

Data part 2

VSCA July 2026

Ngũ

Source: Statista

Int. Finance and Smart Cities

Enterprise and Transactional Data

Scientific Research and Big Data

Cloud Computing and the Shift to Digital Storage

eCommerce

AI-Generated Content

Streaming

AI-Generated Content

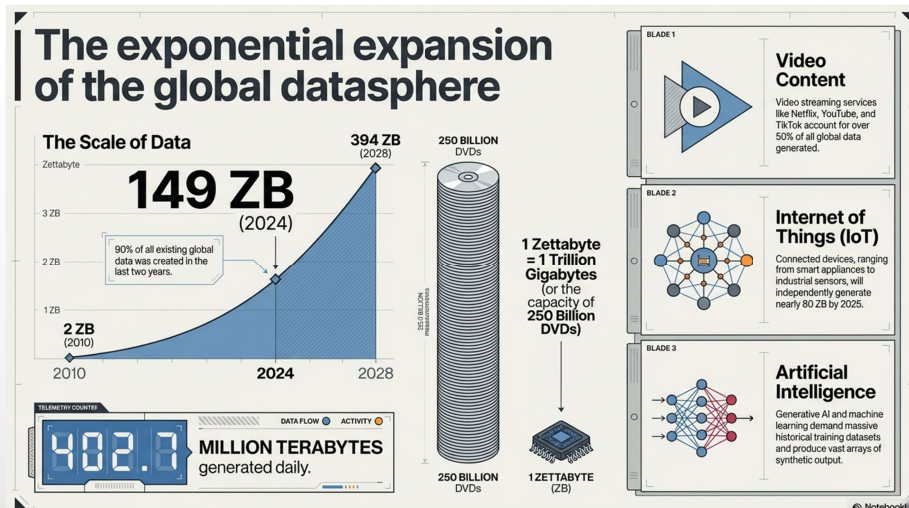
Dữ liệu ở đâu? (Where is the data?) - trên mây (Cloud), dữ liệu (Data Center) ở

Dữ liệu và AI (Artificial Intelligence)-Trí Tuệ nhân tạo

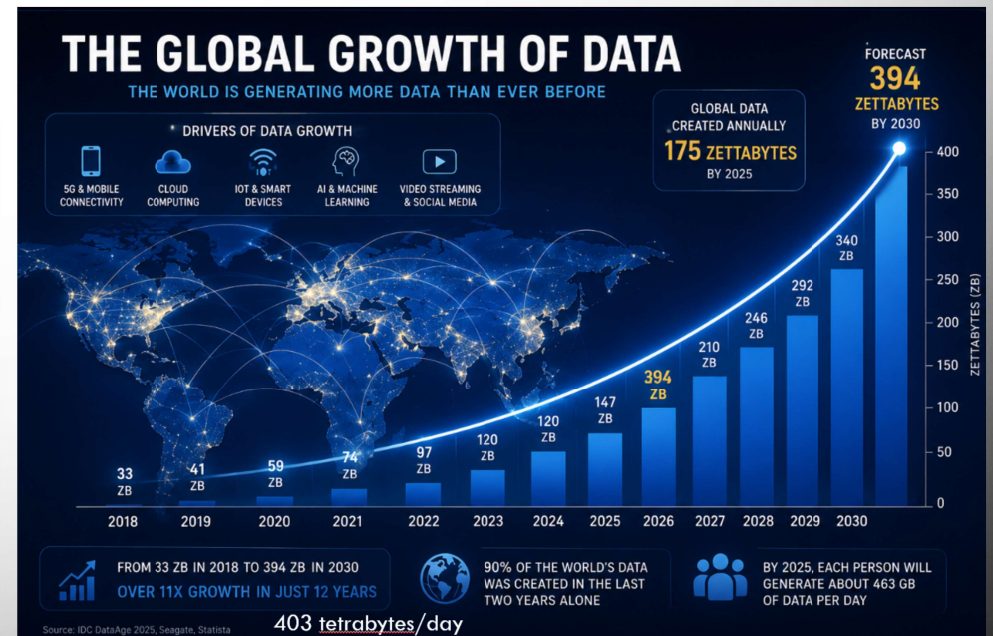
AI tạo sinh Generative-AI tự hành autonomous

