



Tutorial de manejo de Texas Advanced Computer Center (TACC)

Universidad Nacional De Colombia, Sede de La Paz

Programación en lenguajes Estadísticos
Profesor Jose Francisco Ruiz Muñoz

Alvarez Martinez Jairo Alonso, Gutierrez Navarro Ronald Andres,
Palacios Rico Jorge , Torres Bertel Luz Elenis,
Torres Martinez Daniel David

July 1, 2022

Resumen

En el documento se profundizarán los conceptos de un supercomputador, una GPU, las ventajas de las GPU en el análisis de datos, las características de la GPU NVIDIA A100, las características del supercomputador Lonestar 6, qué es Pytorch y por último un acercamiento a la plataforma de Texas Advanced Computing Center, al trabajar en una supercomputadora se señala el paso a paso para ejecutar programas en los lenguajes de programación de R y Python, esto para comparar el proceso de ejecución de los código fuente en una computadora tradicional y una como Lonestar6, por lo que se evidenció que el tiempo en procesar grandes cantidades de datos fue poca mientras que en un computador de mesa de rendimiento medio fue mucha.

Palabras claves: Lonestar6, TACC, R, Python, supercomputador, Pytorch

Abstract

In the document we will deepen the concepts of a supercomputer, a GPU, the advantages of GPUs in data analysis, the characteristics of the NVIDIA A100 GPU, the characteristics of the Lonestar 6 supercomputer, what is Pytorch and finally an approach to the Texas Advanced Computing Center platform, working on a supercomputer, the step by step process to execute programs in R and Python programming languages is pointed out, this to compare the process of executing the source code on a traditional computer and one like Lonestar6, so it was evident that the time to process large amounts of data was little while in a medium performance desktop computer was a lot.

keywords: Lonestar6, TACC, R, Python, supercomputer, Pytorch

1 TACC

1.1 ¿Qué es un supercomputador?

"La supercomputación es una forma de computación de alto rendimiento que resuelve o calcula utilizando una computadora potente, una supercomputadora, lo que reduce el tiempo total de resolución. La supercomputadoras son ordenadores o computadoras de alto desempeño, es decir, son extremadamente potentes y capaces de realizar tareas de cálculo a una velocidad sorprendente que equivale a cientos de veces la velocidad de una computadora de sobremesa o laptop estándar.

Podemos decir entonces que las supercomputadoras son equipos informáticos que están compuestos por cientos de procesadores que trabajan en paralelo y en arreglos combinados, para ofrecer una velocidad y capacidad de cálculo y de procesamiento de datos sorprendentes, de tal manera, que puedan ser utilizadas para fines específicos en donde muchas computadoras trabajando simultáneamente no darían los resultados esperados por los usuarios. La velocidad de estos equipos son medidos en Teraflops que equivalen a billones de operaciones por segundo, lo que da una idea de la potencia y la velocidad de estas colosales computadoras.

Al contrario de lo que se pueda pensar, las supercomputadoras se encuentran integradas por los mismos componentes de las computadoras convencionales. Sólo que sus componentes internos, se encuentran relacionados entre sí de tal manera, que se puede obtener un rendimiento extraordinario de cada uno de ellos.

Las supercomputadoras son utilizadas ampliamente en el campo de la investigación científica, ya que se requiere manipular una enorme cantidad de datos en muy poco tiempo, por lo que el poder de procesamiento de datos incluso llega a quedarse corto en varios campos." [1]

1.2 ¿Qué es una GPU?

"Es un coprocesador dedicado al procesamiento de gráficos u operaciones de coma flotante, para aligerar la carga de trabajo del procesador /CPU central en aplicaciones como los videojuegos o aplicaciones 3D interactivas. GPU es el acrónimo de Graphics Processing Unit y representa precisamente el corazón de una tarjeta gráfica al igual que la CPU lo hace en un PC. Aparte del corazón, también es su cerebro, ya que es la encargada de realizar todos los cálculos complejos que nos permiten disfrutar de nuestros juegos en pantalla. En definitiva es la pieza de silicio que tanto AMD como NVIDIA o Intel fabrican y donde se graban los transistores, por decirlo de manera común, es el llamado chip de la tarjeta gráfica. El concepto es importante desde el punto de vista de la utilidad, ya que como hemos comentado, una GPU no es una tarjeta gráfica en sí misma.

En cambio, una tarjeta gráfica sí integra una GPU en su haber, ya que además de dicha GPU necesita un PCB, resistencias, SMD, capacitadores, controladores de voltaje, reguladores de fase, VRAM, salidas de pantalla y más." [2]

1.2.1 ¿Qué ventajas tiene su uso en análisis de datos?

"Las GPU facilitan la visualización y evaluación de datos en entornos de machine learning e IA Tanto el software con enfoque hacia la inteligencia artificial, el análisis de macrodatos, el machine learning o el deep learning basan su rendimiento en una base de datos. Cuando se dispone de millones y millones de registros es difícil interpretarlos. Para evaluar la calidad de la información, la presencia de errores o la existencia de patrones y tendencias, las GPU son la mejor opción.

El uso de aplicaciones basadas en GPU hace que la presentación de los datos en tiempo real, la visualización o georreferenciación de los datos sea una realidad. En este sentido, si la velocidad de trabajo no es uno de los factores que invitan a usar una GPU en lugar de la usual CPU, puede que la posibilidad de trabajar en mejores condiciones con la información disponible sí aporte valor." [3]

1.3 Características de la GPU NVIDIA A100

Características básicas
Chip gráfico
Modelo GPU GA100
Arquitectura Ampere
Tamaño del chip 826 mm ²
N.º de transistores 54000 millones
Fabricación 7 nm TSMC
Bloques 108
Sombreadores 6912 FP32
Núcleos secundarios 432 núcleos tensoriales
Uds. textura 432
Uds. renderizado 172
Frecuencia base 1095 MHz
Frecuencia turbo 1410 MHz

Memoria
Memoria 40 GB
Interfaz 5120 bits
Tipo memoria HBM2 a 2.43 GHz
Ancho de banda 1,555.20 GB/s

Rendimiento
Tasa texturas 609.12 GTexel/s
Tasa píxel 242.52 GPixel/s
Cómputo FP32 19.49 TFLOPS

Alimentación
Consumo 400 W
Conectores
Información de vídeo
GeekData
GeekRank 60.74 %
Diferencia de potencia +0.00 %

Datos tomados de Geektopia.

1.4 Supercomputador Lonestar 6

1.4.1 ¿Qué es?

"Lonestar6 es el sistema más nuevo de la serie Lonestar de TACC de sistemas informáticos de alto rendimiento que se implementan específicamente para ayudar a los investigadores de Texas. El sistema proporciona un conjunto equilibrado de recursos para respaldar la simulación, el análisis de datos, la visualización y el aprendizaje automático.

1.4.2 Arquitectura del sistema

Lonestar6 se compone de 560 nodos de cómputo y 16 nodos de GPU. El sistema emplea servidores Dell con procesador EPYC Milan de AMD, tecnología HDR Infiniband de Mellanox y 8 PB de almacenamiento basado en BeeGFS en hardware de almacenamiento de Dell. Además, Lonestar6 admite nodos de GPU que utilizan GPU Ampere A100 de NVIDIA para admitir flujos de trabajo de aprendizaje automático y otras aplicaciones habilitadas para GPU.

Los nodos de cómputo están alojados en cuatro gabinetes de refrigerante líquido dieléctrico y 10 bastidores enfriados por aire. Los bastidores refrigerados por aire también contienen los 16 nodos de GPU. Cada nodo de cómputo tiene dos procesadores AMD EPYC 7763 de 64 núcleos (Milán) y 256 GB de memoria DDR4. Cada nodo de GPU también contiene dos procesadores AMD EPYC y dos GPU NVIDIA A100, cada uno con 40 GB de memoria de alto ancho de banda (HBM2)."[5]

1.4.3 Características (Nodos de cómputo)

Lonestar6 aloja 560 nodos de cómputo con 5 TFlops de rendimiento máximo por nodo y 256 GB de DRAM. Especificaciones del nodo:

Nodos de cómputo
CPU: 2 procesadores AMD EPYC 7763 de 64 núcleos ("Milán")
Total de núcleos por nodo: 128 núcleos en dos sockets (64 núcleos/socket)
Frecuencia de reloj: 2,45 GHz (aumento de hasta 3,5 GHz)
RAM: 256 GB (3200 MT/s) DDR4
Almacenamiento local: partición de 144 GB /tmp en un SSD de 288 GB

Datos tomados de Lonestar6 - Texas Advanced Computing Center (TACC).

1.5 ¿Qué es Pytorch?

"PyTorch es una librería open source basada en Python, enfocada a la realización de cálculos numéricos mediante programación de tensores, lo que facilita su aplicación al desarrollo de aplicaciones de aprendizaje profundo. La sencillez de su interfaz, y su capacidad para ejecutarse en GPUs (lo que acelera el entrenamiento de los modelos), lo convierten en la opción más asequible para crear redes neuronales artificiales.

Estas redes neuronales se han convertido, quizá, en la rama más prometedora de la inteligencia artificial, siendo la base de otras tecnologías como los sistemas de traducción automática, de reconocimiento de imágenes, facial, de voz...

Originalmente desarrollado por FAER (siglas de Facebook AI Research), PyTorch ha sido a su vez una pieza fundamental en el desarrollo de relevantes aplicaciones de inteligencia artificial, como el Autopilot de Tesla y el Pyro de Uber.

Con el tiempo, y gracias a una facilidad de uso no reñida con su uso en el ámbito industrial, PyTorch se ha convertido en uno de los frameworks de Deep Learning más populares del mundo, al que sólo hacen sombra Tensorflow y Keras, ambos respaldados por el patrocinio de Google." [6]

1.6 Pasos para crear una cuenta de usuario en TACC

- Accedemos a texas advanced computing center
- Nos dirigimos a use TACC, despues seleccionamos getting started
- Ingresamos a TACC user portal para crear la cuenta
- Diligenciamos los datos personales que nos piden, por ejemplo número de telefono, correo electronico
- Esperamos que nos llegue un código de confirmación a nuestro correo, para confirmar el registro de la cuenta
- Esperar un lapso de 24 horas, para que el personal de TACC revise la solicitud de cuenta
- Dentro de unos dias recibirá información con los pasos adicionales para permitir el acceso a los recursos de TACC
- Para acceder a la supercomputadora es necesario descargar una app movil llamada "Authenticator" en la playstore para usuarios android o en la appstore para usuario IOS, la cual genera un código para validar el acceso cada 20 segundos

1.7 Pasos para transferir archivos

- Abrimos una terminal nueva de ubuntu para transferir archivos a la supercomputadora
- Buscamos la ubicacion del archivo a transferir, mediante el comando `ls`
- Lo transferimos a la supercomputadora mediante el comando `rsync -avtr(Nombre del archivo)(Usuario de TACC)@ls6.tacc.utexas.edu:/ $WORK`.

1.8 Pasos para ejecutar un archivo

- Para acceder a los recursos del supercomputador de modo interactivo usamos el siguiente comando **idev -m 32**.
- Esperar el proceso de verificación ejemplo: nombre de trabajo, host de envío válido, clave de ssh.
- Después de que su trabajo idev comience a ejecutarse aparecerá un símbolo de sistema y puede comenzar su sesión de desarrollo interactivo.
- Con el comando `ls` buscamos el archivo.
- Con el comando `cd$WORK` esto para buscar el archivo en la carpeta Work.
- Luego ejecutamos el comando `ls` y nos muestra los archivos que han sido cargados a la supercomputadora
- Con el comando `cat` imprimimos el archivo, ejemplo: `cat(Nombre del archivo)`.
- Con esto nos aparecerá en pantalla el archivo en nuestra terminal de Ubuntu.

1.9 Resultados

Una vez realizados los pasos anteriormente mencionados, realizamos un código en los lenguajes de programación de R y Python que realice un programa con opciones de:

- Crear dos matrices de 1000x1000 de distribución normal.
- Multiplicar las matrices del punto 1 100 veces.
- Mostrar el tiempo que toma ejecutar en punto 2.

La comparación de estas opciones se llevó a cabo en una computadora de rendimiento medio de las salas de cómputo de la Universidad Nacional de Colombia, sede de la Paz y en un supercomputador Lonestar6 del que tuvimos acceso mediante TACC, al multiplicar dos matrices de dimensiones grandes de manera repetitiva, el tiempo en el que procesa la función es más tardío puesto a la magnitud de los datos, a continuación resultados del tiempo en el que los códigos fuente ejecutaron las opciones de los programas

En el lenguaje de Python:

La multiplicación de matrices de distribución normal de dimensiones 1000x1000 tomó un lapso de tiempo de 4.3457 segundos en Lonestar6 y 6.75 segundos en la computadora de rendimiento medio

En el lenguaje de R:

La multiplicación de matrices de distribución normal de dimensiones 1000x1000 tomó un lapso de tiempo de 0.8523 segundos en Lonestar6 y 2.24 segundos en la computadora de rendimiento medio En conclusión

manejar grandes cantidades de datos requiere un manejo de supercomputadoras como lonestar6, dado a que en computadoras promedio el manejo de datos masivos suele tardar mucho más tiempo.

References

- [1] Noguera, B. (2015, 30 enero). Las supercomputadoras: ¿Qué son y para qué sirven? Culturación. <https://culturacion.com/las-supercomputadoras-que-son-y-para-que-sirven/>
- [2] López, J. (2021, 14 junio). Seguro que sabes lo que es una tarjeta gráfica, pero ¿qué es una GPU? HardZone. <https://hardzone.es/reportajes/que-es/gpu-caracteristicas-especificaciones/>
- [3] 5 puntos que explican porqué el GPU es mejor para Big Data y IA. (2021, 1 septiembre). El Blog de Ikoula. <https://blog.ikoula.com/es/5-puntos-que-explican-porque-el-GPU-es-mejor-para-big-data-y-IA:%7E:text=Las%20GPU%20facilitan%20la%20visualizaci%C3%B3n,en%20una%20base%20de%20datos>
- [4] Expósito, D. B. (s. f.). NVIDIA A100: características, especificaciones y precios. Geektopia. <https://www.geektopia.es/es/product/nvidia/tesla-a100/>
- [5] Lonestar6 - Texas Advanced Computing Center. (s. f.). TACC. <https://www.tacc.utexas.edu/systems/lonestar6>
- [6] Merino, M. (2020, 29 julio). Así puedes aprender a usar PyTorch, la herramienta más accesible para crear redes neuronales. Genbeta. <https://www.genbeta.com/desarrollo/asi-puedes-aprender-a-usar-pytorch-herramienta-accesible-para-crear-redes-neuronales>