



# Huawei Intelligent Computing Product & Solutions

Date: February 2021





HOME RESEARCH OPEN SOURCE STANDARDS CAREERS CONTACT US



Learn More

Learn More

Learn More

## About Us

Futurewei Technologies, Inc., founded in 2001 is a US corporation and an affiliate of Huawei Technologies Company Ltd. engages in research and development of information and communication technologies (ICT).

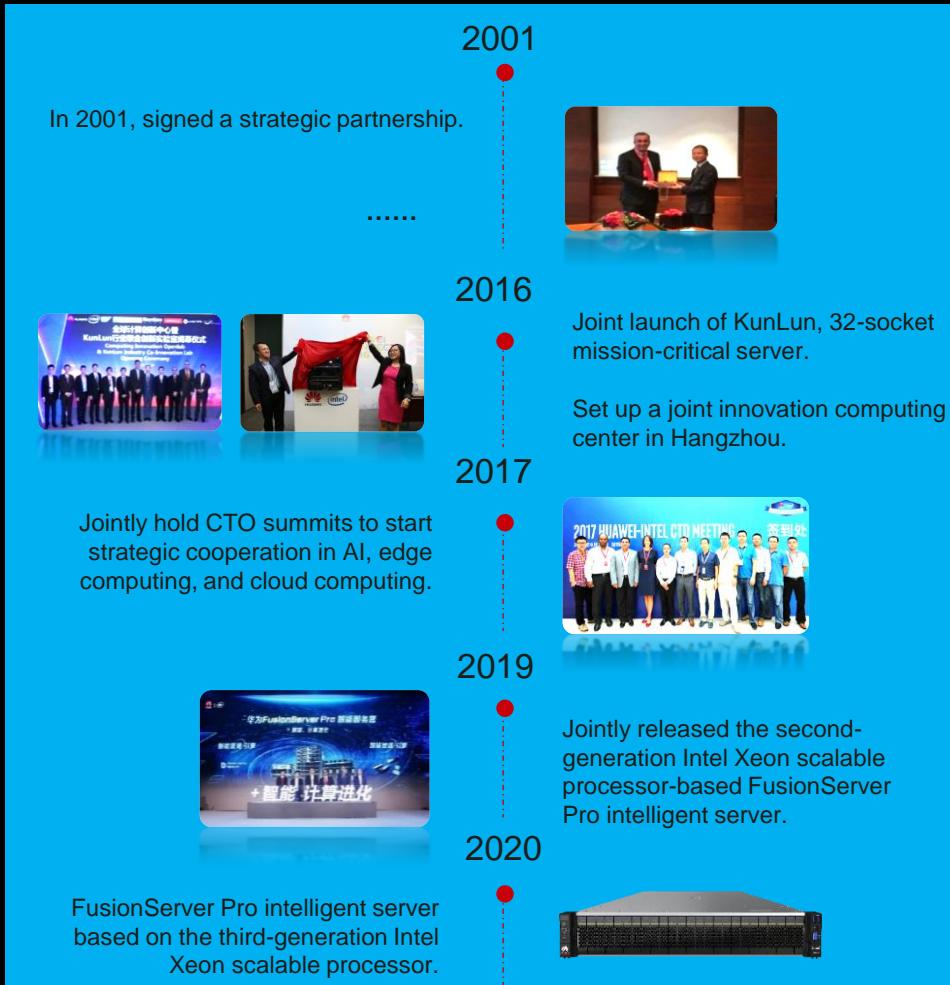
Futurewei maintains ongoing, in-depth collaboration with a broad range of research institutes and forward-thinking companies around the world in multiple domains. Our experts have been actively engaged in standard programs throughout these past two decades.

We are active in open source communities as part of our ongoing efforts to build out the industry and cultivate an ecosystem that promotes shared success.

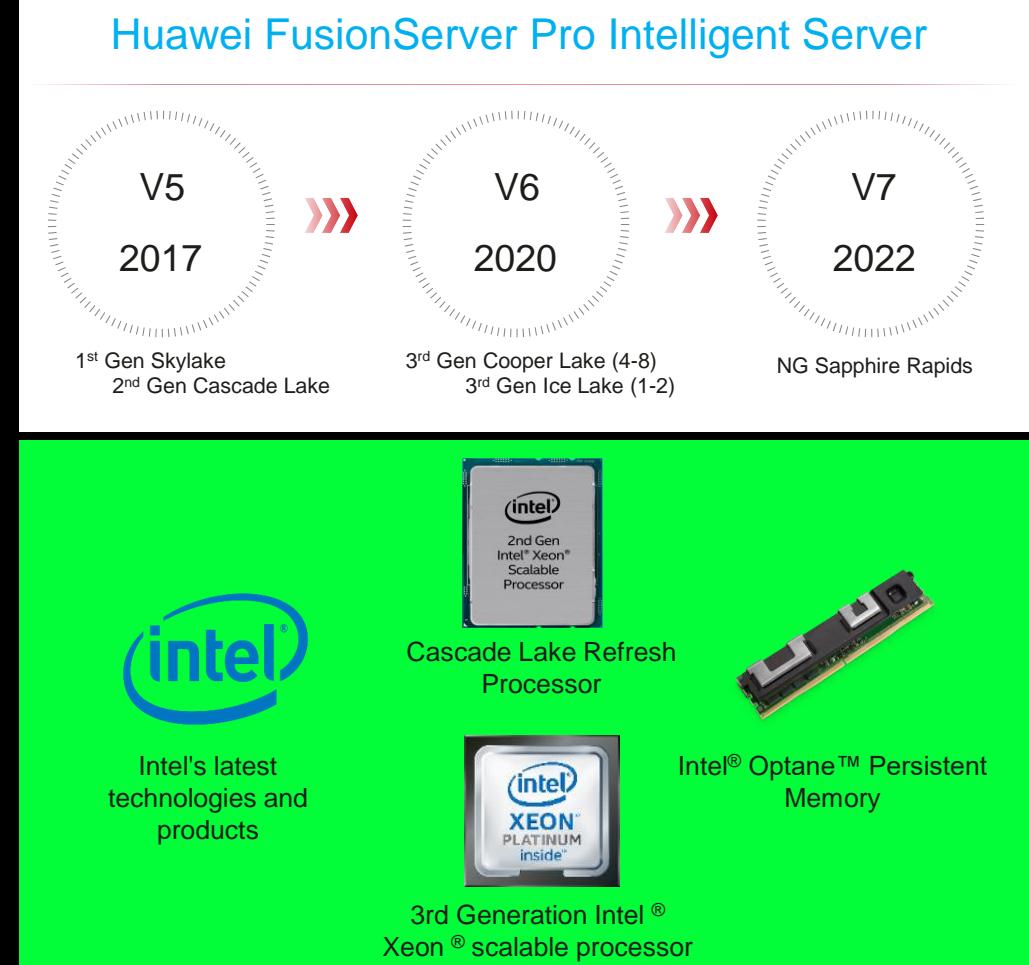
[www.futurewei.com](http://www.futurewei.com)

# 20 Years of Strategic Partnership with Intel Continuously Providing the Latest Technology and Products

Total 5M+ Units Shipped (2012-2020)



## Huawei FusionServer Pro Intelligent Server



# VMware Continues Compatibility Certification & Technical Support

- ✓ Partners such as **VMware**, **SUSE** and **Ubuntu** will continue to work with Huawei to provide compatibility certification and technical support for **FusionServer Pro** servers.

**On May 23, 2020, VMware announced that Huawei FusionServer Pro 2288H V5 and other servers have passed the latest VMware vSphere 7.0 compatibility certification.**

\*VMware vSphere, vSAN, and vCenter will continue to provide compatibility certification and technical support for Huawei FusionServer Pro servers.

The latest 2488H V6 (Copper Lake) has been certified

Huawei Technologies Co., Ltd.	2488 V5	Intel Xeon Platinum 8200 (Cascade-Lake-SP) Series	ESXi	+ 7.0 U1	7.0	6.7 U3	6.7 U2
Huawei Technologies Co., Ltd.	2488H V5	Intel Xeon Gold 6100/5100, Silver 4100, Bronze 3100 (Skylake-SP) Series	ESXi	+ 7.0 U1	7.0	6.7 U3	6.7 U2
Huawei Technologies Co., Ltd.	2488H V5	Intel Xeon Gold 6200/5200 (Cascade-Lake-SP/Refresh) Series	ESXi	+ 7.0 U1	7.0	6.7 U3	6.7 U2
Huawei Technologies Co., Ltd.	2488H V5	Intel Xeon Platinum 8100 (Skylake-SP) Series	ESXi	+ 7.0 U1	7.0	6.7 U3	6.7 U2
Huawei Technologies Co., Ltd.	2488H V5	Intel Xeon Platinum 8200 (Cascade-Lake-SP) Series	ESXi	+ 7.0 U1	7.0	6.7 U3	6.7 U2
Huawei Technologies Co., Ltd.	2488H V6	Intel Xeon Gold 6300/5300 (Cooper-Lake-SP) Series	ESXi	7.0 U1			
Huawei Technologies Co., Ltd.	2488H V6	Intel Xeon Platinum 8300 (Cooper-Lake-SP) Series	ESXi	7.0 U1			
Huawei Technologies Co., Ltd.	5288 V5	Intel Xeon Gold 6100/5100, Silver 4100, Bronze 3100 (Skylake-SP) Series	ESXi	+ 7.0 U1	7.0	6.7 U3	6.7 U2
Huawei Technologies Co., Ltd.	5288 V5	Intel Xeon Gold 6200/5200 (Cascade-Lake-SP/Refresh) Series	ESXi	+ 7.0 U1	7.0	6.7 U3	6.7 U2
Huawei Technologies Co., Ltd.	5288 V5	Intel Xeon Platinum 8100 (Skylake-SP) Series	ESXi	+ 7.0 U1	7.0	6.7 U3	6.7 U2
Huawei Technologies Co., Ltd.	5288 V5	Intel Xeon Platinum 8200 (Cascade-Lake-SP) Series	ESXi	+ 7.0 U1	7.0	6.7 U3	6.7 U2
Huawei Technologies Co., Ltd.	5288 V5	Intel Xeon Silver 4200, Bronze 3200 (Cascade-Lake-SP/Refresh) Series	ESXi	+ 7.0 U1	7.0	6.7 U3	6.7 U2
Huawei Technologies Co., Ltd.	5288X V5	Intel Xeon Gold 6200/5200 (Cascade-Lake-SP) Series	ESXi	+ 7.0 U1	7.0	6.7 U3	6.7 U2

[https://www.vmware.com/resources/compatibility/search.php?deviceCategory=server&details=1&partner=242&evcModel=19&page=1&display\\_interval=50&sortColumn=Partner&sortOrder=Asc](https://www.vmware.com/resources/compatibility/search.php?deviceCategory=server&details=1&partner=242&evcModel=19&page=1&display_interval=50&sortColumn=Partner&sortOrder=Asc)

# 8+ Years Co-Innovation for Build Industry-Leading HANA Solutions

- ✓ Our Latest SAP HANA solution is available for our latest 2488H V6 server with Optane Memory for big Applications



Huawei became SAP's first global technical partner in China  
July 2012



Huawei and SAP jointly launched Tecal RH5885 V2, the first SAP HANA appliance  
Feb. 2013



Strategic alliance partnership:  
Huawei released the certified  
SAP HANA FusionCube  
solution at CeBIT 2014  
Mar. 2014



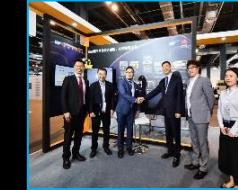
Huawei and SAP signed a  
MoU at CeBIT 2015 to  
deepen collaboration in  
Industry 4.0 and the IOT  
Mar. 2015



As a top-level sponsor,  
Huawei attended the  
SAPPHIRE NOW 2016 global  
flagship conference held by  
SAP in Orlando, USA.  
May 2016



Huawei & SAP Co-  
Innovation Center was  
inaugurated in Zone G of  
Huawei HQ, Shenzhen.  
Aug. 2018



SAP attended HUAWEI  
CONNECT 2018  
as a diamond sponsor  
and major partner.  
Oct. 2018



Huawei and SAP launched  
the FusionServer Pro V6  
SAP HANA solution at  
HUAWEI CONNECT 2020.  
Sep. 2020

2012

2013

HANA  
platform  
adoption

2014

2015

2016

SAP Hybris &  
Ariba contract  
signed

2017

MaxAttention for  
Hybris Billing  
LIVE @ CBG

2018

Hybris Billing  
Rollout @ CBG &  
Cloud

2019

2020

Ren Zhengfei, founder of  
Huawei, met with Jim  
Snabe, Co-CEO of SAP.  
Sep. 2012



Huawei Day was held at Walldorf,  
Germany, the SAP headquarter.  
July 2014



Huawei and SAP announced  
the official opening of the  
Huawei & SAP Co-Innovation  
Center in Shenzhen.  
June 2015



SAP conferred the dkom  
award on Huawei.  
Dec. 2015



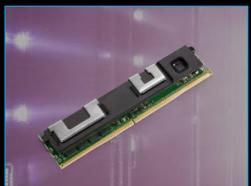
Huawei won the SAP  
HANA® Innovation Award  
at SAPPHIRE NOW held  
in Orlando.  
May 2017



Huawei attended SAP China  
Summit as a diamond  
sponsor to explore market  
opportunities and innovation.  
Sep. 2018



Huawei introduced the  
persistent memory HANA  
solution for larger HANA  
memory capacity.  
May 2019



SAP Max. Attention | SAP and Huawei Day | Huawei & SAP Co-Innovation Center

# Fruitful Results from Huawei-SAP Close Cooperation

400+ multiple joint innovations

Since 2012, SAP and Huawei have jointly developed more than 400 technological innovations



One of SAP HANA's largest customers in China

SAP helps Huawei build the largest SAP HANA-based big data platform in Asia to realize real-time insights in Huawei's finance, supply chain and other fields



One of SAP largest market partners

Huawei and SAP work together to jointly develop the SAP market



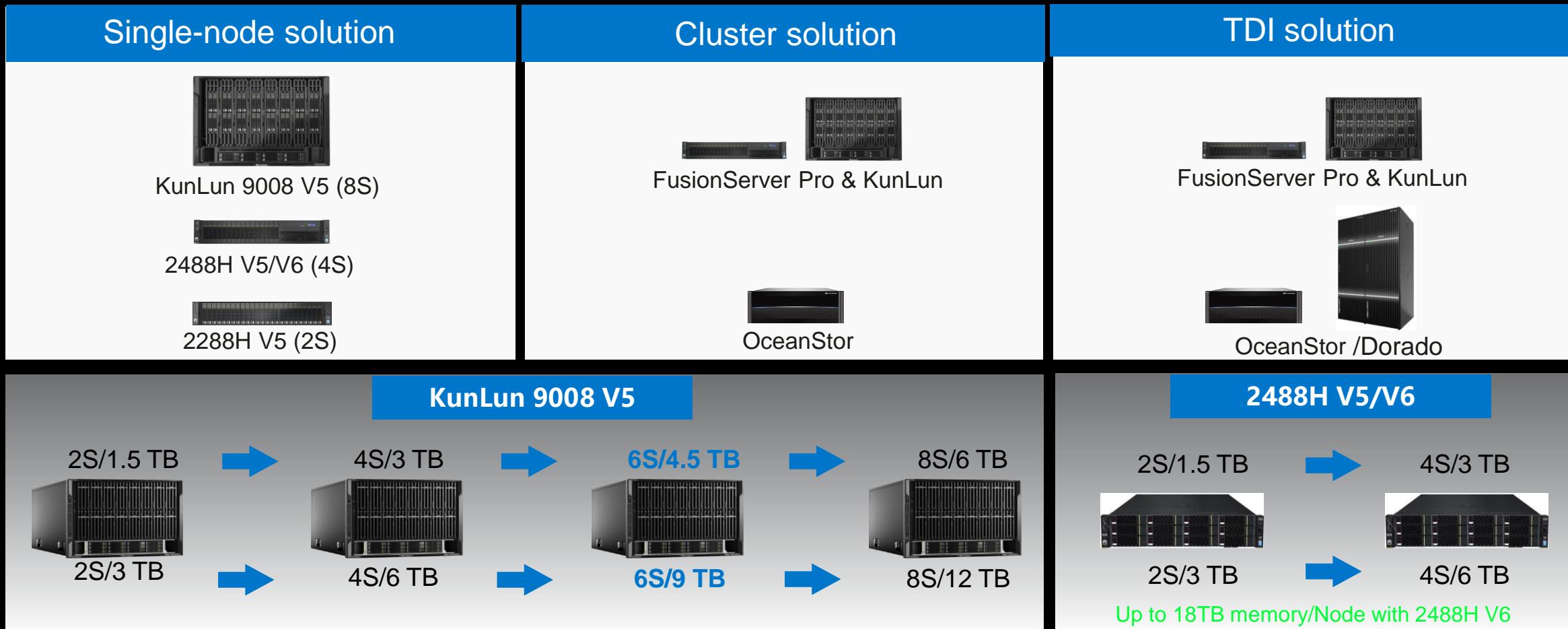
1,600+  
customers

60+  
countries

25+  
industries

# Support All three SAP HANA Models for Flexible Deployment

- ✓ All three SAP HANA Delivery models are supported for flexible deployment
- ✓ With the latest 2488H V6 with Optane Memory, up to 18TB memory per node could be supported



