**Projet De Business Intelligence**

Dans la continuité de nos études dans le domaine de l’informatique nous avons participé au cours de BI (Business Intelligence).

Lors de ce cours il nous a été demandé de réaliser le projet suivant :  
Aider une société informatique à Analyser les données de leur centre d’appels, recevant quotidiennement des appels de leur clients et qui souhaiterai connaitre le déroulement des appels.

Leurs données étant réparties sur plusieurs fichiers ils ne peuvent pas faire bonne usage de ces informations précieuses.

Il nous a donc été confié le travail de développer un projet ETL ainsi que SSIS qui chargera des données dans leur nouveau data warehouse.  
  
le but de notre équipe étant seulement de fournir les données correctement préparées pour l’entreprise, qui les fournira à une autre équipe de data analysts qui seront capable de les utiliser.

Les données à étudier son étalées sur une durée de 3 ans (2018/2019/2020) cependant de nouveaux fichiers pourraient être ajoutés à l’avenir, nous allons donc développer une solution qui pourra s’adapter à ce besoin.

Nous avons commencé par intégrer les données existantes dans une base de données STA qui a pour but de stocker tout sous un même format, qui peut par la suite être formatée dans une base de données ODS. Depuis celle-ci (notre ODS) nous pourrons préparer les données pour les mettre dans notre Data Warehouse (DWH) sous forme de table de Dimension (Table DIM) et table de fait (Table FACT).

Dans ODS Employees nous avons ajouté depuis "UsState" les tables "Name" + "Region" en utilisant les informations similaires de "UsState" (StateCD) et de "Employees" (Site) pour faire le lien entre les tables.

Puis, dans la colonne "Site" d’ "Employees" on a récupéré les informations avant la virgule pour créer la table "City" enfin une fois cela fait nous n’avons plus besoin de la colonne "Site" on peut donc la supprimer de notre ODS.

Dans notre ODS "Calls" nous avons, comme pour "Employees" qui contient les informations de la table STA "Employees" et "UsState", besoin de récupérer les tables des STA "Calls 2018-2021", "Call Types", "Call Charges".

On doit donc commencer par utiliser le conteneur "For Each Loop" dans SSIS, afin de boucler sur le dossier "Calls data" et récupéré toutes les données.

Puis au sein de "Calls" qui contient désormais les informations des "Calls 2018-2021" on va ajouter depuis "Call Types" la table "CallTypeLabel" en faisant le lien entre le "Call Type" et le "CallTypeID".

