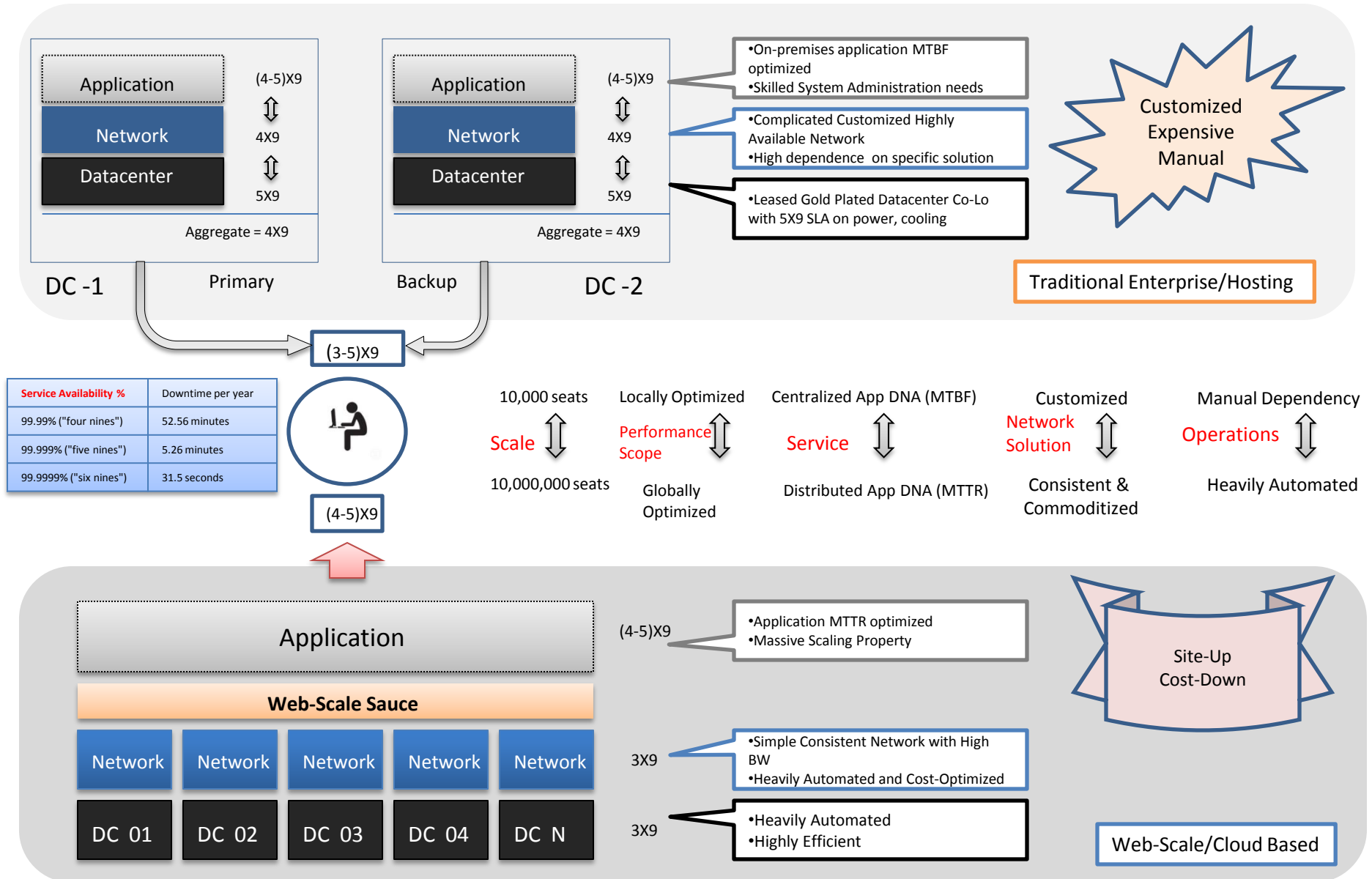# Trends and Challenges of Ethernet Usage in Microsoft Online Infrastructure
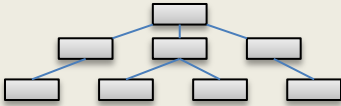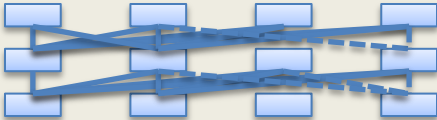
Parantap Lahiri

