

在当今信息超负荷的世界中管理数据

IBM® Redbooks® 观点出版物，
来自 IBM Academy of
Technology



作者：Dan Wolfson (IBM 杰出的工程师)、
Thomas Pflueger (IBM 杰出的工程师) 以及
Vincent Hsu (IBM 员工)

要点

随着企业数据量的增长，您的企业可以通过应对以下趋势来保持竞争力：

- ▶ 以较低的成本增加内存和设备容量
- ▶ 可分配和重新分配硬件资源，以响应工作负载需求的硬件虚拟化
- ▶ 利用硬件技术革新的软件
- ▶ 结合新兴技术以支持广泛分布的计算的大数据

信息基础架构技术中的新兴趋势

当今信息基础架构的大环境正在持续不断地发生改变。收集的数据量在不断增加，各组织正在寻找新的途径来分析、评估和使用所收集到的信息。此外，硬件发展趋势正在提升新信息源捕获和数据管理自动化的经济效益。

从数据中发掘新的意义以及在企业内使用此洞察的能力也许是塑造信息基础架构最重要的趋势。但是，此趋势将如何继续发展？

物联网继续自动收集大量新型信息。移动设备的数量继续激增，提供对现有和新的应用程序与数据的访问。此外，企业继续采用分析和人工智能来理解并使用如此丰富的信息。系统将进化以适应数据工作负载的持续更改和发展。

由各种类型的软件和硬件来处理和管理如此过量的信息的趋势日渐发展。企业期望这些不同类型的软件和硬件必须协同工作以累计、保留、管理、保护和使用现已成为其命脉的信息。此类系统组合带来了挑战和机遇。当这些系统得到有效组合时，可以为组织提供对信息的可靠、安全、值得信任且快速的访问。

因此，存储和信息管理是许多推动变化的新兴趋势的核心。了解这些趋势是为当今企业获取竞争优势的关键。信息管理功能继续在各种软件和硬件之间发展。这些功能可以更低的价格以及商务计算系统中可负担的更改来提高性能。

尤其在硬件领域，内存设备的价格不断降低，容量不断增加，正在重新塑造存储行业。固态驱动器 (SSD)、大量 RAM 和新的 I/O 技术会影响原始系统性能。传统构建系统构造的样式是按处理器、高速缓存和全局内存 (动态 RAM (DRAM)) 进行构建的，这种传统样式正在受到新趋势的挑战。

此外，系统虚拟化可以改善硬件、软件、网络和存储资源的利用。它可以将物理资源作为虚拟资源池来加以处理，从而对其进行动态分配和再分配，以响应工作负载的需求变化。



使用虚拟化和数据降维的存储效率的趋势

存储虚拟化是管理数据增长的成本中的关键要素。虚拟存储器功能（例如，自动精简供应、高空间效率副本和自动供应）可提供高级别的利用率，在某些应用程序环境中分配级别可能超过 100%。

此外，随着处理器功能的增加，存储控制器可以执行数据计算以挤出冗余。此处理能力可以通过无损压缩算法（例如，Lempel-Zev 1 类 (LZ1)）来实现。也可以在小型数据窗口上通过块或文件去重来写入此数据或者在相同数据上使用两种技术的结合来实现此处理能力。

组织可以针对数据备份和长期归档使用压缩和去重技术。针对在线数据的新数据降维技术可以提供直接在存储控制器上执行变换的能力。

有关数据压缩和去重技术如何运作的更多信息，请参阅 IBM 专家 Tony Pearson 所撰写的文章“IBM Data Footprint Reduction Technology for Data Compression and Deduplication”，该文章位于以下地址：

http://www.ibm.com/systems/storage/resource/technology-topics/data_footprint_reduction_techology.html

另外，企业数据可以在世界上任何位置创建，包括分析人员的工作站、店铺中的收银机、有摄像机监控交通的街角或者供应商的工厂。通过策略驱动的自动化流程来移动所需数据是创新的重要领域。

信息管理软件中的趋势

软件继续利用（并且常常驱动）硬件技术的革新，从数据库的新兴功能到分析系统。

通过实施软件功能来应对新的问题，例如在实时数据流上执行分析的需求。例如，使用预测分析有助于您了解并预测实时操作中的趋势和行为。

互联网规模系统（例如，Google、Yahoo 和 Facebook）的兴起部分源于商品硬件的即时可用性。这些计算机用于大型分布式系统中时，可以支持大量信息的收集和计算。此数据用于支持搜索、将广告目标瞄准特定时刻最适合的用户，或者了解其他用户行为的形式（并从中获益）。

