technopc

NVIDIA VERA RUBIN

YENİ NESİL AI SÜPER BİLGİSAYAR

Yedi yeni çip, tek bir olağanüstü
AI süper bilgisayarı.



NVIDIA Vera Rubin

Ajan Tabanlı AI Çağı İçin Tasarlandı

AI, bilişimi kökten değiştiren bir hızla gelişiyor; veri merkezlerini, hesaplama gücünün gelire dönüştüğü AI fabrikalarına dönüştürüyor. Akıl yürütme modelleri ve ajan tabanlı iş akışları, token talebini katlayarak altyapının performans, iş hacmi ve maliyet verimliliği sınırlarını zorluyor. Bu ortamda zekânın üretim hızı ve watt başına performans kârlılığı belirliyor. NVIDIA Vera Rubin platformu, ajan tabanlı iş yüklerinde en yüksek performans, en yüksek verimlilik ve en düşük token maliyetini sunmak için uçtan uca birlikte tasarlandı.

Başlıca Ürünler

> NVIDIA Vera Rubin NVL72

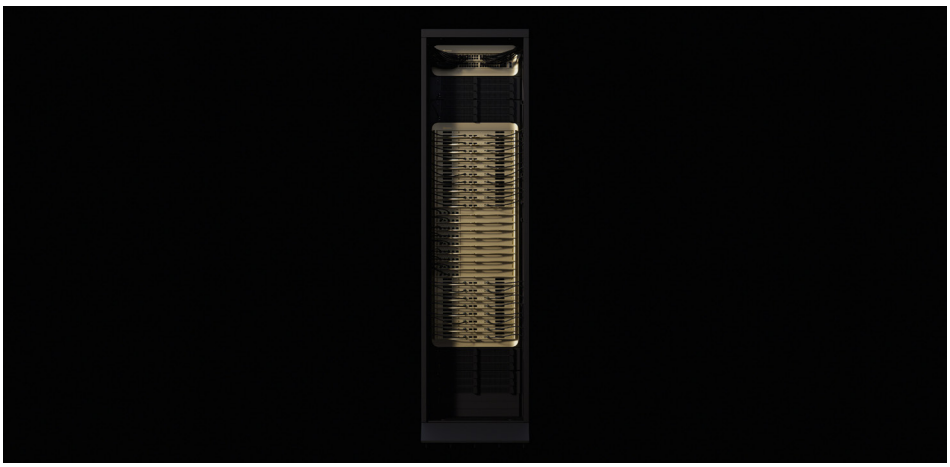
> NVIDIA HGX Vera Rubin NVL8

> NVIDIA HGX Rubin NVL8

> NVIDIA Vera Rubin NVL4

NVIDIA Vera Rubin NVL72

AI İçin Yeni Ufuklar Açıyor



NVIDIA Vera Rubin NVL72

Öne Çıkan Özellikler

> 36 NVIDIA Vera CPU

> 72 NVIDIA Rubin GPU

> 54 TB LPDDR5X bellek

> 20,7 TB HBM4 bellek

> 75 TB'a kadar hızlı erişimli bellek

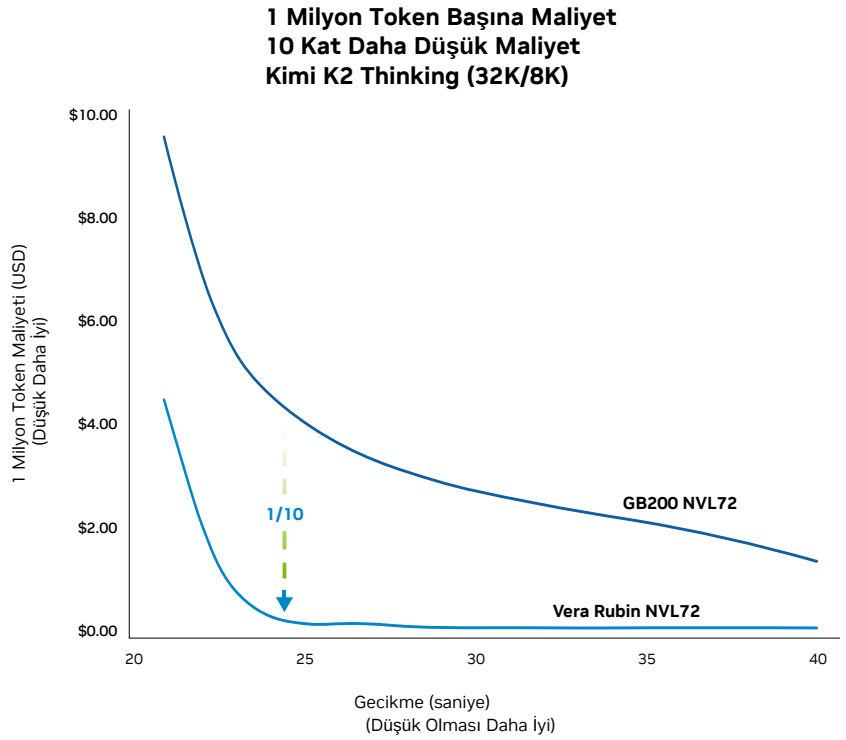
> NVLink alanı: düşük gecikmeli GPU iletişiminde saniyede 216 terabayt (TB/sn)

NVIDIA Vera Rubin NVL72; 72 NVIDIA Rubin GPU, 36 NVIDIA Vera CPU, ConnectX-9 SuperNIC, BlueField-4 DPU ve BlueField-4 STX depolama işlemcileri gibi NVIDIA'nın öncü teknolojilerini birleştirir. NVIDIA NVLink 6 anahtarıyla raf ölçeğinde işlem gücünü artırır; NVIDIA Quantum-X800 InfiniBand ve Spectrum-X Ethernet ile altyapıyı genişleterek AI endüstrisinin büyük ölçekte gelişimini destekler. NVIDIA Groq 3 LPX raflarıyla birlikte kullanıldığında trilyon parametreli modeller ve milyon tokenlık bağlamlar için yeni bir çıkarım performansı sınıfı sunar.

