CA ERwin® Data Model Validator

Implementation Guide

r8

CA Technologies Product References

This document references the following CA Technologies products:

- CA ERwin® Data Model Validator
- CA ERwin® Data Modeler

Contact CA Technologies

Understanding your Support

Review support maintenance programs and offerings.

Registering for Support

Access the CA Support online registration site to register for product support.

Accessing Technical Support

For your convenience, CA Technologies provides easy access to "One Stop" support for all editions of CA ERwin Data Modeler, and includes the following:

- Online and telephone contact information for technical assistance and customer services
- Information about user communities and forums
- Product and documentation downloads
- CA Support policies and guidelines
- Other helpful resources appropriate for your product

For information about other Home Office, Small Business, and Enterprise CA Technologies products, visit http://ca.com/support.

Provide Feedback

If you have comments or questions about CA Technologies product documentation, you can send a message to techpubs@ca.com.

If you would like to provide feedback about CA Technologies product documentation, complete our short customer survey, which is also available on the CA Support website, found at http://ca.com/docs.

CA ERwin Data Modeler News and Events

Visit www.erwin.com to get up-to-date news, announcements, and events. View video demos and read up on customer success stories and articles by industry experts.

Contents

Chapter 1: Introduction

This section contains the following topics:

Database Validation Overview

Welcome to CA ERwin Data Model Validator (CA ERwin DMV), the database validation tool that raises the level of data quality in transactional and data warehouse systems. This software quickly identifies design inconsistencies by systematically scanning CA ERwin Data Modeler (CA ERwin DM) models.

CA ERwin Data Model Validator helps you address a number of critical IT management issues. These include the need to rapidly design and deploy database systems using a limited number of skilled personnel, integrate and maintain database systems with ease over their lifetime, and maintain the highest levels of data integrity across database applications.

The Need For High-Quality Customer Data

eBusiness and customer-relationship management (CRM) have given a new urgency to compiling clean, consolidated customer information. With the massive influx of data resulting from online transactions, organizations are faced with an entirely new source of data quality problems. The need for data quality is critical as organizations increasingly make use of this data to their advantage.

Improve Customer Relationship Management

Organizations cannot operate efficiently without accurate data about its consumers and business partners. Organizations that fail to manage data quality cannot ensure its data is accurate enough for use in its data warehousing, CRM, and eBusiness applications. eBusiness makes data quality management a necessity as new data warehouses and CRM applications are being built to analyze and predict customer behavior or track customer activity. If an organization can't accurately create a consolidated view of the customer, it cannot create predictive models for customers who visit that website.

Useful, Consolidated Data

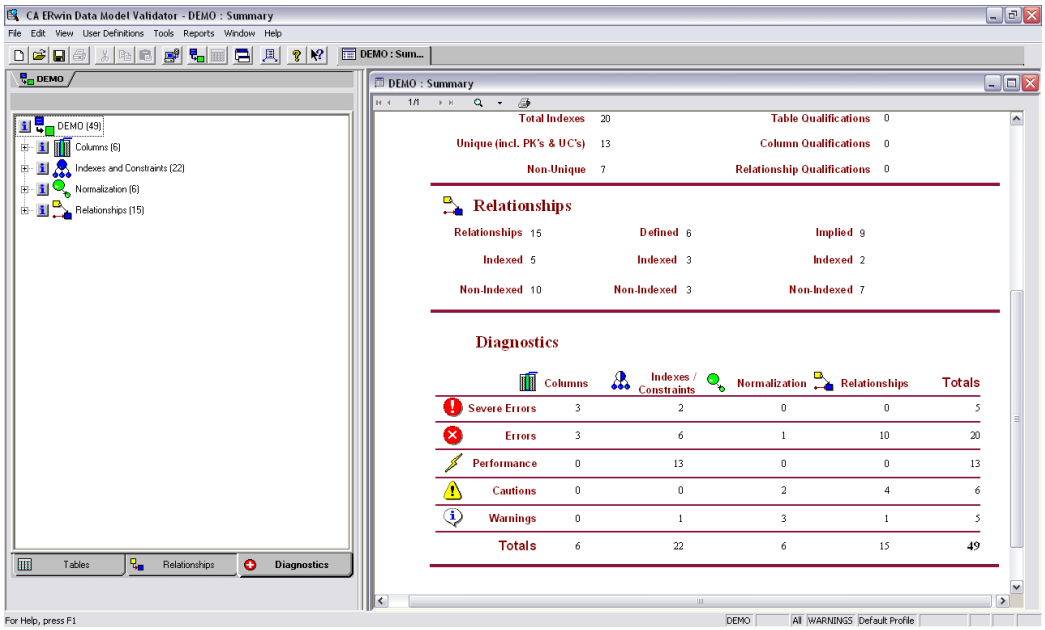
Today's eBusiness architecture is a patchwork of technologies ranging from mainframe applications all the way down to thin-client systems running on the latest handheld devices. Transforming these peninsulas of information into new business opportunities and improved service levels are all critically important to competitive advantage in this new eBusiness environment. The superior capabilities of CA ERwin DMV ensure that data is consistently defined and captured across all your deployment platforms.

Data Quality Made Simple

CA ERwin DMV analyzes databases to identify inconsistencies that adversely affect database integrity and efficiency. It reports on database elements, such as key and domain inconsistencies and normalization analyses. CA ERwin DMV assists you in optimizing data structures and suggests commands for improving the database structure.

CA ERwin DMV provides your organization with an expert, reliable, and consistent validation platform that ensures high levels of data quality throughout your data infrastructure.

