The Data Organization

1251 Yosemite Way
Hayward, CA 94545
(510) 303-8868
info@thedataorg.com

Best Practices

Database Naming Convention

By Rainer Schoenrank

Data Warehouse Consultant

August 2016

Biography

Rainer Schoenrank is the senior data warehouse consultant for The Data Organization. He has degrees in physics from the University of Victoria and computer science from the University of Victoria and California State University Hayward. He has built data warehouses for clients such as Pacific Bell, Genentech, GE Leasing, SGI, PPFA, Brobeck, Bank of America, Clorox and Leapfrog. He can be reached at rschoenrank@computer.org.

Table of Contents

1. INTRODUCTION

1.1 Purpose

The Database Naming Convention describes how to systematically create technical names for the objects within the database and the expectations of the definition of the data identified by the technical name.

By using systematic technical names, the database interface can achieve some level of independence from the DBMS supplier since the structures and objects of the database can be implemented within any DBMS with only minor changes and the table and column names will not require any modification.

1.2 Goals

The primary goal is to create names for database objects that are DBMS independent. Given the current crop of DBMS suppliers, the database object names are less than 31 characters long.

A goal is to enhance understanding by grouping:
- the tables associated with an entity or a measurement
- the database components associated with a table

1.3 Scope

The scope of the naming is limited to a single schema/database.
The convention names the following database objects only:
- Tables
- Table Triggers
- Stored Procedures
- Columns

1.4 Assumptions

Within a single database or schema, the assumptions being made are:

- This is a physical DBMS naming convention. There is no restriction on logical names
- There is no requirement to access tables defined outside of the database
- The technical (physical) name should be UPPER_CASE with underscores (_) between words.
- The format of database table names is <database name>.<table name>
- The format of database non table object names is <database name>.<object name>
- The format of database column names is <database name>.<table name>.<column name>
- Columns in different tables with the same column name mean the same thing. That is, the columns will have the same physical data type and their contents come from the same range of values. For example, the column <logical data type>_CREAT_DT appears in every table in the database. Its physical data type is always TIMESTAMP and it means "the date (and time) on which the table row was created."

2. SCHEMA / DATABASE NAME

The schema/database names will be in accordance with the naming standard provided by the Data Governance organization.

3. TABLE NAMES

3.1 Database Layers

The database for an application can be organized into divisions or layers. The divisions organizes the tables into groups by function to aid understanding. The division codes are usually 2 letter abbreviations such as:

Layer Level	Division Code	Name	Description
1	BM	Business Model	the data described by the logical data model
1	BV	Lookup Data	the data used to validate various table columns
2	DC	Database Configuration	the data used to configure the database and hold the database error log
3	AD	Audit (SOX)	the logs containing the data about the database accesses
3	AU	Authorization (SOX)	the data used to restrict access to tables and rows
4	WF	Workflow	the data used for processing data movement
4	AP	Application Configuration	the data used to configure an application
4	UP	User Preferences	the data used to retain user preferences
Etc.			

3.2 Table Names

Always use singular names when naming tables (entities).

All table names should be spelled out. All tables should have a business name, a technical name and a description. The technical table name should not exceed 21 characters.

During conceptual design the tables will be described using the business names. During the logical data model to physical database transformation the business names will be replaced by technical names.

The table name format is:

<database division>_<conceptual entity name>_<abbreviated business name>

For example:

BM_CUST	- for Customer
BM_CUST_ADR	- for Customer Address
BM_PROD	- for Product
BM_CMPLNT_EVN	- for Complaint Event
BV_UOM_CD	- for Unit of Measure validation

4. TABLE COMPONENT NAMES

The table components are the table triggers and the stored procedures that operate on a single table.

4.1 Trigger Suffixes

Table triggers are used to ensure that the row metadata columns table contents conform to the requirements of Sarbanes-Oxley (SOX).

