



**Programa Ingeniería de Sistemas**

# Herramientas de Software Libre para la obtención de datos Big Data

---

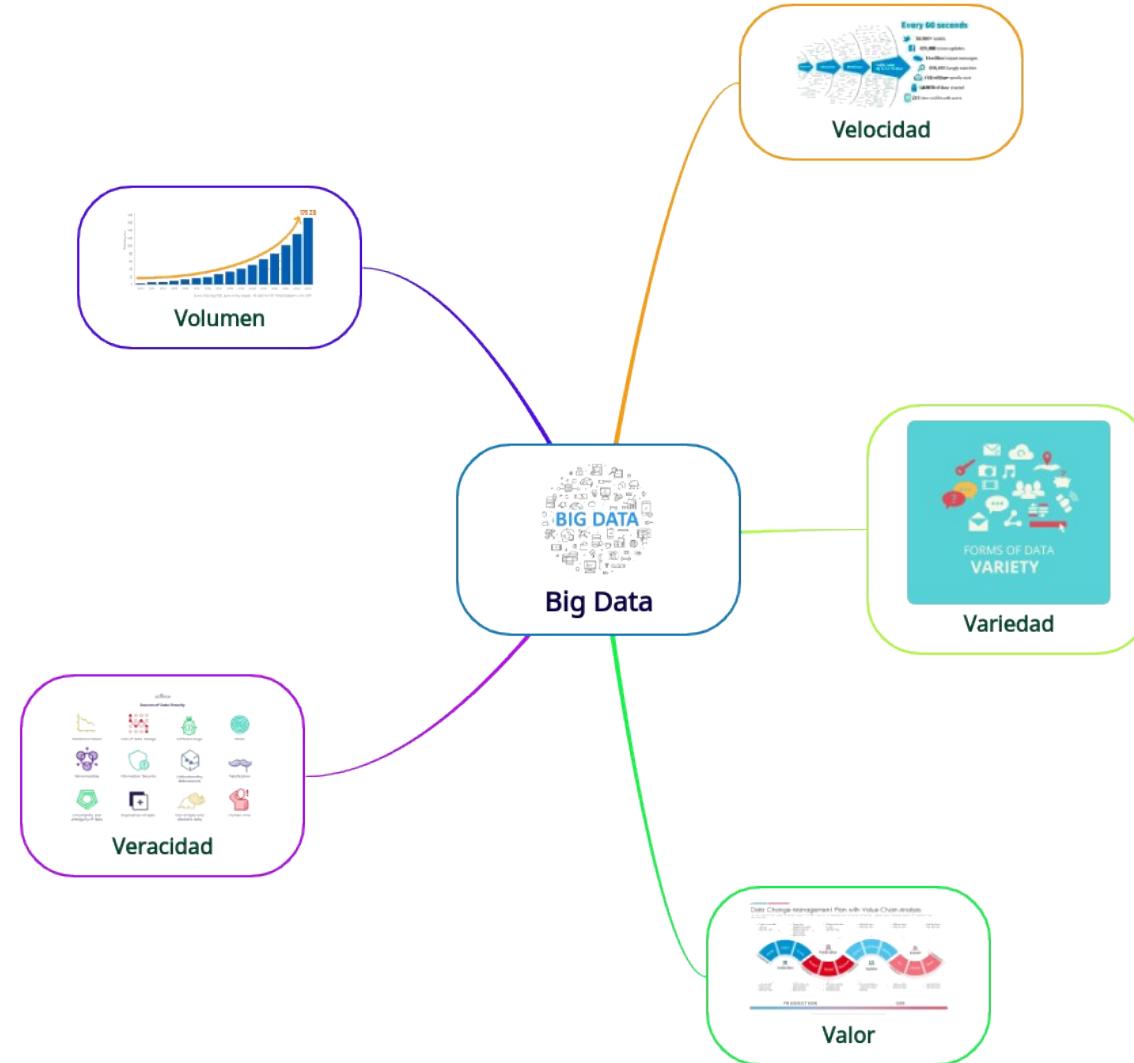
Juan Francisco Mendoza Moreno, PhD



**BIG DATA**

A small black and white Tux the Penguin logo is positioned on top of the letter "I" in the word "BIG".

# ¿Qué es Big Data?

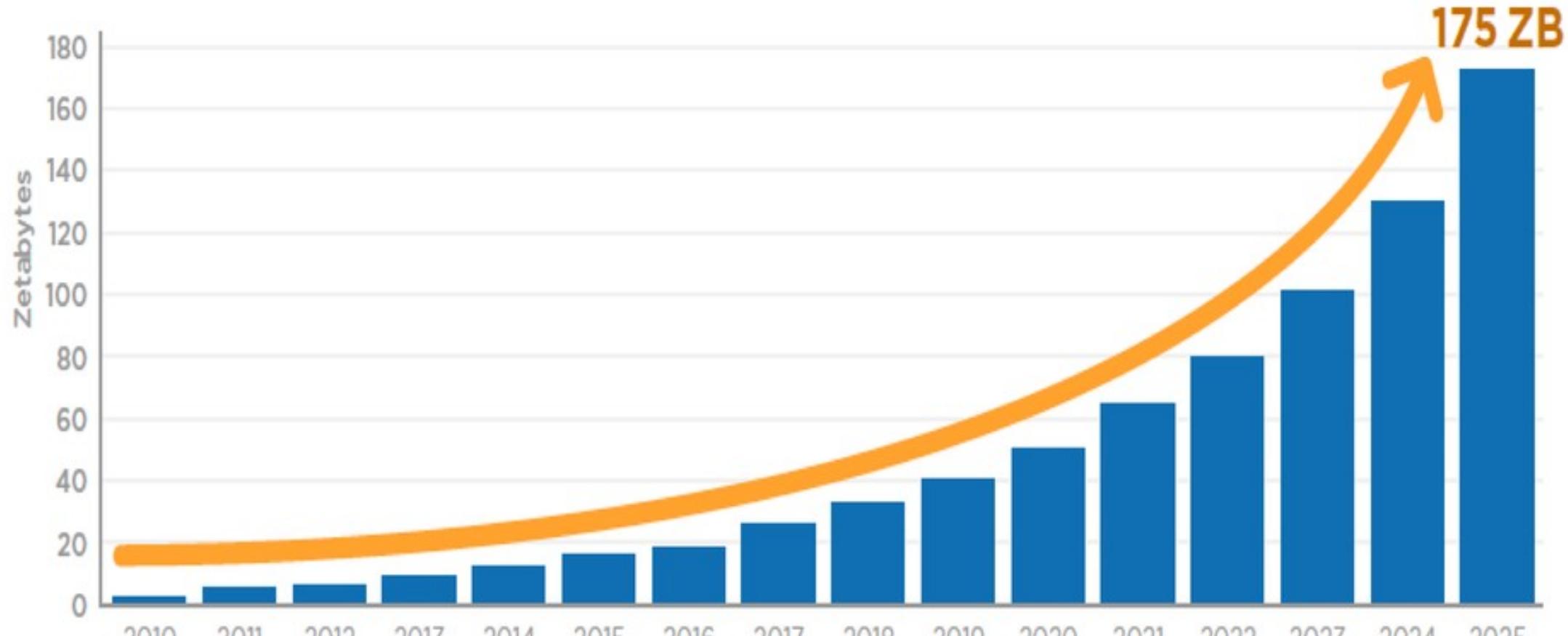


# Datos: El petróleo de nuestra era



Fuente Imagen: [www.OleoShop.com](http://www.OleoShop.com)