NVIDIA Vera Rubin NVL72, üçüncü nesil NVIDIA MGX NVL72 raf tasarımını temel alır ve önceki nesillerden sorunsuz geçiş sağlar. NVIDIA GB200 NVL72 ile karşılaştırıldığında GPU sayısının dörtte biriyle AI eğitimi, milyon token başına onda bir maliyetle AI çıkarımı sunar. Kablosuz modüler tepsi tasarımı ve 80'den fazla MGX ekosistemi iş ortağının desteğiyle bu raf ölçeğindeki AI süper bilgisayarı hızlı kuruluma ve dünya standartlarında performansa olanak tanır.

Çıkarım Maliyetlerini Azaltıyor

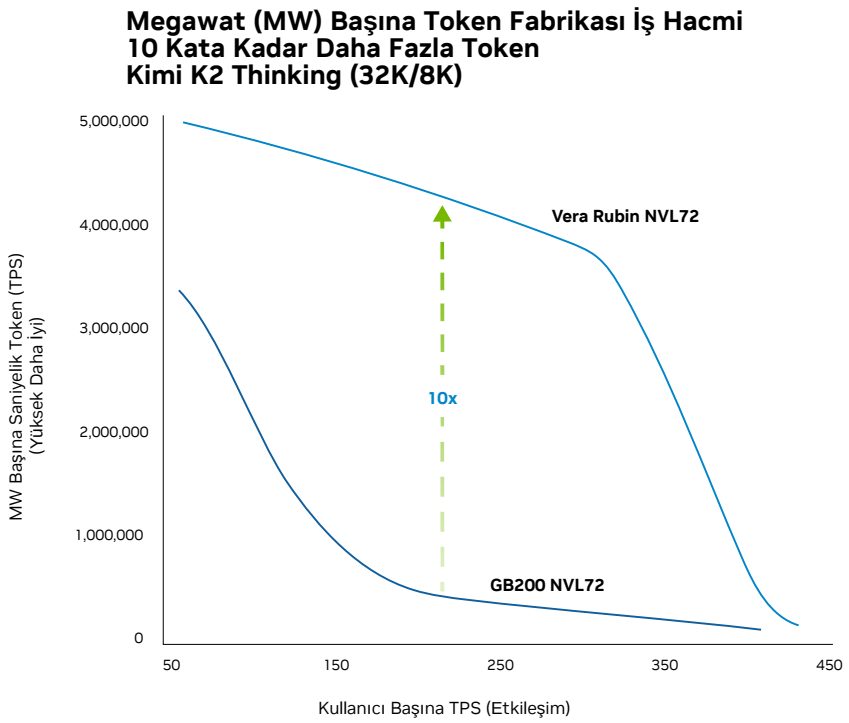
Yüksek etkileşimli ve derin akıl yürütmeye dayanan ajan tabanlı AI uygulamaları, öngörülebilir ve yönetilebilir işletme giderleri gerektirir. NVIDIA Vera Rubin, NVIDIA GB200 NVL72'ye kıyasla milyon token başına maliyeti onda bire düşürür. Maliyet verimliliğindeki bu gelişme, işletmelerin en ileri AI uygulamalarını hızla ölçeklendirmesini ve daha yüksek kârlılığa ulaşmasını sağlar.



LLM çıkarım performansı değişebilir. Milyon token başına maliyet, Kimi K2 Thinking modeli ile 32K/8K ISL/OSL değerlerinde NVIDIA GB200 NVL72 ve NVIDIA Vera Rubin NVL72 karşılaştırmasına dayanır.

Token Fabrikası İş Hacmini En Üste Çıkarıyor

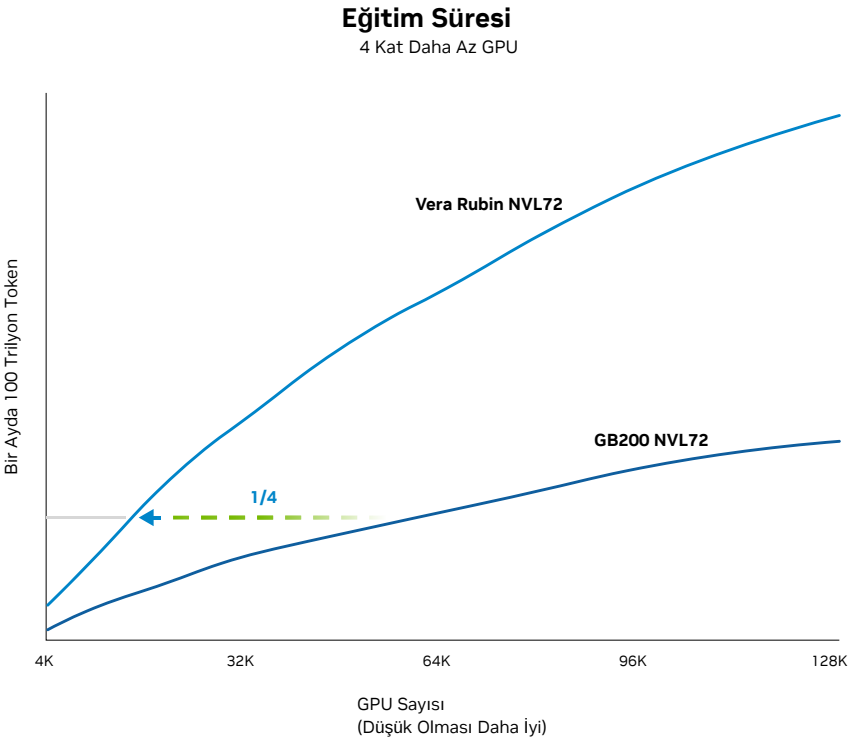
Ajan tabanlı ve akıl yürüten AI sistemlerini ölçeklendirmek, veri merkezindeki her megawattan daha fazla zekâ üretmeyi gerektirir. Kimi K2 Thinking üzerinde NVIDIA Vera Rubin NVL72, NVIDIA GB200 NVL72'ye göre megawat başına 10 kata kadar daha fazla token üretir. İş hacmi ile etkileşim arasındaki sınırı ileri taşıyan bu verimlilik artışı, işletmelerin aynı güç tüketimiyle daha çok kullanıcıya daha gelişmiş ajan tabanlı deneyimler sunmasını sağlar.