# FusionServer Pro V6 SAP HANA Solution is READY

## FusionServer Pro 2488H V6 SAP HANA solution (Single-node solution/Cluster solution/TDI solution)



**112 cores**  
Stronger computing power

**18 TB**  
Larger memory

**NVMe SSD**  
Higher performance

### Stronger computing power, improving SAPS performance by 10%\*

- ✓ 4 x 3rd Gen Intel® Xeon® Scalable processors with a TDP of up to 250 W/CPU and 112 cores, improving SAP BWH Benchmark performance by over 10%

### Larger memory, reducing CAPEX by 20%

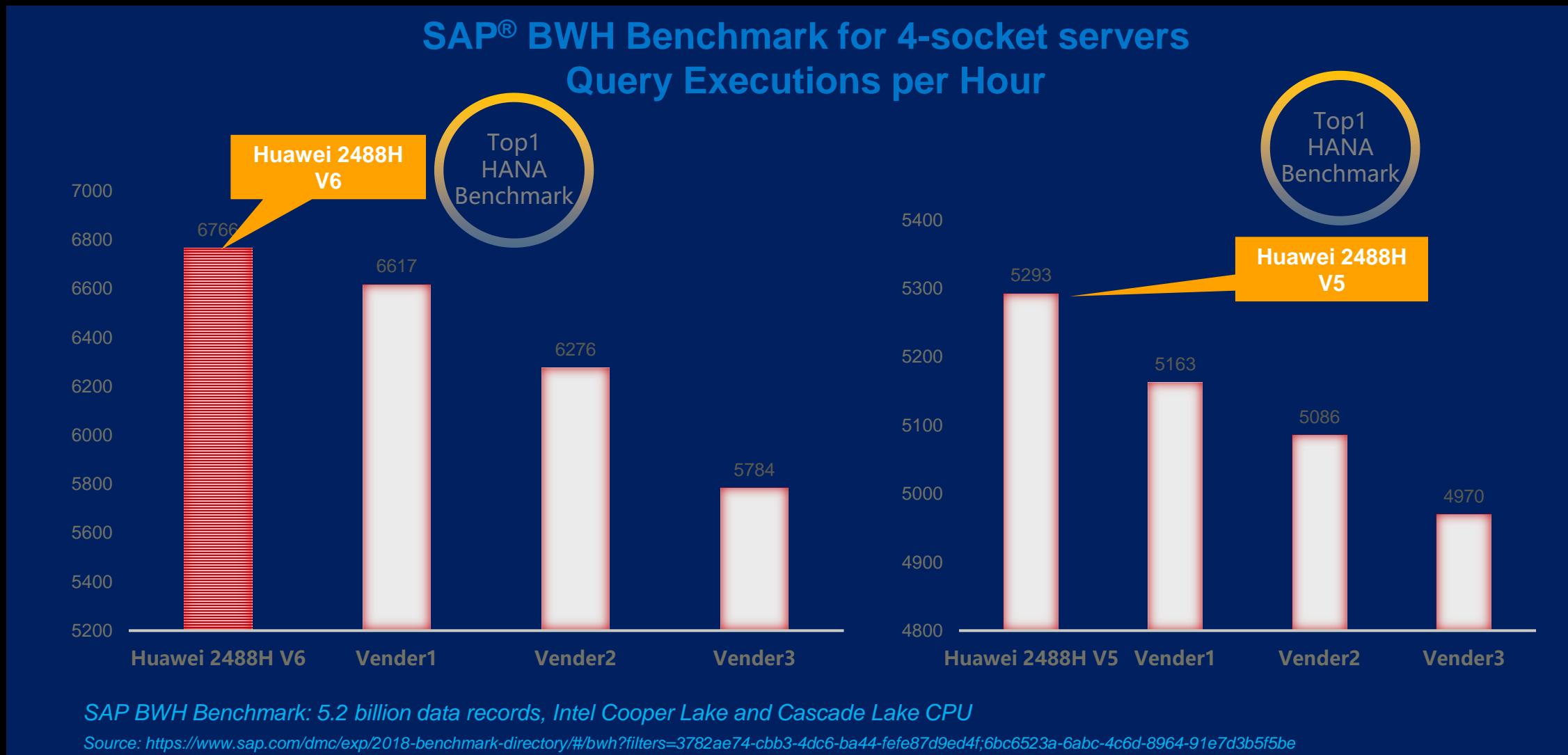
- ✓ 48 DIMM slots, supporting the Intel® Optane™ persistent memory (PMem) 200 series for up to 18 TB memory capacity

### NVMe SSD, boosting performance by 30%

- ✓ Supporting HANA appliances with NVMe SSDs and SAS SSDs, NVMe SSDs yield a 30% higher performance over SAS SSDs.

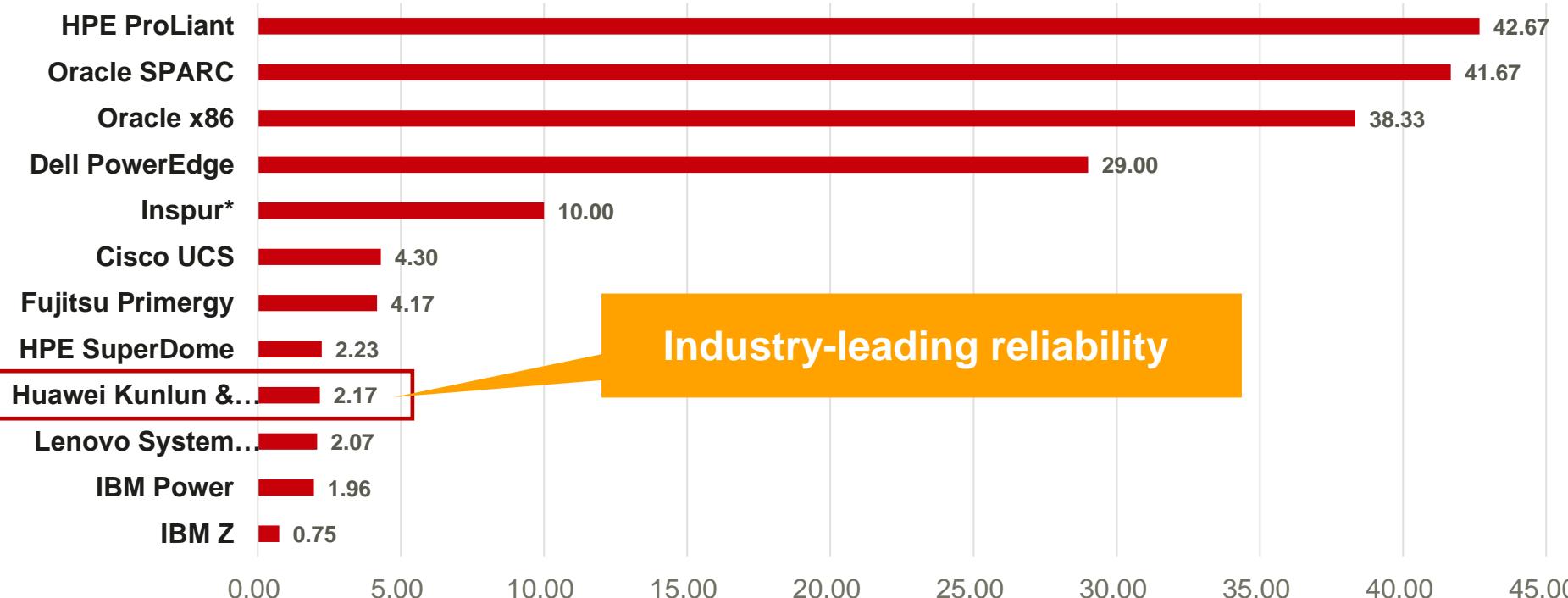
\*SAP Application Performance Standard (SAPS) is an authoritative performance indicator in the SAP industry. Generally, SAPS in HANA projects refers to the score in the SD2 benchmark test. In technical terms, 100 SAPS is equivalent to 2,400 SAP transactions per hour, or 6,000 dialog steps (screen changes) and 2,000 postings per hour.

# Huawei X86 Servers Leading SAP HANA Performance Benchmark



# Huawei KunLun X86 Servers Leading Reliability for SAP HANA

## Unplanned Downtime Due to Inherent system or component flaws per Minute/per Server (2018~2020)



Source: ITIC 2018~2020 Global Server Hardware Server OS Reliability Survey Results

<https://itic-corp.com/blog/2018/08/itic-2018-server-reliability-mid-year-update-ibm-z-ibm-power-lenovo-system-x-hpe-integrity-superdome-huawei-kunlun-deliver-highest-upptime/>

<https://itic-corp.com/blog/2019/11/ibm-lenovo-hpe-and-huawei-servers-maintain-top-reliability-rankings-cisco-makes-big-gains-ibm-lenovo-hardware-up-to-24x-more-reliable-28x-more-economical-vs-least-reliable-white-box-servers/>

<https://itic-corp.com/blog/2020/05/itic-2020-reliability-poll-ibm-lenovo-hpe-huawei-mission-critical-servers-deliver-highest-uptime-availability/>

\* Inspur has the data of 2020 only

# Huawei Server Winning Extensive Customer Recognition

211 of the Fortune Global 500 companies and 48 of the Top 100 companies have chosen Huawei as their digital transformation partner. Our service products and solutions are widely used in the energy, transportation, finance, and manufacturing industries.

**170+**  
Power  
companies



**211**  
G500

**48**  
G100

**300+**  
Financial  
institutes

**1,000+**  
Manufacturing  
companies



**5+ Million Units**  
shipped  
(2012 - 2020)



**260+**  
Oil and Gas

# Huawei Server Honor List



**400+**  
New SPEC test records



**15%**  
lower failure rate than  
industry average\*



**Trusted & Recognized**  
Best Supplier\*\*  
Gartner MQ Challenger



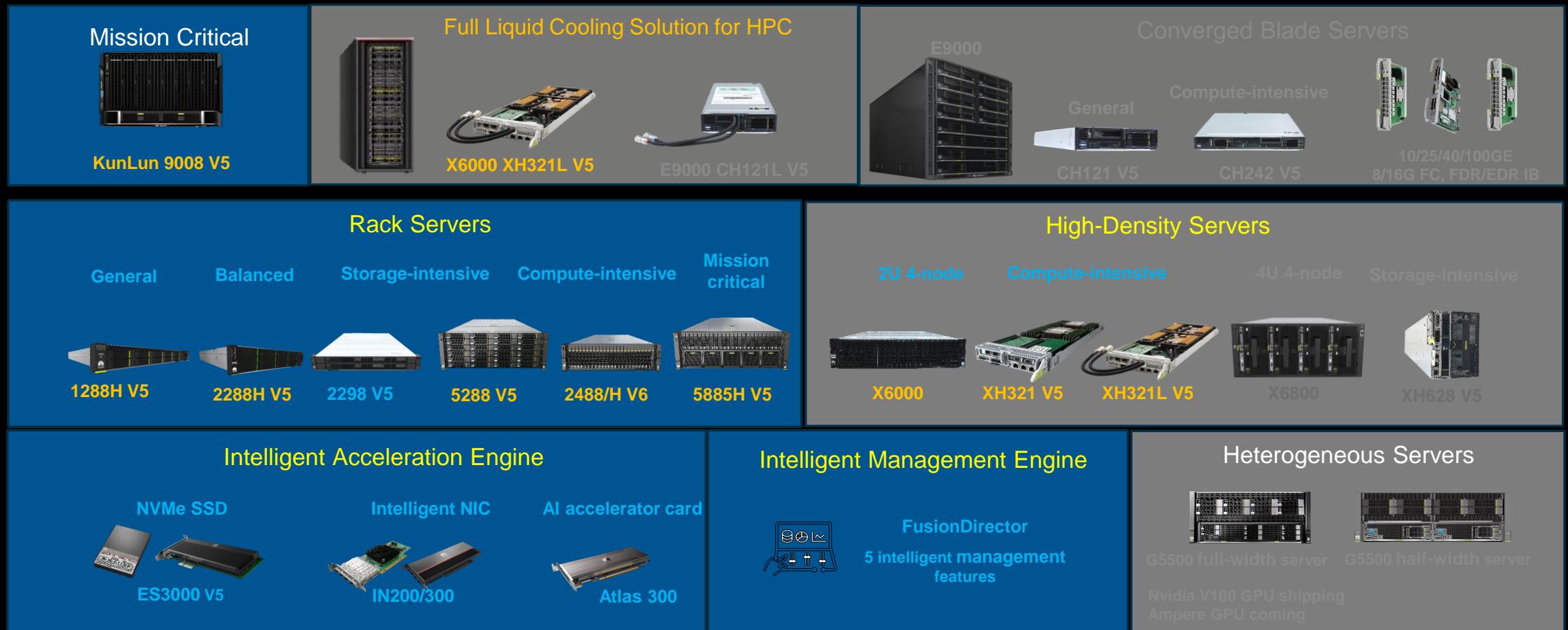
**10000+ Customers**  
across 25+ industries

\* Source: internal statistics of key accounts in the Internet and finance industries