# Volumen

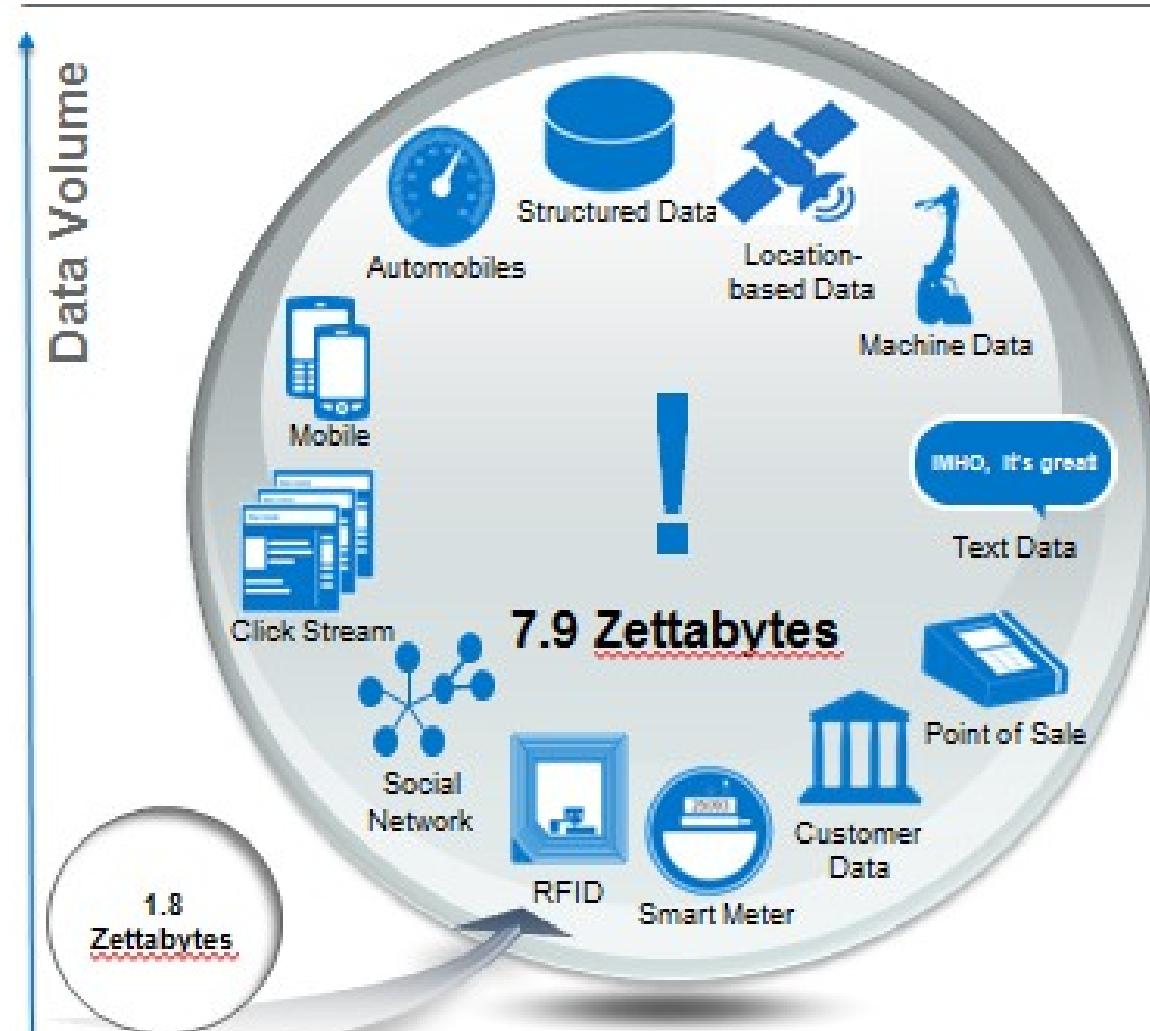


Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

## What Happens in an Internet Minute?



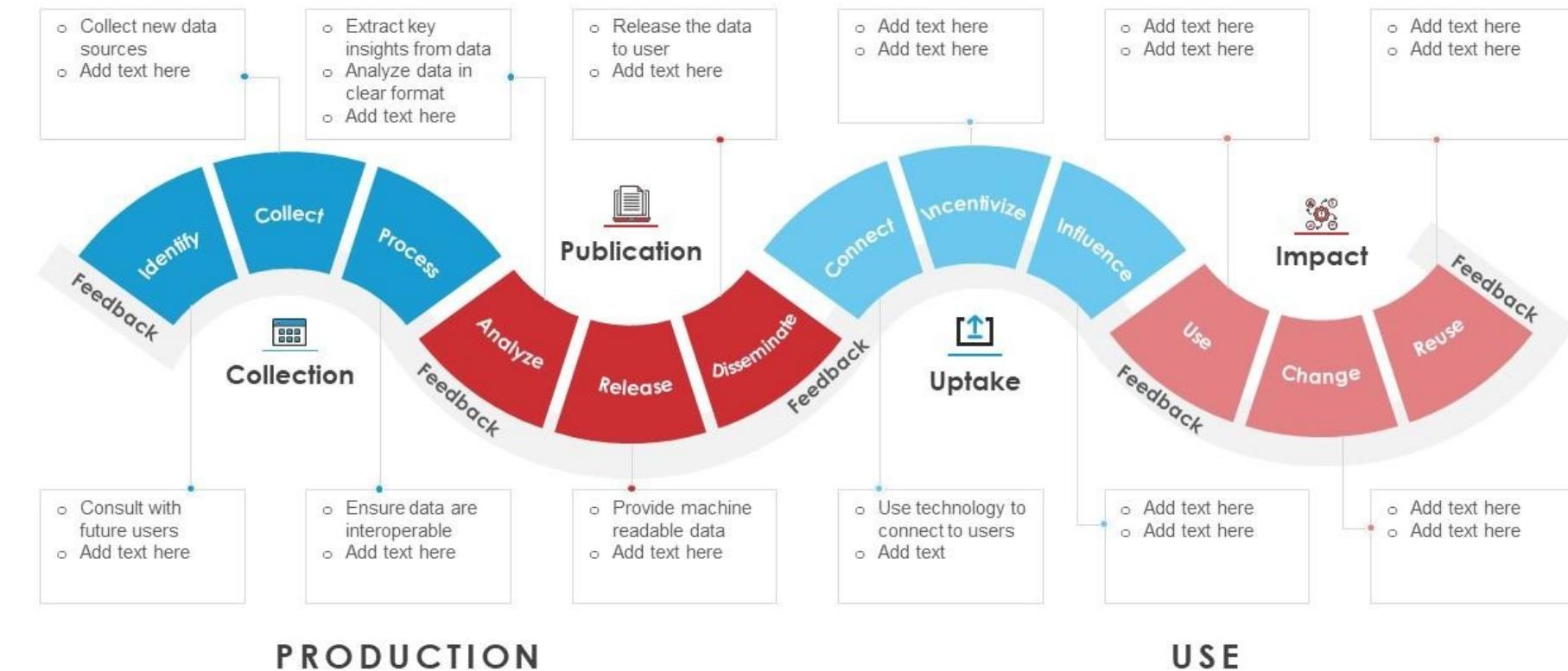
Fuente Imagen: [www.LifeHach.org](http://www.LifeHach.org)



