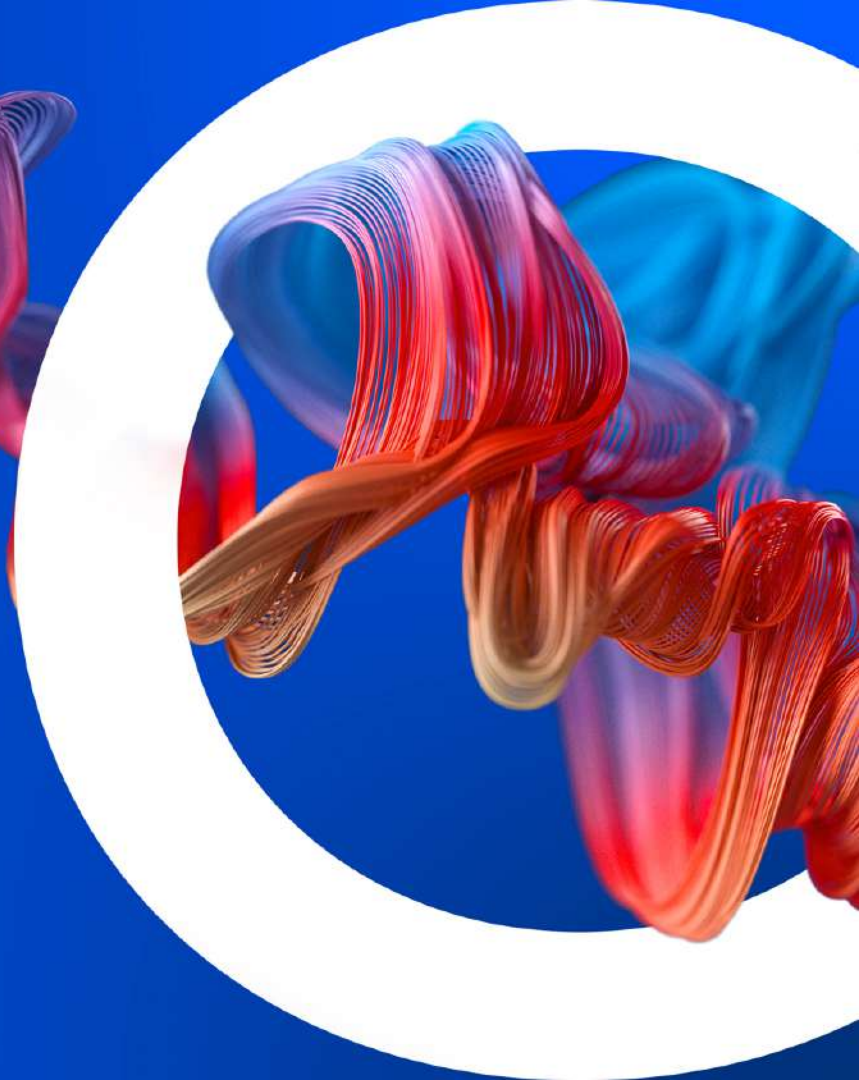
Building out DC Networks in the AI era

Jonas Vermeulen

April 28, 2025



What is Nokia today ?

Network Infrastructure

Mobile Networks

Cloud and Network Services

Nokia Technologies

IP

DWDM

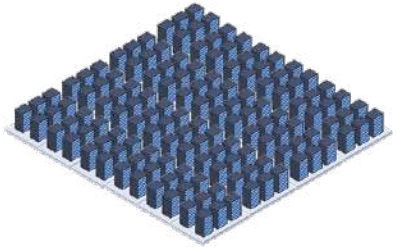
Fiber/DSL

Agenda

- The rise of AI Deployments
- The implications on the network
- Is it all about moving bytes around ?
- Conclusion

Data Centers Are Getting Bigger!

Note: These examples are to show the headlong rush to mega scale data centers is real. The actual ranking of current future data centers is not clear cut – especially in China, where past claims have been challenged.



AI is resetting the expectation of what a "large Data Center" is

2013

Then…largest DC in Europe

Meta's Lulea Data Center

30 MW*

84,000 m²

*267,471 MWh of electricity annually and withdraws 25.4 million litres of water

Source: Baxtel

2025

Largest DC in the World

China Telecom IMIP

150 MW

1,000,000m²

Note – China occupies the first **7 places** in at least one Data Center **Top 10** list

