A Simple AAP/Vector Data Migration Project
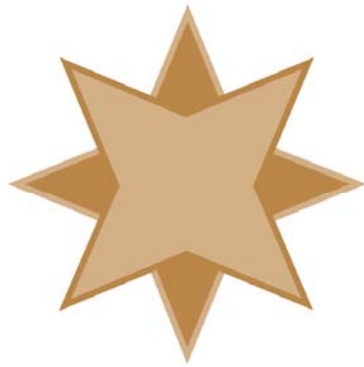
A White Paper From

GOLDSTAR
SOFTWARE

www.GoldstarSoftware.com

For more information, see our web site at
http://www.goldstarsoftware.com

A Simple AAP/Vector Data Migration Project
Last Updated: 09/23/2015

The simplest type of project you can build in Actian Dataflow is a migration project. This type of project simply takes data from a source environment (such as your PSQL database) and replicates it to your target environment (your Vector instance).

Before you can complete these steps, you should be aware of the pre-requisites:
- PSQL v12 Server should be properly installed and configured.
- The AAP/Vector environment should be properly installed to your server.
- The server should have been rebooted after the AAP installation.
- Actian Dataflow should be linked to PSQL by copying the JAR file to the correct location.
- Your PSQL database(s) should be defined in the Actian Dataflow environment.
- Your Vector database(s) should be defined in the Actian Dataflow environment.

If necessary, see the product manuals or the other Goldstar Software white papers about AAP for more information on completing these steps. We can now continue with building the migration workflow.
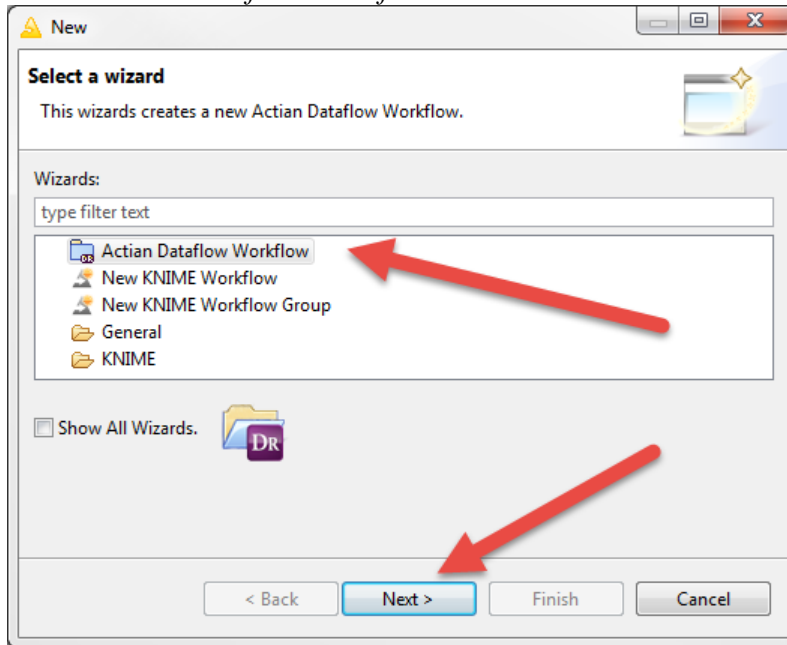
Create a New Actian Dataflow Workflow

We are going to build a workflow to handle a simple data migration of two tables for us. To start this process, we need to create a new workflow first.
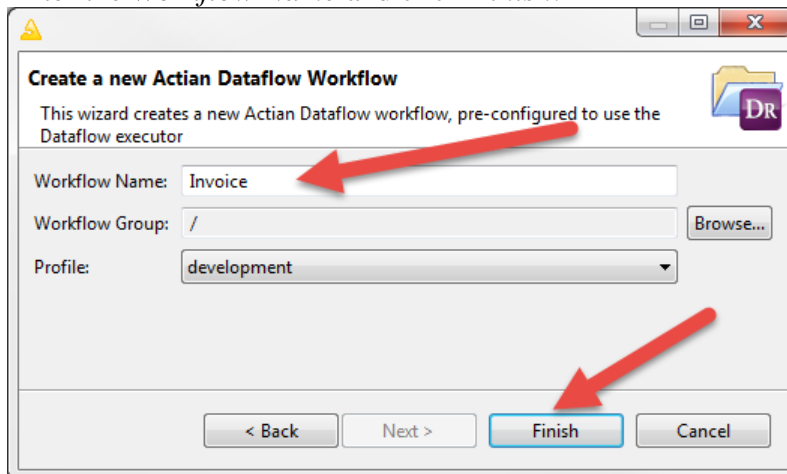1. Launch the *Actian Dataflow (KNIME)* environment.
2. From the menu, select *File*, then *New*.
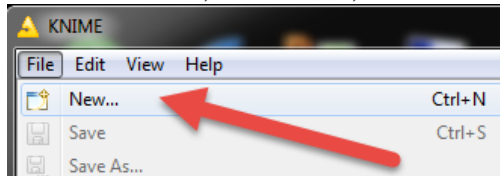
3. Select *Actian Dataflow Workflow* and click *Next*.


