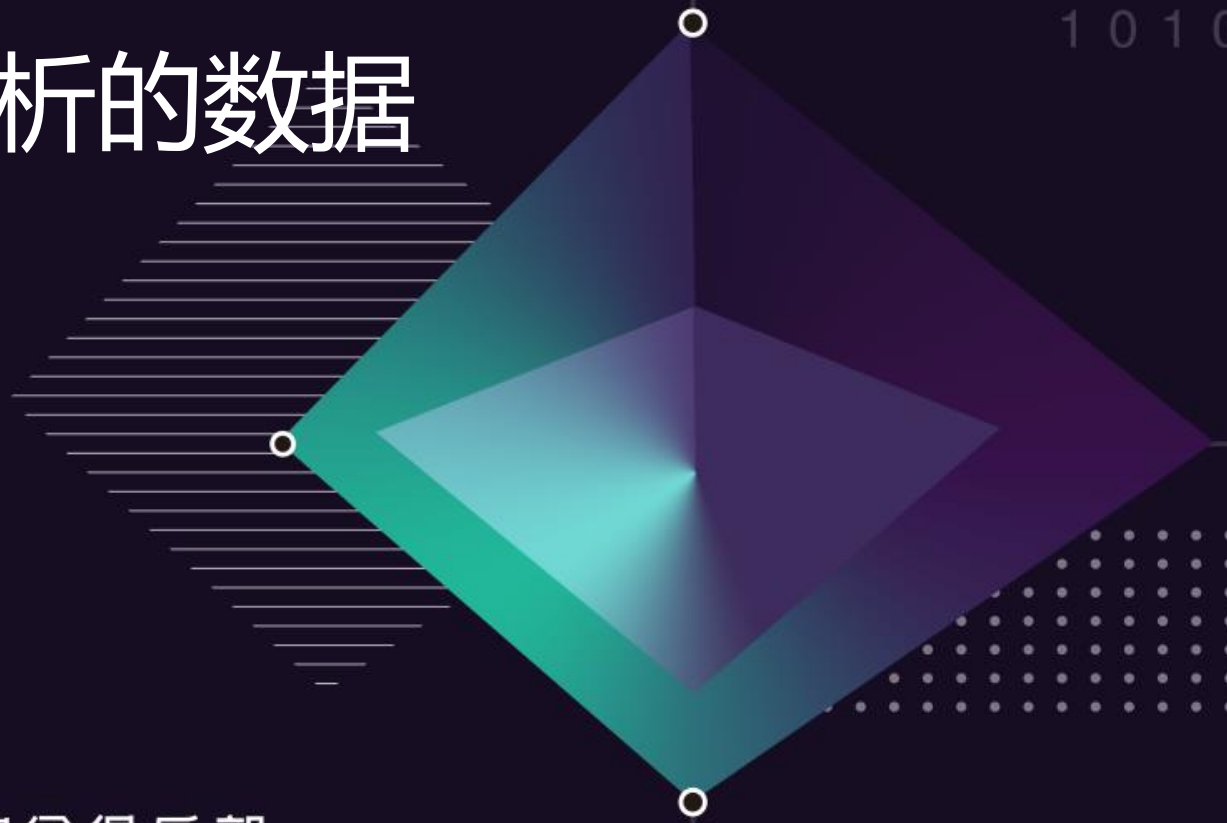
FREEBUF | 企业安全 俱乐部

探索企业安全监控与分析的数据治理之路

张志鹏 安全顾问



2019 | 企业安全俱乐部 数据治理专场

数据的安全靠治理

2018年数据泄露事故top10

编号	组织名称	影响人数	编号	组织名称	影响人数
1	Aadhaar	11亿	6	MyHeritage	9200万
2	Starwood-Marriot	5亿	7	Facebook	8700万
3	Exactis	3.4亿	8	Elasticsearch	8200万
4	Under Armour	1.5亿	9	Newegg	5000万
5	Quora	1亿	10	Panera	3700万

