

SeeBeyond ICAN Suite

eTL Integrator User Guide

Release 5.0



The information contained in this document is subject to change and is updated periodically to reflect changes to the applicable software. Although every effort has been made to ensure the accuracy of this document, SeeBeyond Technology Corporation (SeeBeyond) assumes no responsibility for any errors that may appear herein. The software described in this document is furnished under a License Agreement and may be used or copied only in accordance with the terms of such License Agreement. Printing, copying, or reproducing this document in any fashion is prohibited except in accordance with the License Agreement. The contents of this document are designated as being confidential and proprietary; are considered to be trade secrets of SeeBeyond; and may be used only in accordance with the License Agreement, as protected and enforceable by law. SeeBeyond assumes no responsibility for the use or reliability of its software on platforms that are not supported by SeeBeyond.

SeeBeyond, e*Gate, and e*Way are the registered trademarks of SeeBeyond Technology Corporation in the United States and select foreign countries; the SeeBeyond logo, e*Insight, and e*Xchange are trademarks of SeeBeyond Technology Corporation. The absence of a trademark from this list does not constitute a waiver of SeeBeyond Technology Corporation's intellectual property rights concerning that trademark. This document may contain references to other company, brand, and product names. These company, brand, and product names are used herein for identification purposes only and may be the trademarks of their respective owners.

© 2003 by SeeBeyond Technology Corporation. All Rights Reserved. This work is protected as an unpublished work under the copyright laws.

This work is confidential and proprietary information of SeeBeyond and must be maintained in strict confidence.

Version 20031015130533.

Contents

Chapter 1

System Description	5
Introduction	5
The eTL Integrator Product	5
The ETL Process	6
eTL Supporting Features	6
Business Integration and the eTL Integrator	8
Supporting Documents	9
Writing Conventions	10
eGate Installation Requirements	10
The SeeBeyond Web Site	11

Chapter 2

Interface to eGate	12
Enterprise Designer Components	12
Menu Bar	13
Enterprise Explorer	14
Project Editor	14
Creating Flat File OTDs	14
Importing Metadata Information for flat files	16

Chapter 3

Creating a Simple Project	24
Sample Scenario	24
Using eTL With elnsight	43

Chapter 4

Deployment	48
General Instruction about Creating a Deployment Profile	48
Verify the Output Data	49
Create Environment and Activate the Deployment Profile for eTL	49
Scenario 1	50
Scenario 2	55
Glossary	59
ETL Terms	59
Index	62

Chapter 1

System Description

SeeBeyond's eTL Integrator technology is optimized for very large record sets and build data scenarios that are fully integrated with the SeeBeyond ICAN suite (Integrated Composite Application Network Suite) to unify the domains of eAI (eBusiness and Application Integration) and ETL. The eTL Integrator can be integrated into the enterprise business process or used as a classic, standalone, ETL process.

1.1 Introduction

Extraction Transform and Load (ETL) is a data integration technology that extracts data from several heterogeneous data sources, combines and standardizes the data, then presents or stores the data in a uniform format for informational purposes.

ETL is necessary because many non-modern system architectures evolved over the years in environments where data was typically captured, processed and stored by separate and distinct software applications and databases. As a result, the data residing in the databases of many companies is typically non-standardized.

1.1.1. The eTL Integrator Product

Product Description

SeeBeyond's eTL Integrator technology is optimized for very large record sets and build data scenarios that are fully integrated with the SeeBeyond ICAN suite (Integrated Composite Application Network Suite) to unify the domains of eAI (eBusiness and Application Integration), and Enterprise Information Integration (EII). With these unified domains you can build unprecedented solutions using both message based processing (eGate) and dataset based processing (eTL) technologies.

The eTL Integrator product provides excellent performance at runtime for high volume extraction, and load of tabular data sets, which reduces eGate Collaboration design time. The eTL Integrator can be integrated into enterprise business processes or used as a classic, standalone product. The ETL Process

In an ETL process, data is extracted from data sources. The data is then transformed (or processed), using rules, algorithms, concatenations, or filters or, into a desired state suitable for loading into a database or data warehouse. See the following Figure 1.

Product Usage

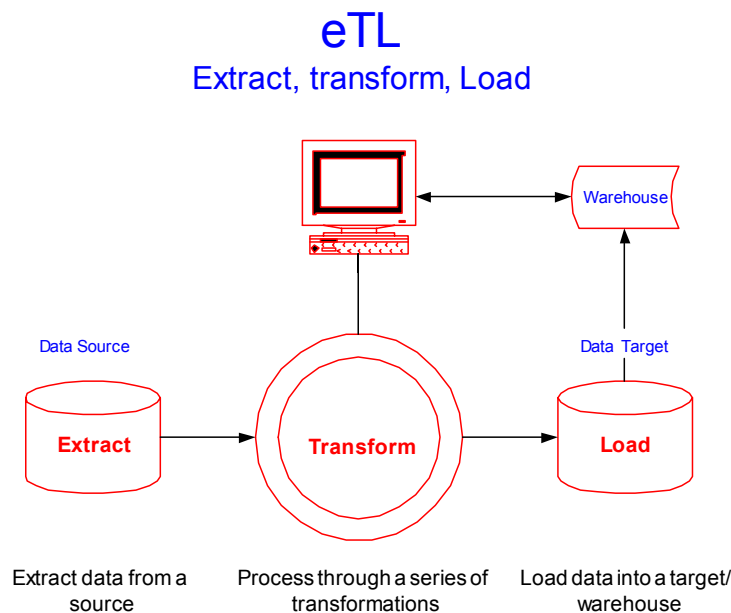
The eTL product can be used to acquire a temporary subset of data for reports or other purposes, or acquire a more permanent data set for the population of a data mart or data warehouse. The product may also be used for conversion of one database type to another or for the migration of data from one database or platform to another.

1.1.2. The ETL Process

In managing databases, extract, transform, load (ETL) refers to three separate functions combined into a single programming tool.

- 1 First, the extract function reads data from a specified source database and extracts a desired subset of data.
- 2 Next, the transform function works with the acquired data – using rules or lookup tables, or creating combinations with other data – to convert it to the desired state.
- 3 Finally, the load function is used to write the resulting data to a target database, which may or may not have previously existed.

Figure 1 The ETL Process



1.2 eTL Supporting Features

Support

eTL Integrator is compatible with the following systems and platforms:

- Oracle, SQL Server, and Flat Files (tabular)

- Multiple sources and multiple destinations
- Standard eGate platform support
- Internet Explorer 6 SP1

Built in Integration Capability

eTL Integrator enables seamless filtering and data transformation.

- Merge/upsert (updates or inserts as appropriate)
- Drag and drop GUI design features (create joins across disparate data sources)
- Validate Collaborations before performing the ETL processes (ICAN Suite provides versioning and history)

Design Tools

User friendly, state-of-the-art, design tools reduce development time and cost.

The eTL Integrator Collaboration editor has two key characteristics that maximize productivity and ease of use:

- 1 GUI based Collaboration editor employs drag and drop design features
 - ♦ User friendly Wizards (easy OTD creation)
 - ♦ Graphical operators (dragged from a toolbar)
 - ♦ Graphical tools (create underlying SQL)
- 2 Tight integration among ICAN Suite business data systems
 - ♦ Web Services interface
 - ♦ Seamless integration with the ICAN Suite

Development Tools

Development is simplified with GUI based development tools that are appropriate for SQL Collaborations. Graphical drag and drop modeling tools enable SQL operations in various categories:

- Number
- Comparison
- Boolean
- SQL Specific
- String

Transformation Capability

eTL Integrator provides all of the common operations in the following areas:

- SQL operators
- Mathematical operators
- String manipulations
- Source date format must match the target date format. (Date format conversions - later release)

- Conditional data transformations

Architecture

Robust business application integration throughout the ICAN Suite makes eTL a more versatile and powerful tool.

- A deployed eTL engine runs as a JCA compliant (J2EE) resource adapter inside the SeeBeyond Integration server.
- The business rules defined by the eTL Collaboration definition are stored in the SeeBeyond Repository.
- At deployment time, the business rules are used to generate the appropriate platform specific SQL.
- eTL Integrator leverages OTDs defined in the Enterprise Designer so you don't have to create OTDs specifically for an eTL Collaboration.

Key Operations and Functionality

An extensive array of operators, filtering, and data manipulation tools offer unlimited data design capability.

- Join
 - ♦ Auto-detect primary key relationships between tables
 - ♦ Between tables from disparate data sources that have no relationship
 - ♦ Supports, inner, left, right, and full outer joins
- Lookup
 - ♦ Extensive list of operators including SQL and string
- Merge
 - ♦ Automatic update if row exists
 - ♦ Automatic insert if row doesn't exist
- Test data and test runs
- Runtime variables (configured by the user)

1.2.1. Business Integration and the eTL Integrator

ETL Technology

- Batch oriented operations are typically restricted to batch windows in a regularly scheduled timeframe.
- Interfaces with data stores (e.g. RDBMS).
- Intended primarily for creating data warehouses.
- Not well suited for online transactions.
- Designed for one-to-one (i.e. point-to-point), integration scenarios.

- Hub-and-spoke architecture (single point-of-failure, hardware forecasting/front-loading required for scalability).
- Typically limited to technical personnel, since business managers usually don't work at the RDBMS row/column level.

SeeBeyond Business Integration Suite Technology

- Real-time oriented (i.e. online transaction support, datamart synchronous and/or asynchronous, publish/subscribe).
- Excellent batch operation (i.e. initial bulk ETL).
- Interfaces with business processes (applications), middleware, component-ware, TP monitors, application servers, web servers and data stores.
- Primarily used to allow the seamless flow of real-time information among disparate applications, supporting a unified business process within the enterprise.
- Manages both inter and intra-enterprise integration (i.e. A2A and B2B). As the business model extends to the internet, the Partner Management facility becomes important, allowing the creation and deletion of business relationships quickly and efficiently.
- Provides Business Process Management functionality, which then drives integration at both the A2A and B2B levels. This allows the integration task to be partitioned among both high-level business managers and the technical personnel that manage the applications and data stores within the enterprise.
- In general, EAI technology (specifically SeeBeyond's e*Gate), provides a superset of the functionality found in ETL tools, allowing organizations to be nimble and agile in today's constantly changing business world.

This chapter includes

- [“Supporting Documents” on page 9](#)
- [“eGate Installation Requirements” on page 10](#)
- [“The SeeBeyond Web Site” on page 11](#)

1.3 Supporting Documents

The following SeeBeyond documents provide additional information about eGate Integrator:

- *eGate Integrator Installation Guide*
- *eGate Integrator Release Notes*
- *eGate Integrator User's Guide*
- *Message Server Reference Guide*
- *eGate Tutorial*

- *SeeBeyond ICAN Suite Deployment Guide*
- *SeeBeyond ICAN Suite Primer*

See the *SeeBeyond ICAN Suite Primer* for a complete list of eGate Integrator documentation. You can also refer to the appropriate Windows or UNIX documents, if necessary.

1.4 Writing Conventions

The following writing conventions are observed throughout this document.

Table 1 Writing Conventions

Text	Convention	Example
Button, file, icon, parameter, variable, method, menu, and object names.	Bold text	<ul style="list-style-type: none"> ▪ Click OK to save and close. ▪ From the File menu, select Exit. ▪ Select the logicalhost.exe file. ▪ Enter the timeout value. ▪ Use the getClassname() method. ▪ Configure the Inbound File eWay.
Command line arguments and code samples	Fixed font. Variables are shown in <i>bold italic</i> .	<code>bootstrap -p <i>password</i></code>
Hypertext links	Blue text	For more information, see "Writing Conventions" on page 10.

Additional Conventions

Windows Systems

For the purposes of this guide, references to "Windows" will apply to Microsoft Windows Server 2003, Windows XP, and Windows 2000.

Path Name Separator

This guide uses the backslash ("\") as the separator within path names. If you are working on a UNIX system, please make the appropriate substitutions.

1.5 eGate Installation Requirements

To simplify these examples, this tutorial assumes you have all of the following eGate components installed on a single Windows system:

- eGate Repository

- Logical Host
- Enterprise Designer

Refer to the *eGate Integrator Installation Guide* for system requirements and installation instructions.

1.6 The SeeBeyond Web Site

The SeeBeyond Web site is your best source for up-to-the-minute product news and technical support information. The site's URL is:

<http://www.seebeyond.com>

Chapter 2

Interface to eGate

The Enterprise Designer is the graphical user interface (GUI) used to design and implement ICAN 5.0 projects. This chapter overviews the features and interface of the Enterprise Designer window.

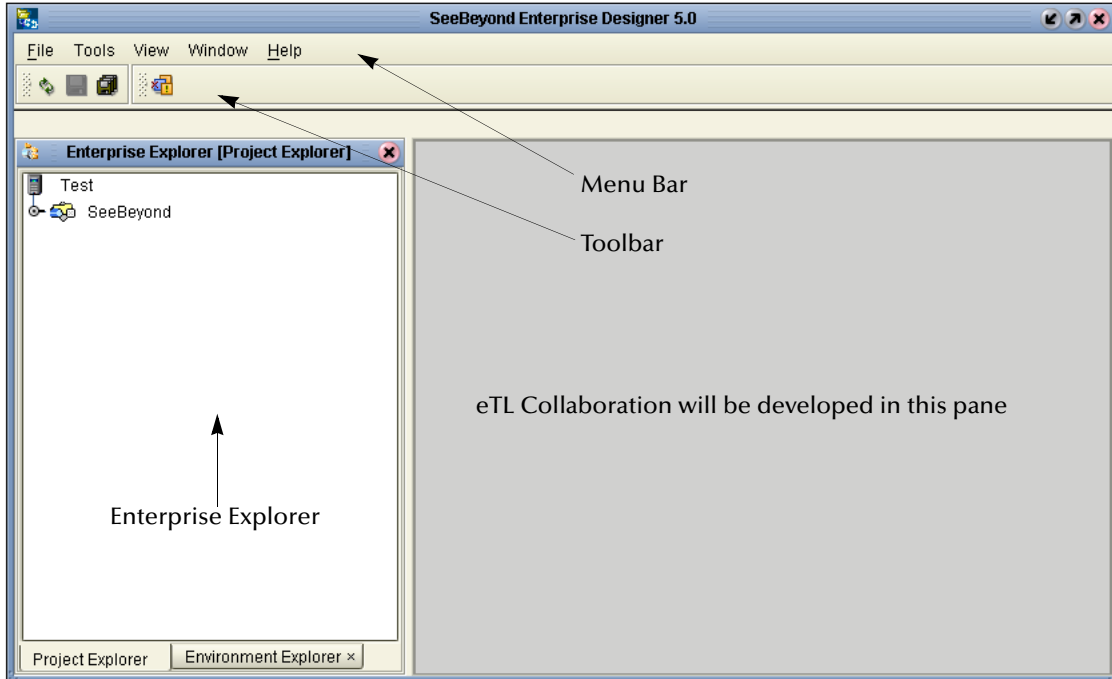
This chapter includes

- “Enterprise Designer Components” on page 12
- “Menu Bar” on page 13
- “Enterprise Explorer” on page 14
- “Project Editor” on page 14

2.1 Enterprise Designer Components

The Enterprise Designer is used to create and configure the components of an eGate Project. Each component of this interface is identified in Figure 2.

Figure 2 SeeBeyond Enterprise Designer Window



Note: This chapter provides a high-level overview of the Enterprise Designer GUI. Refer to the eGate Integrator User's Guide for a more detailed description of the menu bar, toolbar, Enterprise Explorer, and Enterprise Designer.

2.2 Menu Bar

The menu bar shown below in Figure 3 provides access to a series of menu commands used to build and manage a Project. eTL specific tools are explained later in this document.

Figure 3 Enterprise Designer Menu Bar

File Tools View Window Help

The function of each menu is briefly described below:

- **File**—Lists options for import, export, save, save all, and exit.
- **Tools**—Lists options for Impact Analyzer and Update Center.
- **View**—Lists options to view the Environment Explorer and Project Explorer.
- **Window**—Lists options for the various window displays, cascade, tile, etc.
- **Help**—online help for all installed modules

2.3 Enterprise Explorer

The Enterprise Explorer organizes all of the components of a Project into a series of folders and contains the following two tabs:

- **Project Explorer**—Logical configurations designed to help solve a business problem. This branch includes all the components of an Enterprise Designer Project, including, Collaborations, Connectivity Maps, Services, Object Type Definitions (OTD), and Deployment Profiles.
- **Environment Explorer**—Collections of logical hosts and external systems capable of hosting eGate components and information about external systems, which may be involved with an eGate configuration.

2.4 Project Editor

The Project Editor contains the “nuts-and-bolts” of a Project. This part of the Enterprise Designer is empty when you start a new Project. However, as you work through the tutorial, the Project Editor will quickly fill with components and graphical structures representing the various stages of the Project. The types of windows in the Project Editor area include:

- **Connectivity Map**—Contains business logic components, such as Services, Topics, queues, and eWays, that you include in the structure of a Project.
- **OTD Editor**—Edits and tests the OTDs (Object Type Definitions)
- **eTL Editor**—Create eTL Collaborations
- **Deployment Profile Editor**—Edits the deployment profile

2.4.1. Creating Flat File OTDs

In addition to DTD and XSLT based flat files, the ICAN suite also supports tabular data sets in flat files. You can now build OTDs for any kind of structured data, including spreadsheets. flat file OTDs can also be used throughout the ICAN suite.

