企业数据管理解决方案
2008 年 9 月

IBM Information Management software

数据隐私最佳实践：立即行动起来！

目录

执行摘要

在技术驱动的世界中，数据破坏不仅常见，而且会导致惨重的损失。侵犯隐私统计数据表明与数据破坏的发生频率及相关成本正在不断增加，证明行业中的组织需要采取更实用的方法来保护信息，尤其是在很容易受到攻击的非生产（开发、测试和培训）环境中。当非生产环境中的数据被用于开发和测试活动中、被临时员工访问或提供给外部人员时，更容易被破坏。

在史无前例的安全性破坏中，确保机密信息受到保护的最佳方法是开发并实现全面的隐私和安全性战略。一旦组织意识到保护隐私不再是可选项，大多数人会问，“我们应该从哪里开始？”、“要求是什么？”以及“我们的组织应该采取什么措施来实施企业数据隐私和安全性战略？”

本白皮书说明了开发隐私战略和实施第一个数据隐私项目时需要考虑的步骤。使用成熟的数据屏蔽技术，例如 IBM® Optim™ Data Privacy Solution 中提供的技术，可以帮助您的组织在隐私保护中实现最佳实践并使您的隐私项目从头到尾顺利进行。最后，学习一个大型零售公司如何实现 Optim 来开发最佳实践战略和成功的隐私项目，同时克服在企业范围内实施隐私项目时出现的许多难题。

为何保护隐私非常重用？了解事实。

随着黑客的猖獗和身份盗窃的增多，数据隐私破坏正在以我们以前从未想到过的方式影响我们的个人和商业活动。以下数字就是一个证明。根据 Privacy Rights Clearing House，自 2005 年 1 月以来，仅在美国，包含安全性破坏中涉及的敏感个人信息的记录总数是 230,441,730。[1] 该数字每天都在增加。

在这个技术年代，盗窃活动所针对的许多机密信息都位于驱动企业业务计划的业务应用和计算机系统中。如果没有适当的措施来保护隐私和防止破坏，下一个被侵袭的公司就可能是您的公司。

数据隐私首先保护不同类型的敏感应用数据，无论该数据位于整个组织中，还是在生产和非生产（开发测试、培训）环境中。但是，许多公司正在意识到在生产环境中保护隐私的方法对于管理非生产环境中的数据可能不实用或不适当。

哪些区域最易受到攻击？了解风险。

在生产环境与非生产环境中保护隐私的方法应该有所不同。例如，大多数生产环境已经建立了安全性和访问限制来防止免受数据破坏。

标准安全性措施可以在网络、应用和数据库级别应用。通过实现多因素认证方案，例如密钥令牌或者甚至人体生物学特征，可以扩展物理登录访问控制。但是，这些保护措施不能在每个环境中轻松地复制。在生产

环境中保护数据的方法可能无法满足保护非生产环境的独特要求，在非生产环境中，开发人员、测试人员和培训人员需要对实际数据进行的访问更多，而不是更少。

Ponemon Institute and Compuware 在 2007 年进行的一个调查显示许多组织使用实际数据进行测试和开发（69% 的被调查者使用实际数据进行应用测试，62% 的被调查者使用实际数据进行软件开发）。[2] 例如，如果公司正在使用实际数据（例如客户、员工和供应商记录、消费者清单及支付卡、业务合作伙伴和其他类型的机密信息）进行开发和测试，则最终会增加数据曝露的风险。[3] 许多被调查的公司指出软件开发中使用的动态数据将不会被保护。

该调查还指出大约一半的被调查公司将他们的应用测试外包出去并共享动态数据。大多数时候，这些公司实际没有办法知道外包的测试环境中使用的实际数据是否已遭到破坏。惟一可行的解决方案是通过去标识（de-identification）方法来伪装数据。

去标识非生产环境中的数据是系统地删除、屏蔽或转换可能用于标识个体的数据元素的过程。通过数据去标识，开发人员和测试人员可以使用实际数据并生成有效的结果，同时仍然遵从隐私保护规则。去标识之后的数据通常可以用于非生产环境，并能够确保即使数据被盗、曝露或丢失，任何人也将无法使用该数据。

计划数据隐私项目时的考虑事项

行业隐私法规和条例使保护数据隐私不再仅仅是一个可选项。所以公司如何确保他们发送到海外的信息、保存在笔记本中的信息或用于内部开发和测试活动的信息保持受保护状态？为了帮助保护委托给他们

