Achieving Zen ETL for the absolute beginner with expressor Studio – A quickstart tutorial – Part 1

expressor Studio

By

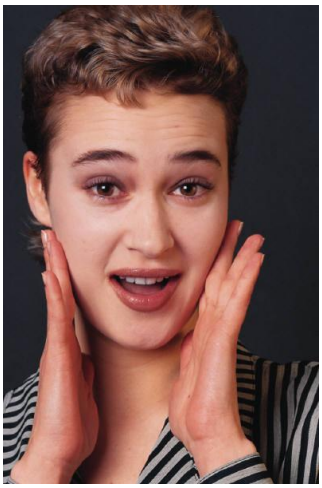
Anil Mahadev

Solutions Architect (Database and Virtualization) Technologies

expressor Studio User

expressor Studio – A quickstart tutorial on getting started with ETL for absolute beginners

For most of us ETL (Extract, Transform and Load) has only been to the realm of Database Architects, DBAs, BI Architects and BI developers. Well not anymore, ETL has gone mainstream and Power Users at companies can now experience the power of ETL, right from their desktops and build ETL applications without much complexity.
Are you serious?
How is it possible you may ask?



Now how do we achieve it?

It is now possible thanks to expressor Studio from expressor software.

In this quick start tutorial, we shall go through the process of downloading expressor Studio Desktop Edition and building a simple ETL package that will read data from a relational database (SQL Server 2008) in this case and push the data to a text file.

How to get expressor Studio Desktop Edition?

You just to point to the following URL in your favorite web browser and download it to your computer.

http://go.expressor-software.com/desktop-edition-etl-tool.html

Apart from expressor Studio, you may want to download SQL Server 2008 R2 Express with Advanced Services from the following website.

SQL Server 2008 R2 32 bit with Tools
https://www.microsoft.com/betaexperience/pd/SQLEXPDBMT32/enus/

SQL Server 2008 R2 64 bit with Tools

https://www.microsoft.com/betaexperience/pd/SQLEXPDBMT64/enus/

Sample Databases (Northwind and pubs)

http://www.microsoft.com/download/en/details.aspx?id=23654

[For SQL Server 2008 R2 upon installation ensure that you have enabled TCP/IP and Named Pipes for connectivity]

Install the sample databases of the Northwind and pubs. This tutorial does not cover the steps in doing so. It presumes that you have already have downloaded and setup the above with enough permissions to create a database and tables with Insert/update permissions.

The installation is a seamless experience and is very fast.

Start your ETL Engines!

Time to build an ETL Project with expressor Studio

Case Study: We need to transport data from the Northwind Products table to a text file.

1) Start expressor Studio Desktop Edition by clicking on Start → All Programs → expressor → expressorStudio as shown in figure 1.1 below

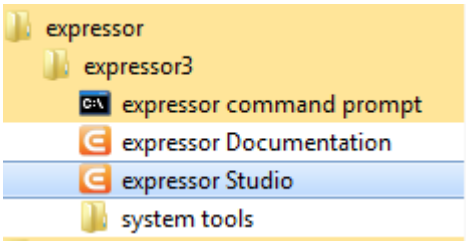


Figure 1.1 – Start Menu to access expressor Studio

2) Once launched, you will be presented with a Project Workspace Wizard as shown in Figure 1.2

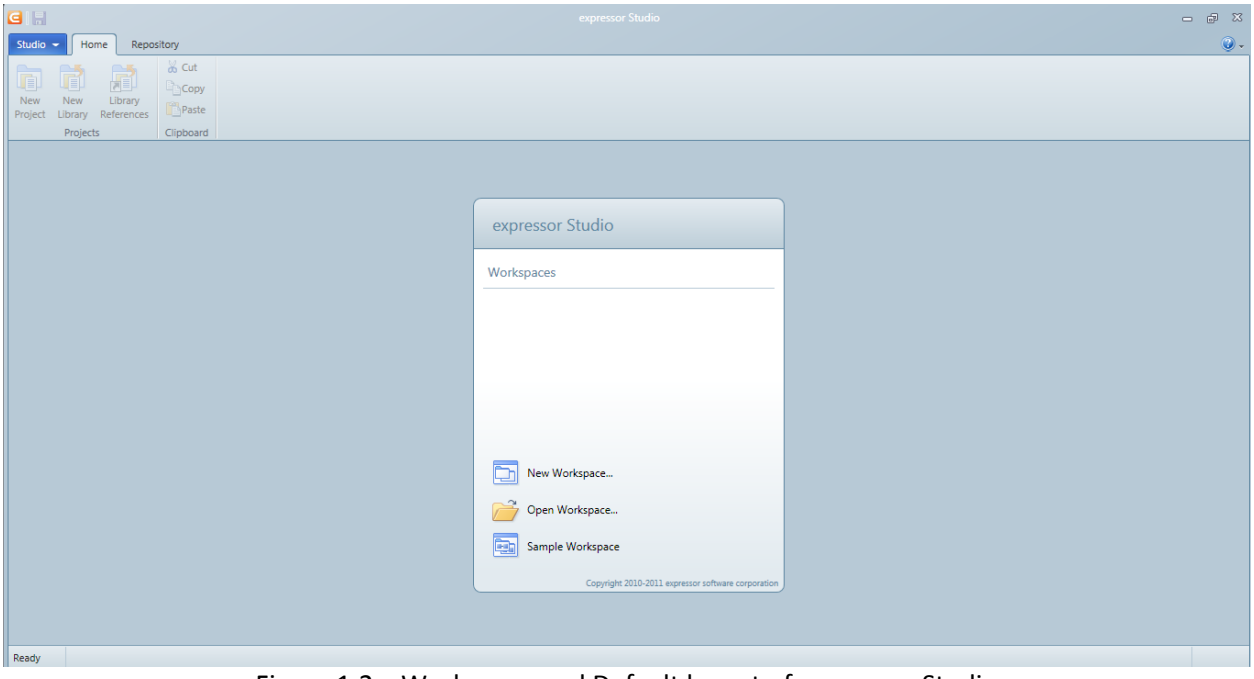


Figure 1.2 – Workspace and Default layout of expressor Studio

3) Click on the New Workspace button to create a new workspace

4) In the new Workspace Dialog box, you may enter
   a) Workspace Name
   b) The location of where the workspace would be stored and created
   c) Description, to indicate the type of workspace

