

金山云助力实现云端数据分析

金山云携手英特尔研发并推出新服务，在中国市场实现差异化云存储业务。

概览:

- 金山云携手英特尔推出金山云 KMR 产品 (Kingsoft Cloud MapReduce) *。
- 英特尔® 至强® 处理器 E5-2690 v4 提供所需性能。
- 英特尔工程师协助进行性能优化。
- 英特尔还协助金山云对这一新服务进行营销推广。
- 金山云在医疗服务、游戏和电子商务领域赢得了多家新客户。

云服务提供商 (CSP) 想要赢得并留住客户，所面临的竞争异常激烈。要想发展壮大，云服务提供商就必须保证不断创新，帮助客户从数据中获得新的洞察，找到创造额外价值的方式，从而帮助客户推动各项业务的发展。

金山云作为中国领先的云服务提供商，提供图片、视频和销售日志等多种数据的云存储服务。客户可以利用金山云的云服务器、庞大的云存储空间、负载均衡和云关系型数据库，动态分配相关资源，以满足自身的业务需要。金山云的云存储服务可以确保客户业务的持续性，改善总体拥有成本 (TCO) 。

金山云的云存储服务规模现已超过 400 PB，每天增加高达 500 TB。金山云不断创新，与英特尔合作，探索、研发并推出了一项大数据分析和深度学习服务，可增强现有金山云存储平台的性能。凭借这一差异化服务，金山云在云服务市场上获得了竞争优势。

挑战

- 探索并推出新的云服务，推动业务差异化，实现进一步增长
- 实现数据中心基础设施现代化，以最适宜的成本提供新的服务
- 获取 Apache Hadoop* 和 Spark* 以及相关技术领域的专业知识和信息

解决方案

- 金山云和英特尔基于金山云的云存储平台和英特尔® 至强® 处理器 E5-2690 v4，联合推出了新的大数据和深度学习服务
- 英特尔工程师利用 Apache Hadoop 和相关技术，在金山云开发和优化新服务的过程中提供了重要帮助

结果

- 平台升级至英特尔® 至强® 处理器 E5-2690 v4，处理能力增强，可以更加合理地利用数据中心的资源
- 新平台在游戏、医疗服务和零售行业赢得了众多新客户

企业过去通常需要

1 到 3

个月才能完成
Apache Hadoop 的部署

现在，在采用了金山云
KMR* 后，只需

 几分钟

即可完成部署

在云存储服务中添加数据分析和深度学习

金山云和英特尔很早之前便建立了深入合作创新的伙伴关系。双方共同发现了一个新的市场机遇，可以利用英特尔® 至强® 处理器 E5 v4 家族，通过新的分析功能来增强金山云的对象存储服务* (KS3)。利用这一增强的服务，客户可从存储在 KS3 云端的数据中获得切实可行的洞察。

更重要的是，这种新服务的成本低，而性能不受影响。本文将向您介绍金山云和英特尔如何通过升级和优化英特尔® 至强® 处理器 E5 家族、集成英特尔® 固态硬盘和英特尔® 网卡等其他技术，携手实现了金山云数据中心的现代化，同时大幅提高了提供新服务的效率。

Kingsoft Cloud MapReduce (KMR) 简介

金山云携手英特尔共同开发并推出了大数据分析平台 Kingsoft Cloud MapReduce* (KMR)，这一平台能够帮助客户快速构建数据分析集群并处理海量数据。KMR 虽然基于 Apache Hadoop，但是也可以通过 Apache Storm* 和 Apache Kafka* 用于分析实时数据。KS3 基于对象存储，可通过 KMR 直接查询。因此，客户可以通过这种方式使用 Apache Hadoop，其成本低于大数据分析通常采用的块存储方式。

KMR 与 KS3 结合使用，让客户能够在云端分析现有的数据。KMR 是 KS3 的增强服务，客户借此可以从现有的数据和平台中获得更多价值。

解决方案的技术组件

- 英特尔® 至强® 处理器 E5-2690 v4。英特尔® 至强® 处理器已安装在本地原始设备制造商生产的服务器上，可满足金山云客户大数据分析解决方案对性能的需求。
- 英特尔® 固态硬盘 (英特尔® SSD) DC S3510 系列。金山云在其整个云解决方案产品组合中均采用英特尔® 固态硬盘，以实现低延迟存储。
- 英特尔® 以太网融合网络适配器 X520-SR2。金山云数据中心内部高速通信均采用英特尔网卡，可加快 KMR 等云解决方案的处理速度。

