**Un conjunto de letras negras en un fondo blanco

Descripción generada automáticamente con confianza bajaUn conjunto de letras negras en un fondo blanco

Descripción generada automáticamente con confianza bajaUniversidad Autónoma de Coahuila**

**Facultad de Ingeniería Mecánica y Eléctrica**

**Modelos Computacionales**

**“Mecanismos para procesamiento de gran volumen”**

**Ing. Jesús Homero Carmona Mendoza**

**Daniela López Rodríguez**

**7°A ISC T.M.**

**Agosto – diciembre 2022**

**24/10/2022**

**Mecanismos para procesamiento de gran volumen**

**Procesamiento batch: para lotes de grandes volúmenes de datos**

Apache Hadoop es un marco de computación distribuida el modelo de Google MapReduce para procesar grandes cantidades de datos en paralelo. Hadoop Distributed File System (HDFS) es el sistema de archivos subyacente de un cluster Hadoop y funciona de manera más eficiente con un número reducido de archivos big data de gran volumen, que con una cantidad superior de archivos de datos más pequeños.

Un trabajo en el mundo Hadoop suele durar de minutos a horas para completarse, por lo tanto, podría afirmarse que la opción Hadoop no es la más indicada cuando el negocio tenga la necesidad de llevar a cabo un análisis en tiempo real, sino más bien en los casos en que sea posible conformarse con una analítica offline.

Recientemente, Hadoop ha evolucionado para adaptarse a las nuevas necesidades empresariales. Los negocios hoy día demandan:

* Menores latencias.
* Minimización del tiempo de respuesta.
* Precisión máxima en la elaboración de decisiones.

Hadoop se ha renovado mejorando su capacidad de gestión gracias a una novedad que se conoce como stream. Uno de los principales objetivos de Hadoop streaming es desacoplar Hadoop MapReduce del paradigma para dar cabida a otros modelos de computación en paralelo, como MPI (Message Passing Interface) y Spark. Con las novedades de la aplicación de técnicas de procesamiento y análisis de datos en streaming se superan muchas de las limitaciones del modelo batch que, si bien puede considerarse demasiado rígido para determinadas funciones, algo que no es de extrañar si se tiene en cuenta que sus orígenes datan de hace más de cuatro décadas; todavía es el más indicado, por la relación coste - resultados, para operaciones como:

* El cálculo del valor de mercado de los activos, que no necesita revisarse más de una vez al día.
* Cálculo mensual del coste de las facturas de teléfono de los empleados.
* Generación de informes relacionados con temas fiscales.

**Procesamiento en stream**

Este tipo de técnicas de procesamiento y análisis de datos se basan en la implementación de un modelo de flujo de datos en el que los datos asociados a series de tiempo (hechos) fluyen continuamente a través de una red de entidades de transformación que componen el sistema. Se conoce como procesamiento streaming o de flujo.

No hay limitaciones de tiempo obligatorias en el procesamiento de flujo, al contrario de lo que sucede con las técnicas de procesamiento y análisis de datos en tiempo real. Por ejemplo, un sistema que se ocupe del recuento de las palabras incluidas en cada tweet para el 99,9% de los tweets procesados es un sistema de procesamiento en stream válido. Tampoco existe una obligación en cuanto al plazo de tiempo de generación del output por cada input recibido en el sistema. Las únicas limitaciones son:

* Se debe disponer de suficiente memoria para almacenar entradas en cola.
* La tasa de productividad del sistema a largo plazo debería ser más rápida, o por lo menos igual, a la tasa de entrada de datos en ese mismo periodo. Si esto no fuese así, los requisitos de almacenamiento del sistema crecerían sin límite.

Este tipo de técnicas de procesamiento y análisis de datos no está destinado a analizar un conjunto completo de grandes datos, por lo que generalmente no presentan esa capacidad, salvo excepciones.

**Técnicas de procesamiento y análisis de datos en tiempo real**

Cuando los datos se trabajan en tiempo real el nivel de procesamiento analítico en línea alcanzado es extremadamente alto y el margen es inferior a segundos. Precisamente por ello, los sistemas real time no suelen usar mecanismos especiales para la atomicidad y durabilidad. Simplemente se ocupan de procesar la entrada tan pronto como sea posible.

