TIBCO® Data Virtualization

Discovery Guide

Version 8.1

Last Updated: February 26, 2019

TIBCO®

Important Information

Contents

Preface

Documentation for this and other TIBCO products is available on the TIBCO Documentation site. This site is updated more frequently than any documentation that might be included with the product. To ensure that you are accessing the latest available help topics, please visit:

- https://docs.tibco.com

Product-Specific Documentation

The following documents form the TIBCO® Data Virtualization(TDV) documentation set:

- *TIBCO TDV and Business Directory Release Notes* Read the release notes for a list of new and changed features. This document also contains lists of known issues and closed issues for this release.

- TDV Installation and Upgrade Guide

- TDV Administration Guide

- TDV Reference Guide

- TDV User Guide

- TDV Security Features Guide

- TDV Business Directory Guide

- TDV Application Programming Interface Guide

- TDV Tutorial Guide

- TDV Extensibility Guide

- TDV Getting Started Guide

- TDV Client Interfaces Guide

- TDV Adapter Guide

- TDV Discovery Guide

- TDV Active Cluster Guide

- TDV Monitor Guide

- TDV Northbay Example

How to Access TIBCO Documentation

Documentation for TIBCO products is available on the TIBCO Product Documentation website mainly in the HTML and PDF formats.

The TIBCO Product Documentation website is updated frequently and is more current than any other documentation included with the product. To access the latest documentation, visit https://docs.tibco.com.

Documentation for TIBCO Data Virtualization is available on https://docs.tibco.com/products/tibco-data-virtualization-server.

How to Contact TIBCO Support

You can contact TIBCO Support in the following ways:

- For an overview of TIBCO Support, visit https://www.tibco.com/services/support.

- For accessing the Support Knowledge Base and getting personalized content about products you are interested in, visit the TIBCO Support portal at https://support.tibco.com.

- For creating a Support case, you must have a valid maintenance or support contract with TIBCO. You also need a user name and password to log in to https://support.tibco.com. If you do not have a user name, you can request one by clicking **Register** on the website.

How to Join TIBCO Community

TIBCO Community is the official channel for TIBCO customers, partners, and employee subject matter experts to share and access their collective experience. TIBCO Community offers access to Q&A forums, product wikis, and best practices. It also offers access to extensions, adapters, solution accelerators, and tools that extend and enable customers to gain full value from TIBCO products. In addition, users can submit and vote on feature requests from within the TIBCO Ideas Portal. For a free registration, go to https://community.tibco.com.

Introducing Discovery

This topic describes the features and architecture of Discovery.

Overview of Discovery

Discovery is a product that reveals hidden correlations in your enterprise data. Discovery uses that knowledge to help you build data models to use for data virtualization and reporting. Working within TDV and Studio, you can use Discovery to scan data and metadata from across data repositories, whether they are applications, databases, or data warehouses. You can then use Discovery's

graphical tools to create data models based on system metadata and discovered relationships. From the data models, you can create views in TIBCO® Data Virtualization(TDV). The following illustration provides an overview of the information flow among the components.
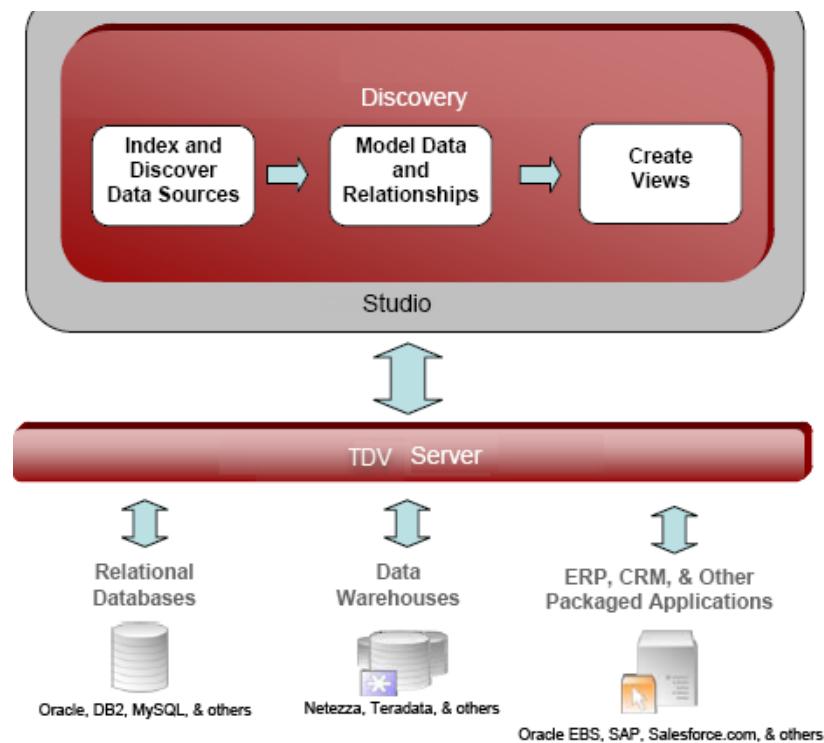
Discovery Key Features and Benefits

Discovery's key features include:

- Automatic discovery of relationships within and between enterprise data sources.

- Interactive display of relationships.

- The ability to leverage relationships to build models for data virtualization.

- The ability to overcome obstacles in creating federated enterprise data models such as:

  — A complex data environment with many heterogeneous data sources.

  — Poor quality data not adhering to standards.

  — Legacy data or data that is not owned by the customer.

  — Lost knowledge about the data.

How Discovery Works

Discovery helps you understand how your data sources relate to each other. Because Discovery understands the underlying data and metadata, it can reveal previously unknown relationships.

Discovery provides the facility to:

- Access enterprise-wide data sources, including relational databases, data warehouses, and packaged applications.

- Index the data sources.

- Discover data relationships within and across data sources.

- Validate data source relationships.

- Create data models using a graphical interface.

- Create views in TDV.

Discovery and TDV

Discovery works with Studio and TDV. Using Discovery with TDV, you can define complex models comprising multiple data sources and SQL views, and create a TDV view based on the model.

Creating a TDV view from a Discovery model lets you take advantage of TDV's key features:

- Data virtualization presents data as if it were available from one source, regardless of how it is physically distributed across data silos. Query optimization and caching provide low latency performance without physical replication.

- Data abstraction simplifies complex data by transforming it from its native structure and syntax into reusable views and Web services.

- Data federation securely accesses diverse operational and historical data, combining it into more complete and meaningful information for a range of applications to consume.

Discovery Concepts

This section introduces key Discovery concepts.

What Is a Model?

A Discovery *model* is an object that contains resource definitions, view definitions, and diagrams that show their relationships: a model diagram that shows tables/keys/columns and relationships and a cross schema diagram that shows the relationships between schemas.

What Is a View?

A TDV *view* is a SQL statement that defines the data to retrieve from one or more underlying data sources. Discovery lets you automatically generate a view based on a model that contains discovered relationships within and across data sources. You can then use Studio and its graphical SQL Editor to refine the SQL views.

What Is a Data Source?

A *data source* is a structured data application such as a database, data warehouse, or packaged software application. Discovery supports most data sources that TDV supports.

What Is an Index?

An *index* is an internal mapping that helps retrieve data faster. Discovery indexes data sources to optimize them for relationship discovery.

What Is a Relationship in Discovery?

A *relationship* is a probable relationship between data in different table columns, based on data values and metadata. A relationship can be:

- Discovered—The relationship is based on the metadata defining a table column and is automatically determined by Discovery.

- User-defined—The relationship is defined by the Discovery Administrator.

- Foreign key—The relationship is determined by Discovery based on a foreign key defined for the tables.

The estimated validity of a relationship is indicated by a Relationship Probability Score (RPS), expressed as a percentage value.

Discovery Architecture

This section provides an overview of some of the key technologies utilized in Discovery.

Indexing

Discovery compiles reverse indexes of the data in the data sources to speed up discovery of and access to it. The indexes are in turn used for relationship discovery.

Accessing the actual data, either when building the index or when performing relationship discovery, requires that Discovery understand the diverse data file formats of all of the data sources. Discovery uses the JDBC standard to access databases.

Discovery uses high-performance federated query algorithms to optimize the select statements generated by the search engine for relationship discovery.

Indexing speed depends on network speed, database constraints, table size, and other traits of the data sources.

Discovery has configuration parameters with which you can control the indexing process. For example, you can configure Discovery to ignore case during indexing, or use data sampling for very large data sources.

Data Relationships

Discovery lets you work with three types of data relationships:

Discovered Relationships

Discovery automatically finds unknown relationships within and between data sources based on:

- Metadata 'introspected' using the TDV data source definition technology. Discovery understands schemas, catalogs, tables, column names, data types, lengths, and primary and foreign key constraints.

- The data values themselves. Implicit relationships are discovered using data correlation between pairs of columns. For columns from two tables to have a JOIN relationship, they must have at least one value in common.

- Discovery assigns a Relationship Probability Score (RPS) based on the number of common values between two columns, whether either column is a primary or foreign key, similar or identical table or column names, and so on.

Discovery finds relationships between single columns across tables. You can also configure it to find:

- Multicolumn compound relationships, where two columns in a table form a unique key that matches the same two columns in another table.

- Relationships based on user-defined data domains that specify patterns in which matching data might be stored.

Discovery lets you validate probable relationships and remove those that are not valid.

Foreign Key Relationships

Discovery finds existing relationships (primary and foreign key relationships) using system metadata in existing schemas. These relationships are discovered and can be displayed in models. Foreign key relationships are always considered to be valid.

User Defined Relationships

Even if a relationship is not found by the Discovery RPS algorithm, users can create new relationships based on their knowledge of the data.

What Is Included in Studio with Discovery

Discovery features that are visible in the Studio user interface include:

- A New Model command in the Studio resource tree.

- A Discovery model workspace with options, controls, and tabs to view and edit information about a model and the resources it contains.

- A Relationship Discovery tab for data sources.

- A Discovery panel in Studio. (Its button is on the left edge of the Studio window.)

- Discovery configuration parameters.

- System tables (ALL_RELATIONSHIP_COLUMNS, ALL_RELATIONSHIPS, and SYS_TASKS).

Note: All of the usual TDV and Studio features are available in the Discovery product.

About Relationship Discovery

Discovery extends the power of data modeling by learning about the relationships between the data sources in your enterprise. Types of relationships are described in the subsections of Data Relationships, page 11. You can evaluate and adjust relationship probability scores yourself. (See About the Relationship Probability Score (RPS), page 23.) You can use a configuration parameter to specify a minimum score. (See Adjusting the Minimum Score for Relationship Discovery, page 73.)

Discovering relationships in large data sources can be time-consuming, so it is best to perform this at times of low system activity.

About Data Sampling

Random sampling of source data increases performance of the indexing and discovery process, although with some reduction in accuracy. In general, Discovery finds relationships that matter while not presenting false negatives for review.

The need to use data source sampling is driven by data volume, available memory, and the bandwidth available between source systems and TDV. If indexing takes too long, you might want to enable data sampling. With data sampling enabled, only a portion of the rows in the table or view are indexed.

How sampling works depends on what is being indexed and the type of data source being discovered.

- Table Data Sampling, page 14

Table Data Sampling

For tables, data sampling is enabled and controlled via two Studio configuration parameters and the table cardinality. Discovery applies the data sampling algorithm only if both of these are true:

• The Sampling Is Enabled configuration parameter is on.

• The cardinality of a table exceeds the data sampling threshold specified by the Sampling Size configuration parameter.

With sampling enabled and the threshold exceeded, the number of rows indexed for a table is calculated using the formula:
Sampling Size/Cardinality=% of rows indexed

For example, if you have a 1 million row table and Sampling Size is set to 100000, then 10% of the table will be indexed.

Which rows are indexed is controlled by a random number generator.

If data sampling is not enabled, data sampling does not occur even if the table cardinality threshold is exceeded.

If a table resides in one of these data sources—Oracle, DB2, MySQL, Netezza, or Microsoft SQL Server—data sampling is pushed to the data source. Otherwise, all data is fetched into TDV for sampling there.

View Data Sampling

For views, data sampling is enabled and controlled by two Studio configuration parameters:

• Sampling Is Enabled

• Sampling Size

See Configuring Data Sampling, page 72.

All rows are indexed up to the Sampling Size threshold. If sampling is enabled, Discovery then starts indexing on a decreasing scale. That is, when the Sampling Size threshold is passed, Discovery begins indexing half as many rows. If the (Sampling Size x 2) is reached, then one quarter as many rows are indexed. Which rows are indexed is controlled by a random number generator.

If data sampling is not enabled, data sampling does not occur even if the threshold is exceeded.

About the Discovery Tasks Panel

This topic contains:

Task Statistics

Task statistics are displayed at the top of the Discovery Tasks tab. Each cell displays in real-time the number of tasks that currently exist in that particular status.

Task Status

Example of the Tasks table with the default columns displayed:

Click to change the columns displayed

The columns in the Tasks tab are explained in the following table:

Tasks Column	Meaning
Type	The task type. Can be any of these values: Discovery Scanner (folder) Indexer (folder) Table Indexer Relationship Scanner (folder) Column Scanner Cross Column Scanner Relationship Analysis Multicolumn Indexer (folder) Multicolumn Table Indexer Multicolumn Relationship Scanner (folder) Multicolumn Relationship Key Scanner Post Process
Name	The paths and names of the resources involved in the task. For Table Indexer tasks, also includes the number of rows and columns in the table.
Status	The task status can be: Running—The indexing or discovery task is running. Finished—The indexing or discovery task has finished. Fail_Waiting—One or more of the task's child tasks has failed but other child tasks are still running. Failed—The indexing or discovery task has failed. Cancel_Waiting—One or more of the cancelled tasks is still running. Cancelled—The indexing or discovery process has been cancelled by the user.
End Time	The time the task finished. While the task is in progress, displays a progress bar for each task.
Duration	The duration of the task.

Tasks Column	Meaning
Time Remaining	The amount of time remaining in this task.
Rows Processed	The number of rows processed by this task.
Start Time	The time the task started. Not displayed by default.
Error Message	The error message if one was generated by this task.
Processing Time Remaining	The amount of time remaining in this task.
Owner	The user who initiated this task.

You can perform any of these actions in the Tasks table:

- Expand and collapse the tree. You can use the + and - buttons to expand or collapse all branches, or you can use the down or right arrows to expand and collapse specific branches.

- Get task details.

- Sort tasks differently by clicking a column header.

- Filter tasks.

- Purge all indexes and relationships.

- Open the Index and Relationship Task Scheduler to index or discover relationships for a data source.

About Indexing

By default, when you index a data source, all tables in the data source are fully indexed. But if time becomes an issue, here are some ways to deal with it:

- If you have an especially large data source, you might want to index it at a time when there is the least amount of data source activity. You might also want to use sampling for large data sources. See About Data Sampling, page 13.

- You can limit the scope of indexing and improve performance by excluding catalogs, schemas, or tables from the data source. See Modifying a Data Source Definition, page 50, for more information.

These factors affect how long the indexing and relationship discovery process might take:

- The number of rows and columns in the data source.
- The TDV machine and processor.
- The maximum Java heap setting.

When Discovery creates the index, it saves it in the TDV installation directory. You can specify a different location of this directory. See Changing the Index Directory, page 70.

About Pattern Expressions

For the pattern expressions entered for Match Expression and Transformation, the expression must be a valid regular expression as defined in:

http://docs.oracle.com/javase/6/docs/api/java/util/regex/Pattern.html

Consider the example below:
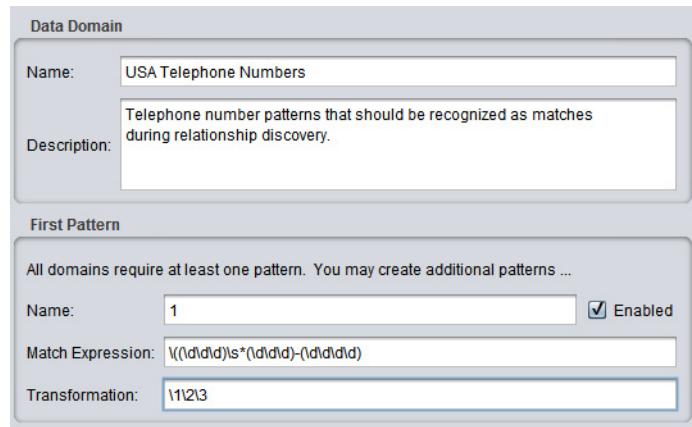
The Match Expression in the example would match a phone number in the format:

(333) 123-1234

Parentheses surround groupings that you want to reuse in the Transformation expression. A backslash marks any parenthesis that is part of the string to match. In the transformation expression, the group of three digits end up in the output of the transformed string by placing \1 in the transformation expression, which means that the first group goes there; similarly for the other groupings. The resulting, transformed phone number is 3331231234. This canonical form of a phone number is applied to all phone numbers so that they can be compared.

The table below shows how you might define telephone number patterns

Phone Number Storage Pattern	Match Expression	Transformation	Canonical Form Used by Discovery
(333) 123-1234	\(\(\d\d\d\)\)\s*(\d\d\d)-(\d\d\d\d)	\1\2\3	3331231234
333.123.1234	(\d\d\d)\.(\d\d\d)\.(\d\d\d\d)	\1\2\3	3331231234
333 123-1234	(\d\d\d)\s(\d\d\d)-(\d\d\d\d)	\1\2\3	3331231234
333-123-1234	(\d\d\d)-(\d\d\d)-(\d\d\d\d)	\1\2\3	3331231234

The formats of all phone numbers in your databases can be thus be normalized in an intermediate step during indexing and comparison, so that cells that contain the same phone number will match.

About the Relationships Discovered

This section focuses on the relationships discovered by Discovery, including:

What Is and Is Not Discovered

At times, it might not be obvious why some relationships are discovered and others are not. The list below explains some of the reasons:

- The number of matches between two columns is less than three so the relationship is not discovered.

  For example, if you have a column that contain the same kind of data but has only two possible values, Discovery does not discover this relationship. This threshold cannot be changed.

- Columns with STRING data types have dissimilar names, but the data has at least three matches, so the columns are discovered. Relationships involving columns with string data types put more emphasis on the data than on the column names. See Adjusting the Weights of the RPS Factors, page 74 for the weights applied to string data types.

- Column names begin and end with the same characters, but the characters between do not match. For example, if CUSTOMER_ID and CUSTOMER_ACCOUNT_ID have matching data, the relationship might not be found or have a low score because the column names are different. The more characters in the column name that do not match, the lower the score.

- Two columns of INTEGER data type might not be found or have a low score. For example, if all the integers in two columns match, the score is still low because they are both of type BIGINT. In this case, Discovery puts more emphasis on the column name comparison.

- If the Min Score value set in the model Diagram is set too high, relationships might be found but not displayed. The default value is 75.

- Matches might not be found because capitalization is different. You can control this using the Use Case Sensitivity for Discovery configuration parameter. See Configuring Case Sensitivity, page 71.

About Multicolumn Relationships

In multicolumn relationships, the combination of two or more columns in one table form a unique key, and have a relationship with the same combination of columns in another table. For example, if both Table A and Table B have FirstName and LastName columns, you might want to combine the multiple columns into single relationship. (The individual relationships are always found.)

Discovery supports discovery of multicolumn relationships. To configure for this, refer to Table Statistics Displayed for Relationship Discovery, page 21. If multicolumn discovery is active, Discovery first indexes and scans for single-column relationships, and then indexes and scans for multicolumn relationships. The corresponding tasks are displayed on the Discovery Tasks panel.

Multicolumn relationships are identified as such in the model Diagram. You can choose to display or not display them. You can also choose whether to combine or separate single-column relationships and multicolumn relationships. See Working with Relationships in the Model Diagram, page 96 for details.

When a relationship is formed by multiple columns, it is indicated as a multicolumn relationship on the model's Relationships tab. (See Working with the Relationships Tab, page 114.) You can also see which columns are participating in the multicolumn relationship on the Relationship Column Details tab. (See Viewing Relationship Details, page 99.)

The multicolumn relationship Discovery feature is controlled with the MultiColumnKey configuration parameters as described in Table Statistics Displayed for Relationship Discovery, page 21. You can access these parameters in Studio by choosing Configuration from the Administration menu. By default, multicolumn relationship discovery is turned off.

Table Statistics Displayed for Relationship Discovery

When Discovery is present in Studio, tables in each TDV data sources display additional details about relationships in the Columns tab. Such data is shown only after relationship discovery has occurred for this data source.

Columns 1-10:

Columns 10-20:

The first three columns are shown for all tables in Studio. (See the *TDV User Guide.*) All other columns are specific to Discovery

Table Columns/ Discovery-Specific Statistic	Description
In Relationship	Indicates if this column is involved in a Discovery relationship.
	Click Display Results to see the relationships of each type in the Result tab.
Relationships	The number of relationships in which this column is involved.
	<#>—The number of relationships for this column.
	Click Display Results to see details about the relationships in the Result tab.
Key Type	If this column is a primary or foreign key, displays the key type:
	PK— Primary key.
	FK— Foreign key.
Null Permitted?	Y if nulls are permitted; blank if nulls are not permitted.
% Null	The percentage of null values in this column.
Total Null	The total number of rows that have a null in this column.
	<#>—The number of nulls in this column.
	Click Display Results to see the data in this table in the Result tab.
Total Non Null	The total number of rows that have a value in this column.
	<#>—The number of rows that have a value in this column.
	Click Display Results to see the data in this table on the Result tab.
Unique Constraint?	Y if the values in this column must be unique; N if not.
% Unique	The percentage of unique values in this column.
Total Unique	The total number of rows that have unique values in this column.
	<#>—The number of rows that have unique values in this column.
	Click Display Results to see the rows with unique data on the Result tab.

Table Columns/ Discovery-Specific Statistic	Description
Total Non Unique	The total number of rows that have non unique values in this column.
	<#>—The number of rows that have non unique values in this column.
	Click Display Results to see the rows with non unique data in the Result tab.
Distinct	The total number of rows that have distinct values in this column.
	<#>—The number of rows that have distinct values in this column.
	Click Display Results to see the rows with distinct values in the Result tab.
Min	The minimum value in this column. If numeric, the lowest numeric value. If alphabetic, the value that begins with the lowest letter in the alphabet where A is low and Z is high.
Max	The maximum value in this column. If numeric, the highest numeric value. If alphabetic, the value that begins with the highest letter in the alphabet where A is low and Z is high.
Mode	Which term happens most frequently.
% Mode	The percentage of time the mode occurs.
Denormalize	The separator to use for denormalization. Typically, a comma, period, space (\s), or tab (\t).

