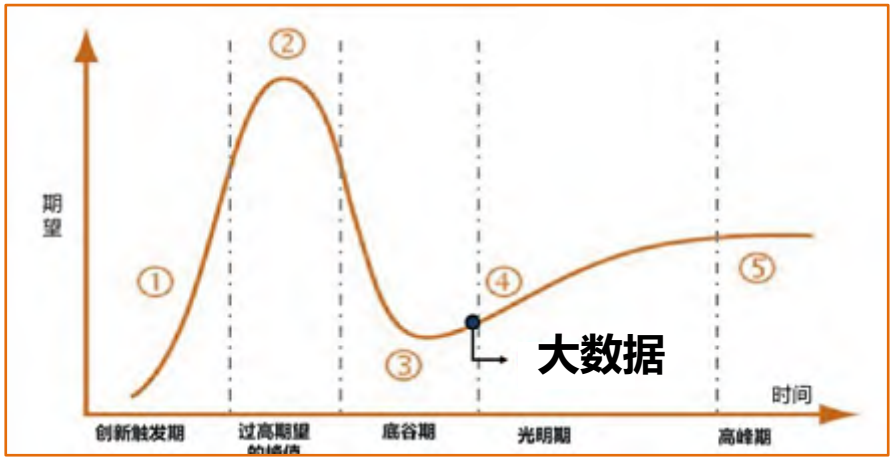
BDTC 2016中国大数据技术大会
Big Data Technology Conference 2016

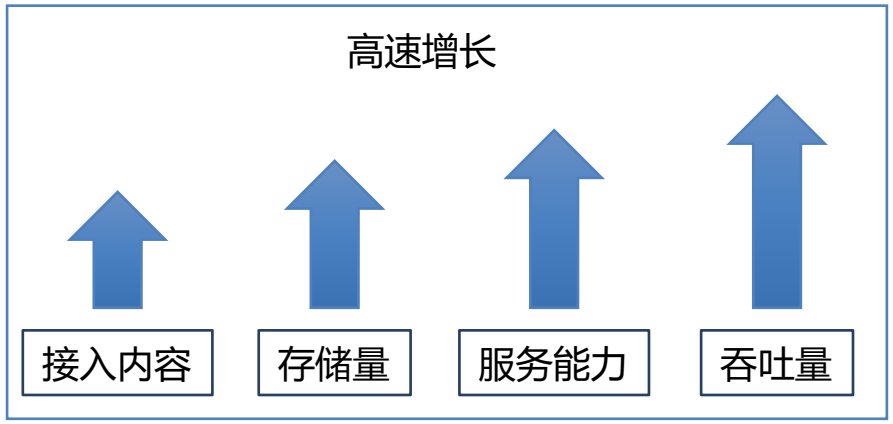
大数据视野下的数据安全防护体系探索

构建可信、可管、可控的大数据安全运行环境

明朝万达
wondersoft.cn

大数据技术成熟度曲线





促进大数据产业发展

●进入成熟期

Gartner技术成熟度曲线报告表明，大数据技术已经进入成熟期，由于大数据应用，具有很强的核心吸引力，必将带动，云计算、物联网、资源虚拟化等相关产业的进一步发展。

●一系列高速增长

由此可以预见：

大数据的接入内容、形式将异常丰富

大数据的存储量将飞速增长

大数据的服务能力将大幅提升

大数据的数据吞吐量将快速增加

大数据发展现状

大数据建设和探索在全国如火如荼的开展....

金融领域大数据
客户画像
精准营销
运营优化
风险控制（反洗钱、反欺诈、贷款风险）
...

能源领域大数据
智能电网
智慧光伏
智慧矿山
...

公安领域大数据
智能交通
大情报研判
重点人员管控
视频监控分析
...

电信领域大数据
智能客服
流量经营分析
个性化精准服务网管优化
...

医疗领域大数据
智能专家诊断
虚假药品分析
早产儿问题分析
...

政府领域大数据
智慧城市
智能政务
...

制造业领域

...

零售业领域

海量数据，带来真正的供给侧革命。通过大数据技术，分析海量数据，推动人工智能与预测。

安全是大数据发展基石

习近平总书记：

在网络安全和信息化工作座谈会上就网络和信息安全发表讲话。

提出没有信息安全就没有国家安全。网络安全是整体的而不是割裂的，是动态的而不是静态的，是开放的而不是封闭的，是相对的而不是绝对的。

2016年中国大数据技术大会

大数据安全，也被作为一个重要的分论坛来进行研讨。

2015.8.19 → 2016.4.19 → 2016.4.28 → 2016.12.10

李克强总理：

主持国务院常务会议，通过了《关于促进大数据发展的行动纲要》。

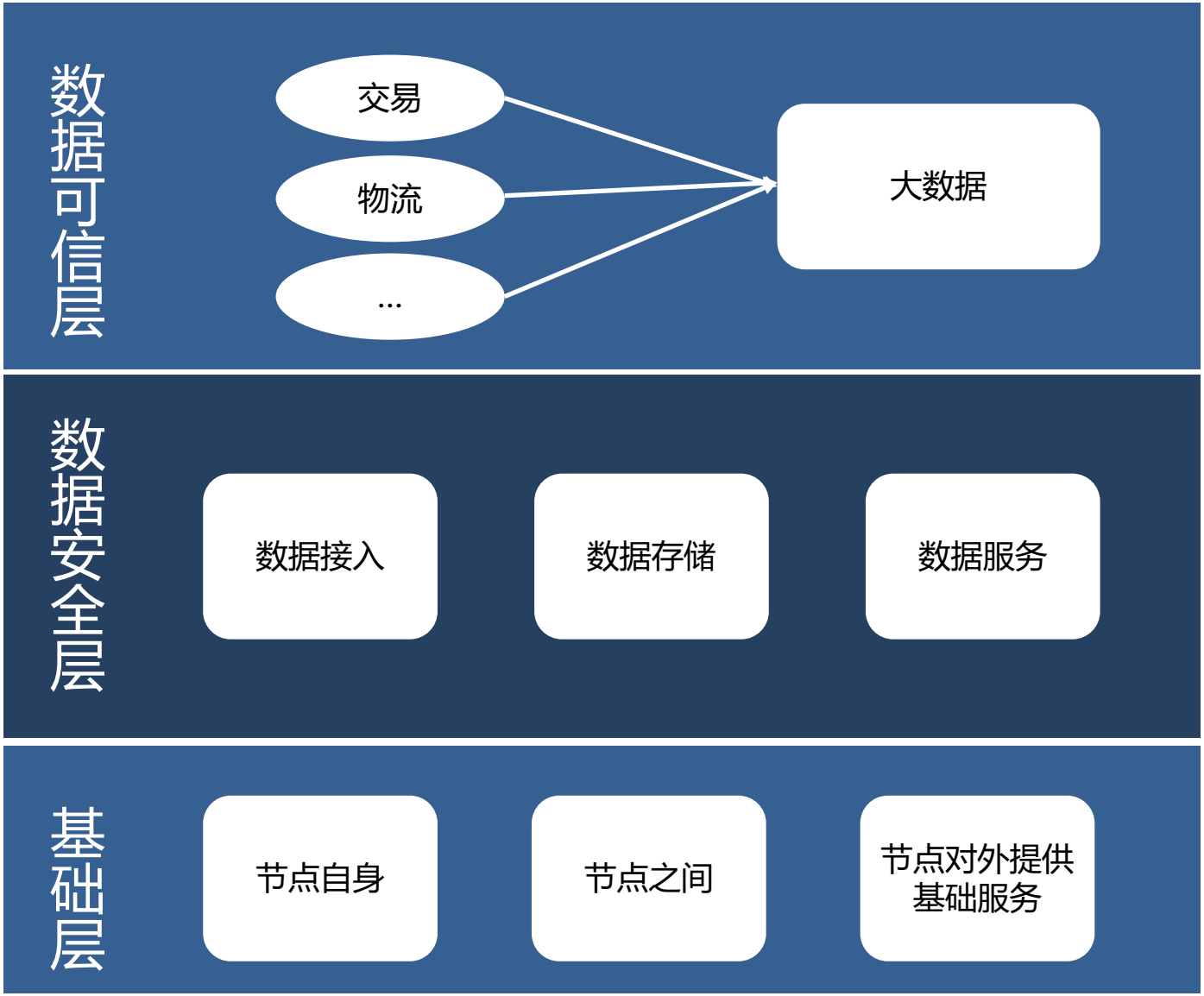
提出共享、开放、安全三个关键词，其中安全是基石，为大数据产业创造一个健康发展的环境。

