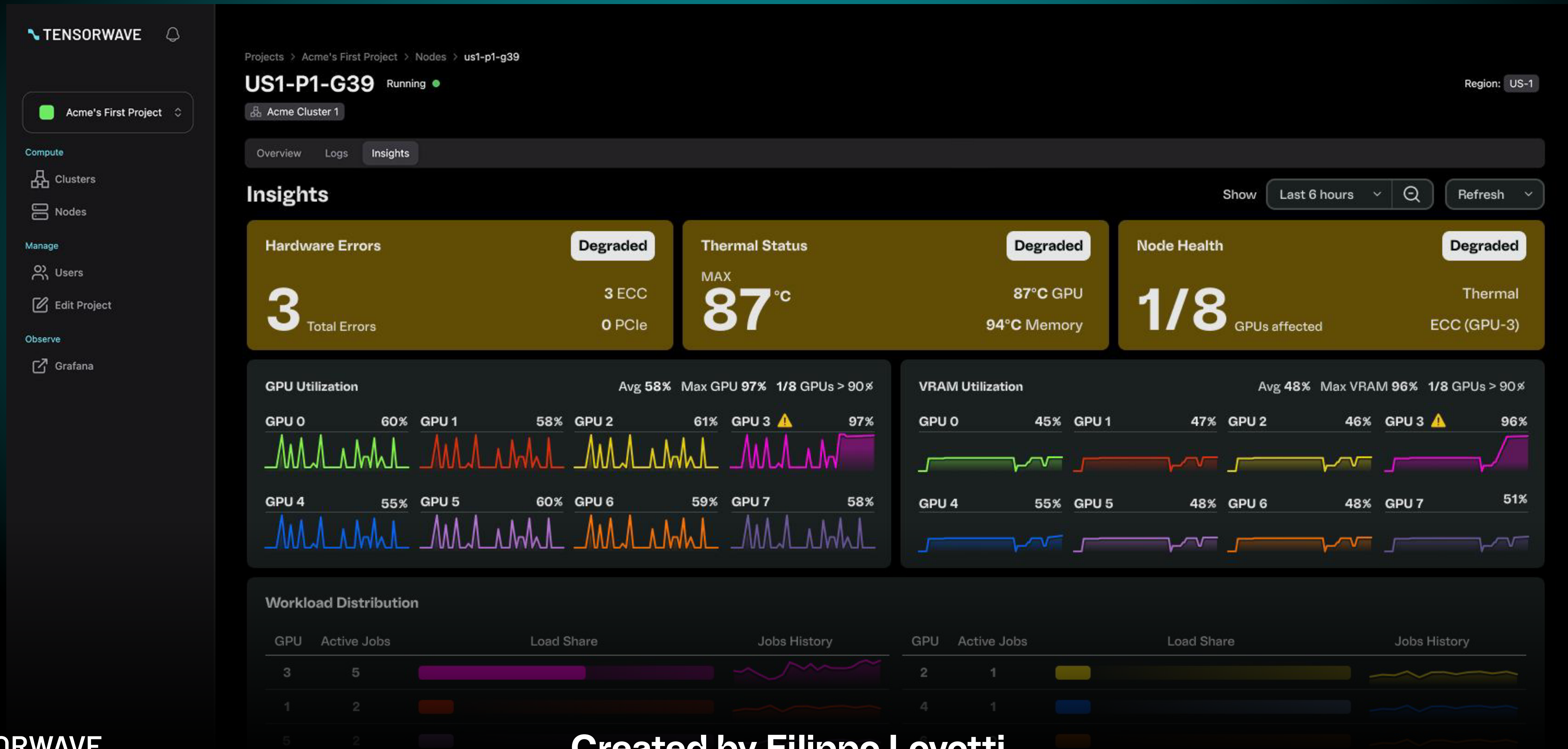


GPU Cluster Monitoring Dashboard Redesign



Strategic Questions

1. What are the top user pain points with the current monitoring experience?

Establishes what is broken now, before proposing solutions, and avoids “designing in the dark”. The answer ensures we design with operational function in mind.