2028

Largest DC Announced

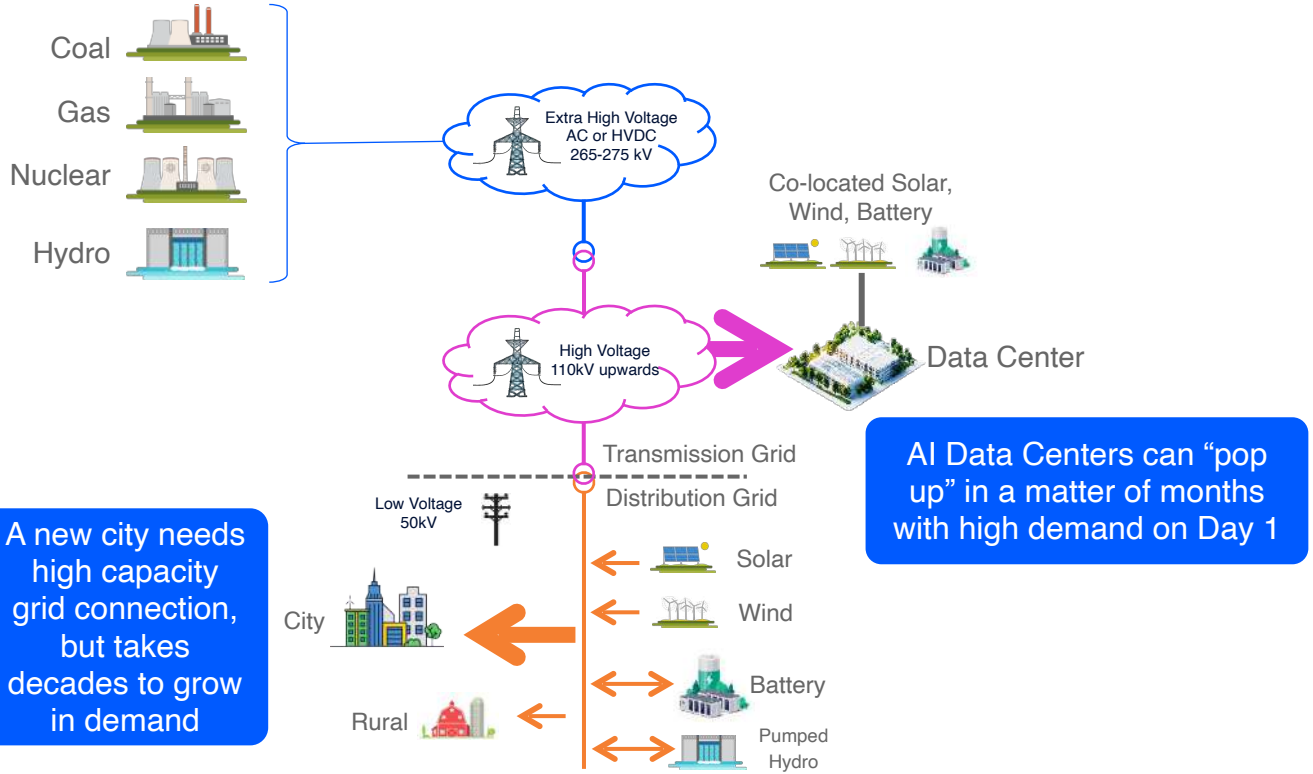
Jeollanam-do Province South Korea

3,000 MW

$35B Investment

Source: Capacity

Regional or National Grids are Crucial



A new city needs high capacity grid connection, but takes decades to grow in demand

AI Data Centers can "pop up" in a matter of months with high demand on Day 1

Think about grid in same way as the internet – a cloud…maybe ☺

The Grid is fed from power stations

And needs different voltage levels for efficient transmission/distribution

Regular customer – like cities and farms – use Distribution Grid

Renewable installations usually feed into the Distribution Grid

Storage systems also have 2-way connection to Distribution Grid

Very high demand users will connect directly to the HV transmission grid

Data Center may have local Solar, Wind and Battery Storage

Needs Grid connection for reliability against intermittent renewables

Is Gigawatt-Scale BYOP Viable? Let's look at xAI's Colossus in Memphis

Today
100,000
H100 GPUs

Tomorrow
200,000
H100 GPUs

100,000 H100s will require **150 MW**

MLGW

Local grid supply*: 8 MW

*As of August 2025

X14 

35 MW

Natural gas powered mobile generators

X4 

64 MW

MLGW has agreed to upgrade site to: 50 MW*

*At a cost of $760,000 to local taxpayers

xAI has committed to fund a **150 MW**, $24M new substation

Article: "xAI cluster is now the most powerful AI training system in the world"

What do AI data center operators really need?

1 Gigawatt Scale

City examples that can be powered with 1 GW*

Dublin	Southampton
Nice	Liverpool
Islamabad	
Amsterdam	Forth Worth
Cologne	Jacksonville
Oslo	Austin
Cartagena	San Jose

*Assumes a population of around 1 million

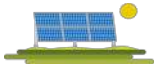
2 24/7/365 Energy Supply


Coal


Gas


Nuclear

24/7/365 Power Sources

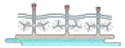

Solar

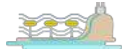

Wind

Intermittent Power Sources


Hydro


Geothermal


Tidal


Wave

Location-Specific Power Sources

3 Compact Installation

Co-locate power and Data Center, or
Distribute to smaller regional ones

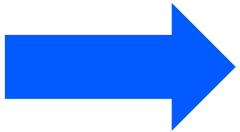
Avoid grid contention

Agenda

- The rise of AI Deployments
- The implications on the network
- Is it all about moving bytes around ?
- Conclusion

The implications on the network

- 100-200K-… GPUs
- Gigawatt Scale
- Compact installations
- Regional installations
- Timeline = Months



