Assignment-2

526-Data Warehouse

Design of star Schema and Analysis

We have used **6 tables** in our start schema.

1. **Dim\_college**
2. **Dim\_ethnicity**
3. **Dim\_program**
4. **Dim\_application**
5. **Dim\_applicant**
6. **Dim date**

Here dim\_date is role playing dimension, dim\_ethnicity is a junk dimension and dim\_college is a bridge dimension.

Here are the queries that are implemented in Pentaho.

**Query for the table input Ethnicity**

SELECT ethnicity\_indian,ethnicity\_asian,ethnicity\_white,ethnicity\_black,ethnicity\_hawaiian,ethnicity\_hispanic FROM admissions\_info;

**Query for the table input Program**

SELECT DISTINCT program, department\_name

FROM admissions\_info

ORDER BY 1,2 ;

**Query for the table input dim\_applicant**

SELECT application\_id, applicant\_first\_name, applicant\_last\_name, country\_of\_citizenship, gender,intended\_entry\_term FROM admissions\_info

**Query for the table input dim\_applicantion**

SELECT application\_id, gre\_quantitative\_perc, gre\_analytical\_perc,gre\_verbal\_perc,STR\_TO\_DATE(submitted\_date,'%m/%d/%Y') AS submitted\_date, STR\_TO\_DATE(last\_updated\_date,'%m/%d/%Y') AS last\_updated\_date ,

CASE prog\_actn WHEN 'DENY' THEN 'Admission Denied'

WHEN 'APPL' THEN 'Applicant Applied'

WHEN 'WAPP' THEN 'Applicant Withrew'

WHEN 'MATR' THEN 'Applicant Matriculated'

END AS admission\_decision

FROM admissions\_info;

We have not implemented transformation for the bridge table:

**Bridge table:**

INSERT INTO dim\_college (application\_id,college\_name\_1,college\_name\_2,college\_name\_3,college\_country\_1,college\_country\_2,college\_country\_3,international\_1,international\_2,international\_3)

SELECT application\_id,college\_name\_1,college\_name\_2,college\_name\_3,college\_country\_1,college\_country\_2,college\_country\_3,international\_1,international\_2,international\_3 FROM admissions\_info

**Query for the table input fact\_application1**

**Fact table:**

SELECT application\_id,STR\_TO\_DATE(submitted\_date,'%m/%d/%Y') AS submitted\_date,STR\_TO\_DATE(last\_updated\_date,'%m/%d/%Y') AS last\_updated\_date,applicant\_last\_name,applicant\_first\_name,email\_address,gre\_quantitative\_perc,gre\_analytical\_perc,gre\_verbal\_perc,ethnicity\_indian,ethnicity\_asian,ethnicity\_white,ethnicity\_black,ethnicity\_hawaiian,ethnicity\_hispanic,

country\_of\_citizenship,gender,department\_name,intended\_entry\_term,program,prog\_actn,1 as applied,

CASE prog\_actn WHEN 'WAPP' THEN 1

WHEN 'MATR' THEN 1

ELSE 0 END AS admitted,

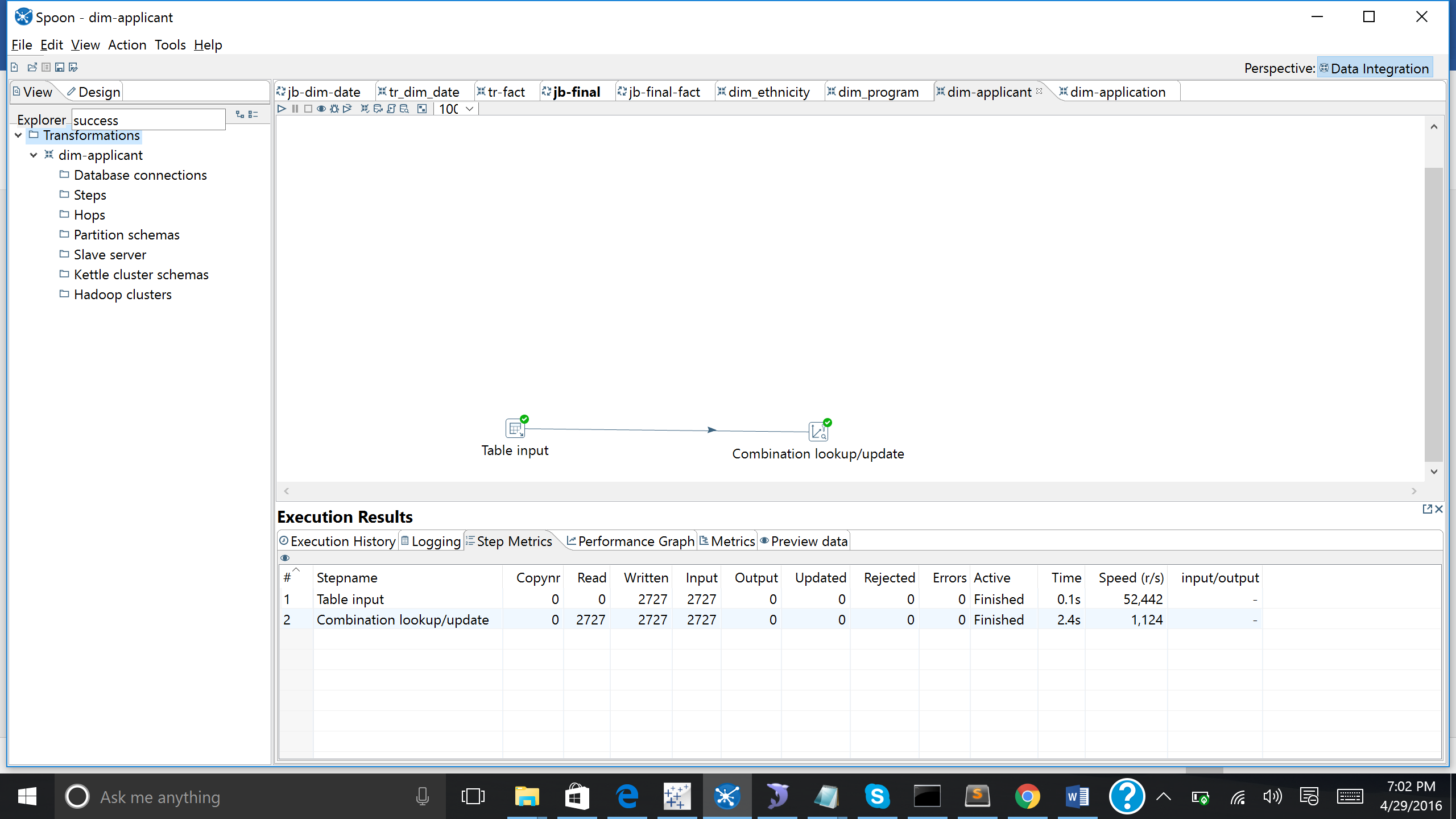
CASE prog\_actn WHEN 'MATR' THEN 1

ELSE 0 END AS accepted

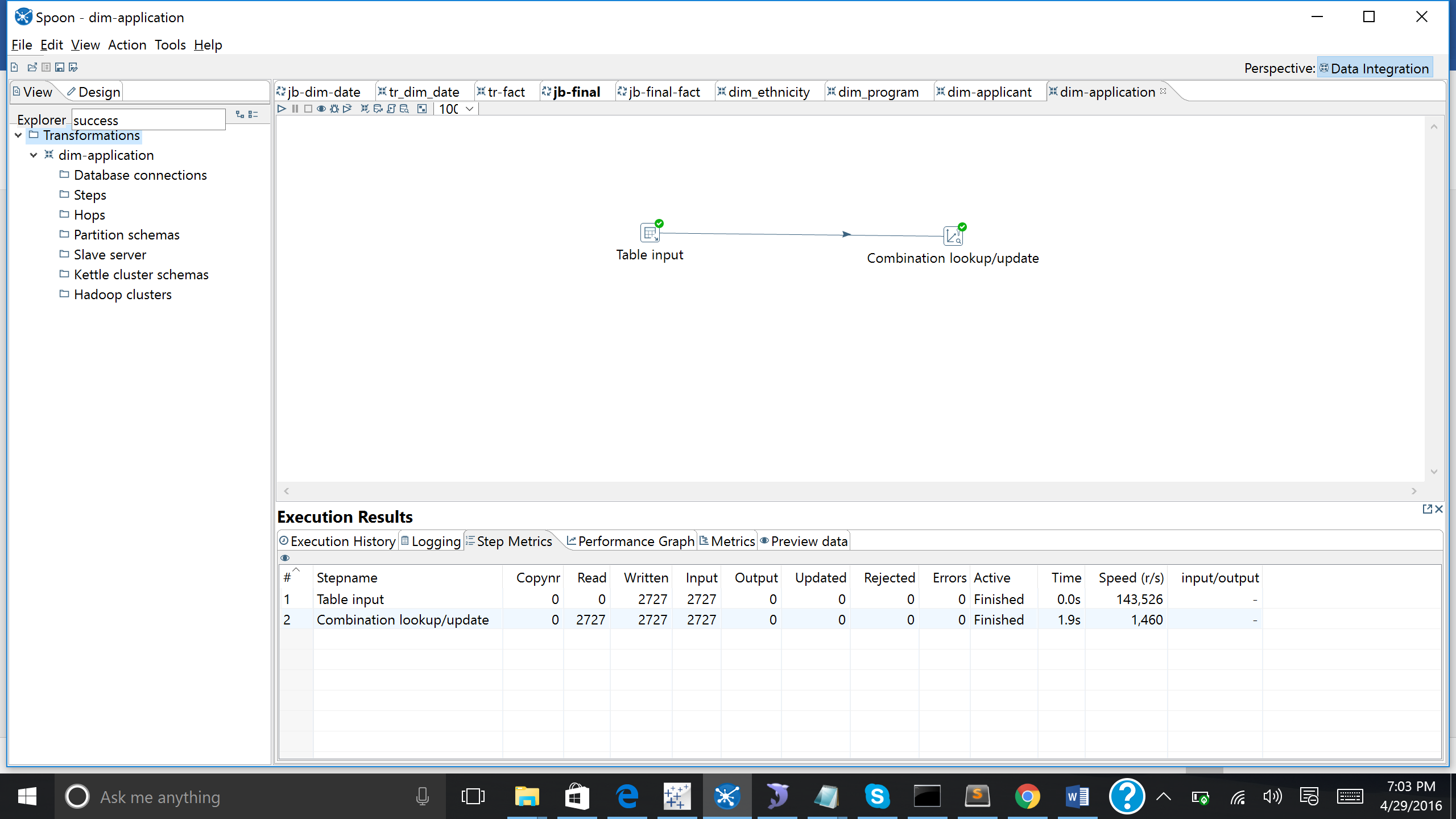
FROM admissions\_info;