About the Relationship Probability Score (RPS)

Discovery controls what relationships are discovered and how they are scored based on Discovery's internal RPS algorithms. Discovered relationships are assigned a RPS based on these factors:

Column Name Comparison Factor
Index Key Factor
Match Percentage Factor
Number of Matches Factor
Schema Locality Factor

The factors are multiplied by their weights (which total 100%) and then added together to arrive at the total score for the relationship using this formula:

Score = columnNameComparisonFactor * 40% + indexKeyFactor * 30% + matchPercentageFactor * 10% + numberOfMatchesFactor * 10% + schemaLocalityFactor * 10%

The factor weights are configurable for non-string data types. See Adjusting the Weights of the RPS Factors, page 74 for information about changing the weights of these factors.

The following table describes these factors.

Factor	Description
Column Name Comparison	This factor is multiplied by its weight to get the name component of RPS. It ranges from 0 to 1, with 1 being an exact match and 0 being no match.
	1.0—The column name of c1 and c2 match exactly.
	0.9—The column names match exactly with non-alphanumeric characters removed. Example: users.user_id has an 0.9 factor when compared with users.userid.

	0.9—One column name ends with the other column name.
	  Example: users.user_id is given an 0.9 when compared with sales.sold_to_user_id
	  Example: term.term_id is given an 0.9 when compared with pymts.pymt_term_id

	0.9—The table name of one column name is part of the other column name. Example: issue.id is given an 0.9 when compared with status.issue
	0.8-0.5—Column values have similar names (to handle misspelling names). Example: cust.user_id is given a factor of 0.5-0.8 when compared with cust.usee_id
Index Key	This factor is multiplied by its weight to get the index key component of RPS. It is in the range from 0 to 1 based on the likelihood that one of the columns in the relationship is a key column:
	1.0—The relationship cardinality is one-to-one, many-to-one, or one-to-many; and either column has more than 90% unique values.
	0.5—The relationship cardinality is many-to-many with less than 90% unique values in both columns.

Factor	Description
Match Percentage Factor	This factor is multiplied by its weight to get the match percentage component of RPS. It is calculated using this formula: [# matches]/ MIN ([# unique values in c1], [# unique values in c2]) [# matches] is the number of unique values in both column1 and column2. See Adjusting the Minimum Unique Percentage, page 74 for information about adjusting the threshold of value uniqueness. Example: If the number of unique values in c1 is 100, the number of unique values in c2 is 50, and the number of unique values appearing in both c1 and c2 is 40. In this case, the factor is equal to 40/MIN(50,100)= 40/50=0.8.
Number of Matches Factor	This factor is multiplied by its weight to get the number of matches component of RPS: 1.0—[number of matches] => 10 else [factor]—[number of matches]/10 By default, if the minimum number of matches is less than 3, the relationship is not discovered.
Schema Locality Factor	This factor is multiplied by its weight to get the schema locality component of RPS: 1.0—Two columns are from the same data source. 0—The columns are not from the same data source.

Getting Started with Discovery

This chapter describes Discovery requirements and support and the general workflow for using Discovery. It also includes a quick tour that shows how to discover relationships and create models.

The following topics are covered in this chapter:

Discovery Requirements and Support

Discovery is bundled with TDV and is automatically installed with it. See the *TDV Installation and Upgrade Guide*.

This section includes summary information about:

Discovery System Recommendations

You must install Studio on Windows to access the Discovery user interface. The recommended TDV Server system configuration for using Discovery is listed below.

System Requirement	Description
Processor recommendations	• Recommended processor—Intel Core i7 CPU 920 @2.67GHZ eight core • Minimum processor—Intel Xeon 5110 CPU @1.60 GHz quad core
Disk space	• Recommended—32 GB • Minimum—16 GB

System Requirement	Description
Physical memory for TDV	• Recommended RAM—16 GB • Minimum RAM—8 GB The physical memory requirements listed here are exclusively for running TDV and do not include the memory you might need to run non-TDV applications.
Operating system platforms	• Windows 64-bit • Linux 64-bit The Discovery user interface is an integral part of Studio and must be installed on a Windows platform. However, you can install TDV and its repository on any supported UNIX platform.
Studio memory	• Default RAM—768 MB • Recommended minimum RAM for Studio with Discovery—512 MB This setting is defined in the <TDV_install_dir>\bin\studio.bat file. The parameter that defines Studio RAM is MAX_MEMORY. For large relationship models, this value needs to be 512MB or larger.

Supported and Unsupported Data Sources

Generally, Discovery supports TDV data sources that are tabular. That is, Discovery can discover relationships between physical tables and physical views but not in procedural objects or XML documents.

Supported data sources

• Relational data sources—The tabular objects of relational sources including physical tables and views are supported.

• LDAP data sources—Supported, because Discovery models LDAP data source content as physical tables.

• Delimited files—Supported, because the content is tabular.

• Adapters—Oracle EBS, Salesforce.com, and Siebel.

Unsupported data sources	• Relational data sources—Procedural objects in relational sources are not supported. HP Neoview 2.3 and 2.4, IBM DB2 z/OS Version 8, and PostgreSQL.
	• Apache Hive/Hadoop
	• WSDL data sources—Not supported, because its content is procedural.
	• XML data sources—Not supported, because its content is not tabular.
	• Custom Java procedures—Not supported, because they are procedural.
	• Adapters—SAP and SAP BW.

See the *TDV Installation and Upgrade Guide* for a complete list of TDV-supported data sources.

Supported Data Types

During introspection, all catalogs, schemas, tables, and columns are discovered. However, Discovery does not support all data types for indexing or relationship discovery. Only columns with supported data types will be indexed by Discovery and returned in Discovery search results.

Supported data types	CHAR, VARCHAR, LONGVARCHAR, TINYINT, SMALLINT, INTEGER, BIGINT, FLOAT, DOUBLE, NUMERIC, DECIMAL, REAL, DATE, TIME, TIMESTAMP, BIT, BOOLEAN
Unsupported data types	CLOB, BINARY, VARBINARY, LONGVARBINARY, BLOB, ARRAY, DISTINCT, DATALINK, JAVA_OBJECT, NULL, REF, STRUCT, OTHER

About Exporting and Importing Discovery Information

Discovery models and metadata, including indexes and relationships, can be exported into a CAR file like other resources. If you perform an export from Studio, Discovery index and relationship files are not included unless you check the Discovery Files check box in the Export dialog box. When you perform a Full Server Backup, all Discovery resources are exported to the backup file by default.

Exporting and importing resources from Studio is documented in the *TDV User Guide*. Exporting and importing using the command line is documented in the *TDV Administration Guide.* The TDV backup_export.sh and backup_import.sh scripts include existing indexes and relationships, as well as models and relationship data, in the CAR file.

Note: When you import a CAR file that contains Discovery index and relationship files, the index status of the imported data sources is updated only if the exported CAR file was created with a 6.1 or later version of TDV server. The index status can be displayed in the Discovery Matrix panel and on the Relationship Discovery tab for an open data source.

Reusing a Discovery Model Created in an Earlier Version

To reuse a Discovery model created in a previous release, you must select the Discovery Files option during the Studio export or full server backup. The export/backup archive then includes the Discovery relationships, indexes, and task information. When you import the archive, Discovery relationships, indexes, and task information are imported only if the Discovery Files option is selected at import.

Note: You can reuse a Discovery model created in earlier versions of Discovery except for Discovery models created in 5.2. You must recreate Discovery 5.2 models.

Overview of the Discovery Workflow

These steps are typical, but not necessarily sequential or required. For example, you might already have existing data sources. Also, the steps are iterative; after you save a model you can revise it, add new data sources, remove data sources, and perform other editing.

Typical Discovery Workflow
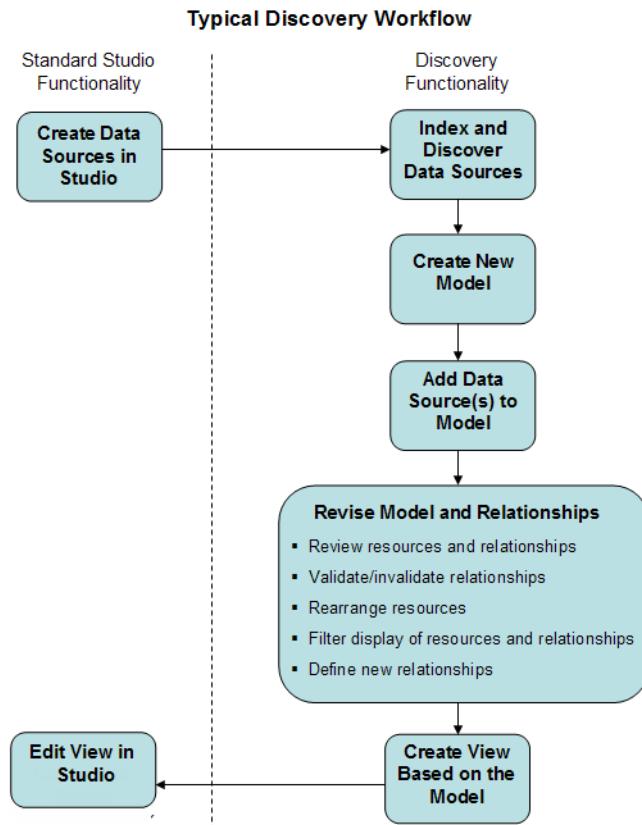
The Discovery workflow is summarized here:

1. Create one or more data sources.

2. Index and discover their relationships.

3. Create a new model in Studio.

4. Add the indexed/discovered data sources to your model.

5. Revise the model and relationships:

— Review the tables and relationships in the model.

— Validate or invalidate relationships.

— Rearrange tables and relationships using the graphical editor.

— Filter the display of resources and relationships using the toolbar controls.

— Optionally, define new relationships.

6. Select resources in the model and create a new view.

7. Edit the view in Studio.

A Quick Tour of Discovery

For this quick tour, we use two of the data sources automatically installed with Studio: ds_inventory and ds_orders. These data sources are installed in your Shared\examples folder. It illustrates many of the things you can do with Discovery, although it is not exhaustive.

In Studio, follow the steps described below to get a quick idea of how you can use Discovery.

To complete the Discovery Quick Tour

Indexing and Discovering the Data Sources

To index and discover a data source

1. Start Studio.

2. In the Studio resource tree, select ds_inventory in the Shared\examples folder.

3. Right-click and choose Open.

4. Click the Relationship Discovery tab at the bottom of the window.

5. Click Scheduler.
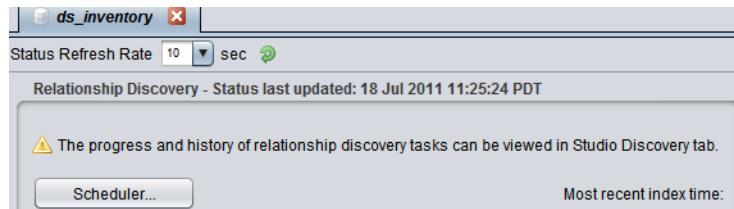
Studio displays the Index and Relationship Task Scheduler dialog similar to the following.
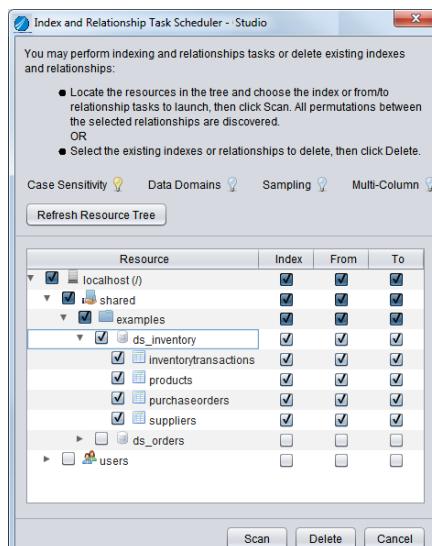
6. In the Index and Relationship Task Scheduler, click Scan.

Discovery indexes and discovers relationships in the ds_inventory data source. When this process is done, the Status on the ds_inventory

Relationship Discovery page displays Finished and the Relationship Discovery tab itself has a green check mark.
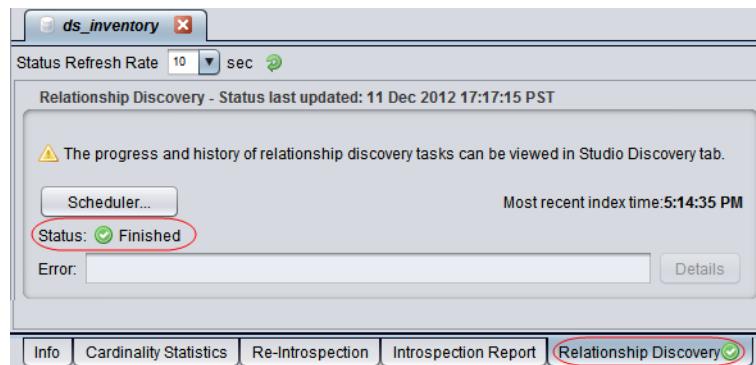
Creating a New Model

To create a new model

1. In the Studio resource tree, select the Shared\examples folder.

2. Right-click and choose New Model.

3. Type a model name and click **OK**.

   Discovery opens the new model in the model workspace.
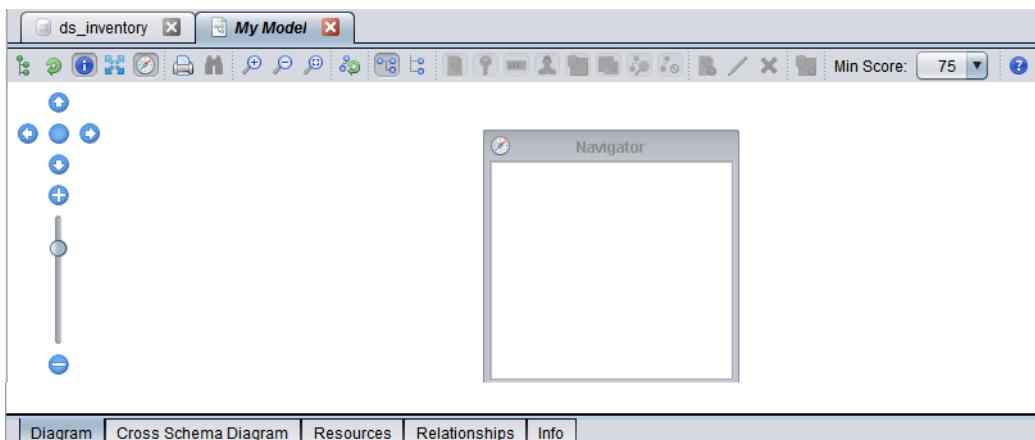
Adding the Data Source to the Model

To add the data source to the model

1. In the Studio resource tree, select ds_inventory and drag it onto the model Diagram.

   Discovery displays the data source in the model, showing the relationships discovered between the tables. Tables in green are views which are dependent on this data source.
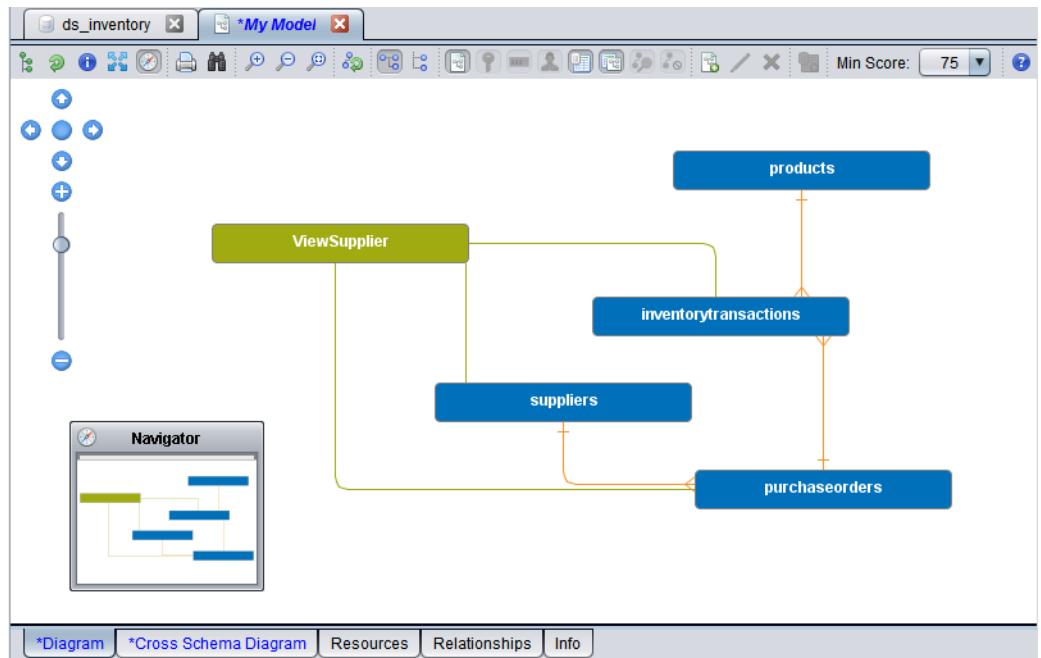
The ds_inventory data source has:

— Four tables with discovered relationships to each other.

— One view that has a dependency upon tables in ds_inventory.

2. Rearrange the tables, view, and Navigator so that you can see all of the elements.

    — Move the Navigator window to another location in the diagram as necessary to see model objects.

    — Select a table and move it.

    — Click Ctrl+A to select all objects and move them.

    — Use the pan and zoom controls to change what is displayed.

Discovering Relationships between Multiple Data Sources

You can display relationships between two data sources in a model, but you must first discover the relationships between them. If a data source has not already been indexed, you also need to index it.

To discover and display relationships between multiple data sources

1. In Studio, click the Discovery tab on the left edge of the window.

2. On the Tasks tab, click Scheduler.

    Discovery displays the Index and Relationship Task Scheduler dialog with no boxes checked.

3. Next to ds_inventory, check the From and To check boxes.

    Note: You already indexed this data source in a previous step so it is not necessary to do it again.

4. Check the box to the left of ds_orders (which automatically checks the Index, From, and To check boxes.)

The Index and Relationship Task Scheduler should look similar to this.
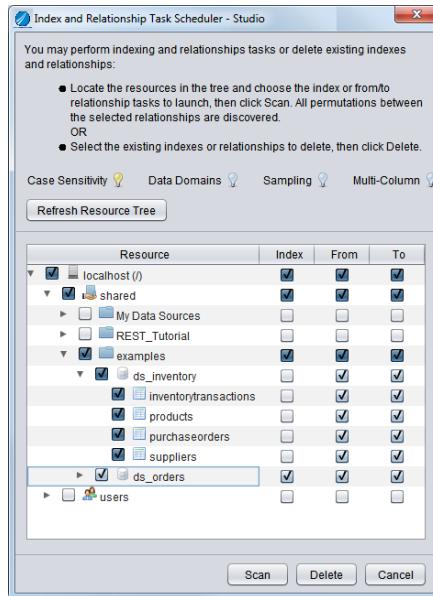
5. Click Scan.

   The Discovery Tasks panel shows the progress of each task in the Status and End Time columns. When the Status for all tasks is Finished, proceed to the next step.

6. In Studio, click the Modeler button on the left edge of the window to display My Model again in the workspace.

7. Click the Refresh button on the model toolbar.

The model diagram is updated to include all tables in ds_inventory as well as all tables in ds_orders that have a relationship with a table in ds_inventory.
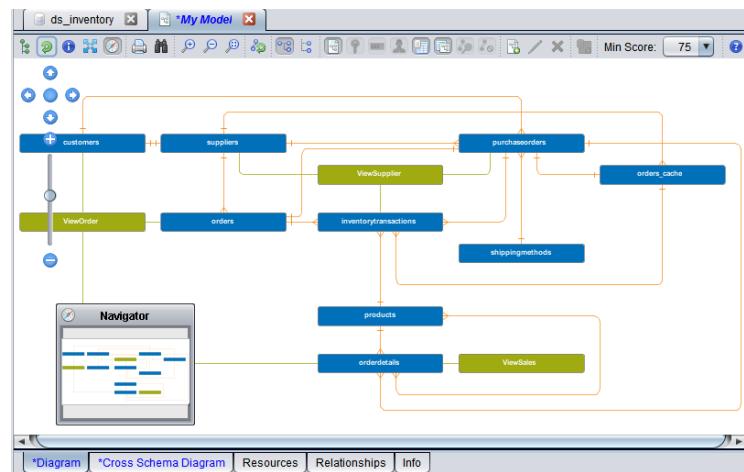
Note: Only one data source, ds_inventory, is included in this model, along with any other resources related to it. That is because some tables in ds_orders are related to tables in ds_inventory, they are now included in the model.

8. Add another resource using this method:

   a. Click the Add/Remove Resources button on the model toolbar.

   b. Expand the Examples directory and add a check mark next to ds_orders.
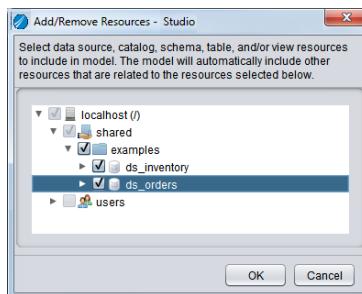
   c. Click **OK**.

The model Diagram tab now shows all resources and dependencies in the two data sources, and the relationships between them.
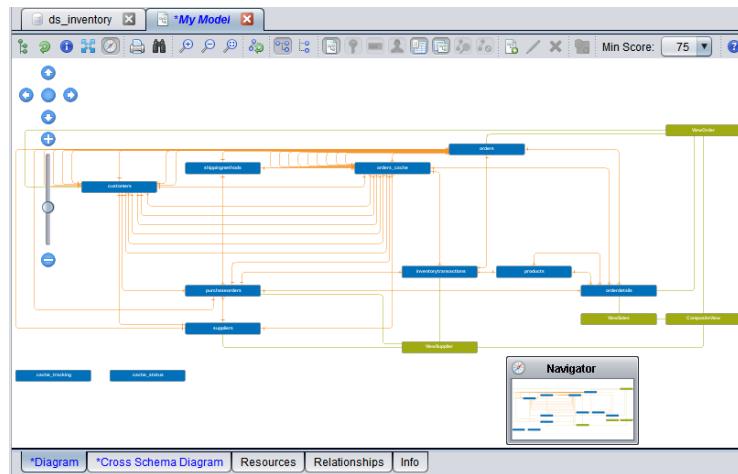
Viewing the Schemas in the Model

You can get a higher level view of the model, and the schemas it contains, in the Cross Schema Diagram.

