**ETL & Data Warehouse Project**

ServiceSpot, an IT company, contacted you in order to help them analyze their call center data.

They receive daily calls from their customers and would like to know how things are going.  
The data is split across multiple files, and they are unable to make good use of it.

Your mission is to develop an ETL project with SSIS that will load data into the new enterprise data warehouse.

As a data engineer team, you will only focus on providing knowledgeable data for the business.  
The reports will then be done by another team of data analysts, who will use your data warehouse.

**Goals of the project :**

* Retrieve data from the many sources
* Design a proper data warehouse
* Apply necessary transformations with SSIS
* Load the data into the data warehouse

**Delivery :**

* Your whole solution
* SQL Scripts to recreate every table
* A Word document explaining your project, including the schema of your data warehouse, and the main decisions you took in order to achieve this project

**Remarks :**

* As of today, you are provided with 4 years of calls data only, but new files (for different years) could be coming in the future, so you have to make sure that your solution will be able to consider the future files well
  + For that purpose, you will have to use the “For Each Loop” container in SSIS, in order to loop over the “Calls data” folder
  + This is one tutorial (amongst many others on the web) that explains how to loop over a folder :
    - <https://www.sentryone.com/blog/how-to-loop-through-files-in-a-specified-folder-load-one-by-one-and-move-to-archive-folder-using-ssis>
  + Please note that in this tutorial, they show you how to store the name of the file, which is not necessary for your project
* The “Call charges” file is provided in a “easy-to-read” format for a human, but not very efficient to deal with for a database. You will need to **unpivot** the data coming from that file, using the “unpivot” transformation in SSIS.
* This is an example that shows quantity of products sold by type and by country :

Une image contenant table

Description générée automatiquement

* This is what “**unpivoting**” the data would be :

Une image contenant table

Description générée automatiquement

* It’s a best practice to separate the “time” data (hour minute second) from the “date” data
* SLA (Service Level Agreement) defines an expected level of service, which can be measured through specific metrics.   
  In this company, it is expected that a call must be answered **before 35 seconds** of waiting time to comply with the SLA. Thus, any call must be considered “**Within SLA**” if the waiting time is below 35 seconds, otherwise it should be “**Outside SLA**”.
  + Make sure that this information is known for each call
* Follow every advice you were taught during the class
* The data warehouse must have only one fact table