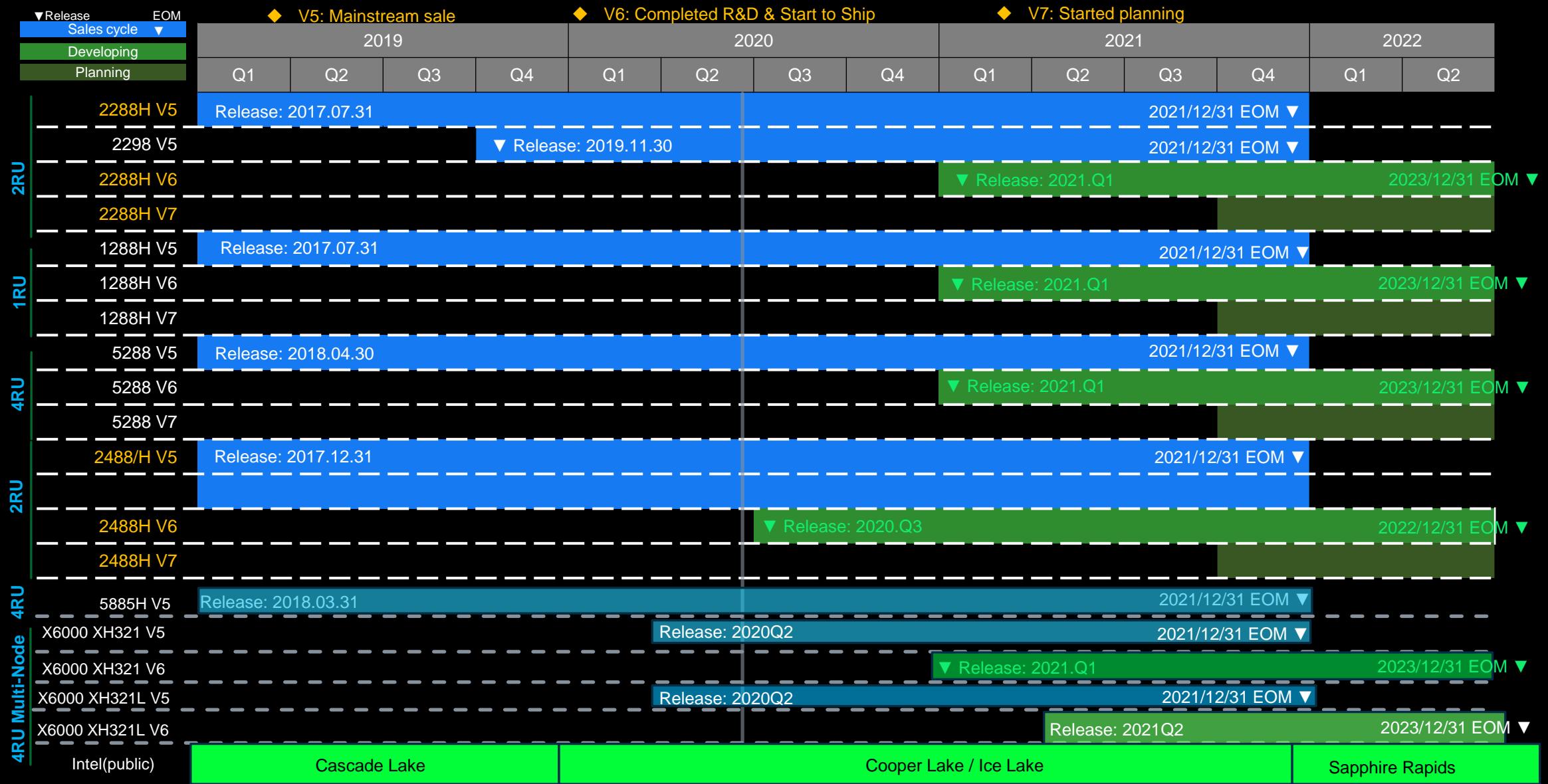
\*\* Awarded by Tencent and Alibaba

# Rich Portfolio to Meet Your Every Intelligent Computing Need

- ✓ Rich Product Portfolio to meet your needs for different application scenarios:
  - Balanced, Computing or Storage Intensive, Air or Liquid Cooling, Rack or Blade, etc.
- ✓ Full series upgraded to the latest Intel Cascade Lake Refresh / Cooper processor, synchronous to Intel CPU Release Schedule



# Continuous Evolution Roadmap Synchronous to Intel Release Schedule



# What Customers Considered When Choosing Huawei Computing

## ✓ Quality Products

- DNA of Telecomm Products, E2E Quality Control

## ✓ Innovative Product Features

- Enabled by Self-Developed ASICs for Management, Perf, Deployment & Diagnosis

## ✓ Knowledgeable & Experienced Team

- Both Hardware & Software, Applications & Operations

## ✓ Long-Term Commitment & Support

- Both spare Parts & Software

## ✓ Local Support

- Quick response & Easy to reach

## ✓ Competitive Pricing for TCO

- Not Only CAPEX, but also OPEX

## ✓ Partnership & Ecosystem

- Strategic Partnership with key players, X86 ecosystem

## ✓ Rich Product Portfolio

- No One Fits All, Rich Models to choose

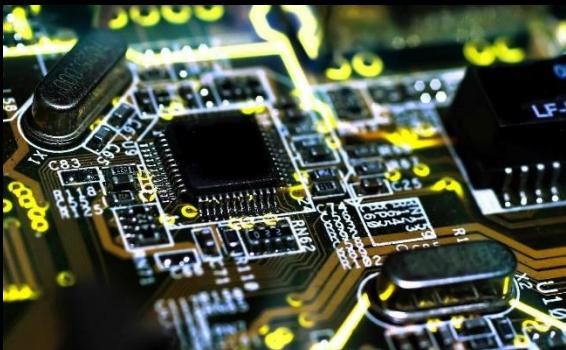
## ✓ Consolidation of Suppliers

- The same supplier for CT & IT, fewer suppliers to Manage

# Quality Product by Design, Process, Sourcing and Testing

Over 30 years of hardware design, development, and manufacturing capabilities coupled with complete product R&D and test processes to maximize server reliability, reduce downtime and data loss, and enhance device maintainability

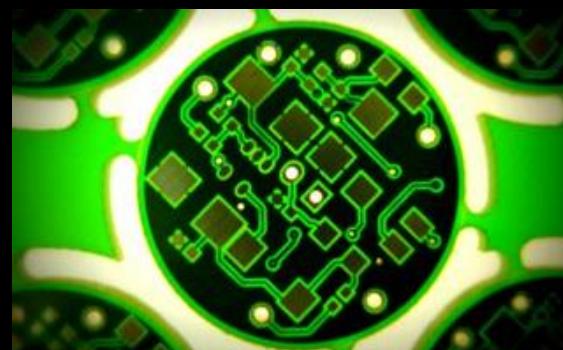
## Component Selection and Process Technique



## High Reliability Design



## Strict Test and Certification



## Enhanced O&M Features



Hard Disk Failure rate

**40% ↓**

Source: Per Internal Testing & O&M Data Collections

Working Temperature

**45°C**

Source: Internal Lab Tested for most products

Machine Failure rate

**15% ↓**

Source: Per a Large Customer Statistics

# Efficient Heat Dissipation, Stable Operation at 45°C

The strict electronic component selection, flow-dividing air ducts and independent air channels, and end-to-end heat dissipation design improve heat dissipation efficiency and prolong the service life of electronic components, allowing servers to run stably even when the air conditioners in equipment room are faulty.

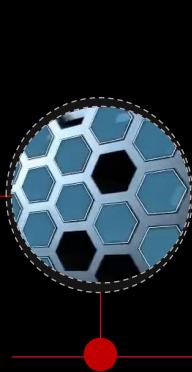
**≥30%**

Reserved margin  
for component  
derating design



**75%**

Cellular board design increases  
porosity rate of the front panel



**30%**

Counter-rotating fans for  
higher wind speeds



**3%**

3D VC heat dissipation with higher efficiency  
than common VC heat dissipation

Flow-dividing air duct and independent  
air channels for BIO modules improve  
heat dissipation efficiency

**9%**

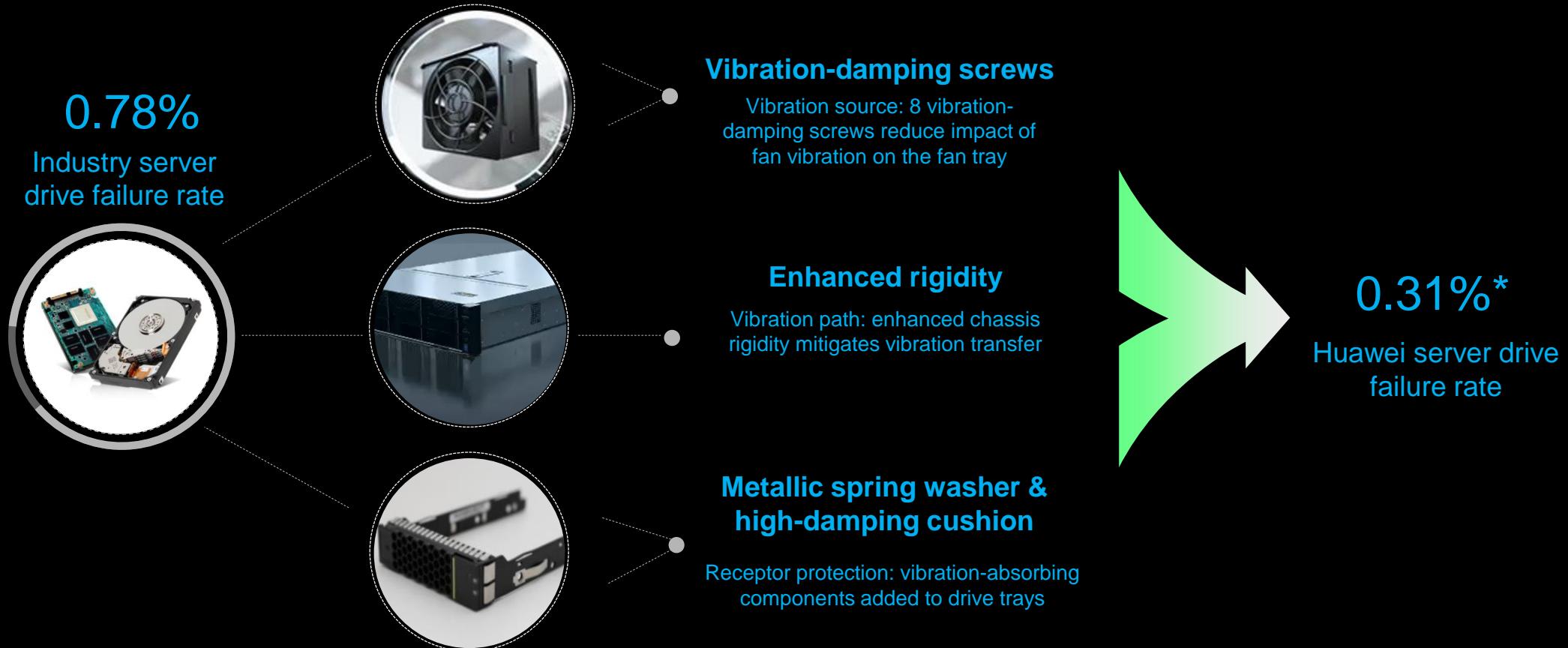


**45°C**

**Stable running at  
high temperature**  
For most models

# Triple Anti-Vibration Designs Reduce HDD Failure Rate by 60%

Triple anti-vibration designs for hard drives: Mechanical hard drives have the highest failure rate among server components. The triple anti-vibration designs for hard drives improve stability and reduce the hard drive failure rate.



\* Based on the fault statistics of 26,000 hard drives from the data center of a major financial customer.

# FusionDirector: 5 Intelligent Management Features Reduce OPEX

## Intelligent Asset Management

Server model, configuration, inventorying, and retirement

Inventorying in seconds, improving space utilization by 10%+



## Intelligent Deployment Management

Automatic deployment in data center cabinets  
10x device rollout efficiency



## Intelligent Version Management

Automatic version matching, one-click update  
Preparation done in seconds, reducing update steps from 20 to 3



## Intelligent Fault Management

Fault prewarning, diagnosis, and locating  
7–30 days in advance for fault prediction,  
93% fault locating accuracy



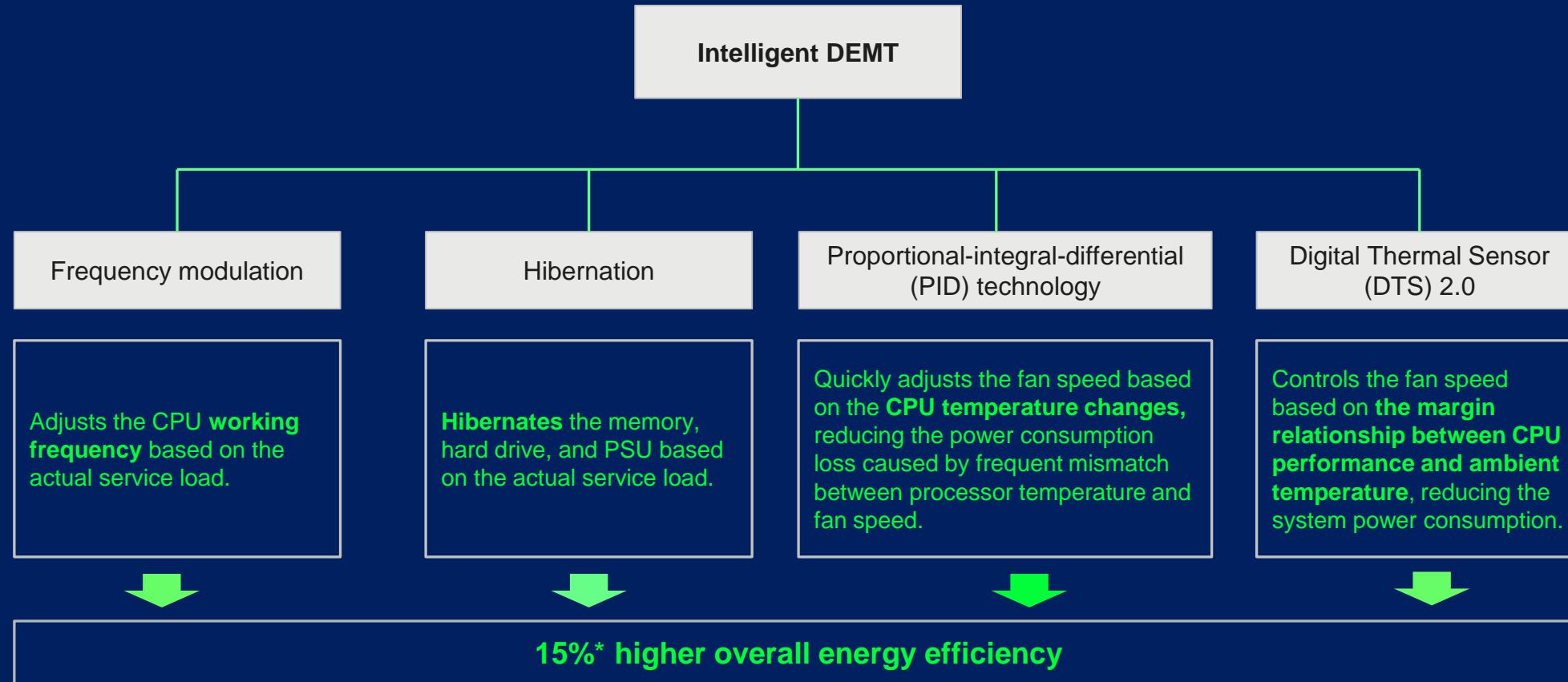
## Intelligent Energy Efficiency Management

Server DEMT, cabinet power capping, and data center linked control  
10% server energy saving, 20% higher deployment density



# Huawei Server DEMT for 15% higher overall energy Efficiency

The Dynamic Energy Management Technology (DEMT) is a set of technologies that intelligently adjust the power consumption of each component in real time, according to multiple input parameters such as load and ambient temperature.



\*Data source: Huawei Server Performance Tuning Lab

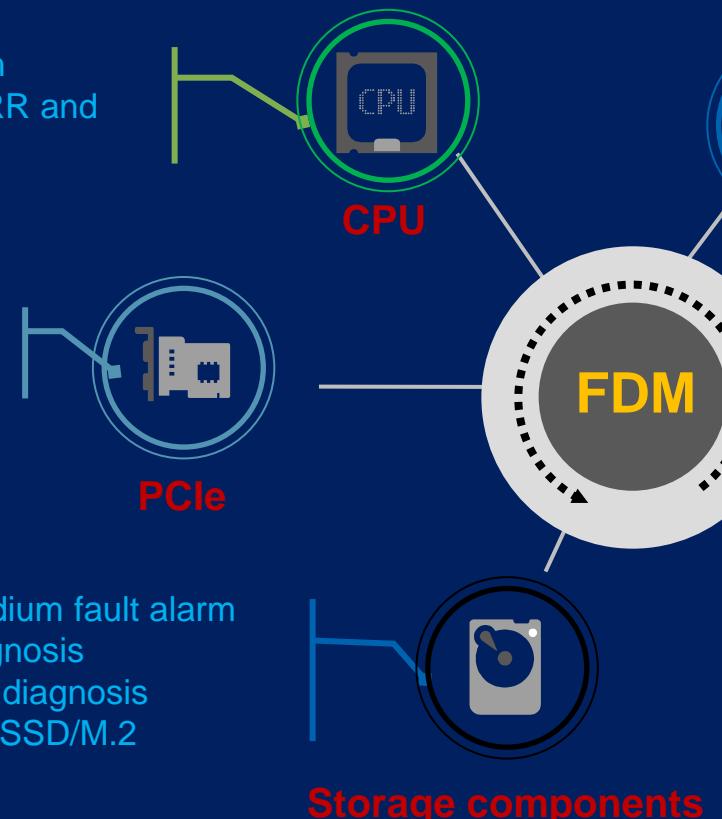
# Huawei FDM Delivers up to 93% Fault Locating Accuracy\*

Fault Diagnosis & Management (FDM) is a set of processes and mechanisms for automatic fault information collection (in-band and out-of-band), analysis, diagnosis, and alarm/pre-warning. This greatly improves the fault handling efficiency and shortens the impact of faults on services.

