数据安全防护白皮书

该白皮书从数据的全生命周期介绍如何对数据进行安全防护，包括数据环境，敏感数据的定位与监控，只要是数据可能存在的地方：传统数据库、文件系统、应用程序、大数据环境等，本文都给出了建议与解决方案。同时提出进行安全分析的标准与方式。

IBM

立即打开了解数据保护的秘籍

目录

数据，作为企业核心资产，越来越受到企业的关注，一旦发生非法访问、数据篡改、数据盗取，将给企业带来巨大损失。数据库作为数据的核心载体，其安全性就更加重要。

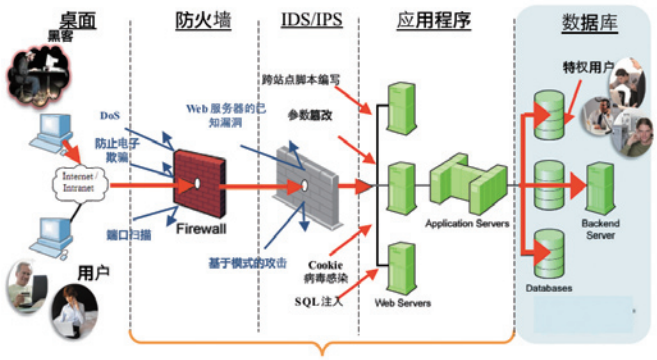
企业核心信息的80%是以结构化信息，即数据形式存在的。数据库中储存著诸如个人身份证号、银行账户、医疗保险、电话记录、客户数据、采购信息、交易明细、产品资料等极其重要和敏感的信息。数据作为企业核心资产，一旦发生非法访问、数据篡改、数据盗取，将给企业在信誉和经济上带来巨大损失，造成的後果可能使灾难性的。因此，需对企业数据进行全方位的保护，建议从以下三阶段，来逐步完成企业数据保护。

第一阶段：企业的基本数据环境及敏感数据保护

保护企业数据的首要任务是保护敏感数据，建议在数据保护的第一阶段，企业根据自己的数据环境、业务特徵，评估和保护数据环境，发现及监控敏感数据。

数据所在的环境安全

只要数据走过的地方，就存在著数据安全问题，因此要全方位保护数据所在的环境安全，网络、用户桌面、机房、应用程序、数据库，每一个环境都不能放过。并能准确表述出是谁在什麼时候以什麼方式访问了哪些数据。



人员安全

人员安全是数据安全中不可或缺的一部分，首先要做的是把不安全人员隔离到数据环境之外。因此，员工的上下班时间，进入机房时间，来访人员信息等都需要进行详细的记录。

同时还要对所有接触到数据的人员提出以下要求：

能设置硬盘密码的电脑必须硬盘密码（包括笔记本电脑和台式电脑）；

所有电脑应设置屏保密码；

有手提电脑密码锁的员工应该在离开办公座位前将密码打乱；

所有包含有机密信息的本地数据库应加密；

安装抗病毒和防火墙软件并正确运行；

仅使用合法软件；

仅授权有业务需要的人员进入敏感系统；

网上邻居共享文件或其他保密文件只允许有业务需要的人员所获取或看到；

当乘坐飞机旅行时，不要将笔记本电脑和其他便携媒介放在检查的行李裏面，当您通过安检时，要对可能被偷的设备保持警觉；

当乘坐机动车辆旅行时，不要将笔记本电脑和其他便携媒介长期放在无人看守的车辆内或尾厢内，要随身携带；

当您必须将笔记本电脑和其他便携媒介留在酒店时，必须将其锁在保险柜裏；

所有电脑必须由公司专门部门来进行安装，必须按照实名来命名个人办公电脑，不得假扮和随意命名；

应用安全

在企业 Web 应用的各个层面，都需要使用不同的技术来确保安全性。为了保护客户端机器的安全，用户需安装防病毒软件；为了保证用户数据传输到企业 Web 服务器的传输安全，通信层建议使用 SSL（安全套接层）技术加密数据；使用防火墙和 IDS（入侵诊断系统）/ IPS（入侵防御系统）来保证仅允许特定的访问。不必要暴露的端口和非法的访问，在这裹都会被阻止；即使有防火墙，企业依然需使用身份认证机制授权用户访问 Web 应用。

但是，即便有防病毒保护、防火墙和 IDS/IPS，企业仍然不得不允许一部分的通讯经过防火墙，毕竟 Web 应用的目的是为用户提供服务，保护措施可以关闭不必要暴露的端口，但是 Web 应用必须的 80 和 443 端口，是一定要开放的。可以顺利通过的这部分通讯，可能是善意的，也可能是恶意的，很难辨别。Web 应用是由软件构成的，那麼，它一定会包含缺陷（bugs），这些 bug 就可以被恶意的用户利用，他们通过执行各种恶意的操作，或者偷窃、或者操控、或者破坏 Web 应用中的重要信息。

因此可以看出，应用安全并不能保证最终的数据安全：

1. 网络脆弱性扫描工具，由于它仅仅用来分析网络层面的漏洞，不了解应用本身，所以不能彻底提高 Web 应用安全性；
2. 防火墙可以阻止对重要端口的访问，但是80和443端口始终要开放，我们无法判断这两个端口中通讯数据是善意的访问还是恶意的攻击；
3. SSL 可以加密数据，但是它仅仅保护了在传输过程中数据的安全性，并没有保护 Web 应用本身；
4. 每个季度的渗透测试，无法满足处于不断变更之中的应用。