找到问题靠漏洞
解决问题靠治理

数据治理是个过程

是指从使用零散数据变为使用统一数据、从具有很少或没有组织和流程到企业范围内的综合数据治理、从尝试处理数据混乱状况到数据井井有条的一个过程

数据治理的重要性

- **数据的增长**
  - ✓ 大数据技术的广泛应用促进数据量的急剧增长
  - ✓ 隐性数据的未知
- **合规的约束**
  - ✓ 法律要求：《网安法》规定明确导致个人隐私泄漏的情形、责任、及导致泄漏的责任主体及法律责任（处罚、有期徒刑、拘役等）
  - ✓ 行业规范：《**银行业金融机构数据治理指引**》...
- **质量的保障**
  - ✓ 管理者需要准确的数据来帮助决策

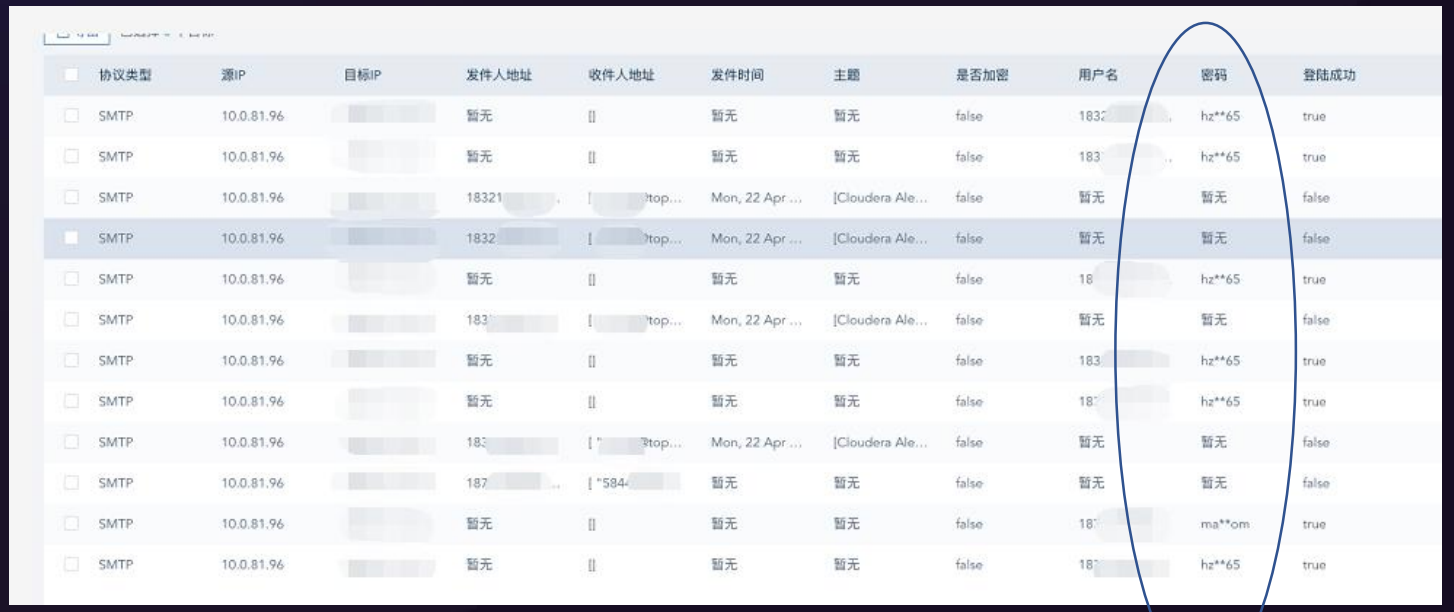
图片来自：数据科学社区

数据治理的漏网之鱼在哪?

