

เตรียมความพร้อมของโครงสร้างพื้นฐาน เพื่อ รองรับ Data Platform & AI ในอนาคต

Sagsarij Akirahchajchaval
Enterprise Sales Manager



- 1 ระบบจัดเก็บข้อมูลที่เหมาะสมเพื่อรองรับ DATA ขนาดใหญ่ สำหรับงาน AI
- 2 ระบบเครือข่ายที่เหมาะสมเพื่อรองรับงาน AI
- 3 Wifi 7 กับ Edge AI



/01

ระบบจัดเก็บข้อมูลที่เหมาะสมเพื่อรองรับ

DATA ขนาดใหญ่ สำหรับงาน AI

ปัจจุบันแนวโน้มของปริมาณข้อมูลเพิ่มขึ้นอย่างต่อเนื่อง



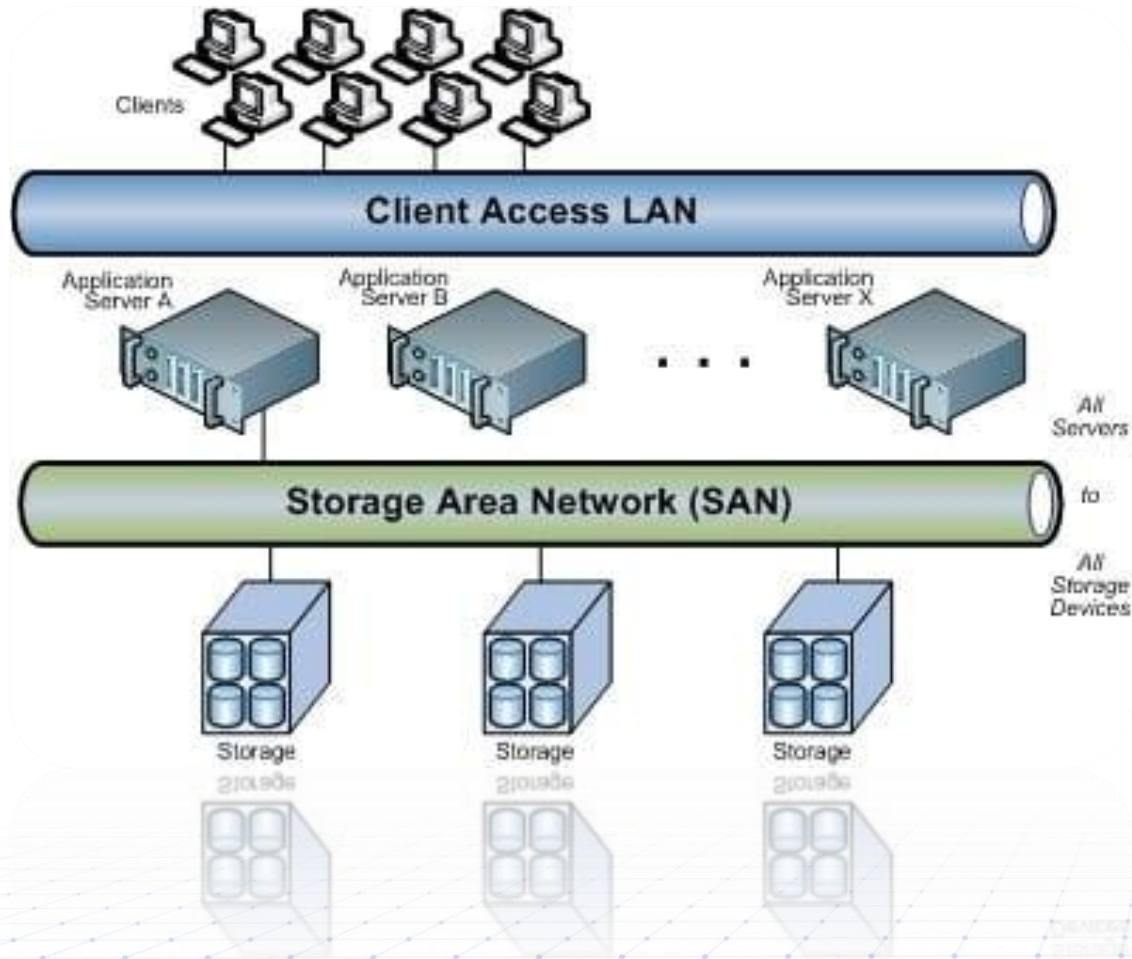
GB

TB

PB/EB

ในยุคดิจิทัล ปริมาณข้อมูลเพิ่มขึ้นอย่างต่อเนื่อง!

