# Desafío Técnico Evaluador NBCH

Consigna:

Se requiere importar en una tabla de base de datos, los feriados locales listados en la planilla Excel adjunta. Teniendo en cuenta que los feriados son indicados por código postal y que cada localidad puede contar con más de un código postal.

De la importación, se deben obtener reportes que indiquen los resultados de éstos, según se indica a continuación.

Tareas:

Debe construirse en la herramienta Kettle de la suite Pentaho un proceso que realice las siguientes tareas:

1. Leer el archivo Excel adjunto (Feriados.xlsx).
2. Obtener la información de la planilla “Feriado Local” y persistirla en una base de datos (Tecnología y estructura de tabla a libre elección). Por cada fila leída deben aplicarse las validaciones incluyentes (indicadas más adelante), y cargar en la tabla las filas que superen todas las validaciones.
3. Generar como salida, un archivo Excel denominado “**reporte.xlsx**” con los mismos datos y formato que el original, pero agregando una columna que contenga el resultado de la carga de cada feriado. Indicando si la misma fue cargada con éxito o fue rechazada por no superar las validaciones (especificando cual o cuales en cada caso). Formato 1º.
4. Generar otro archivo de salida en formato CSV (Utilizando codificación UTF-8) llamada “**total\_provincias.csv**” que contenga la sumatoria de feriados cargados para cada provincia. Ordenado de mayor a menor por la cantidad de feriados. Formato 2º.
5. Generar un tercer archivo de salida (con extensión \*.txt, con registro del tipo FIXED y codificación ANSI) denominado “**total\_localidad.txt**” que contenga la sumatoria de los feriados cargados por localidad. Ordenado por Provincia y Localidad. Formato 3º.

Consideraciones:

* Los datos de la planilla Excel no deben ser modificados manualmente. Debiendo ser procesado de forma íntegra.
* La columna que contenga las fechas de los feriados (punto 2º) debe ser del tipo “date”.
* La identidad de los feriados está dada por las columnas “Fecha” y “Cód. Postal” en conjunto.
* La propiedades de configuración del proceso, como la conexión a la base de datos o la ruta donde se generan los archivos de salida, deben obtenerse de un archivo “.properties” incluido en el proyecto.
* La solución construida debepoder ejecutarse mediante línea de comandos utilizando las herramientas “Pan” y “Kitchen”, permitiendo indicar como parámetro la ruta del archivo a ser importado.
* La solución construida debe ser versionada en un repositorio público (github, gitlab, etc.), indicando el link al mismo mediante correo electrónico.
* La solución debe contener:

1. Código fuente, incluyendo todos los trabajos (.kjb) y transformaciones (.ktr) requeridos para el proceso.
2. DDL de la tabla donde son cargados los feriados.
3. Instrucciones o “Readme” a fin de facilitar la ejecución del proceso desarrollado, indicando comandos para ejecución y cualquier otra información relevante.

* La solución debe ser construida en la siguiente versión de la herramienta Pentaho:

[Pentaho Community Edition Download | Pentaho](https://pentaho.com/pentaho-community-edition/#communityProducts)

Debe descargar la versión indicada como: “pdi-ce-9.4.0.0-343.zip”

Validaciones:

Se deben realizar las siguientes validaciones sobre el archivo de entrada para certificar su integridad y la calidad de los datos provistos.

Las validaciones que se aplican sobre los registros se dividen en:

* Cancelatorias: Aquellas que cancelan el proceso.
* Excluyentes: Aquellas que descartan el registro de la importación.

Cancelatorias: Rechazan el archivo completo y provocan la cancelación del proceso

* El archivo no se localiza en el File System
* La hoja o Sheet donde se encuentran los feriados debe llamarse “Feriado Local”
* La quinta fila del archivo debe contener el encabezado de las columnas, siendo estas “**Fecha**”, “**Cód. Postal”, “P. I.”, “Localidad”, “Provincia”.**

Excluyentes: Descartan el registro de la importación sin abortar el proceso

* La fila se encuentra Vacía (Ignorar)
* Los datos de cada columna deben tener el tipo de dato apropiado y encontrarse dentro del largo máximo soportado (Informar):
  + Fecha: Texto que indica una fecha valida en formato dd/mm/yyyy
  + Cód. Postal: Número entero de hasta 15 dígitos
  + P. I.: Número entero de hasta 5 dígitos
  + Localidad: Texto de hasta 60 caracteres
  + Provincia: Texto de hasta 60 caracteres
* El feriado se encuentra duplicado (Informar).
* El feriado pertenece a una provincia inexistente (Informar).

Puede suceder que existan filas que NO superen múltiples validaciones, en esos casos deben informarse TODAS las validaciones acumuladas que la fila no supere en el campo correspondiente del archivo de reporte.

Formato de los archivos:

Excel de entrada:

|  |  |  |  |
| --- | --- | --- | --- |
| Campo | Tipo Dato | Descripción | Observaciones |
| Fecha | Date | Fecha del Feriado | DD/MM/AAAA  Es importante que el campo "Fecha" se configure en formato texto dentro del Excel. Ya que por defecto se configura en formato de fecha y esto no es soportado ya que el archivo se nos proporciona con este campo en formato texto. |
| Código Postal | Numero Entero | Código Postal |  |
| P.I | Numero Entero | ID interno de Localidad |  |
| Localidad | String | Nombre Localidad |  |
| Provincia | String | Nombre Provincia |  |

Se deben omitir las primeras 5 filas del archivo. Las 4 primeras no corresponden al diseño y la número 5 se corresponde a la cabecera.

Ejemplo:

Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

Formato 1º (Excel de salida):

Mismo formato que el de entrada, pero agregando un campo con el resultado de la carga del feriado.

|  |  |  |  |
| --- | --- | --- | --- |
| Campo | Tipo Dato | Descripción | Observaciones |
| Fecha | Date | Fecha del Feriado | DD/MM/AAAA  Es importante que el campo "Fecha" se configure en formato texto dentro del Excel. Ya que por defecto se configura en formato de fecha y esto no es soportado ya que el archivo se nos proporciona con este campo en formato texto. |
| Código Postal | Numero Entero | Código Postal |  |
| P.I | Numero Entero | ID interno de Localidad |  |
| Localidad | String | Nombre Localidad |  |
| Provincia | String | Nombre Provincia |  |
| Resultado | String | Contiene un texto descriptivo que indica si el feriado pudo cargarse o no y porque. |  |

Se deben omitir las primeras 5 filas del archivo. Las 4 primeras no corresponden al diseño y la número 5 se corresponde a la cabecera.

Ejemplo:

Tabla

Descripción generada automáticamente

Formato 2º (CSV de salida):

|  |  |  |  |
| --- | --- | --- | --- |
| Campo | Tipo Dato | Descripción | Observaciones |
| Provincia | String | Nombre Provincia |  |
| Cantidad | Numero Entero | Cantidad de feriados cargados para esa provincia |  |

Ejemplo:

Texto

Descripción generada automáticamente

Formato 3º (FIXED de salida):

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Campo | Tipo Dato | Inicio | Fin | Largo | Descripción | Observaciones |
| Provincia | String | 1 | 61 | 60 | Nombre Provincia |  |
| Localidad | String | 61 | 121 | 60 | Nombre Localidad |  |
| Cantidad | Numero Entero | 121 | 126 | 5 | Cantidad de feriados cargados para esa localidad |  |

Ejemplo:

