

Sistema Unix, terminal y bash

Capacitación de Analítica Avanzada



Agenda

1

Introducción y
objetivos de la sesión

2

Estructura de unix

3

Comandos básicos de
bash

Recordemos cómo se integra esta sesión en el programa

● Intervenciones teóricas

● Intervenciones prácticas en modo demo/tutorial

#	Módulo	Área	Tipo	Sesión
2.1	Fundamentos para el análisis de datos	Coding	●	Manejo de terminal con <i>bash, shell</i>
2.2	Fundamentos para el análisis de datos	Coding	●	Ambiente de desarrollo: <i>conda, pip, brew, apt get install</i>
2.3	Fundamentos para el análisis de datos	Coding	●	Principios de Python
2.4	Fundamentos para el análisis de datos	Coding	●	Principios de Python
2.5	Fundamentos para el análisis de datos	Data engineering	●	Pandas: <i>toolkit</i> para el procesamiento de datos

Mini-proyecto

Introducción y objetivos de la sesión



Introducción

Conocer la estructura de unix y los comandos básicos de bash permite un entendimiento profundo de las tecnologías que usamos en el análisis de datos, además que son importantes alternativas cuando necesitamos hacer tareas particulares de, por ejemplo exploración de datos o automatización de tareas.



Objetivos

- Entender la estructura sistema unix y la carpeta root del sistema
- Conocer el funcionamiento de la terminal y los comandos básicos del lenguaje bash
- Interactuar con archivos usando expresiones regulares

