



Hello Readers

For some of us March 31st is the end of one year and the beginning of another year.

In a challenging year the value delivered by ERwin to the business has been increasingly brought into sharp focus and this recognition has seen Sandhill being asked to help and support significantly more ERwin new clients.

As this year comes to a close for us, we begin to plan to meet further increases in demand for ERwin and quality support services in the coming year. We will continue to extend our reach into new cities and countries. We will continue to look for ways to enhance the value of ERwin within the ERwin community and we will continue to provide the ERwin support that you appreciate.

In the coming year, we will look to hold additional seminars and workshops. In addition we have updated our Educational Offerings and we are actively looking at holding public ERwin training classes in major cities around the world. Please feel free to contact us if you would like more information on our classes.

As always part of this column is to inform you of the latest version of ERwin. As of this time it is V7.3.9 (build 2386). If you would like a copy of this build and you are on maintenance, please let us know.

Regards

Robert Lutton,
Robert.lutton@sandhillconsultants.com
Editor, ERwin Times
Sandhill Consultants



CA World 2010 - May 16-20 - See you there

If you have not had the chance to register for CA World yet, please click on the link below. CA have now posted the ERwin sessions for this year's CA World (2010) and I have included the session names, times and links for you to view the details of each session when you get a moment.

Sandhill will be there supporting CA with two labs, Session # 4 & # 10. All topics are great for new and experienced ERwin users alike and we look forward to seeing as many folks as possible that can make it. Don't forget to schedule your one on one meetings as well.

Monday

1. [CA ERwin Modeling State of the Union](#) Monday 1:15 - 2:15pm
2. [The Zen of Metadata](#) Monday 2:30 - 3:30 pm

Tuesday

3. [Using CA ERwin Modeling to Assure Data and Metadata Quality](#) Tuesday 9:00 - 10:00 am
4. [Hands On Lab: Applied Dimensional Modeling](#) Tuesday 1:15 - 2:15 pm
5. [The Do's and Don'ts Regarding Data Modeling with CA ERwin DM in an Outsourcing Environment](#) Tuesday 2:30 - 3:30 pm
6. [Strategic Plan for CA ERwin Data Model and CA ERwin Model Manager](#) Tuesday 3:45 to 4:45pm

Wednesday

7. [Optimizing the Design of your Data Warehouse](#) Wednesday 9:00 - 10:00am
8. [A Practical Guide to Enterprise Data Modeling and Model Management](#) Wednesday 1:15 - 2:15 pm
9. [Hands On Lab: The CA ERwin Complete Compare Workshop](#) Wednesday 2:30 - 3:30 PM
10. [Easing the workload by customizing the FE Template](#) Wednesday 3:45 - 4:45
11. [CA ERwin Modeling Global User Community Meeting](#) Wednesday 5:00 - 6:00pm

Thursday

12. [ODBC Reporting with CA ERwin Data Modeler and Introduction to Crystal Reports](#) Thursday 9:00 am to 10:00 am

caWorld'10 | Las Vegas, Nevada, U.S.
May 16-20, 2010

Sandhill Draws Largest ERwin Crowd to Recent Dimensional Modeling workshop in Fort Worth Feb 5, 2010



F1. Left hand side of the class

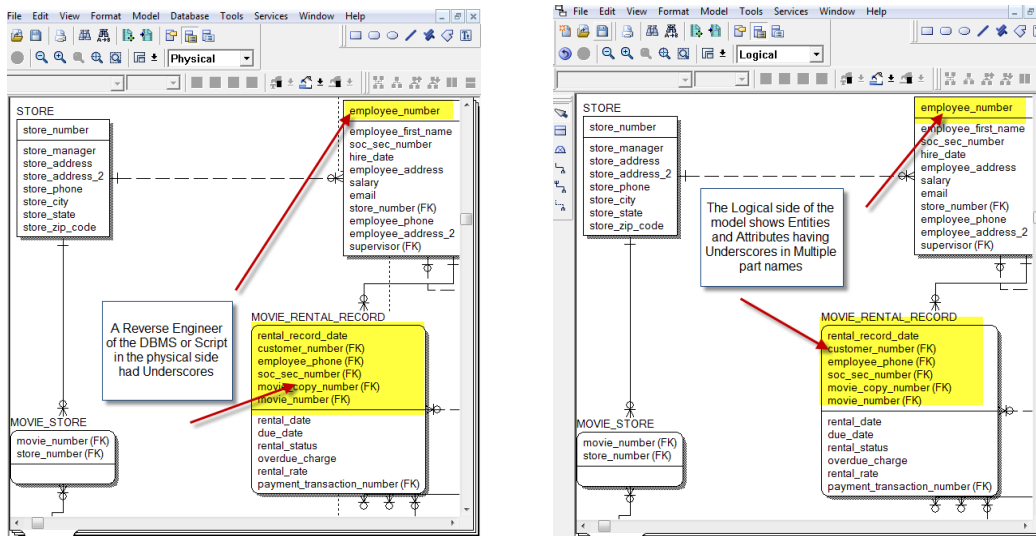
Over 65 folks attended our most recent Dimensional Modeling workshop in Fort Worth, TX, on Feb 5th, 2010.