2. What’s the typical workflow when investigating a performance issue?

We need to understand how users approach monitoring and investigation. The answer reveals their flow and what drives action during an incident.

3. What decisions do users make based on GPU monitoring data?

This question connects the dashboard to real operational outcomes. The answer helps prioritize decision-critical signals over secondary or contextual data.

Additional 1. Which users are passive observers vs active responders?

Clarifies the different roles and hats users wear. This can justify reorganizing tiles, supporting multiple levels of detail, or enabling different dashboard views.

Additional 2. Where and how do users access monitoring data across the product?

Looks at the broader system. Monitoring data may be accessed through alerts, embedded views, or direct navigation, which affects expectations around depth, density, and usage patterns.

New Priorities

Monitoring organized by what matters first, helping users move faster from signal to cause.

Health

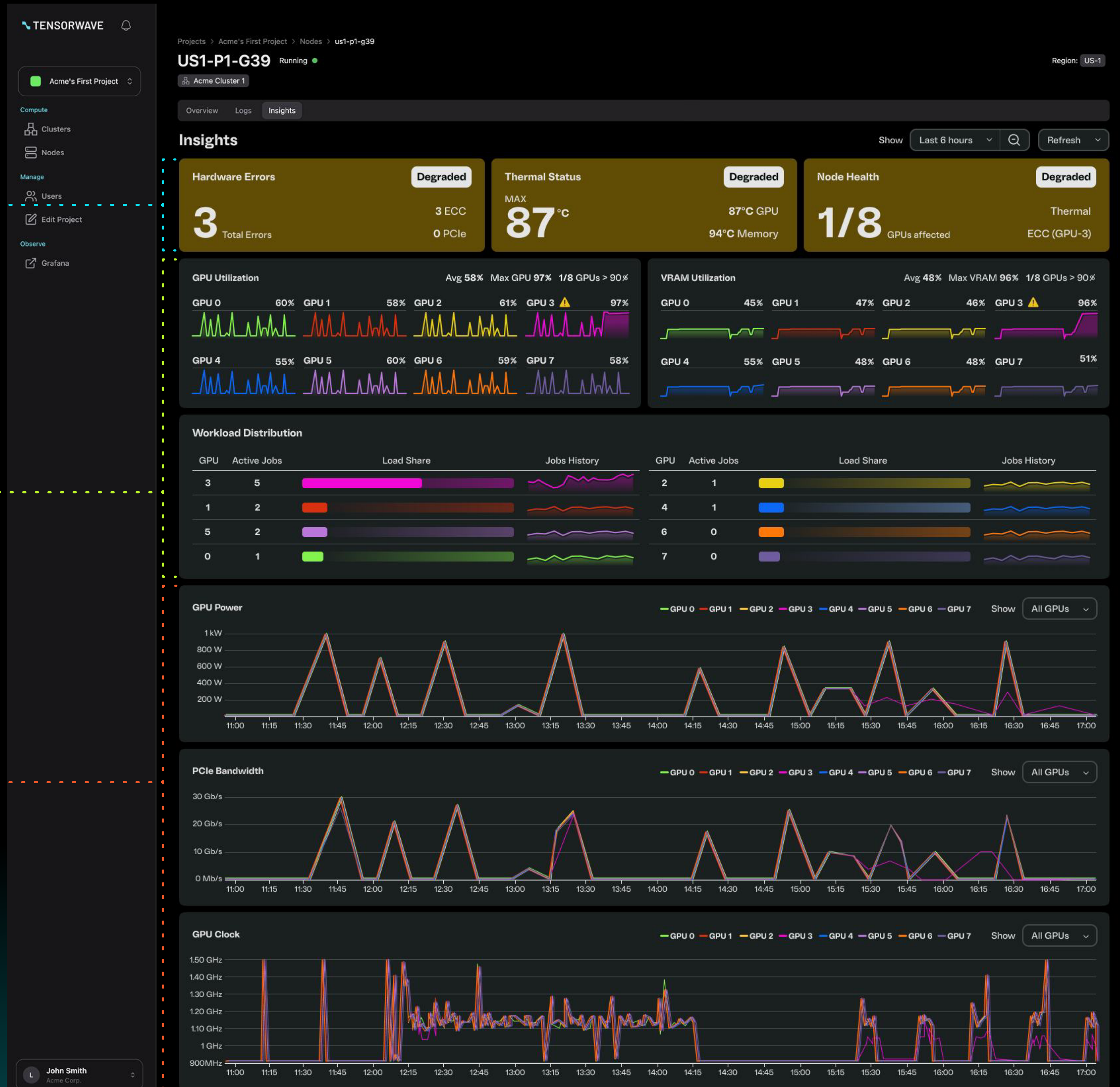
Is something wrong, and how bad is it?