“我们与英特尔合作，学到了他们与其他全球云服务提供商的合作经验。这加快了我們学习的速度，让我们能够快速推出新服务，并借此赢得新客户。”

——倪煜
金山云大数据高级架构师

解决方案架构: Kingsoft Cloud MapReduce

KMR 是一种托管集群服务，构建在数百个双路英特尔® 至强® 处理器 E5-2690 v4 服务器节点之上。所有处理器均采用 14 纳米处理器技术，最多可提供 14 个内核 (28 线程) 和 55MB 高速缓存 (LLC)。每台服务器拥有 12 个英特尔® 固态硬盘数据中心级 S3510 系列驱动器，可为处理器提供极大的数据吞吐量。每台服务器采用一个英特尔® 以太网融合网络适配器 X520-SR2，集群节点因而可以互相共享数据，这同时也保证了严苛的数据中心/云端环境中服务器的灵活性和可扩展性。另外，还采用英特尔® 事务性同步扩展 (英特尔® TSX) 对并行工作负载性能进行了优化。

KMR 基于 Apache Hadoop 和 Spark 等计算框架。客户可以使用金山云 KMR 基于 Web 的管理控制台创建集群，也可以为虚拟机配置内核、内存和磁盘空间。客户可以选择集群的节点数量，也可以在控制台中添加新的节点，逐渐对集群进行扩展。KMR 集成了 Ambari 和 Ganglia 等集群监控管理工具，可以通过基于 Web 的管理控制台配置和使用。除此之外，KMR 还与金山云 KS3、金山云表格数据库服务* 和金山云关系型数据库服务* (KRDS) 等云服务实现了集成，以提供端到端的大数据解决方案。这些数据也可以存储在本地的 Hadoop 分布式文件系统* (HDFS) 中。

由于实现了各项云服务的端到端集成，金山云得以在行业竞争中脱颖而出。例如，KMR 拥有金山云 KS3 的访问接口。KMR 集群中运行的 MapReduce 和 Apache Spark 可以直接将需要处理的数据传出 KS3，然后再将结果写回 KS3。各项云服务实现端到端集成之后，客户可以使用 KS3 确保原始数据和计算结果在集群释放之后永久存储在云端，KS3 这一服务的存储成本相对低廉，数据存储高度可靠。金山云按时间对 Apache Hadoop 集群进行收费，客户因此可以将他们的结果存入 KS3 (对象冷存储)，随时释放 Apache Hadoop 集群，从而节省相关成本。

