GIAC

全球互联网架构大会

GLOBAL INTERNET ARCHITECTURE CONFERENCE

互金行业数据安全之数据脱敏建设

杜亚威　萨摩耶金服　业务研发部技术负责人

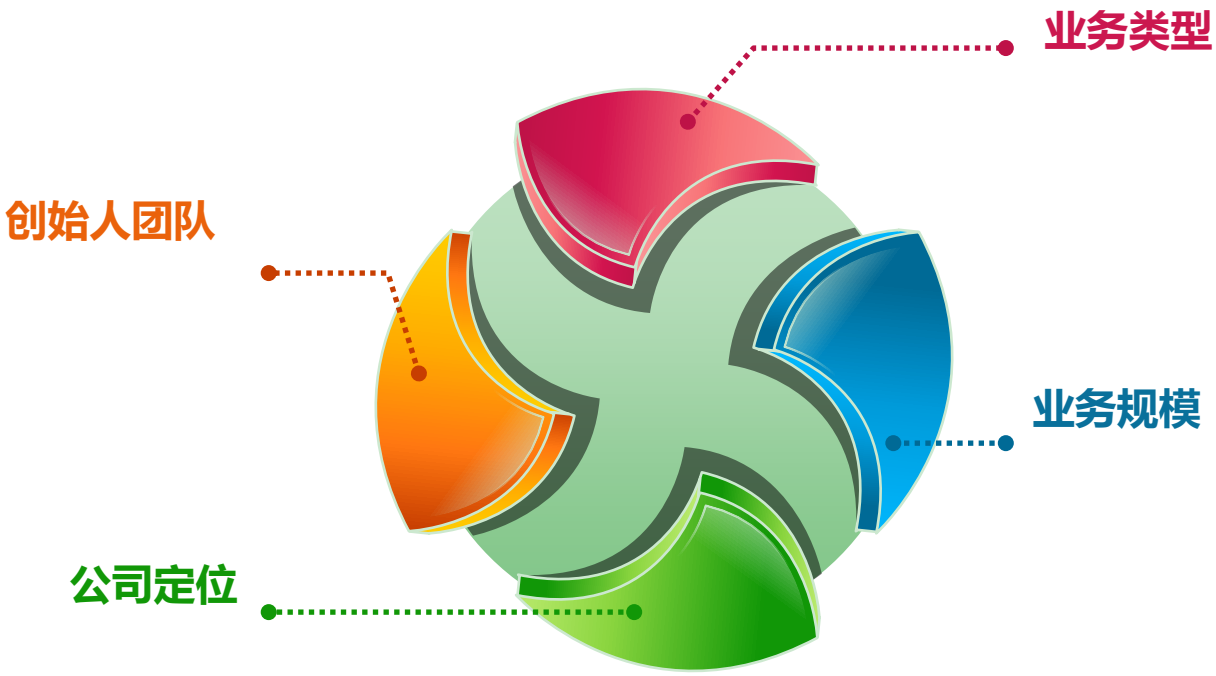
第一部分 背景介绍

第二部分 技术难点与分析

第三部分 技术架构实现

第四部分 未来规划

公司介绍



数据安全背景

一、监管与合规
二、不同的身份需要不同的数据权限
三、数据传输
四、日志敏感信息
五、灵活、全面的审计
六、风控环境数据沙箱

数据防泄漏与保护面临的挑战

一、如何在海量数据应用中自动识别隐私数据
二、如何保护数据防泄漏
三、如何准确高效脱敏

第一部分 背景介绍

第二部分 现状与分析

第三部分 技术架构实现

第四部分 未来规划

现状分析一

一、数据容量
二、数据类型
    1、非结构化Hive、Hbase、Elasticsearch、GFS
    2、结构化 Mysql
    3、其他 Orientdb、Redis
三、技术栈分析

