|  |
| --- |
| **Data Acquisition**  **Week 3 Assignment** |
|  |

**Task**

Set up ETL using SSIS (SSDT) to extract, transform and load necessary data from sources (Yelp and Weather) to your data warehouse.

For this week’s assignment, we set up an ETL workflow to get the necessary data from YelpDB and Weather Data CSV. Addressing the week 2 feedbacks, we decided to do vertical partitioning on the 3 out of 4 tables (business table, user table, and review table) that we will extract from YelpDB. This is because from those 3 tables, we do not need all the columns for our data warehouse that was based on the questions from week 1. We do this with the hope that when we extract the data from YelpDB, it will reduce the time needed to do the aforementioned process.

First of all, we will create a simple data flow node.  
**Graphical user interface, text, application, Word

Description automatically generated**  
  
  
  
  
  
  
Business Table  
In this data flow, we will then create our first extract job, which is from the business table.  
**Text

Description automatically generated with medium confidence**  
  
Then, we need to setup up our source DB, which is the YelpDB  
Graphical user interface, text, application

Description automatically generated  
  
After finish in setting up the source DB, we will define our query to get the data from the business table.  
Graphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generated  
  
  
  
We will also remove any rows that has a null value for the city and name column by using the conditional split node.  
Diagram, timeline

Description automatically generatedGraphical user interface, text, application

Description automatically generated  
  
Then, for the destination Data Warehouse, we set up a table with the following structure  
A picture containing text

Description automatically generated  
Graphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generated  
  
  
  
Category Table  
  
The similar process can also be seen for the category table.  
Chart, box and whisker chart

Description automatically generated  
Graphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generated  
  
  
Text

Description automatically generated with low confidenceGraphical user interface, application

Description automatically generated  
  
  
A picture containing diagram

Description automatically generated  
Graphical user interface, application

Description automatically generatedGraphical user interface, application, Word

Description automatically generated  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
User Table  
  
A picture containing diagram

Description automatically generated  
Graphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generated  
  
  
Timeline

Description automatically generatedGraphical user interface, application

Description automatically generated  
  
  
Chart, box and whisker chart

Description automatically generated  
Graphical user interface, application

Description automatically generatedGraphical user interface, application, Word

Description automatically generated  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
Review Table  
A picture containing graphical user interface

Description automatically generated  
Graphical user interface, application

Description automatically generatedGraphical user interface

Description automatically generated  
  
Diagram, timeline

Description automatically generatedGraphical user interface, application

Description automatically generated  
  
  
  
  
  
  
Diagram

Description automatically generated  
Graphical user interface, application

Description automatically generatedGraphical user interface

Description automatically generated  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
Weather Data  
  
As for the weather data, since it is in the .csv format, the way we extract the data from will be slightly different.  
Text

Description automatically generated with medium confidence  
Graphical user interface, application

Description automatically generatedGraphical user interface, table

Description automatically generated  
  
Then, just like before, we store it in our localhost Yelp data warehouse  
A picture containing text

Description automatically generated  
Graphical user interface, application

Description automatically generatedGraphical user interface, table

Description automatically generated  
  
  
  
Graphical user interface

Description automatically generated