

C. Guía de Convenciones

En la definición de cómo será el Data Warehouse, hay que tener muy en cuenta que si algo queda sin especificarse (documentado u oralmente), más pronto o más tarde se caerán en las mismas inconsistencias que se querían eliminar creando el Data Warehouse. El propio William Inmon lo decía en el libro inicial de la serie sobre Data Warehousing (Building the Data Warehouse, [WI-93]), donde hablaba sobre los beneficios de disponer de una forma de Manual de Estándares, donde se describa que es, cómo se carga y cómo se usa el Data Warehouse.

En este proyecto de Data Warehousing, se manejó una guía de convenciones que debían ser respetadas para toda información que ingresara al Data Warehouse.

La guía tiene dos objetivos.

- Primero es el esquema de la documentación de las tablas del Data Warehouse. Para toda tabla que se desee ingresar al Data Warehouse, habrá que llenar los datos de esta guía.
- Además, se detallan convenciones en nombres de tablas, nombres de atributos, tipos de los atributos, o convenciones de medidas.

En la creación de esta guía hay que ser bastante cuidadoso, puesto que los dos extremos (guía pequeña o guía grande) son peligrosos. Una guía pequeña permitirá ingresar al Data Warehouse información sin detallar de donde procede o sin especificar su semántica. Esto hará que no sea de utilidad. Una guía grande, convertirá la documentación del Data Warehouse en un gigantesco libro en el cual no se podrá encontrar lo que se está buscando.

C.1. Convenciones del Sistema.

Tablas del Data Warehouse

- Nombre

Criterio: Nombres empiezan con DW_

Utilidad: Encontrar entre muchas tablas existentes, cuales pertenecen al Data Warehouse y cuales no.

Obligatorio: Si.

Tipos de los Atributos

- Fechas (día, mes y año)

Criterio: De tipo Date

Utilidad: Poder utilizar las características especiales de las fechas (ejemplo: orden entre fechas y funciones específicas)

- Fechas (mes y año)

Criterio: De tipo Date, con el día puesto en el primero del mes.

Utilidad: Igual al anterior.

- Fechas (Períodos de exámen)

Criterio: Dos campos, uno Anio=VARCHAR2(4), otro Periodo=Integer (1=periodo Febrero, 2=Marzo, 3=Julio, 4=Agosto, 5=Diciembre).

Utilidad: Poder encontrar fácilmente tanto el año como el período del exámen.

- Fechas (Períodos de clases)

Criterio: Dos campos, uno Anio=VARCHAR2(4), otro Periodo=Integer (1=1er. semestre, 2= 2º semestre).

Utilidad: Poder encontrar fácilmente tanto el año como el período de clases.

- Campos de caracteres por los que se van a hacer búsquedas

Criterio: De tipo Varchar2. Las letras están en mayúsculas.

Utilidad: Así se sabe al hacer una búsqueda que los nombres están en mayúsculas (no se tendrá que buscar por Taller, TALLER y TaLiEr, sino que solo por TALLER. Notar que si se compara TALLER con mayusculas(atributo) se pierde el uso de índices).

- Campos que son códigos

Criterio: El nombre comienza con COD_

Utilidad: Así se distingue fácilmente cuales son los códigos y cuales campos descriptivos.

C.2. Documentación

Como mínimo se debería especificar lo siguiente para las tablas del Data Warehouse:

- Nombre de la Tabla
- Descripción de la Tabla
- Para cada atributo:
 - Nombre atributo
 - Descripción
 - Tipo
 - ¿Es clave? (sólo o junto a otros atributos)
 - ¿Puede ser NULL?
 - ¿Referencia otra tabla?
 - Como se obtiene (nombre de tabla fuente y atributo, cálculo)