AI General Intelligence(AGI)-AI Super Intelligence(ASI)

của d



Data Unit	Size (Bytes)	Size (Approx.)
Bit (b)	1	-
Byte (B)	8 bits	-
Kilobyte (KB)	1,000 bytes	10^3 bytes
Megabyte (MB)	1,000,000 bytes	10^6 bytes
Gigabyte (GB)	1,000,000,000 bytes	10^9 bytes
Terabyte (TB)	1,000,000,000,000 bytes	10^{12} bytes
Petabyte (PB)	1,000,000,000,000,000 bytes	10^{15} bytes
Exabyte (EB)	1,000,000,000,000,000,000 bytes	10^{18} bytes
Zettabyte (ZB)	1,000,000,000,000,000,000,000 bytes	10^{21} bytes
Yottabyte (YB)	1,000,000,000,000,000,000,000,000 bytes	10^{24} bytes



Dữ liệu không có cấu trúc

Là dữ liệu không có định nghĩa giúp con người hiểu và xử lý được.

Dữ liệu có cấu trúc

Dữ liệu có cấu trúc thường được định nghĩa là dữ liệu được liên kết bằng cách sử dụng các trường, và có thể được xử lý bằng máy tính. Dữ liệu có cấu trúc có thể được tổ chức theo một số cách khác nhau, chẳng hạn như:

Dữ liệu không có cấu trúc

Dữ liệu không có cấu trúc chiếm một phần đáng kể của tất cả dữ liệu trên thế giới. Loại dữ liệu này không thể được lưu trữ trong cơ sở dữ liệu theo dạng hàng-cột và nó cũng không có mô hình dữ liệu liên quan.

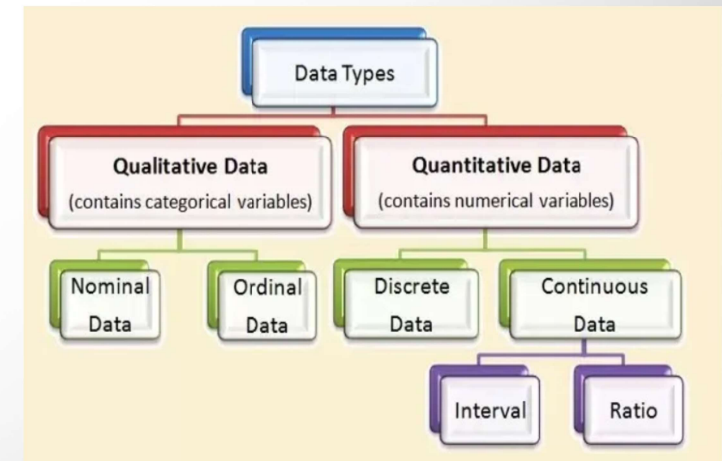
Hình ảnh, tệp văn bản, phim và tệp âm thanh, hình ảnh vệ tinh, nội dung mạng xã hội, câu trả lời từ các cuộc khảo sát mở, bản trình bày, trang web, tệp PDF và bản ghi âm từ các cuộc gọi hỗ trợ khách hàng là tất cả các ví dụ về dữ liệu không có cấu trúc.

Nominal: Dữ liệu được sử dụng để gán nhãn cho các biến không có giá trị định lượng hay thứ tự tự nhiên (ví dụ: màu tóc, giới tính hoặc tên thương hiệu).

Ordinal: Dữ liệu được phân loại theo thứ tự hoặc xếp hạng tự nhiên, nhưng khoảng cách chính xác giữa các giá trị không được biết (ví dụ: điểm chữ cái, xếp hạng mức độ hài lòng của khách hàng hoặc trình độ học vấn).

