

# Building a Data-Centric, Trustworthy Storage Foundation for Diverse Applications

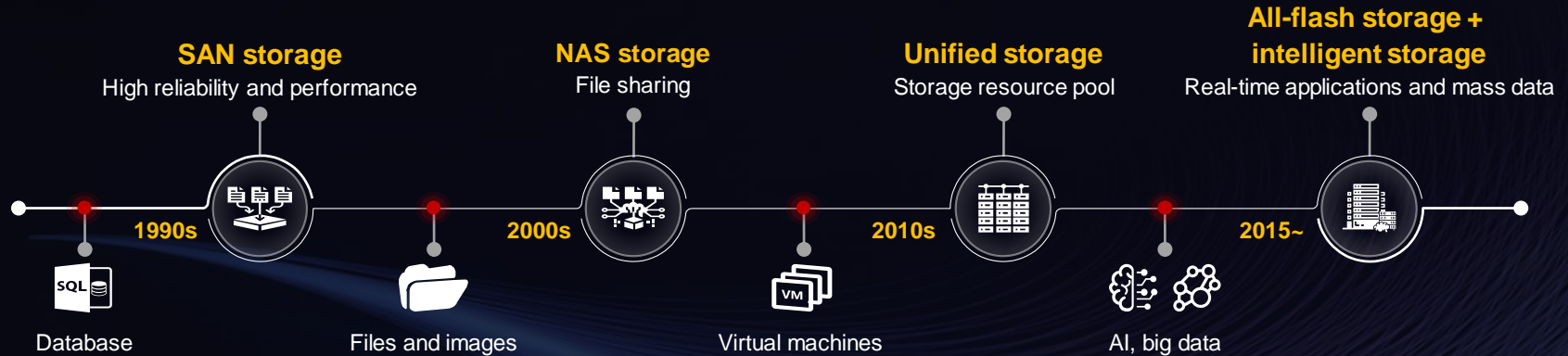
---

Dr. Peter Zhou

Vice President of Huawei

President of Huawei IT Product Line

# Data storage: the optimal foundation for high-value data to grow with new applications



# Four changes to data storage in the new era

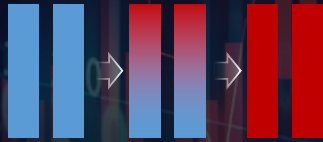
## Ever-changing data applications



**Avg. 110 data applications per enterprise**

Distributed database, big data, AI and HPDA

## Data is getting hotter



Operational analysis report of mobile banking: **days → hours**

Real-time processing:  $T + N \rightarrow T + 0$

## Urgent need for data protection



**\$1.85 million**

Avg. cost of recovery from ransomware attacks

## Energy efficiency: must-have for storage



**Annual energy consumption: 300 kWh/TB**

Energy consumption increases as data volume explodes



**Building a Data-Centric, Trustworthy Storage  
Foundation for Diverse Applications**

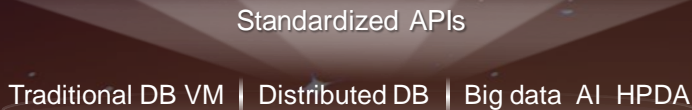
# Building a data-centric, trustworthy storage foundation for diverse applications

Multi-cloud connection



**Diversified data ecosystem**  
Distributed database, big data, AI and HPDA

Diverse data application acceleration engines



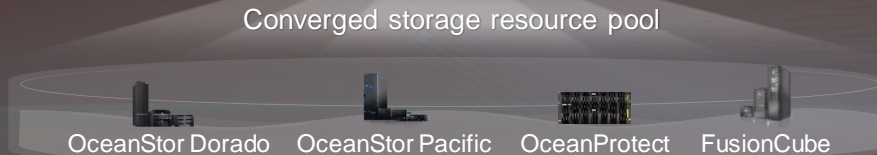
**High level of automation**  
Autonomous-driving storage