This is shown in Figure 1.3 below

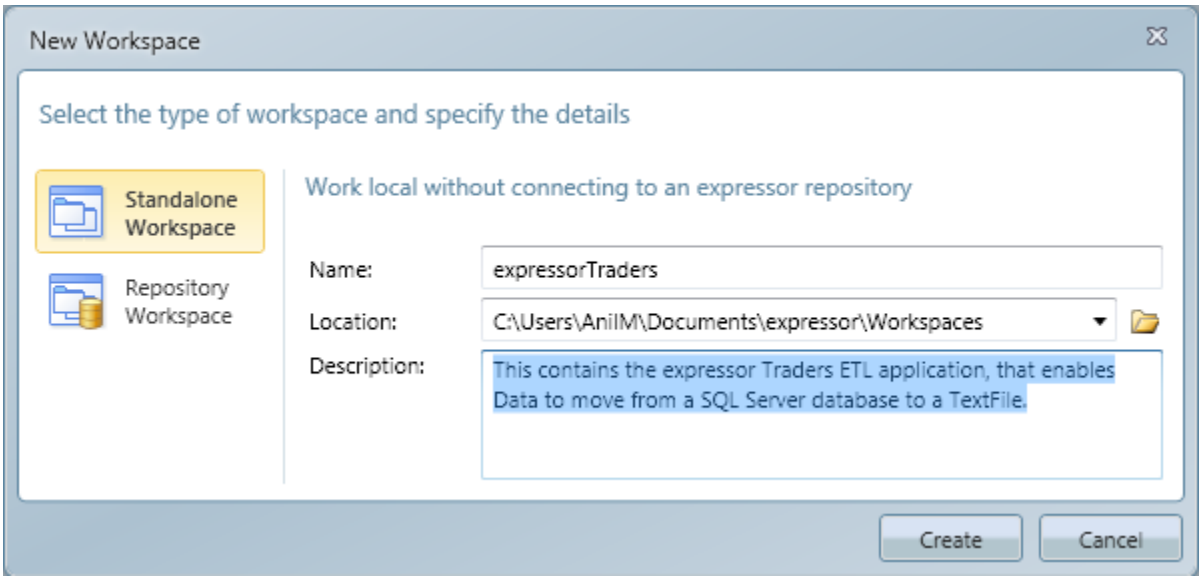


Figure 1.3 – New Workspace Dialog

5) Click on the Create button to create a new Workspace
6) Now that we have a Workspace, it is time to create a new Project
7) To the right of the Workspace window, click on the New Project button as shown below in Figure 1.4

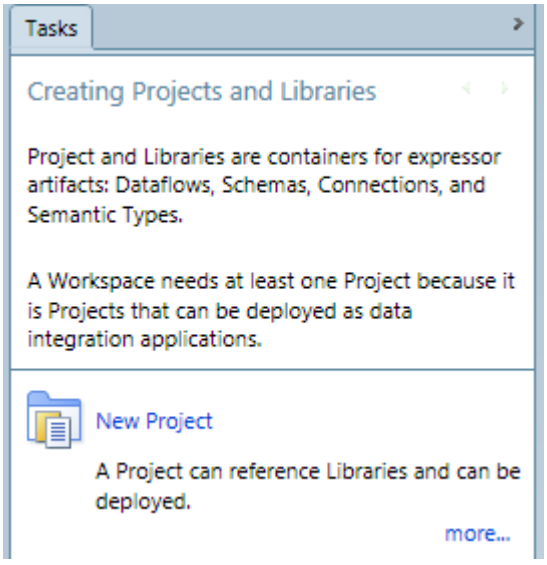


Figure 1.4 – New Project

8) Give the project a name of your choice and a description; click on the Create button to complete the process

9) Once the new project has been created you will be presented with this amazing Interactive Overview Dialog, that clearly helps new users of the ETL world to understand, the basic workflow of an ETL application, as shown in Figure 1.5 below

