腾讯数盾

数隐敏感数据处理系统

CDS-Mask

技术白皮书

目录

[第1章 敏感数据处理现状 1](#_Toc518760506)

[第2章 系统概述 1](#_Toc518760507)

[第3章 产品功能 3](#_Toc518760508)

[3.1 敏感数据自动发现 3](#_Toc518760509)

[3.2 敏感数据变形处理 3](#_Toc518760510)

[3.3 异构数据库同步 3](#_Toc518760511)

[3.4 周期性数据处理 3](#_Toc518760512)

[3.5 敏感数据匿名化 4](#_Toc518760513)

[3.6 数据水印 4](#_Toc518760514)

[3.7 安全多方计算 4](#_Toc518760515)

[第4章 产品价值 6](#_Toc518760516)

[4.1 确保敏感数据安全 6](#_Toc518760517)

[4.2 脱敏数据高可用 6](#_Toc518760518)

[4.3 满足合规性需求 6](#_Toc518760519)

[4.4 确保泄密事件可追溯 6](#_Toc518760520)

# 敏感数据处理现状

大数据时代来临后，数据价值得到了空前的提升。金融通过大数据分析为细分客户推出专属理财产品、医疗通过诊疗信息临床药效、零售企业通过数据挖掘获得消费习惯信息、广告通过客户数据分析修正投放策略。这些收益促进不仅促使企业内部的数据挖掘业务、数据运维业务快速的增长，还开创了企业之间数据共享、协同分析等新场景。

在数据业务扩展的同时，企业数据将受到内外双重挑战。内部数据分析、运维场景容易发生窃密事件，据Verizon2017年发布的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，排名在前4的攻击模式（各种失误、犯罪软件、内部人员/权限滥用、物理偷窃/丢失）涵盖了90%的数据泄密事件，而这4种类型中有3种类型是内部人为的因素导致的数据泄露。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。

而外部数据共享则可能被合作伙伴泄密。在这类场景中，数据具备了共享、发布、外部流转等特性。因为业务原因，数据必须走出安全的环境，在企业之外接受各类未知因素的考验。各类涉及商业秘密和敏感数据信息在其他商家处理、共享和使用过程中面临大量违规越权使用或被用于非法用途等数据信息泄漏的安全风险。因此，大数据时代的数据安全已经不能依靠在企业自身环境中部署一些安全设备就能够实现抵御风险的目标了，必须依靠新的工具才能保障核心数据全生命周期安全。

同时，网络安全法、GDPR等法规对数据使用的强制性要求，也已成为企业不得不考量的重要因素。各国法律针对企业使用用户数据的要求越来越严，合规标准强制要求敏感数据的流转、使用、共享等多个环节必须无懈可击，给企业安全建设带来新的挑战。

# 系统概述

安全防护产品只能解决数据生命周期单个环节的问题，数据流转中任何一个步骤没有部署到位，都将导致数据泄密事故。因此，防护手段无法根治数据安全隐患。

腾讯云数隐敏感数据处理系统以处理数据库文件的方式，对数据库中的敏感数据进行在线的屏蔽、变形、字符替换、随机替换等脱敏操作，达到企业核心数据保密的效果。产品直接在数据内容上进行防护，处理过的数据不管在任何环境、任何流程中都不会造成泄密。在确保安全性的同时，处理过的数据依然保持原始数据的分布特征、数据格式，使统计分析、测试、研发、培训等用途不受影响。

产品首创的数据水印技术，能确保数据外泄之后，泄密的时间、泄密人范围、泄密途径被有效还原。实现泄密事件追溯的效果。

在合规性方面，数隐能够通过匿名化技术，确保脱敏后数据无法被还原。保障企业个人信息使用时严格遵守相关法律法规，满足企业客户合规性需求。

综上所述，整个系统能为客户带来以下价值：

● 减少生产库中敏感数据泄露的风险

● 减少开发、测试和数据交付过程中的数据泄漏风险

● 在泄密事件发生后进行追溯

● 轻松满足隐私数据管理的政策合规性

# 产品功能

## 敏感数据自动发现

数隐敏感数据处理系统支持多达29种内置的敏感数据识别规则，覆盖中美欧等多国有关个人信息保护的法律法规。在满足合规性的基础上，能够对数据集的所有字段进行敏感属性识别，确保隐藏在大段文本中的敏感信息能够得到妥善处理。

