案例研究

英特尔® 傲腾™ 数据中心级持久内存
短视频

快手推荐系统及 Redis 升级存储 借傲腾™ 补上 DRAM 短板

快手
记录世界 记录你

作为短视频领域的领先企业，快手正以实时、高效和精准的视频内容博得海量用户的青睐。随着用户量和短视频作品数量的爆炸式增长，快手也需要不断导入更先进的技术手段，来调整和优化其系统架构。作为短视频系统存储、分发和推荐的核心组件，其存储系统的优化和性能提升，也正面临着非常艰巨的挑战。

为应对短视频应用中高吞吐率、大数据量请求应用场景带来的挑战，快手与英特尔一起，通过深入的技术协作，在国内率先将英特尔® 傲腾™ 数据中心级持久内存产品应用于其推荐系统和 Redis 服务，并通过一系列的软件调优，成功构建起全新的推荐异构存储系统，并优化了 Redis 服务，为其提供了更具优势的存储能力。

来自快手的测试与实践表明，英特尔® 傲腾™ 数据中心级持久内存在新的推荐异构存储系统和升级后的 Redis 服务中，不仅有着与 DRAM 内存相近的性能表现，而且其大容量和非易失性的特性还可帮助系统获得更优的可用性。此外，它相比于 DRAM 内存的成本和容量优势，也可帮助快手有效地降低其推荐系统和 Redis 服务的总拥有成本（Total Cost of Ownership, TCO）。

快手实现的解决方案优势：

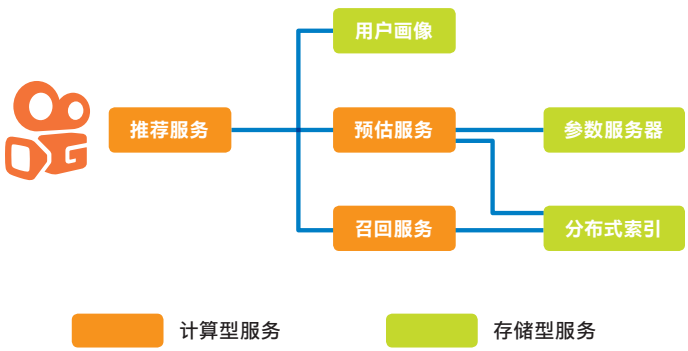
- 引入英特尔® 傲腾™ 数据中心级持久内存构建的快手推荐异构存储系统，除了可以满足请求量、网络带宽、平均处理时延等推荐系统的核心性能指标，在容量和成本方面相比基于 DRAM 内存的方案也更具优势；
- 英特尔® 傲腾™ 数据中心级持久内存非易失性的特性，让快手推荐系统具有更强的可用性，其故障恢复时长可获得高达百倍的缩短[1]；
- 采用英特尔® 傲腾™ 数据中心级持久内存的快手异构存储系统，在满足性能需求的同时，还能帮助快手推荐系统的 TCO 降低 30%[2]。
- 采用英特尔® 傲腾™ 数据中心级持久内存后，快手单 Redis 实例的内存容量获得了一倍以上的提升，其 Redis 服务的 TCO 也降低了 30%[3]。

为超大规模推荐系统重构存储系统

互联网短视频的持续火热，使得更多民众可以利用短视频 APP 普遍参与到短视频的制作与传播中。作为国内短视频服务行业的佼佼者，快手每天都会迎来 2 亿活跃用户以及千万级的短视频上传量[4]。如何在海量用户以及海量短视频内容间构建起一个桥梁，既能让更多用户在屏幕上实时刷出符合自己喜好的作品，又能随时对内容进行点评，或“赞”或“踩”，无疑是快手在后台系统构建时关注的一大焦点。

针对用户希望获取内容实时推荐的需求，快手从诞生伊始，就一直持续、大量地投入于内容推荐系统的建设和技术更新。随着用户和短视频作品数量的不断增长，如何在承载高峰期每秒数十万并发调用量的同时，从上百亿级别的短视频库中，通过千亿参数级别的深度模型将合适的内容推荐给不同的用户，就成为了关键的课题。为此，快手结合最新技术趋势，基于异构设备构建了计算与存储分离的推荐系统架构。