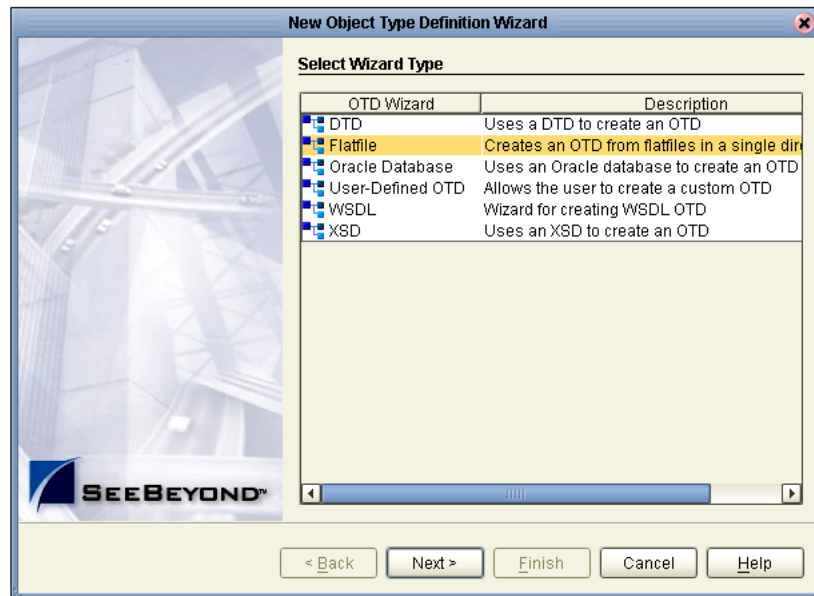
Note: Refer to Chapter 7 of the User Guide for information about collaboration definitions and using the OTD Wizard.)

Creating a flat file OTD

Open the eGate Enterprise Designer.

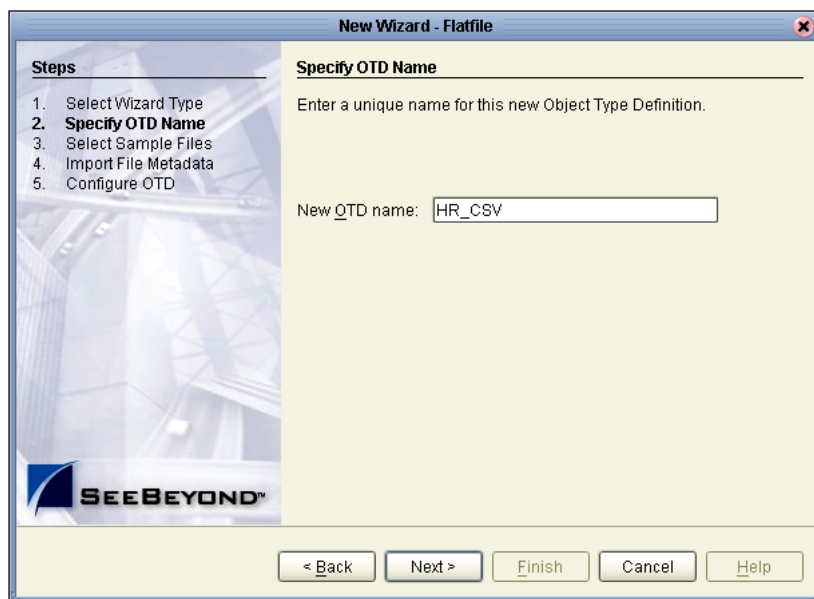
- 1 Right-click on your Project and select **Object Type Definition**.
- 2 Select **flat file** from the list of OTD Wizards and click **Next**.

Figure 4 Select OTD Wizard Type



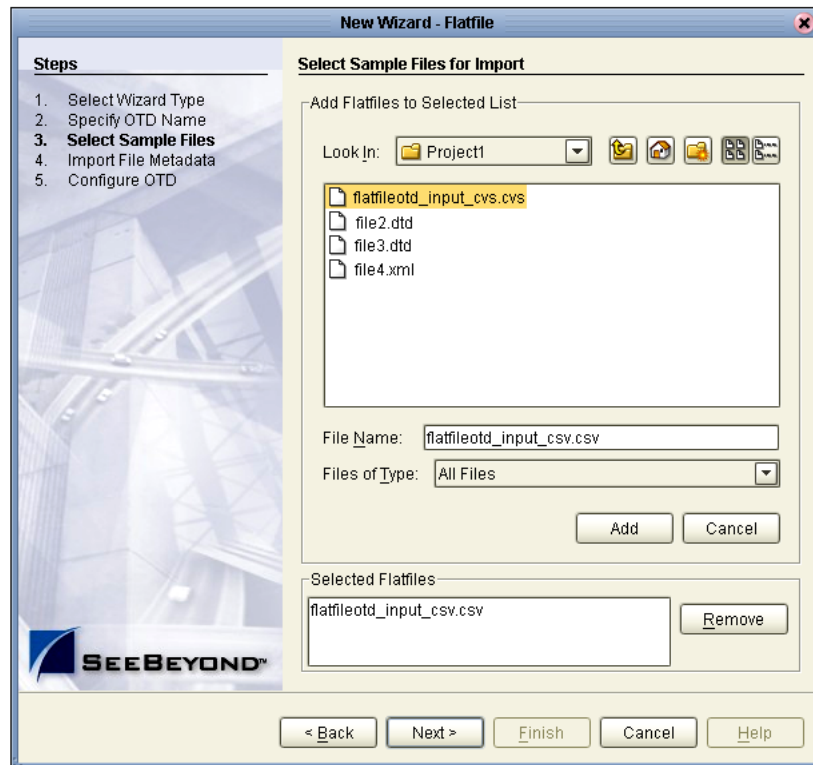
The Specify OTD Name window appears. See Figure 5 below.

Figure 5 Specify OTD Name



- 3 Type a name for your flat file OTD and click **Next**.
The Select Sample Files for Import window appears.

Figure 6 Select Sample Files for Import



In Figure 6 above you can browse to find a file on your computer or on the network.

- 4 Click the drop-down arrow to navigate to the file you want to select.
- 5 Click the **Add** button to select the file.

The selected file(s) appears in the Selected flat files list.

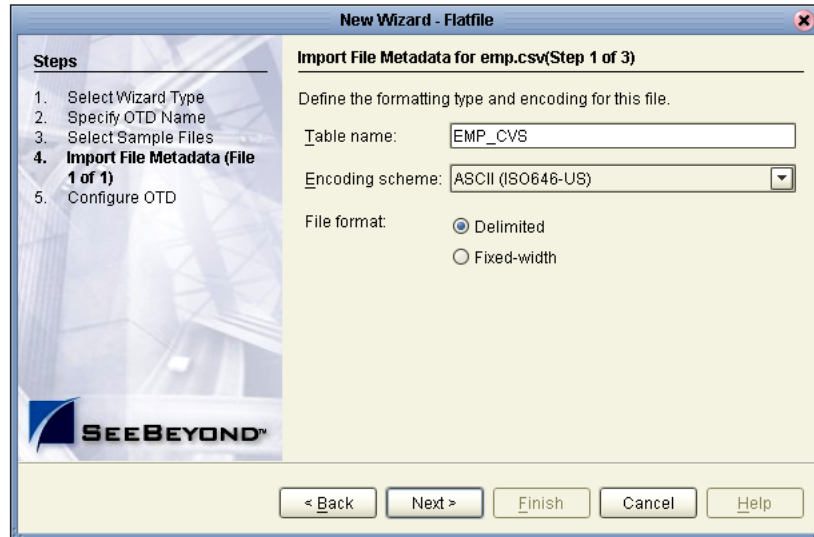
Note: *The file must reside on your computer or on a network location you have permission to access.*

You can select one or more flat files similar to the way you can include multiple tables in an Oracle database OTD. Later in this process, the system will automatically inspect files to assess structure and read sample data.

2.4.2. Importing Metadata Information for flat files

This section describes methods used to supply metadata for parsing selected flat files.

Figure 7 Encoding Scheme and File Format



- 1 Enter a Table name and select an encoding scheme. The default encoding is ASCII.
- 2 Select a File format, **Delimited** or **Fixed-width**.

There are five criteria and delimiters used to instruct the system how to parse your selected flat file. The following are valid for delimited only:

- ♦ Default SQL Type
- ♦ Record Delimiter
- ♦ Field Delimiter
- ♦ Text Qualifier
- ♦ First line contains field names?

In our sample we are configuring the delimited format, Figure 10. The following two figures show examples of fixed width formats:

Figure 8 Fixed Width Parse Properties, Character length

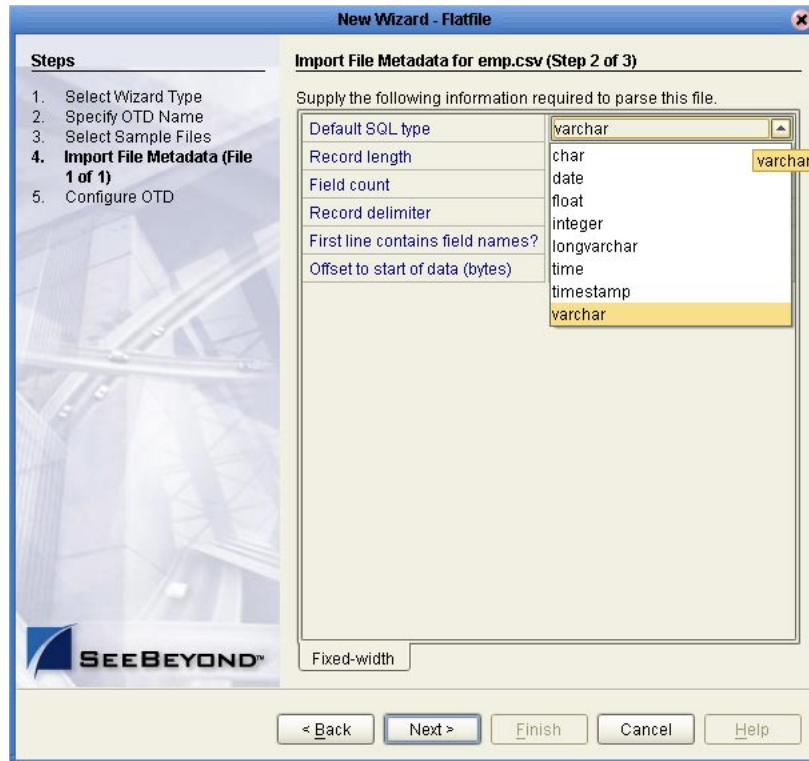
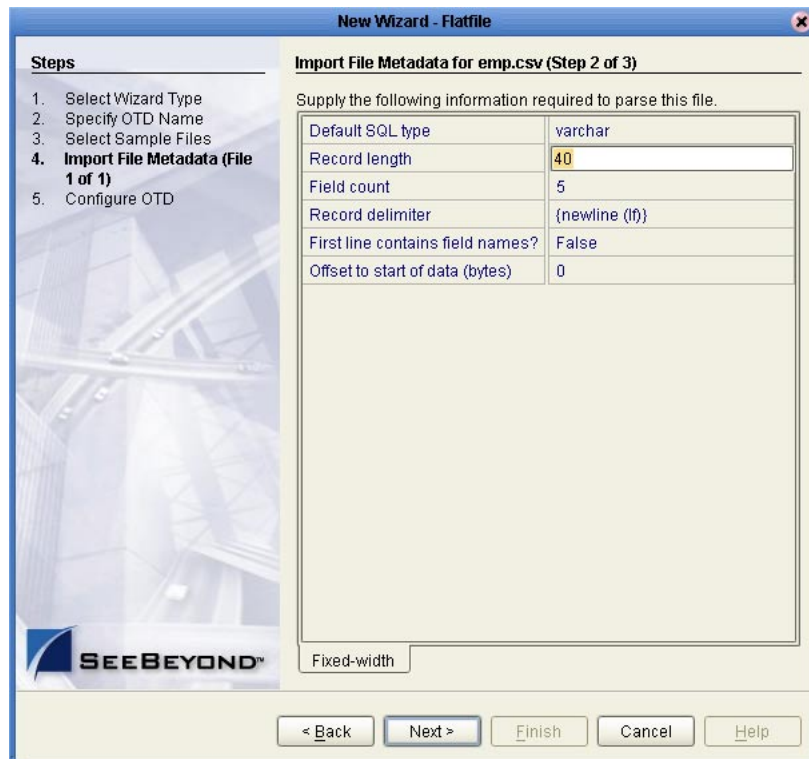
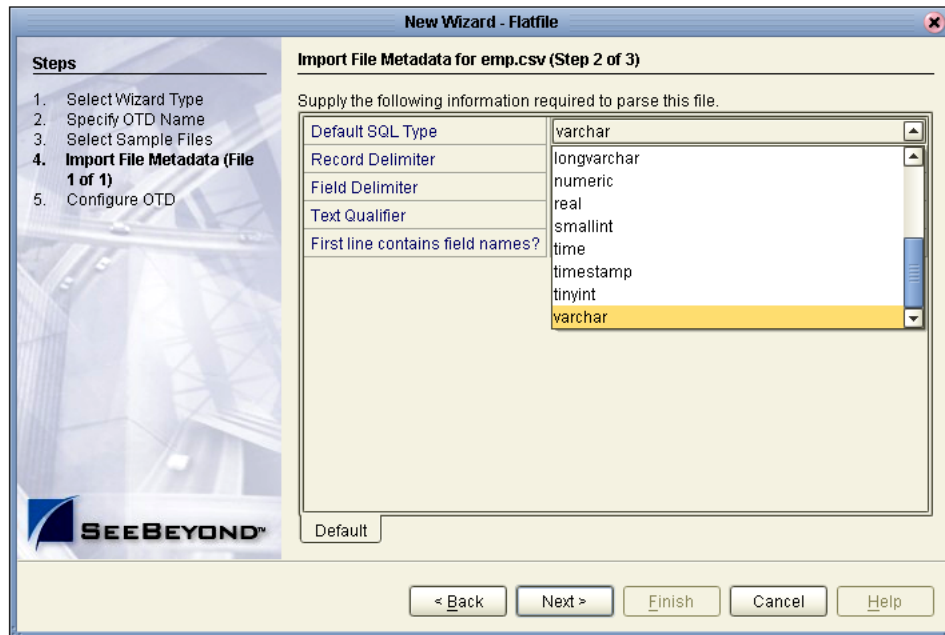


Figure 9 Fixed Width Parse Properties, fixed length



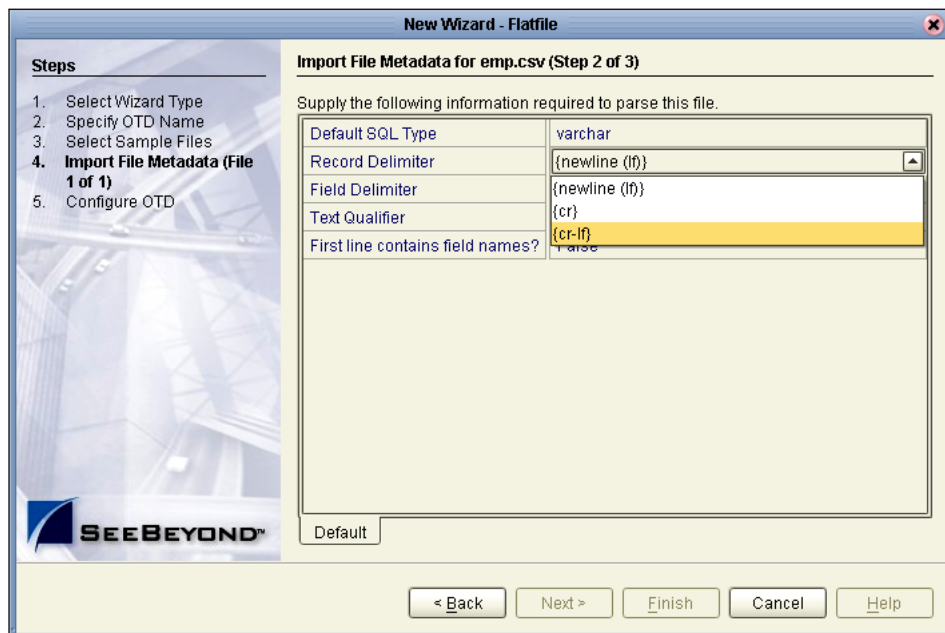
The following examples illustrate the use of each of these tools.

Figure 10 Default SQL Type



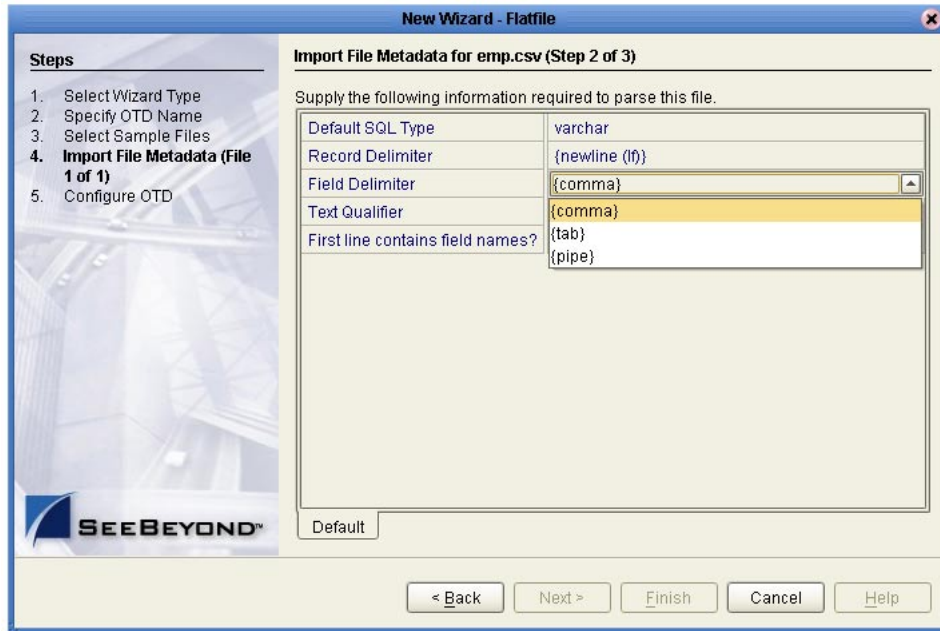
The default SQL type, Figure 10 above, is used for all elements in the flat file OTD structure unless a different type is specified by the end user in a subsequent panel.

