企业数据管理解决方案
2008 年 11 月

IBM Information Management software

# 填补 SAP 应用的数据隐私空白

目录



## 执行摘要

本白皮书说明为什么保护敏感信息和确保隐私性已经成为人们优先考虑的问题。有关信息偷窃和身份盗窃事件的发生频率日益增长的新闻已经将关注重点集中于侵犯隐私及其后果方面。为了应对这些问题，全球已经制订了各种数据隐私法规。尽管这些法规的细节可能各不相同，但如果不能确保数据隐私的遵从性，则可能导致数百万美元的罚款甚至监禁。相关公司还要经受失去客户忠诚度以及毁坏品牌价值的风险。这种影响非常严重，足以使公司破产倒闭。

很多公司依靠关键的 SAP® 应用来支持日程业务运营，因此，确保隐私性以及保护应用数据（无论这些数据位于何处）是至关重要的。但是，在生产环境用来保护数据的相同方法可能无法满足非生产（开发、测试和培训）环境的独特要求。IT 组织如何保护敏感数据，包括员工和客户信息以及企业机密性数据和知识产权？行业分析师建议对数据进行“去标识化处理”或屏蔽，并以此作为保护隐私的最佳实践。但是，选择数据隐私解决方案都有哪些要求呢？

理想的数据隐私解决方案必须提供必要的数据屏蔽技术，以满足最简单和最复杂的隐私要求。屏蔽技术生成的结果必须能反应应用逻辑，还能保证数据的完整性。为了帮助您支持保护数据隐私遵从性的要求，本白皮书讲解面向 SAP® 应用的 IBM Optim™ 数据隐私解决方案提供的全面数据屏蔽技术。

- 应用感知屏蔽功能有助于确保屏蔽数据（如姓名和街道地址）具有与原始信息类似的表达。

- 具有上下文感知功能的预打包数据屏蔽例程可以轻松地对社会保险号、工资数据和电子邮件地址等数据元素进行去标识化处理。
- 持久屏蔽功能能够跨应用、数据库、操作系统和硬件平台一致地传播屏蔽替换值。

通过 Optim，公司可以采用某种方式对数据进行去标识化处理，该方式不仅可在开发、测试和培训环境中有效使用，同时还能够保护数据隐私。

**数据隐私为什么成为了优先考虑的问题？**

信息爆炸已经使访问公共和个人信息成为日常生活的一部分。SAP 应用通常出于合法目的而收集这些信息；但是，由于 Internet 和信息系统的互连性质，敏感数据经常遭到盗窃和滥用。

盗窃身份、侵犯隐私以及对敏感信息进行欺诈性访问等消息不断成为新闻。随着更多的公司认识到数据隐私的需要，并且客户要求增加保护，政府机构正在通过越来越严厉的法律和法规来确保安全。每个公司都必须负责保护机密性员工信息、企业商业智能和敏感的客户数据，以遵守信息控制法规，赢得客户和业务合作伙伴的信任。

**70% 以上的数据泄露实际上源于内部的缺陷。**

**数据隐私遵从性会影响你的业务。**保护数据隐私一项关键商业举措。70% 以上的数据泄露实际上源于内部的缺陷。[1] 这方面的例子包括滥用支付卡号和其他敏感信息的员工，以及那些将机密性数据保存到笔记本电脑上随后其电脑又被人偷走的员工。此外，将应用数据外包到离岸处理环境使人们很难控制对在其他方面不安全的敏感数据的访问，也很难遵守“安全港”指令。

赌注是高昂的。每发生一次事故，企业及其高管都可能面临高达 50 万美元的罚款，甚至可能使当事人锒铛入狱。巨额罚款只是一个说明组织如何受到伤害的例子。其他负面影响包括由于投资者担心而导致股票价格下降，以及由于数据泄露而导致企业声誉受损。不可挽回的品牌损害使公司在公众眼里成为不可信任的公司。如果不采取步骤来保护敏感信息，则公司不仅要经受失去客户和收入的风险，而且还要经受破产的风险。如果失去客户不会对您的业务造成财务影响，请考虑一下，调查泄露事件可能耗费数百万美元。

