Step 1 - Preparation

Download the zipped file, **w4-exercise.zip**, from the Blackboard

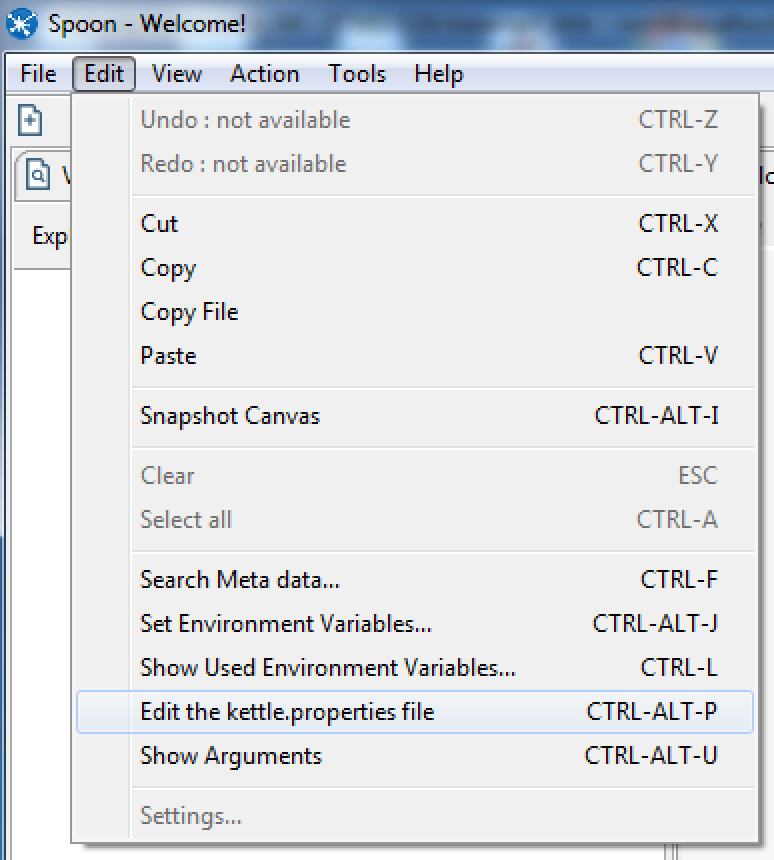
(**Content** 🡪 **Week 04 Dimensional Modeling…** 🡪 **Assignments**).

Extract the files into a folder.

Run the DDL scripts in the **00-w4-ddl-scripts.sql** via your choice of a MySQL client (SQLYog, MySQL Workbench, Sequel Pro, etc). Make sure that there are 5 new tables created as below.

1. datamart\_kbb.dim\_customer
2. datamart\_kbb.dim\_customer\_scd1
3. datamart\_kbb.dim\_customer\_scd2
4. source\_db.city\_state
5. source\_db.customer
6. target\_db.stg\_customer

Launch the PDI and go to **Edit** 🡪 **Edit the kettle.properties file**.

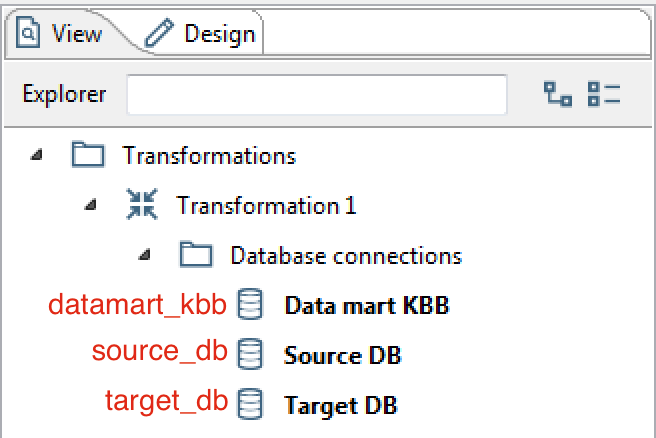


Add two variables **ETL\_USER\_NAME** and **ETL\_USER\_PASS** as the screenshot below (You can add a new variable by right-clicking on any grid and then selecting either the **Insert before this row** or **Insert after this row**). Please make sure that you enter your own database account information in the **Value** column.