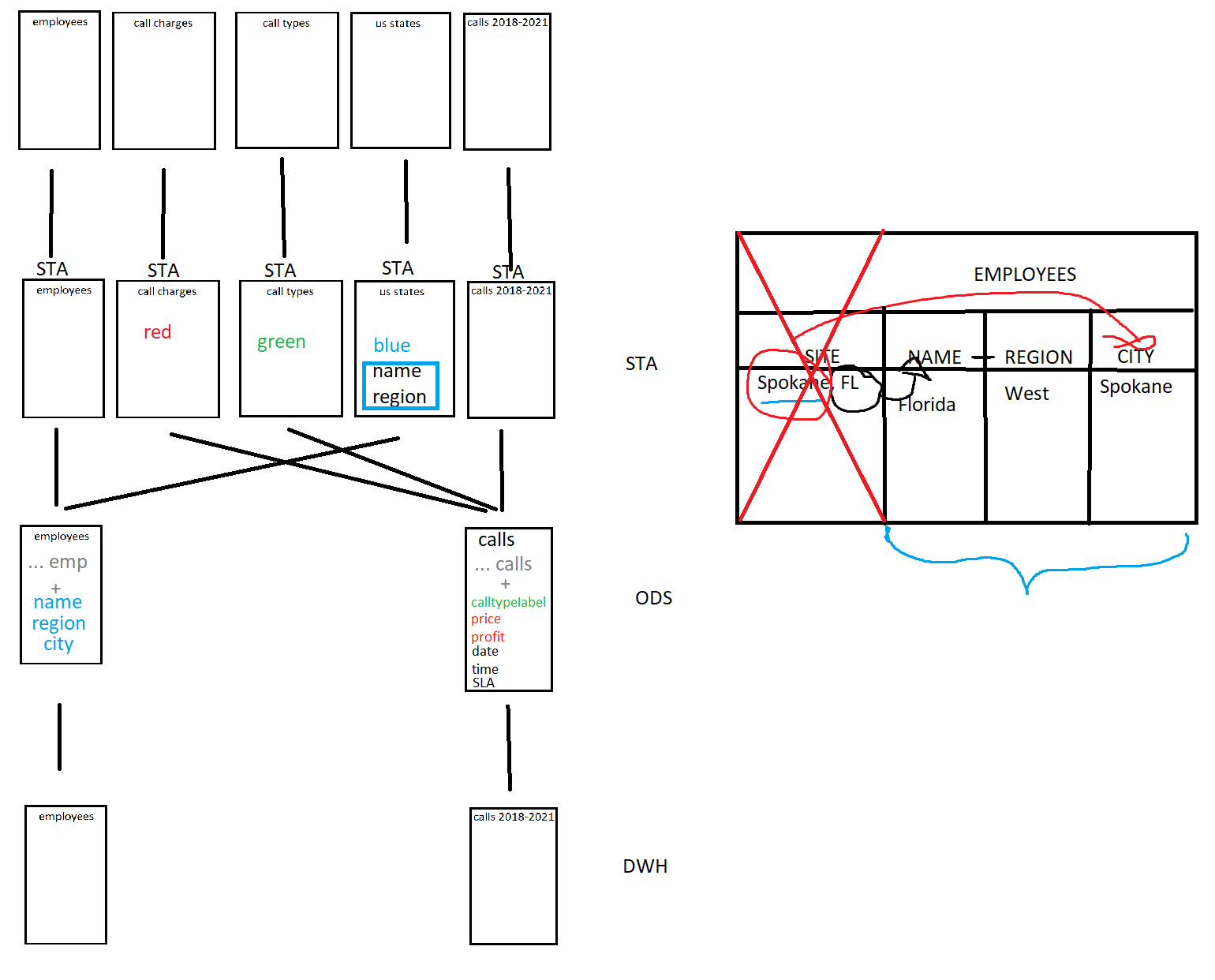
Nous devons modifier le fichier "Call charges" car il était en “easy-to-read” un format lisible par l’être humain qui présente des inconvénients pour la gestion des données, donc dans le SSIS nous l’avons "unpivot", ce qui permet de mettre des lignes en colonnes ou colonnes en lignes.

Grâce au "unpivot" pour modifier les colonnes "Call Charges" on a rendu utilisables dans notre ODS "Calls", les tables de "Call Charges". Etant maintenant utilisables nous récupérons alors les informations de prix pour qu’avec un traitement on puisse mettre cette information dans une colonne "Price" que l’on crée.

Ensuite avec les informations de "Price" et de "CallDuration", on effectue un calcul pour transformer le "CallDuration" de secondes en minutes puis on multiplie le "Price" par la valeur de l’appel en minutes.

Ensuite comme demandé dans le projet nous avons traité les données pour les séparer de la manière suivante : les valeurs de temps "Time" heures:minutes:secondes des données de date "Date".

Enfin comme spécifié dans la documentation, si un appel dure plus de 35 secondes dans "WaitTime" alors on doit spécifier cette information dans une nouvelle colonne nommée "SLA" avec la valeur "Within SLA" sinon, on lui donne la valeur "Outside SLA".



Nous avons choisi de mettre en place le DWH de la manière précédente (image) : en utilisant la table de Fait "FactCalls" comme table de référence alimentée par les tables de Dimension "DimDate" et "DimEmployee".

Dans cette topologie en étoile, nous avons "DimEmployee" qui est la seule Dimension avec "DimDate", lors de la création de "DimEmployee" nous avons décidé de la mettre en place en SCD 1 (la plus courante car elle permet de mettre à jour les données).

SCD 1 fonction de la manière suivante : la table ODS associée à la DIM va être vérifiée afin de savoir si oui ou non, des valeurs des colonnes de la table sont différentes des valeurs actuelles de la DIM, si les colonnes ne sont pas présentes alors la vérification envoie les données dans la DIM, sinon on les revoie vers une nouvelle vérification qui va déterminer les colonnes qui ont changé pour effectuer un UPDATE et modifier les colonnes concernées.

Effectivement, comme le but de l’entreprise qui nous a contacté est de les aider à étudier les données de leur centre d’appels, il paraissait donc logique d’utiliser une table contenant des données en adéquation avec la demande, c’est-à-dire les appels de l’entreprise, d’où le choix de la table Calls pour en faire une table de faits "FactCalls".

Par conséquent les tables restantes deviendraient les tables de Dimensions ayant pour but de contenir les informations complémentaires.