

数据来源：数据库产品上市商用时间



# 第十三届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2022

## 数据智能 价值创新



线上直播 | 2022/12/14-16





# 货拉拉大数据安全体系 建设实践和思考

王海华 货拉拉

# 目录

1

## 背景和挑战

2

## 大数据安全体系

2-1 大数据安全规范

2-2 大数据安全能力建设

2-3 大数据安全治理

3

## 总结与思考

# 1 背景和挑战

# 货拉拉介绍

6+

业务线

352

国内城市

950万

月活用户

66万

月活司机

3+

IDC

1000+

机器数

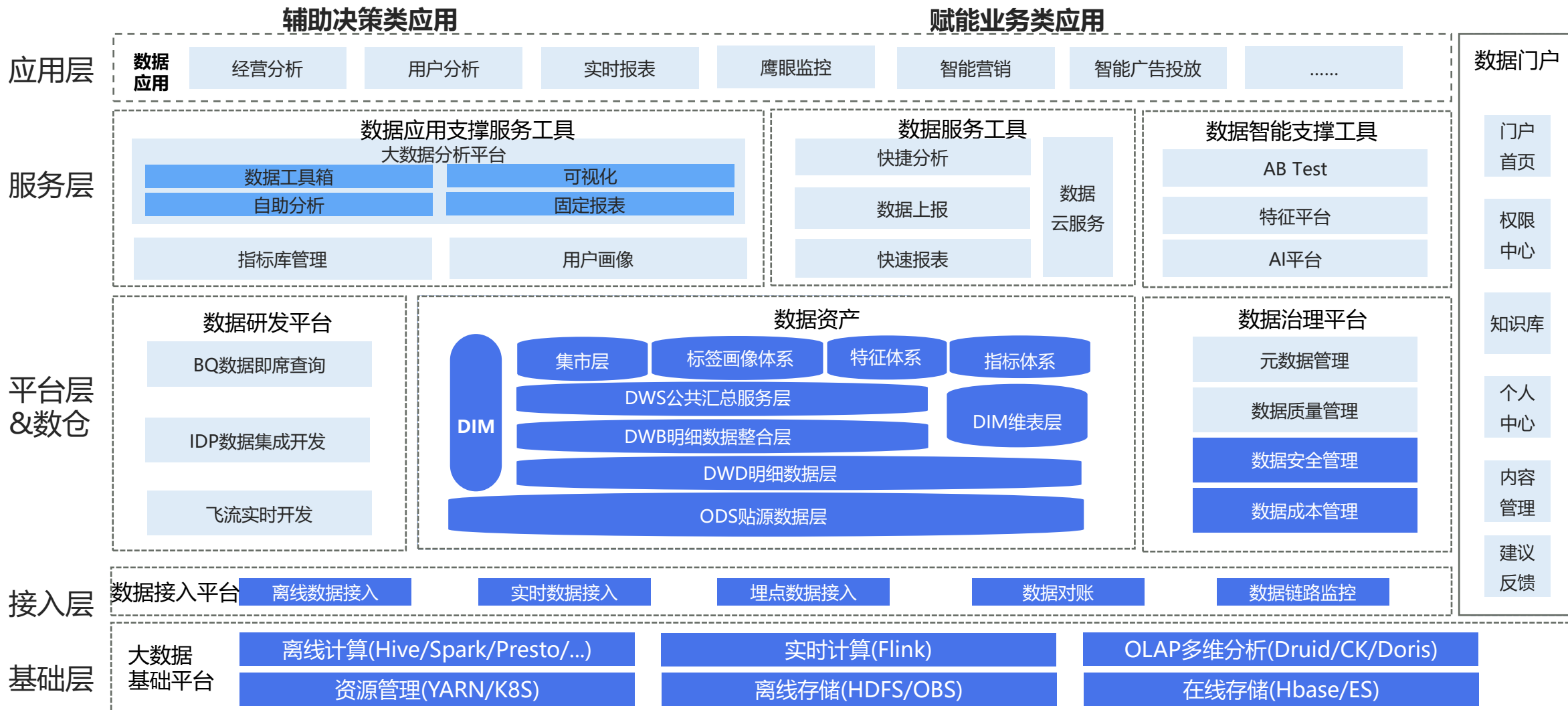
10PB+

存储量

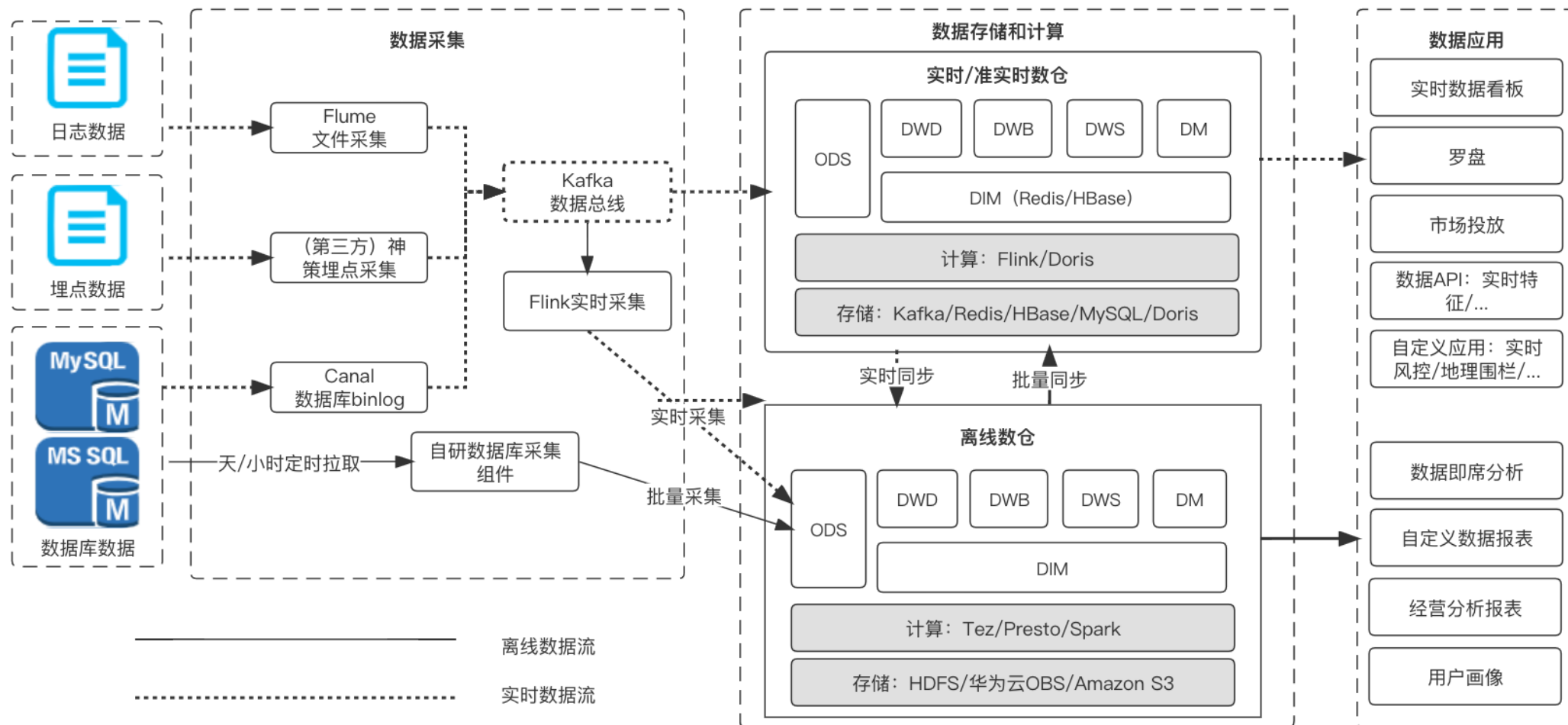
20K

+

日均任务数







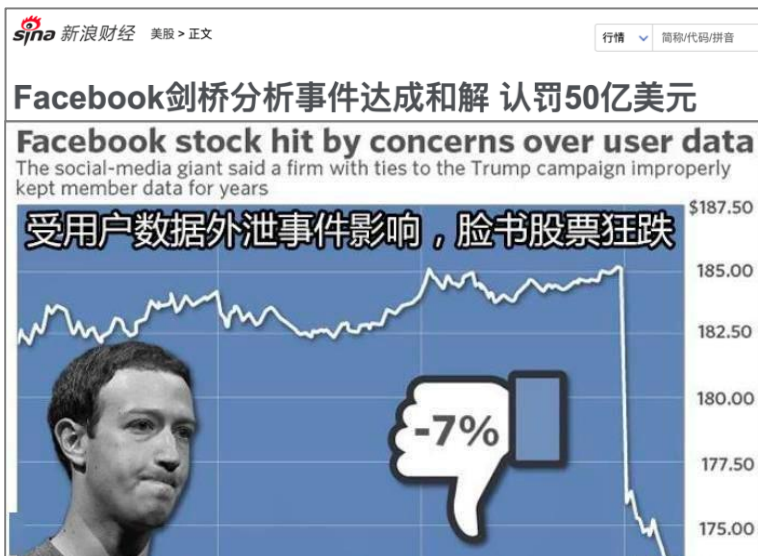
# 为什么要做大数据安全?

## 1 数据资产保护要求

- 商业机密
- 经营数据
- 用户信息

## 2 法律法规要求

- 个人信息保护法
- 网络安全法
- 数据安全法
- 数据安全管理办法



因违反欧盟隐私规定，谷歌脸书或被罚超15亿元

监管 · 隐私护卫队 · 2022-01-06

法国数据监管机构CNIL将对谷歌和Facebook分别处以1.5亿欧元（约合10.81亿人民币）和6000万欧元（约合4.32亿人民币）的罚款。



## 难点与挑战

### 难点



#### 数据资产类型多, 管控复杂度高

- 10P+数据量, 8+数据资产
- 数据集中存储、敏感数据多
- 生命周期长



#### 使用场景多, 攻击面广

- 10+使用场景
- 需要统筹考虑

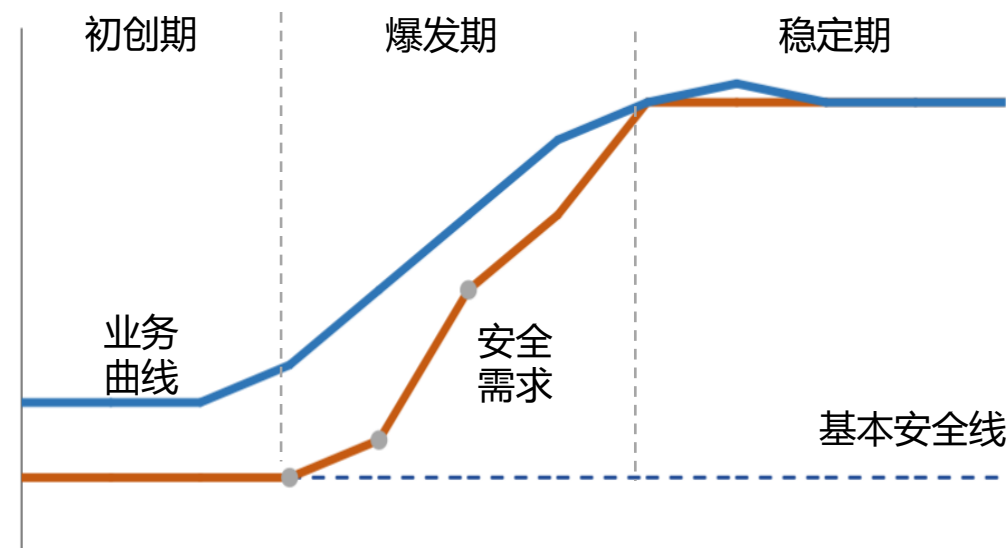


#### 数据产品多, 用户数量大

- 20+大数据产品
- 5000+用户数量

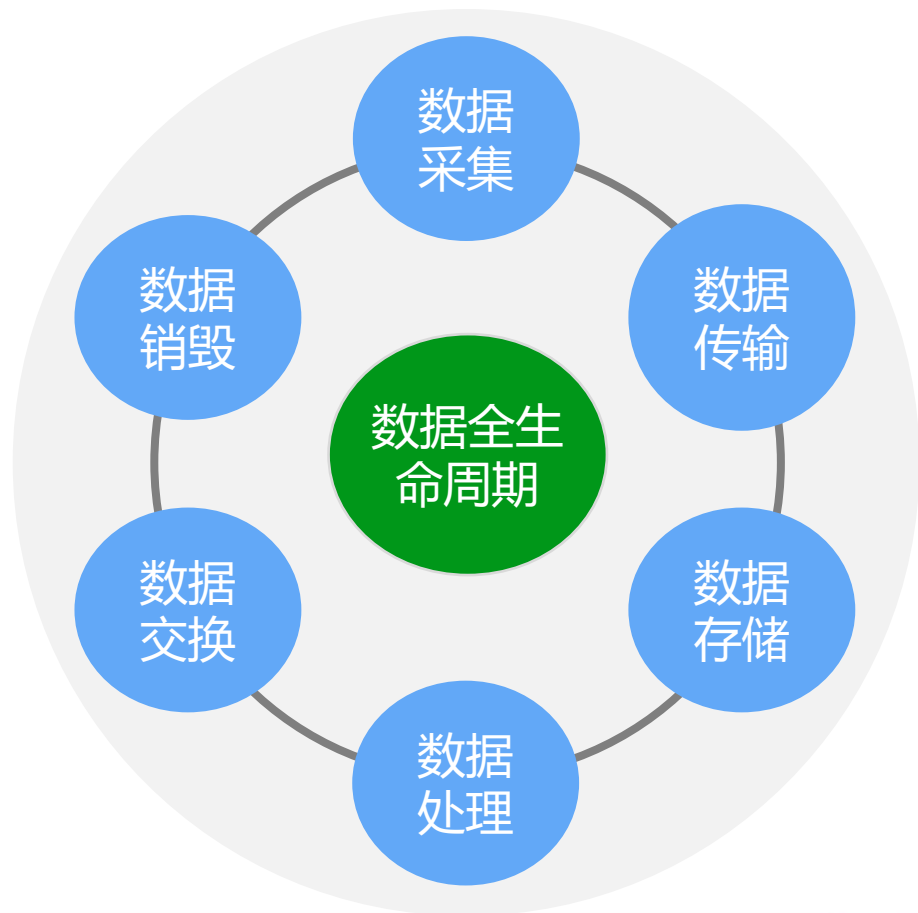
### 挑战

平衡数据安全和业务发展的关系



# 大数据安全体系建设思路

建立**全数据生命周期**的**安全防护体系**，防止数据泄露，满足合规需求



01

PART ONE

**建组织**

建立组织保障



02

PART TWO

**立规范**

有法可依



03

PART THREE

**建能力**

围绕数据生命周期构建安全能力



04

PART FOUR

**做治理**

解决存量安全问题



# 2 大数据安全体系

# 大数据安全体系概览



# 2-1 大数据安全规范



# 数据安全规范 - 敏感分级

数据敏感分类分级结合公司业务场景，同时参考了金融数据安全分类分级标准：  
《金融数据安全数据安全分级指南》（JR/T 0197—2020）

分级名称	定义	被利用价值	使用范围	重要程度
公开数据(C1)	已通过正规渠道正式对外发布的数据，不会对公司造成影响的数据	无价值	外部公开	一般
限制数据(C2)	不适合对外公开，但是对内部人员访问基本无限制的数据，一旦发生泄露，不会对数据主体造成直接损害	低价值	公司内部	敏感
商业秘密(C3)	公司专有或公司保密的，一旦发生泄露，将显著影响相关业务的开展，对数据主体造成直接或者间接损害	中价值可间接利用	公司内部限于相关人员	重要
核心秘密(C4)	具有最高安全属性要求，一旦发生泄露，可能导致公司法律或商业上造成重大影响和损失	高价值可直接利用	公司重要部门特定人员	关键

# 数据安全规范 - 敏感分级

库表：算法定级为主，人为定级为辅

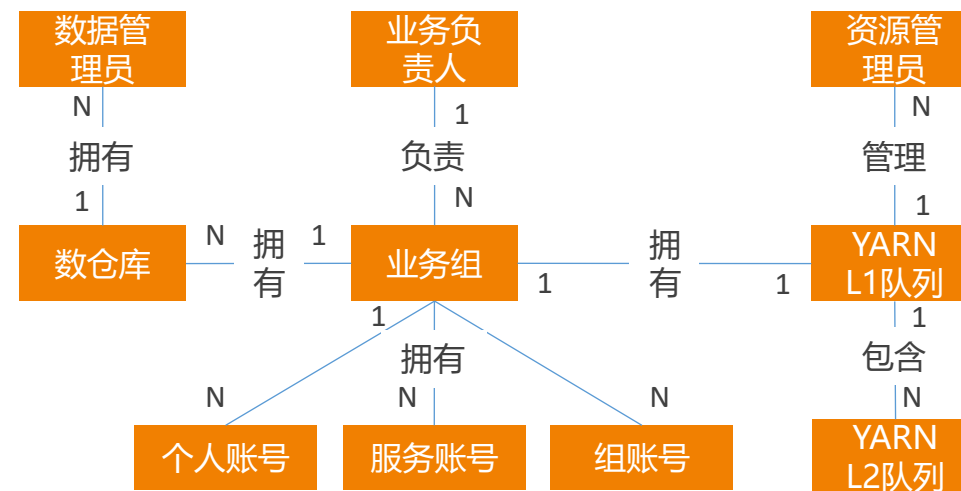
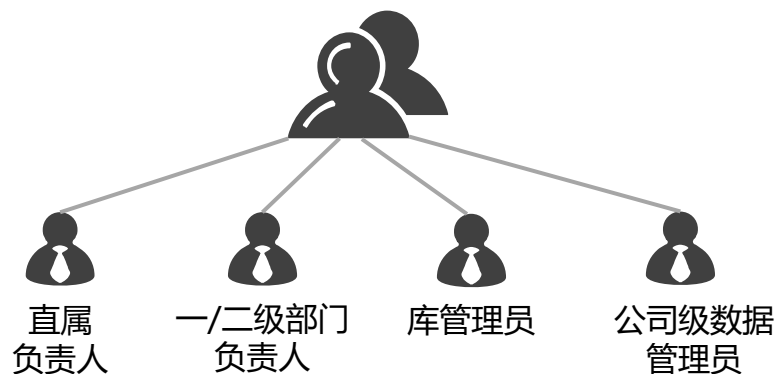
报表：研发人工定级，报表管理员审批