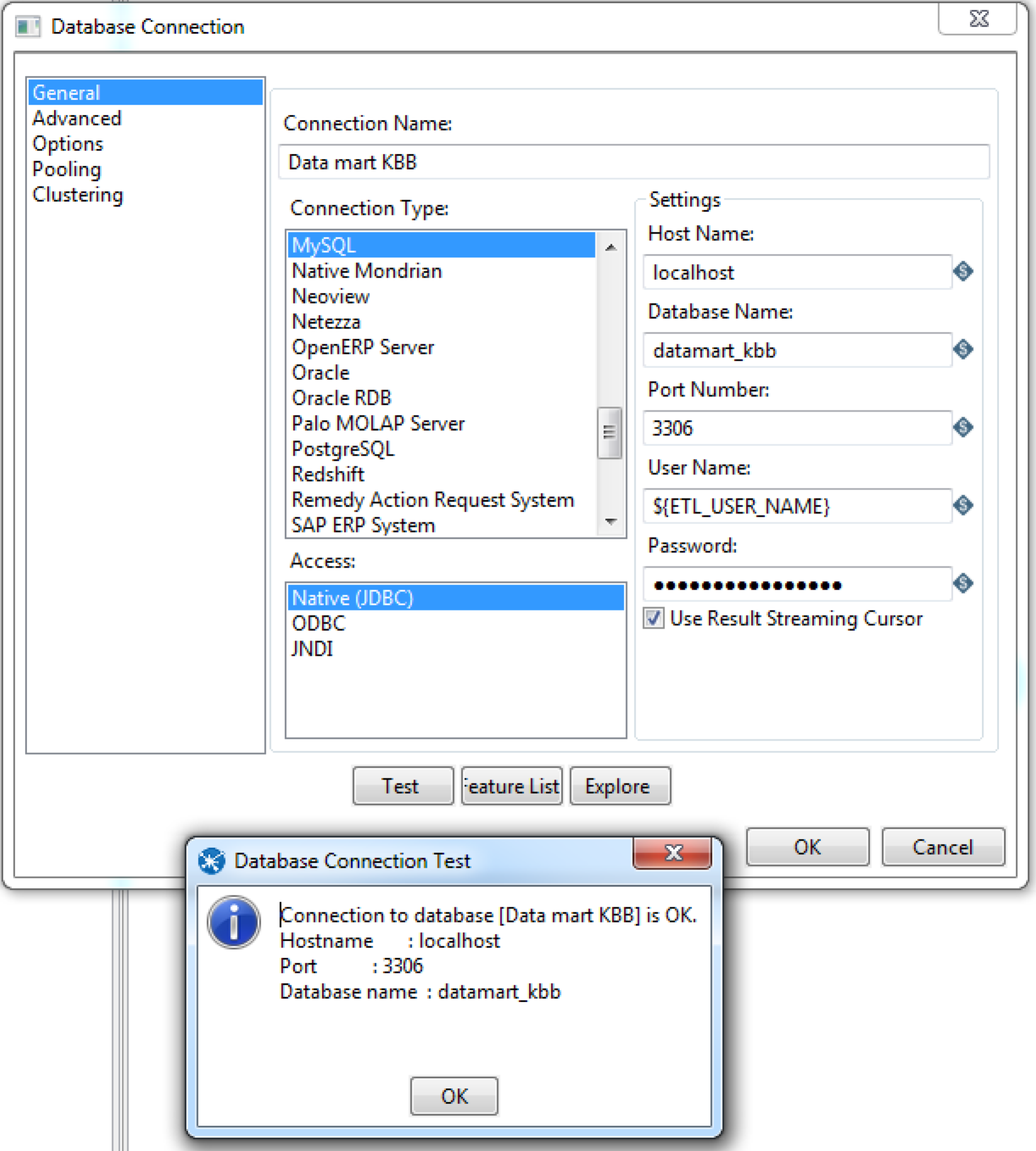


Click **OK** at the bottom.

Now create a new transformation and go to the **Database connections** on the left where you can see your already shared database connections while working on the Week 01 through Week 03 assignments. There you should see three connections as below:



Modify all the connections (right click or double click on each connection) to use the variables you defined in the previous section (**ETL\_USER\_NAME** & **ETL\_USER\_PASS**). You can do that by entering CTRL+SPACE in the **User Name** and the **Password** and selecting the variables. Once you did, please make sure that the connection works (via the **Test** button).

