



# 利用Exadata进行数据脱敏

最小化风险，最大化价值，充分利用Exadata资产

---

**Jim Kong**

Oracle SEHub Solutions



## Safe harbor statement

---

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.



# 内容

---

- 1 数据、数据价值和安全挑战
- 2 Oracle数据脱敏和子集的能力
- 3 在Exadata上进行数据脱敏
- 4 一次数据脱敏性能测试分享



# 数据是当今的资产

世界上最有价值的资源不再是石油，而是数据

## 数据驱动

- 分析和自动化
- 广告和营销预算
- 个性化和改进的体验
- 业务分析和决策
- 政府政策和计划

总体而言，数据有助于改进产品和服务，提供更好的用户体验，并支持和发展业务



个人身份信息数据  
财务数据  
商业机密  
竞争数据  
就业数据  
医疗保健数据  
IT安全数据  
交易数据  
浏览数据...



# 数据可能是一种负担

数字经济的可怕一面

**数据泄露**在全球范围内呈爆炸式增长

- 数据库是涉及违规的最常见资产

数据丢失对受影响的企业来说可能是**灾难性的**

- 因补偿、处罚、法律诉讼、公关、股票价值、恢复成本而产生的成本
- 品牌名誉、客户信任度、知识产权、竞争力
- 整体业务和收入

快速发展、严格的**监管环境**

- 跨行业和地区
- 旨在保护数据和公民隐私的法律



# 敏感数据的激增增加了安全风险



# 各种合规规则要求保护敏感数据

PCI DSS 3.2.1 May 2018



6.4.3 生产数据（实时 PAN）不用于测试或开发目的。

## 欧盟通用数据保护条例



对个人数据应用假名可以降低相关数据主体的风险，并帮助控制者和处理者履行其数据保护义务

## 2021年中国数据安全法，个人信息保护法



使用数据脱敏技术，可以有效地减少敏感数据在采集、传输、使用等环节中的暴露，降低敏感数据泄露的风险，尽可能降低数据泄露造成的危害

**... and more**



# 困境

做还是不做

## 您的愿望

- 从数据中获取可行的见解以做出更明智的业务决策
- 使用实际数据进行开发和分析
- 与开发人员，数据科学家和合作伙伴快速共享数据

解决方案？

## 您的担心

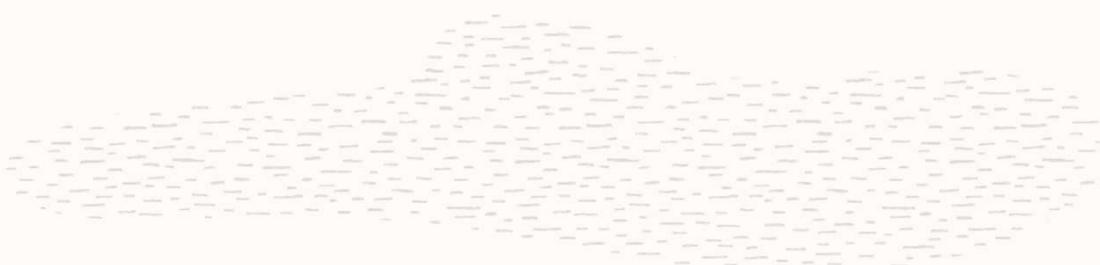
- 避免将敏感数据传播到非生产环境
- 符合数据隐私法规（如GDPR）
- 时间和存储成本最小



# 内容

---

- 1 数据、数据价值和安全挑战
- 2 Oracle数据脱敏和子集的能力
- 3 在Exadata上进行数据脱敏
- 4 一次数据脱敏性能测试分享



# 数据脱敏与子集

## 好的解决方案的特征

### 数据脱敏

- 与敏感数据发现集成，帮助了解现有的敏感数据
- 提供全面而灵活的屏蔽选项，可以生成实际数据以满足各种需求
- 保留应用程序完整性
- 易于使用，无需特别的技术技能