The trigger name format is:

<database division>_<table name>_<trigger suffix>

The trigger suffixes are:

Trigger Suffix	Description
TR_CRT	Trigger executes before the row is created
TR_UPD	Trigger executes before the row is updated

4.2 Stored Procedure Suffixes

Stored procedures are used to implement the programmer's interface to the database, standard stored procedures and business object stored procedures.

Standard stored procedures will access a single table per stored procedure call. They stored procedure will use the table name in the stored procedure name.

Business object stored procedures are used to implement the secure database interface and will access more than one table per stored procedure call. The business object names should not collide with the table names. The business object names will follow the abbreviation rules and the length will not exceed 21 characters.

The database interface stored procedure name format is:

<database division>_<business object/table name>_<suffix>

Copyright © 2016 Rainer Schoenrank

The stored procedure suffixes are:

Database Object	Database Function		Suffix	Description
Function	Read	Table	FN_RD	Reads a row from a database table
Function	Cursor	Table	FN_S01	Reads a set of rows from a database table
Procedure	Create	Table	SP_CRT	Inserts a new row into a database table
Procedure	Update	Table	SP_UPD	Updates an existing row in a database table
Procedure	Delete	Table	SP_DEL	Logically deletes a row in a database table
Procedure	Undelete	Table	SP_NDL	Undeletes a row in a database table
Function	Read	Business Object	BO_RD	Reads a business object
Function	Cursor	Business Object	BO_S01	Reads a set of business object
Procedure	Create	Business Object	BO_CRT	Inserts a new business object
Procedure	Update	Business Object	BO_UPD	Updates an existing business object
Procedure	Delete	Business Object	BO_DEL	Logically deletes a business object
Procedure	Undelete	Business Object	BO_NDL	Undeletes a row business object
Procedure	Process	Business Object	BO_X99	Processes a business object where X is a character and 99 is two digits used to uniquely identify the process

This naming convention ensures that all the database objects associated with an entity will be grouped together when the database objects are listed alphabetically.

5. COLUMN NAMES

5.1 Column Names

During logical data model design the columns (properties/attributes) will be described using the business names of the columns. During physical database design the business names will be replaced by technical names.

The format of the technical column name is:

<abbreviated business name>_<class word suffix>.

Column class word suffixes are given in the section Class Words.

All column names should be spelled out. All columns should have a business name, a technical name and a description. The technical column name should not exceed 31 characters.

5.2 Data Type Suffixes

Data definitions of the logical data model are:
- Explicit as possible. Definitions should be clearly understood by someone who has no prior knowledge of the subject.
- Reflections of the real world definition of their objects.
- Descriptions of what the data is not how it is used.
- Complete enough to ensure a thorough understanding of the data.

Class words are the components of the logical data types. The class word provides a way to classify the contents of a logical data type. Every column must be associated with a class word. Class words must be the rightmost word in all attribute or column names.

The physical column names of the logical data types have column suffixes that are given in the tables below.

Logical Data Type	Description	Component Name	Column Description	Class Word Suffix
Unstructured	An attribute that describes a business data object or data that has not been analyzed for structured contents	Description	Free-form text used to define, comment upon, or describe the object.	DESCR
Structured	Alphanumeric data that can be decomposed into attributes within the structured data object. For example, address, phone number, etc.	Structured		
Address	The description of a place where mail is delivered	Addressee	recipient	_ADRE
		Company	organization	_CO
		Suite	suite or apartment	_APT
		Street	street address	_STRET
		City	name of city	_CITY
		State	name of state	_ST
		Postal Code	postal or zip code	_PSTL_CD
		Country	name of country	_CTRY
Document	The number and line of an accounting document	Document Application	the leapfrog application that created the document	_APPL_CD
		Document Identifier	The document identification string	_DOC_ID
		Document Line Number	the line number in the document	_LN_NBR
Identification	The type and identification number of an externally assigned identifier	Identification Type	the type of identification, e.g. Key, drivers license, passport, etc.	_ID_CD
		Identification Owner	the name of the authority that issued the identification	_ID_OWN_CD
		Identification String	the identification string	_ID_NBR
Person Name	The structure of the name of a person	Title	the person's title, e.g. Mr., Ms., etc.	_TITL_CD
		First Name		_FRST_NM
		Nickname		_NK_NM
		Middle Name		_MDL_NM
		Last Name		_LAST_NM
Phone number	The address of a handset (logical phone line port) on the global telephone network	Country Code	the country portion of the phone number	_CTRY_NBR
		Area Code	the area portion of the phone number	_AREA_NBR
		Exchange Code	the telephone exchange	_EXCH_NBR
		Station Code	the station number	_STATN_NBR
		Extension	the extension of the phone	_EXT_NBR