La cuestión es lo que puede suceder si pierden la entrada. Cuando esto ocurre, hacen caso omiso de la pérdida y siguen procesando y analizando sin detenerse. En según qué entorno, esto no supone un problema, por ejemplo, en un ecommerce, pero puede serlo en el sistema de vigilancia de seguridad de un banco o de una instalación militar. No es bueno que se pierda información, pero incluso la tecnología tiene un límite y, cuando se trabaja de verdad en tiempo real, el sistema no puede dejar las operaciones para volver a arreglar algo que ya es pasado, quedó segundos atrás. Los datos que siguen llegando y el sistema debe hacer todo lo posible para continuar su procesamiento.

En cualquier caso, las técnicas de procesamiento y análisis de datos en tiempo real merecen una seria consideración, antes de su implementación ya que:

* No son tan sencillas de poner en práctica mediante el uso de sistemas de software comunes.
* Su coste es superior al de las opciones streaming.
* Dependiendo del fin para el que se quieran utilizar, puede ser preferible optar por una opción intermedia entre streaming y tiempo real, como la que usa Amazon en su web y que garantiza un resultado que no excede en ningún caso los cien o doscientos milisegundos para el 99% de todas las solicitudes

**Redes neuronales**

Una red neuronal es un método de la inteligencia artificial que enseña a las computadoras a procesar datos de una manera que está inspirada en la forma en que lo hace el cerebro humano. Se trata de un tipo de proceso de machine learning llamado aprendizaje profundo, que utiliza los nodos o las neuronas interconectados en una estructura de capas que se parece al cerebro humano. Crea un sistema adaptable que las computadoras utilizan para aprender de sus errores y mejorar continuamente. De esta forma, las redes neuronales artificiales intentan resolver problemas complicados, como la realización de resúmenes de documentos o el reconocimiento de rostros, con mayor precisión.

Las redes neuronales pueden ayudar a las computadoras a tomar decisiones inteligentes con asistencia humana limitada. Esto se debe a que pueden aprender y modelar las relaciones entre los datos de entrada y salida que no son lineales y que son complejos.

Las redes neuronales pueden comprender datos no estructurados y hacer observaciones generales sin un entrenamiento explícito. Por ejemplo, pueden reconocer que dos oraciones de entrada diferentes tienen un significado similar:

* ¿Puede explicarme cómo hacer el pago?
* ¿Cómo puedo transferir dinero?

Una red neuronal sabría que ambas oraciones significan lo mismo. O sería capaz de reconocer, en términos generales, que Baxter Road es un lugar, pero que Baxter Smith es el nombre de una persona.

Las redes neuronales están presentes en varios casos de uso en muchos sectores, como los siguientes:

* Diagnóstico médico mediante la clasificación de imágenes médicas
* Marketing orientado mediante el filtrado de redes sociales y el análisis de datos de comportamiento
* predicciones financieras mediante el procesamiento de datos históricos de instrumentos financieros
* Previsión de la carga eléctrica y la demanda de energía
* Proceso y control de calidad
* Identificación de compuestos químicos

### **Visión artificial**

La visión artificial es la capacidad que tienen las computadoras para extraer información y conocimientos de imágenes y videos. Con las redes neuronales, las computadoras pueden distinguir y reconocer imágenes de forma similar a los humanos. La visión artificial tiene varias aplicaciones, como las siguientes:

* Reconocimiento visual en los vehículos autónomos para que puedan reconocer las señales de tráfico y a otros usuarios del camino
* Moderación de contenido para eliminar de forma automática los contenidos inseguros o inapropiados de los archivos de imágenes y videos
* Reconocimiento facial para identificar rostros y reconocer atributos como ojos abiertos, gafas y vello facial
* Etiquetado de imágenes para identificar logotipos de marcas, ropa, equipos de seguridad y otros detalles de la imagen

### **Reconocimiento de voz**

Las redes neuronales pueden analizar el habla humana a pesar de los diferentes patrones de habla, el tono, el idioma y el acento. Los asistentes virtuales como Amazon Alexa y el software de transcripción automática utilizan el reconocimiento de voz para realizar tareas como las siguientes:

* Asistir a los agentes de los centros de llamadas y clasificar las llamadas de forma automática
* Convertir las conversaciones clínicas en documentación en tiempo real
* Subtitular con precisión videos y grabaciones de reuniones para aumentar el alcance del contenido

### **Procesamiento de lenguaje natural**

El procesamiento de lenguaje natural (PLN) es la capacidad de procesar texto natural creado por humanos. Las redes neuronales ayudan a las computadoras a obtener información y significado a partir de los datos y los documentos de texto. El PLN está presente en varios casos de uso, entre los que se incluyen los siguientes:

* Chatbots y agentes virtuales automatizados
* Organización y clasificación automáticas de datos escritos
* Análisis de inteligencia empresarial de documentos con formato largo, como emails y formularios
* Indexación de frases clave que indican sentimientos, como los comentarios positivos y negativos en las redes sociales
* Resumen de documentos y producción de artículos para un tema determinado

### **Motores de recomendaciones**

Las redes neuronales pueden hacer un seguimiento de la actividad del usuario para elaborar recomendaciones personalizadas. También pueden analizar todo el comportamiento de los usuarios y descubrir productos o servicios nuevos que interesen a un usuario específico. Por ejemplo, Curalate, una empresa emergente con sede en Filadelfia ayuda a las marcas a convertir las publicaciones en las redes sociales en ventas. Las marcas utilizan el servicio de etiquetado inteligente de productos (IPT) de Curalate para automatizar la recopilación y la selección del contenido social que generan los usuarios. El IPT utiliza las redes neuronales para encontrar y recomendar de forma automática productos relevantes para la actividad del usuario en las redes sociales. Los consumidores no tienen que buscar en los catálogos en línea para encontrar un producto específico a partir de una imagen en las redes sociales. En cambio, pueden utilizar el etiquetado automático de productos de Curalate para comprar el producto con facilidad.

## ¿Cómo funcionan las redes neuronales?

El cerebro humano es lo que inspira la arquitectura de las redes neuronales. Las células del cerebro humano, llamadas neuronas, forman una red compleja y con un alto nivel de interconexión y se envían señales eléctricas entre sí para ayudar a los humanos a procesar la información. De manera similar, una red neuronal artificial está formada por neuronas artificiales que trabajan juntas para resolver un problema. Las neuronas artificiales son módulos de software, llamados nodos, y las redes neuronales artificiales son programas de software o algoritmos que, en esencia, utilizan sistemas informáticos para resolver cálculos matemáticos.

### **Arquitectura de una red neuronal simple**

Una red neuronal básica tiene neuronas artificiales interconectadas en tres capas:

#### Capa de entrada. La información del mundo exterior entra en la red neuronal artificial desde la capa de entrada. Los nodos de entrada procesan los datos, los analizan o los clasifican y los pasan a la siguiente capa.

#### Capa oculta. Toman su entrada de la capa de entrada o de otras capas ocultas. Las redes neuronales artificiales pueden tener una gran cantidad de capas ocultas. Cada capa oculta analiza la salida de la capa anterior, la procesa aún más y la pasa a la siguiente capa.

#### Capa de salida. Proporciona el resultado final de todo el procesamiento de datos que realiza la red neuronal artificial. Puede tener uno o varios nodos. Por ejemplo, si tenemos un problema de clasificación binaria (sí/no), la capa de salida tendrá un nodo de salida que dará como resultado 1 o 0. Sin embargo, si tenemos un problema de clasificación multiclase, la capa de salida puede estar formada por más de un nodo de salida.

### **Arquitectura de una red neuronal profunda**

Las redes neuronales profundas, o redes de aprendizaje profundo, tienen varias capas ocultas con millones de neuronas artificiales conectadas entre sí. Un número, denominado peso, representa las conexiones entre un nodo y otro. El peso es un número positivo si un nodo estimula a otro, o negativo si un nodo suprime a otro. Los nodos con valores de peso más altos tienen mayor influencia en los demás nodos. En teoría, las redes neuronales profundas pueden asignar cualquier tipo de entrada a cualquier tipo de salida. Sin embargo, también necesitan mucho más entrenamiento en comparación con otros métodos de machine learning. Necesitan millones de ejemplos de datos de entrenamiento en lugar de los cientos o miles que podría necesitar una red más simple.

## ¿Cuáles son los tipos de redes neuronales?

Las redes neuronales artificiales pueden clasificarse en función de cómo fluyen los datos desde el nodo de entrada hasta el nodo de salida.

Las redes neuronales prealimentadas procesan los datos en una dirección, desde el nodo de entrada hasta el nodo de salida. Todos los nodos de una capa están conectados a todos los nodos de la capa siguiente. Una red prealimentada utiliza un proceso de retroalimentación para mejorar las predicciones a lo largo del tiempo.

### Algoritmo de retropropagación. Las redes neuronales artificiales aprenden de forma continua mediante el uso de bucles de retroalimentación correctivos para mejorar su análisis predictivo. En pocas palabras, puede pensar en los datos que fluyen desde el nodo de entrada hasta el nodo de salida a través de muchos caminos diferentes en la red neuronal. Solo un camino es el correcto: el que asigna el nodo de entrada al nodo de salida correcto. Para encontrar este camino, la red neuronal utiliza un bucle de retroalimentación que funciona de la siguiente manera:

1. Cada nodo intenta adivinar el siguiente nodo de la ruta.
2. Se comprueba si la suposición es correcta. Los nodos asignan valores de peso más altos a las rutas que conducen a más suposiciones correctas y valores de peso más bajos a las rutas de los nodos que conducen a suposiciones incorrectas.
3. Para el siguiente punto de datos, los nodos realizan una predicción nueva con las trayectorias de mayor peso y luego repiten el paso 1.

Las capas ocultas de las redes neuronales convolucionales realizan funciones matemáticas específicas, como la síntesis o el filtrado, denominadas convoluciones. Son muy útiles para la clasificación de imágenes porque pueden extraer características relevantes de las imágenes que son útiles para el reconocimiento y la clasificación de imágenes. La forma nueva es más fácil de procesar sin perder características que son fundamentales para hacer una buena predicción. Cada capa oculta extrae y procesa diferentes características de la imagen, como los bordes, el color y la profundidad.

## ¿Cómo entrenar las redes neuronales?

El entrenamiento de redes neuronales es el proceso de enseñar a una red neuronal a realizar una tarea. En principio, las redes neuronales aprenden procesando varios conjuntos grandes de datos etiquetados o sin etiquetar. Debido a que utilizan estos ejemplos, pueden procesar entradas desconocidas con mayor precisión.

En el aprendizaje supervisado, los científicos de datos proporcionan a las redes neuronales artificiales conjuntos de datos etiquetados que ofrecen la respuesta correcta por adelantado. Por ejemplo, una red de aprendizaje profundo que se entrena sobre el reconocimiento facial, en un principio, procesa cientos de miles de imágenes de rostros humanos, con varios términos relacionados con el origen étnico, el país o la emoción que describen a cada imagen.

La red neuronal aumenta lentamente el conocimiento a partir de estos conjuntos de datos, que proporcionan la respuesta correcta por adelantado. Una vez que se entrena la red, comienza a adivinar el origen étnico o la emoción de una imagen nueva de un rostro humano que nunca ha procesado.

## ¿Qué es el aprendizaje profundo en el contexto de las redes neuronales?

La inteligencia artificial es el campo de las ciencias de la computación que investiga métodos para dar a las máquinas la capacidad de realizar tareas que requieren inteligencia humana. El machine learning es una técnica de inteligencia artificial que otorga a las computadoras acceso a conjuntos de datos muy grandes y les enseña a aprender de estos datos. El software de machine learning encuentra patrones en los datos existentes y los aplica a datos nuevos para tomar decisiones inteligentes. El aprendizaje profundo es un subconjunto del machine learning que utiliza las redes de aprendizaje profundo para procesar los datos.