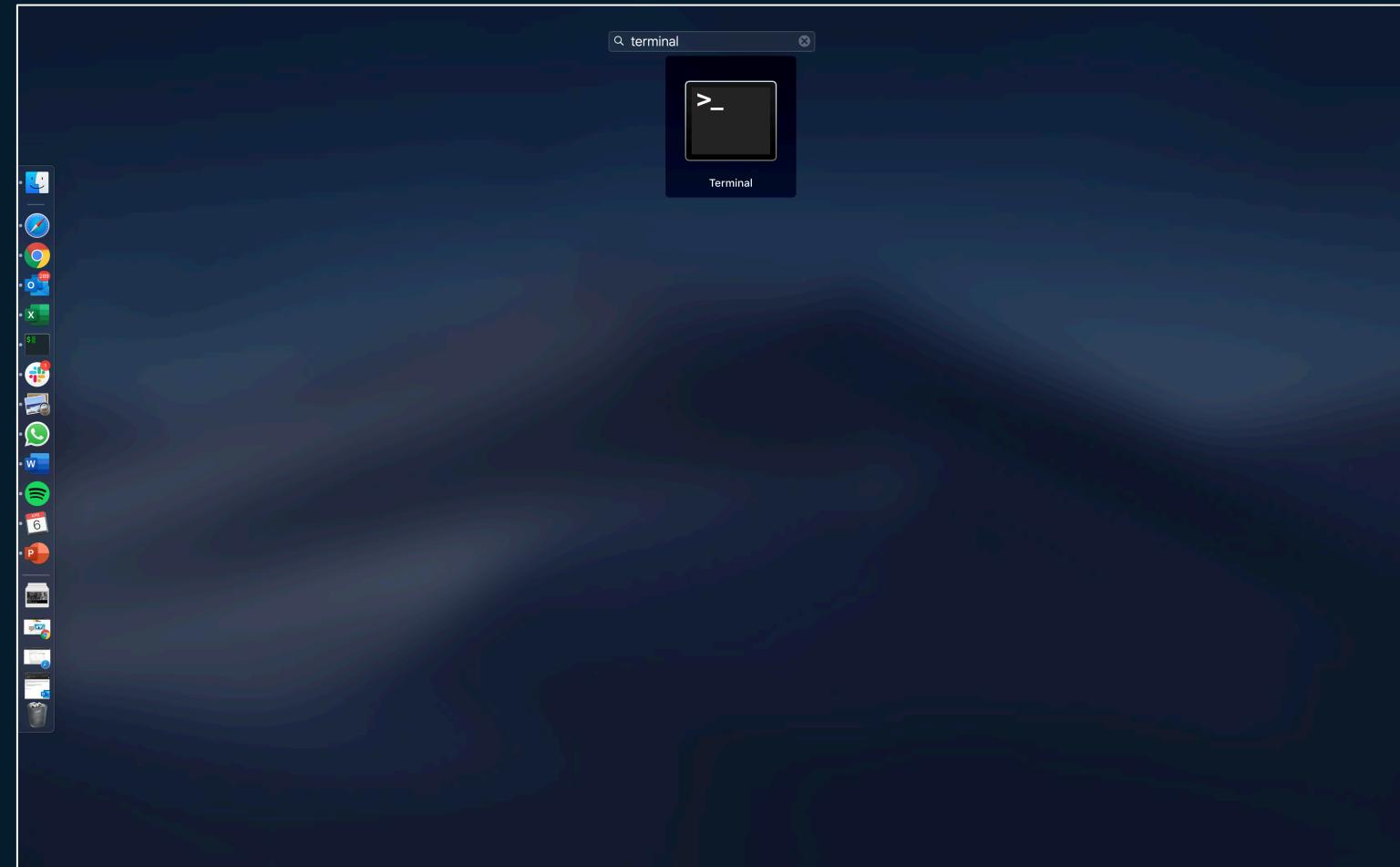
Estructura de *Unix*

La terminal es nuestra plataforma de acceso a este mundo

Tomando como referencia el sistema Mac OS, buscamos en nuestras herramientas la aplicación *terminal*

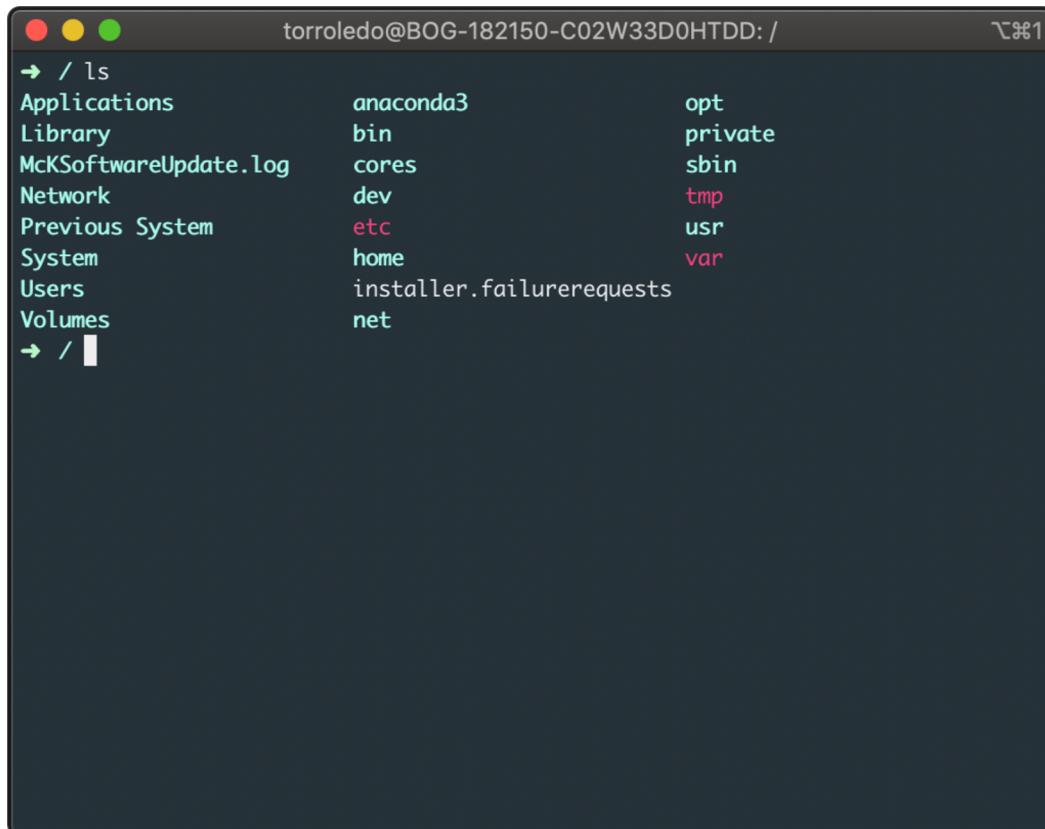
Esta aplicación nativa es la puerta de entrada para comunicarnos con nuestro sistema operativo. Allí podemos:

- Manejar y explorar nuestros archivos
- Ejecutar programas
- Instalar nuevos programas a través de los gestores