网信办张望副局长：

在2016年大数据产业峰会发表主题演讲。

提出数据引领创新，数据驱动发展。数据快速发展所带来的网络安全和数据安全的风险，更是不容忽视，要进一步强化大数据安全的保障能力。

什么是大数据安全


数据可信层
交易
物流
...
大数据
数据安全层
数据接入
数据存储
数据服务
基础层
节点自身
节点之间
节点对外提供基础服务

大数据安全面临的挑战

Volume

数据之巨大，安全管理成本上升
容易被贼惦记

Variety

数据类型之多，安全防护难度增加
给安全带来功能完善的压力

Value

单位数据价值低，安全效能降低
给安全带来新的安全准则

Velocity

处理速度快，对安全手段效率要求高
给安全带来巨大的性能压力

数据的使用者同时也是数据的创造者和供给者，数据间的联系是可持续扩展的，安全边界更加模糊。
开放性会给安全带来新的环境适应性压力

知己–大数据架构模型

业务应用

统计
分析

BI/报表
Ad hoc分析

交互式
Web和手机应用

企业应用

大数据中心

HBase

SPARK | Hive | Pig

YARN

HDFS

数据资源

交易数据

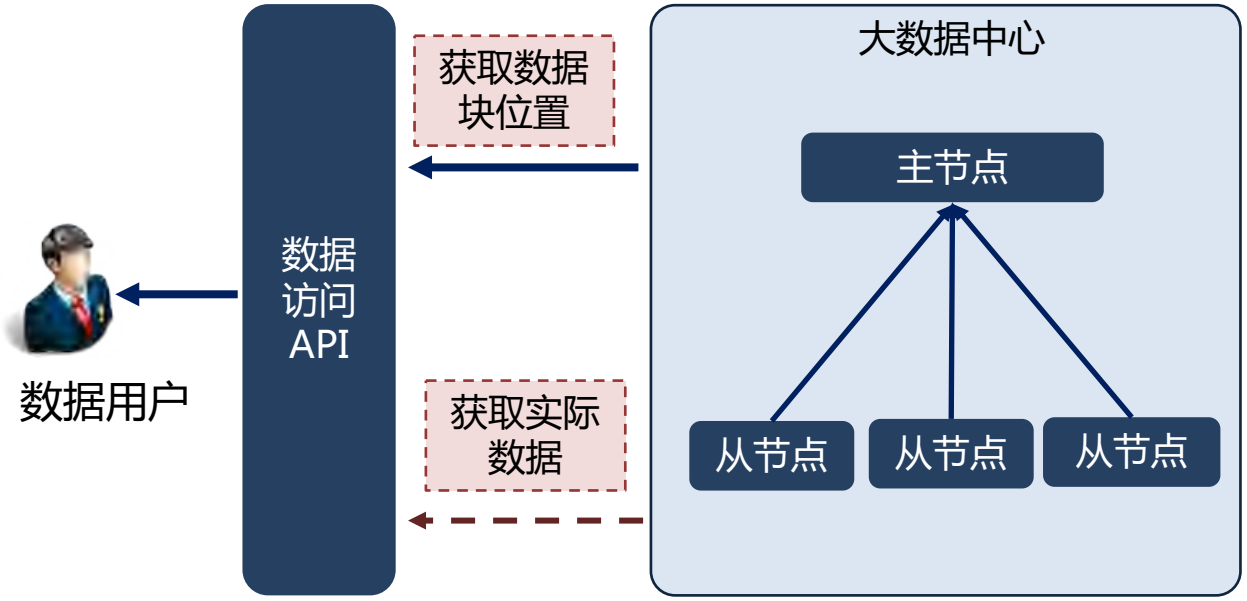
文档
邮件

应用、系统
和设备日志

社会管
控数据

地理信
息数据

大数据生态大多数存储系统采用主从结构



- **主节点：**

1. 存储元数据
2. 通过从节点心跳被动获取数据信息
3. 操作系统级粗粒度、弱权限验证

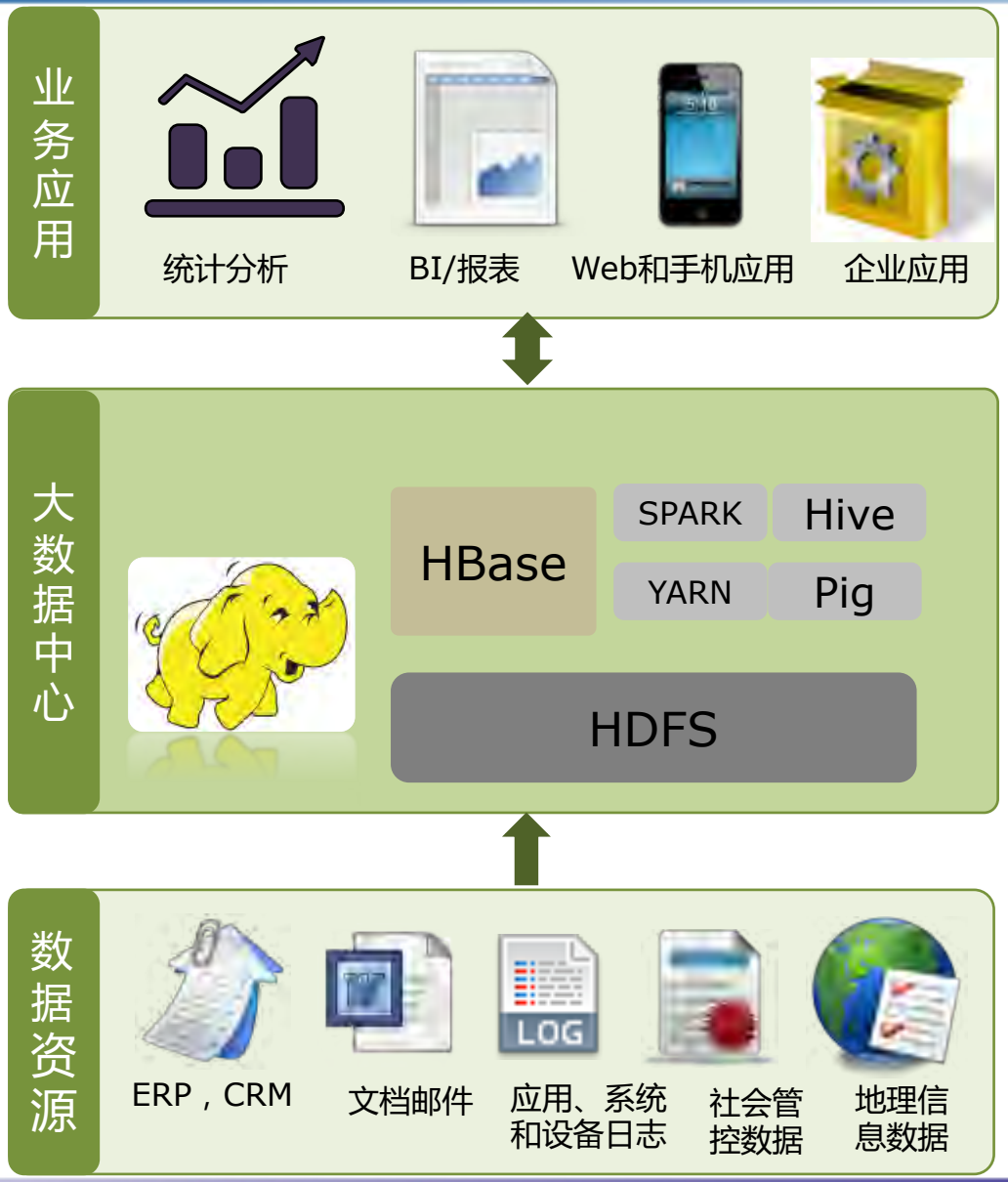
- **从节点：**

1. 存储实际数据
2. 进行实际运算
3. 进行任务调度

- **REST API：**

1. 数据访问接口反映集群拓扑结构
2. 数据明文传输

知彼–脆弱性分析



数据服务环节

- 攻击服务器
- 数据泄漏
- 任意访问数据
- 冒充他人身份
- 窃听通信链路
- 操作抵赖

存储计算环节

- 直接拷贝数据
- 数据篡改
- 加入非法节点

数据接入环节

- 冒充他人身份
- 窃听通信链路
- 操作抵赖

风险	
冒充他人身份	数据泄漏和破坏·、操作行为无法追踪到实际操作者
任意访问数据	增加数据误用风险
篡改数据	影响业务使用
泄露获取数据	敏感数据被恶意用户获取
加入额外节点	冒充合法节点接受任务或者数据等
窃听通信链路	获取会话内容
直接拷贝数据	敏感数据泄漏
攻击服务器	集群瘫痪，无法为合法用户提供服务
操作抵赖	无法定位恶意用户