## 填补非生产环境中的数据隐私空白

根据应用的行业、操作和类型的不同，很多生产和非生产数据库都将处理敏感信息。困难在于在满足业务需要并确保在“需要知道”的基础上管理数据的同时，提供适当的保护。

大多数公司管理着 SAP 应用的多个生产实例。例如，已经实现 SAP HCM 的公司，可能部署一个应用实例来支持它的北美业务运营，部署另外一个应用实例来支持它在欧洲，中东和非洲的业务，部署第三个应用实例来支持其在亚太地区的业务。为了支持应用的开发、测试、培训、备份和其他活动，一个站点可能为每个实例管理 3 到 30 个副本，并包含来自源系统的机密性数据的准确副本。

通过在其 SAP 生产事务处理环境中保护和限制对基础数据的访问，公司可以保护个人信息。严格的控件和仔细设计的界面提供了一个托管视图。遗憾的是，一旦将个人数据复制到测试和开发环境，要保护这些数据就变得不那么容易了。与此同时，开发人员和测试人员对与应用数据交互具有独特的要求。具体说来，他们需要访问有效数据，以便准确地测试和部署他们的 SAP 应用。

*“人力资源和 IT 公司应该避免拿真实的个人数据进行测试，这能够把员工的安全、隐私和机密性至于危险境地。”*

在最近的一项调查中，Gartner 公司对将个人数据用于测试给出了警告，特别是对于那些包含有可能用于身份欺诈的数据的人力资源和工资管理系统。“人力资源和 IT 公司应当避免把真实个人数据用于测试，这足以危害员工的安全、隐私和机密性，而且根据欧盟数据保护法规，此类使用会被认作是违法行为。”[2]

您现在可能会问：“非生产环境确实需要包含生产数据吗？”答案否定的。Gartner 公司和其他的行业分析师一致认为，作为最佳实践，屏蔽数据或对数据进行去标识化处理是一种可行的方法。在非生产环境中对数据进行去标识化处理只是系统地删除、屏蔽或转换可能被用于识别某个个人的数据元素的过程。“使用杂乱无效的数据，能够避免把个人数据置于危险境地。”[3]

## 选择全面的数据隐私解决方案

保护数据隐私不再是一种可选的解决方案——它已经成为明确的法律条文！SAP 客户站点必须制订相应的步骤来跨非生产环境管理这些数据，并且仍然遵守数据隐私法规。高效的隐私保护策略能够确保个人信息的机密性，并且能够跨您的非生产数据库环境改进安全性。但是，您应该在企业数据隐私解决方案中寻找哪些功能呢？

作为公认的最佳实践，对数据进行去标识化处理提供了保护隐私和支持遵从性举措的最高效方式。对机密性数据进行去标识化处理的功能必须使您可以保护隐私，同时仍然能够提供必要的“真实”数据以供在开发、测试、培训或其他合法业务活动中使用。寻找能够提供下列功能的数据隐私解决方案：

- **全面的数据屏蔽技术。**在使用技术屏蔽一些数据的同时，其它诸如银行代码和帐号之类的数据必须是虚拟的而且保持其在上下文里有效。理想的数据隐私解决方案必须提供多种易于使用的屏蔽技术。一些最简单的技术可以屏蔽字符或数值数据，或生成随机或顺序数字，而更高级的屏蔽例程可用于支持复杂的数据隐私要求。

- **支持 SAP 应用逻辑。**数据屏蔽技术必须遵守应用逻辑，并且对查看的人有意义，也就是说，屏蔽数据应该类似于原始信息。数值字段应该保留适当的结构和模式，并且必须保持在许可值范围内，以便功能测试通过所有应用有效性检查。

- **支持业务上下文数据元素。**数据屏蔽技术必须包括相应的功能，以支持特定数据元素的业务上下文。例如，用于准确屏蔽社会安全号码、信用卡号和电子邮件地址的预打包功能是一个明显的优势。

- **保持数据完整性的功能。**数据屏蔽技术必须保持数据的引用完整性。寻找能够自动屏蔽数据元素并且能够跨相关的表以及应用、数据库、操作系统和硬件平台准确传播屏蔽数据元素的功能，以确保有效的结果。如果解决方案不能保持数据的完整性，则处理结果将是不准确的。

简而言之，您需要的数据隐私解决方案应该能够伸缩，以便满足您当前和将来的企业数据屏蔽要求。

### 通过 IBM Optim 应对数据隐私挑战

面向 SAP 应用的 IBM Optim 数据隐私解决方案提供了全面的功能，能够有效地用于非生产环境中的应用数据去标识化处理。Optim 的数据屏蔽技术保持了数据的完整性，并且能够产生反映应用逻辑的一致且准确的结果。

