解决方案简介

英特尔® 傲腾™ 数据中心级持久内存
中国联通沃云

英特尔® 傲腾™ 数据中心级持久内存优化中国联通沃云数据中心资源池应用承载能力

英特尔® 傲腾™ 数据中心级持久内存，是优化中国联通大型云数据中心资源池、提高“沃云”应用承载能力和扩展能力以及探索技术演进的重要力量。

中国联通“沃云”数据中心的承载能力和扩展性常受限于单节点服务器配置，归根结底受限于大规模的采购成本压力，更大内存容量需求和 IO 的限制影响了“沃云”系统的承载能力。通过英特尔第一代数据中心级持久化内存对沃云数据中心节点进行优化，通过以较低采购成本换取更大的内存容量，可以比较充分地提高服务器的有效利用率，提升数据中心单位空间的虚拟机密度。在提高虚拟机数的同时虚拟机内的典型应用如 Redis，MySQL 等性能也能得到良好的展现。

陈硕
OpenStack 研发工程师
中国联通云数据有限公司
英特尔-沃云联合创新实验室成员

快速发展的中国联通云计算平台

作为电信运营商，中国联通“沃云”致力于提供完整的端到端云计算服务平台解决方案，帮助政企客户加快业务转型。2013 年 12 月 12 日在京召开的 2013 云世界大会上，中国联通正式发布旗下云计算业务品牌“沃云”，它是中国联通自主研发的面向企业和政府用户的云计算服务平台，至今已发展到了 5.0 版本，完成了可信云、安全等三级认证，并在电子政务、环保、医疗、教育、金融、旅游等各行业以及联通集团内部、部分省分的私有云得到广泛应用[1]。沃云以多层次的数据中心建设标准，在全国布局，已建设超过十个云数据中心资源池，规模达到 25 万核 CPU，20PB 存储，总带宽 240G，并仍在快速发展[2]。

在快速发展过程中，沃云研发人员不断面对挑战，通过发挥中国联通自身通信网络和数据网络基础设施优势，和提供“沃云”强大的云计算平台支撑能力，满足不同行业客户应用对云计算资源池的不同服务要求和业务增长需求，同时强化资源配置，有效降低采购成本和运维成本，自主可控，按需定制和可持续发展，体现了中国联通“沃云”产品区别于其他云计算解决方案的综合技术实力和优势。

“沃云”平台持续的技术演进

云计算技术一直在快速的技术演进和发展中，“沃云”在新技术的演进中仍需面对大量已建云资源池做维护与发展，它们同样需要技术演进，兼顾老版本和新技术，渐行渐进。通过不断地通过技术、市场竞争和实际行业用户项目的碰撞、总结和体会，我们认识到，“沃云”只有不断紧跟用户需求，采用领先的技术和产品，不断优化和创新，同时发挥中国联通的自身的端、管、云优势地位，才能成就具有自身特色和优势的优秀产品。

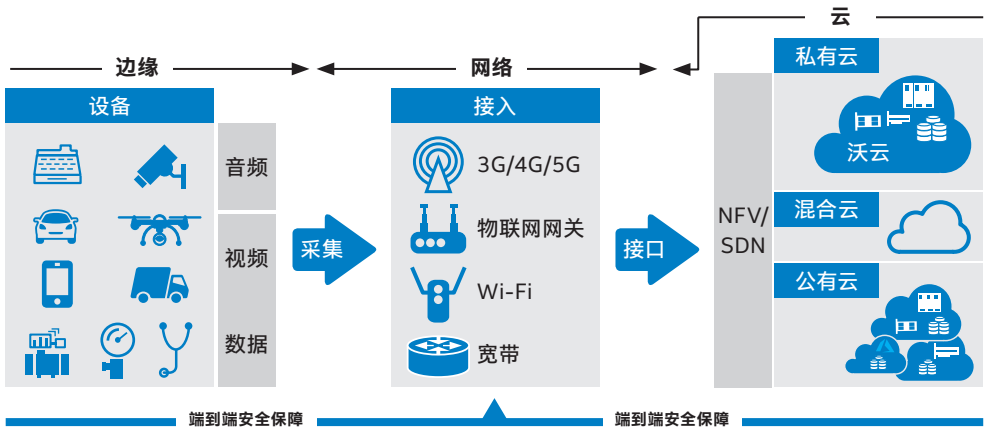

图 1 中国联通“沃云”架构图

“沃云”要有好的承载能力、同时支持应用负载的多样性、提高有限资源的利用率、有持续的技术演进和发展，需要从云计算平台基础设施优化做起，通过结合硬件加速技术、优化硬件资源的配比得到坚实的硬件支撑，同时做好软件虚拟化层和网络层的“沃云”软件优化、内存和存储弹性扩展的优化、云资源池中的资源调度与监控优化，来实现 “沃云”的整体进阶。