存储计算环节

Kerberos

Kerberos，采用客户端/服务器结构与DES对称加密技术，实现了对大数据中心的节点认证

HDFS透明加密

HDFS 提供Encryption zone透明加密，是一种端到端的加密模式，其中加/解密过程对于客户端是完全透明

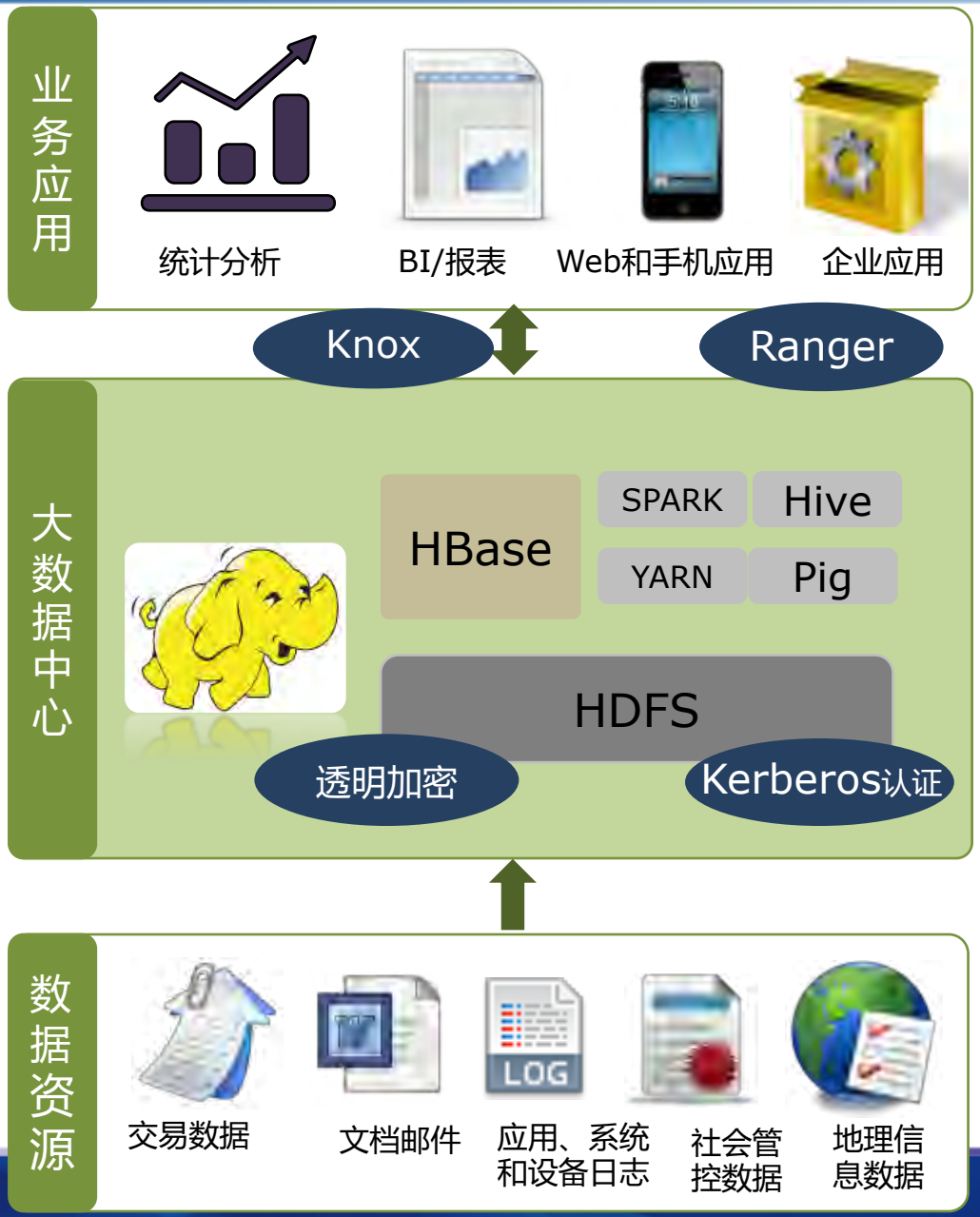
数据服务环节

Ranger

集中式安全管理框架，并解决授权和审计。可对Hadoop生态的组件进行细粒度的数据访问控制

Knox

通过集中式的 REST APIs 访问服务。所有与集群交互的REST API都通过Knox处理，提供基于边缘的安全防护



如何百战不殆

这些工具：

1. 只管理了存储和访问，安全防护全面性不足

2. 点性防护，没有形成整体联动的安全防护体系

可信

用户可信
应用可信
节点可信

可管

访问可管
存储可管
权限可管
使用可管

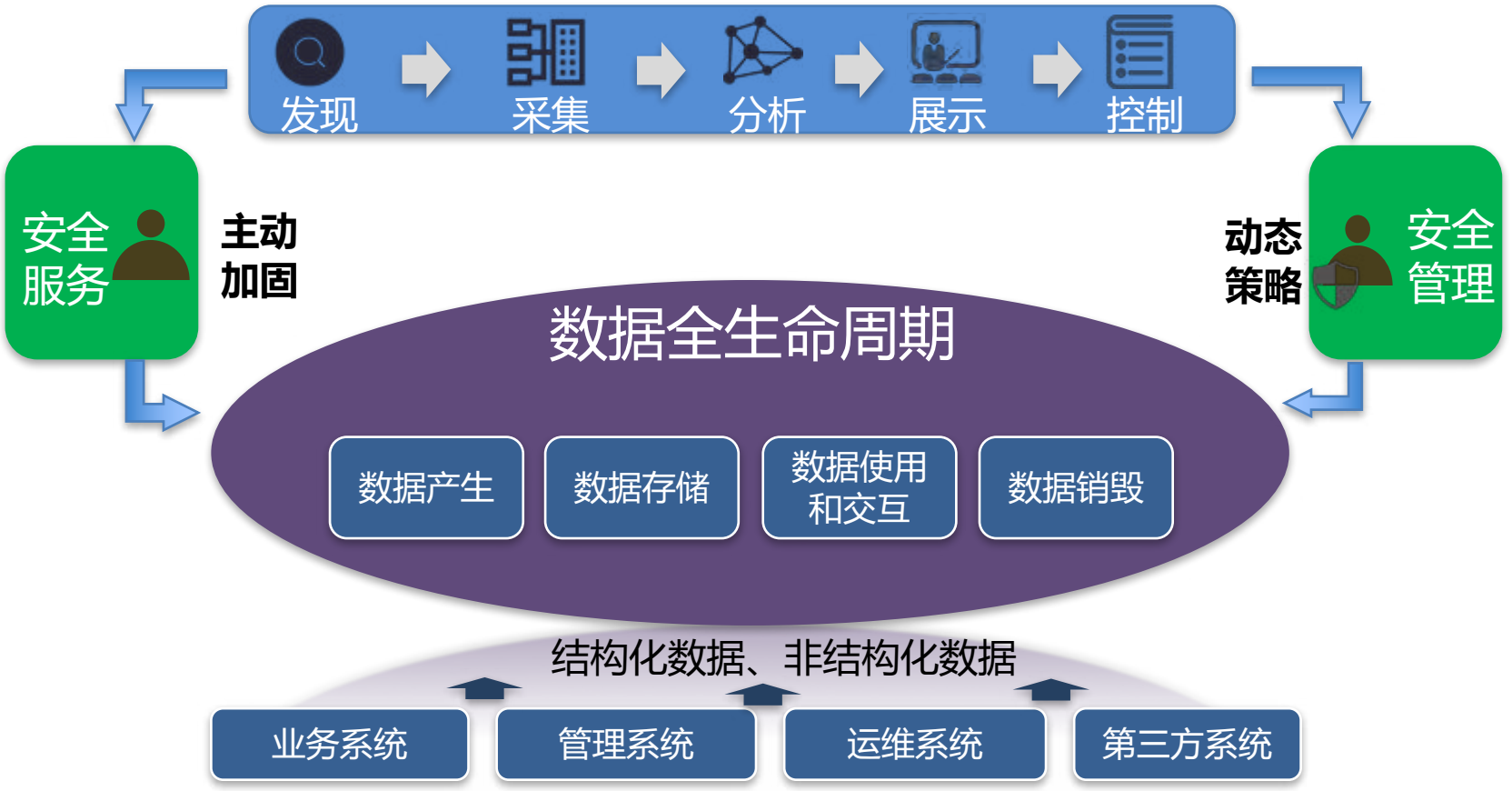
可控

状态可知