指标：研发人工定级，指标管理员审批

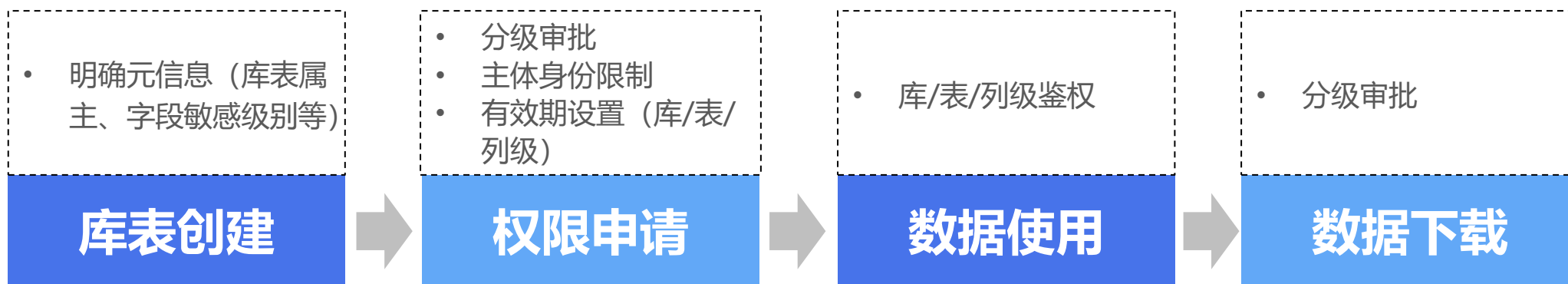


# 库表安全管理规范

## 角色定义



每个数仓库必须有业务组归属

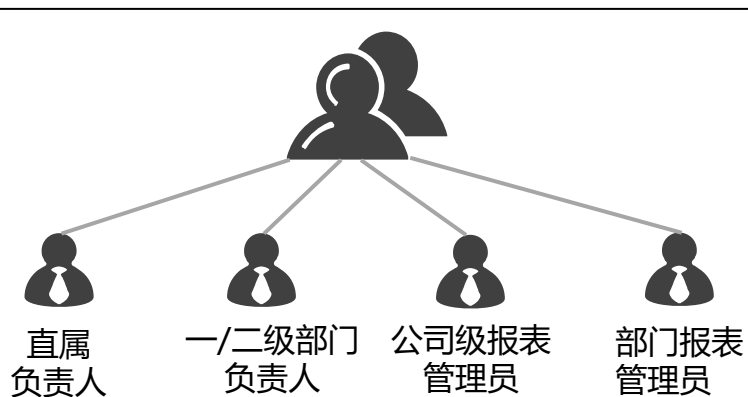


行为审计

序号	发起方	发起方所在部门	库表	库表所属部门	安全级别	备注	审批流程
1	userA	A	db1.t1	A	C1/C2	部门内非敏感数据	userA -> leader3A -> adminA
2	userA	A	db1.t2	A	C3/C4	部门内敏感数据	userA -> leader3A -> adminA -> adminHLL
3	userA	A	db2.t1	B	C1/C2	跨部门非敏感数据	userA -> leader3A -> adminB
4	userA	A	db2.t2	B	C3/C4	跨部门敏感数据	userA -> leader3A -> <b>leader2A</b> -> adminB -> <b>adminHLL</b>

## 库表权限审批流程详情

## 角色定义



- 明确元信息（明确属主、分级等）
- 上线审批

## 报表上线

- 分级审批
- 主体身份限制
- 有效期设置

## 报表使用

- 分级审批
- 下载有效期
- 数据量限制

## 报表下载

## 行为审计

发起方	发起方所在部门	报表所属部门	安全级别	备注	审批流程
userA	A	A	C1/C2	部门内非敏感报表	userA -> leader3A -> adminA
userA	A	A	C3/C4	部门内敏感报表	userA -> leader3A -> adminA -> adminS
userA	A	B	C1/C2	跨部门非敏感报表	userA -> leader3A -> adminB
userA	A	B	C3/C4	跨部门敏感报表	userA -> leader3A -> leader2A -> adminB -> adminS

报表使用流程规范

安全级别	备注	审批流程	单次申请时限	单次申请可下载条数
C1/C2	非敏感报表	userA -> leader3A	1~7天	无限制
C3/C4	敏感报表	userA -> leader3A -> adminS	1~7天	<=上限
C3/C4	敏感报表	userA -> leader3A -> leader1A -> adminS	1~7天	>上限

