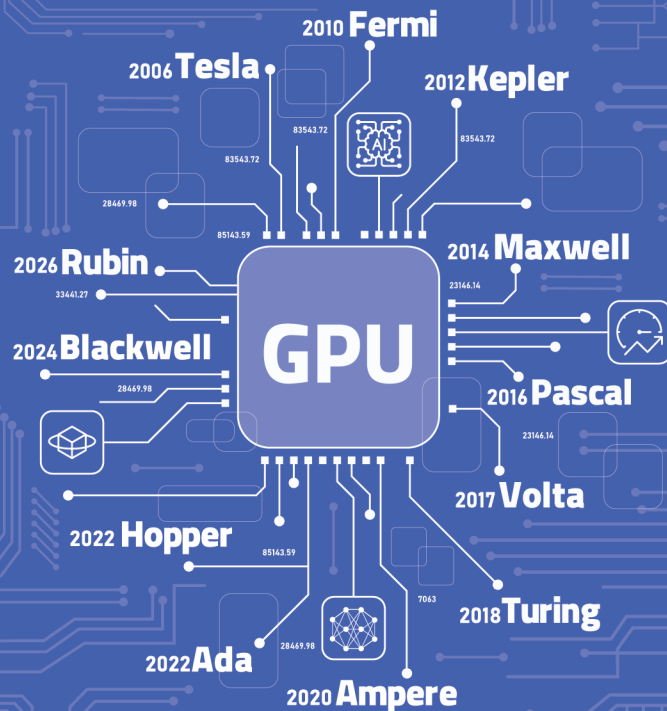


راهنمای معماری GPU

معرفی نسل‌ها، فناوری‌ها و مدل‌های پردازنده گرافیکی



Mana Cloud
مانا کلود

ManaCloud GPU Provider

سرور ابری GPU مانا کلود

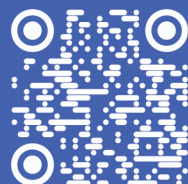
قدرت پردازش حرفه‌ای، بدون نیاز به خرید و نگهداری سخت‌افزار

سرور ابری GPU مانا کلود، برای پروژه‌های نیازمند توان پردازشی بالا و زیرساخت پایدار طراحی شده است. مناسب برای هوش مصنوعی، رندر سه بعدی، پردازش تصویر و تحلیل داده با GPUهای حرفه‌ای NVIDIA.



Think Cloud, Think Mana

ابری فکر کن، ماندگار فکر کن



ارتباط با مانا کلود

آدرس: اصفهان، میدان آزادی، دانشگاه اصفهان،
مرکز نوآوری و توسعه فناوری‌های هوش مصنوعی

تلفن: ۰۲۱-۹۱۰۹۹۹۰۷ | وبسایت: ManaCloud.com

2024 Blackwell

معماری Blackwell 2024

معماری Blackwell نسل تازه‌ای از پردازش هوش مصنوعی و محاسبات شتاب یافته را معرفی کرد. پردازنده‌های RTX مبتنی بر این معماری، با بهره‌گیری از هسته‌های Tensor Core نسل پنجم، هسته‌های RT Core نسل چهارم و فناوری رندرینگ عصبی، توان هوش مصنوعی را با پردازش گرافیکی پیشرفته ترکیب می‌کنند.

1919-2010 | David H. Blackwell



معماری Blackwell به افتخار دیوید بلکول، ریاضی‌دان و آمارشناس برجسته‌ی آمریکایی با پژوهش‌هایی در آمار، احتمال و نظریه بازی، نام‌گذاری شده است.

نوآوری‌ها و قابلیت‌های کلیدی

Tensor Core نسل پنجم: شتاب‌دهی پردازش‌های پیشرفته هوش مصنوعی
RT Core نسل چهارم: Ray Tracing سریع‌تر و Neural Graphics پیشرفته‌تر
Neural Shaders: استفاده از هوش مصنوعی در پردازش گرافیکی
RTX Mega Geomtr: افزایش جزئیات هندسی در صحنه‌های رهگیری پرتو.

GPUهای معماری Blackwell

B200 • RTX PRO 6000 • RTX PRO 4500 • RTX 5090

B200

180GB HBM3e • 8TB/s • SXM6
Training و Inference و زیرساخت‌های چند GPU.

NVIDIA RTX PRO 6000

96GB GDDR7 ECC • 1.792TB/s • PCIe 5.0 x16
AI سازمانی، Rendering، Visual Computing.

NVIDIA GeForce RTX 5090

32GB GDDR7 • 1.792 TB/s • PCIe Gen5
AI، رندرینگ و محتوای سنگین، VFX، Path Tracing.

2022 Hopper

معماری Hopper | 2022

معماری Hopper برای پردازش‌های سنگین هوش مصنوعی و HPC توسعه یافت. Transformer Engine، پردازش مدل‌های بزرگ را سریع‌تر کرده و قابلیت‌های جدید MIG و NVLink نیز اجرای بارهای کاری در مقیاس چند GPU را بهبود می‌دهند.

