

ETL with delimited files;

A delimited file is a text file used to store data, in which each line represents a row or a record, and each row has fields that are separated by a delimiter.

A delimiter is a sequence of one or more characters that is used as a separator for the fields in a row of a delimited file. The following characters are common delimiters:

- Comma (,)
- Semicolon (;)
- Tab
- Space

Types of delimited files;

There are many types of delimited files, but the following are used very frequently.

- CSV file: CSV stands for Comma-Separated Values. It is a delimited file in which a comma (,) is used as a field delimiter.
- TSV file: TSV stands for Tab-Separated Values. It is a delimited file in which the TAB character is used as a field delimiter.

The two most common file extensions for delimited files are .csv and .tsv.

CSV Example

The content of a CSV file *users.csv* is given below:

id,name,age
1,John,35
2,San,30
3,Mary,31

The first line in the file is the header that contains the name of the fields.

There are four rows, including the header, in this CSV file. Each row has three fields that are separated by the delimiter comma (,).

TSV Example

The content of a TSV file *users.tsv* is as below:

id	name	age
1	John	35
2	San	30
3	Mary	31

There are four rows in this TSV file, including the header. Each row has three fields that are separated by the delimiter TAB. The first line in the file is the header that contains the name of the fields.

What is a text qualifier?

A text qualifier is a character used to specify the point at which the contents of a text field should begin and end. Consider that you need to import or read a comma-delimited file, but one of the fields

contains a comma. In such a scenario, a text qualifier can be used to show that the comma should be included within the field and not be used as a delimiter.

The following is the content of a CSV file that contains the id, name, age, and skills. A user can have one or more skills that are again separated by a comma.

```
id,name,age,skills
1,John,35,Java
2,San,30,Java,Python
3,Mary,31,Python
```

The above file will be parsed out like this:

Table 2-16

id	name	age	skills	
1	John	35	Java	
2	San	30	Java	Python
3	Mary	31	Python	

After parsing the above CSV file using a comma delimiter, the skills of the user having id 2 are separated into two different fields.

To avoid this, we can use a text qualifier such as double quotes for skills as below.

```
id,name,age,skills
1,John,35,"Java"
2,San,30,"Java,Python"
3,Mary,31,"Python"
```

Now, the above file will be parsed out like this:

Table 2-17

id	name	age	skills
1	John	35	Java
2	San	30	Java, Python
3	Mary	31	Python

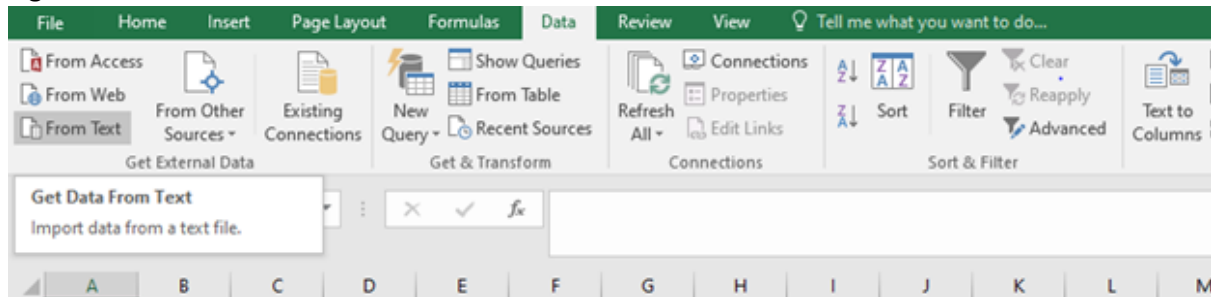
Importing Delimited Files to Excel

The following are the steps to import a delimited file (for example, a CSV file) to Excel.