Fuente Imagen: <https://insightextractor.com/>

## Data Change Management Plan with Value Chain Analysis

This slide shows the value change management lifecycle. It includes the process of collecting and publicize the data and afterwards adopting change management approach for obtaining the data driven outcomes.

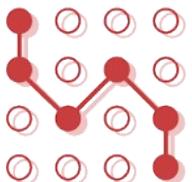


This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

Fuente Imagen: [www.SlideTeam.com/](http://www.SlideTeam.com/)



Statistical biases



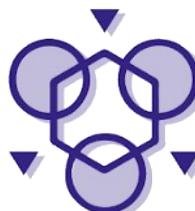
Lack of data lineage



Software bugs



Noise



Abnormalities



Information Security



Untrustworthy data sources



Falsification



Uncertainty and ambiguity of data



Duplication of data



Out of date and obsolete data



Human error

Fuente Imagen: [www.SciForce.com/](http://www.SciForce.com/)

# Fuentes de datos



Fuente Imagen: [www.SlideTeam.com/](http://www.SlideTeam.com/)

# Ciclo de Vida de Big Data



# Comprensión del negocio



Organiza tu propio FLISOL

FLISOL 2022 en:

- Argentina
- Bolivia
- Brasil
- Chile
- Colombia
- Cuba
- Ecuador
- España
- El Salvador
- Guatemala
- Honduras
- México
- Panamá
- Paraguay
- Perú
- Venezuela

Atajos útiles

- Como Ayudar
- Coordinadores
- Organización
- Recursos
- Material Gráfico
- Banners

Historico

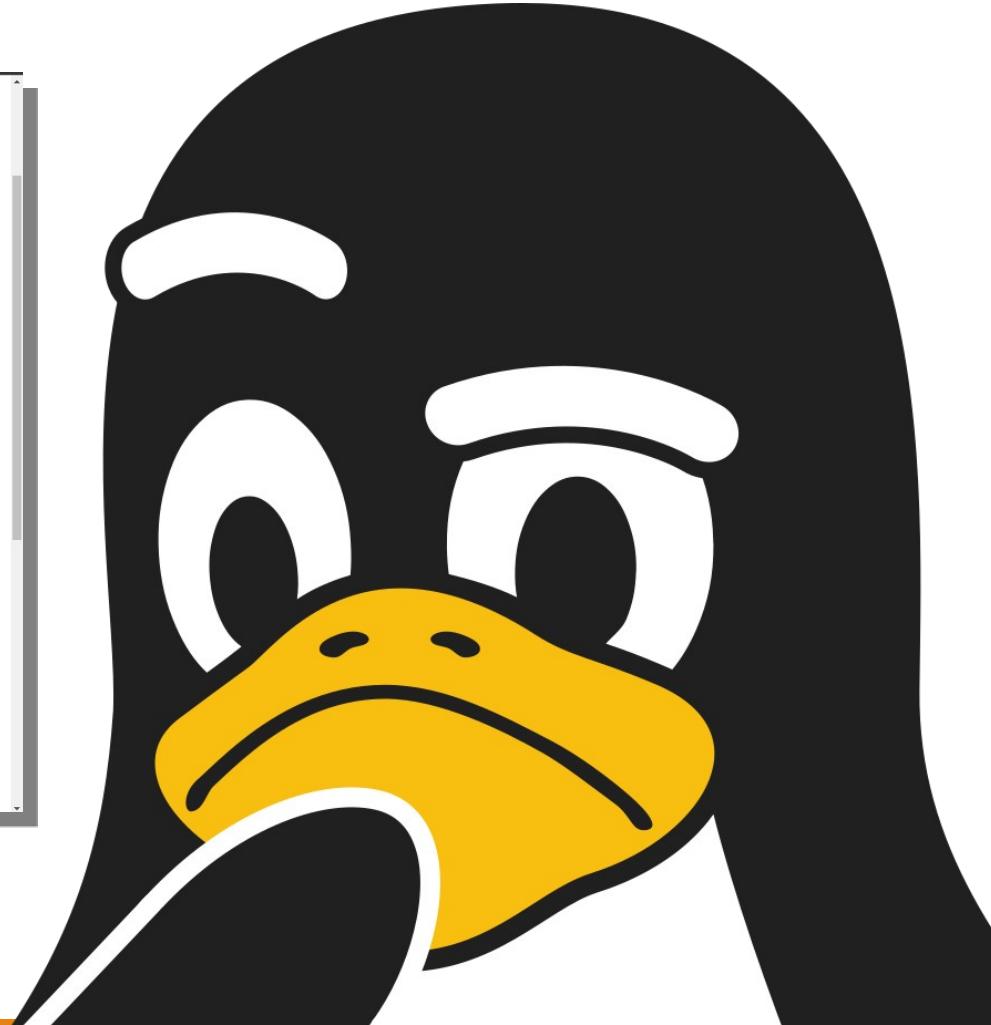
- FLISOL2021
- FLISOL2020
- FLISOL2019
- FLISOL2018
- FLISOL2017
- FLISOL2016
- FLISOL2015
- FLISOL2014
- FLISOL2013
- FLISOL2012
- FLISOL2011
- FLISOL2010
- FLISOL2009
- FLISOL2008
- FLISOL2007
- FLISOL2006
- FLISOL2005

FLISOL 2022 en Bogotá, Colombia

Revive el FLISOL 2022

Puedes ver la [programación](#)