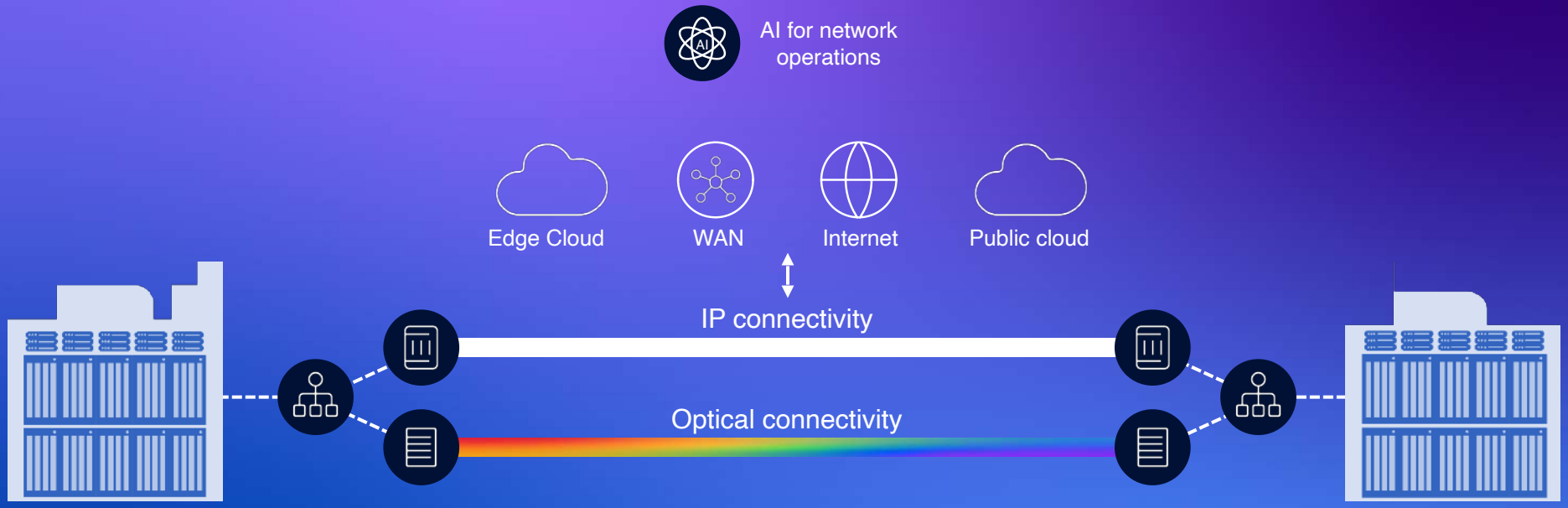
Networking revisited for AI

- Fabrics within the DC
  - Architecture
  - Features
- Datacenter Interconnectivity
- Operations

Making the most out of your GPU investment.

As fast as possible

Networking for the AI era



Data center Networking for AI

Reliable data center switching
Optimized for AI networking needs

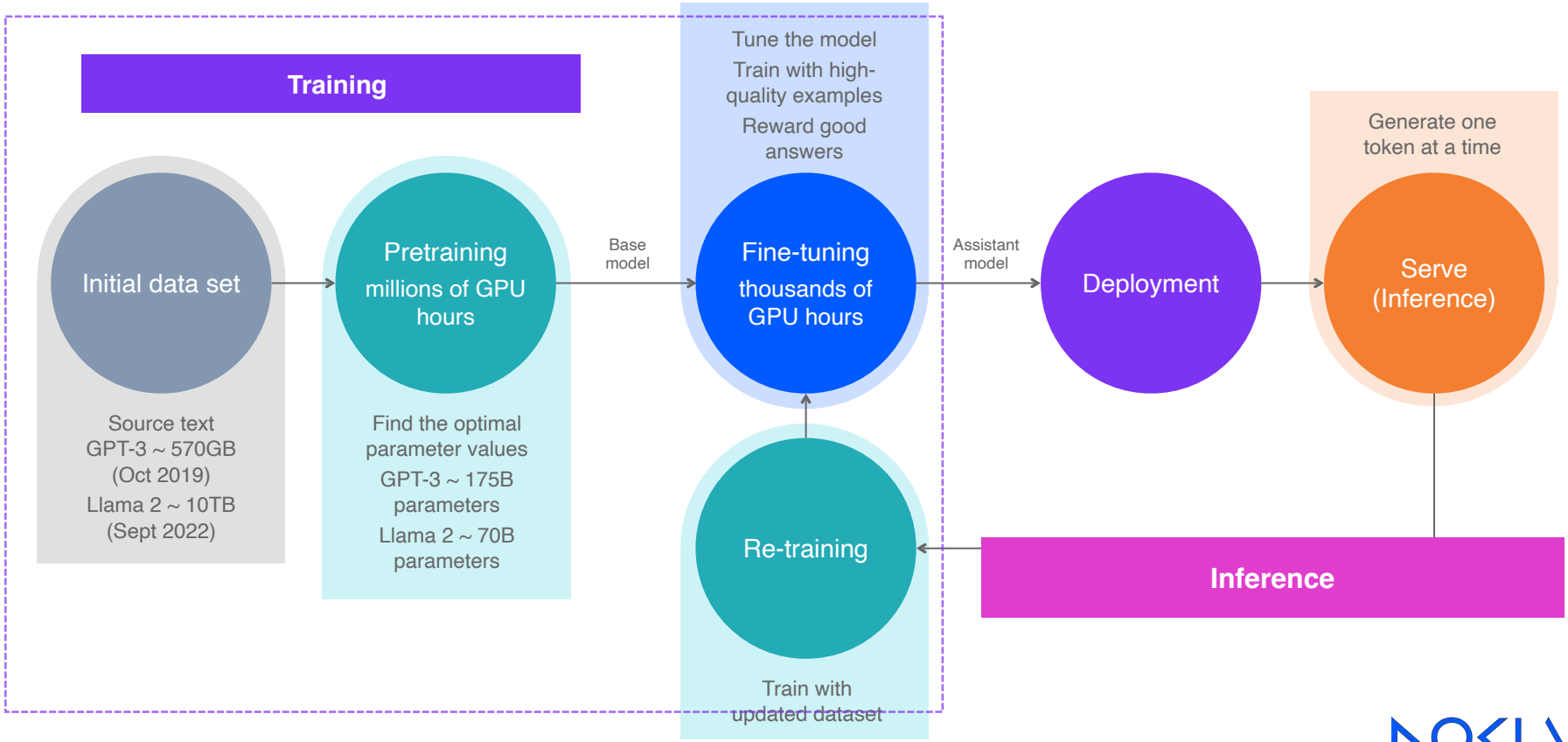
Data center gateway

High performance DC to DC, DC to internet, DC to WAN and DC to clouds interconnect and DDoS mitigation