LLM çıkarım performansı değişebilir. MW başına saniyelik token değeri, Kimi K2 Thinking modeli ile 32K/8K ISL/OSL değerlerinde NVIDIA GB200 NVL72 ve NVIDIA Vera Rubin NVL72 karşılaştırmasına dayanır.

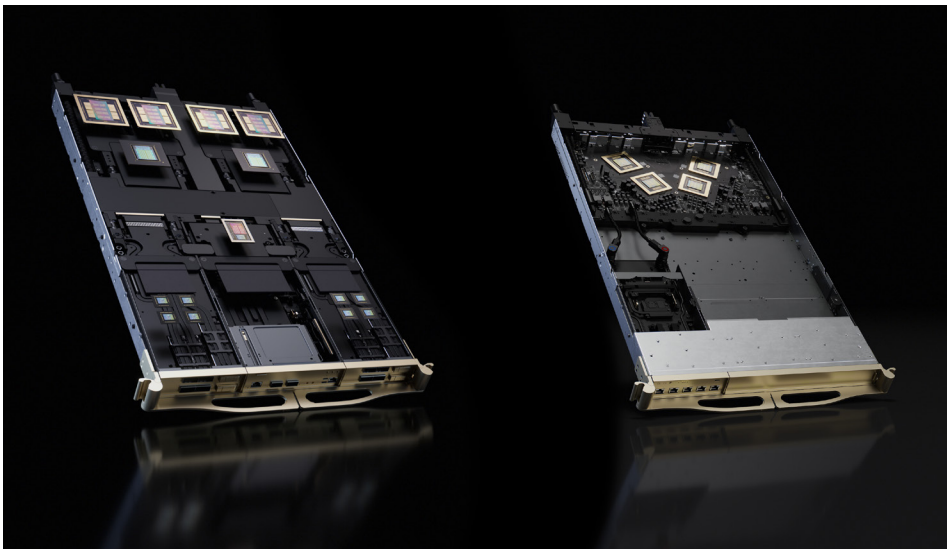
Öncü MoE Modellerinin Ön Eğitimi İçin

En ileri uzmanlar karışımı (MoE) modellerinin ön eğitimi, günümüzün en yoğun hesaplama gerektiren AI iş yüklerindendir ve buna özel tasarlanmış sistem ister. NVIDIA Vera Rubin NVL72, bu modelleri NVIDIA GB200 NVL72'nin dörtte biri kadar GPU ile eğitir. AI fabrikaları kuran kuruluşlar için bu mimari verimlilik; her çeyrekte daha fazla araştırma döngüsü, yeni nesil modellere daha hızlı geçiş ve her AI fabrikasından daha yüksek çıktı anlamına gelir.



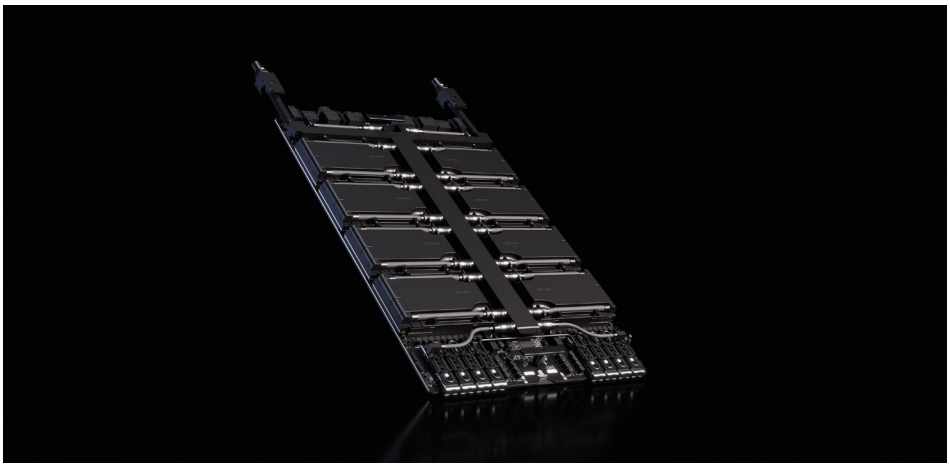
Tahmini performans değişebilir. GPU sayısı; 10 trilyon parametreli bir MoE modelinin, 100 trilyon token üzerinde bir ayda eğitilmesi esas alınarak NVIDIA GB200 NVL72 ve NVIDIA Vera Rubin NVL72 için karşılaştırılmıştır.

Üçüncü Nesil MGX Rafı



Üçüncü nesil NVIDIA MGX raf mimarisi, mekanik, güç ve soğutma tasarımındaki mühendislik yenilikleriyle AI altyapısında üretim standardını belirler. AI fabrikalarının ihtiyaç duyduğu dayanıklılık ve enerji verimliliği için tasarlanan tek genişlikli PCB tabanlı yapı; tamamen modüler, kablosuz, hortumsuz ve fansız işlem ile NVLink anahtar tepsileri sunar. Anahtar tepsileri raf çalışırken değiştirilebilir. Dinamik güç yönlendirme, gücü anlık talebe göre CPU, GPU ve NVLink anahtarlarına dağıtır. Akıllı Güç Dengeleme, raf seviyesindeki kapasitörlerle ani yük değişimlerini yumuşatır. 45°C giriş suyu sıcaklığına uygun %100 sıvı soğutma sayesinde MGX, mevcut sıvı soğutmalı veri merkezlerine kolayca uyarlanır; PUE değerini düşürerek soğutmaya ayrılan enerjinin token üretimine aktarılmasını sağlar.