- Out-of-band collection of CATERR fault information
- Precise locating of MCERR and IERR
- Diagnosis logs

- Accurate diagnosis of PCIe card faults

- Storage medium fault alarm and pre-diagnosis
- Storage link diagnosis
- HDD/NVMe/SSD/M.2



- Memory address error parsing
- Correctable error storm suppression
- Fault locating of the memory link/slot

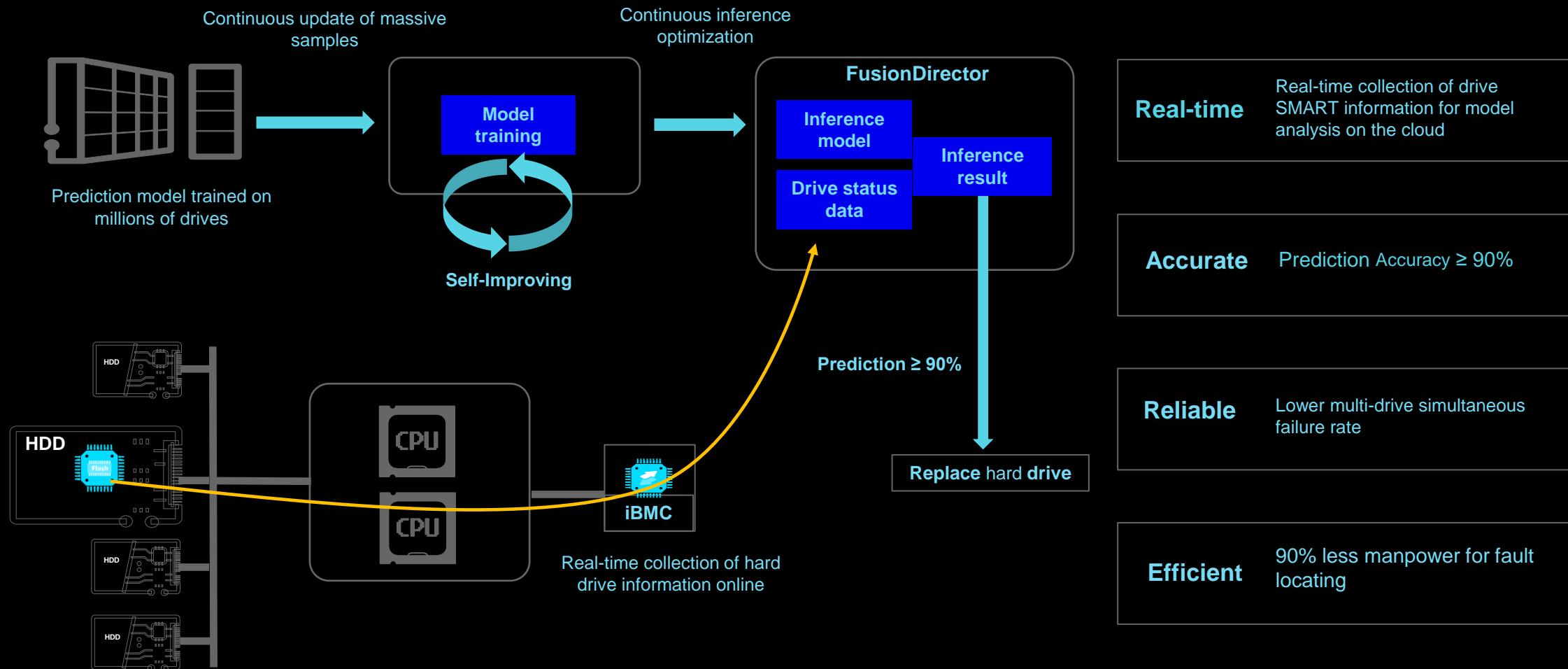
- Fault diagnosis of multiple links, such as network switching links and SAS links
- OS crash information collection

Comparison with Competitors

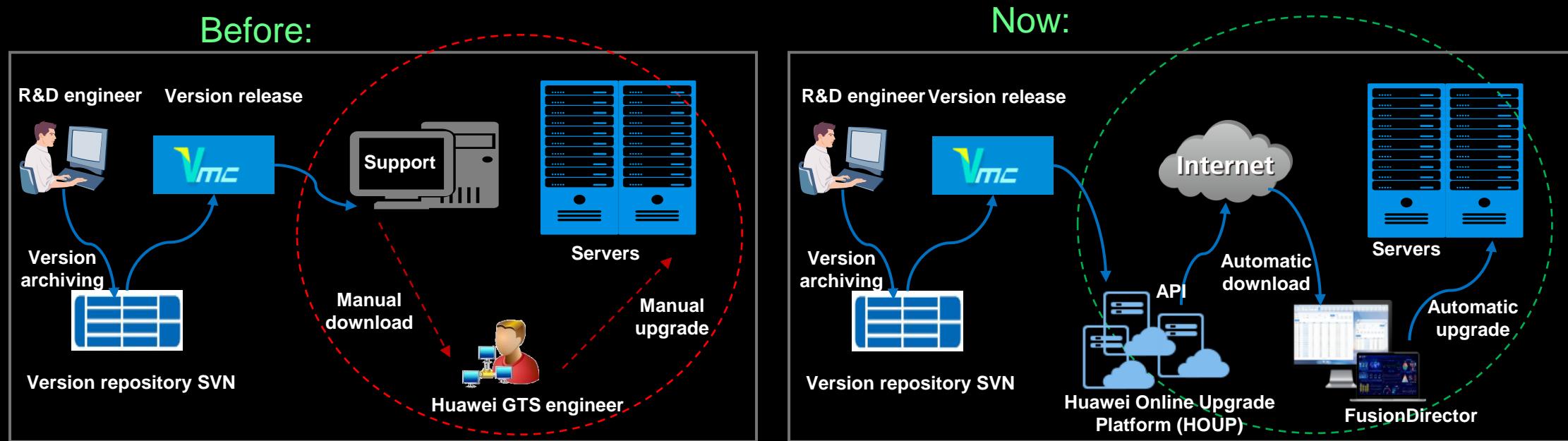
Comparison Item	Competitor	Huawei FDM
CATERR error collection	Live-network faults cannot be collected onsite.	Fault information is automatically collected comprehensively at a time on site. The information collection does not rely on fault reproduction.
CATERR fault locating	Manual analysis, low success rate	Automatic fault locating: The automatic locating success rate is 80%, the total locating success rate is 93%, and the diagnosis scope is wide.
Other troubleshooting capabilities	Rely on the common fault handling capabilities of the OS no prewarning capabilities.	Automatic collection, expert system fault alarm, memory correctable error storm parsing and prewarning

\* Source of data: CATERR fault statistics of the Huawei server O&M team

# Intelligent Fault Prediction: Predicts Risky Hard Drives 7–30 Days in Advance

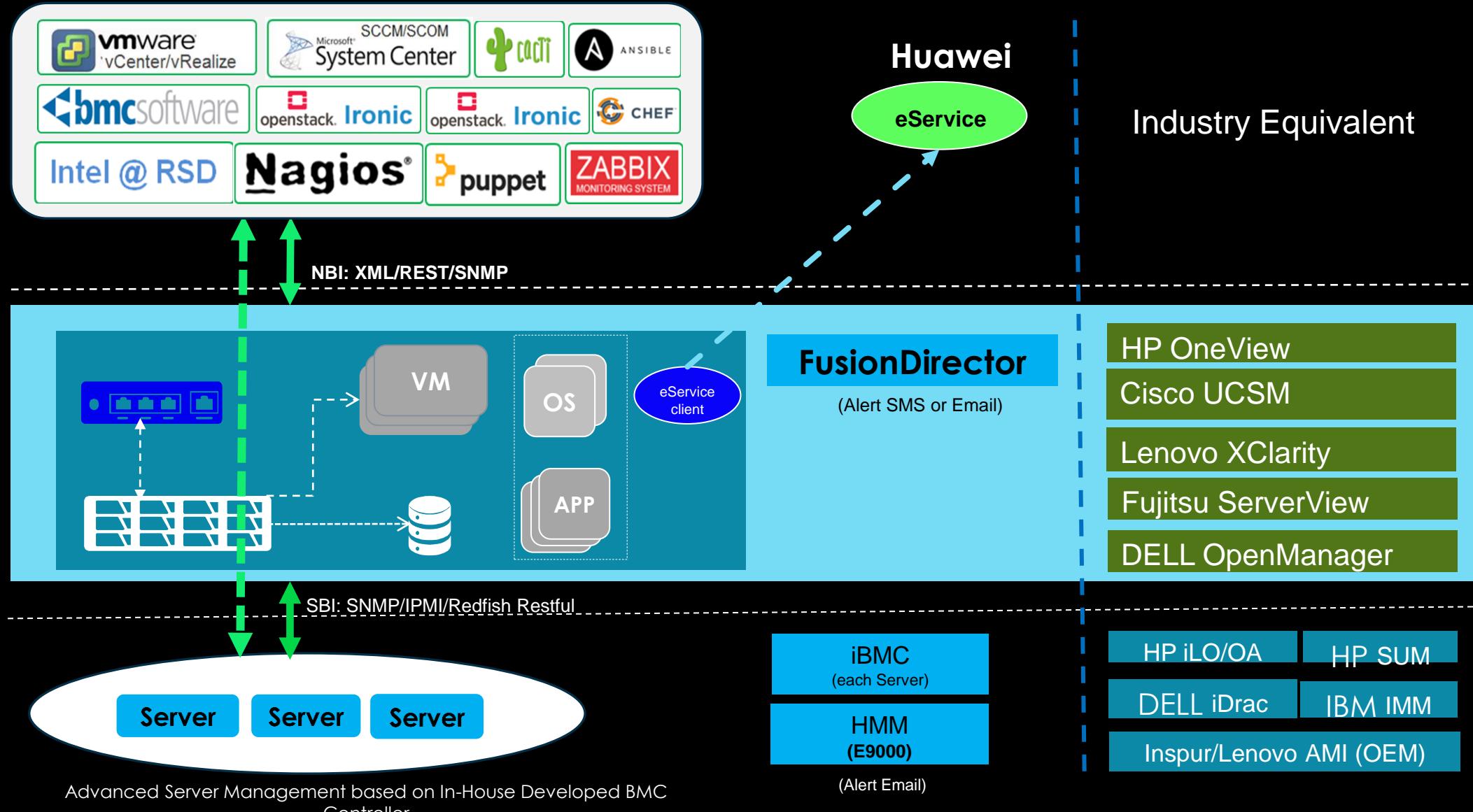


# Intelligent Version Management: Automatic Upgrade Of Device Software



- Automatic download: Automatically detects the firmware and drivers of servers of different models and manages firmware versions comprehensively
- O&M engineers do not need to manually download, upload, or upgrade firmware
- Independent out-of-band upgrade, without occupying service bandwidth

# Open & Standard-Based Server Management



Advanced Server Management based on In-House Developed BMC Controller

# Accelerating X86 Server Performance

ES3000 series intelligent SSDs



- Optimal performance: PCIe Gen4 1.66M Random read IOPS, 7.0GB/s sequential read bandwidth
- All-scenarios: low latency, high performance, large capacity
- Service-aware: intelligent multi-stream, SR-IOV, atomic write

Performance  
100% up

Service life  
20% longer

IN200/300 series intelligent NICs



- Converged: 10GE/25GE/40GE/100GE, 8G/16G/32G FC
- Accelerated: RoCE
- Reliable: industry unique to support firmware upgrade without service interruption

CPU offload  
15%

Latency  
30% less

Atlas 300 AI accelerator card



- Powerful computing: 64 TOPS of INT8
- High-density video parsing: 64-channel HD video real-time analytics

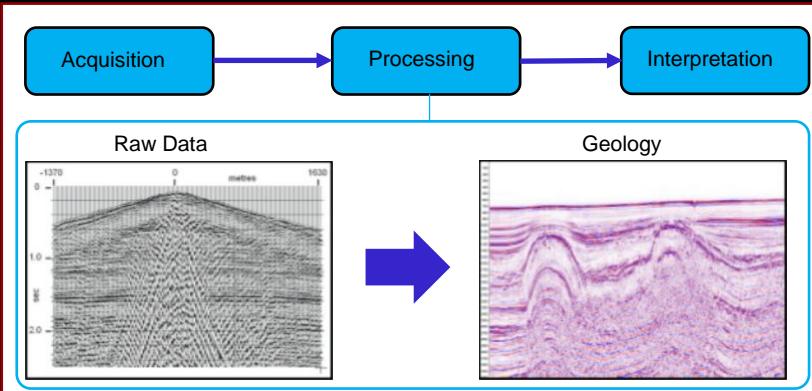
CPU offload  
15%

Performance  
64-Ch HD

# Understanding of Oil & Gas – Exploration, Production & Prediction

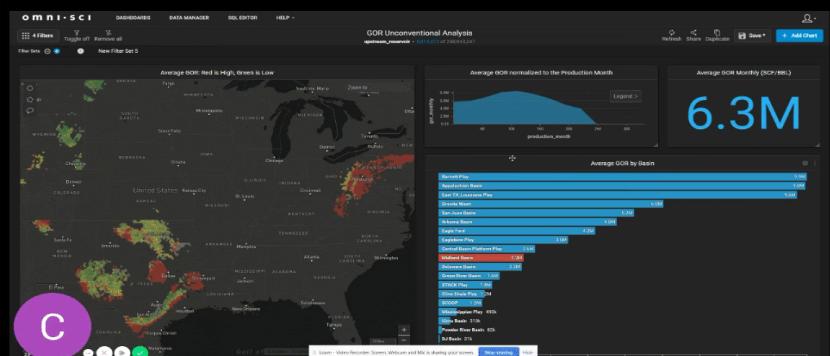
## Exploration - Seismic Data Processing

- Where and How Much is the Oil & Gas?
  - ✓ Build accurate HD earth Subsurface Models
  - ✓ Interpret the Models Automatically



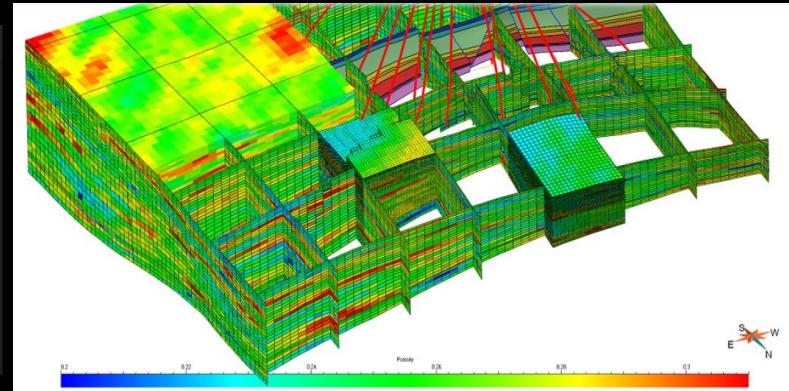
## Production - Drill Operations

- How are the wells going in Real-Time?
  - ✓ Well Operation & equipment status Monitoring
  - ✓ Predictive Maintenance to avoid operation disruptions



## Prediction - Reservoir Simulation

- How Much Oil & Gas Left & How would the reservoir be changing?



