大数据安全实践

屈跃辉

2018-05-12

演讲大纲

- 安全问题背景
- 魅族大数据安全标准体系
- 魅族大数据平台安全架构
- 大数据安全技术
  - 魅族大数据安全技术体系
  - 用户认证与管理
  - 精细化权限控制
  - 元数据管理
  - 数据加密与秘钥管理
  - 监控管理
- 魅族大数据安全管理系统
  - 通用权限系统产品架构设计与展示
  - 安全审计系统产品架构设计与展示
- 总结与展望

安全问题背景

数据、平台、业务快速发展

日新增行为记录	> 430亿	包括ERP、固件及80+业务线
日新增数据量	> 70TB	每年4~6x增长
数据规模	> 30PB	近4年累积
集群设备规模	2000+	Hadoop + Spark + HBase...

安全问题背景

安全问题突出

令人烦恼的安全问题

- 安全规范不完善
- 存在诸多安全漏洞
- 没有认证系统
- 不能够精细化控制数据权限
- 没有对数据进行透明有效管理
- 非法查阅敏感数据
- 异常非法操作（攻击者的挑战）
- ……

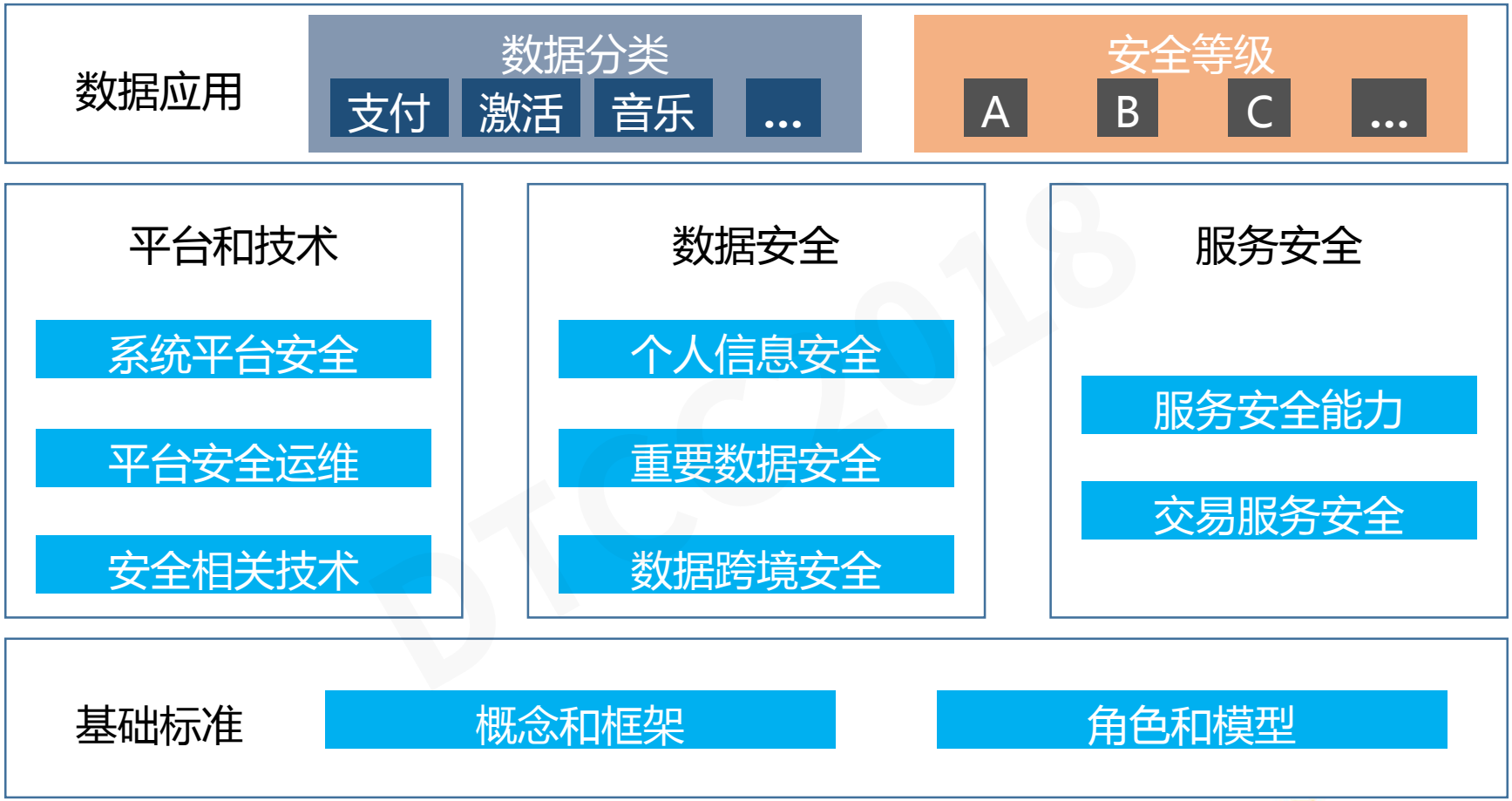
安全问题背景

大数据安全及标准化纳入国家发展战略

2017年4月，全国信息安全标准化技术委员会2017年第一次工作组“会议周”在武汉召开。会上，《大数据安全标准化白皮书》正式发布。

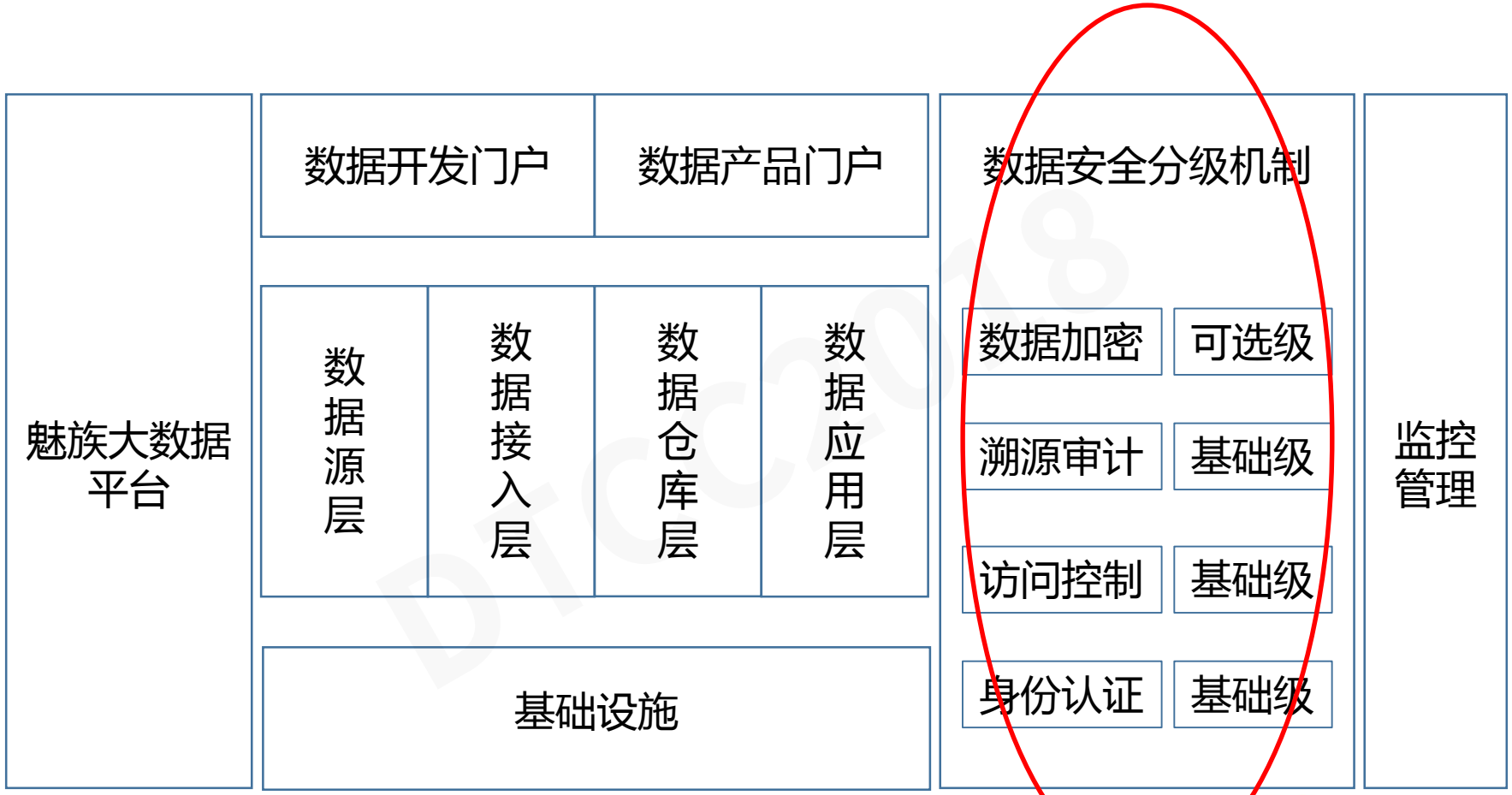
大数据安全标准体系

魅族大数据安全标准体系



大数据平台安全架构

魅族大数据平台安全架构



大数据安全技术

魅族大数据安全技术体系

监控管理

实时分析集群服务审计日志、审计预警用户行为

技术概念：

实时

易扩展

智能分析

Apache Eagle

边界

限制只有合法用户身份的用户访问集群

技术概念：

身份认证

网络隔离

