# Architectural Components of DW

## Source Data

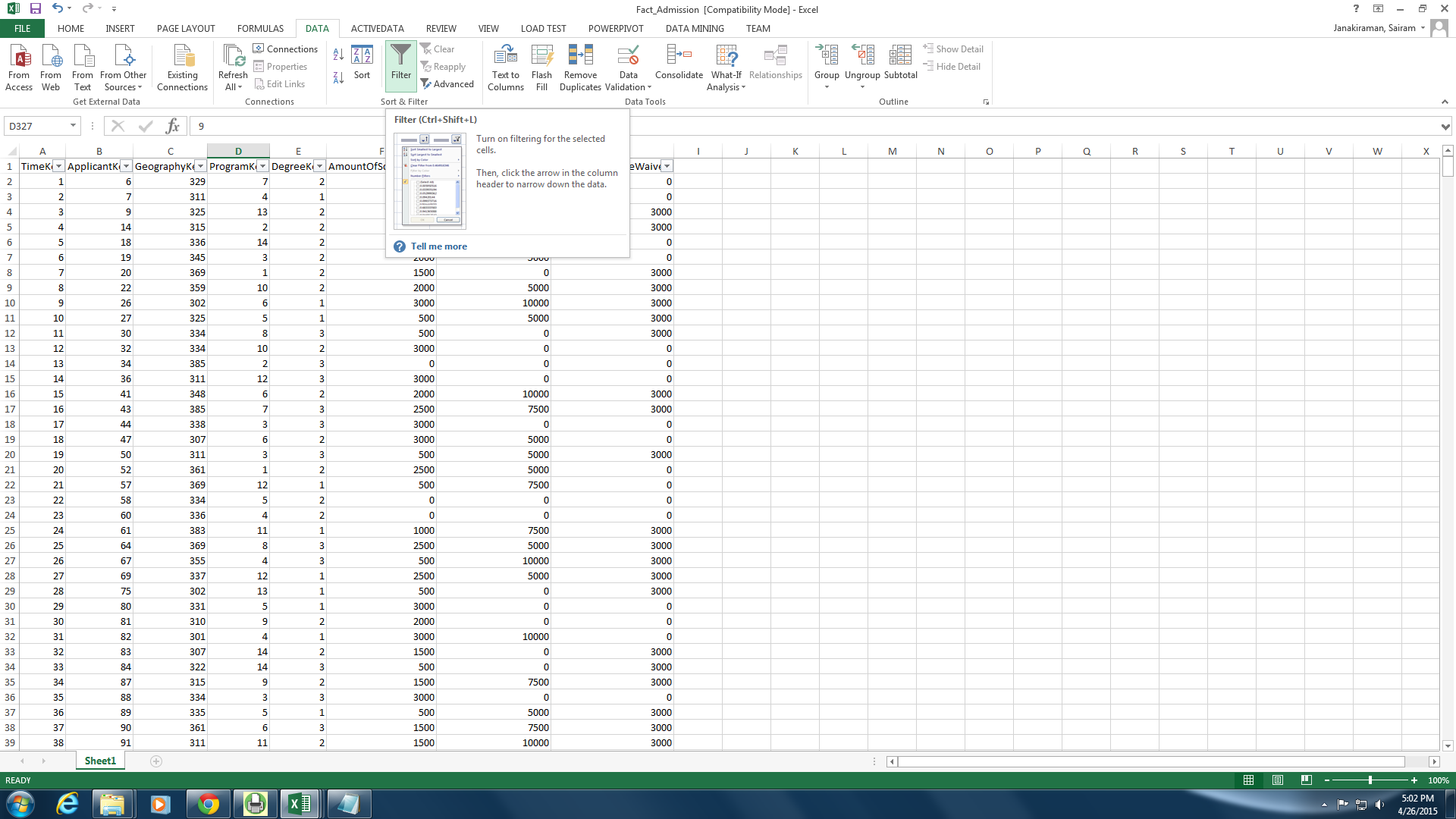
In general, different data sources feed data into a university data warehouse. Applicant data could be from operational data sources while faculty data could be from internal data sources such as files, spreadsheets or documents. While implementing the prototype for Brazos University Data Warehouse, data sources were not readily available. As a result, we had to create our own data sources which could then be used to feed the data for BU data warehouse.

