In order to improve the query performance, you are being asked to set up a new primary DW that is not fully normalized. To partly de-normalize, merge the two tables that store the *state* and *industry* names with the table that stores the *employer* information. In other words, the new database will have two fewer tables. Note that the table containing the *positions* still stays separate and normalized. Further, there are no dates, so you don’t need to create a date table.

**a)** Create the primary DW in **SQL Server Management Studio**, along with appropriate tables inside it. Include in your writeup all CREATE statements that you used for this purpose. What primary and foreign keys did you use?

CREATE DATABASE RecruiterDW;

GO

USE RecruiterDW;

/\* Create tables \*/

CREATE TABLE DimEmployer

(EmployerID varchar(5) PRIMARY KEY NOT NULL,

EmployerName varchar(50) NOT NULL,

StreetAddress varchar(50),

City varchar(50),

StateProv varchar(2),

PostalCode varchar(7),

Country varchar(50) NOT NULL,

ContactFirstName varchar(50) NOT NULL,

ContactLastName varchar(50) NOT NULL,

Position varchar(50),

Phone varchar(10) NOT NULL,

Website bit,

NAICSCode varchar(6) NOT NULL,

Locations varchar(50) NOT NULL,

NAICSDesc varchar(50) NOT NULL);

GO

CREATE TABLE FactPosition

(PositionID int PRIMARY KEY NOT NULL,

PositionTitle varchar(50) NOT NULL,

EmployerID varchar(5) NOT NULL,

Wage Money,

HoursPerWeek int,

Experience bit,

Openings int,

CONSTRAINT FK\_Posi\_Emp FOREIGN KEY (EmployerID) REFERENCES DimEmployer (EmployerID));

