白皮书

英特尔和 Twitter 联合发布

数据中心
Hadoop 性能

利用缓存、快速固态盘和更高算力来提升 Hadoop* 性能和成本效率

通过试验和协作，Twitter 将其 Hadoop* 集群的内核密度提升了高达 6 倍，从而将总体拥有成本 (TCO) 降低了高达 30%，将运行速度加快了高达 50%[1]。

目录

执行概要

存储 I/O 可以说是 Hadoop* 集群的一大严重性能瓶颈，特别是 Twitter 这样的企业所拥有的超大型部署中，单个集群可能包含多达 10,000 个节点，还有将近 100 PB 的逻辑存储。Twitter 公司的典型 Hadoop 集群包含超过 100,000 个硬盘驱动器 (HDD)，但在这些硬盘驱动器的容量随着时间推移而增长的同时，其性能并未得到显著提升，因而这种配置逐渐达到 I/O 性能极限。[2] 因此，仅依靠添加更多容量更大的硬盘驱动器，已经无法解决 Twitter 的扩展难题，事实上，随着每 GB 数据的 I/O 降低，情况反而变得更糟糕。由于空间和电源限制，每节点添加更多磁盘不太可行。

于是，Twitter 的工程师与英特尔工程团队携手合作，开展了一系列试验，结果表明，将 YARN* (Yet Another Resource Negotiator*) 管理的临时文件存储在快速固态盘上，可在现有硬件上实现显著性能提升（运行时间缩短多达 50%）。[3] 该团队还发现，通过消除存储 I/O 瓶颈，他们可以使用更大的硬盘驱动器，同时提高处理器利用率，从而能够使用更高内核数的处理器。这样可以减少所需的硬盘驱动器数量，从而对存储性能产生积极影响，也有利于提高数据中心密度。

通过提高密度，我们可以提升能效、减少机架数、缩小数据中心占用的空间，从而降低总体拥有成本 (TCO)。据 Twitter 预计，整体而言，与原先的生产集群配置相比，通过缓存临时数据和提高内核数，他们可将总体拥有成本降低大约 30%，并将运行速度提高 50% 以上。[1]

业务挑战

Twitter 使用 Hadoop* 来存储数据和执行高级分析，从而获取重要的商业洞察。作为全球的重要 Hadoop 用户之一，Twitter 的 Hadoop 集群由五十万个计算线程和总计超过 300 PB 逻辑存储（每个集群有 30 PB 或更多逻辑存储）构成，由于存在重复，因而产生了数 EB 的物理存储。峰值集群规模可能超过 10,000 个节点，Twitter 每天处理超过 1 万亿个事件。

作者

Dave Beckett
现场可靠性工程师，Twitter, Inc.

Matt Singer
高级硬件工程师，Twitter, Inc.

Milind Damle
高级工程总监，
大数据解决方案和性能工程部门，
英特尔公司

Rakesh Radhakrishnan
高级软件工程师，Hadoop HDFS 专家，
Hadoop PMC 成员，
英特尔公司

Barrie Wheeler
高级应用工程师，
英特尔® 高速缓存加速软件专家，
英特尔公司

编著者

Varun Sampat
高级硬件工程师，Twitter, Inc.

Mark Schonbach
现场可靠性工程师，Twitter, Inc.

Ali Alavi
面向 Twitter 的行业技术专家，
云服务提供商部门，英特尔公司

Mauricio Cuervo
面向 Twitter 的高级客户执行经理，
项目主管，云服务提供商部门，
英特尔公司

Juan Fernandez
NSG 技术解决方案专家，英特尔公司

Uma Gangumalla
高级软件工程师，Hadoop HDFS 专家，
HDFS PCM 成员，英特尔公司

Devaraj Kavali
高级软件工程师，Hadoop YARN 专家，
英特尔公司

David Leone
应用工程经理，附加平台存储软件部门
主管，英特尔公司

Brien Porter
高级项目经理，开源技术主管，英特尔公司

Michal Wysoczanski
高级软件架构师，英特尔® 高速缓存加速
软件主管，英特尔公司

图 1 显示 Twitter 的典型 Hadoop 集群中的数据流。Hadoop 分布式文件系统* (HDFS*) 为每个硬盘驱动器产生大约一个数据流，而映射-归约处理（由 YARN 管理）产生多个数据流来存储临时数据。每个临时数据流都是针对不同的硬盘驱动器的，会与 HDFS 数据流重叠。

