**TECHNICAL DOCUMENT SUMMARY**

**CUSTOMER DATA PROCESSOR**

**1.0.0**

|  |  |
| --- | --- |
| Application Name | **Customer Data Processor** |
| App Version | **1.0.0** |
| Author | **MD. RAFIQUL ISLAM** |
| Document Type | **Technical Document Summary** |

**Overview**

Any business has to deal with data. The volume of data is increasing day by day along with the demand of the data. It is a very big challenge to process large volume of data and it is very expensive in terms of technology as well as money obviously. This is a test project to deal with large volume of data. The main objective is to process data with efficiency and maximum possible speed. Some data processing functions like loading large data, process data with filtering and validation, data storing, data exporting, supply etc. will be tried here to integrate.

**How IT Works**

The App,

* Loads customer data from text file. The text file is comma separated.
* Organizes all the data to get tabular format.
* Filters & removes duplicate records considering all fields of each record.
* Validates all the records considering phone number and email address.
* Uses problem statement provided phone number pattern i.e.:  
  555-555-5555, (555)555-5555, (555) 555-5555, 555 555 5555, 5555555555, 1 555 555 5555   
  to validate phone number.
* Uses combination of RFC 5322 Internet Message Standard and most popular email application like Gmail, Live and Outlook practiced email address pattern to validate the email address.
* Separate invalid data consider phone number and email address.
* Store valid and invalid data into 2 separate tables in a database.
* Export the valid and invalid data into 2 separate text files.
* Export the valid & invalid data into a batch of files including 10k records.
* Uses multiple threads to insert data into database and to export data into files.

**Notes**

* Most of the properties are configurable.
* Support multithreading to run faster.

**Configurable Properties**

* Database Connection string.
* Database table name for valid and invalid data.
* Data column list.
* Column index to be validated.
* Path of input & output files.
* Maximum records for each file of batch.
* Number of maximum threads to insert data.
* Thread counts measurement factor to export data.

**How to Configure before Run**

Open

**Configurable Properties & Settings Values**

|  |  |
| --- | --- |
|  |  |
|  |  |