Figure 11 Specify Record Delimiter



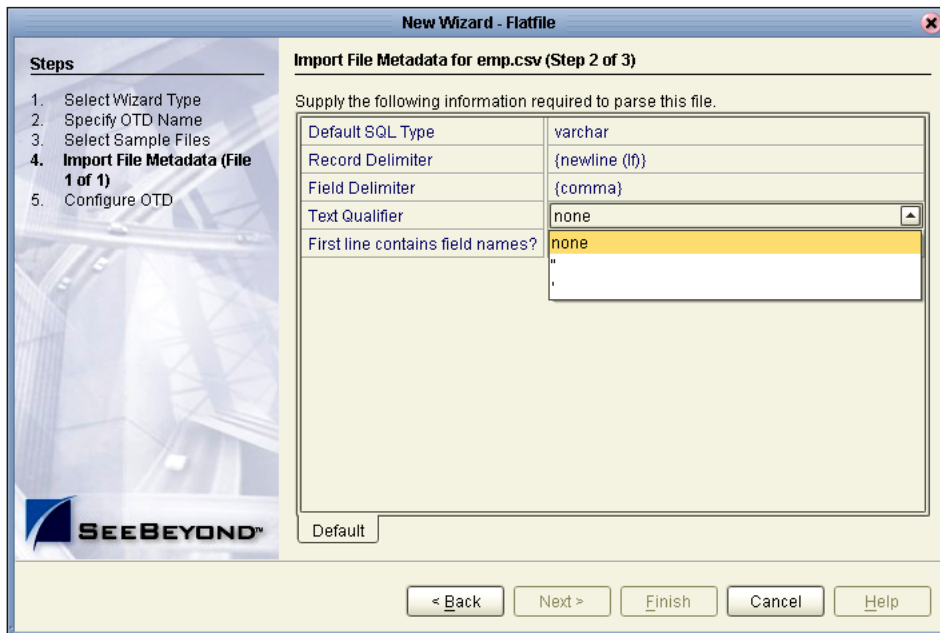
The Record Delimiter, Figure 11 above, allows you to specify how the various records in the flat file are physically separated from each other.

Figure 12 Specify Field Delimiter



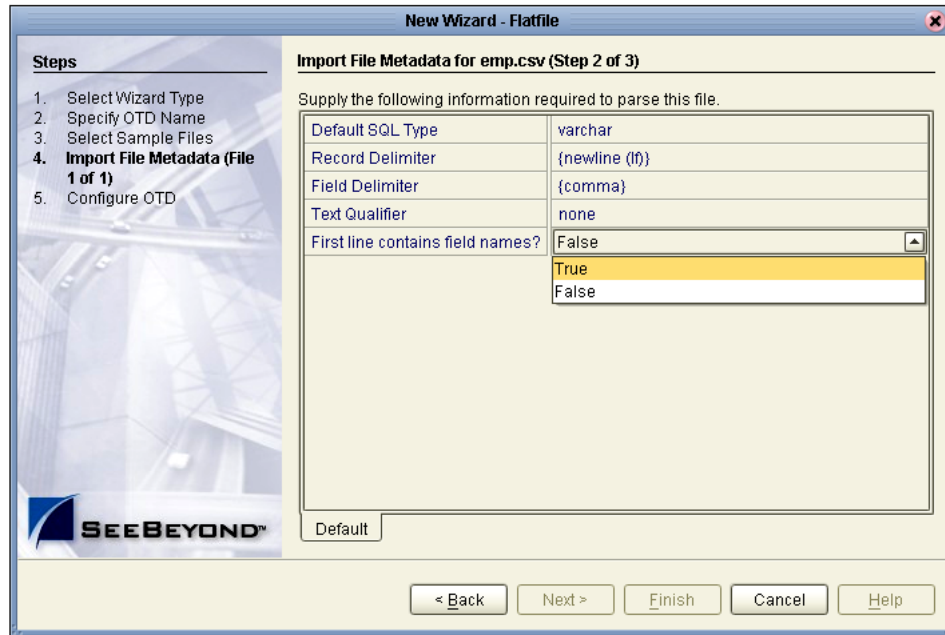
The Field Delimiter, Figure 12 above, specifies how the various elements (fields) in the flat file records are physically separated from each other. The following field delimiters are supported: comma, tab, and pipe (|).

Figure 13 Specify Text Qualifier



The Text Qualifier, Figure 13 above, explicitly specifies how eGate Integrator detects text fields. You can select double quote ("), single quote ('), or none.

Figure 14 First Line Contains



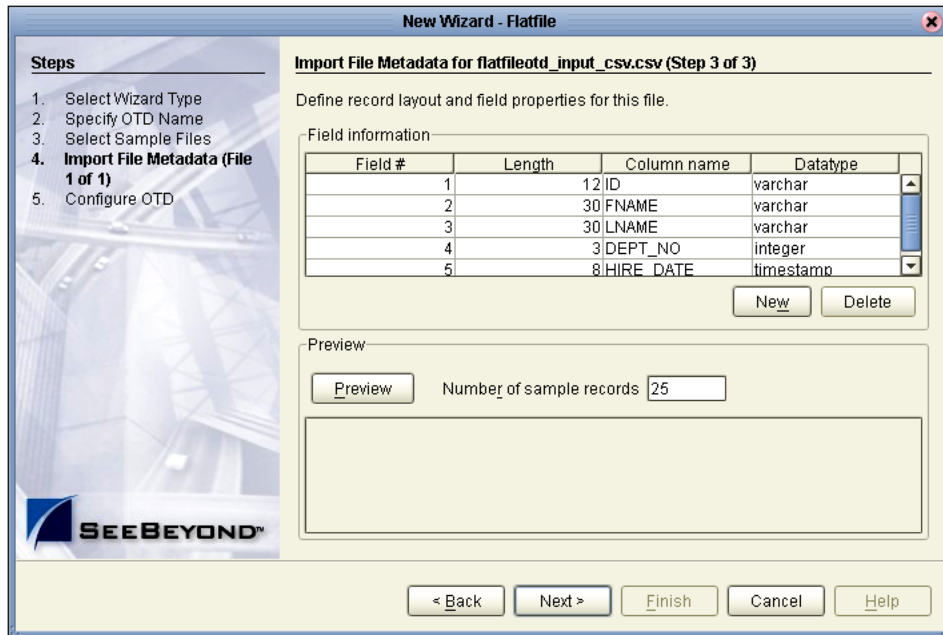
The “First line contains...” offers a True or False selection. See Figure 14 above. You can specify whether the selected flat file includes the names of its fields in its header row.

- True - the names specified in the header row will be used as element names of the new OTD.
- False - the eTL Integrator will dynamically assign initial names to the new OTD elements, which can be changed in the next panel.

Suggested OTD Record Structure

After the parsing specifications have been set, you are ready to define the record layout and field properties for your file. The system displays a suggested OTD record, but you can change the various field properties, including name, length, and data type.

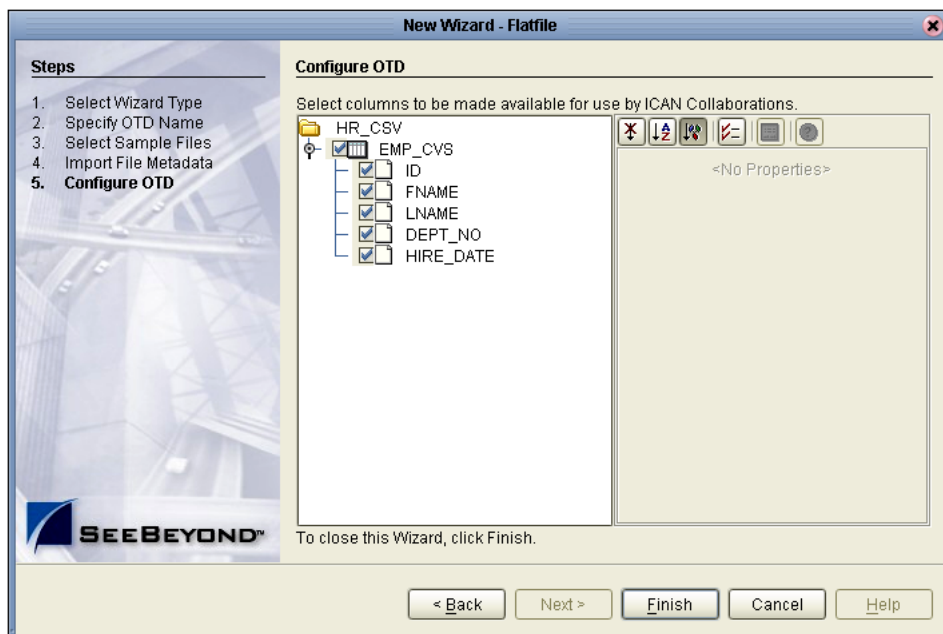
Figure 15 Define Record Layout and OTD Fields



The suggested OTD record properties are displayed, based on your file structure and your previous selections. In the previous Figure 15, the fields **Length**, **Column name** and **Datatype** are editable.

- 1 Click to highlight the field you want to edit.
- 2 Double-click to begin editing.

Figure 16 Select columns for ICAN Collaborations

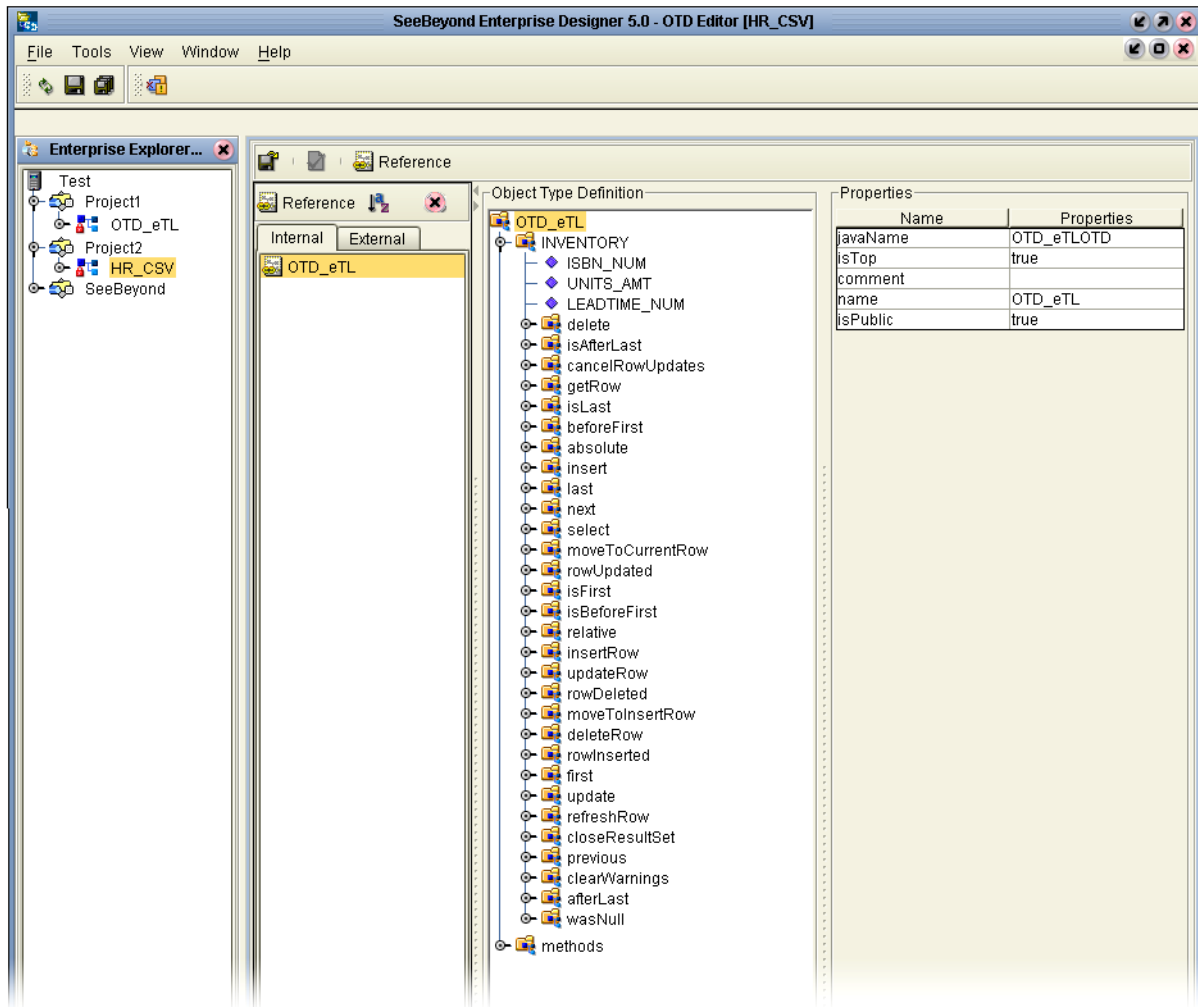


In this final step of the OTD Wizard process, you can select which elements of the new flat file OTD will be made available for use by ICAN Collaborations. See Figure 16 above.

- 1 Click Finish to create the new flat file OTD in the ICAN repository.

The new flat file OTD displays.

Figure 17 Verify New Flat File OTD



Chapter 3

Creating a Simple Project

The following scenario will guide you through the development of a simple project.

3.1 Sample Scenario

There are many ways to create a Project and link modules. This tutorial presents one scenario, using eTL Integrator.

This tutorial/scenario uses a business process that finds customer names in a file and assigns the value to the appropriate eInsight business process attribute as an input into the eTL process. The load orders activity executes an eTL Collaboration which has two data sources:

- 1 a database staging table of orders
- 2 a database table representing the inventory tables for all available products

The Collaboration has one data target (a database table with the exceptional orders which will be shipped directly from the publisher).

The eTL logic will read all orders from the staging table and insert some into the exceptional order table. Depending on the inventory information, individual orders go into specific tables. The eTL Collaboration will look at each entry in the orders table and compare the ordered quantity against current inventory. If the requested quantity is higher, the order is entered into the exceptional order ship table.

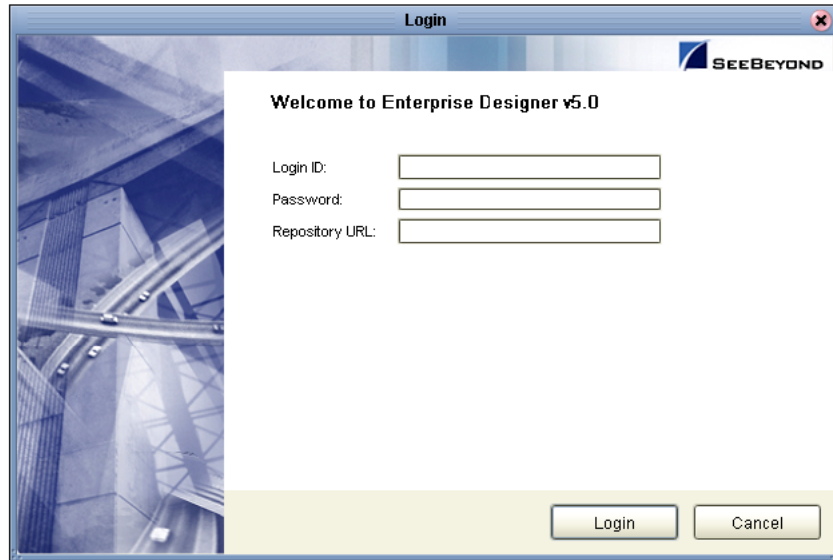
Begin by creating a Project in the Enterprise Designer.

Start the Enterprise Designer

- 1 Start the Enterprise Designer by executing `runed.bat` in your eGate folder (`c:\eGate50\edesigner\bin`), or use your enterprises' start up procedure.

The Enterprise Designer Login dialog box appears.

Figure 18 Enterprise Designer Login



- 2 Type your **Username** and **Password** and then click **Login** to start the Enterprise Designer.

Create and name a Project


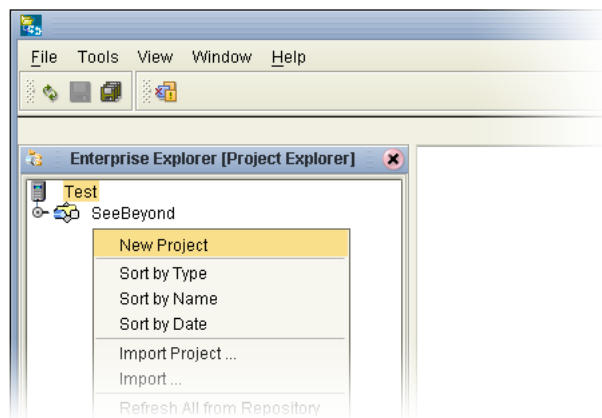
- 1 In the Enterprise Explorer pane of the enterprise Designer, right-click the Repository name (computer icon ) and then click **New Project**.