We would like to extend our thanks to the entire staff at BNSF and for everyone who attended. All the feedback was excellent.

Look for more in other cities in the future.

Replacing Underscores in a logical side of a model when Reverse Engineering a DBMS/ Script?.

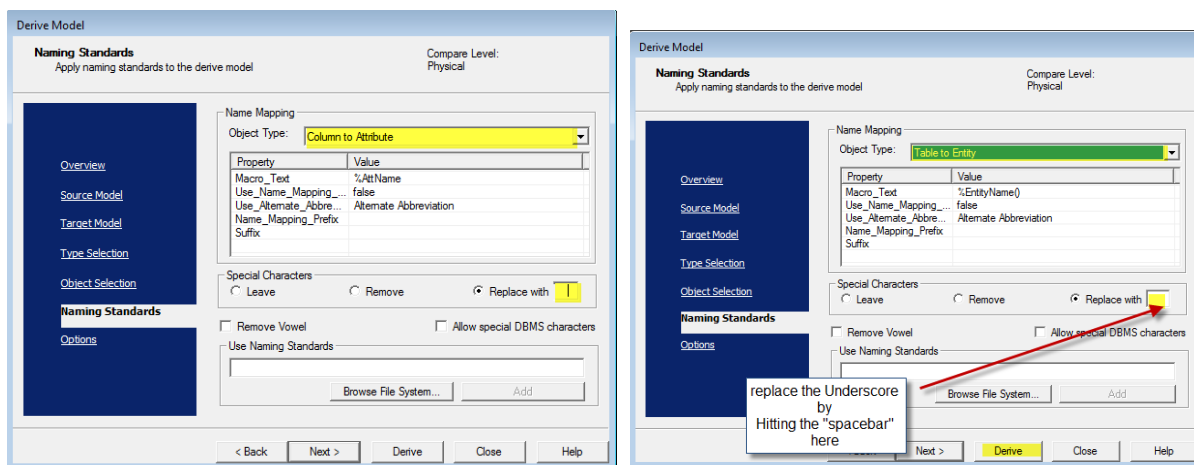
In version ERwin 7.x modelers/Users that R/E (Reverse Engineers) to a Logical /Physical model from the DBMS or SCRIPT resulted with Logical Entities and Attributes with multiple part names with UNDERSCORES logical side of the model.



Organizations have ask if there was a way to create a logical /physical models where the logical side name should have spaces i.e. "MOVIE RENTAL RECORD" removing the underscores (as seen in the diagram above) and replacing them with spaces.

To get do this requires a 3 step process:

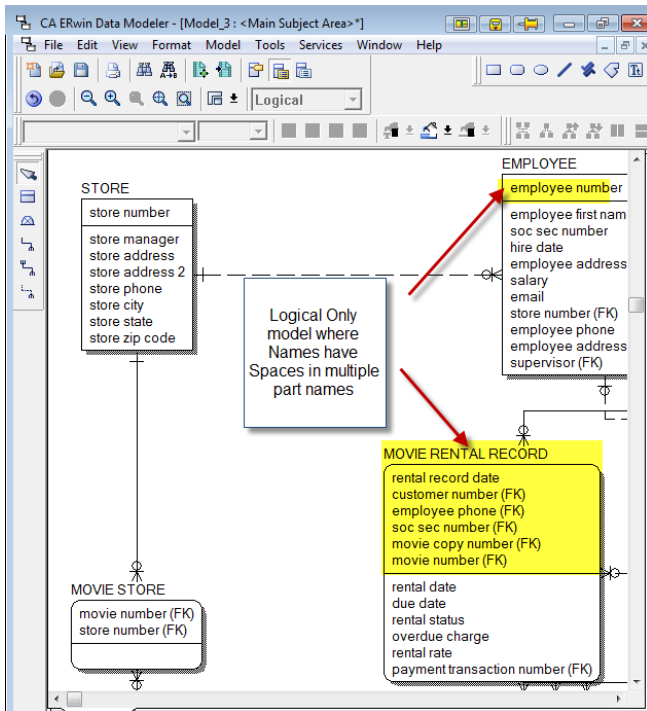
STEP 1: R/E the DBMS or Script to a PHYSICAL ONLY model and then save the physical model. Once the model is save derived a logical model and use the naming standards option to "replace" underscore with "_". The replacement of the "_" to "space" which requires the modeler/user to actually hit the spacebar in the dialog box as shown.



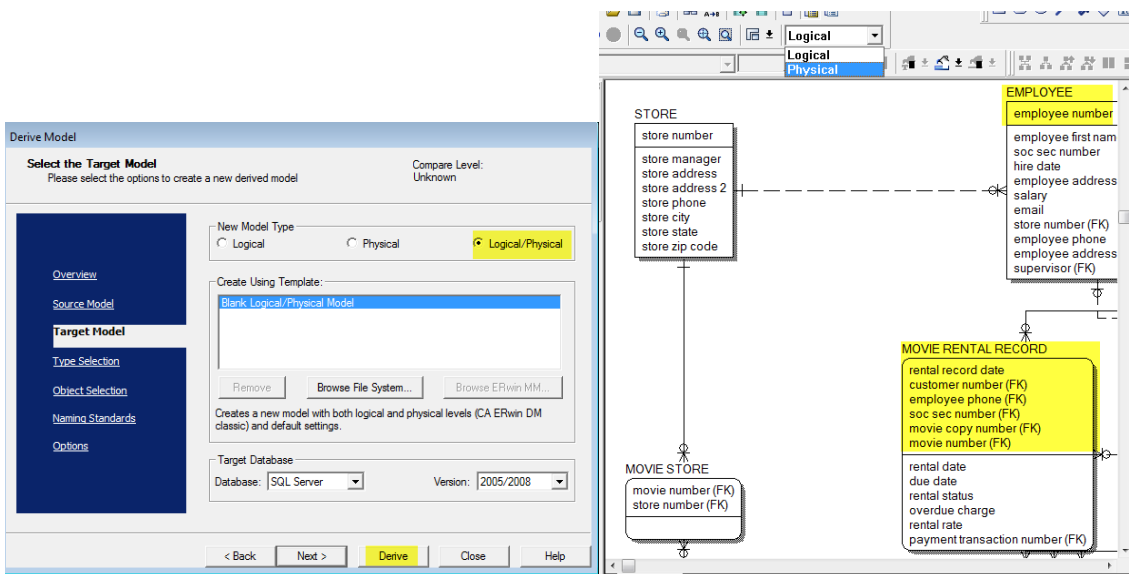
Care should be taken in order to preserve other dual objects that have logical / physical inheritance linked to them such as Relationship that translate to the physical side as "Foreign Key Constraints".

Replacing Underscores in a logical side of a model when Reverse Engineering a DBMS/ Script – Continued

Step 2: Derived a logical only model from the Reverse engineered Physical ONLY model. Save the LOGICAL ONLY model to the file system. The image below shows the derived logical only mode. Note: the Underscores are removed from the model.



STEP 3: Derive a Logical / Physical model from the newly save Logical only model. The result is the logical / Physical model with logical names without underscores!



To discuss more about this article see our linked in topic.

http://www.linkedin.com/groupAnswers?viewQuestionAndAnswers=&qid=1779772&discussionID=15967042&qoback=%2Eanh_1779772

The Data Warehouse Corner...

This issue -Normalization...myths and facts

Welcome to the Data Warehouse Corner. Each issue we'll identify a business problem facing the Data Warehouse / BI data modeller and discuss solutions and implications for the ERwin modeller. In this article we'll identify how normalization and denormalization differ between OLTP and DATA WAREHOUSE models.