- <https://youtu.be/6bXjEeVq0q4> - Infraestructure as Code in Cloud using Terraform
- <https://youtu.be/9vTj9e9ub-M> - Software y hardware libre para una cultura maker
- <https://youtu.be/jSpITqEqnwm> - Programa de Comunidad Wikimedia Colombia "
- <https://youtu.be/1uobkpvuEs0> - Software en el contexto de las elecciones de Colombia
- <https://youtu.be/5oFFcDRWzIQ> - Proyecto FotoE14
- <https://youtu.be/G60YT8HFFM> - Datos abiertos y experiencias de uso de software libre en Bogotá
- <https://youtu.be/DiDISPLAY4zw> - Espacios de aprendizaje seguros - WWCode Bogotá
- <https://youtu.be/ft224vyuOo> - Mapeemos tu casa en OpenStreetMap - Taller
- <https://youtu.be/oZnrPSTXy2l> - El valor las certificaciones Linux y Open Source en un Mercado Global
- <https://youtu.be/nDEwMSola4> - Privacidad y Software Libre en tiempos de persecución
- [https://youtu.be/ur4yhzkP5\\_4](https://youtu.be/ur4yhzkP5_4) - Flutter, Un solo código para muchas plataformas
- <https://youtu.be/ebD14kbM7Rc> - Despliegue de aplicaciones con GitLab CI
- <https://youtu.be/oFEUBGWicRI> - Como aprender a desarrollar en Ubuntu Touch con web technologies sin morir en el intento.
- <https://youtu.be/nf9xFW3Bdw> - A path forward for Latin America
- <https://youtu.be/w8xIxPWXXHU> - Tú eres software libre, yo soy software libre, todos somos software libre.
- <https://youtu.be/w37Dz1wPZU> - Escritura musical para principiantes con MuseScore
- [https://youtu.be/qJ5s\\_LdIEU](https://youtu.be/qJ5s_LdIEU) - Zoomorfismo aplicado a la Robótica
- <https://youtu.be/qnwotqgFKsg> - FIWARE la plataforma para el desarrollo y despliegue de aplicaciones de Internet del Futuro
- <https://youtu.be/vOYStGx404> - Blockchain criptomonedas, NFTs y play to earn
- <https://youtu.be/s01CSxDmvCY> - Matrix Chat Seguro, Autónomo y Federado para Organizaciones
- <https://youtu.be/S7Ap3EsUtBg> - Processing, arte digital.
- <https://youtu.be/5L2LZh2iXg> - Computadora en una caja
- <https://youtu.be/LqQf0Wa360> - Liberación de software y legalidad





shutterstock.com · 1461972224

# WebScraping

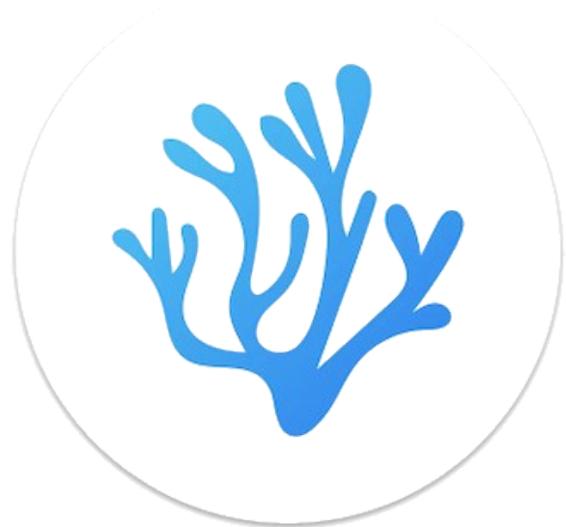
- <https://youtu.be/6bXjEeVq0q4> - Infraestructure as Code in Cloud using Terraform
- <https://youtu.be/9vTJ9e9ub-M> - Software y hardware libre para una cultura maker
- <https://youtu.be/JsSpfTgEqnw> - Programa de Comunidad Wikimedia Colombia "
- <https://youtu.be/1uoKpVuEsQ> - Software en el contexto de las elecciones de Colombia
- <https://youtu.be/5oFFcDRWziQ> - Proyecto **FotoE14**
- <https://youtu.be/G60yT8HFFfM> - Datos abiertos y experiencias de uso de software libre en Bogotá
- <https://youtu.be/DiDLAYIp4zw> - Espacios de aprendizaje seguros - WWCode Bogotá
- <https://youtu.be/fT224vy3uOo> - Mapeemos tu casa en **OpenStreetMap** - Taller
- <https://youtu.be/oznrPSTXy2I> - El valor de las certificaciones Linux y Open Source en un Mercado Global
- <https://youtu.be/ndEwoMSola4> - Privacidad y Software Libre en tiempos de persecución
- [https://youtu.be/ur4YhzkPS\\_4](https://youtu.be/ur4YhzkPS_4) - Flutter, Un solo código para muchas plataformas
- <https://youtu.be/ebDI4KbM7Rc> - Despliegue de aplicaciones con **GitLab CI**
- <https://youtu.be/0fEUBGWlcRI> - Como aprender a desarrollar en Ubuntu Touch con web technologies sin morir en el intento.
- <https://youtu.be/nf9xFIW3Bdw> - A path forward for Latin America
- <https://youtu.be/w8xIxPWXXHU> - Tú eres software libre, yo soy software libre, todos somos software libre.
- <https://youtu.be/w37Dz12wPZU> - Escritura musical para principiantes con **MuseScore**
- [https://youtu.be/jqr5s\\_LdIEU](https://youtu.be/jqr5s_LdIEU) - Zoomorfismo aplicado a la Robótica
- <https://youtu.be/qnwotgqFKsg> - FIWARE la plataforma para el desarrollo y despliegue de aplicaciones de Internet del Futuro
- <https://youtu.be/vOYSTvGx404> - Blockchain criptomonedas, NFTs y play to earn
- <https://youtu.be/s01CSxDmvcY> - Matrix Chat Seguro, Autónomo y Fedarado para Organizaciones
- <https://youtu.be/S7Ap3EsUtBg> - Processing, arte digital.
- <https://youtu.be/5L2ZLgh2iX8> - Computadora en una caja
- <https://youtu.be/LqQf0Wa36j0> - Liberación de software y legalidad
- <https://youtu.be/DbHgNEb-njY> - Fotomanipulacion con Gimp



Fuente Imagen: [www.DailyTech.com/](http://www.DailyTech.com/)



- **Función:** Python es un lenguaje de programación que nos permite trabajar ágilmente e integrar sistemas de forma eficiente.
- **Instalación:**  
<https://www.python.org/downloads/>
- **Uso específico:**
  - Abrir la(s) página(s) web
  - Hacer Web Scraping para obtener el listado de URLs de Youtube



**VSCodium**

- **Función:** La alternativa más fácil a VS Code, construido sin las adiciones propietarias de Microsoft.
- **Instalación:**  
<https://vscodium.com/>
- **Uso específico:**  
Editar los scripts de nuestra solución Big Data



# Requests: HTTP for Humans



**Requests**  
*http for humans*

- **Función:** Es una biblioteca HTTP muy simple y elegante, para Python.
- **Instalación:**  
`python -m pip install requests`
- **Uso específico:**  
Abrir nuestra(s) página(s) objetivo para obtener las URLs de Youtube.