4. Enter the *Workflow Name* and click *Finish*.

5. The new workflow will now be created and opened for you. Of course, it is blank now, but we'll start to put objects in here in the next section.
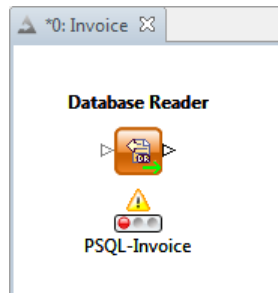
Define Database Readers

The first step in just about any workflow is to define a **Database Reader**. As the name implies, a Database Reader will read existing database records for you, providing data to the remainder of the workflow.

1. In the *Node Repository* window, which is usually in the lower left corner of the KNIME window, drill down into *Actian Dataflow*, *I/O*, and *Read* to expose the

various readers.
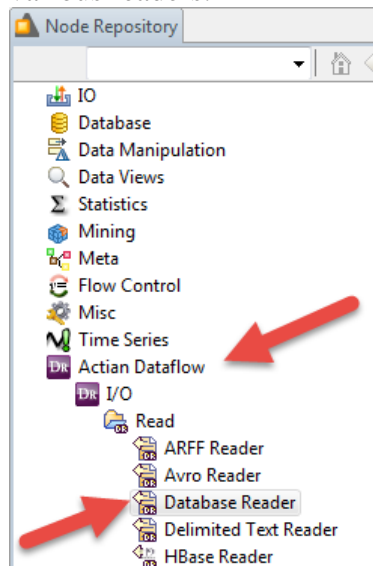
2. Drag the *Database Reader* node onto your workflow diagram. It makes the most sense if you put this on the left side of the diagram, but it can really go anywhere. Double-click the **Node 1** text at the bottom and change it to a meaningful value. In this case, we will call this node **PSQL-Invoice** to document exactly which database and table is being accessed here.


3. You can see the Red light below this node. This means that the node needs to be configured. Right-click the *Database Reader* node and select *Configure*.
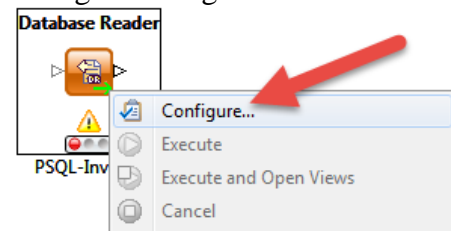

4. The *Reader Configuration* dialog box will open. In the initial release of AAP Express Plus, this dialog box is usually displayed in the wrong size, and some of the controls are all messed up. So, grab a corner and make the box slightly bigger to display all of the controls properly before continuing.
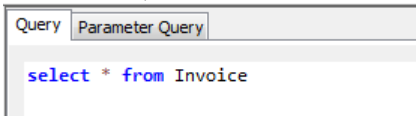
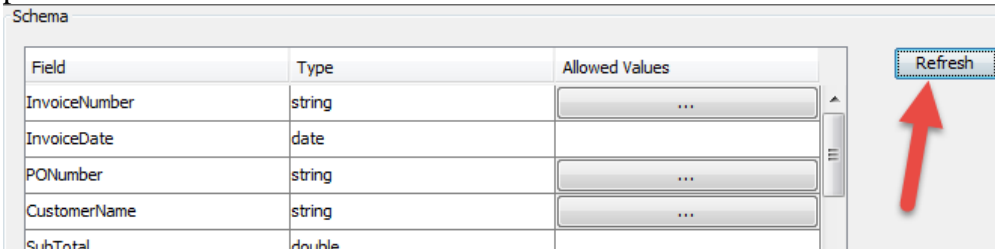5. Select your database *Connection* from the drop-down list provided.
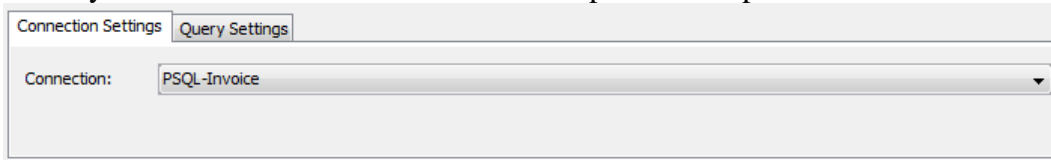
6. Then, in the *Database Browser* section, click the *Fetch Metadata* button, and you should see a list of all of your PSQL tables.
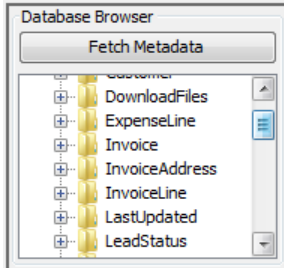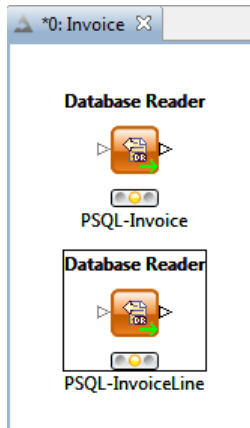