的敏感信息，公司必须考虑在他们的整体隐私和安全性战略中优先实施数据去标识实践。

公司还应根据隐私项目确定要使用的开发方法的类型。基准项目要求和假设在整个项目过程中可能会发生变化。因为隐私项目的规模和复杂性以及遵从性压力，所以组织开发一种在整个过程中具有一定灵活性的方法非常重要。那么开始一个重要的隐私项目时必须要考虑什么呢？表 1 为实现成功的隐私项目提供了 6 个最佳实践。

表 1. 管理成功的隐私项目

步骤	描述
组织	成立一个跨职能的隐私小组来帮助指导工作。
定义要求	定义隐私项目的要求并标识必须保护的应用/硬件/软件/数据的类型。
执行数据盘点	对数据存储、流、流程、依赖性和业务规则进行分析和编目，帮助简化隐私项目的范围。
选择解决方案	选择并实施数据隐私解决方案，该方案可以提供保护所有环境中的隐私所需的技术。
测试、测试、测试	开发项目的原型和方法，然后测试该原型来进行验证。
扩展范围	扩展数据隐私项目，以便包含组织中的其他应用。

这些步骤从更高层面概述了如何管理数据隐私项目。让我们更详细地讲述一下每个步骤。

步骤 1 —— 组织。创建、计划和管理数据隐私项目的第一个步骤是进行组织。通过建立一些基本隐私指令和指导方针，可以帮助您形象化项目的范围并保持正确方向。要管理您的项目并确保执行指令，需要指定一个隐私项目领导人或小组。该小组应该包括： 应用和业务所有者（直接使用应用）；合规经理（确保您的公司遵从隐私规则和条例）；IT 经理（其小组将实现技术来支持您的隐私计划）；业务经理和 QA 经理（如果自动化流程和测试案例全部进行了修改，可能需要他们的参与）；以及直接影响隐私项目的其他任何人。跨职能的隐私小组将要求和激发部门间协作，从而可以表现所有涉及的区域。

在整个项目期间，隐私小组将在整个过程中互相协作来做出项目决策和找到问题的答案，帮助保持项目的焦点和目标状态。建立一个隐私小组，当您在整个企业中扩展数据隐私计划时，还可以为任何其他项目提供支持。

步骤 2 —— 定义要求。一旦建立了基本隐私指令和指导方针，并且已经建立了隐私小组来引导项目，下一步就是了解和定义您组织的隐私和去标识要求。首先，确定有关国家或行业隐私条例的遵从性目标，这些条例包括健康保险流通和责任法案 (Health Insurance Portability and Accountability Act, HIPAA)、支付卡行业数据安全标准 (Payment Card Industry Data Security Standard, PCI DSS) 以及许多其他条例。每个条例都包含您的隐私项目中必须考虑的特定要求。

接下来，编译一系列目标应用，包括物理位置、业务区域、支持数据库和硬件平台。管理或存储机密客户、员工和公司业务数据的应用必须具有高优先级。了解每个应用的用途将可以帮助您确定在程序区域中必须屏蔽的不同数据类型以及原因。然后，确定需要去标识的特定类型的数据并估计在应用、数据库和操作环境中需要屏蔽和传播的数据量。

最后，当屏蔽用于非生产环境的数据时，需要知道屏蔽的替换数据的子字符串 准确反映原始实际数据中应用逻辑的程度。每种类型的数据可能需要考虑不同的屏蔽要求。例如，如果生产数据包含美国社会安全号码，则屏蔽的替换数据是否必须与社会安全号码格式相匹配？

基本上，去标识数据是保护隐私和支持遵从性计划的最有效方法。去标识机密数据的功能必须允许您保护隐私，同时仍然提供必要的“实际”数据来用于开发、测试、培训环境或用于其他合法业务目的。

步骤 3 —— 执行数据盘点。此步骤涉及一些“数据辨析”，因为您需要分析元数据 （有关数据的信息，使数据易于了解、使用和共享）。通过浏览应用来确定和验证内容，将可以更可靠地描述应用中所管理的数据，特别是在遗留数据存储或连续文件中，其中可以“隐藏”个人身份数据。

分析应用组合中的数据流还可以帮助您将数据分成组和层次。对层次结构上层中的数据进行的任何更改都可以级联到下面的层，从而可以缩短项目的时间，减小其规模并降低复杂性。在这样的特定层调查隐私要求可以帮助简化项目，帮助您创建更实际的项目指南来满足最终目标。