## HPC & AI, Edge Computing

- 32-bit FP intensive operations
- FWI algorithm becoming popular but requires 10x compute w/ every frequency doubling
- DNN may for automated image Analysis

## HPDA, Edge Computing

- Real-time Big-Data monitoring & Analysis
- In-Memory Computing w/ SAP HANA or HBM-based GPU will help

## HPC & AI

- Aramco TeraPOWERS is Powerful (X86 CPU based)
- 64-bit FP intensive operations?
- DNN may for automated image analysis & Open-source DEEPSEISMIC is active

# Huawei FusionServer 1288H V5: High-Density Deployment



1RU 2S+24-DIMM Compact w/ medium Storage & I/O Options

- Virtualization
- Cloud computing
- High-performance computing (HPC)
- Web application
- Software-defined storage (SDS)
- AI inference

## High reliability

- Carrier-class component specifications and 100% derating design
- Redundancy design: PSU modules in 1+1 mode, fan modules in N+1 mode, and memory module sparing/mirroring
- 99.99% availability
- 15% fewer failures than the industry average

## Ultimate performance

- 2 x 2nd Gen Intel® Xeon® Scalable processors, up to 28 cores and TDP 205 W per CPU
- 24 DDR4 DIMMs, up to 2933 MT/s; a maximum of 12 Optane™ Persistent Memory (PMem), up to 2666 MT/s; total memory capacity up to 7.5 TB on a single server
- No. 1 in SPECint®\_base2006 and SPECfp®\_base2006 performance tests
- No. 1 in SPECpower ssj2008 Benchmark energy efficiency test

## Unique value

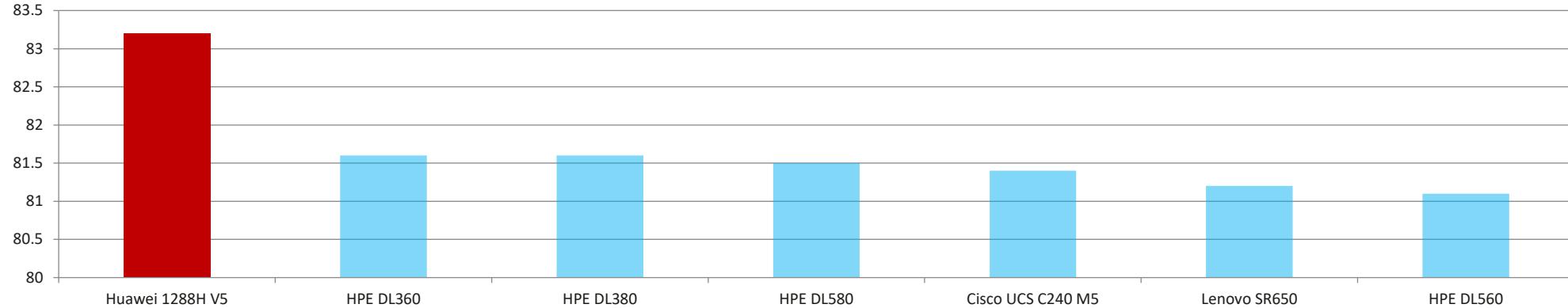
- **FDM:** The patented Fault Diagnosis & Management (FDM) technology provides the fault locating accuracy up to 93% and predicts a fault 7 to 30 days before it occurs, reducing OPEX
- **DEMT:** The patented Dynamic Energy Management Technology (DEMT) reduces energy consumption by 15% without compromising services
- **BSST:** Boot Speedup Storage Technology (BSST), 2 M.2 SSDs as high-speed OS boot drives, industry-unique hot plug and hardware RAID

## Innovative design

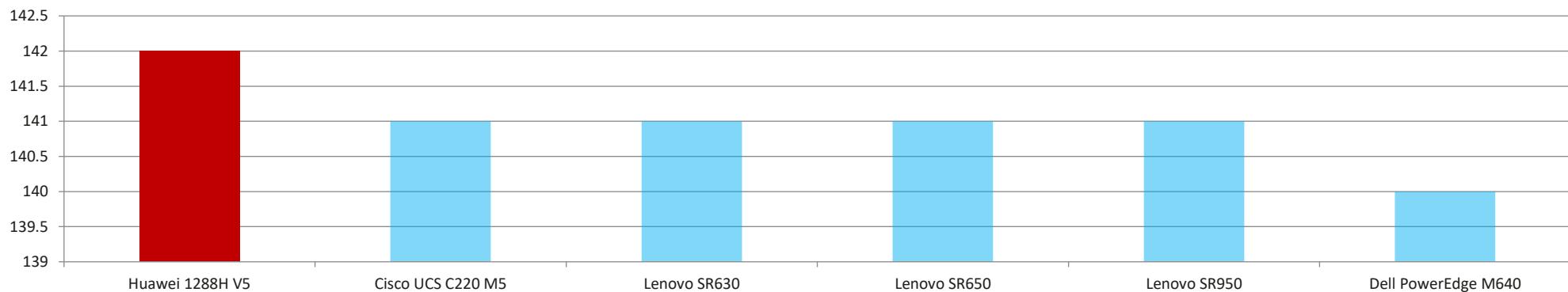
- 2 x 10GE and 2 x GE LOM ports, reducing CAPEX
- Flexible NICs for higher networking capabilities
- 10 x 2.5" SSAS/SATA HDDs or SSDs, providing ultra-large storage capacity
- 8 NVMe SSDs for optimal service experience
- 2 NVIDIA Tesla T4 GPUs for AI inference

# Huawei 1288H V5: Multiple New SPEC Performance Records

**SPECint\_base2006 Test Result**

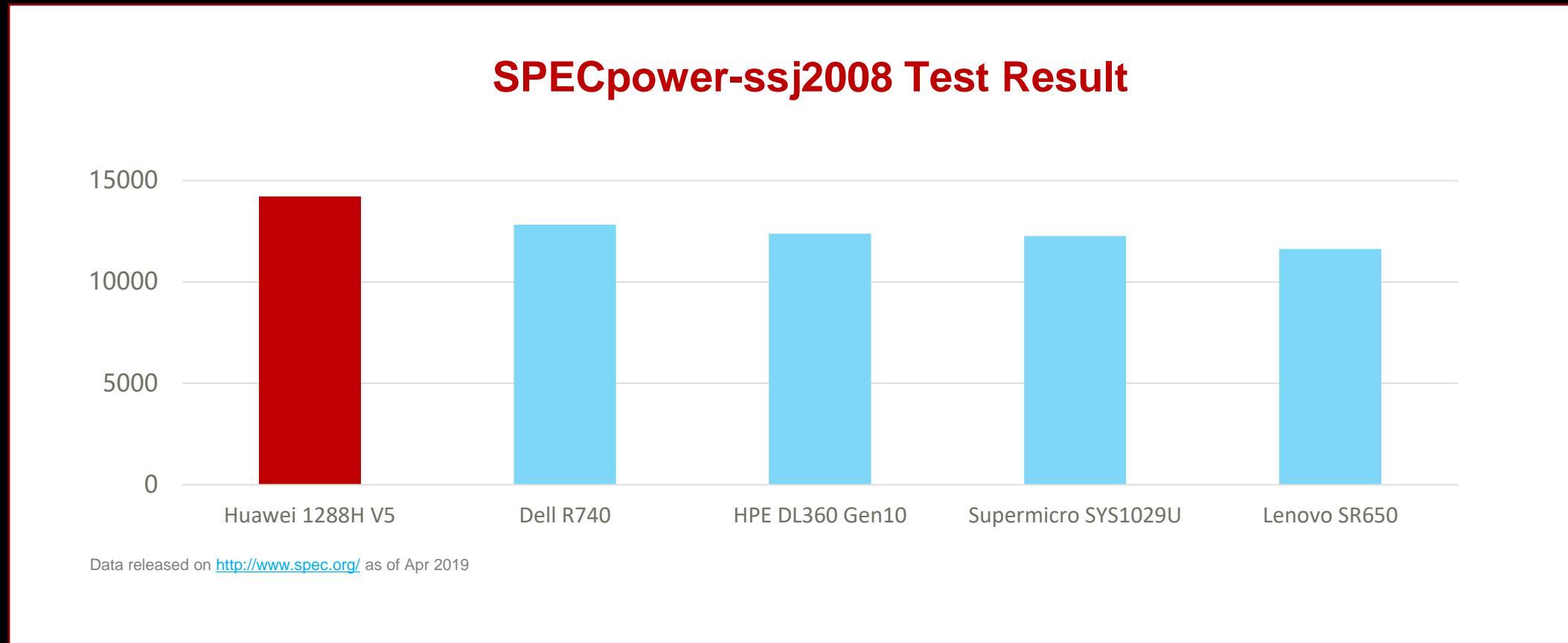


**SPECfp®\_base2006 Test Result**

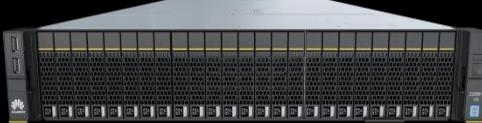


Data released on <http://www.spec.org/> as of Apr 2018

# Huawei 1288H V5: New SPEC Energy Efficiency Records



# 2288H V5:Flexible Configurations Adaptive to Various Workloads



- ✓ 2RU 2S+24-DIMM Ultimate Performance w/ very flexible Storage & I/O Options
- ✓ Many Storage Configurations Available
- ✓ Representing almost 50% of total units shipped

- Virtualization and cloud computing
- Virtual desktop interface (VDI)
- SDS
- HPC
- Big data and hyper-convergence
- Entry-level in-memory database (SAP HANA)
- AI inference or training

## High reliability

- Carrier-class component specifications and 100% derating design
- Redundancy design: PSU modules in 1+1 mode, fan modules in N+1 mode, and memory module sparing/mirroring
- 99.99% availability
- 15% fewer failures than the industry average

## Ultimate performance

- 2 x 2nd Gen Intel® Xeon® Scalable processors, up to 28 cores and TDP 205 W per CPU
- 24 DDR4 DIMMs, up to 2933 MT/s; a maximum of 12 Optane™ Persistent Memory (PMem), up to 2666 MT/s; total memory capacity up to 7.5 TB on a single server
- No. 1 in SPECint®\_base2006 performance test
- No. 1 in SPECpower ssj2008 Benchmark energy efficiency test

## Unique value

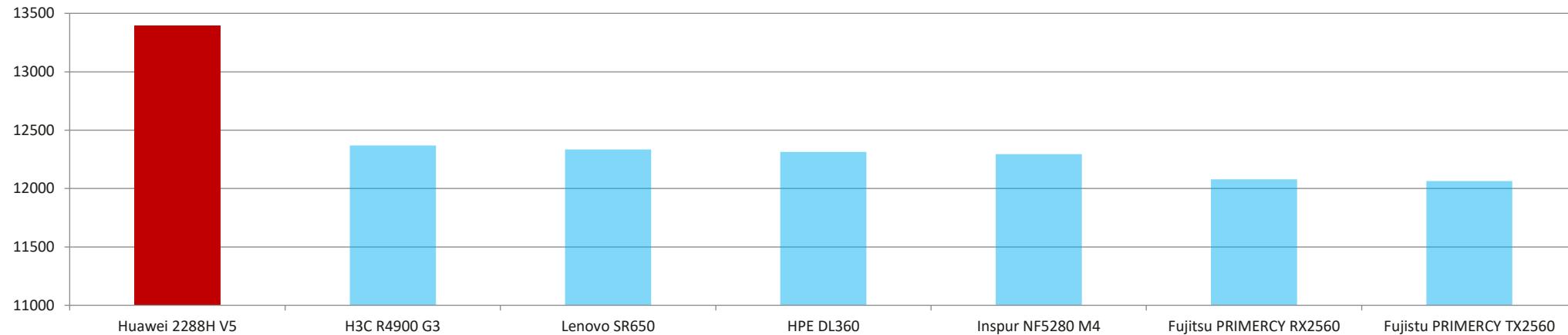
- **FDM:** The patented Fault Diagnosis & Management (FDM) technology provides the fault locating accuracy up to 93% and predicts a fault 7 to 30 days before it occurs, reducing OPEX
- **DEMT:** The patented Dynamic Energy Management Technology (DEMT) reduces energy consumption by 15% without compromising services
- **BSST:** Boot Speedup Storage Technology (BSST), 2 M.2 SSDs as high-speed OS boot drives, industry-unique hot plug and hardware RAID

## Innovative design

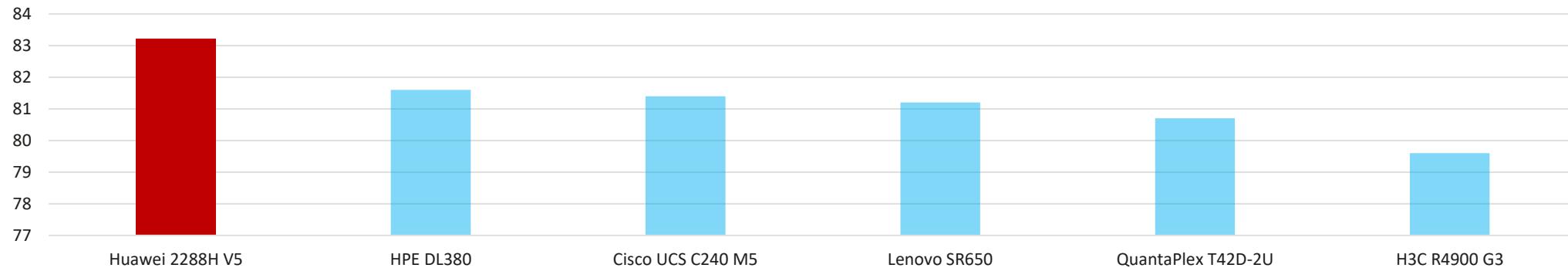
- 2 x 10GE and 2 x GE LOM ports, reducing CAPEX
- Flexible NICs for higher networking capabilities
- 20 x 3.5" or 31 x 2.5" drives for industry-leading storage capability
- 28 NVMe SSDs for optimal service experience
- 7 Atlas 300 inference cards/NVIDIA Tesla T4 GPUs for AI inference
- 2 NVIDIA Tesla V100 cards for AI training

# Huawei 2288H V5: New SPEC Energy Efficiency & Perf Records

SPECpower-ssj2008 Test Result



SPECint®\_base2006 Test Result



Data released on <http://www.spec.org/> as of Apr 2018

# 5288 V5: Ultra-Large Storage Capacity



- Tiered storage of hot, warm, and cold data
- Historical data archiving
- Video surveillance, CDN, and big data
- E-mail OA application

- ✓ 4RU 2S+24-DIMM High Performance w/ Ultra-Large Storage Capacity
- ✓ Dual RAID controller cards for High IOPS

## High reliability

- Carrier-class component specifications and 100% derating design
- Redundancy design: PSU modules in 1+1 mode, fan modules in N+1 mode, and memory module sparing/mirroring
- 99.99% availability
- 15% fewer failures than the industry average

## Ultra-large capacity design

- Hybrid storage architecture for tiered data storage
- 44 x 3.5" SAS/SATA drives + 4 x 2.5" SAS/SATA/NVMe SSDs in a 4U space, providing ultra-large local storage capacity
- Industry-leading storage capacity, up to 765 TB on a single server

## Unique value

- **FDM:** The patented Fault Diagnosis & Management (FDM) technology provides the fault locating accuracy up to 93% and predicts a fault 7 to 30 days before it occurs, reducing OPEX
- **DEMT:** The patented Dynamic Energy Management Technology (DEMT) reduces energy consumption by 15% without compromising services
- **BSST:** Boot Speedup Storage Technology (BSST), 2 M.2 SSDs as high-speed OS boot drives, industry-unique hot plug and hardware RAID