现状分析二

三、脱敏范围

1、PII、相关信息、属性标签
2、其他 如日志

四、算法选择

1、数据加密、
2、数据替换、随机化、偏移与取整、掩码与屏蔽、灵活编码等

五、脱敏类型

1、静态脱敏
2、动态脱敏

技术分析与风险评估

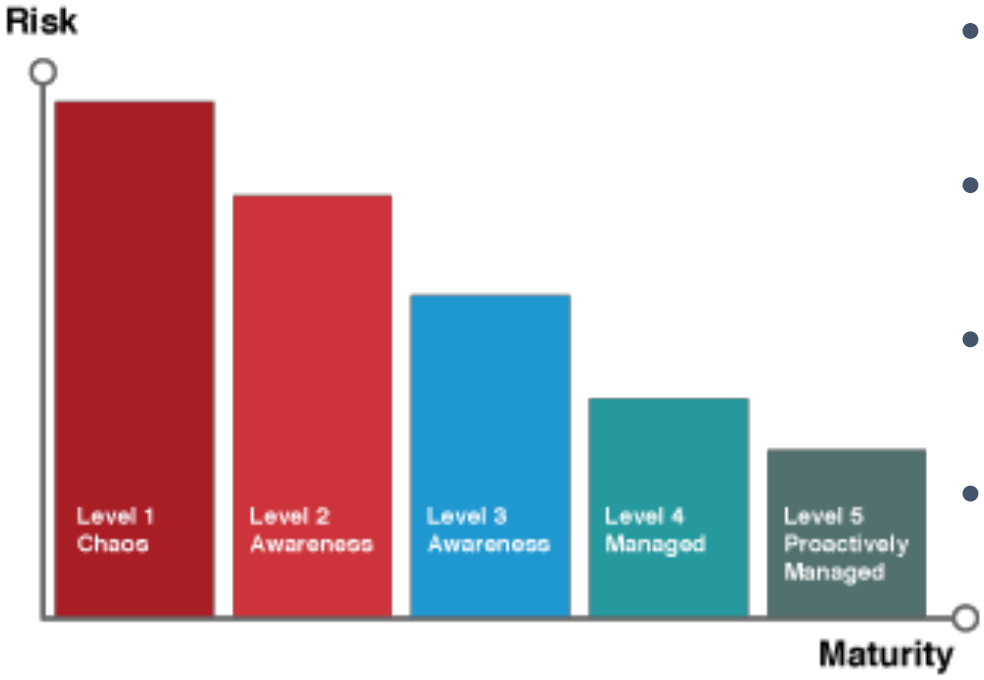
1、列长度
2、密文索引问题
3、密钥版本与存储
4、脱敏算法性能
5、对其他系统影响，如大数据，调用者
6、数据恢复，如容错
7、脱敏范围遗漏
8、角色使用者评估
9、多数据源问题
10、兼容性问题
11、成本

最终方案

类型	脱敏方案	使用者	备注
展现层	1、不同的身份有不同的数据阅读权限(如掩码、完全展现) 2、软防火墙,统一使用自研proxy;禁止原生客户端 3、背景水印	系统/内部业务人员	1、运营支撑系统 2、mysql、hive、es等原生客户端
业务逻辑层	1、系统日志全链路敏感信息加密 2、严格的系统认证与鉴权 3、底层存储敏感信息加密	系统	1、dubbo 2、rocketmq
存储层	1、敏感信息加密; 2、所有权与使用权分开(如DBA); 3、敏感表通过签名实现完整性	系统	结构数据如mysql
	1、兼容mysql协议proxy 2、根据元数据与权限管理，不同的身份具有不同的数据缺陷	内部业务人员/研发/运维	结构数据如mysql
	1、增加Proxy，兼容各种协议 2、增加元数据与权限管理	系统/运维	ES、hbase、hive
	1、增加统一存取proxy 2、底层文件加密 3、浏览时加水印 4、增加元数据与权限管理 5、所有对外文件内存流转时不出隔离区	系统/运维	文件系统GFS
	1、日常数据中心管理方式 2、严格的应用认证与授权	系统/运维	1、OrientDb。如知识图谱系统 2、redis。如名单系统