Estructura del sistema unix

Conocer la estructura del sistema permite que entendamos como personalizarlo



```
torroledo@BOG-182150-C02W33D0HTDD: / → /ls
Applications      anaconda3      opt
Library          bin            private
McKSoftwareUpdate.log cores          sbin
Network          dev            tmp
Previous System   etc            usr
System           home           var
Users            installer.failurerequests
Volumes          net
```

/ - directorio raíz

Similar a C:\ de windows

/bin - binarios del sistema

Ubicación para todos los programas del sistema

/boot – archivos de boot

Archivos necesarios para iniciar el Sistema

/dev – dispositivos

Linux representa dispositivos como archivos

/home – carpeta de inicio

Folder para los archivos de cada usuario

/lib – librerías

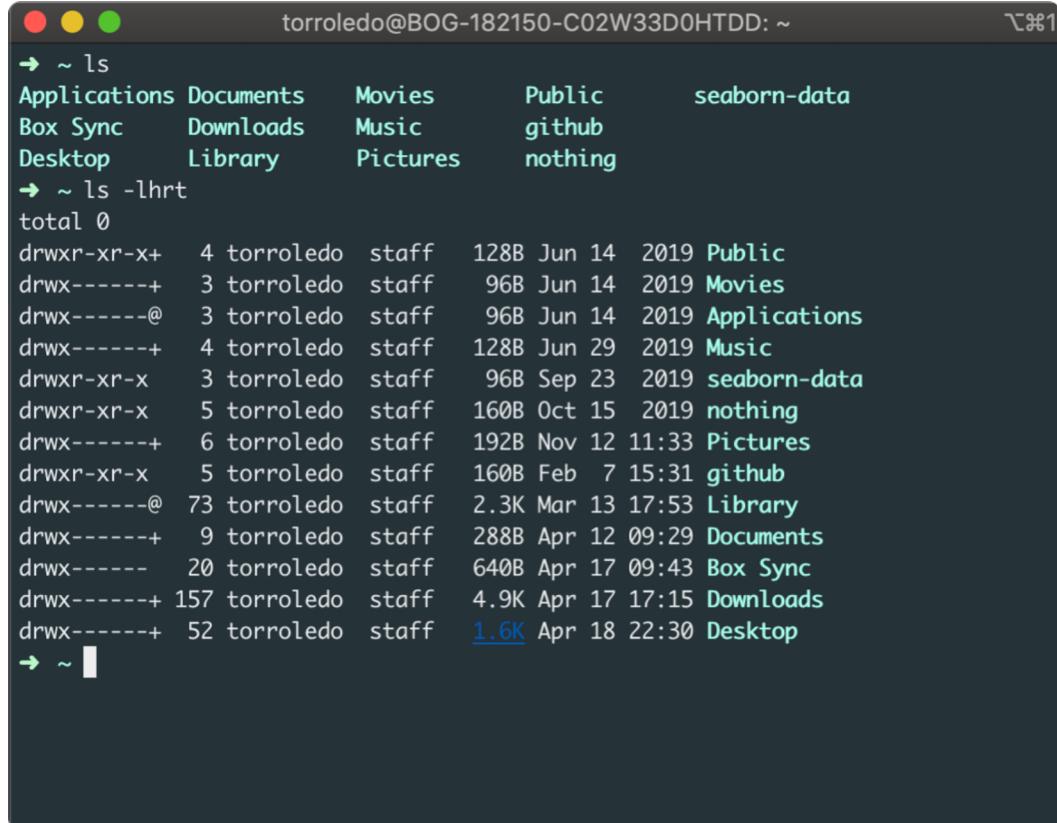
librerías usadas por programas en /bin

/usr

Archivos de binarios para los usuarios

Comandos básico de bash

Comandos básicos de bash para interactuar con folders



A screenshot of a macOS terminal window titled "torroledo@BOG-182150-C02W33D0HTDD: ~". The window shows the output of the "ls" command. It lists several folders: Applications, Documents, Movies, Public, seaborn-data, Box Sync, Downloads, Music, github, Desktop, Library, Pictures, and nothing. Below this, the command "ls -lhrt" is run, showing detailed file information for each folder, including permissions, owner, group, size, date modified, and name. The "Desktop" folder is highlighted in blue.

```
torroledo@BOG-182150-C02W33D0HTDD: ~
→ ~ ls
Applications Documents Movies Public      seaborn-data
Box Sync     Downloads Music   github
Desktop     Library  Pictures nothing

→ ~ ls -lhrt
total 0
drwxr-xr-x+  4 torroledo  staff  128B Jun 14  2019 Public
drwx-----+  3 torroledo  staff   96B Jun 14  2019 Movies
drwx-----@  3 torroledo  staff   96B Jun 14  2019 Applications
drwx-----+  4 torroledo  staff  128B Jun 29  2019 Music