# Beautiful Soup

BeautifulSoup

- **Función:** Beautiful Soup es una biblioteca de Python para extraer datos de archivos HTML y XML.
- **Instalación:**  
`pip install beautifulsoup4`
- **Uso específico:**  
Hacer Web Scraping a nuestra(s) página(s) objetivo para obtener las URLs de Youtube.

# Herramientas del Desarrollador (F12)



Screenshot of Mozilla Firefox browser showing the FLISOL website (<https://flisol.info>) with the developer tools (Inspector) open.

The page content includes:

- FLISOL logo and "Em português | Material gráfico" link.
- Text: "18º Festival Latinoamericano de Instalación de Software Libre el sábado 23 de abril 2022".
- Text: "El FLISOL es el evento de difusión de Software Libre más grande en Latinoamérica y está dirigido a **todo tipo de público**: estudiantes, académicos, empresarios, trabajadores, funcionarios públicos, entusiastas y aun personas que no poseen mucho conocimiento informático.."
- Section: "Revive el FLISOL 2022" with a link to "Bogotá, Colombia".
- Text: "Puedes ver <https://youtu.be/9VTj9e9ub-M> [216.983 x 18]
- Section: "Recursos" with a list of links to various FLISOL events from 2021 to 2010.
- Bottom of the page: "Infrastructure as Code in Cloud using Terraform" link.

The developer tools (Inspector) are open at the bottom, showing the DOM structure and CSS styles for the selected element. The selected element is a link with the class "anchor".

```
<p>
<ul>
  <li>
    <:marker>
    <:xp class="line891">
      <a class="https" href="https://youtu.be/6bXjEeVq0q4">https://youtu.be/6bXjEeVq0q4</a>
      - Infrastructure as Code in Cloud using Terraform
      <span id="line-16" class="anchor"></span>
    </:xp>
  </li>
</ul>
```

Inspector panel details:

- Selected element: `a https href="https://youtu.be/6bXjEeVq0q4">https://youtu.be/6bXjEeVq0q4`
- Style rules:
  - elemento { incorporado }
  - body { style.css:14 padding: 0; margin: 0; min-height: 100%; position: relative; }
  - Heredada de html
  - html { style.css:5 }
- Flexbox section: "Seleccione un contenedor Flex o un ítem para continuar."
- Quadrícula section: "La cuadrícula CSS no está en uso en esta página"
- Box Model section: "position border 0"

# Análisis para obtener la URL

```
▼ <p class="line891">  
  <a class="https" href="https://youtu.be/6bXjEeVq0q4">https://youtu.be/6bXjEeVq0q4</a>
```

- Infrastructure as Code in Cloud using Terraform

```
<span id="line16" class="chor"></span>
```

Etiqueta  
HTML

Clase

href (lo que queremos  
obtener)



# Demo codificación del Web Scraping

A screenshot of the Visual Studio Code (VSCode) interface. The main editor tab is titled 'flisol.py - Flisol - VSCode'. The code itself is a Python script for web scraping. It uses the 'requests' library to get the content of a URL ('https://flisol.info/'), then parses it with 'BeautifulSoup' using the 'html.parser'. It finds all anchor tags ('a') with a 'class\_="https"' attribute and extracts their 'href' values. These URLs are then written to a file named 'urls.txt' in 'w'rite mode.

```
flisol.py
1 import requests
2 from bs4 import BeautifulSoup
3
4 url = 'https://flisol.info/'
5 page = requests.get(url)
6
7 soup = BeautifulSoup(page.content, 'html.parser')
8
9 urls = [a['href'] for a in soup.find_all('a', class_="https", href=True)]
10
11 f = open("urls.txt", "w")
12 for url in urls:
13     f.write(url + '\n')
14 f.close()
```

The status bar at the bottom shows: Lín. 14, col. 10 Espacios: 4 UTF-8 LF Python 3.10.7 64-bit



# Demo ejecución del Web Scraping

The screenshot shows a Microsoft Visual Studio Code (VSCode) interface. On the left is the Explorer sidebar, which lists files and folders related to a project named 'FLISOL'. The 'urls.txt' file is currently selected, highlighted with a blue bar at the bottom of the sidebar. The main editor area displays the content of 'urls.txt', which contains a list of YouTube video URLs. Below the editor are tabs for 'PROBLEMAS', 'SALIDA', and 'TERMINAL'. The 'TERMINAL' tab is active, showing the command '/bin/python /home/jf/Documentos/u/Santoto/research/Flisol/flisol.py' being run in a terminal window. The terminal output shows the command being entered and then completed successfully with the message '[jf@fedora Flisol]\$'. The status bar at the bottom of the code editor shows the file path, line count (Lín. 1, col. 1), character count (Espacios: 4), encoding (UTF-8), line separator (LF), and text format (Texto sin formato).

```
urls.txt - Flisol - VSCode
flisol.py urls.txt
urls.txt
1 https://youtu.be/6bXjEeVq0q4
2 https://youtu.be/9vTJ9e9ub-M
3 https://youtu.be/JsSpfTgEqnw
4 https://youtu.be/1uobKpVuEsQ
5 https://youtu.be/5oFFcDRWziQ
6 https://youtu.be/G60yT8HFFfM
7 https://youtu.be/DiDLAYlp4zw
8 https://youtu.be/fT224vy3uOo
9 https://youtu.be/oznrPSTXy2I
10 https://youtu.be/ndEwoMSoIa4
11 https://youtu.be/ur4YhzkPS_4
12 https://youtu.be/ebDI4KbM7Rc
13 https://youtu.be/0fEUBGWlcRI
14 https://youtu.be/nf9xF1W3Bdw

PROBLEMAS SALIDA TERMINAL ...
/bin/python /home/jf/Documentos/u/Santoto/research/Flisol/flisol.py
● [jf@fedora Flisol]$ /bin/python /home/jf/Documentos/u/Santoto/research/Flisol/flisol.py
● <python /home/jf/Documentos/u/Santoto/research/Flisol/flisol.py
○ [jf@fedora Flisol]$
```

① 0 △ 0 Lín. 1, col. 1 Espacios: 4 UTF-8 LF Texto sin formato



# Subtítulos de Youtube

The screenshot shows a YouTube video titled "Infraestructure as Code in Cloud using Terraform" by "FLISoLBogota". The video is at 4:10 / 52:43. The main content is a presentation slide titled "Contextualization" comparing manual processes to automated ones. A yellow callout bubble points to the subtitle "subtítulos (lo que queremos obtener)".

**Before automation we were doing everything by hand:**

- Set up servers
- Configure networking
- Install and configure software...

**Repeat processes for multiple environments**

- Dev
- Test
- Staging
- Production

High human resources cost  
More effort and time  
More human errors possible

**donde no es que la culpa es del desarrollador no**

**subtítulos (lo que queremos obtener)**

Repetición de Top chat

La repetición del chat en vivo está activada. Los mensajes que aparecieron durante la transmisión en vivo se podrán ver aquí.

