

## 分层内存助力提升虚拟机内存容量并降低总体拥有成本

在内存模式下将大容量英特尔® 傲腾™ 持久内存用于主系统内存，将 DRAM 用于内存缓存，从而提高虚拟机密度，支持更大内存容量的虚拟机，同时降低资本和运营支出。



### 解决方案优势

- 英特尔® 傲腾™ 持久内存可为 VMware vSphere 等虚拟化数据中心基础设施提供比 DRAM 更大的每主机内存容量。英特尔傲腾持久内存提供了更多的扩展空间，帮助企业有效托管需要更大内存容量的虚拟化工作负载，而不是在裸机上运行这些要求苛刻的工作负载。
- 实施采用英特尔傲腾持久内存的分层内存解决方案可以优化总体拥有成本 (TCO)。企业可以通过整合服务器来节省成本，或在配置系统时减少所用的高成本 DRAM，从而降低资本支出。
- 访问存储子系统可能带来的较高延迟让企业无法接受。将更多数据保存在更大的系统内存中通常可以提高工作负载性能。

### 要点综述

随着数据量继续呈爆炸式增长，如今的数据中心需要具有更大内存容量的虚拟机 (参见图 1)。但是，DRAM DIMM 容量的扩展速度无法满足企业对于内存的需求，并且与处理器内核数量的增长不匹配。另外，购买数以 TB 计的 DRAM 的成本过于高昂。

内存分层解决方案是 VMware vSphere 部署的理想选择。与广为接受的分层存储概念类似，内存分层使用价格合理的大容量内存作为主易失性系统内存，并将传统 DRAM 作为内存缓存。大多数虚拟化应用和数据库都是使用 DRAM 和英特尔® 傲腾™ 持久内存 (PMem) 的分层内存架构的理想选择，因为它们经常消耗虚拟机未频繁使用的主机内存。

### 改变应用世界 软件引领云时代创新

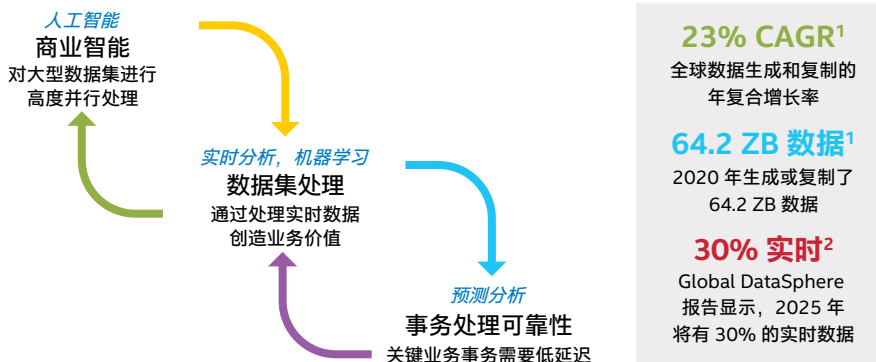


图 1. 企业数字化转型推动数据呈指数级增长。<sup>1,2</sup>

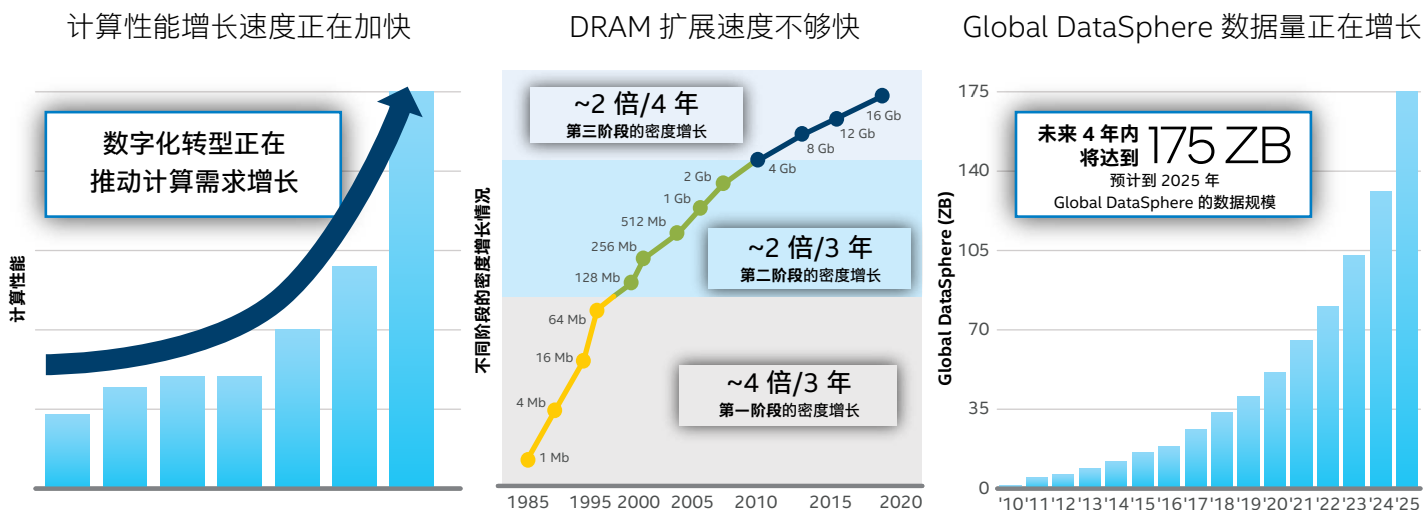


图 2. 计算性能和数据需求正在迅速增长，而内存容量却无法跟上这一需求<sup>3,4</sup>

## 业务挑战：DRAM 容量无法满足工作负载需求

随着越来越多的企业开展数字化转型，市场上出现了三个重要趋势（见图 2）：

- 数字化转型推动了对计算的需求，内存需求也随之增加。
- 业务数据也在指数级增长。然而，传统的 DRAM 无法满足这一需求。随着时间的推移，DRAM 密度的增长速度已经放缓，扩展容量变得越来越昂贵和复杂。也就是说，业务数据增长正在增加对计算能力的需求，但 DRAM 可能成为瓶颈。<sup>3</sup>
- 即使 DRAM 容量增长放缓，内存需求仍在继续增加，导致内存缺口不断扩大。<sup>4</sup>

如今在 VMware vSphere 上运行的关键任务型和数据密集型工作负载需要在内存中存储更多的热数据，但大规模 DRAM 的成本越来越高且容量有限。此类工作负载包括大型内存数据库和实时工作负载。对于这些类型的实时工作负载，将数据尽可能靠近 CPU 至关重要，这样才能保持较低的延迟。

鉴于这些实际情况，企业需要新的内存层来解决这些问题。分层内存系统使用与分层存储相同的方法（参见图 3）。在分层内存场景中，内存模式下的英特尔® 傲腾™ 持久内存（PMem）作为大容量层，而相对少量的 DRAM 作为内存缓存层，以保持高速度和低延迟。

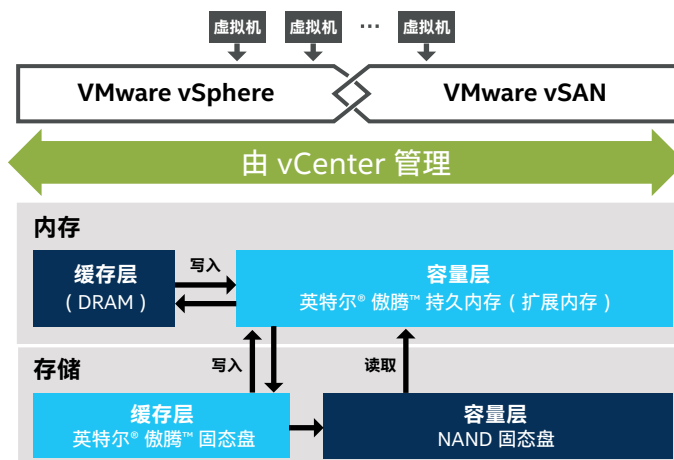


图 3. 在分层内存系统中，英特尔® 傲腾™ 持久内存（PMem）用作主系统内存，而系统的 DRAM 用作内存缓存。

## 内存模式下的英特尔® 傲腾™ 持久内存（PMem）说明

过去，系统中有两个不同的位置来存储数据：易失性内存 (DRAM) 或持久性存储（如固态硬盘）。但如上所述，DRAM 的容量不够大或价格不够经济，无法满足如今更大的内存需求，即使是基于 NVMe 的固态硬盘在现代化的实时工作负载中也会产生不可接受的高延迟。

英特尔傲腾持久内存填补了 DRAM 和存储之间的巨大差距；它是一种兼具记忆和存储特性的全新介质类型。英特尔傲腾持久内存模块看起来非常像 DRAM DIMM，并在相同的内存总线上提供了相同的物理 DIMM 接头。

此外，英特尔傲腾持久内存支持字节寻址，这有助于降低延迟（每个字节都可以被立即设置或重置，这与使用块和页面的存储访问不同）。

英特尔傲腾持久内存可用于两种主要模式。在本文中，我们将关注用于硬件内存分层的内存模式。在内存模式下，使用英特尔傲腾持久内存不需要更改任何软件应用；只需将英特尔傲腾持久内存模块插入服务器上的 DIMM 插槽，VMware ESXi 即可立即使用它们。请注意以下事项：

- 在内存模式下，英特尔傲腾持久内存实际上并不持久化 — 它像 DRAM 一样具备易失性，但容量更大。
- VMware vSphere 将英特尔傲腾持久内存容量识别为 vRAM；DRAM 缓存不计入系统总内存容量。
- 所有内存（DRAM 缓存和英特尔傲腾持久内存容量）均由 CPU 的内存控制器管理，以实现低延迟（参见图 4）。操作系统、内核和虚拟机管理程序无法控制 DRAM 缓存。

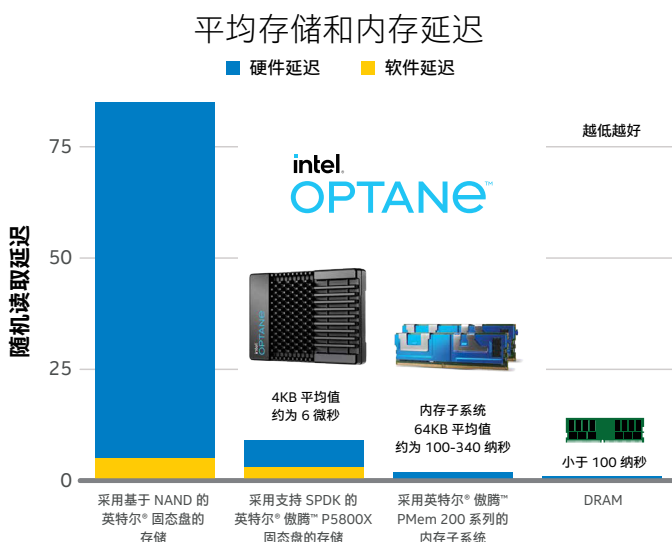


图 4. 由于驻留在内存总线上，因此相比块存储延迟，英特尔® 傲腾™ 持久内存 (PMem) 延迟更接近 DRAM 的延迟。<sup>5</sup>

要了解如何在内存模式下最好地使用英特尔傲腾持久内存，请参见图 5。系统为每个虚拟机分配了一定数量的内存。在分配的内存中，应用正在频繁访问或使用其中一些内存。其余分配的内存要么处于闲置状态（最近没有被访问或使用过），要么处于可用状态（长期没有被虚拟机上的任何应用访问或使用）。活动内存和闲置内存的组合构成了虚拟机所消耗的内存。

在配备英特尔傲腾持久内存的分层内存系统上，DRAM 中的数据是存储在英特尔傲腾持久内存中的活动数据子集。最佳做法是调整 DRAM 的容量，以便所有工作负载的活动内存都适用于 DRAM 的占用空间。通常而言，活动内存小于所耗内存 25% 的主机和应用能够提供与具有单层 DRAM 的主机相同的性能。

内存模式下的英特尔傲腾内存概述：

- 英特尔傲腾持久内存作为易失性内存。
- DRAM 作为内存缓存。
- 企业无需修改软件应用即可获得价格合理的大容量内存。



图 5. 活动内存和闲置内存构成了虚拟机的所耗内存，与可用内存一起构成了为该虚拟机分配的总内存。

## 解决方案价值：分层内存提供出色的成本效益，并支持更多的虚拟机数量

英特尔傲腾持久内存主要在两个方面提供了卓越的业务价值：增加虚拟机密度（每个虚拟机的成本更低），或在相同的虚拟机密度下通过调整 DRAM 容量并利用内存分层来节省资本支出（CapEx）。下面将介绍有关每种场景的更多详细信息，并提供英特尔傲腾持久内存验证点。

