

CIPFP www.fpmislata.com							
Actividad:	Apache NiFi. Ejercicio INE						
Ciclo:	IABD	Modulo:	BDA	Curso:		Agrupación:	1
Alumno/a:						Grupo:	

Actividad

Se quiere realizar un estudio del padrón de una provincia de España, puedes elegir la provincia que desees.

Los datos necesarios se encuentran en las siguientes ubicaciones:

- Datos del padrón por provincias: [INE \(Instituto Nacional de Estadística\)](http://inecensos.inecensos.es/)
 - En este fichero hay filas de datos con totales que debes eliminar. Revisa la estructura y el contenido de los datos
 - NOTA: Descargar los ficheros en formato CSV
- Otros datos ya preparados obtenidos del INE y del CNIG (Centro Nacional de Información Geográfica).
 - Se pueden descargar de <https://doi.org/10.6084/m9.figshare.13505136>.
 - Un fichero con codificación de los rangos de habitantes
 - Un fichero con codificación de municipios

La estructura final de los datos que debes obtener es:

Tabla DatosCenso

- cod_municipio
- municipio
- periodo
- mujeres (Numero de Mujeres)
- hombres (Numero de Hombres)
- nivel_habitantes (Codigo del rango de habitantes)
- capital
- cod_comarca
- comarca
- cod_provincia
- provincia
- cod_comunidad_autonoma
- comunidad_autonoma
- altitud_m
- nivel_altitud
- superficie_ha

CIPFP www.fpmislata.com							
Actividad:	Apache NiFi. Ejercicio INE						
Ciclo:	IABD	Modulo:	BDA	Curso:		Agrupación:	1
Alumno/a:						Grupo:	

Los pasos y acciones que debes realizar son:

1. Revisión y análisis de los datos de entrada:
2. Elabora un plan de trabajo definiendo procesadores y grupos. Dibújalo si te facilita el trabajo.
3. Analiza las comprobaciones previas antes de empezar. Conexiones a BBDD, existencia de ficheros, etc...
4. Ingesta de los datos y transformación
5. Incluye una tabla a modo de log en la que se vayan registrando los procesos que se realizan. La estructura de la tabla debe ser algo así:
 - a. Fecha y Hora (Timestamp)
 - b. Proceso
 - c. Estado (Inicio, Fin)
6. Carga de los datos
 - a. Carga los datos en tablas con la estructura definida al inicio del enunciado de la actividad

Ten en cuenta que:

- ⊄ Deberás crear etiquetas para facilitar la lectura de la ETL
- ⊄ Definir una estructura de procesadores y grupos organizada
- ⊄ Deberás realizar comprobaciones para evitar fallos en el proceso.