NVIDIA HGX Rubin NVL8



Öne Çıkan Özellikler

> 8 NVIDIA Rubin GPU

> 2 TB HBM4 bellek

> NVLink Switch çipi üzerinden GPU'lar arasında 3 TB/sn NVLink

Her Veri Merkezi İçin AI ve HPC Gücünü Artırıyor

NVIDIA HGX platformu; NVIDIA GPU'ların, Vera CPU'ların, NVLink'in, ağ çözümlerinin ve tamamen optimize edilmiş AI ile yüksek başarımlı hesaplama (HPC) yazılım altyapılarının tüm gücünü bir araya getirir. Böylece her veri merkezine en yüksek uygulama performansını ve en kısa sürede sonuç elde etme olanağını sunar.

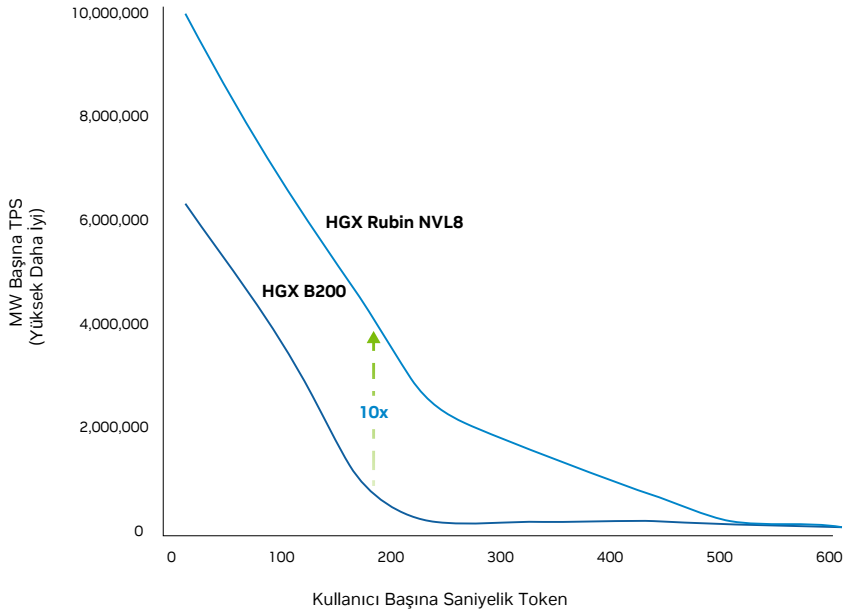
NVIDIA HGX Rubin NVL8, sekiz NVIDIA Rubin GPU'yu altıncı nesil yüksek hızlı NVLink bağlantılarıyla birleştirir. HGX B200'e kıyasla 10 kata kadar daha fazla token fabrikası iş hacmi sağlarken, dörtte biri kadar GPU ile aynı eğitim performansına ulaşır. Rubin tabanlı NVIDIA HGX sistemleri, en zorlu ajan tabanlı AI, veri analitiği ve HPC iş yükleri için tasarlanmıştır. NVIDIA HGX Rubin NVL8; NVIDIA Vera CPU'larla HGX Vera Rubin NVL8 olarak veya x86 tabanlı CPU anakartlarıyla kullanılabilir.

Yeni Nesil Ajan Tabanlı AI'yı Hızlandırıyor

HGX Rubin NVL8 ile Token İş Hacmini Artırın

Ajan tabanlı AI ve akıl yürütme modellerinin büyük ölçekte sunulması, çok yüksek çıkarım iş hacmi gerektirir. NVIDIA HGX Rubin NVL8; 400 PFLOPS ile 5,5 kata kadar NVFP4 işlem gücü, 130 TB/sn ile iki kat bellek bant genişliği ve GPU'lar arası yüksek hızlı iletişim için 24 TB/sn ile 1,7 kat NVLink Switch bant genişliği sunar. Bu mimari yeniliklerle HGX B200'e göre 10 kat daha yüksek token fabrikası iş hacmine ulaşır. AI fabrikaları böylece daha çok kullanıcıya hizmet verir, token gelirini artırır ve token başına maliyeti azaltır.

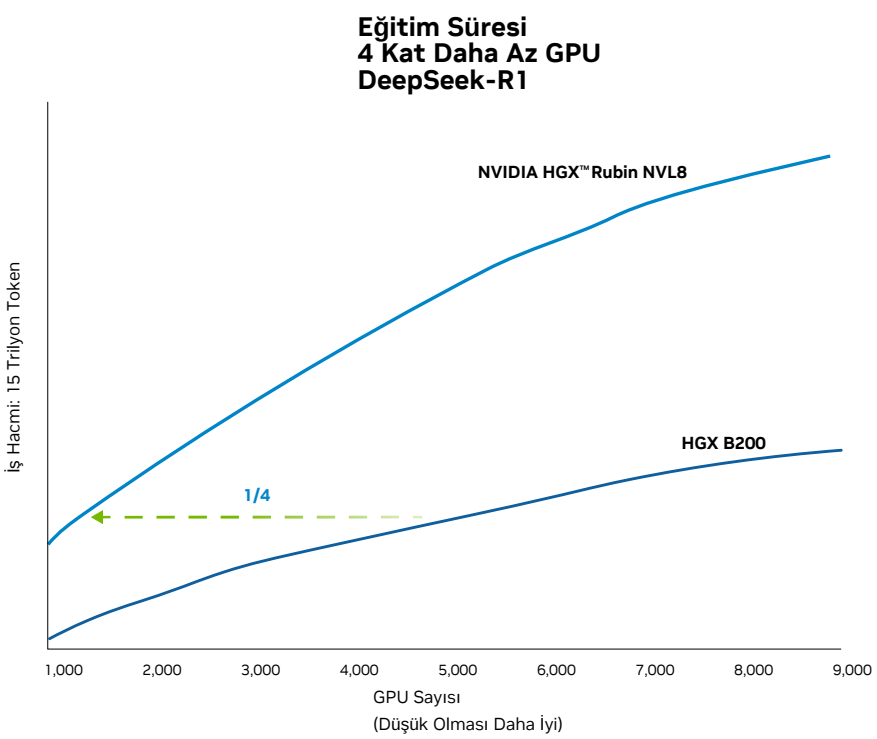
HGX Rubin NVL8 ile Token Fabrikası İş Hacmi
10 Kat Daha Fazla Token
Kimi K2 Thinking (4K/4K)



Tahmini performans değişebilir. FTL <= 500 ms, ISL = 4K, OSL = 4K koşullarındaki Kimi K2 Thinking modeli esas alınmıştır. HGX Rubin NVL8'de seyrek NVFP4; HGX B200'de yoğun NVFP4 kullanılmıştır.

