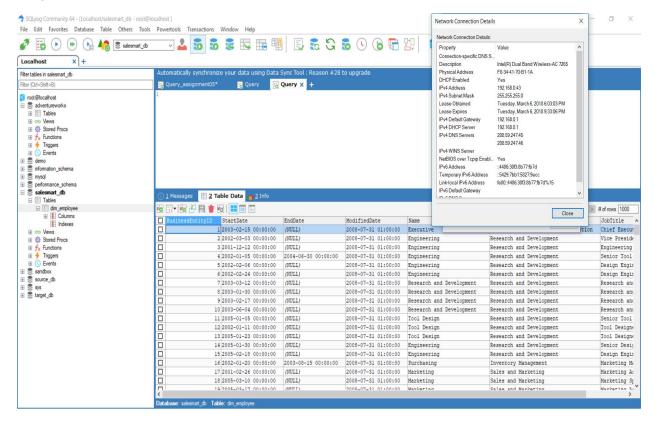
#### **Tableoutput screenshot:**



4. At the end, we will the run the main job 'Assignment05\_pract.kjb'

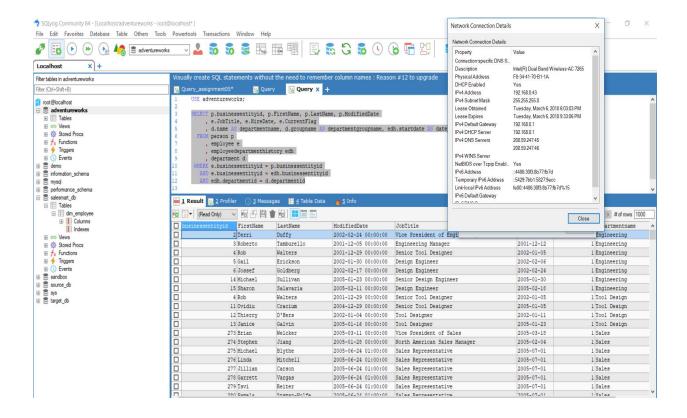


5. We can check the final output of job from database and data can be compared with the query provided.



6. We will compare the job result with the below query result:

```
SELECT p.businessentityid, p.FirstName, p.LastName, p.ModifiedDate
, e.JobTitle, e.HireDate, e.CurrentFlag
, d.name AS departmentname, d.groupname AS departmentgroupname, edh.startdate AS
date_from, IFNULL(edh.enddate, '2999-12-31') AS date_to
FROM person p
, employee e
, employeedepartmenthistory edh
, department d
WHERE e.businessentityid = p.businessentityid
AND e.businessentityid = edh.businessentityid
AND edh.departmentid = d.departmentid
;
```



Hence, we have **denormalize** the four tables in the **source\_db** database and loaded into a **dimension** table called **dim\_employee** in the **salesmart\_db**.