



MASHINALI O'QITISHNI TASHKIL QILISH TEXNIK QURILMALAR

Andijon davlat universiteti, Axborot texnologiyalari kafedrasida o'qituvchi

M.U.Olimov

Kalit so'zlar: *Chuqur o'rganish, NVIDIA DGX stantsiyasi, Edge XT ish stantsiyasi, Deep Learning Workstation, Azure GPU VMlari, Google Cloud TPU*

Ключевые слова: *Глубокое обучение, станция NVIDIA DGX, рабочая станция Edge XT, рабочая станция глубокого обучения, виртуальные машины Azure GPU, Google Cloud TPU*

Keywords: *Deep Learning, NVIDIA DGX Station, Edge XT Workstation, Deep Learning Workstation, Azure GPU VMs, Google Cloud TPU*

Chuqur o'rganish (DL) ish stantsiyasi - bu intensiv hisoblash AI va chuqur o'rganish ish yuklarini qo'llab-quvvatlaydigan maxsus kompyuter yoki server. U bir nechta grafik ishlov berish birliklaridan (GPU) foydalanish orqali an'anaviy ish stantsiyalariga qaraganda ancha yuqori ishlashni taklif qiladi.

Bir necha yil oldingi bilan solishtirganda, ma'lumotlar faniga va sun'iy intellektga bo'lgan talab keskin oshdi, bu esa katta hajmdagi ma'lumotlar va murakkab chuqur o'rganish jarayonlarini qayta ishlashga qodir bo'lgan mahsulotlarni ishlab chiqishga turtki bo'ldi. Ko'pgina ma'lumotlar fanlari loyihalarida xavfsizlik bilan bog'liq muammolar mavjud, bu esa ma'lumotlarni bulutga ko'chirishni qiyinlashtiradi. Bu mahalliy ma'lumotlar markazi doirasidagi AI ish yuklarini ko'paytirishga qodir bo'lgan ixtisoslashgan mahalliy ish stantsiyalari bozorini kuchaytirdi.

Chuqur o'rganish bo'yicha ish yuklarini joylashtirishning to'rtta varianti mavjud:

- An'anaviy bulut provayderlari - bularga Amazon Web Services (AWS), Microsoft Azure va Google Cloud kabi yirik sotuvchilar kiradi.

- Chuqur o'rganish uchun maxsus bulut provayderlari — bular dasturiy ta'minot imkoniyatlari va GPU misollariga e'tibor qaratish kabi chuqur o'rganish ish oqimlarini qo'llab-quvvatlash uchun maxsus mo'ljallangan bulutli takliflardir. Bunga misol Paperspace.

- Oldindan qurilgan, mahalliy chuqur o'rganish serverlari — chuqur o'rganish ish stantsiyalari NVIDIA (masalan, DGX tizimlari) kabi kompaniyalarda mavjud.

- Xususiy DL ish stantsiyalari - bular tashkilot tomonidan noldan qurilgan.

Deep Learning Workstation-ning variantlari:

Bu erda mavjud bo'lgan eng yaxshi chuqur o'rganish ish stantsiyalari mavjud.

NVIDIA DGX stantsiyasi

Bu NVIDIA'ning NVLink texnologiyasiga asoslangan, sakkizta Tesla V100 grafik protsessorli sun'iy intellekt uchun maxsus ishlab chiqilgan birinchi ish stantsiyasidir. U 1 petaFLOPS ishlashiga erisha oladi, bu an'anaviy serverdan yuzlab marta ko'pdir. Ish stantsiyasi ixcham (u stol ostiga sig'ishi mumkin) va suvga asoslangan sovitishdan foydalangan holda jim.



Har bir NVIDIA DGX yechimi singari, DGX Station ham NVIDIA GPU Cloud Deep Learning Software Stack tomonidan quvvatlanadi. Bu sizga DL modellarini sozlashda tezda takrorlash imkonini beradi. Keng ko'lamlı treningni yoqish uchun modelni DGX ma'lumotlar markazida osongina joylashtirishingiz mumkin.

NVIDIA DGX Station integratsiyalashgan dasturiy va apparat yechimini qidirayotgan tashkilotlar uchun javob beradi. NVIDIA qo'llab-quvvatlashni taklif qiladi va yuqori ishlash uchun to'liq stekni optimallashtirishga yordam beradi.

Tegishli kontent: NVIDIA DGX qo'llanmamizda qo'shimcha, kuchliroq DGX modellari haqida ko'proq bilib oling.

Lambda Labs GPU ish stantsiyasi

Bu ikki va to'rtta grafik protsessorga ega o'rta darajadagi ish stantsiyasidir. Bu mashinani o'rganish modellarini mahalliy darajada o'rgatmoqchi bo'lgan kichikroq jamoalar va individual muhandislar uchun javob beradi.