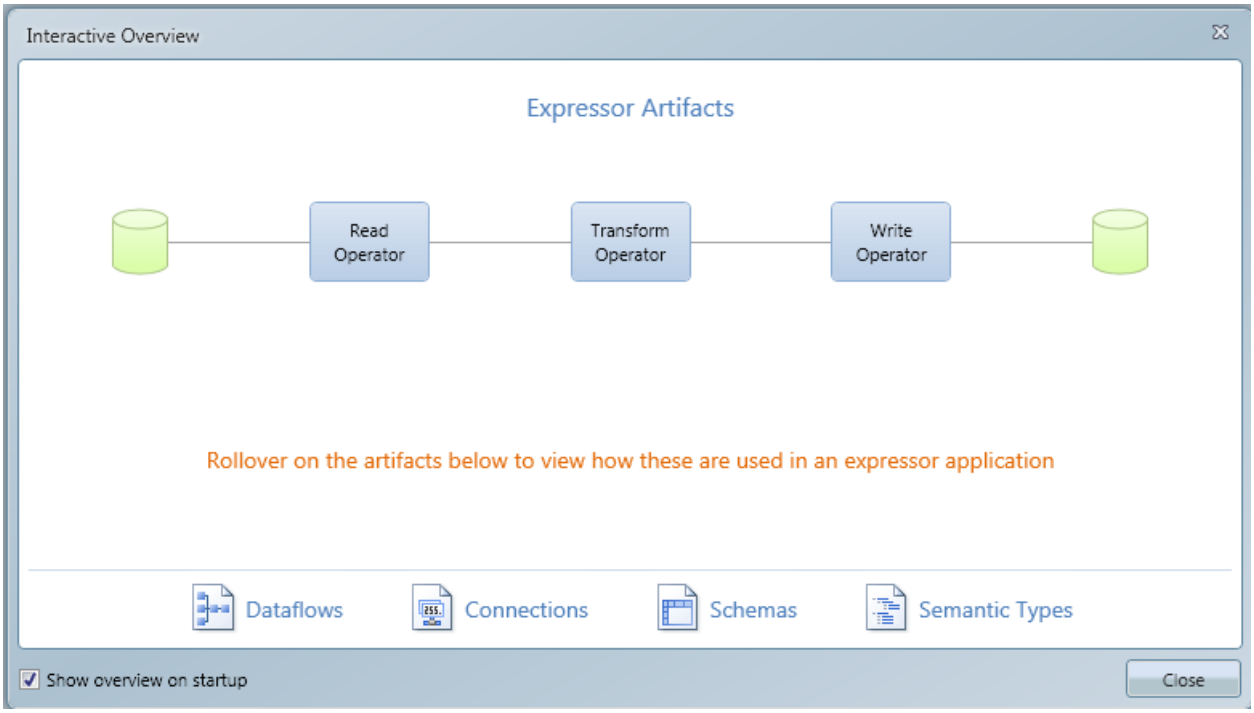


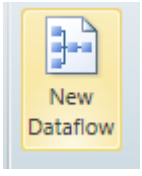
Figure 1.5 – Interactive Overview Dialog

10) Click on the Close button and we shall explore the components needed to build this ETL application

Components of an expressor Studio ETL Application

expressor Studio comprises of Dataflows, Connections, Schemas and Semantic Types. For part I of this tutorial, we shall be focusing on the first two types only.

The User Interface of expressorStudio unlike traditional ETL tools gives end users the rich Office 2010 UI experience, thereby reducing the learning curve needed for beginners.

1) Click on the New Dataflow Button on the Menu Bar 

2) Enter a name for the Dataflow and a description that you so desire and click Create

3) We are now presented with the Dataflow Interface and Input and Output Toolbox window coupled with Transformation sections. We shall deal with transformations in part II of this tutorial

4) Our goal is to read from a database table and push the data into a text file. I have indicated the same as per the figure 1.7 below

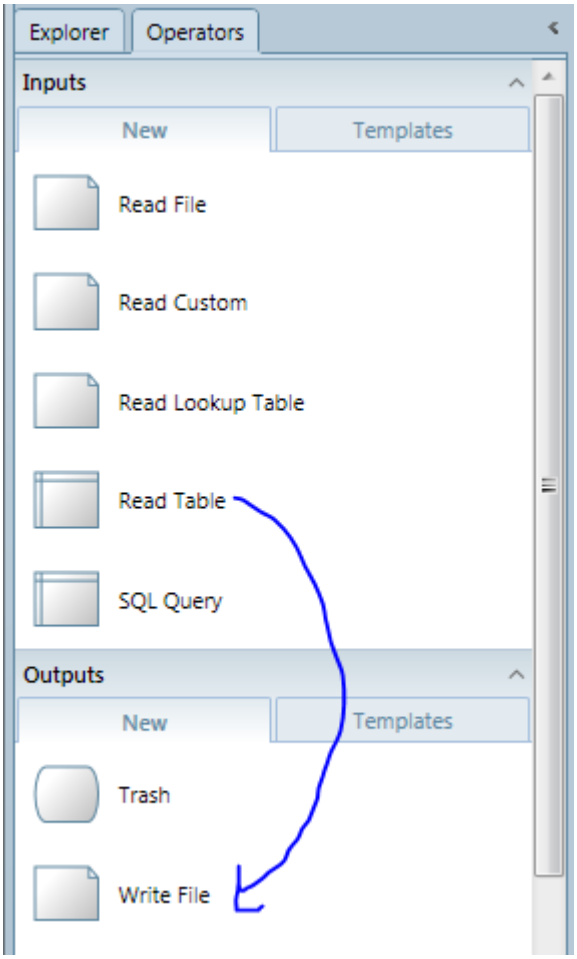


Figure 1.7 – Our objective – Read from a Database Table and push the data to a Text File

5) Drag and drop the following two operators to the blank canvas area to the right → Read Table Operator from the Inputs section and Write File Operator from the Outputs Sections and join them both. You can join them both, by clicking on the right hand side points of the Read Table Operator and drag it to the Write File Operator as shown below in figure 1.8.