步骤 4 —— 选择解决方案。在评估数据隐私技术时，需要查找满足您的要求（如步骤 2 和步骤 3 中所定义）的解决方案。另外，还要计划更改。随着新的隐私法规的制定以及当前法规的执行，您必须能够适应更改。同样地，也将通过每个新的升级和增强来对您的业务应用进行更改。为了跟上这些更改，您的数据隐私解决方案必须能够扩展并提供灵活的功能，使您能够根据需要修改数据屏蔽和隐私保护示例 。

例如，IBM® Optim™ 数据隐私解决方案为去标识应用数据提供了全面的可以在非生产环境中有效使用的功能。通过提供可扩展性和灵活性，Optim 的数据屏蔽技术是一致的且可重复的，可以跨应用、数据库、操作系统和硬件平台进行部署来满足当前和将来需要。

Optim 可以在企业中更广泛、深入地提供数据屏蔽，它不是设计用来解决一个集中的隐私问题的单点解决方案。作为一个解决方案，您可以将 Optim 合并到您的现有和正在进行的业务流程中。

为了使组织能够满足设置最复杂的数据隐私要求，Optim 提供了有效数据屏蔽的以下基础组件：

- **保持应用逻辑。**Optim 的应用感知数据屏蔽功能可以准确了解、捕获和处理数据元素，从而使屏蔽的数据将不验证应用逻辑。例如，姓氏使用随机但有效的姓氏来替换，而不是使用无意义的文本字符串。数字字段保持适当的结构和模式。如果诊断代码是四位数，范围从值 0001 到 1000，则在应用测试环境中，屏蔽值 2000 将是无效的。校验和保持有效性，所以功能测试可以通过应用有效性检查。Optim 还可以在整个测试数据库中一致地传播所有屏蔽的数据元素，以及传播到其他相关应用和数据库。

  例如，Direct Response Marketing Company, Inc. 正在测试它的订单履行系统，需要去标识客户姓名，以便确保安全的测试。Optim 的 Random Lookup 功能允许该公司根据预先定义的“客户信息”表任意生成名字和姓氏。“Lucille Ball”每次出现时都将变为“Elena Wu”等等。基本上，屏蔽将是可重复且可预测的，从而相同的更改可以根据需要在整个非生产环境中一致地显示，来满足您的要求。

- **屏蔽重要数据元素。**Optim 的上下文感知的、预先打包的数据屏蔽示例可以去标识重要数据元素，提供许多公认的数据屏蔽技术，可以用来去标识许多类型的敏感信息。例如，可以屏蔽出生日期来准确反映个人的正确年龄。同样地，可以屏蔽银行帐户信息、国家标识号（例如加拿大的社会保险号或意大利的税号）、福利信息等等。

通过预先打包的 Transformation Library™ 示例，可以准确地屏蔽复杂数据元素，例如社会安全号、支付卡号和电子邮件地址。内置的查找表支持屏蔽姓名和地址。您还可以合特定于具体位置的数据转换示例，这些示例整合了多个相关应用和数据库的处理逻辑并在支持复杂数据屏蔽要求方面提供了高大的灵活性和创造性。

例如，Green Bill Bank 的帐号格式为“999-9999”，其中前三位表示帐户类型（核算、储蓄或货币交易），最后四位表示客户标识号。出于测试目的，这些帐号必须进行屏蔽。Optim 可以利用实际帐号的前 3 位，然后生成连续的 4 位数字来替换实际帐号中的最后 4 位。例如，“001-4750”将变为“001-1000”等等。结果是一个将保持银行的有效帐号格式的虚构帐号。

- **准确地传播屏蔽的数据元素。**Optim 的持久屏蔽功能生成源数据列的已转换的替换值，并在应用、数据库、操作系统和硬件平台中一致而准确地传播这些替换值。持久数据屏蔽功能可以确保可伸缩性，从而可以跨多个非生产环境保护隐私。对于帮助保持数据的引用完整性（甚至在数据进行屏蔽后），将屏蔽的主键值传播到所有相关表是非常必要的，这样能够保持相关数据的完整子字符串完好无缺（参见图 1）。