ระบบจัดเก็บข้อมูลแบบเดิม (Traditional Storage)



ประกอบไปด้วย 3 องค์ประกอบหลัก ได้แก่

- 1 Server
- 2 SAN Switch
- 3 SAN Storage

ข้อจำกัดของระบบจัดเก็บข้อมูลแบบเดิม



1. ประสิทธิภาพ

เมื่อมีการเก็บข้อมูลเป็นหลัก Petabyte จะมีประสิทธิภาพที่ไม่เพียงพอและไม่สามารถตอบสนองข้อมูลเชิงลึกที่เราต้องการได้

2. การเพิ่ม/ขยาย

เมื่อต้องการขยายระบบจัดเก็บข้อมูล จะเป็นการขยายแบบ “Scale-up” ซึ่งประสิทธิภาพในการใช้ทรัพยากรจะได้น้อยกว่าการขยายแบบ “Scale-out” และเมื่อเพิ่มระบบใหม่ ต้องมีการย้ายข้อมูล และจัดการข้อมูล ซึ่งยิ่งขยายมาก การบริหารจัดการก็ยากขึ้น

3. ต้นทุน

ต้นทุนในการขยายระบบสูง เพราะต้องมีต้นทุนของ Controller, Expansion และ Disk

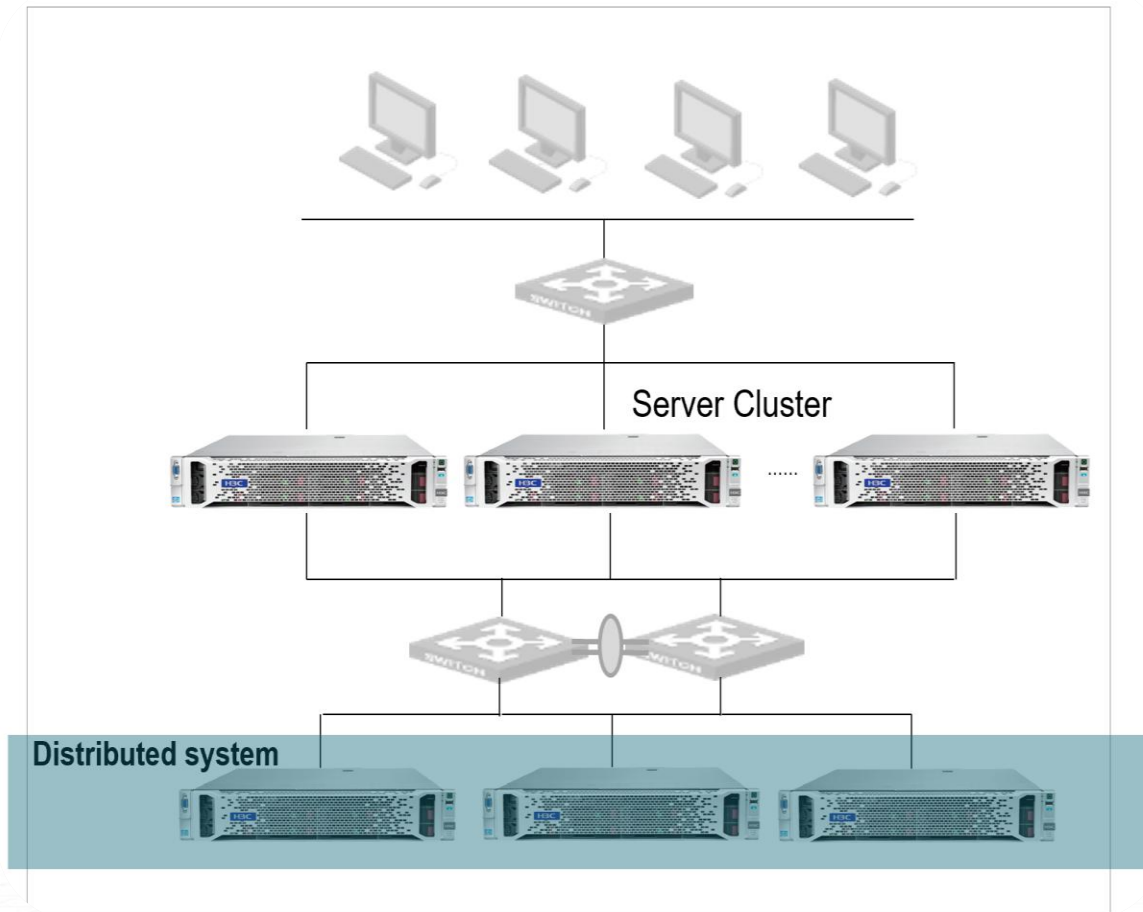
4. การดูแลรักษา

ทำได้ยากมีความซับซ้อน

เทคโนโลยีอะไรที่สามารถทดแทนรูปแบบการจัดเก็บข้อมูลแบบเดิมได้



Software Defined Storage (SDS)



Software Defined Storage (SDS)