Aquila（Kerberos、Ldap、Knox）

访问

定义什么样的用户和应用可以访问数据

技术概念：

权限

授权

Scutum（Apache Ranger）

透明

报告数据从哪来、如何使用的

技术概念：

审计

数据溯源

Apache Atlas

数据

保护集群敏感数据避免数据泄露

技术概念：

加密

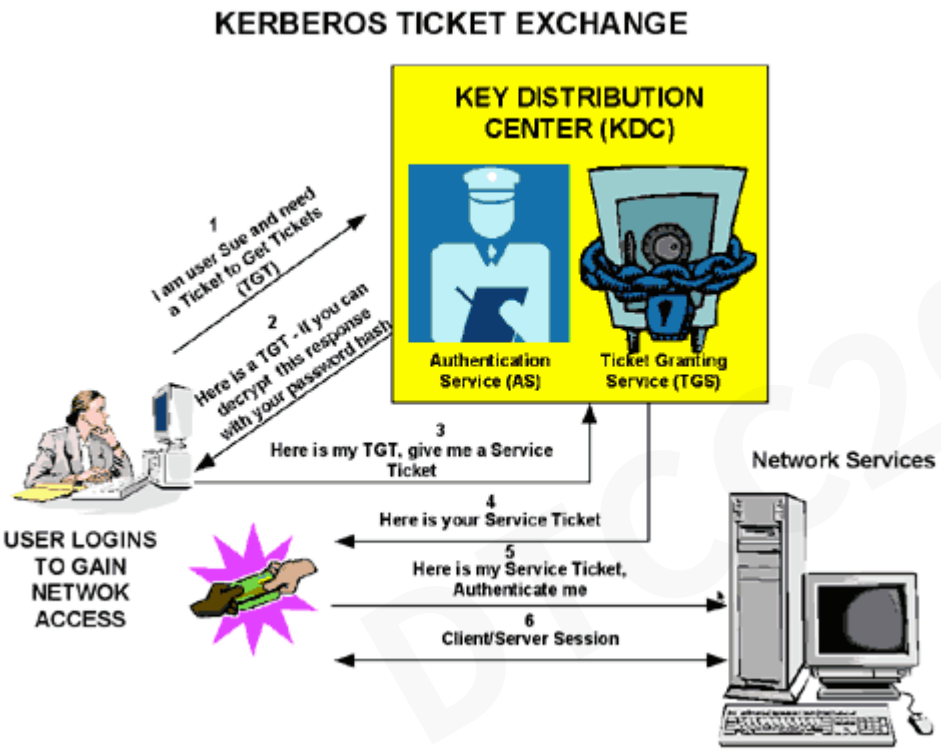
秘钥管理

数据脱敏

Hdfs Encryption & Ranger KMS

魅族大数据安全技术体系架构（图）

大数据安全技术



要点：

- 开源Hadoop生态原生唯一支持
- 服务器到服务器的认证
- Client到服务器的认证

不足：

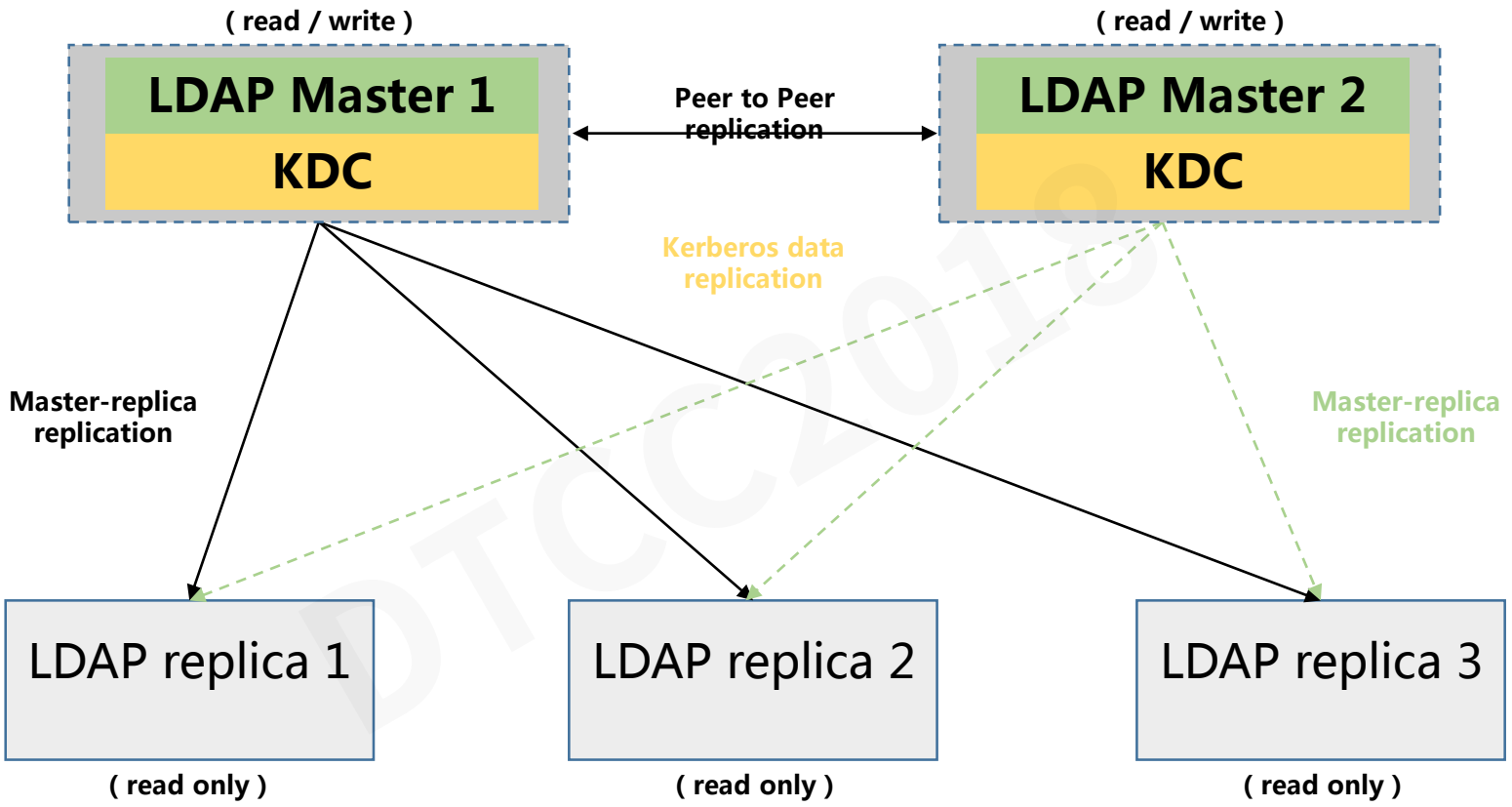
- 服务没有高可用
- 用户及组信息不能集中管理
- 缺乏JAVA API LIB