The following shows a sample Diagnostics Summary:



Data Quality Benefits Your eBusiness

Managing data quality does the following:

- Improve eBusiness efficiency by handling the complex and time-consuming task of validating data designs across heterogeneous platforms
- Improve CRM by ensuring that customer information is captured in a standardized fashion regardless of your implementation environment
- Make an impact on both data warehouses and new CRM applications by facilitating the aggregation and cleansing of transactional information

Distinctive Features

CA ERwin DMV contains the following features:

Diagnostics

The diagnostics are based on the rules of the relational model, including normalization issues relative to your database environment. Diagnostic results are conveniently organized into the following categories:

- Columns
- Indexes
- Constraints
- Normalization
- Relationships

Suggested Changes and Auto-Correct

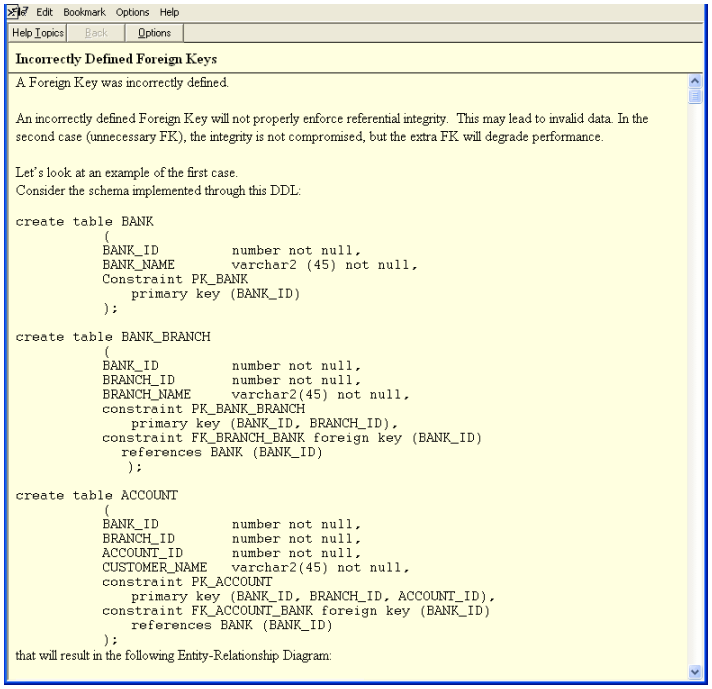
Changes to your database design are recommended based on the results of the diagnostics, providing an efficient and consistent approach to improving database design. In some cases, corrective scripts are automatically generated.

Facilitating Reviews

Several standard reports facilitate database design reviews within a project group. The *Show Me* facility isolates specific design issues within the context of large complicated models. This unique facility enhances productivity by eliminating the task of sifting through complicated database designs. By displaying only relevant database structures, this window provides modelers with a visual context for understanding and addressing the issues at hand. Modelers can also define a subset of the model and work with this subset.

Teach Me Facility

The *Teach Me* facility can help novice database designers better understand the impact of their designs in light of Relational Theory and their particular database environment. This facility explains the problem and instructs the modeler on the implications of their design decisions. Use *Teach Me* to obtain more information about any specific diagnostic provided by CA ERwin DMV, as shown in the following diagram:



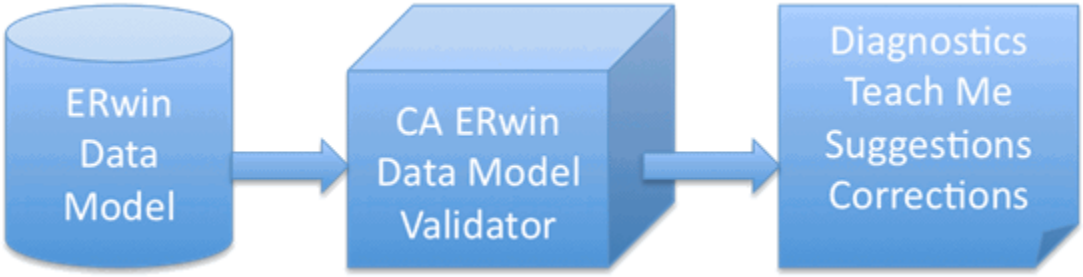
CA ERwin Data Model Validator

CA ERwin Data Model Validator provides a foundation for building, deploying, and managing applications. All of the CA ERwin Data Modeler editions consist of process and project management, change and configuration management, modeling and design, model validation, and knowledge publication and visualization. It strengthens your ability to automate critical application life cycle processes and to thrive in today's increasingly complex and rapidly changing eBusiness climate. These editions help you simplify the complex aspects of analyzing, designing, and implementing applications and business processes by providing a visualization of the relationships between business and technology.

Analyzing Overview

CA ERwin Data Model Validator is a knowledge-based system that analyzes databases to identify inconsistencies that adversely affect database integrity and efficiency. It reports on database elements, such as key and domain inconsistencies and de-normalization issues. It assists the user with optimizing data structures and suggests commands for improving the database structure.

The following diagram illustrates how the process works:



Chapter 2: Installation

This section contains the following topics:

System Requirements Verification

Before you install the software, consult the current CA ERwin Data Model Validator *Release Notes* file for current software and hardware requirements. You must verify that your system requirements meets at least the minimum specified requirements.

Install CA ERwin Data Model Validator

Install CA ERwin Data Model Validator on each client for which you want to use the validation tool.

To install CA ERwin Data Model Validator

1. Insert the installation DVD in the DVD drive.

   If Autorun is enabled, you will be prompted to install. If not, double-click Setup.exe from the DVD drive.

   The Welcome dialog opens.

2. Click Next to continue.

   A License Agreement opens for you to review.

3. Select the option to accept the terms as described in the License Agreement and click Next.

   Follow the instructions provided by the installation wizard. The wizard asks a series of questions you must answer, including:

   - Customer Information
   - Destination folder for the directory to which CA ERwin Data Model Validator will be installed

4. Select the Complete or Custom Setup Type.

   Complete

   Installs all program features.

   Custom

   Installs selected features. A new window opens with options to select specific features for installation. You can also evaluate the space needed to install each feature, and change the destination folder for the files.

   Click Next to continue and then click Install to start the installation.

   The program files are copied.

5. When the install completes, a final screen displays with the following two options:

   - Launch CA ERwin Data Model Validator
   - View the Release Notes

   Both check boxes are selected by default. Click Finish to exit the wizard.

   The installation is complete.

Note: Demonstration models are installed under the directory you specified on the Setup screen.

