

CA ERwin® Data Profiler

CA ERWIN® DATA PROFILER HELPS ORGANIZATIONS LOWER THE COSTS AND RISK ASSOCIATED WITH DATA INTEGRATION BY PROVIDING REUSABLE, AUTOMATED, CROSS-DATA-SOURCE DISCOVERY, ANALYSIS AND PROFILING.

Overview

The costs associated with poorly documented, or completely undocumented, data sources often represent more than 50 percent of a project's overall budget — putting the success of any data warehouse (DW), master data management (MDM) or integration-oriented data management project at risk. CA ERwin Data Profiler can help you significantly reduce the time and effort involved in understanding and cross-referencing these data sources by assisting in the creation of accurate, reusable data models.

Benefits

CA ERwin Data Profiler reduces the time required to document and cross-reference data sources while increasing the accuracy and depth of your data foundation. As a result, you achieve a continual return on investment (ROI) in the form of reduced integration, delivery and maintenance costs, increased data quality and accuracy and the effective leveraging of strategic information assets for competitive business intelligence (BI) and mission-critical information governance.

The CA Advantage

A low-cost alternative that extends the value of data profiling technology, CA ERwin Data Profiler combines proven techniques with value-added capabilities to help you synchronize the data design within your database. With such functionality as primary key (PK) and foreign key (FK) discovery, multi-source overlap and mapping and metadata inference and exchange, the solution helps you better align your business with CA's Enterprise IT Management (EITM) vision.

When System Analysis Puts Strategic Data Initiatives at Risk

For most organizations, data-source discovery, analysis, documentation and comparison are major resource drains during data-intensive projects. Since the expenses associated with these ongoing tasks often represent a large percentage of a given project's overall cost, strategic initiatives are often delayed, reduced in scope or cancelled altogether due to a lack of funding and resources.

It is critical that you always know what data you have, where it is located, how it is structured and how it is related between disparate systems before implementing a data governance, data management or data integration project. Understanding your environment is extremely useful as a first step because it provides insight into your existing architecture and allows a data modeler to leverage the as-is structures to better design the to-be target. It also provides information that is critical for integrating or migrating your legacy data into a new environment.

For a variety of reasons, however, most companies do not have a good understanding of their existing distributed data landscape. As subject matter experts move on, design specifications disappear or become outdated and mergers and acquisitions create massive redundancy and semantic reconciliation issues, you must still have a centralized understanding of where your data is located and how you can leverage it to ensure the success of your strategic data initiatives.

The Power of Reusable, Multi-source Data Analysis

A source system analysis and data profiling solution, CA ERwin Data Profiler is the newest addition to the CA ERwin Modeling family, offering integration with visual data modeling to enable data and metadata discovery, analysis, design, documentation and reuse in a single product family. It can be deployed throughout your organization for use in MDM, data migration, data integration BI and DW projects. The solution delivers robust data profiling management, cross-system overlap analysis and data model reconciliation that includes:

- Support for Open Database Connectivity (ODBC) and flat-file data sources
- Data profiling capabilities to analyze individual columns within tables
- Fully automated discovery of primary and foreign keys and orphaned rows
- Simultaneous cross-source data analysis of up to 20 data sources to identify overlapping and unique attributes and to discover attribute supersets and subsets
- Interactive, side-by-side previewing and comparing of record values within and between different data sources
- The ability to define and persist critical data elements within data sources
- Export of all discovered metadata for reuse with CA ERwin Data Modeler, which allows the design reality exposed in the instance data to be compared with the data model or database metadata