**Screenshot of Execution of Transformation:**

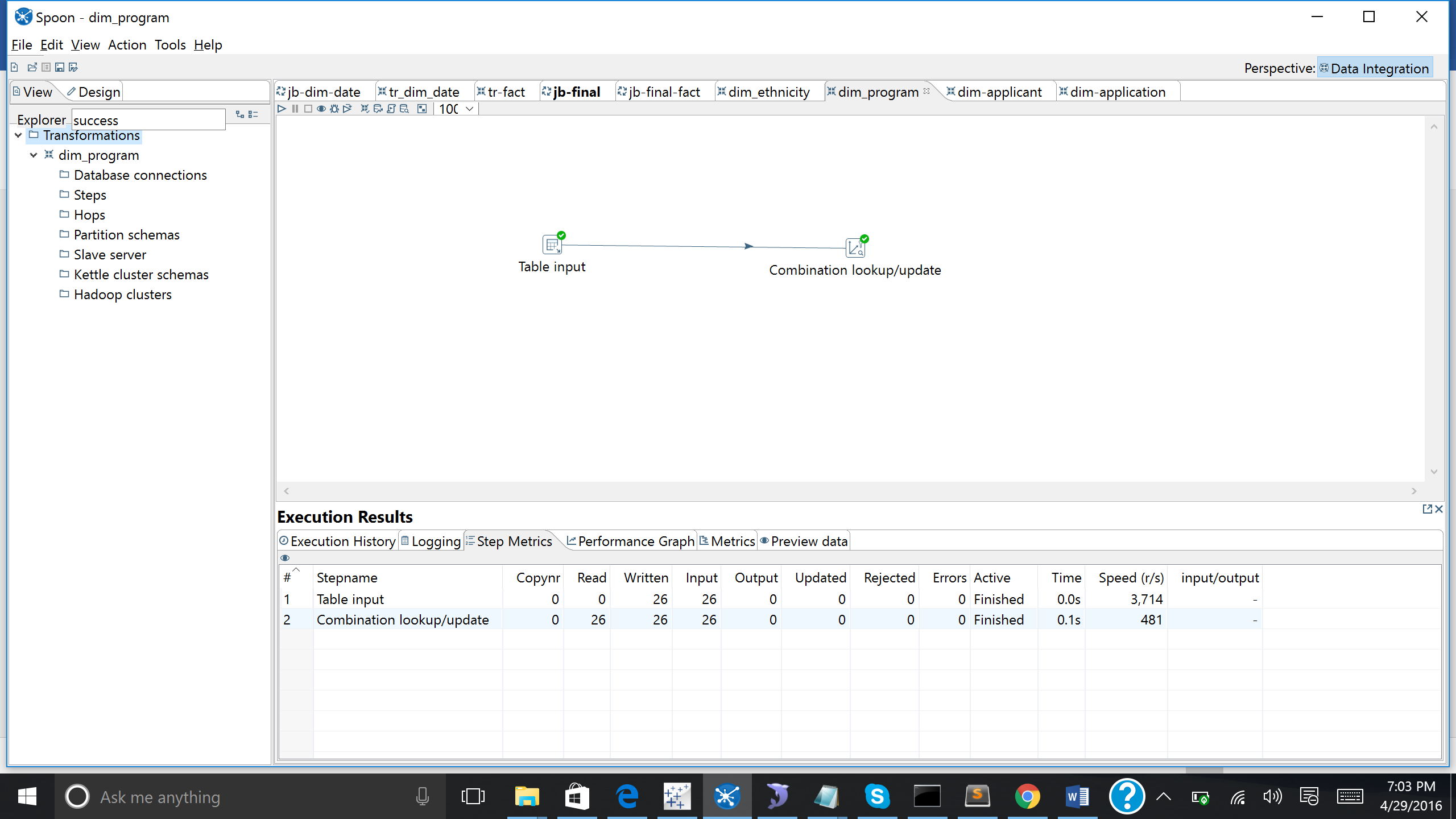
Created Transformation for dim\_applicant



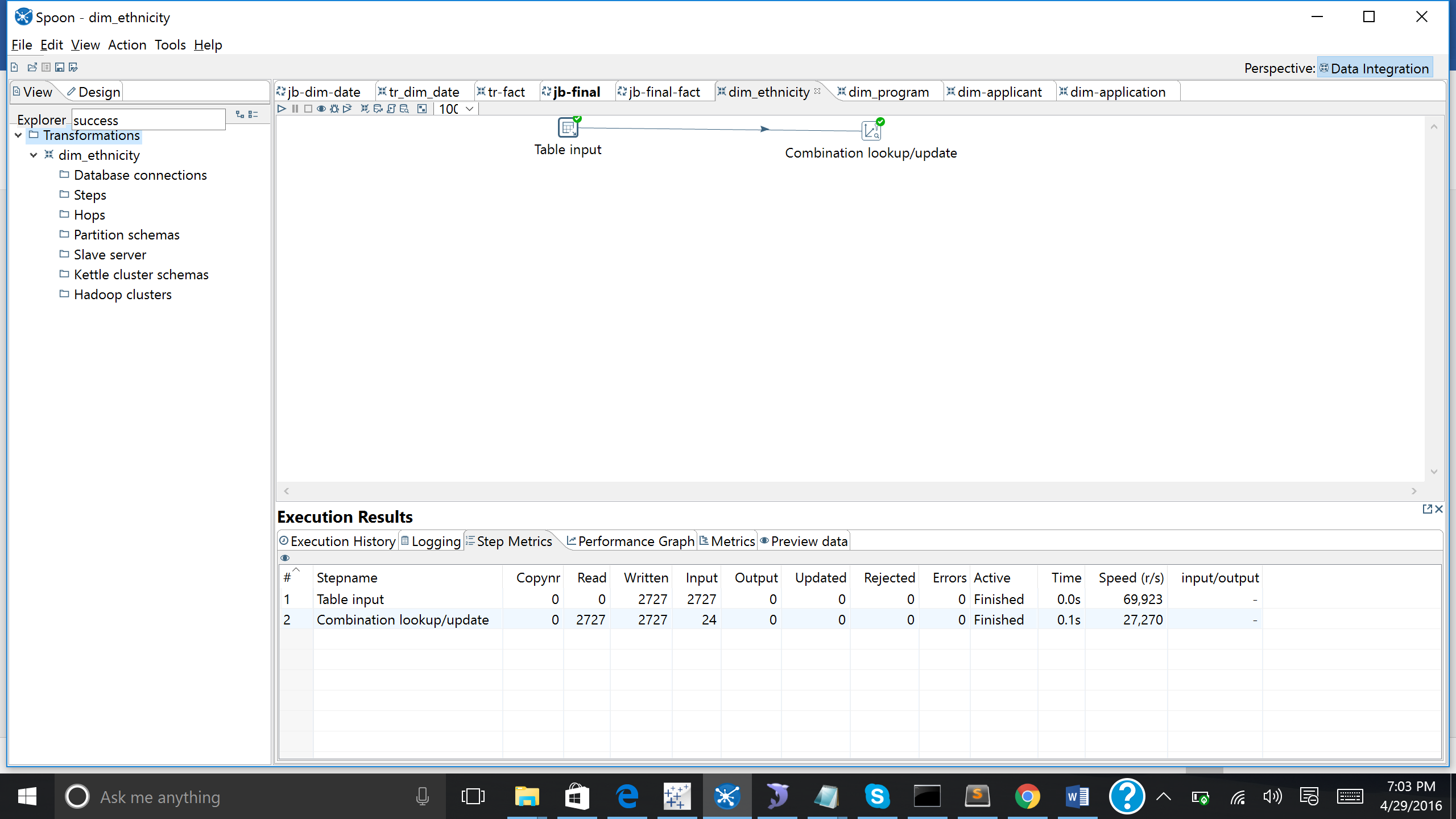
Created Transformation for dim\_application



