ONE MILLION GPU HOURS

A public compute budget forces decisions about which needs deserve electricity, machines and labour. The exercise exposes the political choices that govern a socialist allocation procedure.

Daorde Collective · 12.07.2026

Suppose a public institution receives one million GPU hours for the coming year. It receives a physical quantity of computing capacity that credit cannot expand. Once the institution exhausts it, no capacity remains.

The number looks immense until it is divided. One million hours is enough to keep 114 accelerators occupied for a year. That is slightly more than fourteen eight-GPU DGX H100 systems running continuously. Nvidia's technical sheet specifies an approximate maximum draw of 10.2 kW per system. If all capacity ran on those machines at that maximum, the servers would consume about 1.275 GWh. Cooling and the rest of the facility would come on top.

Renting the hours does not make them immaterial. On 13 July 2026, AWS listed an effective London price of $3.933 per H100-hour in capacity blocks. A million hours would cost $3.933 million before storage, data transfer, support and technical labour. The figure establishes the scale of the resource; a comparison between renting and owning would require other costs.

There are more reasonable uses than capacity, so someone will have to reject applications.

MONEY MAKES THE DECISION TODAY

The market allocates compute through solvent demand. An advertising company can buy one hundred thousand hours if it expects to recover the cost by targeting advertisements more precisely. A hospital without the necessary budget does not enter the queue. Prices coordinate private decisions after decision-making power has already been distributed unequally.

Existing public provision corrects part of the access problem while retaining criteria worth examining. EuroHPC treats requests above 50,000 GPU hours as “large scale”. Under its 2026 industrial call, companies, SMEs and startups may receive capacity free of charge; they retain ownership of generated data and models and may exploit the results commercially. EIC Accelerator awards receive priority over other proposals. Experts score innovation and impact as the decisive criteria.

The access terms establish that the state pays for the machine, a company may appropriate the result, and experts order the queue according to “innovation” and impact.

A socialist institution would have a different mandate: meet agreed needs, reduce social labour, respect material limits and expand common capacities. It needs political power to set that mandate and mathematics to fulfil it.

THE PLAN BEGINS BEFORE OPTIMISATION

Operations research can find an allocation that maximises an objective function subject to constraints. It cannot write the objective on its own. Instruct it to maximise revenue and it will favour uses with the highest monetary return. Instruct it to reduce emissions and it will ask for a definition of emissions, a time horizon and data. If health, language and energy receive numerical weights, someone selected those weights before the program ran.

This distinction separates calculation from planning. Calculation tests whether a combination fits, estimates consequences and discovers incompatibilities. Planning includes the formation of objectives, ownership of the means, decision-making bodies, execution and correction. Planners can use a spreadsheet, but they cannot delegate their legitimacy or responsibility for mistakes to it.

In *Towards a New Socialism*, Cockshott and Cottrell distinguish macroeconomic, strategic and detailed planning. The million-hour budget already presupposes a macroeconomic decision: how much labour, energy and equipment society devotes to compute. Choosing priority areas is strategic planning. Distributing hours, storage and staff among projects is detailed planning. Mixing the three produces absurd debates, such as arguing over queue order without asking why a private firm can enter it.

A WORKED EXAMPLE OPEN TO ARGUMENT

Let us construct an annual plan to state the choices a real allocation would require without claiming to predict the needs of a future society.

Before scoring, the planning body applies four conditions. A project must answer a need recognised by a sector plan or a social initiative with sufficient support. It has to explain why a less compute-intensive method is inadequate. It must have the data, staff and storage required to use the hours. Its results become common property, except for personal data or material whose publication would cause concrete harm.

The planning body excludes commercial advertising, financial speculation, mass police surveillance and fossil-fuel expansion through a political decision about which production deserves no common resources.

The following allocation is a worked example:

Use	**Hours**	**Main condition**
Climate, electricity grid and local adaptation	230,000	Public data and models; coordination with the energy plan

Use	Hours	Main condition
Health, diagnosis and drug discovery	220,000	Independent clinical validation; essential results cannot be privatised
Languages and public knowledge	140,000	Consensually obtained corpora; open models and tools
Materials, transport and industrial production	150,000	Demonstrated material savings; participation by sector workers
Accessibility, care and assistive technologies	80,000	Design with users and evaluation of everyday utility
Reproduction, auditing and safety	100,000	Teams independent from the original result producers
Exploratory proposals	30,000	Access for teams without a supercomputing track record
Emergency reserve	50,000	Publicly reasoned release during the year
Total	**1,000,000**	

These provisional numbers expose conflicts that prices conceal. Allocating 100,000 hours to reproduction reduces the number of new projects but increases reliable knowledge. Reserving 30,000 for inexperienced teams lowers short-term average utilisation and prevents organisations that master the application process from monopolising capacity. An emergency reserve leaves machines uncommitted in advance to respond to an epidemic, wildfire or grid failure.

