Title: ITMD526_Assignment_04

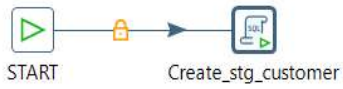
First Name	Last Name	CWID
Arpit	Khandekar	A20409171

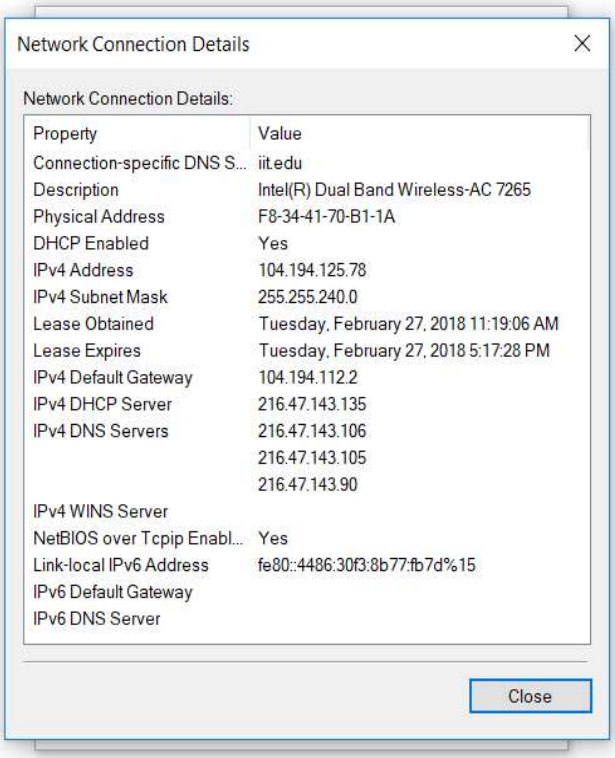
Table of Contents

1. Extracting the excel source to stg_customer table.

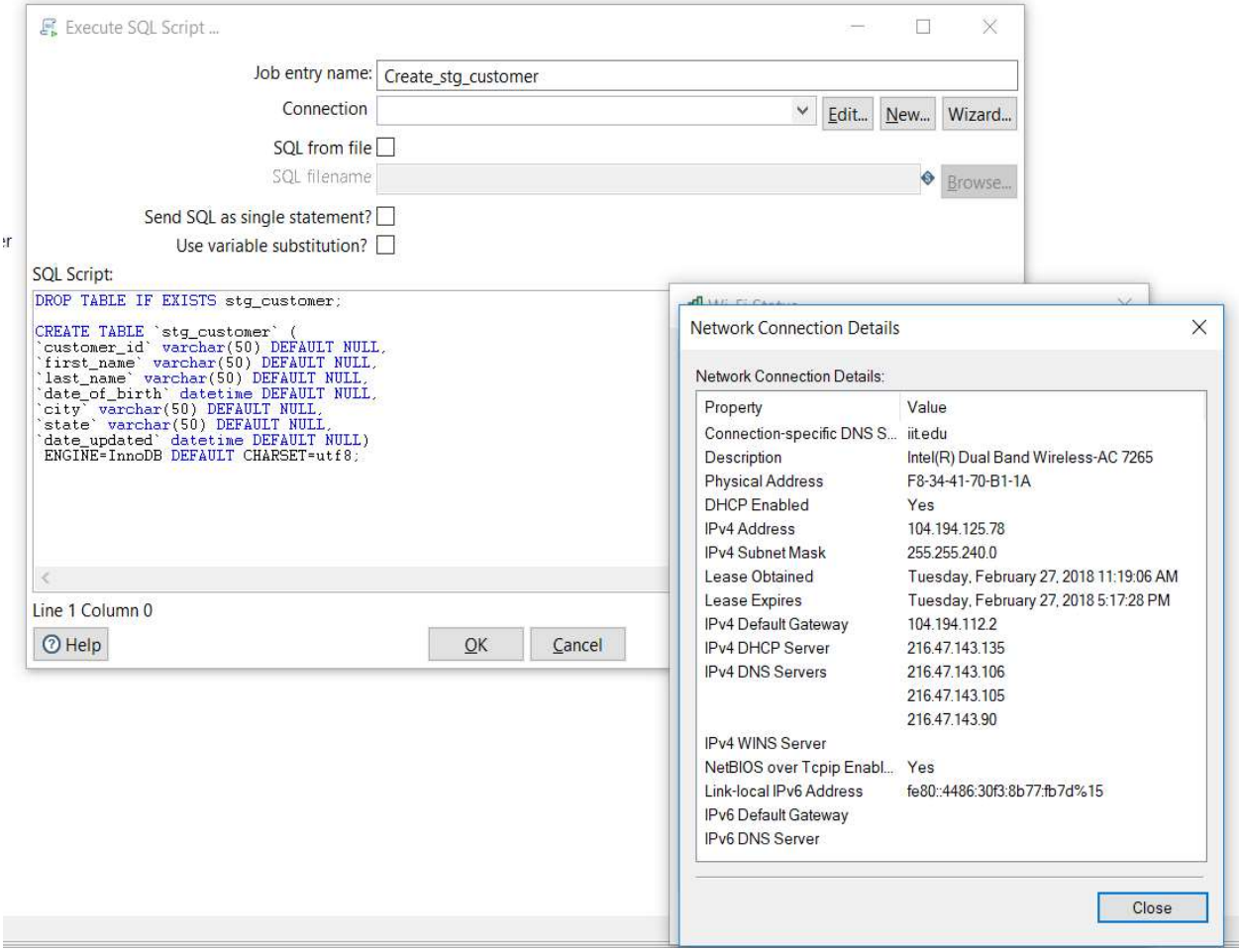
1. First, we need to take new job which start with **Start** input as an input for the job.