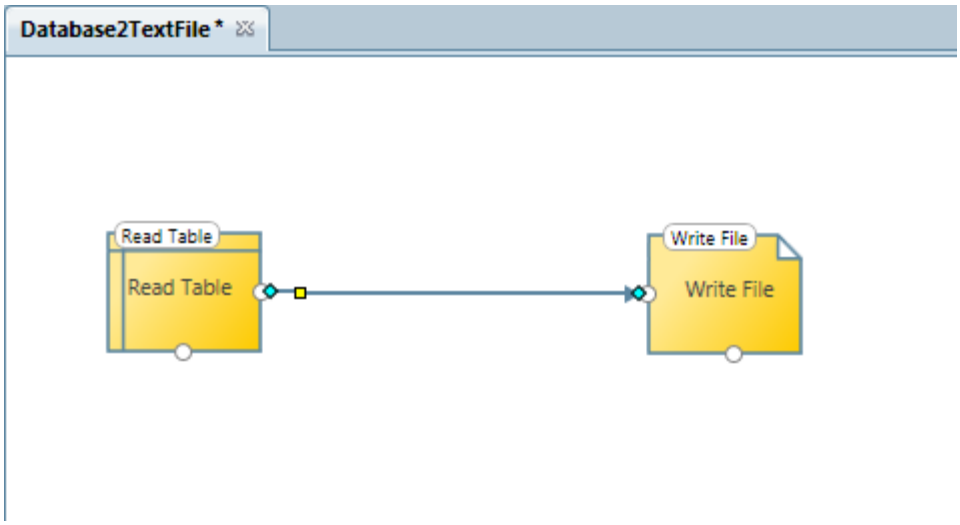


Figure 1.8 – Input and Output Operators connected

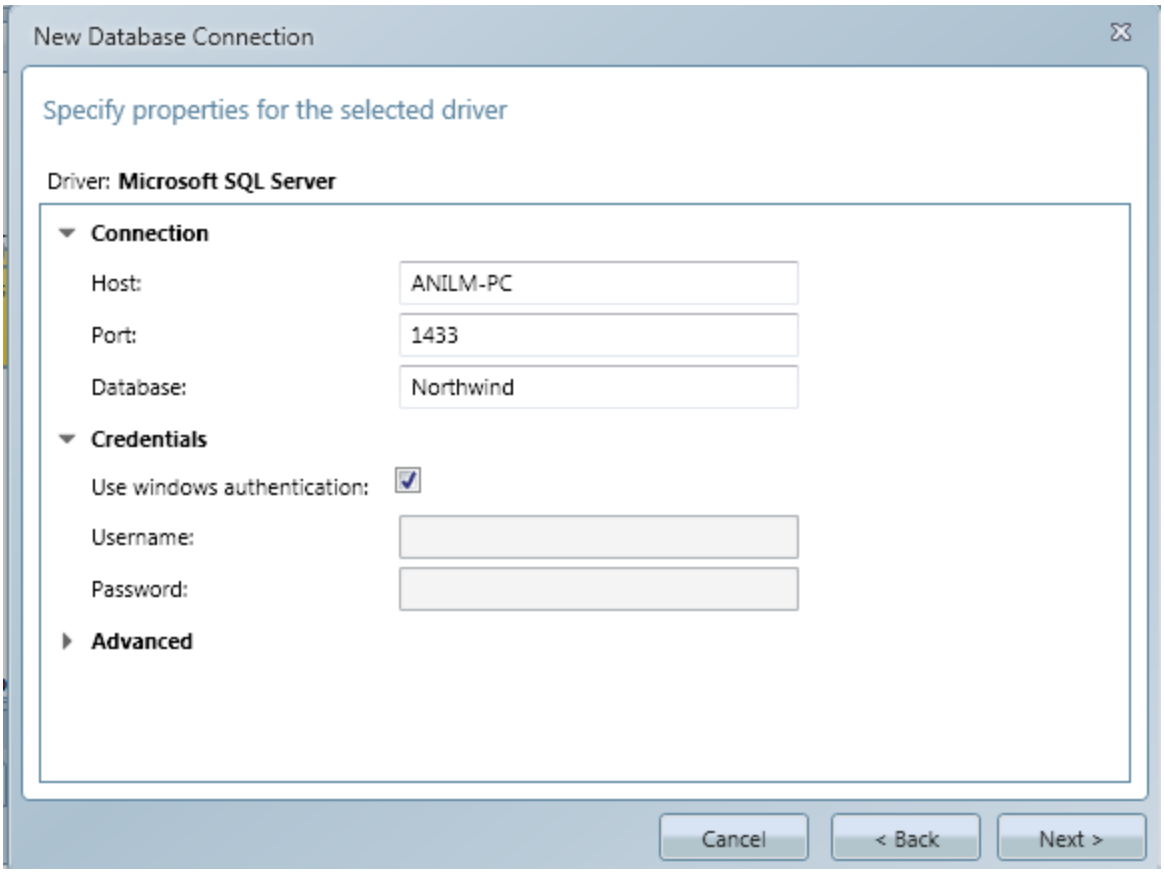
6) The above color in yellow indicates, that we still need to configure the properties of each operator, once configured, they will turn white.

Configuring the Read Table Operator

1) Click on the Read Table Operator and configure the following properties.

   a) Name: Products Table
   b) Connection: Click on the Configure drop down icon → Choose Setup a New Database Connection

   Choose under Supplied Database Drivers → Microsoft SQL Server and click Next
   Enter your SQL Server database credentials as needed.