Lambda Labs GPU ish stantsiyasi 4 tagacha ma'lumot markazi grande NVIDIA GPU (jumladan, A4000, A5000 va A6000), AMD Threadripper yoki Intel Core i9 protsessorlari, 1 TB gacha xotira va 61 TB gacha tashqi xotirani ta'minlaydi.

Lenovo P seriyali ish stantsiyalari

Bu yuqori darajadagi chuqur o'rganish ish oqimlarini taqdim eta oladigan bir qator ish stantsiyalari. Lenovo AI ish stantsiyalari ma'lumotlarni tayyorlash, o'qitish modellari va natijalarni vizualizatsiya qilish kabi chuqur o'rganish ish oqimlarini tezlashtirishga imkon beradi. Siz ilg'or NVIDIA GPU-lar yordamida to'liq tahliliy va ma'lumotlar fanining ish oqimlarini bajarishingiz mumkin. Seriyadan uchta variant mavjud:

- ThinkStation P340 Tiny - chekka asoslangan AI xulosasi uchun mo'ljallangan. i9-10900K protsessor, 32 GB tizim operativ xotirasi, 5 GB operativ xotiraga ega NVIDIA Quadro P2200 GPU va 1 TB xotira bilan birga keladi.

- ThinkStation P520 — chekka hisoblash va AI modellarini ishlab chiqish uchun. Xeon W-2295 protsessori, 256 GB gacha tizimli operativ xotira, 48 GB operativ xotiraga ega NVIDIA Quadro RTX 8000 gacha GPU tanlash, 6 TB gacha saqlash.

- ThinkStation P920 - o'qitish modellari uchun yuqori darajadagi chuqur o'rganish ish stoli. Dual Intel Platinum protsessor, 28 yadrogacha, 1 TB gacha tizim operativ xotirasi, 48 GB operativ xotiraga ega NVIDIA Quadro RTX 8000 gacha GPU tanlash, bortda 4 TB gacha saqlash va qo'shimcha drayverlarga 36 TB gacha qo'shish imkoniyati bilan birga keladi. .

Edge XT ish stantsiyasi

Ushbu ish stantsiyalari qatori 3D tasvir va animatsiyani ko'rsatish imkonini beradi. Bu Kibana yoki Elastic yordamida 3D modellashtirish, video kodlash va ma'lumotlarni vizuallashtirish uchun foydalidir. U ikkita nashrda chiqariladi:

- Edge XTa - AMD Ryzen Threadripper PRO 3995WX protsessor, 256 GB gacha tizim operativ xotirasi, 16 Gb tezkor xotiraga ega NVIDIA RTX A4000 GPU, 1 TB xotira bilan birga keladi.

- Edge XTA - yuqoridagi kabi, Intel i9-10920X CPU bilan.

3XS tomonidan ma'lumotlar fanining ishchi stantsiyalari

3XS Systems kompaniyasining Data Science ish stantsiyalari NVIDIA RTX GPU tezlatgichlari bilan quvvatlanadi. Ma'lumotlar bo'yicha olimlar 3XS ish stantsiyalaridan dastlabki bosqichlarda, korxonada miqyosidagi o'quv uskunasiga o'tishdan oldin foydalanishlari mumkin.

NVIDIA tomonidan boshqariladigan Data Science Workstations NVIDIA CUDA-X AI asosida yaratilgan dasturiy ta'minotni taklif etadi, unda o'n beshdan ortiq ma'lumotlarni qayta ishlash va mashina o'rganish kutubxonalarini mavjud. Bular hisoblash ilovalariga NVIDIA GPU bilan ishlaydigan hisoblash platformasidan foydalanish imkonini beradi.

3XS har biri boshqa GPUga asoslangan ish stantsiyasining uchta asosiy nashrini taqdim etadi - 48 Gb tezkor xotiraga ega Quadro RTX 8000, 24 Gb tezkor xotiraga ega Quadro RTX 6000 va 32 Gb tezkor xotiraga ega Quadro GV100 va ikki tomonlama aniqlik.

Bulutdagi Deep Learning Workstation ning variantlari

AWS GPU misollari

AWS Deep Learning AMI (DLAMI) bulutli chuqur o'rganish uchun oxirigacha yechimlarni taqdim etadi. Xizmat Amazon EC2 mintaqalarining ko'pchiligida mavjud bo'lgan moslashtirilgan mashina nusxasini taklif qiladi.