Microsoft Online Services Division

Paradigm Shifts : On-Premises Vs Web-Scale

DC -1 — Primary

Application	(4-5)X9
Network	4X9
Datacenter	5X9

Aggregate = 4X9

DC -2 — Backup

Application	(4-5)X9
Network	4X9
Datacenter	5X9

Aggregate = 4X9

- On-premises application MTBF optimized
- Skilled System Administration needs

- Complicated Customized Highly Available Network
- High dependence on specific solution

- Leased Gold Plated Datacenter Co-Lo with 5X9 SLA on power, cooling

Customized Expensive Manual

Traditional Enterprise/Hosting

(3-5)X9

Service Availability %	Downtime per year
99.99% ("four nines")	52.56 minutes
99.999% ("five nines")	5.26 minutes
99.9999% ("six nines")	31.5 seconds

(4-5)X9

10,000 seats	Locally Optimized	Centralized App DNA (MTBF)	Customized	Manual Dependency
Scale	**Performance Scope**	**Service**	**Network Solution**	**Operations**
10,000,000 seats	Globally Optimized	Distributed App DNA (MTTR)	Consistent & Commoditized	Heavily Automated

Application

Web-Scale Sauce

Network	Network	Network	Network	Network
DC 01	DC 02	DC 03	DC 04	DC N

(4-5)X9
- Application MTTR optimized
- Massive Scaling Property

3X9
- Simple Consistent Network with High BW
- Heavily Automated and Cost-Optimized

3X9
- Heavily Automated
- Highly Efficient

Site-Up Cost-Down

Web-Scale/Cloud Based

What this means to Network

	Traditional	Web-Scale
Platforms	•Complex Chassis Based System •High power draw per Gig	•Simplistic appliance or basic chassis HW •High Port Density •Lower power draw per Gig •Less proportional space
Ports	•Architecturally lower number of port counts needed •Expensive Port Cost	•High number of port counts in each deployment •Commoditized port cost – Ethernet is the de-facto technology
Availability	•Difficult to upgrade device to introduce feature or fix bugs(Hitless Failover) •Service availability inter-twined network availability	•Code upgrade without uptime impact (Application routes around failures) •Network Robust and Up to date
Architecture	•Basic Tree-topology •Complex Control Plane Protocols •Customization prevents heavy automation	•Heavily inter-connected design •Robust to single failures •Simplistic Control Plane Protocol – Less protocol interaction and dependency •Heavily Automated Deployment and Operations
Bandwidth Needs	•Optimized for N-S bandwidth, typically heavily oversubscribed	•Heavy E-W and N-S bandwidth •Web-Sauce needs very high bandwidth between servers to remain in sync
Network Technology Refresh Cycle	•Disruptive and expensive	•Application has fewer dependencies, so easier to upgrade •Cheaper to upgrade since components are commoditized

Web-Scale Deployment : Network as a Differentiator

Facts

Applications and servers simply perform better with more bandwidth!

Applications continually morph and consume additional available bandwidth

Servers are lot more expensive resource than network

Build Philosophies

Network cannot be the bottleneck starving available server resources for application

Build the largest scaled-out network possible

Reduce network requirements to embrace commodity offerings

- If it doesn't "have to" reside in network then it shouldn't be there. e.g. LB, DPI, Encryption, FW etc.

Develop architecture that can support future technology refreshes

How you can help us?

Evolve technology with large scale DCs in mind

- Ethernet evolution should consider the large scale datacenter cable plants, e.g. 10GBase-SR is 300m but 100G/40G technologies are 100m optimized; cabling plants and distribution frames don't change easily.
- Make choices that lead to higher density switches, less real estate for interfaces on face plate
- Simpler offering with high bandwidth is attractive, delaying evolution cycle for accommodating complex corner cases is not helpful
- Help us build datacenter fabric BW 4-10 times larger than the server NIC BW. Currently they are merging on 10G due to commodity pressure

Leverage the demand cycle of large scale DCs

- Large scale DCs will uniformly adopt a speed the moment it hits the commodity price range.
- At the same time, large scale DC adoption will lead to automatic huge volume uptake driving commoditization

Be cognizant of platform integration needs

- Silicon and optics should attain commodity cost points at similar time, one or the other doesn't help.

Ethernet standards interest

- Primarily focused on using Ethernet as a link technology
- Not high interest in Ethernet network level technology

Packet-Optical Integration

- Provide inter-operable 100GE colored optics for DWDM systems

Thank you