4 Kat Daha Az GPU ile Yeni Nesil AI Modellerini Eğitin

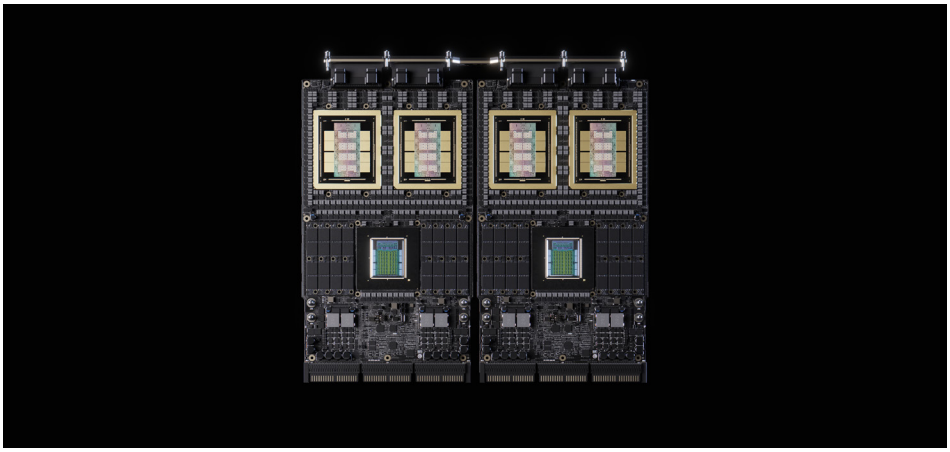
HGX Rubin NVL8, 8 GPU'lu sunucu yapısına çığır açan MoE ön eğitimi getirir ve yeni nesil ajan tabanlı AI modellerini dörtte biri kadar GPU ile eğitir. Bu, HGX B200'e kıyasla 3,8 kata kadar daha fazla NVFP4 eğitim FLOPS'u, 1,4 kat hızlı HBM bellek kapasitesi ve 1,7 kat NVLink bant genişliğiyle mümkün olur. Eğitim verimliliğindeki bu sıçrama, kuruluşların aynı altyapıyla daha fazla model eğitmesine, model geliştirme maliyetini düşürmesine ve AI yatırımlarının getirisini en üst düzeye çıkarmasına yardımcı olur.



Tahmini performans değişebilir. GPU sayısı, DeepSeek-R1 modelinin 4K dizi uzunluğunda 15 trilyon token ile ön eğitimine dayanır.

NVIDIA Vera Rubin NVL4

Otomatik Bilimsel Keşfi ve Ajan Tabanlı AI'yı Hızlandırıyor



Öne Çıkan Özellikler

> 2 NVIDIA Vera CPU

> 4 NVIDIA Rubin GPU

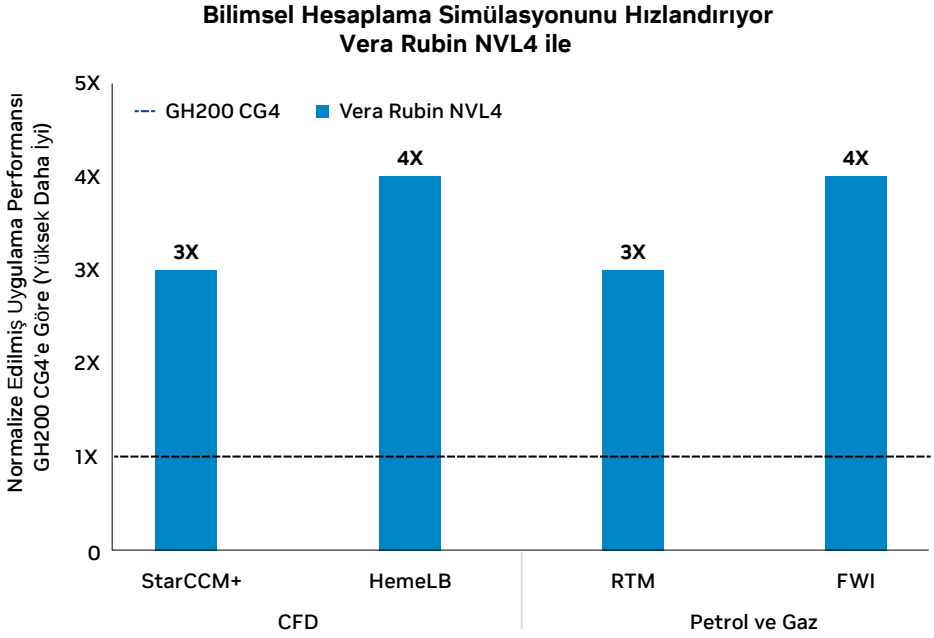
> 4,1 TB'a kadar hızlı erişimli bellek

> Bilimsel simülasyonda 4 kat daha yüksek performans

NVIDIA Vera Rubin NVL4, dört NVIDIA Rubin GPU'yu altıncı nesil NVLink ile çalışan ikinci nesil NVLink köprüsünde birbirine bağlayıp iki NVIDIA Vera CPU ile NVLink-C2C üzerinden eşleştirerek devrim niteliğinde performans sağlar. Sıvı soğutmalı NVIDIA MGX modüler sunucularla uyumludur. Grace Hopper'a kıyasla bilimsel simülasyonda 4 kata, bilimsel AI eğitiminde 6 kata ve bilimsel AI çıkarımında 8 kata kadar daha yüksek performans sunar.