We've all heard about normalization, based on CODDS theory and how it's supposed to be applied in the design phases of application development. First let's describe the requirements that get us to the level of intuitive (3NF) normal form:

1NF – no repeating groups – ensures model flexibility. Remove all repeating attributes or groupings of fields and place them in their own entity.

2NF – all non key attributes depend on the Primary Key (PK) – ensures model consistency. Ensure all non key attributes are dependent on all PK attributes.

3NF – no transitive dependencies. Ensure that no non-key attribute depends on any other non-key attribute.

Each of these "tests" ensure the flexibility, consistency and integrity of the OLTP model, and are applied in series (it is said that you can't be 'in' 3NF until you are in 2NF, but I can find exceptions to this rule). The further you go through normalization (4NF, BOYCE-CODD NF, 5NF), the less the risk that a processing inconsistency, however caused, will compromise the data integrity. Once the model is normalized, you can effectively denormalize the design for performance purposes (you can only denormalize a normalized model).

But let's reflect on OLTP normalization practices in the data warehouse. The worry over integrity and consistency is minimized since the ETL programmers spend all weekend fixing the source system data errors in the ETL layer. So our target data warehouse is as close to 100% accurate as possible, and we will market this aspect back to the business community, noting the data warehouse is the definitive source for answers to business questions. Further to this the data warehouse flexibility will be supported in the model through structural (intentional) redundancy in the Dimensional model designs, which is acceptable given that the ETL folks maximized integrity and consistency on the incoming side to the data warehouse.

How then can we normalize the DATA WAREHOUSE model, without the primary focus being data integrity and consistency, so that it leads to effective denormalization? Let's look at some alternative approaches.

First approach - DATA WAREHOUSE's are Subject oriented – so we can normalize the DATA WAREHOUSE data by grouping it by business subjects. So from all of the source applications, you can model a single CUSTOMER Subject, that comprises all of the required analytical data pertaining to CUSTOMER (there may be many entities in the CUSTOMER SUBJECT).

Second approach - DATA WAREHOUSE's are time variant, that is most analysis involves forecasting the future based upon previous events. The challenge arises when the data's nature or values changes to the extent that historical analysis can no longer be applied to the current state of the data. An approach that will yield much more flexibility is to group (normalize) the data into *classes* of data that change frequently, and separate them from classes of data that never change. This will lead to a more effective application of denormalization tactics, of which Ralph Kimball's Type 1,2, and 3 slowly changing dimensions are the most common examples.

Third approach - We can also normalize based on the nature by which we analyse the data – that is group together data that is analysed at specific times (State based analysis) vs data that is analysed as it happens (Event based analysis). State based analysis requires a balanced periodicity (e.g. every Sunday we publish a new state for the given data), whereas Event based analysis can be published as it happens in the source system (and is sometimes the focus of *operational analytics*).

Within denormalization, we strive for performance and availability, very typical goals with both OLTP and data warehouse systems. With an OLTP model the performance requirements are dictated within the Service Level Agreement obligations (maximum transactions per minute, speed of queries less than x seconds, transaction consistency and data consistency a must).

In the DATA WAREHOUSE we manage performance by firstly managing user expectations (providing realistic turn around times, such as 'no you can't get a 500 million row table to join a 100 M row table to join a 10M row table and run in less than 3 seconds!'), and by utilizing scalable hardware and DBMS objects (MQT's, Indexable VIEWS, Materialized Views). The point then is that denormalization in the DATA WAREHOUSE has similar considerations and approaches, with the recognition that we can focus on approaches that are not so centric to maintaining data integrity.

In this article we've provided the characteristics of 3NF within OLTP models, and identified why we need not be constrained by these requirements within the data warehouse. We've provided 3 alternative approaches to Data Warehouse normalization and discussed the goals of denormalization in each system.

Next time you hear the phrase "that data warehouse model is in 3NF", look closely at the structures and see if there are alternatives that may yield more flexibility for the data warehouse analyst.

Next issue: Slowly Changing Dimensions...who said they were slow?

Article by Joe Cullen, Sandhill Consultants
Joe.Cullen@sandhillconsultants.com