Continuous: Dữ liệu có thể đo lường được, có thể nhận bất kỳ giá trị nào trong một khoảng xác định và thường bao gồm các số thập phân (ví dụ: nhiệt độ, chiều cao hoặc thời gian chính xác).

Discrete: Dữ liệu đếm được, thường bao gồm các số nguyên riêng biệt và tách biệt (ví dụ: số lượng học sinh trong một lớp học hoặc số lượng mặt hàng trong kho).



Dữ Liệu

- Lưu trữ tại chỗ
- Thuê nơi lưu trữ
- Mây công cộng-Public cloud Data Center
- Mây riêng tư-Private cloud Data Center

Các phương pháp lưu trữ data cho các doanh nghiệp

Lưu trữ tại chỗ

Được gọi là phương pháp lưu trữ dữ liệu gốc. Đây là giải pháp dữ liệu tại chỗ thường liên quan đến các máy chủ do tổ chức sở hữu và quản lý. Nếu biết loại dữ liệu và dữ liệu đó rất quan trọng thì việc lựa chọn phương pháp này là hoàn toàn đúng đắn.

Các công ty lớn hơn có thể đặt các máy chủ này trong một trung tâm cơ sở dữ liệu riêng biệt, nhưng trong hầu hết các trường hợp, chúng là các máy trong phòng dữ liệu chuyên dụng được đặt trực tiếp ngay trong văn phòng của các doanh nghiệp.

Thuê vị trí lưu trữ

Nhiều tổ chức vẫn muốn lưu trữ dữ liệu có giá trị của họ trên các thiết bị mà họ sở hữu và kiểm soát, nhưng họ không muốn giải quyết các vấn đề đi kèm với việc quản lý các thiết bị đó. Các yêu cầu về điện năng và làm mát, cũng như tốn nhiều thời gian tích hợp các dịch vụ hay tính năng mới vào cơ sở hạ tầng CNTT, có thể gây khó khăn cho các doanh nghiệp trong việc tự triển khai cơ sở hạ tầng lưu trữ.

Tuy nhiên nếu họ đã xác định được nhu cầu lưu trữ **dữ liệu là gì**, họ sẽ nhận ra các công ty có thể đạt được lợi ích của tính linh hoạt của trung tâm cơ sở dữ liệu trong khi duy trì quyền kiểm soát hoàn toàn đối với dữ liệu của họ bằng cách thuê vị trí lưu trữ ở ngoài trung tâm dữ liệu.

Public Cloud

Các doanh nghiệp vừa và nhỏ không đủ khả năng đầu tư vào phòng dữ liệu lưu trữ dữ liệu đắt tiền.

Giải pháp doanh nghiệp của bạn tiết kiệm chi phí bằng cách chuyển dữ liệu của bạn đến một nhà cung cấp Public Cloud.

Giải pháp Public Cloud có thể dễ dàng mở rộng, cho phép bổ sung thêm tài nguyên lưu trữ hoặc tính toán khi cần thiết. Tính dễ sử dụng của đám mây cũng cho phép nhân viên truy cập dữ liệu từ hầu hết mọi nơi, đây là một lợi ích to lớn đối với các tổ chức có nhiều nhân viên làm việc từ xa. Vì vậy trước khi lựa chọn phương pháp lưu trữ này, bạn nên cân nhắc vào khả năng tài chính và mức độ quan trọng của dữ liệu là gì.

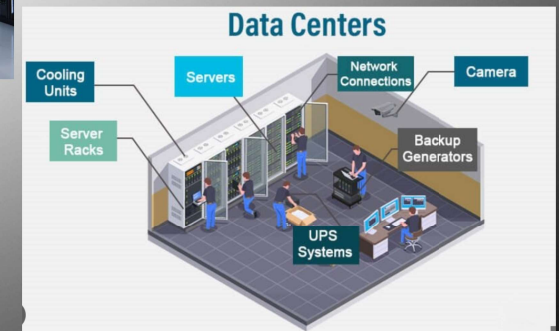
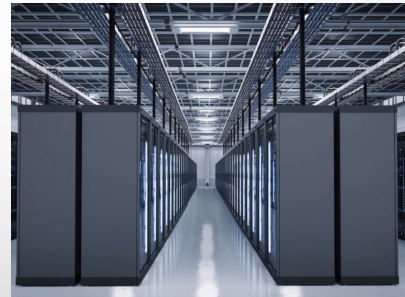
Private Cloud