คือการใช้ Software เพื่อรวม Server หลายเครื่องให้ทำงานร่วมกันเหมือนเป็นระบบเก็บข้อมูลขนาดใหญ่เพียงระบบเดียว

The Future of Storage is Software-Defined

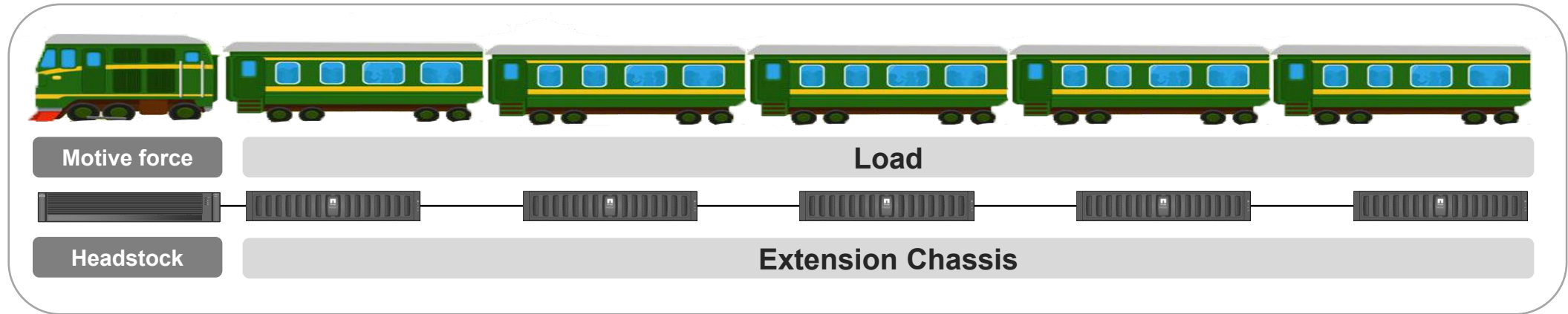


- **Gartner's** Strategic Roadmap for Storage states that **50% of global storage capacity will be deployed as SDS on-premises or in the public cloud by 2024** (up from less than 15% in 2020).
- **IDC** forecasts that the **SDS market will grow at an 8.8% compound annual growth rate (CAGR) over the next 5 years to reach \$26.4 billion in 2024.**