屏蔽数据可以跨多个非生产环境准确地传播，以生成有效的结果。最后，Optim 的数据屏蔽技术是可伸缩的，可以跨应用、数据库、操作系统和硬件平台部署，以满足您当前和将来的需要。Optim 通过提供有效数据屏蔽的基础元素，使公司能够满足最为复杂的数据隐私挑战。

**应用感知数据屏蔽。**Optim 的应用感知数据屏蔽功能能够准确地理解、捕获和处理 SAP 应用数据元素，以便屏蔽数据不会违反应用逻辑。例如，姓氏被替换为虚构的姓氏，而不是没有意义的文本字符串。数值字段会保持适当的结构和模式。同样的道理，如果员工的 ID 号是四位数，并且取值范围为 0001 到 1000，则屏蔽值 2000 在应用测试上下文中将是无效的。校验和（Checksum）仍然有效，因此功能测试会通过所有应用有效性检查。

**上下文感知数据屏蔽。**Optim 的上下文感知、预打包的数据屏蔽例程能够将整个 SAP 应用， 包括 HCM，中的重要数据元素进行去标识化处理。

Optim 提供多种久经考验的数据屏蔽技术，可用于对很多类型的敏感信息进行去标识化处理，例如，出生日期、银行帐号、国民识别号码（如加拿大的社会保险号或意大利的税号）、福利信息、健康保险识别号码等等。

Optim 的 Transformation Library™ 例程可用来准确地屏蔽复杂数据元素，例如，社会保险号、信用卡号和电子邮件地址。内置的查找表支持名称和地址屏蔽。您还可以采用特定于站点的数据转换例程，以整合多个相关的应用和数据库的处理逻辑，并且在支持甚至最复杂的数据屏蔽要求方面提供更高的灵活性和创造性。

**持续的数据屏蔽。**Optim 的持久屏蔽功能能够为源列生成经过转换的替代值，并且能够跨应用、数据库、操作系统和平台一致而准确地传播替代值。持久数据屏蔽功能确保了跨多个 SAP 应用开发、测试和培训环境保护隐私的可伸缩性。

## 久经考验的数据屏蔽技术

Optim 提供了一组全面的、久经考验的数据屏蔽技术，以转换数据或对数据进行去标识化处理。您所使用的方法将取决于您要屏蔽的数据类型以及您想要取得的结果。下文中介绍了一些可通过 Optim 使用的屏蔽技术。

**屏蔽字符和数值数据**。Optim 提供了多种屏蔽字符和数值数据的技术。在简单级别上，可以使用字符串常量来指定用于屏蔽字母数字数据的值。您可以使用包含在引号中的任何字符或数字组合来定义字符串常量。例如，在汽车保险上下文中，可以容易地用“Code60”代替索赔的结算价值。与此类似，子字符串屏蔽技术返回列内容的一个子字符串或一部分。使用包括区号和电话号码头三位的子字符串可提供需要的详细信息，并防止访问实际的电话号码。

顺序屏蔽技术可以与字符或数值数据类型结合使用，并且返回顺序递增的值。例如，在银行应用中，可以使用此技术来屏蔽支票帐号，方法是：简单地指定一个起始帐号，然后按 7 递增每个帐号。

随机屏蔽技术返回一个从用户指定的值范围中随机选择的值，并用于屏蔽字符或数值数据。例如，在测试健康保险应用时，可以生成随机数来屏蔽订户编号、组号、卡号、卡日期和客户编号。

堆 (Shuffle) 屏蔽技术提供了最大限度的随机数据屏蔽。此技术将单个或多个列中的数据重新分配到指定数量的行中，并且可以根据需要跨堆实施惟一性。您可以将此技术应用于几乎任何类型的数据，并且可以轻松地使用它来屏蔽名字、姓氏或两者、地址信息（包括街道地址、市、县以及邮政编码或邮递区号）。

**使用查找值屏蔽数据。**另一种去标识化的方法是使用替代值转换数据。您可以使用查找技术，通过向目标列返回相应的屏蔽值来屏蔽源列中的值。例如，查找表可能将医疗诊断代码转换为虚构代码以达到测试目的。

使用随机查找技术，可以通过为目标列返回一个相应的、随机选择的屏蔽值来屏蔽源列中的某个值。Optim 提供了若干个预定义的查找表，使数据屏蔽变得更加容易：

- **名字查找。**包含 5,000 个以上的名字，用于对个人信息进行去标识化处理。
- **姓氏查找。**包含 80,000 个以上的姓氏，用于对个人信息进行去标识化处理。
- **街道地址/市/州/邮递区号查找。**包含 100,000 个以上的美国地点，用于屏蔽完整地址信息。