### 增加虚拟机密度

VMware ESXi 主机通常受内存限制，而不受 I/O 或 CPU 限制。随着内存分配的增加，VMware ESXi 主机上可添加更多的虚拟机。增加每台服务器的虚拟机数量可提高服务器整合率、CPU 利用率和数据中心效率，并减少数据中心所需要部署的服务器数量。企业将可以降低服务器部署所带来的能源成本、运营成本和软件许可成本，从而获得显著的总拥有成本（TCO）优势。

调整 DRAM 容量，节省资本支出并支持更大的虚拟机通过大幅增加 VMware ESXi 主机系统内存，企业可以负担得起内存容量更大的虚拟机。可能的场景包括：

- 已经虚拟化但内存需求不断增加的应用和数据库。
- 由于内存要求，以前不适合虚拟化的大型裸机应用和数据库。
- 更好地配置新的大型虚拟机，以满足业务需求。

通常的做法是在特定虚拟机上过度配置 DRAM，以满足未来的内存增长需求（因为数据呈指数增长），并处理周期性的工作负载峰值（例如假日在线购物流量增加，或是月/季度/年末结算工作）。

但是，这种做法可能会导致大量系统内存处于闲置状态（图 6）。该快照展示了虚拟机内存资源的 30 天使用图表。在这个真实的示例中，虚拟机的应用只需要将虚拟机所分配的大约 14% 的内存用于活动内存。即使纳入高于此水平 50% 用于负载峰值和安全缓冲，该虚拟机的活动数据的建议 DRAM 容量也仅为所分配内存的 22% 左右——这意味着此虚拟机分配的 DRAM 中大约有 77% 处于闲置状态。在这种情况下，英特尔傲腾持久内存可用于过度配置的内存，从而降低所需的 DRAM 容量并帮助降低总体拥有成本。

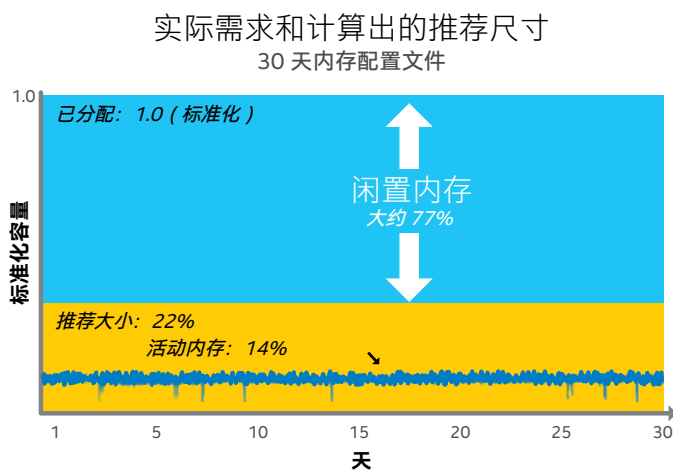


图 6. 集群的累积虚拟机内存利用率配置文件显示，活动数据使用量远低于分配的 DRAM 容量。

## 解决方案架构：面向全新数据领域的大容量内存

基于第三代英特尔® 至强® 可扩展处理器的 VMware ESXi 部署可充分利用第二代英特尔® 傲腾™ 持久内存——英特尔傲腾持久内存 200 系列。这些处理器具有 8 条内存通道，在内存模式下支持高达 8TB 的英特尔傲腾持久内存，相比之下，双路 DRAM 平台仅能容纳 4TB 内存。为实现最佳带宽并保持每内核内存或每虚拟机内存性能，非常有必要保持所有 8 条内存通道处于活跃状态。

图 7 列举了英特尔傲腾持久内存和 DRAM 的几种分层配置。随着工作负载对内存需求的增加，目前配置具有 384 GB DRAM 的服务器的组织可能考虑升级至 512 GB。我们的许多客户使用的是 768 GB 或更高内存，目前正在向 1TB 或更大内存迁移。但借助英特尔傲腾持久内存，系统内存可增加至 1 TB、2 TB 或更高，从而大大降低了预算支出。也可以使用英特尔傲腾持久内存 100 系列（和第二代英特尔至强处理器），如图 7 下半部分所示。

## 采用英特尔® 傲腾™ 持久内存 200 系列的第三代英特尔® 至强® 可扩展处理器

### 2 TB 双路系统

每 CPU 16 个插槽 (8+8)