只要访问可以顺利通过企业的防火墙，Web 应用就毫无保留的呈现在用户面前。

数据安全

从下图数据泄露调查分析报告和对发生的信息安全事件技术分析，总结出信息泄露呈现两个趋势：

1. 黑客通过 B/S 应用，以 Web 服务器为跳板，窃取数据库中数据；传统解决方案对应用访问和数据库访问协议没有任何控制能力，比如：SQL 注入就是一个典型的数据库黑客攻击手段。
2. 数据泄露常常发生在内部，大量的运维人员直接接触敏感数据，传统以防外为主的网络安全解决方案失去了用武之地。单纯的通过网络包来捕获数据行为的安全监控方式也不能监控到直接在数据库服务器端对数据的访问。

Type	Category	All Orgs		Larger Orgs	
POS server	Servers	50%	1%	2%	<1%
POS terminal	User devices	35%	<1%	2%	<1%
Desktop/Workstation	User devices	18%	34%	12%	36%
Automated Teller Machine	User devices	8%	<1%	13%	<1%
Web/application Server	Servers	6%	80%	33%	82%
Database server	Servers	6%	98%	33%	98%
Regular employee/end-user	People	3%	1%	5%	<1%
Mail server	Servers	3%	2%	10%	2%
Payment card	Offline data	3%	<1%	0%	<1%
Cashier/Teller/Walter	People	2%	<1%	296	<1%
Pay at the Pump terminal	User devices	2%	<1%	0%	<1%
File server	Servers	1%	<1%	5%	<1%
Laptop/Netbook	User devices	1%	<1%	5%	<1%
Remote access server	Servers	1%	<1%	7%	<1%
Call Center Staff	People	1%	<1%	7%	<1%

数据库在这些泄露事件成为了主角，这与我们在传统的安全建设中忽略了数据库安全问题有关，在传统的信息安全防护体系中数据库处于被保护的核心位置，不易被外部黑客攻击，同时数据库自身已经具备强大安全措施，表面上看足够安全，但这种传统安全防御的思路，存在致命的缺陷。

安全泄露事件的主要目标是数据，数据泄露的主要起源地是数据库服务器。客户的极具敏感信息存储在数据库服务器上：

- 财务记录
- 客户信息
- 信用卡信息和其它帐户信息
- 个人身份信息
- 就诊记录

企业必须要解决以下问题，保障自身的数据安全：

- 数据库被恶意访问、攻击甚至数据偷窃，而您无法及时发现、追踪并阻截这些恶意行为；
- 数据库遭受恶意访问、攻击後，不能追踪到足够的证据；
- 不了解数据使用者对数据库访问的细节，从而无法保证对数据安全特别是敏感数据的管理；
- 来自内部的威胁：特权用户随意修改配置、改变或盗取数据，没有明确职责分工。

在世界经济全球化的今天，信息安全特别是敏感数据的防护还需要面对越来越多的强制性规定及来审计问题，包括各种财务法律法规（比如 Sarbanes-Oxley Act）、行业规定（比如 Payment Card Industry Data Security Standard [PCI DSS]）和地方性数据隐私法律。每个法律法规都侧重于独特的方面，但是它们都要求组织探测、记录和纠正对敏感数据的未授权访问和更改，包括由特权用户执行的操作，还要求提供安全审计跟踪以便检验合规性。信息安全和数据库经理必须努力实现这些控制方式，尤其是在监视特权用户方面。对企业声誉风险和敏感数据保护的高度关注还促使企业加强内部安全控制。

面对合规审计的要求，企业往往面对以下挑战：

- 针对数据库、应用系统日志的审计只能做事後分析，周期长，且无法进行持续性审计；
- 审计缺乏规范性，无法有效成为公司的安全管理规范且满足外部审计需求；
- 人工审计面对海量数据，无法满足100%可见性，造成审计不完整；
- DBA 权责未完全区分开，导致审计效果问题。

评估数据环境

企业的数据环境多种多样，无论部署在云端还是普通的服务器上，都需要保证数据环境的安全性，必须对各个系统上的数据库配置、认证和权限设置等及可能由此产生的问题进行积极严格的评估和审查。尤其对于云环境的，需要求数据云服务商提供独立第三方的数据访问报告，而不是数据运营商自己的报告来证明数据环境的安全性。

需要经常对数据库进行脆弱性评估，脆弱性评估是帮助保护和增强客户的基础设施的一个重要过程，通过自动化分析一系列已知漏洞；这些深度评估可检测补丁级别和数据库配置，突出环境漏洞，快速修复问题，并维护关键的企业数据，使其免受内部及外部威胁侵害；同时需要保持严格遵守职责分离之间的审计和管理，并不断将其与行业最佳实践进行比较，更新最新研究及广泛且复杂的异构数据库。

脆弱性评估通常至少需要包含以下部分：

- 安全配置评估
  如：数据库配置，操作系统配置，权限及账户设置，共享凭证检测等；
- 数据库基础设施分析及脆弱评估
  如：数据库服务器上的操作系统补丁级别，数据库服务器补丁级别，流量观察等；
- 脆弱报告及建议措施
  对于发现的各类问题，需要有完整的分析与报告，并给出解决方案；
- 合规性验证
  如针对不同行业需要符合 PCI、DSS 及 SOX 等法规要求实施一些控制措施，以防止未经授权情况下修改和访问敏感数据。
- 数据保护服务
  及时处理脆弱报告发现的问题，结合流量分析，设置动态安全保护，如：该系统通过实时观测所有数据库事务发现漏洞，如数据库错误过多(表明可能发生SQL注入攻击)；
- 确立长效机制
  仅只有评估、分析和保护并不完整，需要将所有对于数据库安全的内容自动化，长效化并持续监督其执行与核查。