Figure 19 Create a Project



- 2 Type **Project_eTL** as the name for your project and press **Enter**.
The **Project_eTL** structure appears in the Explorer pane on the left side of the window.

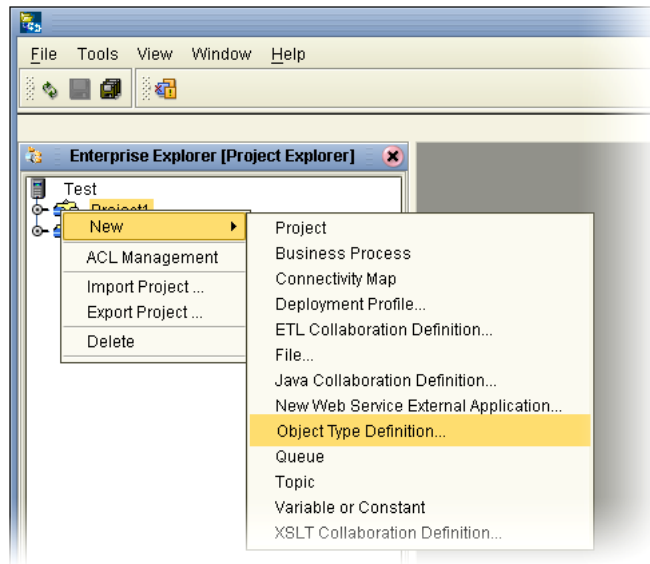
Create a New Object Type Definition

You will be creating definitions for database tables.

- 1 Right-click **Project_eTL**.

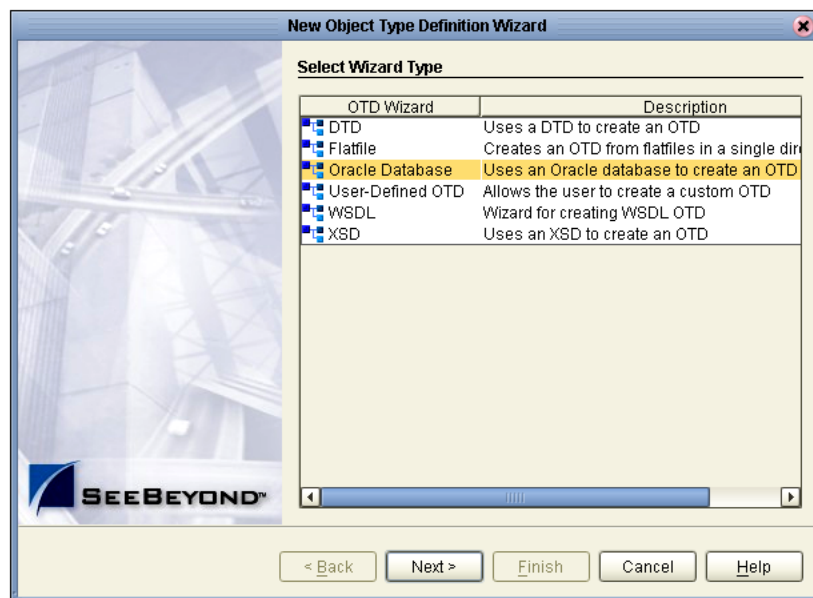
2 Click **New Object Type Definition**.

Figure 20 New Object Type Definition



The **New Object Type Definition Wizard** appears, displaying a list of tools to create OTDs.

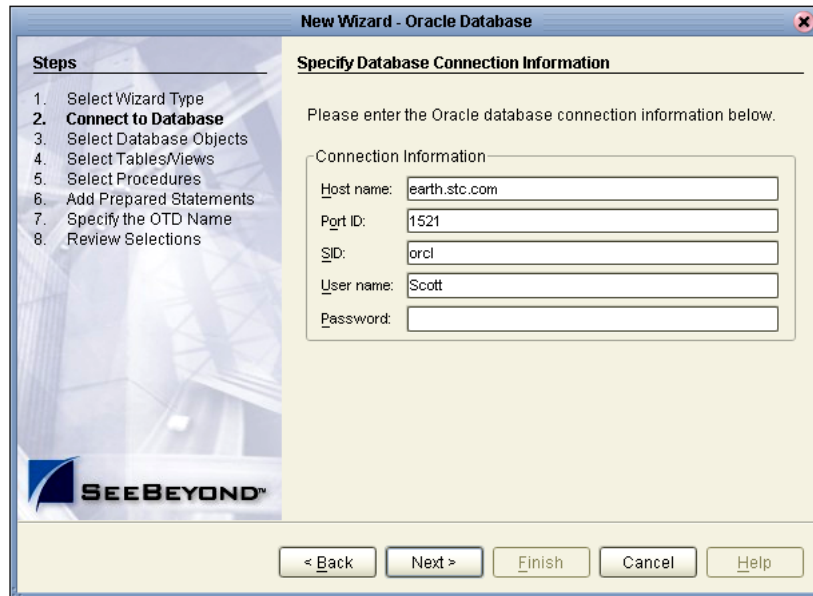
Figure 21 Select Wizard Type



3 Highlight **Oracle Database** and click **Next**.

The **Connect to Database** screen appears, allowing you to type connection information.

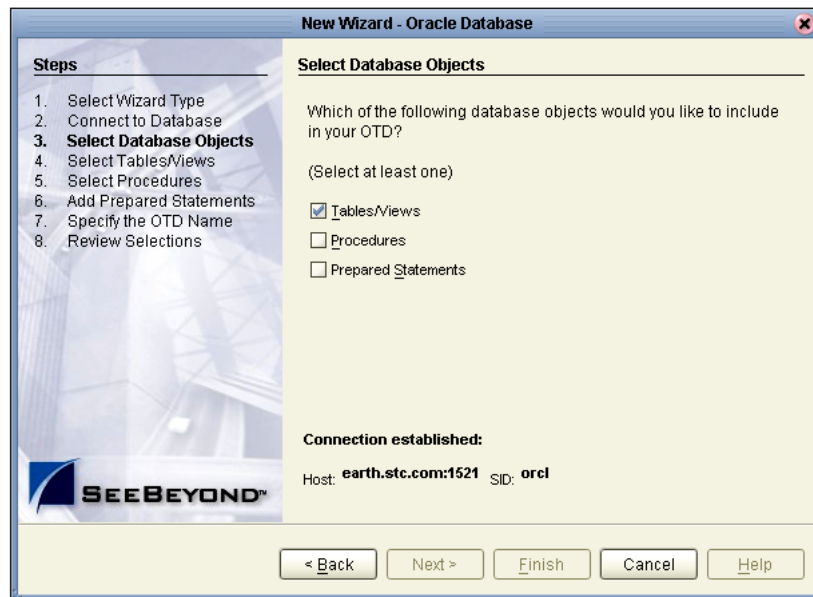
Figure 22 Database Connection Information



Note: The Port_ID is not the eGate port but rather is the database port number.

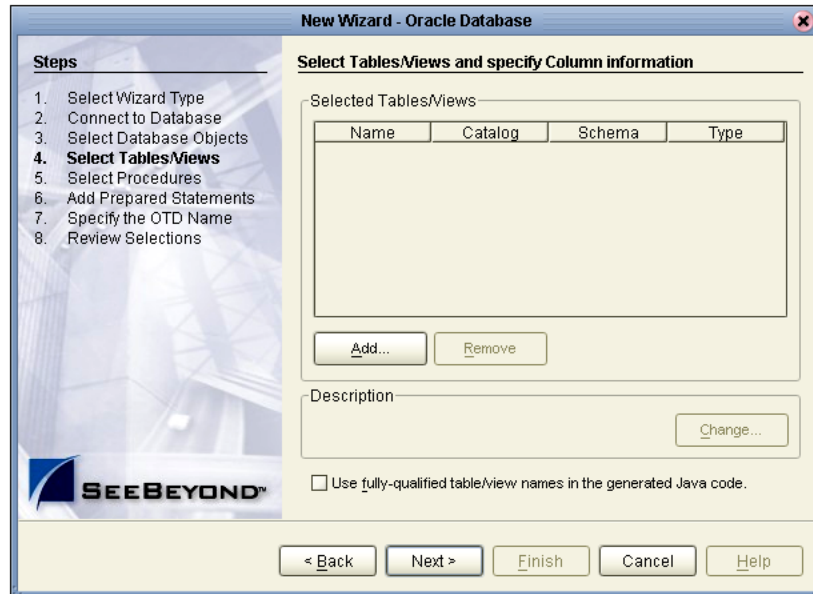
- 4 Type connection information and click **Next**.
The **Select Database Objects** window appears.

Figure 23 Select Database Objects



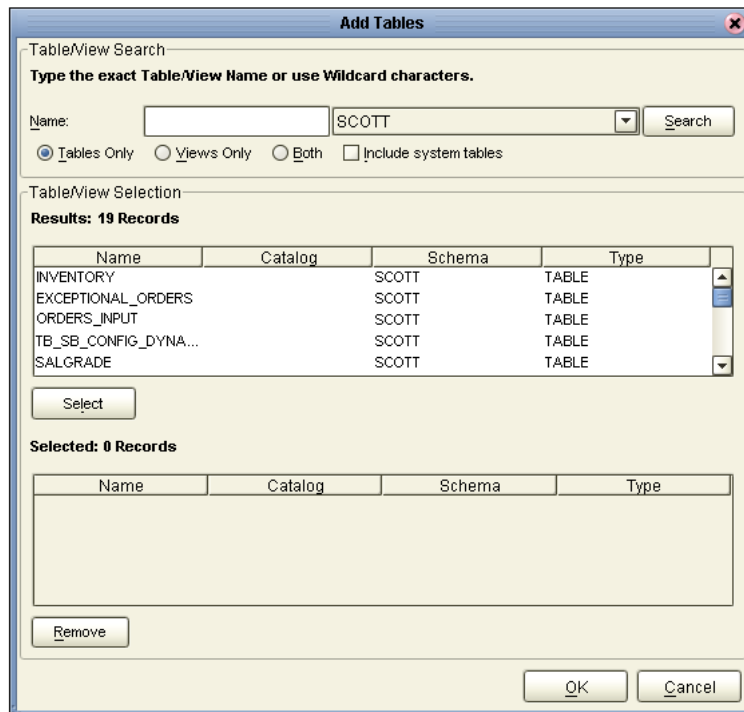
- 5 Check the **Tables/Views** box and click **Next**.
The **Select Tables/Views** window appears.

Figure 24 Select Tables/Views



- 6 Click the **Add** button.
The **Add Tables** window appears.

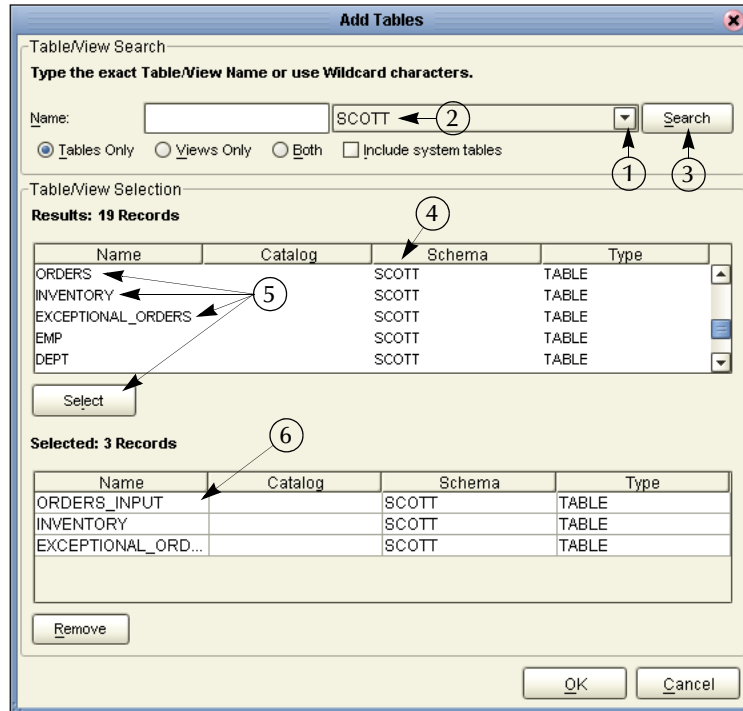
Figure 25 Add Tables



- 7 Click the **drop_down arrow** (number 1 in the following Figure 26).
- 8 Select **Scott** from the list (number 2 in the following Figure 26).

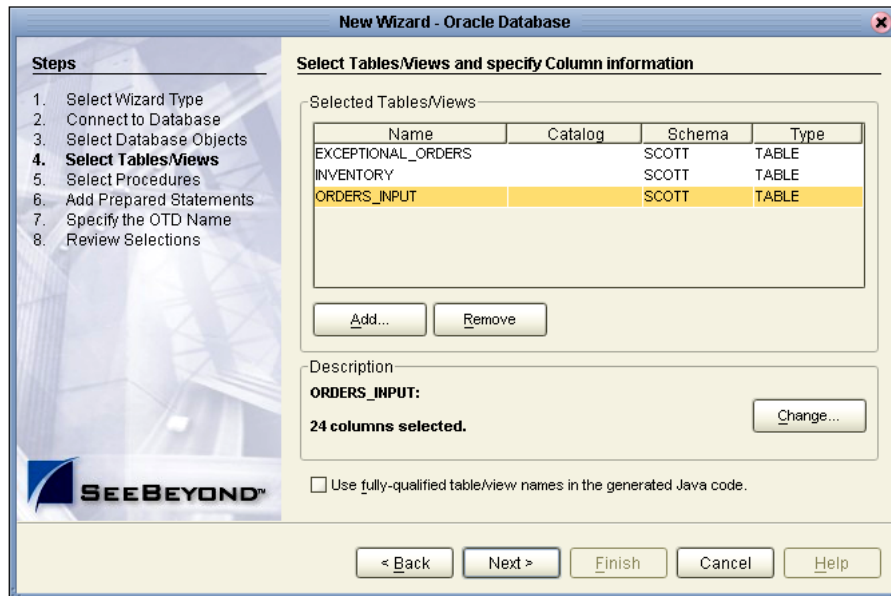
- 9 Click the **Search** button (number 3 in the following Figure 26).
A list of Scott's tables appears in the **Table/view Selection** list (number 4 in the following Figure 26).

Figure 26 Select Tables Using Search Tools



- 10 Next you will select the following tables (number 5 in the previous Figure 26):
 - Orders
 - Inventory
 - Exceptional_orders
- 11 After each selection click the **Select** button. (It will take a minute before the table name appears in the name list below.)
- 12 After you have finished your selections, and they appear in the list (number 6 in the previous Figure 26), click the **OK** button.
The **Selected Tables** window will appear.

Figure 27 Selected Tables

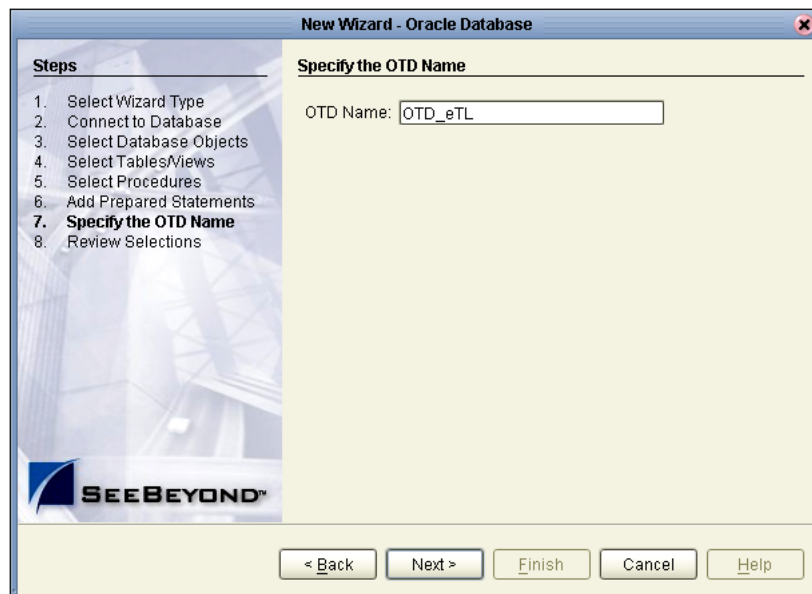


13 Click the **Next** button.

Specify OTD

Next you will specify the OTD name.

Figure 28 Specify OTD Names

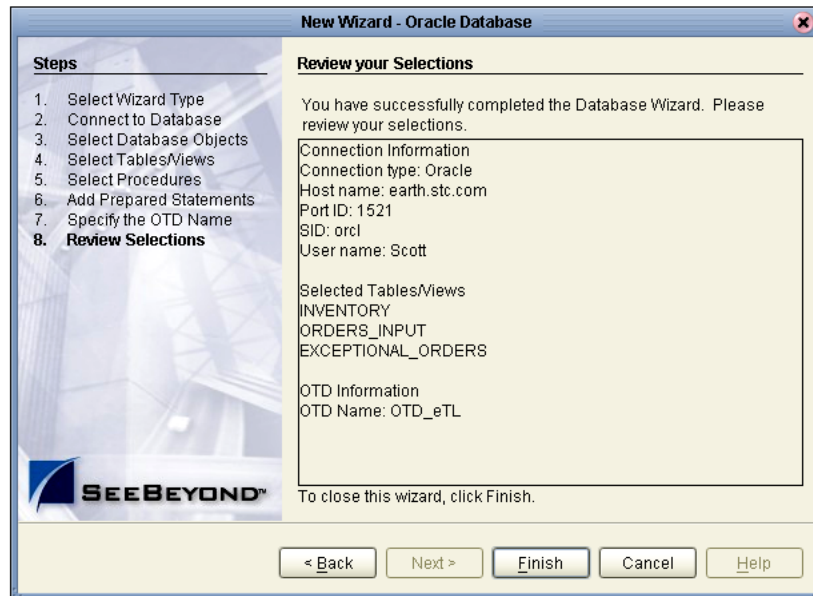


1 Type **OTD_eTL** to specify the OTD name.

2 Click **Next**.

The Review Selections window appears.