挑战：

- 运维管理（如keytab、ticket有效期）

大数据安全技术

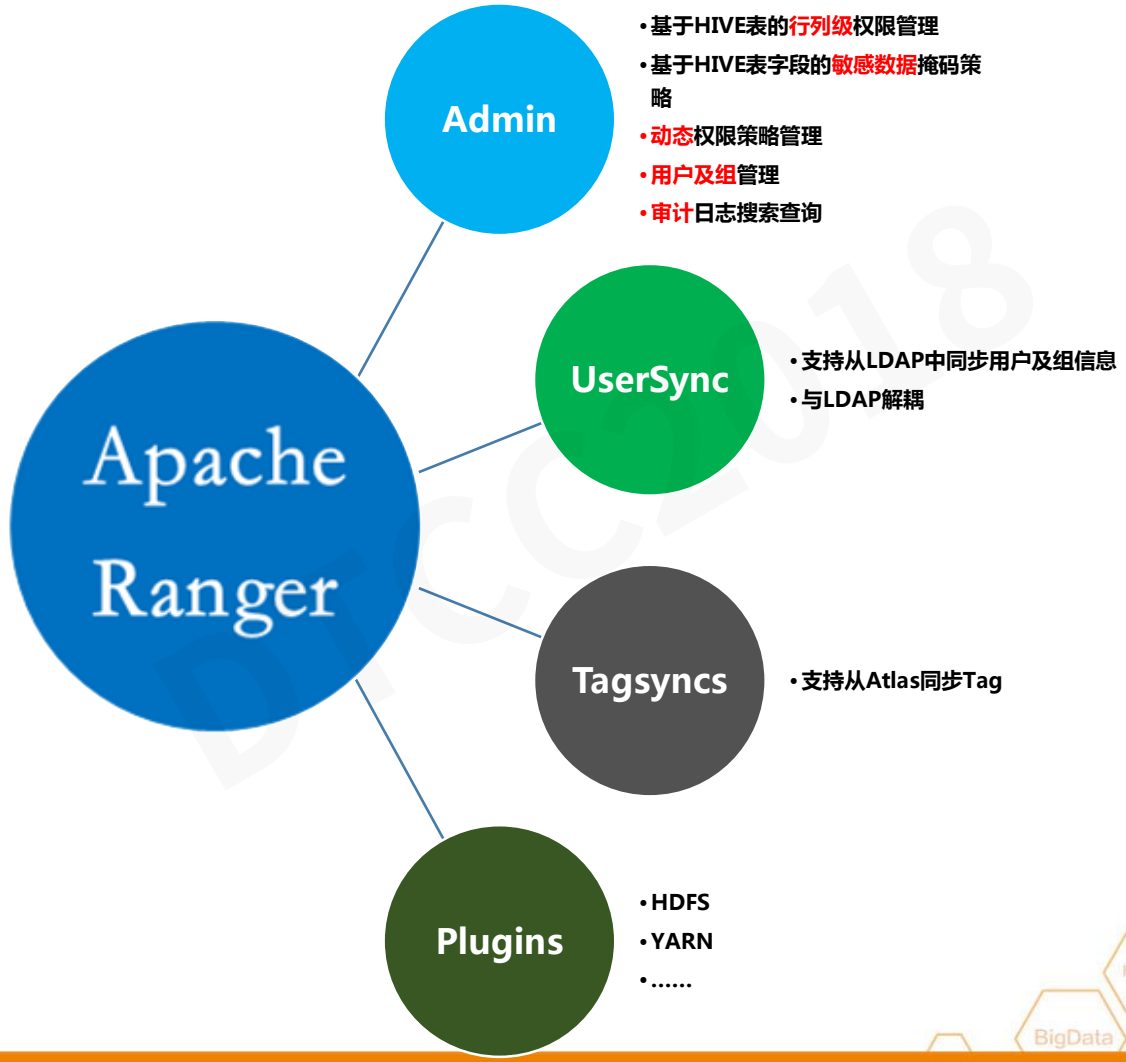
用户认证与管理



引入LDAP重点解决服务的高可用，同时兼顾性能，便利了用户及组信息的注册与管理。

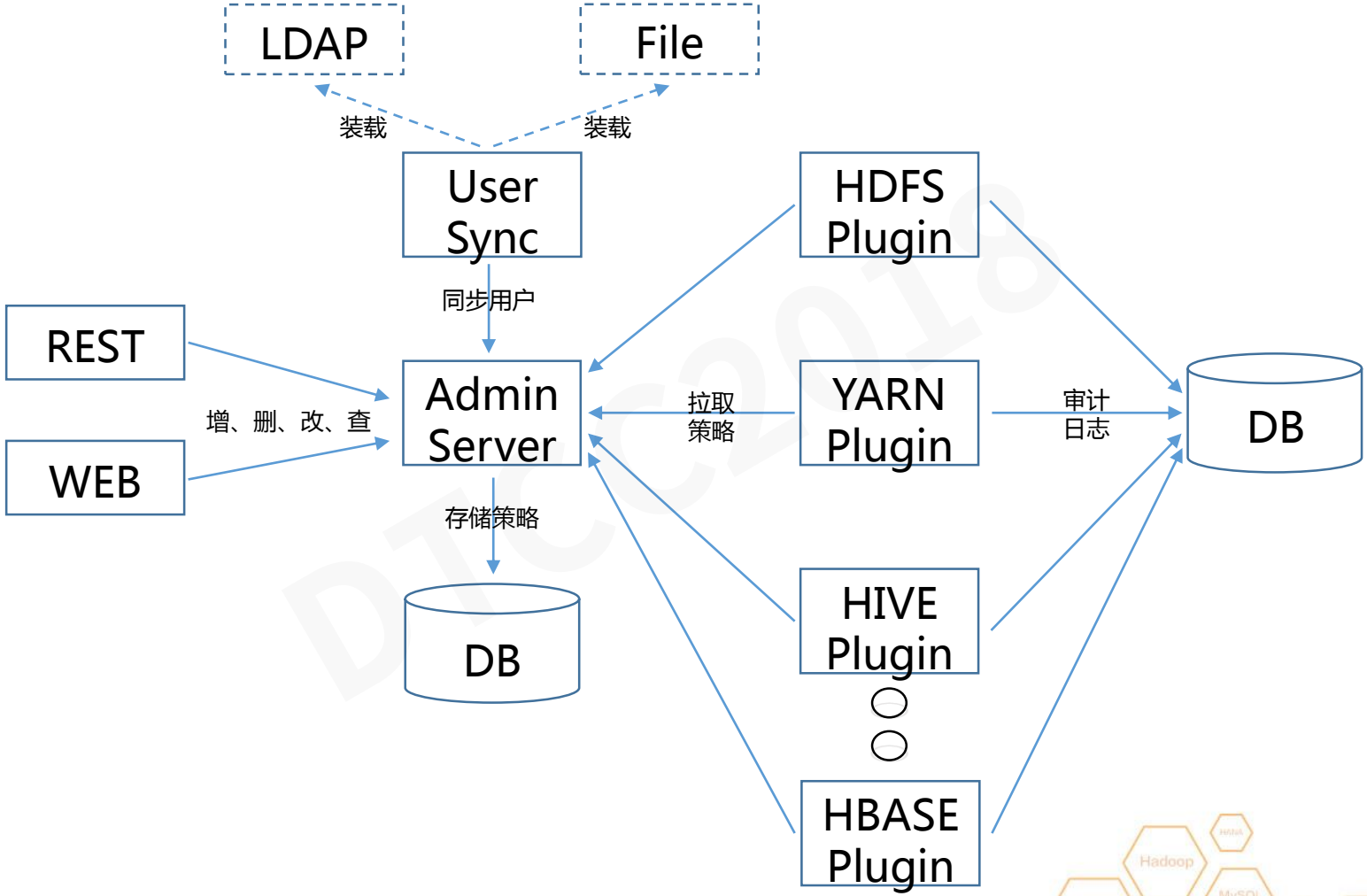
大数据安全技术

精细化权限控制



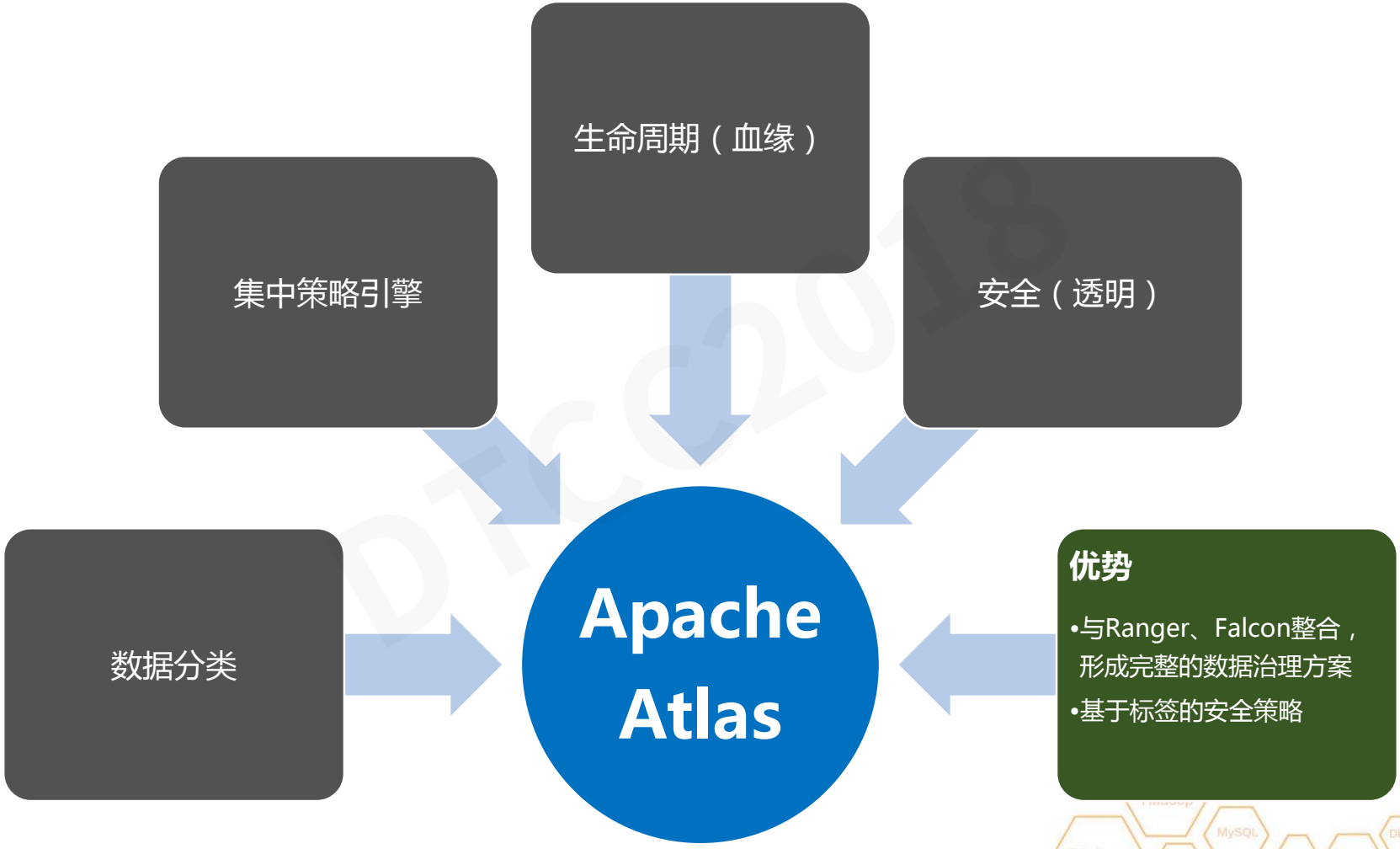
大数据安全技术

精细化权限控制



大数据安全技术

元数据管理



大数据安全技术

数据加密与密钥管理

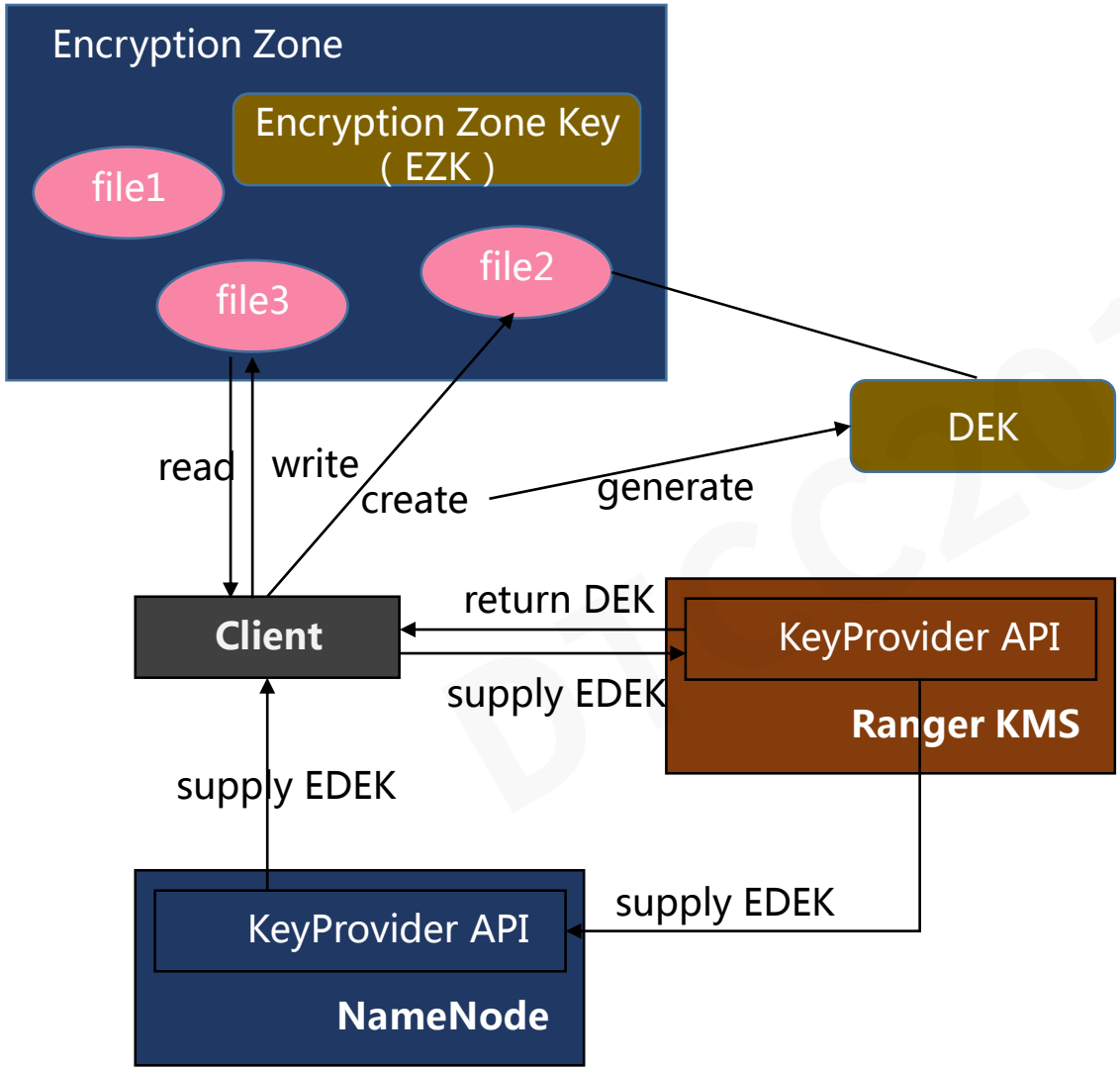
HDFS Encryption（端到端数据加、解密）

- **加密Key**
- **HDFS加密Zone**
  - 加密Zone与创建加密Zone指定的Key相关联
  - 加密Zone内的每个文件仅有唯一加密Key（DEK）
  - HDFS无法访问DEK。DN只能看到一串加密字节。HDFS将“加密数据加密密钥”（EDEK）作为文件元数据存储在NameNode
  - Client访问KMS解密EDEK，得到DEK去读/写数据

Ranger KMS（开源密钥管理服务）

- 密钥管理：提供对存储的加密Zone Key的（增、删、改、查）
- 访问控制策略：控制权限以生成或管理加密Zone Key，创建EDEK存储在HDFS
- 审计：提供Ranger KMS访问事件的完整审计跟踪

大数据安全技术



数据加密与密钥管理

官方示例：

```
# 以普通用户的身份创建一个加密key
hadoop key create myKey

# 以超级用户的身份创建一个空目录,并使之成为加密空间
hadoop fs -mkdir /zone
hdfs crypto -createZone -keyName myKey -path /zone

# 修改此目录权限为普通用户的
hadoop fs -chown myuser:myuser /zone

# 以普通用户的身份进行put上传文件和cat查看文件操作
hadoop fs -put helloWorld /zone hadoop fs -cat /zone/helloWorld
```

