Guía de desarrollo, Anexo 24.01

Proceso de exportación al almacén de datos

Almacén de Datos

GERENCIA DE INFORMÁTICA

JOSEFA VALCÁRCEL, 44
28027-MADRID

Índice General

1 NATURALEZA DE LOS ARCHIVOS DE INTERCAMBIO 3
2 NOMENCLATURA DE LOS ARCHIVOS 3
3 PERIODICIDAD Y NATURALEZA 4
4 ESPECIFICACIÓN DE LA ESTRUCTURA DE LOS REGISTROS 5
5 ESPECIFICACIÓN Y USO DEL FICHERO DE CONTROL 6
6 MECANISMO DE TRANSFERENCIAS DE LOS FICHEROS 7
7 RELACIÓN DE TABLAS Y FICHEROS 8

Índice de Ilustraciones y Tablas

Tabla 1. Ejemplo Periodicidad y Naturaleza 4
Tabla 2. Especificación de la estructura de los registros 6
Tabla 3. Relación de tablas y ficheros 8

1 Naturaleza de los Archivos de Intercambio

La Interfaz de Entrada del *Data Warehouse* está compuesta por ficheros que deben ser generados a partir de la información almacenada en los Sistemas Operacionales existentes en la DGT, siguiendo las especificaciones que se definan.

Los ficheros de intercambio de Almacen de Datos serán ficheros de texto. Dichos ficheros podrán ser tanto de longitud fija como de longitud variable, se empleara el carácter "*pipe*" (|) en el caso de los ficheros de longitud variable.

El delimitador de registro será un avance de línea tipo UNIX (0x0A) o un par retorno de carro/avance de línea tipo Windows (0x0D 0x0A)."

Es necesario, además, que de cara a generar los ficheros que comprenden la interfaz, se tengan en cuenta las siguientes consideraciones:

- Los caracteres del fichero deberán ser ASCII y no podrán contener caracteres de control ni extendidos.
- Cada fichero contendrá un único tipo de registro de información.
- Cada línea del fichero contendrá un único registro de información.

2 Nomenclatura de los Archivos

La nomenclatura a emplear para el nombrado de los archivos tendrá la siguiente estructura:

- Dos o tres caracteres iníciales para indicar la fuente de datos.
- Dos caracteres para indicar el destino de los datos (' wh' *WareHouse*).
- Dos caracteres para indicar la periodicidad del fichero ("**00"** aperiódico, "**01"** diario, "**02"** semanal, "**03"** mensual, "**04**" consolidación)
- Un carácter para indicar la naturaleza de la descarga ("**i"** incremental, "**c"** completa). Se considera descarga completa cuando se proporciona la totalidad de registros contenidos en origen, la descarga incremental se considera como las altas, bajas y modificaciones que se han producido en el periodo de descarga.

- Un carácter '**_**'.
- Hasta cuatro caracteres para identificar al área del modelo al cual pertenecen las tablas en el datawarehouse.
- Dos o tres caracteres numéricos secuenciales identificadores del fichero (01, 02,…).
- Cuando se considere oportuno se podrá añadir un carácter '_' seguido de ocho caracteres (YYYYMMDD) que indicarán el año, mes y día al que pertenecen los datos enviados (fecha de generación del fichero).
- Una extensión '.txt'

Ejemplo:

nowh01i_no001.txt

no = origen: NOSTRA

wh = destino: Warehouse

01 = periodicidad diaria

i = descarga incremental

no = área de negocio en BD: NOSTRA.

001 = identificador secuencial del fichero.

3 Periodicidad y Naturaleza

Como parte de los requerimientos se precisa definir la periodicidad y naturaleza de descarga que tendrán los datos. La periodicidad se refiere al intervalo de tiempo en el cual se descargará cada uno de los ficheros, su naturaleza se refiere al contenido de dicha descarga, ya sea registros incrementales (altas, bajas y modificaciones) ó descarga completa (todas las filas existentes en la tabla origen). Ejemplo:

Nombre	***Periodicidad***	***Ficheros***	***Naturaleza***
NOSTRA	Diaria	nowh01i_no???.txt	Incremental, con ventana de histórico de siete (7) días.
NOSTRA	Diaria	nowh01c_no???.txt	Completa.

Tabla 1. Ejemplo Periodicidad y Naturaleza

4 Especificación de la estructura de los registros

La definición de los registros debe indicar:

- Posición del campo dentro del registro (1,2,3…n).
- Nombre del campo.
- Descripción del campo.
- Si el campo es obligatorio o no.
- El formato exacto del campo en el fichero de exportación:
    1. El formato de las fechas debe ser 'YYYY/MM/DD' o 'YYYY-MM-DD'.
    2. Si la hora es representativa dentro de la fecha en el sistema de origen ('datetime'), el tipo de dato en destino es 'Timestamp(0)' y su formato 'YYYY/MM/DD hh24:mi:ss' o 'YYYY-MM-DD hh24:mi:ss'.
    3. Si se exportan horas (time) el formato será 'hh24:mi:ss'.
    4. Si hay campos numéricos, se debe indicar el formato (escala, precisión, ceros por delante o no, signo ('+/-').
    5. Si hay campos numéricos con decimales, se utilizará el punto '.' como separador de decimales y NO se utilizará separador de grupos (millares, millones etc…)
    6. Si hay valores de dominio, es decir, un campo asociado a un diccionario de datos, todos los códigos y significados deben detallarse (valores admitidos. p. ej.: 'S'=sí, 'N' = no).