1906–1992 | Grace Hopper

معماری Hopper به افتخار گریس هاپر، دانشمند علوم کامپیوتر و از پیشگامان برنامه‌نویسی و توسعه‌ی کامپایلرها، نام‌گذاری شده است.



نوآوری‌ها و قابلیت‌های کلیدی

Transformer Engine: شتاب‌دهی پردازش مدل‌های Transformer و LLM
Tensor Core نسل چهارم: افزایش سرعت آموزش و استنتاج هوش مصنوعی
MIG نسل دوم: مدیریت بهینه‌تر منابع GPU در محیط‌های چندکاربره
NVLink نسل چهارم: ارتباط پرسرعت میان چند GPU

2022 Ada

معماری Ada Lovelace | 2022

معماری Ada Lovelace با تمرکز بر گرافیک حرفه‌ای، رهگیری پرتو و پردازش‌های مبتنی بر AI معرفی شد. این نسل با RT Core نسل سوم و Tensor Core نسل چهارم، برای رندرینگ، Visualization و تولید محتوا بهینه شده است.

1815–1852 | Ada Lovelace

معماری Ada Lovelace به افتخار آدا لاولیس، ریاضی‌دان انگلیسی و از نخستین چهره‌های تاریخ برنامه‌نویسی، نام‌گذاری شده است.



نوآوری‌ها و قابلیت‌های کلیدی

Tensor Core نسل چهارم: شتاب‌دهی پردازش‌های هوش مصنوعی
RT Core نسل سوم: رندرینگ سریع‌تر و Ray Tracing پیشرفته‌تر
SER: بازچینی اجرای شیدرها برای بهینه‌سازی رهگیری پرتو
Encoder AV1: فشرده‌سازی کارآمدتر ویدیو و Streaming

2020 Ampere

معماری Ampere | 2020

معماری Ampere با تمرکز بر هوش مصنوعی، HPC و پردازش حرفه‌ای معرفی شد. این معماری پایه‌ی GPUهایی مانند A100 و RTX A6000 است و از آموزش مدل‌های هوش مصنوعی تا محاسبات علمی و رندرینگ حرفه‌ای را پوشش می‌دهد.

1775–1836 | Andre-Marie Ampere

معماری Ampere به افتخار آندره‌ماری آمپر، فیزیک‌دان و ریاضی‌دان فرانسوی و از پایه‌گذاران علم الکترومغناطیس، نام‌گذاری شده است.

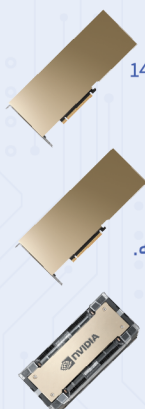


نوآوری‌ها و قابلیت‌های کلیدی

Tensor Core نسل سوم: شتاب‌دهی آموزش و پردازش هوش مصنوعی
TF32: اجرای سریع‌تر یادگیری عمیق با حفظ دقت مناسب
MIG: تقسیم یک GPU به چند نمونه مستقل
RT Core نسل دوم: شتاب‌دهی رهگیری پرتو در GPUهای گرافیکی

GPUهای معماری Hopper

H100 • H200 • H20



NVIDIA H200 NVL

141GB HBM3e • 4.8TB/s • PCIe 5.0 x16 • NVLink 900GB/s
آموزش و استنتاج LLM، هوش مصنوعی مولد و HPC

NVIDIA H100 NVL

94GB HBM3.9 • 3TB/s • PCIe 5.0 x16 • NVLink 600GB/s
استنتاج LLM، هوش مصنوعی مولد و رایانش شتاب‌یافته.

NVIDIA H20

96GB HBM4.0 • 3TB/s • SXM5
AI Inference • LLM • چندکاربره • پردازش دیتاسنتری

GPUهای معماری Ada Lovelace

L4 • L20 • L40 • L40S • RTX 6000 Ada • RTX 4090



NVIDIA L40S

48GB GDDR6 ECC • 864GB/s • PCIe Gen4 x16
Generative AI، Visualization، پردازش AI و گرافیک.

NVIDIA RTX 6000 Ada

48GB GDDR6 ECC • 960GB/s • PCIe 4.0 x16
رندرینگ، VFX، Visualization، سه بعدی و AI حرفه‌ای.

NVIDIA GeForce RTX 4090

24GB GDDR6X • 1008GB/s • PCIe Gen4 x16
GPU رندر سه بعدی، AI، VFX محلی و تولید محتوای سنگین

GPUهای معماری Ampere

A100 • A30 • A40 • A800 • RTX A6000 • RTX A5500 • RTX A5000



NVIDIA A100

80GB HBM2e • 1.935TB/s • PCIe 4.0 x16
آموزش مدل‌های HPC، AI، و مهم‌ترین GPU محاسباتی

NVIDIA RTX A6000

48GB GDDR6 ECC • 768GB/s • PCIe 4.0 x16
رندرینگ حرفه‌ای، CAD/CAE، شبیه‌سازی، پردازش گرافیکی

NVIDIA A40

48GB GDDR6 ECC • 696GB/s • PCIe Gen4 x16
Visual Computing، رندرینگ سروری، مجازی‌سازی GPU