

Переход на NVMeoF в ЦОД – что это дает Заказчику?

Илья Коляда, инженер
kolyada.ilya@huawei.com

Huawei: Ведущий поставщик инфраструктуры информационных технологий и интеллектуальных устройств



HUAWEI — это компания, которая приносит инновации каждому человеку, каждой организации и создает новый интеллектуальный мир



194,000
Работников



96,000+
Научных
сотрудников



170+
Страны и
регионы

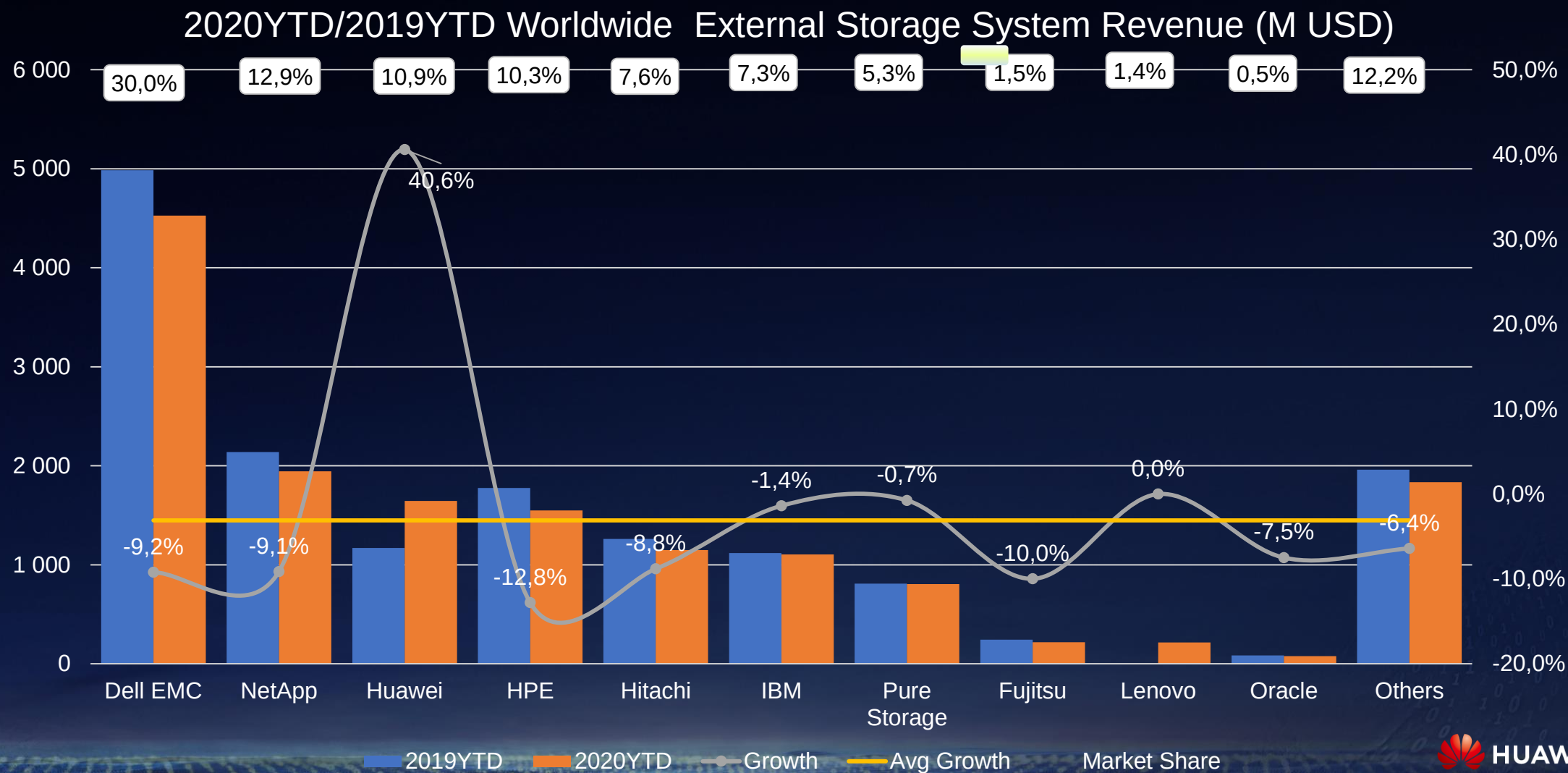


61
в Global 500

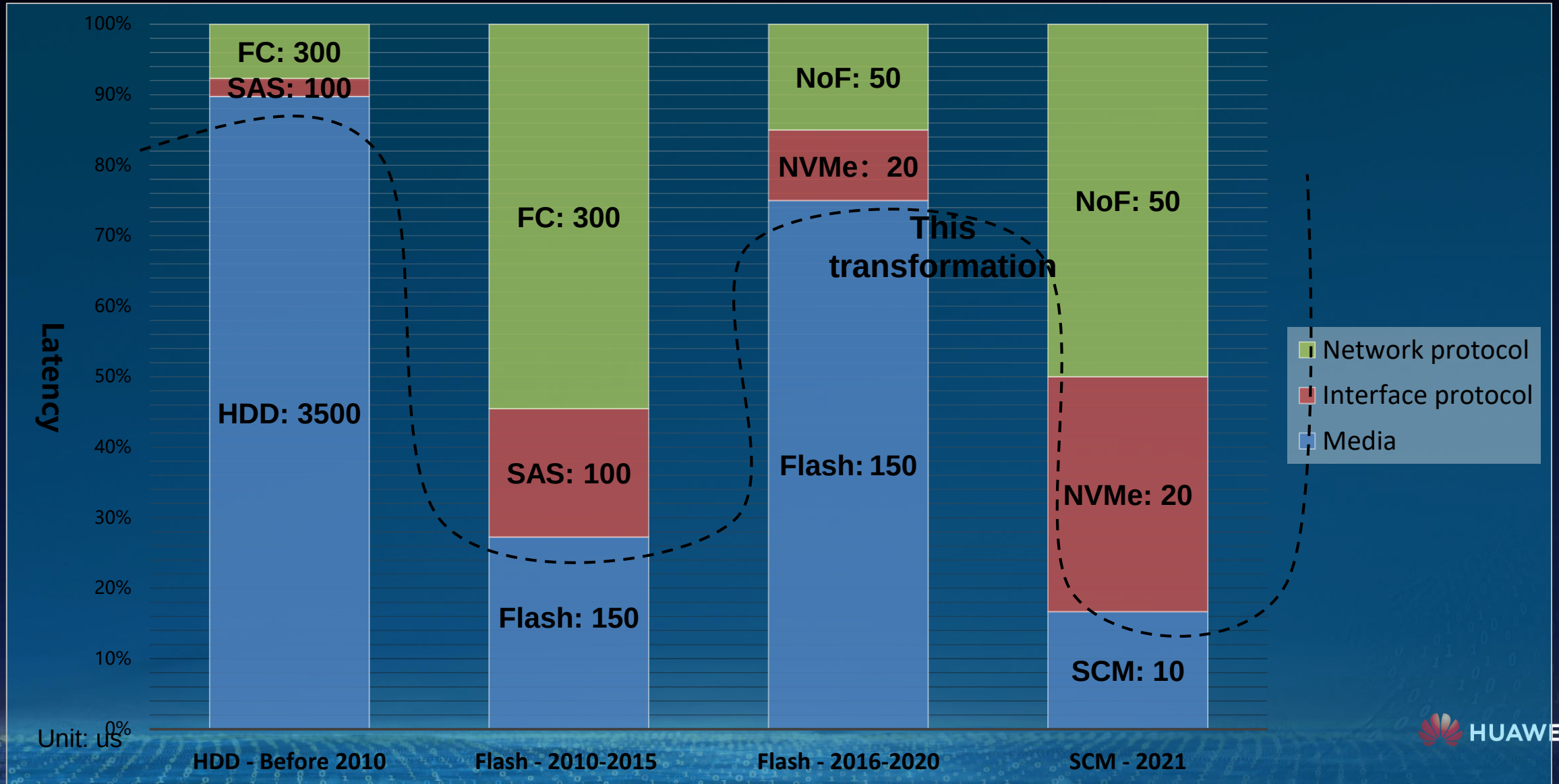
Huawei Storage Has Been Recognized by the Industry



2020YTD Мировой рынок: No. 3 по выручке, No.1 по темпам роста

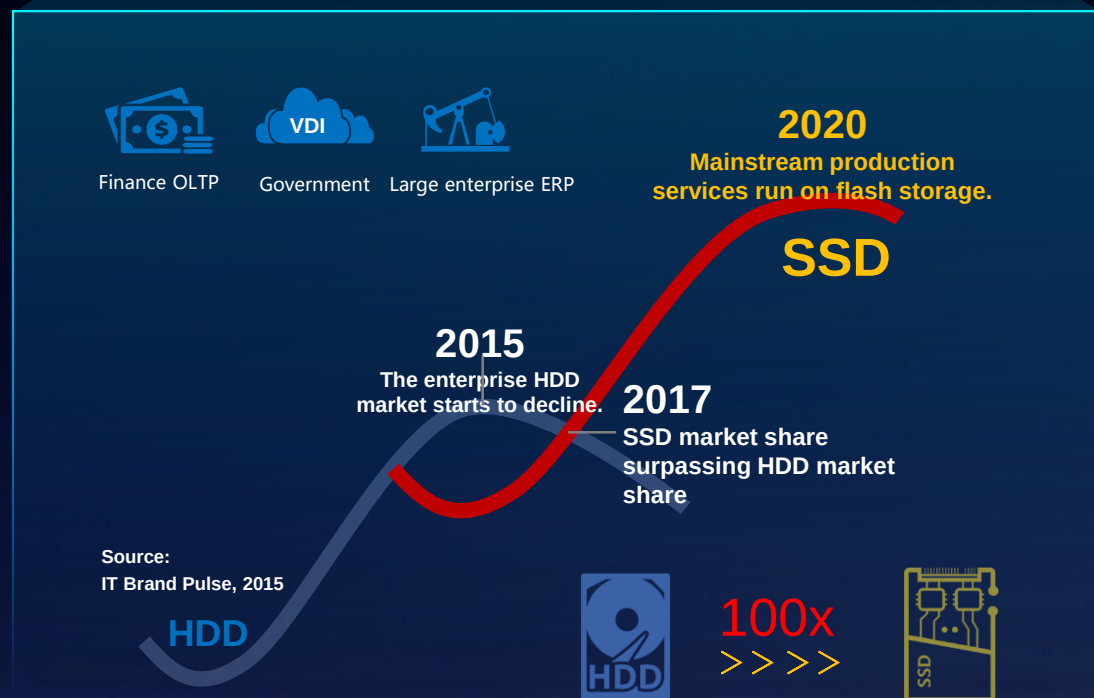


Почему NVMe: Эволюция носителей и протоколов, стимулирующих развитие систем хранения данных