Chapter 3: Analyze Your Database and Improve Data Quality

This section contains the following topics:

Validate the Structure of Your Databases

Today more than ever, time means money. Endless amounts of customer information are pouring into data warehouses from many sources. It is critical that the information contained in these databases maintains its integrity and ability to be deployed across many platforms. That is why the process of developing a data model that maintains data integrity is so important. Using CA ERwin Data Model Validator as your diagnostic technician to validate the structure of all major DBMS databases, you can:

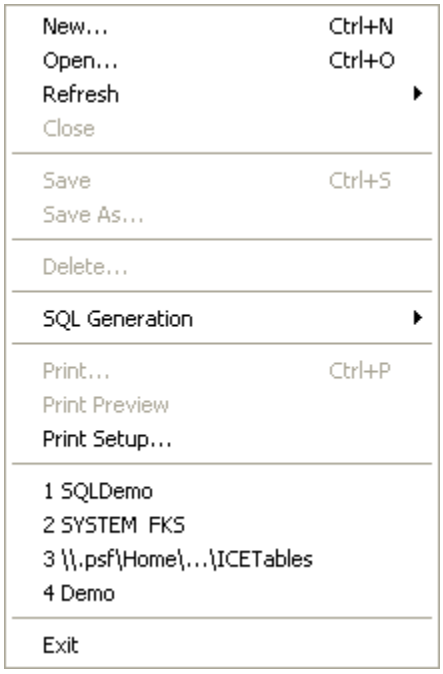
- Produce recommendations to improve database design
- Validate the database by training the user to use relational theory
- Present a graphic and useable tabular representation of the database
- Verify the integrity of database applications already in production
- Incorporate changes into existing database applications
- Review third party applications before purchasing

Examine the Consistency and Integrity of Any Database (New or Production)

To examine the consistency and integrity of any database (new database, production database), you need to use CA ERwin Data Modeler to reverse-engineer the database and save it as a CA ERwin Data Modeler model (.erwin file).

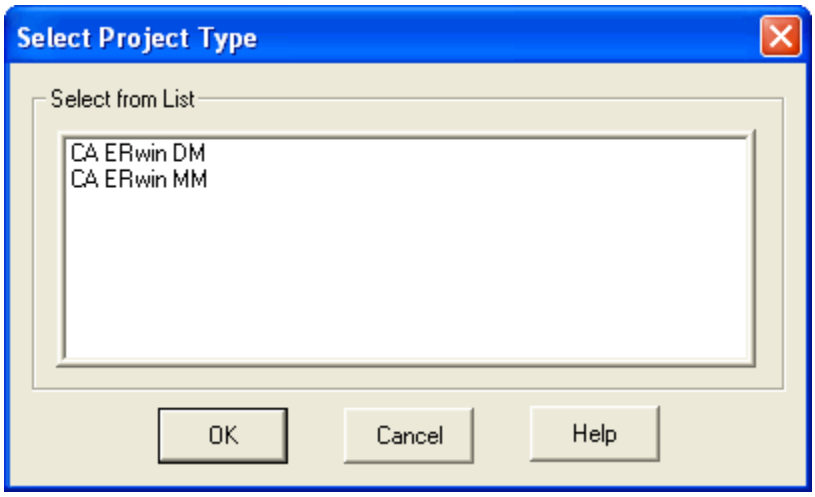
To examine the consistency and integrity of any database

1. Click File, New on the main menu.



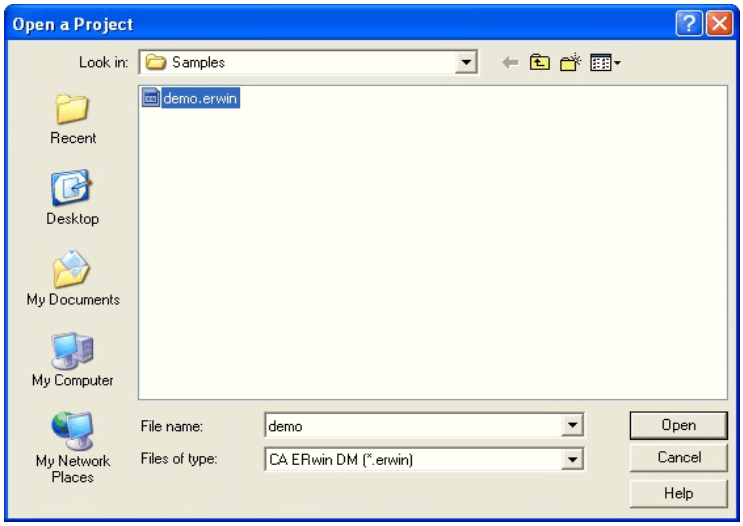
The Select Project Type dialog opens.

2. Select CA ERwin DM and click OK.



The Open a Project dialog opens.

3. Select the CA ERwin DM model that you want to open and click Open.



You are connected to CA ERwin Data Modeler and the Select Tables dialog opens.

4. Select the tables you want to open and click OK.

The model opens.

Analyze the Model

CA ERwin Data Model Validator provides diagnostic messages and suggestions to optimize your database operating efficiency.

The following categories can be diagnosed:

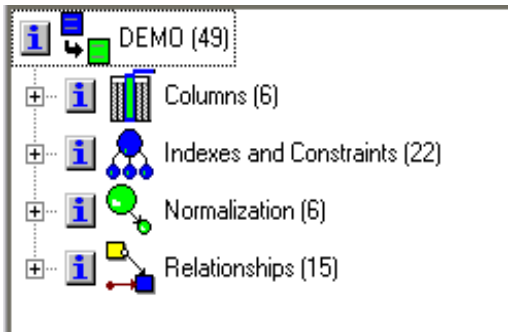
- Columns
- Indexes and Constraints
- Normalization
- Relationships

A corresponding severity level is assigned:

- Severe Errors
- Errors
- Performance
- Cautions
- Warnings

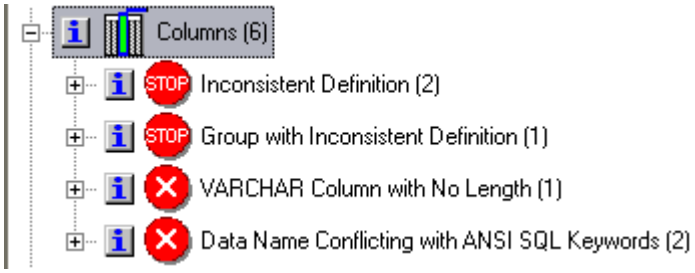
You can begin to analyze a model after you open it. In the example that follows, we use the model DEMO.erwin.