如图一所示，快手在推荐系统中采用了计算与存储分离的架构，其由推荐服务、预估服务和召回服务等计算型服务，以及用户画像、参数服务器和分布式索引等存储型服务组成。前者承担了推荐策略计算、模型预估、视频检索等工作，而后者则为推荐系统中上亿规模的用户画像、数十亿规模的短视频特征以及千亿规模的排序模型参数提供存储和实时更新能力。



图一 计算与存储分离的快手推荐系统架构

众所周知，短视频的典型应用场景是碎片时间，当用户在快手 APP 上随意滑动浏览时，留给推荐系统的处理时长往往只有毫秒级。在计算模块提供高性能策略计算之余，如何让亿级规模的海量存储数据为推荐系统提供实时支持，无疑更富挑战。

因此，快手基于异构设备，针对不同应用场景，采用了多样化的技术实现方式。以分布式索引为例，要在大规模分布式存储集群中高速检索数据，离不开索引的力量。为提升高并发下的索引性能，快手采用了基于内存的键值（Key Value，KV）数据库来构建分布式索引系统。

而作为推荐系统的另一重要基石，快手 Redis 服务的性能表现也会对推荐效果产生显著影响。用户在短视频应用中的行为轨迹会被存储在 Redis 数据库中，并最终构成对用户的精准画像。而 Redis 实例能使用的内存容量越大，就意味着可被存储在内存之中、能实现高速读取的信息越多，用户画像也就更为具体，对用户的内容推荐也会更为准确。

不仅如此，Redis 服务还为快手短视频提供了强有力的社交服务能力的支撑，例如点赞、发表评论以及弹屏等功能。Redis 数据库基于内存的设计，保证了这些社交行为能流畅运作，以输出良好的用户体验。

然而，随着数据量的高速增长，快手基于内存的推荐系统存储服务和 Redis 服务，正在经受越来越大的挑战。一方面受限于物理服务器 DRAM 内存容量规格的限制，各服务实例的内存容量始终难以大规模扩展；而另一方面，DRAM 内存昂贵的价格，也造成快手 TCO 的急剧提升。同时，DRAM 内存易失性的特点，也使系统在故障恢复时需要耗费更多的时间。

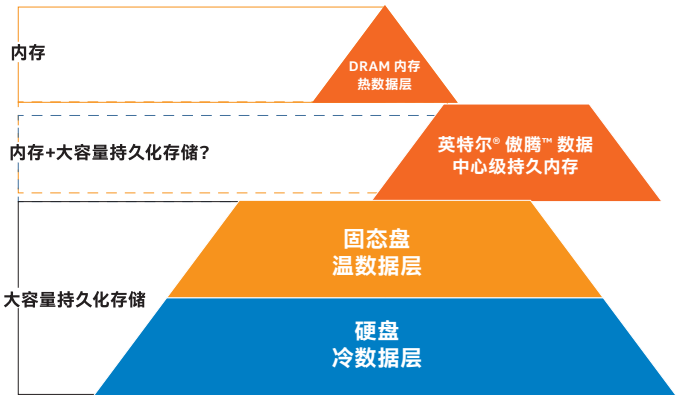
为了破解上述挑战，向用户持续提供更优质的内容推荐服务，快手在利用异构混合计算方案有效提升计算型服务性能之余，也通过与英特尔开展的深入技术合作，引入英特尔® 傲腾™ 数据中心级持久内存产品，对推荐存储系统和 Redis 数据库实施了优化改造。

软硬相济 实现更强存储能力

在传统存储架构中，大容量持久化存储主要由硬盘（Hard Disk Drive，HDD）或固态盘（Solid State Drive，SSD）来承担，而随着

数据应用场景的日益多样化，以及对存储性能提出的更高要求，存储需求层次也变得越来越复杂。采用更多的 DRAM 内存，无疑可以获得更强的性能，但也会带来更高的成本。为此，快手选择全新的异构存储结构，以求在性能、容量和成本三个维度上实现优选。

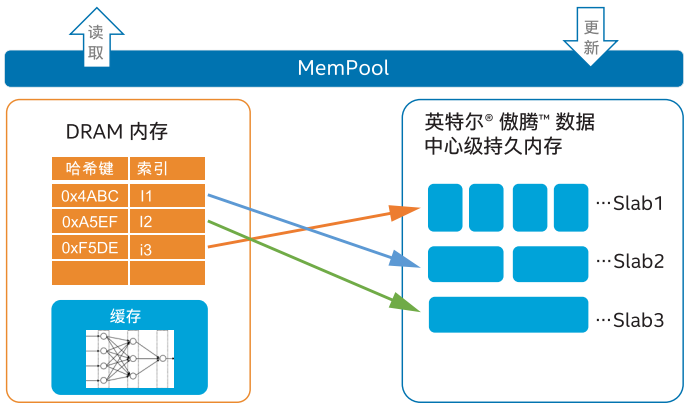
在快手原本的设计中，高性能的 DRAM 内存将承担性能要求最高，但容量要求最小的那部分存储需求，而性能要求较低，容量要求高，且要求持久化的存储任务则交给固态盘和硬盘。但与此同时，快手仍需面对另一种可能的场景，即当存储系统对性能、容量和持久化均有着较高要求时，该如何应对？



图二 英特尔® 傲腾™ 数据中心级持久内存是兼顾内存级性能和大容量持久化存储能力的好选择

如图二所示，基于 3D XPoint™ 存储介质打造的英特尔® 傲腾™ 数据中心级持久内存，显然是快手补上这一环的理想之选。这一创新的内存产品类别，不仅拥有与 DRAM 内存相近的读写性能、访问时延和相比 SSD 更强的耐用性，可在高并发的推荐系统场景中，实现不亚于 DRAM 内存的性能表现；它还可凭借大容量的特性，帮助快手轻松构建起 TB 级的内存数据库。更难得的是，它拥有 DRAM 内存不具备的数据持久性（或称非易失性，基于 App Direct 模式），可为快手的推荐异构存储系统带来更高的可用性。

为了让由 DRAM 内存、傲腾™ 数据中心级持久内存、固态盘以及硬盘组成的异构存储系统发挥更大效能，快手与英特尔等合作伙伴一起，针对推荐系统中的不同场景进行了大量可行性分析和架构设计调研，并根据英特尔® 傲腾™ 数据中心级持久内存的特性，针对分布式索引和参数服务器中的 KV 存储进行了重新设计。



图三 引入英特尔® 傲腾™ 数据中心级持久内存构建的异构存储系统

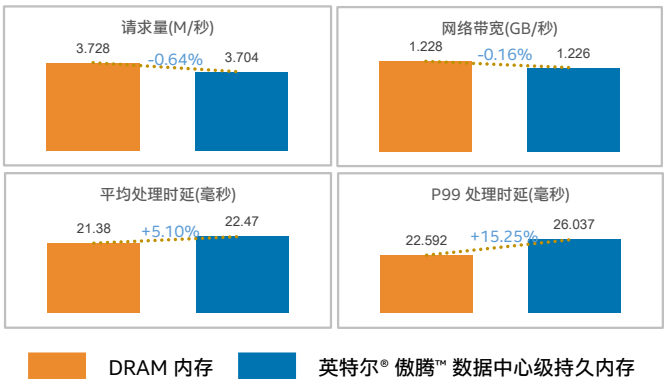
新的设计如图三所示，架构中新增了 MemPool 组件。作为一种缓冲池，该组件能让系统针对不同的访问类型来决定是使用 DRAM 内存，还是使用英特尔® 傲腾™ 数据中心级持久内存；例如在使用参数服务器进行推荐模型预估时，由于模型中的神经网络大小与嵌入表（Embedding Table）相比较小，因此神经网络可被 MemPool 分配进入 DRAM 内存来提高预估的性能。而在分布式索引使用场景中，系统可根据所需索引数据的大小，为其在英特尔® 傲腾™ 数据中心级持久内存中分配不同的 slab（一种内存分配机制），从而提升存取性能和效率。