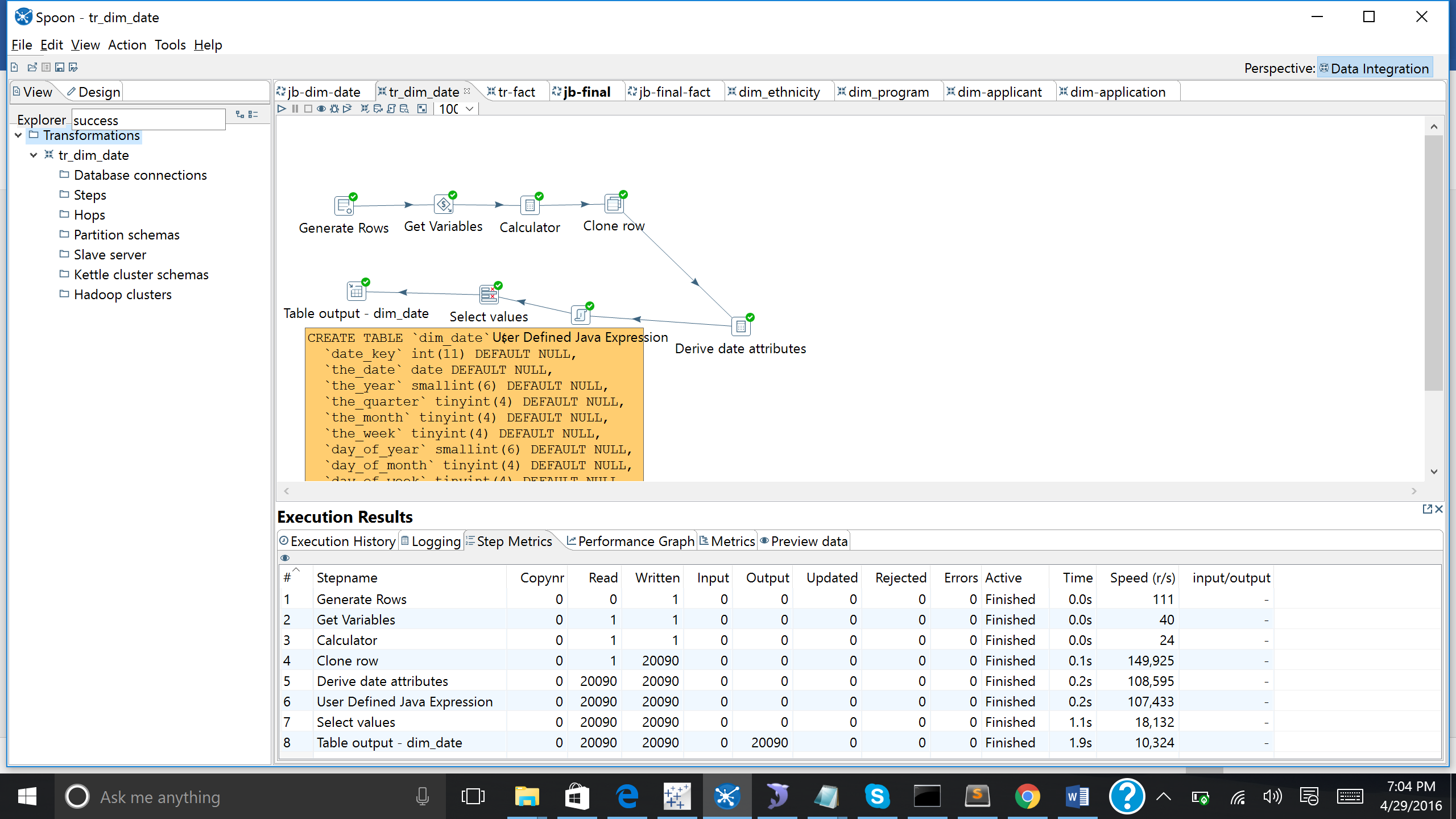
Created Transformation for dim\_program

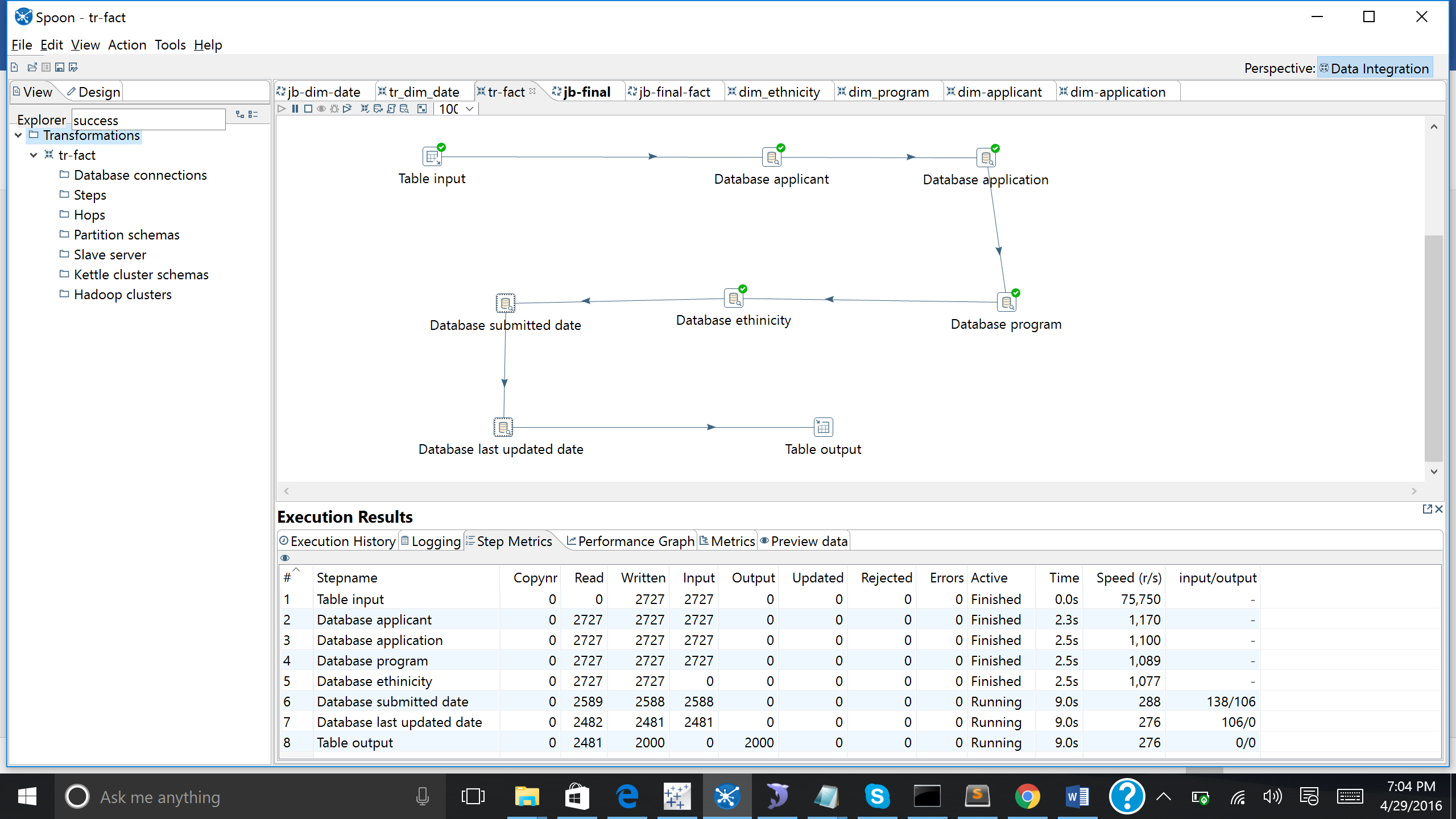


Created Transformation for dim\_ethnicity

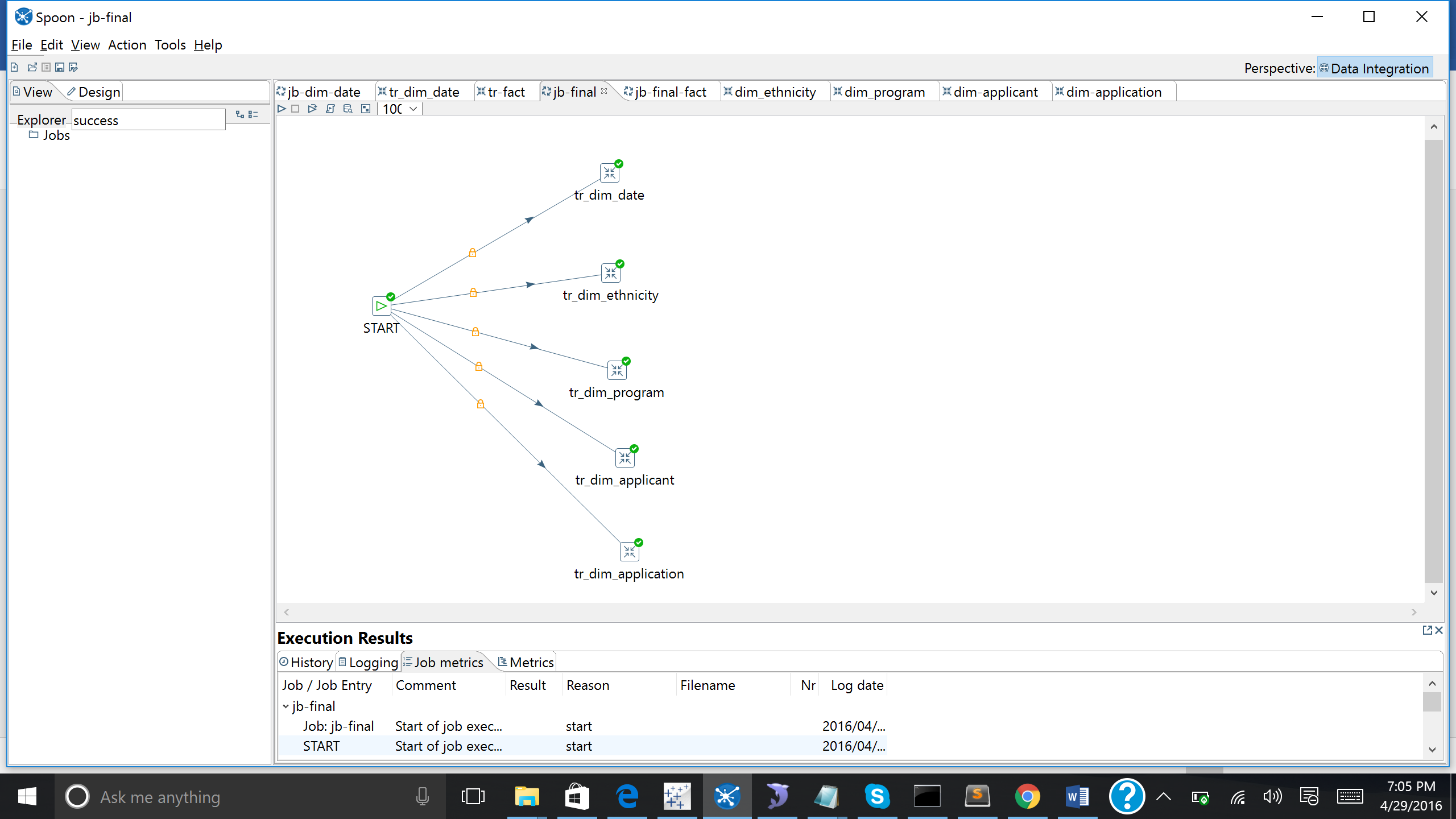


Created Transformation for dim\_date

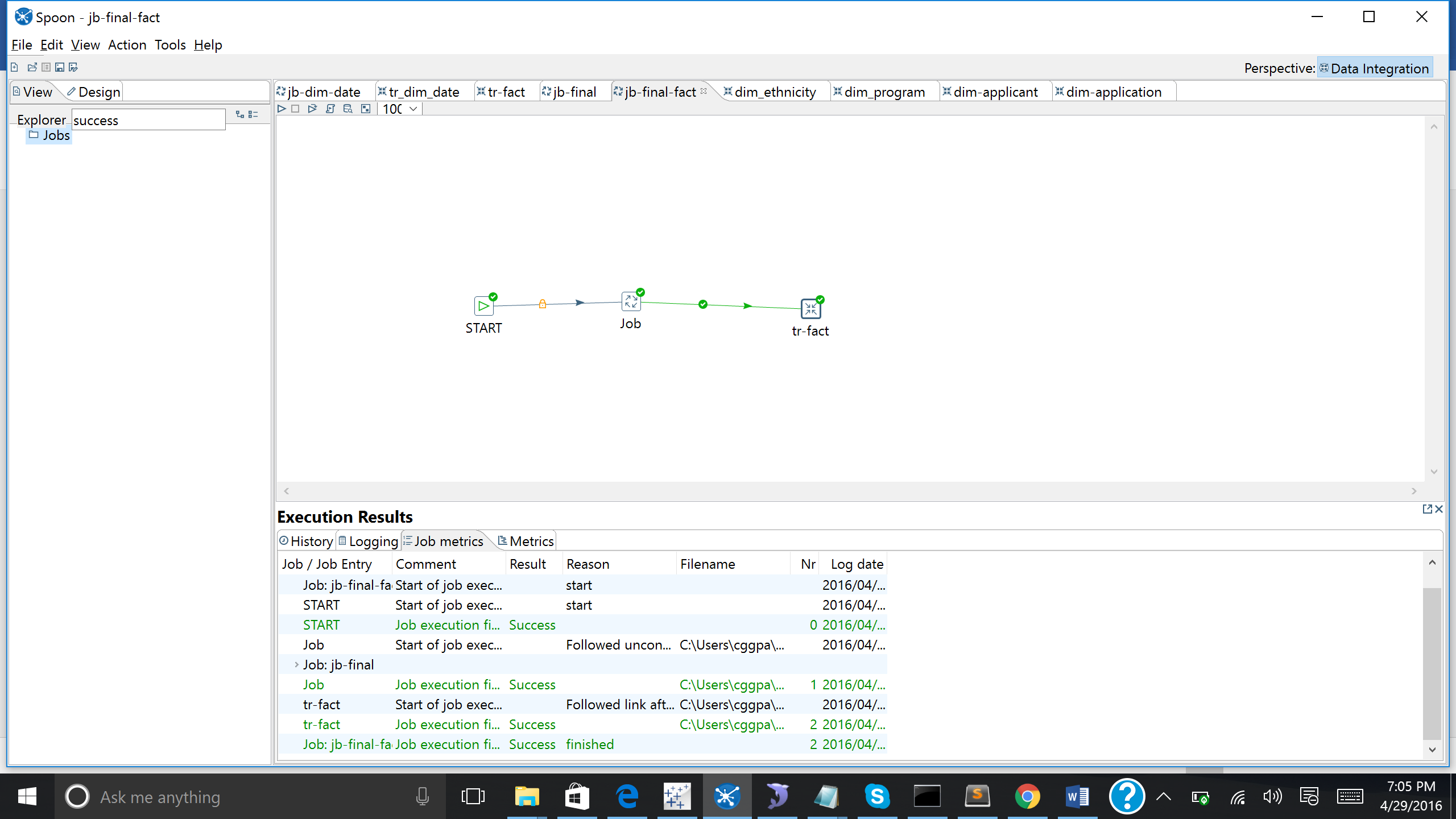