以下趋势已兴起，以支持数据不断增加且多种多样的信息的存储、查询和分析：

- ▶ 大数据将支持大范围分布式计算的新兴技术与推动从各种大量信息源中汲取洞察和意义的现有技术相结合。大数据包含的不仅仅是分析。它代表从多种信息源收集信息的完整的生态系统。它为这些信息做好使用准备、处理信息以从中提取洞察，然后在整个企业中交付此洞察。
- ▶ 对受信任的信息的期望是推动对信息集成和管控的需求，从而提供以下利益：
 - 确保最高质量的信息
 - 通过数据生命周期对数据进行管控
 - 保护所有信息
 - 集成所有数据以提供公共视图
 - 确保单一的知识理解和知识集?
- ▶ 原始信息交付至信息供应链的开头。分析系统的信息供应链模式得到妥善记录。这些模式使用各种技术（例如，抽取，变换和装入 (ETL)）和数据重复来清理、组合信息并将信息装入数据仓库和数据集市。管理信息供应链以在正确的时间将正确的信息交付给正确的消费者，这对于成功的企业是至关重要的。

- ▶ NoSQL 数据库（例如，MongoDB、HBase、Cassandra 和 MemcacheD）放弃了现代关系数据库的丰富性和健全性，以通过使用碎片化来实现广泛的可扩展性和性能。随着移动方式的日趋成熟，术语 NoSQL 已转变为表达不仅仅是 SQL 的意义。此意义确认 SQL 的功能常常仍然是必需的。它还确认可以在现有关系数据库内逐渐适应许多 NoSQL 数据库应对的需求和更简单的方法。
- ▶ 消费者正在寻找更好的途径来优化应用程序工作负载的硬件系统。针对工作负载优化的系统可以改善系统性能和可管理性。伴随此需求，出现了新类型的专家集成系统。这些系统结合了通用系统的灵活性、云计算的弹性以及针对工作负载调优的设备的简单性。

走向高效率的信息基础架构之旅

IT 要求存储器随数据源的增长而继续发展。创新是确保存储系统可以适应此增长并维持可扩展性和可管理性所必需的。存储需求的发展速度比存储技术的成本降低速度更快，这表示即使客户被迫减少成本，IT 成本仍在不断增加。企业必须找到通过技术创新来控制 and 减少成本的途径。

利用来自 IBM 的存储解决方案，您可以从大量数据中生成洞察，从而从根本上改变使用信息的方式。此外，利用 IBM 创新的解决方案，您可以对数量、种类和速度不断增长的数据加以利用，而不会增加复杂性。这些经调整的解决方案可以帮助您为自己的企业选择最佳的技术、服务和融资组合。

考虑适应您环境的存储解决方案时，请记住以下关键的领域：

- ▶ 针对工作负载优化的系统
- ▶ 专家集成系统
- ▶ 存储效率
- ▶ 全局数据访问和移动

针对工作负载优化的系统

实际上，不同的工作负载需要不同的处理、内存和存储器组合。从针对应用程序工作负载优化您的硬件系统开始。针对工作负载优化的系统使用专用硬件来改善系统的性能和可管理性。针对工作负载优化的系统的普及正在通过使用设备结构和硬件虚拟化来驱动专用功能的预集成。

工作负载决定存储器的访问模式、读写的混合、性能目标、创建副本的需求以及数据保护的所需类型。以下类型的工作负载可以提供在线数据访问：

- ▶ 联机事务处理 (OLTP) 和数据库工作负载专注于高事务率和小型记录写入。此类型的工作负载中的数据通常是关键任务数据，存储复制可用于镜像更新，并提供给恢复站点。
- ▶ 业务应用程序工作负载通常包含数据库应用程序，但是 I/O 率较低且写入较少。服务器通常经过虚拟化、复制频率较低并且所使用的数据保护措施成本较使用 OLTP 时更低。
- ▶ 高性能计算 (HPC) 和分析工作负载历史上包含工程应用程序和科学应用程序。但是最近，这些工作负载由业务分析主导。
- ▶ Web、协作和基础架构工作负载通常包含检测来自大量数据或者多个固定内容的副本（例如，电影或歌曲流服务）的洞察。这些应用程序通常部署至虚拟服务器，某些最早的应用程序部署至云环境。