Модернизация всей индустрии флэш-накопителей и требование к более быстрым сетям хранения данных

Новая эра флэш-систем



Сеть Fiber Channel не может соответствовать всем современным требованиям флэш-систем

The storage network becomes a bottleneck in the All-Flash Era

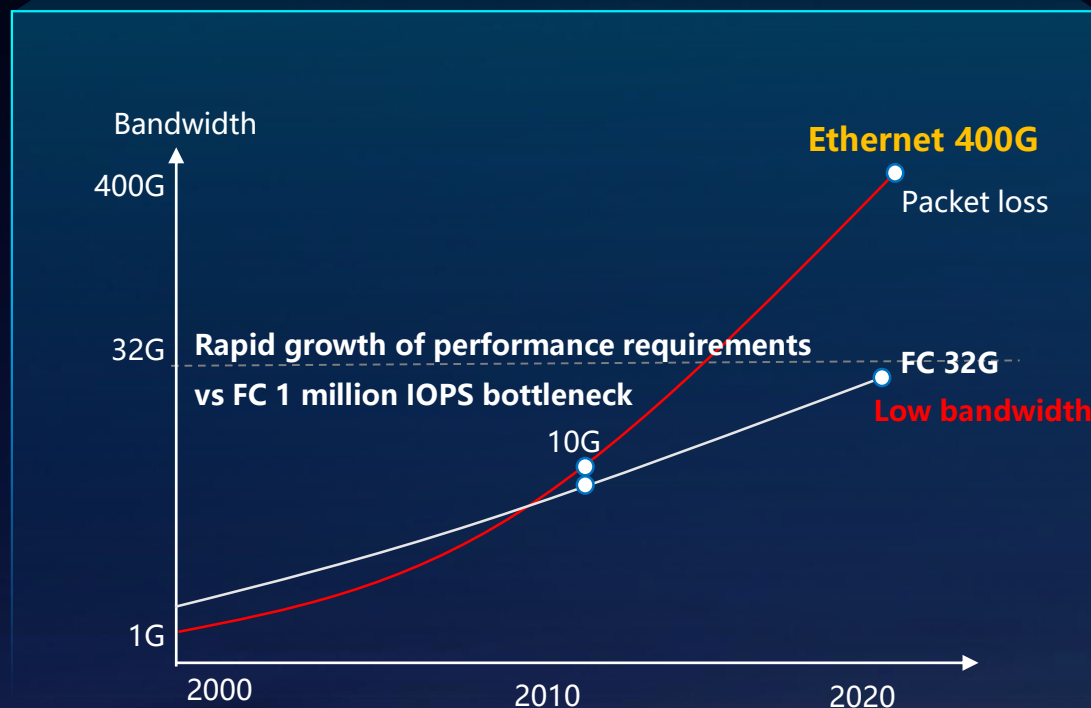


Путь 1:
Upgrading the FC
bandwidth

Путь 2:
Using high-bandwidth
Ethernet

Модернизация всей индустрии флэш-накопителей и требование к более быстрым сетям хранения данных

Сеть Fiber Channel не может соответствовать всем современным требованиям флэш-систем