GO

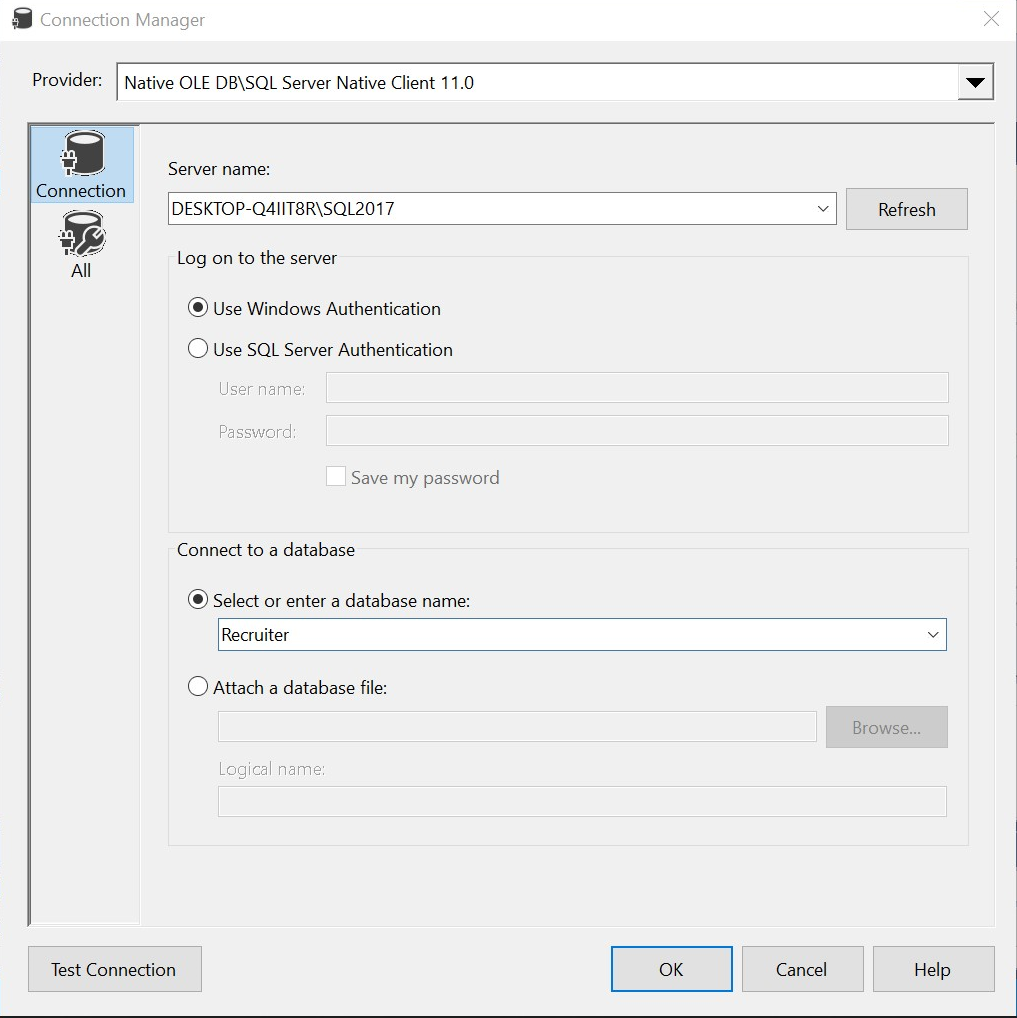


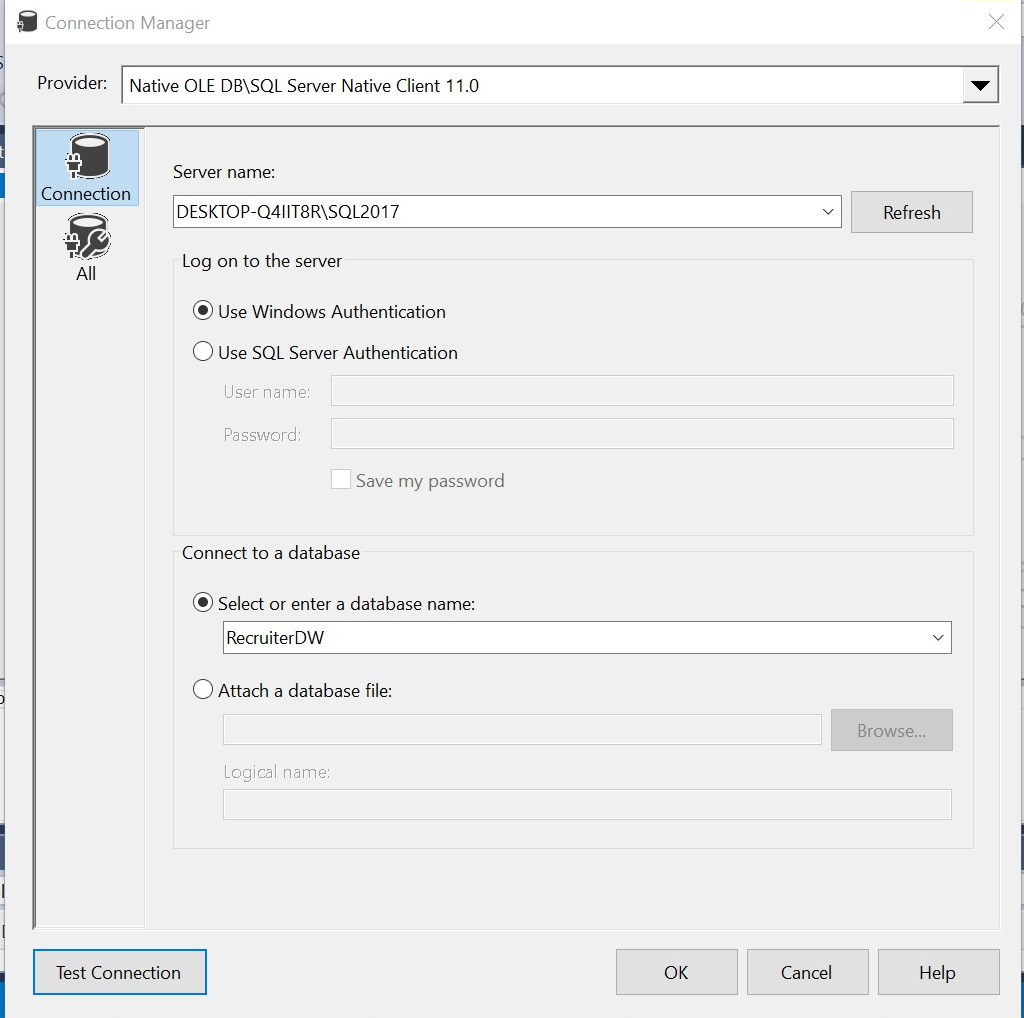
**b)** Does this new primary DW have a table that contains redundant data?

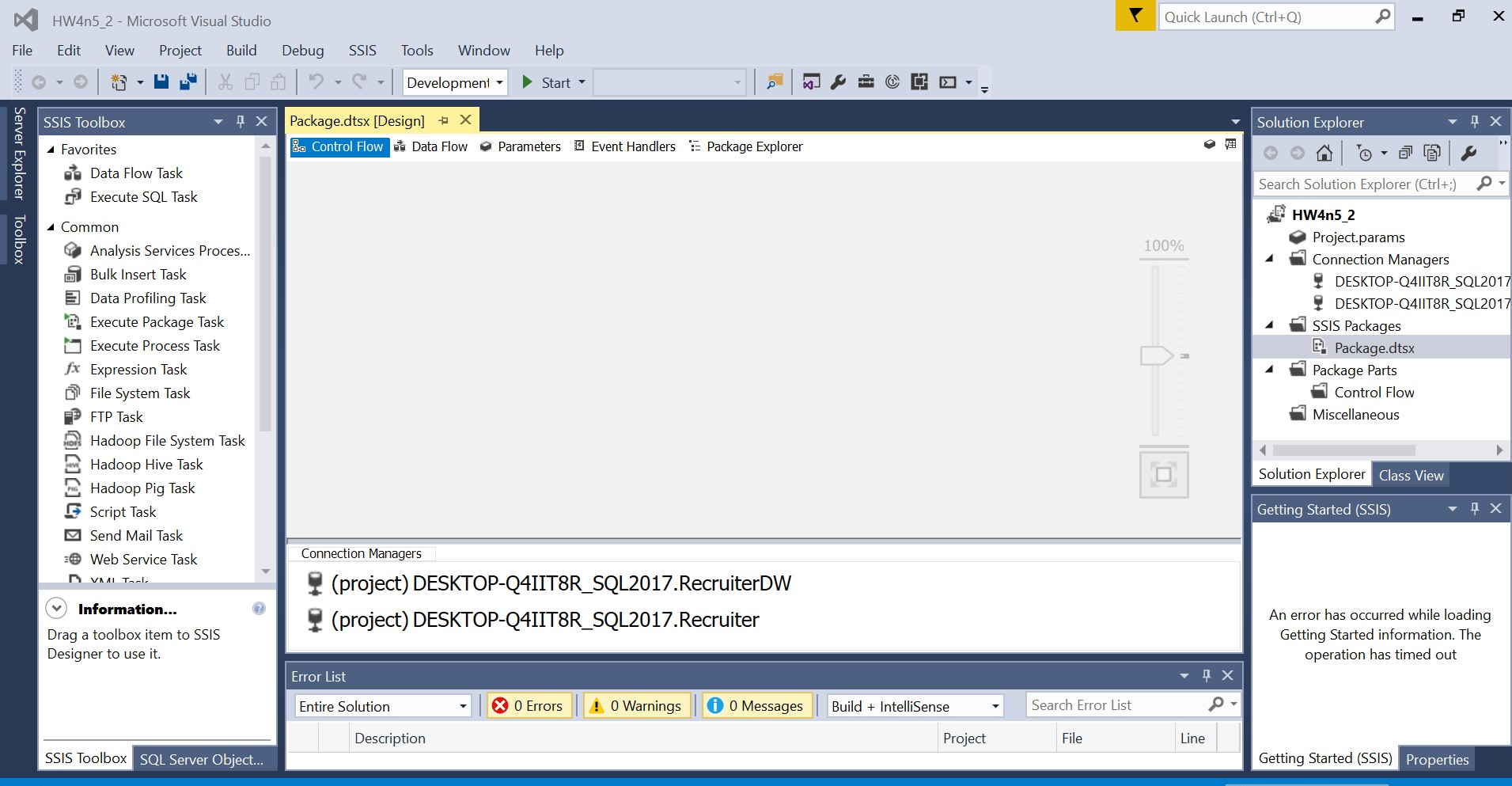
Yes. Employer table contains redundant data in “StateProv” & “Locations” and “NAICSCode” & “NAICSDesc”.