Esta información se detalla en una tabla como la del siguiente ***ejemplo***:

Funcionalidad: Contiene las notificaciones relacionadas con el expediente
Nombre de fichero de carga: `nowh01i_no001.txt`

	Nombre Campo	Descripción	Tipo (BD)
1	IDACUSENOTIF	**Campo obligatorio.** Clave primaria de la tabla. **Formato '999999999999'**	DECIMAL(12,0)

	Nombre Campo	Descripción	Tipo (BD)
2	IDNOTIFICACION	**Campo obligatorio.** Identificador de la notificación a la que pertenecen los acuses. **Formato '999999999999'**	DECIMAL(12,0)
3	XTI_TIPOACUSE	**Campo obligatorio.** Indica el tipo de acuses. '**AD**'=Admisión, '**PD**'=Puesta a disposición, '**LE**'=Lectura, '**RE**'=Rechazo, '**CD**'=Caducidad.	CHAR(2)
4	TIM_ACUSE	**Campo obligatorio.** Fecha en la que se genera el acuse. **Formato 'YYYY-MM-DD HH24:MI:SS'**	TIMESTAMP(0)
5	DAT_MOTIVORECHAZO	Motivo de rechazo que el ciudadano puede incluir cuando realiza esta acción.	VARCHAR(512)
6	DAT_NOMBRELECTURA	Nombre que identifica el certificado con el que se realiza la apertura de la notificación.	VARCHAR(256)
7	DAT_NIFLECTURA	NIF que identifica el certificado con el que se realiza la apertura de la notificación.	VARCHAR(20)

Tabla 2. Especificación de la estructura de los registros

5 Especificación y uso del fichero de control

Es necesario incluir la definición y uso del fichero de control en las transferencias. Para asegurar la correcta finalización de la transferencia se utilizará un fichero de control, el cual se enviará al final de la transferencia del grupo de ficheros.

Este fichero se utiliza tanto para indicar que la transferencia ha finalizado como para enviar los datos que faciliten la verificación de la información enviada.

El fichero de control tendrá un nombre relacionado con el grupo de ficheros de la transferencia e incluirá la partícula '**_control'** y la extensión **'.txt'.**

El fichero de control se tratará como un fichero de texto de campos de longitud variable delimitados por pipes ('|'). El delimitador de registro será un avance de línea tipo UNIX (0x0A) o un par retorno de carro/avance de línea tipo Windows (0x0D 0x0A).

Consistirá en un primer registro con el literal 'FECHA:' seguido de la fecha en formato YYYYMMDDhhmmss. Los siguientes registros consistirán en dos campos, el primero de los cuales indica el número de filas y el segundo el nombre del fichero. Finalmente, un registro con el total de los anteriores indicado por el literal 'total'.

Ejemplo:

$ cat nowh01i_control.txt

FECHA: 20090330153944

152549|nowh01i_no001.txt

150000|nowh01i_no002.txt

…

302549|total

6 Mecanismo de transferencias de los ficheros

La transferencia de los ficheros se realiza a unos servidores intermedios (servidores ETL) que son los que reciben procesan y cargan la información en la base de datos del Almacén de Datos a través de sus propias herramientas.

El Almacén de Datos solicita al departamento de Sistemas la creación de los usuarios correspondientes en estos servidores ETL y, una vez confirmada dicha creación, proporciona a la fuente los datos necesarios para la conexión y transferencia.

Los sistemas emisores dejan los ficheros en un directorio-buzón habilitado al efecto en el sistema receptor utilizando el protocolo FTP. La ruta del directorio-buzón será típicamente `/trabajo1/nombre_fuente` y residirá en el servidor designado para las transferencias (servidor ETL).

7 Relación de Tablas y Ficheros

Nº	Tabla	Fichero de Interfaz
1	NOMBRETABLA1	nowh01i_no001.txt
2	NOMBRETABLA2	nowh01i_no002.txt
…	…	…
n	NOMBRETABLAn	nowh01i_no00n.txt

Tabla 3. Relación de tablas y ficheros

<u>NOTA ACLARATORIA</u>: las fuentes de datos una vez integradas en el DWH deben de mantenerse actualizadas en el tiempo, así comunicando el cambio de interfaz o cambio funcional que sufran dichas fuentes a la jefa de proyecto de almacén de datos de DGT