使用一种增强的随机查找技术，可以轻松地转换目标表中某个行的任何列或所有列中的数据，方法是将该数据替换为从某个查找表中随机选择的整行数据。例如，使用此功能可以屏蔽完整的街道、地址、市、州和邮政编码/邮递区号数据，而不是将一个邮递区号替换为另一个邮递区号。

使用 Optim 的 Transformation Library 屏蔽敏感数据。使用 Optim 的 Transformation Library 可以生成有效的屏蔽值，以对社会安全号码、信用卡号和电子邮件地址进行去标识化处理：

- **社会安全码。**生成遵循美国社会保障署所使用的规则的、有效的、经过转换的数字。例如，在对处理失业救济金的应用进行测试时，可以使用此功能来屏蔽社会安全号码。
- **电子邮件地址。**社会安全码使用字符串常量或名字*/*姓氏列以及字段生成有效的、经过转换的电子邮件地址。例如，可以使用此功能在用于培训新员工的直销应用中屏蔽电子邮件地址。

**保持屏蔽数据的完整性。**迄今为止所介绍的每个方法都可以用来有效地屏蔽数据，以保护机密性。但是，对于如 SAP HCM 之类的关系数据库应用来说，还有一个额外的难题。具体说来，您需要具有相应的功能将屏蔽数据元素传播到数据库中所有相关的表，从而维护引用完整性。

Optim 对键传播提供了完全支持，使您可以向主键或外键列指定一个值并将该值传播到所有相关的表。您指定的值可以是有效的列名称、字符串常量、表达式或其他屏蔽值。例如，有两个相关的 SAP HCM 表（图 1）。个人信息类型 PA0002 是地址信息类型 PA0006 的父类，它的主键列 *PersID* 是一个由 5 个数字组成的数值。*PersID* 代表了 PA0006 里的一个外键。

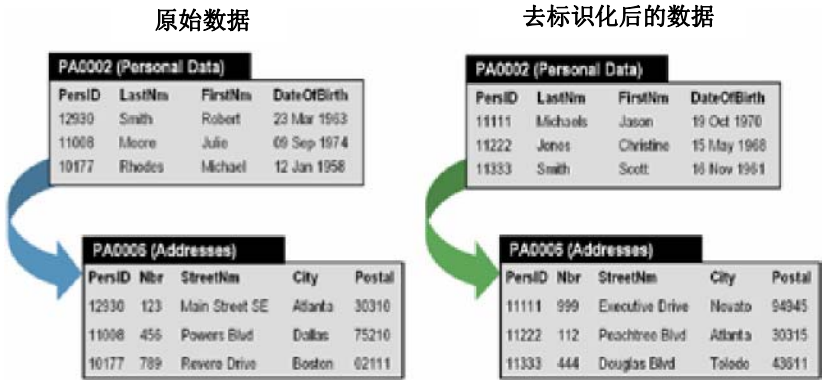


*图 1 即使已经屏蔽数据，Optim 的键传播功能也能帮助保持数据的引用完整性。*

在图 1 的例子中，个人信息类型 PA0002 里的 *PersID*、*LastNm*、*First Nm* 和 *Date Of Birth* 列都被屏蔽了。而在地址信息类型 PA0006 中，*PersID* 和 *City* 列也给隐蔽了。要注意的是个人信息类型（*PersID*）里的主键列的屏蔽数据传播到了地址信息类型里的外键列（*PersID*）。用这种方式，测试数据库里的个人信息类型和地址信息类型之间的关系保持不变。如果不是有着这样的传播屏蔽值的能力，数据的引用完整性将受到严重破坏，从而形成地址信息类型的孤行（图 2）。