Carlos Andrés Páez Reyes

Ocultar la repetición del chat

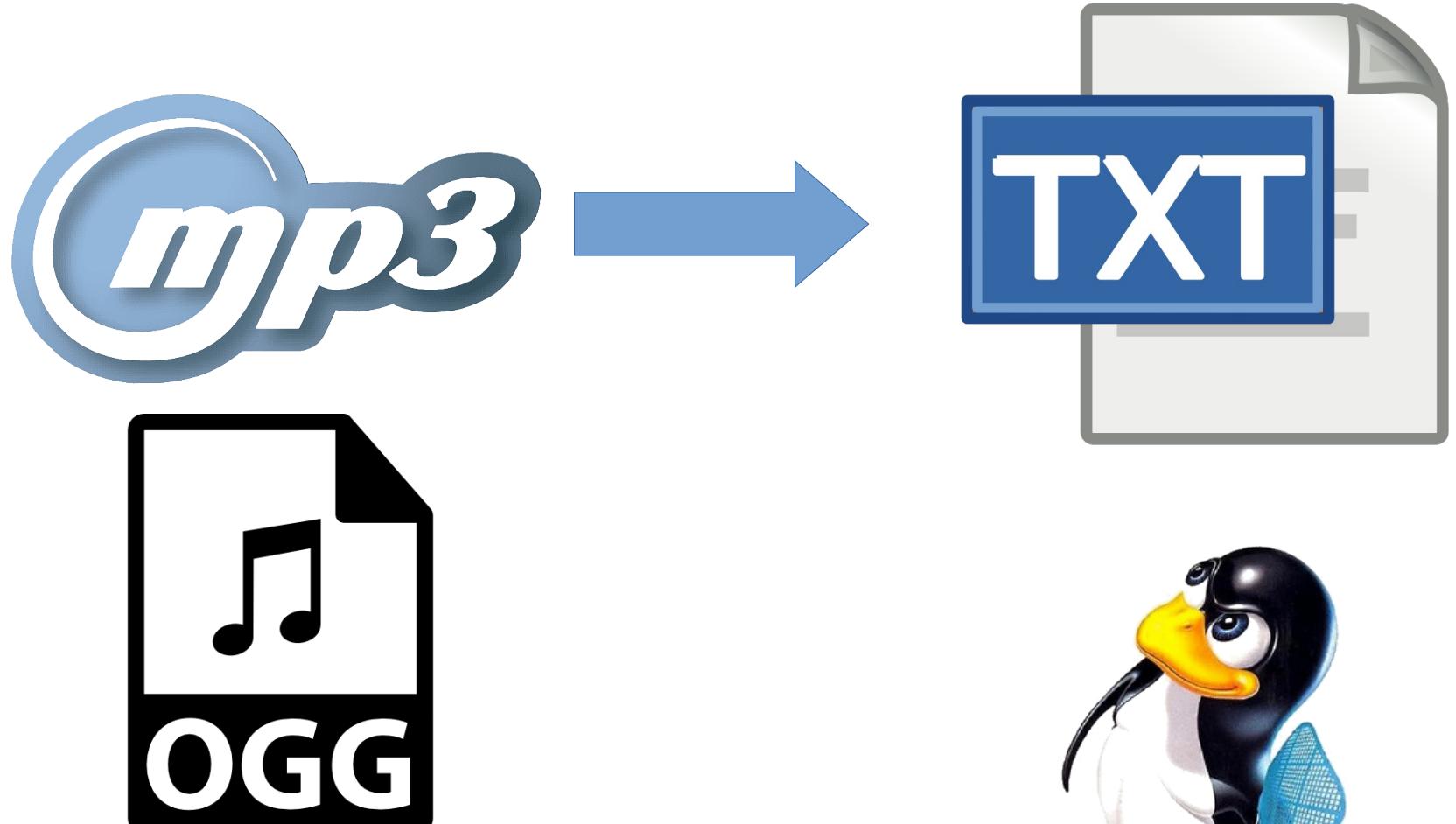
GitOps with GitLab - Infrastructure as Code demo...

36:07

Infraestructure As Code...

# Software de transcripción

- Transcriberry
- Transcriber
- Typingpool
- Transcription Buddy
- Transcription Helper
- VoiceWalker
- Transcribe!
- Express Scribe
- Happy Scribe
- Otranscribe
- Rev
- Express Scribe
- Trint
- Otter
- Temi
- Descript
- Nuance
- Sonix
- Audext
- **Youtube**





# Descarga de subtítulos de Youtube

Infraestructure as Code in Cloud using Terraform - DownSub.com — Mozilla Firefox

Proyecto Fedora - Página d FLISOL2022 - FLISOL Infraestructure as Code in C Infraestructure as Code in C

https://downsub.com/?url=https%3A%2F%2Fyoutu.be%2F6bXjEeVq0q4

HOME HISTORY SUPPORTED SITES CONTACT DONATE LANGUAGE

DOWNSub

https://youtu.be/6bXjEeVq0q4 DOWNLOAD

Infraestructure as Code in Cloud using Terraform  
Duration: 00:52:44

Settings

SRT TXT Spanish (auto-generated) Download Full Video With Subtitle

You May Like Sponsored Links by Taboola

Descargar subtítulos

The screenshot shows a Firefox browser window with several tabs open. The active tab is titled 'Infraestructure as Code in Cloud using Terraform - DownSub.com — Mozilla Firefox'. The URL in the address bar is 'https://downsub.com/?url=https%3A%2F%2Fyoutu.be%2F6bXjEeVq0q4'. The page itself is the DownSub website, displaying the video's title and duration. It includes a 'Settings' section, subtitle download buttons for 'SRT' and 'TXT', and a link to 'Download Full Video With Subtitle'. A large yellow button with the text 'Descargar subtítulos' has a yellow arrow pointing to the 'SRT' download button. The overall theme is related to infrastructure as code and cloud computing.