安全可视
动态可控

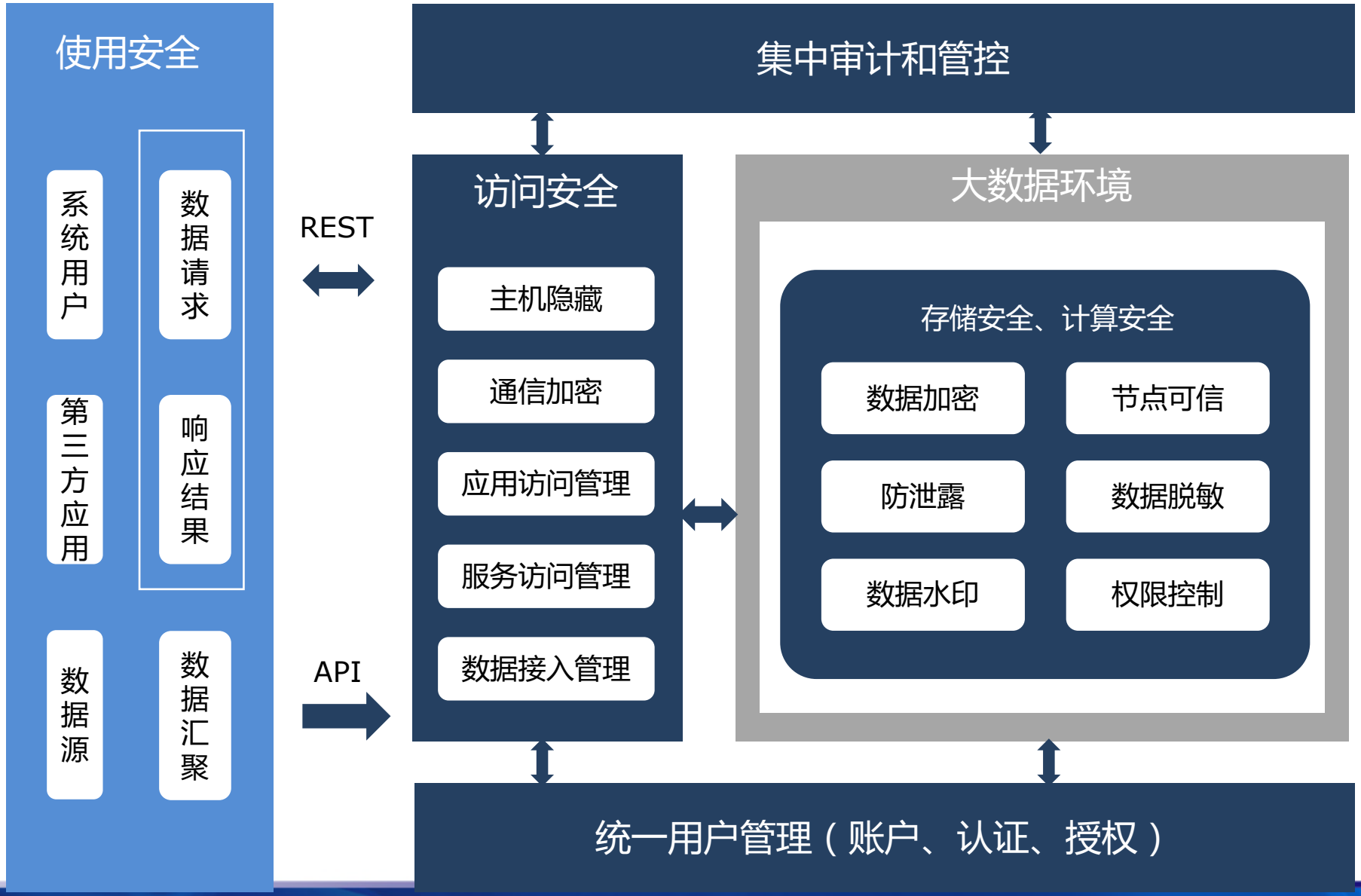
问题

如何构建一个完整的防护体系，从体系角度试图实现上述三大目标

解决传统数据安全手段、机制在大数据环境下的适应性和扩展性问题



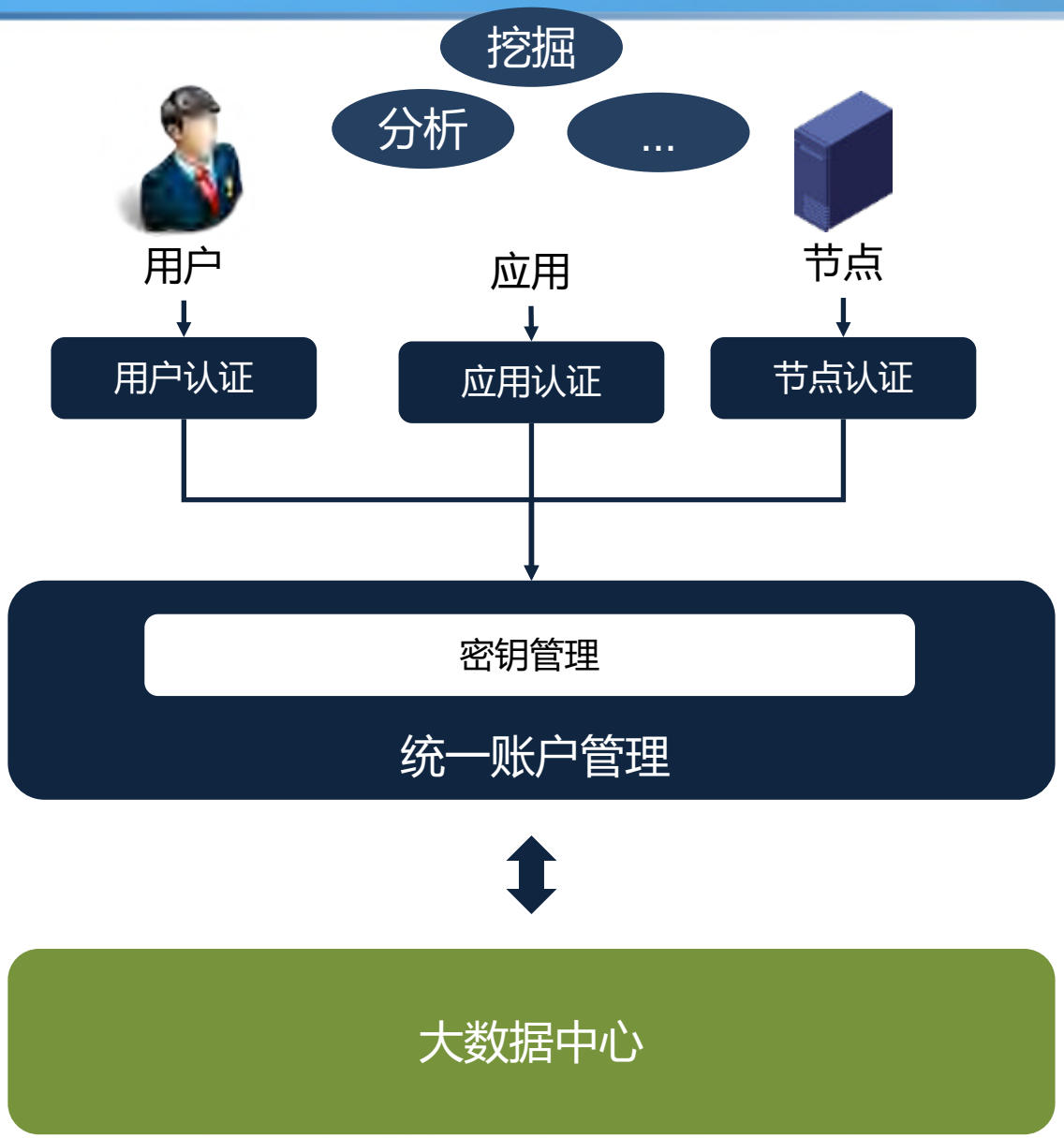
建立可信、可管、可控的数据全生命周期安全防护体系


使用安全
系统用户
第三方应用
数据源
数据请求
响应结果
数据汇聚
REST
API
集中审计和管控
访问安全
主机隐藏
通信加密
应用访问管理
服务访问管理
数据接入管理
大数据环境
存储安全、计算安全
数据加密
节点可信
防泄露
数据脱敏
数据水印
权限控制
统一用户管理（账户、认证、授权）

方法


数据
生命周期
产生
存储和销毁
使用和交互
可信
可管
可控
集中审计和管控
数据接入
管理
数据水印
数据加密
节点可信
防泄露
数据脱敏
权限控制
统一用户管理
（账户、认证、
授权）
主机隐藏
通信加密
应用访问管理
服务访问管理