基于这一目标，中国联通联合英特尔公司成立英特尔-沃云联合创新实验室，并就最新一代的英特尔® 傲腾™ 数据中心级持久内存开展云计算平台上应用的探索、研究和各种努力，希望“沃云”产品更加贴近用户实际应用需求。

“沃云”承载能力以及内存多样性需求

服务器本身是有限资源，同时它的配置扩展性受到数据中心的机位空间、供电、网络接口和带宽、和服务器本身设计的限制。除建设改造数据中心这种大资金投入外，数据中心本身的空间、承重、配电、网络就是一个有限资源，一旦建设完成，就已基本确立了数据中心总体承载能力。提高在单位面积下的计算能力，同时做到降低投入成本，平衡好性能、能耗、空间占用、投资对于“沃云”资源池的建设和管理变得尤为重要。

行业应用类型的多样性也对“沃云”承载能力不断提出挑战，例如一台宿主机服务器最多能够提供多少虚拟机运行并同时有合理的性能保障？这个问题取决于服务器设备配置的投入 (CPU 型号，内存大小，网络带宽，存储性能等)，不同虚机应用对服务器资源的需求不同，和“沃云”宿主机、虚拟机平台软件系统的性能优化深度和广度等因素紧密相关。

大多用户需要高性价比的大内存服务，在提供合理性能的同时尽可能降低成本；有些则需要低延时和超大内存，以期待更多更快的数据处理响应；还有些则需要非易失的内存，应用和数据能够保存在内存中，应用中断能即时重启，希望大型应用系统尽可能快地恢复运行；而另一些应用则需要超快的存储，满足低延时应用存储数据的需要。虚拟化平台、数据清洗平台、流数据处理平台、大数据分析平台、关系型数据库平台，内存数据库平台等等，对 CPU 处理器核数、单核性能、内存容量和数据存储 IO 性能要求都不尽相同，既要满足内存多样性要求，又要简化数据中心服务器典配类型数量，看似矛盾，但又必须面对，是数据中心云资源池基础建设规划中必须考虑的棘手问题之一。

内存扩展是提高“沃云”承载能力和优化资源池的一个关键技术点。通过内存扩展优化做到内存容量、性能、成本的最佳平衡，从而在同样成本和合理性能保证下，创建出更多的虚拟机数。

英特尔® 傲腾™ 数据中心级持久内存是英特尔推出的最新一代 DCPMM 产品，它既具有接近于内存的数据读写性能，又具有固态硬盘的数据非易失性特性和超长擦写寿命，提供了行业领先的高吞吐率、低延时、高服务质量和超高的耐用性，可以对系统吞吐以及应用进行加速，在性能上表现卓越[3]。

在云化服务中虚拟化场景中的内存模型的改变可以为每个云计算节点多增加虚拟机数，带来额外的收入。通过虚拟机优先级在内存和存储级内存中的分类优化，宿主机节点中大量的低优先级虚拟机将不会影响高优先级虚拟机的内存使用性能。

DCPMM 应用提高“沃云”系统承载能力

在英特尔® 傲腾™ 数据中心级持久内存的开发过程中，中国联通基于联合创新实验室与英特尔进行合作，采用新的内存架构，优化其沃云平台的承载能力。该工作包括成本与配置分析，性能分析，设计测试场景，在虚拟化环境下，通过最常用的基准测试工具，对不同压力下的内存读写的带宽、延迟时间 (平均、最高、最低) 测试数据加以细致分析，来观察性能的变化，理解内存扩展技术的机理和合理配置。它还需要开发针对硬件和软件的基准测试技术，以准确灵活地测量配备英特尔® 傲腾™ 数据中心级持久内存的系统的关键绩效指标。

