Stingy Computing

Run less software, buy less hardware

Vijay Kartik
31.07.2025

AI, cryptocurrency mining, etc.

IEA Study Sees AI, Cryptocurrency Doubling Data Center Energy Consumption by 2026

A recent International Energy Agency (IEA) report forecasts that data centers' total electricity consumption could reach more than 1,000 terawatt-hours (TWh) in 2026.

Sean Buckley ▪ March 9, 2024 ▪ 4 min read

AI, cryptocurrency mining, etc.

Forbes

EDITORS' PICK

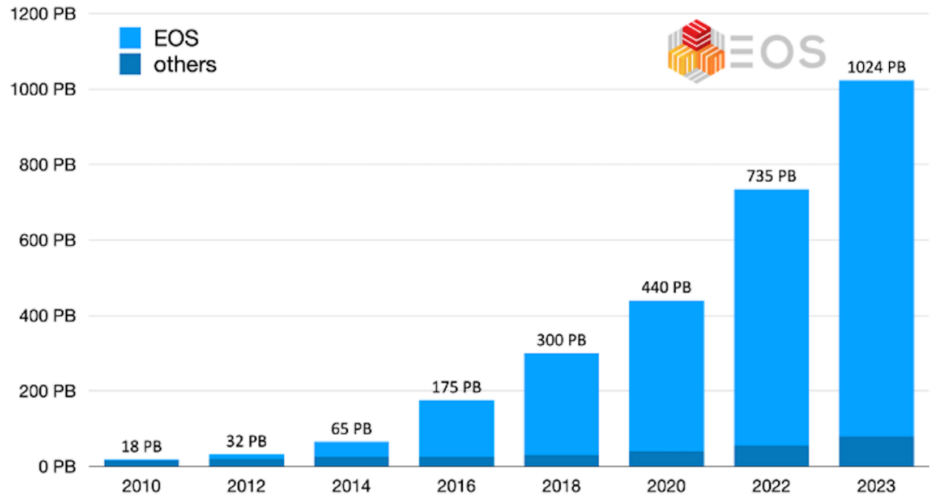
Current Climate: Crypto And AI Have Emerged As Major Energy Hogs

By Alex Knapp, Forbes Staff and Alan Ohnsman, Forbes Staff.

Follow Authors

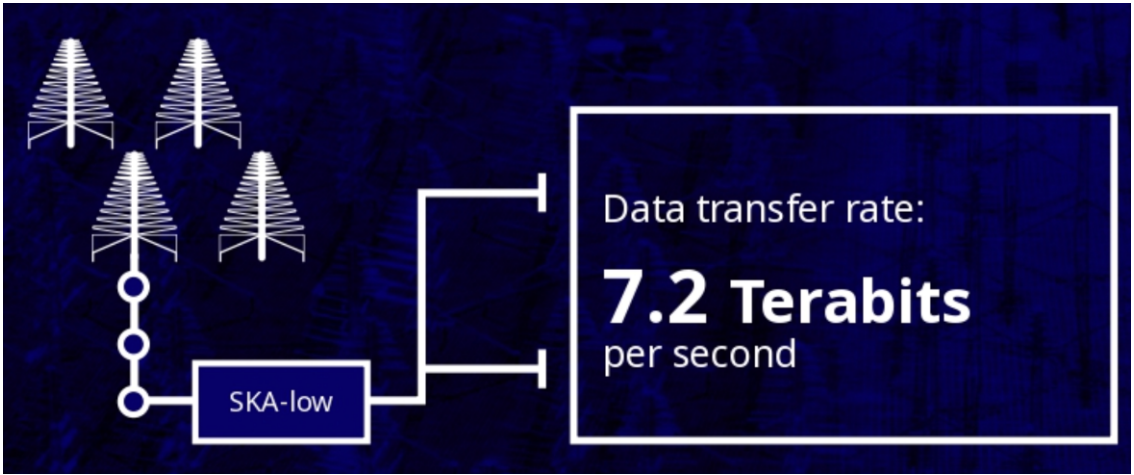
Published Feb 05, 2024, 07:00am EST, Updated Mar 31, 2025, 08:29pm EDT

Large research facilities



This graph shows the capacity evolution of CERN's data store. (Image: CERN)

