3 WAYS TO KEEP YOUR DATA CENTER REFRESH ON TRACK IN A *MEMORY-CONSTRAINED MARKET*

Don't postpone a server refresh because memory is in short supply. Here's how to move forward.

Companies have found that server refresh initiatives can help IT add capacity, improve uptime, meet security requirements, and consolidate while improving energy efficiency and total cost of ownership (TCO). But insatiable demand from AI means DRAM can be in short supply, causing some IT teams to consider postponing refreshes until the memory supply improves.

However, memory procurement may be difficult throughout 2026,[1] and new DRAM fabs will likely not reach full production for several years.[2] Delaying IT modernization can cost more than you expect. Instead, consider how you can right-size memory to help overcome challenges with availability and cost.

1 AUDIT YOUR ENVIRONMENT

Measure the memory resources you're actually using

In the recent past of plentiful, inexpensive memory, it's been a somewhat common practice for IT to overprovision memory to gain flexibility for workload placement. But with memory scarce and costly, this may be an unaffordable luxury.

Audit your on-prem and cloud environments to determine what resources (CPUs, GPUs, and memory) you have and how much you're actually using.

Tip: Measure a substantial duty cycle — weeks to months — instead of a point in time.

Street prices for most categories of DDR5 quadrupled between Q3 2025 and Q1 2026.[3]

AMD

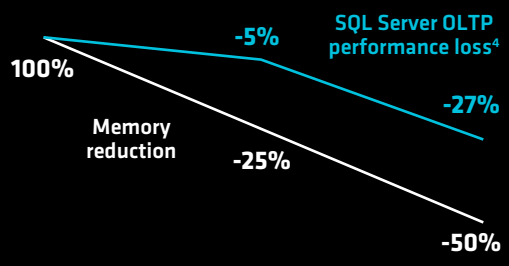
2 IDENTIFY NON-MEMORY-SENSITIVE WORKLOADS

Determine which workloads are not memory-sensitive (and which ones are)

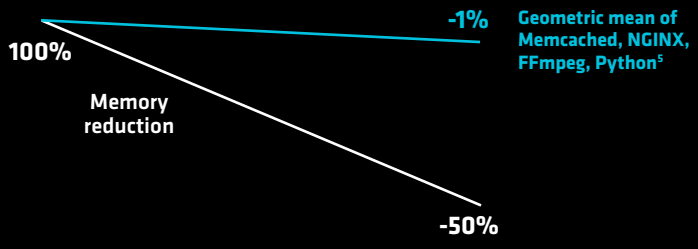
For many workloads, once a workload has enough memory to operate, incremental memory often has little performance impact. If workloads are CPU-, I/O-, or latency-bound, you can explore deploying less DRAM without major performance hits.

Tip: You can test workloads in the cloud for a fast, easy way to evaluate performance impact under multiple memory configuration scenarios.

REDUCTIONS HAVE MORE SIGNIFICANT IMPACT ON MEMORY-SENSITIVE WORKLOADS



REDUCTIONS HAVE MINIMAL IMPACT ON NON-SENSITIVE WORKLOADS



3 RIGHT-SIZE YOUR PROVISIONING

Provide the right amount of memory (plus a buffer) for each workload

Mission-critical and other high-priority workloads should receive a higher buffer for spikes in utilization or growth. Once you find where you can confidently reduce the memory needs of a server, there are two easy paths.

OPTION 1: LEAVE SOME MEMORY SLOTS OPEN

AMD EPYC Server CPUs can support many memory configurations, and not all memory slots must be populated, giving you more flexibility to choose the number of DIMMs and the capacity of each. It's simple to add more DIMMs later during a maintenance window.[6]

OPTION 2: CHOOSE SMALLER-CAPACITY DIMMS

If you can reduce the total capacity of memory in the server needed, but want to maintain maximum memory bandwidth, you can choose a full population of smaller-capacity DIMMs.[6]

WAITING TO REFRESH COSTS MONEY, TOO

Delaying a refresh means extending the life of what is likely inefficient infrastructure and bearing the costs of possibilities such as excess power consumption, maintenance, and downtime with no guarantees that memory or other components will be more readily available or less expensive in the near future.

