2019
05
08-10
北京新云南皇冠假日酒店

数据风云十年变迁

DTCC 第十届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2019

IT168 ChinaUnix ITPUB

数据安全之审计和恢复

青松云安全 鲍磊

内容概要:

- 数据安全现状
- 解决方案
- 机器学习在QDS中的应用

触目惊心的数据安全事件集

- 前沿数控事件

参考URL： https://baijiahao.baidu.com/s?id=1608117984981993366&wfr=spider&for=pc

- 快递公司程序员删库跑路

参考URL：https://baijiahao.baidu.com/s?id=1612288794642928553&wfr=spider&for=pc

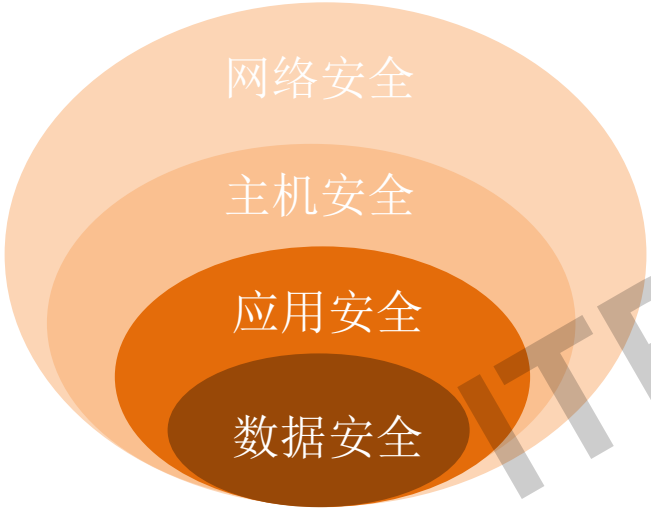
- 华夏银行行内鬼私自修改数据

参考URL： http://bank.hexun.com/2019-02-03/196066368.html

- 90后女生上班遭批评后辞职删公司后台数据泄愤被判刑

参考URL：http://law.eastday.com/n197/n198/u1ai109305.html

数据安全事件的反思



数据安全是核心

- ✓ 以对数据的操作为中心，建立全面、精确的审计和备份机制。对数据库各类操作行为进行监视并记录。及时发出告警信息
- ✓ 一旦发生数据安全事故，应快速、准确的对数据进行恢复。保障业务的持续稳定运行

