**Problem Statement**

To build a classification methodology to determine whether a person defaults the credit card payment for the next month.

**Architecture**



**Data Description**

The client will send data in multiple sets of files in batches at a given location. The data has been extracted from the census bureau.

The data contains 32561 instances with the following attributes:

**Features:**

1. **LIMIT\_BAL**: continuous.Credit Limit of the person.
2. **SEX**: Categorical: 1 = male; 2 = female
3. **EDUCATION**: Categorical: 1 = graduate school; 2 = university; 3 = high school; 4 = others
4. **MARRIAGE**: 1 = married; 2 = single; 3 = others
5. **AGE**-num: continuous.
6. **PAY\_0 to PAY\_6**: History of past payment. We tracked the past monthly payment records (from April to September, 2005)
7. **BILL\_AMT1 to BILL\_AMT6**: Amount of bill statements.
8. **PAY\_AMT1 to PAY\_AMT6**: Amount of previous payments.

**Target Label:**

Whether a person shall default in the credit card payment or not.

1. default payment next month: Yes = 1, No = 0.

Apart from training files, we also require a "schema" file from the client, which contains all the relevant information about the training files such as:

Name of the files, Length of Date value in FileName, Length of Time value in FileName, Number of Columns, Name of the Columns, and their datatype.