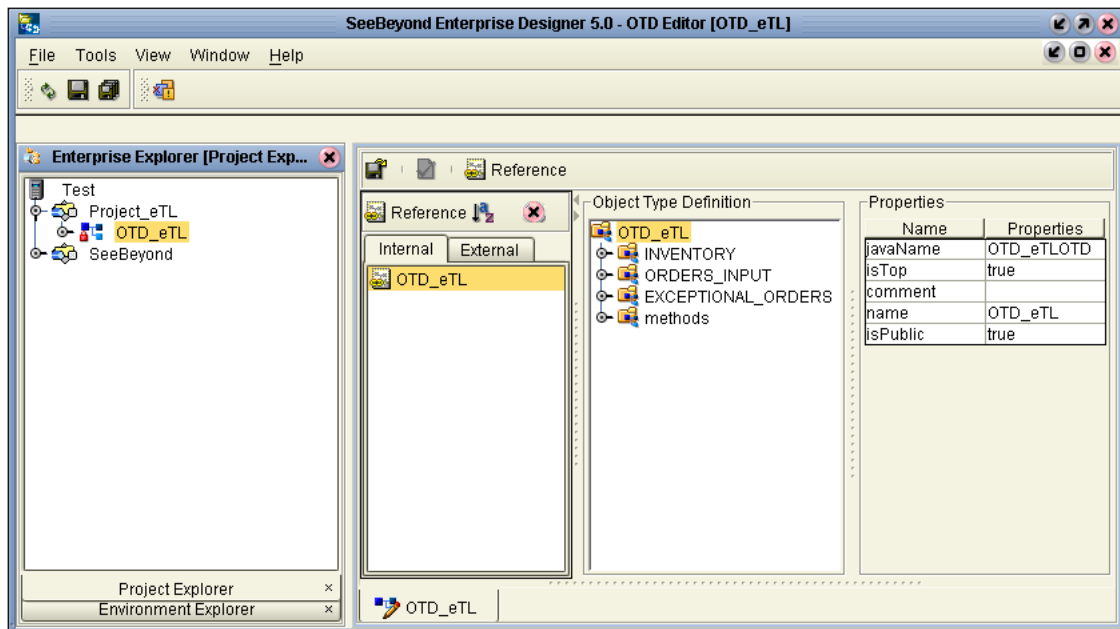
Figure 29 Review Selections



3 Click **Finish**.

The the wizard closes and the Enterprise Designer window reappears.

Figure 30 Enterprise Designer



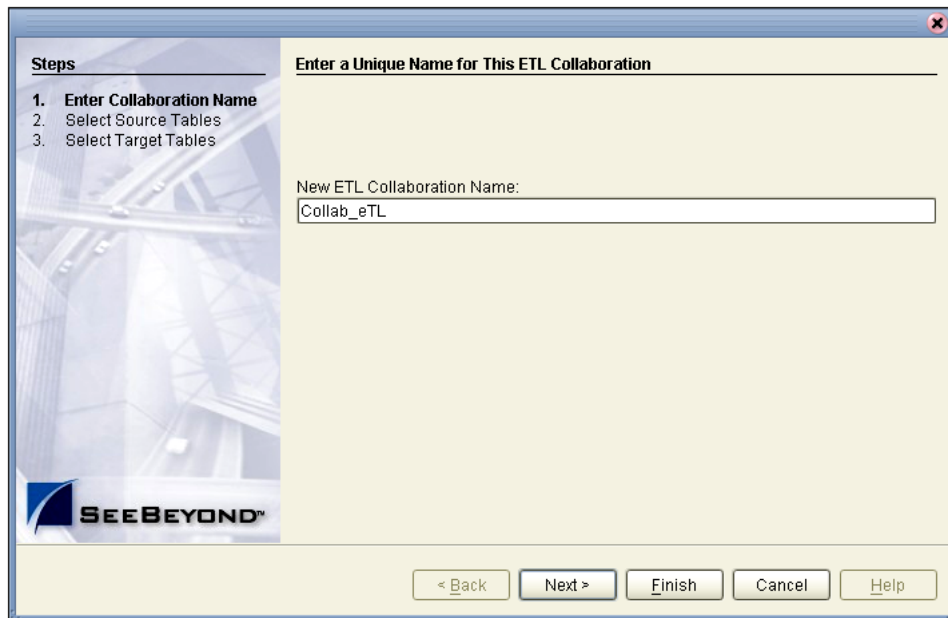
eTL Collaboration

Next you will configure your Collaborations for source and target tables.

1 Right-click **Project_eTL**.

2 Select **New eTL Collaboration**.

Figure 31 eTL Collaboration



3 Type **Collab_eTL**.

4 Click the **Next** button to select source tables (or click **Finish** to create a Collaboration with no source or target tables initially appearing on the designer pane).

Select Source Tables

You can select tables in a multiple table OTD.

5 Highlight your **OTD**.

6 Click the **right-arrow** button to complete your selection.

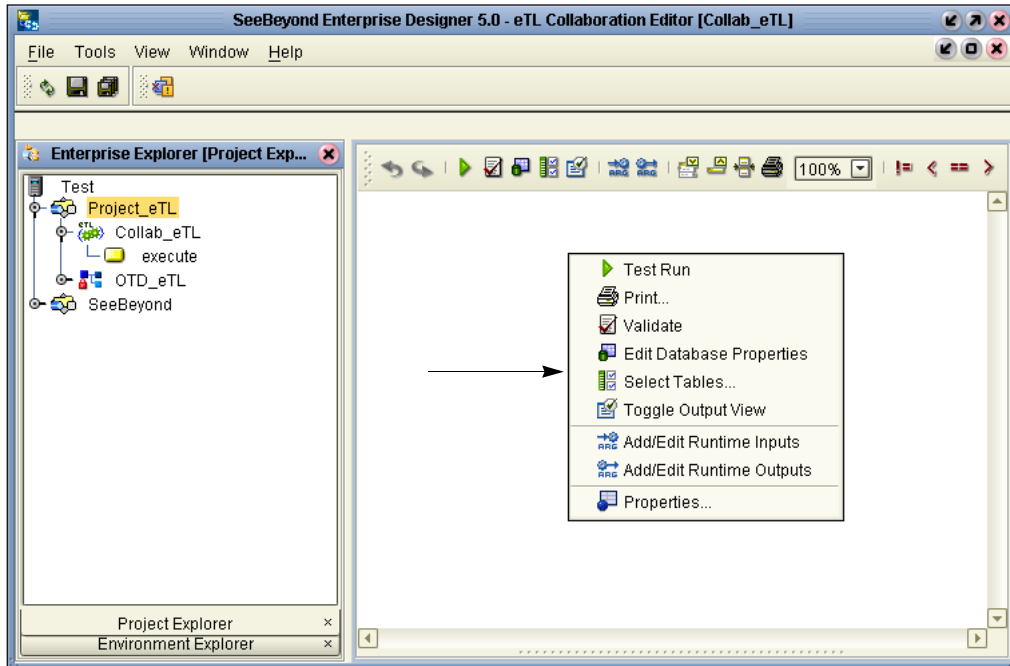
7 Click on the checkbox next to each item to be used as a source table.

8 Click **Next** (then repeat the previous steps to select the target tables).

9 Click **Window** on the menu and click **Close All**.

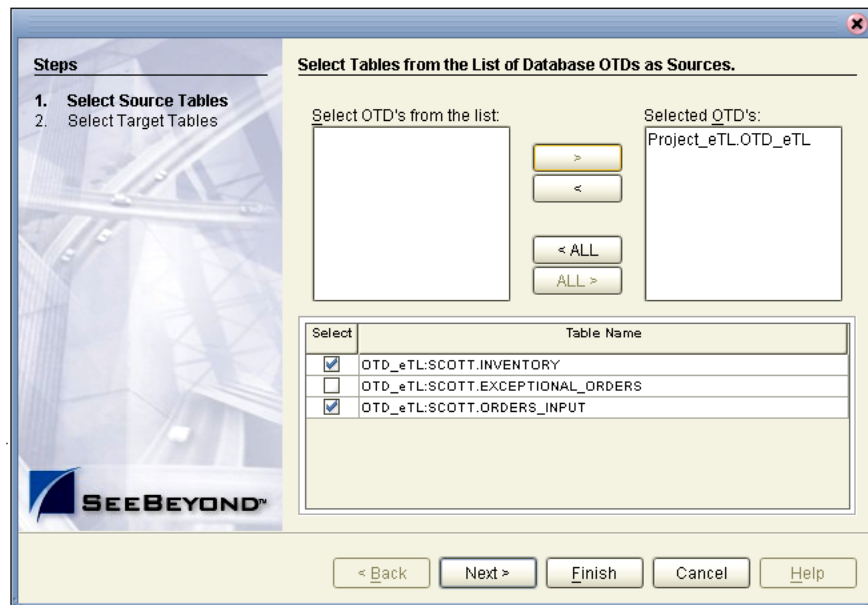
10 Double-click on **Collab_eTL**.

Figure 32 Empty Collaboration designer Pane



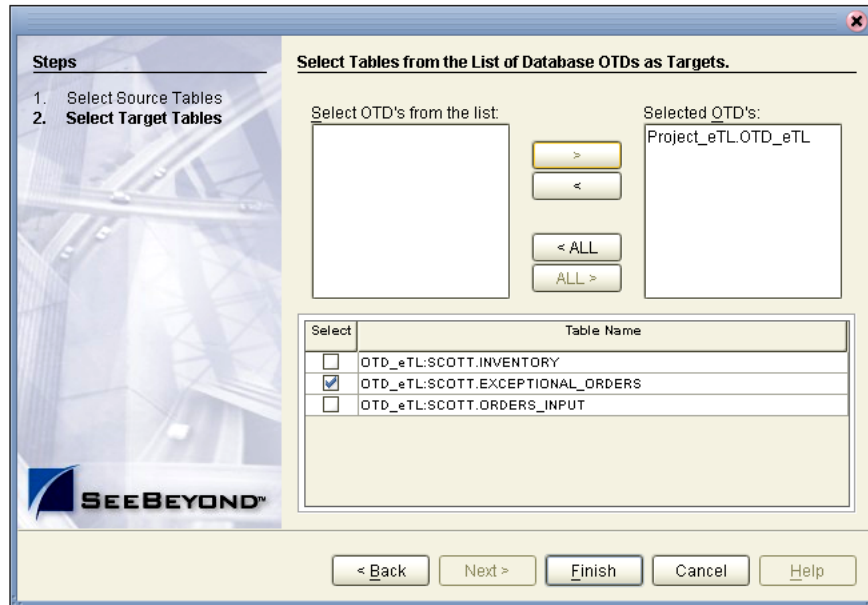
- 11 Right-click on **Collab_eTL**.
- 12 Click **Select Tables...** from the options.
- 13 Navigate to the **Select Source Tables** window (Figure 33).
- 14 Check the table boxes for the source tables, **Inventory** and **Orders_Input**.

Figure 33 Select Source Tables - Check Boxes



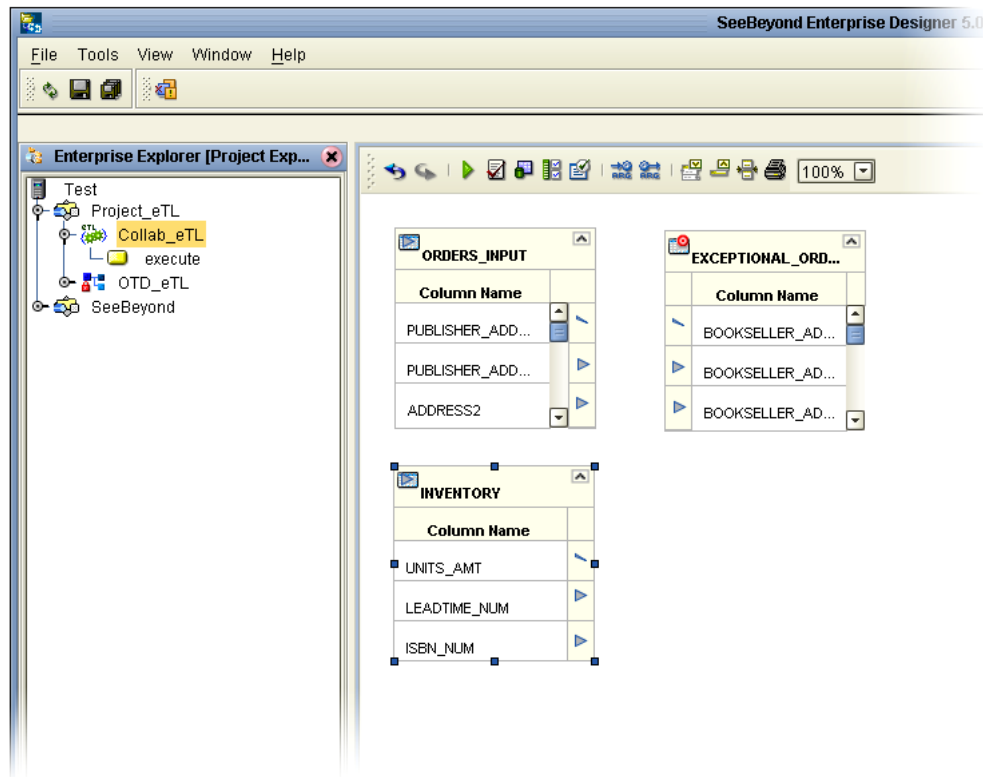
- 15 Click Next.
Select Target Table

Figure 34 Select Target Tables - Check Boxes



- 1 Check the box **Exceptional_Orders** (from the list in Figure 34).
 - 2 Click the **Finish** button.
- Your selected tables will appear in the Enterprise Designer.

Figure 35 Expand the Table Views



3 Click on the graphic “handles” and expand the view of the tables.

Place a Join Operator on the eTL Canvas


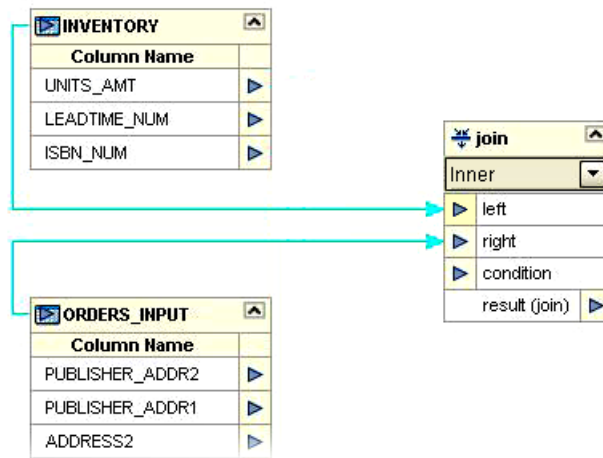
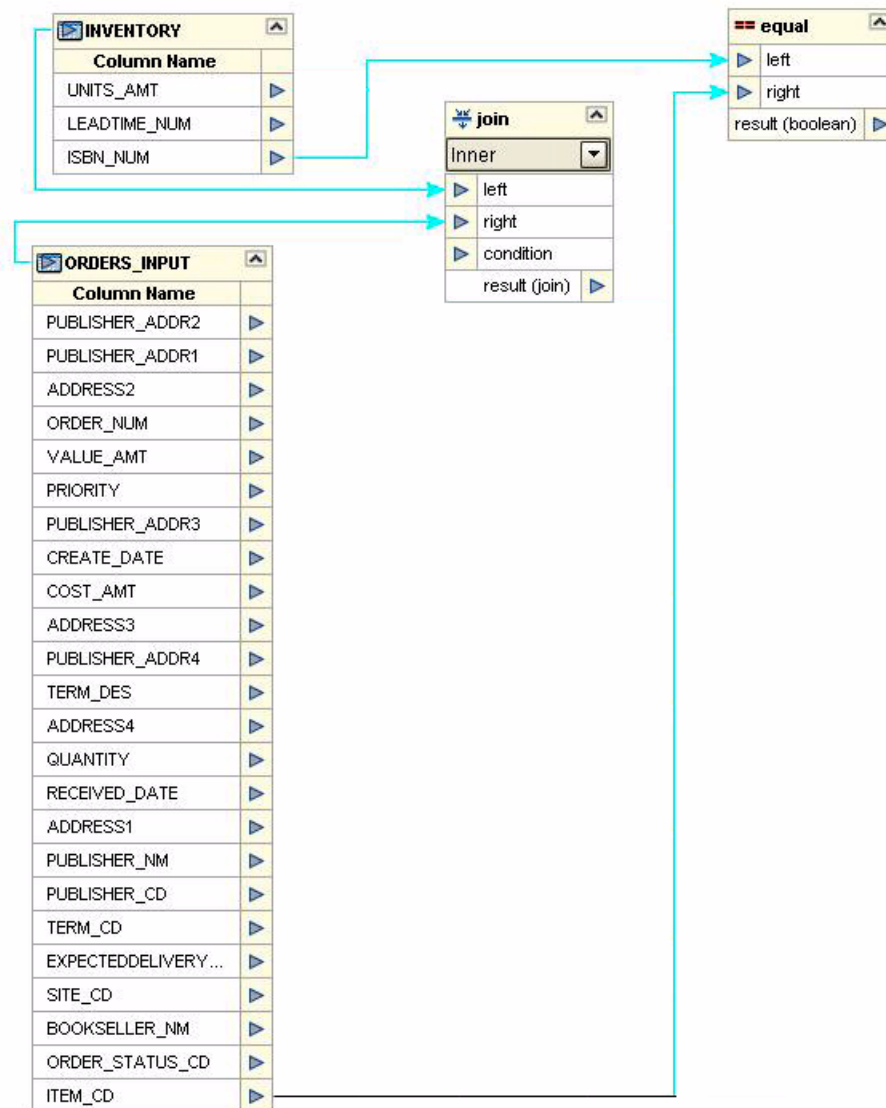
- 1 Click on the 'Join' icon  and drag the join operator to the designer pane.
- 2 Select the **Inventory** table and connect it to the 'left' property of the join operator.
- 3 Select the **Orders_Input** table and connect it to the 'right' property of the join operator.

