Baidu enables technology innovation with an ultra-scalable data platform

Multinational high-tech AI company replaces disks with high-performance tape, storing massive volumes of cold data, cutting costs by 80% and enabling rapid growth in AI, self-driving vehicles and more

IBM

Multinational high-tech AI company replaces disks with high-performance tape, storing massive volumes of cold data, cutting costs by 80% and enabling rapid growth in AI, self-driving vehicles and more

2

As Baidu launches innovative AI services, data volumes are skyrocketing. To reduce costs, increase efficiency and meet data compliance requirements, the company engaged IBM to replace legacy disk storage for cold data with a new solution based on IBM® TS4500 Tape Libraries and IBM Storage Scale software.

From autonomous vehicles to AI to the internet of things (IoT), Baidu is on the cutting edge of digital innovation. In addition to delivering China's largest internet search engine, the company's strategy includes self-driving, Baidu AI Cloud, Baidu Netdisk, and video streaming platforms.

Miao Yu, Senior M http://www.baidu.com/ anager of the Cloud Storage Department at Baidu AI Cloud, explains:

IBM

Multinational high-tech AI company replaces disks with high-performance tape, storing massive volumes of cold data, cutting costs by 80% and enabling rapid growth in AI, self-driving vehicles and more

3

"As more industries upgrade their digital capabilities, our aim is to provide cloud products to customers in industries such as transportation, finance and government. Through the cloud and intelligence capabilities of Baidu AI Cloud, we can help companies unlock cost-efficiencies and create new sources of value."

Across the Baidu platform, data is growing rapidly in volume, velocity and variety. The booming popularity of the company's Baidu AI Cloud and AI offerings has triggered a massive rise in storage requirements. At the same time, uptake of Baidu's smart city, smart home and vehicle automation solutions is rising sharply, further increasing the need for real-time data access and long-term data retention.

"Our data growth shows no sign of slowing," continues Miao Yu. "The latest autonomous vehicles are

Enables
power-consumption
saving of

> 90%

by replacing disks with tapes for cold data storage

Cuts operational
costs for cold
data by

> 80%

compared to previous disk storage platform

IBM

Multinational high-tech AI company replaces disks with high-performance tape, storing massive volumes of cold data, cutting costs by 80% and enabling rapid growth in AI, self-driving vehicles and more

4

equipped with far more sensors than their predecessors, and the amount of data generated per vehicle can be as high as 10 TB per day. Similarly, the fast growth of the smart home category and the widespread popularity of livestreaming in China all bring huge amounts of data, leading to storage challenges. Since 2019, our data volumes have more than tripled, and we now store approximately 100 exabytes [EB] of data."

In the past, Baidu relied on disk storage for long-term data retention. However, its disks had a maximum capacity of 20 TB per drive, limiting storage density, consuming valuable floorspace and increasing costs. Recognizing that this approach was not optimal, Baidu looked for a more resilient, efficient and scalable cold data storage platform that could accommodate continued data growth.

"The combination of IBM TS4500 Tape Libraries with IBM Storage Scale and IBM Storage Defender allows Baidu AI Cloud to reduce costs and scale rapidly to accommodate explosive data growth."
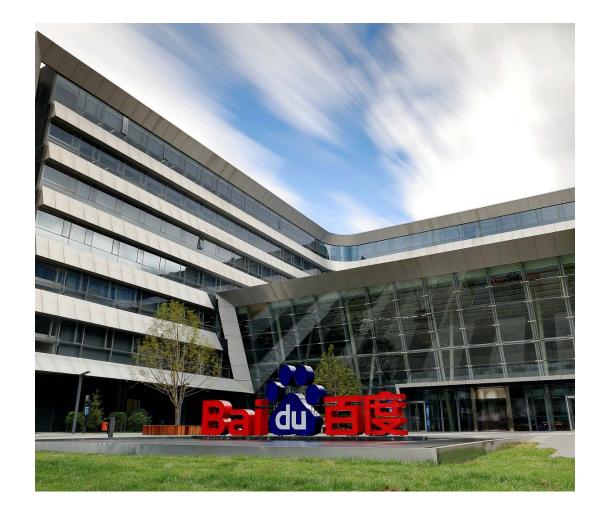
Miao Yu, Senior Manager of the Cloud Storage Department, Baidu AI Cloud

IBM

Multinational high-tech AI company replaces disks with high-performance tape, storing massive volumes of cold data, cutting costs by 80% and enabling rapid growth in AI, self-driving vehicles and more

5

Cutting costs, scaling rapidly

To meet its customers' needs for unlimited data scaling, Baidu AI Cloud engaged experts from IBM to plan, design, deploy and configure a future-ready cold data storage architecture that consists of IBM TS4500 Tape Libraries and the IBM Storage Scale and IBM Storage Defender solutions.