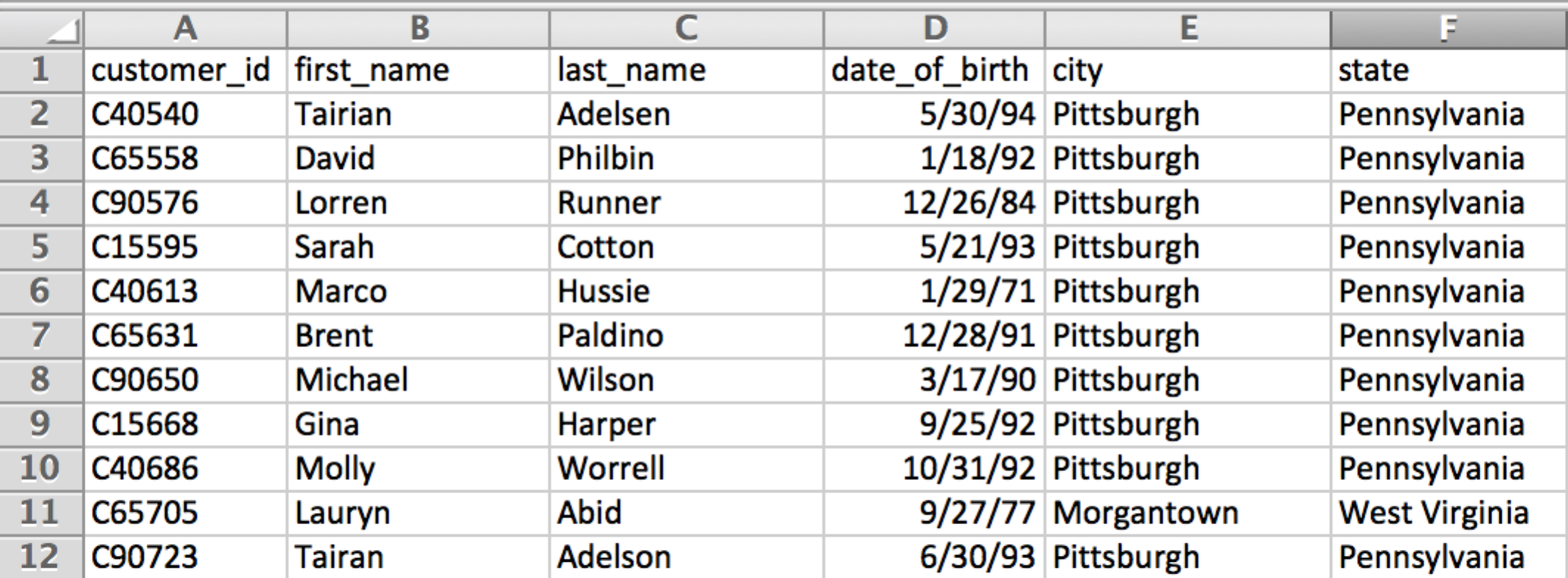


Step 2 - Loading DIM\_CUSTOMER

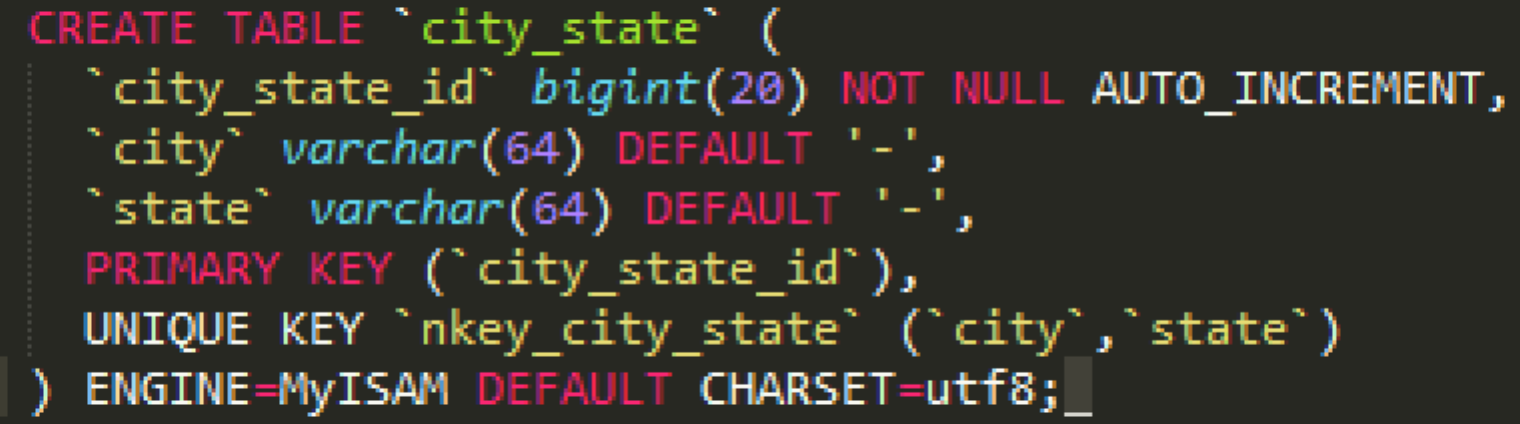