2. Then in Pentaho tool I have drag and drop SQL as scripting input named as **'Create stg_customer'**

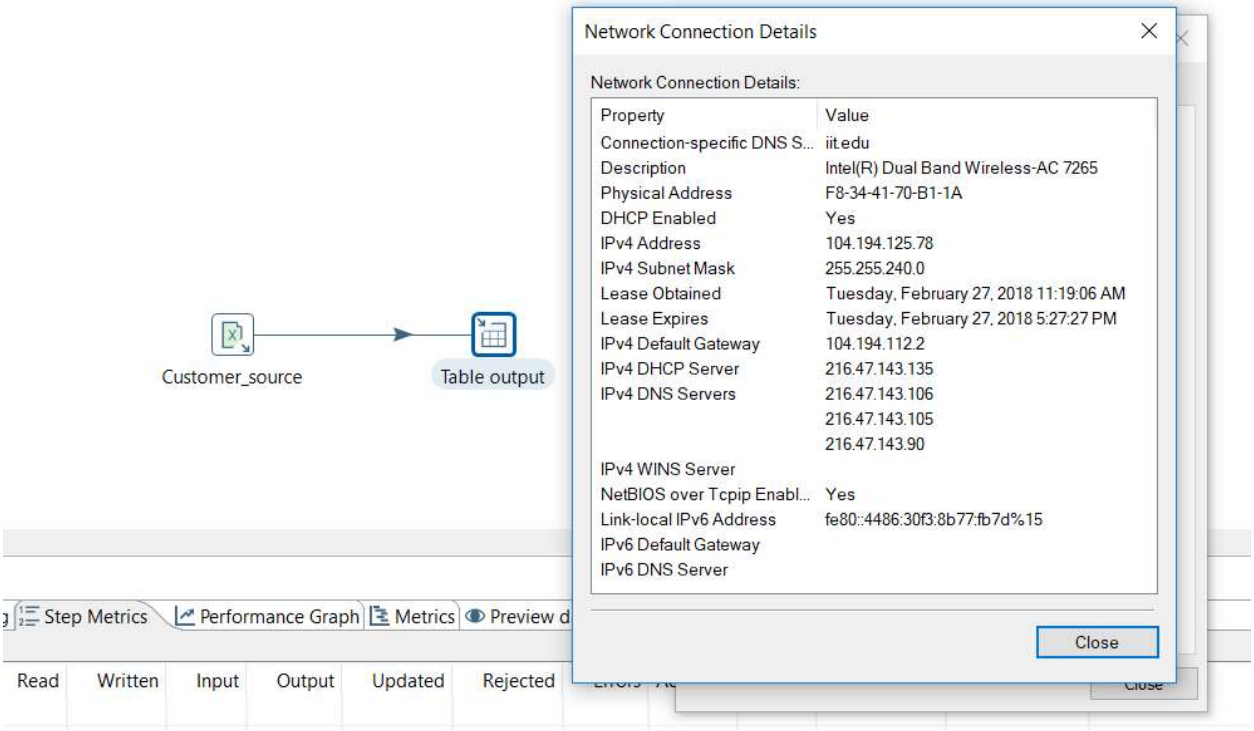




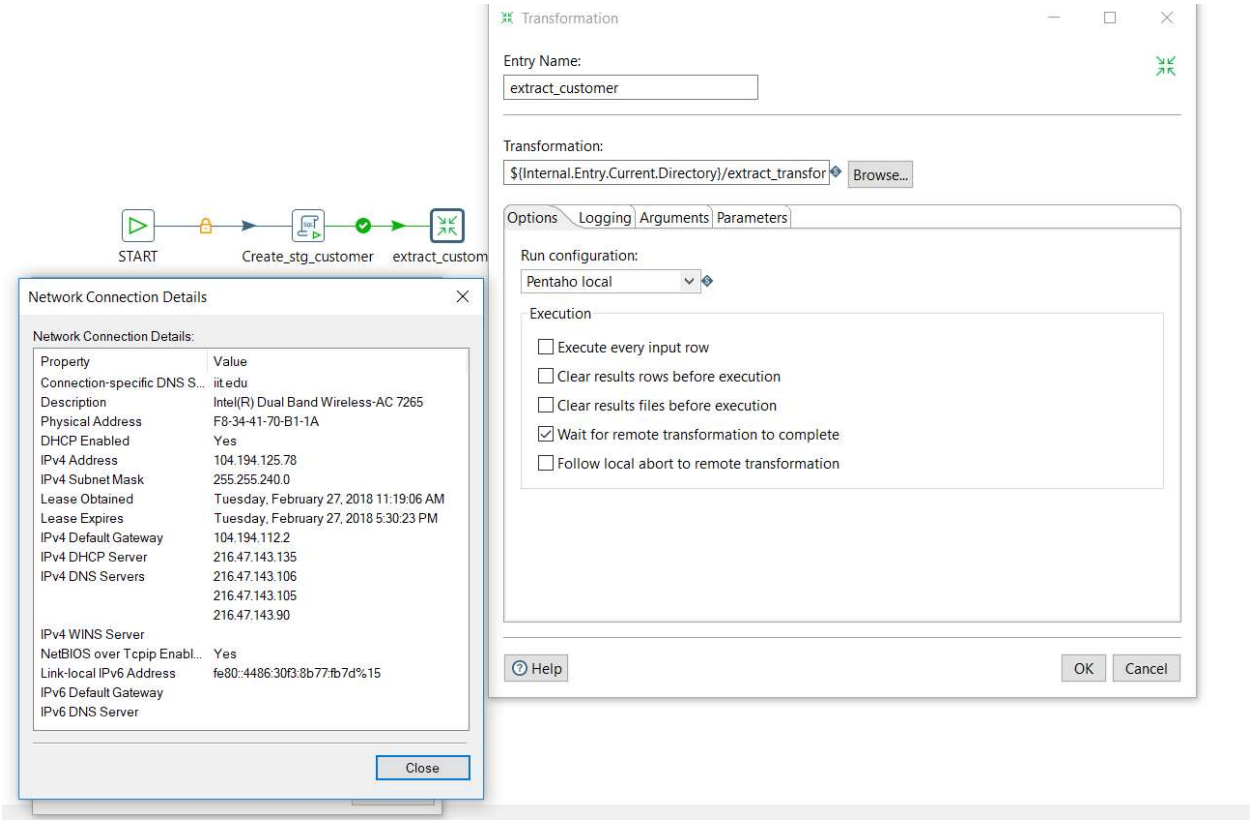
3. We will write **DDL statement** to create stg_customer in SQL script **'Create_stg_customer'**, also we will mention to drop table if it is already existing in database.