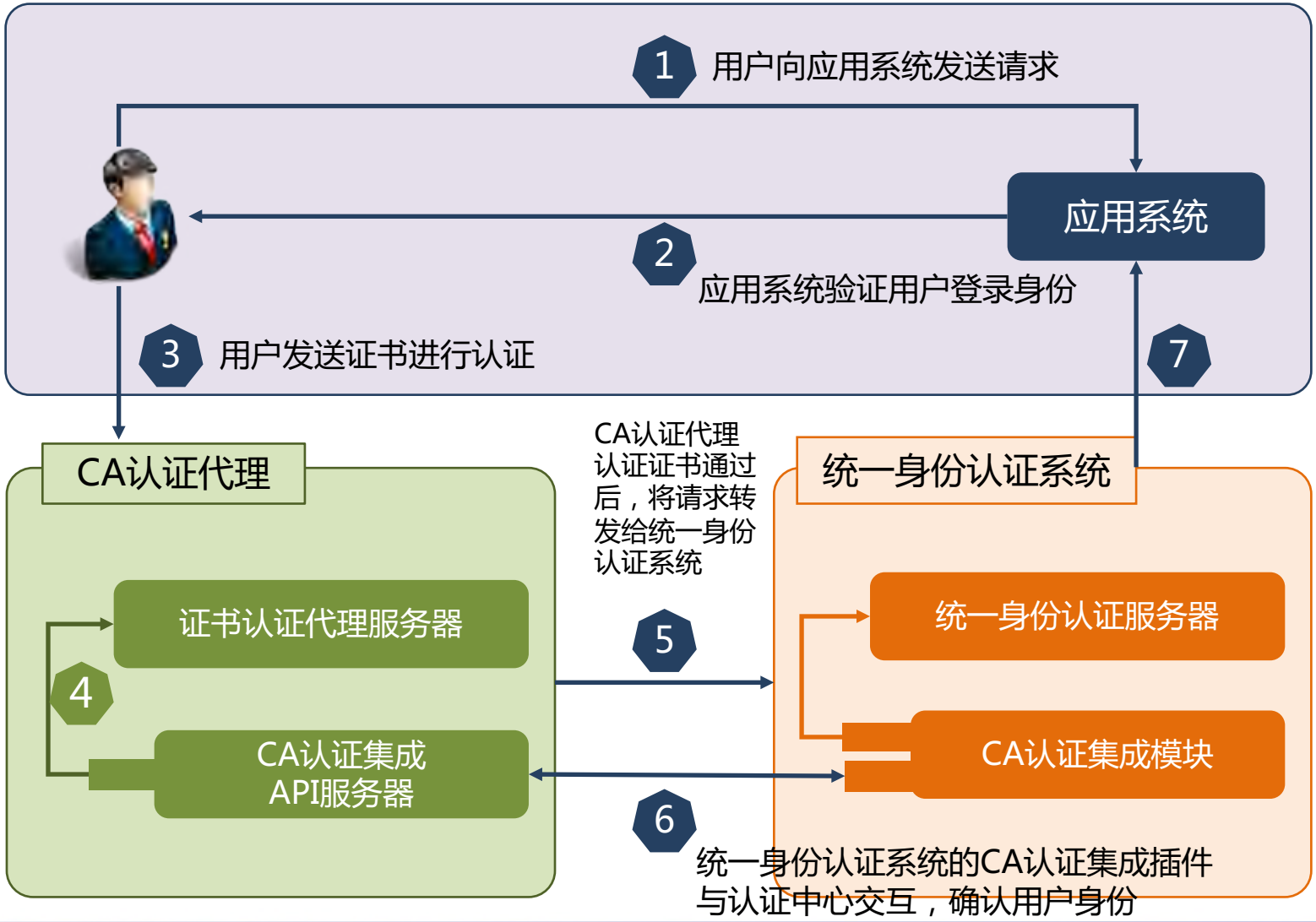
可信概述



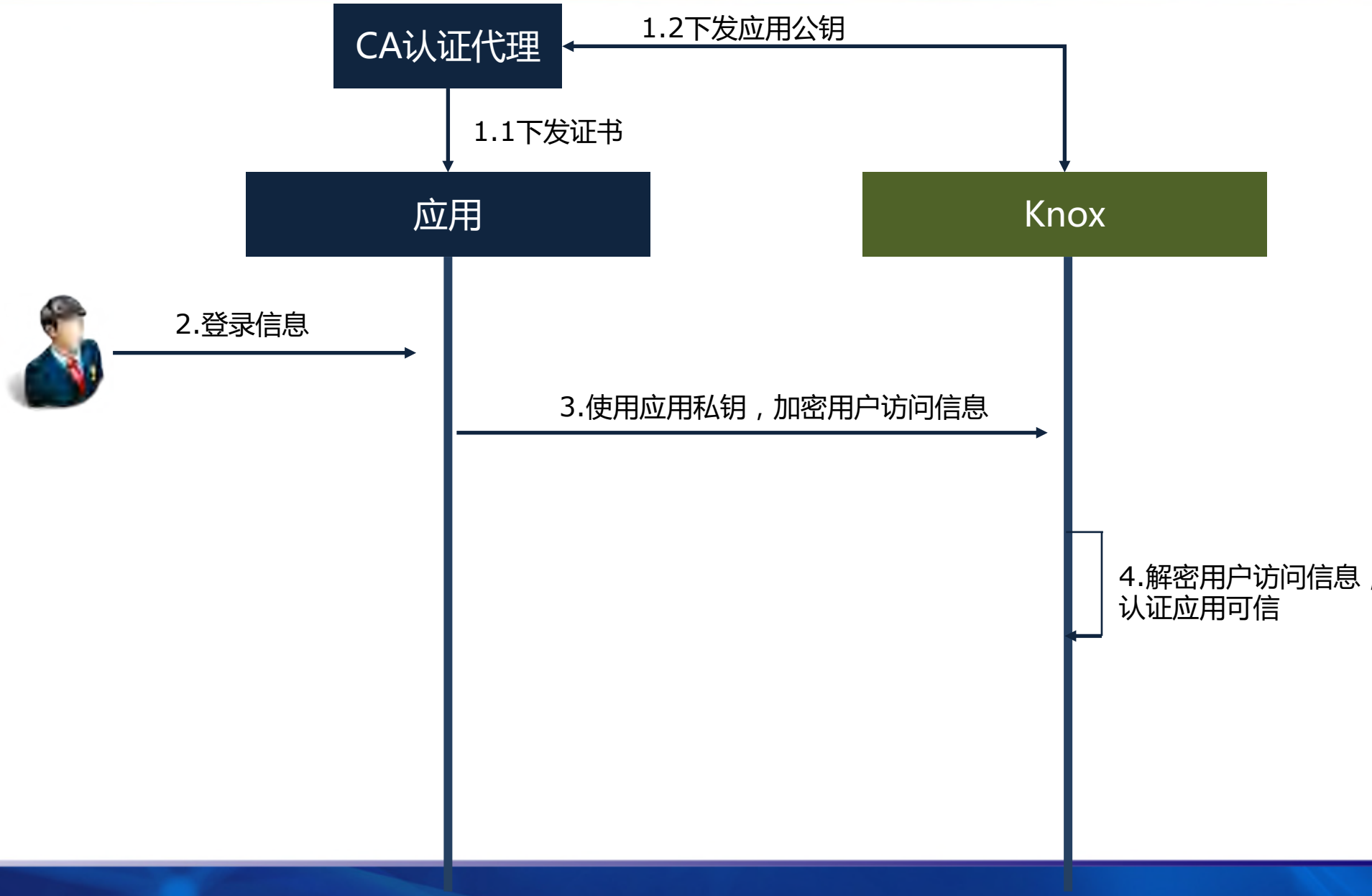
- **用户可信：**
采用CA代理认证与统一身份认证系统，确保用户可信。

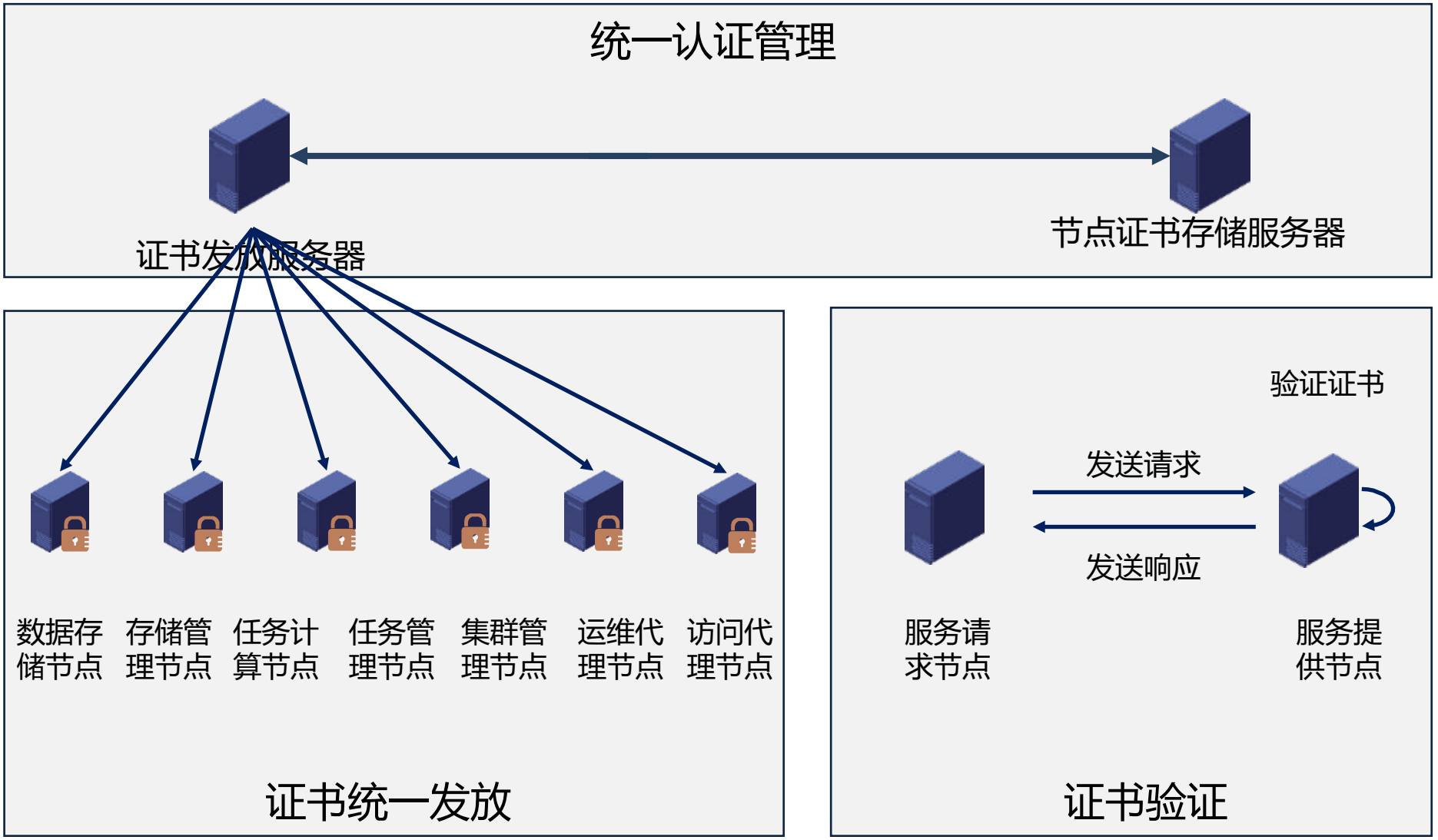
- **应用可信：**
采用密钥认证方式，确保应用可信。

- **节点可信：**
采用向节点发放证书，确保节点可信，防止恶意用户在集群中加入非法节点窃取数据。


1 用户向应用系统发送请求
应用系统
2 应用系统验证用户登录身份
3 用户发送证书进行认证
7
统一身份认证系统的CA认证集成插件与认证中心交互，确认用户身份
CA认证代理
统一身份认证服务器通过调用CA认证插件校验用户身份
CA认证代理认证证书通过后，将请求转发给统一身份认证系统
统一身份认证系统
证书认证代理服务器
4
CA认证集成API服务器
5
统一身份认证服务器
CA认证集成模块
6
统一身份认证系统的CA认证集成插件与认证中心交互，确认用户身份