1. In the open model, click the Cross Schema Diagram tab.

   The model Cross Schema Diagram tab shows the two data source schemas, the relationships between them, and the views that are dependent on the two data sources.
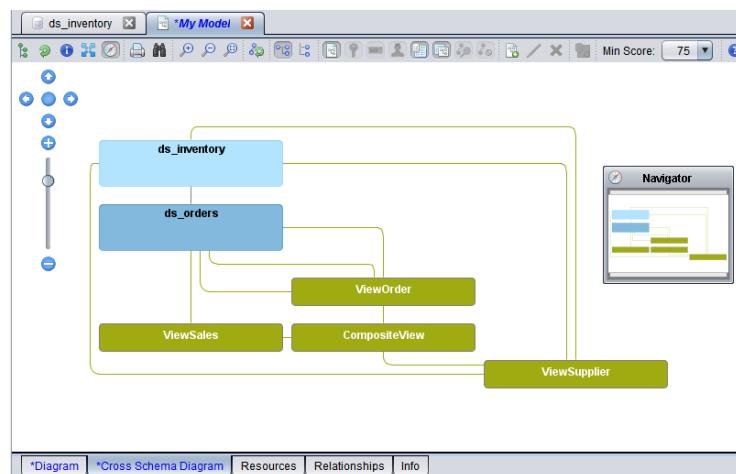
2. Select ds_inventory, right-click, and choose Show Tables with Relationships.

   Discovery displays all relationships within the ds_inventory schema, and all relationships to a table in the ds_orders schema.
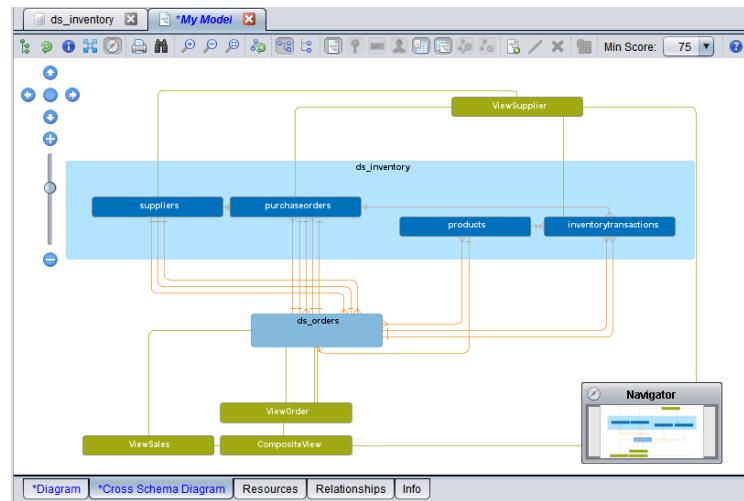
3. Select ds_orders, right-click, and choose Show Tables with Relationships.

   Discovery now displays all relationships within and between the ds_inventory schema and the ds_orders schema.
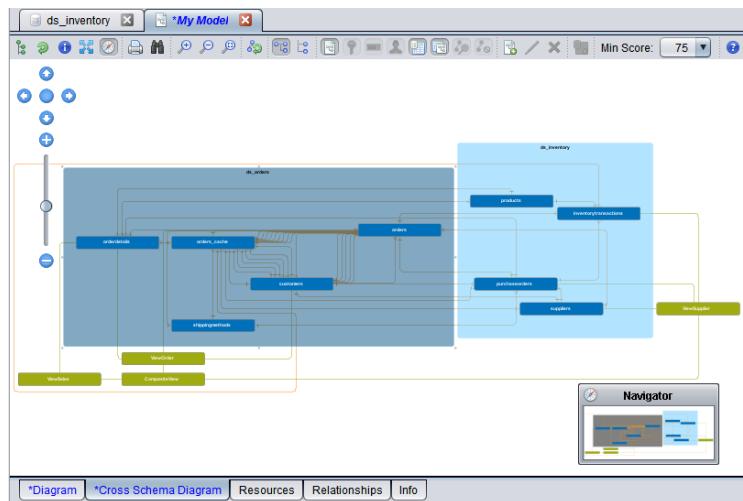
Getting Table and Relationship Information

On both the model Diagram and Cross Schema Diagram, you can drill down for details about the tables and relationships.

1. On the Cross Schema Diagram for My Model, display all tables with relationships.

2. Click the Show Relationship Detail button on the model toolbar.

   Discovery displays the Relationship Details panel under the diagram.

3. Select a relationship line.

   Discovery displays details about the selected relationship.
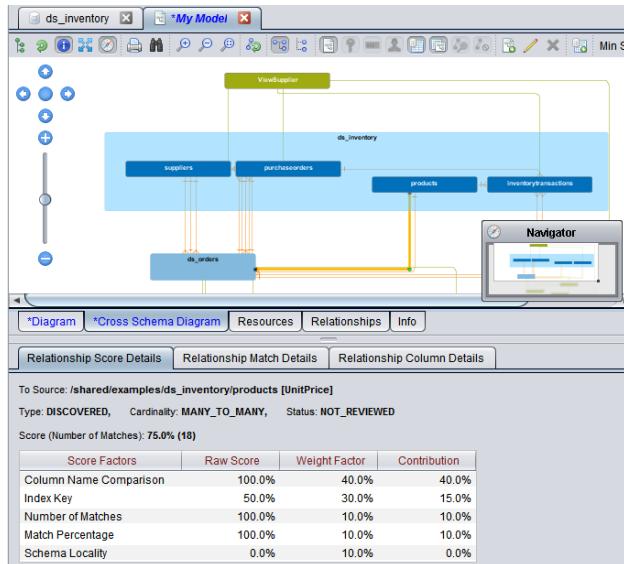
4. Click any other relationship and notice that the relationship details are updated.

   See Viewing Relationship Details, page 99, for more information about the Relationship Details panels.

5. Click the Diagram tab and select a relationship line.

   Discovery updates the Relationship Details tab with information for that relationship.

6. Double-click a relationship line and Discovery opens the Edit Relationship dialog.

   See Editing Relationships, page 104, for more information.

7. In the model Diagram tab, select any table, right-click, and choose View Table Details.

Discovery opens that table in the Studio workspace similar to:

8. Scroll to the right to see all of the table statistics. See Displaying Table Information, page 88, for more information.

Revising the Model and Relationships

You can revise the model visually as well as control the relationships that are included by:

- Rearranging the tables and relationships using the graphical editor.

- Filter ingthe display of resources and relationships using the toolbar controls.

- Validating and invalidating relationships.

- Adding a user-defined relationship.

To revise the model and relationships

1. In the model Diagram, select any table, right-click, and choose an option from the menu to display information about them.

2. Select a table and move it. Notice that all of its relationship connectors move with the table.

3. Select any relationship line and move it to see how it can be rearranged.

4. With the relationship line selected, click the Edit Relationship button on the model toolbar.

Discovery displays the Edit Relationship dialog.

5. Click the Status drop-down arrow and choose VALID to validate this relationship; then click **OK**.

6. Click the New Relationship button on the model toolbar and optionally define a new relationship for this model. See Defining Relationships, page 97, for more information.

7. Click the Hide Views button on the model toolbar to hide the display of the views (green boxes) and their dependencies.

8. Click the white box in the Navigator window and drag to move the entire contents of the diagram.

9. Use the zoom bar or your track wheel to zoom in or out. You can also use the toolbar buttons to adjust the zoom amount.

10. In the model toolbar above the diagram, change the Min Score to 50 to view relationships that have a lower score. Notice that the diagram now shows more relationship lines.

11. Click the Resources tab at the bottom of the window to view all resources included in the model.

12. Click the Relationships tab at the bottom of the window to view all relationships in the model and details about them including their scores.

Creating a View Based on the Model

To create a view based on the model

1. In the model Diagram, click and select several tables that have relationships. Do not select any views or tables that are not connected.

   Note: You can select multiple tables by dragging a box around the tables or by using Ctrl+click.

2. Click the New View button on the model toolbar.

   Discovery displays the Choose Parent Folder dialog.

3. Select the \shared\examples folder and accept New View as the view name.

Discovery creates the new view and opens it in the Studio workspace:
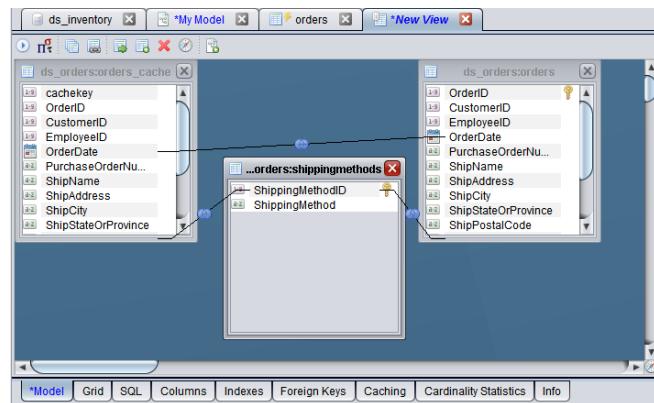
Discovery has automatically defined the join criteria based on known relationships.

Editing the View in Studio

Use Studio's view editing capabilities to edit the view and publish it. See the *TDV User Guide* for working with views.

To edit the view in Studio

1. In the Studio resource tree, open New View, which you created.

2. Click the Show Discovery Relationships button on the model toolbar to display all relationships that were discovered between the tables in the view.

Discovery displays the Add Relationships dialog.
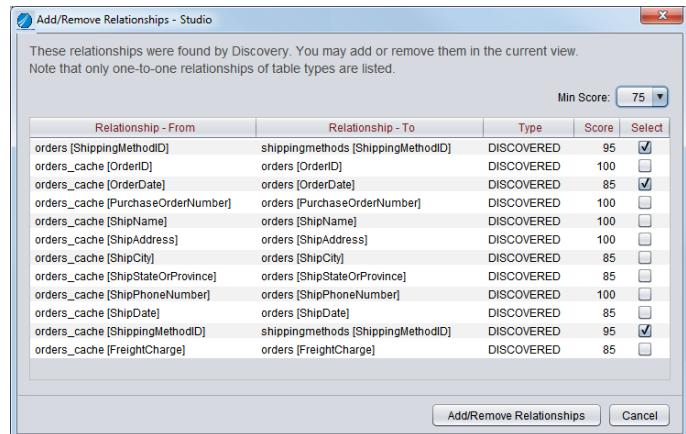
3. Check any additional relationships that you would like to include in this view.

4. Click Add Relationships.

Discovery displays the updated view.
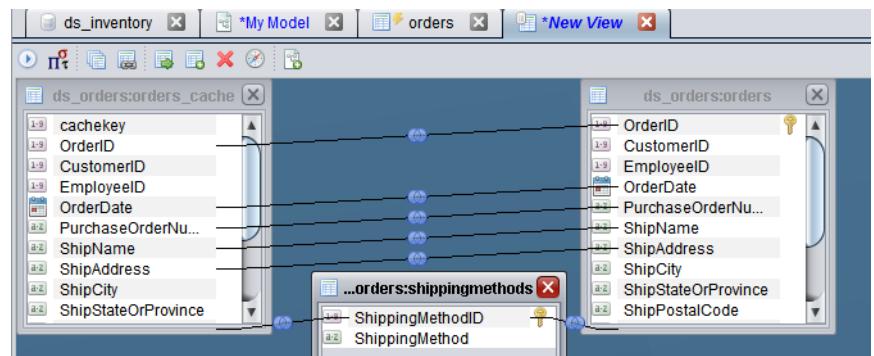
Using and Configuring Discovery

This topic describes how to configure data sources for use with Discovery. This involves defining, introspecting, indexing, and discovering the data sources so that they can be used in Discovery searches.

- Enabling Multicolumn Relationship Discovery, page 47

- Defining a Data Source for Discovery, page 49

- Working with Data Domains in Discovery, page 51

- Running Indexing and Relationship Discovery, page 57

- Reindexing and Rediscovering a Data Source, page 64

- Working with the Discovery Tasks Panel, page 65

- Tracking Index and Relationship Task History, page 68

- Configuring Indexing and Relationship Discovery, page 69

Enabling Multicolumn Relationship Discovery

You can configure Discovery to search for multicolumn relationships between tables. This increases the amount of time required by relationship discovery, but it could be important depending on your data. By default, Discovery searches only for single-column relationships. See About Multicolumn Relationships, page 20 for more information about configuring Discovery for multicolumn relationship discovery.

To configure Discovery to search for multi-column relationships

1. From the Administration menu, choose Configuration.

2. In the Configuration dialog, expand the Discovery/Relationship/MultiColumnKey branch.

3. Select Allow Multicolumn Key and change the value to True.

4. Adjust any of the other multicolumn configuration parameter values described below:

Configuration Parameter	Default	Use to...
Allow Multicolumn Key	False	True—The Discovery Relationship Scanner searches for multicolumn keys. The width of multicolumn key is set in Maximum Columns for Multicolumn Key. False—Only single-column relationships are discovered.
Check Only Many-to-Many Relationships	True	True—The multicolumn keys are only composed from many_to_many relationships. Doing this removes the combinations many-to-one and one-to-many to form a multicolumn key. False—All combinations are checked for relationships. This is more accurate but could take much longer.
Combine Multicolumn Relationships	True	True—Multicolumn relationships between two tables with score over 95 are consolidated into one relationship line on the Model diagram if more than five multicolumn relationships are discovered. False—All individual relationship lines are displayed on the Model diagram.
Maximum Columns for Multicolumn Key	3	The maximum number of columns for creating a multicolumn key. This value is used if Allow Multicolumn Key is True. This value can only be greater than or equal to 2. If this number is too large, Discovery tasks can run a very long time, so use caution when increasing this number.
Maximum Keys to Be Checked for Multicolumn Key	1000	The maximum number of keys to check for possible multicolumn keys between two tables. This value is used if Allow Multicolumn Key is True. If this value is large, it checks more combinations of multicolumn keys. However, it might run out of memory more easily.

5. Click **OK** to apply your configuration changes.

    When a relationship is formed by multiple columns, it is indicated as a multicolumn relationship on the model's Relationships tab. See Working with the Relationships Tab, page 114, for more information.

Defining a Data Source for Discovery

Follow the instructions in the "Working with Data Sources" chapter in the *TDV User Guide* to define a data source. Generally, TDV-supported relational data sources are supported by Discovery. But some TDV-supported data sources are not supported by Discovery. See Supported and Unsupported Data Sources, page 28 for more information.

When you define a data source, its metadata is introspected and stored in the TDV metadata repository. Discovery uses the metadata discovered during introspection to understand the structure and data definitions in the data source.

When you define a data source, you select the resources that you want to introspect and include in the data source definition. This process lets you limit the catalogs, schemas, and tables that are included and introspected at data source definition time. This is important for Discovery, because all resources must be indexed to be included in relationship discovery. Indexing can be time-consuming, so it is a good idea to deselect or filter out resources that you know are not needed.

General steps to configure a data source in Discovery

1. Define the data source (see Defining a Data Source for Discovery, page 49):

    a. Create a new data source and define its access credentials.

    b. Select the data source resources you want to include in your models.

    c. Optionally, filter the objects and relationships to discover in the data source.

2. Index and discover the data source (see Running Indexing and Relationship Discovery, page 57.)

Note: If you scheduled re-introspection for a data source that is used in a Discovery model, you need to reindex and rediscover relationships in that data source.

See these topics for more information about editing data sources in Discovery:

• Modifying a Data Source Definition, page 50

• Deleting a Resource Used by Discovery, page 50

Modifying a Data Source Definition

You can add and remove resources or re-introspect a data source that is used in a Discovery model, but you must reindex and rediscover relationships in the data source, and then refresh the model.

To modify a data source

1. Open the data source in Studio.

2. Optionally, add or remove resources:

   a. Select the Info tab, scroll down, and click Add/Remove Resources.

   b. Select or clear the resources you want to add or remove.

   Studio displays the changes you have made.

   c. Click **Finish** to proceed with introspection and these changes.

3. If necessary, re-introspect a data source.

   Note: Re-introspection does not mean that metadata has changed. You need to reindex and rediscover relationships after re-introspection finds changes to resources that you are interested in modeling.

4. To re-index and rediscover relationships for the data source at this time, select the Relationship Discovery tab and click **Scheduler** (see Running Indexing and Relationship Discovery, page 57, for more information.

Deleting a Resource Used by Discovery

You can delete a resource using Studio's standard delete functionality.

To delete a resource

1. Select the resource in the Studio resource tree.

2. Click the Delete button.

If you delete a resource that is used in a Discovery model, the effects of the deletion depend on the resource type:

• If you delete a data source, the model name changes to red indicating that an underlying data source is no longer valid.

• If you delete a view, the view is removed from the model. You must refresh the model to see the effects of the change.

Working with Data Domains in Discovery

Discovery uses indexes on the data sources to access data efficiently and quickly. Discovery then can discover relationships between data source tables. It is not required to index and discover a data source; you might have defined Primary Key-Foreign Key relationships, or want to configure relationships manually. But to discover relationships efficiently, you need to index and discover the relationships.

This section explains important factors to consider for the indexing and relationship discovery process and describes using data sampling for large data sources.

This topic contains:

Using Data Domains

Discovery lets you index and discover related data that is stored in various patterns. A data domain is defined as a type of data such as phone number, social security number, or date. In each of these examples, the data might be stored in a variety of patterns. For example, if you have phone number data that is stored in different data sources in different patterns such as 800-123-4567 and (800) 123-4567, you can define a phone number domain that lists all the possible patterns in which a phone number might be stored. This enables Discovery to find and match all columns that contain related phone numbers even when their patterns are different.

You can define one or more data domains and a set of patterns for each. If a data domains is enabled during indexing and relationship discovery, Discovery finds and matches data that conforms to the patterns in the domain.

In a model, if data domains were enabled during relationship discovery, the columns that were discovered as a result of domain analysis are indicated.

To use domain analysis for relationship discovery, you must:

- Enable domain analysis. See Enabling Data Domains, page 52, for information about the configuration parameters that control domain analysis.

- Define data domains and their patterns. See Defining Data Domains and Patterns, page 52.

- Editing Domains and Patterns, page 55.

- Deleting Domains and Patterns, page 55.

- Saving and Loading Domain Definitions, page 56.

Enabling Data Domains

To use data domains during indexing and discovery, you must first enable data domains.

Note: If you have created any indexes or done relationship discovery, you must first delete all indexes prior to running indexing and relationship discovery with domain analysis enabled.

To enable data domains

1. From the Studio Administration menu, choose Configuration.

2. Under Discovery, expand the Indexing branch.

3. Set the Use Data Domains for Discovery value to True.

4. Click **Apply** and then **OK** to apply your configuration changes.

See Configuring Data Domains, page 71, for information about domain analysis configuration parameters.

Defining Data Domains and Patterns

For each type of data that might be stored in a variety of patterns, you must define a domain and then define one or more patterns that describe the various ways it might be represented.

To define a data domain

1. In Studio, click on the Discovery tab at the left side of the window.

2. Click the Domains tab.

Discovery displays the Domains tab, where you can define the data domains and their patterns.
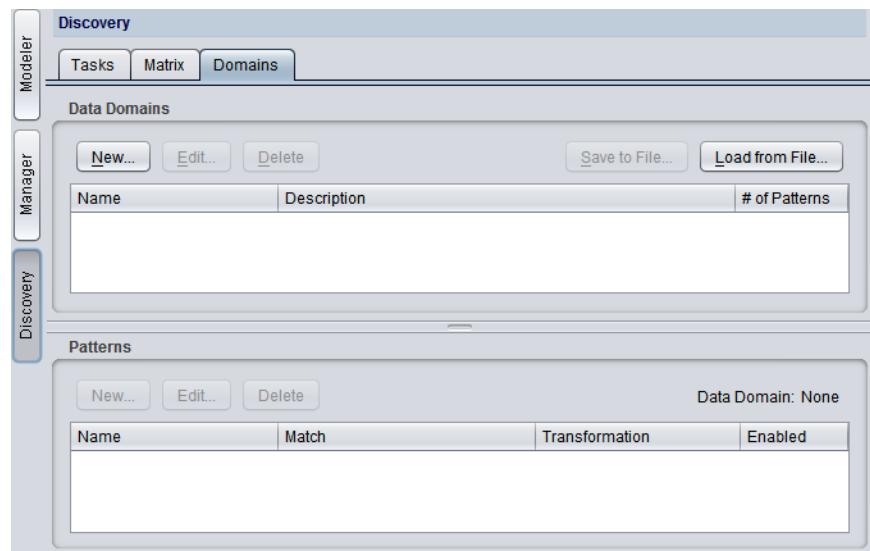
3. Optionally, click Load from File if you have created a domain definitions file. See Saving and Loading Domain Definitions, page 56.

4. Under Data Domains, click New.

   Note: The configuration parameter Use Data Domains for Discovery must be set to True to define data domains. If not, Discovery reminds you to enable data domains.

   Discovery displays the New Data Domain dialog in which you can define the data domain.

5. Under Data Domain, enter a name and description for this data domain.
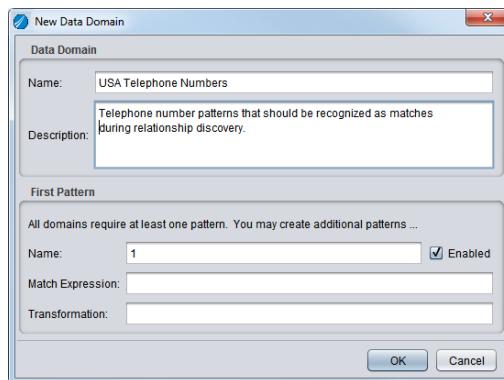
6. Under First Pattern, you must define the first pattern for the new domain:

— Name—Enter a pattern name or number.

— Match Expression—Enter a regular expression that is expected to match the string as it is found in the database.

— Transformation—Enter the replacement string as it would be declared if doing a typical programmatic find/replace operation with regular expressions, where the Match Expression is the "find" expression and the Transformation expression is the "replace" expression.

See About Pattern Expressions, page 18, for more information about how Discovery interprets and uses pattern expressions.

This example illustrates a new domain and its first pattern.
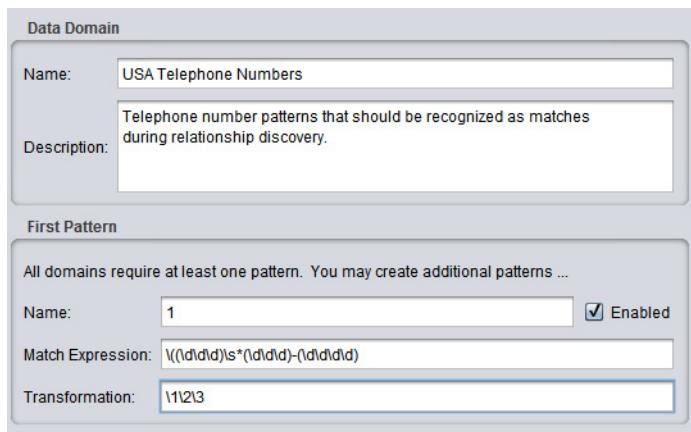
7. Under First Pattern, check Enabled to make Discovery search for data that fits this pattern.