对于数据库环境的脆弱性检测在基于美国国防部创建的计算机互联网安全（CIS）基准、数据库安全技术实施指南（STIG）等行业标准中均有明确的定义和要求。

保护数据环境

绝大多数数据库环境的变更均通过数据库引擎发生。对于大多数数据库类型，数据库控制和配置均通过特定的 SQL 命令，或者由数据库管理员或（数据库）安全管理员执行的存储程序来完成。

在评估数据库整体环境后，需要持续侦测影响数据库安全性的配置变更，对已经达到评估要求的数据库环境进行锁定及保护，以保证符合评估要求的数据库配置等项目。

综上所述，数据库是一项在操作系统级别上安装的程序，并且依赖操作系统的服务。许多配置元素均驻留在操作系统结构内，而不是位于数据中。

这样的例子还包括文件、注册表值及环境变量。这些文件和值大多能控制某些最为重要的数据库安全性方面。数据库身份验证的方法就是一个很好的例子。在几乎所有的数据平台中，管理员均可以通过变动一个这样的值（要麽与 SQL 一同使用，要麽替代 SQL）来变更数据库验证用户身份的方式，显然，如果管理员修改和使用较弱的身份验证方法则可能会出现严重的安全漏洞。因此，必须监控和警惕这种状况的出现。

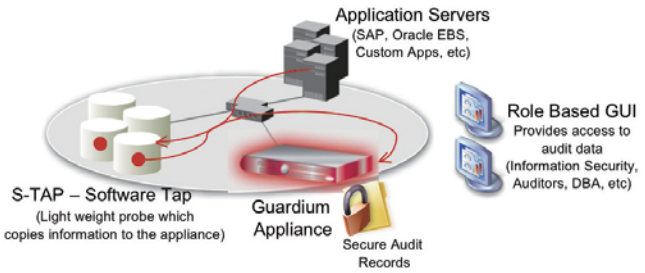
因此，建议通过监控及审核以下内容，达到综上目的：

- 跟踪数据库引擎范围外可能影响数据库安全性的所有变更
- 与数据库活动监控相辅相成，提供全面的数据库监控
- 跟踪可能影响数据库安全状况的数据库配置文件及其他外部对象变更，如
  - ✓ 环境/注册表变量
  - ✓ 配置文件(例如SQLNET.ORA、NAMES.ORA)
  - ✓ 外壳脚本
  - ✓ 操作系统文件
  - ✓ Java 程序等可执行文件
- 实现所有治理及风险管理的必备模块

架构安全

在搭建数据安全监控平台时要做到以下几点：

- 可监控未经过网络的本地数据操作行为；
- 支持职责分离，进行安全监控的负责人可以对数据库没有操作权限；
- 安全监控人员权限管理；
- 收集和标准化数据用于高效存储；
- 所有审计数据统一存储；
- 数据即时显现和高度保护；
- 加密数据库依然可被监控；
- 在进行数据安全监控和保护时，要通过权威的第三方提供监控报告，而且报告内容不可修改。



数据库数据安全

定位敏感数据

对于拥有大量敏感信息和数据的企业而言，如何在使用重要数据的同时，将数据泄露和损失的风险降到最低，始终是一个严峻的挑战。随著企业级业务的快速发展，以及 IT 系统应用的越来越普遍，企业内部已经积累了大量的敏感信息和数据。而这些数据，在企业信息的很多工作场景中都会得到使用，例如，客户信息，行业数据，业务分析、产品开发测试、甚至是一些外包业务等方面，使用的都是真实的业务数据和信息。这些敏感的数据，一旦发生泄漏、损坏，不仅会给企业带来很大的损失，更重要的是会大大影响用户对于企业的信任度。

为了保护信息安全，首先您必须知道敏感数据存在哪裏，发现和定位敏感数据，明确网络环境中的数据库的分布情况。根据定义的敏感数据类型及模式自动的发现包含该敏感数据的所有信息，同时为该敏感信息分类。

要达到上述目标，首先对网段内所有数据库服务器进行扫描，发现所有数据源。其次，定义目标敏感源，如：名字中含有某些内容的数据表；或者数据内容为某一类型：文本类型、数字类型、时间类型等，数据内容匹配某一格式，如：信用卡号。之後，开始扫描所有数据源，包括数据服务器上的所有表，图视，同义词等，定位到与敏感源相关的信息。在某个数据库的某个表的某个字段中含有敏感信息。发现数据库内的敏感数据後，使用智能数据库抓取搜索自定义模式来分类数据。一旦找到敏感对象，自动标记有元数据的分类，如〝合规记录〞，并加入到具有类似性质的项目，从而确保适当的政策会自动应用到具有类似属性的对象中。

监控保护敏感数据

敏感数据对于不同的行业，不同的应用有著不同的涵盖范畴。通常来讲涉及到个人隐私的信息（比如，身份证号，联系方式等）都是敏感信息，不同的行业敏感信息又有各自独有的特点，例如，通信行业中的通话记录，医疗行业的诊疗记录等。

识别，监控，保护数据库系统中的敏感数据是当前数据库安全的一个关键点。

- 识别：指出数据库系统中哪些数据是敏感数据。
- 监控：无论是数据库系统管理层，还是应用层对于敏感数据的访问都应当受到监控，确保敏感数据访问的合法性，合理性，安全性。
- 保护：数据库中的敏感数据应当得到切实有效的保护，例如，归范用户对访问敏感数据的访问权限，敏感数据的展现方式是否需要漂白，脱敏处理等。