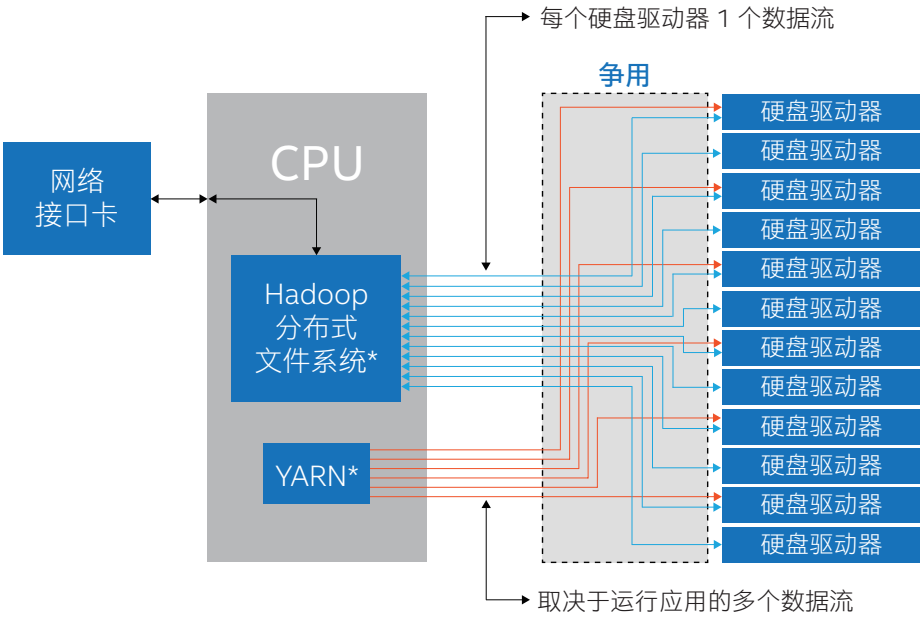

图 1. Hadoop* 集群中的典型数据流导致 HDFS* 数据与由 YARN* 管理的临时数据争用硬盘驱动器。

由于每 GB 成本比较实惠，7200RPM 硬盘驱动器成为 Twitter 的 Hadoop 集群采用的主要存储介质。一直以来，随着存储需求增长而添加更多磁盘似乎都是最佳解决方案。但是，硬盘驱动器的数量逐渐达到了临界量：硬盘驱动器容量随着时间推移而增长，但它们的吞吐量和每秒输入/输出次数 (IOPS) 停滞不前。因此，每 GB 存储的 IOPS 性能指标受到限制，约束了架构和硬件选择，必须为集群添加更多服务器，但这样会推高成本。如何在不大幅增加成本的情况下提升 I/O 性能？这是一个难题。Twitter 工程师展开了研究，结果发现一些长期以来默认的假设并不正确。

错误猜测带来重大发现

Twitter 团队坚信“如果无法测量具体指标，就无法实现性能改进”，因此他们决定结合使用以下方法，测量某个测试集群中的 I/O 和 CPU 占用率：

- 综合性基准测试 (Terasort*)
- 对典型的生产工作负载进行重现（使用 Gridmix*）
- 系统分析工具（英特尔® VTune™ 放大器 - Platform Profiler）

Twitter 的测试集群使用了双路英特尔® 至强® E5-2630 处理器 v4，每路 10 个内核/20 个线程。[4] 该集群包括 102 个节点，分布在六个机架上，使用 25 GbE 连接。同时，英特尔建立了一个相对小一些的实验室（仅九个节点）。这次探索性研究并非一帆风顺，但两个团队的协作和试验揭示了 Hadoop I/O 的一些令人惊奇的内部规律（请参见图 2）。

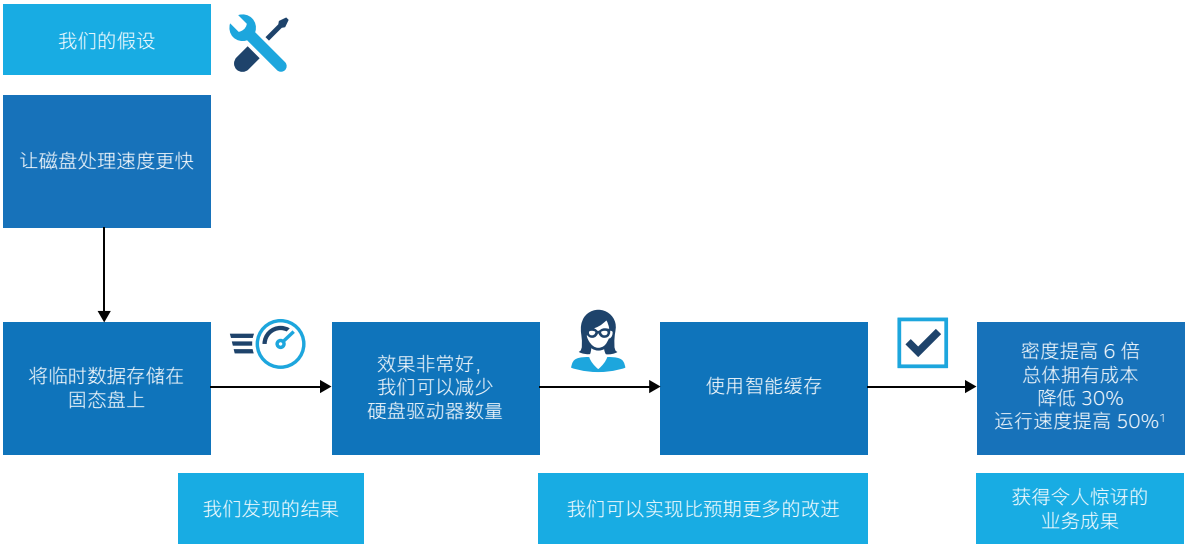

图 2. Twitter 团队的研究过程最初只有一个目标，但通过试验之后，却收获了预期之外的业务优势。

缓存一切并非万全之策

从一开始，Twitter 工程师就假定是大量数据导致了 Hadoop 运行缓慢。经过与英特尔工程师讨论，Twitter 团队决定尝试使用英特尔® 傲腾™ 固态盘 (SSD) 和英特尔® 高速缓存加速软件（英特尔® CAS）对整个磁盘子系统进行缓存。

HDFS 是专为大型数据文件的大批量读取和写入而打造的。Twitter 通常在 HDFS 中使用 512 MB 的数据块，因此，在大多数应用中，它一次可以读取或写入 0.5 GB 数据。运行 HDFS 的工作节点有 12 个驱动器，每个驱动器读取或写入批量数据以供 HDFS 使用。Twitter 工程师依赖所有驱动器的总体 IOPS，来确保将整体性能维持在较高水平。但随着磁盘变大，每 GB 的 IOPS 减小。第一个假设是：虽然读取和写入容量相对于典型缓存驱动器容量比较大，但快速缓存可以缓解 I/O 压力，从而帮助解决这个问题。

在对这种假设进行测试之后，Twitter 团队发现缓存并没有什么帮助。这背后的原因是什么呢？缓存的性能优势通常适用于要被多次访问的数据。但对于 Twitter 的工作负载而言，HDFS 数据写入磁盘后，通常很长时间内不会再使用。因此，缓存不仅无法带来性能提升，甚至测试结果还表明 I/O 性能略有降低（有时这种效应被称为“缓存污染”）。随后，我们展开了更多讨论并做出了第二个假设：如果将临时数据（请参见边栏“Apache Hadoop* 和 YARN* 简介”）存储在固态盘上会怎么样呢？