Figure 36 Join Inventory and Orders_Input



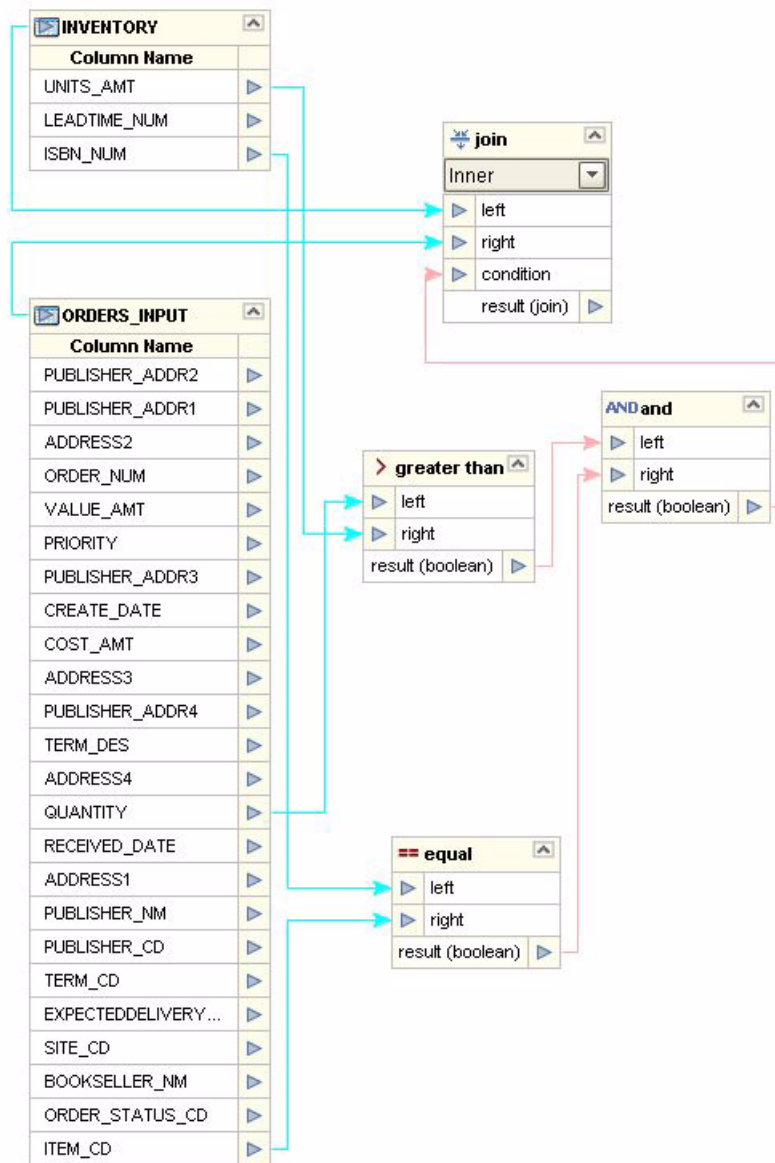
- 4 Place a new 'equal' operator on the eTL designer pane.
- 5 Select the column **Orders_input.item_CD** and connect it to the 'right' property of the 'equal' property of the 'equal' operator.
- 6 Select the **Inventory.ISBN_num** and connect it to the 'left' property of the 'equal' operator.

Figure 37 Equal Operator



- 7 Place a new 'greater than' operator on the eTL designer pane.
- 8 Select column **Orders_input.quantity** and connect it to the 'left' property of the 'greater than' operator.
- 9 Select column **Inventory.units_amt** and connect it to the 'right' property of the 'greater than' operator.
- 10 Place a new 'and' operator on the eTL designer pane.

Figure 38 Inventory and Orders_Input Map



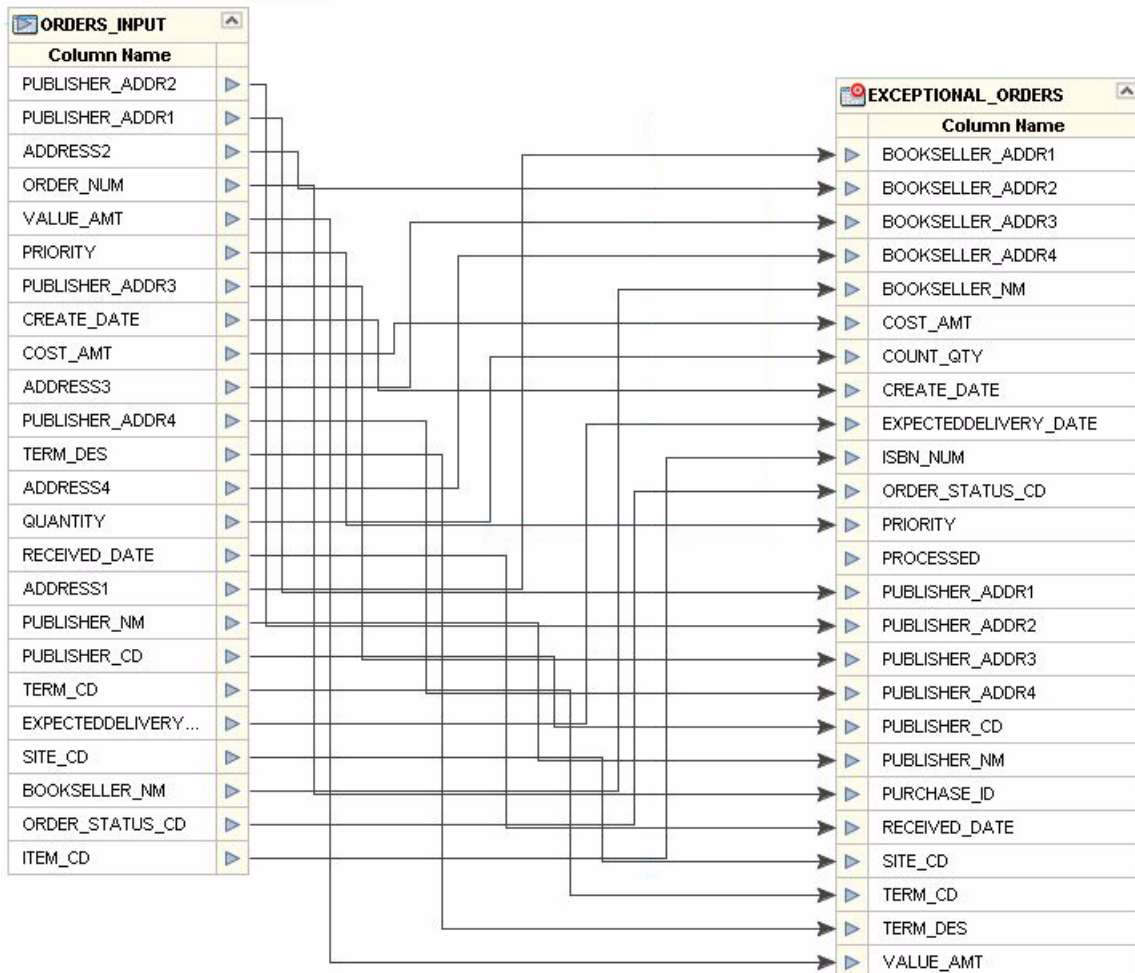
- 11 Select the 'result' property of the 'greater than' operator and connect it to the 'left' property of the 'and' operator.
- 12 Select the 'result' property of the 'equal' operator and connect it to the 'right' property of the 'and' operator.
- 13 Select the 'result property of the 'and' operator and connect it to the 'condition' property of the 'join' operator.

Map Target Table Columns

- 1 Map the following target table columns from the **Order_input** table.
The following are exceptions:

- ◆ Bookseller_addr (1-4_ by address (1-4)
- ◆ ISBN_num by Item_CD
- ◆ Purchase_ID by Order_num
- ◆ Count_Qty by Quantity

Figure 39 Map Source to Target




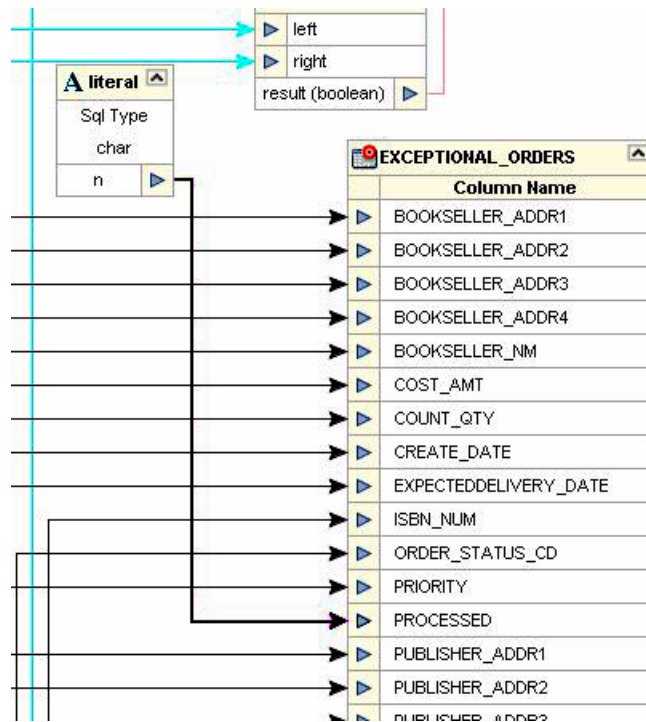
- 2 Place a new 'literal' operator  on the eTL designer pane (click on the icon and drag).
- 3 Enter value 'n' into the 'literal' operator.

Figure 40 Literal Operator



- 4 Connect the 'n' of the 'literal' object to the **Exceptional_orders.processed** column.
- 5 Place a new 'touppercase' operator on the eTL designer pane.

Figure 41 touppercase Operator



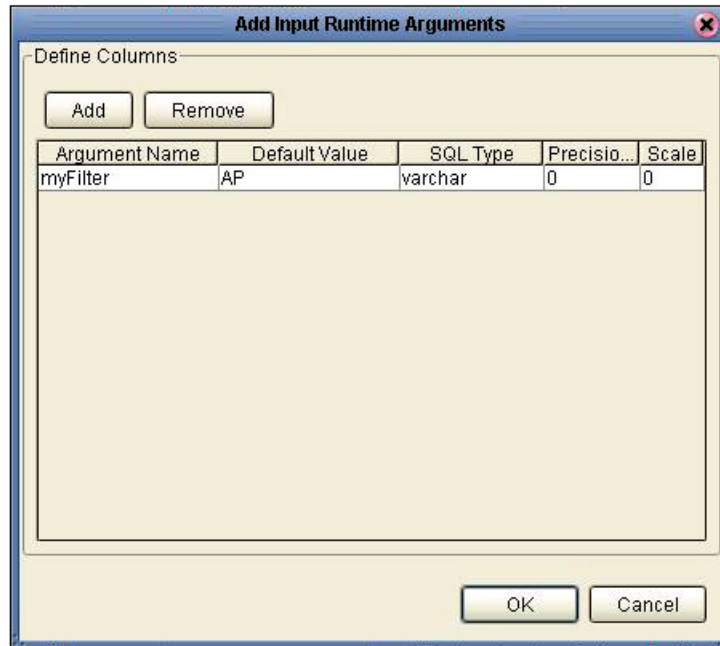
- 6 Select **Orders_input.bookseller_nm** and connect it to the 'string to convert' property of the 'touppercase' operator.

- 7 Connect the 'result' property of the 'toupper' operator to the **Exceptional_orders.bookseller_nm** column.

Set Runtime Input and Output

- 1 Click the **Add/Edit Runtime Inputs** button.

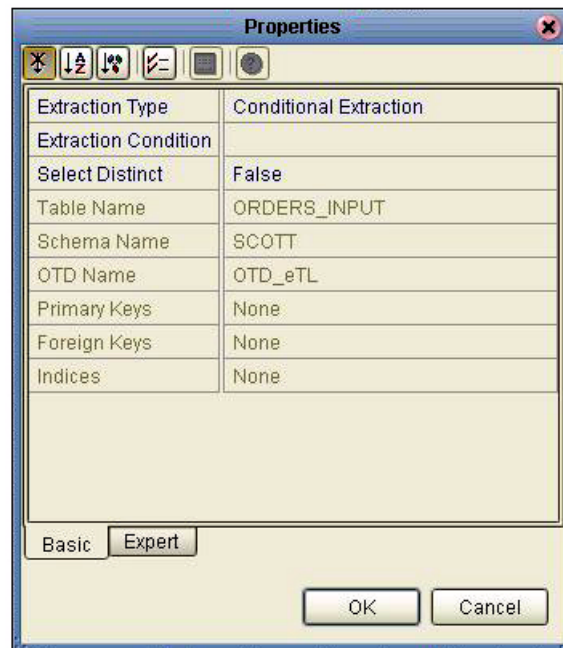
Figure 42 Runtime Inputs



- 2 Name the argument **myFilter**.
- 3 Enter **AP** as the default value.
- 4 Click **OK**.
- 5 Click the **Add/Edit Runtime Outputs** button (accept the default parameters).

Set Properties

Figure 43 Properties



You can define an extraction condition to filter extracted data to a subset.

Validate and Run

- 1 Click the **Validate** button to validate the eTL.
- 2 Click the **Save** button.
- 3 Click the **Run** button to execute the eTL Collaboration.
- 4 Right-click **Exceptional_orders** table and select **Show data**.

Figure 44 Show Data

BOOKSELLER_ADDR1	BOOKSELLER_A...	BOOKS...	BOOKSE...	BOOKSELLER_NM	COS.	COUNT_...	CREA...	EXPE...	ISBN_NUM	ORDER_...	PR...
789 La Sombra	Albuquerque	NM	87103	WHITE SANDS BOOKS	10	250	2002-...	2002-...	0-4545-2110-2	New	5
15 Atlantic Ave.	Cooperstown	NY	20202	BOOKS 'N BATS	8	400	2002-...	2002-...	0-4545-2210-1	New	3
9744 Pittsburgh Row	Pittsburgh	PA	11262	SCATTERGOODS	3	120	2003-...	2003-...	0-4545-2310-7	New	7
200 Campbell St.	River Oaks	MO	63121	SOUTHWEST BOOK CO.	5	20	2003-...	2003-...	0-4545-3221-2	New	5
119 Royal St.	New Orleans	LA	70130	BIG EASY BOOKS	9	400	2003-...	2003-...	0-4545-3421-6	New	2
301 Campbell Ave	Dallas	TX	21971	SOUTHWEST BOOKS	3	20	2003-...	2003-...	0-4545-4369-1	New	7

View the output data in the lower window. Check the database table **Exceptional_orders** to verify that the 'literal' has changed (use show data).

3.1.1. Using eTL With eInsight

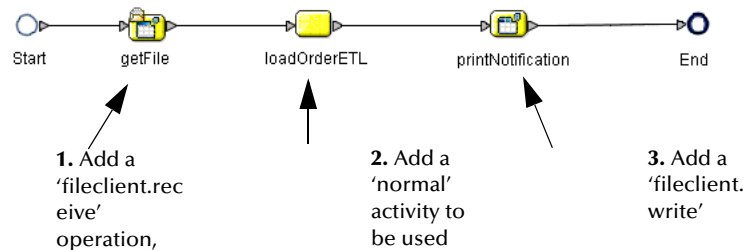
The eTL Integrator produces Collaborations that can be run as “stand alone” or scheduled and used within an eInsight business process. Like other Collaborations in the ICAN suite, eTL Integrator has implemented a Web Services framework to expose its operations and expected inputs/outputs.

Create a Basic eInsight Process

In the following scenario, assume that the ‘exceptional orders’ table has been successfully populated through a previous collaboration.

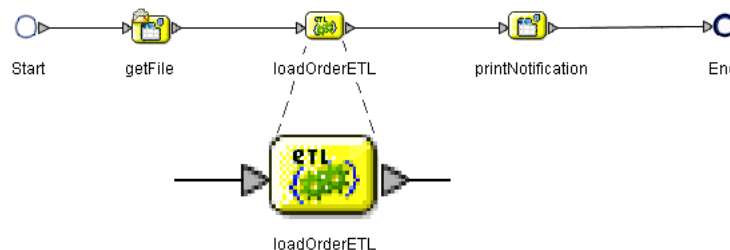
- 1 Create a basic eInsight process.