Execute SQL Script ...
Job entry name:
Create_stg_customer
Connection
Edit...
New...
Wizard...
SQL from file
SQL filename
Browse...
Send SQL as single statement?
Use variable substitution?
SQL Script:
DROP TABLE IF EXISTS stg_customer;
CREATE TABLE `stg_customer` (
`customer_id` varchar(50) DEFAULT NULL,
`first_name` varchar(50) DEFAULT NULL,
`last_name` varchar(50) DEFAULT NULL,
`date_of_birth` datetime DEFAULT NULL,
`city` varchar(50) DEFAULT NULL,
`state` varchar(50) DEFAULT NULL,
`date_updated` datetime DEFAULT NULL)
ENGINE=InnoDB DEFAULT CHARSET=utf8;
Line 1 Column 0
Help
OK
Cancel
Network Connection Details
Network Connection Details:
Property
Value
Connection-specific DNS S...
iit.edu
Description
Intel(R) Dual Band Wireless-AC 7265
Physical Address
F8-34-41-70-B1-1A
DHCP Enabled
Yes
IPv4 Address
104.194.125.78
IPv4 Subnet Mask
255.255.240.0
Lease Obtained
Tuesday, February 27, 2018 11:19:06 AM
Lease Expires
Tuesday, February 27, 2018 5:17:28 PM
IPv4 Default Gateway
104.194.112.2
IPv4 DHCP Server
216.47.143.135
IPv4 DNS Servers
216.47.143.106
216.47.143.105
216.47.143.90
IPv4 WINS Server
NetBIOS over Tcpip Enabl...
Yes
Link-local IPv6 Address
fe80::4486:30f3:8b77:fb7d%15
IPv6 Default Gateway
IPv6 DNS Server
Close

4. Now we will take transformation where excel input '**Customer_source' transformed** to Table output.

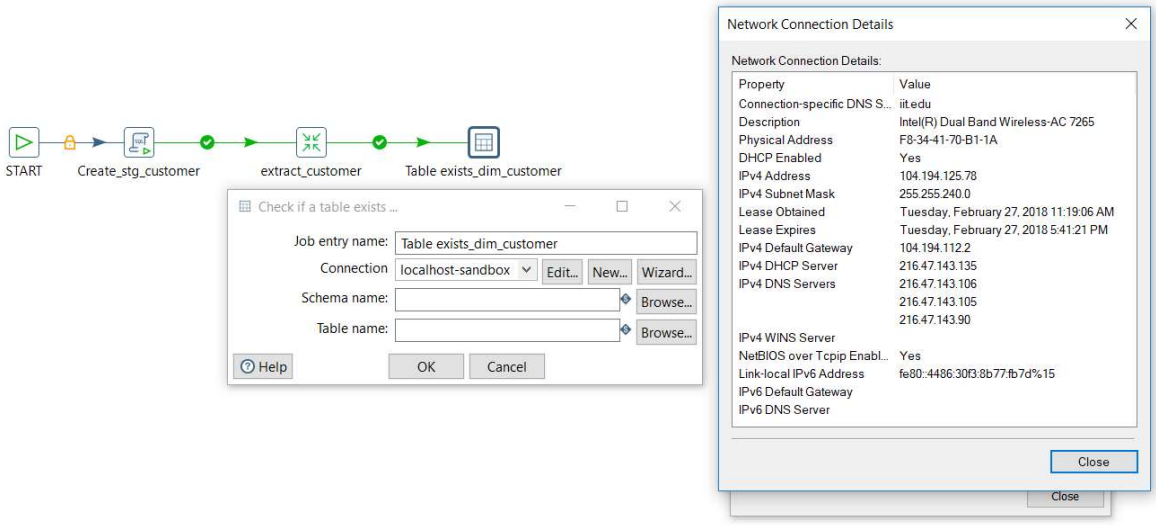


5. We need to **link** this **transformation** to our **main job** which can be done by using **dynamic** path to our job.