8. Click **OK**.

Discovery displays the new data domain on the Domains tab.

9. Optionally, click New under Patterns to define another pattern for this domain and repeat Steps 6 through 8.

You can define as many patterns as needed. You can also disable or enable individual patterns by checking or unchecking the adjacent Enabled checkboxes.

Editing Domains and Patterns

You can edit any existing domain or pattern. See Defining Data Domains and Patterns, page 52, for more information.

To edit a data domain or pattern

1. In Studio, click on the Discovery tab at the left side of the window.
2. Click the Domains tab.
3. Select the domain or pattern you want to edit.
4. Click the Edit button under Data Domains or Patterns.
5. Edit the domain or pattern and click **OK** to save your changes.

Deleting Domains and Patterns

You can delete any existing domain or pattern.

To delete a data domain or pattern

1. In Studio, click on the Discovery tab at the left side of the window.
2. Click the Domains tab.
3. Select the domain or pattern you want to delete.

4. Click the Delete button under Data Domains or Patterns.

5. Confirm the deletion.

Saving and Loading Domain Definitions

You can save domain definitions that you created in Discovery to a comma-separated values (.csv) file. You can also create a .csv file containing domain definitions and load them into Discovery. The format of the domain definitions file is:

	A	B	C	D	E	F
1	Domain Name	Domain Description	Pattern Name	Pattern Match	Pattern Transformation	Pattern Enabled
2	Telephone Numbers	Phone number patterns	1	\(\(\d\d\d\)\)\s*(\d\d\d)-(\d\d\d\d)	\1\2\3	Y
3						
4						

my domain file

The domain name and domain description of every pattern are repeated on every row for every pattern in the domain. Each pattern gets one row in the table. Domains do not have separate rows: they are implied by the existence of patterns with domain names that match each other. Every row in the table should be unique if you consider both the domain name and the pattern name to be the two keys that define uniqueness.

Working with Denormalized Data

Your data source tables might include columns with denormalized data. That is, a table column might include multiple values separated by a delimiter character. By default during relationship discovery, Discovery compares the entire string in a column as a single value. So if a column contains comma-separated values, the individual values would not be compared during relationship discovery. However, you can configure Discovery to compare individual values within a denormalized table column during relationship discovery.

To configure relationship discovery for denormalized data

1. In the Studio resource tree, open the data source table that contains denormalized data.

   See Table Statistics Displayed for Relationship Discovery, page 21, for a display of table statistics included with Discovery.

2. For the column that contains denormalized data, type the delimiter that separates the data values in the Denormalization column (the right column in the table).

Delimiters can be any character of your choice. Typical characters are comma, semicolon, colon, space, or tab. To enter a space, type \s. To enter a tab, type \t.

3. The table must be reindexed and rediscovered to see the results with denormalized data.

   Discovery lets you know if a relationship is based on denormalized data. See Viewing Relationship Details, page 99, for more information.

Running Indexing and Relationship Discovery

Discovery provides two ways for you to run indexing and discovery:

• From the Studio Modeler panel, when you define or open a data source, you can run the Index and Relationship Task Scheduler from the Relationship Discovery tab. By default, all index and discovery check boxes are selected for all resources in this data source. This process is described in Defining a Data Source for Discovery, page 49.

• From the Studio Discovery panel Tasks tab, you can run index and relationship discovery tasks between any known data sources. A model containing any of these data sources then shows all relationships between the selected data sources and, if specified, within the data sources.

This topic contains:

• Indexing and Discovering a Data Source, page 57

• Indexing and Discovering Relationships Across Data Sources, page 60

• Reindexing and Rediscovering a Data Source, page 64

Indexing and Discovering a Data Source

Whenever you need to do it, you can run the indexing and discovery processes on the open data source. You can do this:

• Whenever you add a new data source.

• If you add or remove resources in an existing data source.

• If you re-introspect an existing data source and there are changes as a result.

You can see if a data source has been indexed by opening the data source. The round button on the Relationship Discovery tab is green if the indexing and relationship discovery process has been run on the data source, and it is labeled Finished.

To index and discover a data source

1. Open the data source in Studio.

2. Select the Relationship Discovery tab, which looks like this prior to indexing.
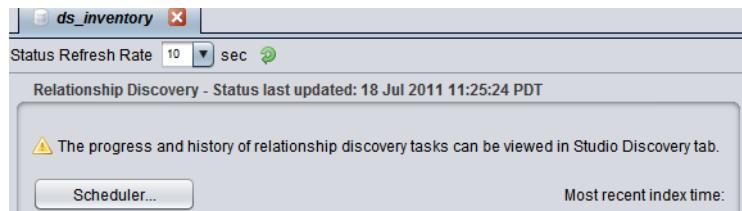
3. Click **Scheduler**.

   The Status field changes to Running, and the percent complete and time remaining are displayed.

   The icon on the Relationship Discovery tab indicates current status

Tab Icon	Meaning
?	Gathering—Discovery is gathering the task status.
White	Not Yet Run—The data source has not been indexed.
Yellow	Running—The indexing and discovery process is running.
Green	Finished—The indexing and discovery process is finished.
Red	Failed—The indexing and discovery process has failed.
Gray	Cancelled—The indexing and discovery process has been cancelled by the user.

For more details about the status of the index and discovery tasks, go to the Discovery Tasks panel. See Tracking Index and Relationship Task History, page 68.

Discovery displays the Index and Relationship Task Scheduler dialog with all check boxes selected to index and discovery relationships within this data source, as in the example below.
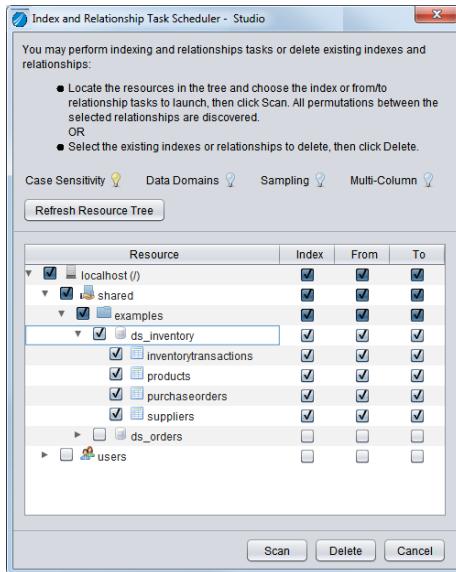
If you want to index and discover relationships between this data source and any others in the list, check the appropriate boxes. See Indexing and Discovering Relationships Across Data Sources, page 60.

Note: By default, the index and relationship discovery is calculated with case sensitivity turned on. If you want to ignore case differences, you might need to change the case sensitivity configuration setting as described in Configuring Case Sensitivity, page 71.

4. Click Scan to index and discover relationships.

The progress and history of the indexing and relationship discovery can be viewed on the Studio Discovery tab.

When the indexing and relationship discovery is successful and complete, the status is Finished. You can also tell if the data source has been indexed and discovered by the green check mark on Relationship Discovery tab.

5. If this data source is included in an existing model, open the model and click Refresh to view this data source and its relationships in the model.

See Working with Models in Discovery, page 77, for more information about how to view, validate, or invalidate the discovered relationships.

Indexing and Discovering Relationships Across Data Sources

For Discovery to find relationships across multiple data sources, you must follow the steps in this section. You can choose to find relationships within a resource or between resources.

About Selecting From and To Relationships

Selecting From and To relationships determines the set of tables between which relationships are discovered. It compares tables in the From side to tables in the To side. For example, assume you have these resources:

Datasource A

Table A1

Table A2

DataSource B

Table B1

Table B2

The table below explains the results of different combinations of selections.

Select	Relationships between these are discovered
From Table A1 and To Table A2	Table A1 <-> Table A2
From Datasource A and To Datasource B	Table A1 <-> Table B1
	Table A1 <-> Table B2
	Table A2 <-> Table B1
	Table A2 <-> Table B2
From and To Datasource A and From and To Datasource B	Table A1 <-> Table A2
	Table A1 <-> Table B1
	Table A1 <-> Table B2
	Table A2 <-> Table B1
	Table A2 <-> Table B2
	Table B1 <-> Table B2

To index and discover relationships across data sources

1. In Studio, click on the Discovery tab at the left side of the window.

    Discovery displays the Discovery panel with the Tasks tab on top.

2. Click **Scheduler**.

Discovery opens the Index and Relationship Task Scheduler dialog, which lists all resources that are known by TDV and are available to be indexed and discovered.
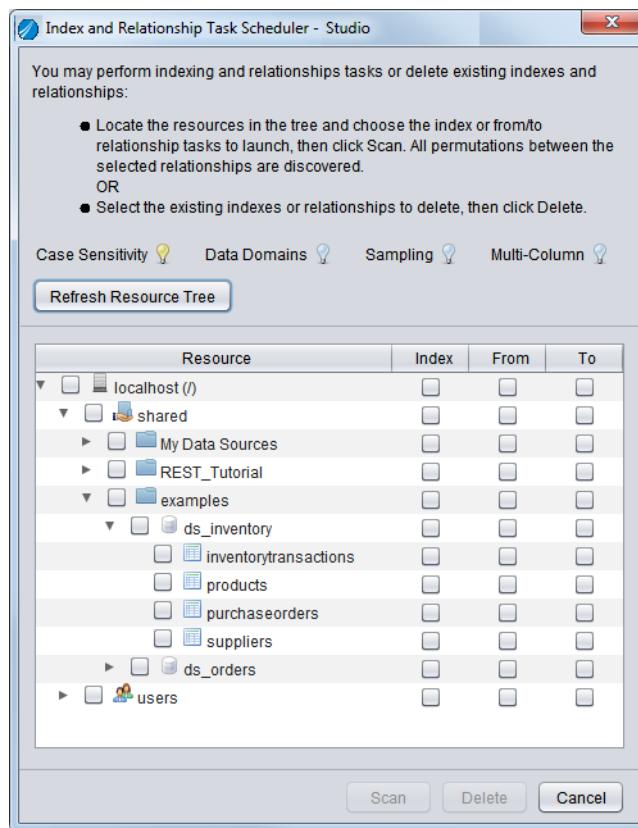
Note: By default, the index and relationship discovery is calculated with case sensitivity turned on (the icon next to Case Sensitivity is enabled). If you want to ignore case differences, you might need to change the case sensitivity configuration setting as described in Configuring Case Sensitivity, page 71.

3. If necessary, expand the resource tree to display the resources you want to index or discover.

4. Check the boxes next to the resources you want to index, discover from-relationships, or discover to-relationships.

   If you want to index and discover everything for a data source, click the check boxes to the right of the data source name. You can also only-index or only-discover relationships for the data source or its individual resources.

   Note: If you re-index a data source or resource, it is a good idea to rediscover relationships based on the new index.

The two groups of objects during relationship discovery are the from-group and the to-group. Relationships are discovered between items in the from-group and items in the to-group. If you want to discover relationships within a resource, you need to check both the From and To check boxes for that resource. See Reindexing and Rediscovering a Data Source, page 64, for more information.

Note: The following conditions apply:

— You must index a resource before you run any type of relationship discovery.

— If you have previously indexed the data source and the index is current, you do not need to reindex the data source before discovering the from/to relationships. For example, if you ran indexing and discovery from the open data source as described in Indexing and Discovering a Data Source, page 57, it might not be necessary to run indexing again.

— The From and To terms used in this case can be reversed when selecting the check boxes.

The example below shows what you would check if you wanted to index and discover all relationships between the two sample data sources ds_inventory and ds_orders.

5. Click Scan.

Discovery launches the tasks you selected and the Tasks tab displays the status of each task.

Discovery

Tasks | Matrix

Statistics

	Running	Waiting	Failed	Successful	Total
Indexing	1	0	0	15	1(
Relationship	0	2	0	13	1!
Total	1	2	0	28	3:

Tasks

Filter: Show All [31]   Scheduler...   Cancel   Clear Tasks   Delete All...

Type	Name	Status	End Time	Durati...	Ti...	R...
Discovery Scanner	Index and Scan /shared/examp...	Finis...		1.002s	0.5...	
Indexer	Index /shared/examples/ds_or...	Runn...		1.002s	0.3...	
Relationship ...	Scan /shared/examples/ds_or...	Run ...		0.000s	0.3...	
Post Process	Post process /shared/example...	Run ...		0.000s	0.5...	
Discovery Scanner	Index and Scan /shared/examp...	Finis...	5:49:13 PM	1.730s	0.0...	891
Indexer	Index /shared/examples/ds_inv...	Finis...	5:49:12 PM	1.153s	0.0...	90
Table Inde...	Index /shared/examples/ds_inv...	Finis...	5:49:12 PM	1.049s	0.0...	35
Table Inde...	Index /shared/examples/ds_inv...	Finis...	5:49:12 PM	0.855s	0.0...	20
Table Inde...	Index /shared/examples/ds_inv...	Finis...	5:49:11 PM	0.452s	0.0...	25
Table Inde...	Index /shared/examples/ds_inv...	Finis...	5:49:12 PM	0.615s	0.0...	10

The progress bar in the End Time column shows you the progress for each task in real time.

Reindexing and Rediscovering a Data Source

If you make changes to a data source that you want to appear in a Discovery model, you need to reindex and rediscover relationships in that data source.

When you re-index or re-discover relationships, the old relationships for the same data source are overwritten.

- To reindex an entire data source, follow the steps in Indexing and Discovering a Data Source, page 57, and then refresh the model for the changes to take effect.

- To reindex selected resources, follow the steps in Indexing and Discovering Relationships Across Data Sources, page 60, and then refresh the model for the changes to take effect.

Working with the Discovery Tasks Panel

The Tasks panel displays all recently run tasks and their status. This topic contains:

Viewing Discovery Tasks Panel

During index and discovery, Discovery performs a series of tasks to gather the data, index the data source, and discover relationships. You can see the status of the discovery tasks, error messages, and other information about this process in the Discovery Tasks panel.

To view the Discovery Tasks panel

1. In Studio, click on the Discovery tab at the left side of the window.

2. If necessary, click the Tasks tab.

Filtering Discovery Tasks

You can filter the types of tasks displayed in the Discovery Tasks table.

To filter Discovery Tasks

1. In Studio, click on the Discovery tab at the left side of the window.

2. On the Discovery Tasks tab, click the Filter drop-down list.

3. Choose an option from the Filter drop-down list as follows.

Filter	Displays
Show All [#]	All tasks.
Indexing [#]	Only indexing tasks.

Filter	Displays
Relationship [#]	Only relationship tasks.
Running [#]	Only tasks that are still running.
Waiting [#]	Only tasks that are waiting to complete.
Failed [#]	Only failed tasks.
Successful [#]	Only successful tasks.
Show History [#]	All tasks that have been launched during this current Studio session.

Deleting All Discovery Indexes and Relationships

You might want to delete all indexes and relationships in these cases:

- To free up disk space and do not think that you want to discover or re-discover relationships involving the index.

- If you change the case sensitivity setting. In this case, the current indexes and relationships are no longer valid and must be deleted prior to re-indexing and re-discovering relationships.

Note: The Delete All tasks function purges the entire indexed and relationship data from the database and the file system. All active tasks (if any) will be canceled and related data will be deleted as well.

To delete all indexes and relationships

1. In Studio, click on the Discovery tab at the left side of the window.

2. On the Discovery Tasks tab, click Delete All.

   Discovery displays a warning and confirmation message.

3. Click Delete All.

   Discovery deletes the indexes and relationships and shows the tasks Discovery ran to perform the deletes.

4. Optionally, you can clear the Delete tasks by clicking the Clear Tasks button.

Clearing Discovery Tasks

You can clear the Tasks table display of all tasks. This is useful if you want to see only the next new set of tasks.

To clear Discovery tasks

1. In Studio, click on the Manager tab at the left side of the window.

2. On the Discovery Tasks tab, click Clear Tasks.

   Discovery displays a confirmation message.

3. Click Yes.

Getting Discovery Task Details

You can view details about a particular task in the Task Detail dialog. This is useful if you want to:

- Display all details for a particular task.

- See the complete text for an error message.

To get details about a task

1. In Studio, click on the Manager tab at the left side of the window.

2. On the Discovery Tasks tab, select a task.

3. Click the tasks detail button.

   Discovery displays a Tasks Detail panel with more information about that task as in this example.
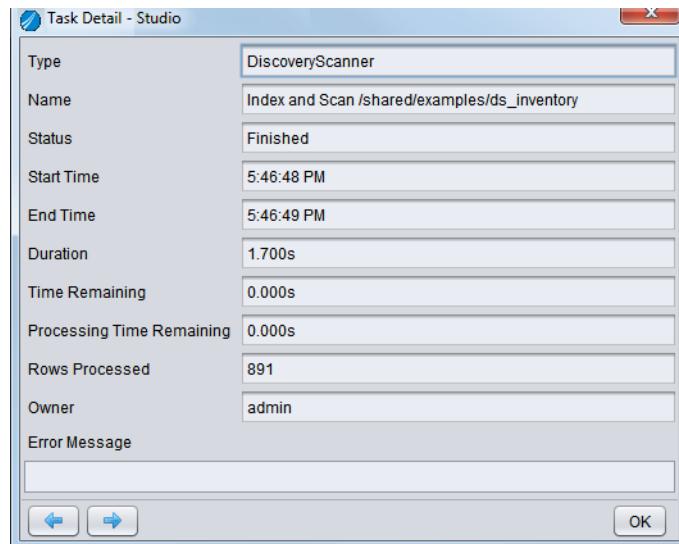
4. YClick the arrows at the bottom of the dialog to view details about the previous or next task.

5. Click **OK** to return to the Tasks panel.

Tracking Index and Relationship Task History

You can track which data sources have been indexed and which combinations of resources have been searched for possible relationships on the Discovery Matrix tab.

To view the index and relationship task history

1. In Studio, click on the Discovery tab at the left side of the window.

2. Click the Matrix tab.

    Studio displays a matrix of all resources in the Studio resource tree, and shows which have been indexed and which combinations have been search for possible relationships.

	Indexed	shared	examples	ds_inventory	ds_orders	cache_status
▼ 🖳 shared	2:58:41 PM	2:58:43 PM	2:58:43 PM	2:58:43 PM	2:58:43 PM	2:58:43 PM
▼ 📁 examples	2:58:41 PM		2:58:43 PM	2:58:43 PM	2:58:43 PM	2:58:43 PM
▶ 🗄 ds_inventory	2:58:41 PM			2:58:43 PM	2:58:43 PM	2:58:43 PM
▼ 🗄 ds_orders	2:58:41 PM				2:58:43 PM	2:58:43 PM
📄 cache_status	2:58:41 PM					2:58:43 PM
📄 cache_tracking	2:58:41 PM					
📄 customers	2:58:41 PM					
📄 orderdetails	2:58:41 PM					
📄 orders	2:58:41 PM					
📄 orders_cache	2:58:41 PM					
📄 shippingmethods	2:58:41 PM					
▶ 👥 users						

The Indexed column contains a date and time for all data resource that have been indexed. If a date and time exists in the intersection of two data resources, it means that they were compared and checked for relationships at that date and time.

3. If necessary, expand and collapse the resource tree to the left and use the scroll bars to see the data.

Getting More Information about Relationships

Studio includes two system tables specific to Discovery. These tables provide more details about the discovered relationships in all resources that have been indexed and discovered, and for which the current user has privileges:

- ALL_RELATIONSHIP_COLUMNS—Lists every relationship, its ID, its "from" column/table/schema/catalog/data source and their IDs, and the same information for the "to" column.

- ALL_RELATIONSHIPS—Lists every relationship, its type, cardinality, status, column information, and score information including the actual factors discovered, such as number of matches.

A summary of this information specific to each model can be viewed on the model's Relationships tab. (See , for more information.) However, the specific factor information discovered for each relationship is available only in the ALL_RELATIONSHIPS table.

Note: For a complete description of these two tables, see the *TDV Reference Guide*.

To view the Discovery system tables

1. In the Studio resource tree, expand the Data Services branch display the system tables: Data Services > Databases > system.

2. Select ALL_RELATIONSHIPS.

3. Right click and select Open from the pop-up menu.

   Discovery opens the table, displaying its columns and column types.

4. Click Show Contents to display the actual relationships table in the lower pane.

5. To see all details about the relationships, scroll to the right.

Configuring Indexing and Relationship Discovery

Discovery provides configuration parameters to help you control some aspects of indexing and relationship discovery. Default settings should be OK in most situations, but you can adjust the settings if necessary to optimize for your particular data sources.

Indexing Configuration Parameters

- Use Case Sensitivity for Discovery—See Configuring Case Sensitivity, page 71.

- Data Domain Sample Size—See Configuring Data Domains, page 71.

- Use Data Domains for Discovery—See Configuring Data Domains, page 71.

- Index Directory—See Changing the Index Directory, page 70.

- Enabling Indexing on Views—See Configuring Indexing and Discovery for Native Oracle Views, page 73.

- Maximum Concurrent Tasks—See Adjusting the Maximum Concurrent Tasks, page 75.

- Sampling Is Enabled—See Configuring Data Sampling, page 72.

- Sampling Size—See Configuring Data Sampling, page 72.

Relationship Configuration Parameters

- MultiColumnKey—See Enabling Multicolumn Relationship Discovery, page 47.

- Minimum Score—See Adjusting the Minimum Score for Relationship Discovery, page 73.

- Minimum Unique Percentage—See Adjusting the Minimum Unique Percentage, page 74.

- Use Discovery Relationships in View Editor—See Configuring Relationship Display in Views, page 126.

- Weight for <x>—See Adjusting the Weights of the RPS Factors, page 74.

Changing the Index Directory

To change the index directory

1. In Studio, choose Configuration from the Administration menu.

2. Expand the Discovery and then the Indexing branch.

3. Select Index Directory and change the location.

4. Click **Apply** and then **OK** to apply your configuration changes.

Note: If you change the index directory, and then use the Delete All button on the Discovery/Tasks panel, the previously generated index in the old index directory is not deleted.

Configuring Case Sensitivity

By default, Discovery indexes-and-relationship discovery treats data values as different if their case does not match. You can specify that Discovery ignore case differences.

Note: If you want to change the case sensitivity configuration parameter as described here, you must delete any existing indexes and relationships before you can apply your change.

To change the case sensitivity setting

1. In Studio, choose Configuration from the Administration menu.

2. Expand the Discovery and then the Indexing branch.

3. Select Use Case Sensitivity for Discovery and change the value:

Setting	Result
True (default)	Case differences are taken into account during indexing and relationship discovery.
False	Discovery ignores case differences during indexing and relationship discovery.