通过引入 DCPMM 实现内存扩展，可以实现在相同的价格下获得比单纯 DRAM 内存更大的内存容量，使“沃云”平台有更多的承载应用数据的空间，并使虚拟机有更均衡的性能发挥。这对大规模的云数据中心来说是非常有利的，可以满足很大一部分需要超大内存容量的应用或服务的需求。

除了工作在内存模式其相对于 DRAM 存储器的成本优势之外，英特尔® 傲腾™ 数据中心级持久内存还可以工作在 APP Direct 模式，这一应用还带来了内存子系统架构的深刻变化，使其成为工作数据和长期存储的主数据层。它可以将类似于 DRAM 存储器的字节寻址能力和类似于存储的持久性合二为一。这种结合意味着它可以直接映射到应用程序地址空间，消除了与传统存储的读写相关的瓶颈。这两种工作模式可以根据实际应用场景进行切换和配置，给实际部署带来极大便利。

总体来说，通过在中国联通“沃云”部署英特尔® 傲腾™ 数据中心级持久内存，对大数据、流式、实时、IMDB 等分布式数据处理分析平台的内存模型做改变，在节点的 IO 性能、内存容量和资金投入之间找到平衡点。

成本可匹配系统配置

在大型数据中心中，要采纳一项革新技术，它的使用成本和带来的收益往往是至关重要的。我们在测试过程中首先重点关注对“沃云”服务器节点配置成本:

“沃云”中一台典型配置的计算节点服务器英特尔® 至强® 金牌 6148 处理器共有 80 vCPU，vCPU 与内存之比按 1:4 计，至少需要配置 320GB 内存，若按 1: 8 计则至少需要配置 640GB 内存，加上宿主机系统所需内存和预留冗余内存，一台两路英特尔® 至强® 金牌 6148 处理器内存的合理配置在 384GB 和 768GB 之间较为合理。CPU 的核数越多，作为虚拟化计算节点，内存容量的需求就可能越大。有些作为大内存节点服务器例如 IMDB 数据库等，内存容量需求就更大 (甚至 1TB 以上)。

根据英特尔推荐的可匹配成本价格配置，384GB DDR4 的价格与 192GB DDR4 + 4×128GB DCPMM 成本价格相当，但是 DCPMM 作为内存模式时可见内存容量为 512GB，比 DDR4 系统内存容量增加了 30%。

综上所述，引入 DCPMM 的内存扩展方案，可以实现在相同的价格下获得比单纯 DRAM 内存更大的内存容量，提升“沃云”