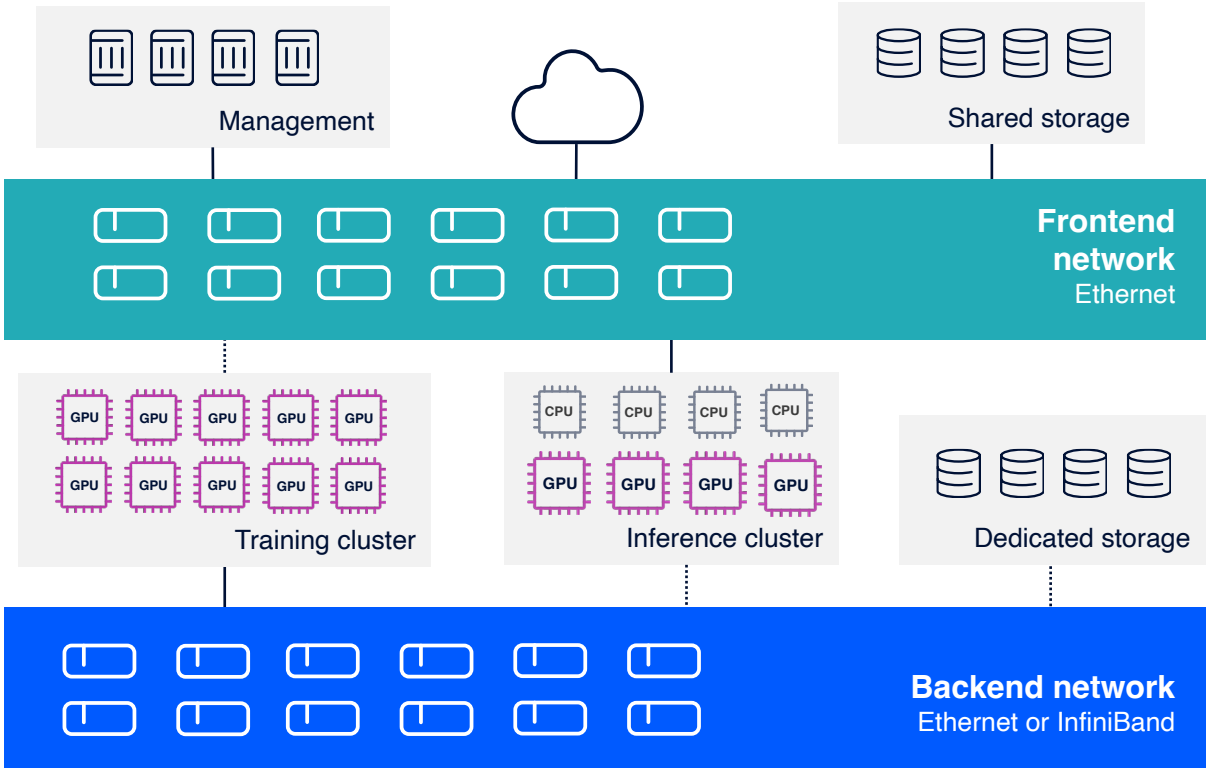
DC Interconnect

Dedicated optical networking enables high bandwidth, very low latency and highly secure data transmission for national/regional AI Grid deployments

The life of a large language model



DC Networking of AI clusters



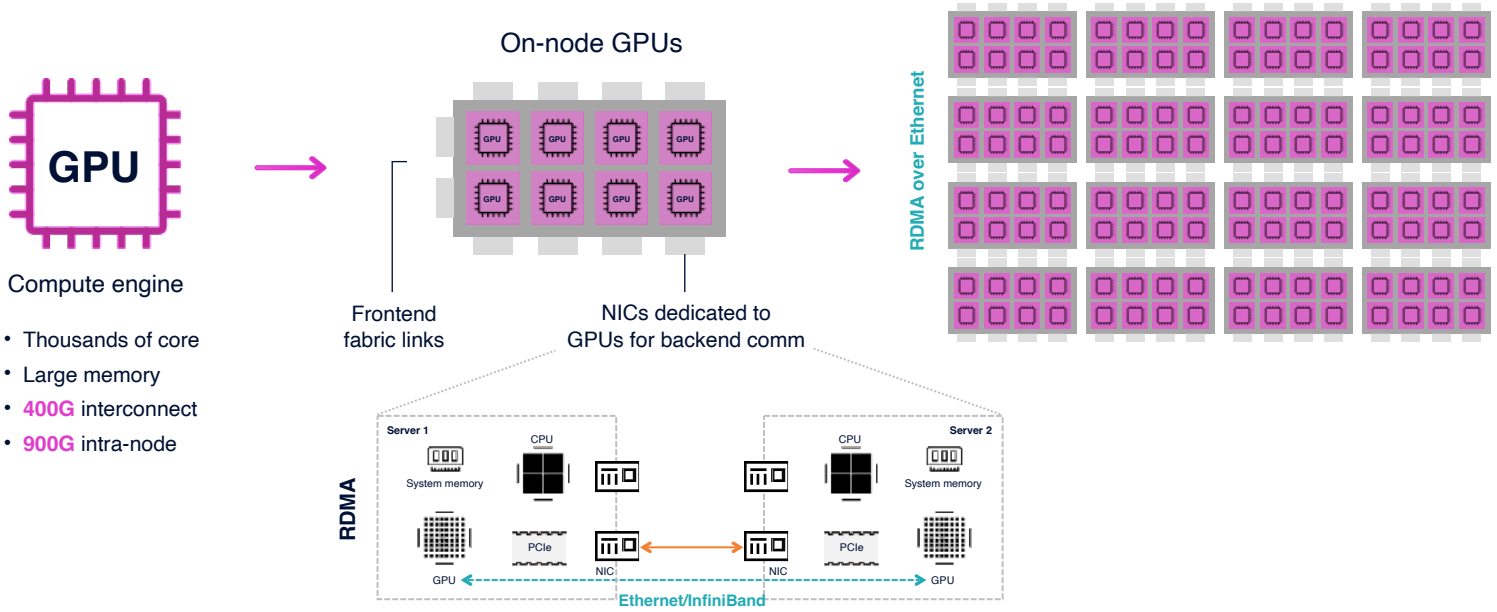
- **Front-end** networks is a typical CLOS-based data center fabrics
- Handles inference requests, cluster management, and data ingress/ egress.