报表下载申请流程规范



# 高敏感数据存储和使用规范

01

PART ONE

## 高敏感数据存储

入仓加密、高敏明文数据  
独立空间存储



02

PART TWO

## 高敏感数据使用

脱敏使用、解密严格审批、解密  
条数限制



03

PART THREE

## 高敏感数据下载

脱敏下载、解密下载严格审批、  
下载条数限制



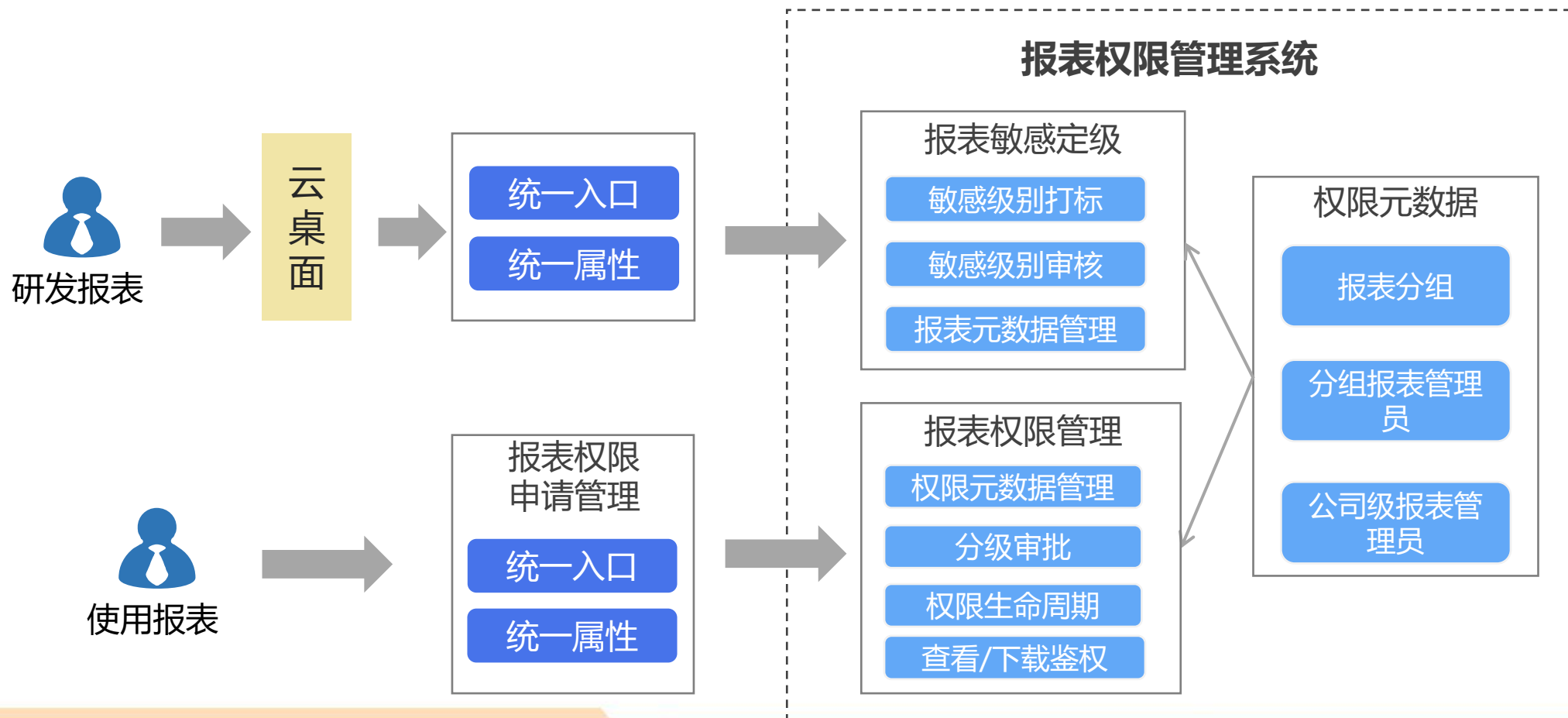
等级	类别	内容	必须加密字段	建议加密字段
C4	身份信息	身份证号、护照号、行驶证号、驾驶证号、照片、手机号	手机号、身份证号、驾驶证号	
C4	银行卡信息	借记卡账号、信用卡账号	银行卡号	
C4	精确定位信息	常用经纬度、精确地址信息(精确到街道)、邮寄详细地址、身份证地址、拉货订单信息	/	/
C4	车辆重要信息	车牌号码(包括用户车辆及车外设备拍摄的其他车辆的车牌号码)	/	车牌号