Created transformation for the fact\_application1 by retrieving the keys from all the tables.

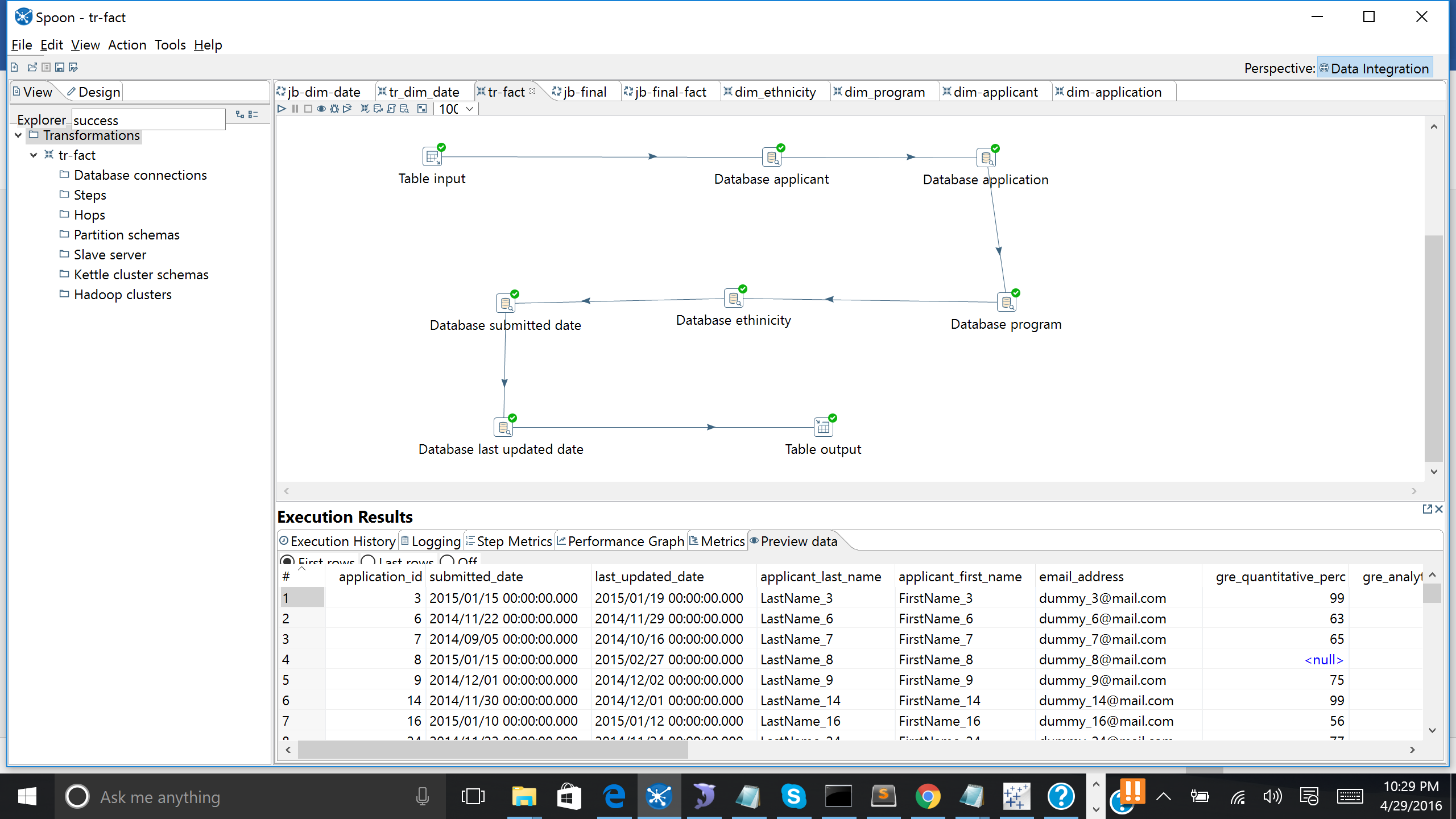
Running the job for all the transformations

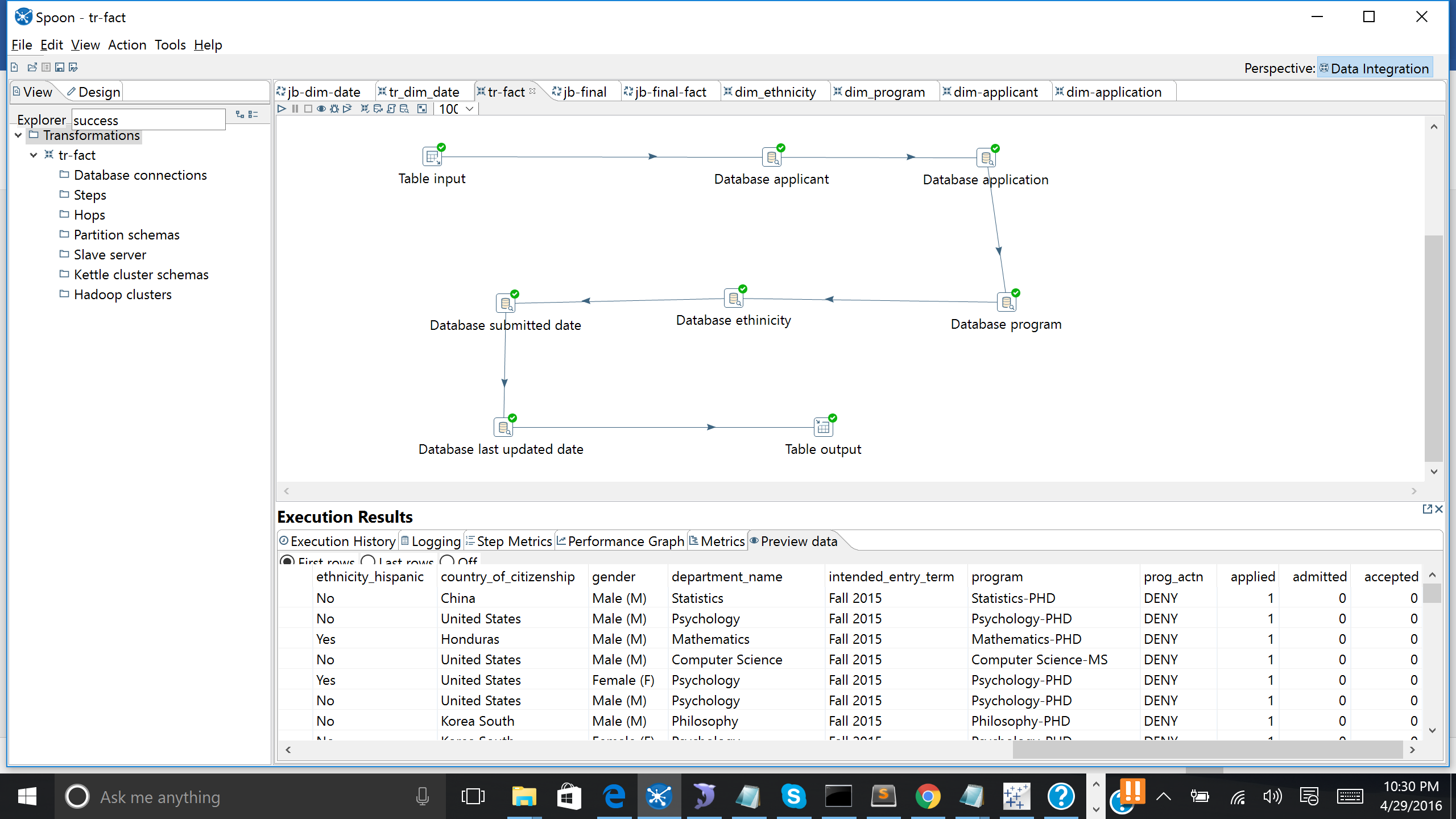


Running the final and the data is inserted into the Fact\_table:



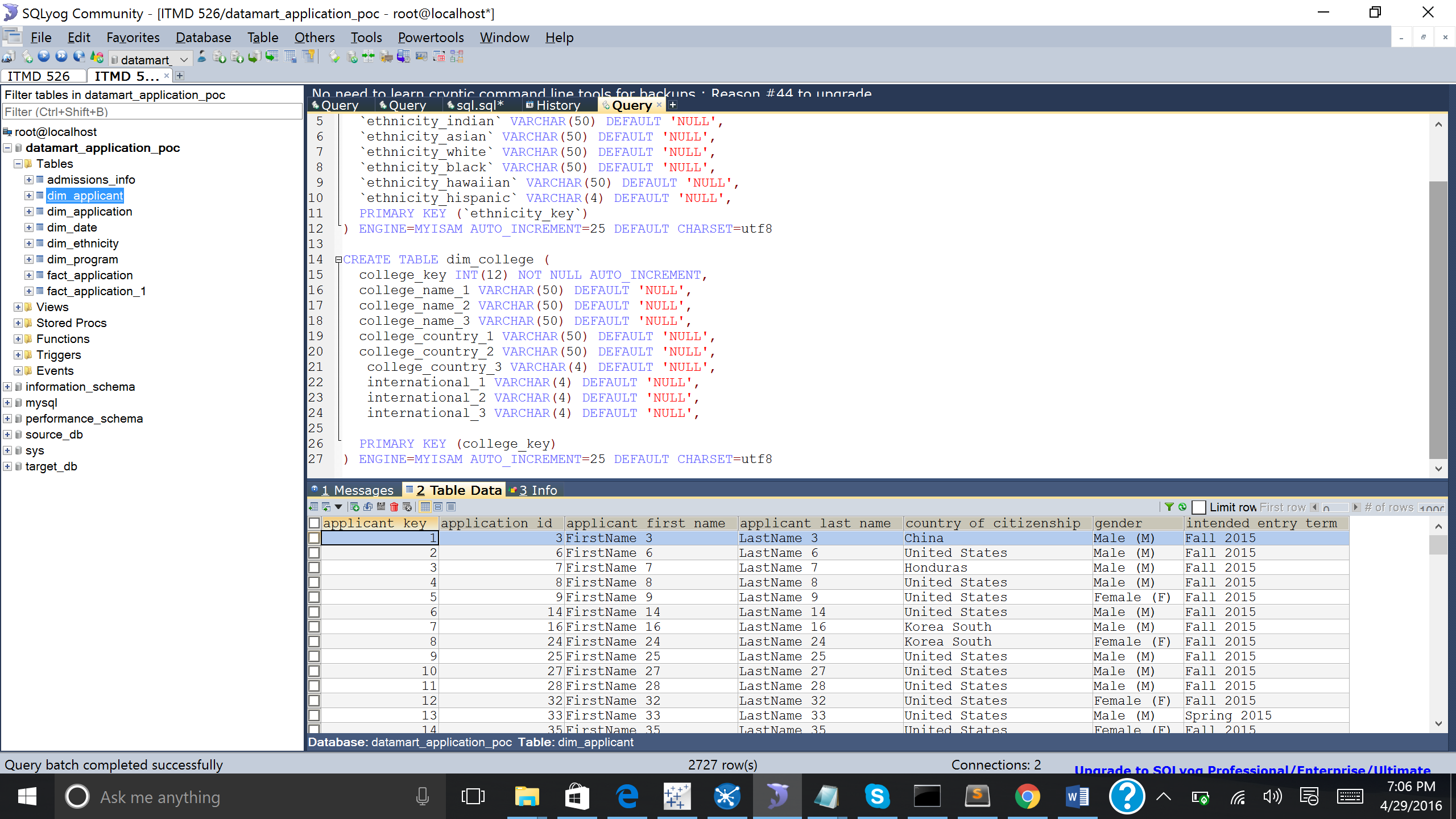
Data in the fact\_Table:



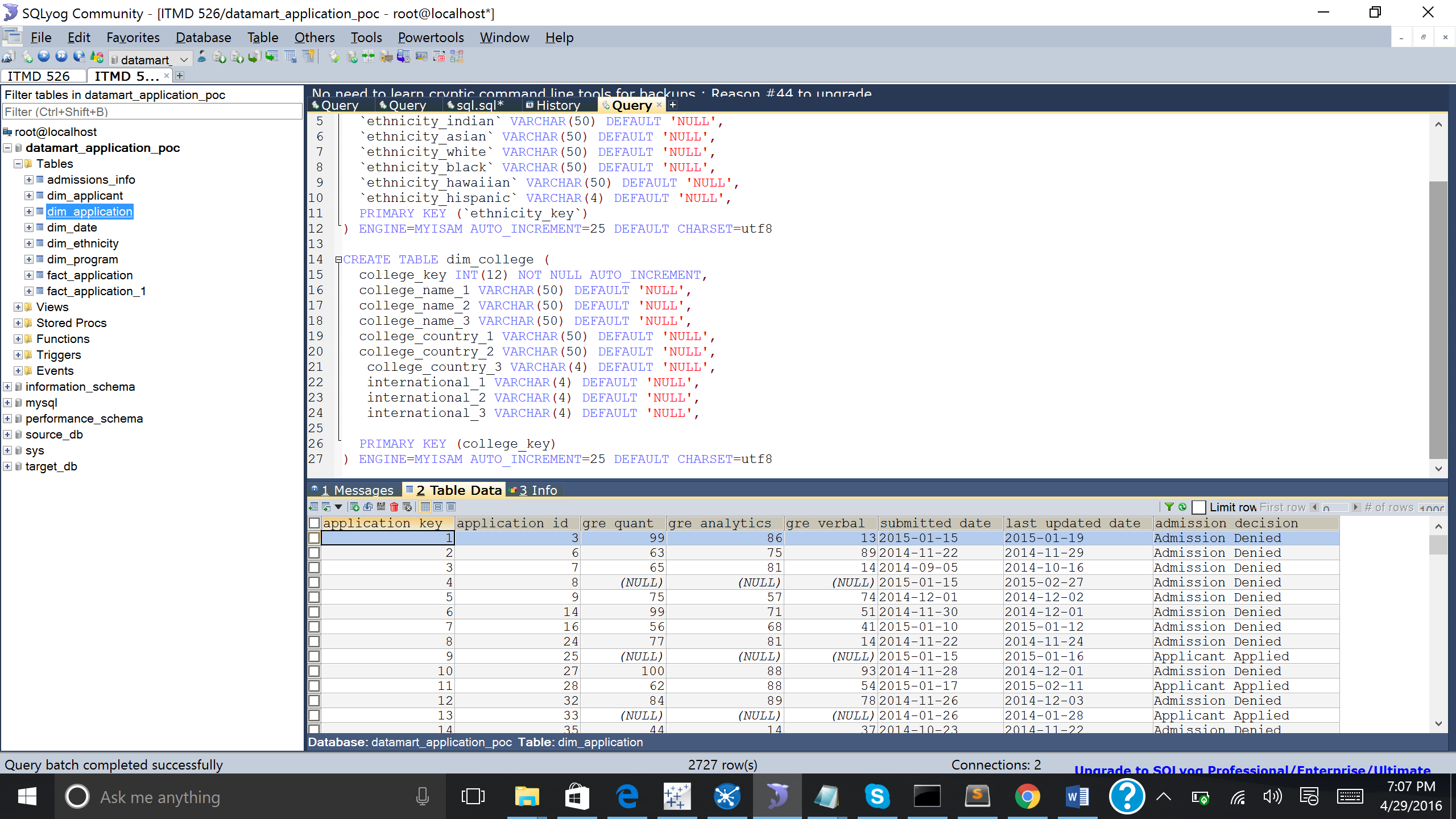


**Data inserted into the Tables after Transformation:**

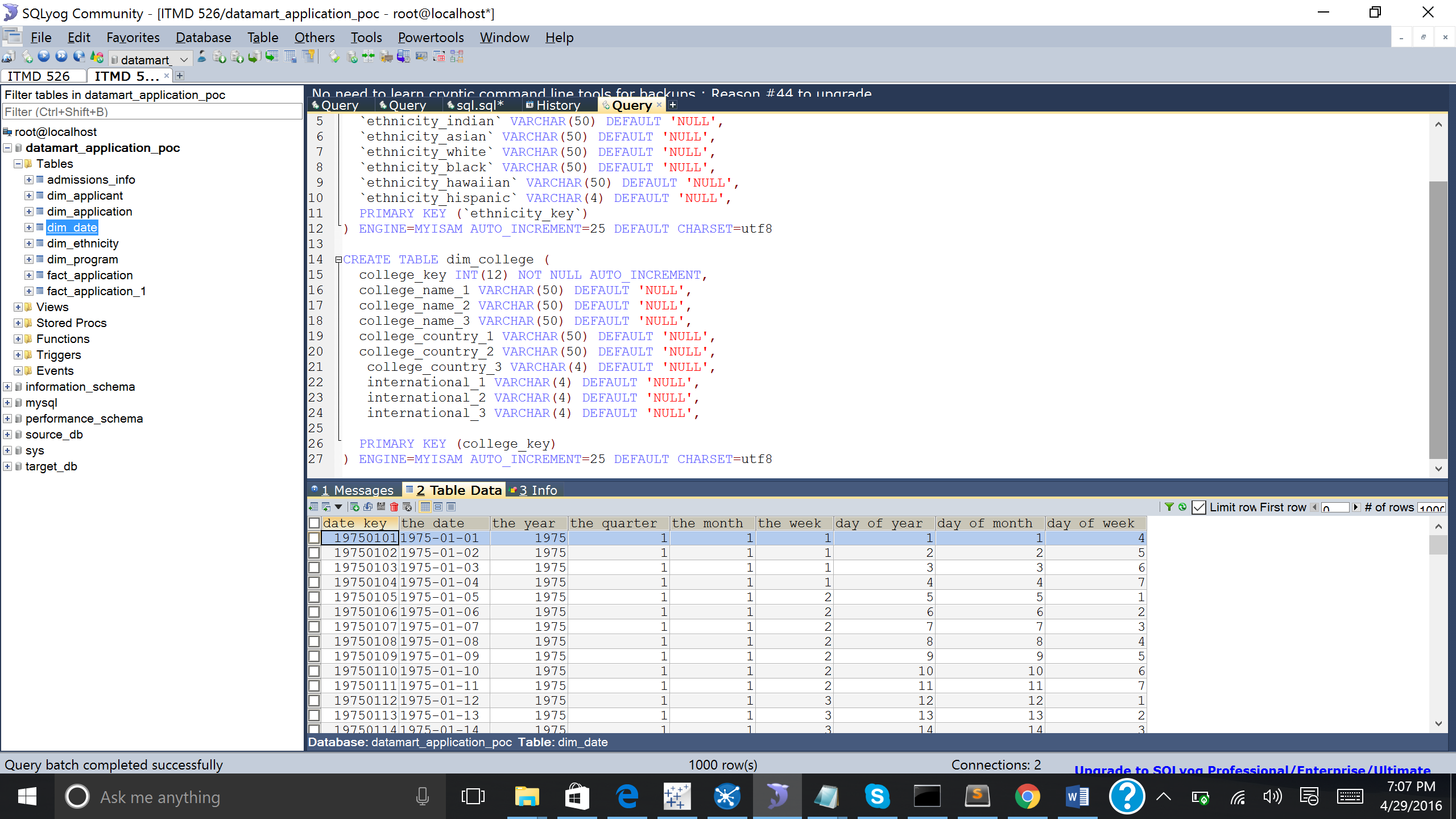
Data got inserted into the table after running the transformation dim\_applicant



