**ETL DataWarehouse Project A picture containing icon

Description automatically generated**

Project made by : **Djime Wague** Professor : **Emerick Duval**

**Ali Semlali**

**Samuel Bunicic**

**Taha Nacer**

**Project Summary :**

* Before starting the project, we decided to see all the data sources to see the diagram on which we will base ourselves in order to do the project. The diagram we have chosen is inspired by what we have learned in the course. It has been designed so that the people who will analyze the data can have the maximum possibility of visualization and analysis. We also integrated a type 1 SCD into the five Dimensions because we concluded that the history of changes and archiving is not important for the data we are processing in order not to take up more space which is not going to serve us.

****

DWH Schema :

**DimDate**

**DimCallCharges**

**DimCalltypes**

**DimEmployee**

**DimUSSTATES**

**FactALLDATA**

DWH Schema in SSIS :

**A screenshot of a computer

Description automatically generated with medium confidence**

* The approach with which we worked to carry out this project is mainly based on the ETL course we had in class. We divided the work on three different types of Packages: **STA - ODS - DWH**. We also did the **Package Scheduler** as taught in class.
* The choice of the diagram of the Fact Table is inspired by what we have seen in class. We decided to have a single Fact Table linked to 5 dimensions. For the dimensions, we decided to choose **SCD (Slow Changing Dimension) TYPE 1.**

A screenshot of a computer

Description automatically generated

* For the STA Packages, we imported the data from the different data sources we have for the project.
* We had to change Project Properties : in particular the change from 64 to 32 bits to be able to import the Excel file. Sometimes, we had to open the task manager to finish the tasks in progress (Debug) to be able to lunch the Debugging like what we did in the course.

For ODS packages, we decided to proceed as follows:

* For the DATA\_2018, DATA\_2018, DATA\_2019, DATA\_2020, DATA\_2021 CSV files, we made a UNION of these tables in the ALLDATA table, then we split the CallTimeStamp column using a derived column. Next, we converted the Date data type. Then we added columns DAY, MONTH, YEAR. Afterwards, we did a lookup process to add a StateCD column to the ALLDATA table from the Employee table. Finally, we checked the data types to inject errors into a Technical Reject table as demonstrated in the course, before doing the resizing and integrating it into the ODS database.

A screenshot of a computer

Description automatically generated with medium confidence

* For the CallCharges table, we did the following process: Unpivot to have a more consistent table. Afterwards, we used a derived column to remove the NULL values that exist in the table, then we checked for errors and integrated them into the Technical Rejects database before integrating the table into the ODS database.

A screenshot of a computer

Description automatically generated with medium confidence

* For the Employee table, we decided to split the Site column to have the name of the state separately and its code separately. Then after we checked the data types and integrated them into the database.
* To be able to facilitate the task of executing the packages, we decided, as we saw in the course, to integrate a Scheduler package which will organize and sequence all the packages of the project.

**A screenshot of a computer

Description automatically generated with medium confidence**