多个层面限制敏感数据的修改、访问

对于敏感数据的访问需要从多个层面加以保护，限制。

- 权限控制：这是敏感数据访问控制的第一层。为不同的数据库用户定义不同的访问权限级别，不同的访问权限对于敏感数据的读写有不同层次的限制。因此，规划数据库用户访问权限是保护敏感数据的第一步。
- 敏感数据访问途径：用户可以通过不同的途径访问敏感数据，比如应用程序，数据库管理平台等。如果只针对应用程序对于敏感数据的访问加以限制，而忽视数据库管理平台的控制，敏感数据的泄漏依然会是系统安全的一大威胁。因此，敏感数据访问途径也是敏感数据管理需要考虑的一个重要层面。

可疑行为阻断

当一个〝可疑〞用户或者合法用户用一种〝可疑〞的访问方式尝试访问数据库的敏感数据时，系统的安全机制应当将其认定为一次的可疑的访问行为，予以阻断。所谓阻断，类似于防火墙的安全机制，将不安全因素隔离在当前系统之外，确保当前系统的安全。数据库系统的安全管理应当涵盖可疑行为的定义，并针对不同的可疑行为采取相应的阻断措施。

关联操作记录及控制

在关系型数据库系统中，不同的数据表中保存不同的数据，而这些数据表之间相关联。直接访问某张数据表的数据看起来没有安全问题，表之间关联访问有可能将一些敏感的数据泄漏出去。例如，在医疗系统中，单独访问患者的就诊信息，和治疗信息都是安全的，但如何将这些信息关联起来访问便可以大致判断病患者的病情等敏感信息。

因此，对于不同的行业，不同的应用系统，需要分析关联操作可能带来的安全隐患，并加以记录及控制。

返回值中敏感数据

为防止第三数据泄漏可能带来的不安全隐患，敏感数据对于所有的数据库用户来说都应当一视同仁的。拥有访问权限的用户访问数据库中的敏感数据时，该如何处理呢？可以肯定的是，敏感数据不能〝赤裸裸〞地直接展现给访问用户，数据库安全机制需要对这些数据加以保护，例如，敏感数据脱敏，漂白，在保证敏感数据安全的前提下，其数据本身始终遵循数据库系统及应用程序的定义。

特权用户的访问记录

在很多情况下，数据库系统特权用户的不正当操作有可能会威胁整个数据库系统的安全。在生产环境中，对于特权用户的访问有严格的审查流程，包括何时访问，执行哪些操作，执行顺序等等。记录审计特权用户的访问记录，可以确保特权用户在正确的时间完成了正确的操作，审查是否有越轨行为的出现，进而保证数据库系统的安全。

动态访问遮蔽

动态访问遮蔽是在实际生产中最常用的功能。一些包含有敏感信息的数据库，在不限制用户访问的情况下，需要对敏感信息进行动态遮蔽。比如，存储有关客户个人信息，例如，身份证号码，电话号码，电子邮件等，应该对这些敏感信息进行部分或者全部遮蔽，来达到数据安全保护的目的。再例如，对一个公司来说，对员工最敏感的工资信息，更需要动态访问遮蔽来保护员工个人信息的隐私。

因此，动态访问遮蔽是在不确定能够排除那些用户，那些访问地址，甚至那些字段为可疑或者有害访问时，关注数据内容本身，抓住敏感信息点，并有针对性地对该部分信息进行动态访问遮蔽，从而达到保护数据安全的目的。

数据库异常监控

数据库异常需要引起重视，一些有特殊权限的账号在短时间内连续登录失败多次，需要提高对该账号的监控，记录该异常发生的时间、IP信息、操作系统用户信息，对账号进行冻结，实时报告该类异常操作。

多次尝试查询不存在表的行为，需要记录并报警，根据异常次数实时阻断该用户的连接，防止非法用户猜测、攻击含有敏感信息的数据库表。

对于数据库的死锁等异常信息，要做到实时报警，确保数据库正常工作。

第二阶段:测试库及非传统敏感数据的保护

测试数据管理

大部分企业的生产环境都建立了安全和访问限制，以防止泄密。标准安全方法适用于网络级别、应用程序级别以及数据库级别。组织甚至可以应用程序开发最佳实践，作为日常流程的一部分，以确保应用程序代码的编写更加安全。但是，这种保护性方法无法适用于所有环境，因为在生产中保护数据的方法无法满足保护非生产环境 （测试、开发和培训）的要求。

在非生产环境中保护数据

在数据落入不当的人手中时确保数据得到保护是最佳的解决方案。身份鉴定防护和屏蔽是确保任何人无法使用被盗、泄密、丢失数据的方法之一。在非生产环境中，数据身份鉴定防护是一个系统化的流程，它可以移除、屏蔽、转换，可以确定个人信息或其他不得公开的机密数据元素。数据屏蔽使开发人员、测试人员和培训人员可以使用符合实际的数据并生成有效的结果，同时遵守隐私保护法规。

加密可以掩饰数据，并将它转化为 `加密´ 格式，以保护数据库内的隐私。要在用户界面中查看、进行报告以及其他开发和测试活动时，数据又将解密回原始状态。大部分情况下，开发人员和测试人员在查看应用程序屏幕、输入数据、运行报告，执行开发和测试活动时将看到解密数据。尽管加密可以在数据从数据库中被盗时提供保护外壳，但是一旦数据解密後将无法防止盗用（例如，如果用户复制数据）。当数据被导出数据库并导入到电子表格或者其他文件格式时，加密不再有效，数据将处在风险之中。只要能够看到数据，就可以复制数据。