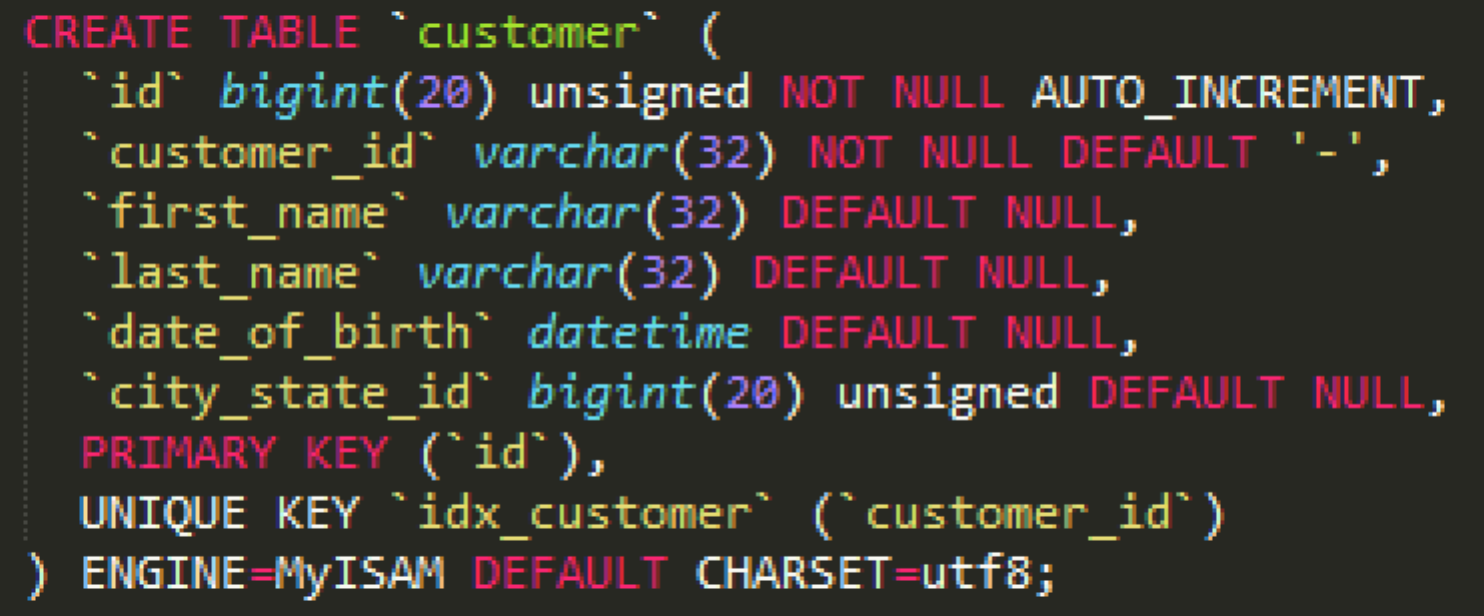
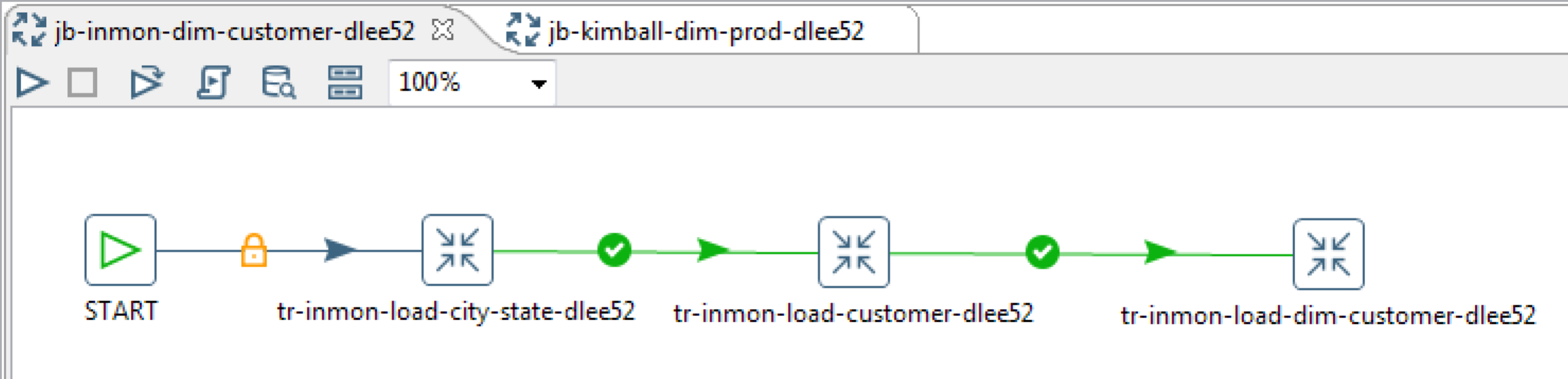
(10 points)

