



白皮书

大数据在裸机上的表现更好

让大数据性能成为优先事项

执行摘要

如今,企业创建并从多种来源收集数量空前的结构化数据和非结构化数据。存储、处理这种“大数据”从中提取价值并非易事。IT 专业人员常常配置公共云服务器来扩展存储和处理能力,以适应这种源源不断的数据流,但这些虚拟化资源无法提供相等裸机服务器所具备的性能和一致性。

IBM Cloud 在虚拟服务器和裸机服务器上测试了大数据工作负载的性能和一致性,以比较两款平台对于用于存储和处理海量数据的应用的适用性。利用这些结果,在为存储密集型和处理器密集型工作负载选择云资源时,IT 专业人员可以做出更好的决定。

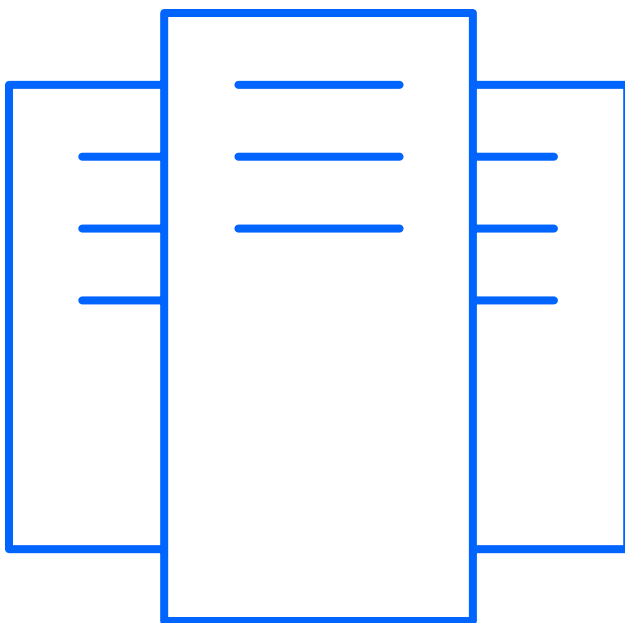


什么是大数据？

随着存储技术的发展以及更大容量变得越来越经济实惠，企业发现了收集和处理更多数据的新方法。从这种信息中，公司发现具有潜在商业价值的洞察。而挑战在于，如何组织和分析数据以制定新业务战略并作出组织决策。

直到最近，最普遍的数据组织和分析工具仍是采用结构化查询语言 (SQL) 的关系数据库管理系统 (RDBMS)。SQL 解决方案使用通常在单台服务器上存储和操作的结构化数据集。当数据集的大小触及现有服务器容量上限时，该解决方案会进行纵向扩展，转向处理性能更强、存储和 RAM 容量更多的更大服务器。这种纵向扩展可能耗费大量时间，并导致成本大幅增加。

当数据从众多来源、在无数架构中更快涌入时，数据库管理员需要最大限度提升他们解决方案的效率和可扩展性。结果，许多人开始利用 NoSQL (Not Only SQL) 数据库，该数据库使用非关系型和非结构化数据集。“大数据”架构让数据能够在多个系统中存储，支持 NoSQL 应用通过逐步增添商品系统进行横向扩展，从而实现按需应变的容量增长和更高的成本效益。



这些大数据架构可以理解海量数据, 但要做到这一点, 数据提出了重大的基础架构要求:

- 适合海量数据的存储
- 按需移动和加载数据的 RAM
- 与解决方案所需性能水平相称的处理能力
- 能够以低延迟连接分布式数据存储以增强性能的网络

为了满足这些要求, 许多企业将云计算资源用作底层基础架构, 以横向扩展他们的大数据环境。在这些环境中, 最常见的构建模块是虚拟化公共云服务器和裸机服务器。

大数据的四个 V

规模 (Volume): 想象 PB 级规模。从网络历史记录, 到公共云, 再到私有内部文档, 企业存储各种信息。

种类 (Variety): 大量的结构化数据和非结构化数据, 包括电子邮件、社交媒体、视频、图像、天气数据、博客等等。

速度 (Velocity): 数据在不断产生, 利用实时查询按需获取有意义的信息。

价值 (Value): 从大数据获取的有意义洞察, 超越了传统智能查询和报告的结果。这些洞察可以转变成预测性分析, 以获取趋势和模式。

裸机服务器与虚拟服务器

将裸机服务器和虚拟服务器当作同一工具箱中的两款工具加以考虑。一款工具并不天生比另一款工具好；每款工具都有它自己的优势和劣势。

裸机服务器向客户提供服务器上原始硬件资源的直接、专有的访问。虚拟服务器是独立的云实例，由可能是公共的（共享）或私有的硬件节点上的虚拟机管理程序配置。

裸机服务器 – 原始马力

对于处理器密集型和磁盘 I/O 密集型工作负载，裸机服务器（有时称作专用服务器）是理想选择。这些服务器是单一租户模式，因此它们完全专用于单个客户。这意味着吵闹的邻居不会影响性能。

此外，由于裸机服务器不在虚拟机管理程序之上运行，工作负载无需支付“虚拟机管理程序税”，即充当操作系统和硬件中间人的虚拟机管理程序所造成的性能下降。

由于没有虚拟机管理程序抽象硬件，相比虚拟机服务器，裸机服务器通常花费更长时间进行部署和配置。当基础架构需要快速扩展时，通常避免使用裸机。为了解决这个缺点，IBM Cloud 设计了自动化的裸机服务器部署和控制功能，最少只需 20 至 30 分钟让特定配置上线，只需 2 至 4 个小时让完全定制的服务器（您选择的处理器、内核、RAM、存储、端口等）上线。

虚拟服务器 – 灵活性与可扩展性

对于规模大幅变化或需要在不断变化的市场中保持敏捷性的应用和工作负载，虚拟服务器是理想选择。虚拟服务器在单租户或多租户公共云环境中的虚拟机管理程序上进行配置。最少只需五分钟，即可按月或按小时部署虚拟服务器资源，因此您可以通过非常快地添加额外服务器进行横向扩展。