被遗忘的数据—安全监控与分析数据

数据治理优先考虑业务数据，
安全监控与分析这类安全数据
往往被忽视

问题一：采集过剩的数据



协议类型	源IP	目标IP	发件人地址	收件人地址	发件时间	主题	是否加密	用户名	密码	登陆成功
SMTP	10.0.81.96		暂无	[]	暂无	暂无	false	1832	hz**65	true
SMTP	10.0.81.96		暂无	[]	暂无	暂无	false	183	hz**65	true
SMTP	10.0.81.96		18321	[top...	Mon, 22 Apr ...	[Cloudera Ale...	false	暂无	暂无	false
SMTP	10.0.81.96		1832	[top...	Mon, 22 Apr ...	[Cloudera Ale...	false	暂无	暂无	false
SMTP	10.0.81.96		暂无	[]	暂无	暂无	false	18	hz**65	true
SMTP	10.0.81.96		183	[top...	Mon, 22 Apr ...	[Cloudera Ale...	false	暂无	暂无	false
SMTP	10.0.81.96		暂无	[]	暂无	暂无	false	183	hz**65	true
SMTP	10.0.81.96		暂无	[]	暂无	暂无	false	18	hz**65	true
SMTP	10.0.81.96		183	[top...	Mon, 22 Apr ...	[Cloudera Ale...	false	暂无	暂无	false
SMTP	10.0.81.96		187	["584	暂无	暂无	false	暂无	暂无	false
SMTP	10.0.81.96		暂无	[]	暂无	暂无	false	18	ma**om	true
SMTP	10.0.81.96		暂无	[]	暂无	暂无	false	18	hz**65	true

- 数据采集设备的明文密码

被遗忘的数据—安全监控与分析数据

问题二：安全数据的复杂无序，无法管理

由安全设备采集的业务方的数据

- 镜像的流量数据（IPS）
- 日志数据（SOC/SEIM）
- 配置数据（SOC/SEIM）
- ……

由安全设备处理后产生的数据

- 安全漏洞数据（漏扫设备）
- 行为攻击数据（IDS、WAF）
- 安全报告数据（态势感知）
- ……

安全设备自身的数据

- 用户数据（账号、密码、邮箱）
- 配置数据（密钥信息）
- ……

被遗忘的数据—安全监控与分析数据

问题三：安全数据壁垒导致的“信息孤岛”