# Herramientas automatización





# Codificar automatización descarga

```
$ getsubtitles.sh
1  firefox -private-window&
2  sleep 2
3  xdotool type https://downsub.com/
4  xdotool key KP_Enter
5  sleep 1
6  xdotool key Tab
7  xdotool key Tab
8  xdotool key Tab
9  xdotool key Tab
10 xdotool key Tab
11 xdotool key Tab
12 xdotool key Tab
13 xdotool key Tab
14
```

The screenshot shows a Firefox browser window with the following details:

- Address bar: <https://downsub.com>
- Page title: Download subtitles from Youtube, Viki, Viu, Vlive and more! - DownSub — Mozilla Firefox
- Navigation icons: Back, Forward, Stop, Home, etc.
- Header menu: Proyecto Fedora - Página d, FLISOL2022 - FLISOL, Infraestructure as Code in
- Content area:
  - Submenu icon (yellow circle labeled 3)
  - Download button: Download subtitles from Y
  - Navigation links: HOME (yellow circle labeled 6), HISTORY (yellow circle labeled 7), SUPPORTED SITES (yellow circle labeled 8), CONTACT (yellow circle labeled 9), DONATE (yellow circle labeled 10), LANGUAGE (yellow circle labeled 11), and another LANGUAGE link (yellow circle labeled 12).
  - Input field: Enter a link to download subtitles. Ex: <https://youtu.be/rN7yhDl1cuk>
  - Download button: DOWNLOAD

13



# Pegar URLs

The image shows a terminal window on the left and a video player interface on the right. The terminal window displays a shell script named `getsubtitles.sh` with the following content:

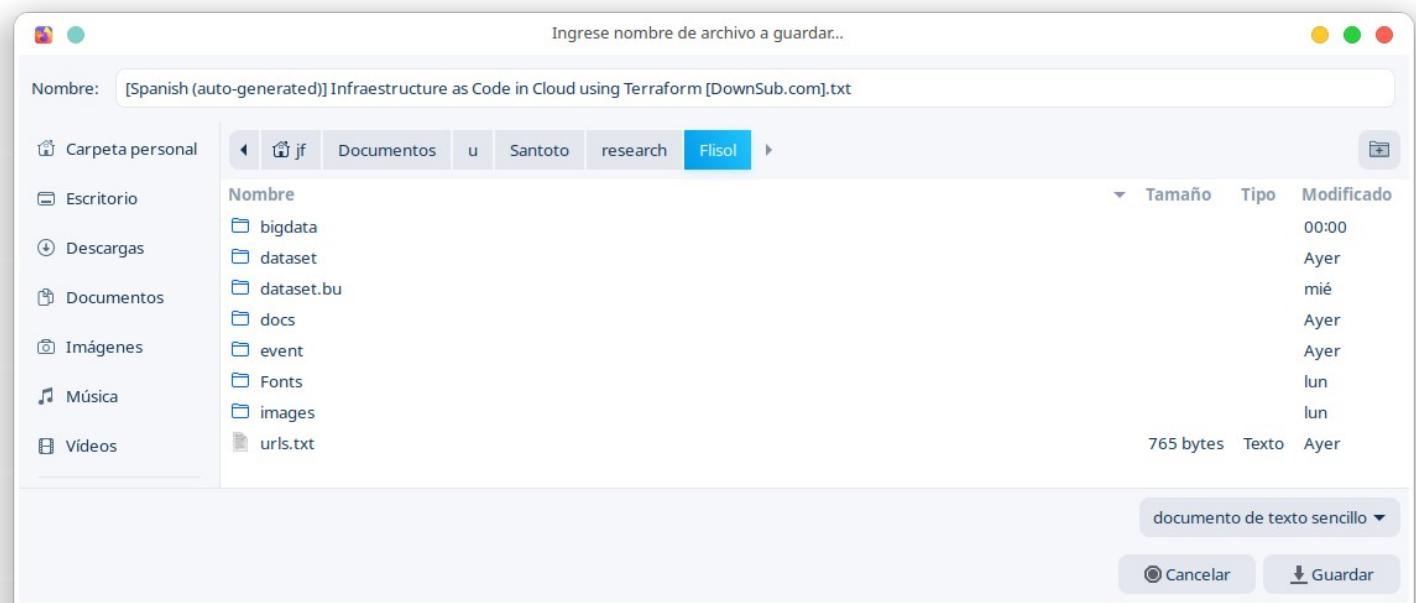
```
$ getsubtitles.sh
15 while read p; do
16     xdotool type "$p"
17     xdotool key KP_Enter
18     sleep 2
19     xdotool key Tab
20     xdotool key Tab
21     xdotool key Tab
22     xdotool key Tab
23     xdotool key Tab
24     xdotool key Tab
25     xdotool key Tab
26     xdotool key Tab
27     xdotool key Tab
28     xdotool key Tab
29     xdotool key Tab
30     xdotool key Tab
31     xdotool key Tab
32     xdotool key Tab
33     xdotool key Tab
34     xdotool key Tab
35     xdotool key KP_Enter
36     sleep 2
```

The video player interface shows a video titled "Infrastructure as Code in Cloud using Terraform" with a duration of 00:52:44. The video thumbnail features the FLISOL 2022 logo. A yellow circle highlights the number 16, which corresponds to the line number in the terminal script. Another yellow circle highlights the number 35, corresponding to the line number in the terminal script. The video player also shows download options for SRT and TXT subtitles, and a button to "Download Full Video With Subtitle".

# Descargar el archivo de subtítulos

```
$ getsubtitles.sh
```

```
37      xdotool key KP_Enter
38      sleep 2
39      xdotool key Shift+Tab
40      xdotool key Shift+Tab
41      xdotool key Shift+Tab
42      xdotool key Shift+Tab
43      xdotool key Shift+Tab
44      xdotool key Shift+Tab
45      xdotool key Shift+Tab
46      xdotool key Shift+Tab
47 done < urls.txt
48 xdotool key Ctrl+w
49
```





# Demo Descargas de los subtítulos



# Dataset de subtítulos

A screenshot of a Mac OS X-style file explorer window. The left sidebar shows standard navigation links like Recientes, Favoritos, Carpeta personal, Escritorio, Descargas, Documentos, Imágenes, Música, Videos, and Papelera. The main pane displays a folder named 'subtitles' containing 22 subtitle files. The files are listed in descending order of size, all being auto-generated Spanish subtitles. Each file has a star icon to its right. The path at the top of the window is 'Carpeta personal / Documentos / u / Santoto / research / Flisol / dataset'.

Nombre	Tamaño	Modificado	
[Spanish (auto-generated)] Blockchain criptomonedas, NFTs y ...	43,0 kB	00:28	★
[Spanish (auto-generated)] Como aprender a desarrollar en U...	86,2 kB	00:27	★
[Spanish (auto-generated)] Computadora en una caja - Mario J...	44,2 kB	00:28	★
[Spanish (auto-generated)] Datos abiertos y experiencias de u...	33,1 kB	00:26	★
[Spanish (auto-generated)] Despliegue de aplicaciones con Git...	35,5 kB	00:27	★
[Spanish (auto-generated)] El valor de las certificaciones Linux ...	39,4 kB	00:27	★
[Spanish (auto-generated)] Escritura musical para principiante...	17,7 kB	00:27	★
[Spanish (auto-generated)] Espacios de aprendizaje seguros - ...	33,7 kB	00:26	★
[Spanish (auto-generated)] FIWARE la plataforma para el desa...	36,8 kB	00:28	★
[Spanish (auto-generated)] Flutter, Un solo código para mucha...	28,5 kB	00:27	★
[Spanish (auto-generated)] Fotomanipulación con Gimp - Tatic...	41,9 kB	00:28	★
[Spanish (auto-generated)] Ilustración Editorial en GNU Linux -...	45,2 kB	00:28	★
[Spanish (auto-generated)] Infraestructure as Code in Cloud u...	44,1 kB	00:26	★
[Spanish (auto-generated)] Liberación de software y legalidad ...	58,6 kB	00:28	★
[Spanish (auto-generated)] Mapeemos tu casa en OpenStreet...	43,2 kB	00:26	★
[Spanish (auto-generated)] Matrix Chat Seguro, Autónomo y F...	53,2 kB	00:28	★
[Spanish (auto-generated)] Processing, arte digital. - Ana Guz...	23,7 kB	00:28	★
[Spanish (auto-generated)] Proyecto FotoE14 - Fundación Corr...	40,1 kB	00:26	★
[Spanish (auto-generated)] Software en el contexto de las elec...	34,1 kB	00:26	★
[Spanish (auto-generated)] Software y hardware libre para una...	36,3 kB	00:26	★