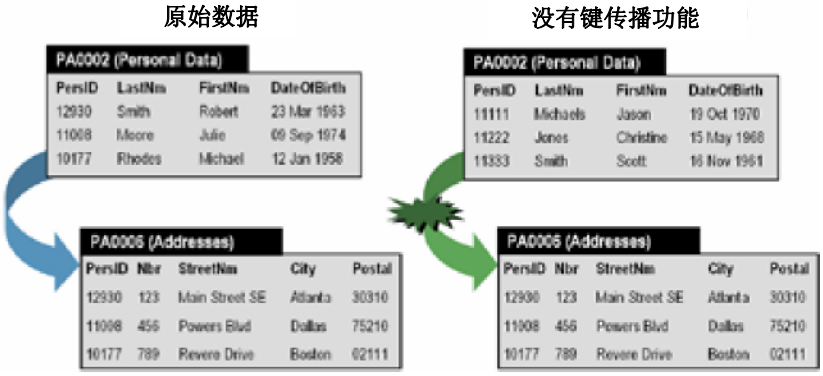


*图 2 没有键的传播能力，关键的数据关系将会受到严重破坏。*

没有准确传播键值的功能有助于保持测试数据库的引用完整性，以支持有效的测试结果。请想象一下当涉及到数百个相关的表，并且必须将键传播到所有相关表时的复杂性。如果没有传播功能，将会产生很多孤表，并且测试数据库很容易毁坏。

**用户定义的屏蔽例程。**当您需要执行更复杂的数据转换时，您可以准备用户定义的退出例程。这些例程只是执行所需数据转换的程序或指令集。在为目标列生成无法使用任何其他方法定义的值方面，退出例程尤其有用。例如，在向应用报告流程提供测试数据之前，测试人员可能需要将年累计销售收入取整。在另一种情况下，可能需要根据客户的地理位置、平均帐户余额和成交量为 Customer ID 代码生成一个值。然后，使用此退出例程生成的 Customer ID 代码将被用于填充目标列。

## 数据隐私最佳实践总结

保护敏感数据的隐私性和机密性的需要跨越了生产和非生产应用环境，并且不受行业和地理界限的约束。尽管很多公司已经采取了有效的措施来保护 SAP 生产环境中的数据，但它们才刚开始将注意力转向非生产数据库环境中的漏洞。

但是，现实中存在很多挑战，因为在生产环境中适用的保护措施未必支持非生产环境的需要。开发、测试/质量控制和培训团队需要现实的数据来准确完成他们各自的活动。对数据进行去标识化处理提供了一种系统地删除、屏蔽或转换可能被用于识别某个个人的数据元素的手段。已经进行去标识化处理的数据在非生产数据库环境中是有效的和可用的。

面向 SAP 应用的 IBM Optim 数据隐私解决方案提供了多种数据转换技术和内置的查找表来屏蔽上下文相关的数据元素，甚至支持自定义的数据屏蔽例程。Transformation Library 提供了为社会安全号码、信用卡号和电子邮件地址生成和传播有效的屏蔽值的功能，能够在确保准确性的同时保护隐私。最重要的是，您可以跨相关的表准确地传播屏蔽数据元素，以帮助保持数据库的引用完整性。在更高层次上，可以跨应用、数据库、操作系统和硬件平台准确地传播屏蔽数据，以保护您的整个企业。

Optim 支持业界领先的数据库管理系统并提供了联合访问功能，使您可以在单个流程中提取和屏蔽来自各种生产数据源的适当数据。Optim 还提供了单一、可伸缩的数据隐私解决方案，该解决方案具有能够轻松适应您当前和将来的要求的灵活功能。实现 Optim 可帮助您遵守数据隐私法规，并且在您的整个企业中保护敏感信息的机密性。

## 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，例如数据增长管理、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 使应用数据管理适应业务目标，以帮助优化性能、降低风险并控制成本，同时提供跨企业应用、数据库和平台进行伸缩的功能。如今，Optim 帮助全球所有行业的企业利用在企业应用数据生命周期的每个阶段管理其数据的能力，充分发挥了企业应用和数据库的商业价值。

## 更多信息

要了解有关 IBM Optim 企业数据管理解决方案的更多信息，请与 IBM 销售代表联系，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
U.S.A.
www.optimsolution.com

在美国印刷

2008 年 10 月



*1 Richard Mogul， "Danger Within – Protecting your Company from Internal Security Attacks，" Gartner，2002 年 8 月。Richard Mogul， "Danger Within – Protecting your Company from Internal Security Attacks，" Gartner，2002 年 8 月。*

*2 Thomas Otter， "Testing Times for HR Systems and EU Data Protection Law，" Gartner Research Publication，ID 号：G00157833，2008 年 6 月 6 日，第 1 页。*

*3 Ibid。第 3 页。*



TAKE BACK CONTROL WITH Information Management

IMW14115-USEN-00