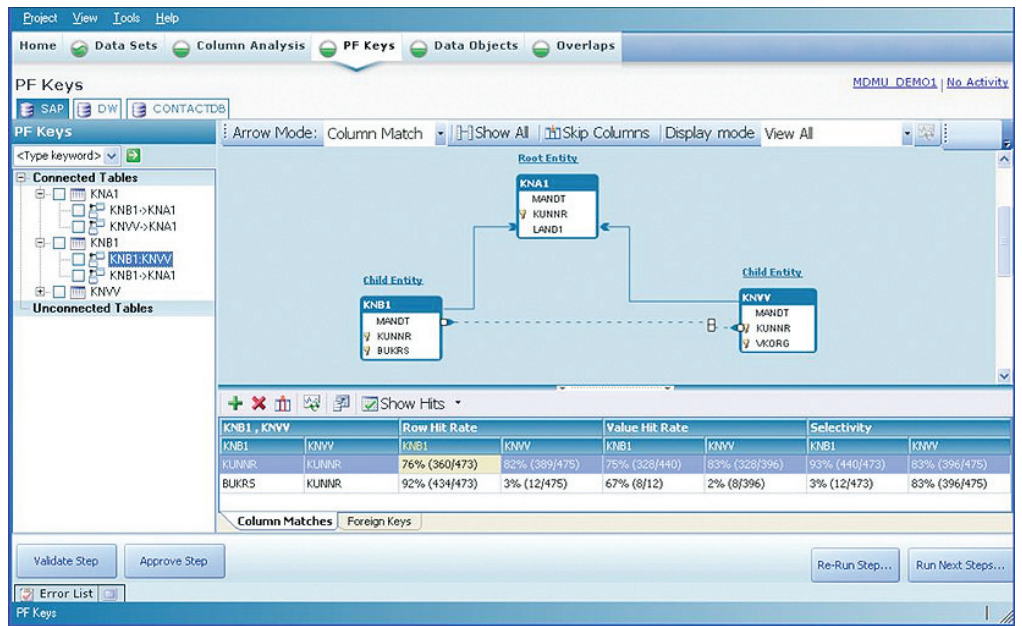
CA ERwin Data Profiler provides an entry-level product that ensures rapid and comprehensive ROI. By leveraging the solution, you can immediately improve the value of the information that drives your business, as well as increase your ability to respond to ongoing and future business needs.

CA ERwin Data Profiler also includes robust PK-FK discovery and cross-system overlap analysis. Integration with CA ERwin Data Modeler enables you to discover, audit and remediate data-related business rules and unknown design inconsistencies — as well as visualize the resultant discovered and inferred metadata in a data modeling environment — providing maximum reusability of the analyst’s work product.

FIGURE A

The automated discovery and visualization of PK-FK relationships accelerates the understanding and documentation of legacy or loosely defined data structures.

AUTOMATED DISCOVERY OF PRIMARY KEY-FOREIGN KEY RELATIONSHIPS



Data Profiling in Your Unique IT Environment

With support for a wide variety of environments, databases and source systems, CA ERwin Data Profiler offers the right combination of flexibility and compatibility, helping you leverage your unique IT components to achieve strategic data initiatives.

SUPPORTED ENVIRONMENT

- Microsoft Windows XP

SUPPORTED DATABASES FOR REPOSITORY AND STAGING AREAS

- Oracle 10g
- IBM DB2 UDB 8.1

SUPPORTED SOURCE SYSTEMS FOR DATA PROFILING AND ANALYSIS

- Relational databases
- Spreadsheets
- VSAM

Decreasing Data Analysis Cycles

By reducing the time required to document cross-reference data sources and increasing the accuracy and depth of understanding of an organization's data foundation, CA ERwin Data Profiler facilitates ongoing ROI with:

- Reduced integration, delivery and maintenance costs
- Increased data quality and accuracy
- Improved utilization of strategic information assets for competitive BI and mission-critical data governance.

CA ERwin Data Profiler includes a host of features and functionality engineered to help you not only gain a better understanding of your distributed data landscape, but also increase your level of control over important business information — so you can effectively pursue and complete strategic data initiatives.

FIGURE B

CA ERwin Data Profiler includes functionality that helps you better understand your distributed data landscape and effectively pursue strategic data initiatives.

FEATURES AND BENEFITS OF CA ERWIN DATA PROFILER

FEATURES	CAPABILITIES	BENEFITS
<ul style="list-style-type: none"> • Column Profiling and Analysis 	<ul style="list-style-type: none"> • Automatically discovers standard column statistics 	<ul style="list-style-type: none"> • Improves speed and effectiveness of analysis cycles through consistent and standardized results
<ul style="list-style-type: none"> • Primary Key-Foreign Key Discovery 	<ul style="list-style-type: none"> • Automatically discovers, defines and visualizes relationships within a single data source 	<ul style="list-style-type: none"> • Provides insight into legacy data structures and documented metadata, so analysts can account for inherent relationships
<ul style="list-style-type: none"> • Cross-system Attribute Overlap Analysis 	<ul style="list-style-type: none"> • Identifies overlapping and unique attributes, as well as attribute supersets and subsets 	<ul style="list-style-type: none"> • Speeds reconciliation of disparate data sources and enables effective identification of transformation requirements
<ul style="list-style-type: none"> • Data Synchronization Analysis 	<ul style="list-style-type: none"> • Validates the global identifiers and ensures that cross-source alignment and synchronization can occur 	<ul style="list-style-type: none"> • Ensures data consistency across existing sources and aids in designing well-aligned target structures
<ul style="list-style-type: none"> • CA ERwin Data Modeler Integration 	<ul style="list-style-type: none"> • Creates CA ERwin Data Modeler data models based on metadata inferred from the instance data and profiling results 	<ul style="list-style-type: none"> • Provides reusable data for documentation, impact analysis and stakeholder visualization purposes