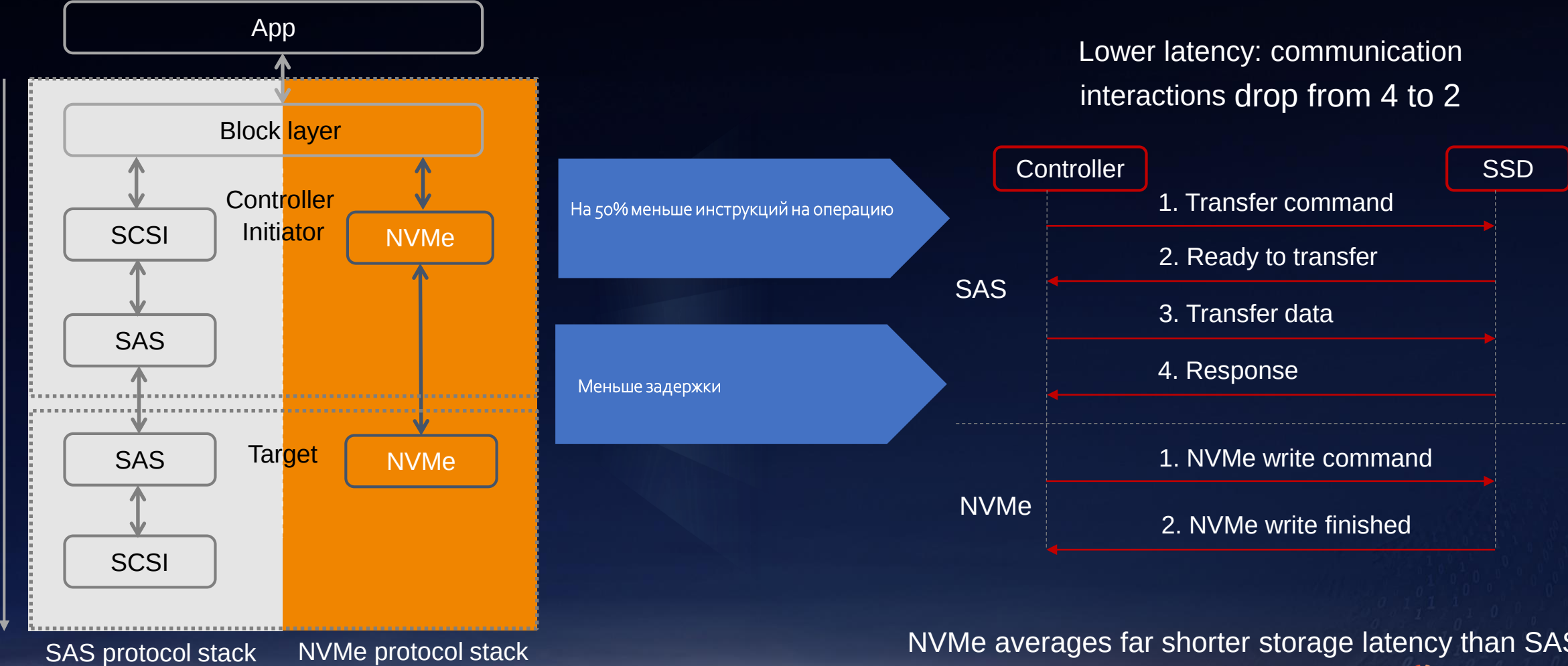


Пропускная способность FC – узкое место

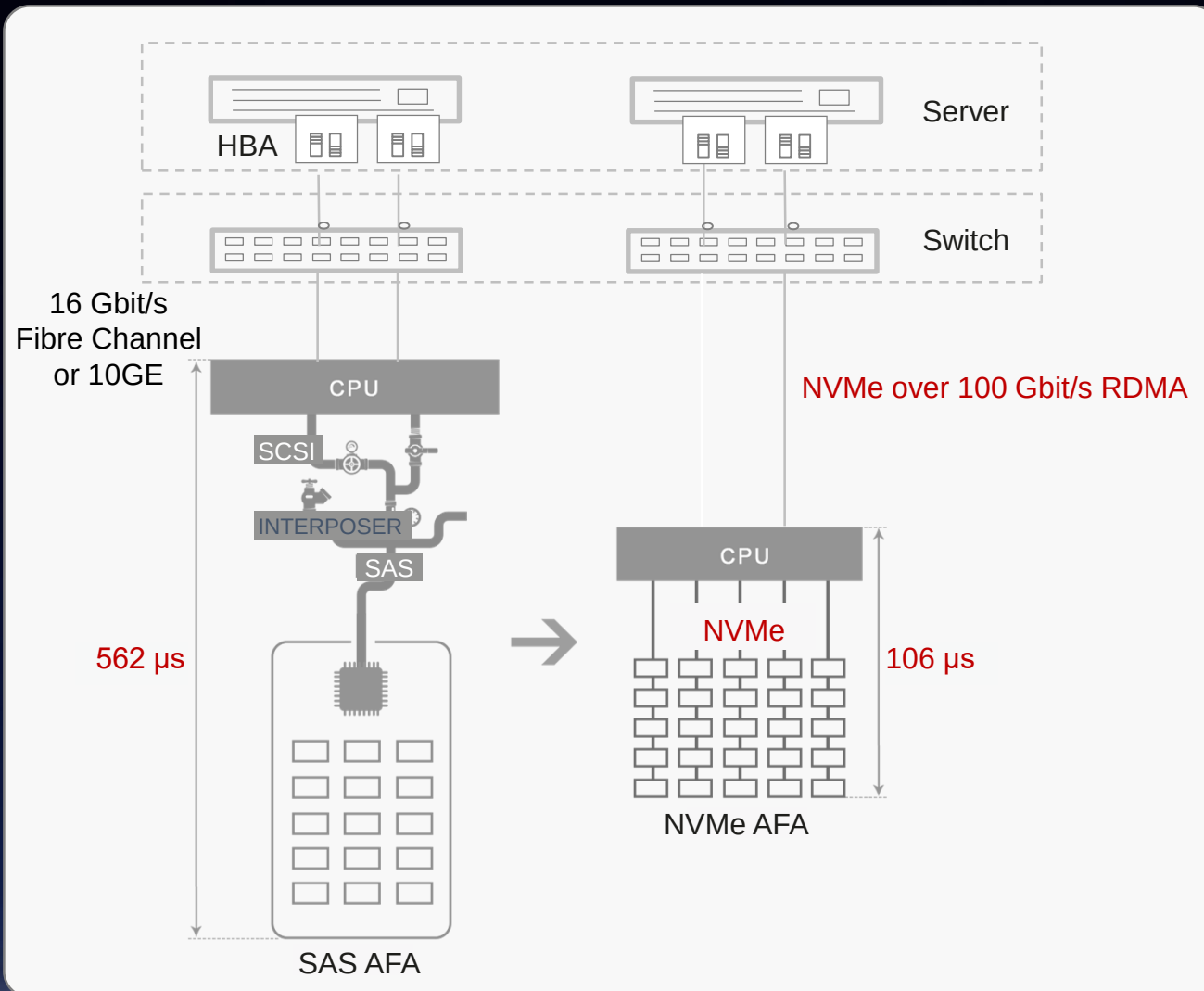
Быстрый рост требований к производительности : **x million, 30%+**
vs.
64G FC IOPS: 2 million

- 1** **Один поставщик** ➡ Риски поставки
- 2** **Высокая стоимость** ➡ Цена устройства высока, каналы WDM заняты legacy каналами
- 3** **Риски обслуживания** ➡ Поставщики пассивно реагируют на неисправности FC в живой сети

NVMe Уменьшает Задержку Обработки протокола и Ускоряет Передачу Данных



All-flash эра переход от SCSI к NVMe

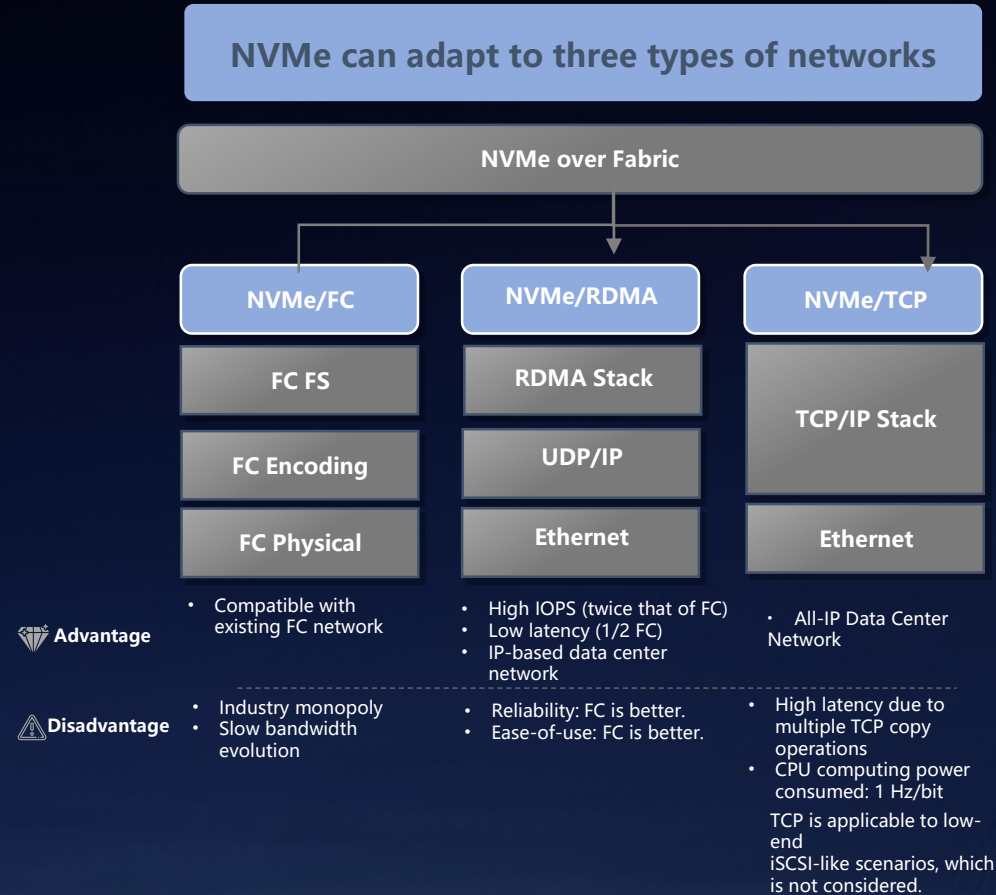
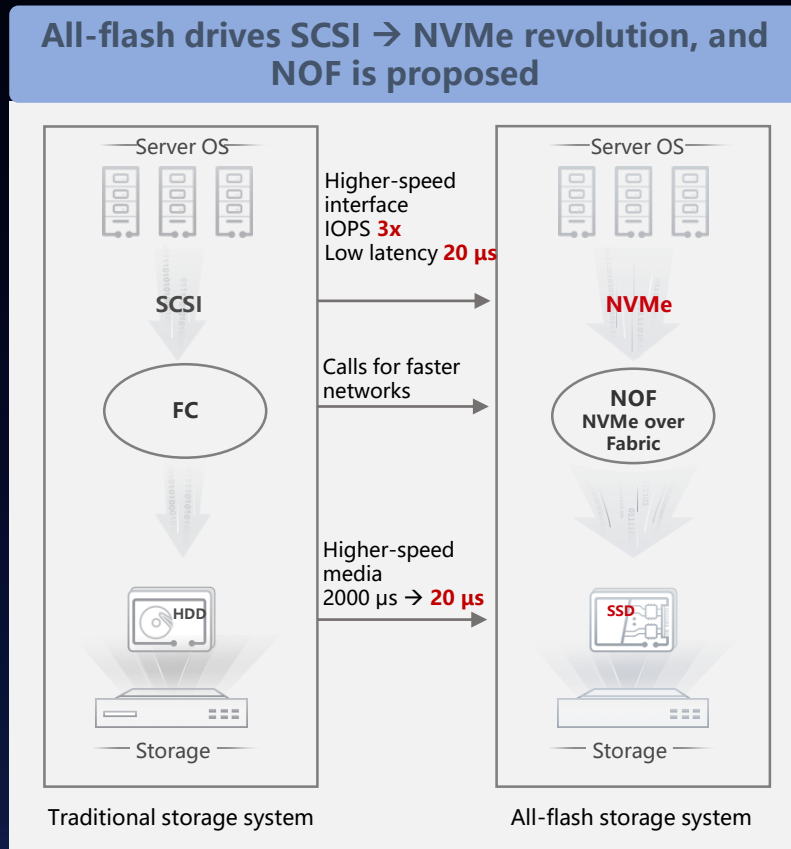