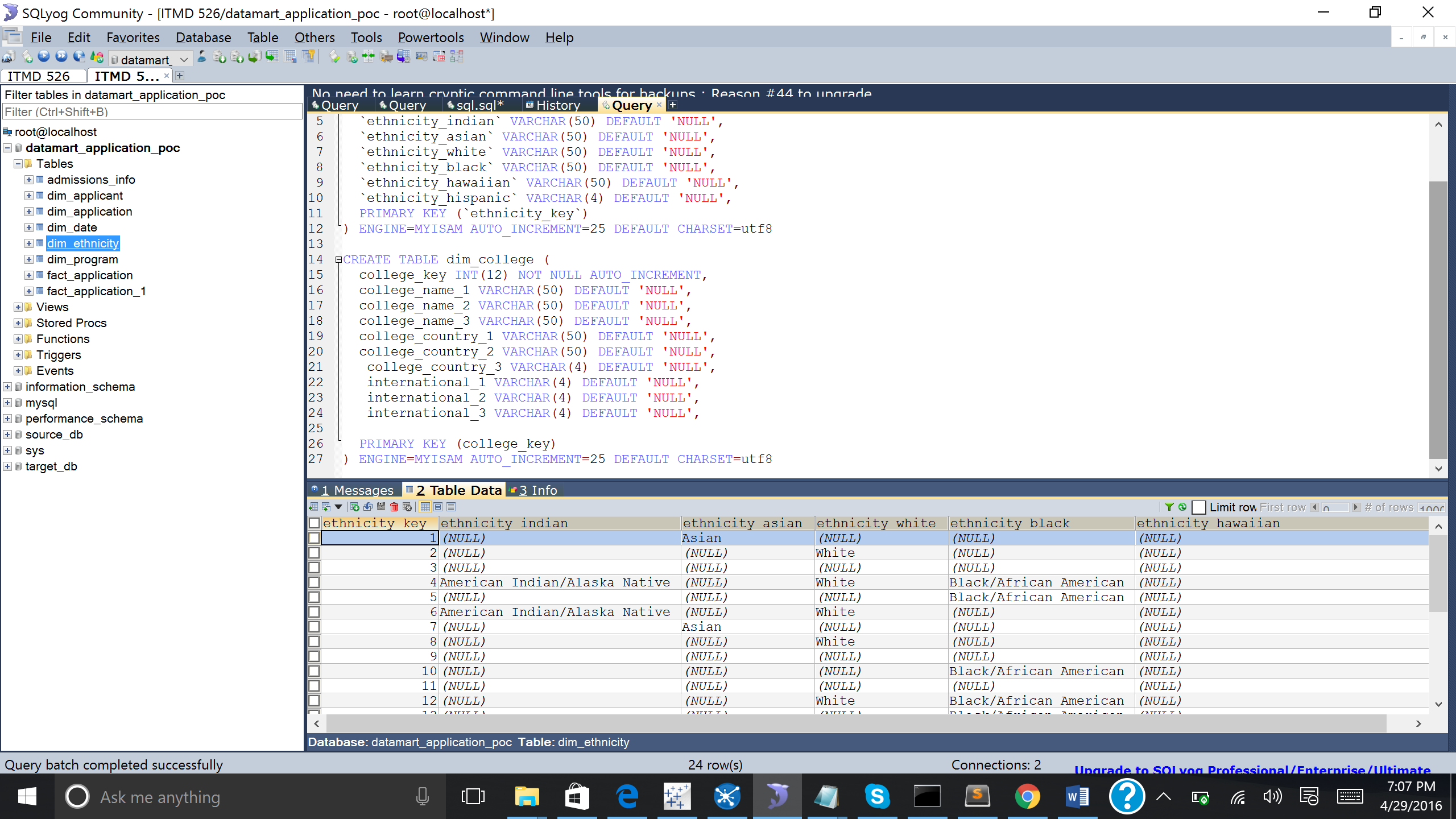
Data got inserted into the table after running the transformation dim\_application



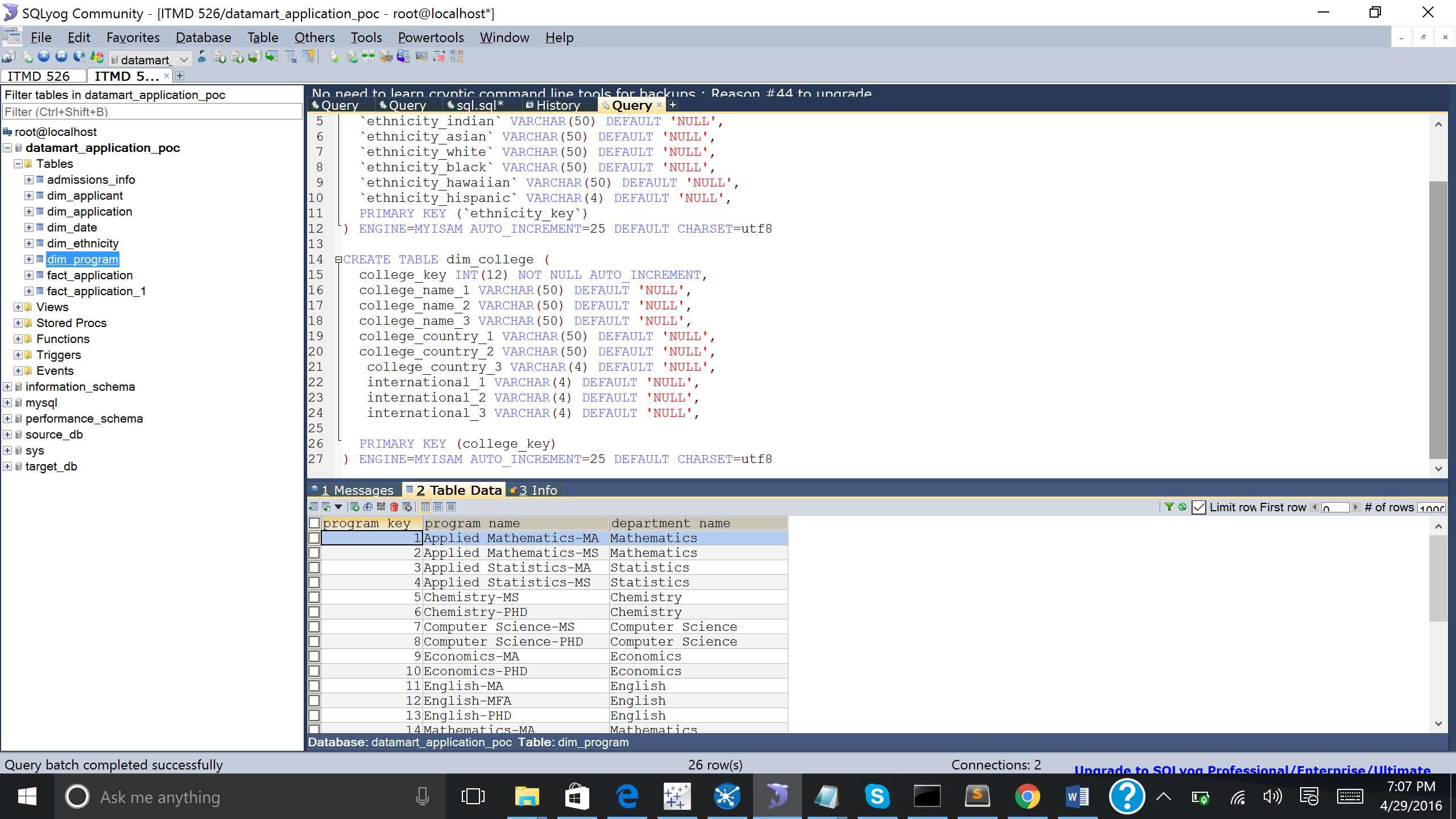
Data got inserted into the table after running the transformation dim\_date



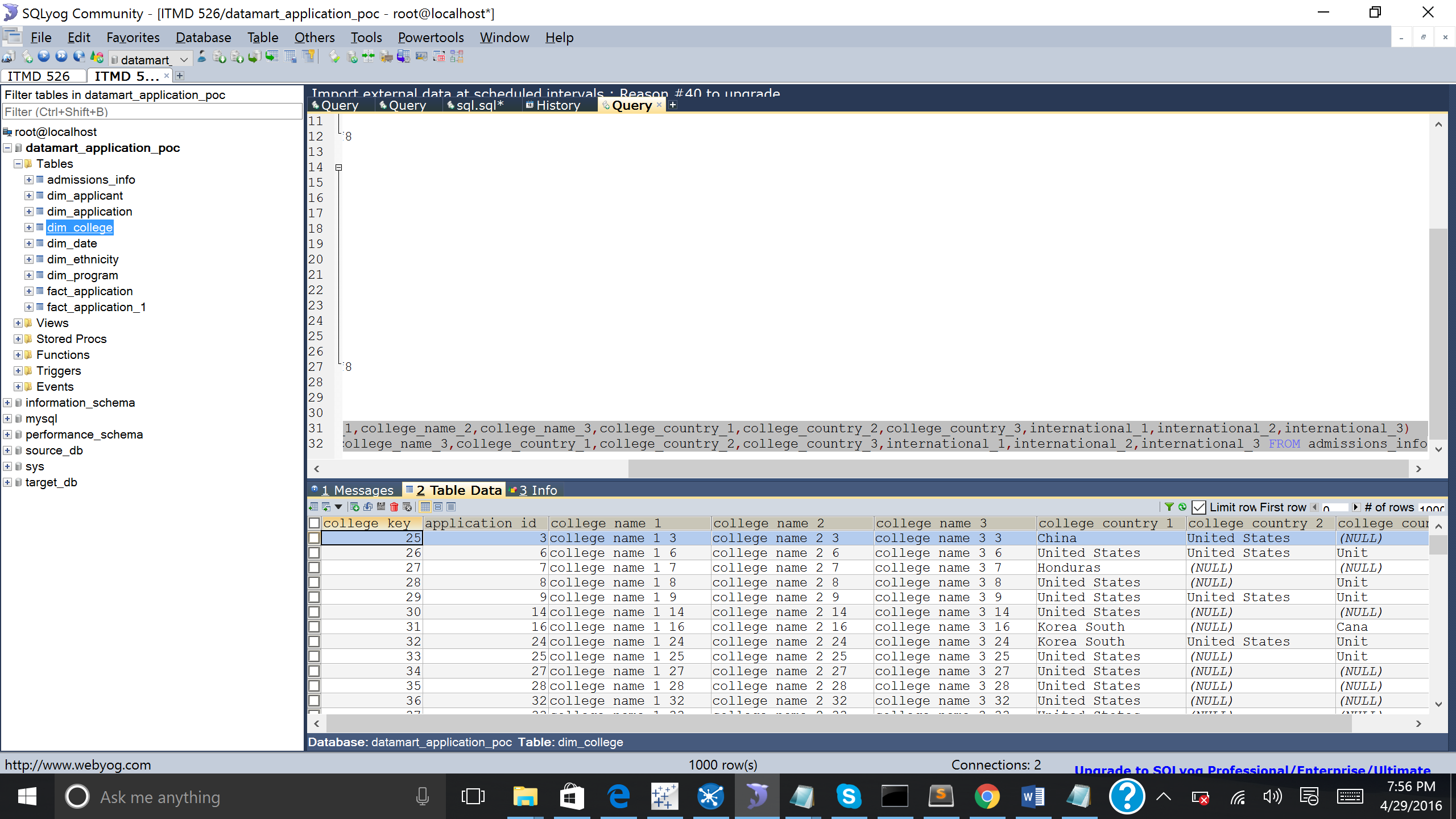
Data got inserted into the table after running the transformation dim\_ethnicity