COLUMN PROFILING AND ANALYSIS Helps reduce analysis cycles and increases their effectiveness through consistent and standardized results. In addition, this functionality enables the automatic discovery of standard column statistics, including:

- Data type
- Value frequency
- Length
- Precision
- Scale
- Formats
- Cardinality
- Selectivity
- Non-null selectivity
- Non-null count
- Min
- Max
- Mode
- Mode %
- Null count
- Blank count

PRIMARY KEY-FOREIGN KEY DISCOVERY Offers fully automated discovery, definition and visualization of relationships within a single data source. Inferred relationships provide critical structural insights into legacy data structures, as well as points of comparison and confirmation for documented metadata — allowing your analysts and designers to account for inherent relationships when building target systems.

CROSS-SYSTEM ATTRIBUTE OVERLAP ANALYSIS Performs an automated cross-compare of all columns across many data sources (up to 20) in order to establish a baseline of overlapping data. By leveraging this functionality, you can discover attribute supersets and subsets, as well as overlapping and unique attributes, to speed reconciliation of disparate data sources and effectively identify and document transformation requirements.

FIGURE C

The automated analysis of attribute overlap and mapping between disparate systems significantly reduces cycles and costs associated with traditional data integration initiatives.

MULTI-SOURCE ATTRIBUTE OVERLAP ANALYSIS



DATA SYNCHRONIZATION ANALYSIS Validates the uniqueness of a potential global identifier in each data source and then confirms that data across sources can be aligned and synchronized using this identifier. This functionality enables you to prototype and test survivorship rules between sources before you move the data into a master structure, helping you ensure quality and consistency across existing sources and well-aligned target structures.

CA ERWIN DATA MODELER INTEGRATION Enables the creation of data models that can be persisted and reused within CA ERwin Data Modeler for documentation, impact analysis and stakeholder visualization purposes. Based on metadata inferred from the instance data and profiling results in CA ERwin Data Profiler, these models help you achieve proper business alignment, improved end-user understanding and reduced analysis cycles in subsequent projects.

The CA Advantage

CA ERwin Data Profiler helps increase the quality of your critical data assets by performing cross-system analysis, generating robust data quality metrics and statistics and validating raw data with database design and architecture. A strong understanding of your existing data infrastructure helps reduce the risk of rework and the cost and time involved in complex data integration, data migration, data warehouse and master data management projects.

The Enterprise IT Management (EITM) Vision

CA ERwin Data Profiler, a key component in the CA ERwin Modeling family of solutions, is also an integral part of CA's Enterprise IT Management (EITM) vision to unify and simplify overall IT management. And with a comprehensive portfolio of modular IT management solutions, CA empowers you to better manage risk, costs and service — helping ensure that IT meets the business needs of your enterprise.

Gaining mastery over business-critical information is a crucial first step in any enterprise-wide IT management initiative. And when integrated as an element of CA's EITM vision, CA ERwin Data Profiler can help you maximize the performance, reliability and efficiency of your overall IT environment. The solution helps you discover, control and manage the data that powers your business, so you can ultimately extend the overarching CA EITM philosophy to operations, storage, security, lifecycle and services management — and unify, simplify and secure IT.

Next Steps

CA ERwin Data Profiler is a proven, flexible solution that provides reusable, automated, cross-data-source discovery analysis and profiling for improved data discovery and control. See how it can enable your organization to achieve strategic data initiatives by reducing the costs and risks associated with data integration. For more information, contact your local CA reseller or go to ca.com/contact/rmdm.

To learn more, and see how CA ERwin Data Profiler can help you unify and simplify data management for better business results, visit ca.com/modeling.

Copyright © 2009 CA. All rights reserved. All trademarks, trade names, service marks and logos referenced herein belong to their respective companies. This document is for your informational purposes only. To the extent permitted by applicable law, CA provides this document "As Is" without warranty of any kind, including, without limitation, any implied warranties of merchantability or fitness for a particular purpose, or non-infringement. In no event will CA be liable for any loss or damage, direct or indirect, from the use of this document including, without limitation, lost profits, business interruption, goodwill or lost data, even if CA is expressly advised of the possibility of such damages. CA does not provide legal advice. Neither this document nor any software product referenced herein shall serve as a substitute for the reader's compliance with any laws (including but not limited to any act, statute, regulation, rule, directive, standard, policy, administrative order, executive order, etc. (collectively, "Laws")) referenced herein. The reader should consult with competent legal counsel regarding any such Laws.

334530109

Learn more about how CA can help you
transform your business at ca.com

