EJERCICIOS DE HBase

Partiendo de los datos de la discografía de álbumes de estudio de Pink Floyd

<http://es.wikipedia.org/wiki/Anexo:Discograf%C3%ADa_de_Pink_Floyd>

Año Album Ranking Certificación EEUU Ranking Certificación UK

1967 The Piper at the Gates of Dawn 131 6

1968 A Saucerful of Secrets — 9

1969 Music from the Film More 153 9

1969 Ummagumma (Disco 2) 74 5

1970 Atom Heart Mother 55 1

1971 Meddle 70 3

1972 Obscured by Clouds 46 6

1973 The Dark Side of the Moon 1 1

1975 Wish You Were Here 1 1

1977 Animals 3 2

1979 The Wall 1 3

1983 The Final Cut 6 1

1987 A Momentary Lapse of Reason 3 3

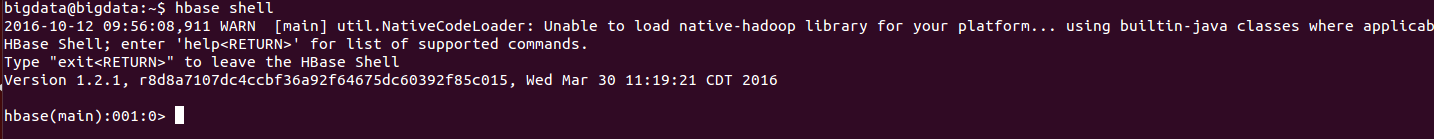
1994 The Division Bell 1 1

2014 The Endless River 3 1

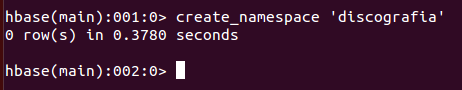
Resolver las siguientes preguntas:

1. Crear un namespace llamado discografía

Lo primero que vamos a hacer es arracar la Shell de hbase:

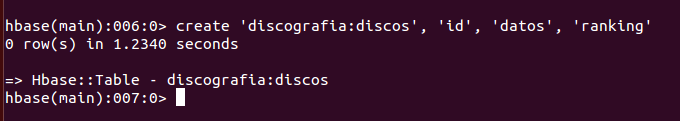


A continuación creamos el namespace “discografía”:



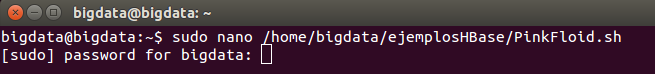
1. Crear una tabla llamada discos que contenga sólo las columnas: grupo, nombre\_disco, anio, ranking (a su vez tendrá identificadores para EEUU y UK)

Creamos la tabla discos con tres family\_column (id, datos y ranking). La familycolumn “id” va a tener el identificador del registro, formado por número y el año del disco (el año no es un valor único). Dentro de datos vamos a guardar las columnas con los datos del disco (grupo, nombre\_disco, anio) y en ranking vamos a guardar el ranking de EEUU y de UK en dos columnas distintas:

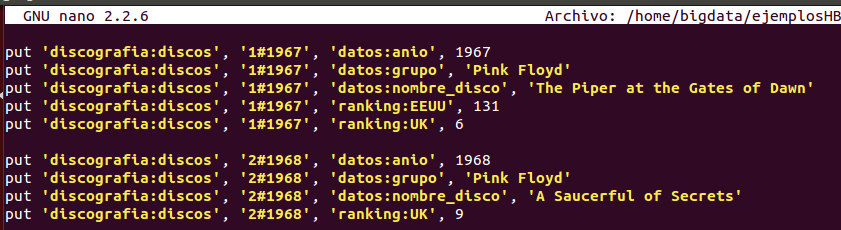


1. Cargar los datos en HBase (se pueden emplear los comandos put o una importación desde PIG). Nota: algunos años hay más de un disco por lo que la clave será el año y un secuencial, y no se disponen de todos los datos de rankings por lo que esos datos nulos no será necesario cargarlos

Para cargar los datos en HBase, vamos a utilizar el comando put. Vamos a crear todos los comando en un script que posteriormente ejecutaremos. En una nueva terminal de Ubuntu creamos el fichero .sh:

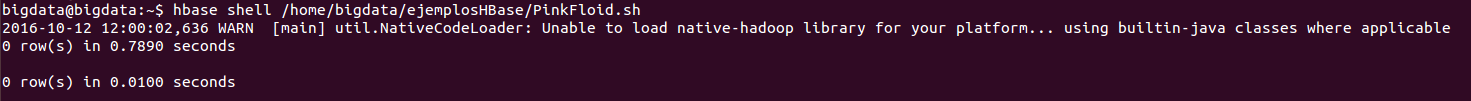


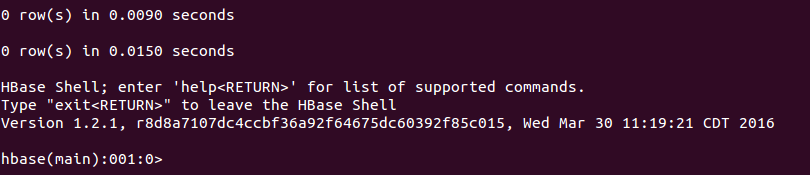
Rellenamos el .sh con las sentencias:





Desde la consola de Ubuntu cargamos el script en HBase

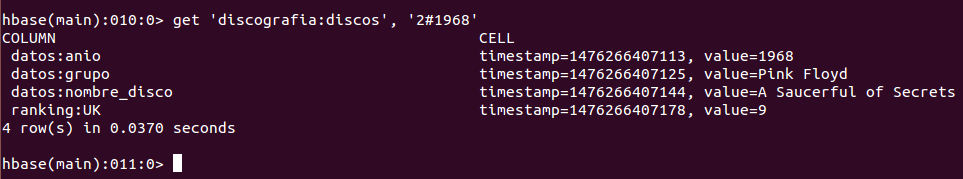




Comprobamos que se han cargado los datos:



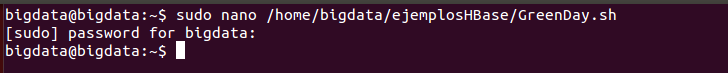
1. Escribir la instrucción y el resultado de consultar el álbum del año 1968



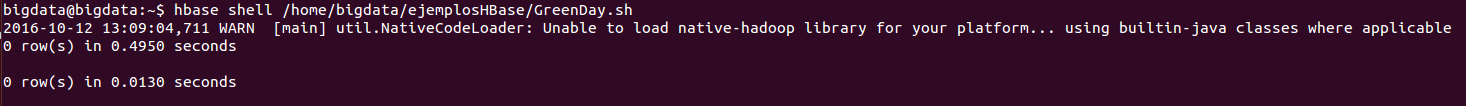
5. Añadir la información de los álbumes de estudio de GreenDay

<https://en.wikipedia.org/wiki/Green_Day_discography>

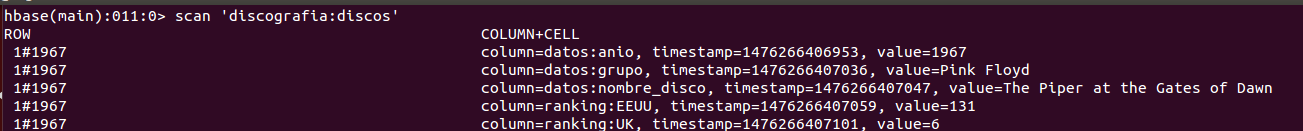
Volvemos a crear un fichero con las instrucciones y lo cargamos en HBase, como hemos hecho anteriormente

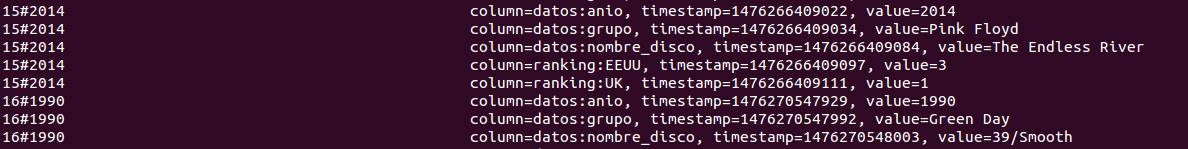






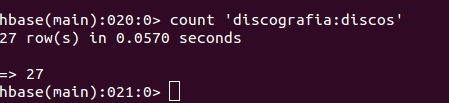
Comprobamos:



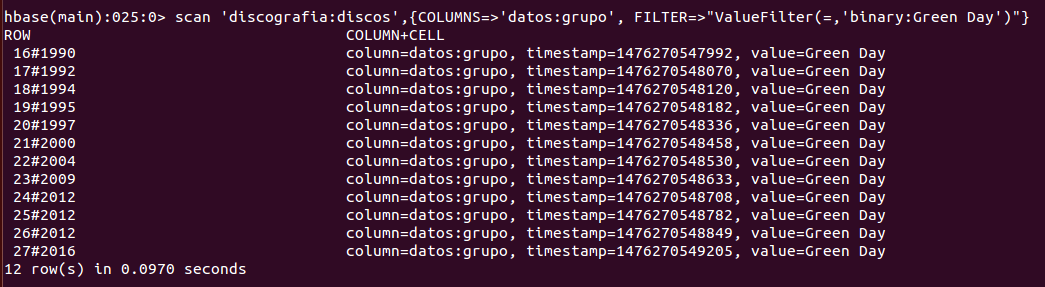


1. Ejecuta varias instrucciones (describiéndolas) para analizar datos anteriores

Número de registros en la tabla:



Registros de Green Day



Registros del año 1994:

