火山引擎 | DataFun.

字节跳动大数据平台安全与权限治理实践

许从余 火山引擎数据平台产品经理

目录 CONTENT

01

字节大数据安全体系现状和难点

治理原则

治理原则

保证合规，兼顾效率

外部

安全合规的风险压力

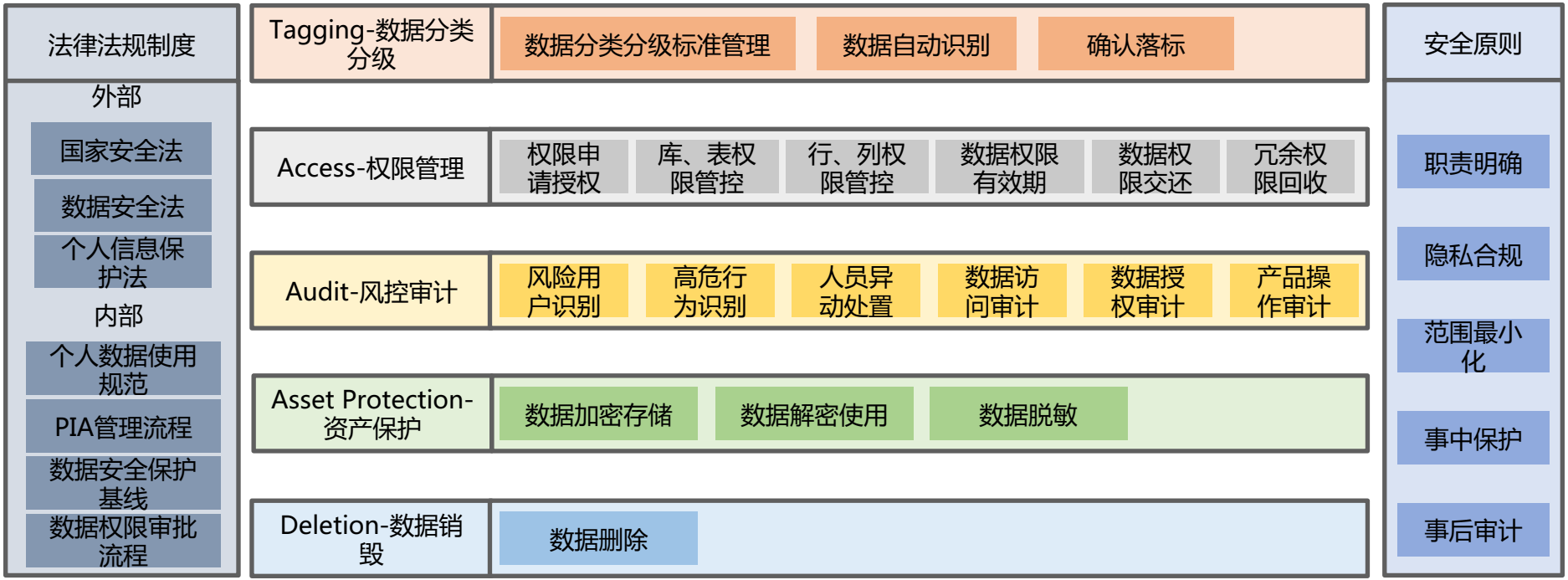
面对政府监管合规要求，暴露出不少问题，例如不能灵活筛选，保留和删除数据等，当没有被很好完成时，核心业务会处于巨大风险中