## Innovative design

- 2 x 10GE LOM ports, reducing CAPEX
- Flexible NICs for higher networking capabilities
- 6 Atlas 300 inference cards/NVIDIA Tesla T4 GPUs for AI inference, an industry-leading feature
- Dual RAID controller cards, doubling IOPS

# Huawei 2488 V5H: High Performance Computing



- ✓ 4S+48-DIMM in 2RU for Ultimate Computing Performance w/ Large Memory
- ✓ 25x2.5" HDDs or NVMe SSDs for high performance storage
- ✓ Up to 11 PCIe slots for expansion

- Virtualization
- Cloud computing
- HPC
- Database
- In-memory database (SAP HANA)
- AI inference

## High reliability

- Carrier-class component specifications and 100% derating design
- Redundancy design: PSU modules in 1+1 mode, fan modules in N+1 mode, and memory module sparing/mirroring
- 65 RAS features
- 99.99% availability
- 15% fewer failures than the industry average

## Ultimate performance

- 4 x 2nd Gen Intel® Xeon® Scalable processors, up to 28 cores and TDP 205 W per CPU
- 48 DDR4 DIMMs, up to 2933 MT/s; a maximum of 24 Optane™ Persistent Memory (PMem), up to 2666 MT/s; total memory capacity up to 15 TB on a single server
- No.1 in SPECrate2017\_int\_base/SPECvirt\_sc 2013 benchmark/SAP B4H benchmark tests
- Full mesh CPU architecture design, improving performance by 14.9%

## Unique value

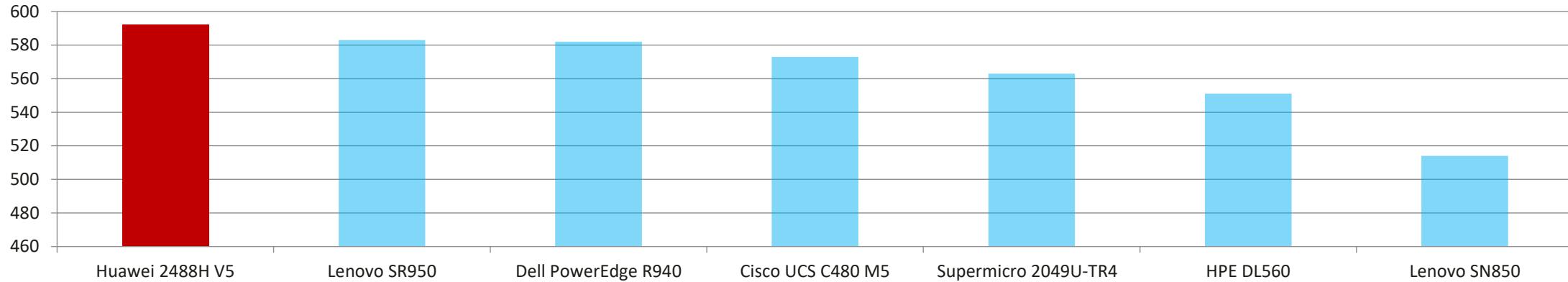
- **FDM:** The patented Fault Diagnosis & Management (FDM) technology provides the fault locating accuracy up to 93% and predicts a fault 7 to 30 days before it occurs, reducing OPEX
- **DEMT:** The patented Dynamic Energy Management Technology (DEMT) reduces energy consumption by 15% without compromising services
- **BSST:** Boot Speedup Storage Technology (BSST), 2 M.2 SSDs as high-speed OS boot drives, industry-unique hot plug and hardware RAID

## Innovative design

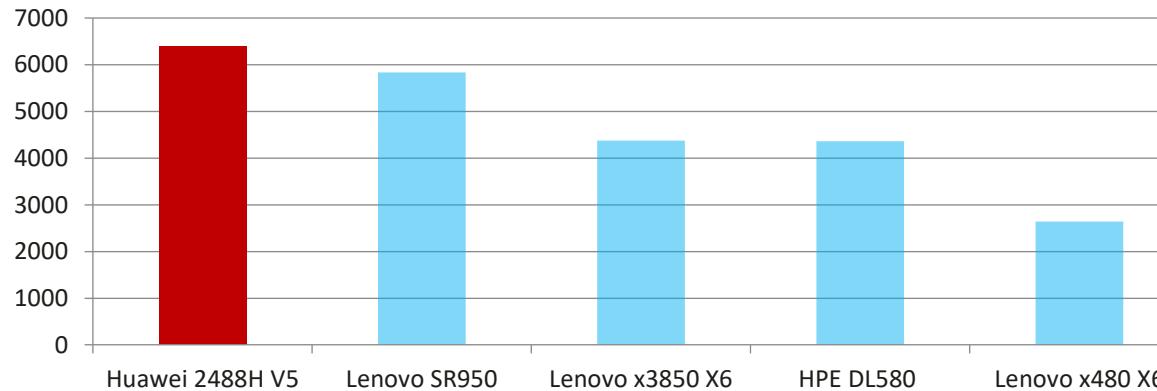
- 2 x 10GE and 2 x GE LOM ports, reducing CAPEX
- Flexible NICs for higher networking capabilities
- 25 x 2.5" drives, providing the industry's largest storage capacity
- 3 NVIDIA Tesla T4 GPUs for image and video inference and analytics
- Up to 11 PCIe 3.0 slots, the most in the industry for diverse applications

# Huawei 2488H V5: New SPEC/SAP HAHA Performance Records

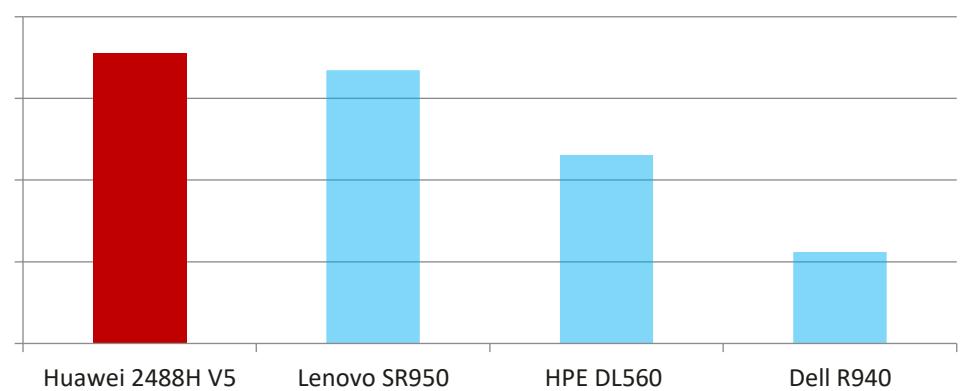
**SPECrate2017\_int\_base Test Result**



**SPECvirt\_sc2013 Test Result**



**SAP B4H Benchmark Test Result**



Data released on <http://www.spec.org/> as of Apr 2018

# 2488H V6: Stable and Reliable Mission Critical Server



- Virtualization
- Cloud computing
- High-performance computing (HPC)
- Database
- In-memory database (SAP HANA)
- AI inference and training

- ✓ 4S+48-DIMM in 2RU w/ Intel latest Xeon CPUs for Ultimate Computing Performance
- ✓ 25x2.5" HDDs or NVMe SSDs for high performance storage
- ✓ Up to 11 PCIe slots for expansion

## High reliability

- Carrier-class component specifications and 100% derating design
- Redundancy design: PSUs in 1+1 mode, fan modules in N+1 mode, and memory module sparing/mirroring
- 69 RAS features
- 99.999% availability
- 15% fewer failures than the industry average

## Ultimate performance

- 4 x 3rd Gen Intel® Xeon® Scalable processors, up to 28 cores and **TDP 250 W per CPU**
- 48 DDR4 DIMMs, up to 3,200 MT/s; 24 Intel® Optane™ persistent memory (Optane™ PMem) modules (200 series), up to 2,666 MT/s; total memory capacity of a single server up to 18 TB
- 2 x 300 W high-performance dual-width GPU accelerator cards, providing powerful computing for AI inference and training

## Unique value

- **FDM:** The patented Fault Diagnosis & Management (FDM) technology provides the fault locating accuracy up to 93% and predicts a fault 7 to 30 days before it occurs, reducing the OPEX
- **DEMT:** The patented Dynamic Energy Management Technology (DEMT) reduces energy consumption by 15% without compromising services
- **BSST:** The Boot Speedup Storage Technology (BSST) uses 2 M.2 SSDs as high-speed OS boot drives, supporting the hardware RAID and the industry-unique hot plug for M.2 SSDs

## Innovative design

- 2 FHFL dual-width GPU AI accelerator cards, such as Tesla V100
- 4 HHHL single-width GPU AI accelerator cards, such as Tesla T4
- 11 PCIe 3.0 slots, including one dedicated PCIe slot for the OCP 3.0 NIC (hot-swappable)
- 25 x 2.5" SAS/SATA drives, up to 24 NVMe SSDs<sup>1</sup>
- Access to iBMC using Type-C, enabling agile O&M

<sup>1</sup> The configuration of 24 NVMe SSDs is estimated to be released in Q4 2020.

# 5885H V5: High Performance and Scalability



- Virtualization
- Cloud computing
- HPC
- Database
- AI inference or training

- ✓ 4RU 4S+48-DIMM Scalable Ultimate Performance & Large Memory Capacity
- ✓ 15x PCIe Expansion slot for 4x T4 or 2x V100 Nvidia GPU cards for AI

## High reliability

- Carrier-class component specifications and 100% derating design
- Redundancy design: PSU modules in 1+1 mode, fan modules in N+1 mode, and memory module sparing/mirroring
- Maintenance of front fan modules without opening the chassis cover
- 65 RAS features
- 99.99% availability
- 15% fewer failures than the industry average

## Ultimate performance

- 4 x 2nd Gen Intel® Xeon® Scalable processors, up to 28 cores and TDP 205 W per CPU
- 48 DDR4 DIMMs, up to 2933 MT/s; a maximum of 24 Optane™ Persistent Memory (PMem), up to 2666 MT/s
- Full mesh CPU architecture design, improving performance by 14.9%

## Unique value

- **FDM:** The patented Fault Diagnosis & Management (FDM) technology provides the fault locating accuracy up to 93% and predicts a fault 7 to 30 days before it occurs, reducing OPEX
- **DEMT:** The patented Dynamic Energy Management Technology (DEMT) reduces energy consumption by 15% without compromising services
- **BSST:** Boot Speedup Storage Technology (BSST), 2 M.2 SSDs as high-speed OS boot drives, industry-unique hot plug and hardware RAID

## Innovative design

- 2 x 10GE and 2 x GE LOM ports, reducing CAPEX
- Flexible NICs for higher networking capabilities
- 4 NVIDIA Tesla T4 GPUs for image and video inference and analytics
- 2 NVIDIA Tesla V100 cards for AI training
- Up to 15 PCIe 3.0 slots, the most in the industry for diverse applications

# Huawei 9008 V5 for SAP HANA: Reliable & Scalable

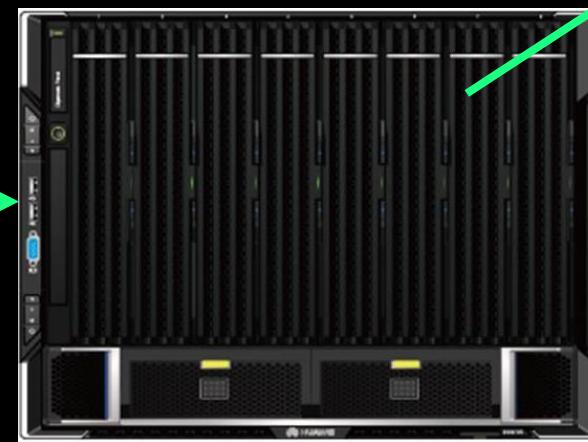


Central Management  
Enclosure (CME) (2U)

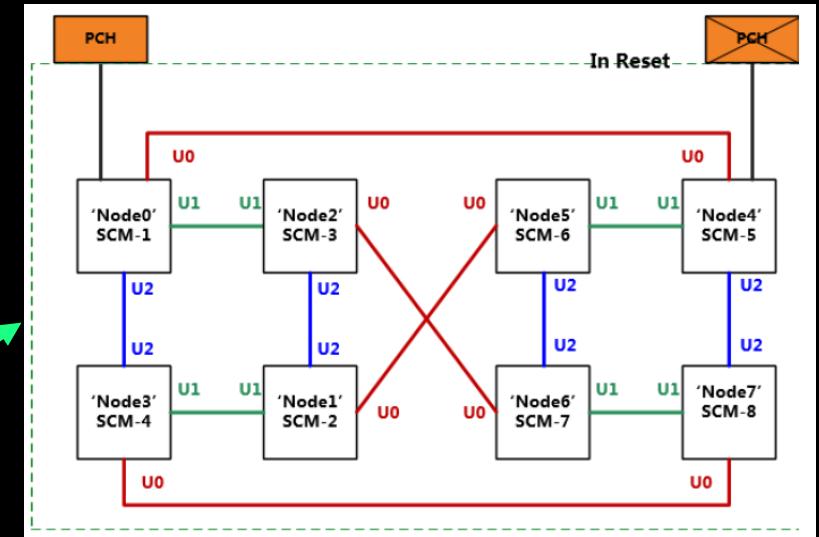
System Computing  
Enclosure (SCE) (8U)

Resource Expansion  
Enclosure (REE) (4U)

- Xeon Scale-up w/ Direct UPI Interconnect
- Up to 8X Xeon Platinum CPUs
- Up to 96 DDR4 DIMMs, Max 12TB DDR4 for V5
- Rich I/O, Storage & Expansion
- Modular Design w/ One CPU + Memory / Blade
- Scalable Power & Cooling for Xeon in 2022+
- Synchronous to Intel Xeon Release

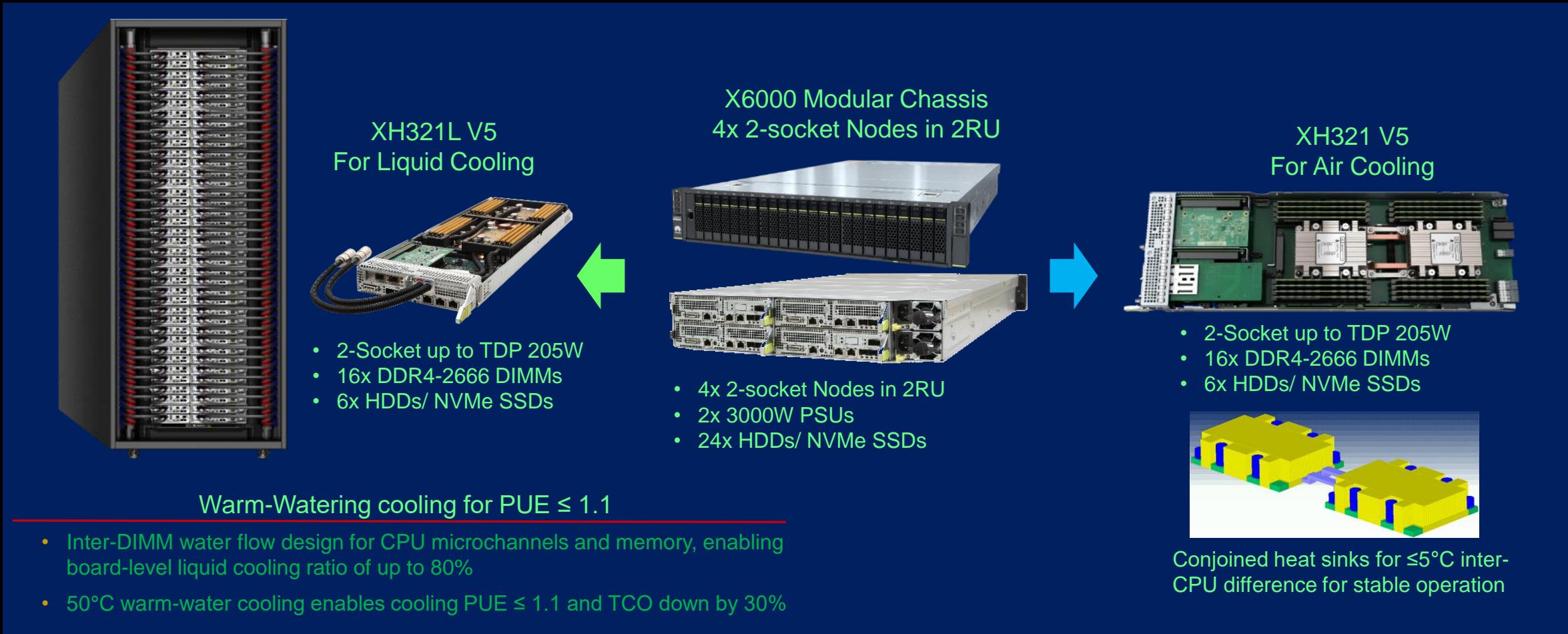


Direct CPU-CPU Interconnect



# X6000 V5: High-Density Modular w/ Air & Liquid Cooling

- ✓ 4-Node 2RU Modular High-Density Design for HPC & High-Density Compute Applications
- ✓ Liquid Cooling Option for Extreme-Low PUE & Density