### 数据子集

- 仅提取相关数据并丢弃不必要的数据
- 提供不同子集选项，以满足各种业务需求
- 保留应用程序完整性
- 与数据屏蔽相集成以提供全面的解决方案

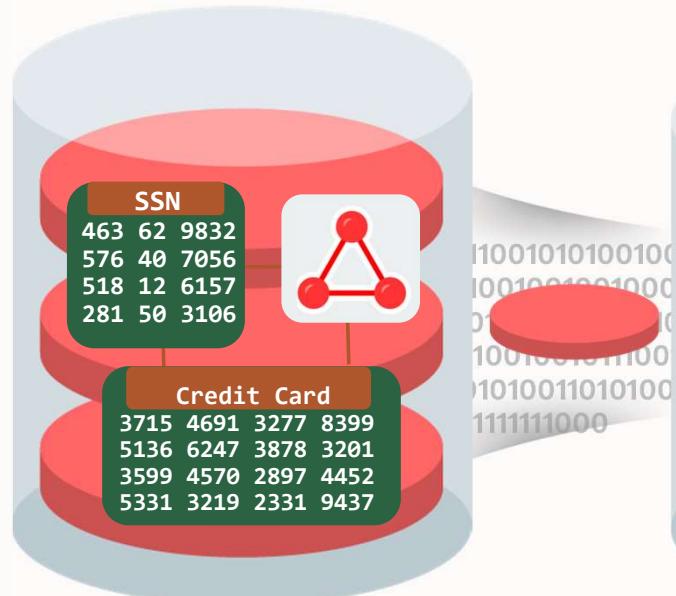


# 数据脱敏和子集

最大限度地减少敏感数据向非生产环境扩散

ORACLE  
ENTERPRISE MANAGER 13c

## 生产环境



## 非生产环境



## 敏感数据发现

全面的脱敏选项

基于目标/条件的子集

数据库内或导出时脱敏

支持云或非Oracle数据库

工作负载捕获和克隆脱敏

预装在企业管理器OEM中



# 敏感数据发现

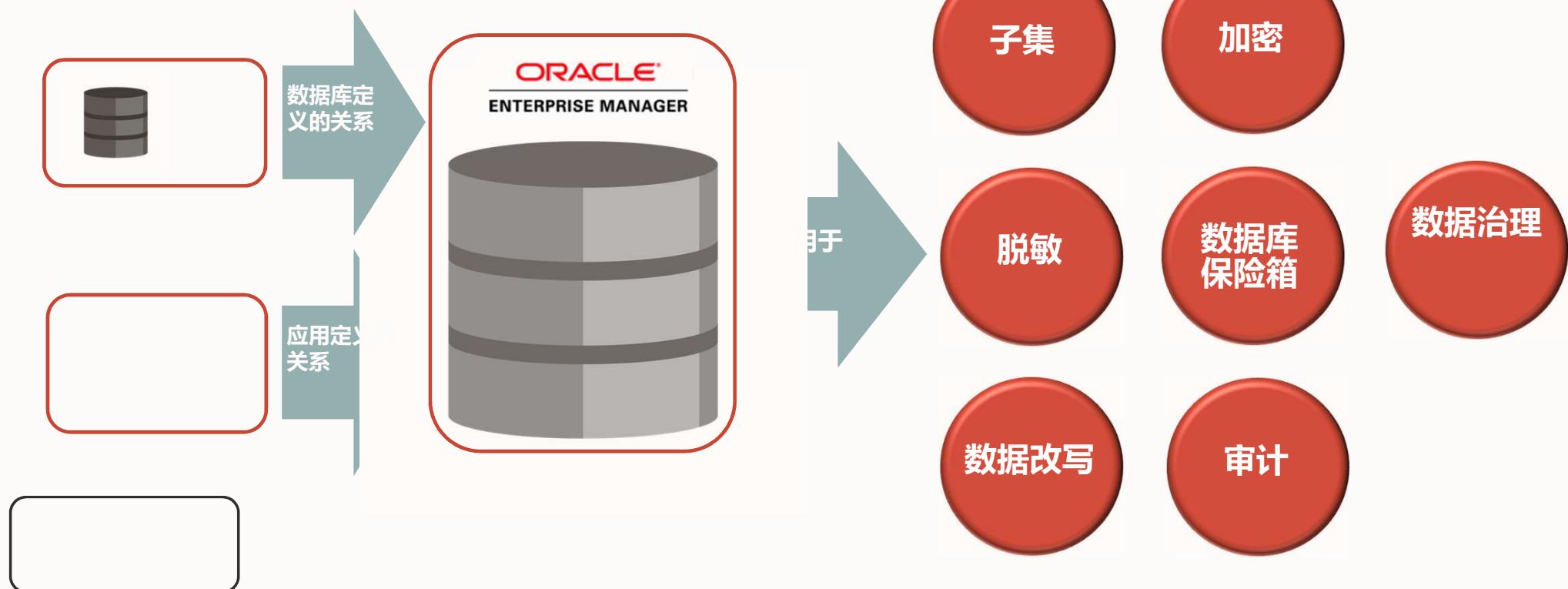
了解有什么风险



身份	特征	IT	财务	医疗	工作	教育背景
身份证号	年龄	IP 地址	信用卡	医院	员工 ID	大学
姓名	性别	用户ID	社保号	保险	职位	年级
Email	籍贯	密码	银行名称	身高	部门	学生号
电话	城市	主机名称	银行账户	血型	受雇日期	金融帮助
护照	地址	定位信息	...	是否残疾	收入	入学日期
...	家庭数据	...	...	是否怀孕	股票	毕业日志
	出生日期			测试结果	...	出勤率
	出生地			...		...
	...					