4. Click **Apply** and then **OK** to apply your configuration changes.

Configuring Data Domains

You can configure the indexing and discovery process to consider data domains that you have defined in Discovery. See Using Data Domains, page 51, for more information about data domains and how to define them.

To configure data domains

1. From the Studio Administration menu, choose Configuration.

2. Under Discovery, expand the Indexing branch.

3. Set these data domain configuration parameters.

Parameter	Default	Result
Use Data Domains for Discovery	False	Enable or disable use of data domains during indexing and discovery. False—Discovery ignores data domain patterns during indexing and discovery. True—Discovery takes the defined data domain patterns into account during indexing and discovery. You must have defined data domain patterns as described in Using Data Domains, page 51, or there is no effect. This setting can only be modified after removing all Discovery indexes and tasks from the Discovery Tasks panel in Studio.
Data Domain Sample Size	64	If Discovery data domains are enabled using the parameter above, Discovery analyzes the data according to this sample size. If the number of entries is less than this value, all data is indexed. Minimum value: 16

4. Click **Apply** and then **OK** to apply your configuration changes.

Configuring Data Sampling

If you have especially large data sources, it can be beneficial to use data sampling. See About Data Sampling, page 13, for more information about how it works and when to use it.

To configure data sampling

1. From the Studio Administration menu, choose Configuration.

2. Under Discovery, expand the Indexing branch.

3. Set these data sampling configuration parameters:

Parameter	Default	Result
Sampling Is Enabled	False	Set to True to enable data sampling.
		False—Disables data sampling regardless of table size. The default.
		True—Enables data sampling on tables and views that exceed the number of rows specified by the Sampling Size configuration parameter.
Sampling Size	100000	Specify the threshold above which data sampling is applied.

4. Click **Apply** and then **OK** to apply your configuration changes.

Configuring Indexing and Discovery for Native Oracle Views

You can enable indexing and relationship discovery for native Oracle views. This feature is provided for Oracle data sources that have native Oracle views and has no effect on TDV views which cannot be indexed or discovered.

To configure indexing and discovery for native Oracle views

1. In Studio, choose Configuration from the Administration menu.

2. Expand the Discovery and then the Indexing branch.

3. Select Enable Indexing on Views and change the value.

   This effects native Oracle views only. The default value is False.

4. Click **Apply** and then **OK** to apply your configuration changes.

Adjusting the Minimum Score for Relationship Discovery

You can change the threshold score above which a relationship is created, in the range from 0 to 100. You can further filter the relationships displayed in the model using the Min Score setting in the model diagram as described in Viewing Relationships Based on RPS Score, page 107.

Note: Use caution when setting this value. If the value is too low, it could involve so many columns that it might not be able to finish relationship discovery. Also, if you set the minimum score to zero, the results would not be very useful because all columns would be related to all other columns.

To change the minimum score for relationship discovery

1. In Studio, choose Configuration from the Administration menu.

2. Expand the Discovery and then the Relationship branch.

3. Select Minimum Score and change the value.

4. Click **Apply** and then **OK** to apply your configuration changes.

Adjusting the Minimum Unique Percentage

You can change the threshold percentage above which a column is considered to contain a unique set of values in the range of 0 to 100. The default value is 90. This setting can affect the Match Percentage Factor when Discovery evaluates relationships. See About the Relationship Probability Score (RPS), page 23, for more information about this factor.

To change the threshold of uniqueness

1. In Studio, choose Configuration from the Administration menu.

2. Expand the Discovery and then the Relationship branches.

3. Select Minimum Unique Percentage and change the value.

4. Click **Apply** and then **OK** to apply your configuration changes.

Adjusting the Weights of the RPS Factors

You can change the weight (importance) of the RPS factors using Studio's configuration parameters. Adjusting the weights affects which relationships and how many relationships are discovered. The five weight percentages ideally add up to 100. However, this is not required, because each weight is calculated according to the following formula: weight = weight percentage / sum of all weights.

The default settings, which apply to non-string data types, are as follows:

Column Name Comparison Factor (40%)
Index Key Factor (30%)
Match Percentage Factor (10%)
Number of Matches Factor (10%)
Schema Locality Factor (10%)

String data types (CHAR, LONGVARCHAR, and VARCHAR) are set so that more weight is given to the data and less to the column name. String factor weight settings, which cannot be modified:

Name Factor (10%)
Relationship Cardinality Factor (30%)
Match Percentage Factor (10%)
Number of Matches Factor (40%)
Locality Factor (10%)

The RPS weights that you can configure as described below apply to non-string data types only.

To change the relationship weight factors

1. In Studio, choose Configuration from the Administration menu.

2. Expand the Discovery and then the Relationship branches.

   The current weights for each factor are displayed.

3. Select the weight you want to change and adjust the value. The minimum value is 0; the maximum value is 100. The five weight configuration parameters should ideally add up to 100.

4. Click Apply and then OK to apply your configuration changes.

Adjusting the Maximum Concurrent Tasks

Generally, the more Discovery tasks that can run concurrently, the faster the results. However, this depends upon your CPU, memory, and how large the data set and tables are.

The maximum number of Discovery tasks that run concurrently is controlled by the Maximum Concurrent Tasks configuration parameter. By default, this parameter is set to 10. Depending on the performance of Discovery in your environment, you might want to adjust this value.

To change the maximum number of concurrent tasks

1. In Studio, choose Configuration from the Administration menu.

2. Expand the Discovery and then the Indexing branches.

3. Select Maximum Concurrent Tasks and change the value.

4. Click **Apply** and then **OK** to apply your configuration changes.

Working with Models in Discovery

This topic describes how to create a model and refine it to display the data entities in which you are interested.

About Models in Discovery

A Discovery model helps you easily view and understand the resources and relationships in your data resources. A model shows resource tables and views and their relationships to each other and lets you review, edit, and filter those relationships. Part of defining the data source in Discovery is running the indexing and relationship discovery process, which is when Discovery learns about these relationships. (See Working with Data Domains in Discovery, page 51.) A relationship can be a discovered, it can be a foreign key relationship, or it can be manually defined.

You can use a Discovery model to:

- View and learn about the tables in a data source and their related views.
- Graphically explore relationships between resources, including tables, columns, and views.
- Graphically explore relationships between database schemas.
- Explore the details about relationships such as type and cardinality.
- Validate and invalidate discovered relationships.
- Add known relationships.
- Rearrange the model elements and print the model.
- Create a new view based on the tables and relationships displayed model

Overview of Working with Models

Working with models is an iterative process and the steps below are not necessarily sequential

1. Create a new model.

2. Add one or more data sources, catalogs, schemas, tables, or views to the model.

3. Explore the tables, views, and their relationships in the Model and Cross Schema diagrams.

4. Validate or invalidate relationships and modify relationship details.

5. Add relationships.

6. Print the model.

7. Create a new view (covered in Creating Views with Discovery, page 121).

Creating a Model

You create a new model using Studio. Initially, the model is an empty container. After you have created at least one data source in Studio, you can add one or more of these resources to a model: data sources, catalogs, schemas, tables, or existing views. When you add a resource to the model, you can graphically view and edit the tables in the data source. If the data source has been indexed and discovered, you can also view and edit the relationships discovered between the model resources.

This topic contains:

To create a model

1. In Studio, select the Shared folder, right-click, and choose New Model.

2. Enter a name for the new model.

3. Optionally, check the Copy privileges from parent folder option.

4. Click **OK**.

   Discovery opens a new empty model in the Studio workspace on the Diagram tab.You can now follow the steps in Adding a Resource to a Model, page 79.
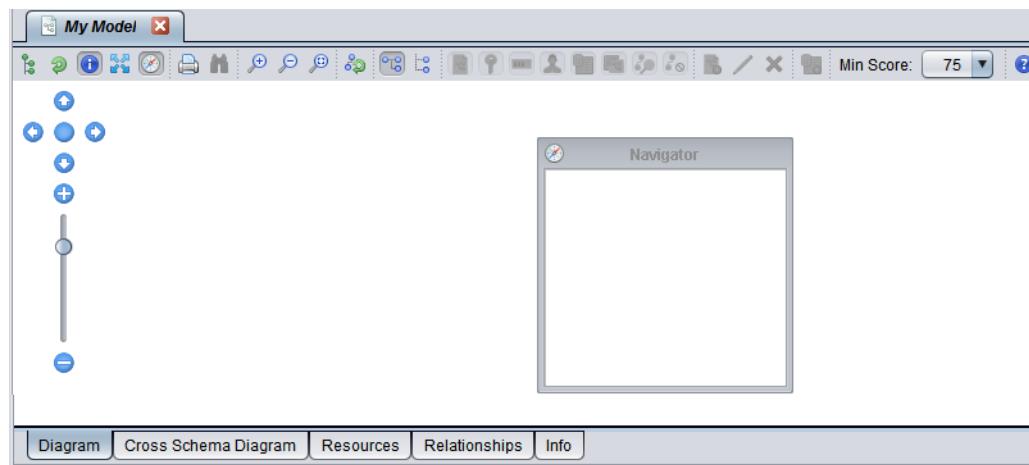
Adding a Resource to a Model

Any supported data source with tabular resources can be added to a model. To see the discovered relationships between the data source tables, you need to index and discover the relationships prior to adding them to the model. (See Working with Data Domains in Discovery, page 51.)

When you add resources to a model, Discovery always displays:

• All EXPLICIT resources—Resources explicitly selected or dragged into the model.

• All IMPLICIT resources—Resources whose parents were selected or dragged into the model.

• All RELATED resources —Resources related to EXPLICIT or IMPLICIT resources via a hard relationship (FOREIGN KEY, USER DEFINED, or DISCOVERED).

• All DEPENDENT resources—Resources related to EXPLICIT, IMPLICIT, or RELATED resources. This is currently limited to views.

Resources in the Discovery model can be classified as EXPLICIT or RELATED among other types. EXPLICIT types refer to the resources added in the model explicitly by users. When user creates user-defined relationships those resources not identified as EXPLICIT can be marked as RELATED. This can occur when a resource is added explicitly and the added resource was related to another resource, not yet in the current model, by "user defined relationships". This secondary resource is marked as RELATED. This "RELATED" link chain can span more than one level. The related links will be discovered until it is fully exhausted.

Consider three resources A, B, and C where user created "user defined relationship" between A and B, and B and C prior. By a transitive nature of these relationships, A is related to C via B. When A is added to a model both B and C will be included as "RELATED" types. Subsequent Discovery model load/save/refresh will reflect this logic. This will also affect deletion of resources as all related resources are removed when necessary.

You can determine why a resource is included using the model Resources tab. See Working with the Model Resources Tab, page 112.

You can add a resource to a model in one of two ways: from the Studio resource tree, or using the Add/Remove resources button in the model Diagram.

To add a resource to a model from the Studio resource tree

1. Select the resource—data source, catalog, schema, table, or view—in the Studio resource tree and drag it onto the model workspace.

   Discovery displays the Diagram tab.
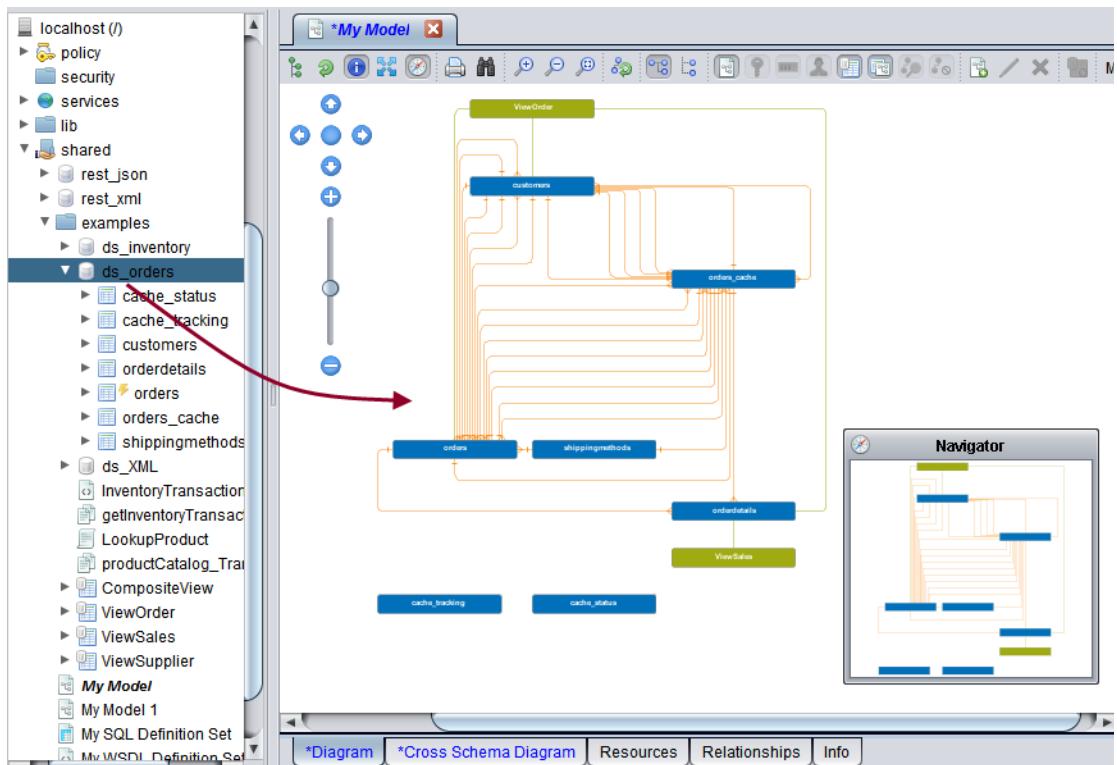

The model Diagram displays these objects:

— Tables in the selected resource (in blue). If you have indexed and discovered the resource, the model shows the table relationships.

— Tables in other indexed/discovered resources that are related to the selected resource.

— Views that are dependent on the displayed resources (in green).

2. Use the controls on the Diagram tab to work with the model. See Working with the Model Diagram, page 84.

Adding and Removing Resources in a Model

You can add and remove resources, including entire data sources or specific tables from within the model. When you remove resources from a model, the resources are still available in TDV and can be re-added at any time. Be sure to use the procedure below to remove resources.

To add or remove resources in a model

1. Open the model.

2. From the model Diagram, click the Add/Remove Resources button.

    Discovery displays the Add/Remove Resources dialog.

3. Expand the tree as necessary to view the currently selected resources.
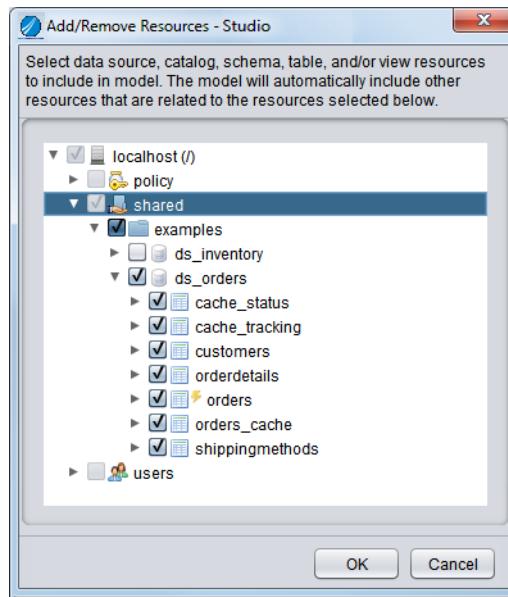
4. Select resources you want to include and deselect resources that you want to remove.

    Note: The selected resources must have been indexed and discovered as described in Running Indexing and Relationship Discovery, page 57, for you to see the relationships between them.

5. Click **OK**.

    The Discovery model immediately reflects your changes.

Renaming a Model

You can rename most objects in the Studio resource tree using the steps below.

To rename a model

1. Select the model in the Studio resource tree.

2. Right-click and choose Rename from the pop-up menu.

3. Type the new name for this model.

Refreshing a Model

If you have made changes to the model like reindexing data sources and you do not immediately see the changes in the model, you need to refresh the model.

To refresh a model

1. With the model open in Studio, click Refresh on the model toolbar.

Saving a Model

If you have made changes to the model, such as adding or removing data sources, or rearranging the model diagram, be sure to save them regularly.

To save a model

1. With the model open in Studio, click the Save in the Studio toolbar at the top of the window.

Note: The current button selections on the model toolbar are saved with the model. See Editing the Model Diagram, page 85 for more information.

Deleting a Model

To delete a model

1. If the model is open, close it.

2. In Studio, select the model.

3. Right-click and choose Delete.

   Discovery asks you to confirm that you want to delete the model.

4. Click Yes to delete the model.

Working with the Model Diagram

The model Diagram tab is where you graphically view and rearrange resource objects and relationships, and define known relationships.

This topic contain:

The model Diagram below shows a simple model and the controls you can use to explore a model.
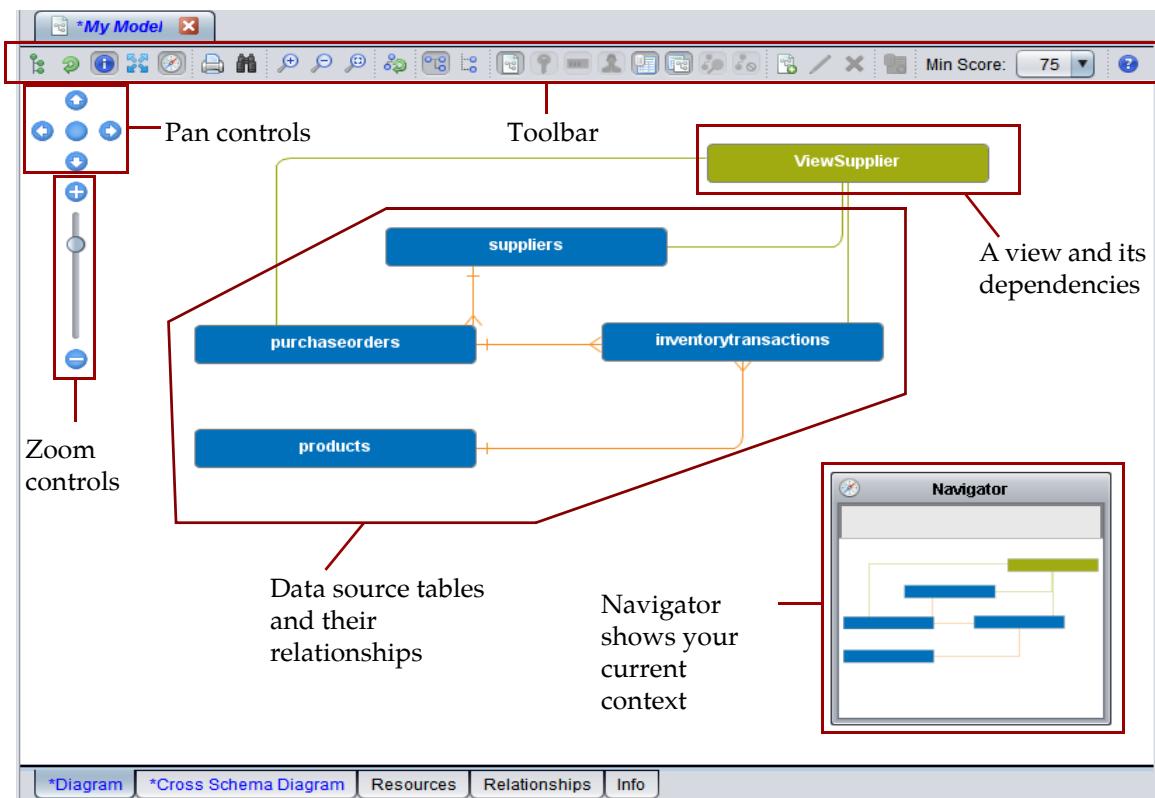
Editing the Model Diagram

Discovery lets you rearrange the model diagram and display details about the resources and their relationships. See these sections for more information:

Rearranging the Tables, Views, and Relationships

You can rearrange the tables, views, and relationship lines on the model diagram in many ways:

- Move any object—Click on the object and drag it to another location.

- Select multiple objects—Ctrl+click or drag a box around to select multiple tables/views and relationship/dependency lines.

- Move around in the diagram—Use the scroll bars or the track wheel on your mouse.

- Click the pan or zoom controls.

Discovery automatically arranges the relationship and dependency lines when you make changes.
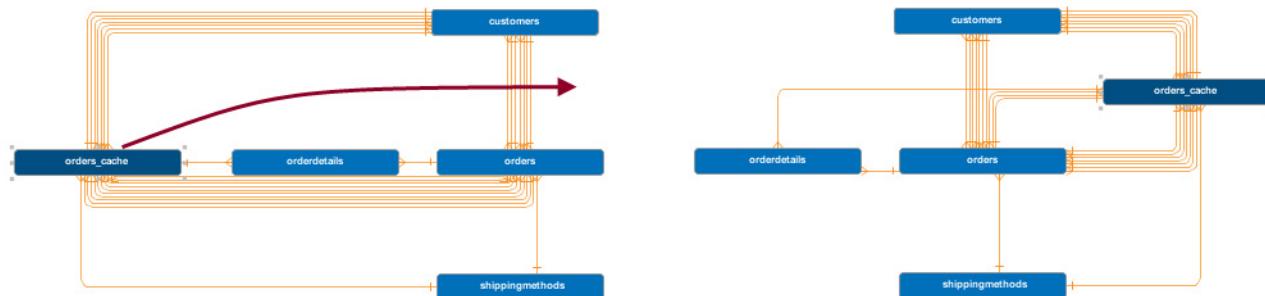
You can also select a corner or piece of a relationship line, and move the relationship line:

Click corner and

Result
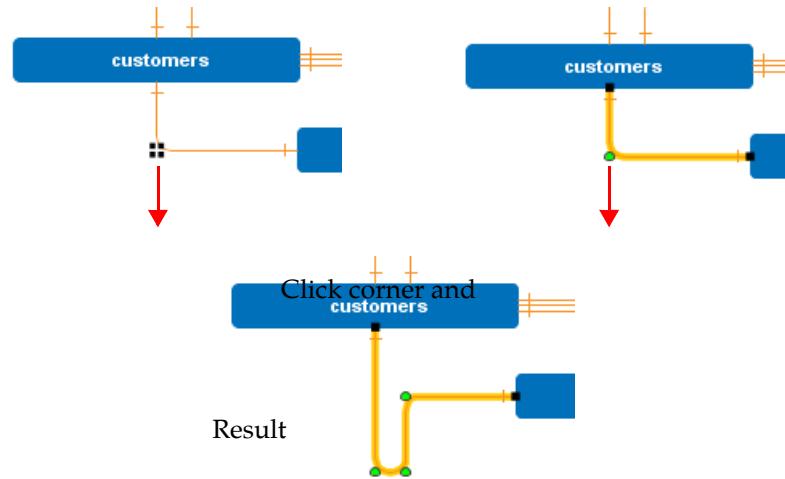
Note: If you save the model after rearranging tables and relationship lines, the table arrangement is saved, but the line changes you made are not saved with the model. Also note that there is no "undo" command. Once you rearrange the objects, you cannot automatically move them back to their previous arrangement unless you had saved it; in this case you can close the model without saving and reopen it. Be sure to save the model when you have an arrangement that you like.