0.1 ms

Test model: 7:3 read/write, 130 μs read
latency, 50 μs write latency

$$\text{Time} = 130 \times 0.7 + 50 \times 0.3 = 106 \mu\text{s}$$

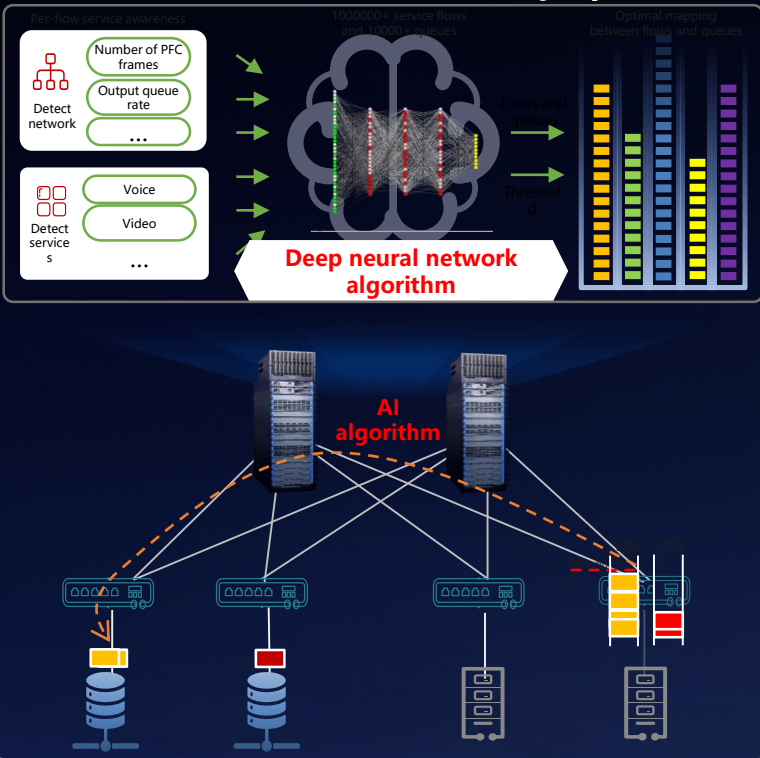
All-flash era transition from SCSI to NVMe



FC or RoCE?

Лидеры отрасли вместе с докторами наук продвигают прорыв в алгоритмах и возглавляют международные стандарты

Первое применения алгоритмов AI для оптимизации сетей в индустрии



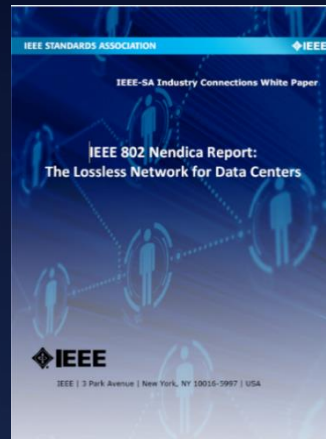
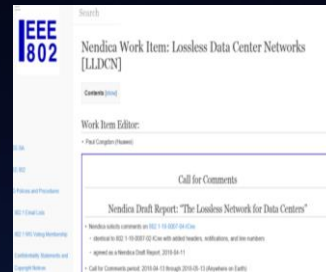
Leading the formulation of IEEE intelligent and lossless international standards



Vice-chairman of IEEE 802



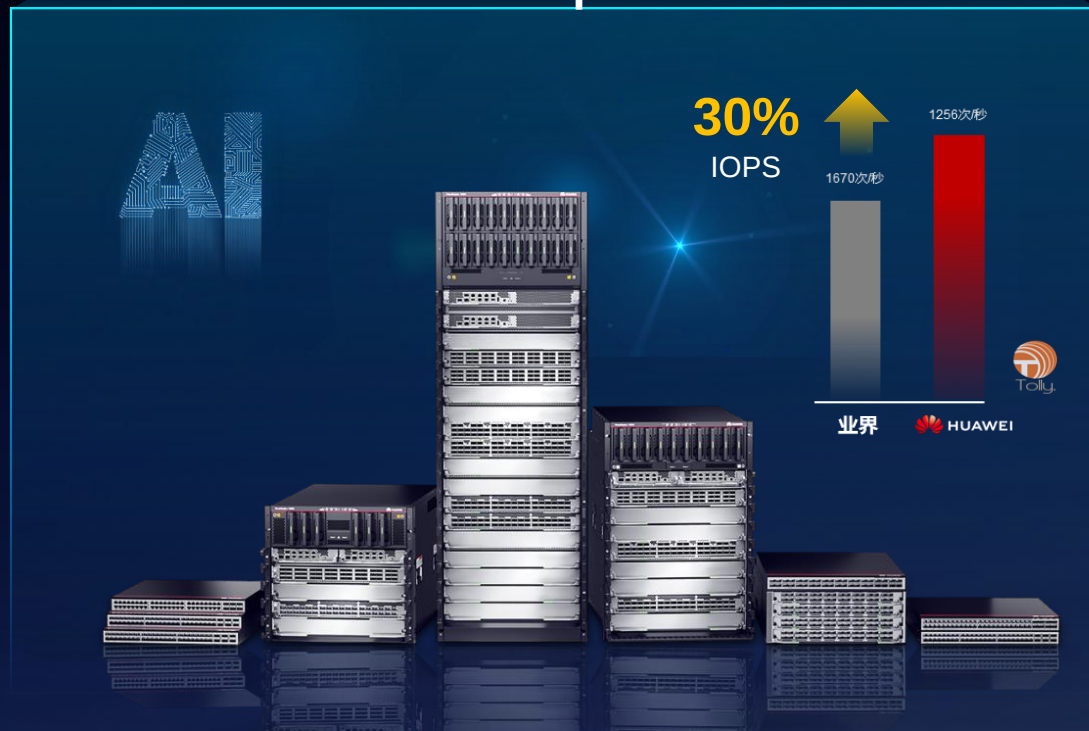
Chair of the IEEE 802.1 Task Group



<https://ieeexplore.ieee.org/document/8462819>

Первая в Отрасли Интеллектуальная Сеть ЦОД Без Потерь

Единственный в отрасли коммутатор Ethernet с нулевой потерей пакетов, интеллектуальный без потерь



Впервые используется в распределенных сетях хранения данных в 2019 году, IOPS: +30%

Enables cloud disks to provide the same performance as local disks.