如何保护我们的数据？

现有的数据安全技术

审计：从源数据库的网卡抓取流量

备份：文件级或系统级进行全量备份

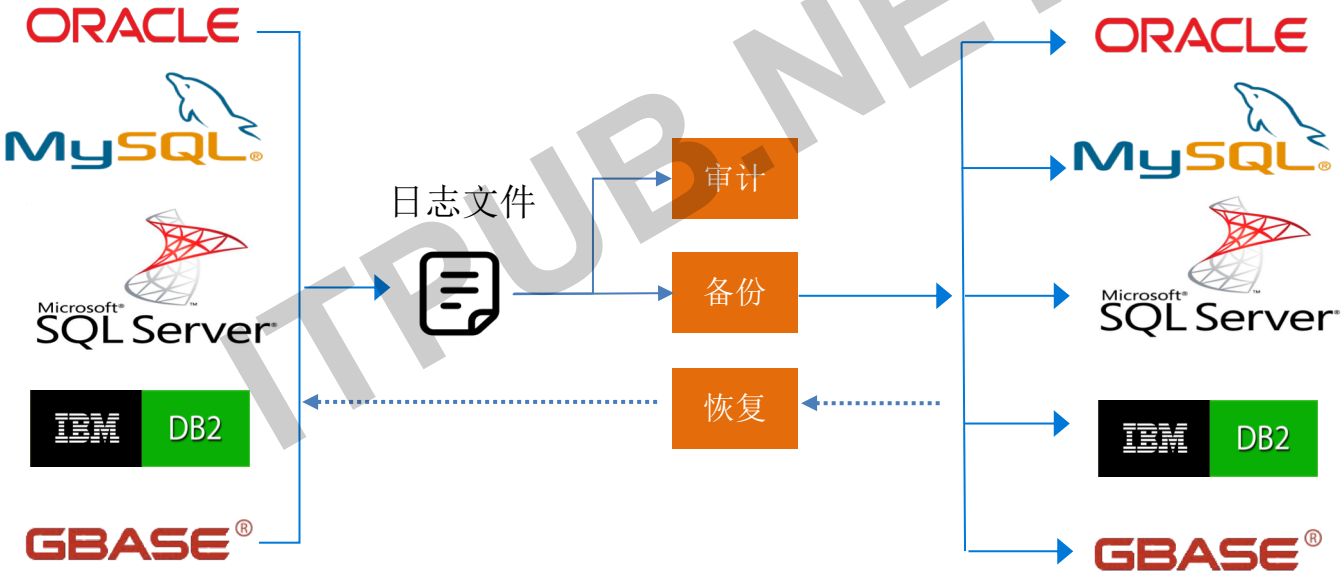
快照：数据库某一时间点的镜像

有绕过风险

颗粒度不精细

影响源库性能

技术解决方案



技术实现概述

基于网络进行日志采集，源库部署agent，属于被动受控端

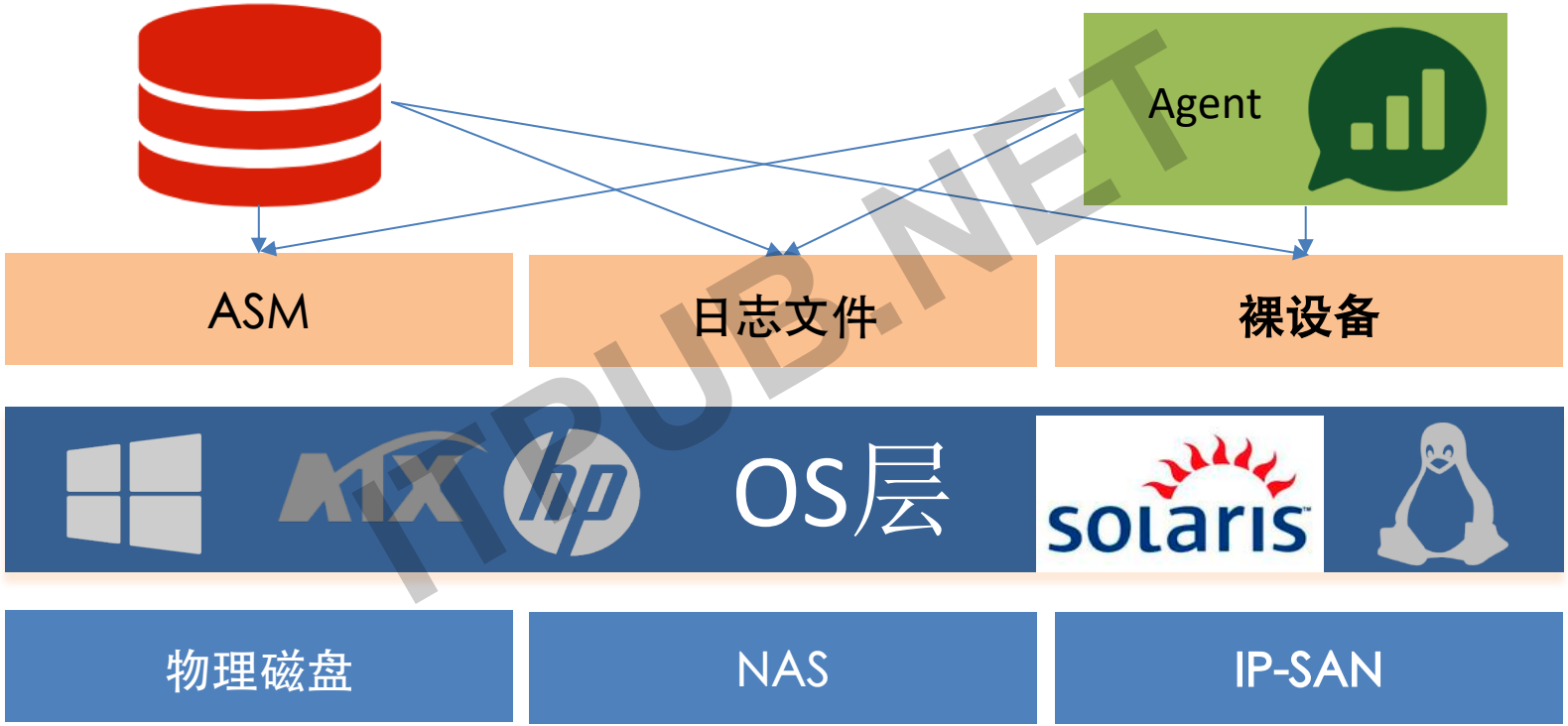
读取并分析数据库日志，解析原始数据

合并数据，关联事务，形成统一格式

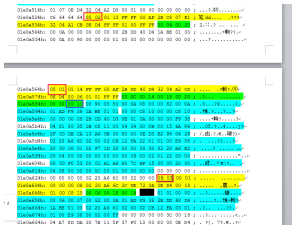
按照既定策略对数据进行分流过滤处理

支持加密压缩输出解析后的真实数据到目的
SQL兼容
二进制格式
JSON等

日志同步技术



基于事务的数据组织方式



完全基于二进制格式日志解析
支持多种平台

DDL解析
DML解析
DCL解析

迅速还原数据操作指令和数据

留存备份
备份库联动
原始数据
commit/rollback

分离数据确定事务
根据DCL决定事务提交与否

目标库
目标库
目标库
丢弃

细颗粒化的数据策略

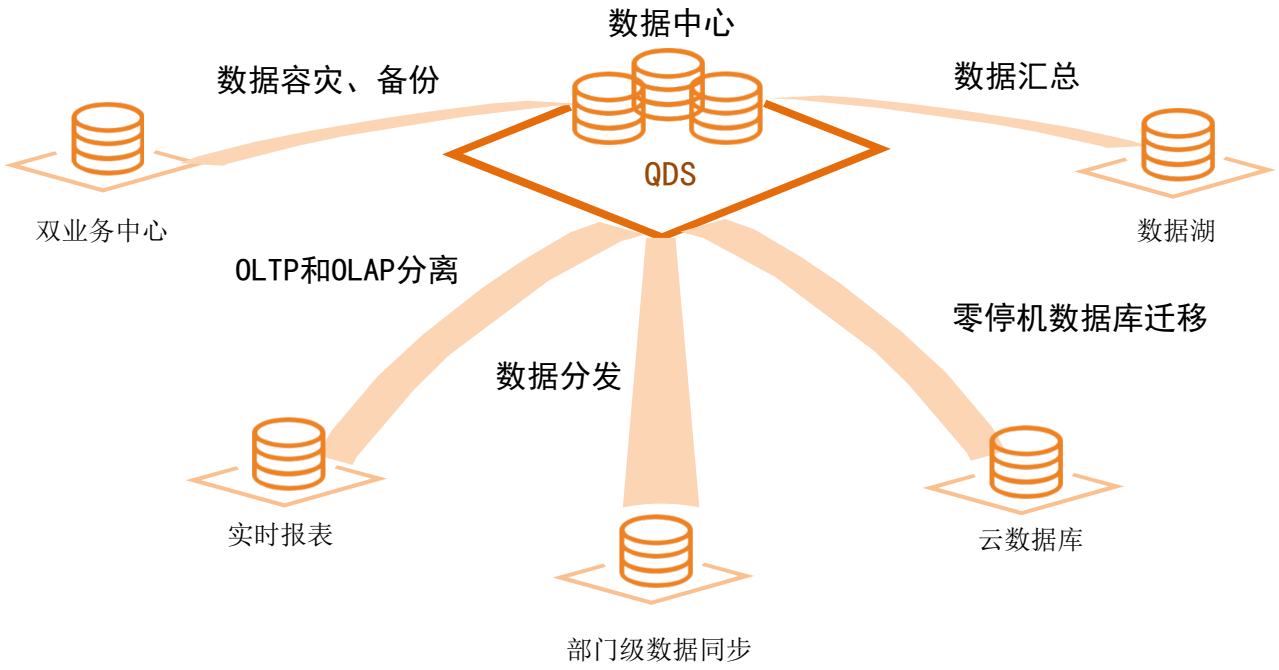
数据时间戳 用户登录时间 用户操作时间	用户 所有者，操作者	表 表名 所在的库	列 列参数	数据内容

数据关联	数据序列化

