(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111931240 A
(43)申请公布日 2020.11.13

(21)申请号 202011000344.4

(22)申请日 2020.09.22

(71)申请人 安徽长泰信息安全服务有限公司
地址 230000 安徽省合肥市高新区习友路3333号中国声谷国际智能语音产业园2-C栋13层

(72)发明人 廉明

(74)专利代理机构 合肥律众知识产权代理有限公司 34147
代理人 魏洁

(51)Int.Cl.
G06F 21/62(2013.01)
G06F 16/23(2019.01)
G06F 21/60(2013.01)
G06F 21/31(2013.01)

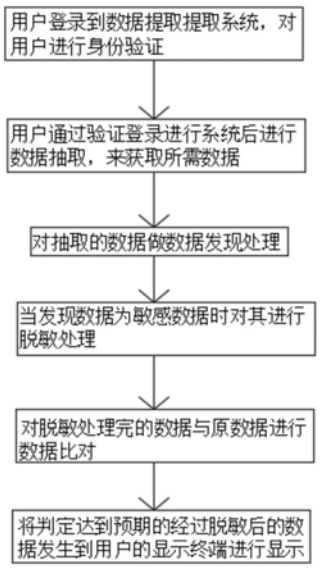
权利要求书2页 说明书4页 附图1页

(54)发明名称

一种用于保护敏感隐私数据的数据库脱敏方法

(57)摘要

本发明公开了一种用于保护敏感隐私数据的数据库脱敏方法,包括以下步骤:步骤一:用户登录到数据提取系统,对用户进行身份验证;步骤二:用户通过验证登录进行系统后进行数据抽取,来获取所需数据;步骤三:对抽取的数据做数据发现处理,识别出数据中的敏感数据类型,记录敏感数据的存放位置;步骤四:当发现数据为敏感数据时对其进行脱敏处理;步骤五:对脱敏处理完的数据与原数据进行数据比对,来判定数据经过脱敏后是否达到预期要求;步骤六:将判定达到预期的经过脱敏后的数据发生到用户的显示终端进行显示。本发明能够更好的对敏感隐私数据进行脱敏处理,满足了用户的不同使用需求。



CN 111931240 A

1.一种用于保护敏感隐私数据的数据库脱敏方法,其特征在于,包括以下步骤:

步骤一:用户登录到数据提取系统,对用户进行身份验证;

步骤二:用户通过验证登录进行系统后进行数据抽取,来获取所需数据;

步骤三:对抽取的数据做数据发现处理,识别出数据中的敏感数据类型,记录敏感数据的存放位置;

步骤四:当发现数据为敏感数据时对其进行脱敏处理;

步骤五:对脱敏处理完的数据与原数据进行数据比对,来判定数据经过脱敏后是否达到预期要求;

步骤六:将判定达到预期的经过脱敏后的数据发生到用户的显示终端进行显示。

2.根据权利要求1所述的一种用于保护敏感隐私数据的数据库脱敏方法,其特征在于:所述步骤二中的数据抽取过程如下:静态脱敏系统采用多种数据抽取方式,面向所有的数据库应用环境提供通用模式的数据处理接口,针对部分数据库应用环境提供了基于极速模式的数据处理接口,通用模式的数据处理接口最大限度的兼容用户的数据库应用场景。

3.根据权利要求1所述的一种用于保护敏感隐私数据的数据库脱敏方法,其特征在于:所述步骤三中的数据发现的过程如下:采用预读取的形式将数据库的结构信息缓存在本地,自动运行的数据扫描任务能够结合数据库结构信息,从数据源中获取少量的样本数据与内置的敏感数据指纹特征进行比对后,识别敏感数据类型,并记录敏感数据的存放位置。

4.根据权利要求1所述的一种用于保护敏感隐私数据的数据库脱敏方法,其特征在于:所述步骤四中的脱敏处理的具体过程如下:在数据抽取、脱敏计算、数据装载的过程中,所有的数据均在内存中进行处理,不在本地磁盘上记录任何敏感数据,坚持在脱敏过程中数据不落地的原则,提高整个数据使用链条的安全防护等级,通过脱敏规则体系,根据实际需要提供数据遮蔽、数据仿真、关键部分替换、随机字符串、重置固定值等多种多样的敏感数据处理方式,隐藏真实有效的敏感信息,自动扫描并发现敏感数据信息在生产环境中的位置、类型等信息,结合预设脱敏规则,提供了多种的敏感数据处理方案。

5.根据权利要求1所述的一种用于保护敏感隐私数据的数据库脱敏方法,其特征在于:所述步骤五中的数据比对过程如下:对脱敏后数据的数据库结构、数据对象、表数量、表内数据量等角度对脱敏任务进行分析对比,帮助用户校验脱敏任务是否完整的按照脱敏方案将所有数据装载到目标库,数据经过脱敏后是否达到预期要求,同时展示脱敏前和脱敏后的数据差异。

6.根据权利要求1所述的一种用于保护敏感隐私数据的数据库脱敏方法,其特征在于:所述步骤一的身份验证的具体过程如下:

S1:用户在登录数据提取系统时需要输入账号和密码进行验证;

S2:将用户输入账号的第一为的时间点标记为T1,将用户输入账号最后一位的时间点标记为T2;

S3:通过公式T2－T1＝$T_{差}$,得到最终的账号输入时长$T_{差}$;

S4:将用户输入密码的第一为的时间点标记为B1,将用户输入密码最后一位的时间点标记为B2;

S5:通过公式B2－B1＝$B_{差}$,得到最终的账密码输入时长$B_{差}$;

S6:再通过公式($B_{差}$+$T_{差}$)/(|$B_{差}$－$T_{差}$|)＝$Bt_{比}$,得到实时验证系数$Bt_{比}$;

S7：当用户输入的账号和密码均正确时对，即提取出实时验证系数$Bt_{比}$与预存的对比系数$Bt_{原}$进行比对；

S8：通过公式$Bt_{比}-Bt_{原}=Bt_{差}$，得到对比系数差值$Bt_{差}$；

S9：当对比系数差值$Bt_{差}$的绝对值为0或者对比系数差值$Bt_{差}$的绝对值小于预设值时即验证通过；

S10：当对比系数差值$Bt_{差}$的绝对值大于预设值时，用户即需要再次输入账号和密码再次采集实时验证系数$Bt_{比}$，并重复步骤S8的过程进行再次验证，再次验证不通过时，即当日无法在登录系统提取数据。

一种用于保护敏感隐私数据的数据库脱敏方法

技术领域

[0001] 本发明涉及景观设计领域，具体涉及一种用于保护敏感隐私数据的数据库脱敏方法。

背景技术

[0002] 数据脱敏是指对某些敏感信息通过脱敏规则进行数据的变形，实现敏感隐私数据的可靠保护。在涉及客户安全数据或者一些商业性敏感数据的情况下，在不违反系统规则条件下，对真实数据进行改造并提供测试使用，如身份证号、手机号、卡号、客户号等个人信息都需要进行数据脱敏。数据安全技术之一，数据库安全技术主要包括：数据库漏扫、数据库加密、数据库防火墙、数据脱敏、数据库安全审计系统。数据库安全风险包括：拖库、刷库、撞库。

[0003] 现有的数据脱敏方法，一般为由内部相关人员进行手工或者编写简单的函数进行脱敏占用大量人力和时间脱敏效率太低，数据准备的时效性低导致数据大量失真，虽然私密数据漂白了，但仿真度太低，影响测试效对客户人员技术水平要求比较高无法进行任务计划、断点续传等，给数据脱敏方法的使用带来了一定影响，因此，提出一种用于保护敏感隐私数据的数据库脱敏方法。

发明内容

[0004] 本发明所要解决的技术问题在于：如何解决现有的数据脱敏方法，一般为由内部相关人员进行手工或者编写简单的函数进行脱敏占用大量人力和时间脱敏效率太低，数据准备的时效性低导致数据大量失真，虽然私密数据漂白了，但仿真度太低，影响测试效对客户人员技术水平要求比较高无法进行任务计划、断点续传等，给数据脱敏方法的使用带来了一定影响的问题，提供了一种用于保护敏感隐私数据的数据库脱敏方法。

[0005] 本发明是通过以下技术方案解决上述技术问题的，本发明包括以下步骤：

[0006] 步骤一：用户登录到数据提取系统，对用户进行身份验证；

[0007] 步骤二：用户通过验证登录进行系统后进行数据抽取，来获取所需数据；

[0008] 步骤三：对抽取的数据做数据发现处理，识别出数据中的敏感数据类型，记录敏感数据的存放位置；

[0009] 步骤四：当发现数据为敏感数据时对其进行脱敏处理；

[0010] 步骤五：对脱敏处理完的数据与原数据进行数据比对，来判定数据经过脱敏后是否达到预期要求；

[0011] 步骤六：将判定达到预期的经过脱敏后的数据发生到用户的显示终端进行显示。

[0012] 优选的，所述步骤二中的数据抽取过程如下：静态脱敏系统采用多种数据抽取方式，面向所有的数据库应用环境提供通用模式的数据处理接口，针对部分数据库应用环境提供了基于极速模式的数据处理接口，通用模式的数据处理接口最大限度的兼容用户的数据库应用场景。

[0013] 优选的，所述步骤三中的数据发现的过程如下：采用预读取的形式将数据库的结构信息缓存在本地，自动运行的数据扫描任务能够结合数据库结构信息，从数据源中获取少量的样本数据与内置的敏感数据指纹特征进行比对后，识别敏感数据类型，并记录敏感数据的存放位置。

[0014] 优选的，所述步骤四中的脱敏处理的具体过程如下：在数据抽取、脱敏计算、数据装载的过程中，所有的数据均在内存中进行处理，不在本地磁盘上记录任何敏感数据，坚持在脱敏过程中数据不落地的原则，提高整个数据使用链条的安全防护等级，通过脱敏规则体系，根据实际需要提供数据遮蔽、数据仿真、关键部分替换、随机字符串、重置固定值等多种多样的敏感数据处理方式，隐藏真实有效的敏感信息，自动扫描并发现敏感数据信息在生产环境中的位置、类型等信息，结合预设脱敏规则，提供了多种多样的敏感数据处理方案。

[0015] 优选的，所述步骤五中的数据比对过程如下：对脱敏后数据的数据库结构、数据对象、表数量、表内数据量等角度对脱敏任务进行分析对比，帮助用户校验脱敏任务是否完整的按照脱敏方案将所有数据装载到目标库，数据经过脱敏后是否达到预期要求，同时展示脱敏前和脱敏后的数据差异。

[0016] 优选的，所述步骤一的身份验证的具体过程如下：

[0017] S1：用户在登录数据提取系统时需要输入账号和密码进行验证；

[0018] S2：将用户输入账号的第一为的时间点标记为T1，将用户输入账号最后一位的时间点标记为T2；

[0019] S3：通过公式$T2-T1=T_{差}$，得到最终的账号输入时长$T_{差}$；

[0020] S4：将用户输入密码的第一为的时间点标记为B1，将用户输入密码最后一位的时间点标记为B2；

[0021] S5：通过公式$B2-B1=B_{差}$，得到最终的账密码输入时长$B_{差}$；

[0022] S6：再通过公式$(B_{差}+T_{差})/(|B_{差}-T_{差}|)=Bt_{比}$，得到实时验证系数$Bt_{比}$；

[0023] S7：当用户输入的账号和密码均正确时对，即提取出实时验证系数$Bt_{比}$与预存的对比系数$Bt_{原}$进行比对；

[0024] S8：通过公式$Bt_{比}-Bt_{原}=Bt_{差}$，得到对比系数差值$Bt_{差}$；

[0025] S9：当对比系数差值$Bt_{差}$的绝对值为0或者对比系数差值$Bt_{差}$的绝对值小于预设值时即验证通过；

[0026] S10：当对比系数差值$Bt_{差}$的绝对值大于预设值时，用户即需要再次输入账号和密码再次采集实时验证系数$Bt_{比}$，并重复步骤S8的过程进行再次验证，再次验证不通过时，即当日无法在登录系统提取数据。

[0027] 本发明相比现有技术具有以下优点：该用于保护敏感隐私数据的数据库脱敏方法，数据通过脱敏系统多次分发并写入到不同的数据应用环境时，还可以保持数据的一致性，并且，满足不同数据在各类应用场景中的敏感识别要求，同时极速模式的数据处理接口能够显著的提升数据的处理性能，提高脱敏业务的时效性，从而让该方法能够满足用户的不同使用需求，并且该系统在用户调取数据前即对用户进行身份的验证，更进一步的提升了该方法的安全性，使得该方法能够更好的对敏感隐私数据进行保护。

附图说明

[0028] 图1是本发明的流程框图。

具体实施方式

[0029] 下面对本发明的实施例作详细说明，本实施例在以本发明技术方案为前提下进行实施，给出了详细的实施方式和具体的操作过程，但本发明的保护范围不限于下述的实施例。

[0030] 如图1，本实施例提供一种技术方案：一种用于保护敏感隐私数据的数据库脱敏方法，包括以下步骤：

[0031] 步骤一：用户登录到数据提取系统，对用户进行身份验证；

[0032] 步骤二：用户通过验证登录进行系统后进行数据抽取，来获取所需数据；

[0033] 步骤三：对抽取的数据做数据发现处理，识别出数据中的敏感数据类型，记录敏感数据的存放位置；

[0034] 步骤四：当发现数据为敏感数据时对其进行脱敏处理；

[0035] 步骤五：对脱敏处理完的数据与原数据进行数据比对，来判定数据经过脱敏后是否达到预期要求；

[0036] 步骤六：将判定达到预期的经过脱敏后的数据发生到用户的显示终端进行显示。

[0037] 所述步骤二中的数据抽取过程如下：静态脱敏系统采用多种数据抽取方式，面向所有的数据库应用环境提供通用模式的数据处理接口，针对部分数据库应用环境提供了基于极速模式的数据处理接口，通用模式的数据处理接口最大限度的兼容用户的数据库应用场景，极速模式的数据处理接口能够显著的提升数据的处理性能，提高脱敏业务的时效性。

[0038] 所述步骤三中的数据发现的过程如下：采用预读取的形式将数据库的结构信息缓存在本地，减少对数据库的并发查询性能消耗，自动运行的数据扫描任务能够结合数据库结构信息，从数据源中获取少量的样本数据与内置的敏感数据指纹特征进行比对后，识别敏感数据类型，并记录敏感数据的存放位置。

[0039] 所述步骤四中的脱敏处理的具体过程如下：在数据抽取、脱敏计算、数据装载的过程中，所有的数据均在内存中进行处理，不在本地磁盘上记录任何敏感数据，坚持在脱敏过程中数据不落地的原则，提高整个数据使用链条的安全防护等级，通过脱敏规则体系，根据实际需要提供数据遮蔽、数据仿真、关键部分替换、随机字符串、重置固定值等多种多样的敏感数据处理方式，隐藏真实有效的敏感信息，自动扫描并发现敏感数据信息在生产环境中的位置、类型等信息，结合预设脱敏规则，提供了多种多样的敏感数据处理方案，脱敏后的数据在测试、分析场景中的仍然具有可用性、规范性以及“真实性”，当相同的数据通过脱敏系统多次分发并写入到不同的数据应用环境时，还可以保持数据的一致性。

[0040] 所述步骤五中的数据比对过程如下：对脱敏后数据的数据库结构、数据对象、表数量、表内数据量等角度对脱敏任务进行分析对比，帮助用户校验脱敏任务是否完整的按照脱敏方案将所有数据装载到目标库，数据经过脱敏后是否达到预期要求，同时展示脱敏前和脱敏后的数据差异。

[0041] 所述步骤一的身份验证的具体过程如下：

[0042] S1：用户在登录数据提取系统时需要输入账号和密码进行验证；

[0043] S2:将用户输入账号的第一为的时间点标记为T1,将用户输入账号最后一位的时间点标记为T2;

[0044] S3:通过公式T2－T1＝$T_{差}$,得到最终的账号输入时长$T_{差}$;

[0045] S4:将用户输入密码的第一为的时间点标记为B1,将用户输入密码最后一位的时间点标记为B2;

[0046] S5:通过公式B2－B1＝$B_{差}$,得到最终的账密码输入时长$B_{差}$;

[0047] S6:再通过公式($B_{差}$+$T_{差}$)/(|$B_{差}$－$T_{差}$|)＝$Bt_{比}$,得到实时验证系数$Bt_{比}$;

[0048] S7:当用户输入的账号和密码均正确时对,即提取出实时验证系数$Bt_{比}$与预存的对比系数$Bt_{原}$进行比对;

[0049] S8:通过公式$Bt_{比}$－$Bt_{原}$＝$Bt_{差}$,得到对比系数差值$Bt_{差}$;

[0050] S9:当对比系数差值$Bt_{差}$的绝对值为0或者对比系数差值$Bt_{差}$的绝对值小于预设值时即验证通过;

[0051] S10:当对比系数差值$Bt_{差}$的绝对值大于预设值时,用户即需要再次输入账号和密码再次采集实时验证系数$Bt_{比}$,并重复步骤S8的过程进行再次验证,再次验证不通过时,即当日无法在登录系统提取数据。

[0052] 此外,术语“第一”、“第二”仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本发明的描述中,“多个”的含义是至少两个,例如两个,三个等,除非另有明确具体的限定。

[0053] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不必须针对的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外,在不相互矛盾的情况下,本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0054] 尽管上面已经示出和描述了本发明的实施例,可以理解的是,上述实施例是示例性的,不能理解为对本发明的限制,本领域的普通技术人员在本发明的范围内可以对上述实施例进行变化、修改、替换和变型。

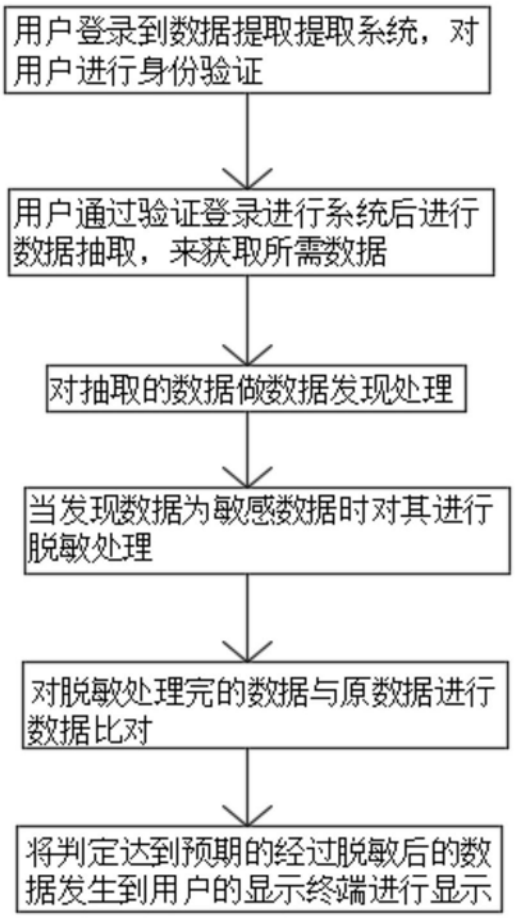

用户登录到数据提取提取系统，对用户进行身份验证
用户通过验证登录进行系统后进行数据抽取，来获取所需数据
对抽取的数据做数据发现处理
当发现数据为敏感数据时对其进行脱敏处理
对脱敏处理完的数据与原数据进行数据比对
将判定达到预期的经过脱敏后的数据发生到用户的显示终端进行显示

图1