1. Create a new validation project and click on the Diagnostics tab. The following screen appears:



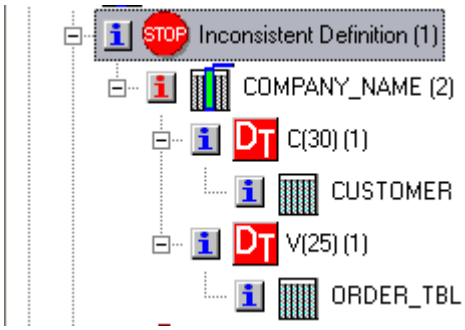
The numbers represent the amount of messages that were found for each category.

2. To obtain more information about the Columns diagnostics, expand the Columns folder. The following information displays:



The numbers represent the amount of messages that were found for each category.

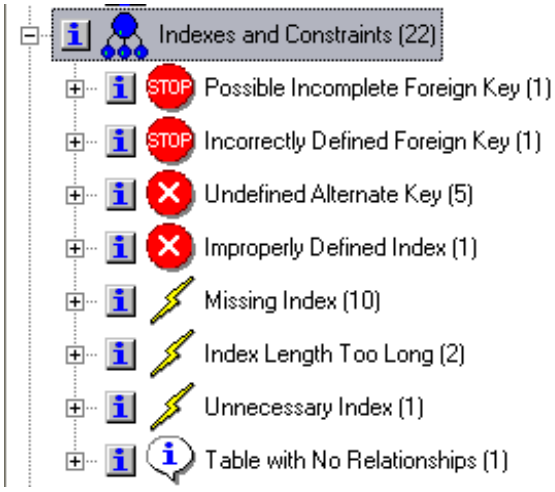
3. Expand Inconsistent Definition.



You can see that COMPANY_NAME has 2 inconsistent definitions. In table ORDER_TBL it is defined as Varchar(25). In table CUSTOMER it is defined as Char(30).

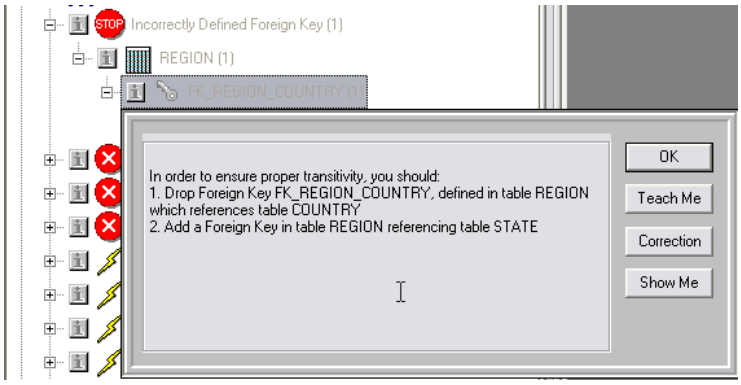
CA ERwin DMV shows that the attribute COMPANY_NAME exists in two different tables with different data types and that this must be avoided.

4. Expand Index and Constraint. You will see the following:

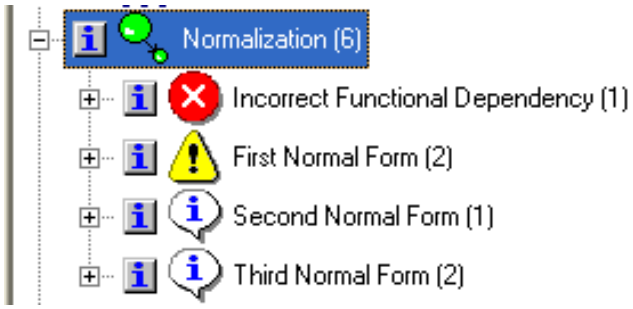


5. Expand Incorrectly Defined Foreign Key and click the information icon.

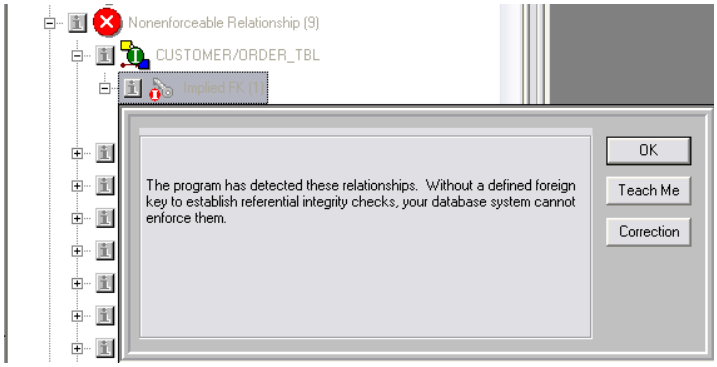
   You can see that the Foreign Key FK_REGION_COUNTRY is not defined correctly. CA ERwin DMV suggests dropping this FK and adding another FK in table REGION referencing table STATE.



6. Expand Normalization. The following appears:

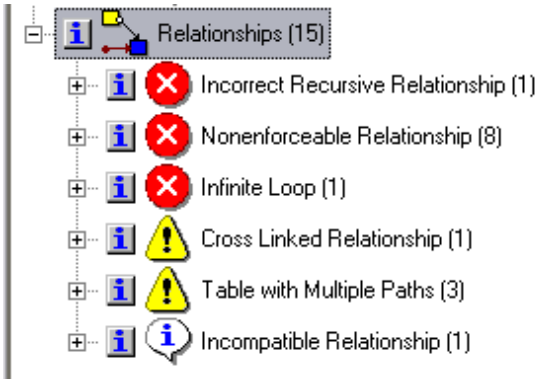


7. Expand Second Normal Form. Click the information icon for table ORDER_ITEM.

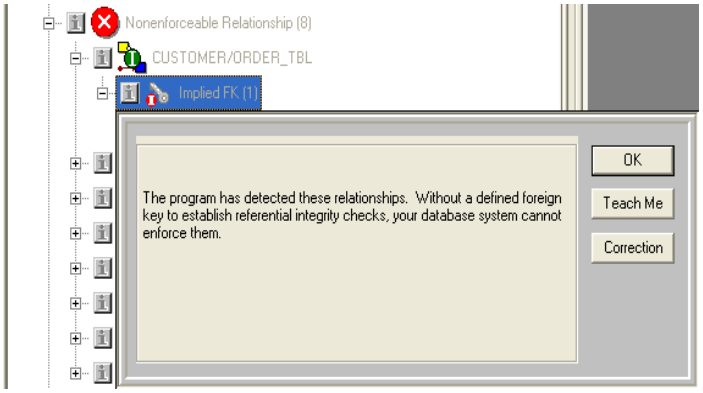


   The Column UNIT_PRICE occurs as a second normal form deviation in table ORDER_ITEM. CA ERwin DMV warns that redundancy must be controlled.

