美光推出全球首款高容量256GB LPDRAM SOCAMM2，為資料中心基礎架構樹立新標竿

March 5, 2026 at 11:00 AM CST

新聞重點：

- **採用業界首款 32Gb LPDDR5X 單晶粒，功耗僅為標準 RDIMM 的三分之一，規格尺寸亦縮小為三分之一**
- **長上下文 LLM 推論的首個token生成時間提升 2.3 倍，獨立CPU應用的每瓦效能提升 3 倍**
- **每個模組容量提升 1.33 倍 — 每顆 8 通道伺服器 CPU 可配置2TB LPDRAM，涵蓋AI及高效能運算（HPC）應用**

2026 年 3 月 5 日，愛達荷州博伊西 —美光科技（Nasdaq： MU）今日宣布開始正式向客戶送樣業界最高容量的 LPDRAM 模組256GB SOCAMM2，進一步鞏固其在低功耗伺服器記憶體領域的領先地位。此一里程碑由業界首款 32Gb LPDDR5X 單晶粒設計推動，為 AI 資料中心帶來變革性的重大突破，提供足以實現全新系統架構的低功耗記憶體容量。

AI 訓練、推論、代理式 AI 與通用運算的融合，正推動更嚴苛的記憶體需求，並重塑資料中心的系統架構。現代 AI 工作負載推動大型模型參數、龐大的上下文視窗及持久性關鍵價值（KV）快取的需求，而核心運算則持續在資料密集度、並行處理和記憶體使用量方面不斷攀升。

面對上述工作負載，記憶體容量、頻寬效率、延遲表現和能源效率已成為系統層級的主要瓶頸，直接影響效能、可擴充性和總體擁有成本。LPDRAM 憑藉上述特性的獨特優勢，使其在日益受限於功耗與散熱條件的資料中心環境中，成為 AI 及核心運算伺服器的關鍵解決方案。 美光正與 NVIDIA 攜手合作，共同設計先進記憶體解決方案，以滿足先進 AI 基礎架構的需求。

美光雲端記憶體業務部門資深副總裁暨總經理 Raj Narasimhan 表示：「美光 256GB SOCAMM2 為 AI 與高效能運算提供最具能源效率的CPU附加記憶體解決方案。此次產品發佈充分展現美光在技術與封裝領域的突破，打造業界容量最高、功耗最低、規格尺寸最小的模組化記憶體解決方案。美光在資料中心低功耗記憶體解決方案領域持續保持領先地位，使我們具備獨特優勢，率先推出 32Gb LPDRAM 單晶粒，協助推動業界加速採用更節能、更高容量的系統架構。」

專為容量、能源效率和工作負載效能最佳化而設計

美光的 256GB SOCAMM2 為各種 AI 及通用運算工作負載提供更高的記憶體容量、大幅降低的功耗以及更快的效能表現。

- **為 AI 伺服器擴充的記憶體容量：**256GB SOCAMM2 較前一代最高容量 192GB SOCAMM2 提升三

分之一的容量，為每顆 8 通道 CPU 提供 2TB 的 LPDRAM，適用於更大的上下文視窗和複雜的推論工作負載。

- **功耗更低，規格尺寸更小：**SOCAMM2 的功耗僅為同級 RDIMM 的三分之一，同時僅使用三分之一的規格尺寸，有效提升機架密度並降低總體擁有成本。[1]
- **提升推論和核心運算效能：**在整合記憶體架構中，與現有的解決方案相比，256GB SOCAMM2 用於 KV 快取卸載時，可將長上下文即時 LLM 推論中首個 token 的生成時間提升 2.3 倍。[2] 在獨立 CPU 應用中，LPDRAM 每瓦效能比主流記憶體模組高出 3 倍以上，適用於高效能運算工作負載。[3]
- **模組化設計提升可維護性和可擴充性：**模組化 SOCAMM2 設計可提升可維護性，支援液冷式伺服器架構，並可隨著 AI 和核心運算記憶體需求持續成長，實現未來容量擴充。

NVIDIA 資料中心 CPU 產品部門主管 Ian Finder 表示：「先進AI 基礎架構需要在每個層面進行極緻的最佳化，才能有效應對高要求 AI 推論工作負載對效能與效率的需求。美光透過 256GB SOCAMM2，以低於傳統伺服器記憶體的功耗，實現大規模記憶體容量與頻寬的卓越成就，為次世代 AI CPU 提供關鍵助力。」

推動業界標準定義並加速低功耗記憶體普及

美光持續在 JEDEC SOCAMM2 規格定義中扮演領導角色，並維持與系統設計業者的深度技術合作，推動整個產業在次世代資料中心平台能源效率和效能方面的提升。

美光 256GB SOCAMM2 已進入客戶送樣階段，並提供業界最完整的資料中心 LPDRAM 系列產品，涵蓋 8GB 至 64GB 元件及 48GB 至 256GB SOCAMM2 模組。

其他資源

- 大規模 LPDDR應用：透過高容量記憶體實現高效 LLM 推論
- 每一瓦至關重要：低功耗記憶體如何改變資料中心
- SOCAMM2 產品網頁
- 資料中心記憶體產品網頁

關於 Micron Technology, Inc.

我們是創新記憶體和儲存空間解決方案的業界領導者，致力於改變世界使用資訊的方式，豐富所有人的生活樣貌。美光持續專注於用戶需求、技術領先、卓越的製造與營運，我們提供高效能 DRAM、NAND 及 NOR 記憶體與儲存產品的完整組合。每一天，我們人員提出的創新推動了數據經濟、人工智慧和 5G 應用程式的發展，激發從資料中心、智慧邊緣到用戶端與行動裝置的多元機會與使用者體驗。欲進一步瞭解 Micron Technology, Inc.（Nasdaq：MU），請瀏覽tw.micron.com。

美光媒體關係聯絡人
Mengxi Liu Evensen
+1 (408) 444-2276
productandtechnology@micron.com

美光投資者關係聯絡人
Satya Kumar
+1 (408) 450-6199
satyakumar@micron.com

---

[1]三分之一的功耗係依據單個 128GB、128 位元匯流排寬度 SOCAMM2 模組與兩個 64GB、64 位元匯流排寬度 DDR5 RDIMM 的功耗瓦數進行比較計算。三分之一的規格尺寸係基於 SOCAMM2 面積（14x90 mm）與標準伺服器 RDIMM 之面積進行比較。

[2]結果依據美光內部測試，使用 Llama3 70B 模型（ FP16 量化）進行即時推論測試，測試條件為 500K 上下文長度及 16 位並行使用者。預計的 TTFT 延遲改善係基於每顆 CPU 配置 2TB LPDRAM 時延遲為 0.12 秒，對比每顆 CPU 配置 1.5TB LPDRAM 時延遲為 0.28 秒。更多測試條件詳情，請參閱本月稍早發布的白皮書：大規模 LPDDR 應用：透過高容量記憶體實現高效 LLM 推論。

[3]美光內部測試，以相同容量的LPDDR5X 和 DDR5 執行 Pot3D 太陽物理 HPC 程式碼之效能量測。