drwxr-xr-x  3 torroledo  staff   96B Sep 23  2019 seaborn-data
drwxr-xr-x  5 torroledo  staff  160B Oct 15  2019 nothing
drwx-----+  6 torroledo  staff  192B Nov 12  11:33 Pictures
drwxr-xr-x  5 torroledo  staff  160B Feb  7  15:31 github
drwx-----@  73 torroledo  staff  2.3K Mar 13  17:53 Library
drwx-----+  9 torroledo  staff  288B Apr 12  09:29 Documents
drwx-----  20 torroledo  staff  640B Apr 17  09:43 Box Sync
drwx-----+ 157 torroledo  staff  4.9K Apr 17  17:15 Downloads
drwx-----+ 52 torroledo  staff  1.6K Apr 18  22:30 Desktop
→ ~
```

whoami

Comando para mostrar nombre de usuario

pwd

Mostrar *path* al directorio actual

ls [-lhrt]

Listar archivos en el folder

cd [folder]

Ir a folder seleccionado

cd ..

Ir al folder anterior

mkdir

Crear folder

rm

Borrar archivos

Exploración de archivos

```
torroledo@x86_64-apple-darwin13: ~/Box Sync/Capability building DE... ✘
(base) ➔ modulo-2 git:(master) ✘ head 2.4-supermarket_sales.txt
Invoice ID|City|Customer type|Gender|Product line|Unit price|Quantity|Tax %|Date|Time|Payment|Rating
750-67-8428|Yangon|Member|Female|Health and beauty|74,69|7|26.1415|1/5/2019|13:08|Ewallet|9.1
226-31-3081|Naypyitaw|Normal|Female|Electronic accessories|15.28|5|3.82|3/8/2019|10:29|Cash|9.6
631-41-3108|Yangon|Normal|Male|Home and lifestyle|46.33|7|16.2155|3/3/2019|13:23|Credit card|7.4
123-19-1176|Yangon|Member|Male|Health and beauty|58.22|8|23.288|1/27/2019|20:33|Ewallet|8.4
373-73-7910|Yangon|Normal|Male|Sports and travel|86.31|7|30.2085|2/8/2019|10:37|Ewallet|15.3
699-14-3026|Naypyitaw|Normal|Male|Electronic accessories|85.39|7|29.8865|3/25/2019|18:30|Ewallet|14.1
355-53-5943|Yangon|Member|Female|Electronic accessories|68.84|6|20.651999999999999|97|2/25/2019|14:36|Ewallet|15.8
315-22-5665|Naypyitaw|Normal|Female|Home and lifestyle|73.56|10|36.78|2/24/2019|11:38|Ewallet|8.0
665-32-9167|Yangon|Member|Female|Health and beauty|36.26|2|3.6260000000000003|1/10/2019|17:15|Credit card|7.2
(base) ➔ modulo-2 git:(master) ✘
```

