ufiSpace

STAYING AHEAD WITH AI

IS YOUR NETWORK READY?

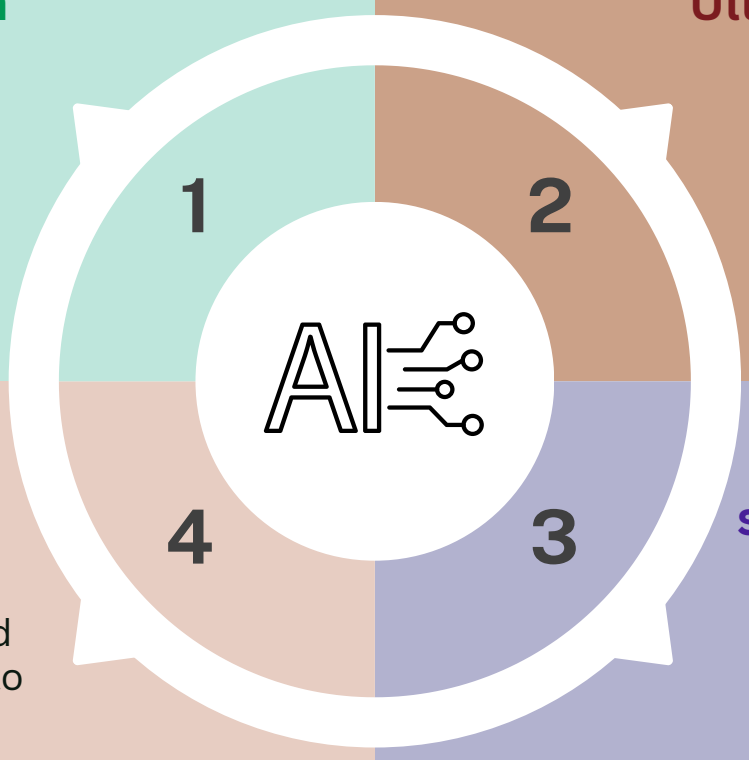
Understanding **AI Model Traffic Patterns**

High Bandwidth

Needs ultra-high bandwidth and High-radix fabric to meet intensive AI workloads.

Ultra Low Latency

GPUs computation cycles require extremely low latency to deliver results in real time.



Burst Traffic

AI training cycles involve elephant and mice flows, leading to fluctuating resource consumption rates.

GPUs synchronization

GPU servers cooperate for parallel computing and synchronized communication

How UfiSpace **DDC can benefit you ?**

UfiSpace Distributed Disaggregated Chassis (DDC) is an innovative networking architecture designed to **create flexible, high-scale routing systems tailored for the AI age.** It features the following key attributes:

Scalability

Enables easy scalability for high bandwith and radix up to **32K* GPUs in 800G**, meeting growing processing demands.

High Utilization

Enables load balancing, spreading cells evenly to all available fabric interfaces to maximize utilization.

Consistent Latency

All packets take a fixed number of hops from the source to any destination node, enabling consistent latency.

Failover

Detects any hardware- level link failure, allowing nano-second-level failover with no impact to job completion.

Low Latency

Various congestion control mechanisms to ensure low latency.

For example, in credit-based congestion control, transmission data is stored in the Virtual Output Queue (VOQ) of the source switch until the destination grants permission for passage.

Experience Comparative Success in the AI Age with UfiSpace DDC Architecture