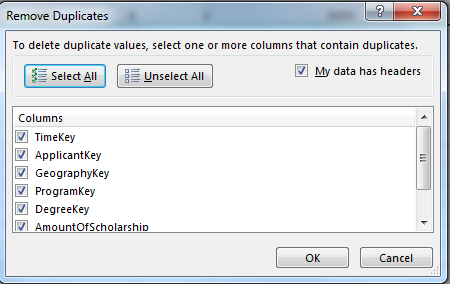
There are many data generators that are available on the internet. We used [www.mockaroo.com](http://www.mockaroo.com) data generator to generate data for admissions, faculty, applicant, facility, registration tables etc. The data for the courses and department tables were extracted from the Texas A&M University website since mockaroo was unable to generate the related data.

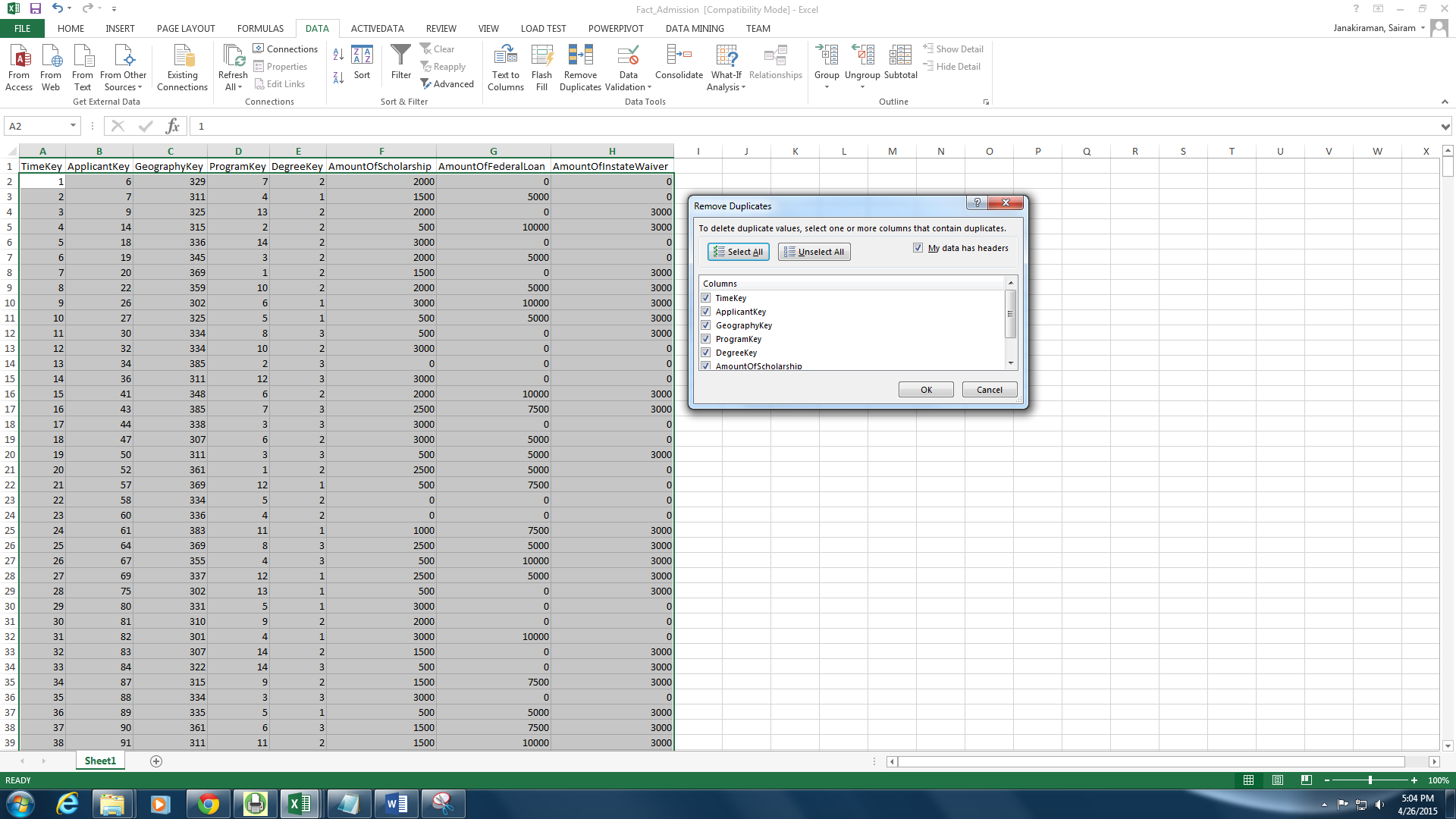
## Data Staging

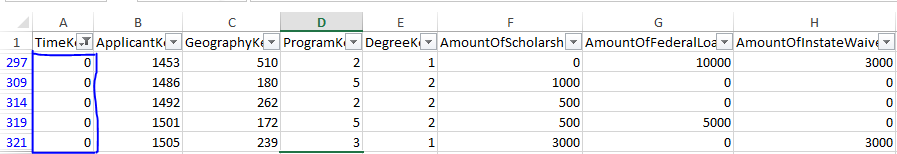
After the data sources were identified, a set of ETL processes were performed to enable the data to be loaded into the data warehouse. Data was extracted from the source system and stored in excel spreadsheets. While following the extraction process, business rules were followed to make sure that the structure of the data follows data warehouse schema. For example, each dimension table was assigned a surrogate key that would be the unique identifier for dimension table.

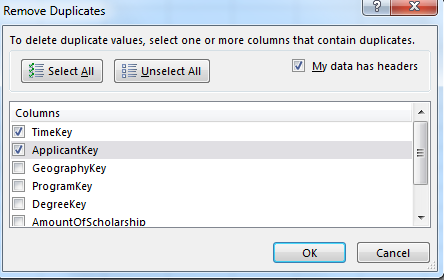
After the extraction process, each table was checked for data and referential integrity in case of a foreign key dependency. As part of the transformation process, aggregate data for the fact tables were added. Since the data was created by the team using external data generators, aggregate data created did not follow business rules. To mitigate this issue, formulas were created in excel, which would follow the business rule parameters. Once all the data were collected, the data was checked for duplicates and missing value using Excel’s *remove duplicate* functionality. The sequence of diagrams below illustrate the data staging process followed for Brazzos University Data Warehouse.