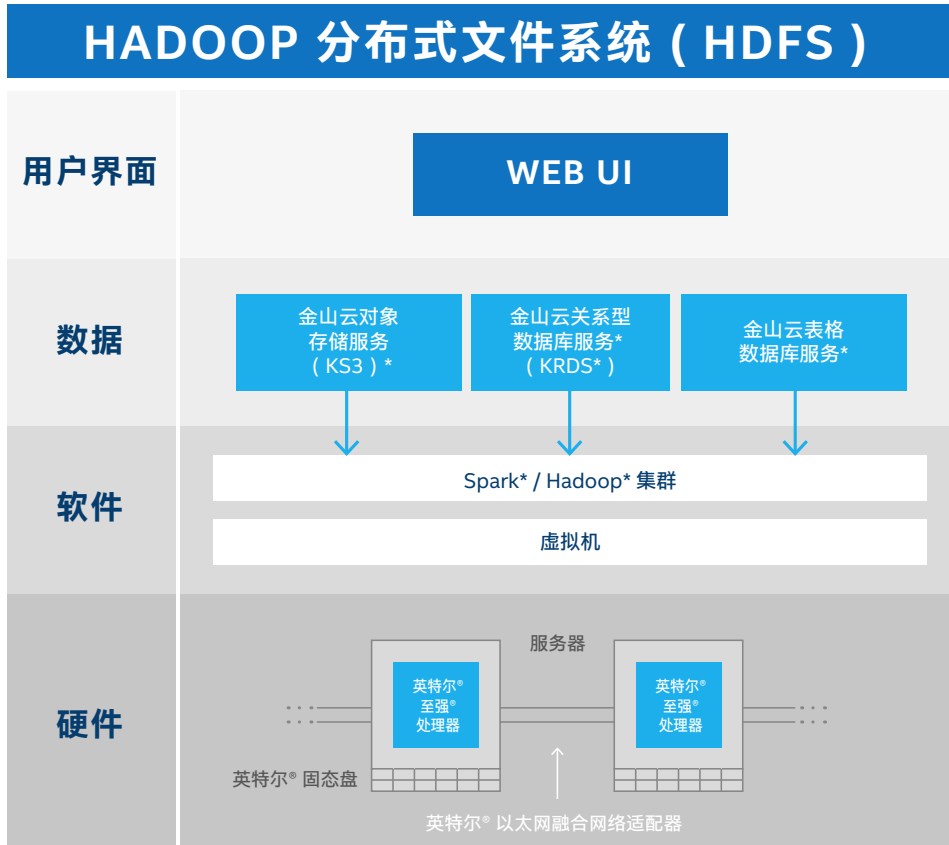


图 1: 金山云大数据分析集群的解决方案架构

实现云端分析

金山云与英特尔共同开发协作，推动其云存储服务成功转型为提供云分析服务。这项新服务帮助金山云吸引了众多新客户和新型工作负载，让金山云在市场竞争中获得了差异化优势。

例如，金山云凭借 KMR 赢得了医疗领域的一个新客户。这家医院有总计 280TB 的数据，临床文档库 (CDR) 包含十亿条患者诊断信息。KMR 部署完毕之后，用户可以在几毫秒之内找到他们需要的信息，也可以快速获得分析结果。

另一新客户运营着一个电子商务网站，产品需求受促销活动的影响非常明显，难以掌控。客户采用 KMR 之后，其电子商务网站与金山云* 相连。KKMR 可快速创建 Storm 和 Kafka 集群，构建实时数据处理系统，将数据处理结果写入 MongoDB 数据库。该客户现已具备数据分析实力，内置云也具有迅速的可扩展性。这一方式的总体拥有成本远低于构建企业自有解决方案，推出速度也远快于后者。

某网络游戏公司的日志文件大小总计有数百 GB。以前，这些文件存储在一个 MySQL 数据库中，并且需要定期查询以生成报

告和关键绩效指标 (KPI)。这种方式的数据存储成本很高，系统的可用性也得不到保障。此外，整合不同的数据集必须由人工操作才能完成。该公司采用 KMR 统一存储和处理日志数据之后，KMR 可以快速搭建集群环境，并根据游戏玩家的数量进行扩展。数据可以存储在 KS3 中，通过 KMR 进行查询，这样能够节省人力，提高效率，大幅降低存储成本。

英特尔工程师的长期合作

金山云成立于 2012 年，此后英特尔便一直与该公司合作，制定解决方案的详细规范，对计算、存储和网络领域的所有设备和平台进行基准测试和优化。双方携手推出了一种软件定义基础设施 (SDI)，对金山云数据中心的基础设施进行了优化和现代化。

金山云大数据高级架构师倪煜表示：“英特尔与多家全球最大的云服务企业合作，他们的工程师非常了解数据分析和云技术。英特尔帮助我们发现了可以涉足的服务领域，与我们一同设计了这一平台。”

金山云已经在英特尔® 架构上实现了标准化，所用处理器已从英特尔® 至强® 处理器 E5-2630 v3 升级至英特尔® 至强® 处理器

E5-2690 v4, 后者是中国市场目前最新的处理器产品。英特尔工程师从一开始就在金山云的数据中心和英特尔实验室中使用实际的基础设施同步进行基准测试工作。英特尔与金山云一同分析了总体拥有成本, 确认数据中心现代化可以为金山云带来可观的利润。

英特尔工程师与金山云一同优化了该公司的网络, 采用了开源数据平面开发套件* (DPDK), 帮助改善虚拟开关和防火墙等网络功能的网络处理性能。

英特尔利用英特尔® 数据分析加速库 (英特尔® DAAL) 帮助金山云优化深度学习工作负载的性能, 同时还利用 Apache Spark 的分布式深度学习库 BigDL 来加快处理速度。

英特尔将继续与金山云合作, 探索更大的优化空间, 发现其他新的服务机会, 帮助金山云进一步打造差异化云业务。金山云目前正在部署英特尔® 至强® 可扩展处理器, 同时正在评估英特尔的其他技术, 希望优化机器学习工作负载的性能。

英特尔不仅为金山云提供了工程设计支持, 还帮助金山云推出新的数据分析服务并进行营销推广, 在英特尔网站上和新闻稿中发布相关文章, 与金山云一起参与相关会议。

倪煜表示: “在搭建 KMR 的过程中, 英特尔不仅提供强大稳定的硬件平台, 还提供了 BigDL 和 Spark 的其他软件优化支持。对于改善 KMR 在数据分析和机器学习应用方面的性能而言, 这些都至关重要。”