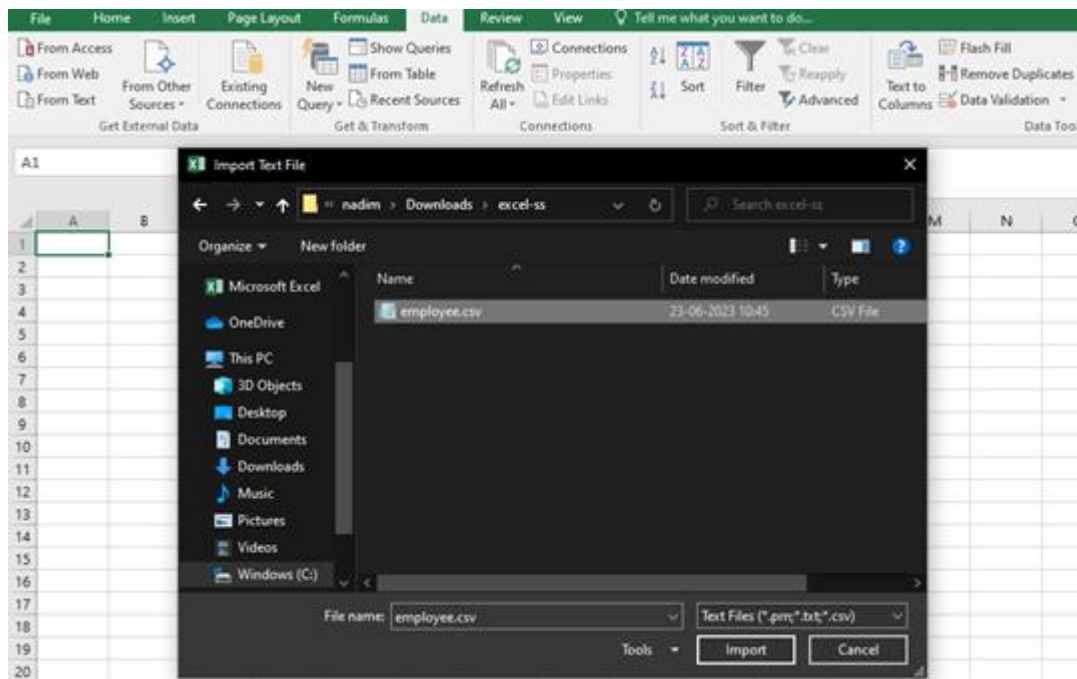
1. Open Excel and click on Data -> From Text as below.

Figure

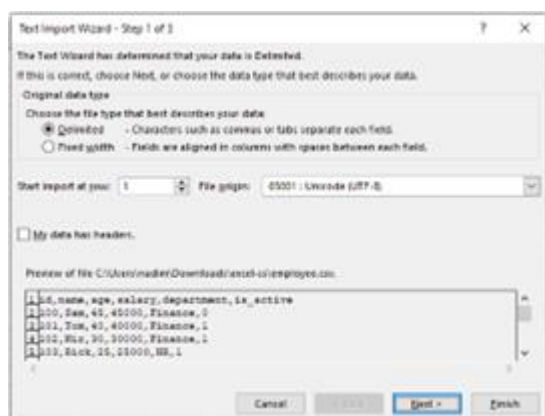
2-2



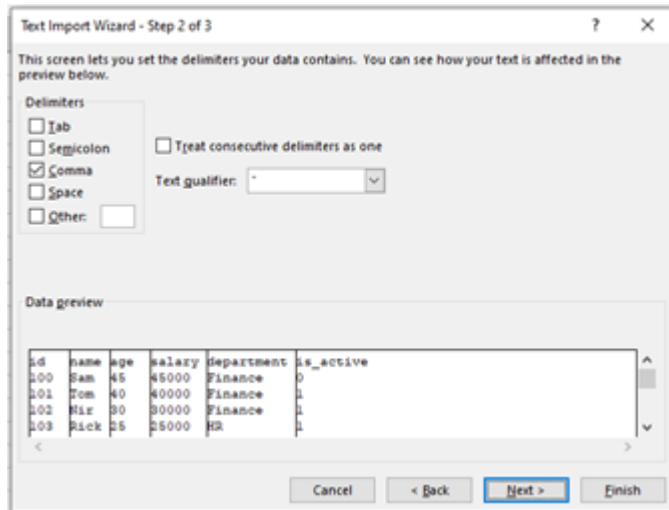
2. A dialog box will be opened to allow you to select the file. Select your delimited file and click on Import.



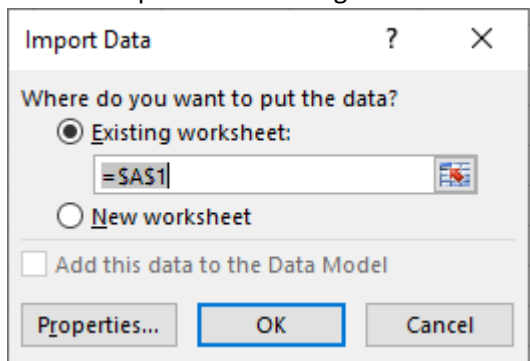
3. Excel will open a preview of the data in the selected file.



Choose the delimiter (for example, comma for a CSV file) and click on **Finish**. **Figure**



A new Import Data dialog box will be opened. Choose either Existing worksheet or New



File importation to Excel Video