IBM 的存储产品服务组合中涵盖所有主要的工作负载类别：

- ▶ IBM 可以最低的客户机成本提供高性能和最优的容量利用率。
- ▶ IBM 提供全局数据放置和自动化数据放置，包括信息生命周期管理 (ILM) 功能。
- ▶ IBM 提供高级虚拟化技术，此技术将存储器与公有云和私有云的用途相匹配。

专家集成系统

构建和调整信息基础架构是 IT 中最复杂的任务之一。您需要以下系统的正确组合：

- ▶ 服务器，考虑处理器能力、大量内存、并行执行和内部适配器带宽等
- ▶ 网络，考虑每秒输入/输出操作 (IOPS) 的优化或延迟以及所使用的技术等
- ▶ 存储器，考虑 IOPS 或延迟、流式方法或随机存取、只读、50/50 读/写、更改频率、压缩和 SSD 或硬盘驱动器 (HDD) 等

由于存储方面与应用程序相关，因此此注意事项需要大量专业知识。同时，您需要软件来利用可用的硬件功能部件，并且需要确保数据的可用性和完整性。

经验显示，设置成功的分析系统可能需要长达一年，以构建所有必需的决策然后对系统进行调优。专家集成系统通过允许您部署预构建的模式来反映设计和部署系统的经验和专业知识，从而缩短此时间。专家集成系统是功能的构建块。当智能和知识直接构建到系统中时，您的团队无需浪费时间来设计、测试或调优其集成解决方案。而是可以使用新的效率和速度级别创建功能。

使用带有集成专业知识的系统有助于您实现更高的敏捷性，从而适应工作负载峰值的需求并交付新的业务功能。您可以通过整合 IT 资源和提升生产力来增加效率。此外，您可以改善简单性，以简化管理、部署和集成。

利用 IBM PureData™ System，设置集成系统的步骤已由该领域的专家完成执行。IBM 可提供以下经优化的 IBM PureData System 模型：

- ▶ IBM PureData System for Transactions
- ▶ IBM PureData System for Analytics
- ▶ IBM PureData System for Operational Analytics

存储效率

在解决应用程序工作负载之后，请考虑存储器层次结构作为性能敏感型应用程序的重要策略。较快的存储设备比较慢的设备价格更昂贵（按每千兆字节美元价格）。通过使用数量适中的快速存储器和自动化，仅将最活跃的数据放置在更快速的存储设备上，以可承受的成本来获取令人印象深刻的系统性能收益。

存储效率是主要的关注项目，因为技术成本削减（每千兆字节美元价格）无法跟上数据增长的速度。许多新应用程序对成本十分敏感，可能无法以较低的存储成本实现。存储虚拟化是开支管理的关键元素。自动精简供应、高空间效率副本和自动供应的虚拟存储器功能可提供高级别的利用率，在某些应用程序环境中分配级别可能超过 100%。随着处理器功能的增加，存储控制器可以通过压缩算法、通过块或文件去重或者通过在相同数据上执行这些技术的组合，来执行数据计算以挤压出冗余。

IBM 在为块存储设备交付复制功能中领先于整个行业已有一段时间，这些功能现在可以支持距离长达 300 千米的块存储镜像制作。最近，IBM 发布了 IBM Active Cloud Engine™，它是按地域分布的功能。

利用通过 Active Cloud Engine 建立的策略，组织可以在 IBM Scale Out Network Attached Storage (SONAS) 或 IBM Storwize® V7000 实例之间复制文件数据，这些实例由远程用户按需存储在高速缓存中或者分发至参与其中的节点的集合。支持同级至同级 (peer-to-peer) 拓扑和集中星型 (hub-and-spoke) 拓扑，如图 1 中所示。