"Different types of data use scenarios put different demands on our storage system," explains Miao Yu. "For example, high-performance computing workloads require high throughput and low-latency read and write access. For other use scenarios, such as storing log files generated by our monitoring systems, we must be able to retrieve data rapidly even if years have passed since it was stored. We looked for a

IBM

Multinational high-tech AI company replaces disks with high-performance tape, storing massive volumes
of cold data, cutting costs by 80% and enabling rapid growth in AI, self-driving vehicles and more

6

cold storage solution with greater cost-efficiency, high IOPS and 24x7 availability—and IBM delivered."

The IBM solution integrates with Baidu AI Cloud's existing distributed storage architecture. IBM TS4500 Tape Libraries for cold data storage are used with legacy solid-state drives for hot data and hard-disk drives for warm storage. Data moves seamlessly across the tiered storage system, enabling high-performance access. What's more, tapes can be retained for over 30 years at a much lower cost than disks.

"We were impressed with the IBM solution," says Miao Yu. "With IBM Storage Scale, we can let data flow freely across our environment, while at the same time simplifying our storage architecture for easier management and maintenance."

To ensure that the new platform met Baidu AI Cloud's long-term requirements for scalability, IBM experts created automated deployment workflows. These allow 20 cold storage nodes to be provisioned in a single batch, significantly accelerating both deployment and expansion.

IBM also helped Baidu better adapt and integrate data from the IBM tape storage solution with its own platform, enabling Baidu to rapidly identify and resolve issues and bring the solution

online. To date, Baidu has deployed 14 IBM TS4500 Tape Libraries at its Yangquan data center, storing more than 2 EB of cold data.

Miao Yu elaborates: "IBM met and exceeded all our core selection criteria, and after a successful proof of concept we were certain that IBM TS4500 Tape Libraries with IBM Storage solutions would be the perfect way to solve the cost and space pressures we faced around disk storage. We have a long and successful history of collaboration with IBM on other projects, and this gave us the confidence that IBM has the technical innovation and services to address the challenges of deploying a large-scale storage solution."

Multinational high-tech AI company replaces disks with high-performance tape, storing massive volumes
of cold data, cutting costs by 80% and enabling rapid growth in AI, self-driving vehicles and more

7

Accelerating technology and service innovation

By replacing disks with tapes, Baidu has met and exceeded the success criteria it established at the start of its cold storage refresh.

"We were originally targeting 11 nines of reliability and an operational cost reduction of at least 50% of the previous disk storage solution, but IBM's storage solution surpassed both those targets," comments Miao Yu. "The IBM cold data storage solution delivers 12 nines of reliability and has reduced our operational costs by 80%— improvements that have far exceeded our expectations."

With cost-efficient, reliable tape storage and a high-performance data platform supporting its AI and big data workloads, Baidu AI Cloud can accommodate fast data growth and expand its innovative offerings and services.

"We must store the massive amounts of data on Baidu AI Cloud—as well as on our search, autonomous driving and other services—for up to three years or more, and make it available for analytics and compliance use whenever needed," explains Miao Yu.

"Today, this and other key data can be stored safely and cost-effectively in our cold storage platform and rapidly retrieved on demand. Tape is an efficient solution because it consumes very little electricity—over 90% less than an equivalent disk storage. So, we can significantly reduce our environmental footprint as well as saving costs."

Miao Yu concludes: "The combination of IBM TS4500 Tape Libraries with IBM Storage Scale and IBM Storage Defender software allows Baidu AI Cloud to reduce costs and scale

IBM

Multinational high-tech AI company replaces disks with high-performance tape, storing massive volumes
of cold data, cutting costs by 80% and enabling rapid growth in AI, self-driving vehicles and more

8

rapidly to accommodate explosive
data growth. Next, we plan to create
a unified tape storage management
platform for the whole of Baidu, and
we look forward to working with IBM
to unlock the full potential of tape
storage across the organization."

Junhua Jiang, Senior Account Manager
– Hyperscale Solutions Sales at IBM,
says: "We look forward to working with
Baidu to enrich their cloud storage
architecture and create a leading
global data service for Baidu Group
and industry clients."

"The IBM cold data storage solution delivers 12 nines of reliability and has reduced our operational costs by 80%—improvements that have far exceeded our expectations."

Miao Yu, Senior Manager of the Cloud Storage Department, Baidu AI Cloud

Multinational high-tech AI company replaces disks with high-performance tape, storing massive volumes
of cold data, cutting costs by 80% and enabling rapid growth in AI, self-driving vehicles and more

9

About Baidu, Inc. and Baidu AI Cloud

Founded in 2000 and headquartered in Beijing, China, Baidu, Inc. (link resides outside of ibm.com) is a leading AI company. The Baidu AI Cloud is Baidu's infrastructure for the smart era. With a full stack of AI technology capabilities, Baidu AI Cloud empowers thousands of clients across multiple industries with its advanced technology and comprehensive solutions.

Solution components

- IBM® Storage Scale
- IBM Storage Defender
- IBM TS4500 Tape Library