Intelligent data management



**Superior performance**  
µs-level latency and efficient data analytics

Optimal data storage foundation



**High reliability**  
0 data loss and 0 service interruption



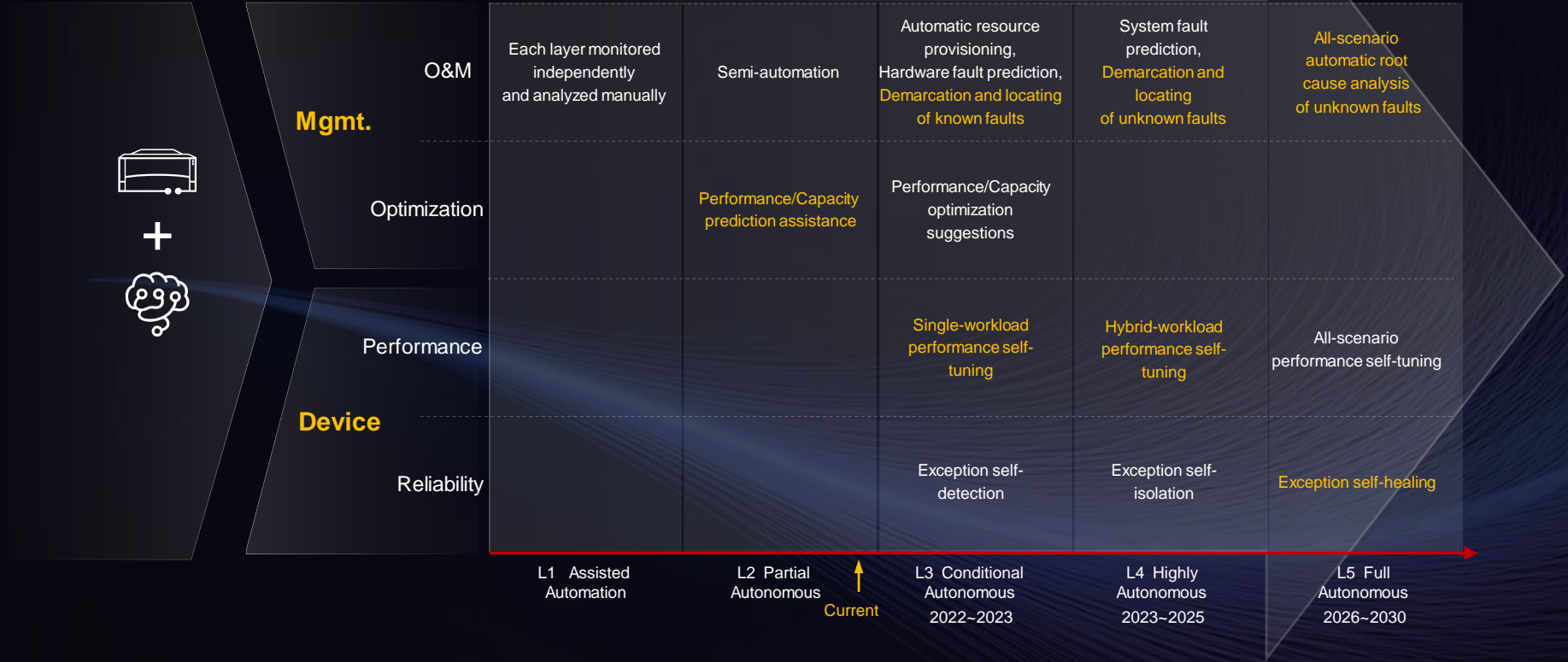
# Reliability: comprehensive reliability ensures zero data loss and zero service interruption



# Performance: $\mu$ s-level latency, diversified analytics loads, and efficient data processing



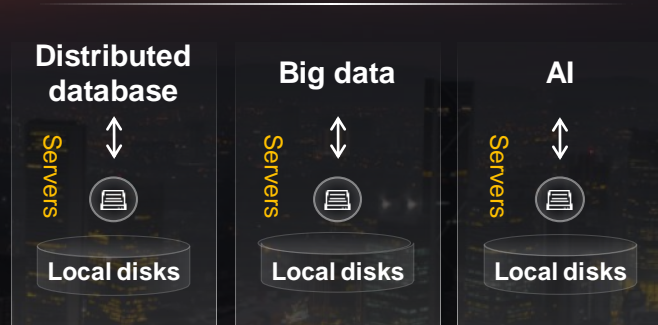
# +Intelligence: leap from L3 to L4 towards autonomous driving storage





# Diversified data ecosystem: building innovative, reliable and cost-effective data infrastructure with professional storage

