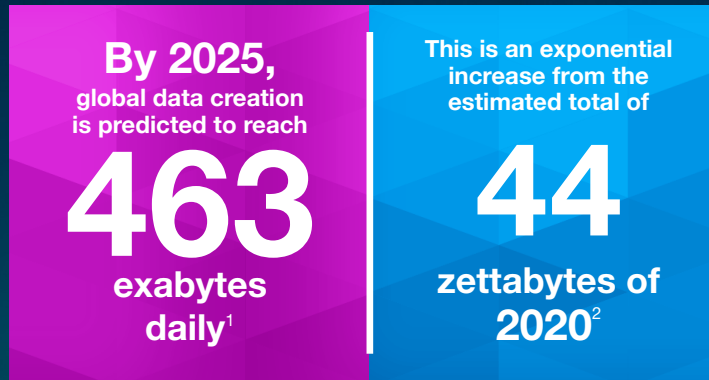# Micron®

Face the Data Deluge with Confidence

How to Meet Evolving Memory and Storage Needs

The World Is Changing. Fast.

These days consumers order groceries online, stream new movie releases from their couch and commute a few steps to another room in their house to work on their laptops. As the world becomes increasingly more digital, the amount of data generated now is practically unfathomable.

By 2025, global data creation is predicted to reach **463** exabytes daily[1]	This is an exponential increase from the estimated total of **44** zettabytes of 2020[2]

But your customers can do more than just prepare to withstand the data deluge. With the right flexible, scalable memory and storage infrastructure in place, they can **leverage this data** to make smart business decisions and capture actionable, insightful intelligence.

The Problem:
The Data Deluge Is Changing Almost Every Industry

Data can get out of hand quickly. As of 2020, 70 percent of data decision-makers report gathering data faster than they can analyze it.[3] But opting to limit data storage is likely not the right decision, either.

Many companies have integrated advanced systems that rely on data to fuel and train them, such as artificial intelligence (AI) and machine learning (ML). If your customer plans on leveraging AI and/or ML, then cutting back on data now might make it necessary to reassess or even reconfigure memory and storage infrastructure too quickly.

In other words: **no data is bad data.** But businesses need to know how to use the information they gather.

What Is the "Data Deluge"?

It's an industry term for the increasing amount of data in the world. Think of it as a flood of information that you can prepare for in advance.

AI and ML Are Adding Complexity

The data deluge is amplified by the rise of AI and ML and complicated by hybrid infrastructures that combine cloud and on-premise server systems. As more industries integrate these technologies, they create even more data. On average, businesses with high-performing AI store twice as much data in their data pipeline (1145TB vs. 537TB) and data lakes (1075TB vs. 516TB) than other organizations.[4]

Moreover, customers who might've once shied away from raising budgets or transforming traditional business might now be seeing the **need to innovate** alongside experts that understand the data deluge. A recent survey found that organizations **look to partners for help** managing and implementing an infrastructure that supports AI/ML.[5]

By 2023
the amount of new enterprise IT infrastructure deployed at the edge

could reach
50%
... up from less than 10% in 2019[6]

By 2024
the AI industry is forecasted to double from

$**50.1** billion to
$**110** billion[7]

By 2025
new enterprise apps using embedded AI

will rise at least
90%[8]

Businesses are embracing AI and ML now.

45% of surveyed IT professionals said their midmarket- or enterprise-level organizations currently have AI/ML project solutions in production.[9]

Organizations with AI/ML solutions in production are twice as likely to identify as being in a solid business position and a leader among competitors.[10]

The Solution:
Flexible, Scalable Memory and Storage

AI demands a new generation of faster, more flexible global infrastructures. The innate parallelism of AI architectures places a greater burden on memory and storage design and performance than ever before.

The good news: Memory and storage solutions are evolving to keep pace with big data and the next-generation technology that leverages it. Consider these needs when building out your customer's infrastructure:

1. Collection

AI processes depend on massive amounts of data. This size can vary, and the data is often in unstructured object or file formats such as videos, images or documents. This data collection requires **scalable storage** capacity like quad-level cell (QLC) SSDs built with NAND memory that can offer a balance between capacity, speed and cost.

2. Processing

Data must be filtered, distilled and efficiently organized to be useful — a problem solved by central processing units (CPU), adding accelerators and graphics processing units (GPU) in a **composable data center infrastructure** and supported by server DRAM, allowing some AI workloads, like huge datasets, to be shifted to dedicated hardware. This step can temporarily hold data while it's being preprocessed and feed it rapidly to the processor. NVMe SSDs can store the data once it is processed and becomes structured so it can be used for AI training.

3. AI Training

This training is extremely resource-intensive: It requires passing pieces of data through the training system hundreds or thousands of times and regularly retraining as new data flows in. Therefore, this process needs **powerful, flexible data centers** that can support high-bandwidth memory repeatedly feeding data at super-high speeds to the GPU or CPU to create the AI algorithm's logical connections.