注:所有方案均可以审计

数据脱敏成熟模型

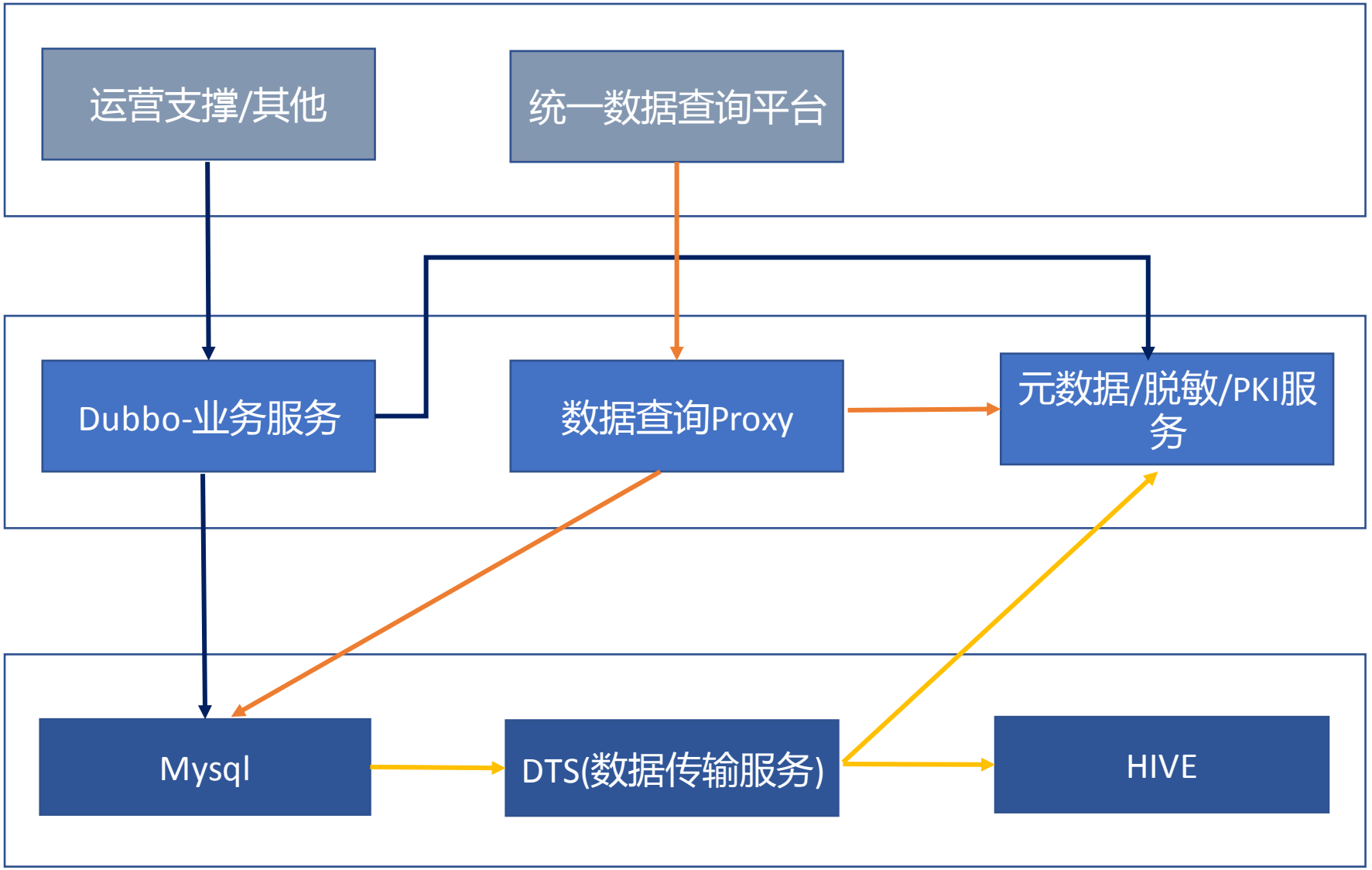


- Level 1 – Sensitive Data Cha
- Level 2 – Sensitive Data Awareness
- Level 3 – Sensitive Data Understood
- Level 4 – Repeatable Sensitiv Data Masking
- Level 5 – Sensitive Data Proactively Masked and Managed

大纲

- 第一部分 背景介绍
- 第二部分 技术难点与分析
- 第三部分 一期技术架构实现
- 第四部分 未来规划

一期技术架构实现


运营支撑/其他
统一数据查询平台
Dubbo-业务服务
数据查询Proxy
元数据/脱敏/PKI服务
Mysql
DTS(数据传输服务)
HIVE

技术架构实现-灰度发布

1、老表全量，新表增量，提供脱敏迁移工具。灰度期间不影响老数据
2、在MYBATIS上做SQL解析，敏感信息在提供在领域模型上提供注解,对于开发者说接入成本相对较低
3、提供各种脱敏算法
4、良好接口兼容性