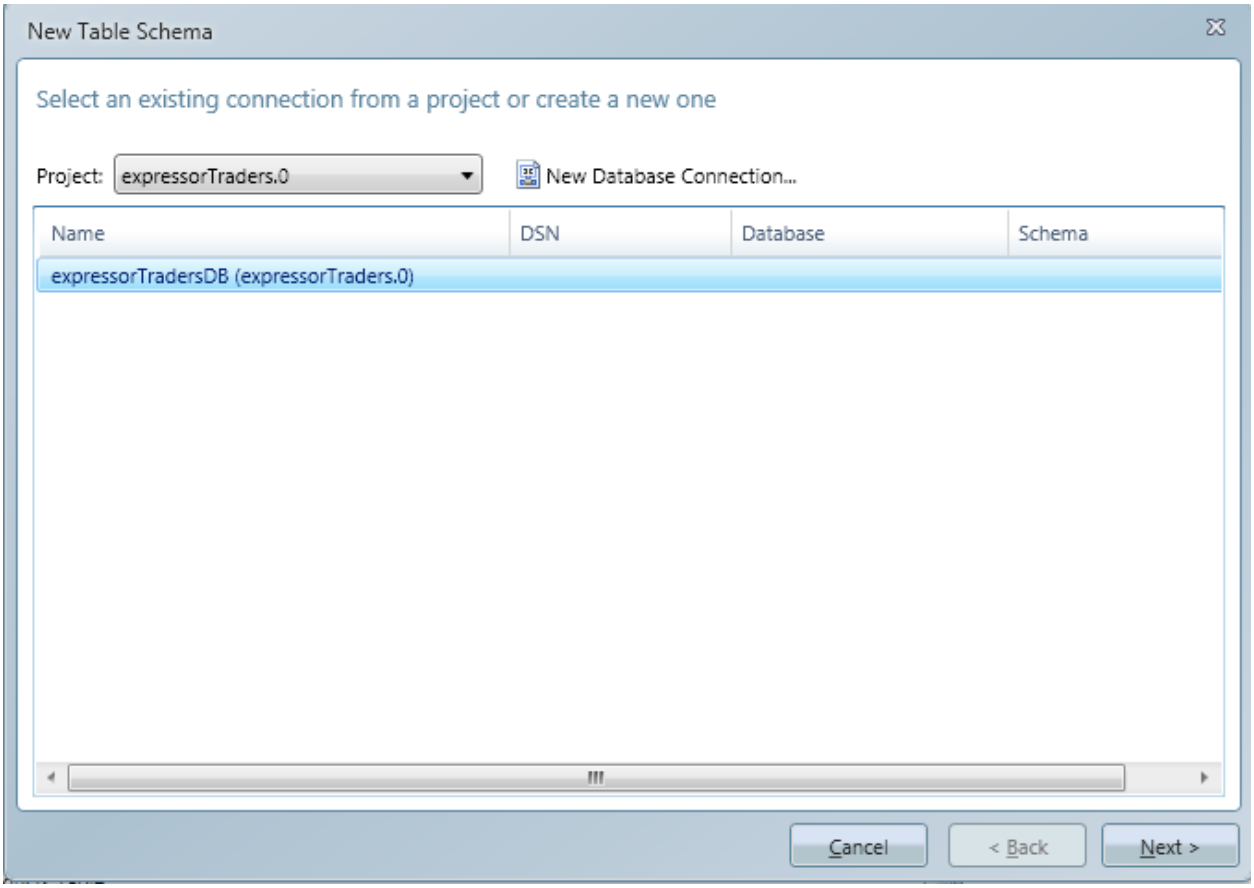


Click Next to continue. Save this connection with a name and finally click on the Finish button.

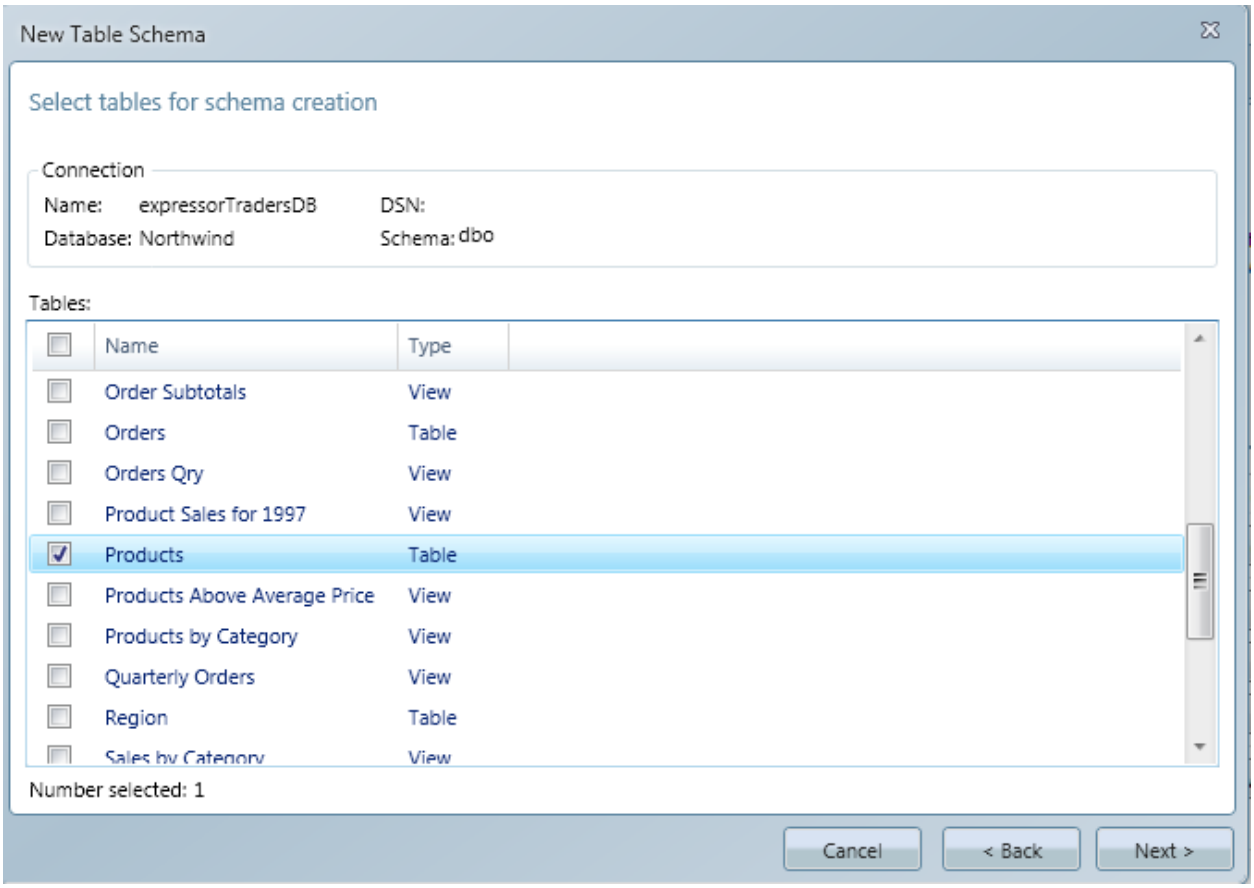
c) We now have the database connection setup; we need to configure our Schema, which is nothing but our Products table from the Northwind database.

Click on the Configure Schema button and choose the first option → create a New Table Schema

In the wizard, make sure that the Database Connection is indeed present as shown below



Click Next to continue to select our schema and tables.
Choose the **dbo** schema and click next
From the list of available Tables, we shall select the Products table as shown below