Data got inserted into the table after running the transformation dim\_program



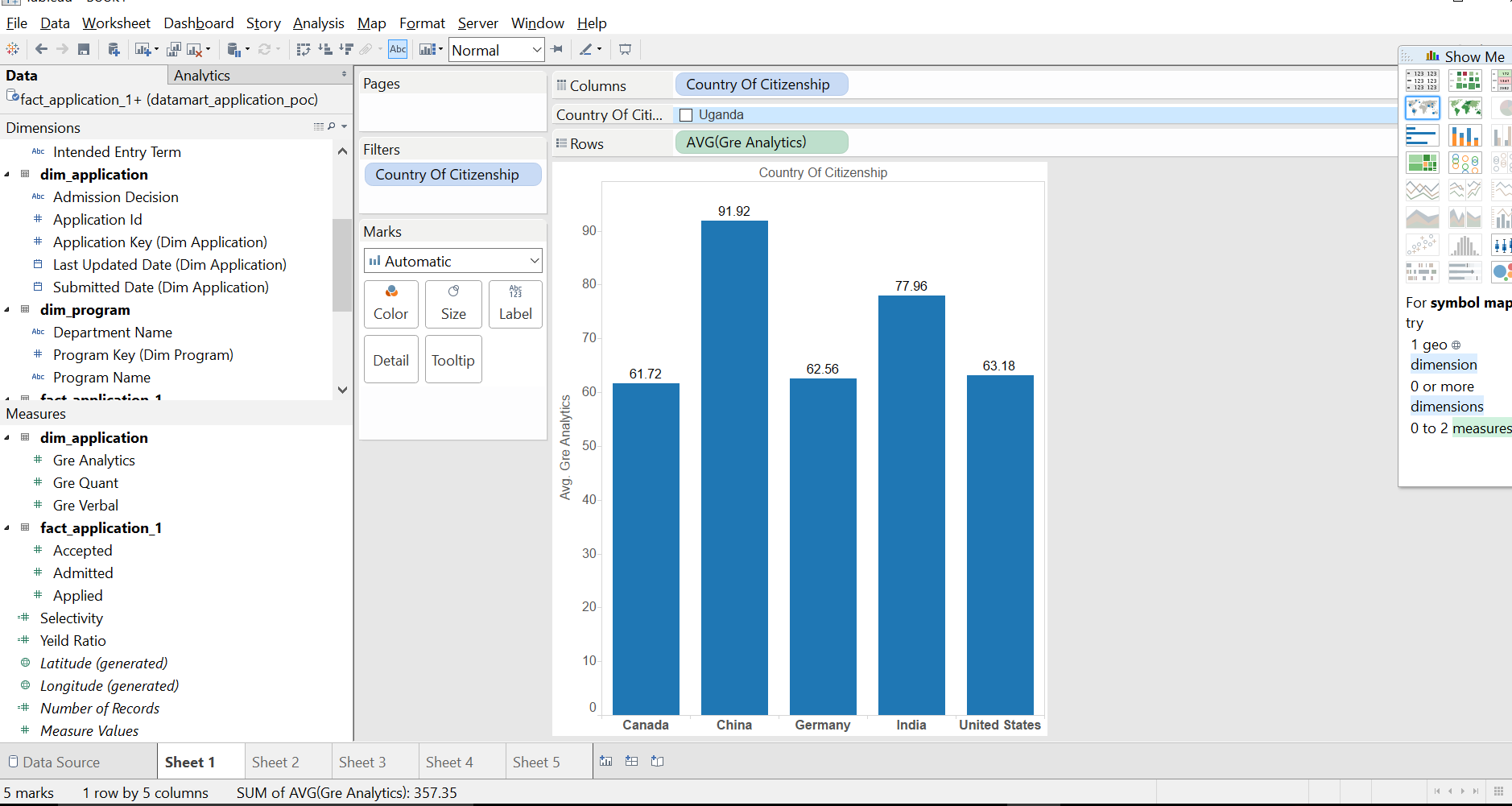
Data got inserted in the Bridge table (did not perform any transformation)



**Analysis and Reporting performed in tableau 9.3**

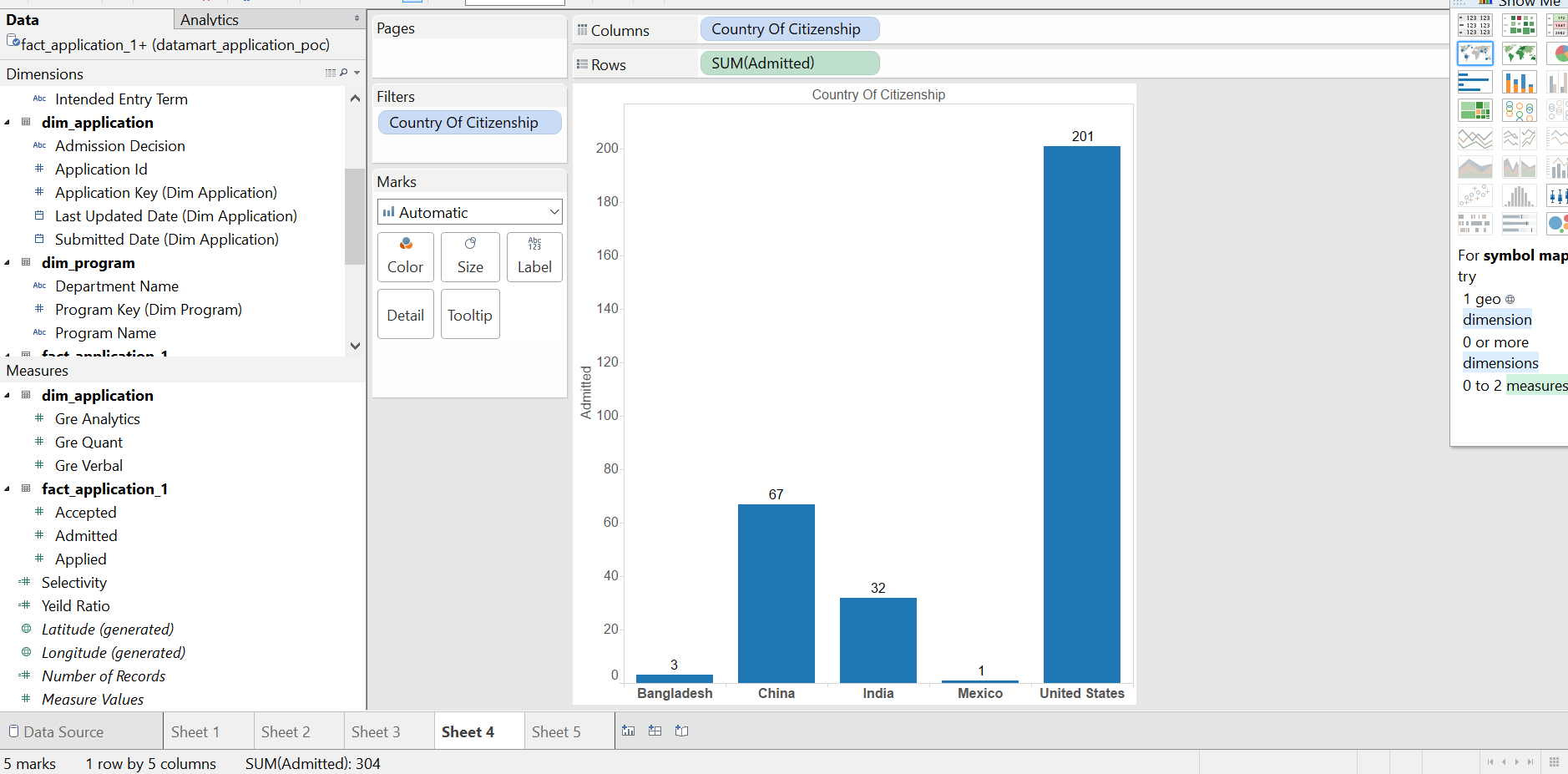
**Average GRE Analystics scored of five countries.**

Below shows the average GRE Analytics score of five countrie i.e., Canada, China, Germay, India and United States. It is seen that china has the highest average GRE analytics score of 91.92 followed by India. Canada has the least GRE Analytics score.



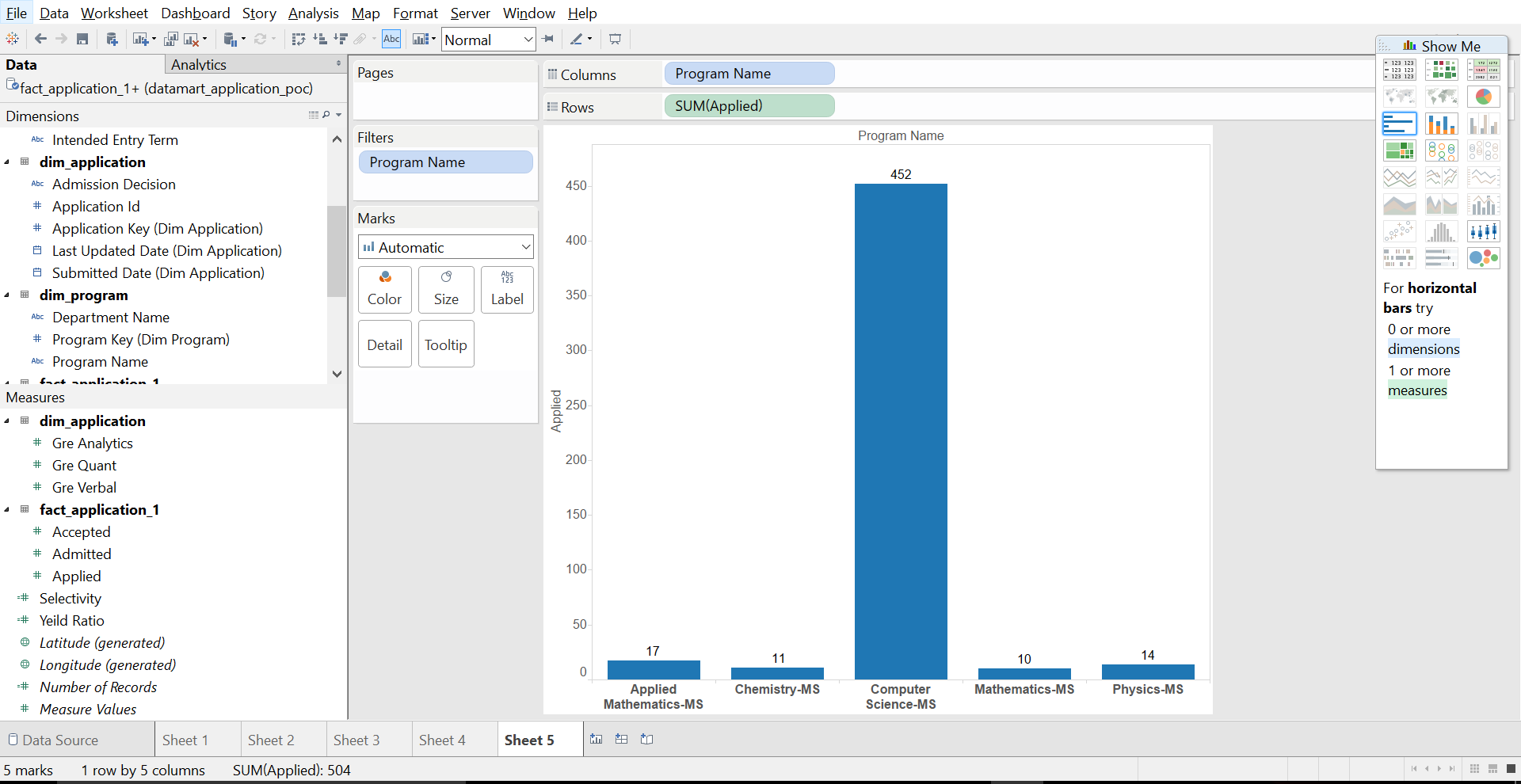
**Number of admits received for five countries**