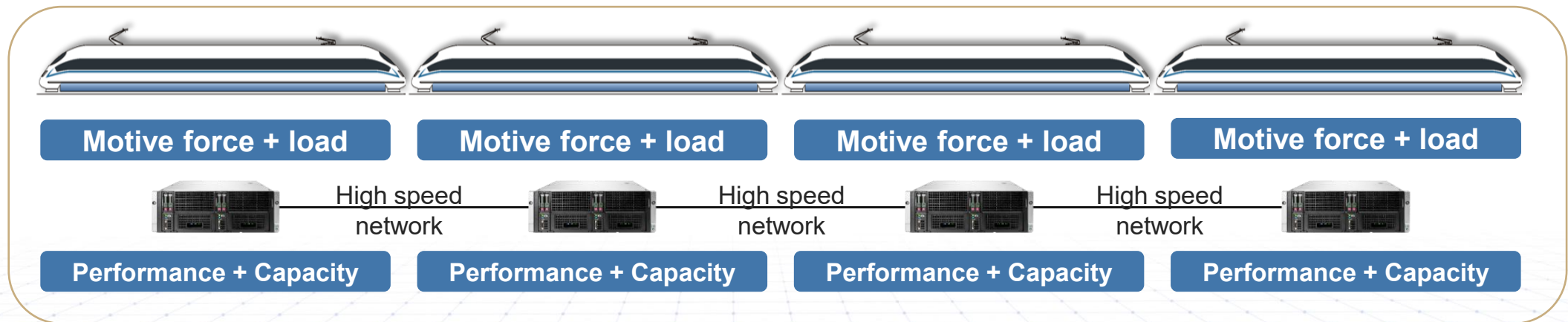
เปรียบเทียบ SDS vs Traditional Storage



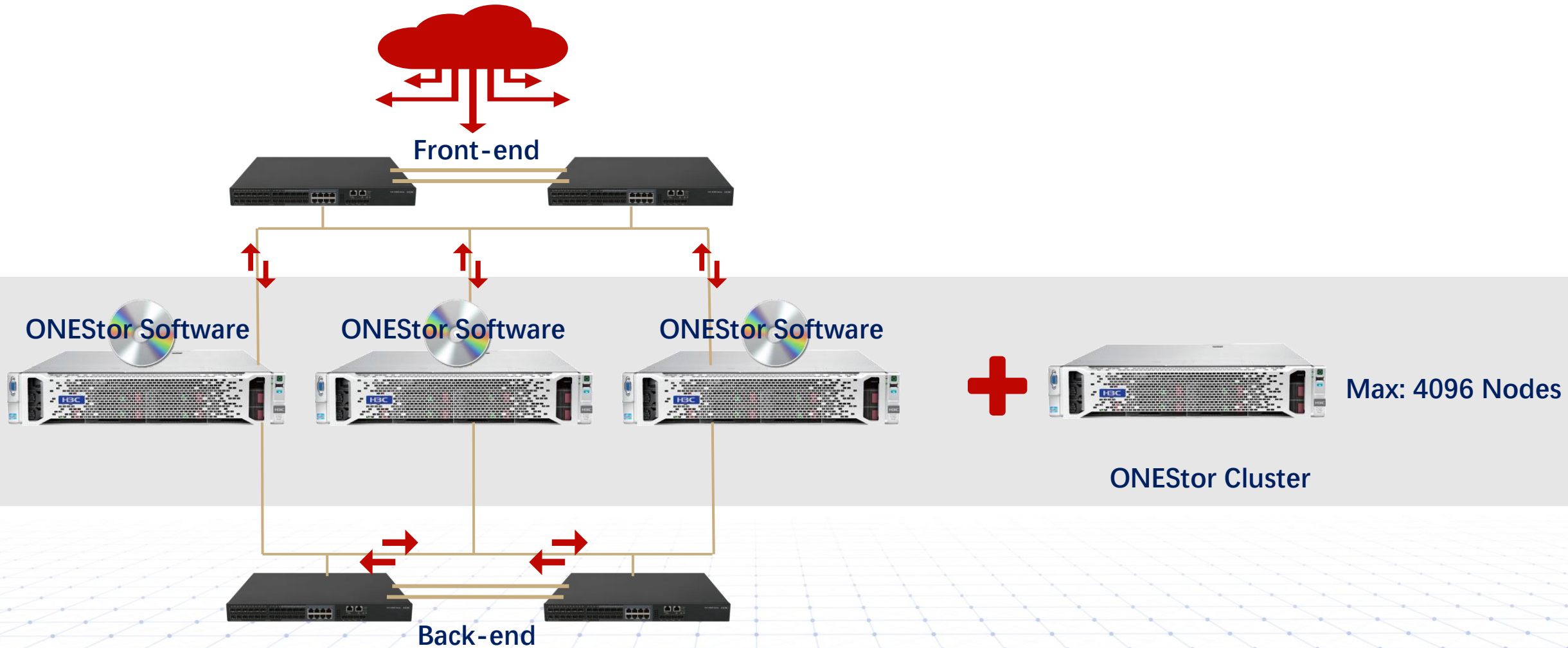
Traditional
architecture



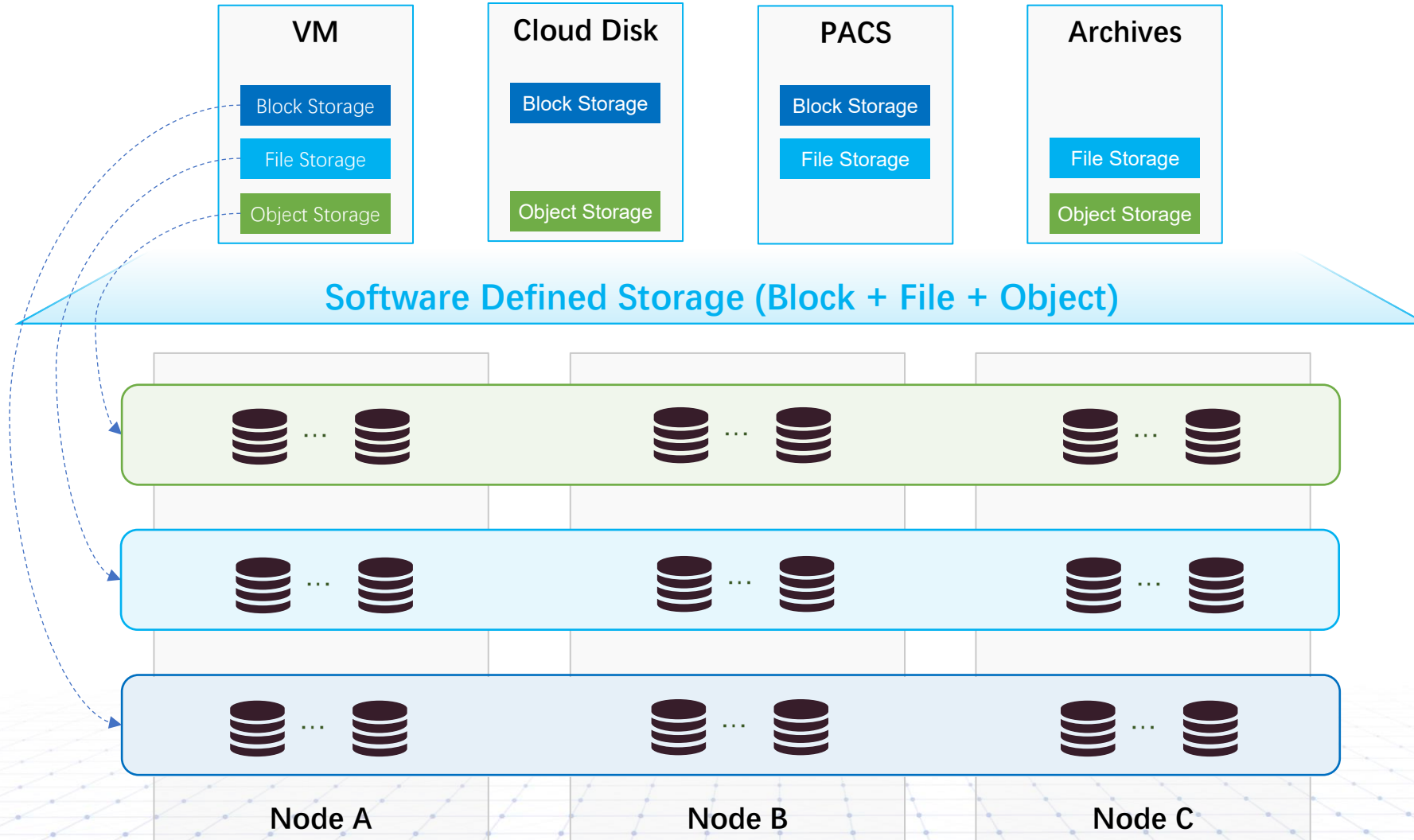
Distributed
architecture



สถาปัตยกรรม Software Defined Storage



Unified Storage Services (Block + File + Object)



ข้อดี Software Defined Storage



Performance

- ประสิทธิภาพจะเพิ่มขึ้นตามจำนวนโหนด (Server) ที่เพิ่มเข้าไปในระบบ
- การทำงานพร้อมกันหลายโหนด (Multi-node concurrency) ช่วยเพิ่มประสิทธิภาพในการอ่านและเขียนข้อมูล



Scalability