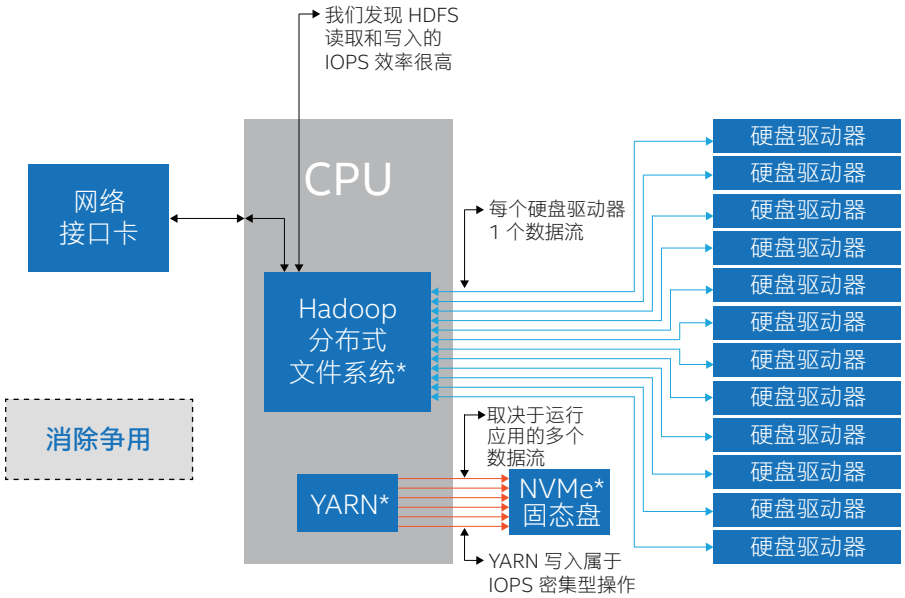

图 3. 使用基于 NVMe* 的固态盘来存储由 YARN* 管理的临时数据，消除了对硬盘驱动器的争用。

将临时数据存储在固态盘上

在与英特尔工程师讨论第二个假设之后，Twitter 决定有选择地尝试将 YARN 临时目录中包含的临时数据存储到固态盘上（请参见图 3）。

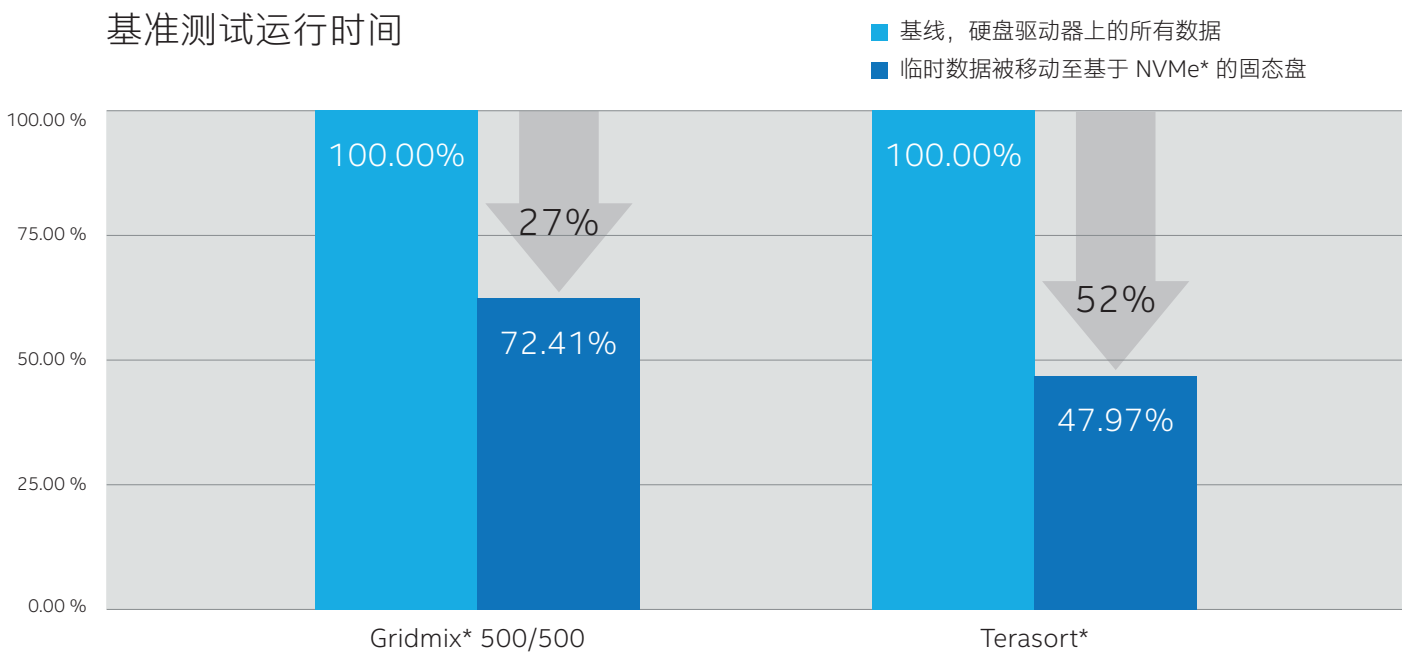

图 4. 将临时数据存储在基于 NVMe* 的固态盘上，显著减少了基准测试运行时间。[5]

结果有些令人惊讶。他们发现，HDFS 数据写入和读取的 IOPS 性能指标非常高，而临时数据的 I/O 密集程度也远超最初的设想。如图 4 所示，仅为测试集群中的每个主机添加一个英特尔® 傲腾™ 固态盘 DC P4800X 来存储临时数据，即可将 Gridmix 的运行时间减少 27.5%，将 Terasort 的运行时间减少 52%。[5]