可信–应用可信


CA认证代理
1.2下发应用公钥
1.1下发证书
应用
Knox
2.登录信息
3.使用应用私钥，加密用户访问信息
4.解密用户访问信息，
认证应用可信



通过节点可信可以防止恶意用户在集群中加入非法节点窃取数据

- **访问行为：**
通过主机隐藏等，确保大数据访问安全

- **数据使用：**
设置用户细粒度数据权限，防止用户越级使用数据

- **数据加密：**
将数据加密存储，确保存储安全

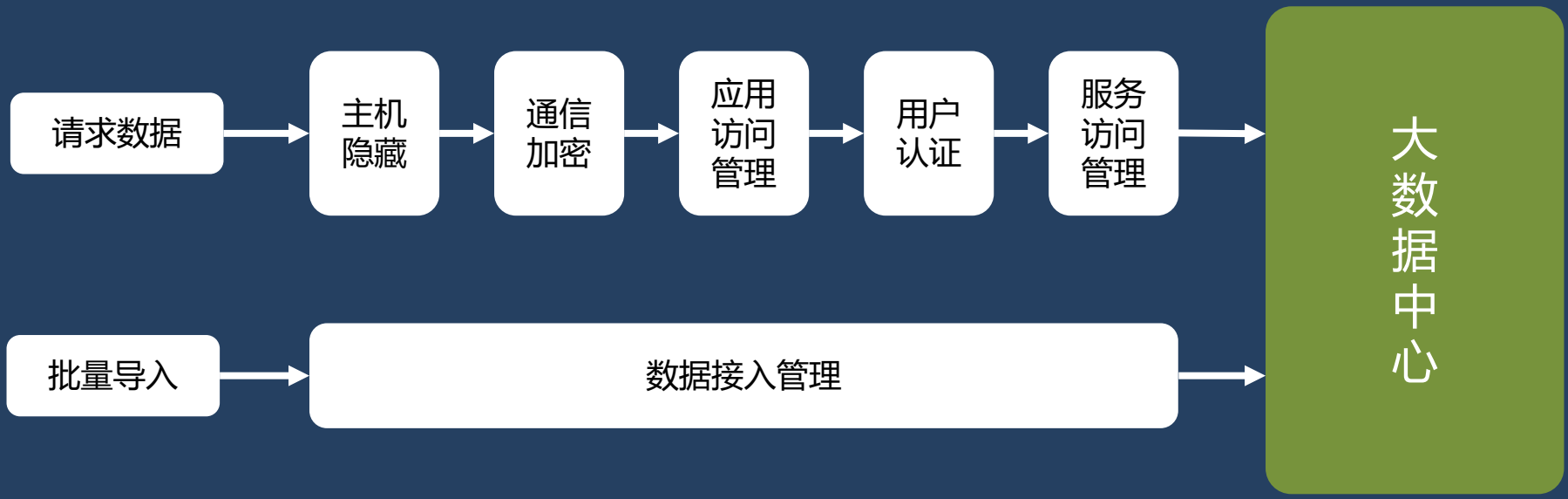
- **存储行为：**
通过权限、加密、水印等措施，确保存储安全

- **数据使用：**
通过水印、脱敏等，确保使用安全

访问安全