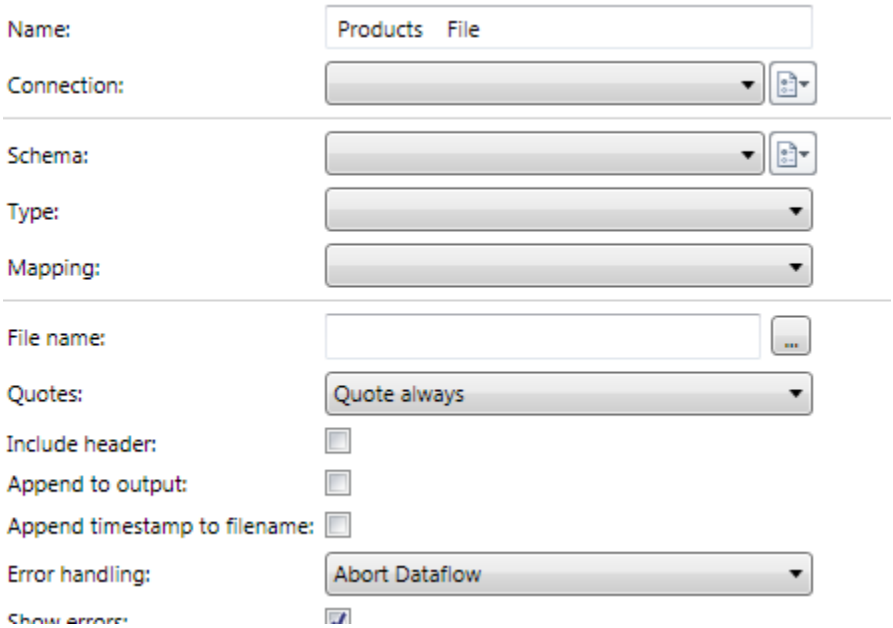
Click Next and Finish to complete the wizard

Now if you have followed the above steps as defined, you should see the Read Table (Products Table) Operator change its color from Yellow to White as shown below

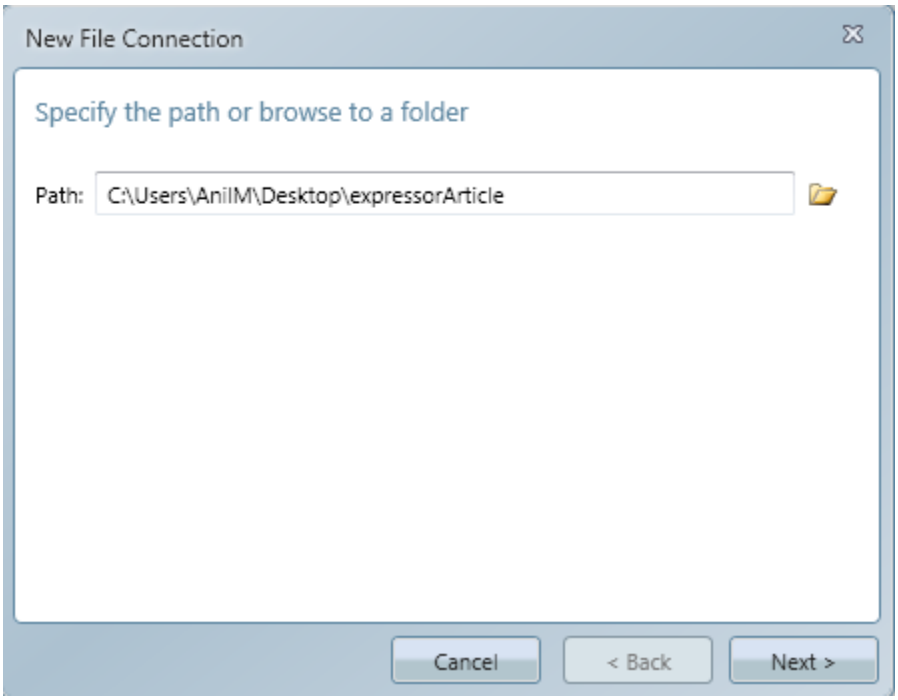
Now leave the other properties as the defaults. We have now successfully configured our Input Operator.

We shall now configure our Output Operator.

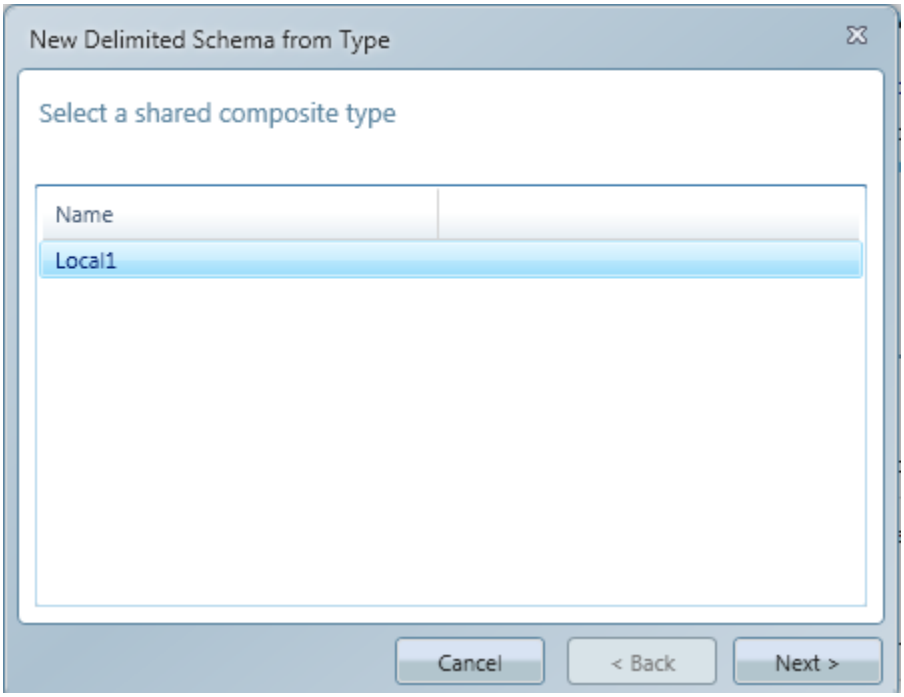
2) Select the Write File Operator and ensure that you are able to see the Properties as shown below



a) Click on the Connection Configure button and choose New File Connection Wizard
b) Choose a File Path where the file will be created as shown below



c) Click next to continue. Assign a name and click Finish
d) Now, let us configure the Schema for the file connection → Click on the Schema Configuration and choose the 3rd option, New Delimited Schema from upstream output
e) You will be presented with an already setup Composite type → Local in this case as shown below. Click Next to continue