7. You next need to build your query into the *Query* window. The easiest form of this query simply uses the "**select * from** <table>" and replaces <table> with your table name, as shown in this example.

8. Click on the *Refresh* button, and the database schema for the SQL statement provided will be filled into the Schema box on the bottom section of the dialog.
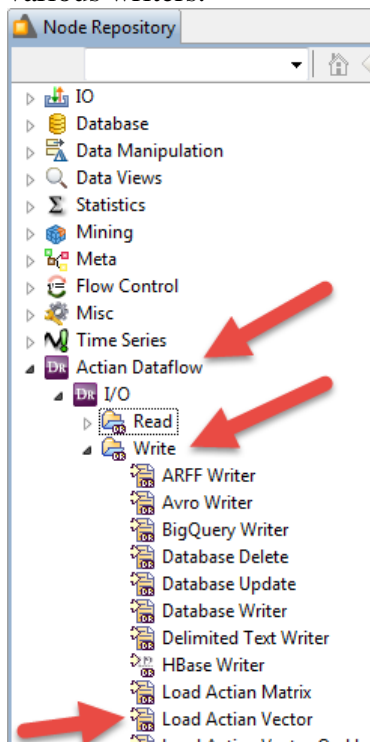

9. When you have finished, click *OK* to close the dialog box, and you will see that the Database Reader has changed its status from Red to Yellow, meaning that it is ready to execute.

10. Repeat the entire process (Steps 1 through 9) for the other tables in your database. For this example, we will create a second reader for the **InvoiceLine** table.
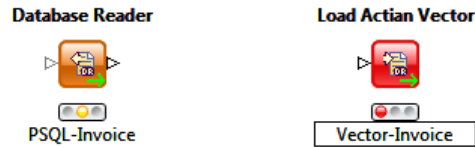
Define Database Writers

Once you have the database readers defined, you can now create the Database Writers. However, as we are going to load the data into Actian Vector, we are going to use the special high-speed data loader.

1. In the *Node Repository* window, which is usually in the lower left corner of the KNIME window, drill down into *Actian Dataflow*, *I/O*, and *Write* to expose the various writers.
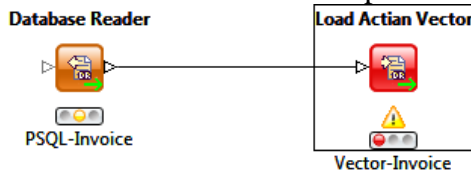

2. Drag the *Load Actian Vector* node onto your workflow diagram. It makes the most sense if you put this to the right of the existing database reader. Double-click the **Node 3** text at the bottom and change it to a meaningful value. In this case, we will call this node **Vector-Invoice** to document exactly which database
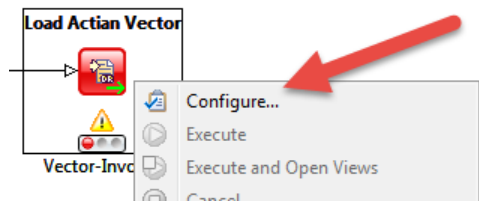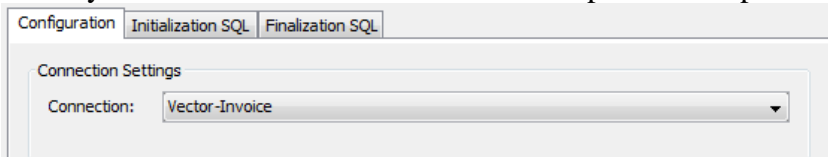
and table is being accessed here.

