Componente Apache Sqoop.

**En esta práctica se trabajará la creación de un grafo relacional de una base de datos, su implementación en SQL y la importación y exportación de datos con MariaDB utilizando Apache Sqoop.**

**Se pide:**

* Leer el enunciado de la siguiente página y realizar un grafo relacional detallando las distintas tablas, atributos, relaciones de tablas así como políticas de actualizado y borrado de datos.
* Realizar un script sql que permita la creación de la base de datos con todas las propiedades descritas en el grafo anterior. **[-createDB]**
* Realizar otro script sql que realice una carga inicial de datos, es decir que pueble todas las tablas de la base de datos. **[-loadData]**
* IMPORTAR todos los datos previos en HDFS desde MariaDB con varias opciones:
  + Sin marcar ninguna opción en específico. **[-import]**
  + Usando filtros. **[-importFilters]**
  + Con formato Avro. Después será necesario ejecutar un comando específico para visualizar los datos en este formato dentro de Hadoop. **[-importAvro]**
  + Con formato Parquet. Después será necesario ejecutar un comando específico para visualizar los datos en este formato dentro de Hadoop. **[-importParquet]**
  + De una manera comprimida utilizando Snappy. **[-importSnappy]**
  + De una manera que luego esos datos puedan ser consultados por Apache Hive. **[-importHive]**
* EXPORTAR los datos a MariaDB desde HDFS. Una de las posibles acciones a realizar para que este proceso funcione es borrar todas las tablas de la base de datos. **[-export]**

**A entregar:**

* Los dos scripts sql anteriormente mencionados.
* Un script bash que ofrezca la posibilidad de hacer todo lo mencionado de manera automatizada. Dependiendo del flag que use el usuario final hará una cosa u otra. Los flags que debe aceptar el script están mencionados en cada opción entre corchetes.

Comentar que el flag para borrar TODAS las tablas de la base de datos es **[-deleteDB]**.

Incluir los comandos de inicialización, copia de archivos jar, instalación de herramientas, entornos virtuales, etc en un flag **[-init]**.

* Una memoria en la que se explique:
  + El grafo relacional desarrollado con sus decisiones de diseño y las políticas de borrado o modificación de datos.
  + Los comandos de sqoop ejecutados para realizar las importaciones y exportaciones que se piden detallando para que sirve cada flag usado.
  + Los comandos del propio Apache Hadoop para leer los datos comptrimidos en formato Avro y Parquet detallando para que sirve cada flag usado.
* Todo lo anterior se debe de comprimir en un fichero (.zip) llamado IAB\_BIU\_Tema\_03\_Practica\_02\_Nombre\_Apellidos (sin tildes, ni ñ ni espacios). Si el grupo lo forma más de una persona poner:
  + IAB\_BIU\_Tema\_03\_Practica\_02\_Nombre1\_Apellidos1\_Nombre2\_Apellidos2

La empresa Distribuciones de Contenidos Digitales Inc. ha lanzado la red de ocio MovieBind, una red orientada al consumo de los productos que ofertan y con funcionalidades sociales para sus miembros.

Para entrar en ella, es necesario registrarse con un nombre de usuario (nickname) que no esté en uso (no se admiten nombres de usuario repetidos), una contraseña (de al menos 8 caracteres) y una dirección de correo electrónico (lógicamente, sólo se permite crear una cuenta de usuario por cada dirección de correo electrónico). De manera voluntaria, los usuarios registrados podrán rellenar los datos de su perfil: nombre de pila, apellido(s), edad, número de móvil, DNI, y fecha de nacimiento. Todos estos datos son opcionales en el perfil (un usuario puede no tener perfil) menos el DNI, que es obligatorio e identificativo (en el perfil), si bien los usuarios que cumplimentan su perfil suelen rellenarlos todos. Los usuarios pueden cambiar los datos de su perfil, pero no pueden cambiar su nombre de usuario si tienen actividad registrada (es decir, si han creado un perfil, si han emitido mensajes o propuestas, si pertenecen a clubes, etc.).

La posesión de una cuenta no implica ningún pago ni obliga a contratar ningún otro servicio, si bien los usuarios pueden, de modo complementario, contratar los servicios de nuestra empresa voluntariamente y así disfrutar de servicios ilimitados de visualización de contenidos digitales. Estos contenidos son películas, que se identifican por título y director (los títulos pueden repetirse, pero siempre con distinto director), y se caracterizan por su reparto (actores y actrices que protagonizan los filmes), su duración, si son en color o B/N, relación de aspecto (ratio), año de estreno, los géneros en que se encuadran, las palabras clave que las describen, calificación por edades, país de producción, idioma de su V.O., su presupuesto, sus ingresos brutos, la información Imdb (link a la página web). Los géneros fílmicos son un conjunto de valores no cerrado y muy poco volátil (p.e., ‘Comedy’, ’Drama’, …).

Para contratar esos servicios, es necesario registrarse (tener una cuenta), rellenar el perfil (al menos, el DNI y el número de teléfono), y elegir uno de los productos que se ofertan (tipo de contrato, cadena de caracteres que se corresponde con un tipo válido en un catálogo). Además de estas informaciones, el contrato también tiene un identificador de contrato, dirección postal, ciudad, código postal, país, fecha de contratación, y fecha de fin (en su caso). El DNI del perfil no puede cambiarse si tiene un contrato vinculado (el resto del perfil sí).

La visualización de contenidos sólo la pueden realizar los usuarios con contrato