Development and test

VDI

Database

Big data



Challenges: Traditional cloud disks have low performance and cannot be scalable.



NVMe Over RoCE

Реверсивные Централизованные Сети Хранения Данных

x86 Server



E2E Latency < 50us

CloudEngine Intelligent Lossless Ethernet Switch



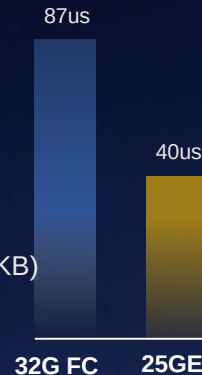
Dorado All flash storage



58% ↓

Latency optimization

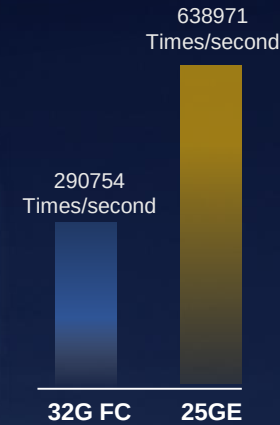
(Average random read/write latency @ 4 KB)



HUAWEI



NVMe Over
RoCE



HUAWEI

2x ↑

IOPS

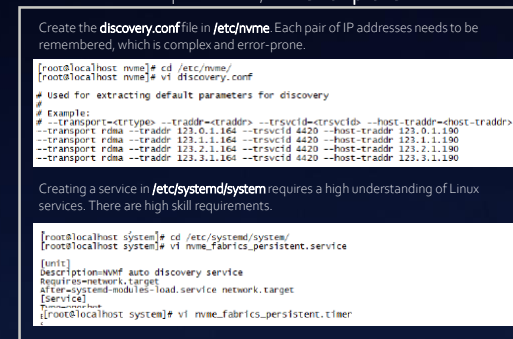
(Storage times per second IOPS@4 KB)

HUAWEI

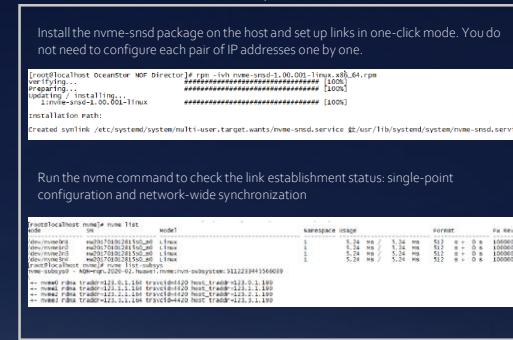
Реверсивные Централизованные Сети Хранения Данных

Единая конфигурация plug-and-play

Standard RoCE: complex configuration, high skill requirements, and **error-prone**



After improvement: one-click installation and link setup,
which is simple and efficient



NVMe Over RoCE, Лучший выбор для сети передачи данных ЦОД в новой эре флэш систем

全网网络多维度对比

— FC

— Huawei Intelligent Lossless DCN

- Выделенная сеть Fibre Channel стоит дорого, и существует только два поставщика.
- Цена каждого порта на интеллектуальном ЦОД без потерь составляет половину от порта FC

- FC необходимо управлять отдельно, что приводит к высоким операционным затратам.
- Intelligent Lossless сеть ЦОД Реализует тройную интеграцию сетей

Cost-effectiveness

Unified management

Reliability

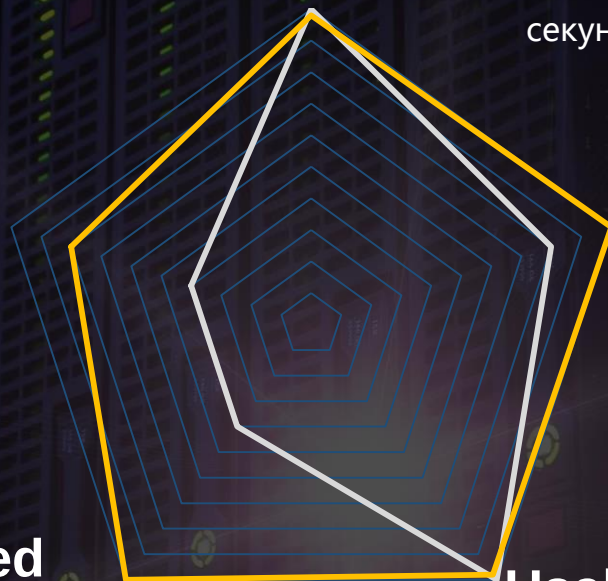
- Время сходимости неисправностей после оптимизации составляет менее 1 секунды.