3. Before you can configure the writer, you must specify its inputs. To do this, use the mouse to click on the output from the *Database Reader* node and draw a line to the *Load Actian Vector* input. These two nodes are now connected.
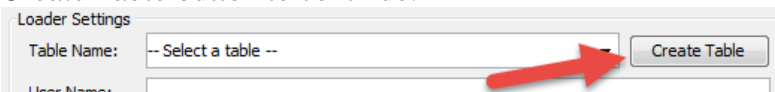

4. You can see the Red light is stil showing below this new node. This means that the node needs to be configured. Right-click the *Load Actian Vector* node and select *Configure*.

5. The *Load Actian Vector Configuration* dialog box will open. In the initial release of AAP Express Plus, this dialog box is usually displayed in the wrong size, and some of the controls are all messed up. So, grab a corner and make the box slightly bigger to display all of the controls properly before continuing.

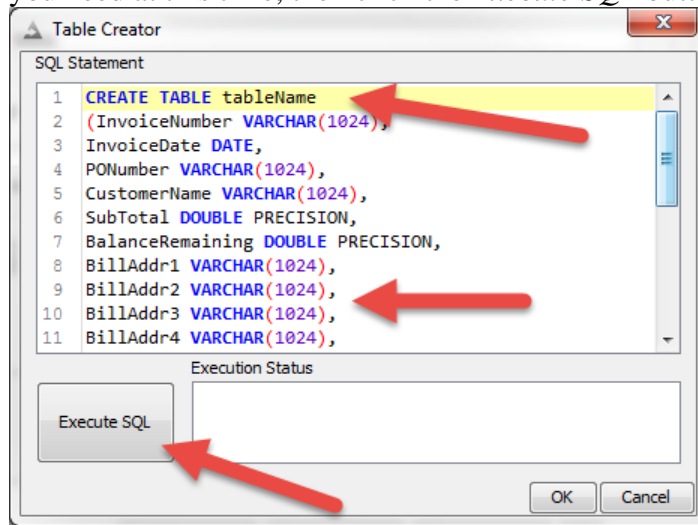6. Select your database *Connection* from the drop-down list provided.

7. Then, in the *Loader Settings* section, you will see the *Table Name* drop box doesn't have a table listed yet. In fact, the table doesn't even exist yet. Click the *Create Table* button to continue.
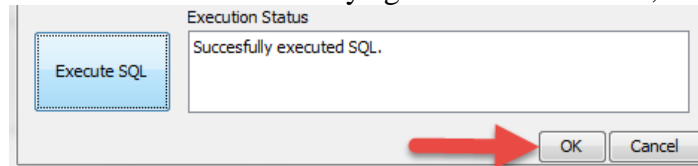

8. The *Table Creator* dialog box will appear, and it will be populated with the fields and data types of the source *Database Reader* node. However, the tableName will not be set correctly. You should first scroll to the top of this *SQL Statement* and change the table name from "**tableName**" to something meaningful. In this example, we'll use "**invoice**". (Vector tables, unlike PSQL, are always lowercase.) This is a good time to review all of your data types and field names to make sure that they are all compatible with Vector. Make any other changes
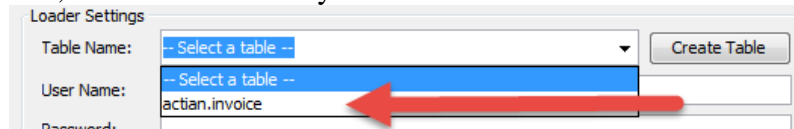
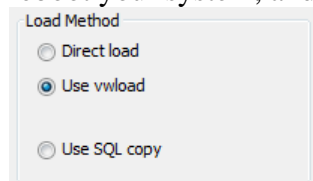you need at this time, then click the *Execute SQL* button.

9. You should receive a message "Successfully executed SQL." If not, correct the CREATE statement and try again. When finished, click *OK* to continue.