内部

业务线的效率压力

面对外部压力，不可避免的会出现一些偏临时，偏刚性的要求和机制，在完成合规要求的同时，必须兼顾内部业务运转的效率

字节跳动大数据安全产品体系



02

细粒度权限管控和治理

细粒度权限模型

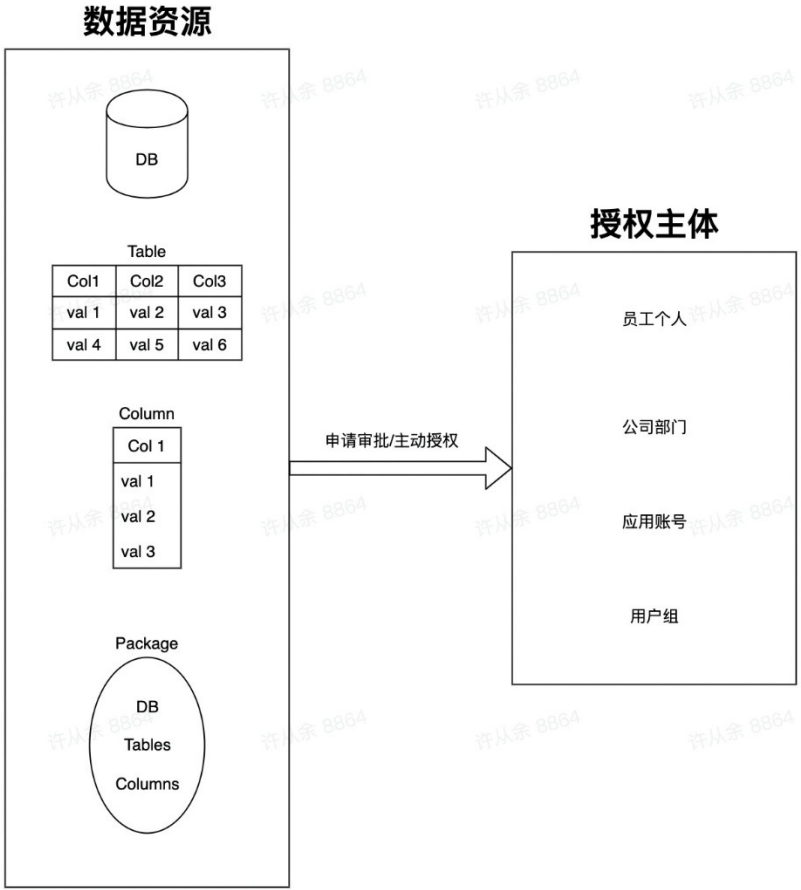
新权限模型特性

列级权限控制　　表/列权限附带行限制　　敏感表/列单独管控

id	name	gender	country	age	race(敏感列)
1	Alice	f	uk	25	black
2	Bob	m	uk	30	white
3	Jack	f	us	22	yellow
4	Lucy	m	us	23	yellow
5	James	m	ca	35	black
6	Lily	f	ca	43	white

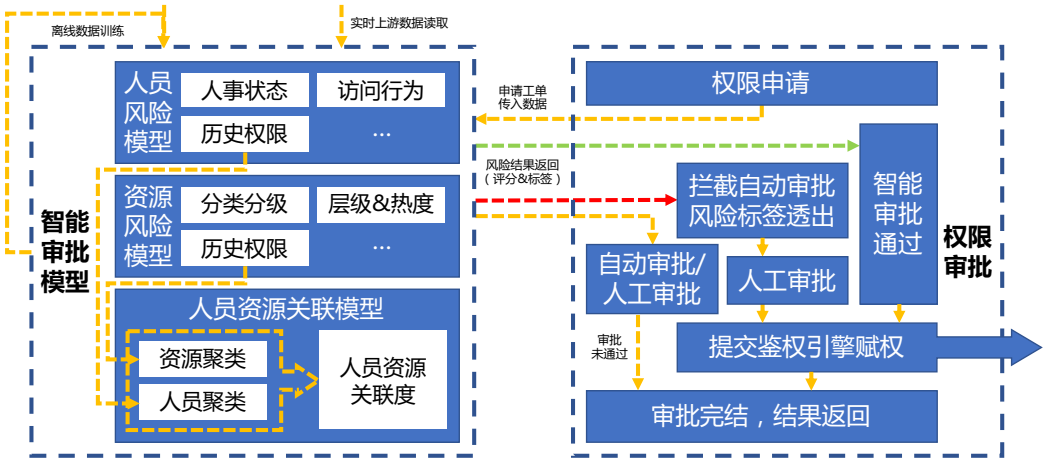
A：DB
B：table
C：table+race(敏感列)
D：id+name+gender
E：table where gender=m and country in (us,ca)
F：country+age+race where country in (uk,us)

灵活的权限授予机制



- 数据资源与授权主体灵活组合
- 审批流灵活定义
- 自动审批 30+%
- 智能风险判断辅助审批

智能审批

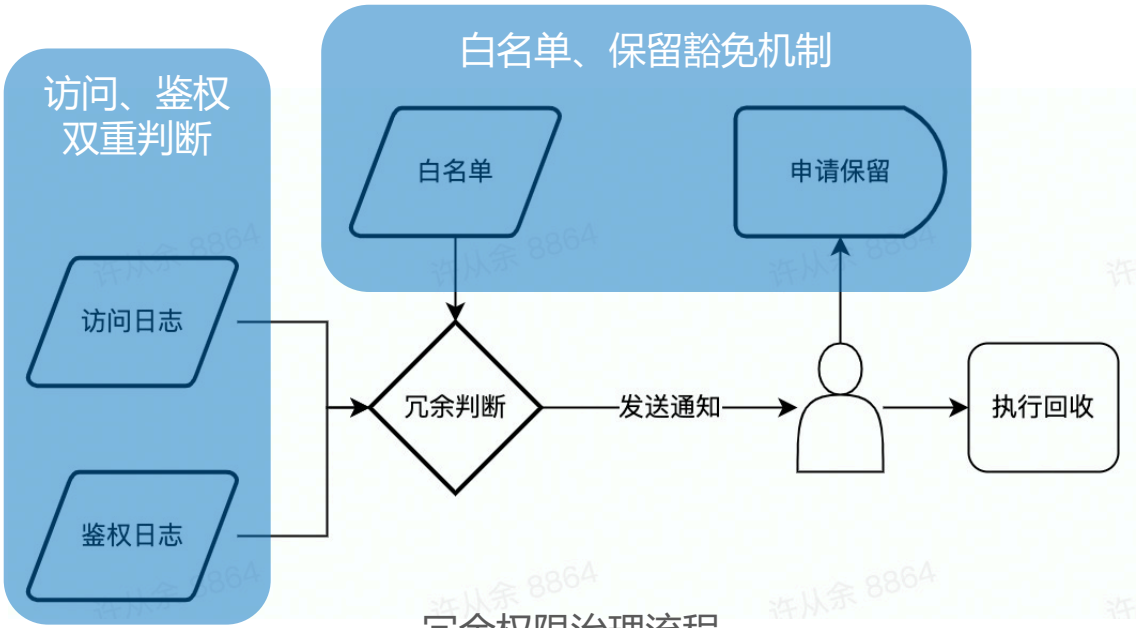


智能审批功能体系

低风险：自动通过
高风险：拦截自动审批，透出风险，人工审批

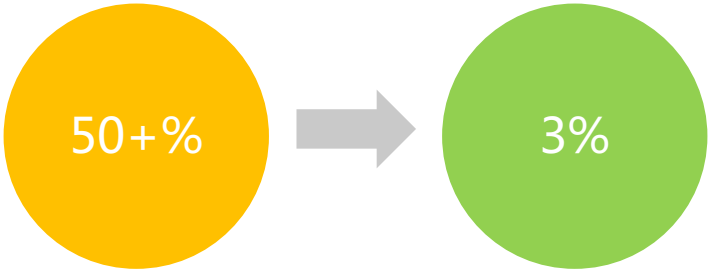
低风险：节约审批时长10万小时
高风险：辅助识别，驳回率高7%

冗余权限治理回收



冗余权限治理流程

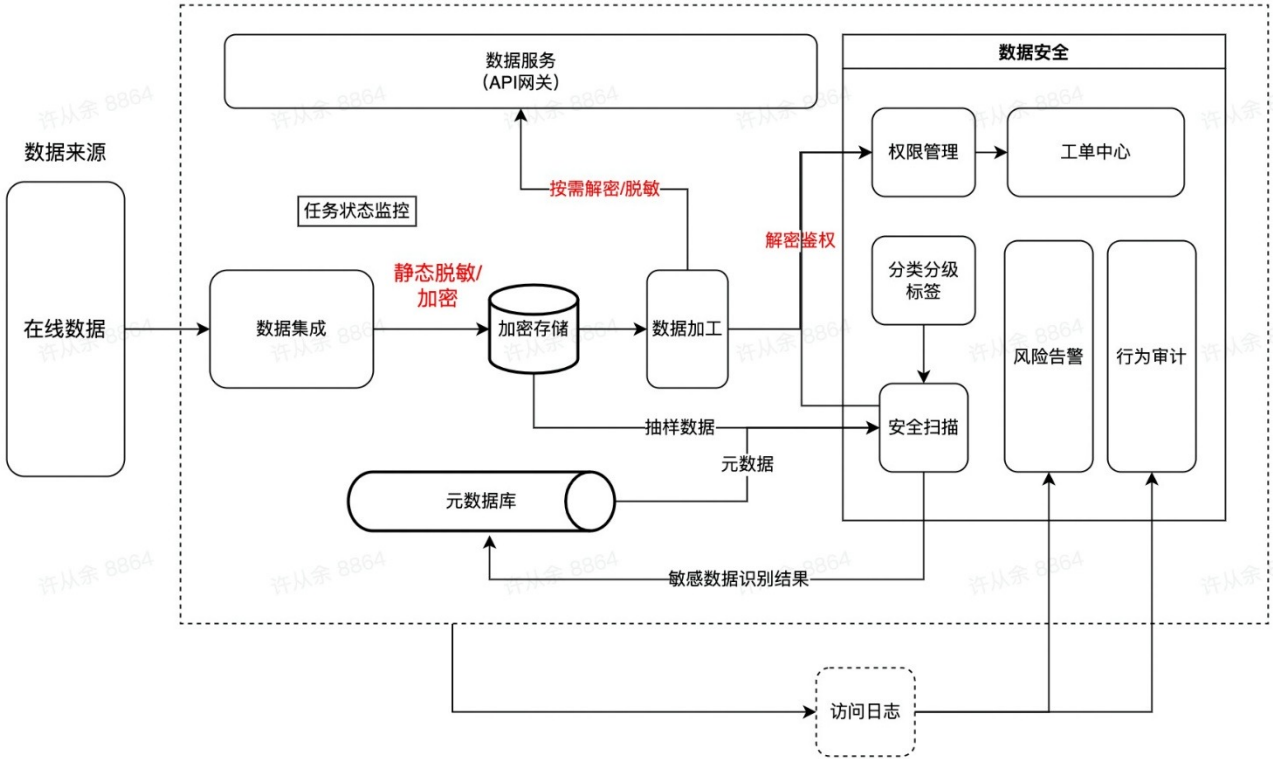
治理效果



03

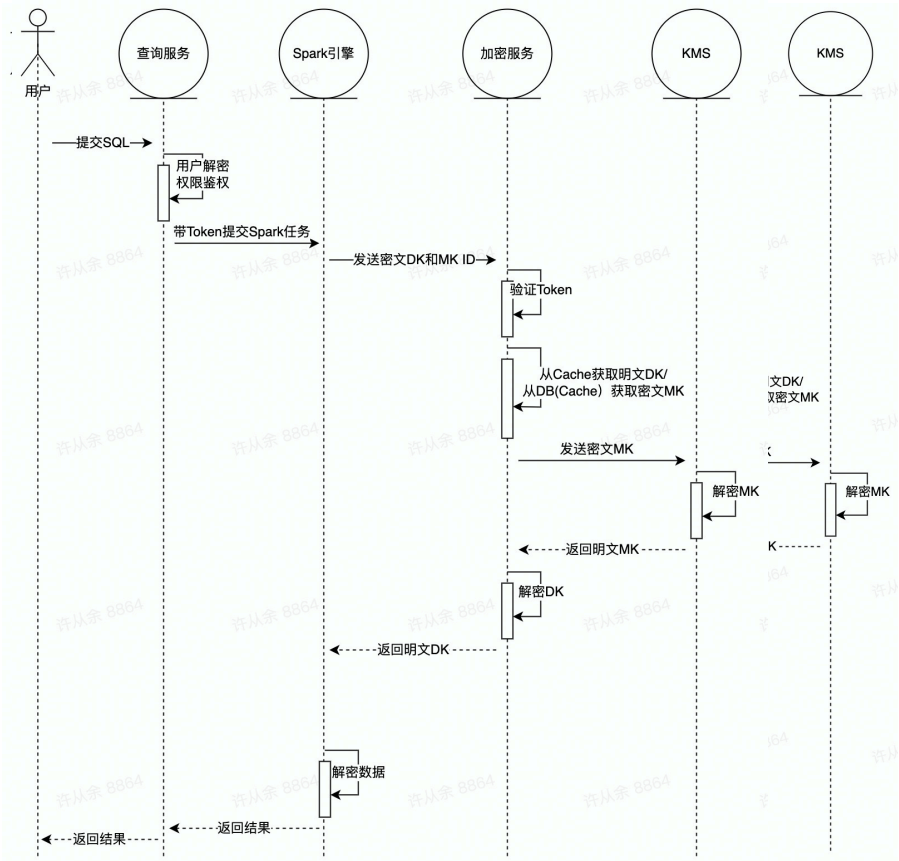
资产保护能力

资产保护应用场景



加密方案介绍

机制	层级	灵活性	强度	兼容性	效率	操作难度	技术难度
数据内容加密	应用级	高	强	低	低	高	低
文件格式透明加密	文件格式	高	中	低	高	低	中
HDFS加密	文件系统	低-中	中	中	中	低	高
磁盘加密	磁盘级	低	中	高	中	低	高



大数据挑战

数据链路长
数据量大