The plan should publish applications, awards, conflicts of interest, actual use and outcomes. It should protect clinical data and information that identifies a person while making decisions of power transparent.

WHO DECIDES

An expert panel is needed to verify that code scales, data exists and an estimate of required hours makes sense. Giving it the final decision would turn a technical constraint into government by specialists.

Decision-making can take place in three stages. Health, energy, science, culture and industry councils define needs and limits within the general plan. Infrastructure workers and specialists review feasibility, consumption and risk. An allocation council with recallable representation from those sectors resolves conflicts and publishes its reasons. Affected parties can appeal factual errors, unequal treatment or procedural breaches.

This composition places political struggle inside a visible institution. An industrial union may defend simulations that reduce accidents; a language community may demand capacity that it could never win by user volume; the health system may contest hours with an energy project. There is no natural unit that makes all these ends equivalent.

Voting can set broad priorities and multicriteria models can test allocations. The plan must show how results change when weights change. If a small adjustment entirely removes one need, the council will have detected a fragile plan.

A “GPU HOUR” IS MISLEADING TOO

An H100, MI300X and B200 do not perform the same work per hour. Even two H100s deliver different results depending on networking, numerical precision, libraries and utilisation. Counting hours without measuring useful work can reward inefficient code or penalise a different architecture.

The GPU hour is a workable initial budget unit because it can be recorded. Evaluation should add energy, occupancy, staff time, storage and a workload-specific measure: simulations completed, samples processed or training to a declared threshold. Projects first receive a small test allocation. Large allocations are released in stages when measurements confirm the assumptions.

Chasing 100 percent utilisation would also be a mistake. Machines need maintenance; jobs fail; research explores paths that do not reach production. A public infrastructure that punishes every negative result will reward conservative proposals and embellished numbers. It should produce knowledge and social capacity even when that lowers its occupancy rate.

LABOUR, ENERGY AND PROPERTY

Operations staff, data engineers, maintenance workers, security specialists, user support and cleaners keep the machines running. Other workers produce and review the data. An allocation without labour hours privileges organisations capable of supplying their own teams. Each award should include or budget technical support and recognise preparation work.

Electricity imposes another constraint. The International Energy Agency estimated that data centres used 415 TWh in 2024 and projects about 945 TWh by 2030 in its base case. Consumption is geographically concentrated and competes for grid connections that take years to build. Our small system can schedule flexible workloads when electricity is abundant, while health or emergencies may require priority. The council can consult hourly carbon intensity, but it must decide which hospital waits.

Property adds another rule. If public compute produces a model that is then enclosed and sold back to the state, society pays three times: for the machine, the research and the licence. Results funded entirely from common resources must return to the commons.

Where protected data is involved, code, documentation, partial weights, evaluations or governed public access may still be opened without releasing the original data. A “confidential” label does not justify turning public support into private capital.

PLANNING MEANS BEING ABLE TO CORRECT

At the end of each quarter, the council compares awarded and used hours, energy, results, additional labour and compliance with conditions. It can withdraw an allocation or transfer capacity. At the end of the year, workers and users assess whether the result changed the service that justified the request. They will not accept a paper as fulfilment if a medical model never enters clinical practice.

Correction requires memory. Each decision keeps its assumptions and the reasons for disagreement. Without that archive, the next plan repeats the application ceremony and once again rewards the best grant writers. With it, society learns the cost of its objectives and which institutions fail to execute them.

One million GPU hours offers an object small enough to stop socialist theory hiding in enormous phrases. It forces us to order concrete machines, a quantity of electricity, workers and incompatible needs. An algorithm can balance the accounts within priorities decided by society.

SOURCES AND METHOD

The equivalences use 8,760 hours per year. The energy estimate uses fourteen complete systems plus the equivalent fraction, at a maximum 10.2 kW for every eight H100 GPUs, following the DGX H100 technical sheet. It is a maximum-power estimate for IT equipment, not a measured consumption figure or a whole-facility estimate. The rental figure uses the public AWS Capacity Blocks price for H100 in London, consulted on 13 July 2026. Access thresholds and property rules come from EuroHPC. Global energy figures come from the IEA ’s *Energy and AI*. The annual allocation is deliberately hypothetical: it demonstrates a procedure and does not describe an existing facility.