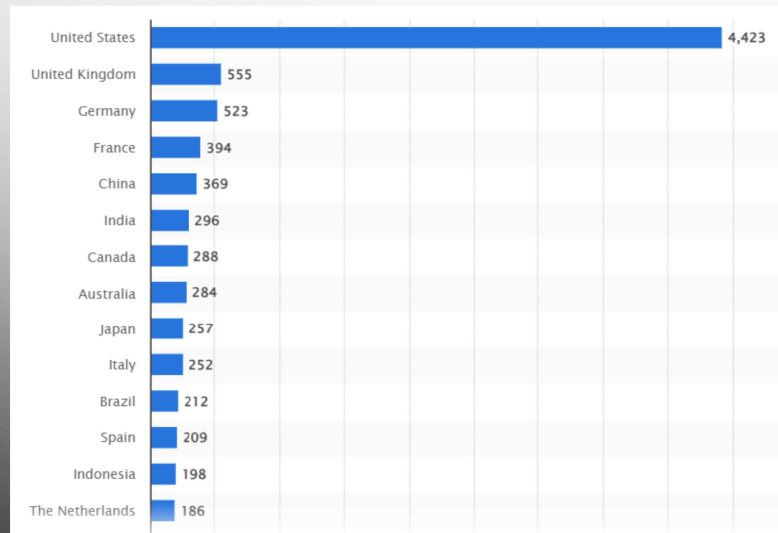
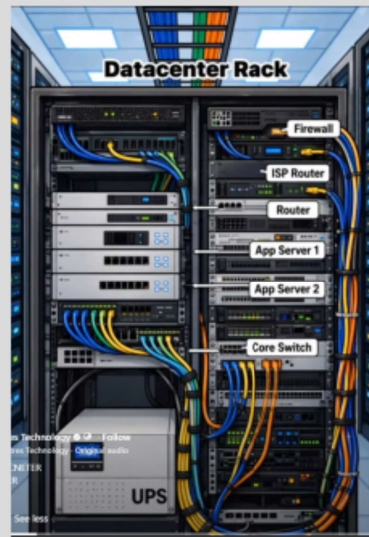
Môi trường của Public Cloud do có tính chung nên khiến việc bảo vệ dữ liệu nhạy cảm khỏi bị truy cập trái phép trở nên khó khăn. Private Cloud được triển khai thông qua một trung tâm dữ liệu ảo hóa cung cấp mức độ bảo mật cao hơn nhiều cho các doanh nghiệp không muốn chấp nhận những rủi ro này, đặc biệt là khi kết hợp với các giao thức mật mã.