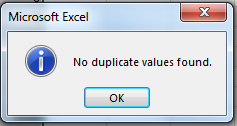






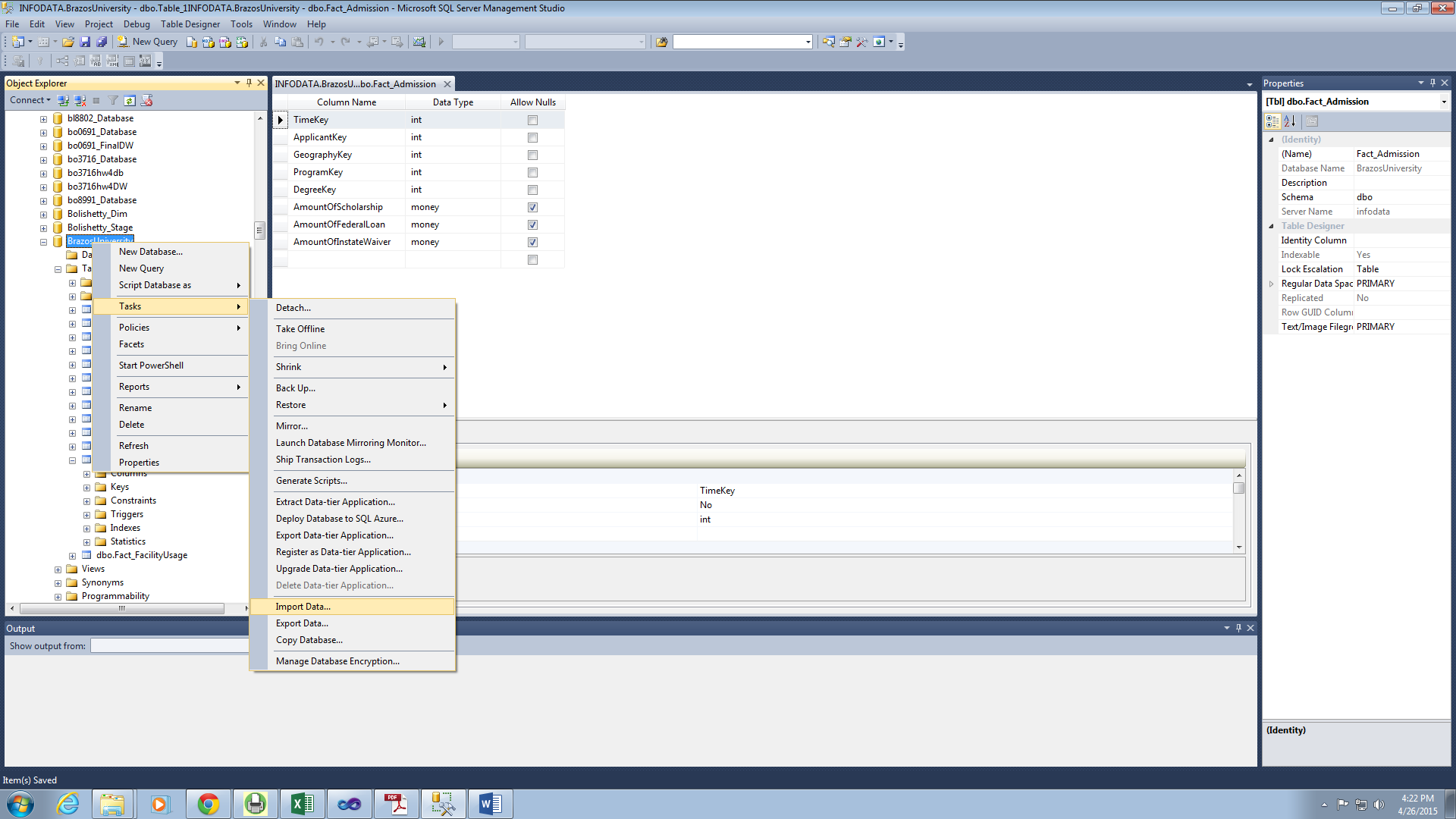




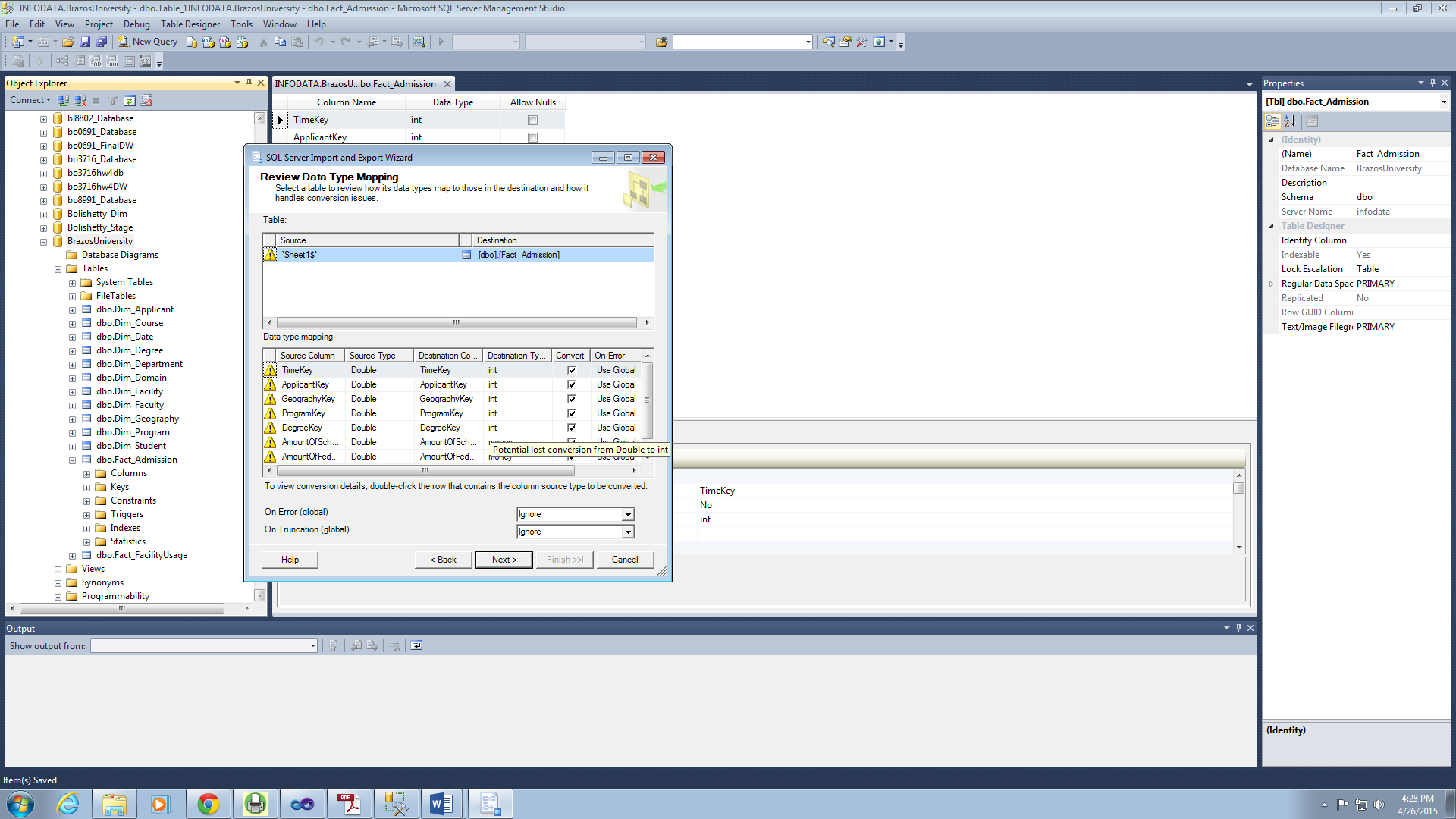


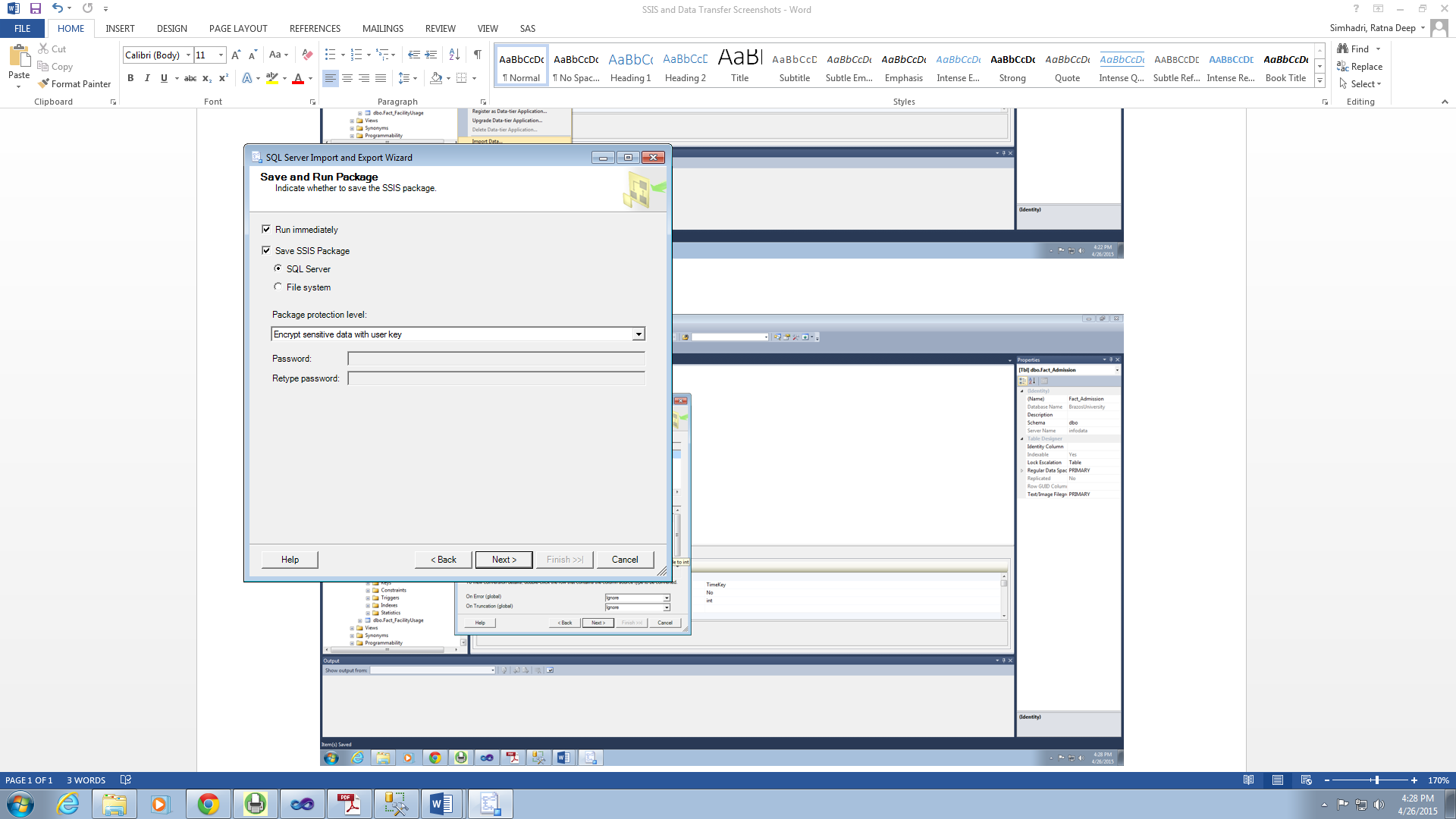