由于 MapReduce* 临时文件和 Hadoop 分布式文件系统* (HDFS*) 不会争用同一个磁盘，因而可以减少运行时间。硬盘驱动器的利用率随之下降，Hadoop 能够更快速地处理数据。应用分析工具表明，在没有固态盘的情况下，硬盘驱动器平均每秒传输 37 MB 数据。有了固态盘，硬盘驱动器的平均流量降低至大约每秒 6 MB。而 Twitter 目前使用的硬盘驱动器的额定处理能力为每秒大约 200 MB，远高于这个数。在 Gridmix 测试过程中，分析工具还发现，CPU 利用率从平均 40% 增加到平均 57%。也就是说，CPU 的工作速度达到了原速度的 1.4 倍，这就使得运行时间得以缩短。

在英特尔实验室测试中，还发现了一个有趣的现象，Gridmix 的运行时间减少了 51.7%，这个数字几乎是在 Twitter 实验室中得出的结果的两倍。[6] 在进行比较时，我们很清楚地看到，英特尔实验室使用了 112 线程配置，采用英特尔® 至强® 铂金 8180 处理器，共有 8 个硬盘驱动器，而 Twitter 使用了 40 线程配置，共有 12 个硬盘驱动器，两者之间存在很大差异。英特尔实验室系统的每硬盘驱动器线程数（14 个）远高于 Twitter 实验室（3.33 个）。通过使用基于 NVMe 的固态盘来消除存储瓶颈，I/O 任务不再受到存储限制，而更多受到计算能力的限制；因此，英特尔的高内核数的测试集群可以更有效地扩展。（有关内核数和 I/O 性能之间关系的详细讨论，请参阅“事半功倍”。）

我们从 Twitter 的测试中得出的一个重要启示是，只专注于低级别基准测试是不够的，还要重视应用基准测试。低级别读取和写入性能数字并不是性能极限的优良指标。实际的工作负载是读取、写入和计算的混合，而且读取和写入也分为不同的类型。Twitter 的 Hadoop 工作负载是 512 MB 文件和相对较小文件的读取和写入的混合。如果不拆分工作负载，这种混合很难优化。

事半功倍

添加固态盘之后，硬盘驱动器的利用率大幅下降（将近 84%），这个数字让 Twitter 团队很直观地产生了一个想法：这样可以减少节点中的硬盘驱动器数量吗？为了解答这个问题，他们对每节点硬盘驱动器数量分别为 12、6 和 3 的集群进行了测试。图 5 显示了令人惊奇的结果：在不增加 Gridmix 运行时间的情况下，可将硬盘驱动器数量减少 75%。[7] 在没有适用固态盘存储临时数据的情况下，将硬盘驱动器从节点中移除后，Gridmix 运行时间将会大幅增加。在只有三个硬盘驱动器且没有缓存的情况下，基准测试的时间增加了 231%。但在使用固态盘的情况下，将硬盘驱动器从节点中移除后，运行时间没有明显变化。

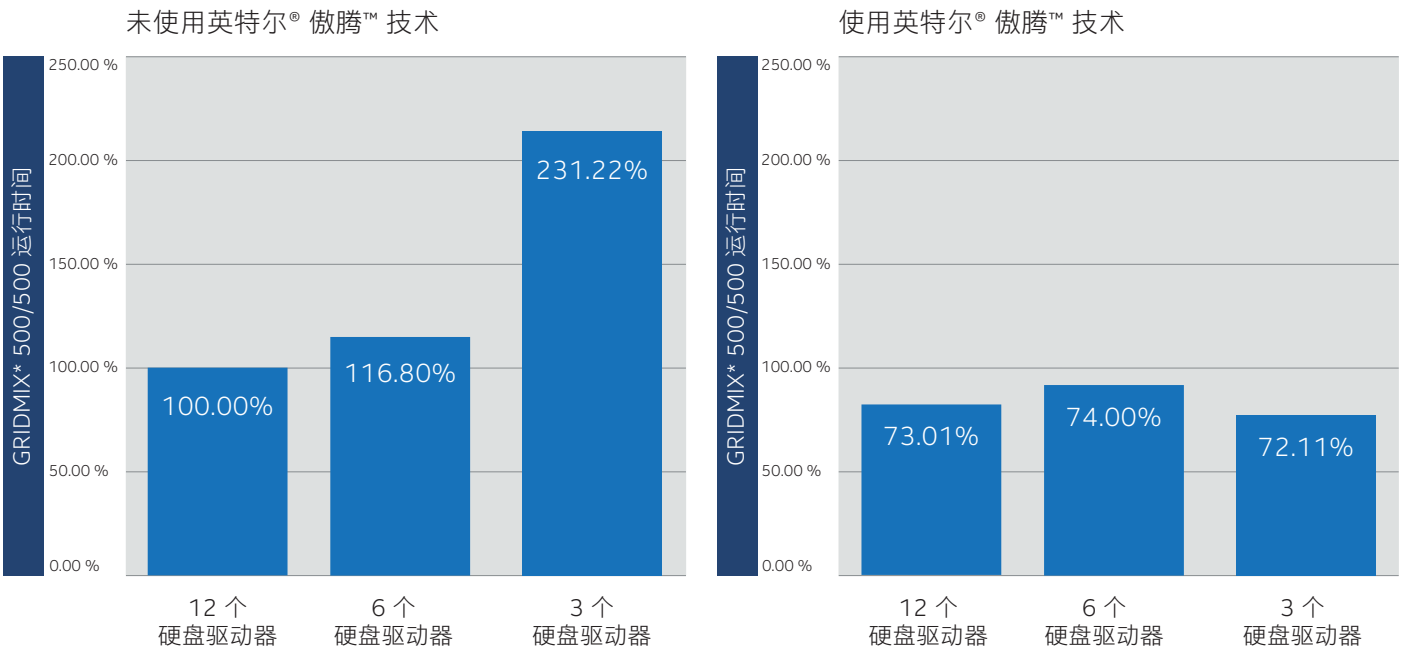

图 5. 使用英特尔® 傲腾™ 技术来存储临时数据，减少了使用的硬盘驱动器数量，而不影响基准测试运行时间。[7]