Refreshing the Layout

If you have rearranged the resources in the model, you can have Discovery organize the model into its current layout type: orthogonal or hierarchical. See for more information.

To refresh the layout

1. Click the Refresh Layout button on the model toolbar.

Changing the Layout Type

Discovery provides two layout types for the model diagram: orthogonal and hierarchical.

- **Orthogonal**—Arranges the resources so that the relationship lines are as short as possible.
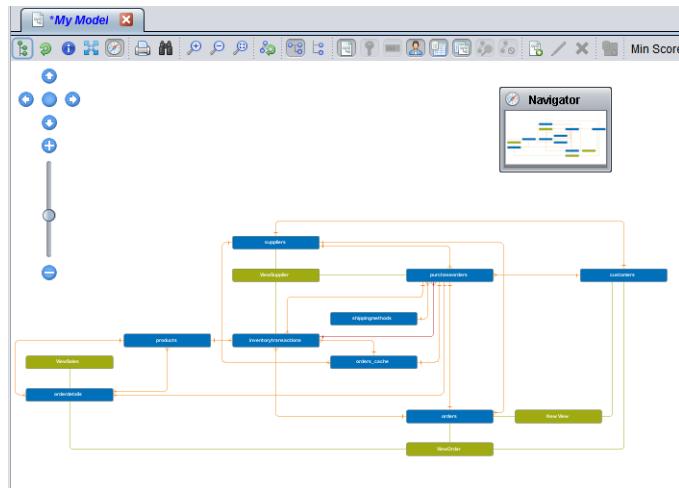
- **Hierarchical**—Arranges the resources in a hierarchy where the resources with the most dependencies are on the top level.
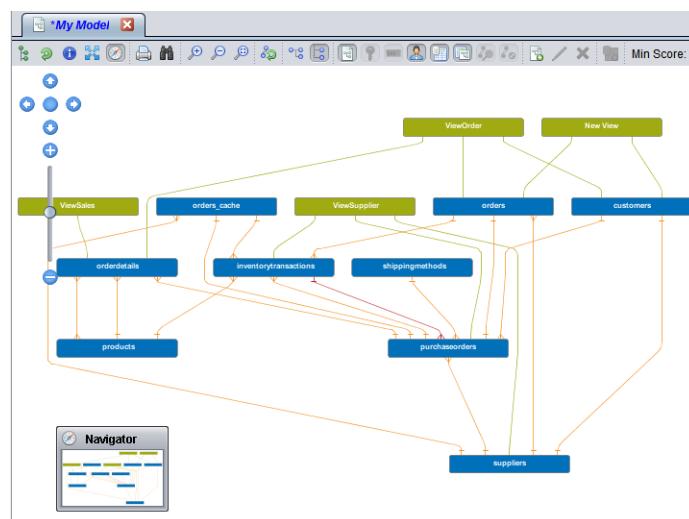
To change the model layout display type

1. Open a model.

2. On the model Diagram, click one of the layout buttons:

             Orthogonal   Hierarchical

The shaded layout button indicates the current layout selection.

Showing and Hiding Tables, Views, and Relationships

The show and hide buttons on the model toolbar control what is currently displayed. When you use one of these buttons to show or hide a table, view, or relationship type, Discovery remembers that state and it is saved as your preference for every model that you subsequently open.

Displaying Table Information

From the Model diagram, you can view a table columns as well as access details about the table.

To view information about a table

1. Expand the table to view its columns in one of these ways:

    — Double-click a table once to see its keys and double-click again to see all columns in the table.

    — Select a table and right click to show the menu of options.
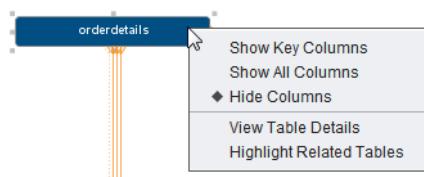
The diamond indicates the current selection.

Show Key Columns	Displays the primary key columns for the selected table, as in this example: An asterisk indicates that this is a primary key column. A number (#) indicates any kind of relationship, including discovered or foreign key. A plus (+) indicates that this is a denormalized column. An @ symbol indicates a data domain relationship.
Show All Columns	Displays all columns for the selected table as in this example: An asterisk indicates that this is a primary key column. A number (#) indicates any kind of relationship, including discovered or foreign key. A plus (+) indicates that this is a denormalized column. An @ symbol indicates a data domain relationship.
Hide Columns	Hides the columns you displayed.
View Table Details	Opens the table in Studio. See Table Statistics Displayed for Relationship Discovery, page 21.
Highlight Related Tables	Highlights and selects all related resources and the relationships between them. This enables you to move all related resources to another part of the diagram.

2. Optionally, click View Table Details to display all table statistics. See Table Statistics Displayed for Relationship Discovery, page 21, for more information about the relationship statistics added for Discovery.

Displaying View Information

Views are displayed in green and are connected by dashed lines to the resources on which they depend. From the Model diagram, you can display a view's columns as well as access details about the table.

To view information about a view

1. Double-click a view or select a view, right click, and choose one of these menu options.

Show All Columns or Hide Columns	Displays or hides the columns for the selected view. When displayed, the column data types are displayed as in this example:

ViewSales	
LeadTime	VARCHAR(32768)
ReorderLevel	DECIMAL(2147483647,0)
Discount	DECIMAL(12,2)
Status	VARCHAR(50)
ProductID	DECIMAL(2147483647,0)
CategoryID	DECIMAL(2147483647,0)

Highlight Related Tables	Highlights and selects all tables that are related to this view as well as their relationship lines. Because the relationships and tables are selected, you can move them as a group.

Note that you can hide the views by clicking the Hide Views button on the model toolbar.

Working with Related Resources

If a model is complex with lots of resources, it can be difficult to see what is related to what. In the Model diagram, you can easily identify all resources related to a selected resource and move the related resources as a group.

To highlight and select related resources

1. In the model diagram, select a table or view.

2. Right-click, and select Highlight Related Tables from the pop-up menu.

    Discovery highlights and selects all related resources and relationships:
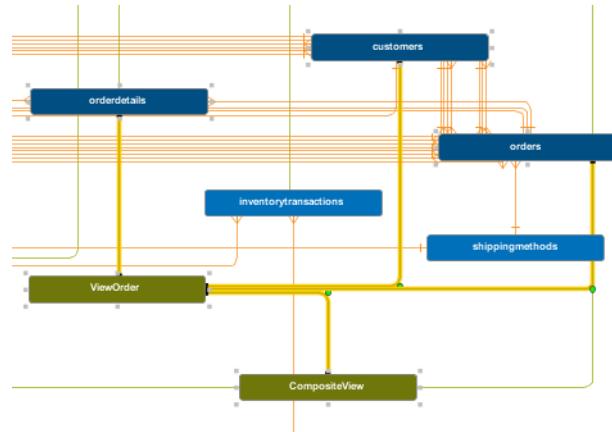
3. Move them as a group to another part of the diagram.

Displaying Relationship Information

A relationship between two nodes in the model is indicated by a connecting line. The ends of the connecting line indicate the cardinality of the relationship using crow's-foot notation.
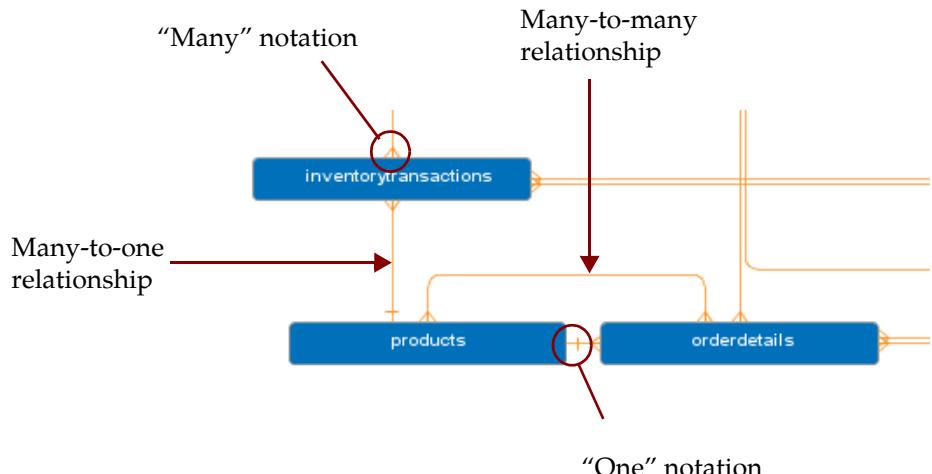
There are four types of relationships: discovered (orange), foreign key, (blue), multicolumn (brown), and user-defined (red). You can hide or redisplay the relationships by type using the buttons on the toolbar.
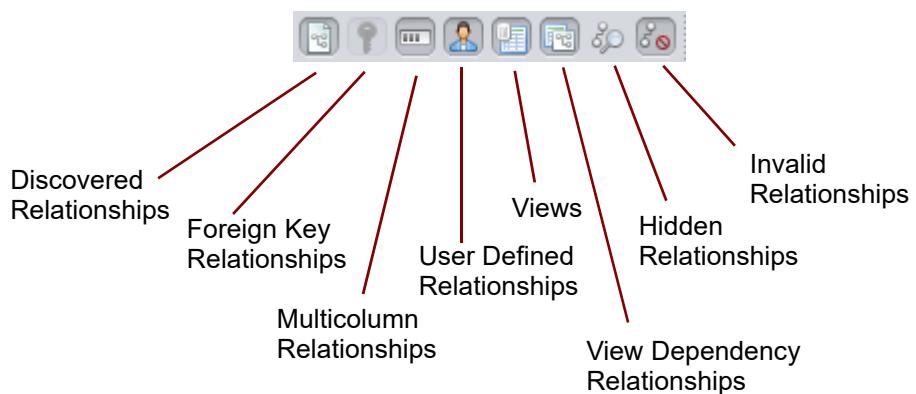
The buttons are available only if the selected model has relationships of that type. If you have marked relationships as hidden or invalid, you can still show them on the model diagram. Hidden relationships are shown with dashed lines, and invalid relationships are shown with dotted lines.

See Working with Relationships in the Model Diagram, page 96 and Working with the Relationships Tab, page 114 for further information.

Displaying View Dependency Information

Views that are dependent on the resources in a model are automatically displayed. The dependencies that a view has on other resources are indicated by green lines between the view and the resources. You can:

- Hide the views by clicking the Hide Views button on the model toolbar.

- Hide the view dependency relationships by clicking the Hide View Dependency Relationships button on the model toolbar.

Using the Navigator

The Navigator gives you a small view of the entire model and shows your current context within that model.

Toggle Navigator

Selected table

Current context; click and drag to change your

Click and drag handle to resize window
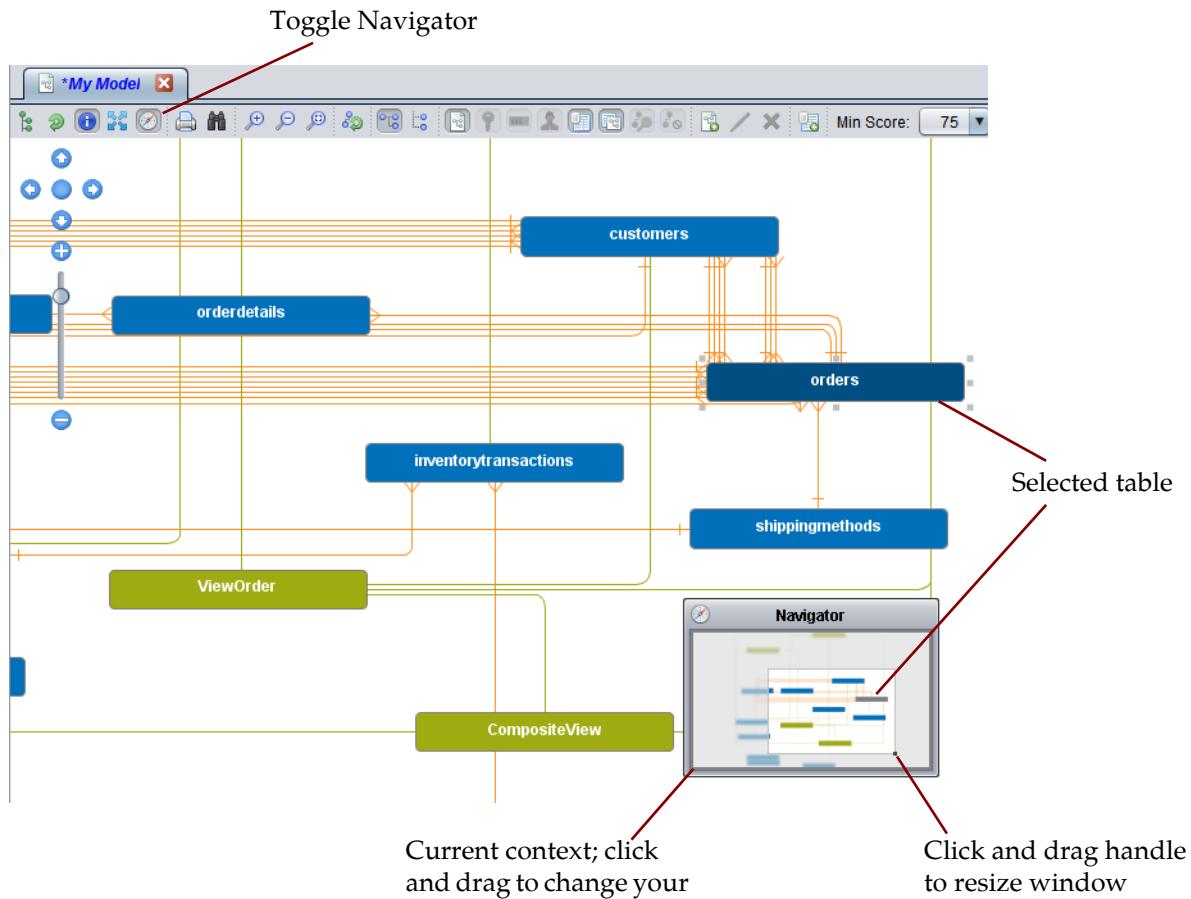
The Navigator:

- Lets you click the current context area and drag it to change your view.
- Displays selected tables in gray.
- Can be resized and moved around the model diagram.
- Can be removed from the model diagram using the Navigator button.
- Lets you select a table in its window to center it in the model diagram.

**Changing Your View Using the Pan Controls**

You can change your view up, down, right, or left using the graphical pan controls in the model diagram. You can also instantly fit the entire model into the current window.

Click to fit entire model
in current window
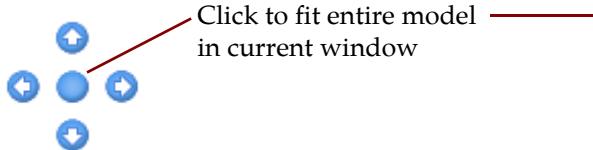
Panning the model in the model diagram can also be done using the keyboard.

Spacebar	Zoom to fit the entire model in the current window.
Arrow Keys	Move the model up, down, right, or left by a set increment.
> (Greater Than)	Zoom in for a closer view.
< (Less Than)	Zoom out for a larger view.
+ (Plus)	Zoom in for a closer view.
- (Minus)	Zoom out for a larger view.

Moving around in the Model Diagram

You can zoom in and out of the model diagram to change your view using buttons in the tool bar or the zoom controls in the model diagram. You can also move mouse track wheel to increase or decrease the zoom.

If the diagram is larger than the window area, scroll bars are made available.

Displaying the Model in a Full Screen

You can display a full screen view of the model using the Full Screen button. Click again to return to the normal view.

Finding Resources in the Model Diagram

Discovery helps you find a particular column or table in the model diagram using a simple search dialog.

To find a table in the model diagram

1. Open the model diagram.

2. Click the Find button or type Ctrl+F.

   Discovery displays a dialog for you to use to locate the columns and tables of interest.
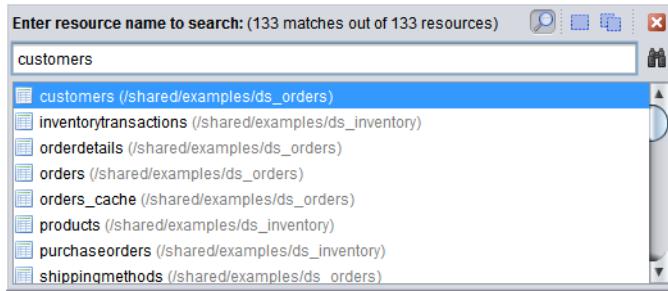
   When you initially open the search dialog, all tables in the model are listed in alphabetical order, followed by all columns in alphabetical order.

   These buttons are provided to help you control zooming and context in the model after a search:

   — Zoom to Found—Controls whether the current zoom amount in the model is remembered after a search. When disabled, the current zoom is retained. When enabled, the zoom returns to the default zoom amount.

   — Select Found Resource Exclusively—Controls whether the found table in the model is selected exclusively. When enabled, the newly found resource is selected and any other already selected nodes (if any) are unselected. This always results in one selected node with the find action.

   — Select Found Resource Cumulatively—Controls whether the found table in the model is added to the current selection. When enabled, the newly found is added to the current selection set. That is, the previously selected nodes are still selected. This is useful when you want to find multiple nodes and retain their selected state.

3. Type the name of the column or table for which you want to search.

   As you type, Discovery limits the columns and tables in the list to those with names that contain the typed characters. Discovery uses Studio's standard search convention. For example, use "*" as a wildcard.

4. When you locate the column or table of interest in the list, there are three ways you can view it in the model:

— Click the Find button.

— Type Enter.

— Double-click the table name in the list. Use this method if you want to keep the Find dialog open.

Discovery locates the table/table that contains the column and moves it to the center of the model display.

Working with Relationships in the Model Diagram

A model shows tables connected with lines that indicate relationships between the tables. If the data source has been indexed and discovered, all discovered relationships that have three or more matches and an RPS greater than 75 are displayed by default. See About the Relationship Probability Score (RPS), page 23 for more information about how relationships are discovered and scored.
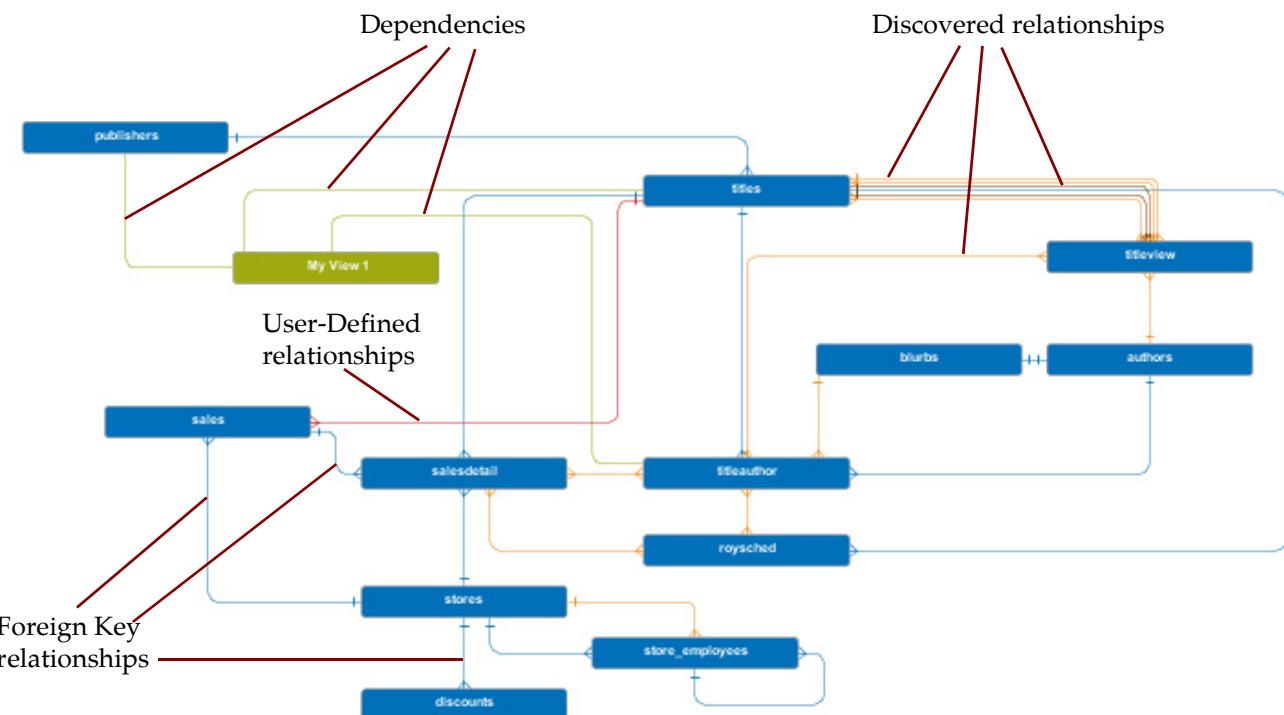
These are the types of relationships which are distinguished by line color and style.

| Relationship Type	Description	Line Color
Discovered	Relationships are discovered using TDV's proprietary algorithms.	Solid orange
User-Defined	User-defined relationships that are specified by the user. Users can define relationships between tables both within and across data sources.	Solid red
Primary Key-Foreign Key (PK-FK)	Defined in the database by the DBA.	Solid blue
View Dependency	The view is dependent on the connected resource.	Dashed green
Multicolumn	Relationships between multiple columns that form a compound key in one table, and matching columns in another table.	Solid brown
Invalid	Relationships that you have marked as invalid.	Dotted orange

Defining Relationships

You can define relationships between resources that you know exist. Discovery calls this type of relationship User-Defined.

Relationship definitions are independent of models. When you define a relationship between two resources, this relationship is displayed in any model that has the same two resources.