**c)** Use **SSIS** to move data to this new primary DW. Provide screenshots of:

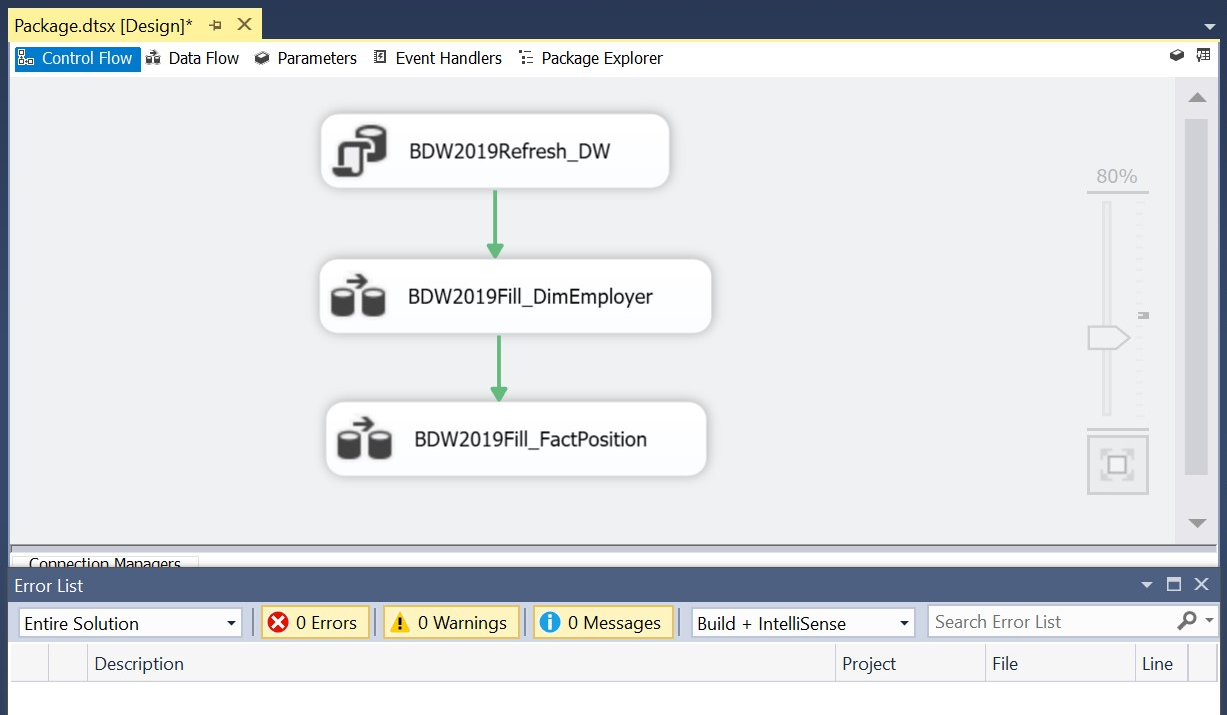
(i) all connections that you set up



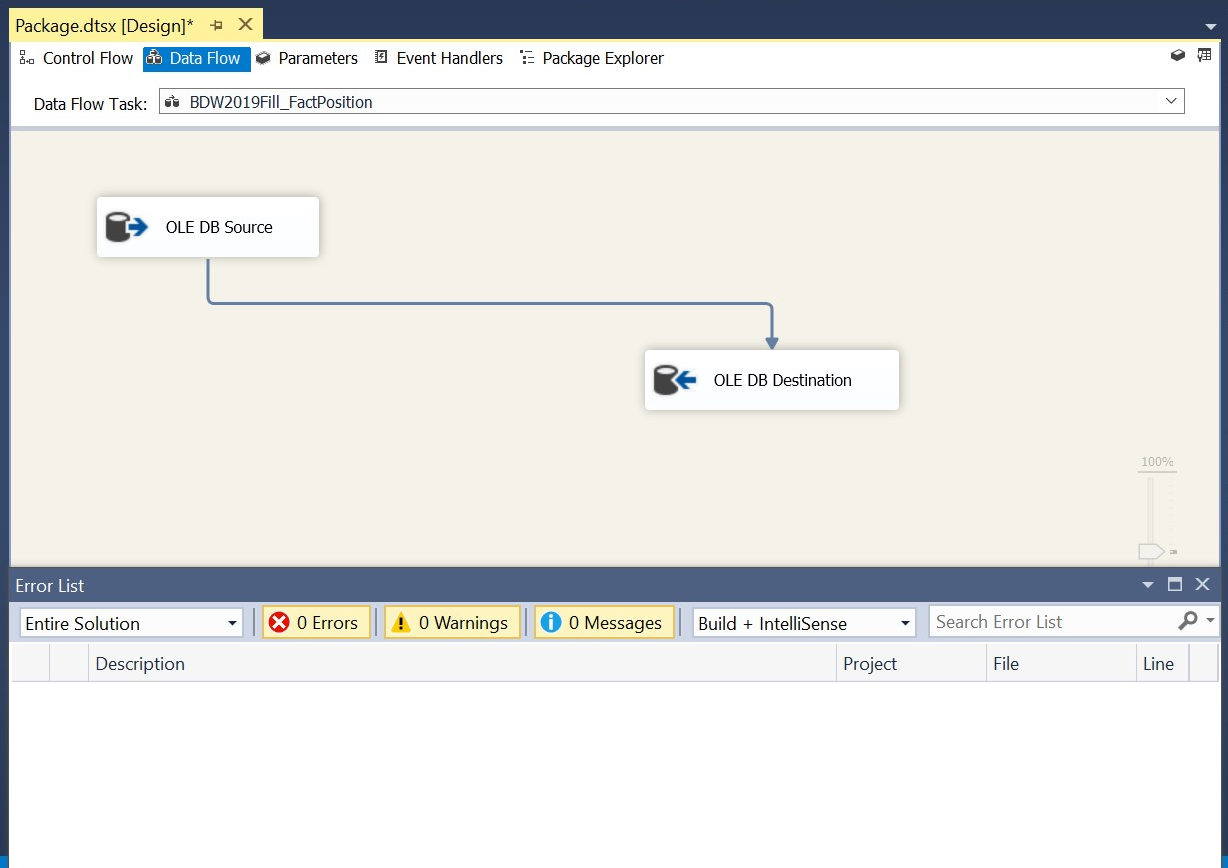


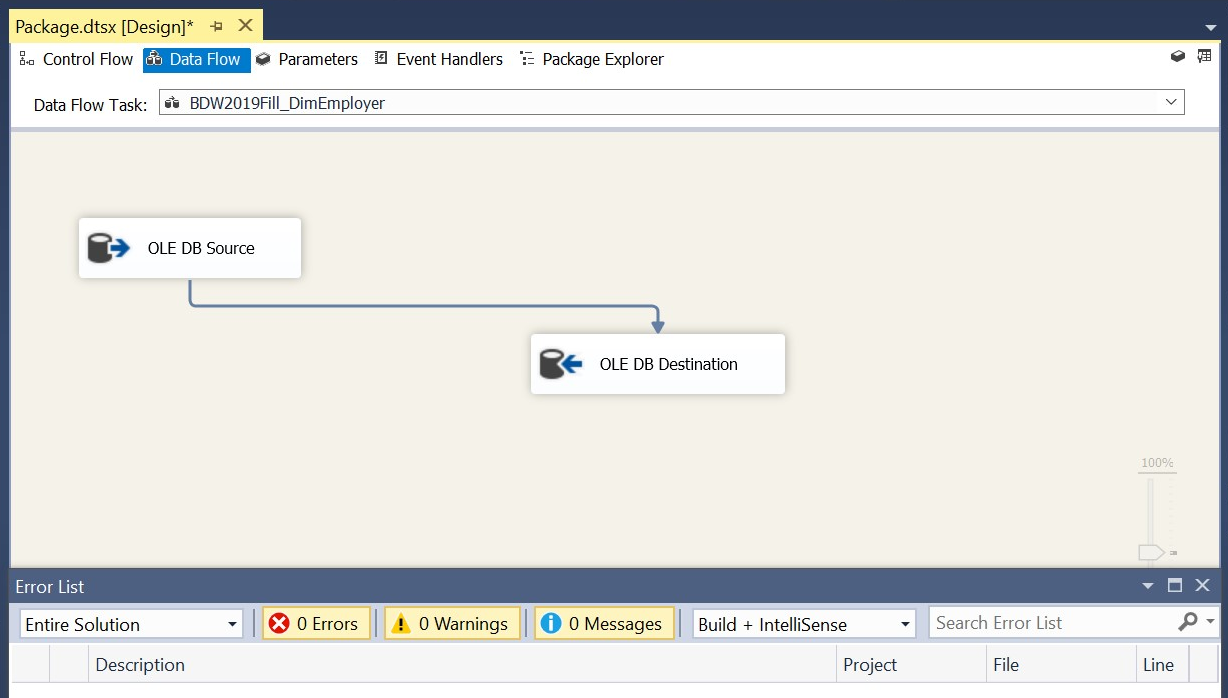