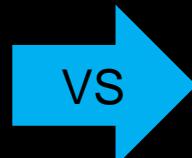


# CPU or Heterogeneous Computing for HPC, HPDA & AI?

- ✓ CPU+GPU/NPU Heterogeneous Computing is dominating the AI training market, but CPU+GPU only for about 25% HPC market
- ✓ CPU is adopting a lot of high-performance technologies used by GPU/NPU to improve HPC/HPDA/AI performance
- ✓ It is expected that CPU-only & CPU+GPU HPC/HPDA/AI solutions will continue co-exist for foreseeable future

## CPU + GPU/NPU

- GPU High-Performance Approaches
  - ✓ 1000s Small Cores
  - ✓ Tensor-Cores
  - ✓ High-Bandwidth HBM (TB/s)
  - ✓ New Data Types: BF16, TF32
  - ❖ Heterogeneous Programming
  - ❖ CPU-GPU Data Transfer overhead



## CPU Only

- CPU High-Performance Approaches
  - ✓ 100s Large Cores
  - ✓ Large Vector (AVX512, SVE) Unit
  - ✓ Matrix-Multiplication Engine
  - ✓ HBM (TB/s) as system memory
  - ✓ New Data Types: BF16, TF32
  - ✓ Homogeneous Programming
  - ✓ No CPU-GPU Data Transfer overhead

# HPC Fabric Evolution Trends: Ethernet-Based Growing

- ✓ Data Center Ethernet port speed in pace with or surpasses InfiniBand technologies

- ❖ IB w/ more In-Network Computing features (such as SHARP)

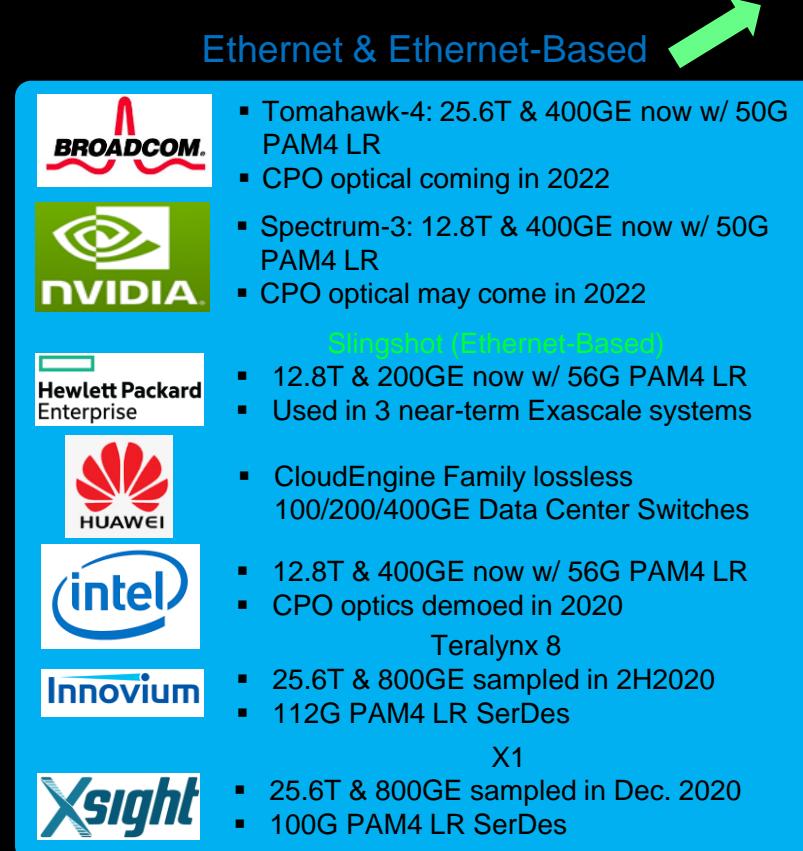
InfiniBand →



- 25.6T & 400GE now w/ 50G PAM4 LR
- CPO optical coming in 2022



- Discontinued 200G OmniPath (IB-based) Development in 2019



Other proprietary Interconnect →

- ✓ Fujitsu Fugaku Tofu-D 6D Mesh

# Top500 HPC Interconnect Trends: Ethernet is No.1 & Growing

- ✓ 25GE & Higher Growing Rapidly, 10GE & Higher Surpass IB a lot (260 vs. 141 in June 2020)
- ✓ In August 2019, Intel discontinued 200G OmniPath development, but acquired Barefoot Networks to enter Ethernet Switch market
- ✓ IB is still dominating Top10 and Top100 Interconnect

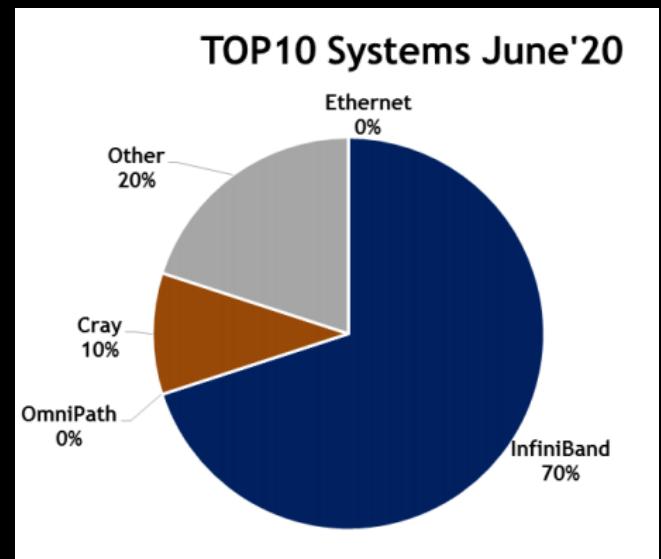
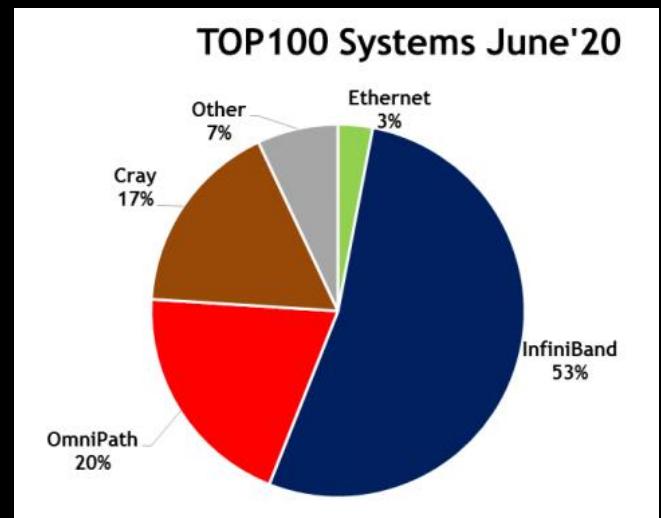
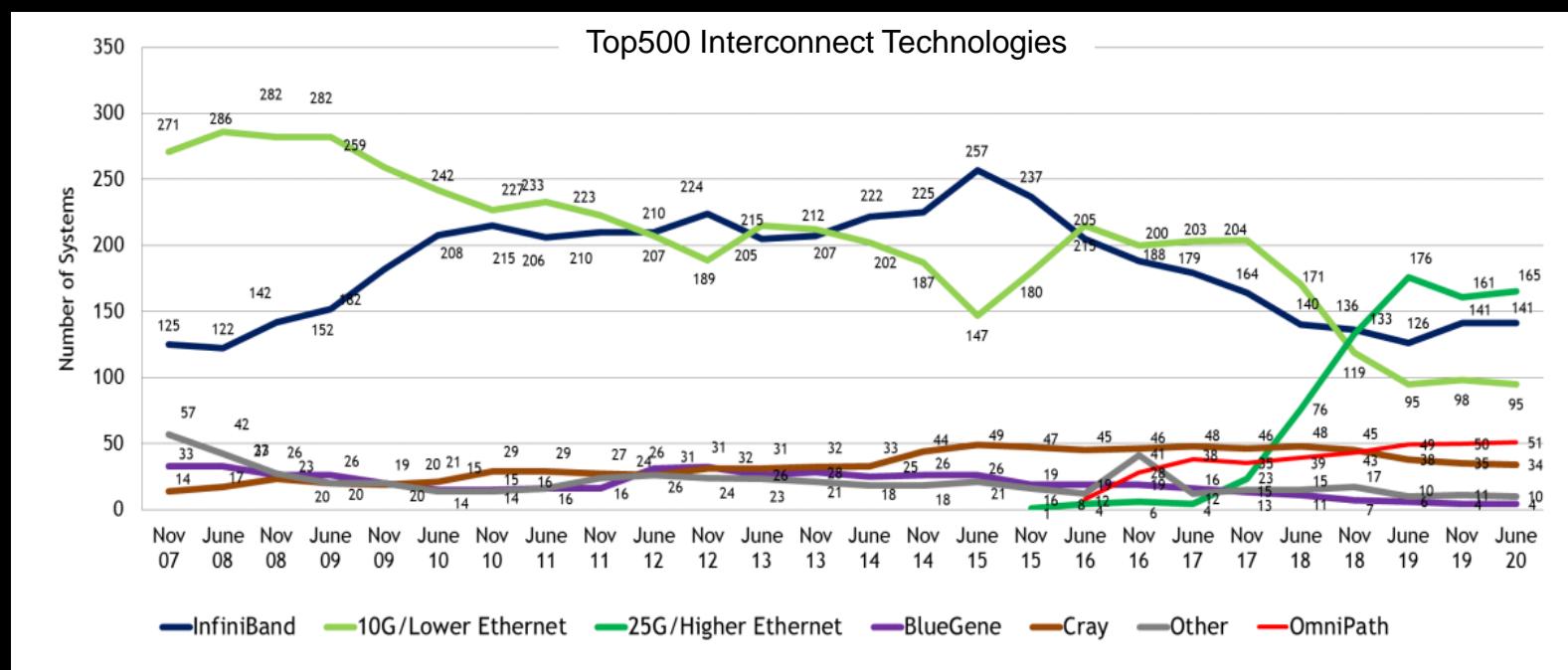


Figure Top500 Interconnect Technologies & Trends (image source & credit: Mellanox)

# Near-Term Exascale Systems Interconnect

- ✓ All the 3 Near-Term USA Exascale Systems adopt Ethernet-based HPE/Cray Slingshot Interconnect, not IB

Frontier at ORNL	Aurora at ANL	El Capitan at LLNL	Fugaku in Japan
<ul style="list-style-type: none"><li>❖ On in 2021-2022</li><li>❖ AMD EPYC CPU + AMD Radeon GPU (future-gen)</li><li>❖ ≥ 1.5 Exaflops DP</li><li>❖ System Interconnect: HPE/Cray <b>Slingshot</b>, an enhanced Data Center Ethernet, 25.6Tbps per Switch chip, 200G/Port</li></ul>	<ul style="list-style-type: none"><li>❖ On in 2022-2023</li><li>❖ Intel Sapphire Rapids CPU + Intel Xe GPU</li><li>❖ ≥ 1 Exaflops DP</li><li>❖ System Interconnect: HPE/Cray <b>Slingshot</b>, an enhanced Data Center Ethernet, 25.6Tbps per Switch chip, 200G/Port</li></ul>	<ul style="list-style-type: none"><li>❖ On in 2022-2023</li><li>❖ AMD EPYC CPU + AMD Radeon GPU</li><li>❖ ≥ 1.5 Exaflops DP</li><li>❖ System Interconnect: HPE/Cray <b>Slingshot</b>, an enhanced Data Center Ethernet, 25.6Tbps per Switch chip, 200G/Port</li></ul>	<ul style="list-style-type: none"><li>❖ On in 2020-2021</li><li>❖ Fujitsu A64FX ARMv8 <b>CPU only</b></li><li>❖ &gt; 400PF DP, &gt;800PF SP, 1.6EP HP</li><li>❖ System Interconnect: Fujitsu Tofu-D 6D Mesh, 28Gx2x10 Per Node</li></ul>



Figure Near-term Exascale System Interconnects (image source & credit: [www.hpcwire.com](http://www.hpcwire.com))



# Ethernet Switch ASIC Evolving Rapidly, CPO coming

- ✓ Broadcom announced industry first Co-Packaged Optics Ethernet Switch ASIC to be available in 2022
- ✓ 400GE Ethernet Switch is shipping, and 800GE is coming soon, in parallel with IB Speed Evolution