8. Expand Relationship:



9. Expand Nonenforceable Relationships and click the information icon  for the Implied FK in the CUSTOMER/ORDER_TBL relationship.



You are warned that there are eight instances of implied relationships that will not be enforced.

10. If you click the Correction button, the scripts creating the Foreign Key are generated. This will enforce the referential integrity between the referenced tables. The scripts are not applied to the database. A text file with the name of the model and .sql extension is generated (in this case DEMO.sql).

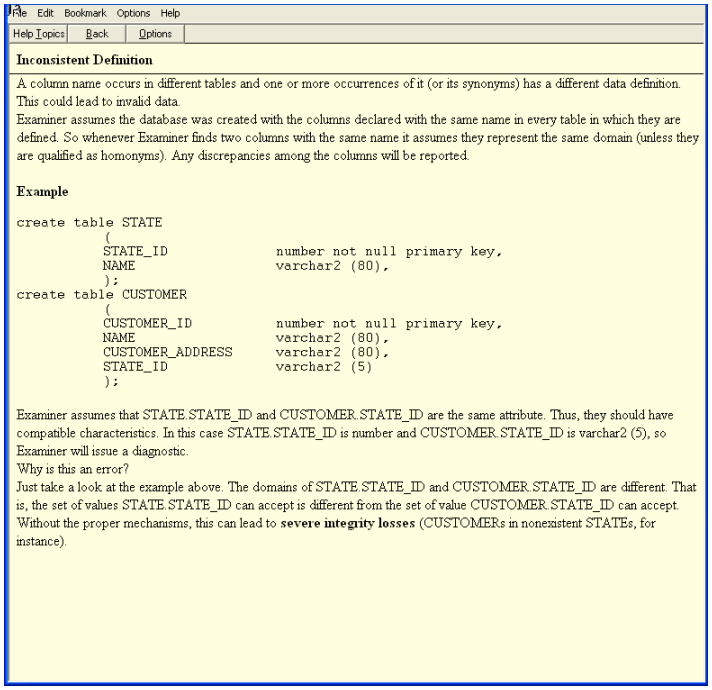
Understanding the Diagnostic Messages and Suggestions

To help you understand the theory behind the messages, we'll take a behind-the-scenes look at how these analytical decisions were reached.

To understand the diagnostic messages and suggestions

1. Click the blue information icon  that precedes a diagnostic message to display the Teach Me button.

2. In this example, we're clicking on the Inconsistent Definition diagnostic that appeared in the Column Diagnostics:



File Edit Bookmark Options Help

Help Topics | Back | Options

Inconsistent Definition

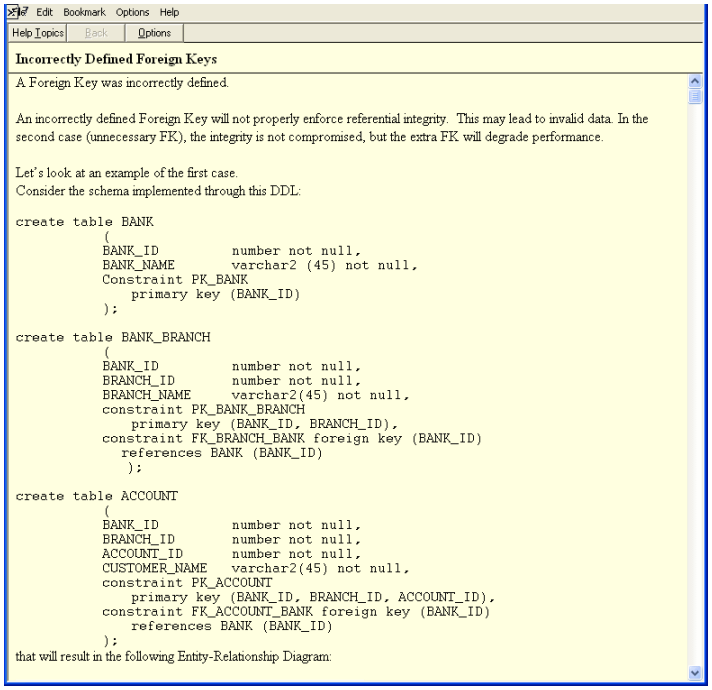
A column name occurs in different tables and one or more occurrences of it (or its synonyms) has a different data definition. This could lead to invalid data.
Examiner assumes the database was created with the columns declared with the same name in every table in which they are defined. So whenever Examiner finds two columns with the same name it assumes they represent the same domain (unless they are qualified as homonyms). Any discrepancies among the columns will be reported.

Example

```
create table STATE
          (
          STATE_ID             number not null primary key,
          NAME                 varchar2 (80),
          );
create table CUSTOMER
          (
          CUSTOMER_ID          number not null primary key,
          NAME                 varchar2 (80),
          CUSTOMER_ADDRESS     varchar2 (80),
          STATE_ID             varchar2 (5)
          );
```

Examiner assumes that STATE.STATE_ID and CUSTOMER.STATE_ID are the same attribute. Thus, they should have compatible characteristics. In this case STATE.STATE_ID is number and CUSTOMER.STATE_ID is varchar2 (5), so Examiner will issue a diagnostic.
Why is this an error?
Just take a look at the example above. The domains of STATE.STATE_ID and CUSTOMER.STATE_ID are different. That is, the set of values STATE.STATE_ID can accept is different from the set of value CUSTOMER.STATE_ID can accept. Without the proper mechanisms, this can lead to **severe integrity losses** (CUSTOMERs in nonexistent STATEs, for instance).

3. Click Teach Me under the Indexes and Constraints diagnostic for Incorrectly Defined Foreign Keys.



Edit Bookmark Options Help

Help Topics | Back | Options

Incorrectly Defined Foreign Keys

A Foreign Key was incorrectly defined.

An incorrectly defined Foreign Key will not properly enforce referential integrity. This may lead to invalid data. In the second case (unnecessary FK), the integrity is not compromised, but the extra FK will degrade performance.

