数据脱敏管理

数据脱敏方案生态对比分析

竞争分析 - 品牌

目前市场存在的脱敏产品

- IBM
  - Optim Test Data Management(TDM)产品
    - 纯软件
    - 专用的TDM产品
  - InfoSphere Information Server(Datastage)
    - 纯软件
    - ETL工具，加上漂白套件实现
- Informatica
  - Test Data Management
    - 纯软件
    - ETL工具PowerCenter外加漂白套件实现
- Oracle
  - OEM Data Masking
    - 纯软件
    - Oracle数据库的企业选件
- 云图数据
  - InfoMask
    - 软硬一体产品

竞争分析

特性	IBM OPTIM	IBM Datastage	Oracle Data Masking	Informatica TDM	其它国产	云图Infomask
即插即用	✗	✗	✗	✗	✗	✓
流程化管理	✓	✓	✗	✓	✓	✓
隐私数据发现功能	✓	✗	✗	✓	✓	✓
隐私数据梳理功能	✗	✗	✗	✗	✗	✓
本地数据漂白算法	✗	✗	✗	✗	✓	✓
异种数据库迁移	✗	✓	✗	✓	✓	✓
良好的扩展性	✓	✓	✗	✓	✗	✓
流程中的数据安全性	✗	✗	✗	✗	✗	✓
版本管理	✓	✗	✗	✗	✗	✓
数据压缩	✓	✗	✗	✗	✗	✓

✓ 完全支持　　✓ 部分支持　　✗ 不支持

产品特征

IBM OPTIM	IBM Datastage	Oracle Data Masking	Informatica TDM	其它国产	云图Infomask
• 纯软件 • 专用的TDM产品	• 纯软件 • ETL工具，加上漂白套件实现	• 纯软件 • Oracle数据库的企业选件 • 通过数据库内部的程序包、函数来实现	• 纯软件 • ETL工具PowerCenter外加漂白套件实现	• 整合的应用平台 • 大多OEM了外企的产品作业核心架构 • 增加了一些行业客户化的功能	• 软硬一体产品 • 专用TDM产品

项目成本

IBM OPTIM	IBM Datastage	Oracle Data Masking	Informatica TDM	其它国产	云图Infomask
• 软件产品 • 服务器 • 存储 • 上述产品安装实施 • 数据梳理 • 二次开发	• 软件产品 • 服务器 • 存储 • 上述产品安装实施 • 数据梳理 • 二次开发	• OEM选件费用 • 上述产品安装实施 • 数据梳理 • 二次开发	• 软件产品 • 服务器 • 存储 • 上述产品安装实施 • 数据梳理 • 二次开发	• 平台产品 • OEM的外企软件 • 服务器 • 存储 • 上述产品安装实施 • 数据梳理 • 二次开发	• 产品费用 • 产品安装实施

流程化管理

IBM OPTIM	IBM Datastage	Oracle Data Masking	Informatica TDM	其它国产	云图Infomask
• 主张标准的抽取、漂白、装载的流程 • 数据需要落地 • 流程无法自动调度 • 非实时漂白	• 由于是ETL工具，其流程完全松散 • 流程控制完全靠二次开发实现 • 流程无法自动调度 • 数据可以落地也可以不落地 • 实时漂白	• 存储过程，完全没有流程化概念 • 只有从克隆库到目标或者直接源端到目标的过程	• 由于是ETL工具，其流程完全松散 • 流程控制完全靠二次开发实现 • 流程无法自动调度 • 数据可以落地也可以不落地 • 实时漂白	• 基本都是抽取、脱敏、装载流程 • 都需要手工建立数据模型才能做 • 部分支持本地脱敏 • 有的需要数据落地，有的不需要	• 主张标准的自动发现、抽取、漂白、装载流程 • 拥有自动化调度功能 • 每个阶段均可单独调度、关联调度、并行调度 • 可定义工作时间暂停 • 数据需要落地 • 非实时漂白

隐私数据发现（数据模型问题）

IBM OPTIM	IBM Datastage	Oracle Data Masking	Informatica TDM	其它国产	云图Infomask
• 仅支持用户自定义正则表达式，准确率极低，完全无法达到用户要求 • 需要手工建模	• 不支持 • 需要手工建模	• 不支持 • 需要手工建模	• 仅支持用户自定义正则表达式，准确率极低，完全无法达到用户要求 • 需要手工建模	• 少数支持自动发现功能，大多不支持 • 使用正则表达式方式或者根据字段名模糊判断，准确率极低 • 需要手工建模	• 内置11种中国地区专有隐私数据类型发现规则,准确率99%以上 • 支持用户自定义正则表达发现 • 支持用户自定义程序发现 • 准确率高，可以省去建模工作

隐私数据梳理

IBM OPTIM	IBM Datastage	Oracle Data Masking	Informatica TDM	云图Infomask
• 无	• 无	• 无	• 无	• 包含一个专业工具，协助客户进行梳理工作，了解数据关系和数据含义，极大提高工作效率

数据漂白脱敏

IBM OPTIM	IBM Datastage	Oracle Data Masking	Informatica TDM	其它国产	云图Infomask
• 包含通用变形函数 • 包含美国专用变形函数 • 中国隐私数据需二次开发，通过用户出口嵌入代码 • 漂白效果：完全看客户化和二次开发的质量 • 需要根据数据模型逐一字段进行手工配置	• 包含通用变形函数 • 包含美国专用变形函数 • 中国隐私数据需二次开发，通过用户出口嵌入代码 • 漂白效果：完全看客户化和二次开发的质量 • 需要根据数据模型逐一字段进行手工配置	• 包含通用变形函数 • 包含美国专用变形函数 • 中国隐私数据需二次开发，形成数据库PL/SQL代码实现 • 漂白效果：完全看客户化和二次开发的质量 • 需要根据数据模型逐一字段进行手工配置	• 包含通用变形函数 • 包含美国专用变形函数 • 中国隐私数据需二次开发，通过用户出口嵌入代码 • 漂白效果：完全看客户化和二次开发的质量 • 需要根据数据模型逐一字段进行手工配置	• 包含通用变形函数，受到OEM外企产品的限制 • 包含少量中国专用隐私数据 • 漂白效果：一般 • 需要根据数据模型逐一字段进行手工配置	• 包含通用变形函数 • 包含11种中国专用隐私数据 • 支持用户自定义隐私数据类型 • 漂白效果：仿真度极高，且允许客户定制规则 • 漂白任务配置完全自动化，无需手工配置

数据装载

IBM OPTIM	IBM Datastage	Oracle Data Masking	Informatica TDM	其他国产	云图Infomask
• 低速：用数据库INSERT命令 • 完全手工配置分发实现并行 • 高速：落地成文件，用户自己通过别的工具装载 • 效率低	• 低速：用数据库INSERT命令 • 通过配置参数实现并行 • 高速：落地成文件，用户自己通过别的工具装载 • 效率较高	• 使用数据泵技术 • 效率较高	• 低速：用数据库INSERT命令 • 通过配置参数实现并行 • 高速：落地成文件，用户自己通过别的工具装载 • 效率较高	• 依赖于所OEM的外企产品 • 效率较高	• 通过数据库bulk insert实现 • 完全自动化实现并行 • 效率较高

异构数据库支持

IBM OPTIM	IBM Datastage	Oracle Data Masking	Informatica TDM	其它国产	云图Infomask
• 不支持	• 支持数据转换 • 对象映射需要全手工配置 • 不支持对象定义转换	• 不支持	• 支持数据转换 • 对象映射需要全手工配置 • 不支持对象定义转换	• 支持数据转换 • 对象映射需要全手工配置 • 不支持对象定义转换	• 支持数据转换 • 支持数据库对象定义转换 • 转换、映射完全后台自动化进行，对用户完全透明

访问控制

IBM OPTIM	IBM Datastage	Oracle Data Masking	Informatica TDM	其它国产	云图Infomask
• 对落地的数据文件进行访问控制	• 无	• 使用Oracle数据库访问控制机制	• 无	• 经过二次开发，已经具有完整访问控制机制	• 通过用户、角色、菜单权限、数据权限等机制实现完整的访问控制

性能

IBM OPTIM	IBM Datastage	Oracle Data Masking	Informatica TDM	其它国产	云图Infomask
• 抽取：默认单线程，需要手工配置多个作业实现并发 • 漂白：通用函数性能缓慢，中国数据漂白依赖二次开发 • 装载：默认单线程，需要手工配置多个作业实现并发 • 总体性能比较慢，代码是在Windows开发的，移植到linux/unix的时候使用了虚拟机，所以效率很低	• 抽取：通过配置实现并发 • 漂白：通过配置实现并发，中国数据漂白依赖二次开发代码质量 • 装载：通过配置实现并发 • 总体性能比较高，但是资源消耗也比较高	• 抽取：无 • 漂白：没有并发，完全看二次开发效率 • 装载：使用数据泵技术传输数据	• 抽取：通过配置实现并发 • 漂白：通过配置实现并发，中国数据漂白依赖二次开发代码质量 • 装载：通过配置实现并发 • 总体性能比较高，但是资源消耗也比较高	• 抽取：支持并发 • 漂白：效率较低 • 装载：通过配置实现并发 • 总体性能一般，漂白脱敏效率不行，抽取和装载效率依赖于OEM的产品性能	• 抽取：自动启用并发 • 漂白：缺省使用并发，漂白效率较高，实测中档机型GM2000最高每秒处理12万个隐私数据 • 装载：缺省启用并发，使用bulk操作

数据安全

IBM OPTIM	IBM Datastage	Oracle Data Masking	Informatica TDM	其它国产	云图Infomask
• 有一定的安全措施，但是所有落地文件均可被单独获取	• 流程完全开放，可以被任意修改，没有很好的安全保护措施	• 漂白前后的数据均为完全访问	• 流程完全开放，可以被任意修改，没有很好的安全保护措施	• 整合型平台，每个系统构件相对松散 • OEM产品则可以绕过应用层直接使用后台的外企产品	• 流程固化，每个阶段的数据均被控制 • 设备为封闭环境 • 数据除了通过自身系统平台进行访问，不允许通过别的途径主动直接访问内部数据

用户界面

IBM OPTIM	IBM Datastage	Oracle Data Masking	Informatica TDM	其它国产	云图Infomask
• C的界面，有点复杂 • 历史版本界面多年没有更新	• ETL工具的开发配置界面	• OEM简单的配置界面，没有任何监控	• ETL工具的开发配置界面	• 风格各异	• 基于WEB的用户门户界面，流程化的使用风格，基本上鼠标就可以做完任务 • CLI的后台配置界面，极少需要用到