touch

Crear un archivo vacío

head

Muestra primera parte del contenido de un archivo

tail

Muestra primera parte del contenido de un archivo

more

Explora los archivos completamente

wc

Cuenta palabras en un archivo

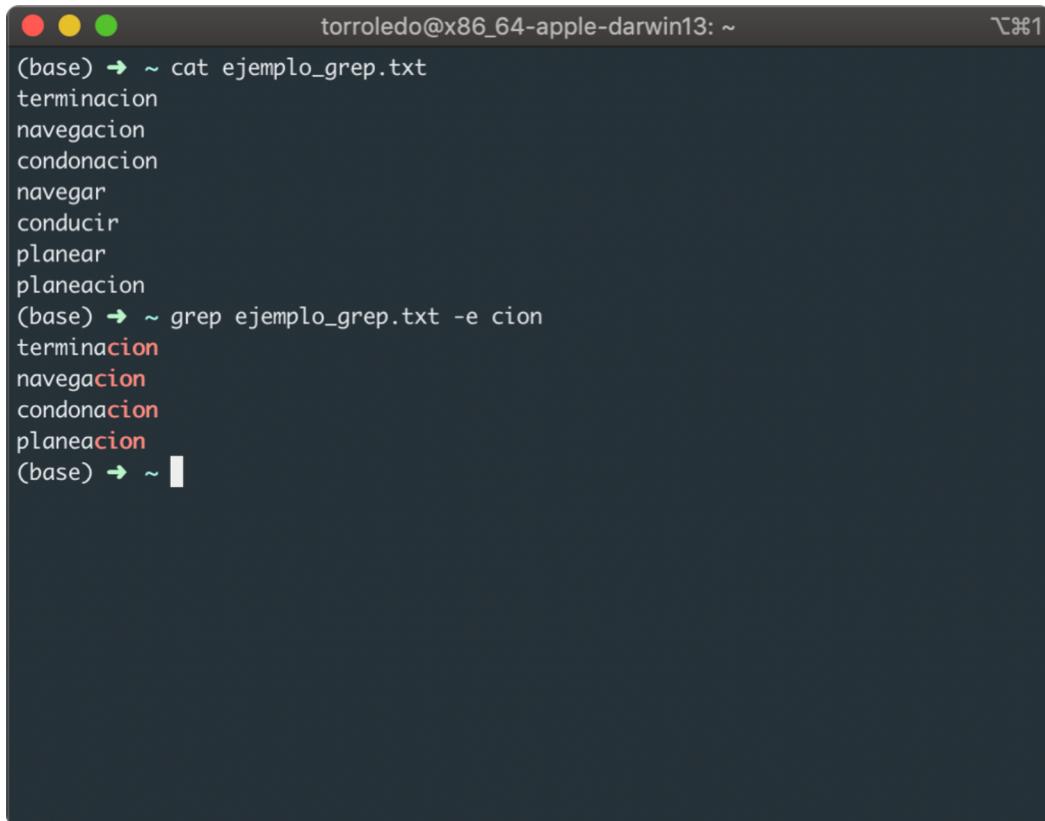
tar

Herramienta para descomprimir archivos

man

Acceso al manual de cada comando

Expresiones regulares



A screenshot of a macOS terminal window titled "torroledo@x86_64-apple-darwin13: ~". The window contains two command-line sessions:

```
(base) ➔ ~ cat ejemplo_grep.txt
terminacion
navegacion
condonacion
navegar
conducir
planear
planeacion
(base) ➔ ~ grep ejemplo_grep.txt -e cion
terminacion
navegacion
condonacion
planeacion
(base) ➔ ~
```

Comando grep

Comando para la ejecución de expresiones regulares o búsqueda de patrones en la data

Opciones:

-e

Búsqueda de patrón

-c

Conteo de coincidencias

-r

Búsqueda recursiva del patrón en todos los folders

-v

Las líneas seleccionadas son las que no cumplen el patrón

¿Cómo podemos descargar archivos de la red?

```
man wget
WGET(1)          GNU Wget          WGET(1)

NAME
Wget - The non-interactive network downloader.

SYNOPSIS
wget [option]... [URL]...

DESCRIPTION
GNU Wget is a free utility for non-interactive download of files from
the Web. It supports HTTP, HTTPS, and FTP protocols, as well as
retrieval through HTTP proxies.

Wget is non-interactive, meaning that it can work in the background,
while the user is not logged on. This allows you to start a retrieval
and disconnect from the system, letting Wget finish the work. By
contrast, most of the Web browsers require constant user's presence,
which can be a great hindrance when transferring a lot of data.

Wget can follow links in HTML, XHTML, and CSS pages, to create local
versions of remote web sites, fully recreating the directory structure
of the original site. This is sometimes referred to as "recursive"
:
```



