

viettel
IDC



datasheet

VIETTEL CLOUD GPU



Nhà cung cấp hàng đầu Việt Nam về Dịch vụ Trung tâm Dữ liệu và Điện toán Đám mây
Website: <https://viettelidc.com.vn/>
Hotline: 1800 8088



TỔNG QUAN

- Viettel Cloud GPU là dịch vụ cung cấp máy chủ ảo tích hợp card đồ họa (GPU) của NVIDIA hướng tới phục vụ các tổ chức, doanh nghiệp có nhu cầu nghiên cứu, phát triển, vận hành các hệ thống, phần mềm ứng dụng Trí tuệ nhân tạo (AI) - Học máy (ML) như trợ lý ảo giọng nói, chấm công khuôn mặt, định danh điện tử (eKYC), camera giám sát, phân tích dữ liệu lớn; các doanh nghiệp sử dụng các ứng dụng kết xuất hình ảnh, video chất lượng cao (render), đồ họa kiến trúc, mô hình 3D, v.v.
- Viettel Cloud GPU được xây dựng và vận hành tại các Trung tâm dữ liệu đạt chuẩn Rated 3 - ANSI/TIA942-B và nhiều chứng nhận tiêu chuẩn quốc tế khác như ISO 9001:2015, ISO 27001:2013, ISO 20001:2018, ISO 27017:2015.

ĐIỂM NỔI BẬT

- Khởi tạo nhiều máy chủ ảo Cloud GPU với NVIDIA Tesla T4 và NVIDIA A30 kết nối với nhau để phân chia workload, tăng hiệu suất xử lý cho các tác vụ.
- Viettel Cloud GPU tương thích với bộ công cụ CUDA Toolkit; hỗ trợ đầy đủ các frameworks & ứng dụng phổ biến cho AI-ML-Data (TensorFlow, Caffe, PyTorch, v.v.).
- Kết hợp và kết nối với các dịch vụ trong hệ sinh thái sản phẩm của Viettel IDC
- Viettel Cloud GPU được trang bị công nghệ Intel Xeon Gold Gen 4 AMX và sử dụng ổ đĩa lưu trữ NVME U.2 với tốc độ IOPS vượt trội giúp tối ưu hóa việc truy xuất và lưu trữ dữ liệu, từ đó nâng cao hiệu suất của các mô hình AI.



Hình 1: Minh họa các lớp hạ tầng GPU, ảo hóa và ứng dụng



TÍNH NĂNG CHÍNH

Kết nối nhiều hạ tầng Cloud GPU

- Khởi tạo nhiều máy chủ ảo Cloud GPU kết nối với nhau để phân chia workload, tăng hiệu suất xử lý cho các tác vụ như Distributed Training.

Tăng tốc xử lý, tiết kiệm thời gian

- Các GPU giúp tăng tốc các tác vụ đào tạo và suy luận của mô hình AI từ 10-30 lần so với chỉ sử dụng CPU, giảm thời gian xử lý từ Ngày xuống còn Giờ, Phút, Giây, khách hàng không còn phải chờ đợi lâu.

Sẵn sàng sử dụng, đáp ứng tốt nhiều nhóm nhu cầu

- Viettel Cloud GPU được cài đặt sẵn hệ điều hành (Windows/Ubuntu/CentOS), NVIDIA Driver và bộ CUDA Toolkit. Dịch vụ hỗ trợ đầy đủ các frameworks và ứng dụng Machine Learning phổ biến (TensorFlow, Caffe, PyTorch),...

Cân bằng tải & tường lửa

- Tùy chọn mua add-on các thành phần Loadbalancer, Firewall với chi phí hợp lý để đảm bảo hiệu suất & bảo vệ hệ thống, an toàn dữ liệu.

Hỗ trợ kết nối băng thông tốc độ cao

- Dễ dàng kết nối với các cụm tài nguyên có sẵn thông qua dịch vụ Viettel Hybrid Connect với băng thông tối đa lên tới 10Gbps, đảm bảo tốc độ đường truyền ở mức hiệu năng tốt nhất.