Bilimsel Keşiflerde Yeni Atılımlar

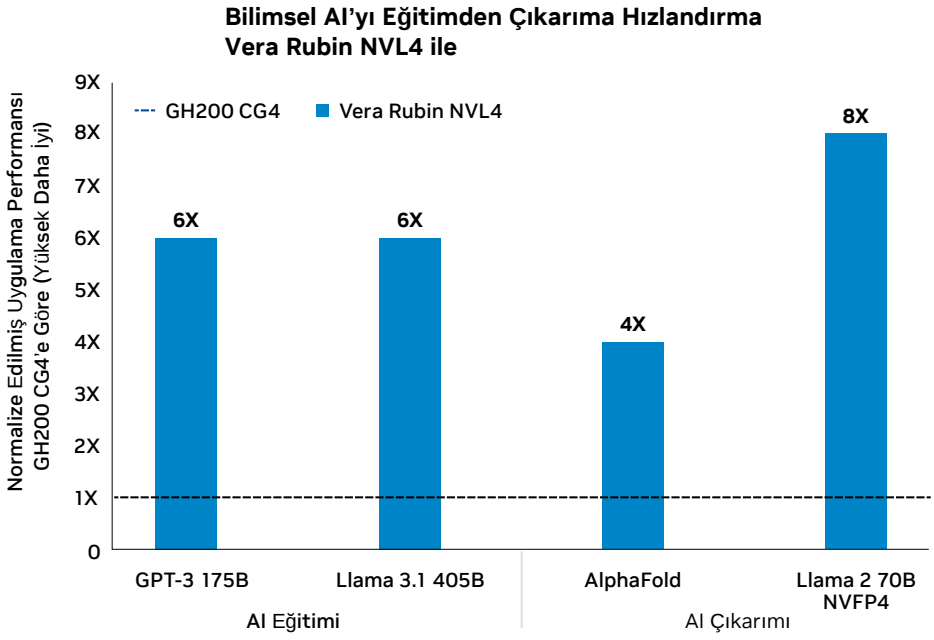
Bilimsel hesaplamanın yeni çağında, çözülmesi gereken sorunlar hem çok büyük hem de çok karmaşıktır. NVIDIA Vera Rubin NVL4, tek bir işlem düğümünde 4,1 TB'a kadar hızlı erişimli bellekle bilimsel simülasyonları hızlandırmak üzere geliştirilmiştir. Hesaplamalı akışkanlar dinamiği ile petrol ve gaz uygulamalarında NVIDIA Grace Hopper'a kıyasla 4 kata kadar uygulama performansı sunar. Böylece araştırmacılar daha büyük modeller kurabilir, simülasyonları daha hızlı çalıştırabilir ve bilimsel keşiflere ivme kazandırabilir.



Vera Rubin NVL4 için tahmini performans değişebilir. Karşılaştırma: Vera Rubin NVL4 ve GH200 CG4 (4 Grace Hopper Superchip). GH200 CG4: GPU başına 200 Gbps ağ, 700 W güç. Vera Rubin NVL4: GPU başına 800 Gbps ağ, 1.800 W güç. Star-CCM+, ticari bir hesaplamalı akışkanlar dinamiği ve çoklu fizik simülasyon aracıdır. HemeLB, kan akışı/hemodinamik modelleme için HPC biyomedikal simülasyon kodudur. RTM ters zaman göçü; FWI tam dalga biçimi ters çözümlemesidir. RTM, yer altı yapılarının görüntüsünü elde etmek için dalga denklemini ileri ve geri çözer. FWI ise sismik dalga biçimleri gözlemlerle örtüşene dek yer altı modelini yinelemeli olarak günceller.

Bilimsel AI: Eğitimden Uygulamaya

NVIDIA Vera Rubin NVL4, alan odaklı modellerde ince ayar, sentetik veri üretimi ve kullanımı, simülasyonların hızlandırılması ve geniş deney alanlarında yüksek iş hacimli çıkarım gibi bilimsel AI iş akışları için tasarlanmıştır. Tek bir işlem düğümünde 4,1 TB'a kadar hızlı erişimli bellek sunar. Grace Hopper'a kıyasla GPT-3 175B ve Llama 3.1 405B gibi modellerde 6 kata kadar hızlı eğitim, Llama 2 70B üzerinde 8 kata kadar hızlı çıkarım sağlar. Böylece araştırmacılar daha büyük modeller geliştirip bunları daha hızlı kullanıma alabilir.



Vera Rubin NVL4 için tahmini performans değişebilir. Vera Rubin NVL4 ile GH200 CG4 (4 Grace Hopper Superchip) karşılaştırılmıştır. CG4 ağ hızı 200 G/GPU, Vera Rubin NVL4 ağ hızı 800 G/GPU alınmıştır. Güç varsayımları CG4 = 700 W/GPU, Vera Rubin NVL4 = 1.800 W/GPU'dur. Hassasiyetler: GPT-3 175B FP8, Llama 3.1 405B FP8, AlphaFold BF16, Llama 2 70B NVFP4. ISL/OSL = 1.024/1.536; uçuş sırasında toplu işleme ve 1.024 parça boyutu kullanılmıştır.

100 MW AI Fabrikası Özellikleri

Sistem Özellikleri	Vera Rubin NVL72 AI Fabrikası
AI Fabrikası Ölçeği	MaxLPS ile 40 bin Rubin GPU
NVFP4 Çıkarım [2]	2 ZFLOPS
NVFP4 Eğitim [3]	1.4 ZFLOPS
FP8/FP6 Eğitim [3]	700 EFLOPS
FP16/BF163	160 EFLOPS
TF32[3]	80 EFLOPS
GPU Belleği | Bant Genişliği	12 PB HBM4 | 800 PB/s
CPU Belleği	30 PB'a kadar LPDDR5X
Hızlı Bellek	42 PB'a kadar