(ii) the control flow



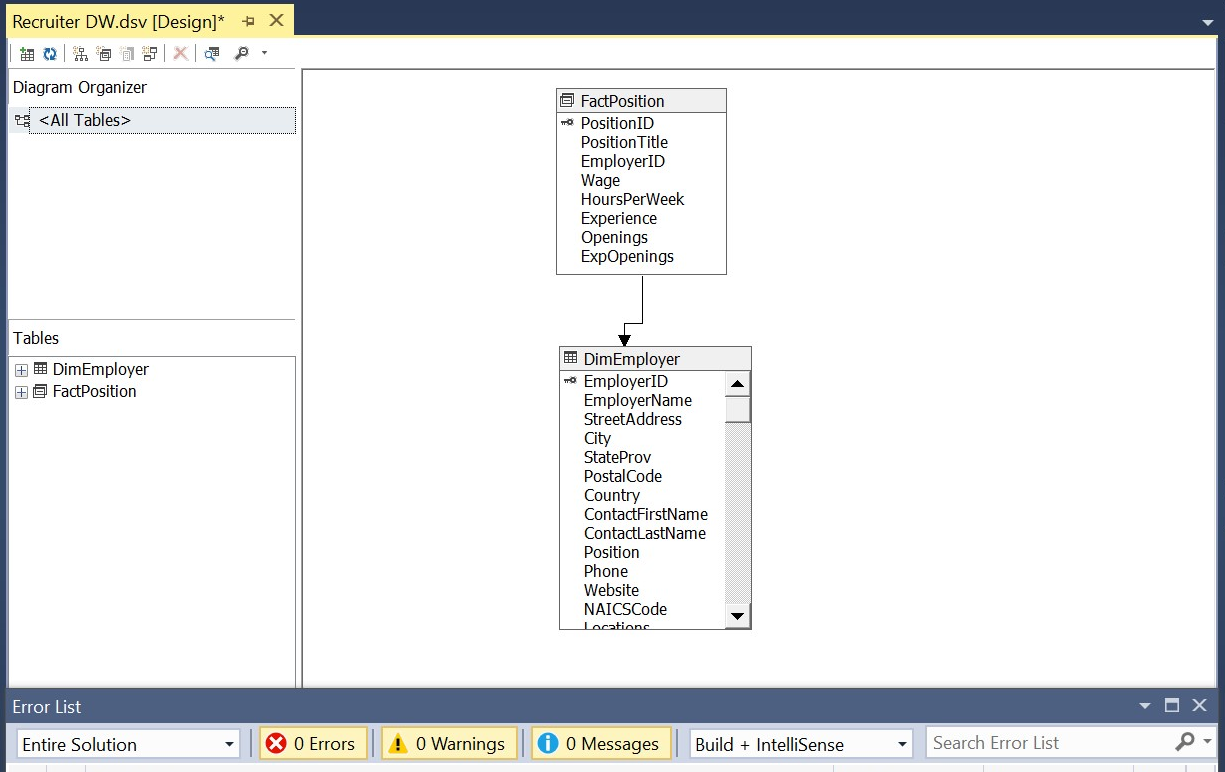
(iii) the source and destination for each data flow task





**d)** Create a cube, this time using your new primary DW as your data source. You do not have to include all your screenshots. Just submit a screenshot of the schema (data source view) after the following is taken care of:

• In the schema, replace the table containing positions with a query that adds **ExpOpenings**, calculated as **Experience\*Openings**, as an additional column.