To define a new relationship

1. Open a model.

2. If necessary, add the resources for which you want to define relationships.

3. Click the New Relationship button.

Discovery displays the New Relationship dialog.
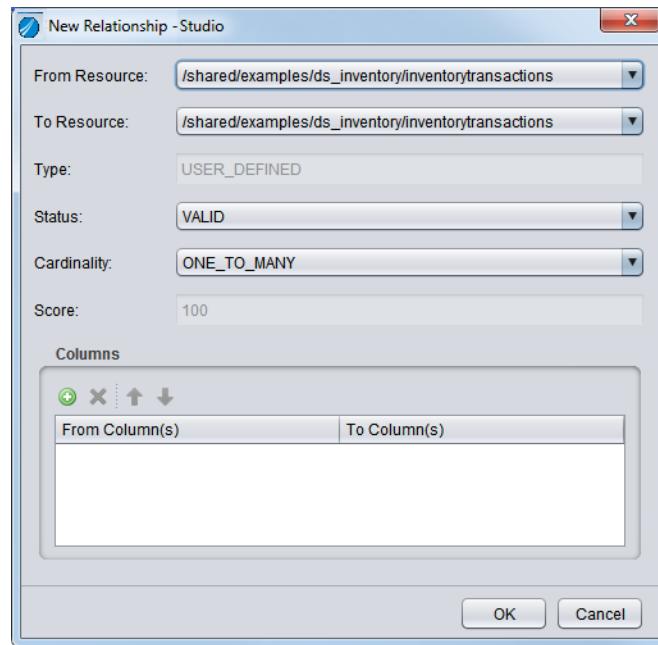
4. Specify the details about the relationship as follows:

Field	Description
From Resource	Choose the from-resource from the drop-down list.
To Resource	Choose the to-resource from the drop-down list.
Type	Always USER_DEFINED when defining a relationship.
Cardinality	Choose the type from the drop-down list:
	UNKNOWN
	ONE_TO_ONE
	ONE_TO_MANY
	MANY_TO_ONE
	MANY_TO_MANY
Score	Always 100 for USER_DEFINED relationships.

5. Under Columns:

   a. Click the Add Column Pair button.

   b. Click in the highlighted field under From Column(s) and choose the from-column.

   c. Click in the highlighted field under To Column(s) and choose the to-column.

   d. Continue adding column pairs.

   e. Optionally, use the arrow keys to change the order of the relationships. This has no effect on the model—only on how relationships are displayed in this list.

6. Click **OK** to save the relationship.

   Discovery displays the new user-defined relationship using a red connecting line. You can see the details about it in any one of these ways:

   — Edit Relationship dialog—Select the relationship, right click, and choose Edit Relationship. See Editing Relationships, page 104.

   — View Relationship Details—Select the relationship, right click, and choose View Relationship Detail. Discovery displays three tabs underneath the model Diagram. See Viewing Relationship Details, page 99.

   — Relationships tab—Click on the Relationships tab. The new relationship is listed in the relationships table in a tabular format. See Working with the Relationships Tab, page 114.

Viewing Relationship Details

You can view details about a selected relationship in three tabs beneath the model diagram. Both the Relationships tab and the Relationship Details tabs display indicators to help you identify columns and relationships of different types. See Relationship and Column Type Indicators, page 102 for examples.

To view relationship details

1. Display the relationship details tabs in one of these ways:

   — Click the Show Relationship Detail button.

   — Select a relationship line, right-click, and choose View Relationship Detail from the menu.

   — Double-click a relationship line.

   — Double-click a blank space anywhere in the model diagram.

Discovery displays the Relationship Details panel beneath the model. If you cannot see it, locate the horizontal bar and drag the bar upwards to expand the panel vertically as illustrated below:
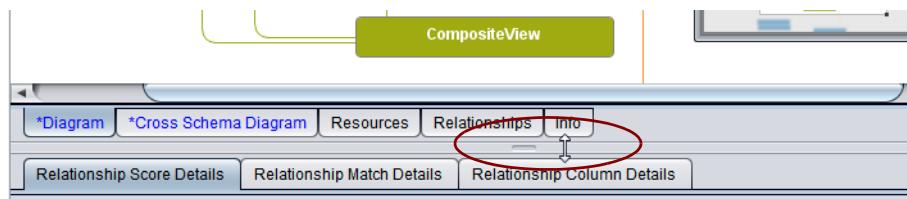
2. Click the Relationship Score Details tab.

The Relationship Score Details tab displays the following information for the currently selected relationship:

— From Source: The path and name of the from source. A yellow asterisk (*) indicates that this is a primary key column. A yellow plus-sign (+) indicates that this is a denormalized column.

— To Source: The path and name of the to source. A yellow asterisk (*) indicates that this is a primary key column. A yellow plus-sign (+) indicates that this is a denormalized column.

— Type: The type of relationship—can be DISCOVERED, FOREIGN_KEY, USER-DEFINED, MULTICOLUMN, or UNKNOWN.

— Cardinality: The cardinality of the relationship, which can be ONE_TO_ONE, ONE_TO_MANY, MANY_TO_ONE, MANY_TO_MANY.

— Status: The status of the relationship, which can be UNKNOWN, NOT_REVIEWED, VALID, or INVALID.

— Score (Number of Matches): The score as a percentage and the actual number of matches.

— Score table: The score factors and their values for this relationship.

— Data Domains (Influencing Matches): Lists the data domains that influenced the discovered relationships.

3. Click the Relationship Match Details tab:

   The Relationship Match Details tab displays the actual data in the tables in the relationship. You can view items that match, items in the from-table that do not match, and items in the to-table that do not match.

   Matches—Displays the actual matched data in the tables in the relationship.

   Unmatches From—Displays the data rows with no matches in the from resource.

Unmatches To—Displays the data rows with no matches in the to resource.
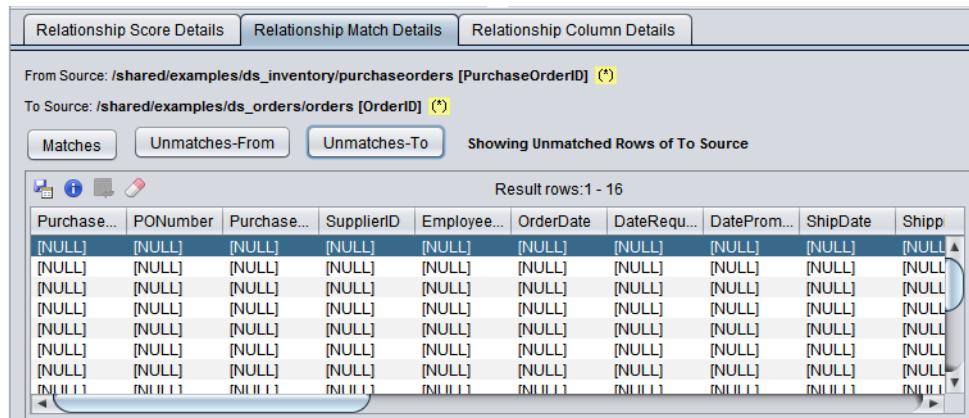
4. Click the Relationship Column Details tab.

The Relationship Column Details tab displays the columns in the From Source and To Source tables including the column name, column data type, and the relationship.

Relationship and Column Type Indicators

The relationship icon in the Relationship column indicates the columns that have the relationship.

If the selected relationship includes multiple columns, there are multiple relationship icons in the From Source and To Source columns with numbers next to them. The numbers indicate which columns match and together form

the multicolumn relationship. The example below shows a multicolumn relationship formed by 14 matching columns

If the relationship includes a denormalized column, it is denoted with (+) mark on the target or source column of the relationship. Note that only one end of source/target pair can be denormalized, not both. See Working with Denormalized Data, page 56 for more information.

The denormalized indicator is displayed both on the Relationships tab and on the Relationship Column Details tab. On the Relationship Column Details tab, the delimiter used in the denormalized column is listed in the Relationship column. The tab and space characters (delimiters that are not visible) are indicated as \t and \s, respectively.

Denormalized column indicator
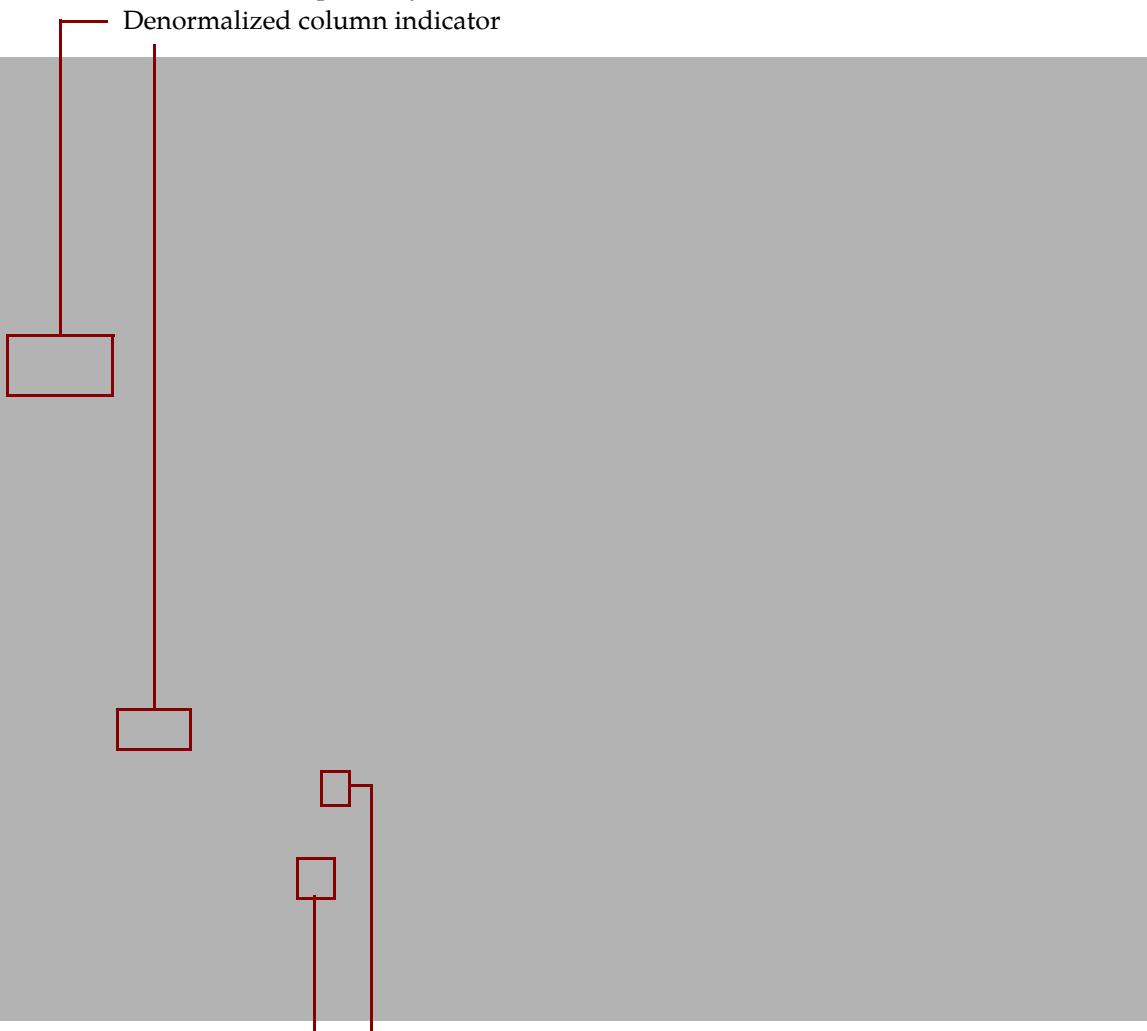
The delimiter in brackets was used in the denormalized column

Editing Relationships

You can edit a discovered or user-defined relationship in the model diagram. What you can change depends on the relationship type:

- Discovered—You can change only the relationship status.

- User-Defined—You can change anything except the relationship type or score.

- Foreign Key—You cannot change anything, because a foreign key relationship is defined by the data source. However, you can view details about a foreign key relationship.

Relationships are independent of models. When you edit a relationship, the relationship is updated on the server and on any model in which it appears.

To edit a relationship

1. Open a model.

2. Select a relationship, and either click the Edit Relationship button or double-click a relationship line.

   Discovery displays the relationship in the Edit Relationship dialog box.
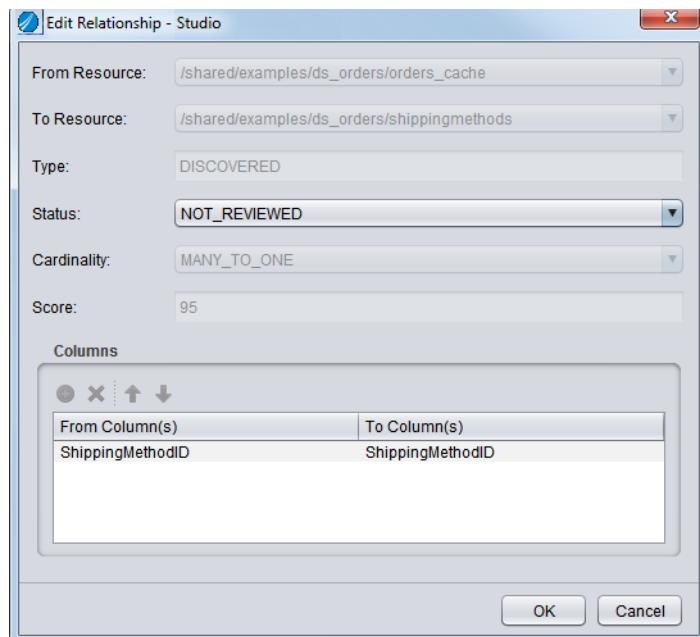
3. Edit the relationship Status based on its type.

Relationship Type	Things You Can Edit
DISCOVERED	The Status defaults to NOT_REVIEWED. You can change this to: VALID—Confirms that the relationship is valid and makes the RPS 100. INVALID—Removes the relationship from the model. UNKNOWN—Marks this relationship as unknown.
USER_DEFINED	Anything except Type and Score. See Defining Relationships, page 97, for information.
FOREIGN_KEY	Nothing; you can only view its definition.

4. Click **OK**.

Validating and Invalidating Discovered Relationships

After relationships have been discovered, you can view them in the model diagram and either validate them (which gives them an RPS of 100), or invalidate them.

To validate or invalidate a discovered relationship

1. Open a model.

2. Select a relationship and click the Edit Relationship button or double-click a relationship line.

   Discovery displays the Edit Relationship dialog.

3. For Status, select either VALID or INVALID.

   INVALID relationships are removed from the model diagram.

   Note: You can still display relationships that are marked as invalid. See Viewing Relationships Based on Type, page 107.

4. Click **OK**.

Deleting User Defined Relationships

You can delete only user-defined relationships. You cannot delete discovered or foreign key relationships.

To delete a user-defined relationship

1. In the model diagram, select the relationship you want to delete.

2. Click the Delete Relationship button.

Viewing Relationships Based on Type

You can display only certain types of relationships using the show/hide relationships buttons. See Displaying Relationship Information, page 91.

The buttons are only available (not dimmed) if the model has relationships of that type.

Viewing Relationships Based on RPS Score

You can display only relationships above a specific RPS score. (See About the Relationship Probability Score (RPS), page 23.) By default, Discovery creates and displays relationships with a score of 75 or higher. You can change the minimum score using the Min Score field at the top of the model diagram.

Printing the Model Diagram

Discovery lets you print the model and provides various ways to control the printed output.

To print a model

1. Arrange the tables in the model in the way you want them to be printed. Note that you can expand the tables and views to display the keys and columns.

2. Click the Print button.

   Discovery displays the Preview dialog.

3. Use the buttons at the top control the printed output as follows.

Use This	To Make This Happen
Page	Change the orientation between portrait and landscape and adjust margins.
Print	Print the model diagram using the standard print dialog.
Zoom In	Zoom in on the print image. This does not increase the size of the image on the page.

Use This	To Make This Happen
Zoom Out	Zoom out on the print image. This does not decrease the size of the image on the page.
Zoom Amount	Select a setting from the drop-down list.
Options	Displays a Print Options dialog with three tabs:

Options continued:

- General tab options let you create a poster-style print of the model diagram by dividing the model into sections that can be pieced together.

  Poster Rows—Divides the image vertically into the number of rows entered and displays how it would appear on that number of pages if printed.

  Poster Columns—Divides the image horizontally into the number of columns entered and displays how it would appear on that number of pages if printed.

  Add Poster Coordinates—Prints the location of each subsection of the model diagram so that you can easily figure out how the sections go together.

  Clip Area—With the Graph option, enlarge the model image as much as possible within the margins. With the View option, increase the blank space around the model image.

- Title tab options let you specify a title for the printed model.

  Text—Specify the title you want to appear at the top of the model diagram.

  Titlebar Color—Click to select the color of the bar behind the title text.

  Text Color—Click to select the color of the title text.

- Footer tab options let you specify a footer for the printed model. If you are printing the model poster-style, the footer will span the pages appropriately.

  Text—Specify the footer text for the bottom of the model diagram.

  Footer Color—Click to select the color of the bar behind the footer text.

  Text Color—Click to select the color of the footer text.

Working with the Cross Schema Diagram

The Cross Schema Diagram tab displays every schema currently included in the model. You can:

- View the resources (schemas and views) and the tables in each schema.

- View all relationships, both within schemas and between schemas.

- Create relationships and edit existing relationships.

To work with the Cross Schema Diagram

1. Open a model.

2. Select the Cross Schema Diagram tab at the bottom of the window.

   Discovery displays a diagram that shows all schemas that are currently in the model. In this example, the two schemas are shown as blue boxes.
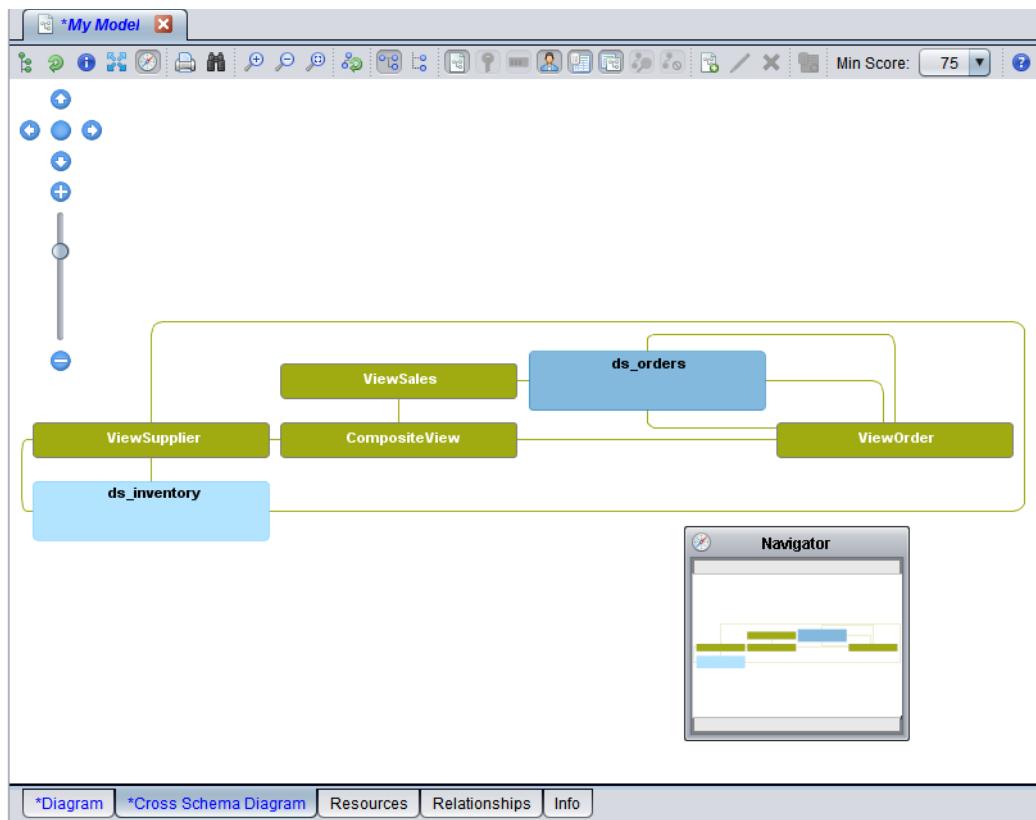
3. Optionally: select a schema, right-click to display the pop-up menu, and choose one of the options.

ds_inventory

Show Tables With Cross Schema Relationships
Show Tables With Relationships
Show All Tables
◆ Hide Tables

Note: The diamond indicates the current selection.