Teknik Özellikler

Sistem Özellikleri	NVIDIA Vera Rubin NVL72	NVIDIA Vera Rubin Superchip	NVIDIA Vera Rubin NVL4
Yapılandırma	72 NVIDIA Rubin GPUs | 36 NVIDIA Vera CPUs	2 NVIDIA Rubin GPUs 1 NVIDIA Vera CPU	4 NVIDIA Rubin GPUs | 2 NVIDIA Vera CPUs
NVFP4 Çıkarım [2]	3,600 PFLOPS	100 PFLOPS	200 PFLOPS
NVFP4 Eğitim [3]	2,520 PFLOPS	70 PFLOPS	140 PFLOPS
FP8/FP6 Eğitim [3]	1,260 PFLOPS	35 PFLOPS	70 PFLOPS
FP16/BF16[3]	288 PFLOPS	8 PFLOPS	16 PFLOPS
TF32[3]	144 PFLOPS	4 PFLOPS	8 PFLOPS
FP32	9,360 TFLOPS	260 TFLOPS	520 TFLOPS
FP64	2,400 TFLOPS	67 TFLOPS	134 TFLOPS
GPU Belleği | Bant Genişliği	20.7 TB HBM4 | 1,400 TB/s	576 GB HBM4 | 38.5 TB/s	1.1 TB HBM4 | 77 TB/s
NVIDIA NVLink	Altıncı Nesil		
NVLink Bant Genişliği	216 TB/s	6 TB/s	12 TB/s
NVLink-C2C Bant Genişliği	65 TB/s	1.8 TB/s	3.6 TB/s
CPU Çekirdek Sayısı	3.168 özel NVIDIA Olympus çekirdek | 6.336 izlek	88 özel NVIDIA Olympus çekirdek | 176 izlek	176 özel NVIDIA Olympus çekirdek | 352 izlek
CPU Belleği	54 TB'a kadar LPDDR5X	1,5 TB'a kadar LPDDR5X	3 TB'a kadar LPDDR5X
Ağ Bant Genişliği (Ölçek Genişletme)	32.4 TB/s	0.9 TB/s	1 TB/s
Toplam NVIDIA + HBM4 Çipi	1,296	30	60
GPU (100 MW) [3]	40 bin GPU		-
Güç	MaxP: 227 KW MaxQ: 187 KW MaxLPS*: 156 KW	-	-
Giriş Sıcaklığı	45°C		

*NVIDIA DSX'in bir parçası olan NVIDIA MaxLPS; sabit veri merkezi güç bütçesinde AI fabrikası çıktısını artırmak için tasarlanmış çip, termal, sistem ve yazılım teknolojilerini kapsar. Üç katmanda optimizasyon sağlar: kullanılmayan gücü geri kazanmak için dinamik GPU güç tahsisi, watt başına performansı yükselten yazılım optimizasyonu ve 45°C soğutucu giriş sıcaklığında sıvı soğutmayla PUE'yi iyileştiren termal verimlilik.

Teknik Özellikler

Sistem Özellikleri	NVIDIA HGX Vera Rubin NVL8	NVIDIA HGX Rubin NVL8
Yapılandırma	8 NVIDIA Rubin SXM ve tek soketli Vera CPU	8 NVIDIA Rubin SXM
CPU | Bellek	88 özel NVIDIA Olympus çekirdeği | 1,5 TB'a kadar LPDDR5X	x86 CPU[5]
NVFP4 Çıkarım [2]	400 PFLOPS	
NVFP4 Eğitim [3]	270 PFLOPS	
FP8/FP6 Eğitim [3]	130 PFLOPS	
FP16/BF16[3]	32 PFLOPS	
TF32[3]	16 PFLOPS	
FP32	1,040 TFLOPS	
FP64	265 TFLOPS	
GPU Belleği | Bant Genişliği	2 TB HBM4 | 130 TB/s	
NVLink Switch Bant Genişliği	24 TB/s	
NVIDIA NVLink	Altıncı Nesil	
Ağ Bant Genişliği [4]	1.8 TB/s	
Giriş Sıcaklığı	45°C	

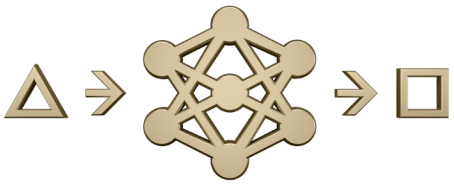
1. Özellikler, NVIDIA DSX ve MaxLPS kullanan büyük ölçekli AI fabrikasına dayanır.
2. Seyrek hesaplama özelliği.
3. Yoğun hesaplama özelliği.
4. Çift yönlü bant genişliği.
5. CPU ve bellek özellikleri OEM ürünlerine göre belirlenir.

NVIDIA Vera Rubin'in Teknolojik Yeniliklerini Keşfedin

Transformer Motoru

Rubin GPU, NVFP4 performansını doğruluktan ödün vermeden artırmak için donanım hızlandırmalı uyarlanabilir sıkıştırma sunan yeni bir Transformer Motoru (TE) içerir. Bu sayede 50 petaFLOPS'a kadar NVFP4 çıkarımı sağlar. NVIDIA Blackwell ile tam uyumluluk, önceden optimize edilmiş kodların Vera Rubin platformuna sorunsuz geçişini mümkün kılar.





Üçüncü Nesil Gizli Hesaplama

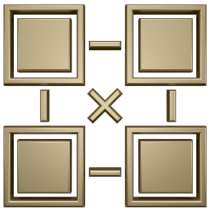
NVIDIA Gizli Hesaplama'nın üçüncü nesli, NVIDIA Vera Rubin NVL72 ile güvenliği tüm raf ölçeğine taşır. Platform; 36 NVIDIA Vera CPU, 72 NVIDIA Rubin GPU ve bunları bağlayan NVIDIA NVLink ağı genelinde birleşik, güvenilir bir yürütme ortamı oluşturur. CPU, GPU ve NVLink alanlarının tamamında veri güvenliğini korur. Kriptografik uygunluk kanıtı sağlayan doğrulama hizmetleriyle büyük ölçeği üst düzey korumayla birleştirerek en büyük özel modelleri, eğitim verilerini ve çıkarım iş yüklerini güvence altına alır.

