Sample planning profile

Two NVIDIA A100 80GB PCIe cards, one node, rented. October 26 2026, 04:08 to 14:53 UTC, 10.7 hours. The load was a script, not a customer job: training, inference, idle and resident phases at random. A power budget was in force all night, so this profile shows the load as the controller shaped it. The controller held at least one card below its limit on 22 percent of samples.

This is what DMS hands a grid planner or a study consultant after a measurement run: how the GPU load on a node behaves, in planning terms, with the measurement basis stated. A site's profile covers its own nodes and workloads. The figures here describe one node under a scripted load and say nothing about any real site.

How it was measured

Each card's power was read ten times a second through NVIDIA's management library and summed per node. These are the GPUs' own device readings. They leave out CPUs, memory, networking, cooling and power supplies, and they are not a revenue meter or a grid operator's measurement at the connection point.

On this driver the "instant" power field read the same as the default reading on 99.97 percent of samples. The largest change between two consecutive readings, a tenth of a second apart, was 345 W, so the readings were not heavily averaged. 100 percent of the 10.7 hours was covered, with no gaps.

Level

The node's rating is 600 W, two cards at 300 W each. At idle (the 5th percentile) it drew 136 W, and its median was 386 W. The 95th percentile was 445 W, just under the 480 W budget the controller held.

One reading of 734 W came at the instant a card woke from idle. It lasted a single tenth of a second. Whether it is real current or a sensor settling stays unconfirmed until it is checked against a wall meter.

The load spent 43 percent of the time between 20 and 30 percent of its rating (idle), and 39 percent between 60 and 70 percent.

Swings

The largest change inside any window, as a share of the 600 W rating:

Window	Largest change	Largest fall	99th percentile change
0.5 s	475 W (79%)	342 W	12 W
1 s	475 W (79%)	345 W	16 W
5 s	476 W (79%)	345 W	201 W
60 s	571 W (95%)	454 W	327 W

Steps

A step is any change over 120 W (20 percent of the rating) inside one second.

Kind	Count	Per hour	Median size	Steepest
Rise from idle (a job starting)	26	2.4	251 W	2,095 W/s
Rise between busy levels	27	2.5	123 W	3,447 W/s
Fall to idle (a job ending or crashing)	25	2.3	227 W	2,065 W/s
Dip that recovered within 30 s	47	4.4	97 W	2,502 W/s
Fall between busy levels	5	0.5	29 W	1,815 W/s

Dips recovered after a median of 0.9 seconds.

Swings by frequency

Band	RMS of node power
0.01 to 0.2 Hz	21.3 W
0.2 to 1 Hz	5.3 W
1 to 3 Hz	2.3 W
3 to 5 Hz	1.5 W

The bands above 0.2 Hz are marked unconfirmed in this profile. On this driver the instant field reads the same as the default reading, and published measurements on A100 and

H100 differ on what window a reading covers. They get checked against a meter before anyone relies on them.

Scaled to a 100 MW site of identical nodes

These are bounds, not a forecast.

If every node moved together, as in one big training job	If jobs were unrelated and their swings partly cancelled
The largest 5-second change would be 79 MW	It would be about 0.2 MW

A real site sits between the two, depending on how many nodes share a job. ERCOT's draft rule on power variation counts changes over 10 MW in any 5 seconds, measured at the connection point, where cooling and power supplies are included.

Files that come with a profile

- The full profile as signed JSON.
- A one-second CSV of node power, absolute and per unit, for playback in a study tool.
- 0.1-second windows around the largest job start, drop to idle and dip.

Dynamic Margin Systems, liam@dynamicmarginsystems.com