Utilization

Where is the pressure coming from?

Diagnostics

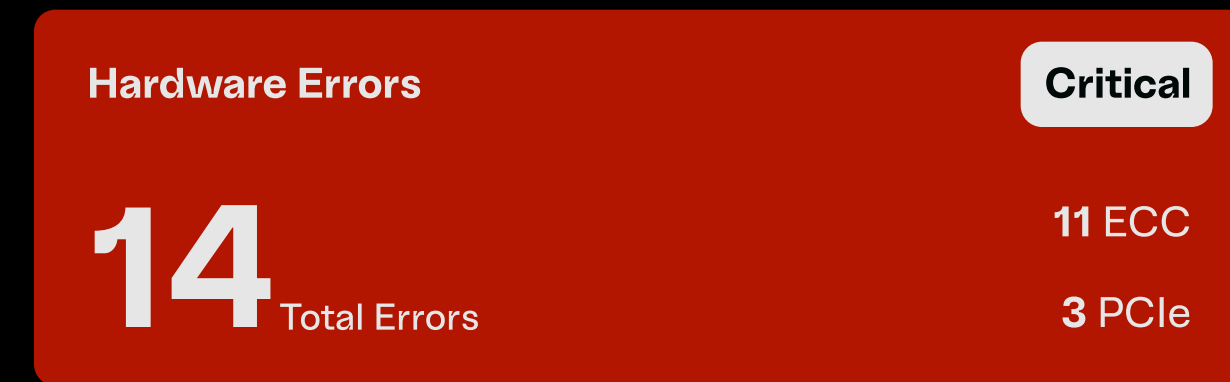
What's causing this GPU to behave differently?



Interactions

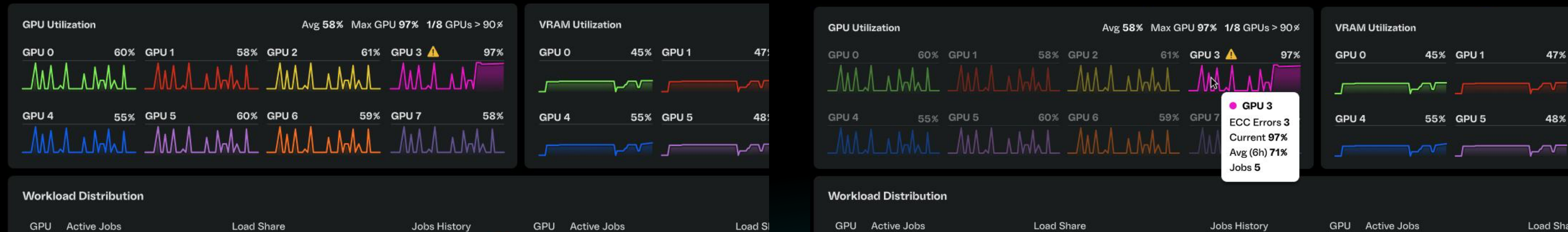
Health States

The new Health tiles show the status of the Node by color and state. The design optimizes speed over comprehensive information.