Figure 45 Basic Process



- 2 Add normal activity to the eInsight business process and rename it ‘LoadOrderETL.’
- 3 In the Project Explorer pane of the Enterprise Designer, select the ‘execute’ action of the eTL Collaboration that needs to be included into the eInsight business process and drag it on top of the ‘loadOrderETL’ process activity.

Figure 46 eTL Collaboration in eInsight



Note: The ‘loadOrderETL’ icon (in Figure 45) is replaced by the eTL Collaboration icon (Figure 46).

Note: You can drag and drop the execute object from the Project Explorer.

- 4 Double-click the ‘loadOrderETL’ icon to open the actual eTL Collaboration.

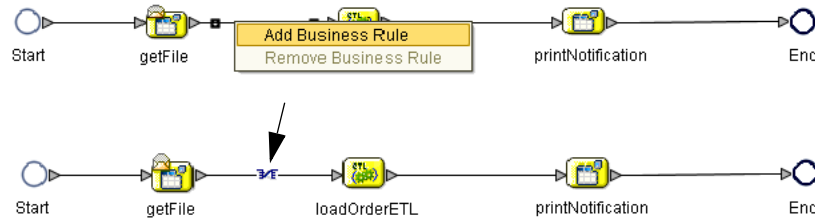
This allows you to view and modify the Collaboration.

Define Business Rule so eTL can Receive its Input

Map the business rule to enable eTL to receive input from a previous activity. In this scenario, the 'getFile' activity.

- 1 Define inbound mapping for the eTL Collaboration.

Figure 47 Map First Business Rule



- 2 Right-click the connection between 'getFile' and 'loadOrderETL' and select the 'Add Business Rule' option.

The business rule, represented by the Mapping symbol (referred to as 'M' in eInsight) appears.

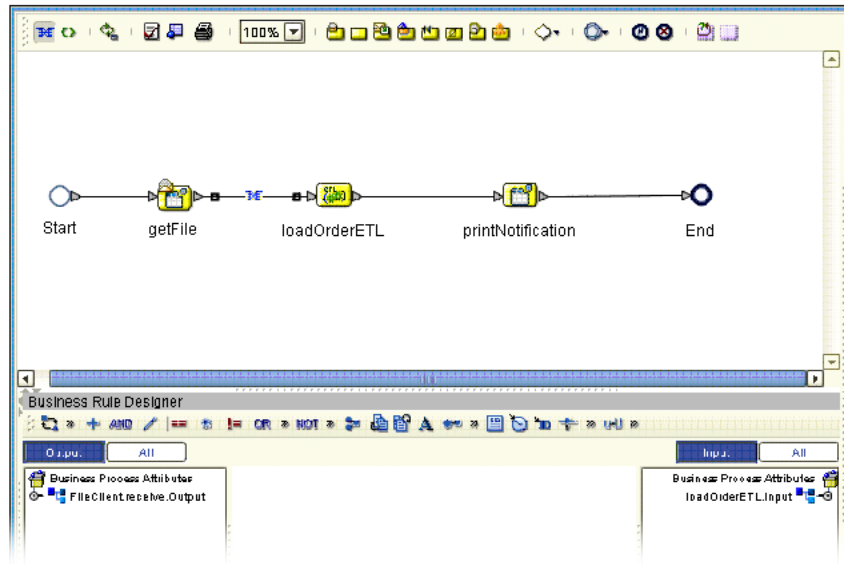
- 3 Click the 'M' icon in the eInsight toolbar to invoke the Transformation Designer.

Define Inbound Mapping for the eTL

- 1 Select the newly created business rule, (the 'M' of the 'getFile' 'LoadOrderETL' connection).
- 2 Select the 'M' icon of the eInsight toolbar.

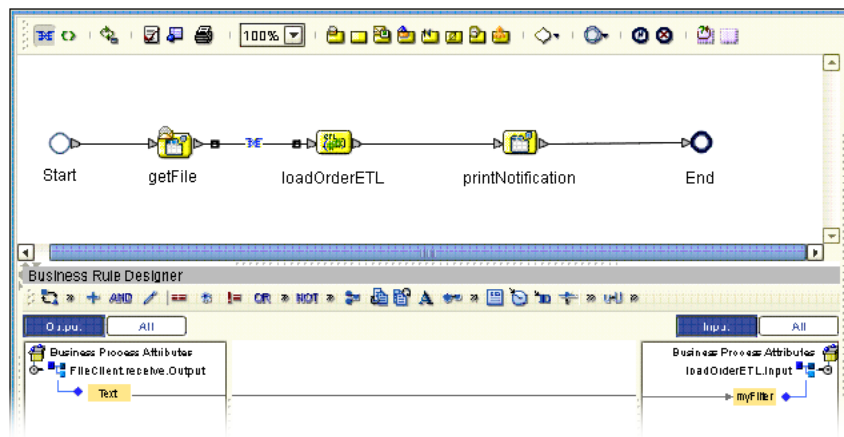
The Transformation Designer appears.

Figure 48 Define Inbound Mapping



- 3 Expand the 'FileClient' operator in the left pane of the Transformation Designer.
- 4 Expand the 'loadOrderETL' operator in the right pane of the Transformation Designer.
- 5 Connect 'text' data in the left pane to 'myFilter' data in the right pane. See the following Figure 49.

Figure 49 Define Mapping - Connect Text



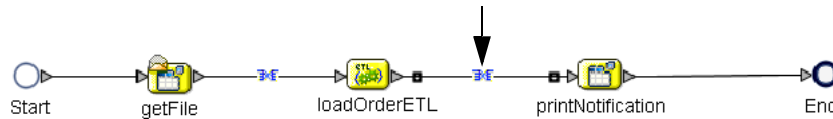
The eTL Collaboration's parameter ('myFilter' data element) is now supplied by the contents of the 'read' file. This embeds the eTL Collaboration into the business process.

Define Outbound Mapping for the eTL

We will now define a business rule (mapping) to make it possible for the eTL Collaboration to convey its output to a subsequent activity, in this case to 'printNotification.'

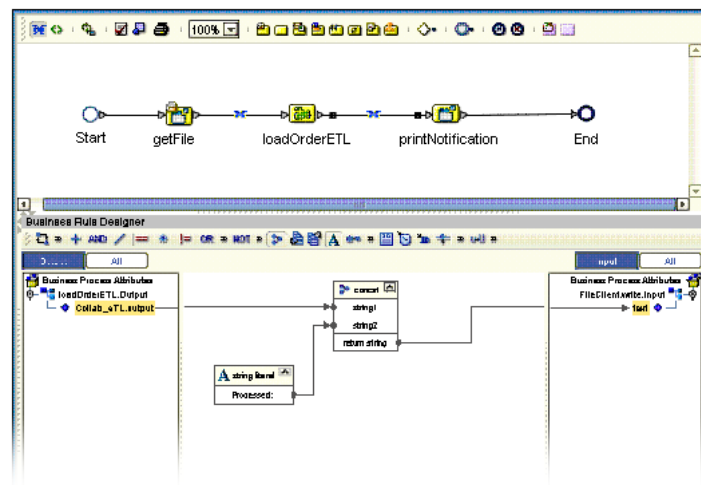
- 1 Right-click the connection between 'loadOrderETL' and 'printNotification,' and select the 'Add Business Rule' option. See Figure 50 below.

Figure 50 Add Outbound Business Rule



Map the business rules as shown in the following Figure 51.

Figure 51 Outbound Mapping



- 2 Expand the 'loadOrderETL' view in the left pane.
- 3 Expand the 'printNotification' view in the right pane.
- 4 Place a new string literal in the Business Rules Designer window.
- 5 Enter 'Processed' as a literal value and click OK.
- 6 Place a new 'concat' operator in the Business Rules Designer window.
- 7 Connect the output data element in the left pane to the 'string1' property of the new concat operator.
- 8 Connect the string literal operator to the 'string2' property of the concat operator.
- 9 Connect the 'return' property of the concat operator to the text element in the right pane.

Complete the eInsight Business Process

- 1 Click the 'Validate' option in the eInsight toolbar.
- 2 Click the 'Synchronize' option in the eInsight toolbar.
- 3 Click the 'Show Business' option in the eInsight toolbar to show the generated BPEL source code.

The business process exposes itself as a Web Service. You now have embedded your eTL Collaboration into a business process that can be invoked as a Web Service.

Chapter 4

Deployment

After a Project has been completed it must be deployed. This section explains that process.

4.1 General Instruction about Creating a Deployment Profile

Before you can map components, you must first create a Deployment Profile.

New Deployment Profile

A Deployment Profile contains information about how Project components will be deployed in an Environment, and it also maps components to the Environment.

- 1 Click the Project Explorer tab to return to the Project Explorer pane of your Project.
- 2 Right-click **Project1**, and then click **New Deployment Profile**.

After creating your deployment profile, you will drag and drop components into the server dialog boxes.

Drag and Drop Collaborations

- 1 Drag-and-drop your Collaborations into your Integration Server.
- 2 Drag-and-drop your topics into your Message Server.
- 3 Drag-and-drop your external files into your External File Servers.
- 4 After populating your Environment with all the available components, there are no components listed in the middle pane. All your components reside in the servers.
- 5 Deploy your Project

When you activate a Project a deployment file will be created that will later be picked up by the Logical Host. The Project you created with Enterprise Designer will be “code generated” into a package that will be passed to the Repository. The Logical Host will pick up this package from the Repository.

If you want to do a “hot deploy,” which means that the Logical Host will pick up the current package in real time while it is running, you could click **YES** in the Activate Dialog box when prompted “Do you wish to apply Logical host(s) immediately?” (See below).

Activate Environment

- 6 Click the **Activate** button. The Activation in Progress message appears.

Activating the Deployment Profile may take a few minutes.

Run the Bootstrap and Management Agent

The Bootstrap process executes your and begins the process of polling your input data. The Bootstrap process is performed from a command prompt. Bootstrap will pick up the deployment profile the first time it runs; after that you would redeploy.

The Bootstrap command is case sensitive on Windows.

To run the Bootstrap

- 1 Open a Windows command prompt as shown in the following. (Click **Start**; click **Run**; type `cmd`.)

- 2 Navigate to where you installed the logicalhost; for example, `eGate50\logicalhost\bootstrap\bin`, then type the following command:

```
CD \eGate50\logicalhost\bootstrap\bin
```

- 3 To start the Bootstrap process, type the following command:

```
bootstrap -e environment_name -l logicalhost_name  
-r repository_URL -i username -p password
```

environment_name is the name of your environment (for example, *TutorialTest*),

logicalhost_name is the name of your Logical Host (for example, *LogicalHost1*),

repository_URL is the full URL of your Repository including the Repository name (for example, *http://labserver:9000/Test*),

username is your user name, and

password is your password.

- 4 Press **Enter**.

The Bootstrap process takes a few minutes to execute. The Management Agent starts the components in the Project.

4.1.1. Verify the Output Data

When you click **OK** to apply your activation to the logical host(s), the Management Agent runs for a few minutes and creates your output.

4.2 Create Environment and Activate the Deployment Profile for eTL

This section explains how to create and activate the Deployment Profile. Before you can create the profile, you must create an Environment. Many of the steps in this section are

summarized. For a complete, simplified, end-to-end example of project development and deployment, refer to the *eGate Tutorial*.

Create an Environment and Logical Host

An Environment is a collection of physical resources and their configurations that are used to host Project components. An Environment contains logical hosts and external systems.

eTL Collaborations are either executed as ‘stand-alone or they are ‘invoked’ by an eInsight business process.

The following two scenarios will help explain the process.

4.2.1. Scenario 1

Deploy a Stand-alone Process

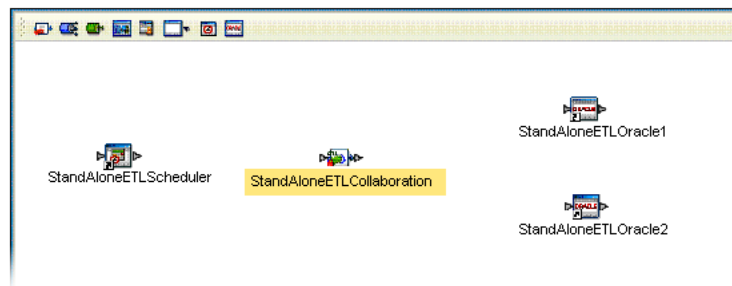
This is a continuation of [“Using eTL With eInsight” on page 43](#).

The following examples were created with an Oracle database.

- 1 Create a new Connectivity Map in Enterprise Designer.
- 2 Add two Oracle external application icons.

This will implement the database logic as defined in the eTL Collaboration.

Figure 52 Create Connectivity Map

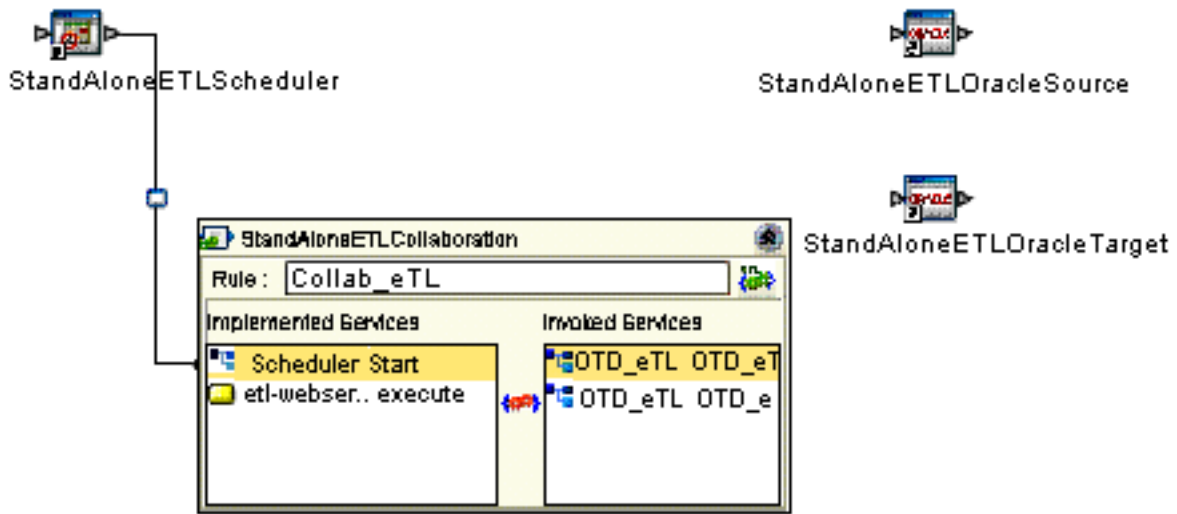


- 3 Add one Service icon.
- 4 From the Project Explorer, drag the eTL Collaboration into the second service icon.
- 5 Add a Scheduler icon.

Define the eTL Scheduler

- 1 Double-click the eTL Collaboration Service icon.

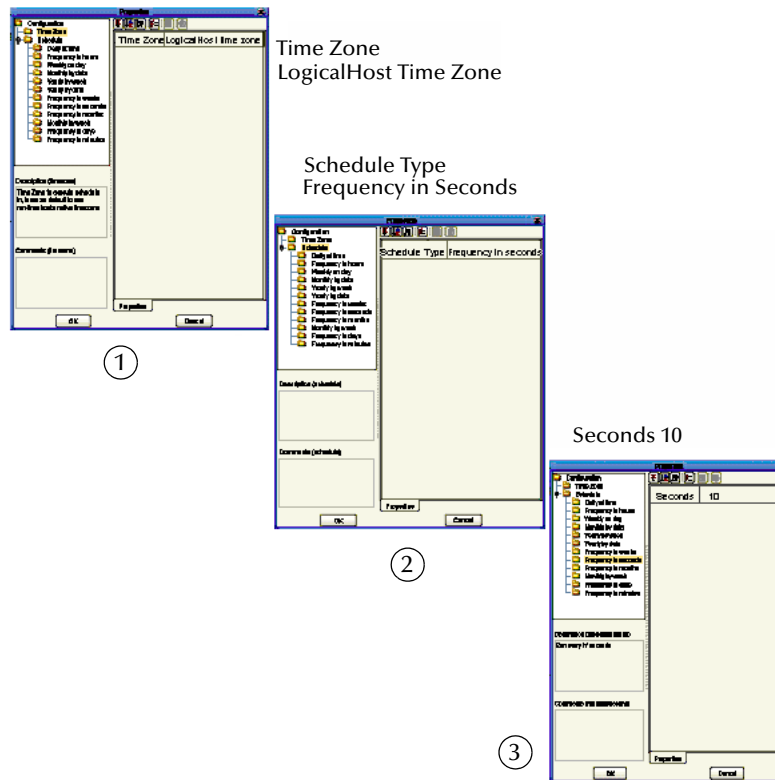
Figure 53 Define eTL Scheduler



- 2 Connect the Scheduler Service from the left pane of the eTL Collaboration to the Scheduler icon.
- 3 Double-click the new connection (line).

Its property sheet displays. There are many options available; this scenario covers only a few.

Figure 54 Define Schedule Criteria



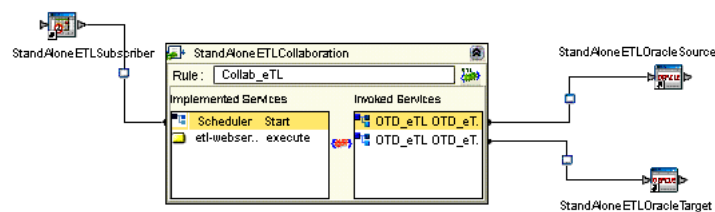
- 4 Select a timezone, where the LogicalHost resides, from the dropdown list. (See graphic 1).
- 5 Select a schedule type (interval type). For example, Daily at time, Frequency in seconds, or Frequency in hours. (See graphic 2.)
- 6 Select an interval value such as, midnight, Monday 8 AM PST, every hour or every 10 seconds. (See graphic 3.)