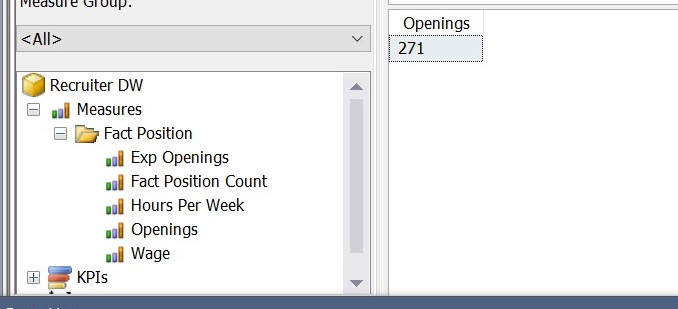


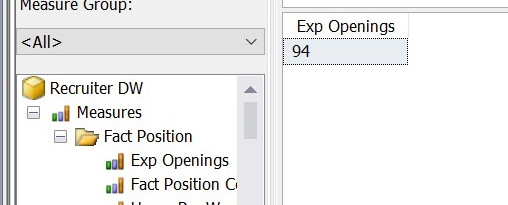
* What schema do you see, star or snowflake?

It is star schema. The fact table FactPosition is normalized, and the dimension table DimEmployer is denormalized.

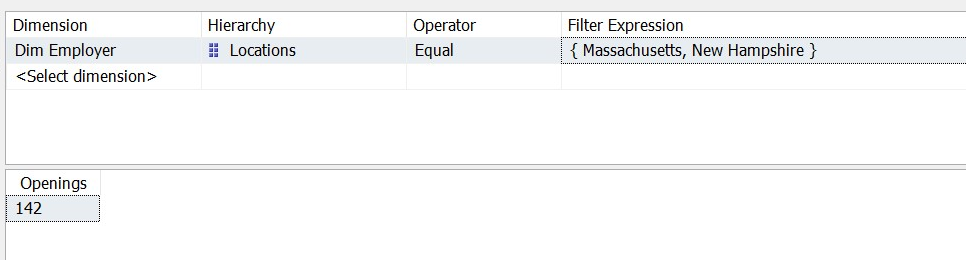
**e)** Time to deploy and browse the cube! In this OLAP session, I expect you to do the following:

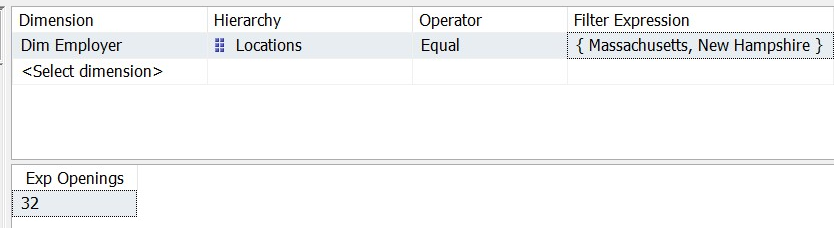
* First, display the overall total number of openings as well as the overall total number of openings requiring experience. Provide a screenshot of the cube browser window.



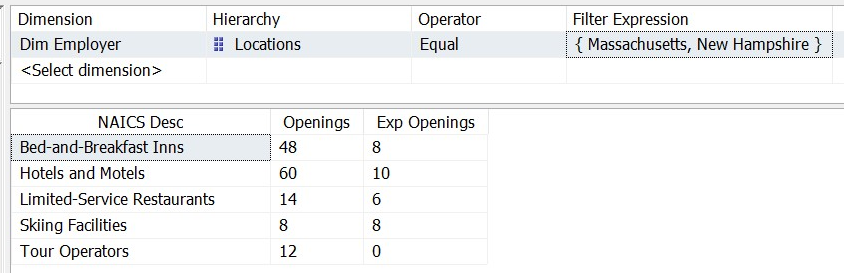


* Now, suppose that you are interested in only two states, namely, Massachusetts and New Hampshire. Report the same information as in part (i), but just for these two states (that is, the total for the two states **combined**, and not separate numbers for each state). Again, provide a screenshot of the cube browser window.





* Finally, suppose that you are interested in breaking down the totals in part (ii) by industry. Please show this break down. Note: I expect to see several rows, each with three columns, namely, the NAICS description, the number of openings, and the number of openings requiring experience. Again, provide a screenshot.



* Which OLAP operation did you perform in part (ii)? Which operation did you perform in part (iii)?

In part (ii), I perform filter operation to choose only Massachusetts and New Hampshire.

In part (iii), I perform drill down operation to see details in each industry.

* Transfer the cube to an Excel sheet, and repeat part (iii) in Excel. As usual, include appropriate screenshots.