This section is to load the customer information in the source excel file, **customer\_raw.xls**, into the **dim\_customer** table via both the Inmon’s and the Kimball’s approaches. This should be very similar to what we did for the Week 02/03 class exercises. If you have not participated in the week 02/03 classes, please check out the videos (the second part in particular) and the exercise files in BB.

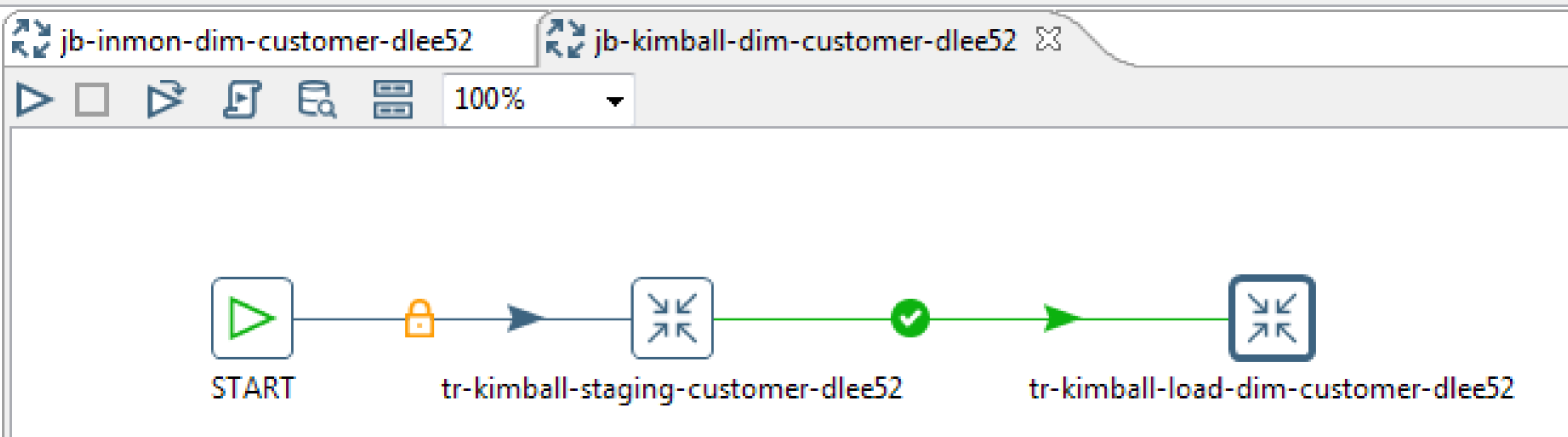
The contents of the source excel file are as below:



## For the Inmon’s approach, you will create…

1. **tr-inmon-load-city-state-yourID.ktr**: This transformation is to populate the **source\_db.city\_state** table (normalized). Please note that the **city** and the **state** columns are the composite natural key as you can see from the DDL of the table below:  
   
2. **tr-inmon-load-customer-yourID.ktr**: This transformation is to populate the **source\_db.customer** table (normalized). The **customer\_id** is the natural key. Here is the DDL of the table:  
   
3. **tr-inmon-load-dim-customer-yourID.ktr**: This transformation is to populate the **datamart\_kbb.dim\_customer**. The two normalized tables (source\_db.city\_state & source\_db.customer) in the previous steps are the sources in this transformation.
4. **jb-inmon-dim-customer-yourID.kjb**: This is a job to include all those three transformations you created above. For your reference, this is how my job looks like:  
   