## Data Storage

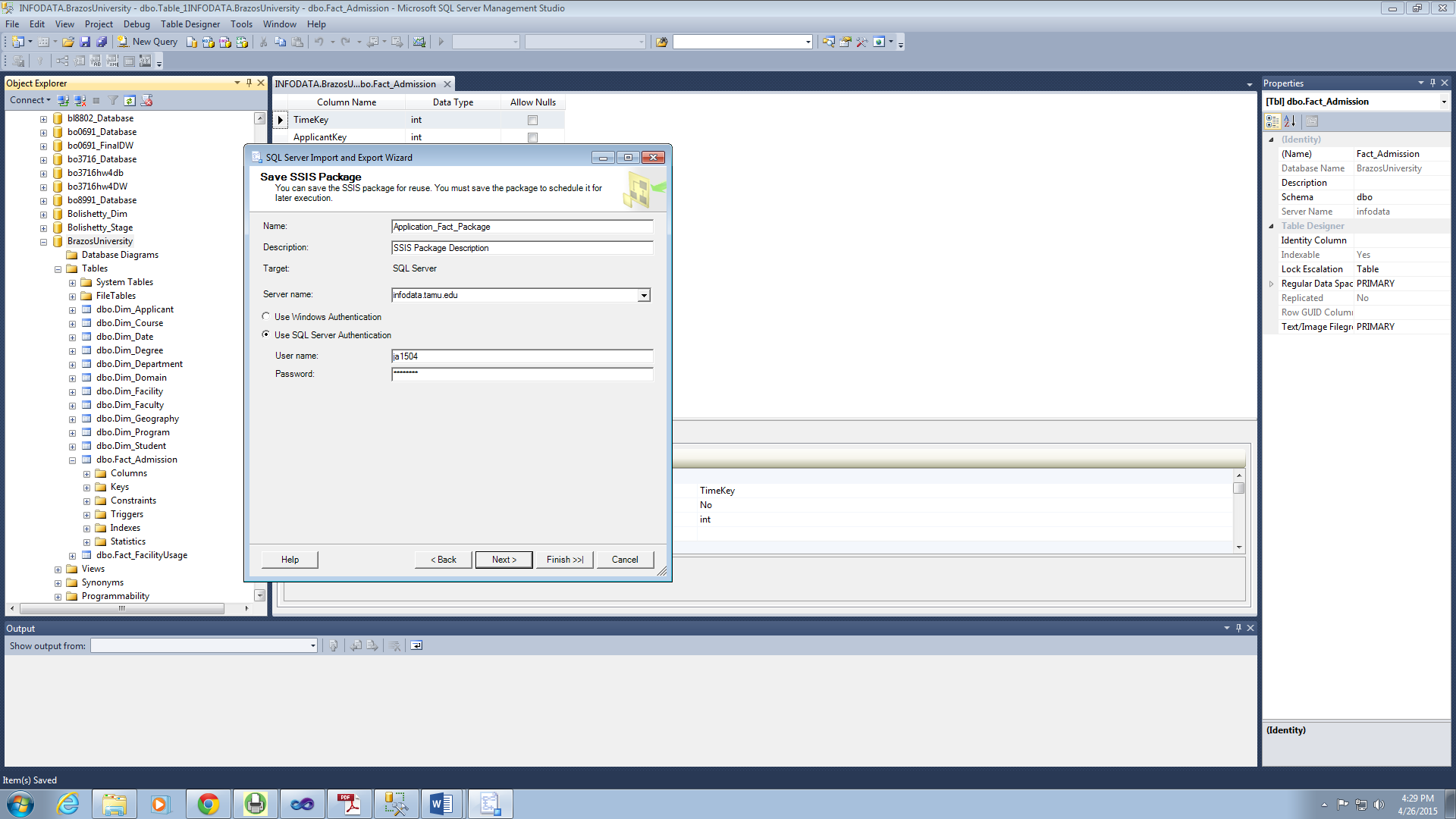
Once the transformation process is complete, duplicates of each data file is created to provide backup and recovery options if needed. Before loading the data into the data warehouse any remaining primary and foreign key referential are identified and resolved. The data sets are created and consolidated as excel sheets. Now these excel files are imported through the import option in the *SQL Server Management Studio*. Below are the screenshots of the steps that were followed to load the data into the data warehouse.