Right-sizing memory can reduce not only your memory needs and costs but also impact power consumption and TCO. The result can change the outlook on a refresh, making it more affordable and accessible.

See how much a refresh and right-sized memory can save with AMD EPYC Advisory Tools.

You can also model your memory capacity needs and compare costs with our Cloud Advisory Tools.

AMD

1. AI Server Demand to Drive Memory Contract Price Increases in 2Q26 as CSPs Secure Supply via Long-Term Agreements, TrendForce, March 31, 2026.
2. How and When the Memory Chip Shortage Will End, IEEE Spectrum, February 10, 2026.
3. Memory Price Trends, PCPartPicker, April 16, 2026.
4. 9xx5-282: MySQL TPC-H derivative workload (SQL Server OLTP Brokerage) estimate based on internal AMD measurements as of 04/01/2026. The MySQL TPC-H workload is derived from TPC-Benchmark™ Standard, and as such is not comparable to published TPC-H™ results, as the results do not comply with the TPC-H Benchmark Standard.
   Workload configs: MySQL 8.0.39, SF3000
   1P AMD EPYC 9655 powered reference system (96 total cores), Ubuntu 24.04.2 LTS (Linux 6.8.0-60-generic), BIOS RVC100DB, SMT=on, Mitigations=off, Power Determinism, Memory Configurations: 12x128GB DDR5-6400 (100% capacity), 12x96GB DDR5-6400 (~75% capacity),
   12x64GB DDR5-6400 (~50% capacity)
   DIMM Score Relative
   128 2065198.0 1.00
   96 1949138.6 0.95
   64 1500947.6 0.73
   Workload performance with reduced memory capacity depends on workload characteristics and memory sensitivity. Results shown are limited to the specific workloads tested and should not be generalized to other workloads. Variables effecting these specific results include but are not limited to system configurations, software versions and BIOS settings. TPC, TPC Benchmark, and TPC-H are trademarks of the Transaction Processing Performance Council.
5. 9xx5-281: Geometric Mean of 4 select non-memory bound Phoronix workload results based on Phoronix Test Suite paid testing as of 04/01/2026.
   Workload Configs: Memcached - Set To Get Ratio: 1:10 (Ops/sec), nginx - Connections: 500 (Reqs/sec), Timed FFmpeg Compilation - Time To Compile (sec), Timed CPython Compilation - Build Configuration:
   Released Build, PGO + LTO Optimized (sec)
   1P 32C AMD EPYC 9355P-powered production system, BIOS 3.8, 2 x 3841GB SAMSUNG MZWLO3T8HCLS-00A07, 2 x Broadcom NetXtreme BCM5720 PCIe, Ubuntu 26.04 6.19.0-9-generic (x86_64), GCC 15.2.0, SMT=on, Memory Configurations: 12x64GB DDR5-6400 (at 6000 MT/s), 6x64GB DDR5-6400 (at 6000 MT/s)
   Workload 50% Memory 100% Memory Relative Normalized (HIB)
   memcached 5581297.57 5601003.16 0.996
   nginx 348480.18 347401.43 1.003
   ffmpeg 22.912 22.679 0.99
   python 203.185 202.906 0.999
   Geomean 0.997
   Workload performance with reduced memory capacity depends on workload characteristics and memory sensitivity. Results shown are limited to the specific workloads tested and should not be generalized to other workloads. Variables effecting these specific results include but are not limited to system configurations, software versions and BIOS settings. Testing not independently verified by AMD.
6. Always consult your vendors' memory population guides for complete guidance on allowable memory configurations.

AMD
together we advance_