Let's look at an example of the first case.
Consider the schema implemented through this DDL:

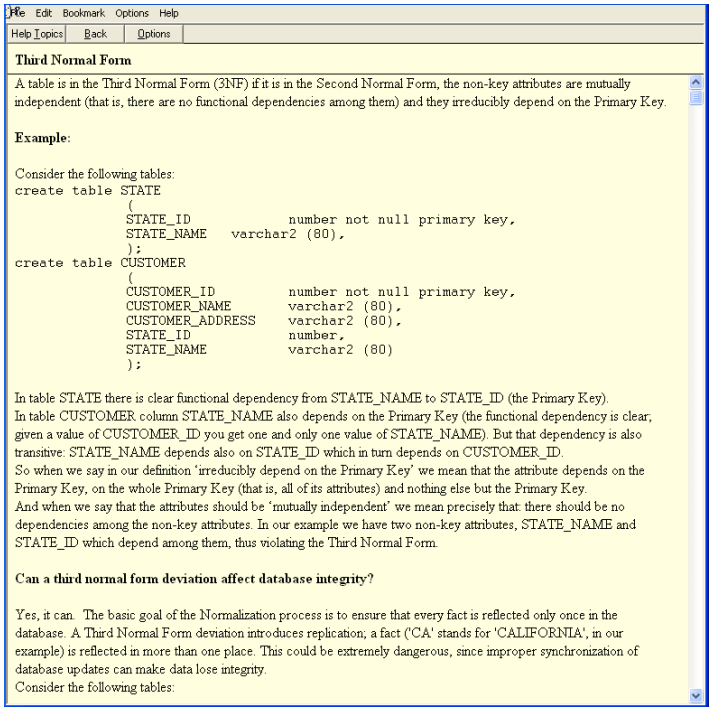
```
create table BANK
          (
          BANK_ID         number not null,
          BANK_NAME       varchar2 (45) not null,
          Constraint PK_BANK
              primary key (BANK_ID)
          );

create table BANK_BRANCH
          (
          BANK_ID         number not null,
          BRANCH_ID       number not null,
          BRANCH_NAME     varchar2(45) not null,
          constraint PK_BANK_BRANCH
              primary key (BANK_ID, BRANCH_ID),
          constraint FK_BRANCH_BANK foreign key (BANK_ID)
            references BANK (BANK_ID)
              );

create table ACCOUNT
          (
          BANK_ID         number not null,
          BRANCH_ID       number not null,
          ACCOUNT_ID      number not null,
          CUSTOMER_NAME   varchar2(45) not null,
          constraint PK_ACCOUNT
              primary key (BANK_ID, BRANCH_ID, ACCOUNT_ID),
          constraint FK_ACCOUNT_BANK foreign key (BANK_ID)
              references BANK (BANK_ID)
          );
```

that will result in the following Entity-Relationship Diagram:

4. Click Teach Me under the Normalization diagnostic for the Third Normal Form for the following information:



File Edit Bookmark Options Help

Help Topics | Back | Options

Third Normal Form

A table is in the Third Normal Form (3NF) if it is in the Second Normal Form, the non-key attributes are mutually independent (that is, there are no functional dependencies among them) and they irreducibly depend on the Primary Key.

Example:

Consider the following tables:

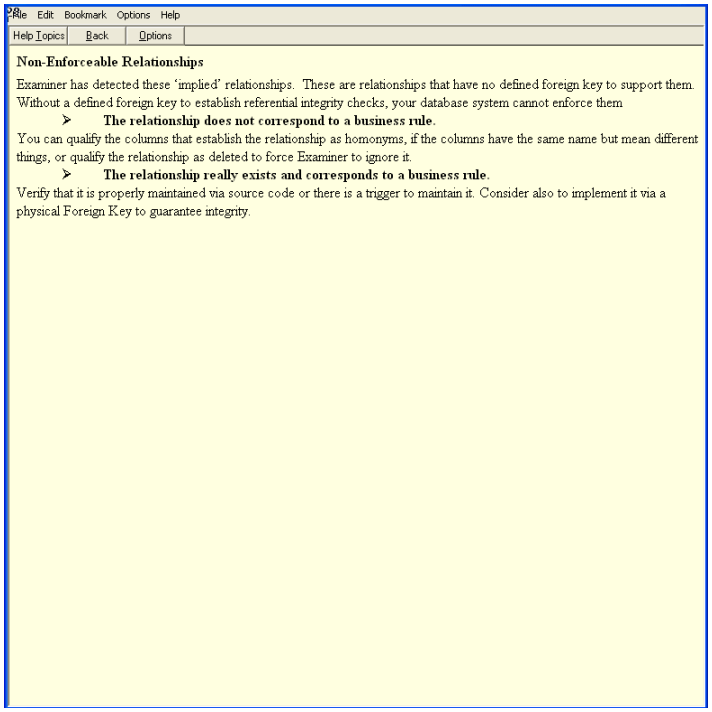
```
create table STATE
             (
             STATE_ID            number not null primary key,
             STATE_NAME   varchar2 (80),
             );
create table CUSTOMER
             (
             CUSTOMER_ID         number not null primary key,
             CUSTOMER_NAME       varchar2 (80),
             CUSTOMER_ADDRESS    varchar2 (80),
             STATE_ID            number,
             STATE_NAME          varchar2 (80)
             );
```

In table STATE there is clear functional dependency from STATE_NAME to STATE_ID (the Primary Key).
In table CUSTOMER column STATE_NAME also depends on the Primary Key (the functional dependency is clear; given a value of CUSTOMER_ID you get one and only one value of STATE_NAME). But that dependency is also transitive: STATE_NAME depends also on STATE_ID which in turn depends on CUSTOMER_ID.
So when we say in our definition 'irreducibly depend on the Primary Key' we mean that the attribute depends on the Primary Key, on the whole Primary Key (that is, all of its attributes) and nothing else but the Primary Key.
And when we say that the attributes should be 'mutually independent' we mean precisely that: there should be no dependencies among the non-key attributes. In our example we have two non-key attributes, STATE_NAME and STATE_ID which depend among them, thus violating the Third Normal Form.

Can a third normal form deviation affect database integrity?