- Show Tables with Cross Schema Relationships—Displays only relationships between the selected schema and other schemas in the model.
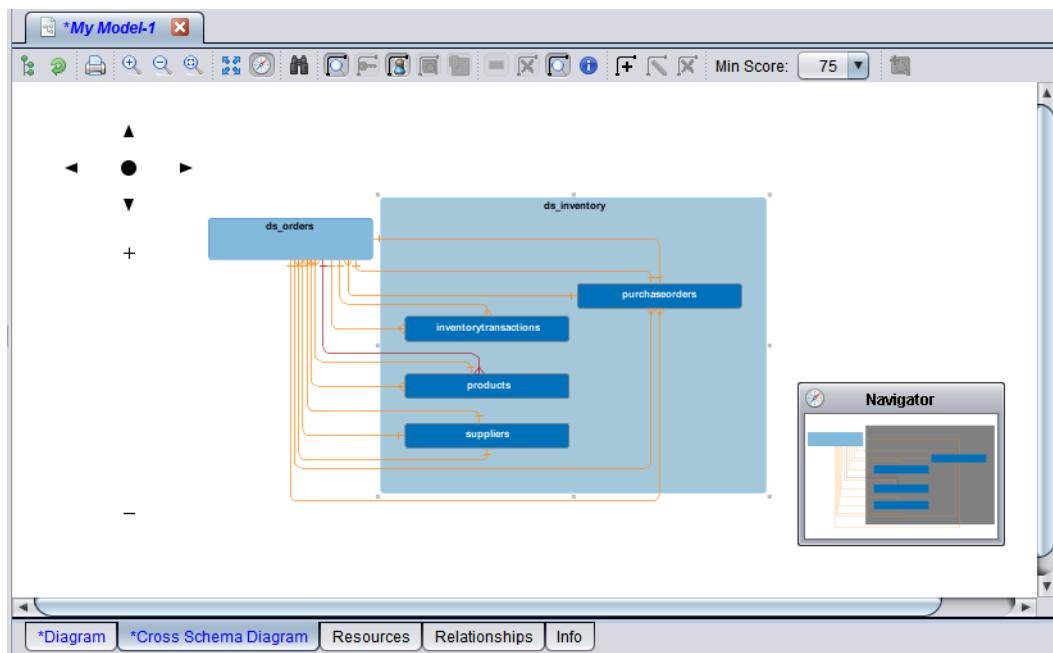
- Show Tables with Relationships—Displays all relationships both within the selected schema and between it and other schemas in the model.
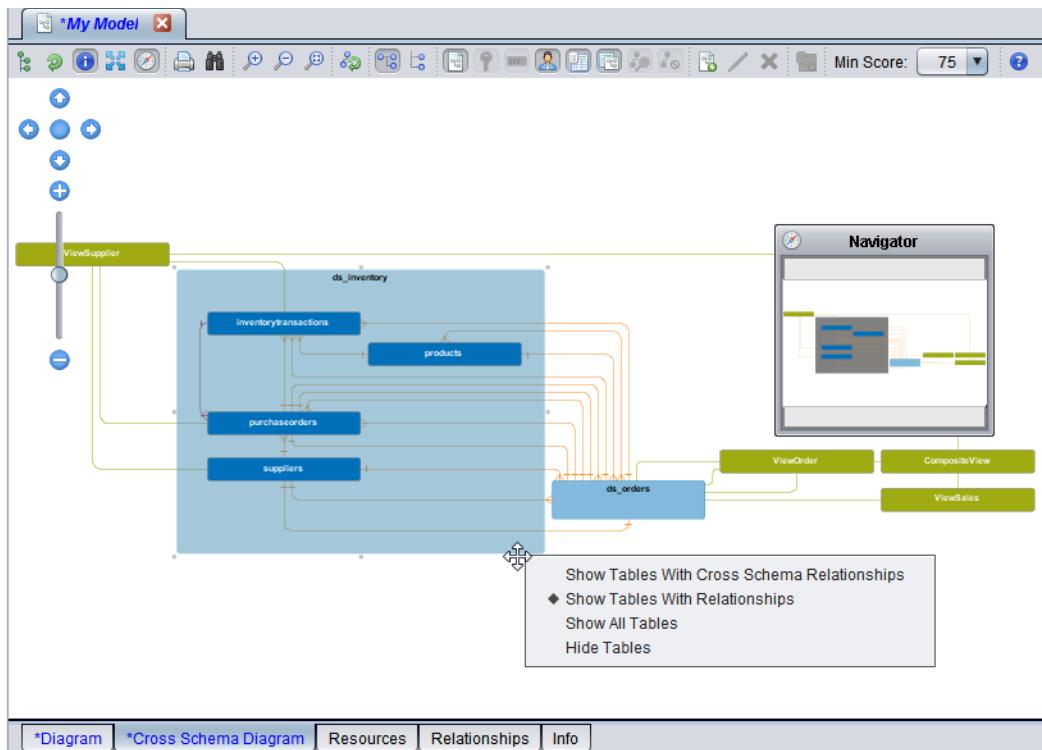
Double-clicking any table shows its key columns. Double-clicking again displays all columns.
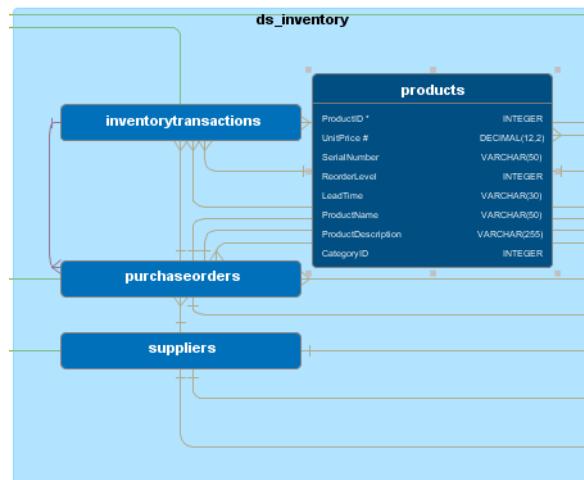
You can also do most of the same things in the Cross Schema Diagram that you can do in the model Diagram tab including:

— Rearrange tables and relationships in the schema (see Rearranging the Tables, Views, and Relationships, page 85).

— Create or edit a relationship (see Defining Relationships, page 97 or Editing Relationships, page 104).

— Get relationship details (see Viewing Relationship Details, page 99).

— Print the current display (see Printing the Model Diagram, page 107).

Working with the Model Resources Tab

The Resources tab displays all resources that are currently included in the model including data sources, catalogs, schemas, tables, and views. The Resources tab lists the resources in tabular format and lets you:

• View the path for each resource.

• Discover how the resource was included—by explicitly being added or because of its relationship with an explicitly added resource.

• Locate a resource by column sorting.

To work with the model Resources tab

1. Open a model.

2. Select the Resources tab at the bottom of the window.

   Discovery displays the Resources table.

The columns in the Resources table provide this information.

Column	Description
Name	Name of the resource.
Path	Path to the resource.
Type	resource type: DATA_SOURCE, CATALOG, SCHEMA, Table, or View.

Column	Description
Inclusion	One of the following:

- EXPLICIT—You explicitly selected the resource in the Add/Remove Resources dialog, or dragged the resource into the model.

- IMPLICIT—You selected a parent or dragged a parent of the resource into the model. For example, if you add an entire data source to a model, all of its resources (tables, etc.) are implicitly added and are listed as IMPLICIT in this column.

- RELATED—The resource is included because it is related to an EXPLICIT or IMPLICIT resource.

- DEPENDENT—Indicates a view resource that is dependent on a related resource.

Working with the Relationships Tab

The model Relationship tab gives you details about the relationships of the resources in the model. The Relationships tab lists the relationships in tabular format and lets you:

- View details about the data source tables and their relationships.

- View details about column-to-column relationships.

- Learn about a relationship type, cardinality, status, and score.

- Filter the relationships by type or minimum score.

- Click on the column headers to change the sort column and sort order.

- Ignore columns so that they do not appear in a model. See Ignoring Certain Columns, page 117.

- Save relationship information to a Microsoft Excel file. See Saving Relationships to a File, page 118.

To work with the model Relationships tab

1. Open a model.

2. Select the Relationships tab at the bottom of the window.

Discovery displays the Relationships table

Click to view a relationship's details   Relationship type   Minimum Score filter

The columns in the Relationships table contain this information.

Column	Description
From Path	The full path of the "from" resource.
From Resource	The name of the "from" resource.
From Column(s)	The names of the "from" columns in the resource.
From Type	The data type for this column (VARCHAR, INTEGER, DECIMAL, DATE, etc.).
To Path	The full path of the "to" resource.
To Resource	The name of the "to" resource.
To Column(s)	The names of the "to" columns in the resource.
To Type	The data type for this column (VARCHAR, INTEGER, DECIMAL, DATE, etc.).
Type	The relationship type, which can be: DISCOVERED USER_DEFINED FOREIGN_KEY DEPENDENCY UNKNOWN—Should not appear under normal circumstances.

Column	Description
Cardinality	The relationship cardinality, which can be:
	ONE_TO_ONE
	ONE_TO_MANY
	MANY_TO_ONE
	MANY_TO_MANY
	UNKNOWN
Status	USER_DEFINED and FOREIGN_KEY relationships always have a status of VALID.
	DISCOVERED relationships can have one of these values:
	• INVALID—The user marked this relationship as invalid. It will disappear from the model diagram after marking invalid.
	• VALID—The user marked this relationship as valid. Its score is set to 100.
	• NOT_REVIEWED—The relationship has not been reviewed and marked as VALID or INVALID.
	• UNKNOWN—Should not appear in normal circumstances.
	See Editing Relationships, page 104 for information about marking relationships as VALID or INVALID.
Score	The Relationship Probability Score (RPS) can be 0-100%, to indicate the likelihood that the relationship is valid. The score depends in part on the relationship type:
	DISCOVERED—The relationship probability score for this relationship. See About the Relationship Probability Score (RPS), page 23.
	USER_DEFINED—Always 100, because this relationship is known to exist.
	FOREIGN_KEY—Always 100, because this relationship is defined in the data source.
	DEPENDENCY—Always 100, because this view depends on the resource.
Hidden [check box]	Displays a check mark if this relationship is currently hidden.
Multicolumn	Displays an icon indicating that this relationship includes multiple columns. See About Multicolumn Relationships, page 20, for more information.

3. Optionally, use one of the filter buttons to determine which relationships are displayed.

4. Optionally, display details about this relationship by clicking the Show Relationship Details button. See Viewing Relationship Details, page 99, for more information.

5. Optionally: select a relationship, right-click, and choose one of these menu options.

Choose This Option	To Do This
Create View	Open a dialog that lets you specify the location in the Studio resource tree in which to save the new view.
Mark Valid	Mark this relationship as VALID.
Mark Invalid	Mark this relationship as INVALID. When displayed, relationships you have marked as invalid are shown connected by a dotted orange line in the model diagram.
Mark Hidden	Mark this relationship as hidden. The relationship continues to be listed in the Relationship table, but it is not displayed in the model diagram unless you click the Show Hidden Relationships button. Relationships marked as hidden are shown connected by a dashed line.
Edit Relationship	Open the relationship in the Edit Relationship dialog. See Editing Relationships, page 104.

Ignoring Certain Columns

You can ignore some columns. When you do this, relationships to them do not appear in the model.

To ignore certain columns

1. Run a full indexing scan on the data source resources you want in the model, including the tables with columns that you want to ignore.

2. Create a model.

3. Drag in tables and data sources that you have indexed.

4. Select the Relationships tab for the model.

5. Click From Column(s) in the table header to sort by the from-column names.

6. Find the columns you want to ignore. (They should all be adjacent.) Select them all, and then right-click and select Mark Invalid in the context menu.

7. Repeat the preceding two steps for the To Column(s) header.

8. Go back to the Diagram or Cross Schema Diagram and hide the invalid relationships by selecting the Hide Invalid Relationships toolbar button.

These steps eliminate all relationships involving the ignored column in the model display.

Invalidated columns are invalid across all subsequent models you create. However, if a subsequent model exposes these columns, you need to invalidate the new ones if you do not want to include them.

Saving Relationships to a File

You can save the relationship information displayed on the Relationships tab to a comma-separated value (.csv) file.

To save the model relationship information to a .csv file

1. Open a Discovery model.

2. Click the Relationships tab.

3. Click the Save To File button on the model toolbar.

   Discovery opens a dialog for you to specify where to save the file.

4. Click **OK** to save the file.

   Discovery saves a .csv file.

|	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Column N	From Path	From Reso	From Colu	From Type	To Path	To Resour	To Colum	To Type	Type	Cardinalit	Status	Score	Hidden	Multicolumn
2	1	/shared/e	orders		ShipName	VARCHAR	/shared/e	customers	Company	VARCHAR	DISCOVER	ONE_TO_(	NOT_REVI	89.40% N	N
3	2	/shared/e	orders		ShipCity	VARCHAR	/shared/e	customers	City		VARCHAR	DISCOVER	MANY_TO	NOT_REVI	84.00% N	N
4	3	/shared/e	orders_ca	ShipCity	VARCHAR	/shared/e	customers	City		VARCHAR	DISCOVER	MANY_TO	NOT_REVI	84.00% N	N
5	4	/shared/e	orders_ca	ShipName	VARCHAR	/shared/e	customers	Company	VARCHAR	DISCOVER	ONE_TO_(	NOT_REVI	89.40% N	N
6	5	/shared/e	orders_ca	ShipCity	VARCHAR	/shared/e	orders		ShipCity	VARCHAR	DISCOVER	MANY_TO	NOT_REVI	85.00% N	N

About the Model Info Tab

The model Info tab displays the ownership and lock information of the current model. This is the same information as Studio provides for other resource types. You can add notes about the model to track its purpose and status.

To display the model information

1. Open a model.

2. Select the Info tab at the bottom of the window.

    Discovery displays the model information.
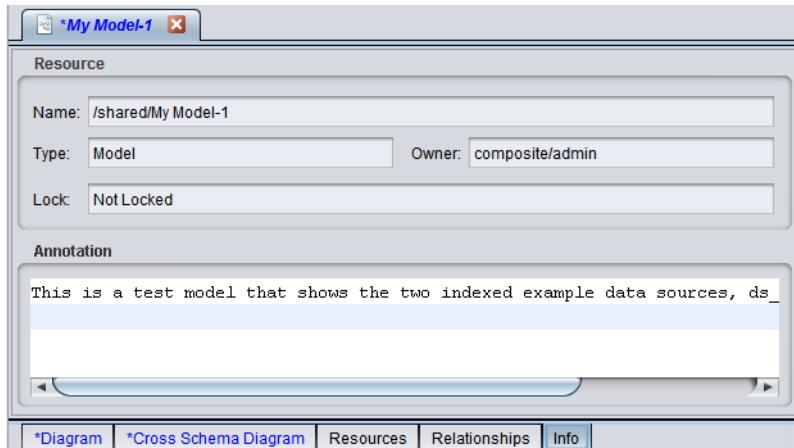
    The Info tab provides this information:

Column	Description
Name	Name and path of the model.
Type	Resource type, which is always Model.
Owner	User who created this model.
Owner Domain	Domain in which this model was created.
Lock	Indicates if the resource is locked.
Annotation	An area for notes about the model.

See Resource Management Basics in the *TDV User Guide* for more information about resource locks.

Creating Views with Discovery

This chapter describes how to create and work with SQL views in Discovery.

After you have explored the resources and their relationships in a Discovery model, you can use the model diagram to select the tables you want to include in a composite view. You can then use Studio's view editing and publishing features to work with the new view.

The following topics are covered in this chapter:

- Creating a New Composite View from a Discovery Model, page 121
- Working with Discovered Relationships in Views, page 123

Creating a New Composite View from a Discovery Model

Discovery makes it easy for you to create a new composite view in TDV, based on relationships in a discovery model. All tables to be included in the new view must have at least one relationship to another table in the view. You can create a view based on a single relationship or on multiple relationships.

To create a new view based on a single relationship in a model

1. Open a model.

2. Select a relationship in the model.

3. Right-click and choose Create View Using Relationship from the pop-up menu.

Discovery opens a dialog for you to name the view and then displays the view in Studio.

To create a new view based on multiple relationships in a model

1. Open the model in Studio.

2. In the model Diagram, rearrange and edit the model as described in Working with Models in Discovery, page 77, and locate the resources that you want to include in the view.

3. Select the resources to include in the view, keeping these things in mind:

   — All tables should have at least one relationship to another table, unless you are creating a view from a single table that is unrelated to anything else.

   — Do not select any views; views cannot be included within a new SQL View.

   — To select multiple resources:

   — Ctrl+click the resources.

   — Hold down the left mouse button and draw a box around the resources.

   Discovery displays the highlighted resources and makes the New View button available.
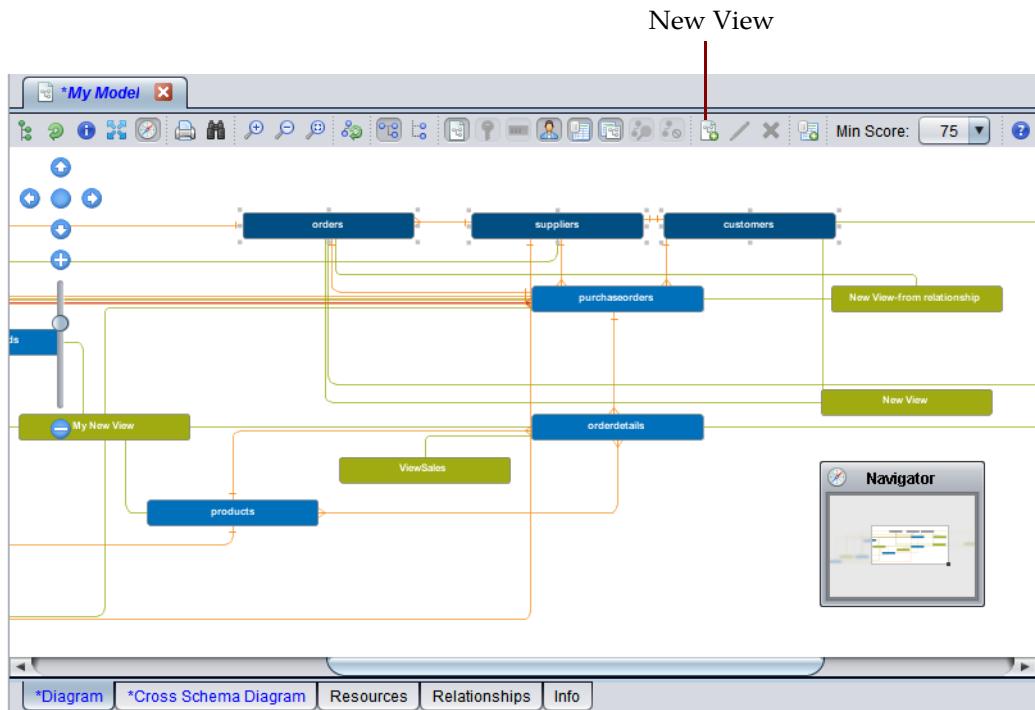
4. Click the New View button.

   Discovery displays a dialog box for you to choose the location and name the view.

5. Choose the location and enter a view name.

   Discovery creates the new view, opens it in the workspace. and adds it to the Studio resource tree.

Note: Creating a view can take awhile if the view and its SQL are complex.

Studio displays the new view in the Studio workspace on its own tab.
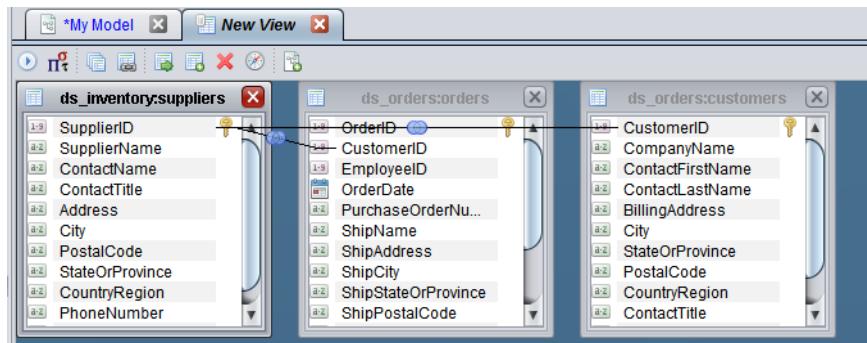
All resources and their relationships are defined. The SQL tabshows the SQL that Discovery created to define this view.

6. You can now use Studio's extensive view editing capabilities to refine and publish the view.

See the *TDV User Guide* for information on views and publishing.

Working with Discovered Relationships in Views

Composite views are integrated with Discovery's features. When you work with views in Studio, you can take advantage of the knowledge gained by Discovery when data sources are indexed and discovered. Discovered relationships between tables can be listed along with the scores so that you can evaluate if you want to include them in the view.

### Adding Discovered Relationships to a View

If Discovery has been used to index and discover relationships among data resources, Studio gives you the opportunity to include those relationships in the view when you add those resources to it.

Note: If you add a relationship using the Add Relationships dialog, you can only remove it via the view Model display. To do this, hover over the relationship, right-click, and choose Remove from the pop-up menu.

To add discovered relationships to a view

1. Create or open a view.

2. Drag a table into the view from an indexed and discovered data source.

   If the table has relationships to tables that are already in the view, Studio displays the Add Relationships dialog.
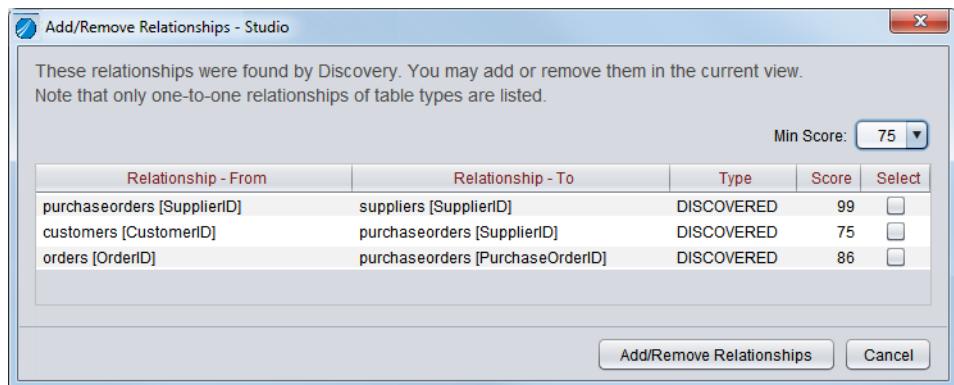
Note: The Add Relationships dialog is displayed by default only if the Use Discovery Relationships in View Editor configuration parameter is set to True.

3. Optionally, limit the list to relationships above a certain RPS score by selecting a different value for Min Score.

4. Select relationships to include in the view and click **Add Relationships**.

## Showing Discovered Relationships in a View

If any of your views contain data sources that have been indexed and discovered using Discovery, you can display a list of one-to-one relationships between tables in the view. (Some relationships might already be displayed in the view.)

To display a list of all one-to-one discovered relationships between tables in

a view

1. Open an existing view in Studio.

   Discovery displays the tables in the view.
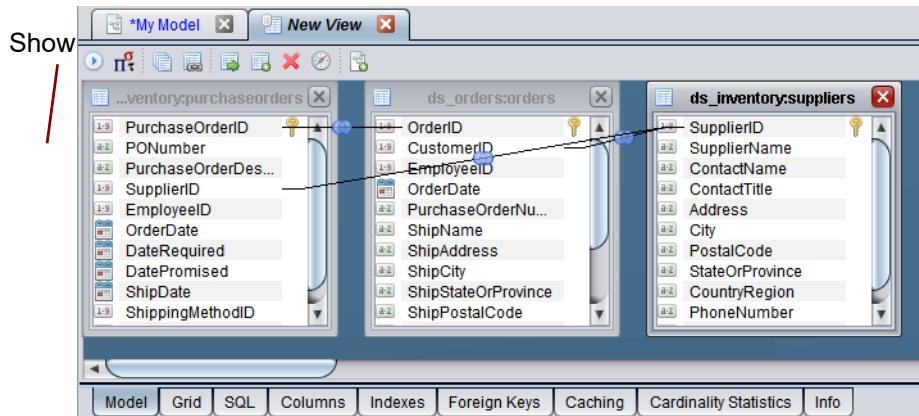
Show

2. Click Show Discovery Relationships to display relationships in this view.

   Studio displays a list of one-to-one relationships that have been discovered between the tables in this view.

Relationships with check marks are already displayed and included in this view. These cannot be deselected.

3. Optionally, select more relationships to include in this view and click Add Relationships.

Configuring Relationship Display in Views

By default, when you add a table to a view, Studio looks for discovered relationships between tables in the view, and gives you the opportunity to include them in the view. You can configure Studio to show or ignore these discovered relationships.

To configure which discovered relationships Studio is to show in views

1. Choose Administration > Configuration to open the Configuration dialog.

2. Expand the Discovery > Relationship branch.

3. Select Use Discovery Relationships in View Editor and choose:

   — True (default)—Automatically display discovered relationships when a table is added to a view.

   — False—Do not display discovered relationships when a table is added to a view.