### **Comparación entre machine learning y aprendizaje profundo**

Los métodos tradicionales de machine learning requieren la intervención humana para que el software de machine learning funcione suficientemente bien. Un científico de datos determina de forma manual el conjunto de características relevantes que el software debe analizar. Esto limita la capacidad del software, lo que hace que su creación y administración sean tediosas.

Por otro lado, en el aprendizaje profundo, el científico de datos solo proporciona datos sin procesar al software. La red de aprendizaje profundo obtiene las características por sí misma y aprende de forma más independiente. Puede analizar conjuntos de datos no estructurados como documentos de texto, identificar qué atributos de los datos deben priorizarse y resolver problemas más complejos.

Por ejemplo, si se entrenara un software de machine learning para identificar la imagen de una mascota de forma correcta, debería seguir estos pasos:

* Encontrar y etiquetar de forma manual miles de imágenes de mascotas, como gatos, perros, caballos, hámster, loros, etc.
* Indicar al software de machine learning qué características debe buscar para poder identificar la imagen mediante la eliminación. Por ejemplo, podría contar el número de patas, comprobar la forma de los ojos, la forma de las orejas, la cola, el pelo, etc.
* Evaluar y cambiar de forma manual los conjuntos de datos etiquetados para mejorar la precisión del software. Por ejemplo, si su conjunto de entrenamiento cuenta con demasiadas fotos de gatos negros, el software identificará de forma correcta a un gato negro, pero no a uno blanco.
* En el aprendizaje profundo, sin embargo, las redes neuronales procesarían todas las imágenes y determinarían de forma automática que necesitan analizar primero el número de patas y la forma de la cara, y mirar las colas en último lugar para identificar con éxito el animal en la imagen.

**Docker**

Docker es una plataforma que proporciona software en contenedores usando virtualización a nivel de sistema operativo (Kernel Linux). Su utilidad principal es distribuir y ejecutar software contenerizado, consiguiendo portabilidad.

Docker nos permite realizar un empaquetado de aplicaciones en estructuras llamadas **contenedores**. Estos contenedores incluyen todo lo necesario para que el software se ejecute correctamente, como bibliotecas, herramientas del sistema, scripts, etc.

Los contenedores se encargan de aislar el sistema de ficheros, los procesos y la utilización de recursos del sistema. De esta forma no interfiere con sus vecinos en entornos cloud multitenant, en los que interesa que múltiples aplicaciones (tenants) compartan el mismo host para ahorrar en recursos y aumentar la eficiencia del sistema.

Un contenedor es una abstracción en la capa de aplicación que empaqueta el código y las dependencias juntos. En lugar de virtualizar toda la máquina física, los contenedores virtualizan solo el sistema operativo del host.

En los últimos años, esta tecnología ha experimentado un crecimiento impresionante. El sector tecnológico ha sabido utilizar las ventajas de los contenedores para aislar y optimizar los entornos de ejecución y han surgido sistemas de orquestación de contenedores muy populares como [**Kubernetes (K8s)**](https://aprenderbigdata.com/kubernetes/)**.** Podemos decir que los contenedores son el futuro del despliegue en el cloud.

También resuelve el problema común al que se enfrentan los desarrolladores de software: el código funciona en una máquina y no en otra. Todo el tiempo consumido en resolver estos problemas se puede evitar usando contenedores, que aseguran que el software se ejecute de la misma manera en todos los entornos.

Docker usa partes del Kernel de Linux como los namespaces para crear contenedores sobre el sistema operativo y desplegar automáticamente software. Docker se ha hecho muy popular ya que es una herramienta que facilita mucho la creación y la ejecución de aplicaciones con contenedores. Resulta muy útil en el desarrollo, proporcionando el entorno a medida (librerías, dependencias, etc.) que necesita la aplicación.

Docker CE es gratuito. También existe la versión certificada, llamada Docker EE, que incluye el soporte, así como utilidades extra.

