

Protecting TTFT as inference scales.

800G is a planning milestone for keeping distributed AI performance consistent across corridors, handoffs, and clusters.

800G isn't a fit for every build today—but it's worth preparing for if you're scaling distributed inference.

TTFT benchmarks:

TTFT = time to first token

<200_{ms}

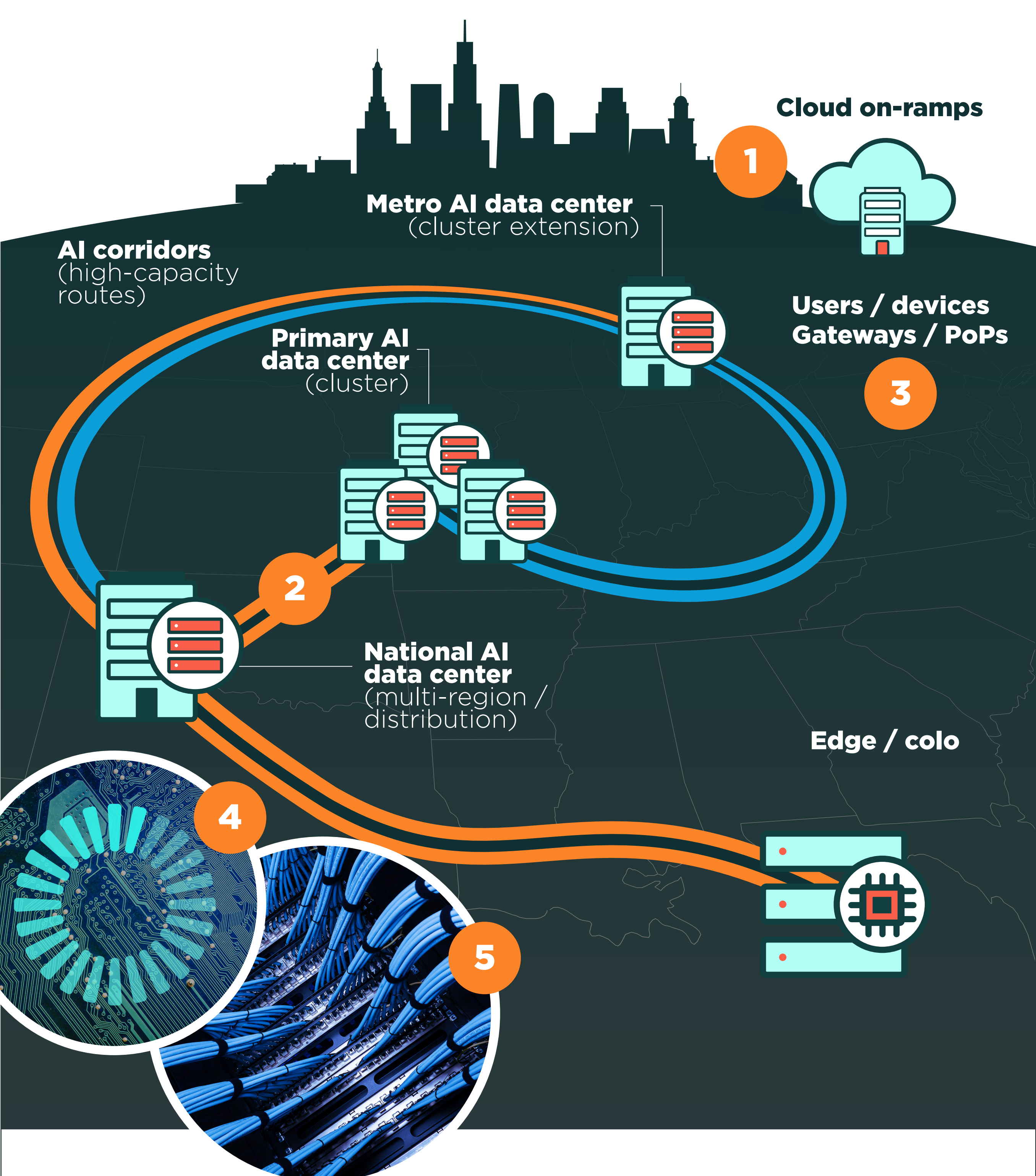
feels instant

<500_{ms}

feels responsive

<1_s

is noticeable



1

Inference at scale

As usage grows, inference becomes a persistent load; performance issues show up as **TTFT impact**—latency variability, jitter, and timeouts—before they show up as clean outages.

2

Distributed clusters drive “east-west” traffic

Training workloads drive heavy east-west traffic, and AI clusters rely on high-speed fabrics to keep pipelines saturated.

3

Handoff points break performance first

Gateways/PoPs, on-ramps, and interconnect points are common choke points; if these aren't designed for scale and resiliency, users feel the degradation first.

4

GPU utilization risk

When transport can't move data fast enough, expensive compute waits; higher-capacity corridors reduce the risk of under-utilization as clusters scale and data movement intensifies.

5

Efficiency and complexity pressure

Scaling capacity often increases ports, optics, and cabling—adding operational complexity and consuming space and power budgets; higher-capacity links can reduce components per unit of throughput.



Where this matters first: AI corridors, handoff-heavy paths, and inter-city routes connecting distributed compute.



Planning lens: design for volatility—assume demand jumps, then jumps again.

Lumen helps AI builders design and scale high-capacity connectivity across metros and long-haul routes—so performance stays consistent as inference grows.

[Download the 800G Whitepaper](#)



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.

lumen.com | 1-877-453-8353