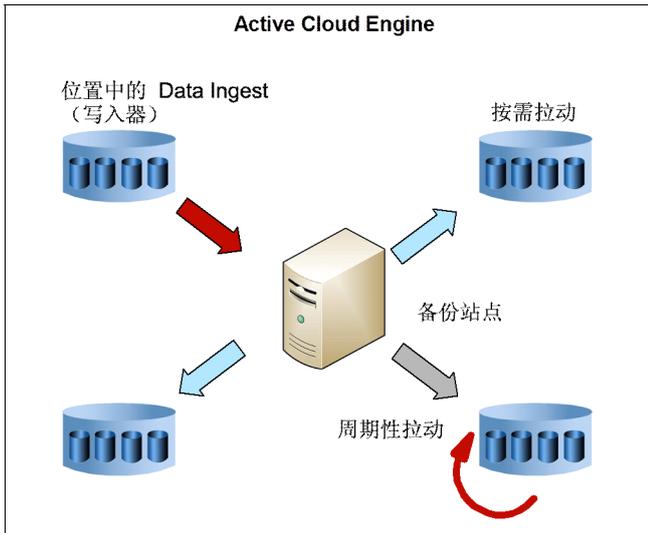


图 1 IBM Active Cloud Engine

当 **Active Cloud Engine** 与本地存储效率优化和虚拟化功能结合时，会成为可创建经优化、有弹性且全局基础架构的强大工具。

IBM 存储基础架构管理软件可以交付以下利益：

- ▶ 它有助于通过集中、简化和自动化存储任务来减少存储环境的管理复杂性。这些任务通常与存储硬件、复制服务、数据保护服务、容量管理和安全合规性报告关联。
- ▶ 它可通过在发生时更快地通知管理员并从更少的控制台管理更多的环境，从而改善服务级别。
- ▶ 它可通过帮助存储管理员以更少的工作管理更多的信息来控制运营成本。

全局数据访问和移动

最后，请记住，由于在世界上任何一个角落都可以创建企业数据，您的数据必须具有全局访问权和移动能力，并且必须通过策略驱动的流程来移动。

业务数据的数量和速度都在加速增长，并且越来越多的数据正在被识别为关键任务数据。与此同时，数据可用性和响应时间预期正在持续提升。部署并运行可靠且高度可扩展的事务数据库系统昂贵且耗时。需要新的方法来交付快速、可靠且可扩展的数据服务。

在当今业务环境中，您的信息管理策略必须支持大数据，以帮助您安全地存储此数据并有效地加以访问。您需要可以过滤来自几乎任何互连设备的大量数据的信息基础架构解决方案。您还需要对动态数据加以分析，以决定是否存储任何数据然后实际上将其与传统数据仓库进行集成。

为何选择 IBM?

企业已准备好针对 IT 采用新的方法。IT 中的发展（例如，虚拟化或云）可有助于推动公司向前发展。但是信息爆炸丝毫没有减速的迹象，任何获益可能都始于和结束于数据中心。要使 IT 在整个业务中提供真实的价值，需要更智慧的计算方法。

随着信息量的增加，查找、管理和适当使用此信息的需求将在整个企业中推动连续不断的创新流。某些创新可能导致现有技术的独一无二的组合，从而提供新的功能。此外，其他创新可能反映技术领域中更基本的增强。

针对存储器和工作负载优化的系统将支持并带领新的经济实惠的信息技术的繁荣。内存内处理、更快且更大的存储器以及更高密度的计算能力将使当前的不可能转变为将来的可能，并带来实际的商业利益。

IBM 信息基础架构解决方案可帮助您跟上大数据的需求，并应对围绕存储效率或数据保护的挑战。使用 IBM 存储解决方案可以引领您的公司改善生产力、服务交付和减少风险，同时精简成本。

可提供更多信息的资源

有关本文中突出显示的概念的更多信息，请参阅以下资源：

- ▶ IBM 信息管理解决方案
<http://www.ibm.com/software/data/>
- ▶ IBM 信息管理解决方案门户网站
<https://www.ibm.com/developerworks/wikis/display/im/Information+Management+Solution+Portal>
- ▶ IBM developerWorks® 的 IBM 信息管理软件的技术资源
<http://www.ibm.com/developerworks/data>
- ▶ IBM 大数据平台
<http://www.ibm.com/software/data/bigdata/enterprise.html>
- ▶ 针对大数据的信息集成
<http://www.ibm.com/software/data/infosphere/information-integration-big-data/index.html>
- ▶ 智慧的 IT 方法：IBM PureSystems™
<http://www.ibm.com/ibm/puresystems/us/en/op-ad.html>
- ▶ IBM PureApplication™ System
http://www.ibm.com/ibm/puresystems/us/en/pf_pureapplication.html
- ▶ IBM PureData System
http://www.ibm.com/ibm/puresystems/us/en/pf_puredata.html
- ▶ IBM PureSystems 概述 , TIPS0892
<http://www.redbooks.ibm.com/abstracts/tips0892.html?Open>

