



CONTENIDO

APARTADO A

Práctica con PIG

Mediante un script de PIG, encontrar las cinco películas (código, título y número de votos) más votadas (recuento de votos, no media).

1. Describe informalmente los pasos que darás para llegar a la solución, por ejemplo:
 - a. Extraer todos los registros de u.data.
 - b. Extraer el identificador y título de u.item.
 - c. Convertir.....
2. Implementa en PIG el script necesario para obtener la información deseada.
3. Muestra la salida del mismo por pantalla y almacena también su resultado en un archivo en HDFS.

CONTENIDO

APARTADO B

Mediante un script de PIG, encontrar las diez películas mejor valoradas (código, título y media de puntuación) por los usuarios (ahora sí, media de todos los votos recibidos).

1. Describe informalmente los pasos que darás para llegar a la solución.
2. Implementa en PIG el script necesario para obtener la información deseada.
3. Muestra la salida del mismo por pantalla y almacena también su resultado en un archivo en HDFS.

CONTENIDO

APARTADO C

Práctica con PIG

Mediante un script de PIG, encontrar las cinco películas más antiguas con una valoración media por encima de 4 puntos.

1. Implementa en PIG el script necesario para obtener la información deseada.
2. Muestra la salida del mismo por pantalla y almacena también su resultado en un archivo en HDFS.

CONTENIDO

APARTADO D

Práctica con PIG



Mediante un script de PIG, encontrar la película mejor valorada por cada una de las ocupaciones (student, writer, doctor, etc.)

1. Implementa en PIG el script necesario para obtener la información deseada.
2. Muestra la salida del mismo por pantalla y almacena también su resultado en un archivo en HDFS.

CONTENIDO

APARTADO E

Mediante un script de PIG, encontrar el promedio de valoraciones por décadas, guardarlo en HDFS como un archivo csv. Posteriormente lo descargaremos a nuestro ordenador y con EXCEL hacer un gráfico de barras con los datos del ficchero

1. Implementa en PIG el script necesario para obtener la información deseada.
 - Analizar el archivo u.item (información de películas) y extraer la fecha de estreno utilizando el operador SUBSTRING.
 - Agrupar las películas por década (por ejemplo: 1970, 1980, 1990, etc.).
 - Calcular el promedio de rating por década.
 - Guardar los resultados en CSV para graficar posteriormente.
2. Abrir el archive en Excel y generar un gráfico de barras con los datos.