# 应用数据模型

## 敏感数据发现



# 数据脱敏

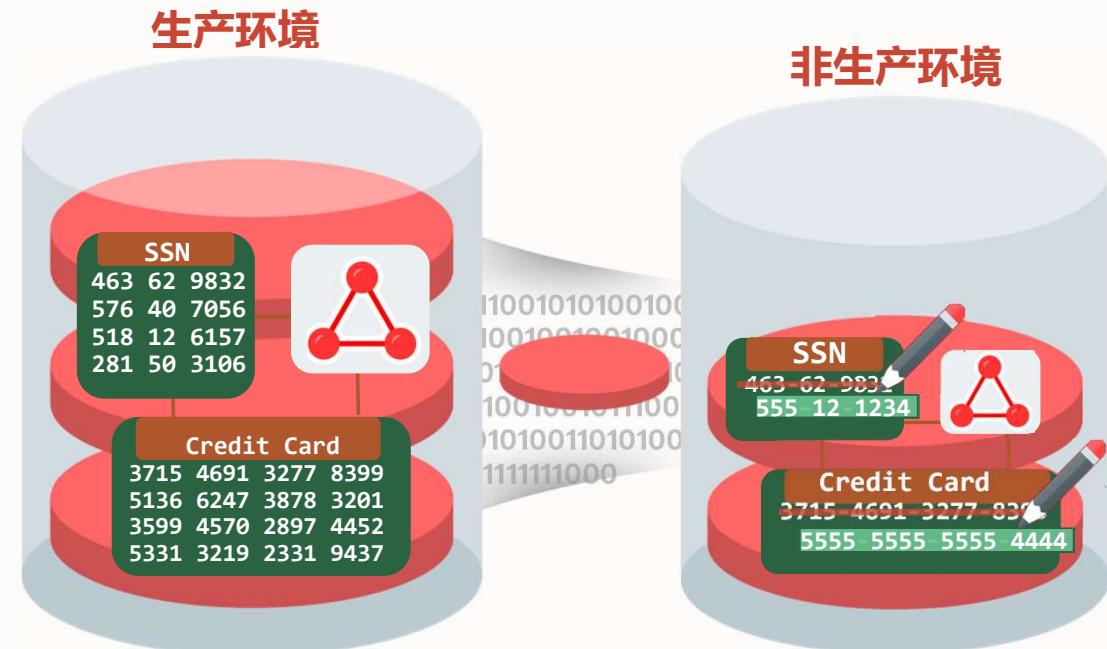
通过脱敏敏感数据将风险降至最低

用虚构但真实的数据替换敏感数据，这些数据对分析、开发等仍然可用

保持数据完整性，帮助确保应用程序继续使用被脱敏的数据

提供全面灵活的脱敏选项，以满足不同的需求

开箱即用的脱敏模板



# 数据脱敏

## 全面且灵活的脱敏格式

### 常见的预定义脱敏格式

- 信用卡号 Credit Card Number
- 社保号码 Social Security Number
- 保险号码 National Insurance Number
- ... 更多

### 灵活地定制脱敏格式

- 固定 number / string
- 随机 numbers / strings / dates / list
- 替换、加密、乱序、空值
- 用户自定义PL/SQL 函数
- ... 更多

样本有助于预览和验证脱敏数据

Format	Description
American Express Credit Card Number	~10 billion unique American Express card numbers
Discover Card Credit Card Number	~10 billion unique Discover card numbers
MasterCard Credit Card Number	~10 billion unique MasterCard card numbers
Visa Credit Card Number	~10 billion unique Visa card numbers
Generic Credit Card Number	~10 billion unique generic credit card numbers
Generic Credit Card Number Formatted	~10 billion unique generic credit card numbers, formatted
National Insurance Number Formatted	~10 billion unique National Insurance numbers, formatted
Social Insurance Number	~1 billion unique Social Insurance numbers
Social Insurance Number Formatted	~1 billion unique Social Insurance numbers, formatted
Social Security Number	~71 billion unique Social Security numbers
Social Security Number Formatted	~71 billion unique Social Security numbers, formatted
ISBN (Ten Digit)	~1 billion unique ISBN (Ten Digit) numbers
ISBN (Ten Digit) Formatted	~1 billion unique ISBN (Ten Digit) numbers, formatted
ISBN (Thirteen Digit)	~2 billion unique ISBN (Thirteen Digit) numbers
SQL Expression	Custom SQL expression for generating sensitive data
Substitute	Replace specific characters or patterns in the data
Substring	Extract a portion of the data as the output
Table Column	Reference a column from a database table
Truncate	Remove all characters from the data except the first few
User Defined Function	Custom function written in PL/SQL to generate sensitive data



# 数据脱敏

脱敏变换以满足不同的业务用例

<b>条件脱敏</b>	根据条件以不同方式脱敏多行数据 示例：根据国家/地区脱敏国家标识符
<b>确定性脱敏</b>	跨多个数据库将数据脱敏为相同的一致值 示例：跨架构和数据库一致地脱敏员工标识符
<b>复合脱敏</b>	确保相关列之间的脱敏值保持相同的关系 示例：将州、邮政编码和国家等地址字段脱敏为一个组
<b>格式保存</b>	脱敏数据，同时保留其格式，例如长度和特殊字符 示例：在保留空格和连字符的同时脱敏税务标识符
<b>可逆脱敏</b>	使用加密密钥加密和解密数据 示例：从合作伙伴处接收到处理后的数据后取消脱敏
<b>洗牌</b>	打乱列中的值 示例：组织中员工的随机年龄
<b>扰动</b>	在用户提供的范围内生成随机值 示例：在指定数据范围内生成随机日期

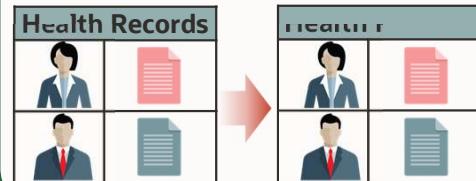


# 数据脱敏 示例

基于条件脱敏



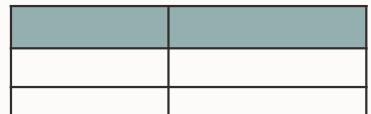
打乱顺序



生成确定性输出

HR

在保留格式的同时生成随机值




脱敏LOB列



# 数据脱敏

## 脱敏模板

### Oracle E-Business Suite 模板

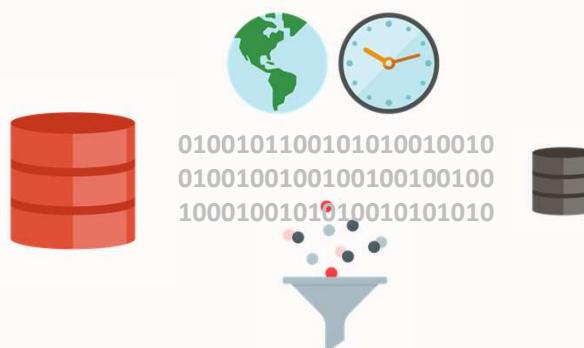
- 适用于特定版本
  - HCM, FIN, ATG, MOC, MSC, OKL, OPM, PRC, PROJ, and TCA
  - 大约 1000 列
- 脱敏关键的 PII 元素
  - 姓名、位置、电话、账户、驾驶执照、身份证件、网站、薪酬、国籍、健康信息、审计信息……
- 包括预定义的数据模型和脱敏规则
- 包含在产品许可中

EBS Masking MOS Note [2076834.1](#)



# 数据子集

为什么要取数据子集？



通过仅与内部和外部团队共享相关数据来最大程度地降低风险



降低非生产环境中的存储成本



从 SaaS 中提取订阅者数据



对数据子集进行研究和分析



提取数据子集作为电子发现请求的一部分

# 数据子集

基于目标或条件的子集

相关表的大小

基于条件



Extract ASIA Sales



表分区



# 数据子集

## 预览和验证结果

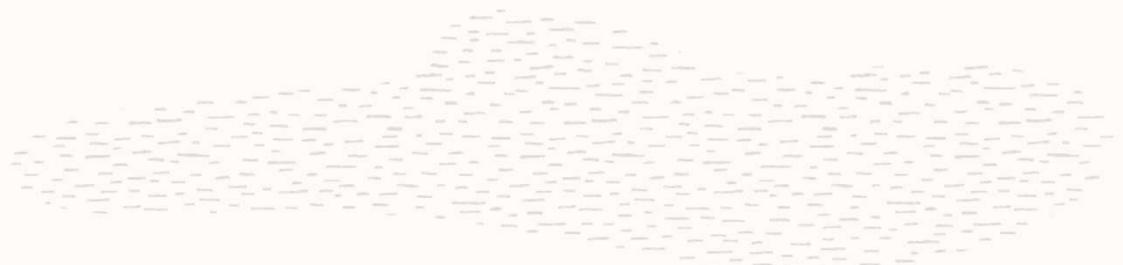
Space Estimates						
Name	Object Rule	Source Size		Estimated Subset Size		
		MB	Rows	MB	Rows	%
Applications and Objects		0.01	215	0.01	113	53
HR(HR)		0.01	215	0.01	113	53
EMPLOYEES	DEPARTMENT_ID=80	0.01	107	0	59	55
DEPARTMENTS		0	27	0	16	59
REGIONS		0	4	0	3	75
JOB_HISTORY		0	10	0	10	100
LOCATIONS		0	23	0	3	13
COUNTRIES		0	25	0	3	12
JOB		0	19	0	19	100