大纲

第一部分 背景介绍

第二部分 技术难点与分析

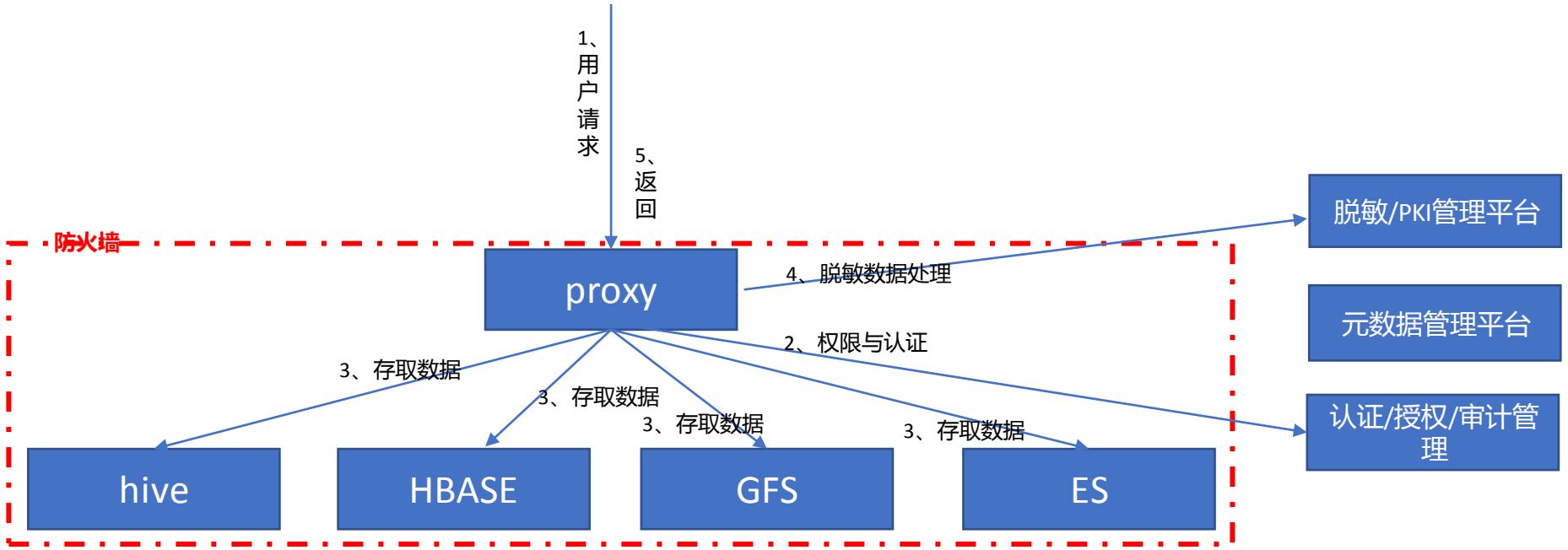
第三部分 一期技术架构实现

第四部分 未来规划

密钥与认证管理

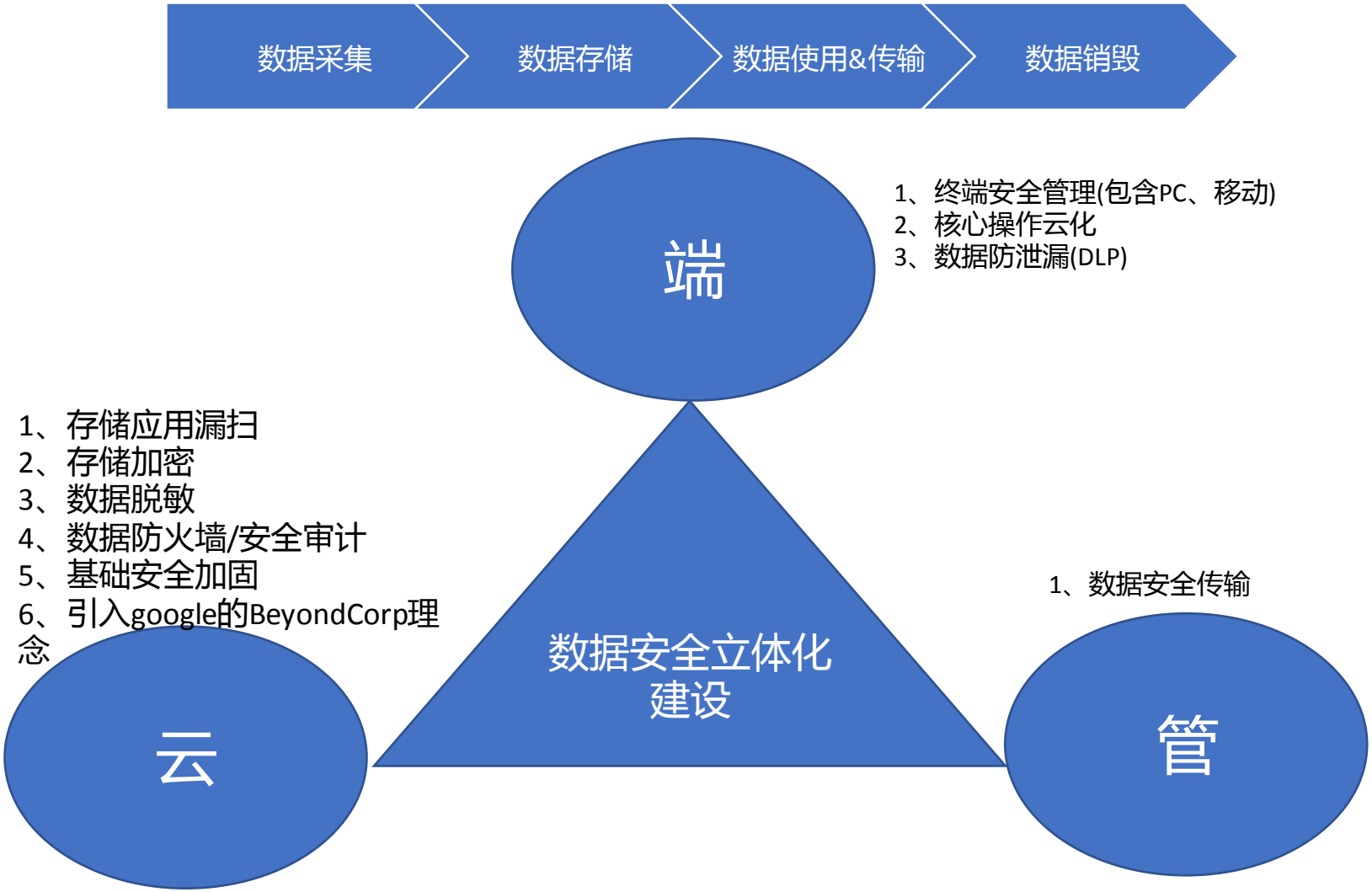
一、HSM/KMS
    1、密钥存储
    2、密钥生命周期
    3、密钥保护
    4、三级密钥体系
二、增强认证与授权
    1、引入Kerberos机制

三、非结构化数据全量脱敏



1、支持多协议，如es、hive、hbase、gfs等，可嵌套使用
2、全局的元数据管理
3、精细化ACL与RBAC模型,统一认证授权与审计管理
4、灵活的脱敏处理方式。如文件级加密与图片、合同水印等功能
5、智能隐私数据发现。

数据安全立体化联动


数据采集
数据存储
数据使用&传输
数据销毁
端
1、终端安全管理(包含PC、移动)
2、核心操作云化
3、数据防泄漏(DLP)
1、存储应用漏扫
2、存储加密
3、数据脱敏
4、数据防火墙/安全审计
5、基础安全加固
6、引入google的BeyondCorp理念
数据安全立体化建设
1、数据安全传输
云
管

智能隐私数据定级与识别

1、 数据特征学习以及自然语言处理
2、 基于某种规则

管理与认证

1、 ISO/IEC27018
2、 GDPR


用
守


SAMOYED

GIAC 全球互联网架构大会

GLOBAL INTERNET ARCHITECTURE CONFERENCE

关注公众号获得
更多案例实践