- 数据过滤
- 数据分发分流
- 时间策略
- 自定策略

- 向量化
- 人工智能策略
- 报警处理
- 逻辑检查

QDS应用场景



- ✓ **数据容灾、备份**
  打通数据中心与云之间的壁垒，数据彼此独立满足数据安全的需求
- ✓ **OLTP和OLAP分离**
  快速构建BI数据库系统。实现数据业务的快速响应
- ✓ **数据分发**
  实现数据共享，打破信息孤岛
- ✓ **零停机数据库迁移**
  无缝数据迁移，保障在线业务持续稳定运行
- ✓ **数据汇总**
  构建数据湖，在此基础上实现对数据的监管和智能分析

机器学习安全检测技术优势

大多数主流传统安全产品“基于特征码”技术，在与黑客对抗中已经显示出了弱势。在攻防技术不断迭代更新的过程中，特征码技术不足以保证抵抗黑客攻击，无法充分保护用户利益。而基于机器学习的技术已从图形声音向更广阔的信息处理等领域拓展，未来必将整体革新信息处理的生产力，机器学习非常适用于数据安全使用场景。

特征码　缺点：人力成本高、规则维护困难、防护滞后、规则更新慢

机器学习　优势：无需维护、发现0day攻击、节约成本、增强安全分析能力、多种模型算法、准确率高

机器学习在QDS中的应用

收集对数据库操作的数据为样本，使用ETL产品进行初步处理。经过特征工程和向量化，通过卷积神经网络(CNN)进行训练。使用Tensorflow Serving部署到生产环境提供服务


THANKS