NVIDIA Gizli Hesaplama Hakkında Daha Fazla Bilgi

Altıncı Nesil NVLink ve NVLink Switch

Altıncı nesil NVLink, 72 NVIDIA Rubin GPU'yu tek performans alanında birleştiren yüksek hızlı GPU bağlantı mimarisinde büyük bir sıçrama sunar. Rubin GPU, NVIDIA Blackwell'in performansını ikiye katlayarak GPU başına 3 TB/sn bant genişliği ve düşük gecikmeyle toplam 216 TB/sn bağlantı sağlar. Toplu işlemlerde ağ yoğunluğunu %50'ye kadar azaltan NVIDIA Scalable Hierarchical Aggregation and Reduction Protocol (SHARP) ile birlikte bu yeni nesil bağlantı, dünyanın en büyük modellerinde eğitim ve çıkarımı büyük ölçekte hızlandırır.

NVIDIA NVLink ve NVLink Switch Hakkında Bilgi





İkinci Nesil Güvenilirlik, Kullanılabilirlik ve Servis Kolaylığı (RAS) Motoru

NVIDIA Vera Rubin, gelişmiş güvenilirlik özellikleriyle raf ölçeğinde dayanıklılık sağlar. Rubin GPU'lar, kesinti olmadan proaktif bakım ve gerçek zamanlı sağlık denetimleri sunan ikinci nesil RAS motoruna sahiptir. Vera CPU'lar, SOCAMM LPDDR5X bellek modülleri ve sistem içi CPU çekirdeği testleriyle servis kolaylığını artırır. Modüler, kablosuz tepsiler NVIDIA Blackwell'e göre 90 kat hızlı montaj ve kolay bakım sağlar. Akıllı dayanıklılık özellikleri ve yazılımla tanımlanan NVLink yönlendirmesi kesintisiz çalışmayı destekleyip bakım yükünü azaltır.

NVIDIA Vera CPU

NVIDIA Vera CPU; hızlandırılmış sistemlerde veri hareketi ve ajan tabanlı akıl yürütme için tam gizli hesaplama desteğiyle tasarlanmıştır. NVIDIA GPU'larla birlikte veya analitik, bulut, orkestrasyon, depolama ve HPC iş yüklerinde bağımsız çalışabilir. NVIDIA tasarımı 88 çekirdeği, 1,2 TB/sn'ye kadar LPDDR5X bellek bant genişliğini ve NVIDIA Scalable Coherency Fabric'i birleştirerek veri ve bellek yoğun işlerde öngörülebilir, enerji verimli ve Arm uyumlu performans sunar. Yerleşik NVIDIA NVLink-C2C bağlantısı, yüksek bant genişlikli ve tutarlı CPU-GPU bellek erişimiyle sistem verimliliğini artırır.

NVIDIA Vera Hakkında Daha Fazla Bilgi



NVIDIA Mission Control ile Temel İşleri Otomatikleştirin

Bir AI fabrikası kurmak yalnızca altyapı satın almak değildir. İş yükleri, altyapı ve tesisler genelinde görünürlük, otomasyon ve dayanıklılık gerektiren sürekli bir operasyonel sorumluluktur. NVIDIA Mission Control, Vera Rubin sistemlerinin yapılandırılmasından tesis entegrasyonuna, kümelerin ve iş yüklerinin yönetimine kadar tüm süreçleri hızlandıran isteğe bağlı bir yazılım altyapısı sunar. İşletmelere soğutma ve güç olayları üzerinde daha fazla kontrol sağlayarak altyapı dayanıklılığını yeniden tanımlar.

AI'ya Hazır Kurumsal Platform

NVIDIA AI Enterprise; AI geliştirme mikroservislerini, çatılarını ve kütüphanelerini gelişmiş GPU orkestrasyonu ve altyapı yönetimiyle birleştiren, tam destekli ve üretime hazır ticari bir yazılım paketidir. Önde gelen açık kaynak araçlarını ve AI modellerini güvenle kullanıma alın; üretkenliği ve yatırımların değer üretme hızını artırın, kaynak kullanımını optimize ederek AI iş yüklerini büyük ölçekte çalıştırın.

NVIDIA AI Enterprise'ın parçası olan NVIDIA NIM, temel modellerin her bulutta veya veri merkezinde hızla devreye alınmasını ve verilerin güvenle korunmasını sağlayan, kullanımı kolay mikroservisler bütünüdür. İş süreçlerinde AI kullanan kuruluşlar, pilot aşamadan üretime sorunsuz geçmek için NVIDIA AI Enterprise'ın güvenlik, destek, yönetilebilirlik ve kararlılık özelliklerinden yararlanır.

NVIDIA Vera Rubin platformuyla NVIDIA AI Enterprise, AI'ya hazır altyapının kurulmasını kolaylaştırırken yatırımların daha hızlı değer üretmesini sağlar.

NVIDIA Launchpad'in uygulamalı laboratuvarlarında NVIDIA AI Enterprise ile AI iş akışlarını keşfedin.

TECHNOPC / KURUMSAL AI ALTYAPISI

AI GPU Sunucuları

Yüksek Performanslı Hesaplama

Kurumsal AI Sistemleri

TEKNOLOJİ ODAĞI

NVIDIA Vera Rubin

PROJENİZİ BİRLİKTE PLANLAYALIM

Kurumsal teknoloji çözümleri ve proje değerlendirmeleri

technopc.com.tr

technopc

© 2026 NVIDIA Corporation ve bağlı kuruluşları. Tüm hakları saklıdır. NVIDIA, NVIDIA logosu, BlueField, ConnectX, DSX, Grace Hopper, HGX, MGX, Mission Control, NIM, NVLink, Scalable Hierarchical Aggregation and Reduction Protocol (SHARP), Spectrum-X ve SuperNIC, NVIDIA Corporation ve bağlı kuruluşlarının ABD ve diğer ülkelerdeki ticari markaları veya tescilli ticari markalarıdır. Diğer şirket ve ürün adları, ilgili hak sahiplerinin ticari markaları olabilir. 5697927. Eki26
Teknik kaynak: NVIDIA Vera Rubin (2026). Ürün ve teknoloji markaları ilgili hak sahiplerine aittir.