Yes, it can. The basic goal of the Normalization process is to ensure that every fact is reflected only once in the database. A Third Normal Form deviation introduces replication; a fact ('CA' stands for 'CALIFORNIA', in our example) is reflected in more than one place. This could be extremely dangerous, since improper synchronization of database updates can make data lose integrity.
Consider the following tables:

5. Click Teach Me under the Relationships diagnostic for Nonenforceable Relationships, to learn the rules that govern "implied" relationships:



File Edit Bookmark Options Help

Help Topics | Back | Options

Non-Enforceable Relationships

Examiner has detected these 'implied' relationships. These are relationships that have no defined foreign key to support them. Without a defined foreign key to establish referential integrity checks, your database system cannot enforce them

- **The relationship does not correspond to a business rule.**

You can qualify the columns that establish the relationship as homonyms, if the columns have the same name but mean different things, or qualify the relationship as deleted to force Examiner to ignore it.

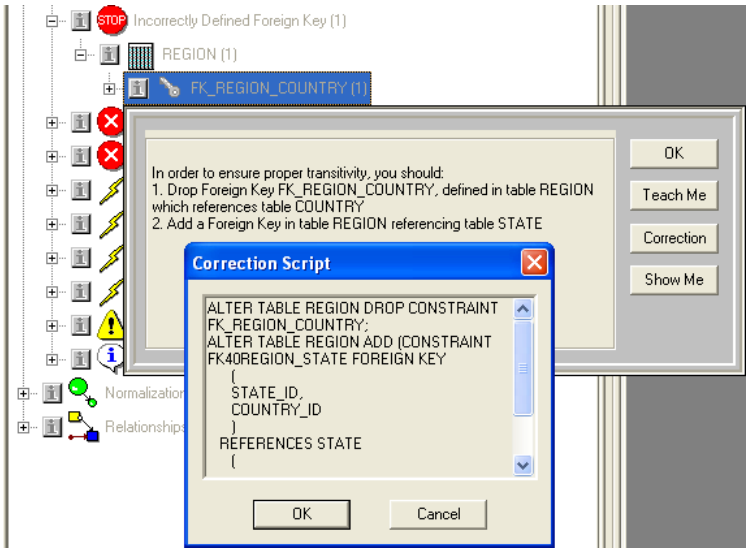
- **The relationship really exists and corresponds to a business rule.**

Verify that it is properly maintained via source code or there is a trigger to maintain it. Consider also to implement it via a physical Foreign Key to guarantee integrity.

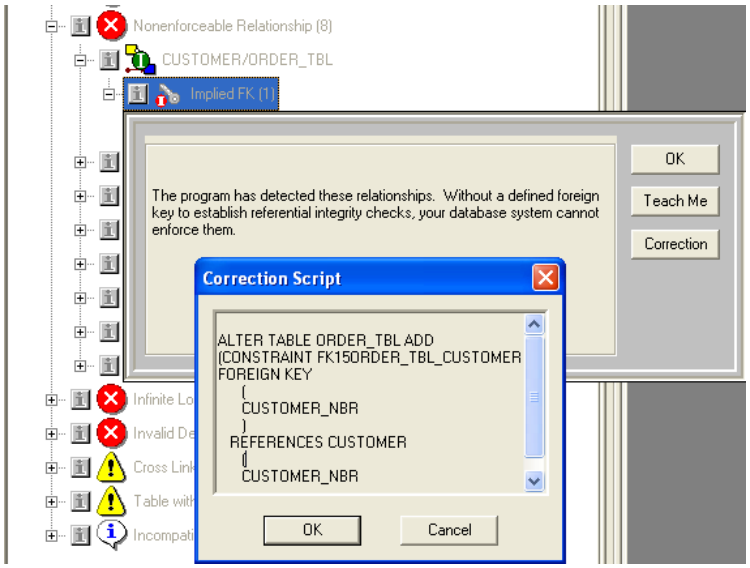
Use Scripts To Solve Problems

In certain situations, SQL/DDL scripts are generated to correct detected problems. For instance, the following example shows that the Foreign Key FK_REGION_COUNTRY should be dropped, and a new FK in table REGION referencing table STATE should be created. All you need to do is click the Correction button. Then the script for corrective action is generated.

The following diagram shows an example:

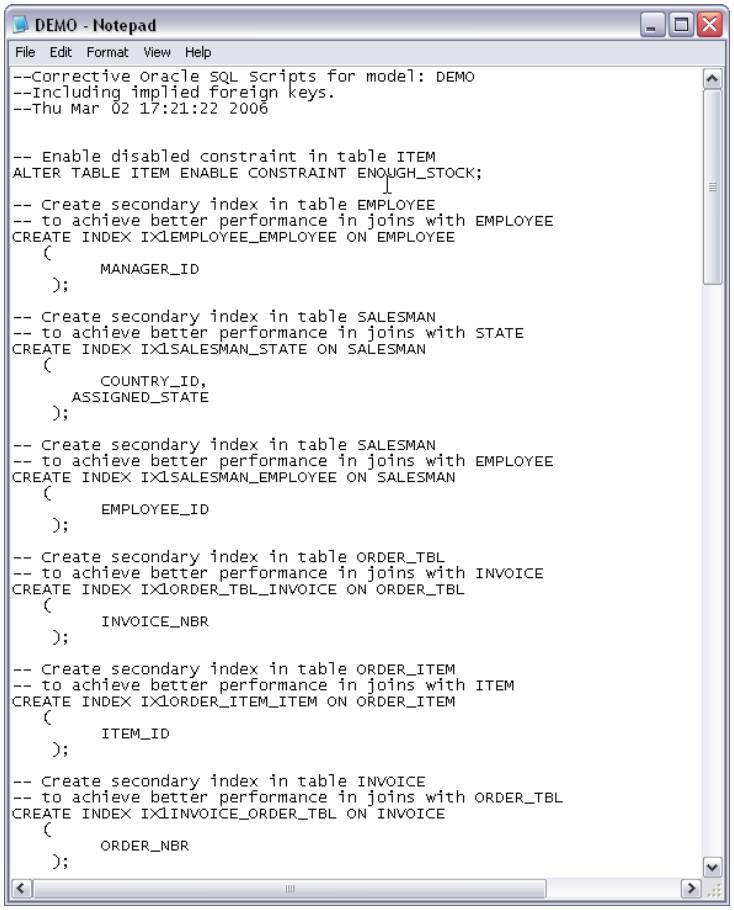


In the next example, the scripts are generated to implement the Foreign Keys and the implied relationships so that the database system will be able to enforce these references:



SQL Script Generation

To generate all the corrective scripts, click File, SQL generation, New File. The result looks like the following:



Appendix A: Frequently Asked Questions

This section contains the following topics:

Models

Question: What kinds of models can I validate with CA ERwin DMV?

