

Introducción a Unidad 2



Daniel Millán & Nicolás Muzi

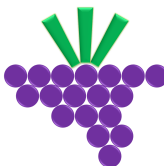
Marzo 2022, San Rafael, Argentina



UNCUYO
UNIVERSIDAD
NACIONAL DE CUYO



FACULTAD DE
**CIENCIAS APLICADAS
A LA INDUSTRIA**



Curso basado en uno propuesto por *William Knottenbelt*, UK, 2001



Este es un **anexo** a la Unidad 2 que trata sobre **awk**:

- **awk** (Aho, Weinberger and Kernigan)
- **awk**, es un programa de Unix que interpreta programas escritos en el lenguaje de programación AWK.



13. Procesamiento avanzado de archivos de texto

awk

- AWK es un lenguaje de programación diseñado para procesar datos basados en texto, ya sean ficheros o flujos de datos.
- Debido a su densa notación, frecuentemente es usado para escribir programas de una línea.
- AWK fue una de las primeras herramientas de Unix (versión 3) y permite añadir funcionalidad a las tuberías de Unix.
- El lenguaje AWK es estándar en casi todo OS tipo Unix.
- AWK es mencionado en las Single UNIX Specification (especificaciones básicas de unix) como una de las utilidades necesarias de todo OS Unix.



13. Procesamiento avanzado de archivos de texto

awk

- **awk** es útil para manipular archivos que contienen columnas de datos de la misma longitud.
- Como en **sed**, podemos pasar a **awk** declaraciones directamente por línea de órdenes en una terminal, o podemos escribir los comandos en un archivo de secuencia de órdenes (*script*) también se utiliza el término guión de órdenes.
- Q: ¿Cuándo ud piensa que esto puede resultar ser útil?



13. Procesamiento avanzado de archivos de texto

awk

- Por ejemplo, si tenemos un archivo “nbateams.txt” con los resultados para cada equipo de baloncesto de la NBA durante la temporada 2015, que contiene información para cada equipo de:
 - los partidos jugados,
 - los goles realizados,
 - los intentos, etc (ver “README_nbateams.txt”):

Team	GP	FGM	FGA	2FG	2FGM	2FGA	3FG	3PM	3PA
Boston_Celtics	74	39.1	89.3	70.7 %	30.4	63.1	29.3 %	8.7	26.2
Golden_State_Warriors	74	42.9	88.2	64.6 %	30.0	57.0	35.4 %	12.9	31.2
Chicago_Bulls	75	38.6	87.9	75.9 %	30.8	66.7	24.1 %	7.8	21.2
...
Minnesota_Timberwolves	74	37.4	81.0	80.1 %	32.0	64.9	19.9 %	5.4	16.1
Utah_Jazz	75	36.0	80.6	70.9 %	27.6	57.1	29.1 %	8.4	23.5



13. Procesamiento avanzado de archivos de texto

awk

- Por ejemplo si deseamos imprimir en la salida estándar sólo la primera y segunda columna:

```
$ awk '{print $1 $2}' nbateams.txt
```

Team	GP
Boston_Celtics	74
Golden_State_Warriors	74
Chicago_Bulls	75
...	...
Minnesota_Timberwolves	74
Utah_Jazz	75

- Aquí **\$n** representa el enésimo campo o columna de cada línea en el archivo de datos. Si se escribe **\$0** se hace referencia a toda la línea.



13. Procesamiento avanzado de archivos de texto

awk

- Además es posible escribir un guión de órdenes “nbateams.awk” para calcular el *Field Goal Percentage*

```
$ cat > nbateams.awk
BEGIN { teams = -1 }
{ teams++; FGM=$3; FGA=$4}
{ if (teams>0) print FGM/FGA*100 }
END { print "The number of NBA teams is "teams }
(Ctrl-d)
$ awk -f nbateams.awk nbateams.txt
43.785
...
44.665
The number of NBA teams is 30
```

La cláusula **BEGIN** se ejecuta una vez al comienzo del *script*, la cláusula principal una vez por cada línea, la cláusula que calcula FGP sólo si el número de “teams” es mayor que 0 y la cláusula **END** de finalización una vez al final de la secuencia de órdenes.