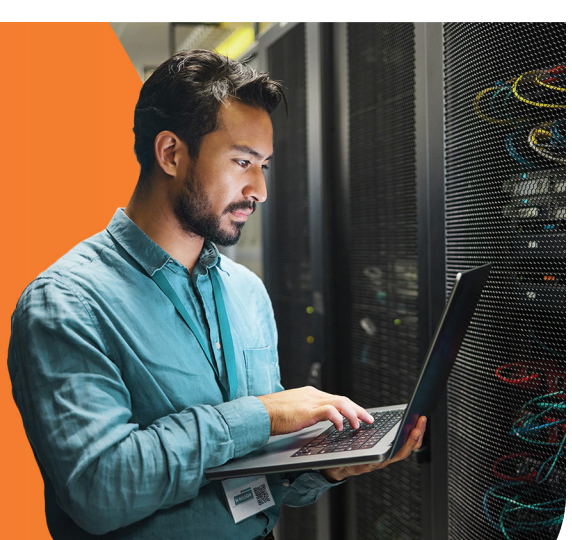
LUMEN

Powered by ciena

Neoclouds / AI Builders

Backbone for AI:
Design for performance today—and volatility tomorrow



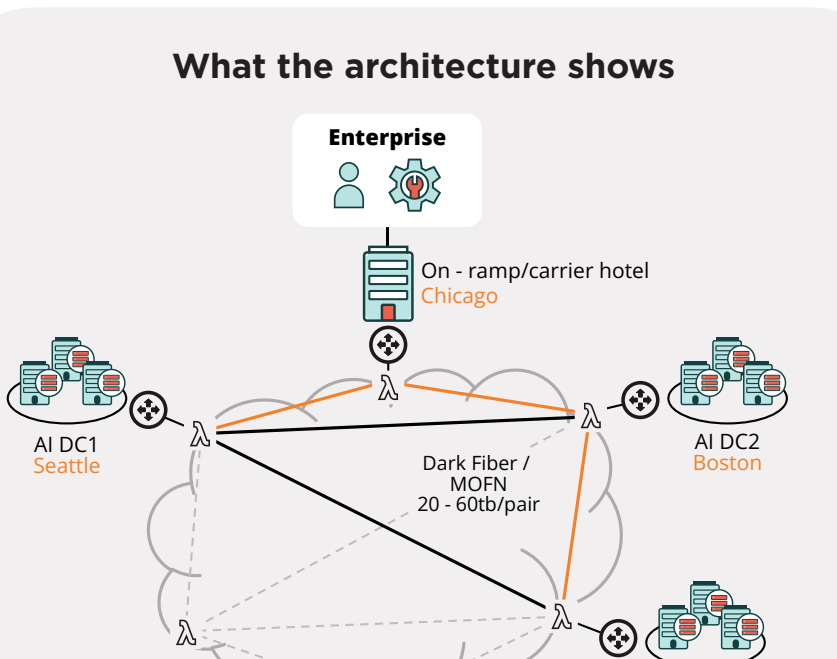
AI infrastructure moves fast—and it rarely grows in a straight line. New models, new regions, new customers, and new product launches create step-function demand. The backbone has to scale without forcing redesign. Most builders are forced to **build where power is** and **serve where users are.** The backbone is what makes that workable.

AI puts pressure on the backbone in two ways:

- **Training** drives massive, irregular movement of datasets and model artifacts across sites.
- **Inference** scales with adoption and becomes a persistent load. Performance issues show up as variability (latency spikes, timeouts), not always as clean outages.

When the backbone isn't designed for this, builders feel it as delayed expansions, inconsistent inference experience, and escalating operational friction.

What the architecture shows



AI builders operate across a distributed path: AI compute sites, inter-city backbone routes, gateways/PoPs, edge/colo environments, cloud on-ramps, and end users. Those handoff points are where performance and resiliency decisions show up first—especially as traffic shifts and demand spikes.

LUMEN®

Powered by ciena

Decisions that matter early

1 Where inference must be served

- Which experiences can't tolerate variability?
- What latency range is acceptable for those experiences as usage scales?

2 How you scale capacity without redesign

- How quickly can you add capacity when demand jumps—then jumps again?
- Which routes are long-haul inter-city vs metro, and what's your scaling plan for each?

3 Backbone model by route, not by ideology

- Where do you need speed to capacity (fast turn-ups, rapid upgrades)?
- Where do you need maximum control for sustained high volume?
- Where do hybrid managed options help reduce operational overhead without sacrificing performance characteristics?

4 Resiliency built around inference performance

- Protected can improve availability, but shared geography can still fail together.
- Diverse reduces correlated failures and is often the baseline for experience-critical inference.
- Active-active protects performance stability when "degradation" is effectively a failure.

✔ Quick checklist (use before adding your next site)

- ☐ Do we know which inference paths are experience-critical?
- ☐ Do we have true geographic diversity on those paths?
- ☐ If a route fails, does inference degrade—or stay stable?
- ☐ Can we add capacity quickly without re-architecting?
- ☐ Quick test: if a single construction cut could slow inference, you need true geographic diversity on that path.

What to do next

Map where inference must be fast and consistent, map where training traffic moves between sites, then choose connectivity and resiliency route-by-route (metro vs inter-city) based on performance impact—not habit.

Download the AI Backbone Playbook

About Lumen

Lumen is guided by our belief that humanity is at its best when technology advances the way we live and work. We deliver a fast, secure platform for applications and data to help businesses, government, and communities deliver amazing experiences.

Get in touch

lumen.com | 1-877-453-8353

LUMEN®