## For the Kimball’s approach, you will create…

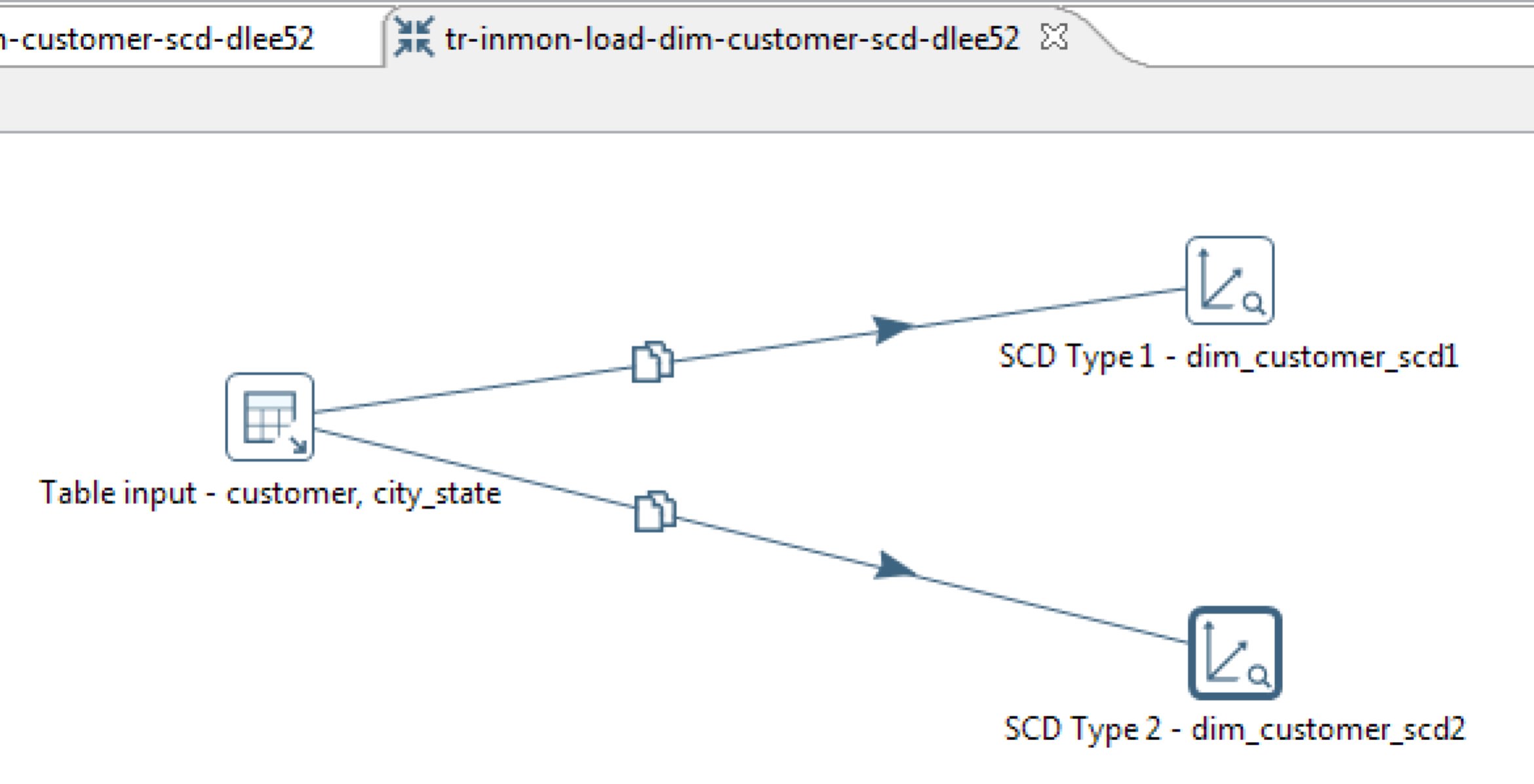
1. **tr-kimball-staging-customer-yourID.ktr**: This transformation is to stage the source excel file into the **target\_db.stg\_customer**.
2. **tr-kimball-load-dim-customer-yourID.ktr**: This transformation is to populate the **datamart\_kbb.dim\_customer**. The **target\_db.stg\_customer** is the source input in this transformation.
3. **jb-kimball-dim-customer-yourID.kjb**: This is a job to include those two transformations above. For your reference, this is how my job looks like:  
   