数据漂白是保护测试数据的有效手段

由于业务上的需要，很多企业需要在非生产环境中使用生产系统的真实业务数据，比如在测试系统中用真实数据运行各项测试用例、在培训系统中用业务数据进行演示等等。在应用于非生产环境时（开发人员、测试人员和培训人员需要访问真实数据），在生产环境中保护隐私的标准方法可能不再有效。而往往在非生产环境中不具备生产环境中的各项数据保护措施，如权限控制、访问控制等等，使得在这些环境中对生产数据的保护面临严峻的挑战。再加上测试环节的复杂性，如引入第三方测试夥伴，数据分发过程等，数据安全更加难以保证，数据泄露的风险很高。

大部分组织已经有正式的应用程序开发生命周期流程。大部分组织已经意识到创建数据管理和数据治理战略的需求。在生命周期中使用定义明确的实践集合保护数据可以确保它在开发、测试和培训活动中仍然能够得到保护。从相关行业测试数据管理以及数据隐私保护项目的运行经验看，在非生产系统中使用真实数据进行测试前，先将数据漂白然後再分发、使用是保护数据的有效方式。数据漂白（又叫数据脱敏或数据变形）是指在非生产环境中将涉及个人隐私或其它需要保护的敏感业务数据，进行移除、屏蔽以及转换的一系列系统过程，目的是将敏感数据进行脱敏处理，以达到对其保护的目的。数据漂白需要明确数据漂白范围、确定数据漂白需求，以及实现有效的数据漂白技术。

常见敏感数据的类别

什麼样的数据是敏感数据？这个问题的答案，当然是根据不同业务系统的特点、测试背景而各不相同。通常可能被视为敏感数据的业务数据包含客户个人隐私数据以及某些关键的敏感业务数据：

- 个人隐私数据 ：姓名，身份证号码，地址，电话号码 / 联系电话，邮箱地址，所属城市，邮编号码等。
- 其他敏感业务数据（关键业务数据）：组织机构名称，客户账号信息 （ 银行账号等 ） ，营业执照号码，交易日期，识别码 （ 交易号、系统唯一 ID 等 ），密码类 （ 如账户查询密码、取款密码等 ）。

具体数据漂白项目中涉及的敏感数据范围，需要由业务人员以及漂白项目人员共同确定。明确敏感数据范围之後，下一个问题需要确定数据漂白的具体需求，即明确数据漂白要做到什麼程度、漂白成什麼样的目标数据。

数据漂白需求以及常见的三对矛盾

数据漂白对数据安全、数据质量、漂白效率等各个方面都有不同的要求，而各个需求之间有相互关联的特点：

- 去隐私化与保持真实性：数据漂白的基本原则是对隐私数据进行脱敏处理，然而过度的处理往往会带来负面的问题：数据失去了原有的格式以及含义，不满足测试对数据保持仿真度的要求。比如在客服培训系统中如果将客户姓名〝张三〞转换为随意汉字组合或者乱码序列〝A%#LK 嗑〞，将会使培训人员不易识别甚至带来困惑，从而影响培训效果。因此，数据漂白在脱敏保证数据安全的同时，也需要保持原来业务数据的特点才能保证其在测试场景中的可用性。
- 保持业务规则与漂白通用性：保证数据业务规则不变是数据漂白的基本要求，包括：保持漂白数据的数据关联性以及业务语义不变等。其中数据关联性包括主外键关联性、关联字段的业务语义关联性等。保持主外键关联是保证数据一致性的重要保证：如客户表和订单表通过客户号码进行关联，如果将客户表的客户号进行漂白，而订单表的客户号没有漂白或者漂白为不一样的数值，会出现数据不一致性。而业务语义关联是指不同字段没有直接的主外键关联，但在业务含义上确实有相互关联的，比如，客户地址与邮编号码有语义关联，客户表中的客户名称字段与邮寄地址表中连络人的客户名称有语义关联。通常数据库中的冗馀设计容易带来这样的语义关联。而真实情况往往更复杂，比如关联数据跨越不同数据表以及不同的数据库，为保证漂白数据的一致性带来难度。业务语义不变则是许多测试案例的硬性指标，如保证数据格式不变、保证业务数据在合理的取值范围等。总体上来讲，业务规则是千变万化的，数据漂白需要根据不同的漂白需求设计不同的漂白流程，但如果数据漂白方案设计没有考虑通用性，会带来管理上的问题以及不能适应漂白需求的变化，最终导致数据漂白项目不可控。因此数据漂白设计时需要综合考虑业务规则以及漂白通用性。
- 漂白结果一致性与不可逆需求：在数据漂白项目中，同样的数据进行多遍漂白的情况是很常见的：如分别对不同的子系统进行异步漂白；或者对一个系统按照不同条 件进行任务划分，使同样的数据（不同时期的客户订单中同一客户的姓名数据等）出现在不同的任务中。因此保证不同漂白过程中漂白结果的一致很重要，而做到漂白结果最理想情况是能做到数据值的一一映射。然而实现一一映射却会带来另外一个问题：被变形的数据有可能会被逆转，得到真实的原始数据，从而失去数据漂白的意义。综合考虑这两个因素，一个有效解决手段是实现可控规则的一一映射，即：将实现数据映射规则的人员和实施数据漂白的人员分开，设计人员负责实现映射算法并保证算法的不可被逆推，保证在有效取值范围内实现一一映射，同时保留 外部能影响映射的参数。而实施人员在实施数据漂白项目时才指定控制参数，进而得到唯一的映射规则。