- **Back-end** networks are dedicated to GPU-to-GPU, optionally GPU-to-Storage communications
- A separate storage network may be considered for large clusters

DC Networking of AI clusters

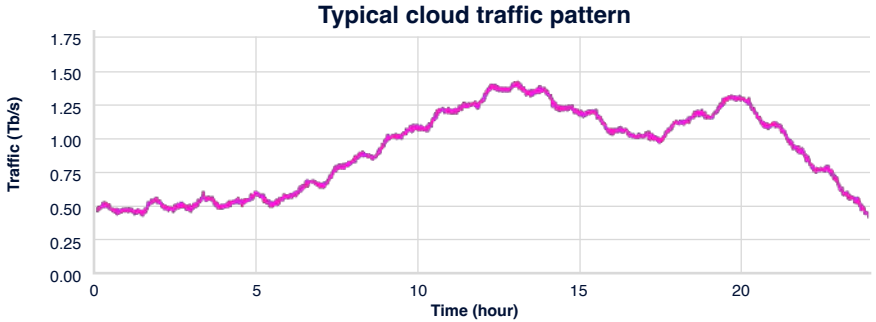
GPU-based workloads

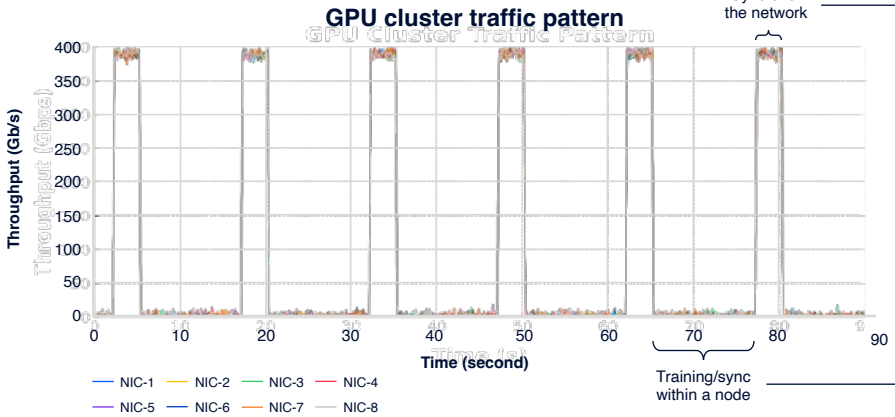
The **backend** fabric connects one BIG compute cluster



Ethernet (RoCEv2)	MAC	IP	UDP	Base transport header+	Data payload	Invariant CRC	FCS
InfiniBand	Local routing header	Global routing header		Base transport header+	Data payload	Invariant CRC	Variant CRC

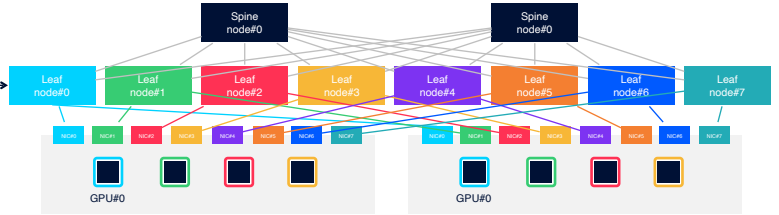
AI network traffic is different from traditional apps



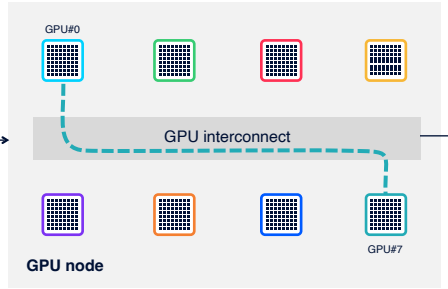


Characteristic	Traditional cloud	AI GPU training
Traffic pattern	Gradual, continuous waves	Sharp, **periodic bursts**
Flow structure	Many small flows	Few large flows
Network requirements	Tolerates some packet loss, needs redundancy	**Lossless** transport, minimal **tail latency**
Throughput per interface	Moderate	High (200G-400G)

Sync over the network



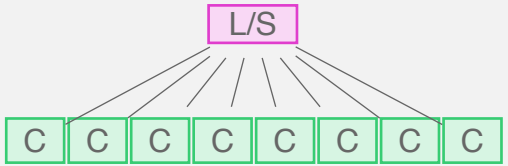
Training/sync within a node



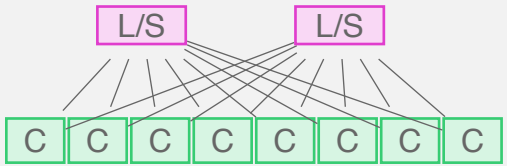
High-speed interconnect (NVLink, InfinityFabric …)