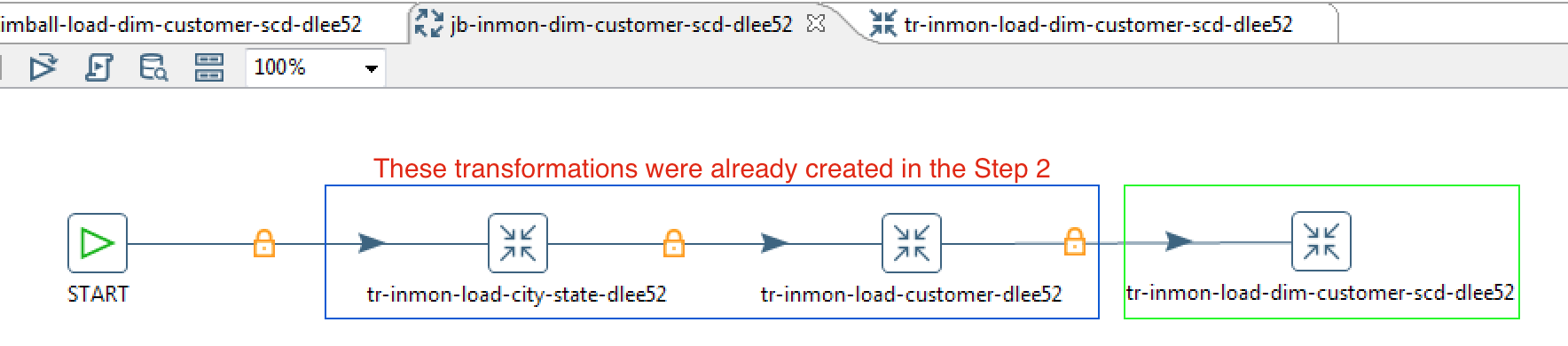
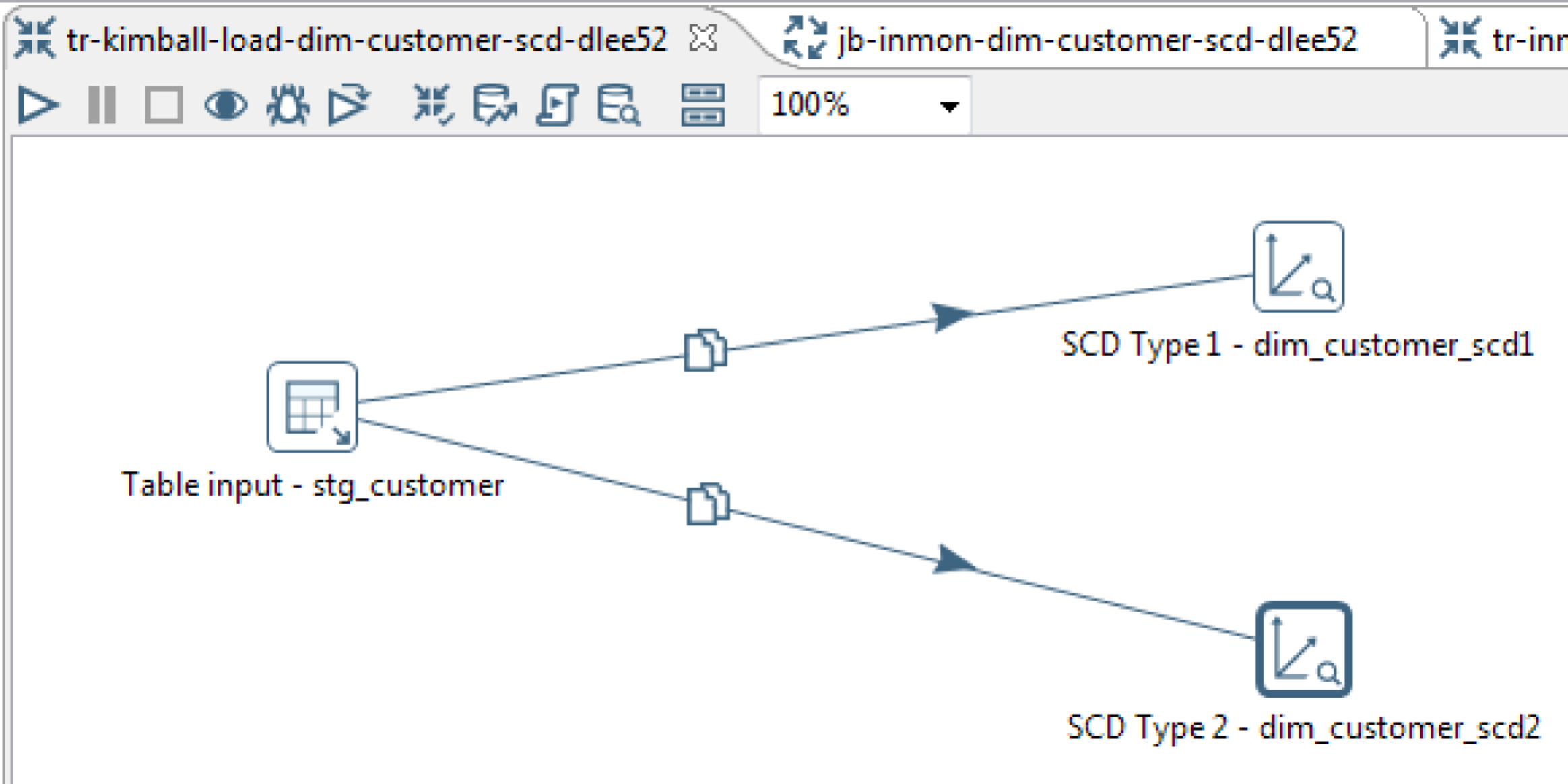
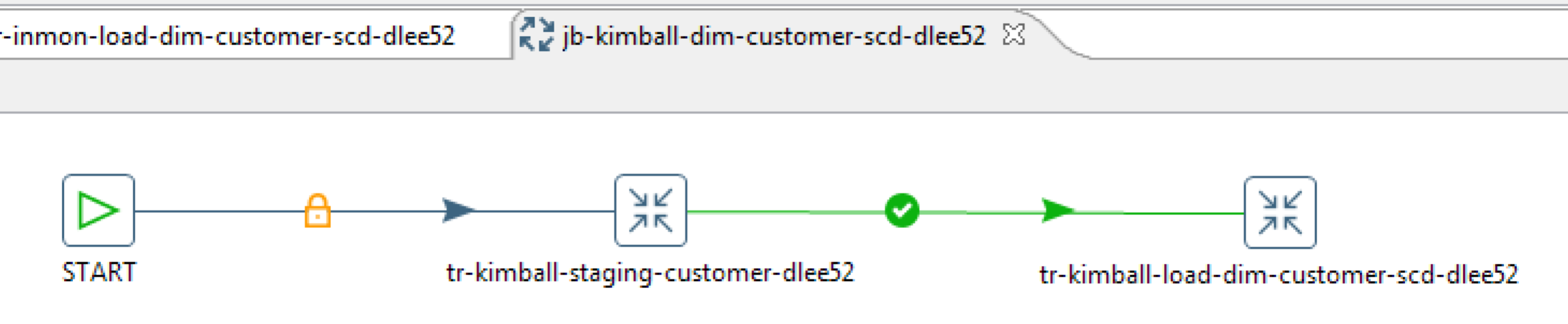
Step 3 - Loading DIM\_CUSTOMER\_SCD1 and DIM\_CUSTOMER\_SCD2 (10 points)