为了进一步研究内核数和 I/O 性能之间的关系，Twitter 团队试验了几种配置。他们使用 EMON 工具（英特尔® VTune™ 放大器 - Platform Profiler 的一部分）来仿真 CPU 从基线 10 内核/20 线程系统扩展到 20 内核/40 线程系统的情况。[8] 这些测试的部分结果显示在图 6 中。由于具备更高的算力，还有固态盘用于存储临时数据，与线程数较少的 12 个硬盘驱动器的基线配置相比，它可将硬盘驱动器数量减少 75%，将运行时间减少大约 40%。换言之，在优化存储子系统之后，Twitter 存储集群的瓶颈在于 CPU，而不是 IOPS。添加了固态盘用于存储临时数据后，即可实现 CPU 驱动的扩展。

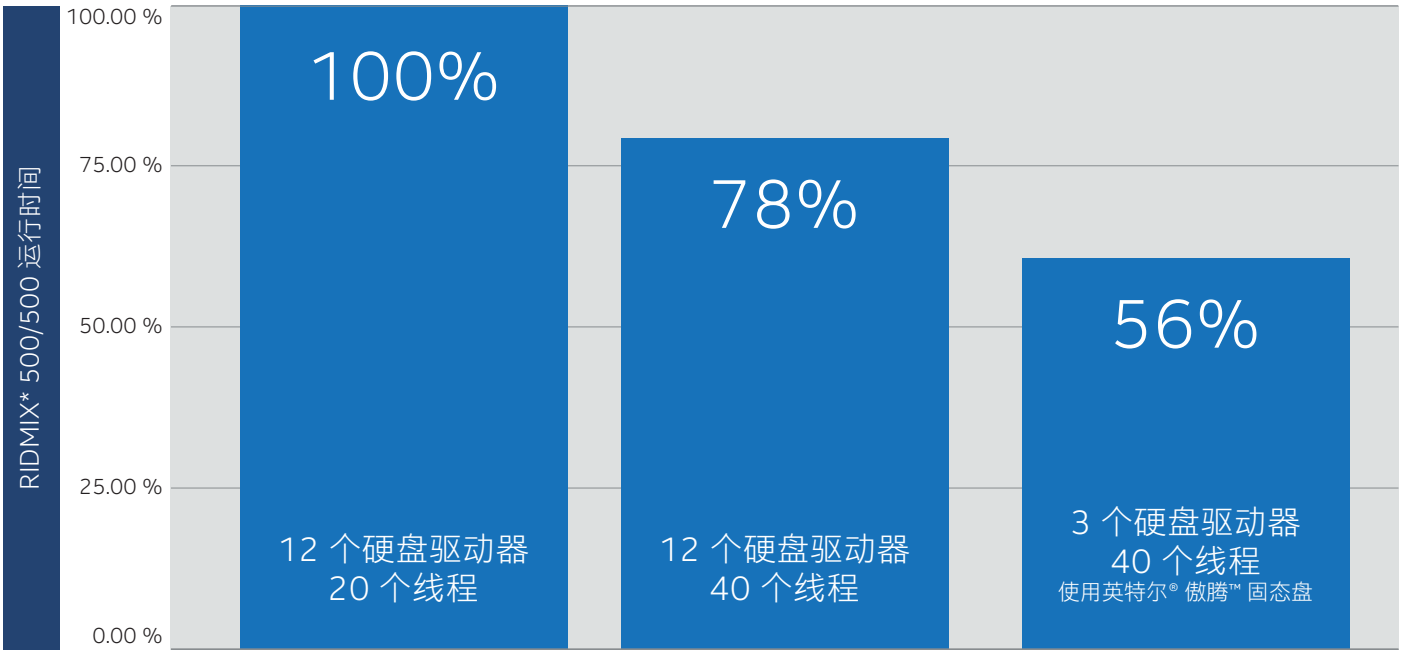

图 6. 凭借更高的计算能力和英特尔® 傲腾™ 技术，可以减少四分之三的硬盘驱动器，同时将基准测试运行时间减少大约 40%。[8]

该团队对测试结果进行了透彻分析，并向英特尔软件和服务 (SSG) 工程师进行了咨询。通过从所有数据进行推断，该团队确定基于英特尔® 至强® 金牌 6262 处理器的 24 内核系统最适合 Twitter 的集群。

值得注意的是，这款第二代英特尔® 至强® 可扩展处理器的低功耗特性对于这一部署决策至关重要，因为它能够在良好的热设计功耗 (TDP) 下提供高内核数。

密度是实现成本节约的一大因素

通过使用快速固态盘来智能缓存临时数据，并结合更好的 CPU 处理能力，带来了巨大的业务优势，超出了 Twitter 团队最初的设想。看到初步的结果后，他们认为集群密度可能增长至原先的二倍甚至三倍。但如果采用计划配置（请参见表 1），那么相对于原有集群配置，计算密度实际上有望提高 6 倍。

密集化可以减少硬盘驱动器和服务器的数量（避免资本支出），还能降低驱动器维护、用电和散热成本（节省运营支出），减少数据中心占用的空间，从而实现成本节约。例如，如果采用计划配置，Twitter 集群的硬盘驱动器数量可从超过 100,000 个减少至只有 20,000 个。这意味着风扇、电源和其他可能出现故障的运转部件的数量都将得以减少。此外，Twitter 预期硬盘驱动器数量的减少还能直接降低由于硬盘驱动器故障产生的运营成本。

经过评估后，Twitter 研究了这种配置的其他方面和未来需求，决定采用计划配置。Twitter 确定了适当的 CPU 扩展规模，预计将有一些工作负载产生更高计算量，因而产生更多的 YARN 临时空间需求，因此他们决定使用遥测数据达到以下目标:

- 必须增加线程密度。
- 必须增加临时数据的空间。
- 缓存固态盘必须达到至少每节点 6.4 TB（所见的最大数据量的第 95 百分位数）。[9]
- 减少的硬盘驱动器数量不应该是测试结果所示的最大数字，以便为 IOPS 和带宽留下一定余量。

Twitter 团队计划部署五个搭载英特尔® 至强® 金牌 6262V 处理器的机架，以在生产负载上充分验证新硬件的性能。整体而言，采用新的计划配置，Twitter 预期可将总体拥有成本降低 30%。

