**Tidy ETL Cloud Application**

* **Project Description:** Tidy is a cloud-based application. The motivation behind the project is to understand the ETL Process. For bulk data import/export for various sections like Customers, Suppliers, Materials, Material Prices, and Users. Tidy Customers download the empty Template CSV file and upload their data in rows ordered by columns and headers. Sometimes customers upload their data with different file formats and data types, therefore receiving errors. The Tidy Support team helps the customer resolve these errors but it is a tedious and time-consuming (and therefore costly) process. It is also very frustrating for the customer and the risks that they will leave Tidy. **The objective of this project is to solve this problem in a repeatable, reliable, cloud-application-like method.**
* **Project Challenges:** There is an issue with data transparency meaning that support (or any automated AI-like process, cannot “see” the data the customer is trying to upload. It is a very user-intensive, “awkward”, the process for Tidy see the data the customer is trying to upload. It normally involves them e-mailing the file over to support. This becomes very time-consuming and costly. Currently, it is too much of a manual approach to sort out the errors. Also, how many customers simply leave in frustration because they are not willing to spend the time to work with Tidy support?
* **Project Objective:** To design and prototype a new, revised version of the Data Import/Export Section. The data can be in any file format (only the CSV file). For this, design and prototype a new ETL Application which collects data from different file sources and moves the data from the Customer’s site to Tidy’s cloud environment. ETL (Extract, Transform, and Load) is one of the essential techniques in the data processing. There is so much data currently in on-premise applications that needs to be moved to the cloud. Solving this problem will have massive value inefficiency and error reduction for Tidy and its customers by first uploading the customer’s data in its “raw form” directly into the cloud and then work on that data. This means that the data probably needs to be “staged” as it is worked on in a process similar to that of recycling waste products. The data, like recycling waste, need to be collected, then progressively sorted and cleaned so it can be reused for a different purpose.
* **Requirements:**

1. I want to import data from different data sources such as CSV, Excel Workbook, Pdf and Doc file. But I can only import data in the CSV template provided on software.(Already shared the “Suppliers csv template”)
2. I want to select which fields in my raw data map (shared file) to fields in the Tidy Software(provided csv template). (I want the import tool to extrapolate data from the spreadsheets which are not in a table format).
3. The ETL portal should show the uploaded file (even if failed) and the list of errors in which a system administrator can download the file, make changes according to the list of errors and re-upload the file.

I want to populate that ETL solution for other bulk import/export data tabs such as:

1. Assemblies
2. Departments
3. Activities
4. Project/Quote Time Plan
5. Project/Quote Material Plan

The main challenges for these are, they require data in some fields that are existing in Tidy, so this data would need to be made available as a selection in the tool. For example, to import materials into a material plan, only materials that are currently in Tidy can be selected.

Notes:

I want to design and implement prototype in the Microsoft .net framework.