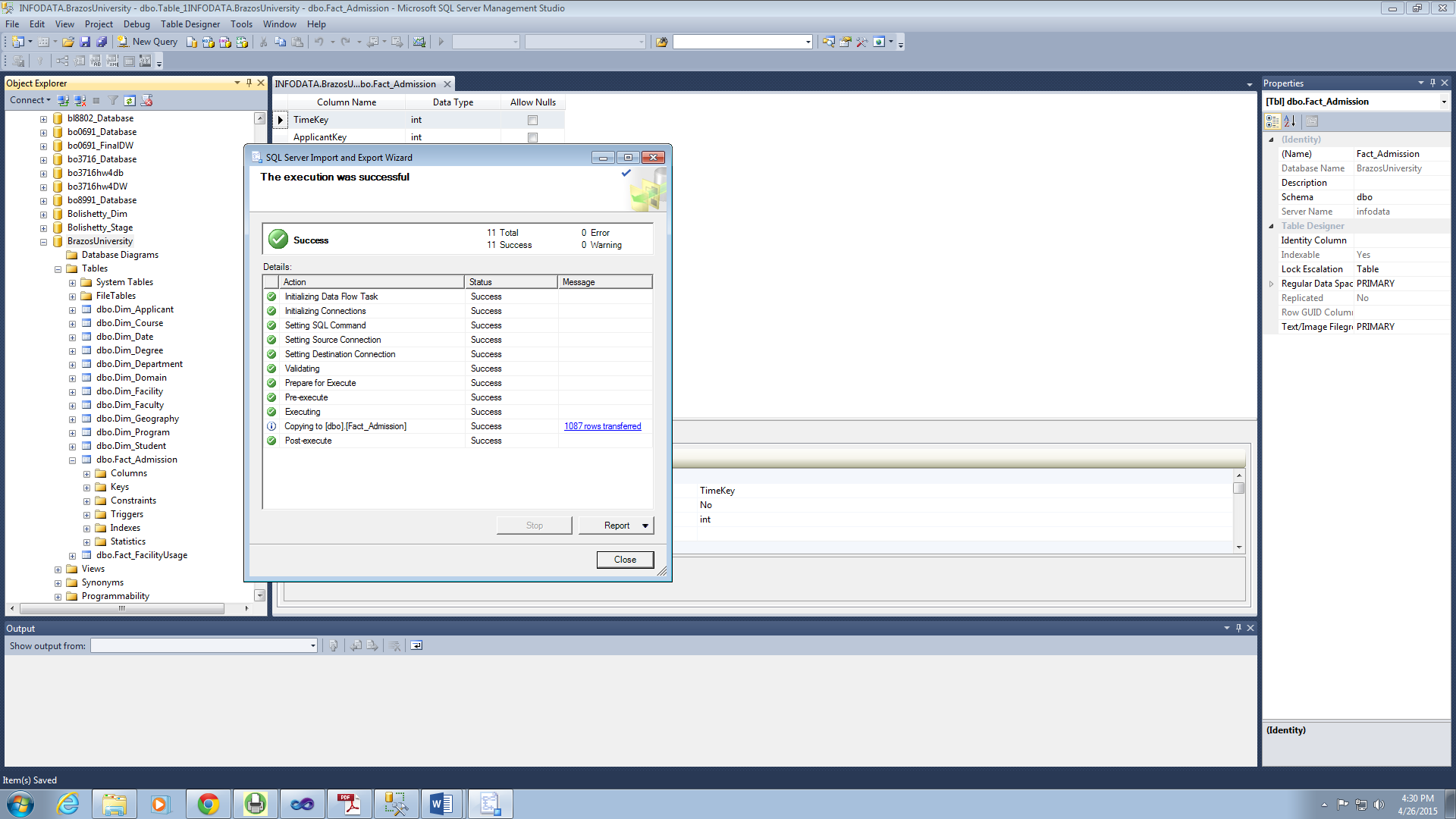


After importing the files, for each table a data type mapping is performed. This helps in defining the data types for each attribute of a table as well as assign foreign key referential for a particular attribute.









## Information Delivery

Once the data is loaded into the data warehouse, there has to be a process which allows the users to access the data. The strength of any data warehouse architecture is manifested through the robustness and flexibility of the information delivery component. To achieve this, we have provided online analytical processing (OLAP) to help the users to generate reports that would be useful in making business decisions for Brazzos University.