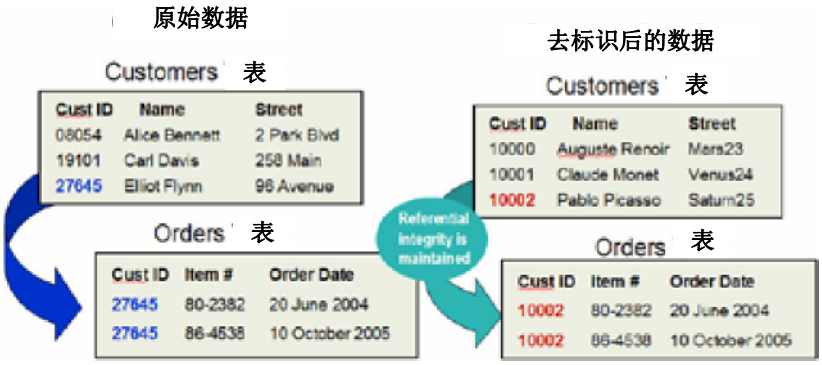

图 1. Optim 中可以帮助保持数据的引用完整性的重要传播功能

图 1 显示了一个包含两个相关表的简单示例，其中 Customers 表是 Orders 表的父级，它的主键列 Cust_ID 是一个 5 位的数值。Cust_ID、Name 和 Street 列将进行屏蔽。姓名“Elliot Flynn”已经屏蔽为“Pablo Picasso”。使用连续屏蔽技术将 Elliot Flynn 的原始 Cust_ID 从 27645 转换为 10002。屏蔽的 Cust_ID 值将从 Customers 表传播到所有相关表，且测试数据库中 Customers 和 Orders 表之间的重要关系保持不变。

一旦已经设计了隐私战略并已经选择了解决方案，接下来的步骤就是实现。Optim 的独特应用感知和上下文感知技术以及它的传播功能提供了成功实现所需的强大数据屏蔽资源。

步骤 5 —— 测试、测试、测试。一旦已经开发了数据隐私战略，下一步就是确定“它是否起作用？” 需要为一个或一组选择的应用构建原型屏蔽方案，

以确保您的屏蔽技术和测试活动将按预期执行。此方案应该与您的应用测试要求相关联。需要验证是否已经有效地设计了隐私战略来满足定义的要求。例如，将原始数据与去标识的数据进行比较，以确保已经应用了您所做的更改。最后，验证您的隐私战略适合于应用和非生产环境。

然后，测试该原型，从而验证是否应用了适当的屏蔽技术。您是否已经确保包含敏感数据的所有区域都进行了屏蔽？所有敏感数据是否将在应用和测试环境中被屏蔽？您的隐私战略是否将使您保持遵从行业和联邦隐私法律和条例？

步骤 6 —— 扩展范围。后实现（Post-implementation）可以扩展数据隐私项目来包含您组织中的其他应用集。主动保护您组织中的隐私是非常重要的。坚持在您组织中的其他存储或管理敏感信息的区域中实施数据隐私战略。这些区域是否正变为数据破坏的目标？

保护隐私的需要不限于您公司的一个区域。扩展隐私项目的范围来包括包含敏感信息的各种区域将是一种保护您的信息资产的更有效方法。

保护隐私提供了一种管理敏感数据并使业务受益的最佳实践。数据去标识为任务关键型应用测试提供了“安全沙箱”，帮助公司保护隐私以及维护客户忠诚度。Optim 提供了灵活的功能，可以进行扩展来支持当前和将来的数据隐私要求，帮助使数据去标识成为您的整体数据隐私战略的一部分。使用 Optim，组织将处于有利形势来满足本地、州/省、国家及国际的隐私条例以及法律和行业要求。

要考虑的其他事项

隐私项目（ 包括隐私项目小组、流程、环境和规则）将因场所的不同而有所变化，并不是每个项目都是相同的。确保有效的数据保护和安全性可能取决于许多因素。下面是您开始自己的数据隐私项目时可能要考虑的一些其他事项：

- 考虑将隐私保护流程与测试流程分离。在类似生产的测试环境中实现隐私流程将有助于提高您组织中可以获得的最高安全水平。
- 考虑在项目小组中执行严格的“职责分离”。如果仅基于需求来访问去标识的数据，则某个人反转屏蔽过程的风险就会降低。
- 考虑通过轻微修改您的数据隐私“处方”来采取措施保护内部安全性。在将数据移到生产环境中之前对其进行轻微更改可以增加额外的保护层。
- 考虑第三方审核来证明您的隐私流程能够防篡改，例如，您能够执行职责分离，您的路线图正确，以及您的时间表提供了足够的“缓冲时期”来填补该流程中的任何空白。