概括来讲，敏感隐私数据的保护实施方案要具备以下一些功能：

- 身份鉴定防护功能，使用符合实际但虚假的数据屏蔽机密应用程序数据；
- 应用程序感知屏蔽功能，可以确保屏蔽的数据类似于原信息的结构和特徵；
- 上下文感知、预打包的数据屏蔽实用程序，可以轻松保护能鉴别身份的数据元素，比如信用卡编号、社保编号和电子邮件地址；
- 持久屏蔽功能，跨应用程序、数据库、操作系统和硬件平台持续传播屏蔽的替代值；
- 屏蔽数据的参照完整性，确保成功测试和开发；
- 帮助维护对国家和全球数据隐私规则和要求的遵从性；
- 实现和使用简单。

文件系统安全

监控重要文件操作

随著各种电子设备的迅速推广，以及互联网应用的快速发展，每时每刻都会有大量的文件带著各种信息在网络上传播,确保这些文件的安全访问，特别是监控一些包含敏感信息的重要文件，是每个企业和个人实现数据安全不可或缺的重要元素。

但是在日常文件管理中，容易出现下面这些安全隐患：

- 一些重要文件被不小心删除；
- 重要文件都是什麼时候被谁修改了，修改历史不能追溯；
- 敏感信息的内容有变化，不知道到底有哪些文件包含敏感信息；
- 非管理员用户还是可以看到一些重要文件；
- 有黑客攻击重要文件，如何最快知道。

其实系统中的文件存在脱离不开以下五大因素：

1. 文件路径 Where：在主存储之外，无论是大容量的存储系统，还是便携的移动存储，文件在系统中都必须以具体的文件路径才能进行各种操作。
2. 操作用户 Who：从文件的生存周期看，从创建到消失，都离不开各种用户主动发起的操作。
3. 时间属性 When：现在各个操作系统都会清楚记录文件的创建时间和最近修改时间，从安全管理角度看，文件的任何修改甚至消失，时间纪录尤为重要。
4. 文件操作 Action：依照不同用户的不同需求，文件操作使文件在系统中活跃起来。
5. 内容 What：文件中包含的信息是它存在于系统中的必要条件，内容的变动必然来自于上述四个因素的联合作用。

基于上述五个因素，可以从下面几个方面考虑针对重要文件的监控：

1. 监视的细粒性：文件在系统中发生的任何变化，都可以追溯到在什麼时间，是什麼用户以何种操作修改了哪个具体文件；同时也可以有方法通知到相应的安全人员。
2. 控制的灵活性：无论是针对单个重要文件，还是一些聚集在某些目录下的多个文件；无论是单个读写命令，还是各种文件操作；无论是指定单个用户或者多个用户，还是针对所有用户，均可以对重要文件进行控制。
3. 控制的时效性：每个文件在系统中都有访问控制属性，随著文件内容的变更或者安全要求的修改，在不同时期对某些文件的控制会有不同需求。
4. 敏感定义的扩展性：每个系统所产生文件的信息不同，同时不同行业对什麼是敏感信息的定义也会不一样，应该允许不同系统下监视不一样的重要文件。
5. 监视的完整性：重要文件有可能来自于单个用户的简单操作，也有可能产生于某些系统的输出，而往往这些文件散落在系统各个文件路径中，如果全部监控，对系统的性能和其它非重要文件的操作必然有影响。应确保将包含敏感信息的所有这些重要文件都定位出来，以便集中监控起来。

为了防患于未然，对于重要文件的监控，建议采取以下几个阶段实现：

第一阶段是准备计划阶段：基于现有系统，将重要文件划分出来，在文件系统下进行有效的组织；界定业务用户和系统用户权限，从而找到相对应的重要文件权限。

第二阶段是预启动查漏阶段：按照一定安全规则将文件系统监控起来，评估系统上各个应用的运行情况，有针对性地对一些重要文件进行操作，验证监控功能。按照预定义扫描文件系统，以免有新的重要文件产生。

第三阶段是监控运行阶段：基于定期检查文件监控日志，建立相应的文件安全分析，基于新的需求从而丰富安全规则。

文件加密

敏感数据不仅可能在流动的过程中被窃取或篡改，也可能在〝静止〞的状态被拷贝和篡改，导致信息泄露。那麼如何保护〝静止〞的敏感数据呢？ 通常采用的防护办法是〝加密〞，谈到加密就涉及到秘钥的管理。所以目前比较高级的加密保护系统是一个服务器管理多个文件服务器。服务器负责统一生成密钥并负责管理密钥，并使用指定的密钥对用户指定的敏感数据文件或目录进行加密或解密，还根据用户设定的安全策略来限制哪些用户可以访问哪些敏感数据（解密过程对用户是透明的），哪些应用程序不允许访问敏感数据，哪些用户只能在特定的时间（如工作时间）访问敏感数据等，所有触发安全策略的访问行为都被记录下来，供日後审计使用。在实际使用中，可以对数据库的归档文件进行加密，该文件仅供管理员用户使用，其他用户无法访问该文件（因为该用户看到的是密文）。也可以对HADOOP系统中的数据节点上的文件进行加密，防止敏感信息在文件系统级别被泄露。

应用程序安全

企业的安全需求多种多样，为了满足不同的应用需求，仅仅通过数据库安全监控已经并不能满足企业需求，例如：

- 不是所有的应用都是基于数据库的，包括一些敏感图片等；
- 应用中展示的信息可以是通过某种方式计算出来的，不是单纯的从数据库裹拿取数据；
- 应用中的隐含信息包含敏感内容，如网页中的标签；

屏蔽的应用界面

敏感信息可能存在于应用界面的某一列。需要能根据该列的内容特徵进行屏蔽，某一<table>标签中的某个<td>

```
<table class="bz_buglist" width="100%" cellspacing="0" cellpadding="4">
```