GPU fabric topologies

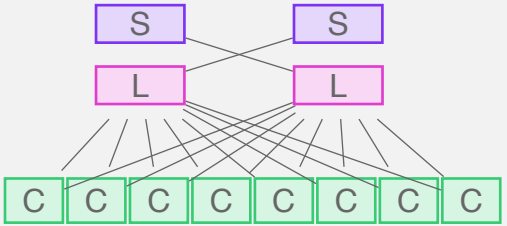
Single rail

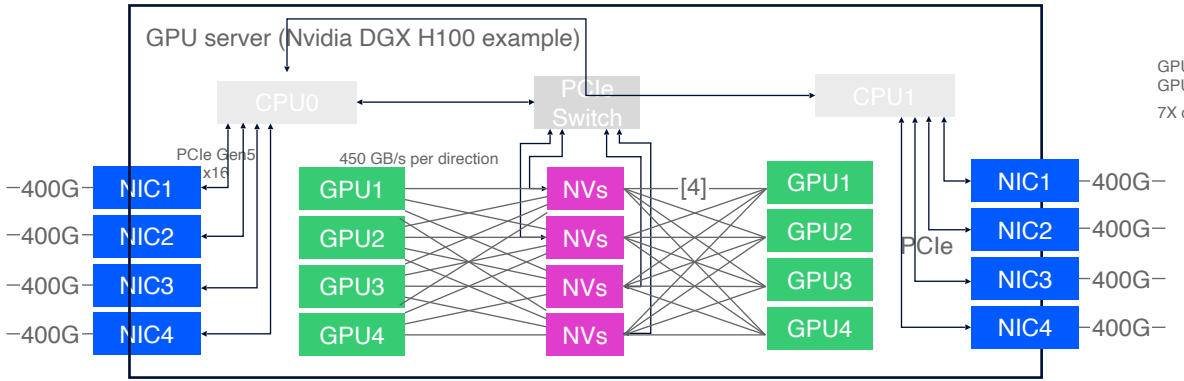


Rail only - Scaled



Rail optimized fat tree



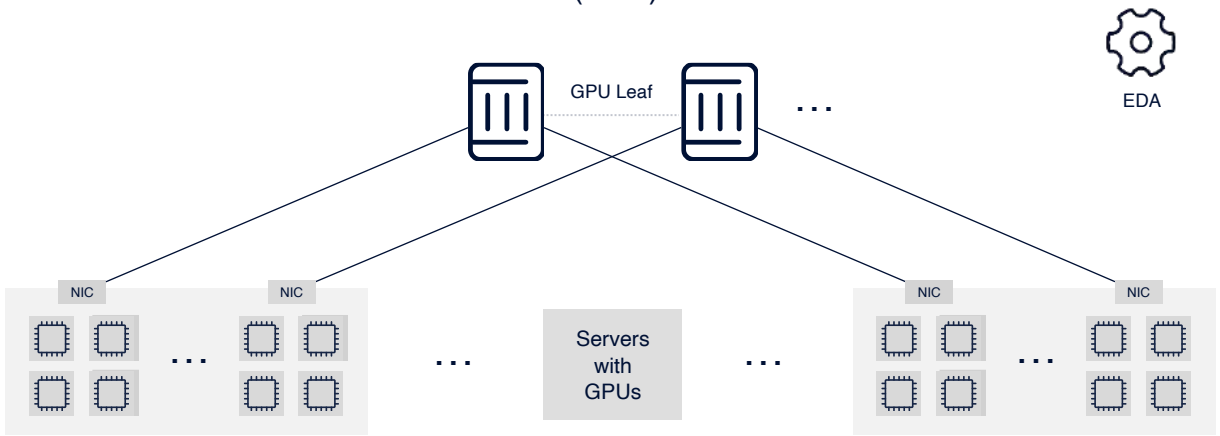


GPU b/w towards NIC: 63 GB/s
GPU b/w towards other GPUs: 450 GB/s

7X difference

Nokia reference design based on Ethernet

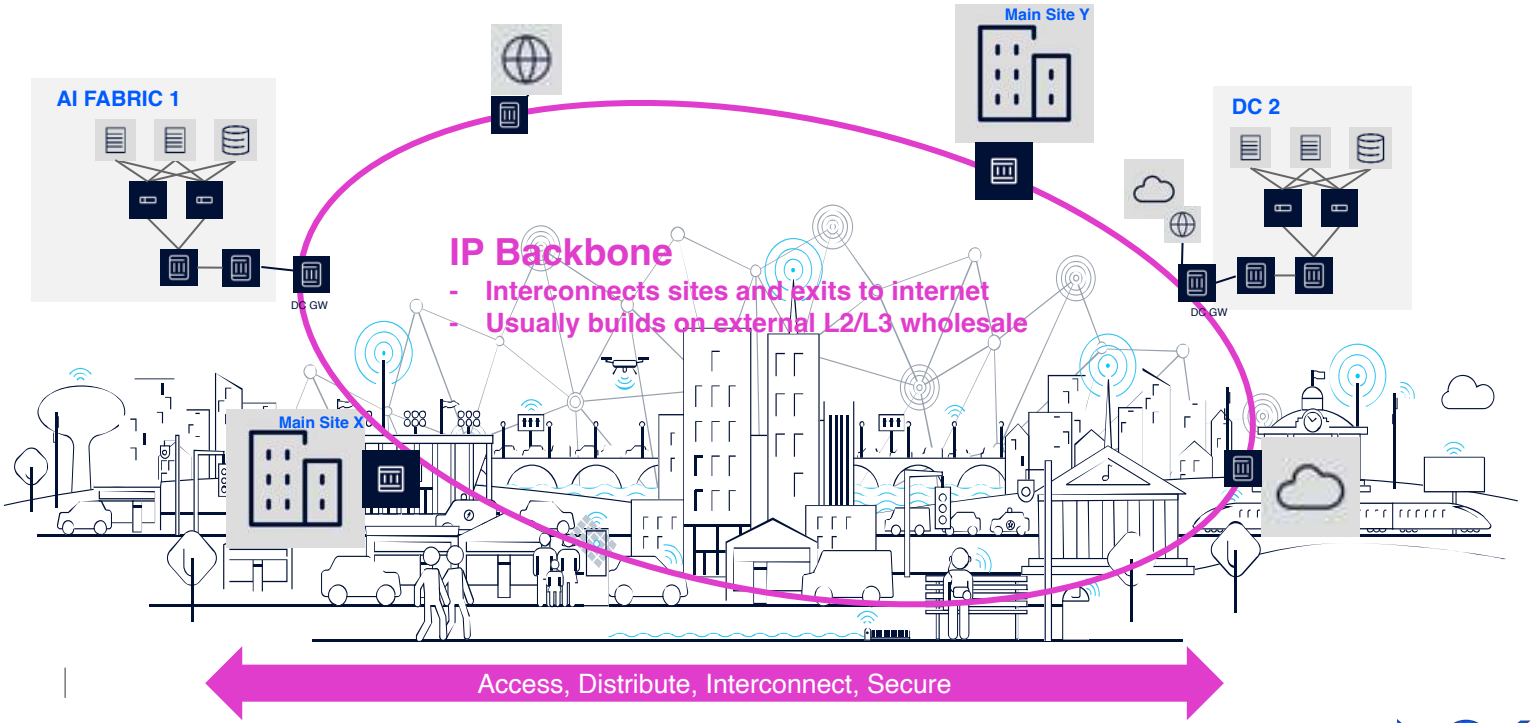
Platforms: Nokia 7220 IXR-H4/H5 (Tomahawk), Nokia 7250 IXR-6e/10e/18e (Jericho 2C+/3)
Automation: Event-driven Automation (EDA)