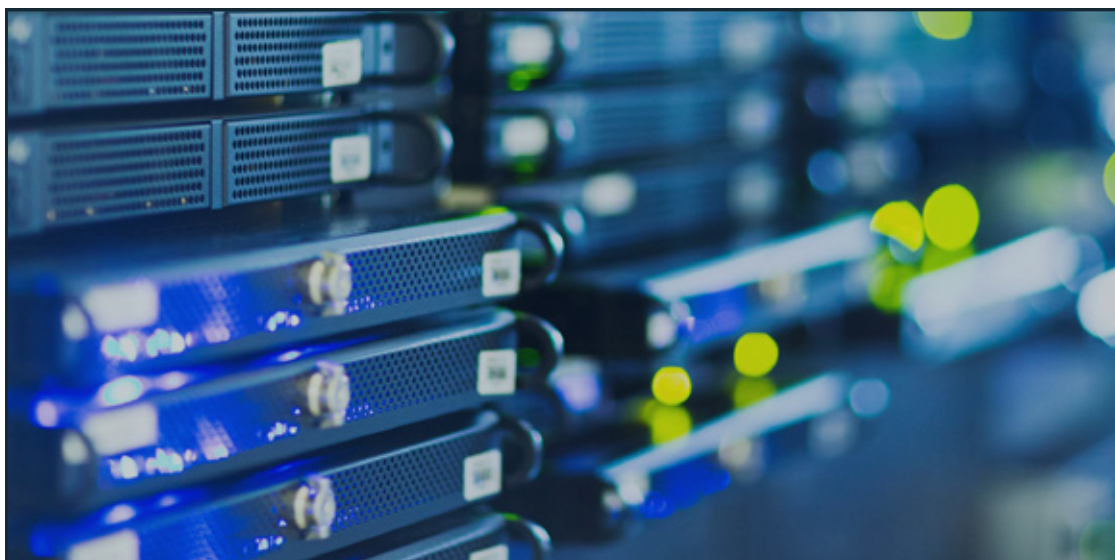
裸机与虚拟 – 配合使用

IBM Cloud 在单个统一的云环境中配置裸机服务器和虚拟服务器, 让客户可以选择和控制将支持他们各种不同工作负载的资源。

企业需要变革。我们的解决方案旨在让您专注于今天的需求, 无需担心未来几天、几周或几个月的需求将如何变化。

IBM Cloud 平台和基础架构旨在为具备彻底的可扩展性而设计和构建:

- 按需添加裸机服务器和虚拟服务器
- 必要时进行收缩以降低成本
- 按小时或按月订购, 以适合不同项目的时间表
- 无长期合同



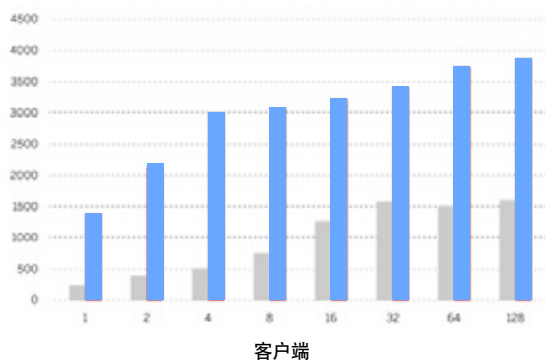
大数据 - 性能

为了确定大数据应用更适合使用裸机服务器还是虚拟服务器, 我们安排了基准测试, 以衡量这两个平台的性能和一致性。为了精确地衡量性能, IBM Cloud 工程师配置了相等的裸机和虚拟测试环境, 以利用其免费提供的基准测试工具查询和更新 MongoDB 数据集 (详见附录 A)。

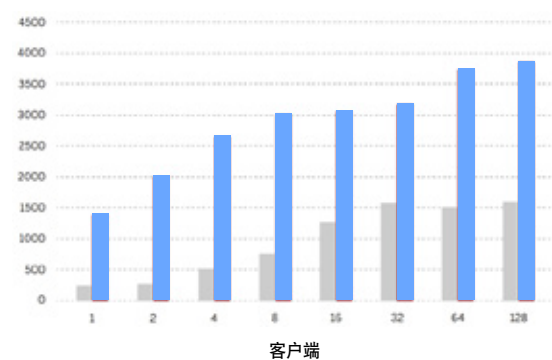
根据涉及的并发客户端数量, 此基准测试工具记录每个集群的每秒读取和写入操作。测试结果意义重大。就平均读取和写入而言, 每个裸机环境的表现都超过了相等的虚拟服务器。