声明

本信息是针对在美国提供的产品和服务编写的。

IBM 可能在其他国家或地区不提供本文中讨论的产品、服务或功能特性。有关您当前所在区域的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或暗示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务，都可以代替 IBM 产品、程序或服务。但是，评估和验证任何非 IBM 产品、程序或服务，则由用户自行负责。

IBM 公司可能已拥有或正在申请与本文档内容有关的所有专利。提供本文档并未授予用户使用这些专利的任何许可。您可以用书面方式将许可查询寄往：IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

本条款不适用英国或任何这样的条款与当地法律不一致的国家或地区：International Business Machines Corporation “按现状”提供本出版物，不附有任何种类的（无论是明示的还是暗示的）保证，包括但不限于暗示的有关非侵权、适销和适用于某种特定用途的保证。某些国家或地区在某些交易中不允许免除明示或暗示的保证。因此本条款可能不适用于您。

本信息中可能包含技术方面不够准确的地方或印刷错误。此处的信息将定期更改；这些更改将编入本资料的新版本中。IBM 可以随时对本资料中描述的产品和/或程序进行改进和/或更改，而不另行通知。

本信息中对非 IBM Web 站点的任何引用都只是为了方便起见才提供的，不以任何方式充当对那些 Web 站点的保证。那些 Web 站点中的资料不是 IBM 产品资料的一部分，使用那些 Web 站点带来的风险将由您自行承担。

IBM 可以按它认为适当的任何方式使用或分发您所提供的任何信息而无需对您承担任何责任。

涉及非 IBM 产品的信息可从这些产品的供应商、其出版说明或其他可公开获得的资料中获取。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

本信息包含在日常业务操作中使用的数据和报告的示例。为了尽可能完整地说明这些示例，示例中可能会包括个人、公司、品牌和产品的名称。所有这些名字都是虚构的，若现实生活中实际业务企业使用的名字和地址与此相似，纯属巧合。

此处包含的任何性能数据都是在受控环境中测得的。因此，在其他操作环境中获得的数据可能会有明显的不同。有些测量可能是在开发级的系统上进行的，因此不保证与一般可用系统上进行的测量结果相同。此外，有些测量是通过推算而估计的，实际结果可能会有所不同。本文档的用户应当验证其特定环境的适用数据。

版权许可：

本信息包括源语言形式的样本应用程序。这些样本说明不同操作平台上的编程方法。如果是为按照在编写样本程序的操作平台上的应用程序编程接口 (API) 进行应用程序的开发、使用、经销或分发为目的，您可以任何形式对这些样本程序进行复制、修改、分发，而无须向 IBM 付费。这些示例并未在所有条件下作全面测试。因此，IBM 不能担保或暗示这些程序的可靠性、可维护性或功能。

本文档“REDP-4945-00”创建或更新于 November 26, 2013。



商标

IBM、IBM 徽标和 ibm.com 是 International Business Machines Corporation 在美国和 / 或其他国家或地区的商标或注册商标。这些术语和其他 IBM 已注册商标的术语在本信息中首次出现时都使用适当的符号（或）加以标记，以表示在本信息发布时由 IBM 在美国注册或拥有的普通法商标。这些商标也可能是其他国家或地区的注册商标或普通法商标。以下 Web 站点上的“Copyright and trademark information”部分中包含了 IBM 商标的最新列表：ibm.com/legal/copytrade.shtml



以下术语是 International Business Machines Corporation 在美国和 / 或其他国家或地区的商标：

Active Cloud Engine™
developerWorks®

IBM®
PureApplication™
PureData™

PureSystems™
Redbooks®

Redbooks (徽标)  ?®
Storwize®

以下术语是其他公司的商标：

其他公司、产品或服务名称可能是其他公司的商标或服务标记。