**Componentes de Docker**

**Imagen:** es un paquete ejecutable, ligero e independiente, que contiene todo lo necesario para ejecutar cierto software. Una imagen se compone de múltiples capas, una por cada dockerfile definido.

**Docker Registry:** es el lugar para almacenar imágenes y distribuirlas fácilmente. Existen registros públicos como [Docker Hub](https://hub.docker.com/) o Docker Cloud. También se pueden almacenar en el registro local privado.

Contenedor: es una instancia en tiempo de ejecución de una imagen. Consiste en una aplicación aislada construida a partir de una o más imágenes. Los contenedores no empaquetan sistema operativo, sino que usan el del host.

Dockerfile: es un documento de texto con los comandos necesarios para ejecutar la imagen.

Arquitectura de Docker

Docker funciona con una arquitectura de tipo cliente servidor. El cliente se comunica con Docker Daemon, la pieza responsable de construir, desplegar y distribuir los contenedores. Ambos componentes pueden ejecutar en el mismo sistema o de forma remota comunicándose a través de la API.

**Componentes principales**

El cliente de Docker es la interfaz de línea de comandos que transmite las instrucciones de los usuarios. Después de haber transmitido una orden desde el cliente hasta el demonio, éste se descarga la versión que necesite del Registry, a continuación, crea el contenedor a partir de la imagen y lo ejecuta en el host.

Ventajas

Portabilidad: Permite conseguir el mismo funcionamiento en todas las máquinas y sistemas, ya que contiene todas sus propias dependencias. Resulta muy útil en las tareas de desarrollo, facilitando el entorno a medida que necesita la aplicación en cualquier ordenador. El uso de Docker incrementa el ritmo de producción de software y la productividad de los desarrolladores.

Además, al no contener el sistema operativo completo, la tecnología de contenedores es bastante más ligera que una máquina virtual (VM), por ejemplo, las generadas con VirtualBox o VMWare. Además, es compatible con [Vagrant](https://aprenderbigdata.com/introduccion-vagrant/), que nos permite gestionar los contenedores y las máquinas virtuales de una manera más flexible y unificada.

Esquema de componentes en Docker y en Máquinas Virtuales

Esto se debe a que se construye la imagen y los contenedores por capas reutilizables. En concreto, contiene solamente las librerías, los ficheros y las configuraciones necesarias para ejecutar la aplicación.

Las máquinas virtuales son más costosas de gestionar que los contenedores. Cada máquina virtual ejecuta un sistema operativo completo dentro del sistema operativo del host. Esta característica hace a los contenedores más eficientes y rápidos para ejecutar aplicaciones aisladas de su entorno.

Para conseguir el aislamiento de los contenedores, hacen uso de utilidades Linux. Para aislar los procesos se usa el concepto de namespaces y para aislar los recursos de un proceso se usan control groups (cgroups).

Para todos los que trabajamos con Big Data, es evidente que uno de los mayores retos es la configuración y puesta a punto de un ambiente funcional. Por mucho tiempo, la solución rápida para esto han sido las imágenes de sistemas para su virtualización. Sin embargo, existe una mejor solución: Los contenedores de [Docker](https://www.docker.com/).

Un contenedor Docker que tiene el ambiente Big Data más usado en la actualidad:

* Apache Hadoop (HDFS & YARN).
* Apache Spark.
* Apache Hive (on Spark).
* Python 3 y herramientas para HDFS.

Para instalar directamente desde los repositorios de Docker:

docker build -t sequenceiq/hadoop-docker:2.7.1.

docker pull sequenceiq/spark:1.6.0

**WSL**

El Subsistema de Windows para Linux (WSL) es una característica del sistema operativo Windows que permite ejecutar un sistema de archivos Linux, junto con herramientas de línea de comandos y aplicaciones de GUI de Linux, directamente en Windows, junto con el escritorio y las aplicaciones tradicionales de Windows.