大数据安全技术

监控管理

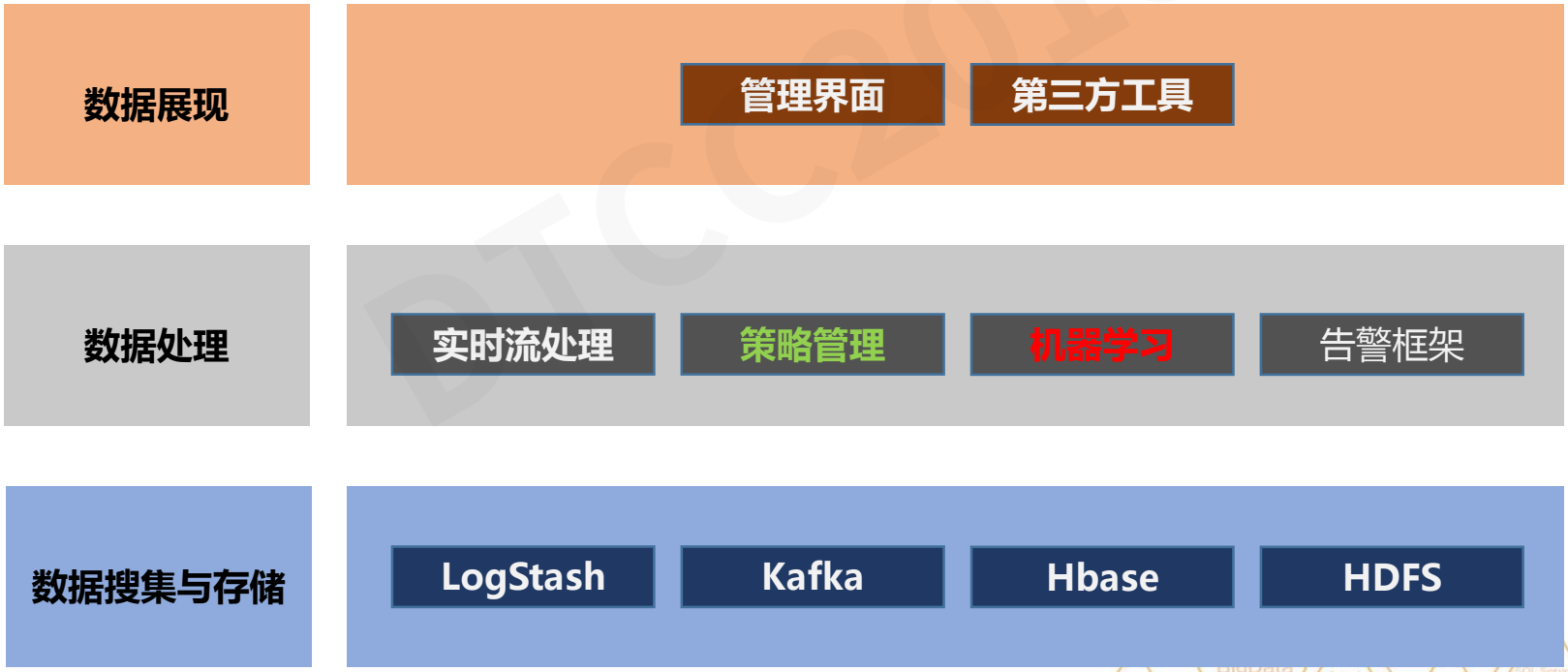
Apache Eagle核心能力：

- 监控Hadoop中的数据访问流量
- 检测非法入侵和违反安全规则的行为
- 检测并防止敏感数据丢失和访问
- 实现基于策略的实时检测和预警
- 实现基于用户行为模式的异常数据行为检测

Apache Eagle主要优势：

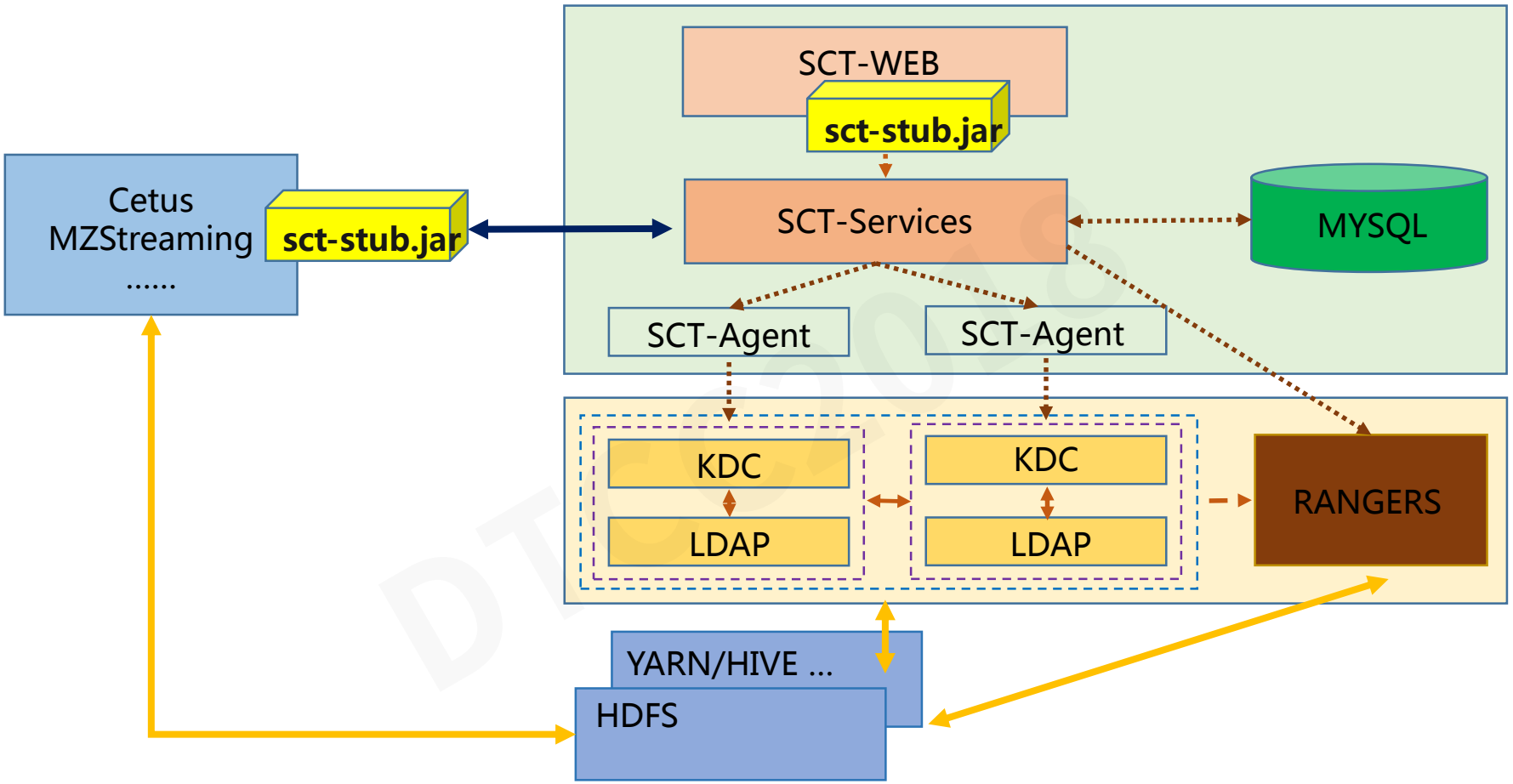
- 实时
- 易扩展
- 使用成熟的Hadoop生态技术
- 一套现成可用的告警策略规则引擎
- 机器学习算法