Optim 实现 – 用户视角

作为一个成长中的折扣零售商，Marzan Corporation 在整个美国和英国拥有并经营着多家连锁零售店。随着它 1979 年在新泽西创办了第一家店， Marzan 这些年已经成为有名的提供日常用品的“廉价但别致”的一站式商店。从家用货品到服装以及电子到汽车设备，Marzan 零售店为客户提供了一个便利、时髦、有趣且廉价的购物场所。

各种应用组合支持业务。为了帮助支持和推动其正在进行的业务以及支持其供应商关系，Marzan 依赖于 PeopleSoft® Enterprise 供应链应用，在内部称为 IZZI，该应用管理供应商订单和库存活动。IZZI 包含供应商名称和 ID 等敏感信息，运行在 Oracle® 数据库上。

Marzan 还使用 Siebel® CRM 应用（称为 Jasper）来处理客户订单。Jasper 在 DB2 开放系统环境中运行，包含敏感客户订单信息，例如客户姓名、地址、电话号码和支付卡号。其中某些信息对于帮助 Marzan 进行促销活动来扩展其客户基础是非常重要的。

最后，Marzan 使用称为 Centz 的自定义财务应用，该应用运行在 IBM System z™ 上。该客户开单应用捕获并存储 Marzan 信用卡应用信息以及客户开单数据，可能包括社会安全号、姓名和地址。

数据隐私挑战。提供高质量的产品和客户服务对于 Marzan 始终具有高优先级。为了帮助提供更好的客户服务，Marzan 希望为它的客户提供改进的功能来在线访问帐户信息和执行在线订单输入。实现此目标需要改进现有 Jasper 和 Centz 应用。每个应用都包含从在线输入收集的敏感数据，包括支付卡号、社会安全号、姓名、街道地址和电话号码。增强功能将允许更完善的在线、面向客户的帐户功能、可访问性和安全性。

而且，因为 IZZI 应用存储供应商名称和 ID 等供应商订单信息，Marzan 要求该信息应该进行屏蔽，以确保该信息可以安全地用于测试和开发活动。Marzan 知道它将必须跨数据库和应用执行测试，所以它还希望确保数据的引用完整性保持不变。

最后，作为大型零售连锁店，Marzan 必须保持遵从 PCI DSS (Payment Card Industry Data Security Standard) 条例。PCI DSS 要求处理支付卡的大型零售商和公司屏蔽应用测试环境中使用的个人身份客户信息。Marzan 必须保证来自其 Centz 和 Jasper 应用的用于开发和测试目的的任何数据都进行了屏蔽且可以安全用于非生产环境。

探求解决方案。Marzan 需要一个可以满足企业中所有隐私要求的解决方案。为了支持和实现其整体数据隐私计划，Marzan 决定购买并实现 IBM Optim Data Privacy Solution。购买企业数据管理屏蔽解决方案的决策是由一个决策小组提出的，该小组包含应用开发人员和测试人员、IZZI、Jasper 和 Centz 用户以及隐私专业人员。

Marzan 建立了一个隐私小组来引导多方面的项目。该小组希望主动保护客户信息。Optim 具有保护 Marzan 企业中敏感信息的功能。Optim 的用于大型机和开发系统应用数据的独特屏蔽和转换功能可以帮助保护应用、数据库、操作系统和硬件平台中的隐私。

成功的实现带来了企业效益。成功实现 Optim 后，Marzan 的去标识项目小组部署了其努力呈现的隐私项目方法。从 IZZI 开始，Marzan 采取措施在企业中去标识非生产环境中的敏感数据。

作为一种在 IZZI 中去标识供应商名称的方法，Marzan 使用了 Optim 的查找技术来使用替代值转换数据。通过将相应屏蔽值返回目标数据列，该小组可以屏蔽源数据列中的某个值，从而将实际供应商名称转换为虚构名称以用于测试和开发目的。Optim 的查找表确保 Acme Pencil Company 的“Dave Acme”将在非生产环境中去标识为“Michael Craft”。

IZZI 中的供应商 ID 长度为 6 个字符，第一个为基于供应商名称第一个字母的字母字符，第二个为基于供应商所在城市的字母字符。最后 4 位是数字，必须在 1000 到 6999 范围内。所以位于 AZ 的 Tucson 的 Acme Pencil Company 将具有类似于 AT1453 的供应商 ID。Optim 的应用感知功能确保屏蔽的供应商 ID 保持应用逻辑并返回类似于 CD2047 的屏蔽供应商 ID，而不是类似于 CD8945。