# Herramientas reemplazo cadenas caracteres

```
man(1)      800 General Commands Manual      man(1)

NAME
     sed --- stream editor

SYNOPSIS
     sed [option] command [file]...[file]...[file]...[file]...[file]...[file]...[file]...

DESCRIPTION
     The sed utility reads the specified files, or the standard input if no files are specified, and writes the input as associated by the command. The input is then written to the standard output.

     A single command may be specified as the first argument to sed. Multiple commands may be specified by using the -e option. All commands are applied to the input in the order they are specified regardless of their origin.

     The following options are available:
       -e      Insert regular expressions as extended (POSIX) regular expressions rather than basic regular expressions described in the International General Usage Facility Description with Formats.
       -n      The lines listed as parameters for the -w command are truncated before any processing happens by default. The -n option causes sed to only report each file with a command.
       -c      Causes sed to print the command line number and the command itself.
       -d      Deletes lines matching the pattern.
       -f      Causes sed to read commands from the file(s) specified as arguments. The file(s) must contain valid sed commands.
       -i      Edits the files in place; each file of input is written to the standard output after all of the commands have been applied to it. The -i option suppresses this behavior.
       -l      The following command is as follows:
              [address]...[function](arguments)
       -s      Whitespace may be used to separate the first address and the function portions of the command.
       -t      Normally, sed will copy a line of input, not including its terminating newline character, into a pattern space, unless there is something left after a 'd' function, applies all of the commands to addresses that select that pattern space, copies the pattern space to the output space, appends a newline, and deletes the pattern space.
       -u      Some of the functions use a context address or part of the pattern space for subsequent retrieval.
       -v      Used Addresses
              An address or two addresses, one of which must be a number (the counts begin from sequentially across input files), a dollar ('$') character that addresses the last line of input, or a context address which consists of a regular expression preceded and followed by a colon.

       -w      Whitespace or newlines, but not carriage returns, may be inserted into the pattern space.
```

# Stream Editor





# Código depuración de nombres de archivos

The screenshot shows a VSCode interface with a terminal tab open. The terminal window displays a bash script named `clean_filenames.sh`. The script content is as follows:

```
$ clean_filenames.sh
1 #!/bin/bash
2 for f in *.txt
3 do
4     mv "$f" "$(echo "$f" | tr -d '[]' | tr -d '()' | sed -e 's/Spanish auto-generated //' | sed -e s'/ DownSub.com//')"
5 done
6
```



# Dataset depurado

The screenshot shows a file explorer interface with the following details:

- Path:** /s/u/Santoto/research/Flisol/dataset/subtitles
- Left sidebar:** Includes links to Recientes, Favoritos, Carpeta personal, Escritorio, Descargas, Documentos, Imágenes, Música, Vídeos, Papelera, and Otras ubicaciones.
- Right pane:** A list of files under the "subtitles" folder, each represented by a document icon and its name.

Nombre
Blockchain criptomonedas, NFTs y play tc - David Vega.txt
Como aprender a desarrollar en Ubuntu Touch con web technologies.txt
Computadora en una caja - Mario Josué Pérez Cruz.txt
Datos abiertos y experiencias de uso de software libre en Bogotá - Ana Carolina Escobar.txt
Despliegue de aplicaciones con GitLab CI - Mario García.txt
El valor de las certificaciones Linux y Open Source en un Mercado Global.txt
Escritura musical para principiantes con MuseScore - Laura Losada.txt
Espacios de aprendizaje seguros - WWCode Bogotá - Ing Isabel Yepes.txt
FIWARE la plataforma para el desarrollo y despliegue de aplicaciones de Internet del Futuro.txt
Flutter, Un solo código para muchas plataformas - David García.txt
Fotomanipulacion con Gimp - Tatica Leandro.txt



# Ejemplo archivo de subtítulo

Datos abiertos y experiencias de uso de software libre en Bogotá - Ana Carolina Escobar.txt \* — KWrite

+ Nuevo Abrir Guardar Guardar como Cerrar Deshacer Rehacer

```
1 leonard si te escuchó perfectamente y
2 permite un momentico
3 ok
4 bueno muy buenas a todos nos encontramos
5 en otra edición del flisol bogotá año
6 2022 y nos encontramos con ana carolina
7 carolina escobar de alta consejería
8 buenos dias sana hola buenos dias cómo
9 estás bien por favor si quieres puedes
10 presentarte visto bueno mi nombre es ana
11 carolina escobar hago parte de el equipo
12 de trabajo de la alta consejería
13 disfruta el tic
14 en especial y en particular del equipo
15 de gobierno abierto que es una
16 estrategia de distrito que busca
17 permear el modelo de transparencia
18 colaboración participación entre todas
19 las entidades con muchas apuestas en las
20 distintas líneas en la que el tema que
21 vamos a tratar hoy tenemos 222 apuestas
22 importantes o dos temas que estamos
23 manejando que es datos abiertos en la
24 línea en el componente de transparencia
25 y por la invitación que nos ha hecho
26 crisol pues es una indagación de como en
27 el distrito venimos usando el tema de
28 software público qué desafíos representa
29 esto para la entidad para la entidad
30 como alcaldía mayor de bogotá
31 y bueno pues nada para fortalecer lazos
32 con la comunidad para conversar para
33 abrirles las puertas al diálogo a
34 conocer lo que están haciendo desde la
35 academia desde los grupos organizados de
```

1:1 INSERTAR es\_CO Tabuladores débiles: 4 UTF-8 Normal

# Herramientas Populares de Big Data



**Este es el inicio...  
Pero la historia con Big Data  
continuará en un próximo episodio...**

# Juan Francisco Mendoza Moreno



@jfmdozam



jfmendozam



jfmendozamoreno

# ¿Preguntas?

# Gracias

FLISOL 2022, Puro software Libre!

FLISOL | Programa Ingeniería de Sistemas |  
FUP | POPAYÁN 2022