虚拟服务器与裸机服务器

并发客户端的平均每秒读取操作



并发客户端的平均每秒写入操作



虚拟服务器

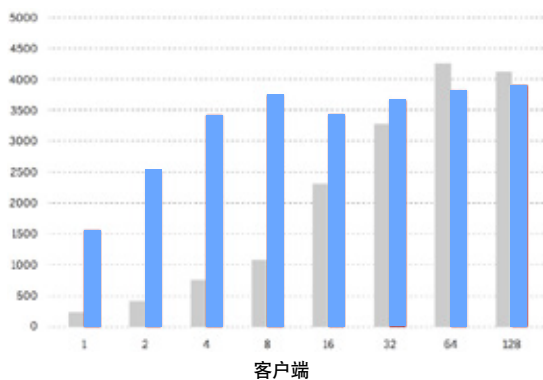
裸机服务器

由于裸机环境可以直接利用服务器的硬件资源, 而不必与其他用户争夺资源, 裸机服务器提供的性能最高是相等虚拟服务器的六倍。

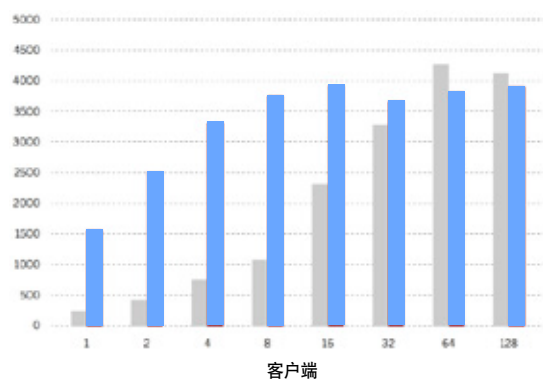
在测量平均每秒读取和写入操作时, 基准测试工具还记录了每个环境的峰值性能, 这些结果也值得注意 (出于不同的原因):

虚拟服务器与裸机服务器

并发客户端的平均每秒读取操作



并发客户端的平均每秒写入操作



虚拟服务器

裸机服务器

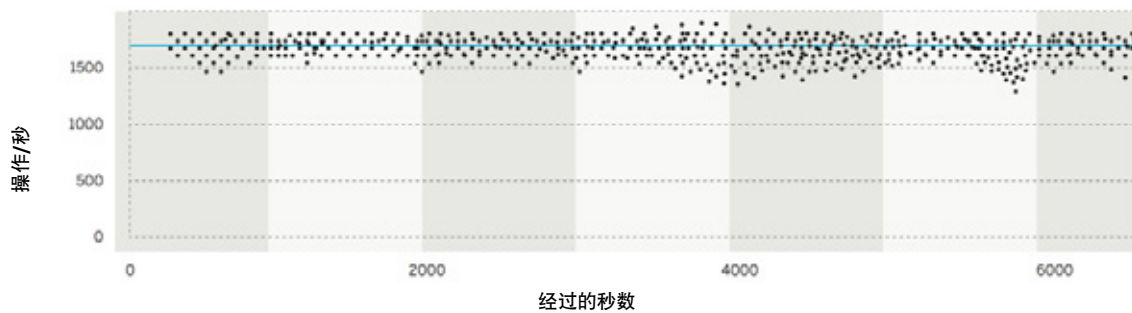
裸机环境的每秒读取和写入操作的峰值结果非常接近为该环境记录的平均读取和写入操作, 而在虚拟服务器环境中, 峰值结果与平均结果差别很大。在这两个情形中, 虚拟服务器实际上取得高于裸机服务器的峰值。参照虚拟服务器环境记录的平均每秒操作, 这些结果实际上凸显了大数据工作负载的另一个关键性能指标: **一致性**。

大数据 - 一致性

性能只有在保持一致性时才有意义。在我们的性能测试中, 虚拟服务器环境可能创下了每秒 4,500 次读取操作的峰值记录, 但平均而言, 该环境每秒 1,500 次读取操作。如果一个环境的性能在两秒之间的差异如此巨大, 就很难构建一个环境来处理日益增长的工作负载。为了比较裸机服务器与虚拟服务器的结果一致性, IBM Cloud 工程师利用 Basho Bench, 配置了两个五节点 Riak 集群和模拟的负载部署 (详见附录 B)。这项测试观测并绘制两小时时段内的每秒操作情况:

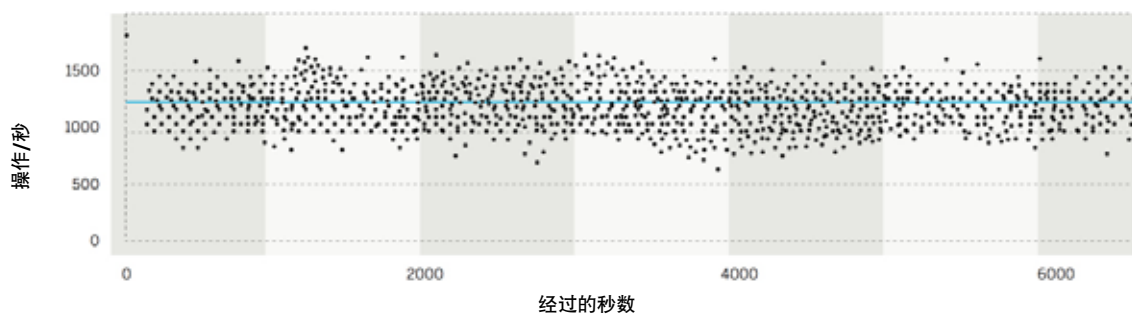
裸机服务器:

有负载的每秒操作 (2 小时)



虚拟服务器:

有负载的每秒操作 (2 小时)



在整个测试中，裸机环境的平均每秒操作更高，但更引人注目的是，结果更紧密地凝聚在平均值的周围。当虚拟服务器环境中的性能在两秒之间的差异巨大时，容量规划成为一个难题。当决定是扩展还是收缩环境时，应采用什么统计数据？如果您构建环境以适应最差记录的结果，当性能良好时，您将过度配置资源。如果选择让容量基于最佳结果，很可能导致环境性能不足。而让容量基于平均结果，无异于是在这两种选择之间抛硬币。

企业依赖 **一致的结果** 预测趋势、编制预算，
并作出重要决策。云基础架构环境的规划也不例外。

大数据需要裸机

快速、轻松部署的承诺可能让人难以抗拒。一些应用最适合运行在公共云环境的虚拟服务器上，但大数据不属于这类应用。

应注意：

- 运行大数据等高 I/O 工作负载的云环境的两个最重要的特征是 **性能** 和 **一致性**。
- 裸机服务器经过配置和优化，可在 **提供和处理海量数据** 时 **提供无以伦比的性能结果**。
- 当多名用户共享同一虚拟服务器主机节点时，负责处理高 I/O 工作负载的虚拟服务器可能 **受到其他用户资源使用的负面影响**。
- 裸机服务器 **资源位于本地，不是共享的**，因此工作负载表现的一致性远高于在共享和/或联网虚拟服务器环境中的表现。
- 虚拟服务器可以快速配置，并以远快于裸机服务器的速度横向扩展，但对于无需突发性的工作负载，则可以受益于裸机服务器的性能和一致性。

什么让 IBM Cloud 成为 大数据工作负载的理想提 供商？

无与伦比的技术：IBM Cloud 为您提供最高性能的云基础架构。无论您的大数据在全球扩展，还是在本地扩展，我们遍布全球的数据中心和一流的裸机服务器及虚拟服务器都可以胜任这项工作。

无缝的网络：我们的高速网络集成公共的、私有的和内部的管理网络，从而提供更高速度，这对于分析和传输大数据必不可少。

全面的管理和自动化：我们在一体化的自动化平台上开发了不同类型的云解决方案。每个服务器、存储设备以及管理和安全服务都可以通过单个管理系统加以控制，并可以供我们的 API、客户门户甚至是移动应用访问。

在裸机服务器上运行您的大数据。我们的 IBM Cloud 专家将帮助您打造高性能的云基础架构，以充分满足您的大数据需求。

探索 IBM Cloud 裸机服务器和虚拟服务器，网址：

<http://ibm.co/bare-metal>,

并进一步了解按订单开发的大数据解决方案和应用特定的

Riak、Hadoop、MongoDB 最佳实践，网址：

<http://ibm.co/big-data>.

还有其他疑问？

咨询专家：

<http://ibm.co/contact-us>

或致电：214-442-0600.