- สถาปัตยกรรมแบบกระจายศูนย์ (Distributed) ของ SDS ทำให้มีความสามารถในการขยายระบบได้มาก สามารถขยายความจุได้ง่ายถึงระดับเพตะไบต์ (PB - Petabyte)



Cost-effective

- ใช้ Server มาตรฐานทั่วไป (x86 servers) ทำให้การขยายระบบในอนาคตมีต้นทุนที่คุ้มค่ากว่าการใช้ฮาร์ดแวร์เฉพาะทาง



Necessary

- Cloud Data Center
- Massive Data
- การจัดเก็บข้อมูลแบบ Object Storage หรือ File Storage
- ระบบที่ต้องการรองรับการเข้าใช้งานพร้อมกันจำนวนมาก (High concurrency scenarios)



/02

ระบบเครือข่ายที่เหมาะสมเพื่อรองรับ

AI Workload

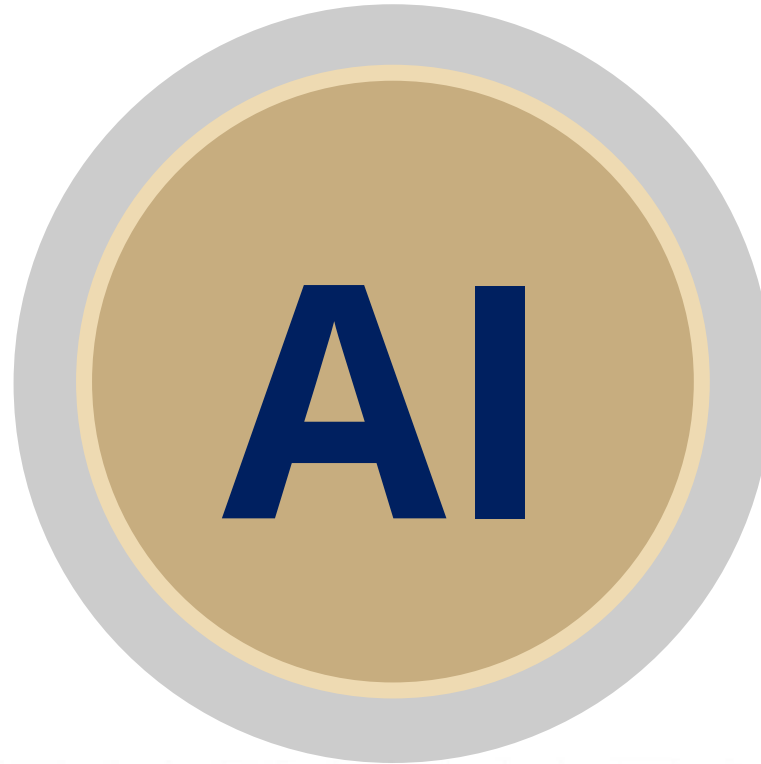
แนวทางการเตรียมความพร้อมองค์กรสู่ยุค AI



1. กำหนดวิสัยทัศน์และกลยุทธ์
AI (Strategy & Vision)