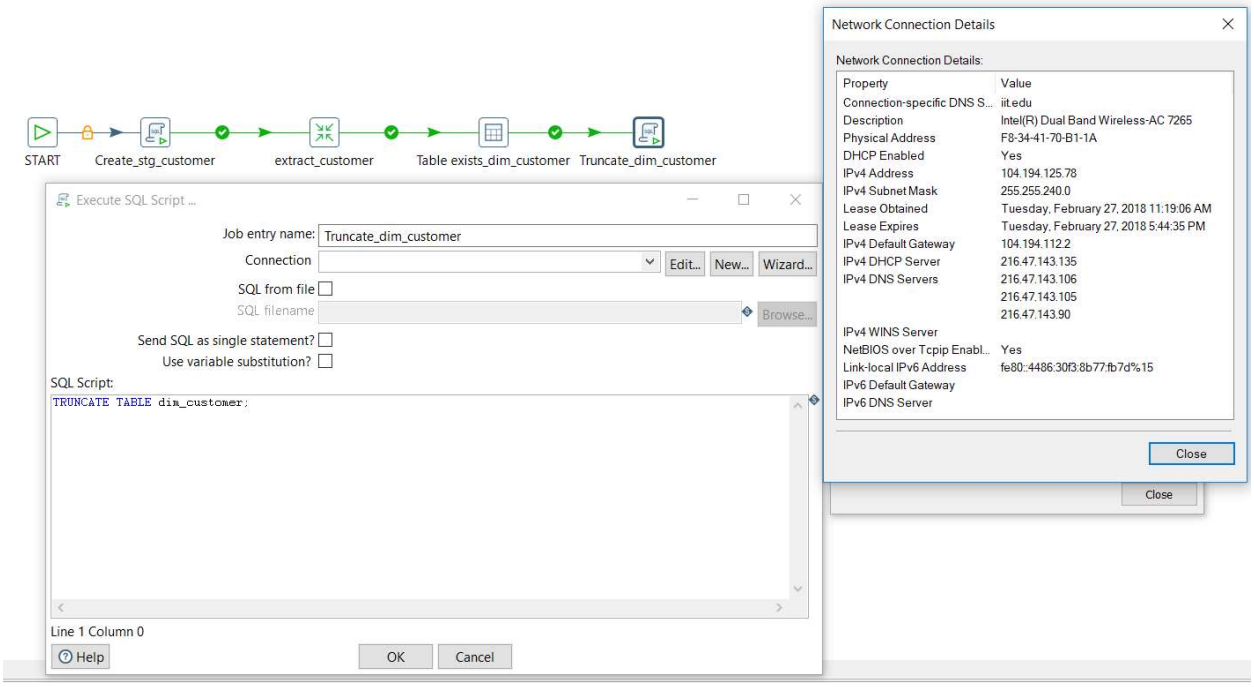

START
Create_stg_customer
extract_custom
Network Connection Details
Network Connection Details:
Property
Value
Connection-specific DNS S...
iit.edu
Description
Intel(R) Dual Band Wireless-AC 7265
Physical Address
F8-34-41-70-B1-1A
DHCP Enabled
Yes
IPv4 Address
104.194.125.78
IPv4 Subnet Mask
255.255.240.0
Lease Obtained
Tuesday, February 27, 2018 11:19:06 AM
Lease Expires
Tuesday, February 27, 2018 5:30:23 PM
IPv4 Default Gateway
104.194.112.2
IPv4 DHCP Server
216.47.143.135
IPv4 DNS Servers
216.47.143.106
216.47.143.105
216.47.143.90
IPv4 WINS Server
NetBIOS over Tcpip Enabl...
Yes
Link-local IPv6 Address
fe80::4486:30f3:8b77:fb7d%15
IPv6 Default Gateway
IPv6 DNS Server
Close
Transformation
Entry Name:
extract_customer
Transformation:
${Internal.Entry.Current.Directory}/extract_transfor
Browse...
Options
Logging
Arguments
Parameters
Run configuration:
Pentaho local
Execution
Execute every input row
Clear results rows before execution
Clear results files before execution
Wait for remote transformation to complete
Follow local abort to remote transformation
Help
OK
Cancel

2. Read the staging table to build a dimension called dim_customer

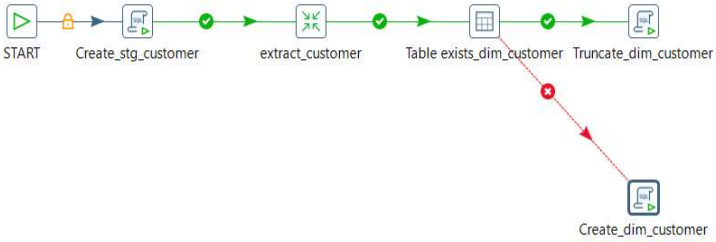
1. Now, we will take table exists" **Table exists_dim_customer**" to our job just to check if table "**dim_customer**" already exist, if table **exist** then **truncate** the table "dim_customer" entries else **create** a **new** table "dim_customer"

