ภาพรวมของ Accelerated computing (GPU) ในปี 2023

โดย มติชน มณีกาศ Head of research center, VulturePrime

VULTURE PRIME

รู้หรือไม่ GPU สามารถเร่งการประมวลผลทำให้คุณคำนวนได้ไวขึ้นมากกว่า 100 เท่า

I. หลักการทำงานของ GPU

เดิมที่แล้ว GPU ถูกออกแบบมาเพื่อช่วยประมวลผลในงานที่เกี่ยวกับรูปภาพหรือวีดิโอ แรกเริ่มจากมีความต้องการสร้างสื่อบันเทิงในรูปแบบเกมส์ ทำให้เริ่มมีการใช้งาน CPU มาประมวลผลในงานเกี่ยวกับภาพและวีดิโอ

แต่เนื่องจากข้อจำกัดของการออกแบบ CPU ทำให้ประสิทธิที่ได้มานั้นไม่เป็นที่น่าพอใจ เนื่องจาก CPU ถูกออกแบบมาเพื่อเป็นหน่วยประมวลผลกลาง ไม่ได้ออกแบบมาเพื่อประมวลผลเฉพาะทาง ทำให้เริ่มเกิดการคิดค้นหน่วยประมวลผลพิเศษสำหรับประมวลผลรูปภาพและวีดิโอขึ้นมาโดยเฉพาะ โดยให้ชื่อว่า Graphic Processor Unit (GPU)

GPU นั้นออกแบบมาเพื่อประมวลผล Matrix และการประมวลผลแบบขนาน ในความละเอียดการประมวลผลที่ต่ำกว่า CPU แต่ยังเพียงพอต่อความต้องการในการใช้งาน ต่อมาจึงเริ่มนำมาใช้ในวงการ Game entertainment หรือวงการ Visual effect สำหรับสร้างภาพยนต์ เนื่องจากประสิทธิภาพของ GPU ที่ทำได้เทียบกับ CPU นั้นสามารถทำประสิทธิภาพที่สูงกว่า 20-50 เท่า จากนั้นทำให้ GPU เริ่มเป็นที่แพร่หลายในอุตสาหกรรมดังกล่าว

II. วิวัฒนาการของ GPU

GPU ถูกใช้เรื่อยมาในอุตสาหกรรมบันเทิงจนถึงประมาณปี 2015 Google ได้เปิดตัว open-source ชื่อ Tensorflow โดยมีความสามารถในการสร้าง Neural network ซึ่งสามารถใช้ GPU ในการประมวลผล ทำให้ผลลัพธ์ในการประมวลผล Neural network ทำได้เร็วขึ้นมากกว่า 100-500 เท่า และด้วยความเข้าถึงได้ง่าย (Tensorflow-Python) และความแพร่หลายของ Tensorflow ทำให้นักวิจัยเริ่มใช้ GPU เข้ามาช่วยประมวลผลในงานวิจัยมากขึ้นเรื่อย ๆ

ในปี 2018 ในงาน Nvidia GTC 2018 (Europe) ได้มีการประกาศเปิดตัว RAPIDS open-source framework ออกแบบมาเพื่อทำให้สามารถใช้งาน GPU ในงานที่หลากหลายมากยิ่งขึ้น นอกจากงาน Visual entertainment และ AI แล้วนั้น ยังสามารถใช้ GPU มาช่วยในงานสร้างแบบคำนวนสมการคณิตศาสตร์เพื่อจำลองผลลัพธ์การทดลองวิทยาศาสตร์, ใช้ในการประมวลผลข้อมูลขนาดใหญ่ (Big data) ในหน้าที่ของการทำ Filter, Reduce หรือ Compute หรือใช้ในงานประมวลผล SQL เพื่อให้ค้นหาข้อมูลได้รวดเร็วยิ่งขึ้น

III. บทบาทของ GPU ในภาคธุรกิจ

สำหรับภาคธุรกิจนั้นได้เริ่มนำ GPU ในเกรด Data center เข้ามาใช้อย่างแพร่หลายมากยิ่งขึ้น นอกจากประสิทธิภาพที่เหนือกว่าอย่างมากและความต้องการการประมวลผลระดับสูง (High performance compute) ส่งผลให้ GPU เป็นตัวเลือกอันดับแรกอย่างไม่เป็นที่น่าสงสัย เนื่องจากประสิทธิภาพในการประมวลผลของ GPU นั้นถูกเพิ่มขึ้นมากกว่า 2 เท่าในเวลา 2 ปี

กระแสของ AI ที่เป็นที่ถูกพูดถึงมากขึ้นเรื่อย ๆ และการเปิดตัวของ Generative AI อย่าง chatGPT ทำให้ภาคธุรกิจมีความต้องการเป็นเจ้าของ AI ด้วยทรัพยากรของตัวเอง หรือบางบริษัทที่มีประสบการณ์ในการใช้งาน GPU มาก่อนหน้านี้ สามารถใช้ GPU ในหลากหลายกรณีอย่างเช่น การทำ Autonomous Vehicle, Route optimization, Fraud Detection, Network graph analytic หรือ Recommendation system เป็นต้น ส่งผลให้ GPU ยังคงมีความต้องการอย่างต่อเนื่อง

Super computer ในยุคหลังจากปี 2020 ได้ใช้ GPU มาเป็นหน่วยประมวลผลหลักแทนที่ CPU ไม่ว่าจะเป็นเครื่อง Frontier อันดับ 1 ใน TOP 500 ปัจจุบัน (03/2023) ก็ใช้ GPU เป็นหน่วยประมวลผลหลัก และ LANTA Super computer ล่าสุดของประเทศไทยก็ใช้งาน GPU ของ Nvidia ในการเป็นหน่วยประมวลผลหลัก ฉะนั้นแล้วเราคงยากที่จะปฏิเสธถึงกระแสการใช้งาน GPU ในชีวิตประจำอย่างหลีกเลี่ยงไม่ได้

IV. Conclusion

- GPU นั้นสามารถประมวลผลเฉพาะทางได้เร็วกว่า CPU 100-500 เท่า
- ความต้องการในการประมวลมากขึ้น ส่งผลให้ GPU เป็นที่ต้องการอย่างต่อเนื่อง
- GPU เริ่มทดแทนการประมวลผลในงานทั่วไปมากขึ้น เช่นงานค้นข้อมูล Database หรือ SQL
- GPU ใน 4-5 ปีที่ผ่านมา ถือเป็นช่วง Early adopters หลังจากนี้จะเริ่มเป็นช่วง Early majority

ใครอยากลองเล่น GPU ลองรัน code ใน Google colab ลิ้งค์ด้านล่าง

เป็น Google colab ตอนใช้สอน workshop ในงาน Techsuace Global Summit 2022

https://link.vultureprime.com/nvidia-colab