Se trata principalmente de una herramienta para desarrolladores, especialmente desarrolladores web, aquellos que trabajan en proyectos de código abierto o se implementan en entornos de servidor Linux. WSL es para cualquier persona que quiera usar Bash, herramientas comunes de Linux (sed, awk, etc.) y marcos de trabajo primero de Linux (Ruby, Python, etc.), pero también disfruta del uso de herramientas de productividad de Windows.

WSL le permite ejecutar Linux en un shell de Bash con su elección de distribución (Ubuntu, Debian, OpenSUSE, Kali, Alpine, etc.). Con Bash, puedes ejecutar aplicaciones y herramientas de Linux de línea de comandos. Por ejemplo, escribe lsb\_release -a y presiona Entrar para ver los detalles de la distribución de Linux que se está ejecutando actualmente

También puede acceder al sistema de archivos del equipo local desde el shell de Bash de Linux; encontrará las unidades locales montadas en la /mnt carpeta . Por ejemplo, la unidad C: se monta en /mnt/c

### **¿Por qué usar WSL en lugar de Linux en una VM?**

WSL necesita menos recursos (CPU, memoria y almacenamiento) que una máquina virtual completa. Asimismo, WSL también te permite ejecutar aplicaciones y herramientas de línea de comandos de Linux junto con la línea de comandos de Windows, aplicaciones de escritorio y de Store, así como la posibilidad de acceder a los archivos de Windows desde Linux. Esto te permite usar las aplicaciones de Windows y las herramientas de línea de comandos de Linux en el mismo conjunto de archivos, si así lo deseas.

### **¿Por qué usar, por ejemplo, Ruby en Linux en lugar de en Windows?**

Algunas herramientas multiplataforma se crearon suponiendo que el entorno en el que se ejecutan se comporta como Linux. Por ejemplo, algunas herramientas suponen que pueden obtener acceso a rutas de acceso de archivos muy largas o que existen archivos o carpetas específicos. Esto suele provocar problemas en Windows, que a menudo se comporta de forma diferente de Linux.

Muchos lenguajes, como Ruby y Node.js, a menudo se proscriben y funcionan bien en Windows. Sin embargo, no todos los propietarios de bibliotecas Ruby Gem o node/NPM portan sus bibliotecas para que admitan Windows, y muchas tienen dependencias específicas de Linux. Esto a menudo puede dar como resultado que los sistemas que se crearon con herramientas y bibliotecas de este tipo sufran errores de compilación y, a veces, del entorno de ejecución o tengan comportamientos no deseados en Windows.

Estos son solo algunos de los problemas que han provocado que muchas personas pidan a Microsoft que mejore las herramientas de línea de comandos de Windows, lo que nos llevó a asociarnos con Canonical para permitir que las herramientas de línea de comandos nativas de Bash y Linux se ejecuten en Windows.

### **¿Qué significa esto para PowerShell?**

Al trabajar con proyectos de OSS, existen numerosos escenarios en los que es enormemente útil usar Bash desde un símbolo del sistema de PowerShell. La compatibilidad con Bash es complementaria y fortalece el valor de la línea de comandos en Windows, permitiendo que PowerShell y la comunidad de PowerShell aprovechen otras tecnologías populares.

### Acceso a la unidad C:

Los puntos de montaje de las unidades de disco duro de la máquina local se crean automáticamente y proporcionan un acceso sencillo al sistema de archivos de Windows.

**/mnt/<letra> de unidad/**

### **¿Cómo uso un archivo de Windows con una aplicación de Linux?**

Uno de los beneficios de WSL es que puedes acceder a tus archivos mediante aplicaciones o herramientas de Windows y Linux.

WSL monta las unidades fijas de la máquina que están en la carpeta /mnt/<drive> en tus distribuciones de Linux. Por ejemplo, la unidad C: se monta en /mnt/c/

Si usas tus unidades montadas, puedes editar el código (por ejemplo, C:\dev\myproj\ usando [Visual Studio](https://visualstudio.microsoft.com/vs/) / o [VS Code](https://code.visualstudio.com/)) y compilar o probar ese código en Linux accediendo a los mismos archivos a través de /mnt/c/dev/myproj.

### **¿Los archivos de la unidad de Linux son diferentes de la unidad de Windows montada?**

1. WSL controla los archivos en la raíz de Linux (es decir /, ) que se alinean con el comportamiento de Linux, incluidos, entre otros:
   * Archivos que contienen caracteres de nombre de archivo de Windows que no son válidos
   * Vínculos simbólicos creados para usuarios que no son administradores
   * La opción de cambiar atributos de archivo a través de chmod y chown
   * Distinción de mayúsculas y minúsculas en archivos o carpetas
2. Los archivos de las unidades montadas se controlan mediante Windows y tienen los siguientes comportamientos:
   * Compatibilidad para la distinción de mayúsculas y minúsculas
   * Todos los permisos se establecen para reflejar mejor los permisos de Windows

### **¿Cómo desinstalo una distribución de WSL?**

Para quitar una distribución de WSL y eliminar todos los datos asociados a esa distribución de Linux, ejecute wsl --unregister <distroName> donde <distroName> es el nombre de la distribución de Linux, que se puede ver en la lista del wsl -l comando .

Además, puede desinstalar la aplicación de distribución de Linux en la máquina igual que cualquier otra aplicación de la tienda.

### **Cambio el idioma de visualización de WSL**

La instalación de WSL intentará cambiar automáticamente la configuración regional de Ubuntu para que coincida con la configuración regional de la instalación de Windows. Si no quieres que esto pase, puedes ejecutar este comando para cambiar la configuración regional de Ubuntu una vez finalizada la instalación. Tendrá que volver a iniciar la distribución de WSL para que este cambio surta efecto.

### **Problemas con el Internet**

Algunos usuarios han informado de problemas con aplicaciones de firewall específicas que bloquean el acceso a Internet en WSL. Los firewalls que dan error son:

1. Kaspersky
2. AVG
3. Avast
4. Symantec Endpoint Protection
5. F-Secure

En algunos casos, si desactivas el firewall podrás acceder. En otros casos, simplemente tener instalado el firewall parece bloquear el acceso a Internet.

### **Acceso a un puerto desde WSL en Windows**

WSL comparte la dirección IP de Windows, ya que se ejecuta en Windows. Así pues, puedes obtener acceso a cualquier puerto en localhost; por ejemplo, si tienes contenido web en el puerto 1234, puedes obtener acceso https://localhost:1234 al explorador de Windows.

### **Copia de seguridad de mis distribuciones de WSL o moverlas de una unidad a otra**

La mejor manera de realizar copias de seguridad o mover las distribuciones es a través de los [comandos de exportación e importación](https://learn.microsoft.com/es-es/windows/wsl/basic-commands) disponibles en Windows versión 1809 y posteriores. Puedes exportar toda la distribución a un tarball mediante el comando wsl --export. Después, puede volver a importar esta distribución en WSL mediante el wsl --import comando , que puede asignar un nombre a una nueva ubicación de unidad para la importación, lo que le permite realizar copias de seguridad y guardar los estados de (o mover) las distribuciones de WSL.

Tenga en cuenta que los servicios de copia de seguridad tradicionales que copian los archivos de copia de seguridad en las carpetas de AppData (como Copias de seguridad de Windows) no dañarán los archivos de Linux.

## Prerrequisitos

Debe ejecutar Windows 10 versión 2004 y posteriores (compilación 19041 y posteriores) o Windows 11.

**Referencias**

<https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/los-3-principales-tipos-de-tecnicas-de-procesamiento-y-analisis-de-datos>

<https://aws.amazon.com/es/what-is/neural-network/#:~:text=Una%20red%20neuronal%20es%20un,lo%20hace%20el%20cerebro%20humano>

<https://aprenderbigdata.com/introduccion-docker/#:~:text=Docker%20es%20una%20plataforma%20que,ejecutar%20software%20contenerizado%2C%20consiguiendo%20portabilidad>

<http://blog.leonelatencio.com/contenedor-docker-herramientas-de-big-data/>

<https://learn.microsoft.com/es-es/windows/wsl/faq>

<https://learn.microsoft.com/es-es/windows/wsl/install>