Computer Center Room



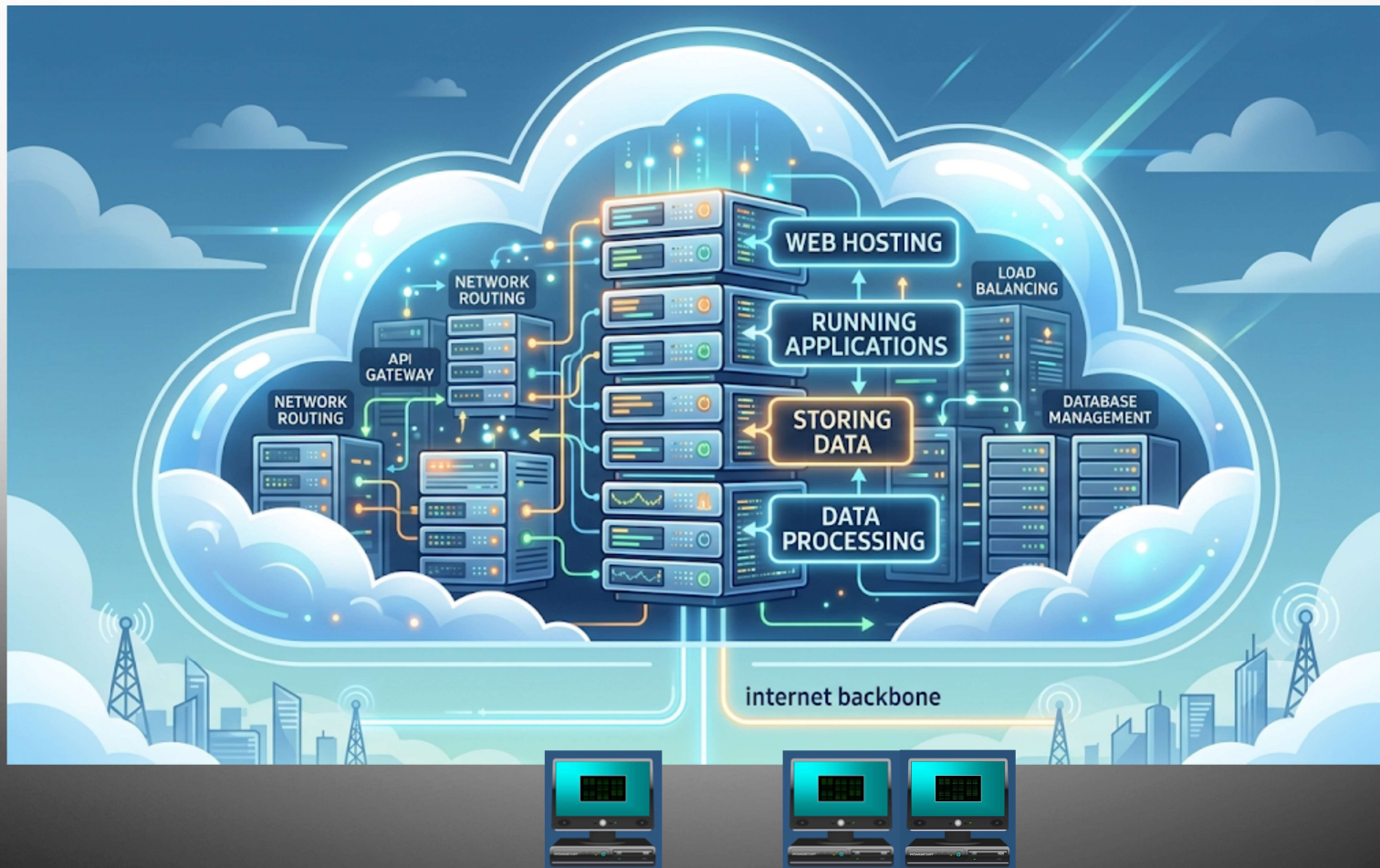
Data Center





Global Total Number of Data Centers around 11800

Cloud Computing (IaaS, PaaS, SaaS)



- (AWS) Amazon Web Services
- Microsoft Azure
- Google Cloud Platform
- Alibaba
- IBM

Alphabet / Google

Ingests colossal quantities of data: high-resolution map layouts, multi-spectral video streams (YouTube), and trillions of user-generated web nodes. (10-15 Exabytes)

Meta Platforms

Curates continuous engagement archives, real-time spatial user tracking, advertising logs, and complex social graphs across platforms.

Amazon AWS

Anchors the digital infrastructure of modern commerce, enterprise backup, database nodes, and cloud-native application stacks. (100s exabytes)

- **Building**



Computing types

- **Hardware, software**

- **Maintenance**

- **Power-Electricity**

- **Cooling**

- **Scaling**

- **Reliability, Business Continuity-backup, second site...**

- **Others**

IaaS-Infrastructure as a service all the hardware
PAAS-Platform , All the hardware-OS, Middle ware
SAAS-Software. Cloud providers provide everything including applications

Intelligence Repositories: Intelligence bureaus deploy multi-megawatt facilities, engineered to handle massive streams of intercepted network parameters.