业务部门各自为战，安全部门独木难支，

安全数据渐渐被孤立。

问题四：数据权限最小化应用

Eg. 业务分析中采集了个人敏感数据；

问题五：数据销毁不谨慎

安全的数据容易被遗忘在存储介质中

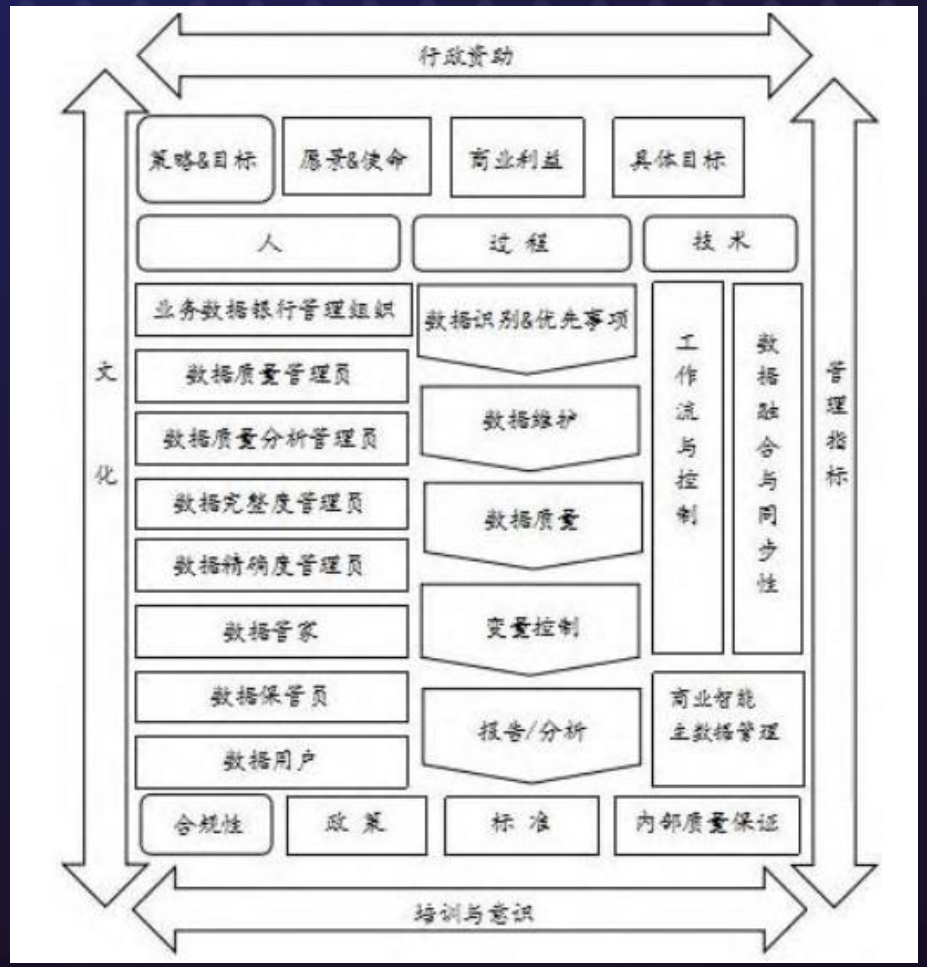
解决这些数据的数据治理应满足企业对数据的治理要求

一些的套路

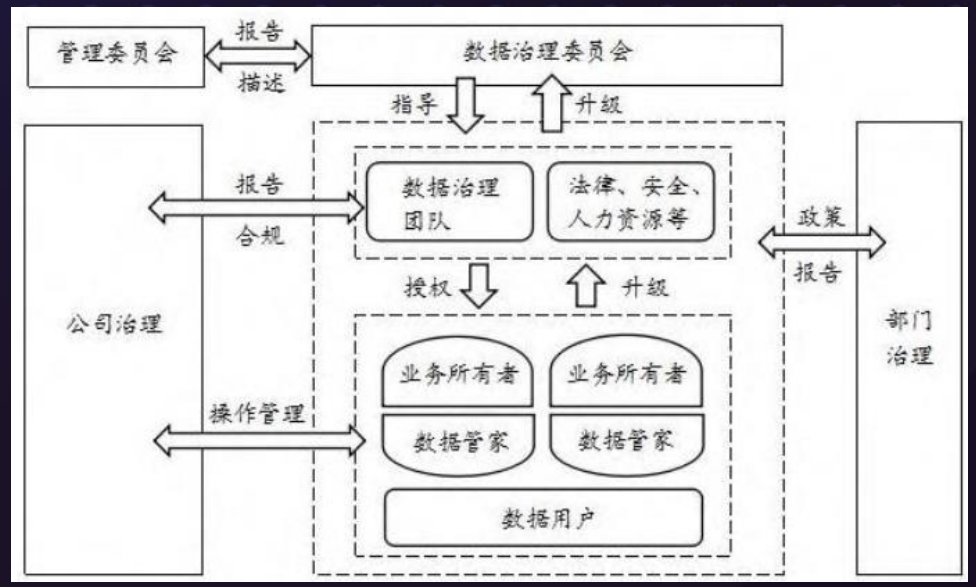
+

一些的工具

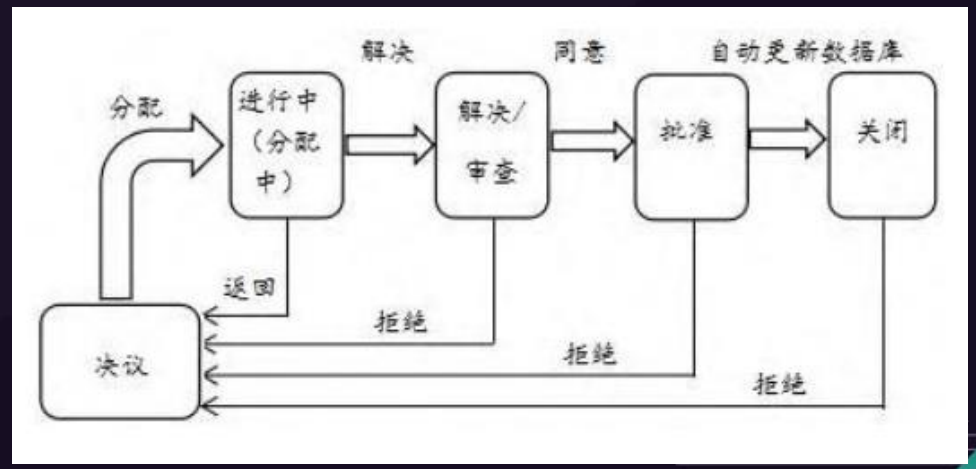
数据治理模型