- Rail-optimized, **lossless** backend network design
- Design approach that supports
  - **Large** ‘000s GPU deployments
  - **Compact** installations (e.g. 64-128 GPU)
- **AI-aware** fabric automation

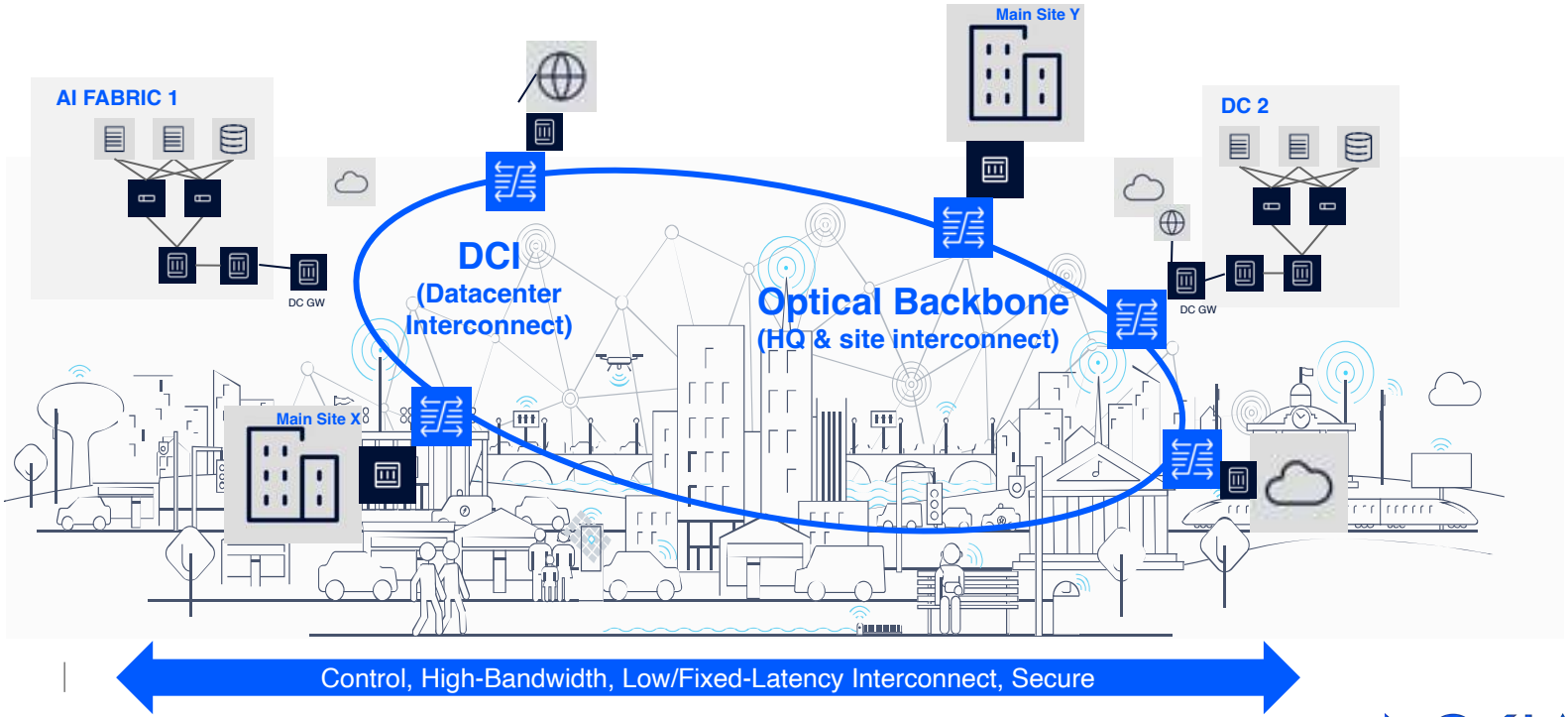
Building an AI-Interconnect-Network

Key use cases for IP



Building an AI-Interconnect-Network

Key use cases for DWDM



Agenda

- The rise of AI Deployments
- The implications on the network
- Is it all about moving bytes around ?
- Conclusion

Is it just about choosing the right architecture and platforms ?