10. Back at the Configuration dialog box, you can now select the *Table Name* drop box, and select the newly-created table.
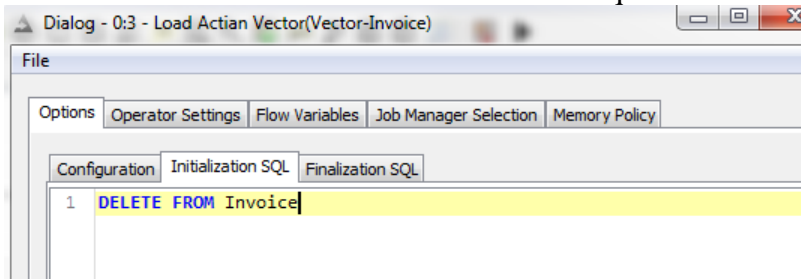

11. **IMPORTANT**: In our own testing, we have found that the "*Direct load*" method does not work correctly for many data transfers. You want to make sure that your *Load Method* is ALWAYS set to "*Use vwload*". If this option is grayed out, then you may not have rebooted since you installed AAP. Please save your data flow, reboot your system, and come back to this point before continuing.
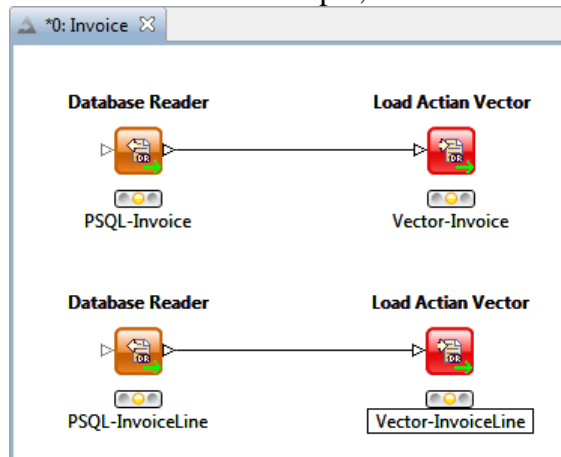

12. (Optional) If your source and target data tables don't use similar schemas, then you may need to map fields from the source to the target. Click on the *Map Fields* button, and you can define those mappings. For this simple example, we don't need to do anything with this. By default, AAP will map all fields.

13. (Optional #2) If you are going to run a Dataflow many times, the system will keep inserting records into Vector without removing them first. To force the system to clear any records from the table before writing, select the Initialization SQL tab

and enter a DELETE statement for the table in question.

14. When you have finished, click *OK* to close the dialog box, and you will see that the *Load Actian Vector* node has changed its status from Red to Yellow, meaning that it is ready to execute.
15. Repeat the entire process (Steps 1 through 13) for the other tables in your database. For this example, we will create a second writer for InvoiceLine.
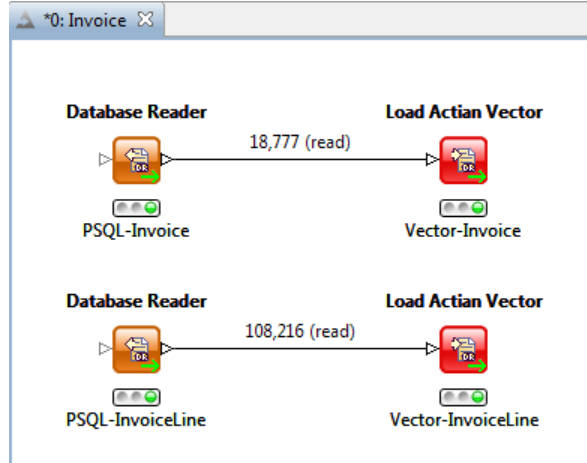

Execute All Executable Nodes

Once all readers and writers are properly configured, it is time to run the migration.

1. In the tool bar, click on the *Execute All Executable Nodes* button.

2. Dataflow will now execute the readers and writers, completing the data move. If everything executes properly, you should see a list of the records that were passed

on each connection, as well as see all of the lights turn green, indicating success.
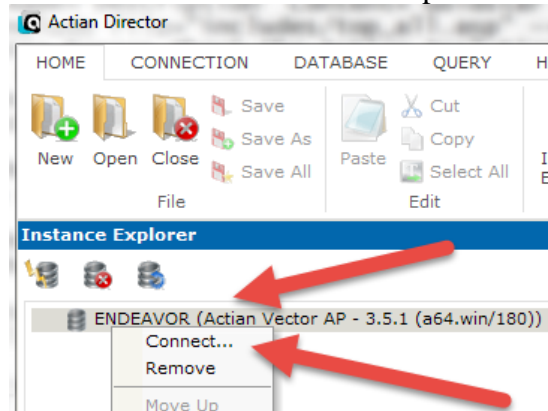