Continuous Ingestion: Satellite imagery, international network lines, and security surveillance archives generate consistent ingestion traffic

Classified Infrastructure: Utilizing air-gapped private sovereign cloud environments to support highly secure, real-time analytics.

Public Sector & Weather Services

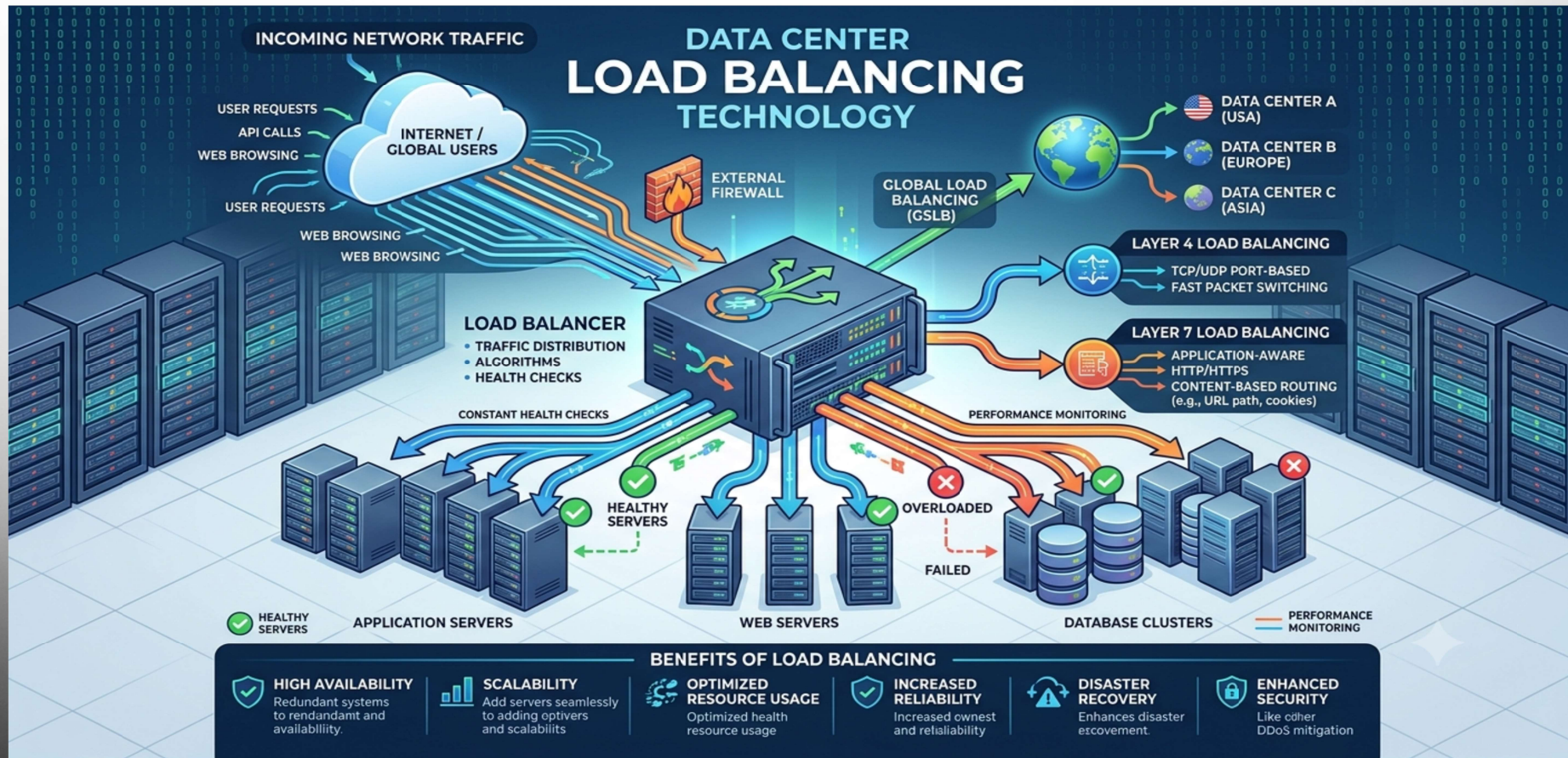
NOAA & Climate Models

The National Oceanic and Atmospheric Administration processes dozens of terabytes of new telemetry daily from remote radar networks. These pipelines feed complex climate systems representing some of the most data-rich and oldest continuous databases on the planet.

sea-buoy arrays, and satellites.