Drive human error to zero

Making it work	Making it predictable	Making it for you
Reliability 	**Predictability** 	**Integration** 
Can be enabled with • Resilient architecture • Platform system-design • Quality-first software	Can be enabled with • Net-Ops: Intent-Based roll-outs based on blueprints • Git-Ops • AI-Ops:	Can be enabled with • Open and Easy integration with related ecosystems • Pluggable systems

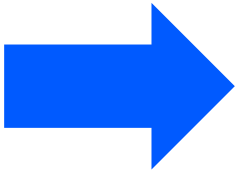
Speed of deployment comes as a by-product of avoiding errors along the way

Operating critical AI Fabrics

Automating and avoiding fat fingers at scale

Predictability

90% of all outages are caused by human error



Making it predictable

Net-Ops - Intent-Based networking

Git-Ops – Version and control

AI-Ops - Streamed telemetry as continuous operational input

Agenda

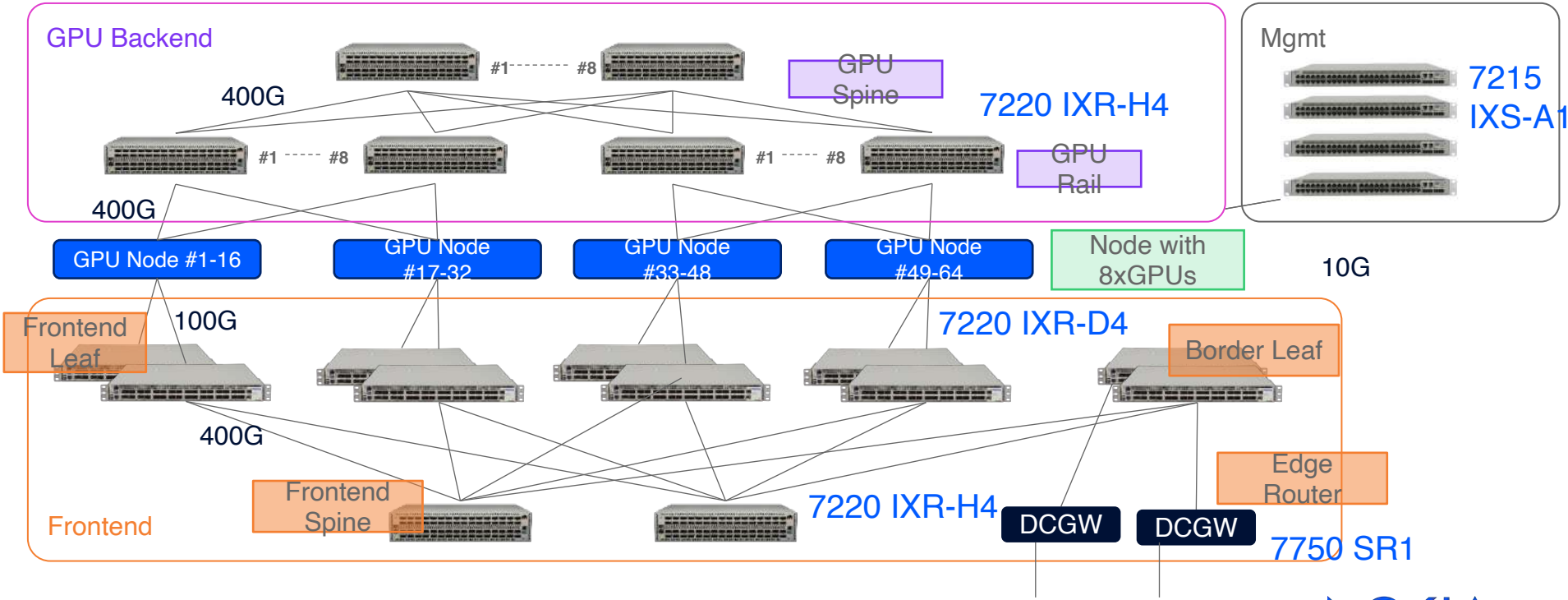
- The rise of AI Deployments
- The implications on the network
- Is it all about moving bytes around ?
- Conclusion

Deployment example

AI Training cluster at Nscale

NSCALE

Press release

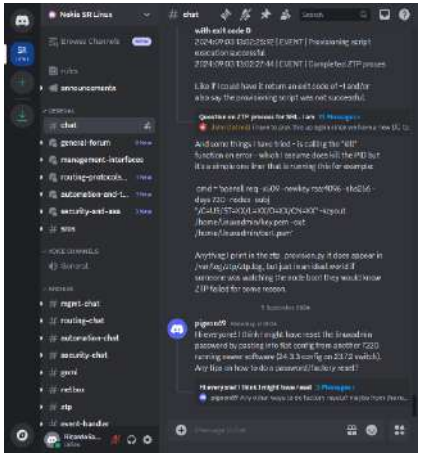


NOKIA

How to learn SR Linux?

learn.srlinux.dev

Discord channel



containerlab.dev

learn.eda.dev

EDA Playground

Questions ?

Talk later here or at
Nordic Datacentre Forum
(28 May, Nokia HQ Espoo)


NOKIA

Copyright and confidentiality