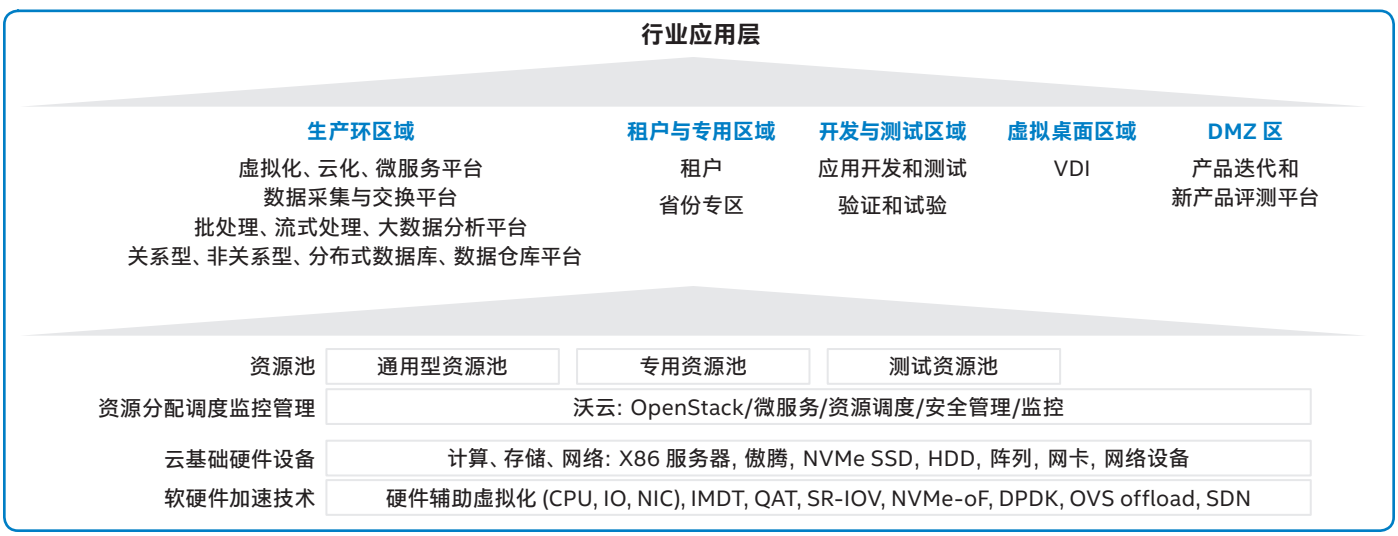

图 2 “沃云”资源池架构图

平台的承载应用数据的空间，并使虚拟机有更均衡的性能发挥，可以在很大程度上满足大规模云数据中心对超大内存容量的应用或服务的需求。

	基准配置 (DRAM)	DCPMM 配置 (内存模式)
CPU	2 x 英特尔® 至强® 金牌 6148 处理器 @2.30GHz 20 核心	2 x 英特尔® 至强® 金牌 6230N 处理器 @ 2.30GHz 20 核心
内存	DDR4 2666 384GB (24 x 16GB)	DDR4 2666 192GB (12 x 16GB) + 4 x DCPMM 128GB
硬盘	1x 2TB NVMe SSD + 1 x 960GB SATA SSD	1 x 2TB NVMe SSD + 1 x 960GB SATA SSD
OS	CentOS Linux release 7.6.1810/4.20 GNU/Linux	CentOS Linux release 7.6.1810/4.20 GNU/Linux
测试对象	• Stream Triad • Redis • MySQL	• Stream Triad • Redis • MySQL
测试负载	Memtier_benchmark, Sysbench	Memtier_benchmark, Sysbench

表 1 测试服务器配置对比

基准内存带宽测试

中国联通“沃云”服务器上物理机上的 stream Triad 基准内存带宽性能压力测试数据如图 3 所示。数据显示，在使用英特尔® 傲腾™ 数据中心持久内存的情况下，物理机上基本达到跟 DDR4 相当的内存带宽性能，在实际生产过程中符合高性能的要求。

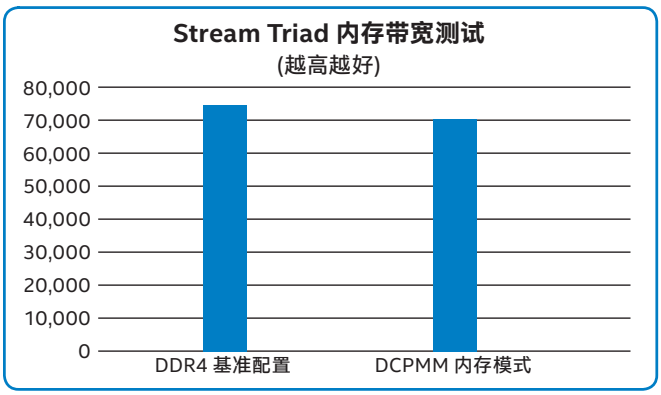

图 3 stream Triad 基准带宽性能测试 (单 socket)

虚拟机 Redis 数据库性能测试

在虚拟化性能测试中，通过服务器启动多虚机进行系统测试整机性能的最大吞吐。每个虚机中启动单实例 Redis，采用 memtier_benchmark 测试虚机中的应用性能并通过 TPS 评价。

按照虚机配置，每台虚机配置 vCPU 为 4 核心，最大内存为 32GB，实际消耗内存为 24GB。根据实际内存最大 90% 进行估算，DDR4 基准配置可以启动约 14 个虚拟机，DCPMM 内存模式可以启动 19 个虚拟机。

memtier_benchmark 是 Redis Labs 推出的一款命令行工具，它能够产生各种各样的流量模式，可以对 Memcached 和 Redis 实例进行基准测试。这个工具提供了丰富的自定义选项和报表功能，通过命令行界面就能够轻松地使用。测试键访问模式符合常见的高斯分布的钟型曲线，根据 Redis 内存数据实际应用的场景，读写比例设置为 4: 1。