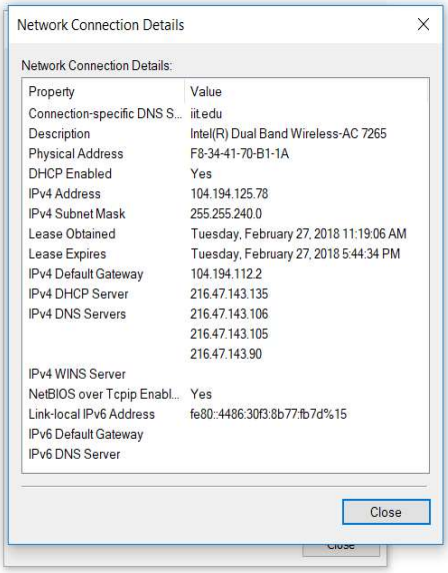


2. If table "dim_customer" exists then truncate table dim_customer by taking one sql scripting in pentaho tool.

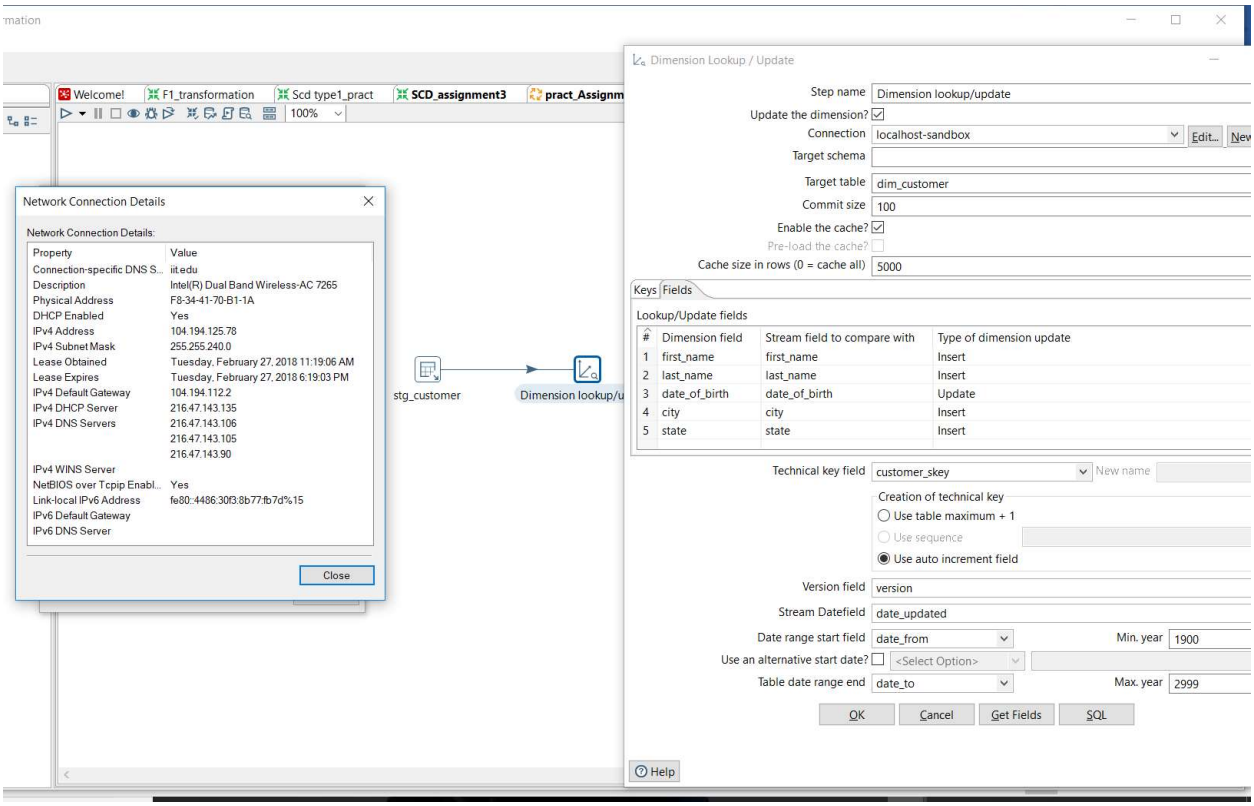


3. If table does not exists, then **create new** table dim_customer by taking another SQL scripting in the Pentaho.



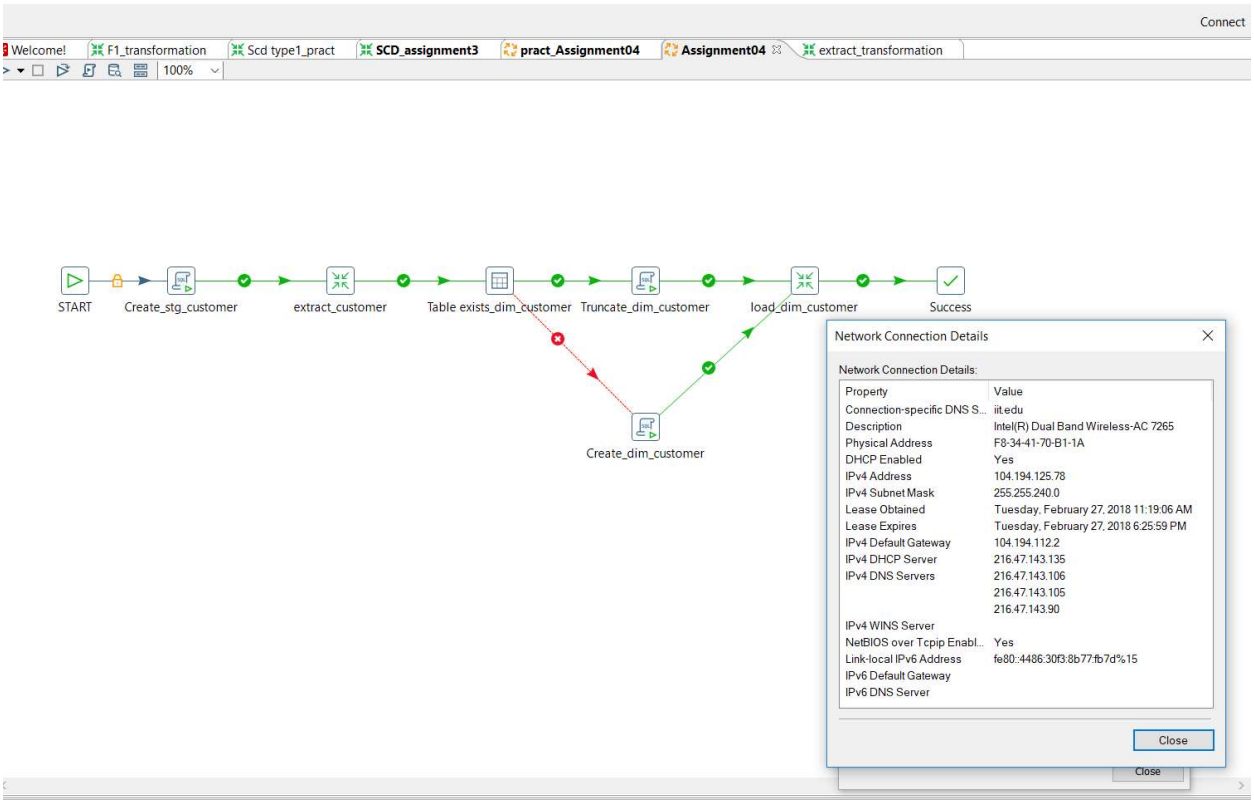