同时，产品还支持自定义敏感数据类型的自动发现，满足用户个性化的数据保护需求。

## 敏感数据变形处理

产品支持屏蔽、变形、移位、格式保留加密、令牌化、洗牌、强加密算法等多种脱敏算法。敏感数据可根据需求进行不同的脱敏处理。如对身份证进行格式保留，确保业务系统能够正常识别其属性；又如，对外键进行令牌化，确保处理后的字段与外键表之间保持一致性。

## 异构数据库同步

数隐敏感数据处理系统能够对MySql、MariaDB、TXSQL、TDSQL、SQLserver、Oracle、postgreSQL等数据库的数据进行脱敏处理。脱敏采用静态同步方式，将源数据库的数据抽取后，经过变形处理在装载到目标数据库中。此类方式灵活可靠，能完成异构数据库之间的同步。即目标数据库与源数据库不需要是同类型、同品牌的数据库，也可以完成脱敏任务。

## 周期性数据处理

对于企业网络而言，生产环境中的数据是不断变化的，需要脱敏的数据也是不断增加的。为此，数隐敏感数据处理系统支持周期性任务、全量任务以及增量任务。能够通过多种任务模式，将敏感数据灵活的从生产环境导入到开发测试环境中。对于频率较高的业务，数隐可用增量模式、短周期的快速更新脱敏数据；对于数据完整性要求较高的业务，则可通过全量模式、长周期的进行更新。

## 敏感数据匿名化

传统的脱敏技术只对数据进行一对一替换。这时，如果有获取脱敏数据的黑客，将其他外部数据集引入进行对比，依然有可能将脱敏后的数据进行破解还原，从而导致数据泄露。如以下医疗信息看似个人信息已脱敏，实则容易被破解：

姓名 年龄 单位 职位 所患病症

XX 99 中山大学 教授 癌症

上述信息姓名被隐藏了，但如果攻击者已知中山大学仅一位在职的99岁教授王某，那么姓名这一栏毫无疑问就是特指王某，那么王某患癌症的敏感信息就被破解还原了。

数隐敏感数据处理系统则可以通过匿名化方式，将部分标识数据进行泛化处理，使得某一段区间的数据行数一定不小于某个值。如上述例子，如果将多个90多岁的教授年龄均进行泛化处理，使得数据集有3个95岁的病患信息，则攻击者无法判断哪个数据对应的是王某。

GDPR、网络安全法等信息安全法规，对于个人信息有明确的“不可还原”要求。传统的脱敏算法遇到上述社工攻击依然会被还原，而数隐匿名化技术则可助力企业真正意义上完成数据安全合规性建设。

## 数据水印

数据脱敏后，虽然在泄露时能够极大的降低损失，但作为数据泄露事件，造成的影响还是需要尽可能降到最低。而通过水印技术可以将泄露的数据集进行外泄时间和嫌疑人的定位，缩小排查范围，保障泄密企业快速追查责任人。

通过快速的责任追查，能够了解泄密途径和数据的传播范围，从而更好的弥补漏洞，删除传播到外网的数据，最终更大程度的减少泄密事件所产生的负面影响。

## 安全多方计算

数隐敏感数据处理系统能够为云平台多个租户之间数据共享分析计算的场景进行安全保护。通过多个租户间线下协商密钥，并且将敏感数据加密上传的方式，数隐能够将多个租户的核心数据安全的放入隔离环境中进行联合分析，同时保障参与计算的任何一方都无法破解上传的数据。

这种安全共享分析计算的流程，既能够保护用户的敏感数据，又能使各企业融合关键信息，实现数据价值的提升。

# 产品价值

## 确保敏感数据安全

数隐敏感数据处理系统具备匿名化等高级脱敏技术，在数据误差不影响可用性的前提下，能够有效保障数据安全。相比传统脱敏技术，从根本上杜绝了撞库攻击的风险。

## 脱敏数据高可用

数隐敏感数据处理系统支持令牌化、格式保留加密等多种脱敏算法，能够保证多个有关联的表脱敏后数据内容完全相同，也能确保各类有严格格式要求的字段在脱敏前后格式保持一致。从技术上杜绝脱敏后数据无法装载，装载后报错等一系列异常情况，确保脱敏数据的高可用性。

## 满足合规性需求

真正意义上实现数据不可逆的脱敏。严格满足各国法律法规要求，保障用户个人信息不被泄露，确保企业的全球业务无法律风险。使业务部门在拓展新区域时无后顾之忧。

## 确保泄密事件可追溯

数隐敏感数据处理独特的数据水印系统，使得不同场景获得的数据具有不同的标记。泄露数据可以根据水印标记回溯泄密人和泄密途径，确保泄密事件被精确定位。