用户多

需满足

高数据一致性
高数据可用性
高效率数据重写
密文具备可识别性

性能优化

用户鉴权交由权限引擎
DataKey缓存

04

数据删除能力

数据删除介绍

删除需求场景

账号删除
滚动删除
其他

性能的提升

基于Bytelake降低覆写总量15倍
Bytelake格式转换速度提升10倍
提升覆写速度80%

调度和系统的优化

系统能力
计算能力
存储能力
数据库能力

大数据删除技术挑战

- 传统HDFS数据删除只能通过覆写文件的方式达成，删除一个用户数据就需要覆写该表所有hdfs文件，对系统I/O消耗巨大。
- 数仓存储在HDFS之上，主要格式：列存储，而对于用户数据遗忘权的满足需要对行级别数据的删除，删除效率低，开销大。
- 离线表的数量庞大。对HDFS的存储资源、磁盘IO、网络吞吐、计算资源、ETL调度系统都会有极大挑战。
- 对业务资源的抢占。
- ETL任务脏读、幻读和不可读的问题

欢迎联系我们

扫码关注

“字节跳动数据平台”微信公众号

扫码添加小助手

进入“字节跳动数据平台”官方交流群

火山引擎 | DataFun.

非常感谢您的观看

火山引擎 | DataFun.