4. Furthermore, we need to load a transformation using table as input ”stg_customer” and Dimension lookup/update as output connecting to same database schema ‘**localhost-sandbox’** Having target table as “**dim_customer**”.

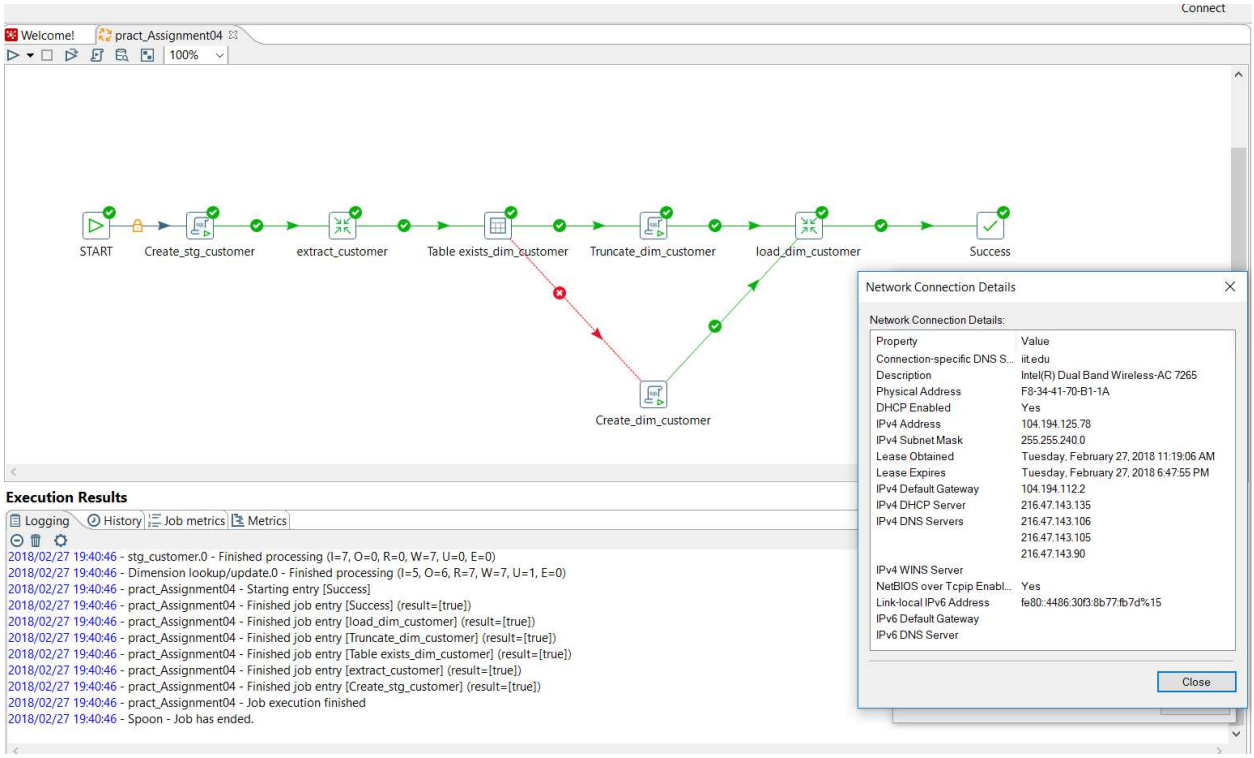


Taking type of dimension as **Update** (**SCD type1**) for date_of_birth column, while for other column as **Insert (SCD type2).**