NASA Deep Space Telemetry

Active planetary probes, orbiters, and advanced space telescopes transmit continuous streams of multi-wavelength observational imagery. The telemetry output is securely cataloged into open-source repositories, driving scientific discoveries and spatial models around the globe.



Intelligence Processing: Claude helped analysts digest and summarize vast volumes of intelligence data, satellite imagery, and battlefield reports much faster than humanly possible. [[1](#), [2](#)]

Target Mapping: Integrated with Palantir's Maven Smart System, it was used to identify, map, and prioritize military targets before and during operations. [[1](#), [2](#)]

Simulations: The AI assisted planners by running and evaluating potential battlefield scenarios to minimize risks. [[1](#), [2](#)]

Political and Ethical Controversy

Ongoing Tensions: The use of the AI models sparked a severe standoff between the Pentagon and Anthropic over how autonomous weapon systems and military surveillance algorithms should be handled. [[1](#), [2](#)]

Official Ban: Citing supply-chain risks and disputes with Anthropic's leadership over the scope of the technology's deployment, the U.S. government ordered all federal agencies and the Department of Defense to phase out the use of Claude and similar systems. [[1](#), [2](#)]

Civilian Casualties: The employment of these AI platforms for rapid target prioritization came under heavy scrutiny—most notably following a deadly missile strike on an Iranian elementary school in late February, leaving Anthropic leadership publicly expressing uncertainty over how their models influenced those specific battlefield decisions. [[1](#)]

For further details on how emerging technologies are shaping modern armed conflicts, you can read more via the in-depth coverage in [Responsible Statecraft](#) or the original reporting from [The Wall Street Journal/Bloomberg](#).

Artificial General Intelligence

A hypothetical type of AI that can **understand, and apply cognitive skills** to **virtually any intellectual task**, matching or **exceeding human capabilities**. Unlike today's **narrow AI**, which excels at specific jobs, AGI would **seamlessly transfer knowledge** between domains and **adapt to unfamiliar problems**.

Trí tuệ Nhân tạo Tổng quát (AGI) là một dạng AI giả định có khả năng học hỏi, thấu hiểu và vận dụng các kỹ năng nhận thức vào hầu hết mọi tác vụ trí tuệ, với năng lực ngang bằng hoặc vượt trội so với con người. Khác với các hệ thống AI chuyên biệt (narrow AI) hiện nay vốn chỉ xuất sắc trong những công việc cụ thể, AGI có thể linh hoạt chuyển giao kiến thức giữa các lĩnh vực khác nhau và thích ứng với những vấn đề hoàn toàn mới lạ.

Artificial superintelligence (ASI) is a hypothetical AI system that vastly surpasses the brightest human minds in every cognitive domain, including creativity, general wisdom, social **skills**, and problem-solving. Unlike today's task-specific AI, an ASI could continuously rewrite and optimize its own code to become exponentially smarter.

Siêu trí tuệ nhân tạo (ASI) là một hệ thống AI giả định, sở hữu năng lực vượt xa những bộ óc xuất chúng nhất của nhân loại trên mọi phương diện nhận thức, bao gồm khả năng sáng tạo, trí tuệ tổng quát, kỹ năng xã hội và khả năng giải quyết vấn đề. Khác với các hệ thống AI chuyên biệt hiện nay, ASI có thể liên tục tự viết lại và tối ưu hóa mã nguồn của chính mình để trở nên thông minh hơn theo cấp số nhân.