# 内容

---

- 1 数据、数据价值和安全挑战
- 2 Oracle数据脱敏和子集的能力
- 3 在Exadata上进行数据脱敏
- 4 一次数据脱敏性能测试分享

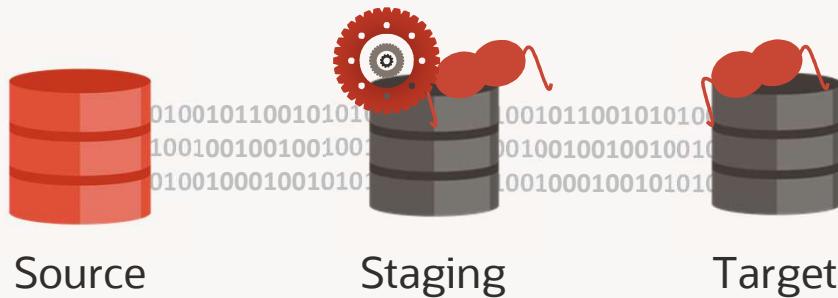


# 数据脱敏的两种模式

根据场景灵活选择

## 数据库内脱敏 In-Database Mode

Clone → Mask/Subset → Replicate



## 导出时脱敏 In-Export Mode

Clone → Mask/Subset → Replicate



对生产环境影响最小

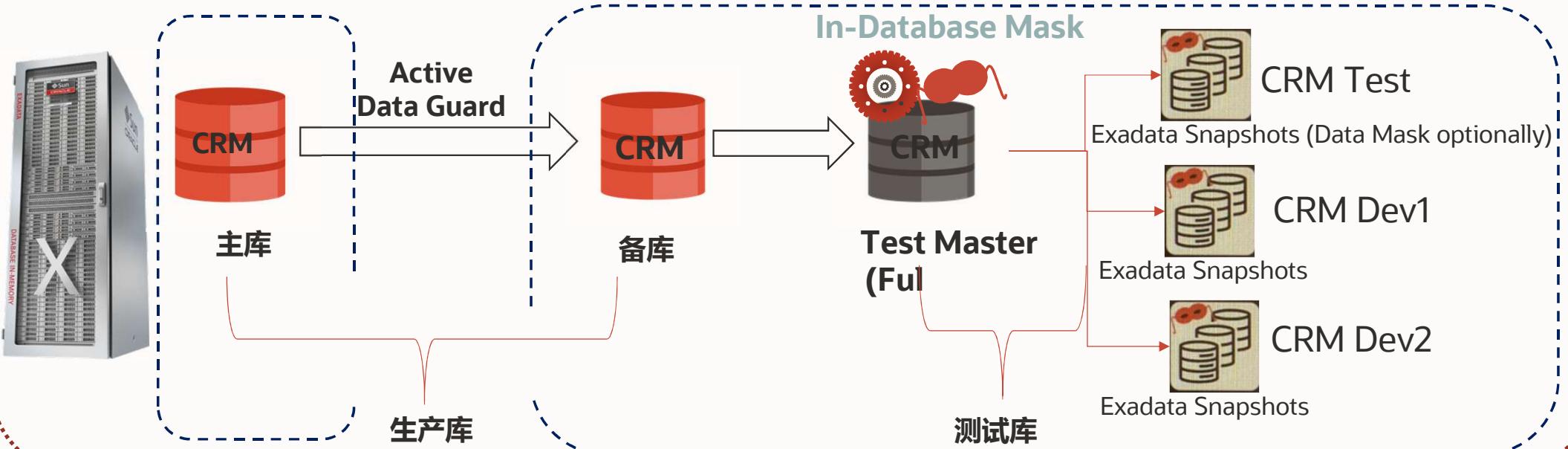
敏感数据保留在生产范围内



# 在Exadata上进行数据脱敏/屏蔽

屏蔽敏感数据 保障数据安全

## Exadata Environment



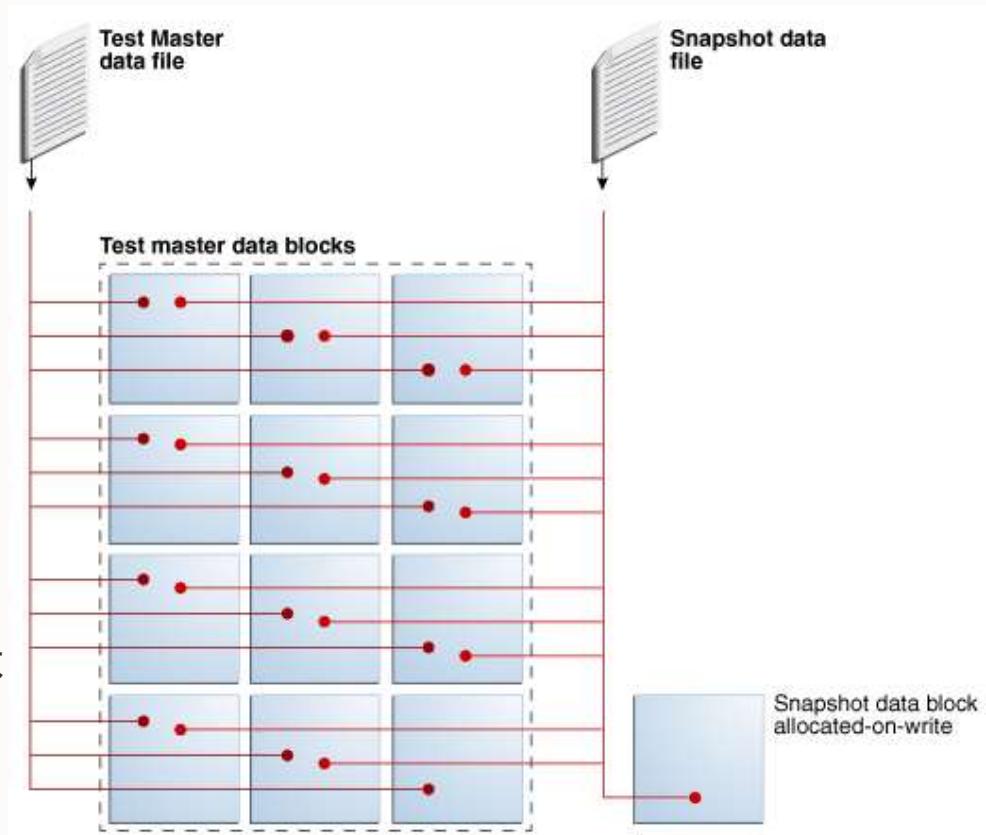
最小化对生产环境的影响

最大化数据库副本的价值

充分利用Exadata空间和计算能力

# Exadata 如何加速Data Masking

- ✓ **空间节省**, Exadata Snapshot 基于sparse disk group, 仅仅在变更的数据上占用额外存储空间, 而对用户/应用看来是一个独立的数据库
- ✓ **性能加速**, Exadata Sanpshot作为一个数据库对象, 能充分利用Exadata特有的smart scans, smart flash cache, resource management等特性
- ✓ 依靠Exadata**强大的硬件配置**, 优化数据更新的SQL读写能力
- ✓ 通过**PDB结合Test/Dev生命周期**, 方便快捷针对不同应用需求分发数据, 灵活调整masking规则



# 总结

## 主要优势

- 在不影响敏感数据的情况下最大化数据的商业价值
- 通过不扩散敏感数据将安全风险降至最低
- 通过脱敏敏感数据来避免违规
- 通过子集降低测试/开发的存储成本
- 通过Exadata Sparse Clone 加速脱敏性能
- 充分利用Exadata Rack资源，实现资源最大化利用

## 主要特征

- 自动发现敏感数据并保持参照完整性
- 提供全面而灵活的脱敏选项
- 支持基于目标和条件的子集
- 与企业数据管理工具(OEM)集成
- 与其他数据库组件集成使用



# 数据脱敏 与其他组件集成使用

集成	好处	
Database Lifecycle Management Pack	集成的克隆和脱敏	
Oracle Cloud Management Pack	集成的快照克隆和脱敏	
Real Application Testing Pack	在数据库中脱敏	
Oracle Data Integrator	脱敏增量数据	



# 内容

---

- 1 数据、数据价值和安全挑战
- 2 Oracle数据脱敏和子集的能力
- 3 在Exadata上进行数据脱敏
- 4 一次数据脱敏性能测试分享

## 测试环境

- Oracle Exadata X8 cluster
- 数据库版本: Oracle Database 19.11.0.0.0 4节点 RAC
- CPU: 每个节点 32个 vCPU , Intel(R) Xeon(R) Platinum 8270CL CPU @ 2.70GHz
- RAM: 每个节点100G
- 1T 的测试数据
- Oracle enterprise manager cloud control 13c release 5
- Oracle Linux Server release 7.9



## 数据库系统参数设置

- PARALLEL\_DEGREE\_POLICY = 'AUTO'  
Oracle 数据系统智能决策并发执行的程度
- 目标表上NOLOGGING设置  
数据脱敏时避免写入redo日志，避免性能损耗
- PGA\_AGGREGATE\_LIMIT = 70G  
增加PGA可使用空间至70G (每个RAC节点内存为100G)



## 目标说明

- 目标表 stcs.gl\_jornl\_dtl\_f\_x

列数: 72列, 行数: 4亿, 表大小: 245G

- 目标表上的敏感列

电话号码, 如 15677778888

身份证号, 如 425995564916875712 (18位)

住址, 如 广西壮族自治区 石家庄市 中阳县 荒河镇 (假地址信息)

[REDACTED], 如 广西壮族自治区 台州市 璧山区 中国进出口银行分行 (假地址信息)



## 测试内容

- 共进行5轮测试，每轮测试运行三次，取平均运行时间作为测试结果

第一轮，单独在**电话号码**列上进行脱敏，运行三次，取平均值

第二轮，单独在**身份证号码**列上进行脱敏，运行三次，取平均值

第三轮，单独在**住址**列上进行脱敏，运行三次，取平均值

第四轮，单独在**银行账户地址**列上进行脱敏，运行三次，取平均值

第五轮，同时在**电话号码，身份证号码，住址和银行账户地址**列上进行脱敏，运行三次，取平均值



# 1. 脱敏规则 - 电话号码

- 要求随机数脱敏，符合电话号码规则，11位数字

脱敏策略：固定数字 + 随机位数

ORACLE Enterprise Manager Cloud Control 13c

dbm02 / PDB2

Oracle 数据库 ▾ 性能 ▾ 可用性 ▾ 安全性 ▾ 方案 ▾ 管理 ▾

数据掩码定义 > 编辑掩码定义: round1\_mask\_sourceDB\_on\_bigger\_size > 定义列掩码

所有者 STCS  
列 PHONE 对象 EOM\_EMPLOYEE\_D\_X  
数据类型 VARCHAR2(120)

确定 取消

默认情况下，表中的所有记录将使用指定的格式进行掩码。可以选择使用条件标识记录的多个子集。每个子集均可使用相应的掩码格式进行掩码。这些子集将按指定它们的顺序进行掩码。即便子集与后续条件匹配，也不会再次对其进行掩码。

添加条件

导入格式 格式条目 随机位数 添加

全部展开 全部隐藏

选择	条件	格式条目属性				示例	移去
		属性	值	属性	值		
	▼ 条件						
<input checked="" type="radio"/>	▼ 默认条件					15454086132	
	固定数字	固定数字	<input type="text" value="1"/>				
	随机位数	起始长度	<input type="text" value="10"/>	结束长度	<input type="text" value="10"/>		

确定 取消

# 电话号码脱敏- 测试1 – 脱敏运行时间

作业

作业活动 > 执行: dbm02\_PDB2

切换到经典视图 自动刷新 15 秒

页面刷新于: 2021-9-16 2:03:06 PDT

执行: MASKING\_JOB\_192

概要

状态 ✓ 已成功  
重复 立即  
调度 2021-9-16 1:54:30 GMT-07:00  
启动 2021-9-16 1:54:35 GMT-07:00  
显示更多

结束 2021-9-16 2:00:43 GMT-07:00  
用时 6 分, 8 秒

dbHostCreds ORACLE (oracle/\*\*\*\*\*\*\*\*\*)  
dbCreds NC\_DBM02\_PD\_2021-09-07-203056  
(stcs/\*\*\*\*\*\*\*\*\*)

名称 MASKING\_JOB\_192  
类型 数据掩码  
所有者 SYSMAN  
目标 dbm02\_PDB2

操作 操作 ID CC19AC7D6ADDCAA1E053EE0C200A5D74  
脚本 /u01/app/oracle/product/19.0.0.0/dbhome\_1/dbs...

dbm02\_PDB2的任务列表

查看 停止步骤

状态 全部 目标名称

名称	类型	状态	执行时间
DeployFmtLib	步骤	✓	
PutSqlFile	步骤	✓	
Command	步骤	✓	
DropFmtLib	步骤	✓	

DropFmtLib

状态 ✓ 已成功  
显示更多

输出日志

输出

Successfully dropped format library package.  
Successfully dropped DM package.  
Successfully dropped DM type.  
Success

34 Copyright © 2021, Oracle and/or its affiliates



# 脱敏规则 - 电话号码 - 测试1 - 主机 (CPU, 内存)

CPU 占用率	
查看	<input checked="" type="checkbox"/> 显示最后已知值
时间	CPU 占用率 (%)
2021-9-16 0:56:14	4.51
2021-9-16 1:01:14	4.38
2021-9-16 1:06:14	4.31
2021-9-16 1:11:14	4.25
2021-9-16 1:16:14	4.32
2021-9-16 1:21:14	4.31
2021-9-16 1:26:14	5.12
2021-9-16 1:31:14	5.08
2021-9-16 1:36:14	5.03
2021-9-16 1:41:14	5.15
2021-9-16 1:46:14	4.98
2021-9-16 1:51:14	4.75
2021-9-16 1:56:14	23.72
2021-9-16 2:10:17	4.51
2021-9-16 2:14:08	4.57
2021-9-16 2:14:39	4.57

内存使用量	
查看	<input checked="" type="checkbox"/> 显示最后已知值
时间	内存使用率 (%)
2021-9-16 1:51:14	42.87
2021-9-16 1:56:14	56.22
2021-9-16 2:01:14	56.90
2021-9-16 2:06:14	45.43
2021-9-16 2:11:14	45.45
2021-9-16 2:16:14	45.43
2021-9-16 2:21:14	43.33
2021-9-16 2:26:14	43.79

此次脱敏资源使用情况：

CPU: 23.72% , 增长 ( $23.72\%-4.75\% = 18.99\%$ )

RAM: 56.9%, 增长 ( $56.90\%-42.87\% = 14.03\%$ )

注：数据库空闲期CPU 使用率： 4.75%

内存使用率： 42.87%



# 电话号码脱敏 – 结果分析

脱敏策略说明：固定数字 + 随机位数

固定数字：脱敏时只取指定的数字，不是用查找或映射表

随机数位：在指定数位范围内生成唯一的数值（若指定的数位不足，无法保证生成唯一的数值）

例子：电话号码11位数字，脱敏规则：固定数字取1，剩余10位数字随机生成

15248740762

指标	测试1	测试2	测试3	平均值
电话号码脱敏耗时	6分8秒	4分53秒	4分54秒	<b>5分18秒</b>
CPU使用率	23.72%	44.91%	48.1%	<b>38.91%</b>
RAM使用率	56.9%	58.64%	59.5%	<b>58.35%</b>



## 2. 脱敏规则 – 身份证号码

- 要求保留身份证号前6位，剩余号位随机生成，共18位

脱敏策略：子字符串 + 随机位数

ORACLE Enterprise Manager Cloud Control 13c

dbm02 / PDB2

Oracle 数据库 性能 可用性 安全性 方案 管理

数据掩码定义 > 编辑掩码定义: round1\_mask\_sourceDB\_on\_bigger\_size > 定义列掩码

所有者 STCS  
列 ID\_CARD\_NUMBER

对象 EOM\_EMPLOYEE\_D\_X  
数据类型 VARCHAR2(100)

确定 取消

默认情况下，表中的所有记录将使用指定的格式进行掩码。可以选择使用条件标识记录的多个子集。每个子集均可使用相应的掩码格式进行掩码。这些子集将按指定它们的顺序进行掩码。即便子集与后续条件匹配，也不会再次对其进行掩码。

添加条件

导入格式 格式条目 随机位数 添加

全部展开 全部隐藏

格式条目属性				
属性	值	属性	值	
示例	662911706867696470	移去		
选择 条件				
▼ 条件				
① ▼ 默认条件				
子字符串	起始位置	1	长度	6
随机位数	起始长度	12	结束长度	12

确定 取消

# 身份证号码脱敏 – 结果分析

脱敏策略说明：子字符串 + 随机位数

子字符串：通过起始位置和长度作为输入，使用 SUBSTR 从输入列中提取该数据，并将其用作掩码值。

随机数位：在指定数位范围内生成唯一的数值（若指定的数位不足，无法保证生成唯一的数值）

例子：身份证号码共18位，脱敏规则：子字符串取（开始位1，结束位6），剩余12位取随机数

321732862098548155

指标	测试1	测试2	测试3	平均值
身份证号码脱敏耗时	4分52秒	4分57秒	5分14秒	<b>5分1秒</b>
CPU使用率	32.01%	43.58%	46.44%	<b>40.68%</b>
RAM使用率	57.42%	57.42%	60.16%	<b>58.33%</b>



### 3. 脱敏规则 – 住址

- 要求使用地址替换的方式

脱敏策略：表列

- 预先创建地址信息表，通过省，市，乡镇，地区等信息通过笛卡尔积

```
CREATE SEQUENCE stcs.china_address_sequence START WITH 1;  
  
create table stcs.chinese_address_random as (  
    select stcs.china_address_sequence.nextval,t1.province||"||t2.city ||"||t3.area||"||t4.town address_info  
    from stcs.china_province t1, stcs.china_city t2,stcs.china_area t3,stcs.china_town t4  
    where rownum<=4000000000);
```

The screenshot shows the Oracle Database Masking tool interface. The main window displays a table with columns: 表名 (Table Name) set to STCS.CHINESE\_ADDRESS, 列名 (Column Name) set to ADDRESS\_INFO, and a condition entry field containing '宁夏回族自治区石家庄市新巴尔虎左旗桑梓镇' (寧夏回族自治区石家庄市新巴爾虎左旗桑梓鎮). A red box highlights this entry. The bottom right corner of the dialog has '确定' (Confirm) and '取消' (Cancel) buttons.

# 住址脱敏 – 结果分析

脱敏策略说明：表列

表列：该方法会创建一个映射表，将原始列值映射到用户指定表中的列值。并且在将每个值映射到原始列之前，用户指定的表中的值使用随机排序。

例子：

原始列值：北京市石家庄市吉利区唐塔街道

用户指定表中的列值：河北省石家庄市郵城县班竹镇

指标	测试1	测试2	测试3	平均值
住址脱敏耗时	6分24秒	6分28秒	5分45秒	<b>6分12秒</b>
CPU使用率	35.03%	36.43%	24.6%	<b>32.02%</b>
RAM使用率	58.18%	57.54%	58.9%	<b>58.21%</b>



## 4. 脱敏规则 – 银行账户地址

- 要求采用固定格式填充，如 广西壮族自治区宜昌市海南XXXXXXXXXX11
- 脱敏策略：子字符串 + 固定字符串 + 表列（表STCS.MASK\_DICT中存放“分行”字符）

dbm02 / PDB2

Oracle 数据库 | 性能 | 可用性 | 安全性 | 方案 | 管理

数据掩码定义 > 编辑掩码定义: mask\_chinese\_bank\_account\_address\_test > 定义列掩码

所有者 STCS  
列 BANK\_ADDRESS      对象 GL\_JORNAL\_DTL\_F\_X\_TEST3  
数据类型 VARCHAR2(200)

默认情况下, 表中的所有记录将使用指定的格式进行掩码。可以选择使用条件标识记录的多个子集。每个子集均可使用相应的掩码格式进行掩码。这些子集将按指定它们的顺序进行掩码。即便子集与后续条件匹配, 也不会再次对其进行掩码。

添加条件

导入格式	格式条目	表列	添加				
全部展开	全部隐藏						
选择	条件	格式条目属性				示例	移去
▼ 条件		属性	值	属性	值		
① ▼ 默认条件						广西壮族自治区保定市万柏XXXXXXXX分行	
子字符串		起始位置	1	长度	12		
固定字符串		固定字符串	XXXXXXX				
表列		表名	STCS.MASK_DICT	列名	bank_branch		

确定 | 取消

# 银行账户地址脱敏 – 结果分析

脱敏策略说明：子字符串 + 固定字符串 + 表列（表STCS.MASK\_DICT中存放“分行”字符）

子字符串：通过起始位置和长度作为输入，使用 SUBSTR 从输入列中提取该数据，并将其用作掩码值。

固定字符串：该方式不使用查找或映射表。它为字符串列分配一个固定的字符串值。

表列：该方法会创建一个映射表，将原始列值映射到用户指定表中的列值。并且在将每个值映射到原始列之前，用户指定的表中的值使用随机排序。

例子：

原始列值：**广西壮族自治区九江市西丰县中国人民银行分行**

脱敏后：**广西壮族自治区九江市西丰XXXXXXXX分行**

指标	测试1	测试2	测试3	平均值
银行账户地址 脱敏耗时	2分19秒	2分39秒	2分42秒	<b>2分6秒</b>
CPU使用率	31.11%	37.56%	39.88%	<b>36.18%</b>
RAM使用率	52.83%	47.35%	47.42%	<b>49.2%</b>



## 5. 电话号码+身份证+住址+银行地址同时脱敏 样本数据展示

```
select t.jornl_det_idn id, t.phone, t.id_card_number, t.chinese_address, t.bank_address,
t1.phone, t1.id_card_number, t1.chinese_address, t1.bank_address
from stcs.gl_jornl_dtl_f_x t inner join stcs.gl_jornl_dtl_f_x_test1 t1 on t.jornl_det_idn = t1.jornl_det_idn
where t.jornl_det_idn in ('1585946304', '-1454954242', '1418414834', '1414116774', '-1321128718');
```

原数据

脱敏后数据

Query Result		Query Result 1		Query Result 2	
SQL		All Rows Fetched: 5 in 20.649 seconds			
ID	PHONE	ID_CARD_NUMBER	CHINESE_ADDRESS	BANK_ADDRESS	
1	1585946304	12352720740	327747969797627681	广西壮族自治区 秦皇岛市 湖口县 东柳街道	广西壮族自治区九江市西丰县中国人民银行分行
2	-1454954242	16062887764	518605119894419636	广西壮族自治区 石家庄市 南漳县 白衣镇	广西壮族自治区汕尾市合江县中国进出口银行分行
3	1418414834	13980433632	398005402440548632	广西壮族自治区 唐山市 元氏县 铁西街道	广西壮族自治区邢台市灯塔市交通银行分行
4	1414116774	15031316485	755200195819408176	广西壮族自治区 唐山市 吉州区 陆港街道	广西壮族自治区大理白族自治州普宁市中国进出口银行分行
5	-1321128718	17962165174	907458414690567344	广西壮族自治区 唐山市 岱山县 兴林镇	广西壮族自治区保山市三都水族自治县中国农业发展银行分

电话号码脱敏：  
保证手机号位  
11位，起始数  
位1

身份证号码脱  
敏：保留前六  
位，后12位随  
机生成

住址信息脱敏：  
从其他地址信  
息表中随机替  
换

银行账户地址：保  
留前12个字符，接  
着以字符“X”进行  
替换，最后以“分  
行”结尾

# 电话号码+身份证号+住址+银行地址脱敏- 结果分析

- 目标表上的敏感列脱敏（**4列敏感列， 4亿行记录，大小245G**）

**电话号码**，随机生成11位电话号码，如 15677778888

**身份证号**，保留前6位，随机生成12位数字，如 425995564916875712 （18位）

**住址**，替换

原始列值：北京市石家庄市吉利区唐塔街道

用户指定表中的列值：河北省石家庄市郵城县班竹镇

**银行账户地址**，固定值替换，如

原始列值：广西壮族自治区九江市西丰县中国人民银行分行

脱敏后：广西壮族自治区九江市西丰XXXXXXXX分行

指标	测试1	测试2	测试3	平均值
脱敏耗时	14分52秒	15分5秒	16分13秒	<b>15分39秒</b>
CPU使用率	52.24%	46.26%	59.70%	<b>52.73%</b>
RAM使用率	63.31%	59.49%	59.77%	<b>60.86%</b>

# 脱敏测试结果汇总

敏感列	脱敏策略	源数据格式	脱敏后
电话号码	固定数字 + 随机位数	12352720740	15058801334
身份证号码	子字符串 + 随机位数	618605119894419636	618605536230496146
住址	表列	广西壮族自治区邢台市 灯塔市交通银行分行	山西省石家庄市郵城县 查干哈达苏木
银行账户地址	子字符串 + 固定字符串 + 表列	广西壮族自治区邢台市 灯塔市交通银行分行	广西壮族自治区邢台市 灯塔XXXXXX分行

脱敏方式	平均耗时	CPU平均使用率	内存平均使用率	目标表记录数
电话号码 (单独脱敏)	5分18秒	38.91%	58.35%	4亿行
身份证号码 (单独脱敏)	5分1秒	40.68%	58.33%	
住址 (单独脱敏)	6分12秒	32.02%	58.21%	
银行账户地址 (单独脱敏)	2分6秒	36.18%	49.2%	
电话号码+身份证+住址+银行地址 (同时脱敏)	15分39秒	52.73%	60.86%	



# Thank You

---



ORACLE

O

Our mission is to help people  
see data in new ways, discover insights,  
unlock endless possibilities.

