【提示】

今儿 19:00-21:00 远中悦来·阿基米德会议室 第六期班委邀请 百度数据安全leader **孙硕** 分享《企业数据安全思考与建设实践分享》。欢迎大家参加！

百度分享内容要点：

- 我们对监管合规与业务挑战的理解

- 我们关注的数据安全风险

- 安全管理思路及几个关键安全能力的构建

- 分享一套数据治理与隐私合规解决方案

leader、技术、产品、大数据仓库。

**会议内容：**

* **挑战：**

1.AI时代-大数据--分析产生信息数据。

2.监管与法律合规。

    网络安全法：日志留存6个月。（可追溯）

    开始：要求BAT、之后网易。（以金融公司的标准要求）

3.安全防御体系

    No DB-->DB-->BI/DW--->Big Data

高核心数据资产、个人信息。

数据安全对抗发生在外部（Web）及**内部安全边界（）**。

**数据堂--案例**

帮助查询公民个人信息--->用户地址、上网等等。也贩卖：方言库。导致被调查。

其他挑战：

* 数据过度采集和隐私权限滥用（头条）
  + 用户不能控制不上传通讯录
* 隐私数据泄露（twitter，美团外卖）
  + 密码打到日志里等等
* 对生态伙伴的不当数据共享（Facebook）
* 数据的流动困难（Netflix，推荐算法开源数据）
  + 开源数据已经脱敏，但是依然可逆向。定位某些用户。

**风险关注点：**

* 隐私数据权利场景
  + 购买数据的时候，需要确定提供方隐私数据来源合法。
  + 合作伙伴数据的利用。
* 设计敏感数据的线上应用场景
  + 黑产，根据掌握的数据，利用找回密码（打星的手机和邮箱）碰撞
* 大数据应用场景
* 数据合作场景
* 数据存储访问场景
* 关键人员岗位场景
  + token对应敏感数据（信用卡）--内部接触的都是token，出去的时候再还原
  + 数据流转都是token，数据进入数据库都是加密的。通过用户id来定位。
* 办公环境数据使用场景

**风险**：合规风险、PR风险、数据泄露风险。

底层：存储层、数据处理层使用相同的系统。

**数据发现**。

摈弃装agent模式，流量识别、

目标：数据在安全、可控的环境下快速流动。

* 可控
  + 4A
  + 细粒度-字段级别、列级别、行级别。
  + 脱敏等过
* 可审计
  + 异常使用行为
* 可追溯
* **关键能力**

* 线上系统的安全质量
* 产品隐私合规
* 数据统一管理
  + 一个员工各种访问的权限。
* 数据二次流转
  + 避免原始数据泄露（不做原始数据交换，SMC保护计算过程，不保护计算结果，结合数据核实）
  + 相关性分析（需要和第三方交互时使用的。）
* **挑战：**

* **数据治理和隐私解决方案**

数据识别---数据分类---数据保护。

产品安全是基础，做好了才做数据安全。

产品安全层面：

* 开发层面：
  + 把安全嵌入到框架中。
  + 统一公共仓库。只能从公共仓库拉代码。（检测统一仓库代码）
* 上线前（强制性）：
  + 静态代码分析（白盒），纯自研（90%以上），两三个人两三年，支持常见开发语言。
  + 扫描器（黑盒+灰盒，系统运行时获取的信息）
* 检测上线系统，是否有安全通过的标。
* 第三方库也检查。
* 线上系统嵌入一个包，检测是否异常。

* 团队，10个多人，提供统一服务。
  + 总体三年。三四个，蓝军。只要发现入侵路径就好。不用发现所有的渗透测试点。

**问题：**

数据管理。

业务线分散，很多业务线，很多服务器，资产。

* 业务统一：统一认证、统一授权、
  + 数据地图、数据资产产品团队和各业务数据平台负责人。
  + 统一schema、
  + 平台化、保留比较好的、比较强的数据平台。小业务进行迁移。
  + 所有的数据在网络上是通的。
* 大多数统一。

数据发现:

**质子大数据：**

自研系统，类似HUE的系统，动态脱敏。

权限控制到字段。（怎么实现的？）

法务内部安全红线，内部人员不能查用户敏感数据。