表 1. 实现密集化和性能提升的计划配置

	原有配置	计划配置
处理器	英特尔® 至强® E3-1230 处理器 v6（单路，4 核）	英特尔® 至强® 金牌 6262V 处理器（单路，24 核）
内存	32 至 64 GB	192 GB
硬盘驱动器 (HDD)	12 个 1 TB 或 2 TB 的硬盘驱动器	8 个 6 TB 硬盘驱动器
引导磁盘	英特尔® S4500 240 GB	英特尔® S4510 240 GB
缓存固态盘（YARN* 存储，临时数据）	N/A	1 个英特尔® 固态盘 DC P4610 6.4 TB（基于 NVMe* 的高性能固态盘）
计算	1 倍	6 倍
存储	1 倍	每节点 3 倍至 6 倍
机架减少系数	1 倍	4 倍
计算扩展	1 倍	6 倍
缓存软件	N/A	英特尔® 高速缓存加速软件（英特尔® CAS）
网络	1 GB 至 10 GB	25 GB

协作实现智能缓存

在 Twitter 的初始测试中，使用了基于 NVMe 的固态盘，临时数据被直接发送至固态盘（没有缓存），这样做的原因是对工作负载的初始分析表明，所有临时数据可以存储在 6.4 TB 的固态盘上。但经过更多研究，Twitter 工程师发现，随着通过 Twitter 集群的数据量持续增长，分析工作负载也会增加。因此，临时数据将远远超出 6.4 TB（高达 12 TB 甚至更多）。Twitter 团队与英特尔® CAS 工程师携手合作，利用英特尔® CAS 现有的功能，在缓存设备存满时，将数据存储至另一个驱动器。这样可以避免基于 NVMe 的 YARN 专用固态盘空间不足所导致的作业失败。

由 YARN 管理的临时数据通常要发送到 Hadoop 配置的目录中，因此 Twitter 团队要求英特尔® CAS 能够支持目录特定缓存。英特尔团队修改了英特尔® CAS，并执行了密集性能测试，成功地满足了这个要求。这种特性确保所有临时数据能够按目录隔离，并传输至缓存设备。这样，临时数据能够发挥缓存的全部优势，另外还能充分利用英特尔® CAS 保护数据的能力，当缓存设备存满时，临时数据可转存至硬盘驱动器。大多数作业只产生少量的临时数据，可以完全依赖于固态盘的容量，但超出该容量的大型作业仍然要充分利用缓存的优势。因此，临时数据的最大容量不受限于缓存固态盘的容量。但是，要大幅提升性能，该驱动器的容量必须在大部分时候大于临时数据量。

Apache Hadoop* 和 YARN* 简介

Apache Hadoop* 软件系统是一个框架，让我们能够使用简单的编程模式，对计算机集群上的大数据集进行分布式处理。其设计目的是从单台服务器纵向扩展到数千台服务器，每台服务器均可提供本地计算和存储。该系统可检测和处理应用层的故障，在计算机集群顶部提供高可用性服务，而单独的计算机则容易出现故障。YARN* 是 Apache Hadoop* 的资源管理器和作业调度工具。实质上，您可将 YARN 视为 Hadoop V1 中的资源管理的抽象化，允许使用除 MapReduce* 之外的不同计算框架，例如 Spark*。

在集群架构中，YARN 是软件层，驻留在 Hadoop 分布式文件系统* (HDFS*) 与运行应用的处理引擎之间。在 YARN 顶部使用框架的应用（例如 MapReduce）会在运行作业时创建临时文件，例如映射输出。在作业运行时，应用将这些临时数据写入到磁盘，在作业完成后，它们会清除临时数据。一般来说，所有临时数据文件都非常小。由于文件较小，而且要进行重复访问，这两个因素结合，使得临时文件非常适合缓存到单独驱动器（而不是写入到 HDFS 使用的主硬盘驱动器）。此外，缓存还可减少对硬盘驱动器的争用。

转换 Hadoop 集群的最佳实践

在这个过程中，Twitter 认识到了以下几点：

- 对长期以来默认的假设提出质疑；这会让您有意想不到的发现。例如，Twitter 团队从未预期到他们可以减少系统中 75% 的硬盘驱动器，而不影响性能。
- 使用清晰定义的流程进行高效的试验：测量、试验、学习和重复。
- 在测量方面，可以使用先进的分析工具，结合强大的可视化工具，从而轻松了解在测试和生产集群中发生的真实情况。
- 务必测量多个级别的读取和写入性能，而不只限于低级别。有一点必须注意，您的特定工作负载可能与 Twitter 的工作负载存在很大区别。具体来说，您应该知道其中发生了哪些类型的读取和写入。这可为您的优化工作提供指引。
- 与能够为您提供新思路的其他专家展开协作。例如，英特尔团队分享了自己的测试结果，帮助解释某些问题发生的原因，在规划新的集群配置时，帮助 Twitter 达到甚至超出他们的扩展和能效目标。

未来举措

试验没有尽头，您可以不断地学习和改进。展望未来，Twitter 团队计划进行更多试验，探究以下问题：

- 在基于 NVMe 的固态盘上，最合适的缓存容量是多少
- 固态盘耐用性需求
- 硬盘驱动器、线程、基于 NVMe* 的固态盘三者之间的最佳平衡

结论

图 7 总结了 Twitter 团队发现的关键要点。在将 MapReduce 进程产生的临时数据转存至快速固态盘之后，需要的硬盘驱动器得以减少。增加每个磁盘的计算线程，可以进一步增强性能。在整个发现过程中，Twitter 和英特尔工程师通过协作取得了丰硕的成果，致力于寻找能够解决 Twitter 难题的解决方案。例如，英特尔® CAS 的目录特定缓存功能就是此次协作的直接成果。Twitter 和英特尔将继续携手合作，分享研究成果，进一步优化 Twitter 的 Hadoop 集群。