在圈定要屏蔽的模块或内容时，需可以指定用什麽内容来对圈定部分进行替换。

ID ▼	Sev	Pri	OS	Assignee	Status	Resolution
24417	nor	P4	Othe		REOP	---
40631	maj	P2	Othe		NEW	---
46406	nor	P2	Othe		NEW	---
~~42566~~	~~nor~~	~~P5~~	~~Othe~~		~~CLOS~~	~~INVA~~
40261	nor	P2	Othe		RESO	DUPL
45795	nor	P2	Othe		RESO	FIXE
46228	nor	P2	Othe		NEW	---
~~22734~~	~~nor~~	~~P4~~	~~Othe~~		~~CLOS~~	~~FIXE~~
43977	cri	P1	Othe		VERI	FIXE
~~44019~~	~~nor~~	~~P3~~	~~Othe~~		~~CLOS~~	~~WONT~~

除了对文字进行屏蔽外，还要对一些特殊格式的敏感信息进行屏蔽和替换，如图片，表格等。

移除应用中的隐含信息

在应用程序的代码裹，有时标签中包含敏感信息，因此需要对可以查看的源代码进行全面扫描分析，对敏感信息进行屏蔽。如下

应用程序内容屏蔽权限设置

信息安全很重要的一点是权责分离，因此要设定不同的用户权限来实现不同的任务：

- 应用用户：应用程序的使用者，但不允许看到所有信息，用户正常使用应用程序，但是敏感信息需被屏蔽。
- 系统管理员：配置系统以符合某些隐私规则，创建、编辑基于内容和基于上下文行为屏蔽规则。基于内容的规则，适用于文本结构，通过简单的正则表达式定义哪些文字被屏蔽，基于上下文行为的规则，适用于屏幕上的结构或者布局，定义 JavaScript 行为，在一个单一信息上执行，通过选择工具，告别〝手工〞编写脚本，实现可视化的创建规则，易于自定义规则。
- 安全保密员：跟踪检查规则的遵守情况，制作查看报表。

应用程序安全与数据库安全相结合

信息安全不仅限于应用程序安全或者数据库安全，它需要综合考虑各种场景，有信息的地方就需要考虑信息安全问题，因此应用程序的安全也需要满足一下要求：

- 通用性：在协议层实现屏蔽，支持大范围的应用；
- 可拓展性、易用性：支持基于环境的屏蔽（不仅仅局限于正则表达式）；
- 完整性：和数据库安全无缝集成，可通用同一套策略
- 扩宽性：可以简单地扩宽到云或者移动端用户实例

归档及大数据安全

监控记录归档

对于数据安全监控数据，要做到定期归档，同时要对归档数据进行加密保护，防止归档数据丢失或被篡改。

归档後的数据要方便恢复，并能把恢复後的数据和当前已有数据合并起来，方便联合分析、追溯问题。

大数据安全

大数据不仅意味著海量的数据，也意味著更复杂、更敏感的数据，这些数据的大量汇集使得在一次攻击中可以获得更多的信息，信息泄露风险大大增加。大数据并没有传统关系型数据库那麼完善，自身的安全管理机制还不健全，存在很多安全隐患。因此，无论使用的是开源 Hadoop，还是企业级大数据，都要对其进行信息安全管理。

大数据环境多种应用的监控

大数据环境的多样性，给大数据安全带来了更大的挑战。需要同时在各节点上对多种大数据应用进行监控：HDFS，Hbase，Hive，Impale，MapReduce 等。从不同角度，细粒度的对大数据环境进行全方位保护，根据事件发生的时间、地点、人物和方式对其进行跟踪、分析，防止特权用户进行非法或可疑的行为，同时抵挡欺诈用户或外来者的攻击，做到事中报警。

为了方便追溯问题，要做到事後审计。覆盖所有数据库活动的连续、详细的追踪记录，并进行实时的语境分析和过滤，从而实现主动控制，生成审计员需要的具体信息。生成的结果报表使所有数据库活动详细可见，如登录失败、权限升级、计划变更、在非规定时间或来自非法程序的访问、敏感数据表访问等，这些活动是否合规一览无馀。例如，监控所有的：

- 安全异常事件，多次登录失败和错误的执行命令；
- 非授权用户的访问；
- 敏感数据的访问。

大数据安全与其他数据库安全联合分析

大数据环境下应用多样化，敏感信息的存放通常也不仅限于某一应用。在进行安全分析时，应对多种应用监控结果进行联合分析，防止同一攻击者通过不同应用获取多样信息，从而进行信息叠加。如：Hbase 中的药品信息和 Hive 中的病人信息如果被同一个人获取到，会导致病人隐私泄露。因此，需要实时分析并监控，多应用操作行为。

如果企业中，大数据环境与传统数据库环境同时存在时，要同时监控大数据和传统数据库，并生成联合统一的审计报告，防止多数据库查询结合查询，利于追溯问题。如：在 DB2，Oracle，Hadoop 中同时存有相关敏感信息，在监控时，需智能判断出不同数据库访问行为是否为同一人操作。并跟踪该操作人员的所有行为。

第三阶段：数据安全智能分析

智能的数据安全防御

过去几年，大量政府机构和知名公司都遭到高级的有针对性攻击，许多高级威胁的受害者拥有最新的检测和防御系统，但仍未能阻止威胁。通常情况下，企业对于高级威胁往往准备不足，原因在于：

- 过度依赖基于签名的安全技术，敏捷性不高，只能检测出已知的恶意文件与通信内容，因而无法对零日漏洞及攻击做出有效回应
- 对于自动化安全方案存在错误认识，认为方案部署结束後就一劳永逸，享受由其提供的完整保护与攻击预防、检测功能。事实上，没有经验丰富的安全分析人士介入其中，没有哪种自动化工具能够真正在攻击活动出现时做出及时响应。人的介入在安全分析流程中的重要意义不可替代。