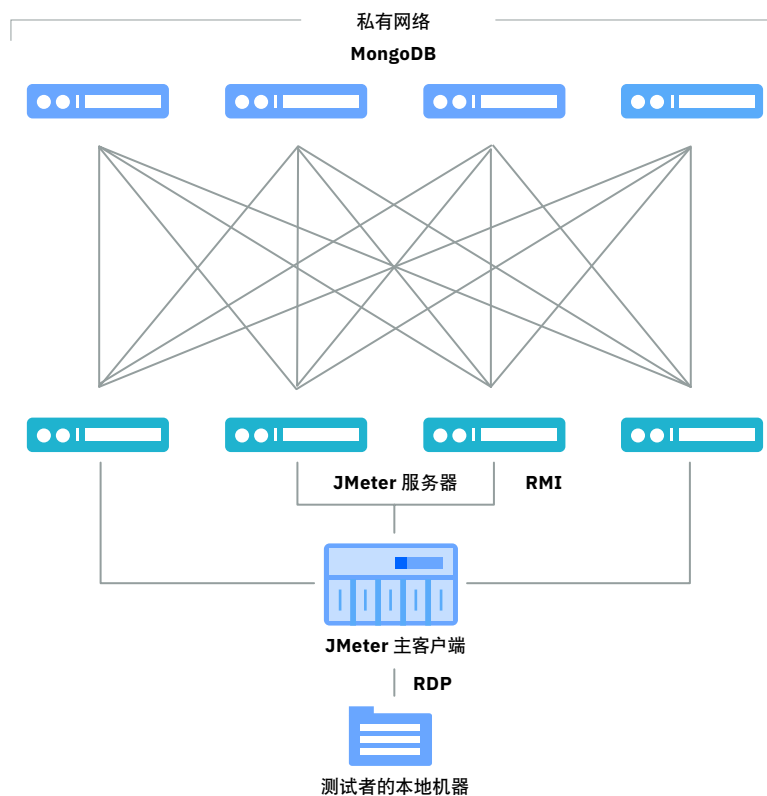
附录 A

大数据性能测试方法 — MongoDB

512kb 文件的数据集被预先加载到每台服务器的单个 MongoDB 实例中。创建的数据集相比可用内存存在各种大小，从而产生大于（2 倍）和小于可用内存的数据集。测试还确保数据集在测试过程中得到足够频繁的更改，以防止查询将所有数据缓存到内存中。

一旦数据集创建好，包含 4 个内核和 16GB RAM 的 JMeter 服务器被用于驱动基准测试工具 MongoDB 的“benchrun”。以下图表说明了我们如何设置测试环境。

这些 Jmeter 服务器充当在 MongoDB 实例上产生流量的客户端。每个客户端按每个更新六个请求的比例随机生成查询和更新请求（测试中的更新更求用于确保数据未被完全缓存到内存中，且从不从磁盘操作读取。）这些测试旨在从指数级增长的客户端在服务器上创建极高负载，直到系统资源变饱和，我们记录 MongoDB 应用的性能结果。



测试配置

- 数据集（32GB 的 .5mb 文件）
- 6:1 查询：更新操作的 200 次迭代
- 并发客户端连接按指数从 1 增长到 128 个
- 测试持续时间长达 48 小时

附录 A (续)

大数据性能测试方法 — MongoDB

裸机服务器与虚拟服务器

	裸机服务器节点	虚拟服务器节点
内核	两个 6 核 Intel 5670 CPU	26 个虚拟计算单元
操作系统	64 位 CENTOS	64 位 CENTOS
RAM	36 GB RAM	30 GB RAM
RAID	2 个 64 GB SSD RAID1 (Journal Mount)	2 个 64 GB 网络存储 RAID1 (Journal Mount)
SAS	4 个 400 GB SSD RAID10 (Data Mount)	4 个 300 GB SSD RAID10 (Data Mount)
网络	1 GB 网络 绑定	1 GB 网络

