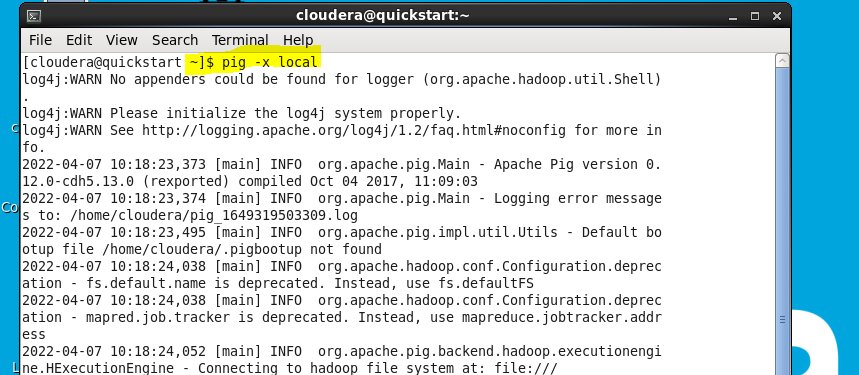
PIG

PIG es una plataforma de alto nivel para crear programas MapReduce utilizados en hadoop

Pig Latin es un lenguaje de flujo de datos

Ejercicios PIG

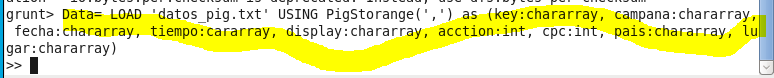
1.Copiar en local file sistema de la MV el fichero datos\_pig.txt en la ruta /home/cloudera/ejercicios/pigy abrir el fichero para revisar su contenido.



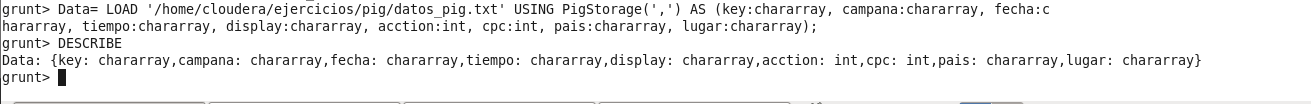
2.Arranca el Shell de Pig en modo locala.“pig -x local”

pig -x local

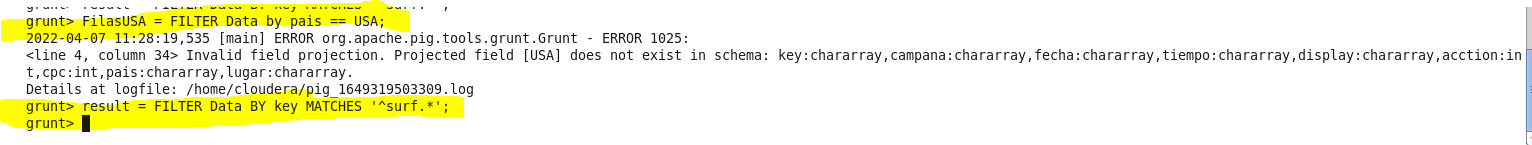
3.Carga los datos en pigen una variable llamada “data”. Los nombres de las columnas deben ser (key,campana, fecha, tiempo, display, accion, cpc, pais,lugar). Los tipos de las columnas deben ser chararray excepto acction y cpcque son int. A.



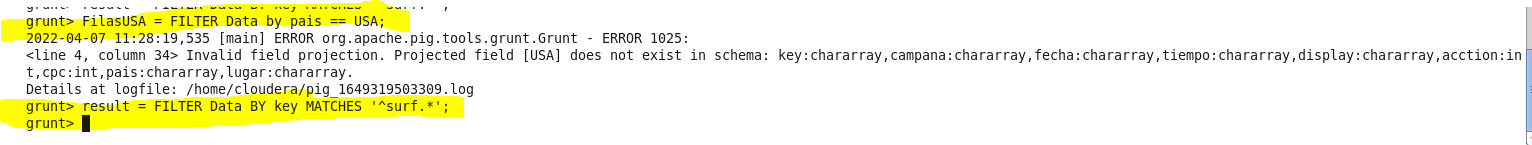
4.Usa el comando DESCRIBE para ver elesquema de la variable “data”



5.Selecciona las filas de “data” que provengan de USAa.



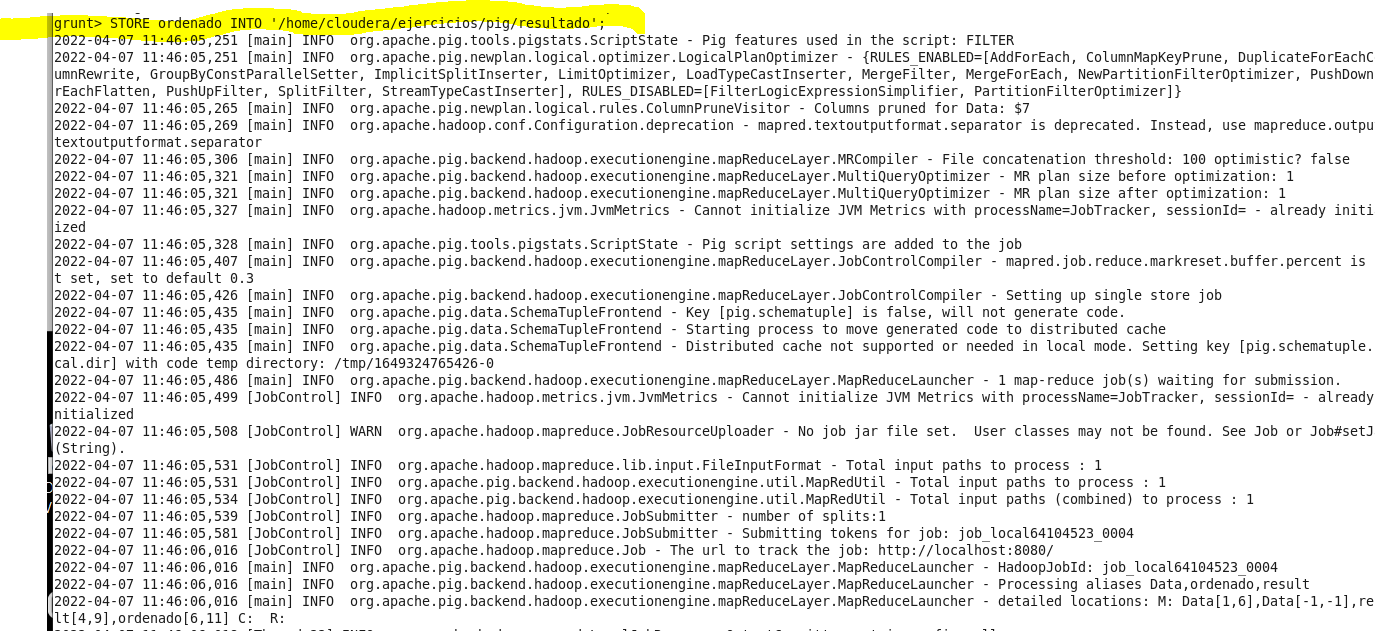
6.Listar los datosque contengan en su key el sufijo surf:



7.Crear una variable llamada “ordenado” que contenga las columnas de data en el siguiente orden: (campaña, fecha, tiempo, key, display, lugar, accion, cpc)a.



8.Guarda el contenido de la variable “ordenado” en una carpeta en el local file system de tu MV llamada resultado en la ruta /home/cloudera/ejercicios/piga.



9.Comprobar el contenido de la carpeta.