AWS DLAMI tarkibiga NVIDIA cuDNN, NVIDIA CUDA va mashhur chuqur o'rganish ramkalarining so'nggi versiyalari kiradi. Siz uni bir nechta namuna turlari bilan ishlatishingiz mumkin, jumladan, faqat protsessorli kichik nusxalar va yuqori quvvatli multi-GPU nusxalari:

- Amazon EC2 P3 nusxalari - 8 tagacha NVIDIA Tesla V100 GPU.
- Amazon EC2 G3 nusxalari - 4 tagacha NVIDIA Tesla M60 GPU.
- Amazon EC2 G4 nusxalari - 4 tagacha NVIDIA T4 GPU.
- Amazon EC2 P4 nusxalari - 8 tagacha NVIDIA Tesla A100 GPU.

Azure GPU VMlari

Azure GPU-optimallashtirilgan virtual mashinalar (VM) seriyasini taklif etadi. Azure GPU VM-lari bir nechta o'lchamlar, jumladan, bir nechta, bitta yoki kasrli GPUlar bilan foydalanish uchun mavjud. Ushbu o'lchamlar, ayniqsa, grafikalar va hisoblashni talab qiladigan ish yuklari uchun mo'ljallangan. Bu erda bir nechta variant mavjud:

NCv3 seriyali va NC T4_v3 seriyali

Ikkala seriya ham, ayniqsa, hisoblash intensiv GPU tezlashtirilgan ilovalar uchun optimallashtirilgan. Masalan, OpenCL va CUDA asosidagi simulyatsiyalar, chuqur o'rganish va AI. Bu seriyalar orasidagi asosiy farqlar:

- NC T4 v3-seriyasi - NVIDIA Tesla T4 GPU hamda AMD EPYC2 Rim protsessoriga ega bo'lgan xulosa ish yuklariga qaratilgan.
- NCv3 seriyali - NVIDIA Tesla V100 GPU xususiyatiga ega yuqori unumli hisoblash (HPC) va AI ish yuklariga qaratilgan.

ND A100 v4 seriyali

Seriya chuqur o'rganish bo'yicha treninglarni kengaytirish va HPC ilovalarini tezlashtirishga qaratilgan. U 200 Gigabit InfiniBand MDR orqali ulangan har biri 40 GB xotiraga ega sakkizta A100 GPU-ni taqdim etadi.

NV seriyali va NVv3 seriyali



Ushbu seriyaning o'lchamlari masofaviy vizualizatsiya, o'yin, oqim va kodlashni yoqish va qo'llab-quvvatlash uchun maxsus ishlab chiqilgan va optimallashtirilgan. Bundan tashqari, DirectX va OpenGL kabi ramkalardan foydalanadigan VDI stsenariylari uchun ideal. NV seriyali va NVv3 seriyali VMlar NVIDIA Tesla M60 GPU tomonidan quvvatlanadi.

Google Cloud GPU va TPU

Google Cloud ikkita asosiy chuqur o'rganish variantini taklif qiladi - GPU va TPU.

Google Cloud GPU

Google Cloud Compute Engine GPU'larni taklif qiladi, ularni VM nusxalariga qo'shishingiz mumkin. Google Cloud GPU'lari misollaringizda ishlaydigan muayyan ish yuklarini, xususan, ma'lumotlarni qayta ishlash va mashinani o'rganishni tezlashtirishga yordam beradi.

Google Cloud sizga NVIDIA GRID texnologiyasidan grafik talab qiladigan ish yuklari, jumladan 3D renderlash va 3D vizualizatsiya, shuningdek virtual ilovalar uchun virtual ish stantsiyalarini yaratishda foydalanish imkonini beradi. Siz NVIDIA K80, P4, P100, V100, A100 va T4 kabi bir nechta GPU-lardan foydalanishingiz mumkin.

Google Cloud TPU

GPU-lardan tashqari, Google Cloud sizga tenzor ishlov berish bloklaridan (TPU) ham foydalanish imkonini beradi. TPU - bu maxsus integral mikrosxemalar (ASIC) qurilmasi, ayniqsa mashinani o'rganish ilovalarining hisoblash talablarini bajarish uchun mo'ljallangan.

Cloud TPU mahsulotlari mashinani o'rganish bo'yicha muhandislar, ishlab chiquvchilar, tadqiqotchilar va Google Cloud'da mashina o'rganish modellarini ishga tushirishni tanlagan ma'lumotshunoslar uchun kengaytiriladigan bulutli hisoblash resurslarini taklif etadi. Cloud TPU 100 petaflopsdan ortiq unumdorlikni ta'minlaydi, masalan, 8 yadroli TPU v2 tugunidan 2048 yadroli to'liq TPU v3 tuguniga o'tish imkonini beradi.