3. If any errors were returned, check the *KNIME Console* in the lower right corner of the screen for information. If you need additional details about the errors, select *View/Open KNIME Log* from the menu.
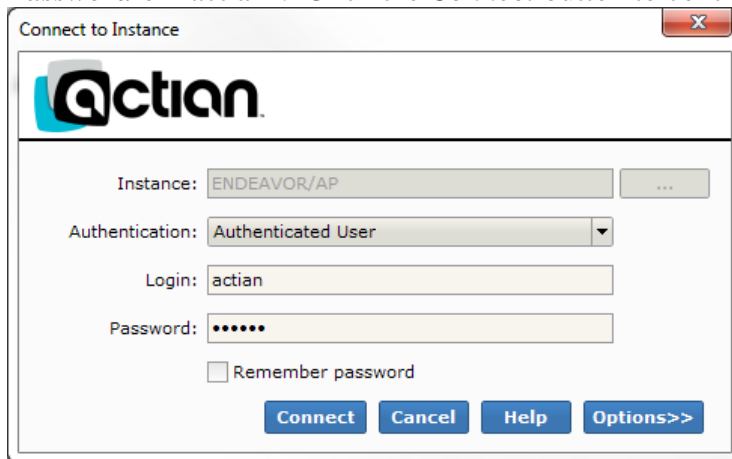
Verify the Data in Vector

If no errors are returned, then the data should now be accessible in the Vector database. You will want to verify that the data has all migrated properly before redirecting your queries to the new database.

1. Start the *Actian Director (x64)* application. Right-click the computer name listed there and select the *Connect...* option from the menu.
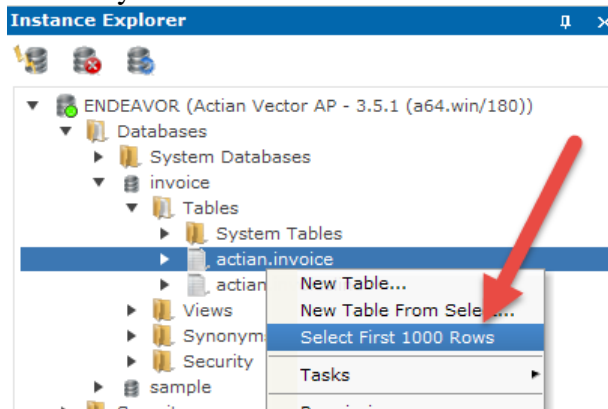

2. Enter the login credentials for your connection. If you followed the Actian directions for installing AAP, then you likely have a *Login* of "actian" and a
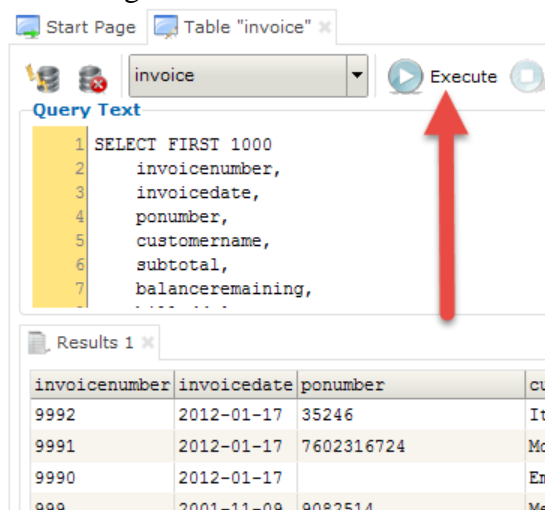
Password of "actian". Click the *Connect* button to continue.

3. In the tree view, drill down into your database until you see the tables. Right-click on your table name and select *Select First 1000 Rows*.

4. This will run a simple SELECT query against your table, showing all fields. If needed, you can modify the query to suit your needs and click the *Execute* button to run it again.

Now that you know the data is there, you can start submitting queries to Vector and enjoy the benefits of the high speed reporting engine included with PSQL v12!

Information Provided By **Goldstar Software Inc.**
http://www.goldstarsoftware.com

If you are not comfortable with this, we do offer a complete set of installation services to handle this for you. For more information, contact us at 1-708-647-7665 or visit our web site at http://www.goldstarsoftware.com.