# 2-2 大数据安全能力建设

## 覆盖库表权限全生命周期，支持列级细粒度鉴权

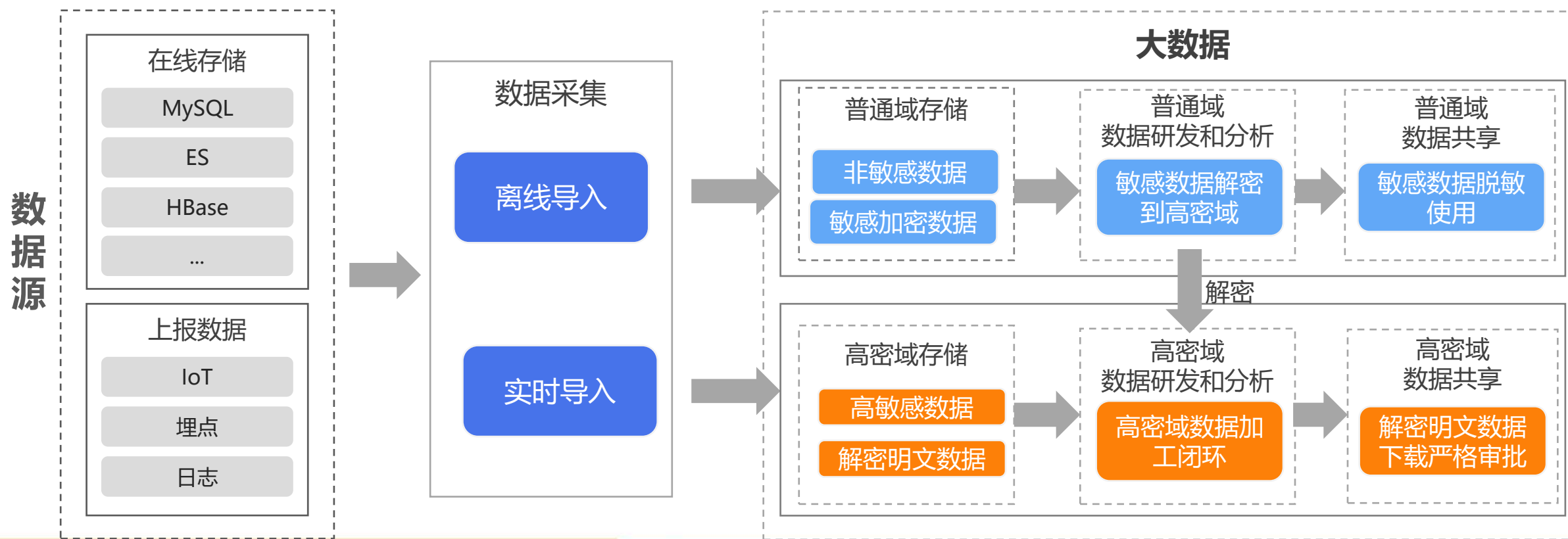


覆盖报表**权限全生命周期**，包含**研发、上线、使用**等过程



# 高敏感数据加密和脱敏

- 在线加密敏感数据导入加密兼容
- 高敏数据和解密数据独立空间存储，严格审批





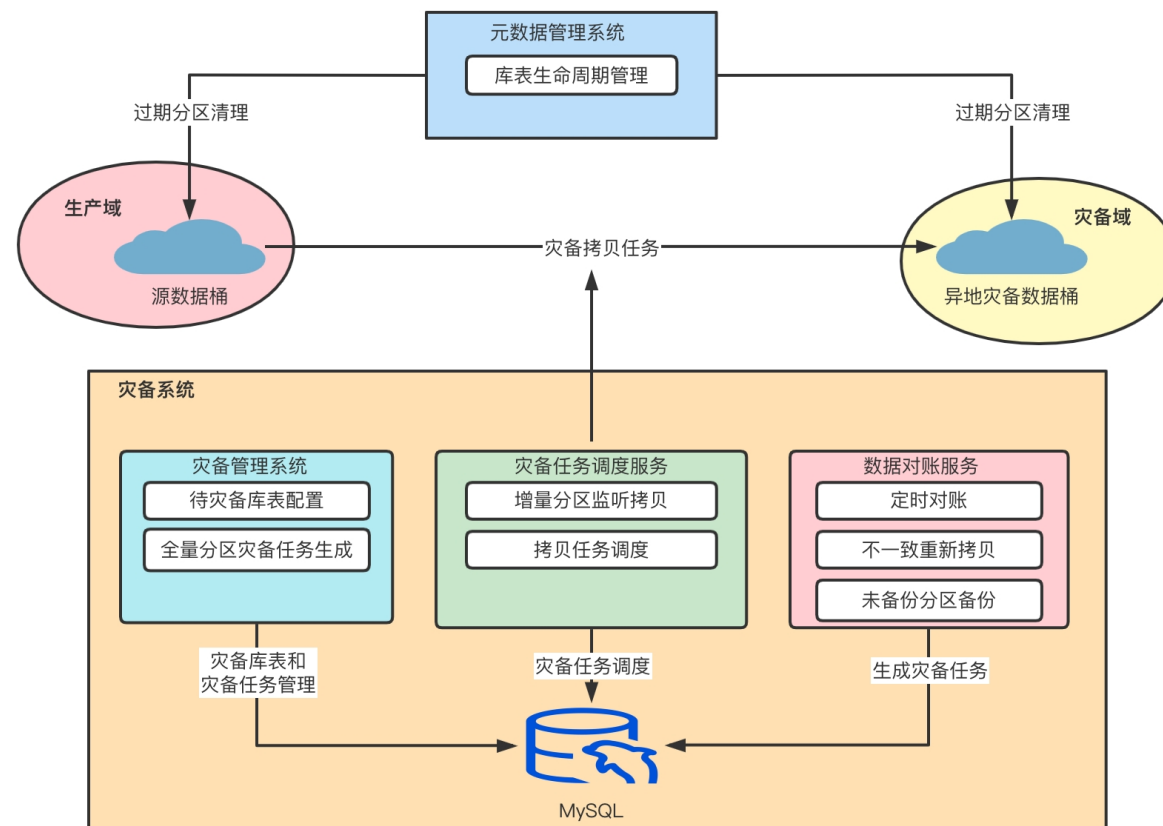
# 数据灾备能力

大数据系统所面临的风险和威胁无处不在：

- 人为误删
- 云机房故障
- ...