Performance

- E2E задержка уменьшена на **50%+**
- IOPS производительность увеличена в **2x**

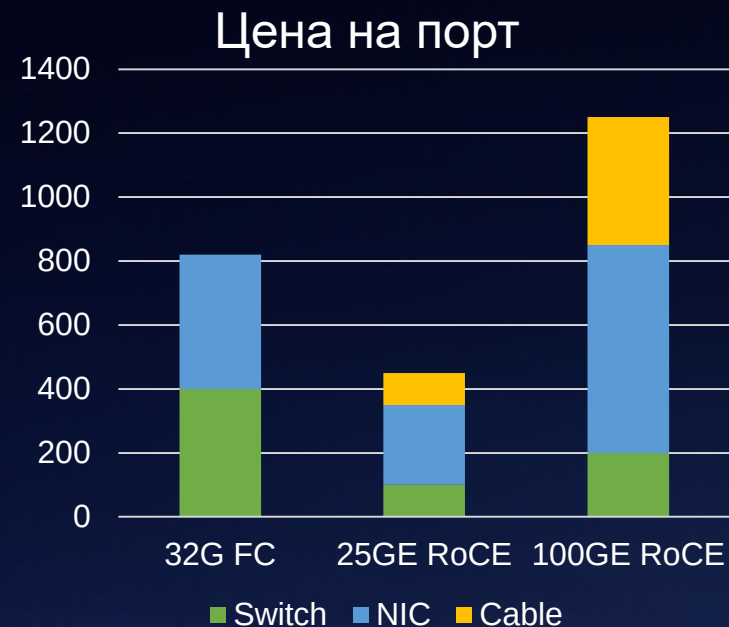
Usability

- Интеллектуальная сеть ЦОД без потерь оптимизирована и удобна в использовании.

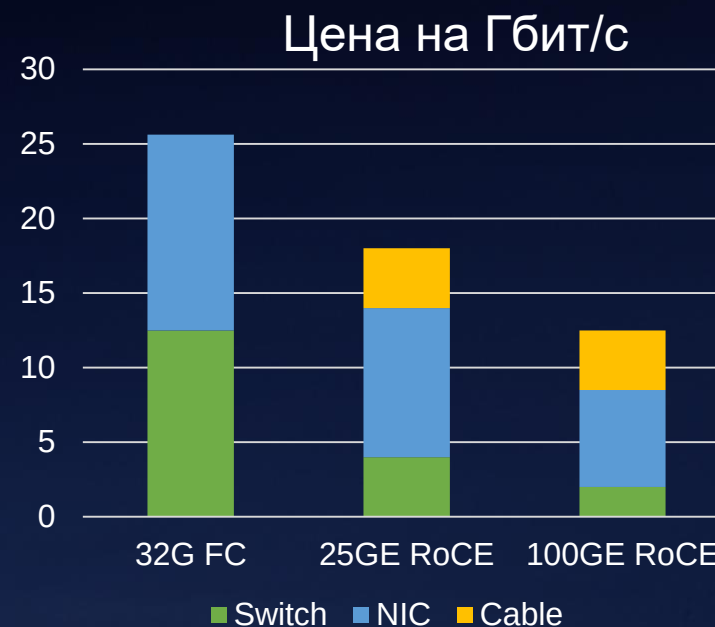


Относительная стоимость владения существенно ниже FC

Стоимость порта 25GE составляет половину от стоимости порта 32G FC.



Цена за 1 Гбит/с на 100GE составляет половину от цены 32G FC.



Huawei E2E NVMe Over RoCE Solution Launch



Инновации в области хранения данных

SMARTMATRIX
GLOBAL CACHE
INTELLIGENT DAE
HIGH DENSITY DAE STORAGE FEDERATION
HIGH DENSITY DAE RAID 2.0+ END-TO-END NVME
ONE-SECOND NDU STORAGE ACTIVE/ACTIVE CLUSTER
ENGINE LEVEL FAULT TOLERANCE
FLASHEVER SELF-DEVELOPED CHIPS SMARTMATRIX
SELF-DEVELOPED CHIPS ONE-SECOND NDU RAID 2.0+ GLOBAL CACHE
PALM DRIVE ONE-SECOND NDU STORAGE FEDERATION END-TO-END NVME
SYMMETRIC ACTIVE/ACTIVE CONTROLLER
SMARTMATRIX STORAGE ACTIVE/ACTIVE CLUSTER HIGH DENSITY DAE
PALM DRIVE FLASHEVER PALM DRIVE END-TO-END NVME
SELF-DEVELOPED CHIPS **AI ENGINE**



OceanStor Dorado

High-end competitiveness
Rock-solid architecture + leading AI tech

General-purpose mid-range SAN storage
High-end advantages + mid-range storage form factor

At least 192G for initial configuration
of entry-level AFAs



Dorado 3000 V6

Two levels for attack
defense



Dorado 5000 V6

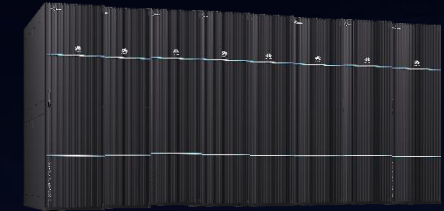
First-in-class



Dorado 6000 V6



Dorado 8000 V6



Dorado 18000 V6

Tolerates failure of 3 out
of 4 controllers/Full
interconnection/Delivery
by enclosure
Entry-level high-end
storage system

Building competitiveness based
on high-end architecture
Breaking into core applications
based on NVMe/AI tech

	Entry-Level	Mid-Range			High-End	
Model	3000	5000	6000	8000	18000	
Height / Controllers of Each Engine	2U/2C	2U/2C	2U/2C	4U/4C	4U/4C	
Controller Expansion	2-16	2-16	2-16	2-16	2-32	
Cores in Each Controller	32	64	96	128	192	
Maximum Disks	1200	1600	2400	3200	6400	
Cache/Dual Controller	128/192G	256G/512G	512G/1024G	512G/1024G/2048G	512G/1024G/2048G	
Front-end ports	8/16/32G FC, 1/10/25/40/100G Ethernet, 25/100G RDMA					
Back-end ports	100G Ethernet					

**Объединяйтесь, чтобы
двигаться навстречу
интеллектуальной эре**

**Прозрачные сети хранения
данных и вычислительные
потенциалы**