Answer: Models created with CA ERwin DM (.erwin files), models from a script file with SQL/DDL statements, and database schemas from the DBMS Catalog or Dictionary.

Question: Does CA ERwin DMV process CA ERwin DM subject areas?

Answer: Yes. When you create a new project from an .erwin file, CA ERwin DMV will detect the subject areas and at this time you may select to process one subject area or none. Each subject area will create a sub model, associated to the main model.

Question: Can I drop a table from CA ERwin DMV analysis, without dropping the table from the CA ERwin DM model?

Answer: Yes. You can do this in several ways. One way is to go to User definitions, Table Processing. Another way is to right-click on the table on the Table tree and select Drop Table. A third way is to right-click on the tree on the entity-relationship (ER) diagram and select Drop Table.

Question: Can I add knowledge about my specific model to CA ERwin DMV?

Answer: Yes. There are several things you can do. You can define synonyms, define homonyms, and qualify columns, tables and relationships. Synonyms allow the user to define that two or more columns have different names but have the same meaning and therefore should be treated as the same. Homonyms are the opposite of Synonyms. Homonyms are when two or more columns have the same name, but they have different meanings and therefore should be treated differently.

Question: Are there any limits to the number of tables in the model?

Answer: No. There are no limits other than the ones imposed by your machine (for example, memory and speed). Models have been processed with well over 2000 tables. In this case, it is suggested that the model be processed initially with the FKs Only option. Later you may want to run the model again using the ALL relationships instead of the FKs Only option. This procedure will save you time.

Implied Relationships

Question: Can I detect relationships that are not defined using the Foreign Key mechanism?

Answer: Yes. CA ERwin DMV will infer implied relationships, based on same name attributes. For example, if in table CLIENT there is a PK based on column Cli_ID and in the ORDERS table there is a non-key column attribute Cli_ID, CA ERwin DMV will detect this relationship as being a one-to-many (1:N) from CLIENT to ORDERS. It is a 1:N because on the CLIENT side the column is unique (it is a PK) and on the ORDERS side the column is non-unique (non-key attribute).

Question: If I do not have a good naming convention how can I take advantage of the implied relationships detection?

Answer: CA ERwin DMV allows you to define synonyms for the attributes. If in one table you use column Cli_ID and in another table Cli#, you can define these as synonyms. Go to User Definitions and select Synonyms. This way CA ERwin DMV will consider these two as the same.

Question: What do I have to do to make CA ERwin DMV detect implied relationships?

Answer: CA ERwin DMV is installed with the implied relationship detection as a default. If you want to change this setting, go to Tools, Preferences, and Model Settings and select the proper option for Relationship Processing. The option Foreign Keys Only indicates that no implied relationships will be detected. The option ALL indicates that implied relationships will be detected.

Question: Do I have any overhead using implied relationships?

Answer: There is always some overhead detecting implied relationships, but this is only noticeable in very large models. CA ERwin DMV must compare every attribute in every table with the attributes in every other table.

Diagnostics

Question: Can I select which diagnostics to run so I can define a quality standard for my department?

Answer: Yes. CA ERwin DMV allows the user to define (a) which diagnostics to run and (b) the severity level associated with each diagnostic. You must go to Tools, Preferences and in the Diagnostics Tab, click the Select Individual Diagnostics button. You can make the changes for a specific run or you can save them as a profile that can be associated with any model.

Question: Are there any user-defined diagnostics?

Answer: Yes. CA ERwin DMV provides some user-defined diagnostics. You must go to Tools, Preferences and in the Diagnostics Tab, click the Select Individual Diagnostics button. Then go to the Index and Constraint Diagnostics Tab. You may select the user-defined diagnostics you want and set the corresponding thresholds.

Question: Can I hide any diagnostics that I am not interested in?

Answer: Yes. Just right-click on the diagnostic that you want to hide and select Hide This Diagnostic. You cannot hide individual occurrences of the diagnostic, but you may hide the whole diagnostic.

Question: How can I see the diagnostics by severity instead of by category?

Answer: Right-click on the diagnostics tree and select Order by Severity. If you want to change back to diagnostics by category, right-click on the diagnostics tree and select Order by Category.

Question: Can I print the diagnostic reports?

Answer: Yes. You can print not only the diagnostic reports, but you can also print several other reports produced by CA ERwin DMV. Go to Reports and select what you want.

Entity-Relationship (ER) Diagrams

Question: Can I change the display of the ER diagram?

Answer: Yes. Choose Tools, Preferences, Model Display Settings and select the options you want. You may select to display the table and all its columns (All Columns), the table name and the PK name (Table Name Only) or the table and all its indexes (All Indexes). You may also select to display the data types (Show Data Types) and the user defined data types (Show User Defined Data Types).

Question: How can I display lines instead of connector boxes in the ER diagram?

Answer: CA ERwin DMV can draw the diagram using connector boxes only, lines only, and a combination of lines and connector boxes. To display lines only, unclick the Use Connector Boxes in the Tools, Preferences, and Model Display Settings. To display connector boxes only, click the Use Connector Boxes in the Tools, Preferences, Model Display Settings and set to zero (0) the Connector Distance. To display lines to all adjacent tables and connector boxes in all other relationships, click the Use Connector Boxes in the Tools, Preferences, Model Display Settings then set to one (1) the Connector Distance. This change will only take place after you go to View, Auto Arrange and select one of the options.

Question: Can I navigate through the ER diagram?

Answer: Yes. If you have a connector box, just click on it and CA ERwin DMV will take you to the table that the connector box points to. If there is a line representing the relationship, click on the line and select where you want to go (to the parent or the child table), by selecting Scroll to Related Table. You can also right-click on any table and select Scroll to Related Table.

Question: Can I change a line to a connector box and vice-versa?

Answer: Yes. If you have a connector box, right-click on the connector box and select Change to Line. If you have a line, right-click on the line and select Change to Connector.

Question: Can I synchronize the diagnostic message with the ER diagram to better understand the diagnostic?

Answer: Yes. First the ER diagram must be open on the left side of the screen. Then you select any diagnostic and click on the specific table related to the diagnostic; then right-click on the table and select Scroll to Table in Model.

Index