Connect Source and Target

Resume mapping the Connectivity map.

- 1 Double-click the eTL Collaboration Service icon, if not already open.

Figure 55 Connect Source and Target



- 2 Connect the source Oracle OTD Service to the source external Oracle system icon.
- 3 Connect the target Oracle OTD Service to the target external Oracle system icon.
- 4 Double-click the connection between the source Oracle OTD Service and the source external Oracle system icon.

The connection's property sheet displays.

- 5 Select the Outbound Oracle eWay option (because this is an outbound operation for the eTL Collaboration).
- 6 Click the OK button.
- 7 Double-click the connection between the target Oracle OTD Service and the target external Oracle system icon.

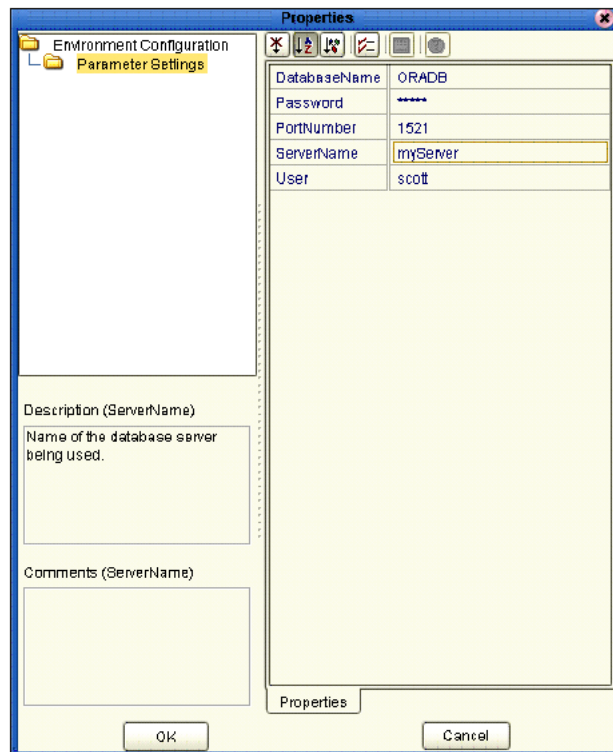
The connection's property sheet displays.

- 8 Select the Outbound Oracle eWay option.
- 9 Click the OK button.

The connection property sheet is displayed.

Note: *The environment properties of the source database are determined at the time the solution is activated via a deployment profile; therefore, you don't have to specify the actual database name here.*

Figure 56 Properties



- 10 Define the Oracle external system if not already done.

Create an Environment

This involves the following steps:

- 1 Switch to the Environment Explorer view and create a new environment.
- 2 Select the Oracle External system.
- 3 Enter the property values.

Note: *The Oracle External system will be used in the Deployment profile.*

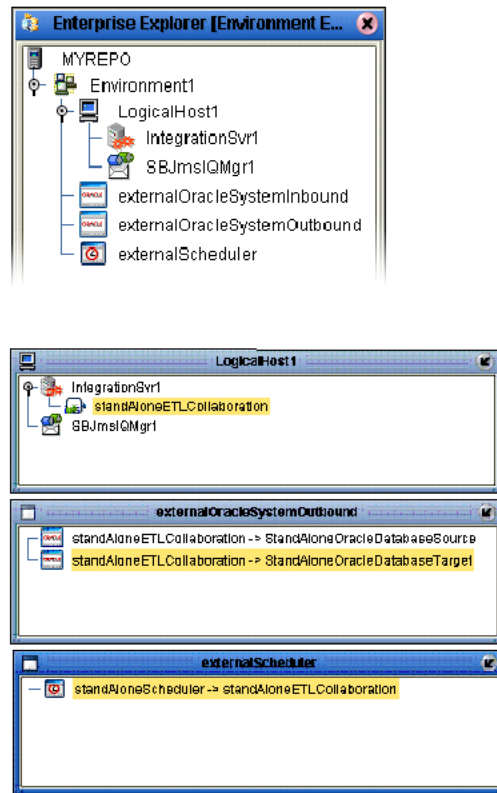
Create a Deployment Profile for Scenario 1

Prerequisites:

- Create an environment
 - Create at least one logical host
 - Create at least one Integration server within the logical host
 - Create at least one Oracle external system
- 1 From the Project Explorer of the Enterprise Designer, create a new deployment profile.
 - 2 Drag the eTL Collaboration Service to an Integration server.
 - 3 Drag the Oracle Services to an external Oracle system.

See the following Figure 57.

Figure 57 Environment - Deployment Profile



- 4 Start your logical host, if not already running.
- 5 Activate your deployment.
- 6 Verify the results.

4.2.2. Scenario 2

Deploy as an Invoked Service

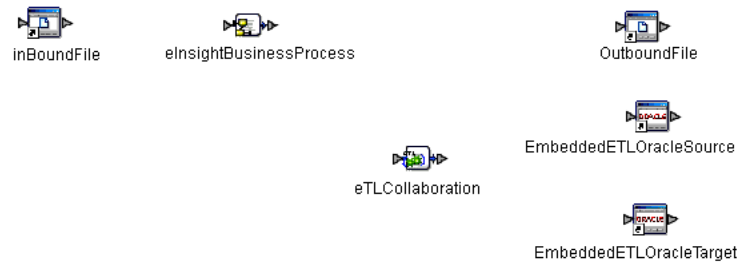
The example illustrated in this section explains how to invoke a service from an eInsight business process.

The following examples were created with an Oracle database.

- 1 Create a new Connectivity Map in Enterprise Designer.
- 2 Add two external application icons (inbound and outbound).
- 3 Add two Oracle external application icons.

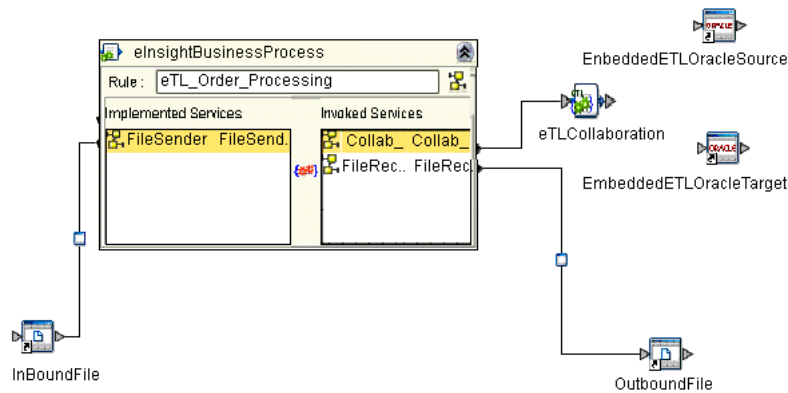
This will implement the database logic as defined in the eTL Collaboration.

Figure 58 Connectivity Map for an Invoked Service



- 4 Add two Service icons.
- 5 From the Project Explorer of the Enterprise Designer, drag the eTL business processes into the first Service icon.
- 6 From the Project Explorer of the Enterprise Designer, drag the eTL Collaboration into the second Service icon.

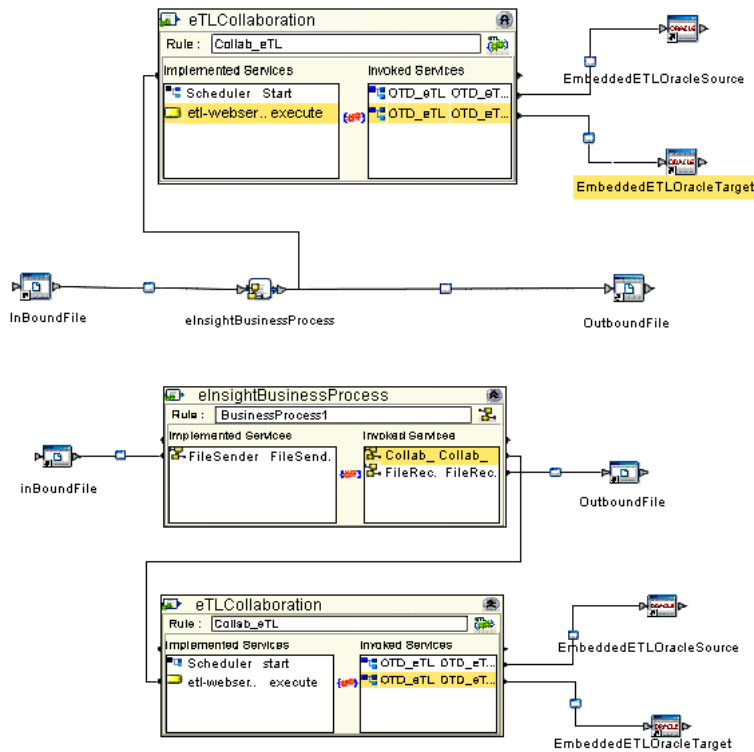
Figure 59 Connect Inbound and Outbound Services



Double-click the eInsight business process Service icon.

- 7 Connect the implemented Service, in the left pane, File Sender, to the inBoundFile icon.
- 8 Connect the eTL Collaboration invoked Service to the eTL Collaboration icon.
- 9 Connect the outbound file Service to the OutboundFile icon.

Figure 60 Connect to External Services



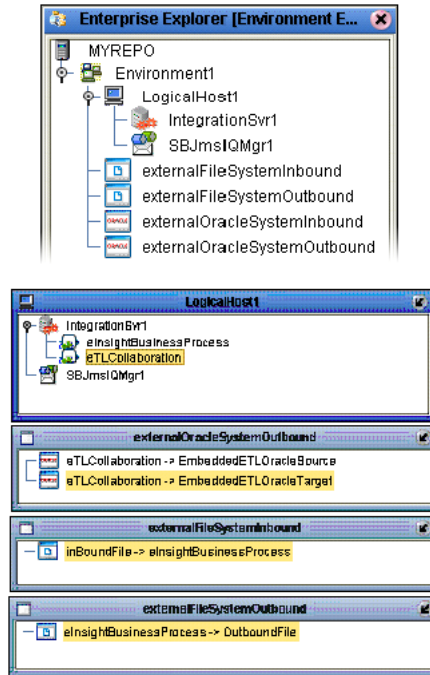
- 10 Connect the Oracle source OTD to the Oracle source external system.
- 11 Connect the Oracle source OTD to the Oracle target external system.

Create a Deployment Profile for Scenario 2

Prerequisites:

- Create an environment
 - Create at least one logical host
 - Create at least one Integration server within the logical host
 - Create at least one Oracle external system
 - Create at least one File external system
- 1 From the Project Explorer of the Enterprise Designer, create a new deployment profile.

Figure 61 Environment Servers



- 2 Drag the eInsight business process Service into the Integration server.
- 3 Drag the eTL Collaboration Service to an Integration server of a logical host.
- 4 Drag the Oracle Services to an external Oracle system.
- 5 Drag the (flat) file (read and write) Services to an external file system.
- 6 Start your logical host, if not already running.
- 7 Activate your deployment.
- 8 Verify the results.

Glossary

The following terms, although not all inclusive, are common to ETL.

5.1 ETL Terms

Common ETL terminology.

BI	Business integration (also Business Intelligence).
CRM	Customer Relations Management
Data Cleansing	Data must be cleansed of errors in structure and content before it is useful in data warehousing and integration; this means transforming data for accurate and effective use in a database or data management system by cleansing “dirty” or redundant data.
Data Dictionary	Defines the organization of a database and lists all files in the database, the number of records in each file, and the names and types of each field. The data dictionary is often hidden from end users. Although the dictionary doesn’t contain actual data, it does contain essential information for managing the database.
Data Integrity	Refers to the accuracy and validity of data. Data integrity can be compromised in many ways, including human error through data entry, or through faulty logic in programming. Computer viruses, software bugs and many other factors can also compromise data integrity.
Data Mapping	In relational databases (RDBMSs) data mapping is the relationship and data flow between source and target objects. Mapping involves structuring the relationship between source and target objects.
Data Mart	A smaller, focused, database designed to help managers make business decisions. (A data warehouse is a larger, enterprise, database(s).)
Data Mining	Used to synthesize or isolate unique data patterns to predict future behaviors or to filter data to select patterns that help discover previously unknown relationships among data. Commonly used by marketers who acquire and distill consumer information.
Data Transformation	Data transformation is necessary after extracting data from legacy data formats, or any format that requires cleansing. Data is transformed for efficient use for Business-to-Business Enterprise Data Integration.
Data Warehouse	A copy or view of enterprise transaction data (sometimes non-transaction data) that is used for reporting. The data is often summarized and always structured for queries and analysis.

Dimension Table	Dimension tables describe the business entities of an enterprise; also called lookup or reference tables.
Dirty Data	Dirty data contains, but is not limited to, incorrect data including spelling errors, punctuation errors, incorrect data referencing, incomplete, inconsistent, outdated, and redundant data.
Drill Down	To move from summary to more detailed data by “drilling down” to get it. In database terminology this might mean starting with a general category and drilling down to a specific field in a record.
EPR	Enterprise Resource Management
ETL	Extract, Transform, Load. Extract is the process of reading data from a source database and extracting the desired subset of data. Transform is the process of converting the extracted data from its previous form into the desired form. Load is the process of writing the data into a larger database.
Extraction	Data are extracted from a source using software tools. This first step in ETL initially “gets” the data.
Fact Table	A fact table typically contains two types of columns: those containing facts and those that contain foreign keys to dimension tables. Fact tables contain detail facts and/or summary facts.
Join	Matches records, which are joined by a common field, in two tables in a relational database. Often part of a Select query.
Metadata	“Data about data.” Metadata describes “how,” “when,” and “who” about structure and format, of a particular set of data. ETL tools are used to generate and maintain a central metadata repository.
Non-normalized Data	Non-normalized data cannot be cross-referenced accurately, if at all, and causes manageability issues. Non-normalized data may be converted to normalized data.
Normalized Data	Normalization is a common database design process used to remove redundant or incorrect organization and data. The design and normalization of the database will create a maintainable data set that can be cross-referenced. Normalized data is not only easier to analyze but also easier to expand. Normalization involves removing redundancy and correcting incorrect data structure and organization.
OLAP	Online analytical processing.
Query	A request for information from a database. There are three query methods: Choose – With this easy-to-use method, the database system presents a list of parameters from which you can choose. This method is not as flexible as other methods. Query by example (QBE) – With this method, the system lets you specify fields and values to define a query. Query language – With this method, you have the flexibility and power to make requests for information in the form of a stylized query using a query language. This is the most complex and powerful method.
Raw Data	Data that has not been turned into “information,” through processing. Although factual and “real” raw data is unorganized.
Relational Database (RDBM)	Short for relational database management system, most often referred to as RDBM. Data is stored in related tables. Relational databases can be viewed in many different ways. In this system a single database can be spread across several tables. (RDBM differs from flat-file databases where each database is self-contained as a single file or table.)

Repository	A specific location where data is stored and maintained.
Staging Data	Data that is to be processed before entering the warehouse.
Table	Refers to data arranged in rows and columns, like a spreadsheet. In relational database management systems, all information is stored in tables.
Transformation	Data that are extracted from databases are transformed into a desired form, using various tools that cleanse, merge, purge, aggregate, calculate, audit, remove redundancy, standardize, etc.
Warehouse	(See data warehouse)

Index

A

Architecture 8

B

Bootstrap
see also deployment 49

C

Connectivity Map
defined in a project 14
conventions
path name separator 10
Windows 10

D

Deployment
activate 49
bootstrap and management agent 49
deployment profile 48
eTL deployment 50
eTL, connect source and target 52
eTL, deploy a stand-alone process 50
eTL, deploy an invoked service 55
general instructions 48
integration and message servers 48
run the bootstrap 49
verify output 49
Deployment Profile Editor
defined in a project 14
document
conventions 10
Documents
supporting documents 9

E

Enterprise Designer
create and configure components 12
editor 14
GUI 13
menu bar 13
project 14
project editor, is in the right pane 14
Enterprise Explorer
organize components, in the left pane 14

the left pane of the Enterprise Designer 13
Environment Explorer
accessed from view menu 13
collection of logical hosts 14
ETL process
the ETL process, with graphic 6
Extraction Transform Load
ETL, eTL for the SeeBeyond product 5

F

Flat files
see OTD, flat files 14

I

Installation requirements 10

M

Menu Bar 13

O

Operators and transformation tools
and 37
equal 36
greater than 37
join 35
literal 39
touppercase 40
OTD
creating a flat file 14
flat file, field delimiter 20
flat files, selection criteria 16
suggested record displays, based on your file
structure 22
OTD Editor
defined in a project 14

P

Project
connect to database 26
create a new sample project 24
create new object type definition 25
Enterprise Designer window 31
new eTL Collaboration 32
properties 42
runtime inputs 41
runtime outputs 41
select database objects 27
select tables for Connectivity Map 33

Index

- select tables using search tools 29
- specify OTD name 30
- tables and views 29
- using eTL with eInsight 43
- using with eInsight, business rules 44
- using with eInsight, inbound mapping for eTL 44
- using with eInsight, outbound mapping 46
- using with eInsight, validate 47
- validate and run 42

R

- Requirements 10

S

- SQL
 - default type 19

W

- writing conventions 10