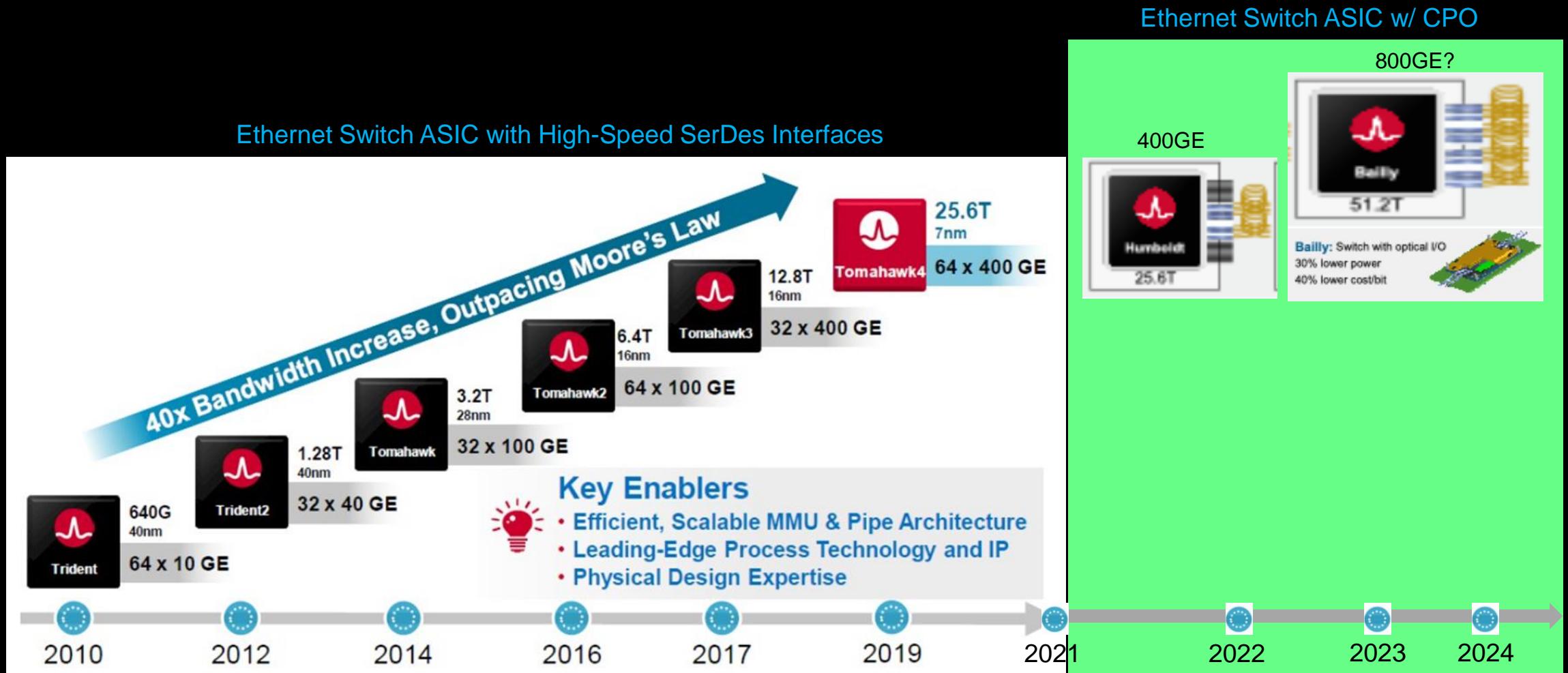
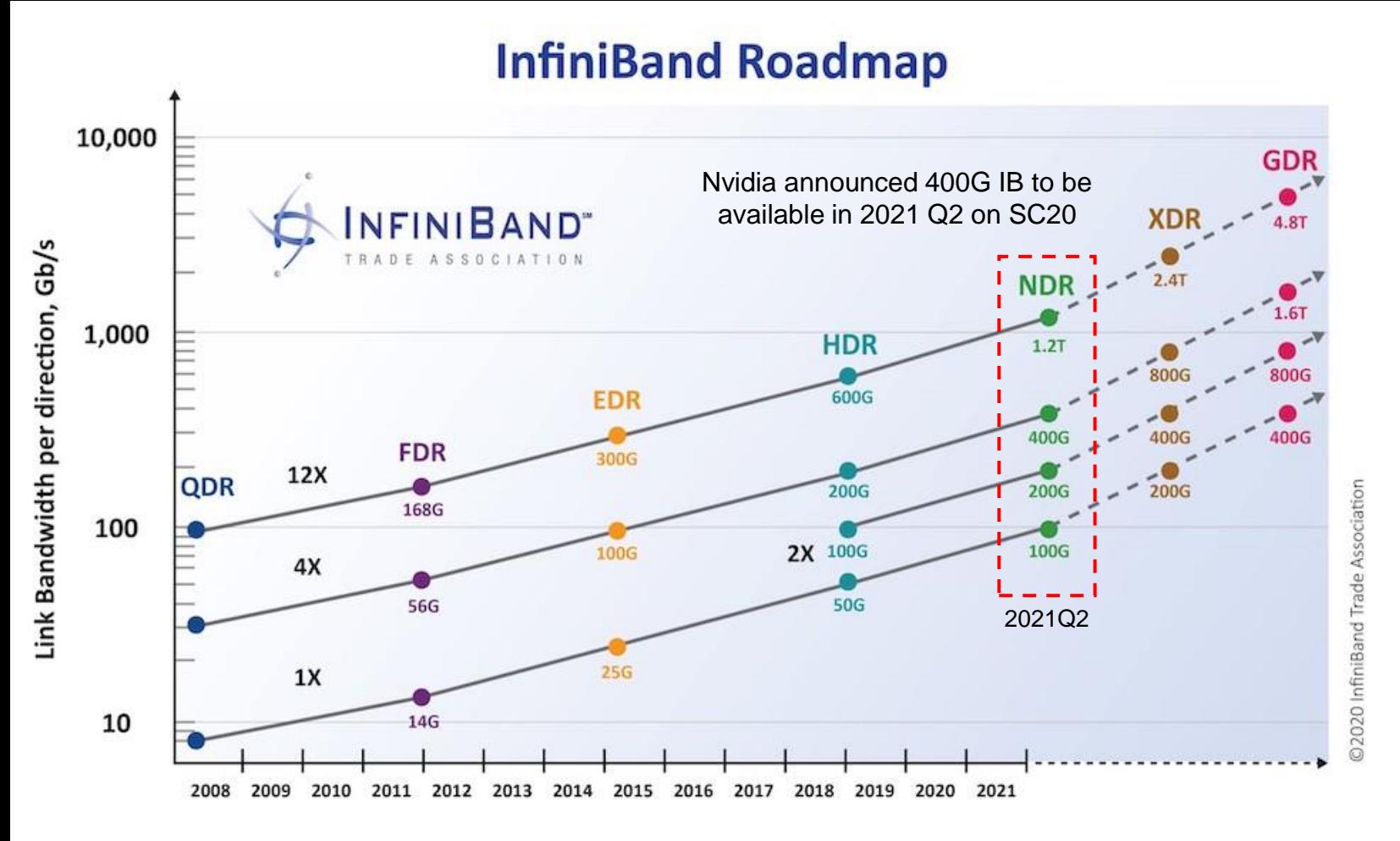


Figure Broadcom Data Center Ethernet Switch Evolution (Basic image source & credit: Broadcom)

# IB: Speed Evolution Parallel w/ Ethernet

- ✓ Speed evolution parallel with Ethernet, used to be ahead of Ethernet, but IB w/ more In-Network Computing features (SHARP)
- ✓ Intel discontinued 200G OmniPath (IB) development in 2019, leaving Nvidia/Mellanox the ONLY IB vendor
- ✓ For the same data rates, IB is still delivering better performance for some applications



# Why Ethernet-Based HPC Interconnect Becoming Popular?

- ✓ Latency caused by bad Congestion Control is much higher than Switch's P2P Latency where IB is a little better; Verified by Industry and our internal HPC & AI application benchmarkings
- ✓ Why Ethernet for HPC Popular: Comparable Performance, Unified Fabric for Lower TCO, No Need for IB talents & Costly High-Performance IB-Ethernet GW for storage, Native Network for NVMe-over-Fabric

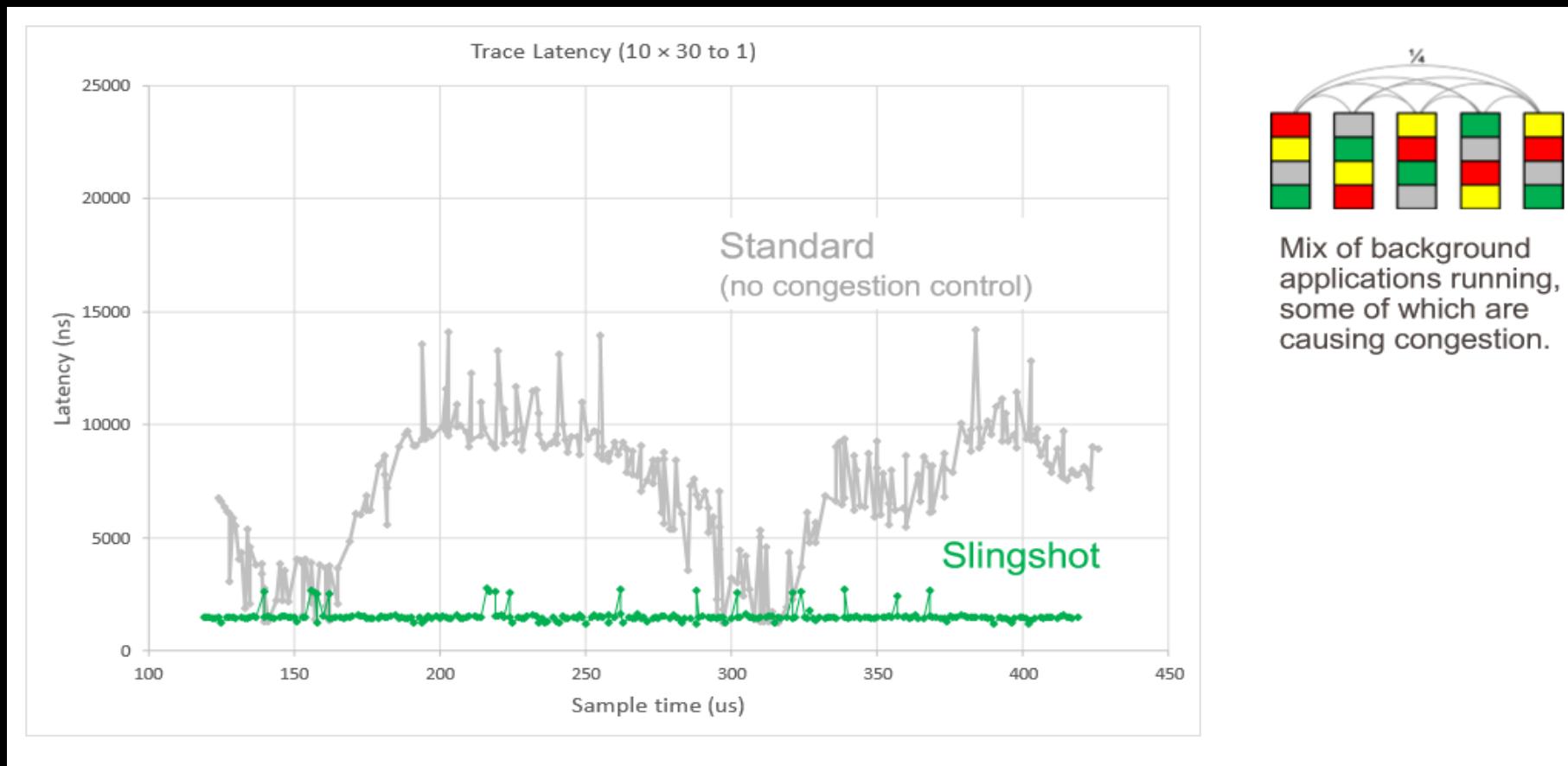


Figure Congestion Control is Key to End-to-End Latency (image source & credit: Cray via [www.Nextplatform.com](http://www.Nextplatform.com))

# Energy HPC POC w/ Huawei Lossless Data Center Switch

- ✓ We are doing a HPC POC with our lossless & enhanced Congestion-Control 100GE Switch with one of our energy customers in Middle East; and the initial result showing minor performance difference with the same speed IB
- ✓ Similar Result for Our performance benchmarking with GROMACS, LAMMPS, VASP, Quantum Espresso

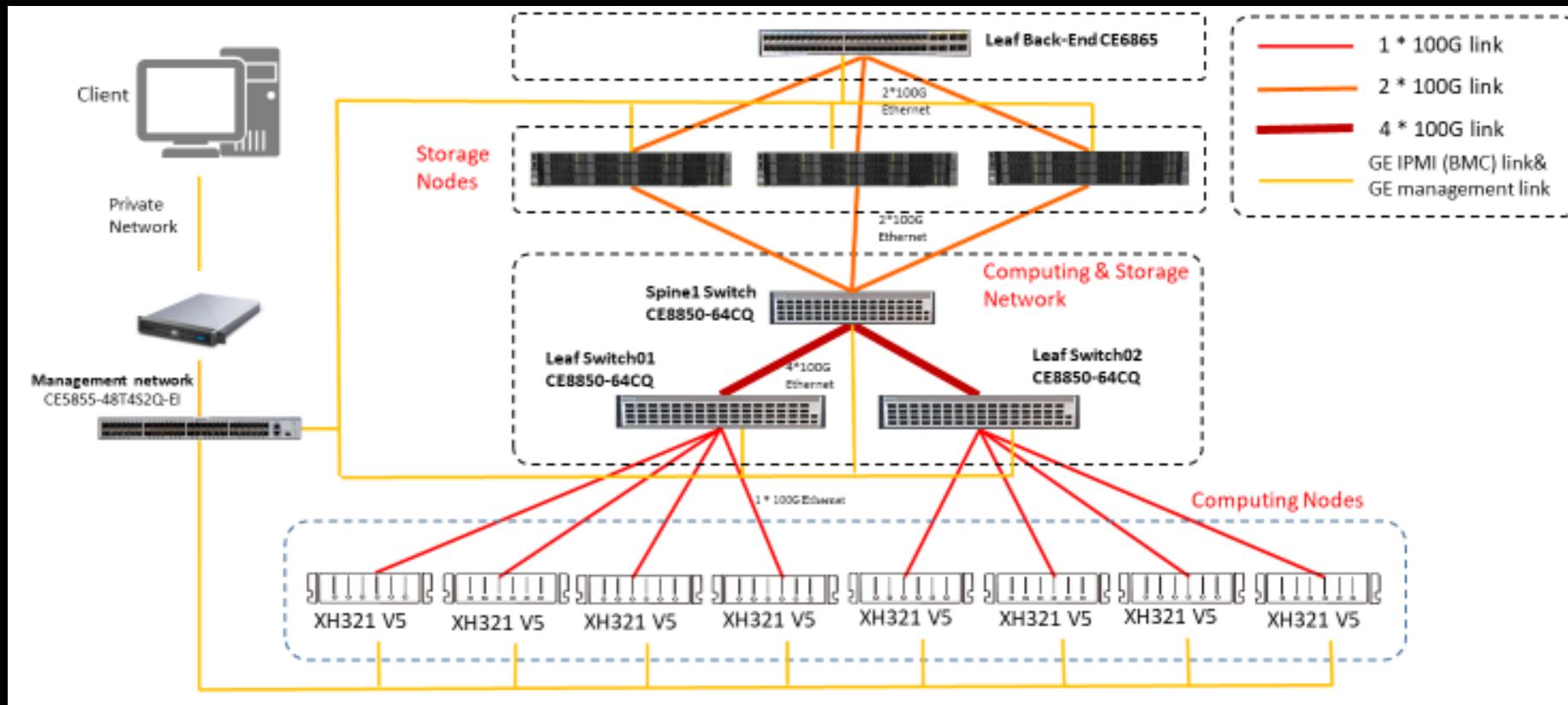


Figure Energy HPC Application POC with Huawei Lossless Data Center Switches (image source & credit: Huawei)



2.6  
PFLOP  
S

## Huawei Helps **Compute Canada** Build Supercomputing Platform to Propel Science Research

2.6 PFLOPS compute scale  
CPU & heterogeneous accelerated hybrid HPC cluster  
Diversified Workloads including Oil & Gas



# Huawei V5 Server Powers BGP HPC for Seismic Data Processing



## BGP of CNPC HPC Cluster:

- ✓ 256 Nodes of RH2288H V5
- ✓ 149 TFLOPS
- ✓ All IP Unified Interconnect
- ✓ Huawei OceanStor9000 file Storage for 40GB/s throughput



CEPSA, an Oil & Gas Company in Spain, Builds Efficient ERP Systems with Huawei SAP HANA Appliance, ensuring stable operation of ERP core systems.



# Huawei Servers Bolster Digital Transformation for the Iraq Branch of Gazprom



Gazprom is the world's largest natural gas company. Gazprom provides almost all the natural gas needed in Central and Eastern Europe and the former CIS countries. Gazprom ranked No. 49 on Fortune Global 500 2018.

Gazprom adopts Huawei 2288H V5 Rack Server for their Digital Transformation for its High Performance, Flexible Configurations, Reliability, 96% PSU efficiency, long-term 45°C working capability and FDM features

# Take-Aways

- ✓ CPU, especially Intel X86, will continue play a key role in HPC, Big Data Analytics and AI Applications
- ✓ CPU, especially Intel X86, are improving HPC, HPDA & AI Performance with Denser Vector (AVX512) & Matrix units
- ✓ New CPUs with HBM as System Memory and enhanced Vector and Matrix acceleration units will come soon, providing competitive performance to GPU for HPC, HPDA & AI with simpler programming model
- ✓ Data Center Ethernet is evolving in speed parallel to or surpassing InfiniBand, and is taking more market share in HPC, HPDA and AI applications
- ✓ In working with partners like Intel, VMware, SAP, Huawei has been offering a rich and innovative X86 Product Portfolio, delivering industry leading performance, reliability and manageability with elegant engineering, complete end-to-end process, strict sourcing and rigorous testing; as well as lossless & AI-based ECN flow-control CloudEngine Data Center 100/200GE Switch products

Thank you