接下来，Marzan 对 Jasper、Siebel CRM 应用应用了去标识技。Optim 的上下文感知数据屏蔽示例在应用测试和开发环境中去标识了重要数据元素，例如社会安全号、信用卡号和出生日期。为了执行准确的开发和测试活动，Marzan 的开发人员需要处理真实的 16 位信用卡号。Optim 的智能屏蔽功能生成了上下文有效但被屏蔽了的信用卡号。发生数据破坏时，屏蔽呈现对盗窃者无用的信号卡号，但是确保它们可以有效用于非生产环境（参见图 2）。

图 2. Optim 根据发行者的格式要求生成有效、惟一但进行了去标识的支付卡号。

类似的上下文感知屏蔽技术也应用到了 Centz 财务应用中。该应用捕获并存储客户开单信息，所以 Marzan 使用了 Optim 的内置查找表来屏蔽 Centz 中存储的姓名和地址。所以一旦进行了去标识，姓名“Beth K. Smith”的所有实例都将变为“Claire P. Hamill”。Optim 对这些更改进行了相应传播，以保持引用完整性并确保 Marzan 的非生产环境中的隐私保护。

最后，为了对其在线客户帐户访问程序进行应用改进，Marzan 需要创建联合测试和开发环境。Optim 提供了联合访问功能，允许开发人员在一个流程中提取和屏蔽各种数据源中的适当测试数据。Optim 的子字符串功能提供了自动且可重复的功能，用于处理来自 IZZI 和 Jasper 应用的联合提取数据。

使用 Optim，Marzan 为其测试和开发环境创建了实际且规模合适的应用数据子字符串。Optim 屏蔽了敏感客户信息，例如客户姓名、地址、电话号码、支付卡号和付款历史记录明细。去标识的数

据可以用于测试和开发环境。然后，这些数据将在非生产环境中准确地传播，同时保持引用完整性来支持可靠测试。

Optim 的用于保护开发和测试环境中的客户信息的功能，还有助于满足规定的 PCI DSS 要求。因为个人客户和支付卡信息位于 Marzan 的供应链应用中，所以 Optim 提供了屏蔽机密数据的功能。结果，Marzan 降低了可能导致罚款的法律风险并减少了对客户忠诚度以及公司努力建立的信任的丧失。

通过使用子字符串、随机或连续数字替换方法、算术表达式、日期期限和其他技术，Marzan 使用上下文准确但虚构的数据替换了实际客户数据，从而生成准确的测试结果。该数据可安全地用于非生产环境，但是对于盗窃者或黑客没有任何帮助。

基本上，Optim 帮助减少了与潜在隐私破坏相关联的风险和成本，这保持并加强了 Marzan 的第一流的声誉。Marzan 的客户和业务用户正在受益于更可靠且功能丰富的应用。更重要的是，Marzan 已经保持了其高水平的客户服务，正在获得更多的收入机会带来的效益。

关于 IBM Optim

IBM® Optim™ 企业数据管理解决方案关注关键的业务问题，比如数据增长管理、数据隐私遵从性、测试数据管理、电子发现、应用升级、迁移和退役。Optim 使应用数据管理与业务目标保持，帮助优化性能、降低风险并控制成本，同时交付在不同的企业应用、数据库和平台中伸缩的能力。如今，Optim 正在帮助全球各行业的企业利用在其生命周期的每个阶段管理企业应用数据的能力，发挥其企业应用和数据库的商业价值。

更多信息

要了解 IBM Optim 企业数据管理解决方案的更多信息，请联系 IBM 销售代表，或者访问：http://www-01.ibm.com/software/cn/data/data-management/optim-solutions/。

IBM Software Group
111 Campus Drive
Princeton, NJ
08540-6400
USA

www.optimsolution.com

在美国印刷

2008 年 9 月

[1] *Privacy Rights Clearing House，http://www.privacyrights. org/ar/ChronDataBreaches.htm#2008。*

[2] *Compuware and The Ponemon Institute LLC。“测试数据的不安全性：看不见的危机”。Compuware .com。2007 年 12 月：第 3 页。*

[3] *Ibid，第 4 页。*

TAKE BACK CONTROL WITH Information Management

IMW14072-SEN-00