2. เตรียมความพร้อมด้านข้อมูล
(Data Readiness)

3. พัฒนาบุคลากรและทักษะ
(Talent & Skills)



4. ลงทุนในเทคโนโลยีและโครงสร้าง
พื้นฐาน
(Technology & Infrastructure)

5. กำหนดกรอบจริยธรรมและธรรมา
ภิบาล AI
(Ethics & Governance)

6. สร้างวัฒนธรรมองค์กรที่พร้อมรับการ
เปลี่ยนแปลง
(Culture & Change Management)



ประเมินโครงสร้างพื้นฐานปัจจุบัน :

- ตรวจสอบว่าระบบ IT และโครงสร้างพื้นฐานปัจจุบันสามารถรองรับการใช้งาน AI ได้หรือไม่ (เช่น ระบบประมวลผล, ระบบจัดเก็บข้อมูล, เครือข่าย)



พิจารณา Cloud Computing :

- Cloud Platform มักมีความยืดหยุ่นและเครื่องมือ AI สำเร็จรูปที่ช่วยให้เริ่มต้นได้ง่ายขึ้น



เลือกเครื่องมือและแพลตฟอร์ม AI:

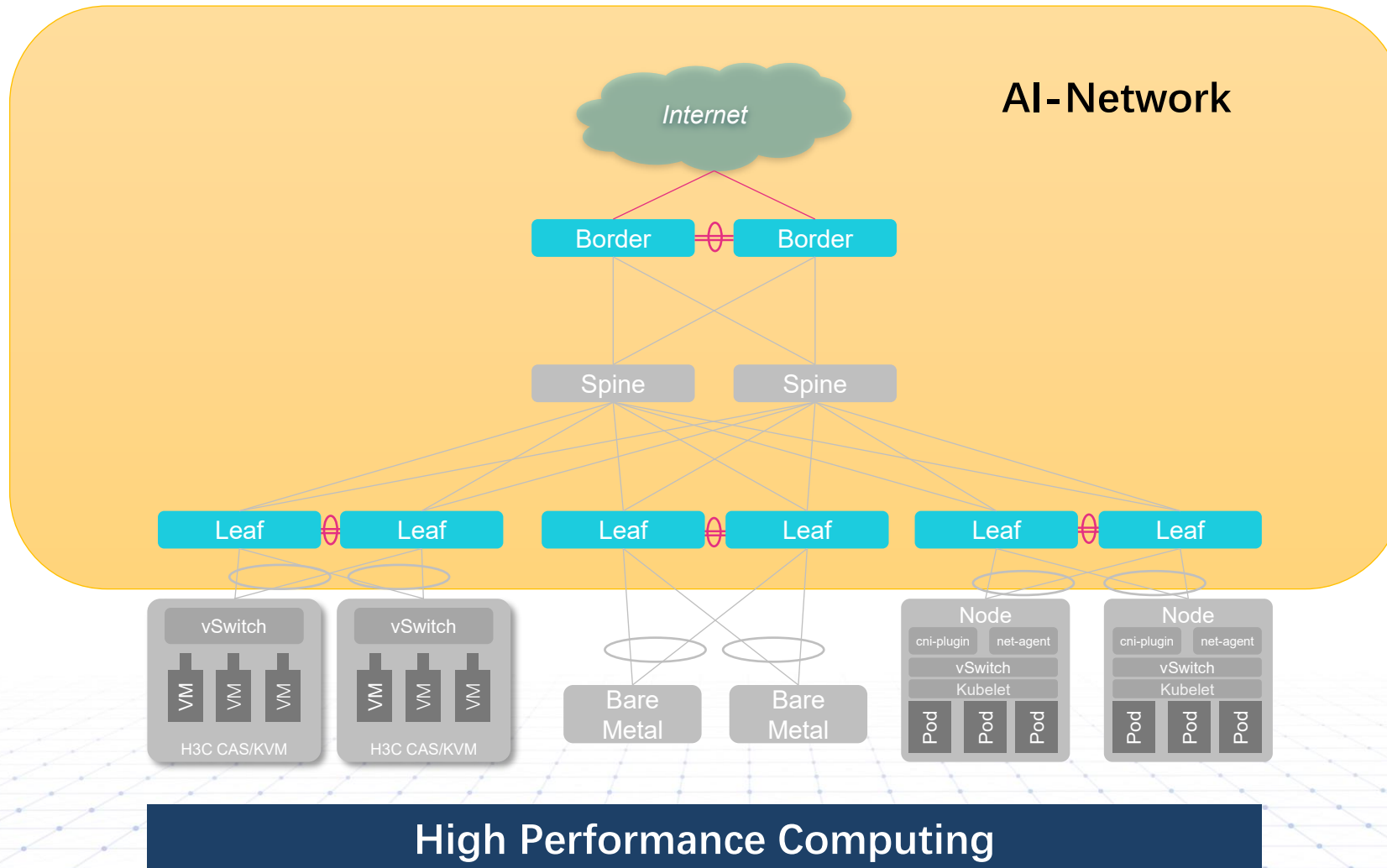
- ศึกษาและเลือกใช้เครื่องมือ (Tools) หรือแพลตฟอร์ม (Platforms) AI ที่เหมาะสมกับความต้องการและงบประมาณขององค์กร