附录 B

大数据性能测试方法 — Riak

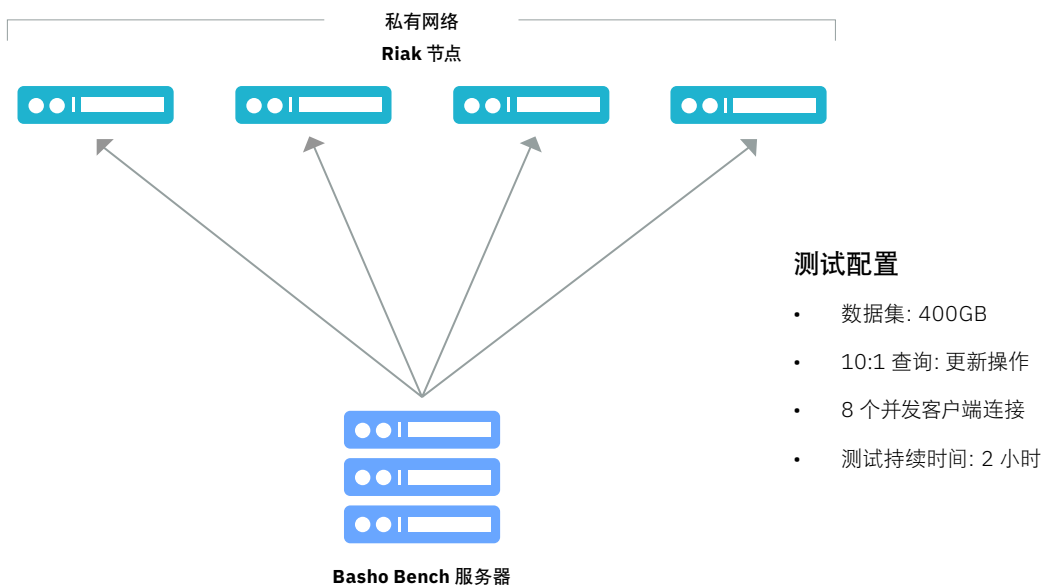
在裸机服务器和公共云虚拟服务器上，利用 Riak 1.3.1 创建五节点集群。要优化 Riak 性能，在每个服务器的操作系统级别进行微调（运行 CentOS 64 位）：

```
Noatime  
Nodiratime  
barrier=0  
data=writeback  
ulimit -n 65536
```

通用的 Noatime 和 Nodiratime 设置消除了读取过程中的写入需求，从而提高性能并减少磁盘损耗。barrier 和 writeback 设置不那么通用，可能与您通常设置的不一样。虽然这些设置在磁盘出现故障时存在极小的数据丢失风险，应记住 Riak 解决方案以五节点环形式部署，在环中的多个节点中存在数据冗余性。

此外，考虑到每个节点还利用 RAID10 存储阵列进行部署，在整个解决方案中，单个磁盘出现故障时的微小数据丢失风险将不会影响整个数据集（因为相关数据有许多冗余副本可用）。鉴于涉及的风险极少，这两种设置带来的性能提升具有合理性。

当所有节点经过微调并配置到集群中时，我们设置了 Basho 的测试工具 Basho Bench，以远程模拟部署上的负载。通过配置一些 worker 以利用一个 driver 类型来生成负载，Basho Bench 让您可以为 Riak 集群创建可配置的测试计划。它被封装为一个 Erlang 应用，其中包含配置文件示例，您可以通过修改示例，为并发性、数据集大小和您的测试持续时间创建详细信息。结果能以 CSV 数据形式显示，还有一个可选的图像软件包让您可以生成图形。我们测试环境的简化图形看起来如下图所示：



附录 B (续)

大数据性能测试方法 — Riak

Riak - 一致性测试

裸机 5 节点集群与虚拟服务器 5 节点集群

	裸机服务器节点	虚拟服务器节点
内核	两个 6 核 Intel 5670 CPU	26 个虚拟计算单元
操作系统	64 位 CENTOS	64 位 CentOS
RAM	36 GB RAM	30 GB RAM
RAID	4 个 300 GB 15K SAS RAID10	4 个 300 GB 网络存储
SAS	1 GB 网络 - 绑定	1 GB 网络