Utilization Spark Lines on Hover

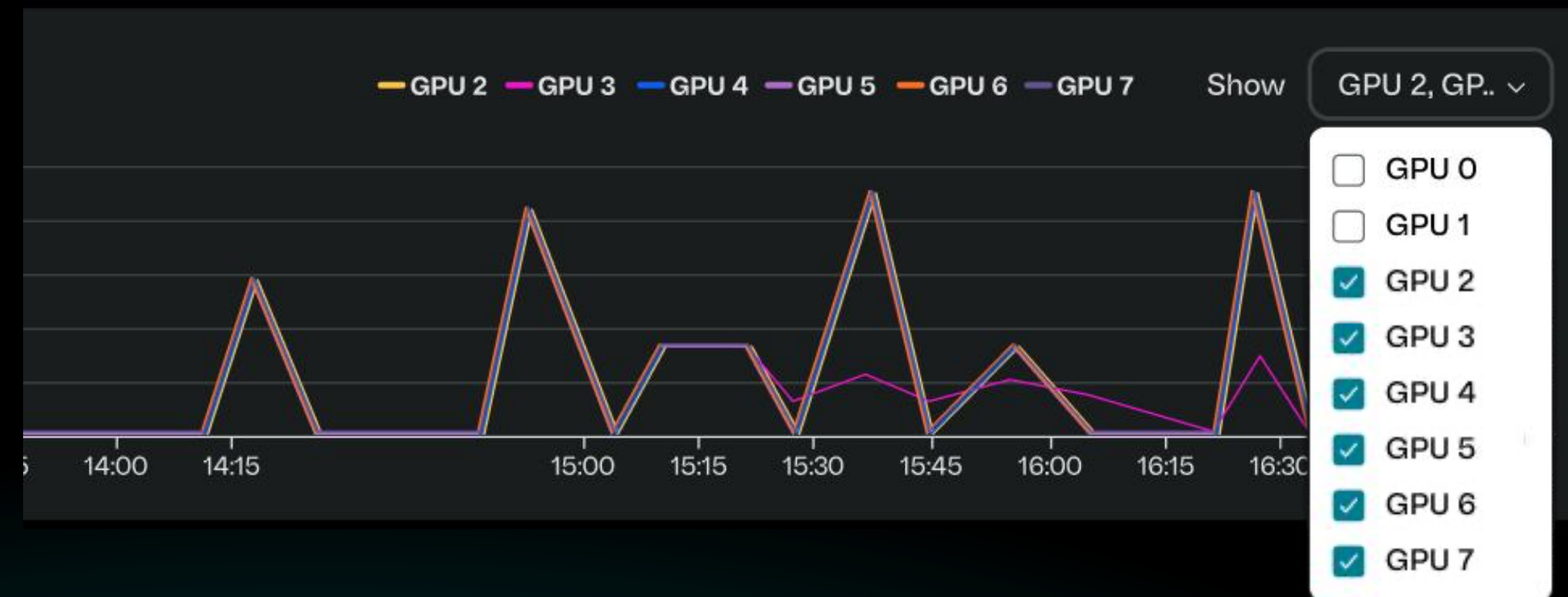
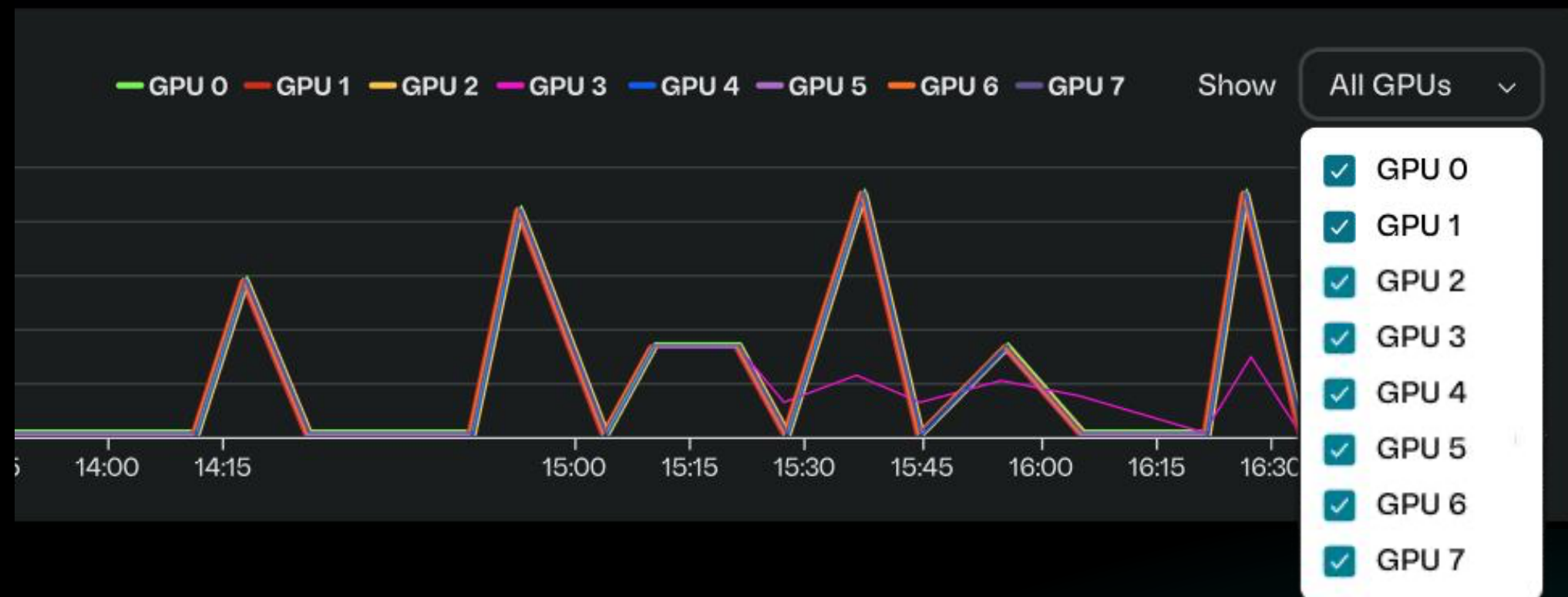
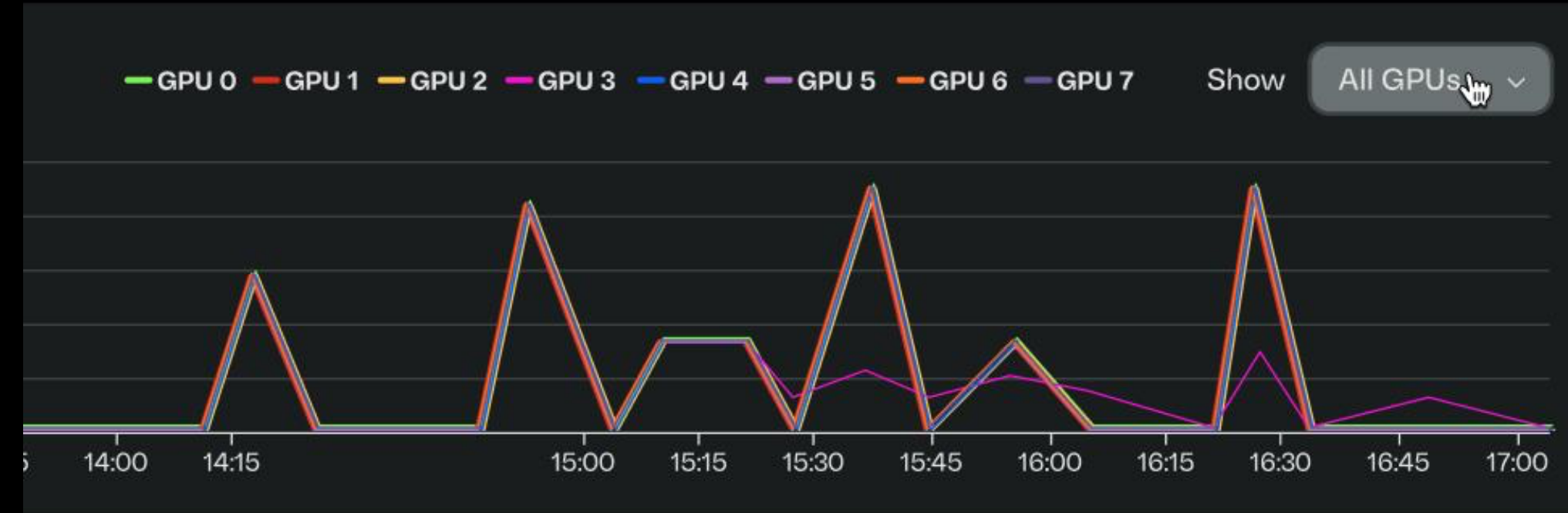
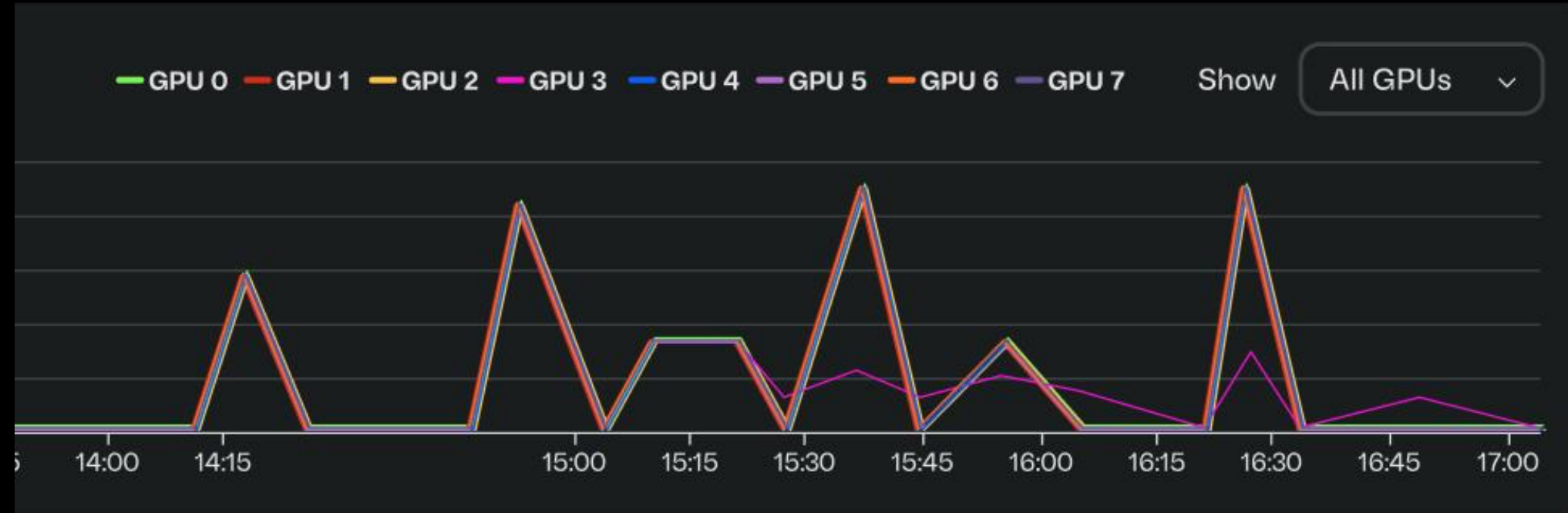
Sparklines provide quick trend context at a glance without pulling focus from comparison or current state.



Interactions

Diagnostics dropdown filter

The dropdown filter lets users isolate one or more GPUs during investigation while preserving time-based context across diagnostics.



Design Rationale

1. Information Hierarchy: How did you prioritize what information is most critical?

I prioritized what helps users detect and localize problems first, then pushed detailed diagnostics lower so users don't have to interpret everything at once.

2. Visualization Choices: Why did you choose specific chart types for different metrics?

I chose chart types based on how information is consumed at each stage of investigation. Health uses clear state-based indicators to communicate at a glance, Utilization relies on compact visuals, and Diagnostics use time-series charts where understanding behavior over time matters.

3. Key Design Decision: What was your biggest design change from the original and why?

The biggest change was organizing the dashboard around investigation flow, so users move from "something's wrong" to "this GPU is the problem" faster.