# Takehome - Data Engineering related Role

We offer this takehome assignment for a few reasons:

1. We would like to get a sense of your data processing abilities as well as your approach to solving a problem.

2. We would like you to get a sense for something that a data engineer might be expected to do.

3. It is an opportunity for you to sharpen your data processing skills and perhaps learn something new!

You should be able to complete this takehome within a few hours (more if you are also getting up to speed with Spark and/or Scala).

You should use Scala and Spark RDD (not Data Frames or SQL) to complete the task as this is what we use for data engineering.

Spark is open source and can be downloaded from: http://spark.apache.org/downloads.html

You can browse Spark documentation here: http://spark.apache.org/docs/latest/

You can install Scala directly: http://www.scala-lang.org/download/install.html and use Spark REPL or you can use an IDE such as IntelliJ:

https://www.jetbrains.com/idea/

## The Assignment

The primary purpose of the assignment is to produce produce a simple data process from user files.

The main goal is to define a new phone marketing campaign for top 1000 users in order to upsell them with a new compelling product.

### The Data

The data is fictitious and has been generated by a program but is representative of something that could have to be done in real life.

We have created the following files:

\* The `transactions.txt` file contains information about transactions.

The structure of the file is the following:

Customer ID, transaction amount, transaction date

\* The `users.txt` file contains information about users.

The structure of the files is the following:

Customer ID, Customer full name, email list, phone list

\* The `donotcall.txt` file contains phone numbers that should not be used in the campaign

### The Analysis

The selection of the top 1000 users should include the following steps:

. Parsing of the files

. We wish to contact users using phone numbers that are not in the do not call list.

As this is a phone campaign, there should be at least one phone number per user in the output.

. Selecting the top users based on highest transaction amount for year 2015

. Save the campaign in a file using the following structure:

Customer ID, Customer name, phone list that can be used to contact the user, total transaction amount

## Submission of Work

Please include your Scala file used in solving the problem along with the solution campaign file.

Feel free to explain some of the design decisions you made and some of the problems you may have encountered.

Have a good work!