发现

1 将临时数据存储在 NVMe* 固态盘上

仅这一项措施即可显著改变磁盘访问模式。

2 采用高密度驱动器

将临时数据转移至固态盘后，我们不再需要那么多硬盘驱动器。

3 每个磁盘具有更高计算能力

我们设计的下一个平台必须为系统中的每个硬盘增加更多计算线程。

图 7. 从 Twitter 的 Hadoop* 集群优化工作中发现的关键要点。

了解更多信息

您可能会发现以下资源对您有用：

- 面向云服务提供商的英特尔® 资源
- 英特尔® 固态盘 DC D7 系列
- 英特尔® 傲腾™ 固态盘 DC P4800X 系列
- 英特尔® 高速缓存加速软件（英特尔® CAS）

[1] 基准：单路英特尔® 至强® 处理器 E3-1230 v6（4 核）；32 至 64 GB RAM；1 个 1 TB 或 2 TB 的硬盘驱动器；英特尔® S4500 240 GB 引导磁盘；1 GbE 至 10 GbE 以太网；无缓存。
测试：单路英特尔® 至强® 金牌 6262 处理器（24 核）；192 GB RAM；英特尔® S4500 240 GB 引导磁盘；8 个 6 TB 硬盘驱动器；1 个英特尔® 固态盘 DC P4610 6.4TB；25 GbE 以太网；使用英特尔® 高速缓存加速软件进行缓存。

操作系统：Twitter CentOS* 6 Derivative，内核版本 2.6.74-t1.el6.x86_64（基于上游 4.14.12 内核），BIOS 版本：D3WWM11，微代码版本：0xb000021

[2] Backblaze，2018 年 9 月，“Hard Disk Drive (HDD) vs Solid State Drive (SSD): What's the Diff? ”（硬盘驱动器 (HDD) 与固态硬盘 (SSD)：差异何在？），https://www.backblaze.com/blog/hdd-versus-ssd-whats-the-diff/

[3] 基准：双路英特尔® 至强® E5-2630 v4 @ 2.2 GHz（每路 10 个内核/20 个线程）；128 GB RAM；12 个 6 TB 7200 RPM SATA 硬盘驱动器；1 个 SATA 固态盘引导磁盘；25 GbE 以太网；分布在 6 个机架上的 102 个节点。工作负载：Gridmix* 和 Terasort*。Gridmix 分数：3309 秒；Terasort 分数：5504 秒
测试：双路英特尔® 至强® E5-2630 v4 @ 2.2 GHz（每路 10 个内核/20 个线程）；128 GB RAM；12 个 6 TB 7200 RPM SATA 硬盘驱动器；1 个 SATA 固态盘引导磁盘；1 个基于 NVMe* 的 750 GB 英特尔® 傲腾™ DC P4800X 固态硬盘；25 GbE 以太网；分布在 6 个机架上的 102 个节点。工作负载：Gridmix 和 Terasort。Gridmix 分数：2396 秒；Terasort 分数：2640 秒

操作系统：Twitter CentOS* 6 Derivative，内核版本 2.6.74-t1.el6.x86_64（基于上游 4.14.12 内核），BIOS 版本：D3WWM11，微代码版本：0xb000021

[4] 请注意，测试集群使用了比 Twitter 的生产 Hadoop* 集群更高的内核数，在Twitter 的集群中，每个硬盘驱动器仅提供 4 个内核/8 个线程。

[5] 基准：双路英特尔® 至强® E5-2630 v4 @ 2.2 GHz（每路 10 个内核/20 个线程）；128 GB RAM；12 个 6 TB 7200 RPM SATA 硬盘驱动器；1 个 SATA 固态盘引导磁盘；25 GbE 以太网；分布在 6 个机架上的 102 个节点。工作负载：Gridmix* 和 Terasort*。Gridmix 分数：3309 秒；Terasort 分数：5504 秒
测试：双路英特尔® 至强® E5-2630 v4 @ 2.2 GHz（每路 10 个内核/20 个线程）；128 GB RAM；12 个 6 TB 7200 RPM SATA 硬盘驱动器；1 个 SATA 固态盘引导磁盘；1 个基于 NVMe* 的 750 GB 英特尔® 傲腾™ DC P4800X 固态硬盘；25 GbE 以太网；分布在 6 个机架上的 102 个节点。工作负载：Gridmix 和 Terasort。Gridmix 分数：2396 秒；Terasort 分数：2640 秒

操作系统：Twitter CentOS* 6 Derivative，内核版本 2.6.74-t1.el6.x86_64（基于上游 4.14.12 内核），BIOS 版本：D3WWM11，微代码版本：0xb000021

[6] 英特尔测试。
基准：1 个名字节点 (name node)（2 个英特尔® 至强® E5-2699 v4 @2.20 GHz，128GB DDR4-2666 ECC，240 GB 的英特尔® 固态盘 DC S4600 用作引导盘，2 个英特尔® 以太网控制器 10-Gigabit X540-AT2 版本 01）；9 个数据节点 (data node)（2 个英特尔® 至强® 铂金 8180 处理器 @ 2.5 GHz，128 GB DDR4-2666 ECC， 240 GB 的英特尔® 固态盘 DC S4600 用作引导盘，4 个英特尔® 以太网控制器 X710/X557-AT 10GBASE-T 版本 02，8 个硬盘驱动器 Seagate 7200RPM SATA ST4000NM0085）。Gridmix* 分数：5592 秒
测试：1 个名字节点 (name node)（2 个英特尔® 至强® E5-2699 v4 @2.20 GHz，128GB DDR4-2666 ECC，240 GB 的英特尔® 固态盘 DC S4600 用作引导盘，2 个英特尔® 以太网控制器 10-Gigabit X540-AT2 版本 01）；9 个数据节点 (data node)（2 个英特尔® 至强® 铂金 8180 处理器 @ 2.5 GHz，128 GB DDR4-2666 ECC，240 GB 的英特尔® 固态盘 DC S4600 用作引导盘，4 个英特尔® 以太网控制器 X710/X557-AT 10GBASE-T 版本 02，8 个硬盘驱动器 Seagate 7200RPM SATA ST4000NM0085，1 个基于 NVMe* 的 1.6 TB 英特尔® P4600 固态盘和 1 个基于 NVMe 的 750 GB 英特尔® 傲腾™ P4800X 固态盘用于存储临时数据）。Gridmix 分数：2702 秒
软件：操作系统：Twitter CentOS* 6 Derivative，内核版本 2.6.74-t1.el6.x86_64（基于 4.14.12 内核），应用：Apache Hadoop* 2.9 复制系数 3，网络接口绑定：2x10 Gbps 接口绑定 20 Gbps 模式 4 LACP，英特尔® 高速缓存加速软件版本 3.9（缓存 YARN* 目录和元数据），Supermicro* X11DPU BIOS 版本 2.0a，微代码版本：0x200003a