在这些主要的设计之外，快手还针对英特尔® 傲腾™ 数据中心级持久内存的特性，实施了多种调优方案。首先在数据读取场景中，通过采用非一致内存访问（Non-Uniform Memory Access Architecture，NUMA）节点绑定的方式，避免英特尔® 傲腾™ 数据中心级持久内存在进行数据存取时，在不同的 NUMA 节点间切换，以实现更好的读写性能；同时，无锁和零拷贝（Zero Copy）技术的加入，也避免了临界区对英特尔® 傲腾™ 数据中心级持久内存的频繁访问，以及降低数据访问对内存带宽的占用，从而提升存储系统的整体效能。与此同时，英特尔® 傲腾™ 数据中心级持久内存的非易失性特性，也能让新设计的索引系统获得分钟级别的故障恢复速度，与之前小时级别的恢复速度相比，提升达百倍之多[1]。

而在 Redis 服务中，英特尔® 傲腾™ 数据中心级持久内存大容量的特性，令快手单 Redis 服务器内存容量达到了 TB 级，使单 Redis 实例的内存容量从 4GB 扩展到 8GB，实例的内存容量得以翻番，为业务的进一步发展提供了更强的硬件基石。

满足性能需求的同时降低 TCO

为验证全新的快手异构存储结构在采用英特尔® 傲腾™ 数据中心级持久内存，并实施了一系列软件调优后的实际性能表现，快手与英特尔一起，使用真实线上请求数据，对采用了英特尔® 傲腾™ 数据中心级持久内存的各个相关系统，如推荐系统中所用到的索引系统，进行了一系列的模拟压力测试。



图四 基于英特尔® 傲腾™ 数据中心级持久内存的索引系统压力测试结果

测试围绕着推荐系统中核心的请求量、网络带宽、平均处理时延以及 P99 处理时延四大核心性能指标展开。测试结果如图四所示，可以看出，在这四个核心指标上，英特尔® 傲腾™ 数据中心级持久内存均获得了与 DRAM 内存相近的性能表现，尤其是在网络带宽指标上，双方差值仅为 0.16%[5]。

在实现相近性能的同时，傲腾™ 数据中心级持久内存大容量和非易失性的特性，以及相比 DRAM 内存更为实惠的价格，让快手的成本得到了有效控制。来自快手的估算显示，通过引入英特尔® 傲腾™ 数据中心级持久内存，快手推荐系统以及 Redis 服务的 TCO 均获得了 30% 的降低[2,3]。

结语

作为国内率先将英特尔® 傲腾™ 数据中心级持久内存引入推荐系统的互联网企业，快手以其卓越的技术创新能力，为异构存储结构在推荐系统中的构建和运用，以及大容量 Redis 服务在短视频服务中的运用，进行了有意义的探索，并获得了丰硕的成果。

着眼未来，快手正与英特尔探讨成立联合实验室，以借助英特尔各种创新产品和技术，来驱动自身业务的创新及其数据中心的升级和演进。上述英特尔® 傲腾™ 数据中心级持久内存的应用，其实就是双方在这个联合实验室筹备过程中推进的第一个项目。今后快手还将继续携手英特尔，挖掘英特尔® 傲腾™ 数据中心级持久内存在旗下其他业务场景或服务中的应用价值，推动对各类数据处理和存储系统实施优化和革新。

[1] 数据援引自: https://36kr.com/p/5232799
[2] 数据援引自: https://36kr.com/p/5232799
[3] 成本结果引自快手内部测算，如欲了解更多详情，请联系快手
[4] 数据援引自: https://36kr.com/p/5232799
[5] 测试结果援引自快手内部评测，如欲了解更多详情，请联系快手