|  |
| --- |
| **Data Acquisition**  **Week 4 Assignment** |
|  |

**Task**

**Graphical user interface, text, application

Description automatically generated**

For this week four’s assignment, I have made several changes from last week’s ETL workflow. That will be explained along the way as I explained the main topic for the week 4 assignment, which are the finance data in MongoDB and Teleport dataset.

First of all, I will show you the control flow of the entire ETL process.

Graphical user interface, application

Description automatically generated

In this control flow, the ETL process is divided into 4 parts. The first data flow will do ETL job for the dimension tables, like weather and category. The second one will do ETL job for finance business and tip data, whereas the third and fourth data flow will do ETL job for review and user related data respectively.   
  
1. First Data Flow

Graphical user interface

Description automatically generated  
In this data flow, we extracted the category table from yelp db and store it in our datawarehouse. We also extracted the weather data from the csv file and store it in our DWH.

2. Second Data Flow

Graphical user interface

Description automatically generated

This data flows will cover the most of the question for this week’s assignment. First of all, we setup a connection with the mongoDB.

Graphical user interface, text, application

Description automatically generated

This can be done by using the ZS MongoDB source connector. Then the finance data from the MongoDB will be store in 2 outputs, the finance table and in the business table (storing the FK). The next thing that we do is creating a BI question so that we can use the dataset in Teleport and store it in our datawarehouse. The question is “What city has the largest business balance?”. Based on that question, we decided to get the /cities GET endpoint to get all the cities data.

Text

Description automatically generated Graphical user interface, application

Description automatically generated

3. Third Data Flow

Graphical user interface

Description automatically generated

This data flow is just used to store the review table data into our data warehouse.

4. Fourth Data Flow

Graphical user interface, text, application, Word

Description automatically generated

This data flow is just used to store the review table data into our data warehouse.

The .sln file can be seen in the following GitHub link: