

# Archivado de software público a escala masiva

---

Adrián Alves Morales – UO284288

Luis Manuel Solares García – UO282631

# ¿En qué consiste?

- Preservación de software de interés público en grandes cantidades.
- Su objetivo es garantizar que el software esté disponible para la posteridad y esté al alcance de todos.



# Acceso y uso del Software Público archivado

→ Puede variar dependiendo del producto.

→ Descarga directa, solicitud previa...

→ Restricciones por derechos de autor.



# Ejemplos de archivos de Software Público

→ Internet Archive.

→ Proyecto GNU.

→ **Software Heritage Project.**



# Software Heritage

- Fundada por Roberto Di Cosmo.
- Archivo universal de todo el código fuente disponible públicamente.
- Surge de la necesidad de una fuente de Software unificada persistente.



# Arquitectura



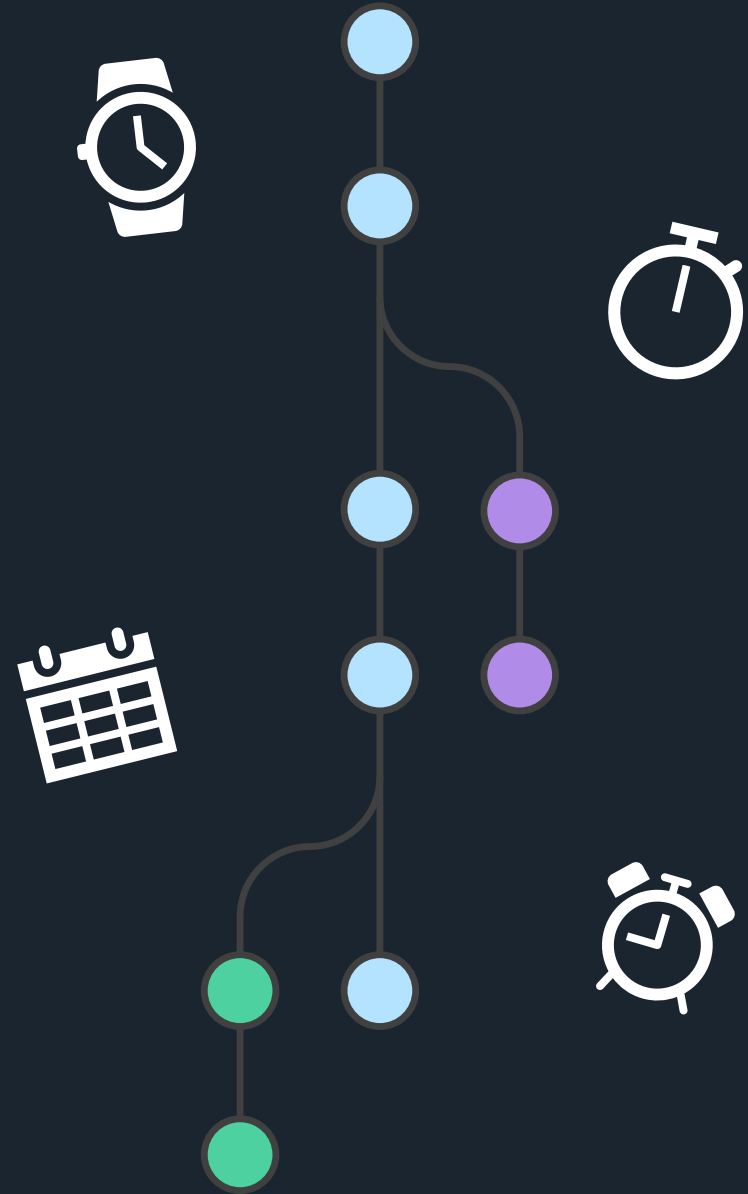
→ ¿Cómo almacenamos repositorios de diferentes orígenes?

→ Los almacenamos automáticamente.

→ Con una línea, tenemos acceso a 180 millones de repositorios.