金山云还与英特尔合作开发并推出了金山云深度学习平台* (KDL), 帮助客户从数据集中获得更深刻的洞察。这一系统可以提供人工智能云服务 (AlaaS), 帮助客户使用应用程序编程接口 (API) 进行深度学习训练和推理; 也可以用于图像分析、图像识别、视频识别和语音识别等应用场景。KDL 以容器为构建基础, 支持 TensorFlow、Caffe 和 MXNet 深度学习框架。客户无需搭建深度学习环境, 就能专注于构建深度学习模型和运行训练及推理工作负载。所有数据都可以存储 KS3 中。英特尔帮助金山云优化了英特尔® 至强® 处理器上的深度学习性能。金山云利用英特尔® 至强® 可扩展平台大幅增强 KDL 的性能, 与上一代英特尔® 至强® 处理器相比, 性能提升了一倍。

英特尔助力金山云实现业务增长

金山云与英特尔紧密合作, 现已对其数据中心基础设施进行了优化, 并打造了差异化云端分析服务。金山云凭借 KMR 现已在医疗服务、电子商务和游戏行业赢得了众多新客户。虽然行业竞争异常激烈, 但这一新服务预计将会帮助金山云留住现有客户, 在客户已有的可信数据平台上添加强大的分析功能。

经验总结

云服务提供商可从金山云的成功中获得以下经验:

- 客户之所以选择云端大数据处理, 是因为这种解决方案灵活、高效、便于部署和易于扩展。事实证明, 金山云的客户最青睐灵活的云服务
- 云服务提供商可从与英特尔的合作中获得工程设计专业知识, 包括 Apache Hadoop* 和 Spark*, 以及云计算中广泛使用的其他技术
- 升级至新一代英特尔® 至强® 处理器可显著提升吞吐量, 大幅增加服务器上虚拟机的数量。这样可以降低总体拥有成本, 帮助云服务提供商从数据中心资产中获得更多价值

金山云简介

金山云成立于 2012 年, 目前是中国领先的云服务提供商之一。其业务范围包含云服务器、大型云存储、负载均衡和云关系型数据库, 可帮助客户实现动态资源分配, 解决严峻的业务挑战, 提高业务稳定性。金山云总体存储规模现已超过 400 PB, 数据量每天增加多达 500 TB。

www.ksyun.com

寻找适合贵公司的解决方案。请联系您的英特尔代表或访问 intel.cn/CSP。

了解更多信息

- 英特尔® 至强® 处理器 E5 家族
- 英特尔® 固态硬盘 DC S3510 系列
- 英特尔® 以太网融合网络适配器 X520-SR2
- 数据平面开发套件 (DPDK)
- 英特尔® 数据分析加速库 (英特尔® DAAL)
- BigDL



性能测试中使用的软件和工作负载可能仅在英特尔微处理器上进行了性能优化。诸如 SYSmark 和 MobileMark 等测试均系基于特定计算机系统、硬件、软件、操作系统及功能。上述任何要素的变动都有可能测试导致测试结果的变化。请参考其他信息及性能测试 (包括结合其他产品使用时的运行性能) 以对目标产品进行全面评估。有关更多信息, 请访问 <https://www.intel.cn/content/www/cn/zh/benchmarks/benchmark.html>。

基准性能测试结果均在实施近期软件补丁和固件更新 (旨在解决称为“Spectre”和“Meltdown”所造成的漏洞) 之前完成。实施更新后, 这些结果可能不再适用于您的设备或系统。

此处提供的所有信息可在不通知的情况下随时发生变更。关于英特尔最新的产品规格和路线图, 请联系您的英特尔代表。

¹英特尔技术特性和优势取决于系统配置, 并可能需要支持的硬件、软件或服务得以激活。

产品性能会基于系统配置有所变化。没有计算机系统是绝对安全的。更多信息, 请见 <https://www.intel.com/content/www/cn/zh/products/processors/xeon.html>, 或从原始设备制造商或零售商处获得更多信息。

© 2018 英特尔公司版权所有。英特尔、英特尔标识、至强是英特尔公司在美国和/或其他国家的商标

* 其他的名称和品牌可能是其他所有者的资产。

0118/CAT/RW/PDF

337023-001EN