4. Execution

Typically, as the trained AI model is inferring (making predictions), enterprises must continually evaluate its accuracy. The resulting trend or feedback analysis is usually part of an analytics process that uses data captured by devices and inference results. Although specifications depend on the use case, memory that provides a good balance of **high performance and low power consumption** can be ideal when executing inference on small remote or mobile devices.

Use Case:
Media Streaming

People don't just watch TV anymore — they binge it. And as more and more people stream their favorite movies or TV shows on demand, the companies that keep these customers happy and engaged must face their industry's own data deluge. These companies can accomplish this feat by using AI/ML powered by dynamic memory and storage solutions.

How AI/ML Is at the Core of Media Streaming

Personalization

Ever wonder how your streaming service seems to know you better than anyone? To make accurate recommendations, the engine captures and consolidates the following data:

- **Your movie-watching history**
- **Preferred film genres**
- **Movie synopses**
- **Casts and crews**
- **Movie reviews or social comments**
- **And even screenplay text**

It's a lot of data, and a massive amount of memory and storage are required to handle these workloads. **But the results are tailored content recommendations with eye-catching, individualized thumbnails that inspire clicks.**

Streaming Quality

Streamers want their content now, no excuses. ML can help predict what a consumer base might want to watch — and when. To improve streaming speeds, some tech optimizations happen behind the scenes.

Services use ML predictions to move the required data to an edge server that's closer to the end-user.

Such predictions enable companies to understand which servers to use for the **fastest download speeds,** minimizing poor-quality experiences.

Use Case:
Media Streaming

One size never fits all. Therefore, it can be critical in streaming to build flexible and efficient infrastructures that allow data to flow at speeds that enable consumers to enjoy their preferred content, however and whenever they want it.

Building the Right Infrastructure

This infrastructure can take many different shapes based on a service's unique needs. However, the following solutions can be used to support AI/ML workloads, including within media streaming:

- **Multichip packages, fast and vast storage, and high-performance and high-capacity memory power AI training and inference engines. This infrastructure works in the cloud or can be embedded in mobile/edge devices.**

- **Innovative memory solutions such as flash storage, DRAM, GDDR graphics memory, accelerators or FPGAs help serve up large datasets and volumes of data in real time, perfect when training AI systems.**

- **High-performance, low-power memory chips can be an option to accelerate inference for AI. These chips can help enable edge devices and keep them smart.**

Micron works together with key collaborators to build tailored solutions that adapt to your customers' particular needs — like how to enable a platform that lets everyone watch a brand-new, breakout TV show at the same time. Micron labs continually test memory and solutions to ensure exceptional, real-world performance.

Start Enabling Next-Generation Data Centers

The key to storing tomorrow's data is balancing your customer needs with the memory and storage infrastructure necessary to safeguard and enable it. You need experts in memory and storage who can guide you and your customers to the right solution to reach these objectives:

✔ **Maximize the infrastructure that your customer uses or expects to use.**

✔ **Meet tomorrow's challenges by offering agility and flexibility.**

✔ **Seamlessly enable cloud, on-premise or hybrid.**

Integrate the Future of Memory and Storage — Alongside an Industry Leader

Micron memory and storage have been foundational in AI's transformation into highly adaptable self-training and ubiquitous ML systems. Micron's fast, vast storage and high-performance, high-capacity solutions power AI/ML training and inference engines at scale, whether in the cloud or embedded in mobile and edge devices.

Find out how Micron helps you stay ahead.

Contact us to learn more

SOURCES

1. Jeff Desjardins, "World Economic Forum – How much data is generated each day?" Visual Capitalist, April 17, 2019 (accessed September 15, 2021).

2. Jeff Desjardins, "World Economic Forum – How much data is generated each day?" Visual Capitalist, April 17, 2019 (accessed September 15, 2021).

3. "Overcome Data Challenges With Data-As-A-Service," Forrester Consulting, May 2021 (accessed October 20, 2021).

4. ESG Research, AI/ML Storage Trends, September/October 2020 (accessed September 15, 2021).

5. ESG Research, AI/ML Storage Trends, September/October 2020 (accessed September 15, 2021).

6. Frank Gens, Meredith Whalen, Deepak Mohan, et al., "IDC FutureScape: Worldwide IT Industry 2020 Predictions," IDC, October 2019 (accessed September 15, 2021).

7. Stacey Soohoo and Michael Shirer, "Worldwide Spending on Artificial Intelligence Is Expected to Double in Four Years, Reaching $110 Billion in 2024, According to New IDC Spending Guide," IDC, August 25, 2020 (accessed September 15, 2021).

8. Frank Gens, Meredith Whalen, Deepak Mohan, et al., "IDC FutureScape: Worldwide IT Industry 2020 Predictions," IDC, October 2019 (accessed September 15, 2021).

9. ESG Research, AI/ML Storage Trends, September/October 2020 (accessed September 15, 2021).

10. ESG Research, AI/ML Storage Trends, September/October 2020 (accessed September 15, 2021).

version 01.13.2022