造成核心数据丢失

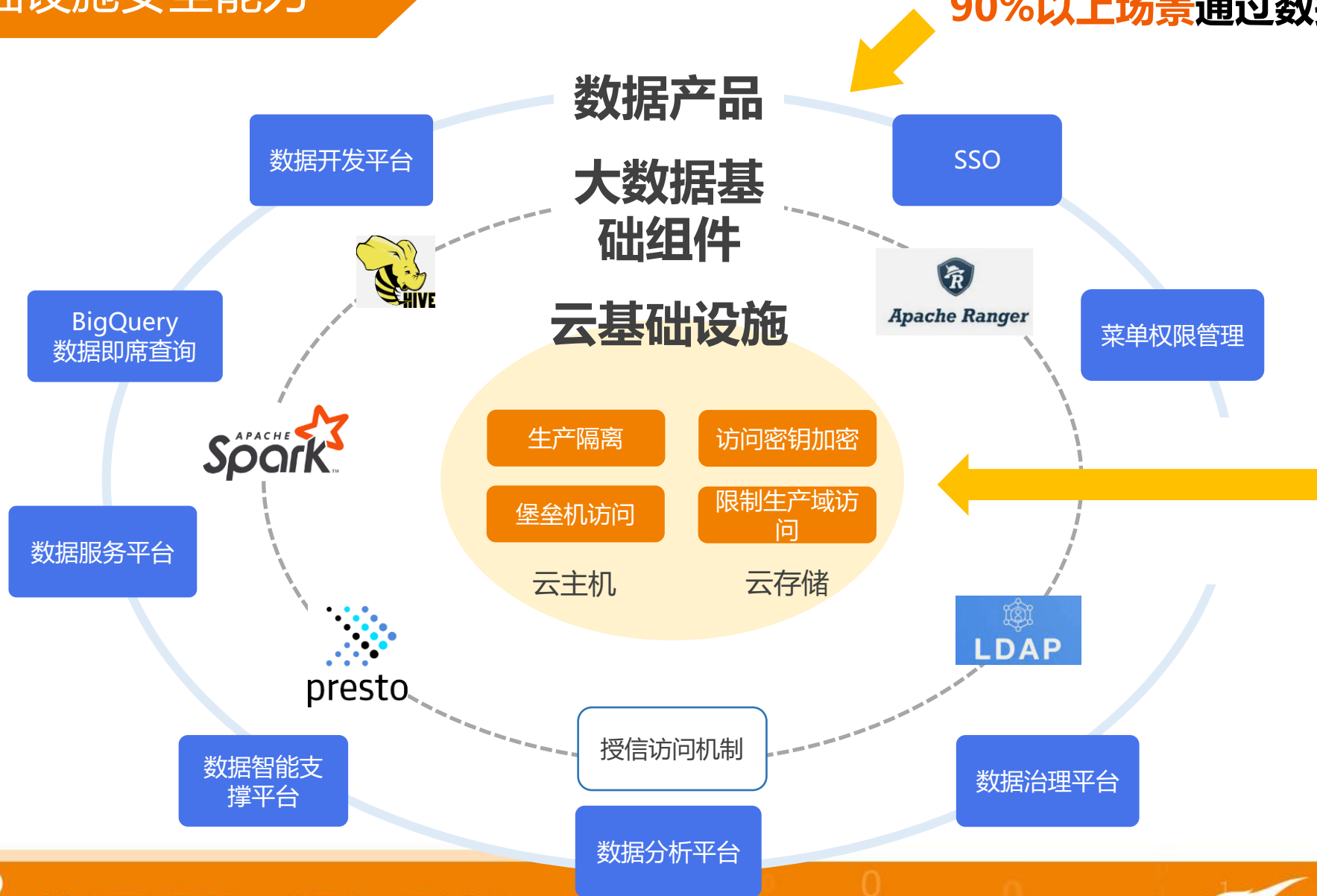
数仓重要原始/结果数据异地备份



## 基础设施安全能力

90%以上场景通过数据产品进行

10%以下场景直接通过基础组件或者云基础设施



# 2-3 大数据安全治理



开展专项治理



逐部门  
逐项治理



治理结果

数仓库表安全治理

- 库治理：每个库必须有部门归属、数据敏感分级
- 不合理权限收归：研发不得拥有非本部门的库级权限
- 权限有效期覆盖和批量过期

- 库表敏感分级覆盖
- 库全部有归属
- 不合理库表权限全部回收

报表安全治理

- 报表归属和敏感分级覆盖率提升
- 非岗位必需敏感报表权限回收
- 跨部门报表权限批量回收

- 报表敏感分级和归属覆盖
- 不合理报表权限全部回收

高敏感数据治理

- 高敏感数据全量加密、或迁移高密域
- 脱敏函数推广和替代解密
- 非数据研发人员数据研发权限回收

- 高敏感数据全部标识
- 不合理研发系统权限全部回收

# 3 总结和思考



## 大数据安全防护体系：全数据生命周期覆盖 + 有效防护方法（规范、能力、治理）

### 数据安全规范

- 数据分类分级
- 库表安全管理规范
- 报表安全管理规范
- 高敏感数据存储使用规范

### 数据安全能力

- 数仓库表安全
- 数据报表安全
- 高敏感数据存储使用安全
- 数据灾备和基础设施安全

### 数据安全治理

- 数仓库表安全治理
- 报表安全治理
- 高敏感数据治理

立规范

建能力

做治理






**安全是业务1前面的那个0，安全投入需要跟业务投入取得平衡**



**大数据安全需要数据生命周期全局和体系化保障，不能只靠局部点突破**



**借鉴业界和专业安全团队最佳实践，同时结合公司实际情况可落地能解决问题**

-  **对标行业，安全能力成熟度提升：中->高**
-  **安全攻防，避免事故样本太小能力效果不佳问题**
-  **产品能力完善，部分线下 -> 全面线上化**

# THANKS

SQL Server  
vertica  
D B 2  
G B a s e  
O r a c l e  
达梦数据库  
神舟通用  
KingbaseES

2010

2014

2018

openGauss  
OceanBase  
ArkDB  
RASESQL  
HotDB  
StellarDB  
QianBase xTP  
云树Shard  
GoldenDB  
DolphinDB  
MatrixDB  
DynamoDB  
SinoDB  
FastData  
Galaxybase  
KunDB  
GDB  
GaussDB  
PolarDB  
TiDB  
Spacture  
Sequoiadb  
OushuDB  
ArgoDB  
开务数据库  
GreatDB  
UbiSQL  
MongoDB  
TDSQL  
Tapdata  
StarRocks