ISACA数据治理模型



HESA数据治理模型



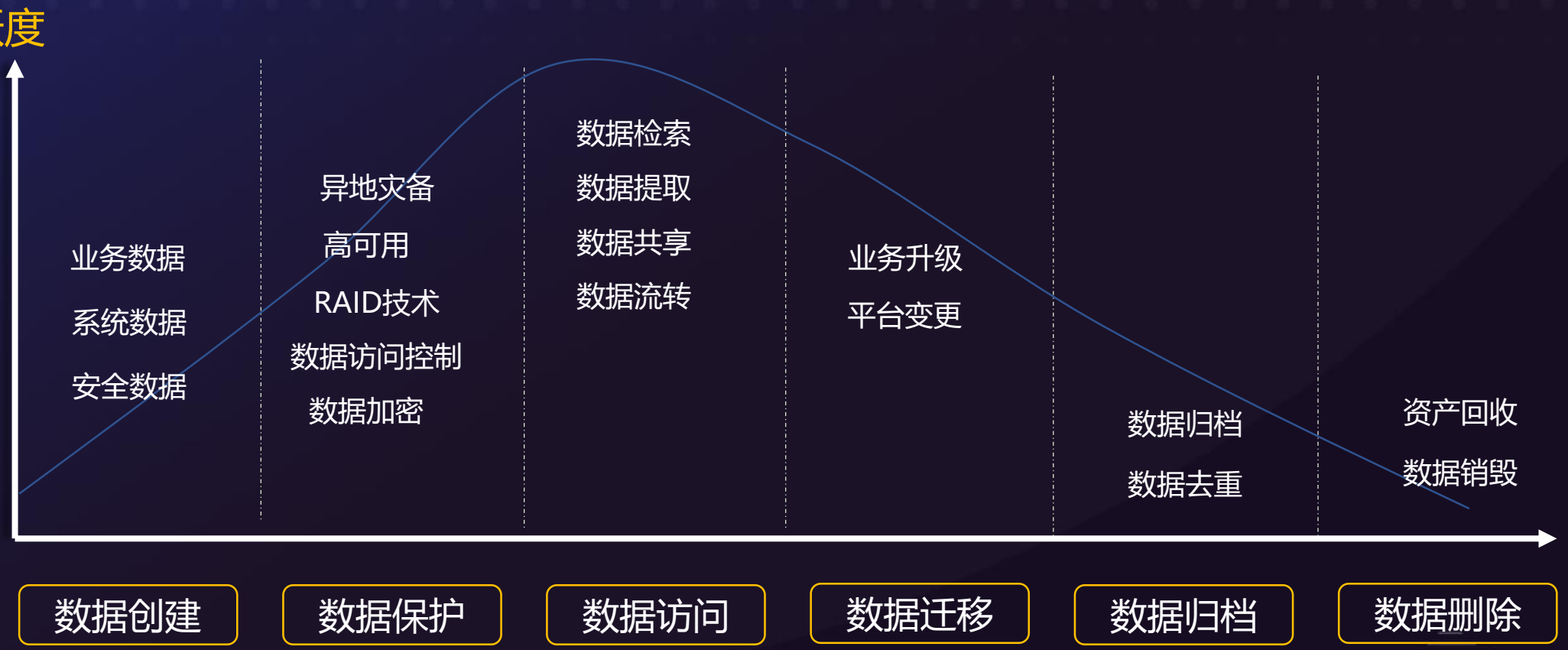
Information Builder 数据治理模型

对比结果

模型名称	ISACA	HESA	Information Builder
组织范围	国际	英国	美国
特点	行政资助、文化、管理指标、培训与意识培养四个角度；体现人主导的作用；顶层设计，内部实施，内外结合的原则。	强调数据治理模型与组织的设计与管理结构密切相关；强调分配的关键角色，而非任务。	数据的保密性、质量和完整性是核心；采用重复渐进的方法，配以7个步骤辅助实施。
优势	充分考虑人的重要性，内外结合，十分全面	依托于组织的设计和管理的结构	容易操作，具备较高的可行性
不足	比较复杂，难以全套落地	较为复杂，依赖定义角色	过于简单，遵从7个步骤，较为死板

各有所长，事无尽善　　按需择取，灵活配置

数据生命周期


活跃度
业务数据
系统数据
安全数据
异地灾备
高可用
RAID技术
数据访问控制
数据加密
数据检索
数据提取
数据共享
数据流转
业务升级
平台变更
数据归档
数据去重
资产回收
数据销毁
数据创建
数据保护
数据访问
数据迁移
数据归档
数据删除