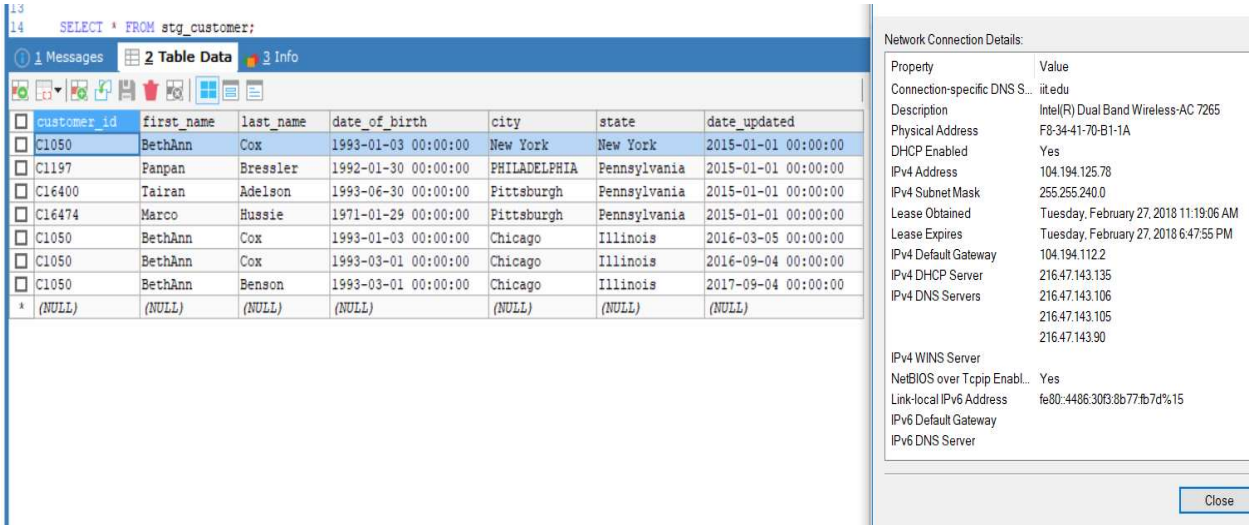
5. Every Job ends with **Success**, drag and drop to our main output sheet.



6. Running the job **successfully without** any **errors**.



7. Checking the **entry** from the **database**, where we found that records are inserted to **stg_customer** table.



```
SELECT * FROM stg_customer;
```

1 Messages | 2 Table Data | 3 Info

customer_id	first_name	last_name	date_of_birth	city	state	date_updated
C1050	BethAnn	Cox	1993-01-03 00:00:00	New York	New York	2015-01-01 00:00:00
C1197	Panpan	Bressler	1992-01-30 00:00:00	PHILADELPHIA	Pennsylvania	2015-01-01 00:00:00
C16400	Tairan	Adelson	1993-06-30 00:00:00	Pittsburgh	Pennsylvania	2015-01-01 00:00:00
C16474	Marco	Hussie	1971-01-29 00:00:00	Pittsburgh	Pennsylvania	2015-01-01 00:00:00
C1050	BethAnn	Cox	1993-01-03 00:00:00	Chicago	Illinois	2016-03-05 00:00:00
C1050	BethAnn	Cox	1993-03-01 00:00:00	Chicago	Illinois	2016-09-04 00:00:00
C1050	BethAnn	Benson	1993-03-01 00:00:00	Chicago	Illinois	2017-09-04 00:00:00
(NULL)	*(NULL)*	*(NULL)*	*(NULL)*	*(NULL)*	*(NULL)*	*(NULL)*

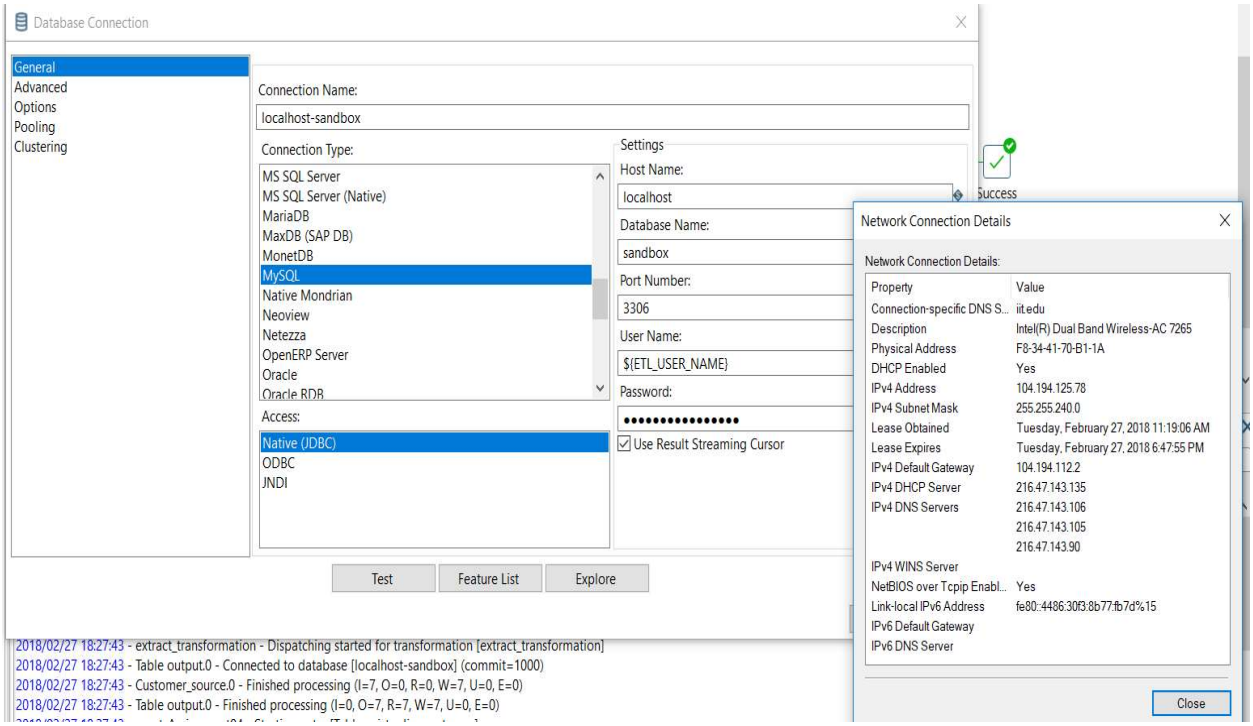
Network Connection Details:

Property	Value
Connection-specific DNS S...	iit.edu
Description	Intel(R) Dual Band Wireless-AC 7265
Physical Address	F8-34-41-70-B1-1A
DHCP Enabled	Yes
IPv4 Address	104.194.125.78
IPv4 Subnet Mask	255.255.240.0
Lease Obtained	Tuesday, February 27, 2018 11:19:06 AM
Lease Expires	Tuesday, February 27, 2018 6:47:55 PM
IPv4 Default Gateway	104.194.112.2
IPv4 DHCP Server	216.47.143.135
IPv4 DNS Servers	216.47.143.106
	216.47.143.105
	216.47.143.90
IPv4 WINS Server	
NetBIOS over Tcpip Enabl...	Yes
Link-local IPv6 Address	fe80::4486:30f3:8b77:fb7d%15
IPv6 Default Gateway	
IPv6 DNS Server	

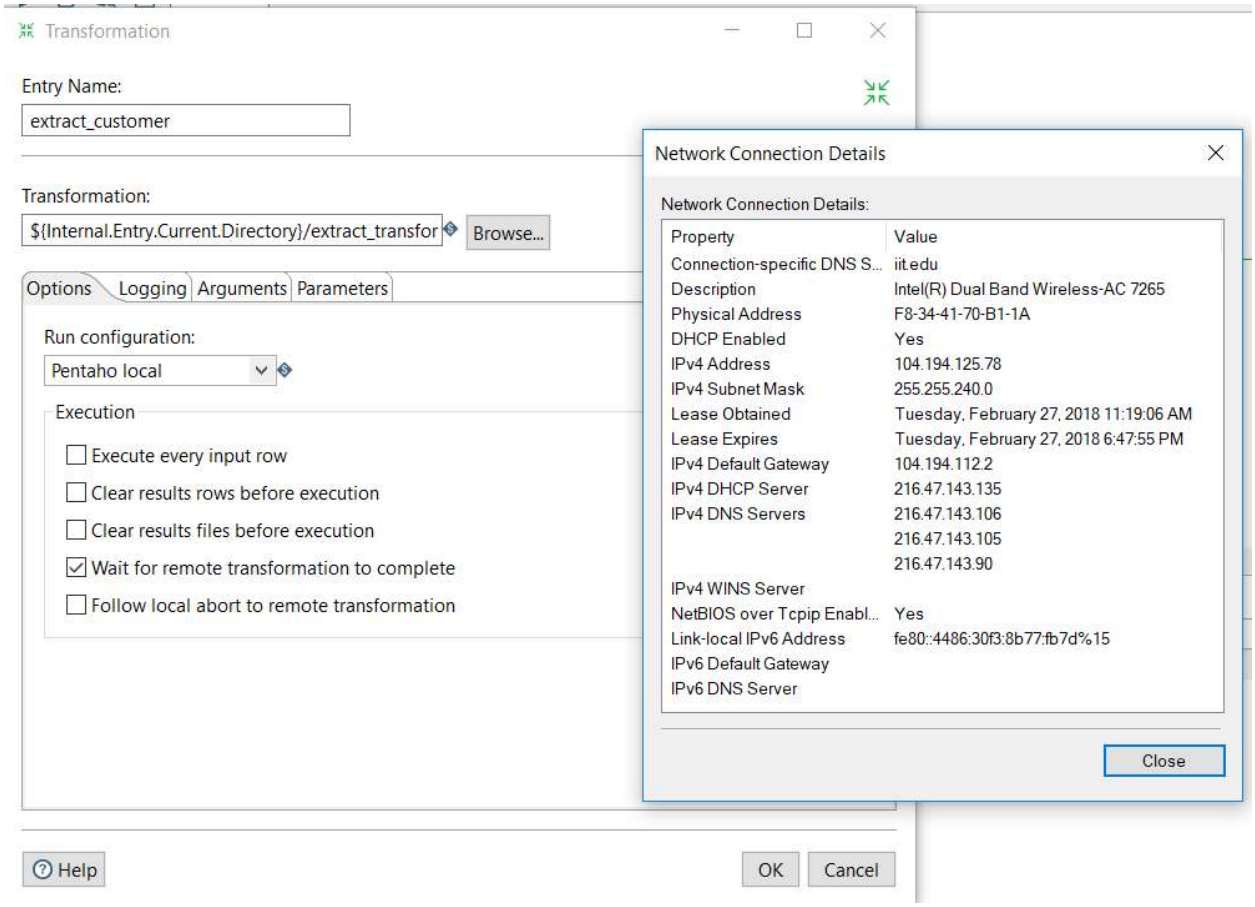
Close

3. Jobs has been created which by following guidelines:

1. Job has been created by using environment variables as **ETL_USER_NAME** and **ETL_USER_PASS** for database connections.
2. Database is connected to **sandbox** database having server as **localhost**.



3. Transformations are linked with **Internal.Entry.Current.Directory** to jobs, **instead** of using **absolute** path.



4. Zip file includes main Kettle job "**pract_Assignment04.kjb**" and two transformation as **extract_transformation.ktr** and **Dimension_Customer_transformation.ktr**