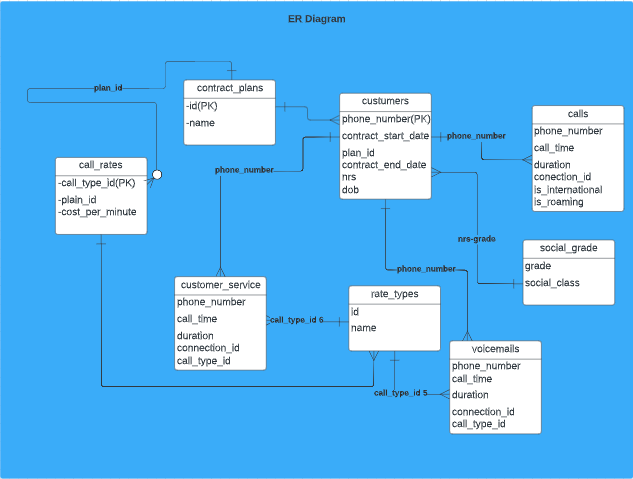
# Customers and Plans Data Model and Fact Tables

### **Objective: Create a Model that can answer the following questions** A. Identify how valuable a customer is to the company relative to other customers

B. Build up a picture of their customers’ profiles  
C. Determine whether a customer’s behaviour patterns have changed recently  
D. Identify the call plans which bring in the most revenue

### **The raw data:**

Down below we have a simplified representation of the raw data imported from our csv files.  
  


# 

# In our raw data, we can see that customers have 3 types of what we can call “connections”, that involves calls, voicemails and customer\_service.

# Rate types for voicemails and customer\_service are always the same whereas it varies for calls and needs to be inferred given if it’s international, roaming and what time of the day it’s.

# We can see that we have information for our customers regarding their social\_grade, age and plans that they signed for but we don’t have neither address nor sex which might be missed for our analysis further on.

# Our customers are uniquely identified by phone\_number so, we should use this more as an unique account instead of a unique person, in case of a person having more than one phone number registered with us.

### **Star schema 1 - Explain Everything with detail but low performance**

# 

### The first schema proposed was the one above, it could answer all the questions and be future-proof as we could even bring later on new definitions of what is a change in customer behaviour. The reason we decided to not go with the schema is that, although it would be relatively simple to build, its performance wouldn’t be great taking into account only the questions that we have at hand right now for the business. And we can always build a new star schema if necessary. As we will not use this approach right now, I’ll not further explain how we could leverage this schema. **Star Schema 2 - Answer the questions**

For the Star Schema 2 in the image above, although it’s less extensible than 1, is capable of meeting all current business requirements with a better performance than 1 since, for the customer-related questions, it doesn’t need to iterate over many rows and for plan related questions it only has to sum the customer totals.

The customers\_facts table will represent each customer (or account better saying) in the system. It will have aggregated values for time and money spent on calls over the time and for the last month, it also has information about user profile.

Below you can see how we are going to answer each one of the business questions:

**A. Identify how valuable a customer is to the company relative to other customers**

For that we can just look at the total\_spend field on the customers\_facts table for each customer. We can also check if this customer contract has ended to see if he is still with us.

# **B. Build up a picture of their customers’ profiles**

# We can look at the customer contract\_plans, his age, social grade, average call time, spent per call and percentages of peak, international, roaming, voicemails and customer\_service calls.

# **C. Determine whether a customer’s behaviour patterns have changed recently** We can look to see if there is a difference between the average and the average from the last month in these customer calls, or in the percentage of type of calls vs. the percentage of type of calls for the last month. Since it’s for last month, we will have numbers only for customers that had activity in the last month (calls, customer\_service or voicemails), if the customer doesn’t have any activity for the last month, those statistics will not show his behaviour in the last month that he was our customer. This can be covered in the future though.

# **D. Identify the call plans which bring in the most revenue**

# For the raw revenue figures, we can simply sum the total\_spend of all the customers for a given contract plan. As we don’t have here figures of operational costs, the best we could do regarding estimating net revenue, is to calculate how many cents-per-minute we make on each plan. We could do that by getting the total spend divided by the total call time and summing all those entries for a given plan. As future work, We could categorise this by call type as well like international, roaming, peak, etc, given that those calls probably have different operational costs.