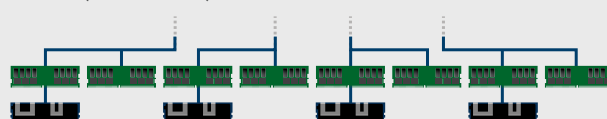
最大内存容量和带宽



### 1 TB 双路系统

每 CPU 12 个插槽 (8+4)

高带宽，更大容量，可显著节省 DIMM 插槽上的主板空间



DRAM (16 GB 或 32 GB DIMMs, 取决于工作负载缓存需求)

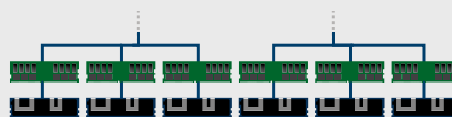
英特尔® 傲腾™ PMem (128 GB DIMM)

## 采用英特尔® 傲腾™ 持久内存 100 系列的第二代英特尔® 至强® 可扩展处理器

### 1.5 TB 双路系统

每 CPU 12 个插槽 (6+6)

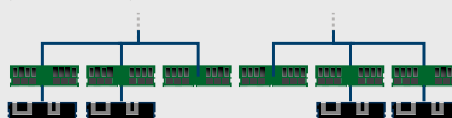
最大内存容量和带宽



### 1 TB 双路系统

每 CPU 10 个插槽 (6+4)

高带宽，更大容量，可显著节省 DIMM 插槽上的主板空间



DRAM (16 GB 或 32 GB DIMMs, 取决于工作负载缓存需求)

英特尔® 傲腾™ PMem (128 GB DIMM)

图 7. 使用英特尔® 傲腾™ 持久内存的分层内存配置可帮助组织达到自身的性能和 TCO 目标。

## 将英特尔傲腾持久内存用于 VMware vSphere 的最佳实践

VMware vSphere (vSphere 6.7 EP10 及更高版本) 支持基于第二代英特尔至强可扩展处理器的英特尔傲腾持久内存 100 系列。vSphere 7.0U2 及更高版本支持基于第三代英特尔至强可扩展处理器的英特尔傲腾持久内存 200 系列。支持多种配置，而且 VMware 提供多种最佳实践支持将英特尔傲腾持久内存用于 VMware 软件。例如：

- DRAM 大小建议如下：
  - DRAM 与英特尔傲腾持久内存的比例为 1:8，也就是 DRAM 容量为英特尔傲腾持久内存容量的 12.5%
  - DRAM 与英特尔傲腾持久内存的比例为 1:4，也就是 DRAM 容量为英特尔傲腾持久内存容量的 25%
- 系统必须每插槽安装 4 根英特尔傲腾持久内存 DIMM。
- 如果稳定状态下主机的活跃内存存在主机配置的 DRAM 容量范围内，将可获得最高性能。即使 DRAM 高速缓存未命中，内存延迟仍然在几百纳秒范围内。<sup>6</sup>
- 验证服务器平台是否按照服务器 OEM 推荐的平衡配置 BIOS 设置运行。

请阅读 VMware 知识库，“vSphere 支持英特尔傲腾持久内存 (PMem)”，获取关于最佳实践和支持配置的完整介绍。如欲更详细地了解能满足这些要求的服务器平台，请联系您的 OEM 服务器厂商了解支持的配置。



## 结论

通过使用分层内存打破 DRAM 限制扩展内存容量，VMware ESXi 或类似的环境可享有广泛优势。（私有云或混合云中的）大多数虚拟化、通用的工作负载对活跃内存的利用率较低，非常适合采用分层内存结构。这类工作负载的例子包括虚拟桌面基础设施 (VDI)、数据库等。

当技术趋势要求虚拟机在不影响性能的情况下支持更大的规模，英特尔内存技术的突破提供了一个机会，支持打破传统架构的限制，整合服务器资源，大幅降低 TCO。英特尔傲腾持久内存是一款经济适用的解决方案，可直接利用现有的 DDR4 内存插槽。IT 部门可以适当调整 DRAM 投资规模，扩展每台服务器的总内存，容纳更大型的工作负载，提高虚拟机密度，并显著提升资源利用率。

致力于推进数字化转型，对基础设施进行现代化改造、提高运营效率，以满足日益增长的数据需求的企业，可以充分利用英特尔傲腾持久内存，发挥分层内存系统的优势。

## 了解更多信息

以下是可能会对您了解更多信息有所帮助的资源：

- [第三代英特尔® 至强® 可扩展处理器](#)
- [英特尔® 傲腾™ 持久内存](#)
- [英特尔 + VMware 协作页面](#)
- [英特尔 + VMware 虚拟化解决方案](#)
- [采用英特尔® 傲腾™ 技术的 VMware 解决方案](#)
- [VMware 知识库文章 — vSphere 支持英特尔傲腾 PMem](#)

