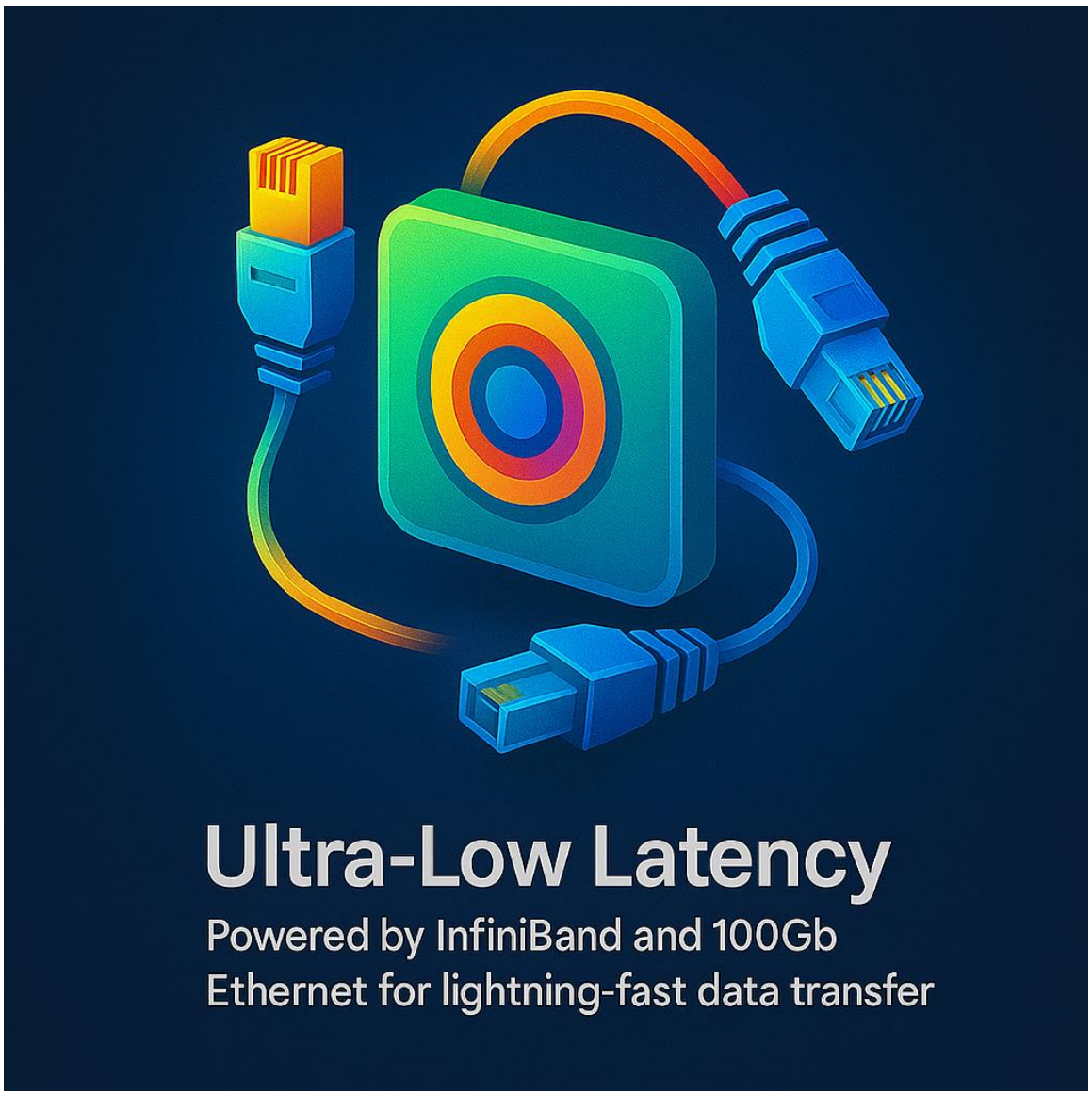
System Base Labs

A Carbon-Neutral Company

SBL

Today's AI Startup. Engineering the Intelligence of Tomorrow



White Paper: Ultra-Low Latency – Powered by InfiniBand and 100Gb Ethernet

Authored by: Aleiman Shankar Rao
Founder & CEO, System Base Labs
www.systembaselabs.com

System Base Labs

A Carbon-Neutral Company

SBL

Today's AI Startup. Engineering the Intelligence of Tomorrow

Introduction

Time is the true cost in AI. At System Base Labs, we've eliminated latency bottlenecks using InfiniBand networking and 100Gbps Ethernet backbones. Our architecture is engineered for sub-millisecond latency, enabling real-time collaboration between AI modules, distributed training nodes, and live inference endpoints.

What is Ultra-Low Latency?

Ultra-low latency refers to the minimal delay between input and response in a computing system. In AI systems, it is critical for enabling real-time inference, training, and interaction across distributed environments.

How We Achieve It

We utilize high-speed networking technologies like InfiniBand and 100Gb Ethernet to ensure data transfer between nodes is virtually instantaneous. Our infrastructure uses RDMA (Remote Direct Memory Access), kernel bypassing, and dedicated switching fabrics to minimize processing overhead and jitter.

Key Features and Advantages

Near-zero jitter across clusters

Jitter—variation in packet arrival times—is nearly eliminated using synchronous communication patterns and quality-of-service routing in InfiniBand. This ensures reliable AI module synchronization and steady throughput.

Seamless high-speed data replication

With our 100Gbps Ethernet backbone, data is instantly duplicated across storage and compute zones. This provides redundancy and rapid model checkpointing during training cycles, improving uptime and resilience.

Real-time response for edge and cloud fusion

Edge nodes processing real-time data (like agricultural sensors or diagnostic imaging devices) communicate instantly with centralized AI systems. This supports real-time feedback, hybrid deployment, and global inference pipelines.

Conclusion

Think faster. Train smarter. Deploy instantly. At System Base Labs, ultra-low latency isn't a luxury—it's a foundation. Our high-speed architecture ensures your AI workflows can match the speed of thought.

- **System Base Labs**: GPU Compute Infrastructure (India)
- **Shankar AI HQ**: Research, Integration & Strategic Ops (San Francisco)

Contact: partners@systembaselabs.com

#ShankarAI #SystemBaseLabs #OpenAI #AugmentedAI #EnterpriseLLM #SustainableCompute