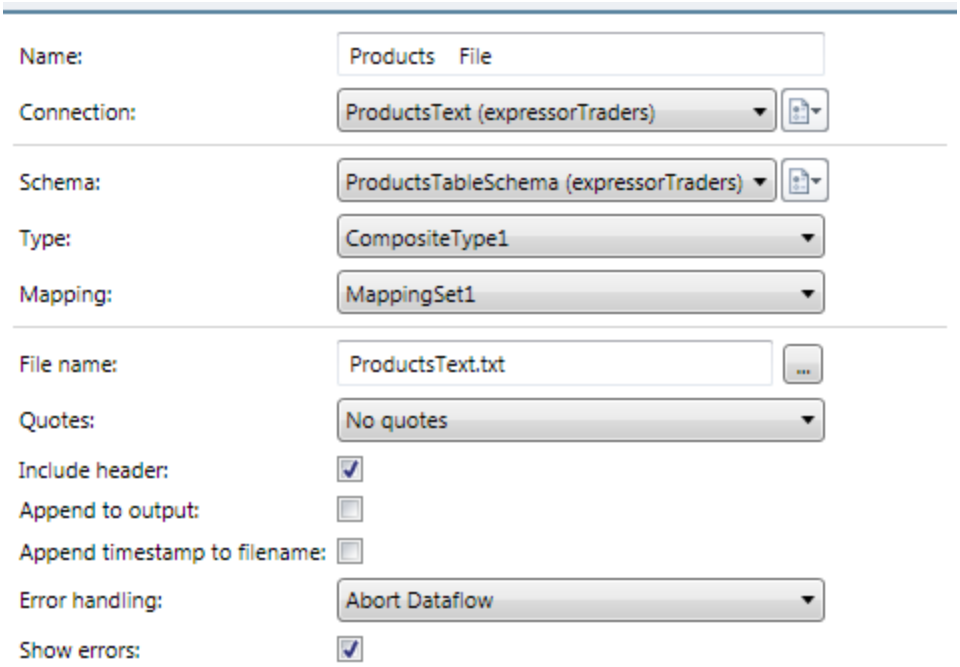
f) Assign a name to the schema and click Finish

g) Now comes the crucial part of specifying the output file; Click on the Filename ellipse button. I have already created a blank text file within the expressorArticle folder called **ProductsText.txt**, you may also create the same

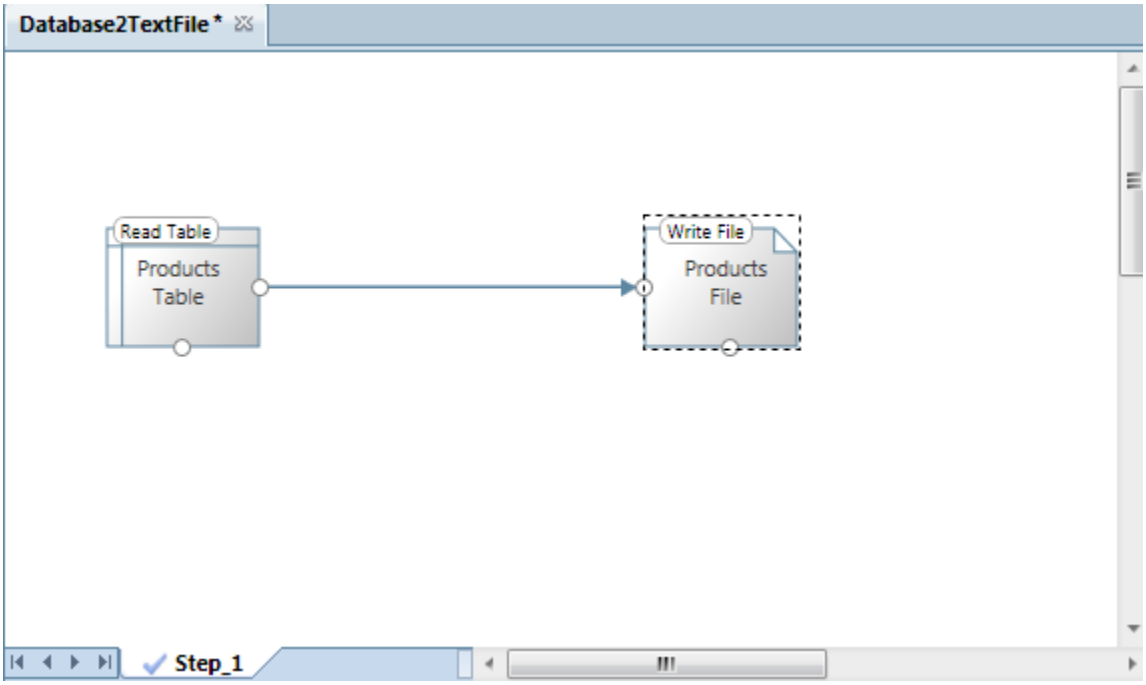
h) Quotes: Choose No Quotes
i) Check the Include Header option

Finally leave the remaining as Defaults.