Logical Data Type	Description	Component Name	Column Description	Class Word Suffix
Enumerated	An attribute that is constrained to a list of allowed values. For example, color, unit of measure, currency, etc.	Code		_CD
Indicator	An attribute whose values can only be yes, no or unknown. The default value is unknown	Indicator		_IND
Ordinal	An attribute used to order a series of business data objects	Sequence Number		_NBR _CNT
Monetary Amount	An attribute that represents an amount of money in a business transaction. Implemented as three columns in the database	Amount	The value of the monetary amount.	_AMT
		Currency	The currency of the monetary amount	_CRNCY_CD
		Rate	the frequency at which the monetary amount is paid	_RATE_CD
Measured Quantity	An attribute that represents a physical measurement on the business data object. Implemented as two columns in the database	Quantity	The value of measurement that was made.	_QTY
		Unit of Measure	unit of measure of the measurement	_UOM_CD
Percentage	An attribute that records the number of parts in or to every hundred.			_PCT
Table Key	A system-generated number used to uniquely designate each occurrence of one business data object			_KEY _ID
Event Timestamp	the date and time of a recorded event. Includes year, month, day, hour, minutes, seconds			_TMSTP
Calendar Date	A specific chronological point using calendar conventions.			_DT
Clock Time	The time within the date			_TM
Time Interval	The specification of a time interval with a start and an end. The implementation requires two columns in the data base	Start Date	start of time interval	_BGN_DT
		End Date	end of time interval	_END_DT

6. ABBREVIATION PROCEDURE

This section details how to determine if a word or term is approved, how to create an abbreviation if the term is not approved, and how to get approval for the abbreviation.

If required, a term may have two abbreviations: a standard abbreviation and a "tiny" abbreviation. The tiny abbreviation will only be created if a column name cannot be shortened without dropping terms. The standard abbreviation is to be used whenever possible.

6.1 Abbreviations

For frequently used or long terms, standard abbreviations are recommended to help keep name lengths reasonable. In general, abbreviated variable names greater than 30 characters can be difficult to read. When using abbreviations, make sure they are consistent throughout the entire application. Randomly switching between Cnt and Count within a project will lead to unnecessary confusion.

A listing of the currently approved abbreviations is found in the EXCEL file:

StandardAbbreviations.xlsx

6.2 Creating a New Abbreviation

A procedure for creating new abbreviations is:

1. One, two, three and four letter words are not abbreviated.

2. Five letter words are abbreviated in some cases. They may be abbreviated with abbreviations in common usage.

3. Words of six or more letters are abbreviated.

4. To create an abbreviation for a word, vowels should be removed, except where they are the first letters.

5. In abbreviations, redundant (double) consonants should be removed.

6. Strive to use language, company and the industry standard abbreviations.

7. Use acronyms for abbreviations of multiple words where they are standard to the company or industry, e.g., "DC" for Distribution Center.

8. The business term should be easily identifiable from the abbreviation. If the abbreviation could denote more than one word, try adjusting the abbreviation.

9. Do not include special characters in the abbreviation.

10. Create an abbreviation that is as short as possible but is still recognizable as a business term.

If the rules do not produce an acceptable abbreviation, adjust the abbreviation accordingly.