คำนึงถึงความปลอดภัย:

- วางแผนและลงทุน ในเรื่อง Cyber Security สำหรับระบบ AI และข้อมูลที่เกี่ยวข้อง

ระบบเครือข่ายที่เหมาะสมเพื่อรองรับ AI Workload



RDMA over Converged Ethernet (RoCEv2)



เป็น Network Protocol ที่ช่วยให้สามารถ ส่งข้อมูลโดยตรงจากหน่วยความจำของเครื่องหนึ่งไปยังอีกเครื่องหนึ่ง ผ่านเครือข่าย Ethernet โดยไม่ต้องใช้ CPU มาก ซึ่งช่วยลด Latency และเพิ่มประสิทธิภาพของ Application ที่ต้องการความเร็วสูงและการรับส่งข้อมูลขนาดใหญ่

ข้อดี

- Latency ต่ำ ($\sim 1-2 \mu s$) : เร็วกว่าการใช้ TCP/IP ทั่วไป
- ใช้ CPU ต่ำกว่า TCP/IP : เพิ่มประสิทธิภาพโดยให้ NIC จัดการข้อมูลเอง
- รองรับ Layer 3 (IP) : สามารถใช้กับเครือข่ายที่ต้องมี Routing ได้
- รองรับ QoS และ Flow Control : ป้องกัน Packet Loss และ Network Congestion
- ใช้ได้กับ Ethernet มาตรฐาน : ต้นทุนต่ำกว่า InfiniBand

Tolly Test Report – RoCE vs IB



#225111

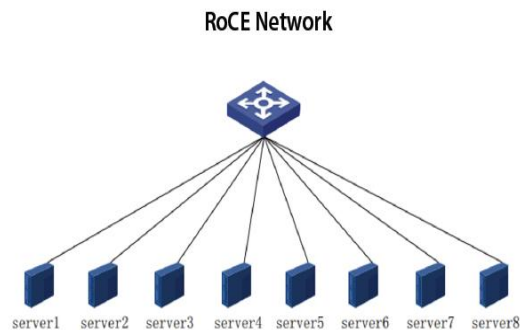
January 2025

Commissioned by
New H3C Technologies Co., Ltd.

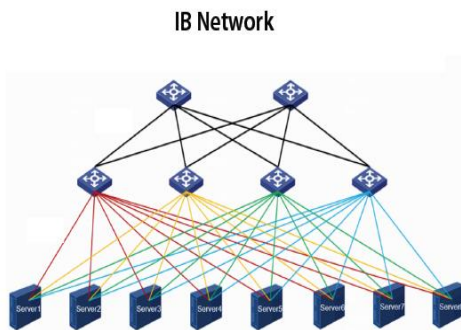
64 GPU AI Computing Performance Comparison Test

H3C RoCE Network (S12500CR Series Switches) vs. InfiniBand Network

AI Model Training Test Network Topology



Each line represents 8 × 400G Ethernet links



Between Spine and Leaf, each line represents 8 × 400G links;
Between Leaf and servers, each line represents 2 × 400G links.

Note: Each server is equipped with 8 NVIDIA H20 GPU cards.

Source: Tolly, November 2024

Figure 1

*Report link: <https://reports.tolly.com/publications/detail/225111>.