Khả năng kết hợp

- Viettel Cloud GPU và các dịch vụ tính toán (Viettel Cloud Server, Viettel Virtual Private Cloud), lưu trữ (Viettel File Storage) và sao lưu (Viettel Cloud Backup) để phục vụ tốt nhất nhu cầu sử dụng thông qua một kết nối bảo mật, an toàn.

Khởi tạo nhanh chóng, nâng cấp linh hoạt

- Viettel Cloud GPU có thể khởi tạo & nâng cấp tài nguyên linh hoạt bất kỳ khi nào khách hàng cần, không phải chờ đợi lâu như khi tự triển khai hạ tầng vật lý.



TRƯỜNG HỢP SỬ DỤNG (USE CASE)

Xử lý hình ảnh và video

- Các mô hình AI được sử dụng rộng rãi cho Nhận diện khuôn mặt, Phát hiện đối tượng và Xử lý hình ảnh đều hoạt động hoàn hảo. Nó phù hợp với các doanh nghiệp và tổ chức có nhu cầu triển khai hệ thống và ứng dụng cho đăng nhập FaceID, eKYC, Camera giám sát AI, v.v.

Trợ lý ảo & Chatbot

- NVIDIA A30 phù hợp với các mô hình AI NLP như BERT, GPT và CNN-Text, đáp ứng nhu cầu cho các ứng dụng như Chuyển văn bản thành giọng nói, Chuyển giọng nói thành văn bản, Trợ lý ảo Chăm sóc khách hàng, v.v.

AI Tạo nội dung (GenAI)

- Các mô hình AI tạo nội dung cho văn bản, hình ảnh, video và mã đều có thể được hưởng lợi từ việc sử dụng NVIDIA A30.

Xử Lý Video và Truyền Thông

- Viettel Cloud GPU hỗ trợ giải mã và mã hóa video, làm cho nó phù hợp cho các ứng dụng như streaming video, xử lý video AI, và các giải pháp truyền thông trực tuyến.

Triển khai AI với kiến trúc microservice

- Với khả năng tương thích cao với Docker và Kubernetes, Viettel Cloud GPU rất lý tưởng cho các môi trường đám mây phục vụ cho việc triển khai các ứng dụng AI và microservices.

Ứng Dụng VR/AR

- Với hiệu suất cao trong việc xử lý hình ảnh và mô hình 3D, A30 có thể chạy các ứng dụng thực tế ảo và tăng cường một cách mượt mà, mang lại trải nghiệm người dùng tốt hơn. Điển hình như trong các trò chơi VR hoặc ứng dụng AR cho giáo dục.



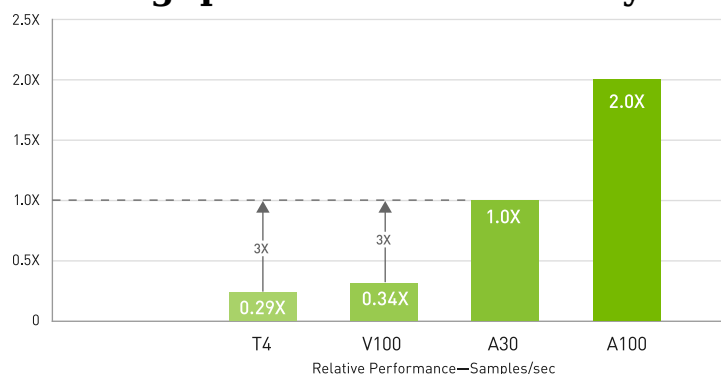
DEVICE SPECIFICATIONS

Device	GPU NVIDIA A30	QUALCOMM NPU Cloud AI 100 Standard	GPU NVIDIA Tesla T4
Interface	PCIe Gen 4	PCIe Gen 4	PCIe Gen 3
vRAM (GB)	24	16	16
Memory Bandwidth (GB/s)	933	137	300
TDP (W)	165	75	70
FP16 peak (TFLOPS)	165 330 *	175	65
INT8 peak (TOPS)	330	350	130
Interconnect	PCIe Gen4: 64GB/s Third-gen NVIDIA® NVLINK® 200GB/s**	PCIe Gen4	PCIe Gen 3
Purpose	Inference, fine-tuning	Inference	Inference, fine-tuning