寻找适合贵公司的解决方案。请联系您的英特尔代表或者访问采用英特尔® 傲腾™ 技术的 VMware 解决方案。

<sup>1</sup> 23% 的年复合增长率和数据复制免责声明：<https://www.idc.com/getdoc.jsp?containerId=prUS47560321>

<sup>2</sup> 实时数据免责声明：<https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

<sup>3</sup> “DRAM 扩展：3D NAND 技术 — 对企业存储应用的影响”，J. Yoon, IBM；2015 闪存峰会。

<sup>4</sup> 数据扩展：《数据时代 2025》，由 Seagate 赞助发布，采用了 IDC Global DataSphere 的相关数据，2018 年 11 月。

<sup>5</sup> **NAND**：基于截至 2020 年 3 月 16 日英特尔对 6.4 TB 英特尔® SSD P5600 的测试：英特尔® 至强® 金牌 6139 处理器 (2.30 GHz)，BIOS：SE5C620.86B.00.01.0014.070920180847 (英特尔® 服务器主板 S2600WFT)，双插槽，32 GB RAM (DDR4-2137)，RAM 填充：4 条通道中的 1 条，填充 4 个 DIMM 插槽，PCIe 连接：CPU (未连接 PCH 通道)，芯片组：英特尔® C612 芯片组，交换机/ReTimer 型号/厂商：英特尔® G4SAC 交换机 (PCIe Gen4)，NVMe 驱动程序：kernel 4.17.74 (原生)，C 状态：禁用，英特尔® 超线程技术 (英特尔® HT 技术) 禁用，CPU 调节器 (通过 OS)：性能模式，OS：CentOS 7.5，核心：4.14.74，FIO 版本：3.5。

**英特尔® 傲腾™ 固态硬盘**：基于截至 2020 年 9 月 25 日英特尔对 1.6 TB 英特尔® 傲腾™ SSD P5800X 的测试。英特尔® 至强® 金牌 6254 处理器 (3.10 GHz，30 MB，160 W，每路 18 核，双路)，BIOS：SE5C620.86B.02.01.0009.092820190230，32 GB RAM (DDR4)，RAM 填充：不适用，填充的 DIMM 插槽：4 个插槽，PCIe 连接：CPU (未连接 PCH 通道)，芯片组：英特尔® C610 芯片组，交换机/ReTimer 型号/厂商：英特尔® G4SAC 交换机 (PCIe Gen4)，OS：CentOS 7.5.1804，核心：4.14.74，FIO 版本：3.5；NVMe 驱动程序：Inbox，C 状态：禁用，英特尔® 超线程技术禁用，CPU 调节器 (通过操作系统)：性能模式。增强型英特尔® SpeedStep® 技术，英特尔® 睿频加速技术和 P 状态禁用；IRQ 平衡服务 (OS) 关闭；操作系统中设为 SMP affinity；队列深度 1 (QD1) 利用 I/O 轮询模式，且 ioengine=pvsync2/hipri。

**英特尔® 傲腾™ 持久内存**：英特尔傲腾 PMem 200 系列测试由英特尔在单 DIMM 配置中执行，时间截至 2020 年 9 月 25 日。第三代英特尔® 至强® 可扩展处理器；芯片组 LBG B1，26 核，单路，DDR 速度：2,666 MT/秒；256 GB 英特尔傲腾 PMem 模块，15 W 内存配置，1 条通道，32 GB DDR4 (每插槽 6 个)；英特尔傲腾 PMem 固件：2.2.0.1516；BIOS 0017.P23；已知最佳配置 (BKC) 版本 WW38 BKC，Linux OS Fedora 版本 29 4.20.6-200.fc29.x86\_64；Spectre/Meltdown 已安装补丁 (1, 2, 3a, 4)，性能调优服务质量禁用，IODC=5 (AD)。英特尔傲腾 PMem 使用 DDRT 内存接口和 MLC 测量性能。

**DRAM**：<https://gist.github.com/jboner/2841832> (见第 7 行)。

<sup>6</sup> 请参见尾注 5。