Large research facilities

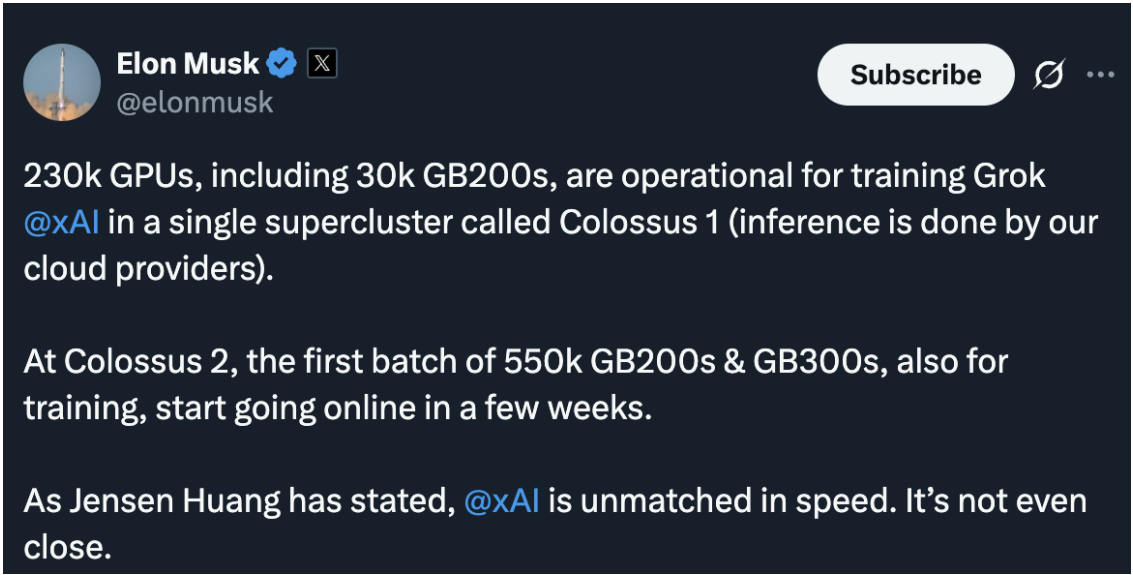


Source: SKAO press pamphlet

Resource shortage?

Buy more | **Consume less**

Buy more


Elon Musk
@elonmusk
Subscribe
230k GPUs, including 30k GB200s, are operational for training Grok @xAI in a single supercluster called Colossus 1 (inference is done by our cloud providers).
At Colossus 2, the first batch of 550k GB200s & GB300s, also for training, start going online in a few weeks.
As Jensen Huang has stated, @xAI is unmatched in speed. It's not even close.

Consume less

Do fewer things

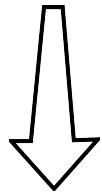
- Pre-select batch jobs
- Reuse processed results
- Do nothing (jobs/day is not a metric)

Do same things in less time

- Produce efficient software
- Energy consumption $\propto$ Running time
- Memory is 'cheap'

Computing efficiency checklist

- Optimize your software
- Implement checkpointing
- Revisit precision needs
- Ask someone else to (re)write your software



Return on investment

Sustainable computing

Minimize resource consumption

- Use less energy to run
- Use old hardware for longer

Promote user autonomy

- Provide users with choice
- Remove barriers to usage and ownership

Greenfield efforts towards sustainable computing

- PULP (RISC, CH+IT)
  - Parallel Ultra-Low Power software design
- TinyML (AI/ML, UK)
  - Optimized ML workloads in $\mathcal{O}$(mW)
  - *Kind of* superseded by 'Edge AI'
- PIM (IoT/ML, USA)
  - Processing In Memory
  - non-von Neumann, experimental
- Bits&Bäume (DE)
- KDE Eco (DE)
- BE4FOSS (Ended, DE)

Quantifying sustainability

Measure → Analyze → Certify

Measurement tools

- Running Average Power Limit (RAPL)
- Turbostat
- Power Measurement Toolkit (PMT)
- pyJoules
- eco2AI

Labels/Scores

- SEIC - Server Idle Energy Coefficient (UBA)
- SCI - Software Carbon Intensity (GSF)
- Blauer Engel (DE-UZ 215)

♪ It's not easy being green ♪

Measure: What? How?

Optimize: For?

Act: Educate? Incentivize?

Source: Wikimedia/CC-BY-SA-4.0

FIN.

Abstract

With ever-increasing data volumes pushing the climate footprint of computing at large-scale research facilities to record-high levels, we need to take a very critical look at our current software algorithms and drastically reduce the energy consumption of data analysis and storage solutions. The main questions to ask are - Is AI an unquestionably good idea for science? After an experiment, do we need all the data, all the time, for ever? Can we perform reproducible science while being stingy with our energy consumption? There are collaborative efforts in the software development community to address energy challenges - and ErUM-Data could benefit from joining and/or getting inspired by such efforts.