You should now have a screen similar to as shown below



finally you should now see both Operators in White as shown below



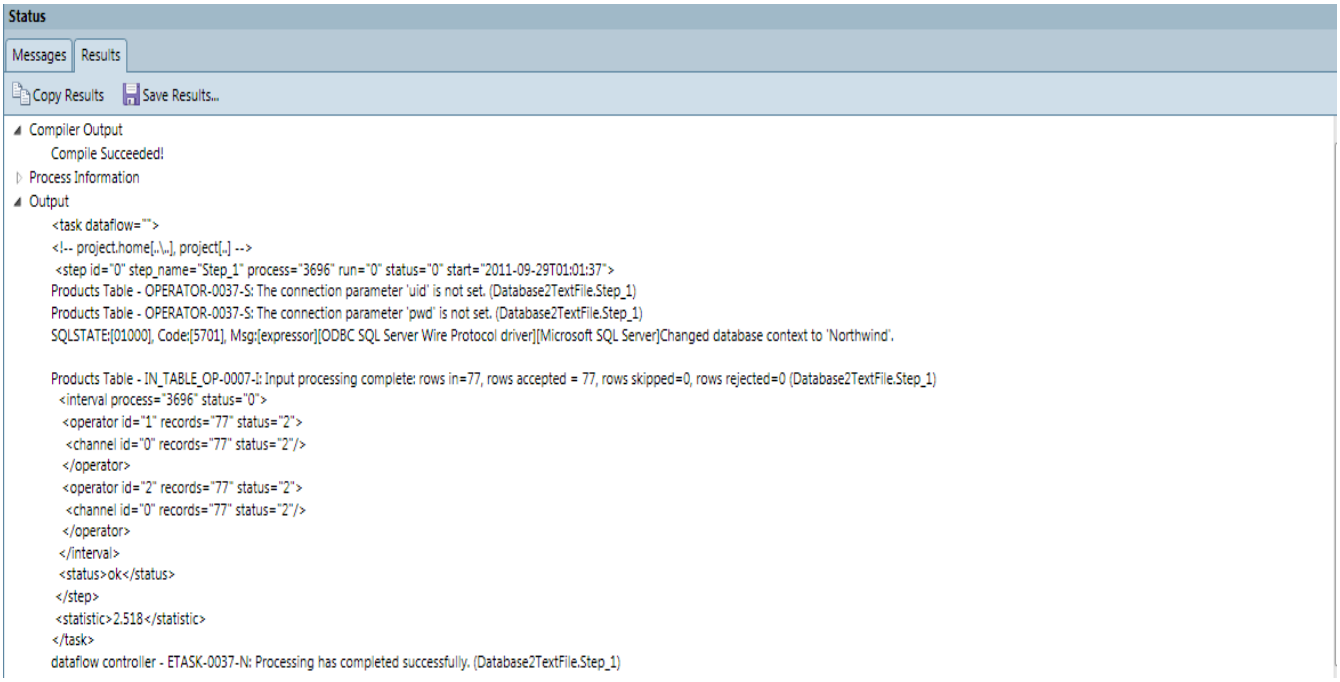
Executing the Project

Now, assuming that everything we have done so far is inline, pat yourselves on the back for a job well done!

Click on the Start Button 

to begin the process of sending the Products Data from the Database to the text file.

As you can see, we have successfully executed the Project as shown below



to verify that the text file has indeed all the records, let us now open the file as shown below.

ProductsText - Notepad

File Edit Format View Help

```
ProductID,ProductName,SupplierID,CategoryID,QuantityPerUnit,UnitPrice,UnitsInStock,UnitsOnOrder,ReorderLevel,Discontinued
1,Chai,1,1,10 boxes x 20 bags,18.0000,39,0,10,0
2,Chang,1,1,24 - 12 oz bottles,19.0000,17,40,25,0
3,Aniseed Syrup,1,2,12 - 550 ml bottles,10.0000,13,70,25,0
4,Chef Anton's Cajun Seasoning,2,2,48 - 6 oz jars,22.0000,53,0,0,0
5,Chef Anton's Gumbo Mix,2,2,36 boxes,21.3500,0,0,0,1
6,Grandma's Boysenberry Spread,3,2,12 - 8 oz jars,25.0000,120,0,25,0
7,Uncle Bob's Organic Dried Pears,3,7,12 - 1 lb pkgs.,30.0000,15,0,10,0
8,Northwoods Cranberry Sauce,3,2,12 - 12 oz jars,40.0000,6,0,0,0
9,Mishi Kobe Niku,4,6,18 - 500 g pkgs.,97.0000,29,0,0,1
10,Ikura,4,8,12 - 200 ml jars,31.0000,31,0,0,0
11,Queso Cabrales,5,4,1 kg pkg.,21.0000,22,30,30,0
12,Queso Manchego La Pastora,5,4,10 - 500 g pkgs.,38.0000,86,0,0,0
13,Konbu,6,8,2 kg box,6.0000,24,0,5,0
14,Tofu,6,7,40 - 100 g pkgs.,23.2500,35,0,0,0
15,Genen Shouyu,6,2,24 - 250 ml bottles,15.5000,39,0,5,0
16,Pavlova,7,3,32 - 500 g boxes,17.4500,29,0,10,0
17,Alice Mutton,7,6,20 - 1 kg tins,39.0000,0,0,0,1
18,Carnarvon Tigers,7,8,16 kg pkg.,62.5000,42,0,0,0
19,Teatime Chocolate Biscuits,8,3,10 boxes x 12 pieces,9.2000,25,0,5,0
20,Sir Rodney's Marmalade,8,3,30 gift boxes,81.0000,40,0,0,0
21,Sir Rodney's Scones,8,3,24 pkgs. x 4 pieces,10.0000,3,40,5,0
22,Gustaf's Knäckebröd,9,5,24 - 500 g pkgs.,21.0000,104,0,25,0
23,Tunnbröd,9,5,12 - 250 g pkgs.,9.0000,61,0,25,0
24,Guaraná Fantástica,10,1,12 - 355 ml cans,4.5000,20,0,0,1
25,NuNuCa Nuß-Nougat-Creme,11,3,20 - 450 g glasses,14.0000,76,0,30,0
26,Gumbär Gummibärchen,11,3,100 - 250 g bags,31.2300,15,0,0,0
27,Schoggi Schokolade,11,3,100 - 100 g pieces,43.9000,49,0,30,0
28,Rössle Sauerkraut,12,7,25 - 825 g cans,45.6000,26,0,0,1
29,Thüringer Rostbratwurst,12,6,50 bags x 30 sausgs.,123.7900,0,0,0,1
30,Nord-Ost Matjeshering,13,8,10 - 200 g glasses,25.8900,10,0,15,0
31,Gorgonzola Telino,14,4,12 - 100 g pkgs,12.5000,0,70,20,0
32,Mascarpone Fabioli,14,4,24 - 200 g pkgs.,32.0000,9,40,25,0
33,Geitost,15,4,500 g,2.5000,112,0,20,0
34,Sasquatch Ale,16,1,24 - 12 oz bottles,14.0000,111,0,15,0
```

Congratulations!! You have built your very first ETL Application without much difficulty with expressor Studio.

expressor Studio makes building ETL applications a breeze. If you have any feedback with regard to this tutorial, feel free to e-mail me anil.mahadev@gmail.com with the Subject indicating “expressor Studio Article”

So stay tuned for Part II, where we shall explore Transformations and many more cool features.