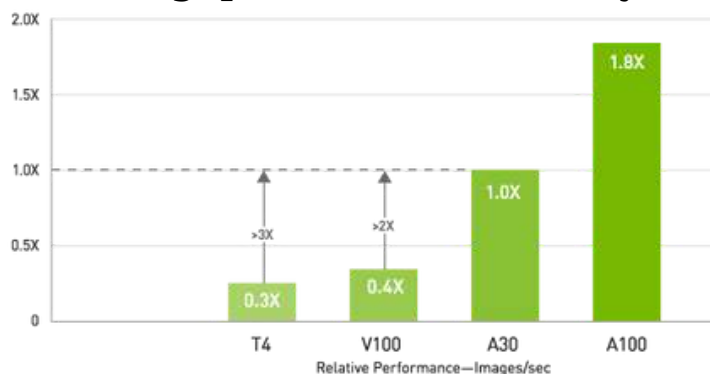
*: with sparsity

**: near future

BERT Large Inference (Normalized) Throughput for <10ms Latency



RN50 v1.5 Inference (Normalized) Throughput for <7ms Latency





TẠI SAO CHỌN VIETTEL IDC?

Đơn vị dẫn đầu trong lĩnh vực công nghệ điện toán đám mây

- Công ty TNHH Viettel - CHT (tên giao dịch: Viettel IDC) thành lập năm 2008 - nhà cung cấp dịch vụ điện toán đám mây tiên phong với hệ sinh thái đa dạng nhất Việt Nam với sứ mệnh cung cấp nền tảng tin cậy để khách hàng sáng tạo, bứt phá và chuyển đổi số một cách toàn diện. Dải sản phẩm của Viettel IDC bao gồm 50+ dịch vụ trải dài trên nhiều nhu cầu từ tính toán, lưu trữ, kết nối tới xử lý đồ họa, giám sát,...

Hạ tầng chuẩn Quốc tế

- Hạ tầng được đặt tại các Trung tâm dữ liệu của Viettel IDC đạt chuẩn TIA-942 Rated 3 Constructed Level, các chứng chỉ quốc tế ISO 9001: 2015, 50001: 2018, các chứng chỉ về bảo mật, an toàn thông tin: ISO 27001: 2013, 27017:2015 (chuyên cho các dịch vụ Cloud) và PCI DSS đảm bảo đáp ứng các tiêu chí khắt khe nhất của khách hàng về hạ tầng, chất lượng dịch vụ và an toàn, bảo mật thông tin.
Ưu điểm về hạ tầng multi-site trải dài trên ba miền Bắc Trung Nam của Viettel IDC giúp khách hàng dễ dàng lựa chọn khu vực phù hợp để thuận tiện kết nối với các dịch vụ khác trong hệ sinh thái, giảm thiểu tối đa độ trễ đường truyền.



Công nghệ hiện đại, phù hợp xu thế

- Viettel IDC cam kết áp dụng những công nghệ hiện đại, tốt nhất cho Khách hàng. Với tôn chỉ đổi mới sáng tạo là kim chỉ nam, chúng tôi lựa chọn & triển khai các công nghệ chất lượng hàng đầu, liên tục cập nhật các xu hướng, giải pháp công nghệ mới nhất cho không chỉ riêng cho Viettel Cloud GPU mà cho toàn bộ hệ sinh thái sản phẩm dịch vụ.

Đội ngũ chuyên gia và kỹ sư giàu kinh nghiệm

- Đội ngũ hỗ trợ kỹ thuật 24/7 cùng số lượng kỹ sư, chuyên gia hùng hậu với bình quân hơn 5 năm kinh nghiệm, thâm niên vượt trội và kiến thức sâu rộng trải dài trên nhiều lĩnh vực công nghệ khác nhau sẽ giúp giải quyết các bài toán của Khách hàng một cách toàn diện và hiệu quả nhất.