	Redis Benchmark 虚拟机配置
vCPU	4
内存	24GB
虚机数量	DDR4 基准配置 14个 / DCPMM 内存模式 19 个
测试 workload	Redis 4.0, Data Size 1024B, RW ratio 4R/1W
测试负载	memtier_benchmark --ratio=1:4 -d 1024 -n 26000000--key-pattern=G:G --key-minimum=1 --key-maximum=26000001 --threads=1 --pipeline=64 -c 3 --hide-histogram -s <server ip> -p <port>

表 2 Redis 测试虚拟机配置

实际测试结果显示 (测试数据如图 4 显示)，在保证服务质量的前提下，尽管单虚机的性能 DCPMM 内存模式会略低于 DDR4 基准配置的性能，但是整体整机的吞吐在多虚机助力的情况下，系统整体吞吐性能比 DDR4 基准配置还高 10% 以上。

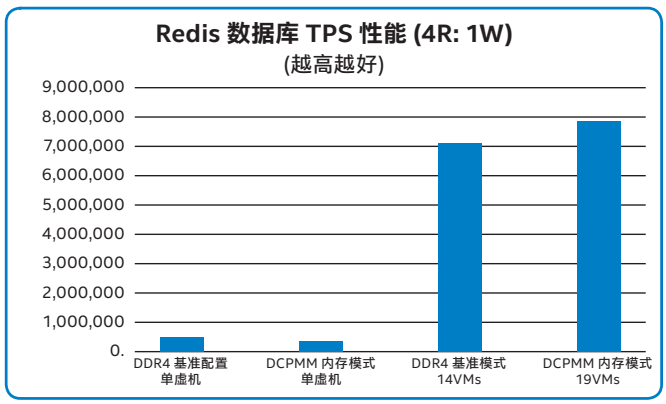

图 4 Redis 内存数据库性能测试

虚拟机 MySQL 数据库性能测试

在 MySQL 数据库的性能测试对比中，工作人员使用英特尔® 傲腾™ 数据中心持久内存与物理内存在大缓存页场景下，观察

MySQL 数据库吞吐能力 (测试数据如图 5 显示), 鉴于 MySQL 吞吐能力依赖于 IO 读写性能, 通过增大缓存页可以进一步提高 MySQL 数据库的实际吞吐, 而这正依赖于实际内存的大小。在相同配置场景下, DDR4基准配置在运行15个虚拟机的场景下, 已经基本将系统内存消耗完毕, 这基本上达到了系统吞吐的极限。而在 DCPMM 内存模式配置下, 由于系统内存尚有余量, 可以进一步增开虚机, 从而将系统整体吞吐进一步提高约 20%。

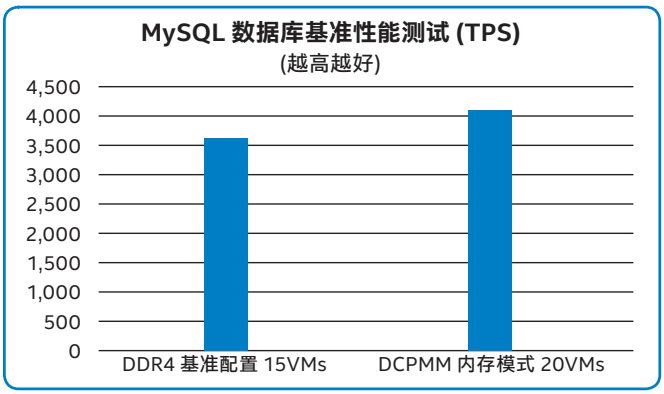

图 5 MySQL 数据库性能测试

	MySQL 数据库虚拟机配置
vCPU	4
内存	32GB
虚机数量	DDR4 基准配置 15个 / DCPMM 内存模式 20 个
测试 workload	MySQL 5.5.56
测试负载	SysBench 1.1.0