Apache Eagle框架概览：



魅族大数据安全管理系统

通用权限系统产品架构设计

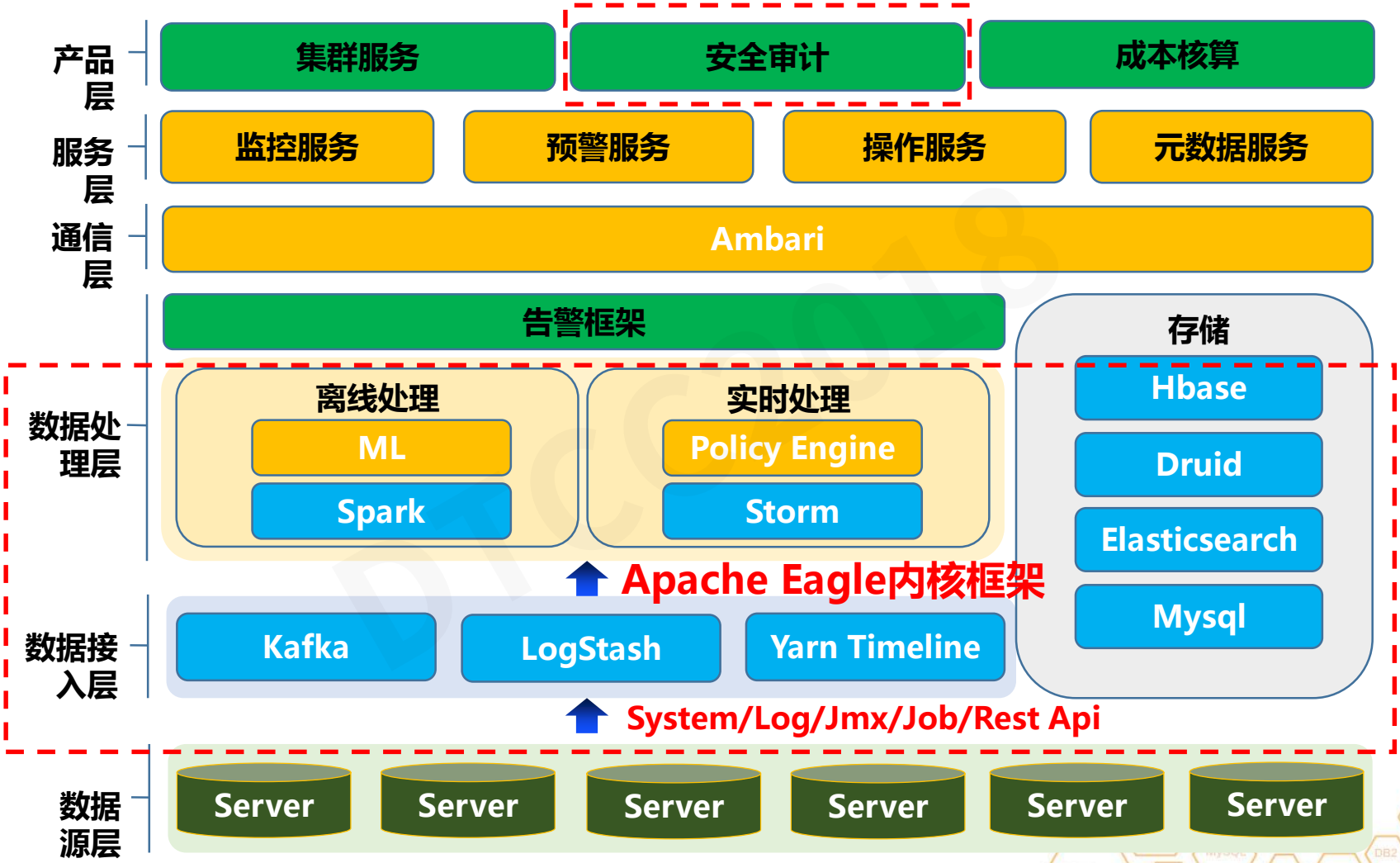


魅族大数据安全管理系统

通用权限系统产品展示

魅族大数据安全管理系统

安全审计系统产品架构设计



魅族大数据安全管理系统

安全审计系统产品展示

1、添加Hive敏感表字段访问预警策略

2、实时搜集Hive行为日志并处理，触发告警

ID	告警时间	消息时间	应用名称	策略名称	用户	源	描述
1	2017-11-02 08:30:17	2017-11-02 08:30:17	hiveQueryLog	queryPhoneNumber	hive	hiveAccessLogStream	The Policy "queryPhoneNumber" has been detected with the below infor... tion: timestamp="1[redacted]939" sensitivityType="PHONE_NUMBER" re source="/xademo/customer_details/phone_number" command="SELECT" user="hive"
2	2017-11-02 07:09:00	2017-11-02 07:09:00	hiveQueryLog	queryPhoneNumber	hive	hiveAccessLogStream	The Policy "queryPhoneNumber" has been detected with the below infor... tion: timestamp="1[redacted]3691" sensitivityType="PHONE_NUMBER" re source="/xademo/customer_details/phone_number" command="SELECT" user="hive"

总结与展望

THANKS


HANA
Hadoop
SQL
MySQL
DB2
BigData
Spark
SQL Server
NewSQL
Oracle
MPP
NoSQL
OLAP

讲师申请
联系电话（微信号）：18612470168

关注“ITPUB”更多
技术干货等你来拿~

与百度外卖、京东、魅族等先后合作系列分享活动

微学堂

让学习更简单

微学堂是以ChinaUnix、ITPUB所组建的微信群为载体，定期邀请嘉宾对热点话题、技术难题、新产品发布等进行移动端的在线直播活动。

截至目前，累计举办活动期数60+，参与人次40000+。

ITPUB学院

ITPUB学院是盛拓传媒IT168企业事业部（ITPUB）旗下
企业级在线学习咨询平台
历经18年技术社区平台发展
汇聚5000万技术用户
紧随企业一线IT技术需求
打造全方式技术培训与技术咨询服务
提供包括企业应用方案培训咨询（包括企业内训）
个人实战技能培训（包括认证培训）
在内的全方位IT技术培训咨询服务

ITPUB学院讲师均来自于企业
一些工程师、架构师、技术经理和CTO
大会演讲专家1800+
社区版主和博客专家500+

培训特色

无限次免费播放
随时随地在线观看
碎片化时间集中学习
聚焦知识点详细解读
讲师在线答疑
强大的技术人脉圈

八大课程体系

基础架构设计与建设
大数据平台
应用架构设计与开发
系统运维与数据库
传统企业数字化转型
人工智能
区块链
移动开发与SEO

联系我们

联系人：黄老师
电　话：010-59127187
邮　箱：edu@itpub.net
网　址：edu.itpub.net
培训微信号：18500940168