# 数据脱敏的“用”、“护”之道

## 一. 引子

“资本只有在流动中才带来价值，单纯存放起来只会贬值”，这似乎已经成了放之四海皆准的共识。在信息化大潮愈演愈烈的当下，数据和信息不啻为一种“新型资本”，和传统资本具有的特性相似，数据资产的价值在流转、共享、整合利用中逐渐显现并越发放大。举例来说：银行数据部门掌握的用户储蓄和消费信息，对内共享可以完善业务部门的信息化系统开发；对外则可为政府部门、征信机构等提供有效参考；甚至可以成为零售或制造业制定战略目标的有效参考依据。

然而数据的共享和流转带来的红利，远远无法冲散遮盖在数据保管者和拥有者心头的乌云。这种担忧来自对敏感信息在共享环节可能发生的泄密风险，更是来自敏感数据“合理使用”并且“安全防护”两者之间的矛盾。

于是，数据脱敏技术和专业脱敏产品应需而生，这类专业产品可以按照不同数据使用场合，对敏感数据进行变形处理，在脱敏处理的同时，不改变数据的类型、格式、含义、分布等使用特征，让用户不再因为深陷对安全的顾虑，而不得不割舍掉数据分享和流转带来的价值。

## 二. 典型、新型脱敏应用场景举例

场景一：高敏感度的社保数据如何安全使用

某省社保数据中心，业务系统的开发和测试全部交由外包人员完成。对于敏感数据，决策者陷入“给”还是“不给”的两难选择。假如不使用生产数据，而是由开发测试人员杜撰测试数据来进行上线前的模拟和程序功能设计，会导致测试用例不全面、功能覆盖不完整，可能造成系统上线后的风险和隐患。但是，如果直接将生产数据交给开发测试人员使用，又面临敏感数据泄露的风险。这种场景下，对敏感数据进行脱敏是唯一可行的办法。可以直接在生产环境下应用静态脱敏技术，对社保数据中参保人的姓名、联系方式、金额、年限等敏感信息进行脱敏处理，生成新的数据库，提供给开发测试方，不仅兼顾用户数据的可用性，又满足了用户数据使用的安全性需求。

场景二：如何对不同身份的访问者实时提供不同的脱敏数据

某运营商搭建的数据集中管理平台中，客户信息、通话记录、资费信息等全部集中在大数据分析环境中，面向内部提供数据查询分析服务，面向外部如监察机构、公安机关、政府部门等则提供数据检索接口。由于不同访问者的身份、权限差异，同样的数据对不同访问者的脱敏策略各不相同。 为了实时、 安全地向各方提供数据， 依靠静态脱敏技术已经不再合适。此时，动态脱敏服务器便成了连接数据中心和外界使用者之间的通道，对不同身份、不同权限的用户配置实时数据脱敏规则，让其可以恰如其“ 份 ”的访问数据。

结合两个典型场景可以看出，无论是静态脱敏还是动态脱敏，在解决敏感数据使用、共享的需求中，扮演着同样至关重要的角色。然而，从使用场景和发展方向来看，二者又会沿着截然不同的道路各自前行。

# 三． 静态脱敏——融入安管流程

静态脱敏技术的应用，其价值在于打造一份全新的、“ 高度仿真 ”的数据库，供非安全环境下使用。凭借着低门槛、易部署等特性，静态脱敏技术率先被用户所接受。

适用行业

在过去的一年中，这种数据处理方式先后被银行、证券、保险、社保、央企等行业所采纳，成为数据共享中的重要工具。

成熟的静态脱敏产品， 应当能够自动甄别目标库中的敏感数据， 如用户信息、 联系方式、消费信息等，并将其保存成敏感数据字典。基于数据字典中梳理出的数据分布及数据关系，按照不同场景配置不同的脱敏算法，为不同身份的访问者发放数据。

这种数据处理机制，可以解决外包、测试、开发以及数据分析场景中，满足数据使用需求的同时，保证数据的安全性。然而，要想将静态脱敏应用切实有效地落地，需要考虑的重点是，如何与现有的安全管理流程有效融合。

可以接触生产数据的人员，通常存在三类不同的身份 ：

首先是安全部门负责人， 这类人关注的重点是数据的安全合规， 那么在静态脱敏流程中，让其参与敏感数据的梳理无疑是最佳选择。同时，脱敏系统可以根据数据特征自动发现敏感数据和数据间的关联关系，辅助安全负责人对梳理结果进行核实确认，最终形成针对特定数据库的敏感数据字典，为后续数据脱敏与数据发放奠定坚实的基础。

第二类人员是脱敏数据的实际使用者，包含测试、开发及分析人员，他们更关心的是，脱敏后的数据质量能否满足工作需求。这类人群可以参与制定脱敏策略，确保同一份敏感数据能够满足不同使用需求，并按照他们的详细要求进行掩码处理。

第三类人员为数据运维部门，他们承担着日常数据维护、数据迁移等工作，发放敏感数据通常也由这类人员完成。引入静态脱敏产品进行数据流程管理后，运维人员可以根据不同需求场景，将源库中的敏感信息，按照已指定的脱敏策略进行数据变形，并发放至目标库。

将敏感数据静态脱敏应用于数据管理制度，可以对上述三类人员的工作内容进行有效整合，并为各环节工作提供相应的技术保障，将数据的安全管理落到实处。

# 四． 动态脱敏——数据共享之闸

如果说基于“ 迁移并构建新数据 ”的静态脱敏技术，为敏感数据的移交使用提供了技术保障，那么基于网络层的动态脱敏技术则为实时数据共享开辟了新的前景。前文场景二中提到，数据集中管理平台需要对外提供数据检索接口，满足来自外部的数据使用请求。为了满足数据访问的实时性和精确性，需要根据访问者的身份特征，在网络层实时动态地进行查询内容的掩码返回，确保在不改变底层数据存储的基础上，实现敏感数据的安全使用。

相对静态脱敏技术而言，动态脱敏是完全不同的技术路线，采用截然不同的处理机制。为了实现网络层的数据掩码返回，需要对不同数据库进行精确的协议解析。这种解析不仅要做到对数据库访问请求的有效拆解、准确判别访问者身份和访问的数据库对象，还要能够处理数据库返回结果，对其中的敏感信息进行有效的变形替换。动态脱敏的数据掩码技术与静态脱敏一致，也应提供部分遮蔽、随机替换、变更顺序等脱敏方式，并且遵循“ 高度仿真 ”这一原则。

由于动态脱敏设备作用于网络层，串接在访问者和敏感数据之间，因此产品需要兼顾高性能与稳定性，不会因脱敏处理导致性能的明显下降，同时兼容全系列数据库协议；另一方面，需要具备高并发条件下的抗压能力，拥有完善的容灾机制；这些都将成为衡量动态脱敏产品成熟度的标准。

在动态脱敏产品的实际应用场景中，用户可将该设备作为数据分享的唯一通路使用，在此前提下， 除了根据不同访问者身份配置脱敏策略， 使其恰如其“ 份 ”地访问权限内的数据，还应当提供对风险操作的判断能力与拦截能力，能够根据对象类型、操作特征、返回行数等因素阻断高危操作，充分发挥“ 数据共享之闸 ”的作用。

# 五. 总结

在数据安全领域，“禁止”和“防护”固然重要，但是如果背离了数据共享和合理使用的前提，那么数据的价值将大幅度下降。数据脱敏技术，正是兼顾数据“用”、“护”之道的有效手段。无论动态脱敏还是静态脱敏，在数据安全领域，将越发不可替代，真正为用户铸造安全、可靠、高效的数据使用环境。