[7] 基准：双路英特尔® 至强® E5-2630 v4 @ 2.2 GHz（每路 10 个内核/20 个线程）；128 GB RAM；12 个、6 个和 3 个 6 TB 7200 RPM SATA 硬盘驱动器；1 个 SATA 固态盘引导磁盘；25 GbE 以太网；分布在 6 个机架上的 102 个节点。工作负载：Gridmix*。Gridmix 分数（12 个硬盘驱动器）：3309 秒，Gridmix 分数（6 个硬盘驱动器）：3865 秒，Gridmix 分数（3 个硬盘驱动器）：7651 秒
测试：双路英特尔® 至强® E5-2630 v4 @ 2.2 GHz（每路 10 个内核/20 个线程）；128 GB RAM；12 个、6 个和 3 个 6 TB 7200 RPM SATA 硬盘驱动器；1 个 SATA 固态盘引导磁盘；1 个基于 NVMe* 的 750 GB 英特尔® 傲腾™ DC P4800X 固态硬盘；25 GbE 以太网；分布在 6 个机架上的 102 个节点。工作负载：Gridmix*。Gridmix 分数（12 个硬盘驱动器）：2416 秒，Gridmix 分数（6 个硬盘驱动器）：2448.5 秒，Gridmix 分数（3 个硬盘驱动器）：2386 秒

操作系统：Twitter CentOS* 6 Derivative，内核版本 2.6.74-t1.el6.x86_64（基于上游 4.14.12 内核），BIOS 版本：D3WWM11，微代码版本：0xb000021

[8] 基线（12 个硬盘驱动器，20 个线程）：双路英特尔® 至强® E5-2630 v4 @ 2.2 GHz（每路 10 个内核/20 个线程，但有一半内核关闭），128 GB RAM，12 个 6 TB 7200 RPM SATA 硬盘驱动器，1 个 SATA 固态盘引导磁盘，25 GbE 以太网；分布在 6 个机架上的 102 个节点。工作负载：Gridmix*。Gridmix 分数：4227 秒
测试（12 个硬盘驱动器，40 个线程）：双路英特尔® 至强® E5-2630 v4 @ 2.2 GHz（每路 10 个内核/20 个线程，所有内核活动），128 GB RAM，12 个 6 TB 7200 RPM SATA 硬盘驱动器，1 个 SATA 固态盘引导磁盘，25 GbE 以太网；分布在 6 个机架上的 102 个节点。工作负载：Gridmix*。Gridmix 分数：3309 秒
测试（3 个硬盘驱动器，40 个线程，采用基于 NVMe* 的缓存）：双路英特尔® 至强® E5-2630 v4 @ 2.2 GHz（每路 10 个内核/20 个线程，所有内核活动），128 GB RAM，3 个 6 TB 7200 RPM SATA 硬盘驱动器，1 个 SATA 固态盘引导磁盘，1 个基于 NVMe* 的 750 GB 英特尔® 傲腾™ DC P4800X 固态盘，25 GbE 以太网；分布在 6 个机架上的 102 个节点。工作负载：Gridmix*。Gridmix 分数：2386 秒

操作系统：Twitter CentOS* 6 Derivative，内核版本 2.6.74-t1.el6.x86_64（基于上游 4.14.12 内核），BIOS 版本：D3WWM11，微代码版本：0xb000021

[9] 6.4 TB 远大于可用的最大容量英特尔® 傲腾™ 数据中心级固态盘 (1.5 TB)。因此，虽然这些测试使用了英特尔® 傲腾™ 数据中心级固态盘来缓存临时数据，但对于计划生产配置，Twitter 选择了英特尔® 固态盘 DC P4610，因为它实现了基于 NVMe* 的固态盘与 Twitter 需要的高容量之间的适当平衡。

解决方案提供商：

性能测试中使用的软件和工作负荷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。更多信息，详见 www.intel.cn/benchmarks。

由 Twitter 测试。详情请参阅配置信息披露。

性能测试结果基于 2018 年 9 月 26 日进行的测试，且可能并未反映所有公开可用的安全更新。详情请参阅配置信息披露。没有任何产品或组件是绝对安全的。

优化声明：英特尔编译器针对英特尔微处理器的优化程度可能与针对非英特尔微处理器的优化程度不同。这些优化包括 SSE2、SSE3 和 SSSE3 指令集和其他优化。对于非英特尔微处理器上的任何优化是否存在、其功能或效力，英特尔不做任何保证。本产品中取决于微处理器的优化是针对英特尔微处理器。不具体针对英特尔微架构的特定优化为英特尔微处理器保留。请参考适用的产品用户与参考指南，获取有关本声明中具体指令集的更多信息。

英特尔并不控制或审计第三方数据。 请您审查该内容，咨询其他来源，并确认提及数据是否准确。

描述的成本降低情景均旨在在特定情况和配置中举例说明特定英特尔产品如何影响未来成本并提供成本节约。情况均不同。英特尔不保证任何成本或成本降低。

英特尔、英特尔标识、傲腾、至强和 VTune 是英特尔公司或其子公司在美国和/或其他国家的商标。

Twitter、Tweet、Retweet 和小鸟标识是 Twitter, Inc. 的注册商标。

*其他的名称和品牌可能是其他所有者的资产。

 0319/CAT/RW/PDF