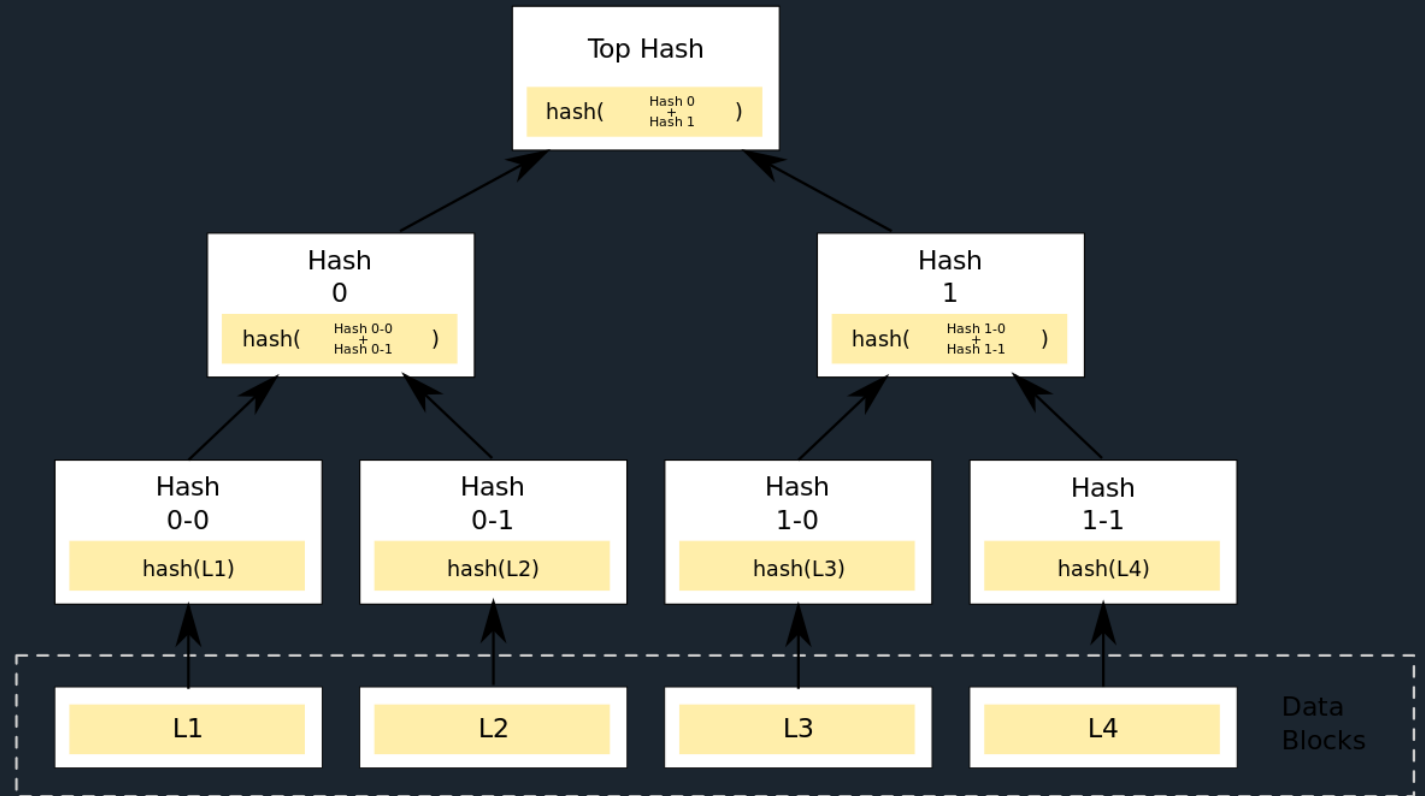
# ¿Cada cuánto almacenamos los datos?

- Forma normal (regular).
- Save code now.
- Acuerdos.



# Estructura del Grafo

- Grafo Merkle.
- Permite deduplicar datos de los repositorios archivados.





# EL SWHID

- Estándar de claves intrínsecas relacionadas a repositorios, directorios, etc.
- Permite acceder a la información con un solo identificador.



# Desafíos del archivado de software

- Selección y priorización de software y versiones.
- Elección de las tecnologías adecuadas.
- Gestión de obsolescencia.
- Protección de privacidad.
- Necesidad de mantenimiento.



# FIN

Gracias por vuestra atención

