

Inspección de datos

Herramientas para tocar los datos
en conjuntos completos en la memoria.

- fdisk eficiente
- Cut archivo.txt : vuela toca
- less Archivo.txt : programador organiza
- Esperio : Avanza program
- grep patron : Busca hora debida
- ? patron : Busca hora otros
- Slicing
- head -n 5 archivo.csv :
un encabezado y primeros 5 filas
- tail -n 5 archivo.txt : monitorea un tiempo real de log

2) tuber y redirección

manipulación de entradas / salidas
estándar

- operador de redirección

• > : sobrescribe archivo con
stdout

• >> : agrega stdout al final
del archivo

• >>> : redirige todo errores

& > : reduce los a (stdout +
stderr)

• < : alimenta en comando con un
archivo

• Pipes (|)

- Combinar las salidas de un comando
con la entrada del siguiente

• Ejemplo dato:

Cat new-det.csv | grep "2024" | head
(Filtra por año y muestra muestra

③ wrangling y transformación

Limpieza y estructuración de datos
Crushers

• Filtrado vertical

- isla por archivos delimitados
(CSV, TSV)

- Cut -d -f 1,3 data.csv : Extrae
columnas 1 y 3 usando
Comas como delimitadores

• Ordenamientos - Sort

Sort -n : orden numérico (Líos esto
10 va antes que 2)

Sort -r : orden inverso

Sort -t / -K2 : ordena por la
segunda columna de
en CSV

Estructuras : -WC

WC -1 : Conteo de registros (line count)
La métrica más rápida para

saber el volumen

1

Filtros avanzados grep y regex

el motor de búsqueda de patrones

- Comando grep clave

- grep "error" log.txt

- grep -v "null": invertir selección

- grep -i : ignora mayúsculas / minúsculas

grep -c : Cuenta ocurrencias (mos rápido que grep | wc -l)

grep E : habilita regex extendido
(necesario para |, +, ?)

Cheat Sheet de regex

- anclaje

- ^ Dato : empieza con "Dato"

- CSV \$: termina con "CSV" /

- Comprobacion

- Cualquier caracter unico

- * : 0 o mas veces

- + : 1 o mas veces (Solo con -E)

- ? : 0 o 1 vez

- Clases tipos

- [0 - 9] : Cualquier digito

- [^0 - 9] : Cualquier cosa distinta de numeros en digitos

- gray / gray : operador "OR" logico
(solo con \in)