## Server-based data application architecture: Simple resource planning



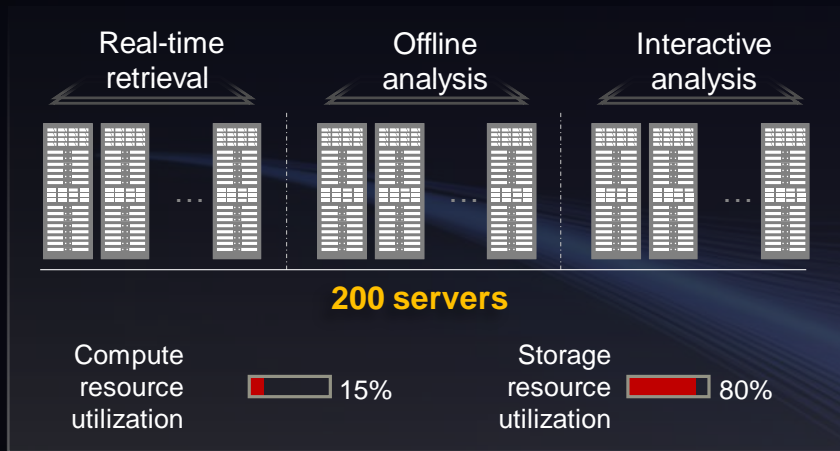
the gap between the computing power lifecycle and data lifecycle ↑  
computing and storage resources should be planned and maintained flexibly & independently

## Storage-compute decoupling architecture: High reliability and on-demand expansion

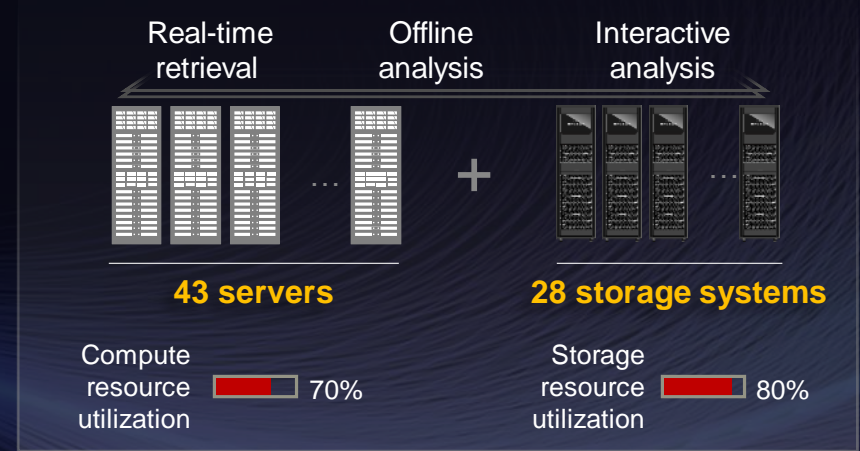


# Big data: storage-compute decoupling enables independent resource expansion, 30% TCO saving

## Storage-compute coupling



## Storage-compute decoupling for big data



# Building diverse data application acceleration engines to accelerate data processing by 10x

MySQL

## Distributed database acceleration engine

### Data multi-write enabling engine

Shared buffer pool

Global distributed lock management (DLM)

### Data sharing acceleration engine

Log Store

Page Store

Operator acceleration

- Failover in seconds, RPO = 0, RTO < 10s
- Dedicated read/write interfaces improve performance by 2x
- On-demand expansion reduces TCO by 30%

Spark | Flink | Kafka

## Big data acceleration engine

### Big data acceleration client

Memory access interface

Shared memory pool

### Metadata gateway

Global namespace

Unified metadata of data lakehouse

### Operator acceleration engine

Cache data prefetching

Eviction/Filter

- Global data management and analytics across data centers and locations
- Unified metadata formats save data storage space by 30%
- Dedicated operator acceleration increases big data processing efficiency by 5x

TensorFlow | PyTorch

## AI acceleration engine

### AI inference training acceleration engine

Dataset pipeline task scheduling

Global shared memory

### AI data management acceleration engine

Feature extraction, preprocessing, and associated operator offload

Feature store

- Dedicated feature storage and feature processing operator offload acceleration increase AI training efficiency by 10x

# Building a low-carbon society with green storage



## Green production

### Green energy

PV power systems

### Renewable materials

aluminum and tin



## Green product

### Improved hardware density

7.68TB SSD vs 1.2TB HDD, high-density hardware 120 disks/5U

### Reduced data volumes

72:1 deduplication and compression ratio, 22+2 Elastic EC algorithm

### Higher resource utilization

multi-protocol convergence, data center storage resource pool



## Green enabler

### Green finance

The all-flash storage saves 30% of power consumption

### Green manufacturing

The all-flash storage slashes electricity expenditure significantly and OPEX by 70%

# Green, acceleration, and innovation start a new journey of Huawei Data Storage



## Green

Green production  
Green product  
Green enabler



## Acceleration

$\mu$ s-level latency  
Diversified analytics workloads  
Efficient data analytics



## Innovation

Full-lifecycle intelligent data management  
Diversified Data Acceleration Engine