Below shows the number of admits given to the students belonging to five different countires. It is seen that students from United States were given the most admits i.e., 201 followed with China (67), India (32), Bangladesh (3) and Mexico (1) which is been the least.



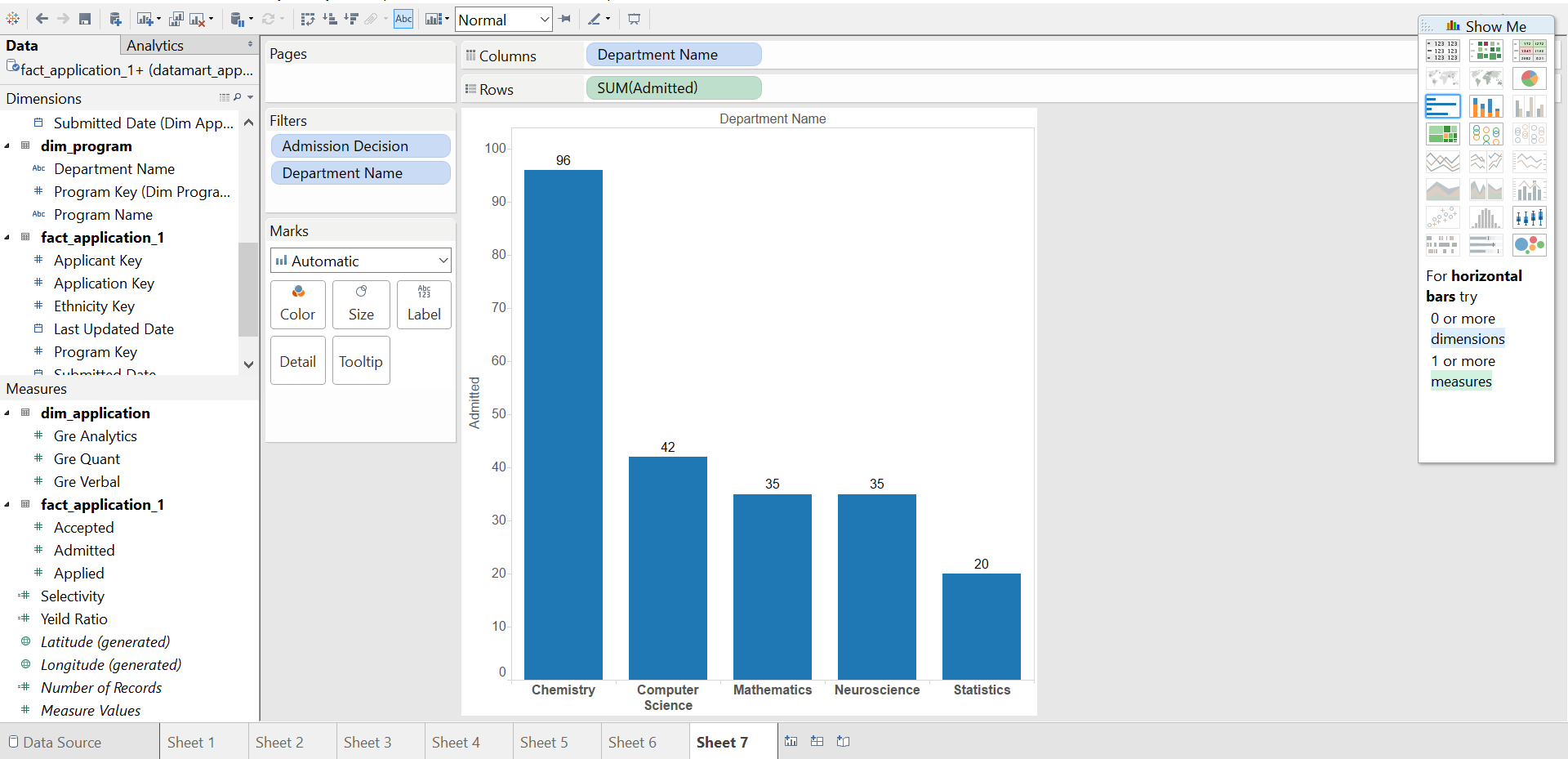
**Program name with the highest number of applications received:**

Below figure shows the students applied to five different program for MS program. It is seen that computer science MS has the highest number of applications received with count of 452. Mathematics-MS has the least number of applications received when compared with the other courses i.e., Computer Science, Applied Mathematics, Chemistry and physics.



**Highest number of admissions received/given for a department:**

Below figure shows the admissions given to students by five different departments i.e., Chemistry, Computer Science, Mathematics, Neuroscience and Statistics. Chemistry department has given many admits when compared with the other departments. The chemistry department has 96 admits and Statistics has 20 which is the least.



**Number of applications received for five departments:**

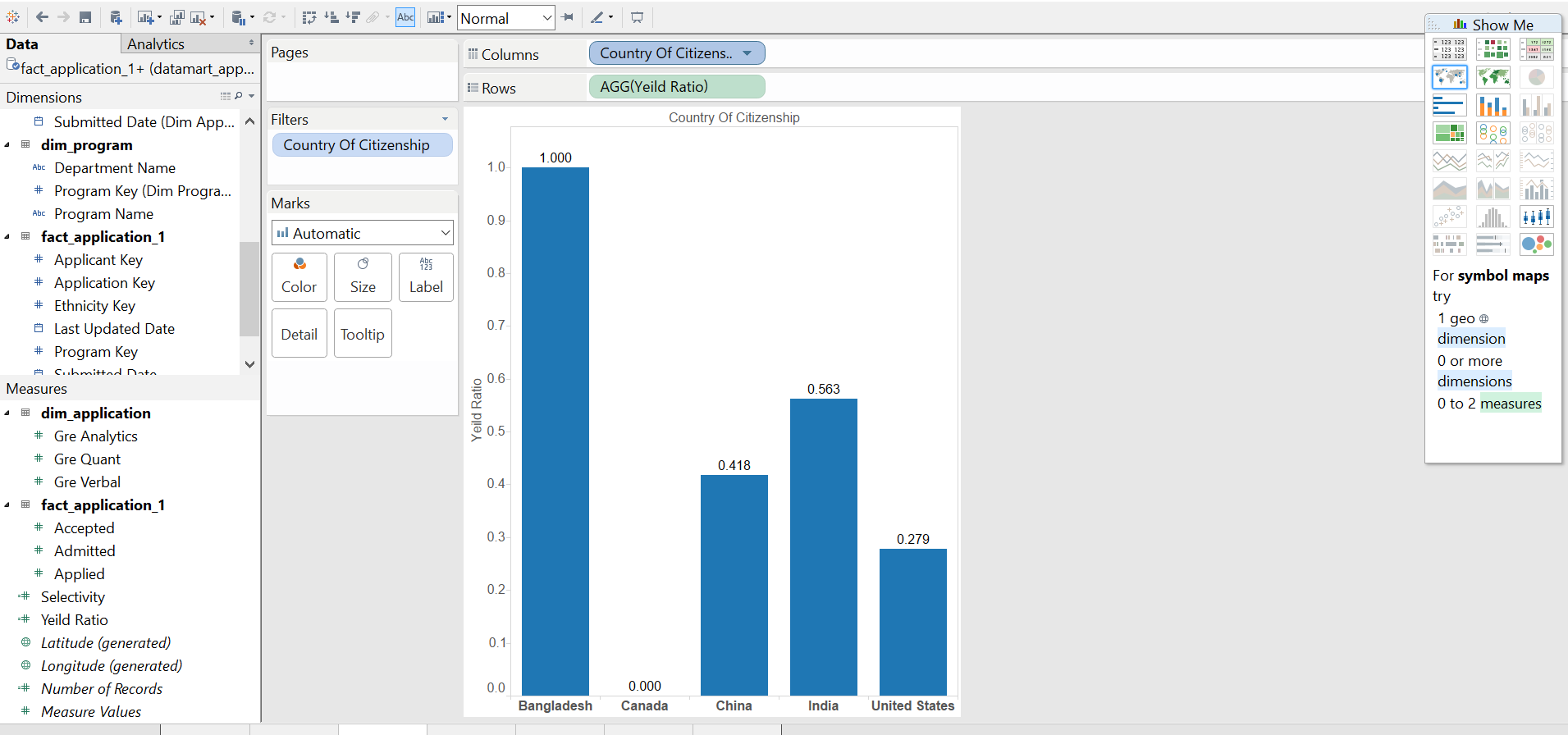
Below shows the number of students applied to five different departments i.e., Chemistry, Computer Science, English, Neurosciences and Statistics. Computer Science department has the highest applications applied by the student i.e., 579 followed with Statistic (304), Chemistry (237), English (242) and Neuroscience being the least (162)



**Two Calculated measured in the BI Application Layer:**

**Yield Ratio: Number of Accepted/Number of Admitted.**

Below shows the yield ratio for the five countries i.e., Bangladesh, China, India, Malaysia and United States. It is seen that yeild ratio is one for the country Bangladesh and least for Canada.



**Selectivity: Number of Admitted/ Number of Applied.**

Below shows the selectivity plotted against five different countries i.e., Bangaladesh, China, India, Mexico and United States. We can see that Bangladesh has the highest number of selectivity i.e., 27.27 followed with United States and least China