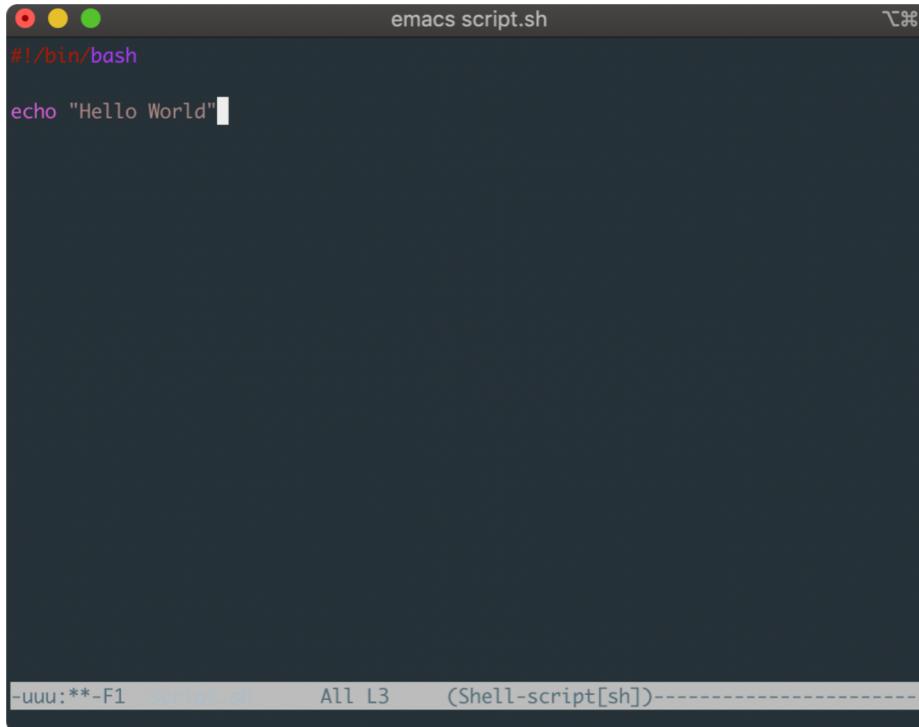
Función wget

Comando para descargar contenido de la red. Es útil cuando queremos automatizar un proceso que usa información disponible en línea

Alternativamente tenemos **curl** que permite transferir archivos hacia o desde un servidor (descargar)

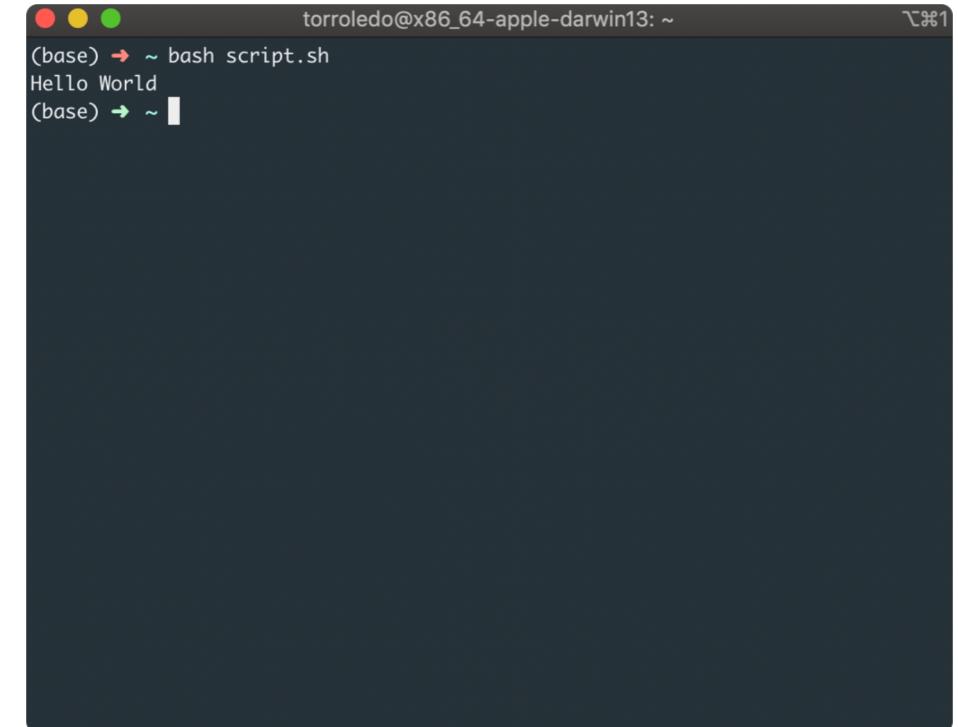
Crear scripts de ejecución en bash

Todas estas acciones y más se pueden convertir en un programa



```
#!/bin/bash
echo "Hello World"
```

Se debe crear el script de código con las instrucción **#!/bin/bash** para definir que tipo de script queremos correr y cuál es el interprete para ejecutarlo



```
torroledo@x86_64-apple-darwin13: ~
(base) ➔ ~ bash script.sh
Hello World
(base) ➔ ~
```

Corremos el programa usando el comando **bash**

Creación de contenido

Autores



Iván Torroledo

Data science fellow

Ivan_Torroledo@mckinsey.com

Bogotá



Simon Tamayo

Data science specialist

Simon_Tamayo@mckinsey.com

Bogotá