In this section, you are to load the customer information into the SCD Type 1 and 2 dimensions, which is very similar to what we did last week (week 03). If you have not participated in the week 04 class, please check out the class exercise files or the video (the second part in particular).

You will create…

1. **tr-inmon-load-dim-customer-scd-yourID.ktr**: This transformation uses the two normalized tables (**source\_db.city\_state** and **source\_db.customer**) as the input sources and loads into two different SCD type tables as below:  
   1. **datamart\_kbb.dim\_customer\_scd1**: treats all the attributes other than the surrogate and natural keys (**first\_name, last\_name, date\_of\_birth, city**, and **state**) as SCD Type 1 (**Punch Through** in PID’s terminology).
   2. **datamart\_kbb**.**dim\_customer\_scd2**: treats those **first\_name, last\_name, city**, and **state** attributes as SCD Type 2 (**Insert** in PDI’s terminology)

For your reference, this is how my transformation looks like:  


1. **jb-inmon-dim-customer-scd-dlee52.kjb**: This job includes the two transformations created in the Section 2 and the one created in the previous step. For your reference, this is how my job looks like:  
   
2. **tr-kimball-load-dim-customer-scd-yourID.ktr**: This transformation uses the staging table (**target\_db.stg\_customer**) as the input source and loads into the same SCD type tables (**datamart\_kbb.dim\_customer\_scd1** and **datamart\_kbb.dim\_customer\_scd2**) .  
   For your reference, this is how my transformation looks like:  
   
3. **jb-kimball-dim-customer-scd-dlee52.kjb**: This job includes the transformation created in the Section 2 and the one created in the previous step. For your reference, this is how my job looks like:  
   

Step 4 - Submit your Week 04 homework

*Submit those eight transformations, the two jobs, and your* ***screenshots in a single zip file****. As for the screenshots,* ***no snapshot is needed for the Step 2****. Just include the snapshots of the Step 3 results only. To be more specific, the snapshots must demonstrate how the changes in the* ***customer\_raw.xls*** *file got reflected in the* ***dim\_customer\_scd1*** *and the* ***dim\_customer\_scd2*** *tables.*

*The zipped file must follow the naming convention below:*

*yourID\_week04.zip*

*(e.g. dlee52\_week04.zip)*