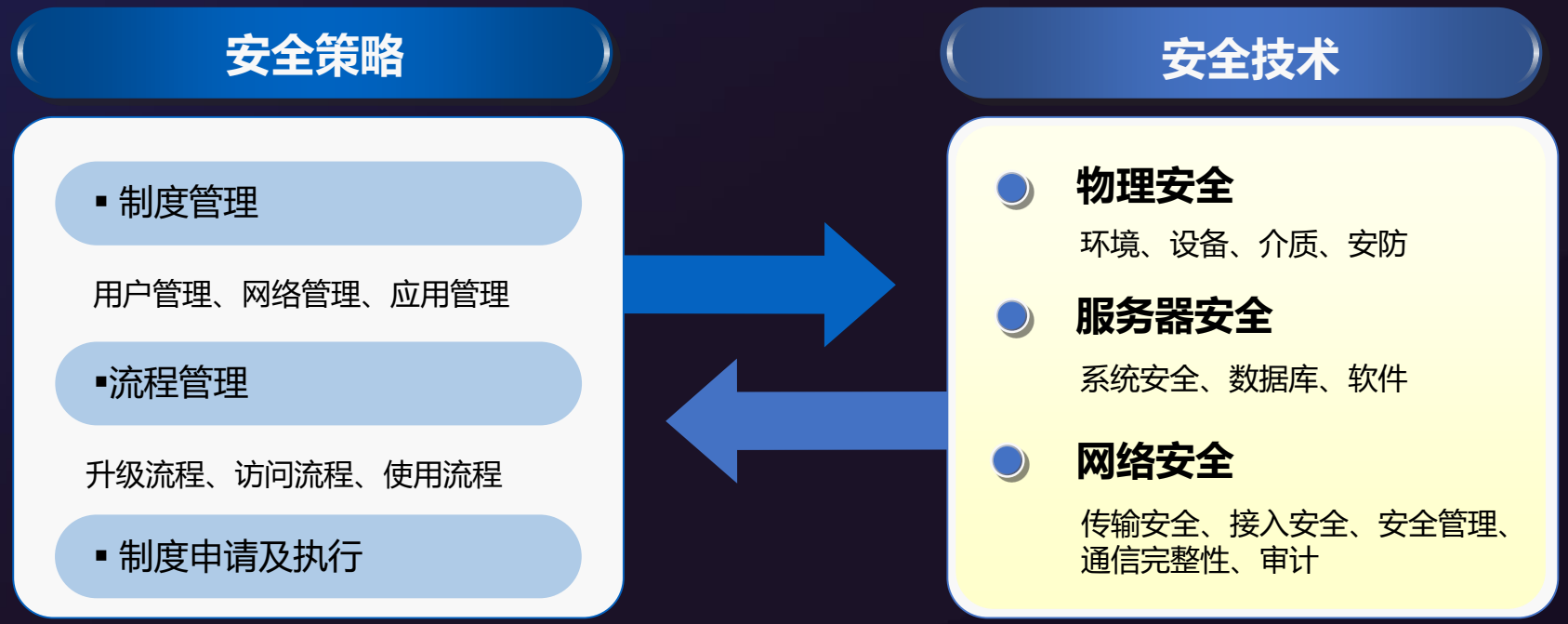
定义数据的共性信息

分类：来源+内容+用途

分级：数据价值、敏感度、影响（CIA）、扩散范围

编码：用来描述数据具体对象与特征，具备唯一性

数据的安全


安全策略
▪ 制度管理
用户管理、网络管理、应用管理
▪流程管理
升级流程、访问流程、使用流程
▪ 制度申请及执行
安全技术
物理安全
环境、设备、介质、安防
服务器安全
系统安全、数据库、软件
网络安全
传输安全、接入安全、安全管理、
通信完整性、审计

安全数据治理的实施流程

体系构建

设计可实施的治理模型
设计可落地的企业级标准

现状梳理

元数据梳理与采集
构建企业数据资料库

应用实践

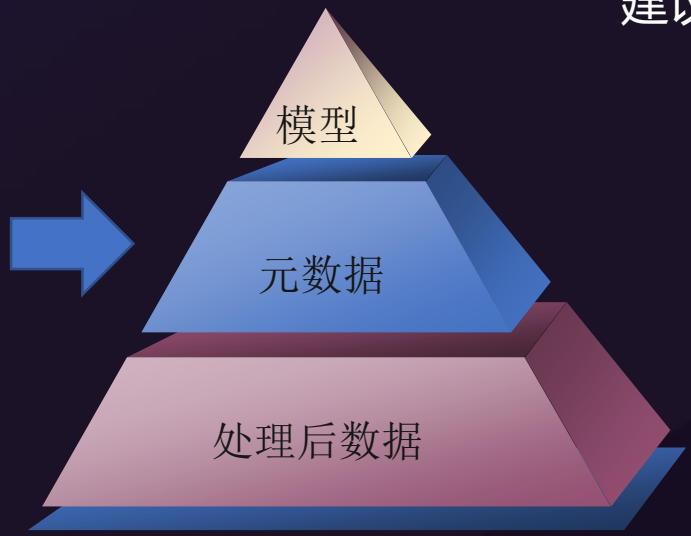
选择必要的工具及技术
实施、监督与管理

有效的数据治理框架

总则	数据治理架构	数据管理	数据质量控制	数据价值管理	数据监督管理
• 法律依据 • 适用范围 • 总体要求 • 监督依据	• 各部门规章制度 • 数据的生命周期与安全边界 • 数据流程规范	• 制定数据战略 • 数据管理制度 • 监管统计制度 • 数据标准 • 信息系统 • 数据共享 • 数据安全 • 数据存储 • 应急方案	• 质量控制整体要求 • 业务制度 • 技术工具 • 考核指标 • 整改制度 • 质量控制	• 风险管理有效性 • 风险监控 • 内容控制 • 数据价值实现要求	• 监督方式 • 监督措施

数据资产梳理

需要搜集和保留哪些数据?
从哪里可以获取这些数据?
谁拥有或管理这些数据?
数据的特点是什么?
梳理后的数据要变成什么样子?



建议:

- 通过自动化工具进行资产数据采集
- 建立数据模型，对数据进行标签化处理
- 建立数据关联关系

数据治理过程中的常用工具

项目	工具
大数据技术	HDFS/Hive/Hbase/Sqoop/Flink/Flume/Spark/Pig
数据库组件	Oracle/MySql/DB2/Subase/SqlServer/PostgreSQL
治理建模工具	ERWin/Powerdesigner/IBM infosphere
ETL工具	Primeton DI/Datastage/PowerCenter

最后

理想的框架、合适的工具、优秀的人才?
不能盲目依从，只有合适的才是事半功倍的组合

THANKS