表 3 MySQL 测试虚拟机配置

效果: 面向未来的云计算

云计算带来了人与人、人与物、物与物的广泛连接, 5G 技术进一步推动了物联网、边缘计算等概念走向落地, 对于中国联通云计算平台而言, 各种业务都在快速往云计算平台迁移, 计算量大幅增长, 这意味着对“沃云”计算平台资源池是一个巨大的挑战, 同时也是一个重要的商业契机。面向正在高速发展的云计算平台, 中国联通希望在云计算技术上做业界的领跑者, 同时面向 5G 时代对云计算平台进行持续的技术创新。

通过搭载英特尔® 傲腾™ 数据中心级持久内存, 中国联通得以在快速增长的云计算业务中, 在控制“沃云”服务器成本的前提下, 快速提升平台资源池。同时能够无缝迁移各种云计算平台上的各种业务需求, 解决了业务爆发式发展与 IT 基础设施承载能力不足的矛盾。

由于中国联通“沃云”服务器性能的提升, 计算资源得到了保证, 可以为客户提供全方位的、及时的、友好的服务能力和服务渠道, 保证良好的客户体验。同时, 中国联通“沃云”服务器可以有效支撑大数据、流式、实时、IMDB 等分布式数据处理分析平台的各类高吞吐业务的及时上线, 通过系统高可用以及保障系统高性能, 始终让沃云在竞争激烈的市场中处于有利地位。

展望未来, 中国联通与英特尔携手合作, 在各种新技术领域联合创新, 以进一步提高沃云计算平台的性能优势。

结论

随着 5G 时代的加速到来, 人工智能、物联网和增强型云服务等因素提高了电信网络上的数据量及其多样性, 特别是随着云业务的持续高速增长, 电信运营商业务也在面临严峻的考验。

英特尔® 傲腾™ 数据中心级持久内存 (DCPMM) 具有大容量、低成本、和持久性存储的特点, 能够为大数据分析、内存数据库等应用带来巨大的性能提升, 同时能够为用户降低 IT 成本, 简化基础设施, 英特尔® 傲腾™ 数据中心级持久内存 (DCPMM) 正在成为构建新一代数据中心和数据分析平台的最佳解决方案。

我们看到在中国联通“沃云”数据中心服务器中采用英特尔® 傲腾™数据中心级持久内存 (DCPMM) 在大幅节省部署成本的同时, 也获得了巨大的内存容量提升。虽然数据中心持久化内存的延迟比 DRAM 有所增加, 但是我们可以看到在实际应用中由于内存容量的扩大, 实际整机的总体系统整体吞吐性能仍然有一定提升, 可以很好地解决计云计算平台中大内存和高性能吞吐场景中的系统瓶颈, 并且能显著改善由于虚机数增加或虚拟机大内存需求造成的物理内存不够引发的虚拟机计算性能陡降问题, 并能够提供相同的 SLA (Service Level Agreement) 保证。

DCPMM 作为一种新型的硬件解决技术方案, 通过未来更多的应用场景优化及业务整合方案的完善, 具有成为一种高性价比的系统解决方案的可能。

此外, 我们也非常期待下一代平台的的可持久化内存 Barlow Pass 的出现, 它可以提供更高的带宽, 更好的延迟, 更能充分发挥 3D Xpoint™ 高速性能, 给用户带来更好的应用体验。

采取下一步行动

了解更多关于英特尔® 傲腾™ 数据中心级持久内存的信息:

www.intel.com/optanedcpersistentmemory

了解更多关于英特尔® 至强® 处理器的信息, 请访问:

www.intel.com/xeon

联系中国联通沃云销售人员或注册免费试用

孙意欣 sunyx2@chinaunicom.cn

https://www.wocloud.cn/

作者

陈硕
联通云数据有限公司 Openstack 研发工程师, 英特尔-沃云联合创新实验室成员

刘中
联通云数据有限公司研发总监, 英特尔-沃云联合创新实验室成员

郑春阳
英特尔数据中心部门资深平台应用工程师

胡自强
英特尔行业技术专家, 英特尔-沃云联合创新实验室成员

[1] 中国政务云发展白皮书 (2018 年) 云计算开源产业联盟

[2] http://www.woclouddata.cn/zhuzhan/aboutus/index.html

[3] Intel Launches Optane DIMMs Up To 512GB: Apache Pass Is Here! https://www.anandtech.com/show/12828/intel-launches-optane-dimms-up-to-512gb-apache-pass-is-here