RoCE vs. IB - NCCL Ring-AllReduce Results

Size	Bus Bandwidth (busbw) - GB/s		
	H3C S12500CR switch RoCE network	NVIDIA QM9700 switch InfiniBand network	RoCE vs. IB
4MB	19.20	19.30	-0.52%
8MB	37.99	36.52	+4.03%
16MB	71.77	71.48	+0.41%
32MB	114.19	116.11	-1.65%
64MB	165.53	165.37	+0.10%
128MB	214.24	211.91	+1.10%
256MB	255.24	255.08	+0.06%
512MB	311.39	288.36	+7.99%
1GB	349.08	354.15	-1.43%
2GB	369.31	361.13	+2.27%
4GB	368.26	368.68	-0.11%
8GB	371.41	369.19	+0.60%
16GB	372.59	371.22	+0.37%
Average	232.60	230.698	+0.82%

Tolly Test Report – RoCE vs IB



RoCE vs. IB Llama3

Llama3 is an open-source large language model developed by Meta AI, with Llama3 70B referring to a model with 70 billion parameters.

Tolly engineers conducted Llama3_70B training using the topology in Figure 1, comparing the IB network (with NVIDIA QM9700 switches) and the RoCE network (with H3C S12500CR switches). In the test results, a lower average iteration time is

better, and a higher samples/s value is preferred.

The results indicate that the training performance of the RoCE network is nearly identical to that of the IB network. Detailed results can be found in Table 2.

RoCE vs. IB Summary

The test results for NCCL and the large language model Llama3 demonstrate that, under the same workload scenarios, RoCE delivers performance comparable to IB and provides a consistent user experience.

RoCE vs. IB - Large Language Model (LLM) Training Performance Results

Network	Number of Nodes	Model	Number of Iterations	Average Iteration Time (ms)	samples/s
NVIDIA QM9700 switch InfiniBand network	8	Llama3_70B	10	17,724	14.44
H3C S12500CR RoCE network	8	Llama3_70B	10	17,665	14.49

Source: Tolly, November 2024

Table 2



/03

WiFi 7 กับ Edge AI

Wi-Fi 7 คือ



Wi-Fi 7 เป็นมาตรฐานเทคโนโลยี Wi-Fi รุ่นใหม่ มีอีกชื่อที่รู้จักกันคือ 802.11be ซึ่งมีเป้าหมายในการเพิ่มการรองรับปริมาณการใช้งานให้สูงขึ้น

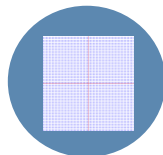
Wider Channel Bandwidth

320
MHz

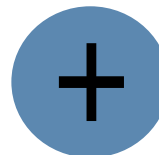


Automatic frequency coordination

4096 QAM



Wi-Fi 7



16x16 MU-MIMO

Multi-Link Operation



Tri-Band Operation
(2.4 GHz, 5 GHz, 6 GHz)

ตารางเปรียบเทียบ Wi-Fi 7 กับ Wi-Fi 6



Standard	802.11n Wi-Fi 4	802.11ac Wi-Fi 5	802.11ax Wi-Fi 6/6E	802.11be Wi-Fi 7 R1
Operating frequency	2.4GHz, 5GHz	5GHz	2.4GHz, 5GHz (Wi-Fi 6E 6GHz)	2.4GHz, 5GHz, 6GHz
Theoretical maximum speed.	600Mbps	6900Mbps	9600Mbps	46Gbps
Channel bandwidth	20/40 MHz	20/40/80/160/80+80 MHz	20/40/80/160/80+80 MHz	20/40/80/160/ 240/320 /80+80/ 160+80/160+160 MHz
MIMO	4X4	8X8	8X8	16X16
Modulation mode	64-QAM	256-QAM	1024-QAM	4096-QAM

Wi-Fi 7 กับ Edge AI



ความต้องการที่เพิ่มขึ้นสำหรับ AI Application ที่ต้องมีคุณสมบัติแบบ Real-time, Latency ต่ำ และคำนึงถึง privacy ทำให้จะมีการนำ Edge AI ไปใช้เพิ่มมากขึ้น

จุดเด่นของ Wi-Fi 7 ในเรื่อง MLO , 320 MHz Bandwidth และ 4096-QAM จะช่วยในเพิ่ม Performance ของ Edge AI และ user experience

เทคโนโลยี Wi-Fi 7 จะช่วยทำให้สามารถ run multi-modal Generative AI และ language models ที่ Edge ได้

Computer vision application จะสามารถ process ข้อมูลที่ local ได้ อุปกรณ์ Edge ที่รองรับ Wi-Fi 7 สามารถ process งาน Computer vision ที่ซับซ้อนด้วยความเร็วและมีประสิทธิภาพ

Wi-Fi 7 ยังช่วยให้ Application เช่น Facial recognition, Object detection และ Automated surveillance ทำงานได้อย่างมีประสิทธิภาพ



Summary



Software-Defined Storage

AI/HPC Networking (RoCE)

Wi-Fi 7

ถ้าคุณกำลังมองหา Virtualization Platform ใหม่ เพื่อทดแทน **Mr. V**



Virtualization Platform

License per CPU socket and Perpetual License



H3C

Highly

Creative

Caring

Credible



THANKS

www.h3c.com