2012年 Verizon 调查报告指出，99%的违规会导致数据在几天甚至更短时间内受损，而85%的违规需要几星期甚至更长时间才会发现。

为了解决以上问题，信息安全领域的大数据分析的价值备受青睐。智能的数据安全分析系统通常由如下功能模块组成：

1. 数据收集：实时捕获对敏感数据的访问事件（特别是特权用户对敏感数据的访问），并将所有访问事件汇总到存储系统中。
2. 数据提取：根据调度时间，定期从存储系统中抽取必要的数据，生成通用格式的文件供分析系统处理。
3. 分析用户的访问行为，建立安全模型：根据已有的数据，从用户、数据服务器以及与时间、访问对象和访问方式的组合等多个维度，来分析用户的访问行为，建立完整的数据安全模型。该安全模型通常在最初时间建立，然後再定期通过机器学习的方式不断修正。
4. 实时分析用户的访问行为，并与安全模型比较，得出分析结果： 根据评分高低，将风险事件可分为〝严重〞和〝中等〞；根据评分的类型，将风险事件分为〝新事件〞（如某用户去访问某个从未访问的对象，或者以新的方式去访问某对象），〝高频率事件〞（如某用户访问某服务器的次数突然大幅度增加）和〝失败事件〞
5. 分析结果展现：所有分析结果以可视化的方式展现，高风险事件以红色标示，中等风险事件以黄色标示。同时，还为所有分析结果生成索引，方便快速搜索，特别是日後调查取证。另外，还可以根据时间段实时统计用户的访问行为，如访问了多少数据服务器及服务器列表，哪些时间访问（具体到分钟），以及如何访问敏感数据等
6. 用户响应：任何自动化的系统都不是万能的，需要一定的人工干预，所以提供了用户反馈界面，经验丰富的安全分析人士可以对异常事件及时做出响应，比如忽略还是告警。

问题追溯

多系统多数据库安全监控结果联合分析

数据的价值不是独立存在的，通常，把从各种系统各种数据库中得到的数据结合起来时，会获取到更多有价格的信息。因此在进行数据安全监控时，也需要把各系统的监控结果联合分析。

跨数据库分析

企业数据库的多样性，要求在进行数据安全监控分析时，不能只独立分析某一数据库，需智能的分析出该用户的一系列操作。这就要求数据安全监控与分析支持多平台多数据库。如：DB2、Oracle、Hadoop、MySQL 等各种数据库监控结果联合分析。

跨系统分析

数据安全不应只局限于数据库安全，应把数据库安全分析结果与其他系统安全分析结果相结合。

- 数据库表内数据与数据库环境：某人先修改了数据库配置文件，使数据库不在进行日志记录，之後有查询了某些数据表；
- 数据库表内数据与操作系统环境：某人先修改了注册表，使一些监控信息无效，再查询某些数据表；
- 数据库表内数据与文件：数据库表中查询出文件所在的位置，再对该文件进行替换或修改。

这种联合查询的案例并不是少数，并且带来的危害要远大于单独攻击某一系统，而且难于发现。因此在做数据安全分析时，要全面覆盖各种系统，并支持各系统分析结果的无缝连接。

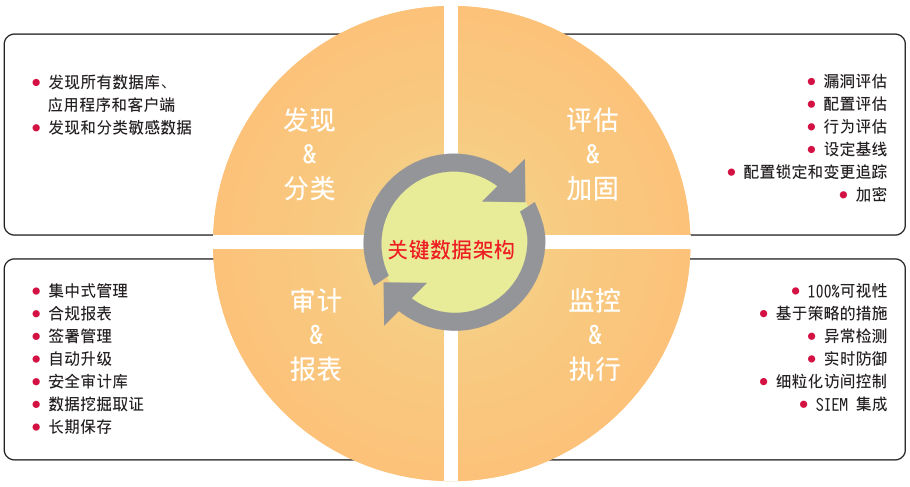
安全监控结果与其他信息联合分析

除了数据库监控结果和操作系统监控结果相结合外。信息安全还应该全方位考虑其他多种因素，如人员基础信息：

- 安全监控结果与人员刷卡考勤记录的联合分析：某人在公司的时间与其操作数据库的记录；
- 安全监控结果与人员生理信息的联合分析：某人心跳信息与其操作数据库的记录；
- 安全监控结果与已有历史报表的联合分析

全方位保护数据

发现敏感数据，评估数据环境，加固数据环境，监控数据安全并实施采取安全措施，自动生成细粒度审计报表，全方位，全周期确保敏感信息不被泄露。



台灣國際商業機器股份有限公司
110台北市松仁路7號3樓
軟體事業處
技術諮詢熱線：0800-000-700

台湾印制
2015 年 06 月