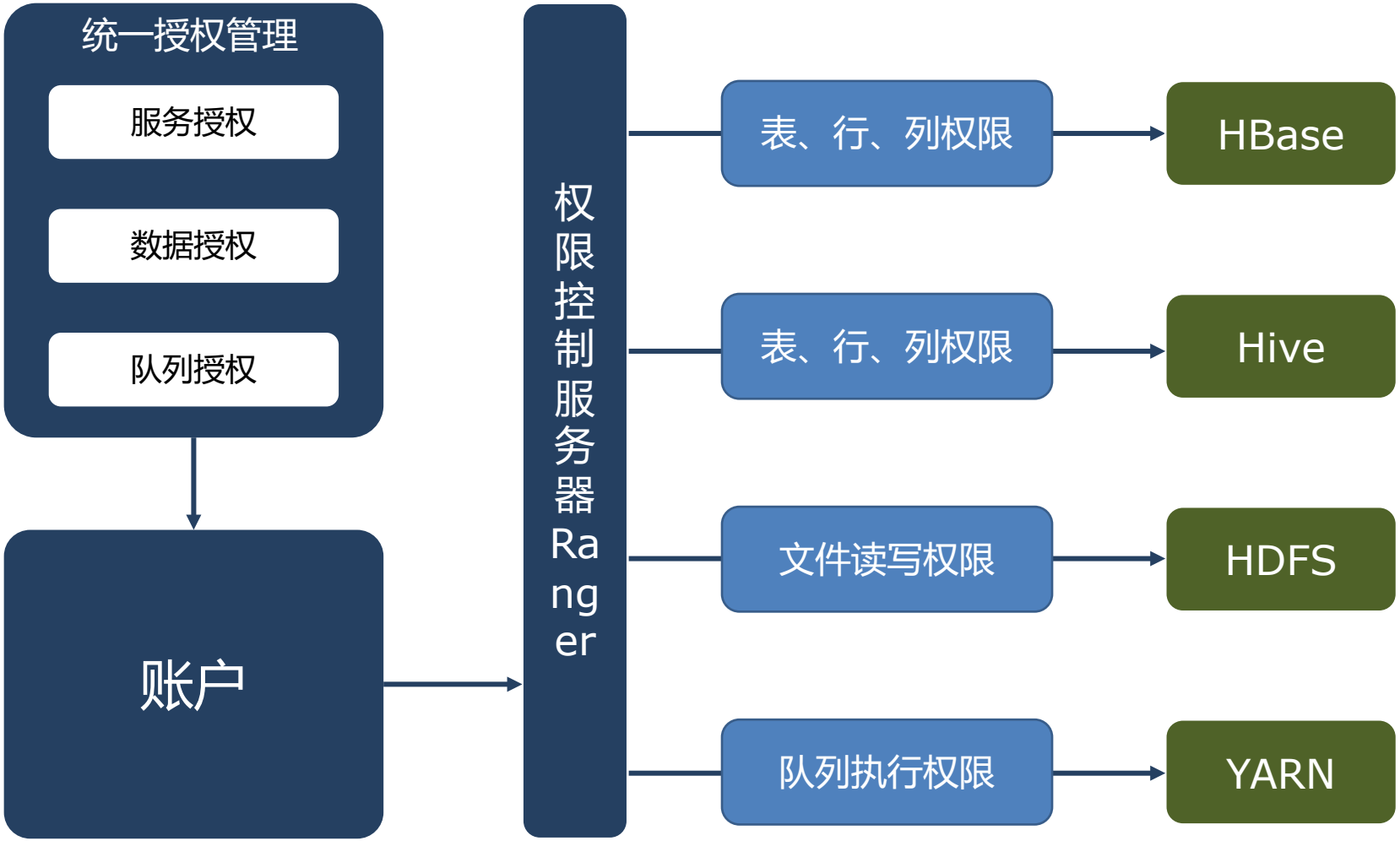
主机隐藏 | 通信加密 | 应用访问管理 | 服务访问管理 | 数据接入管理

确保大数据的访问安全

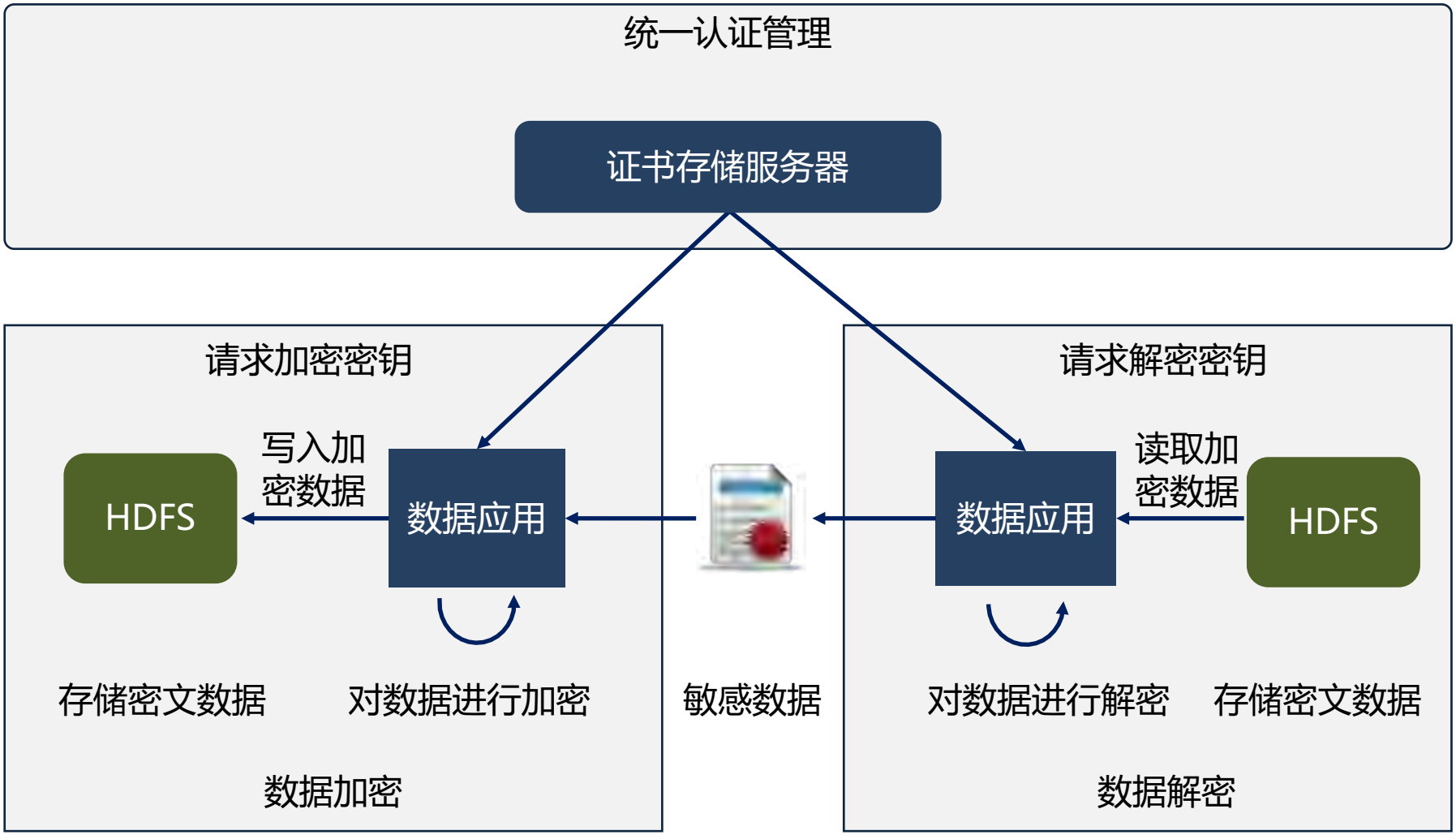


存储计算安全

数据水印 | 数据脱敏 | 数据防泄密 | 节点可信 | 权限控制 | 数据加解密

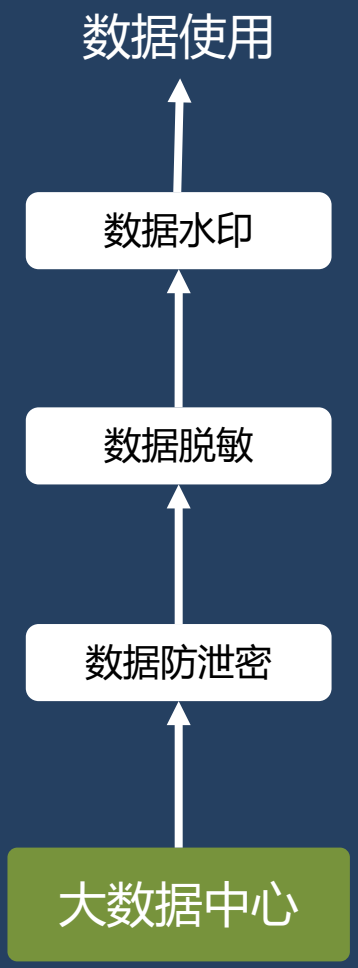


设置用户细粒度数据权限，防止用户越级使用数据，以及其它对数据的任意访问行为



通过端到端加解密的方式，使敏感数据在存储和传输过程中都以密文形式存在，从而防止直接拷贝文件和网络窃听方式的数据泄漏

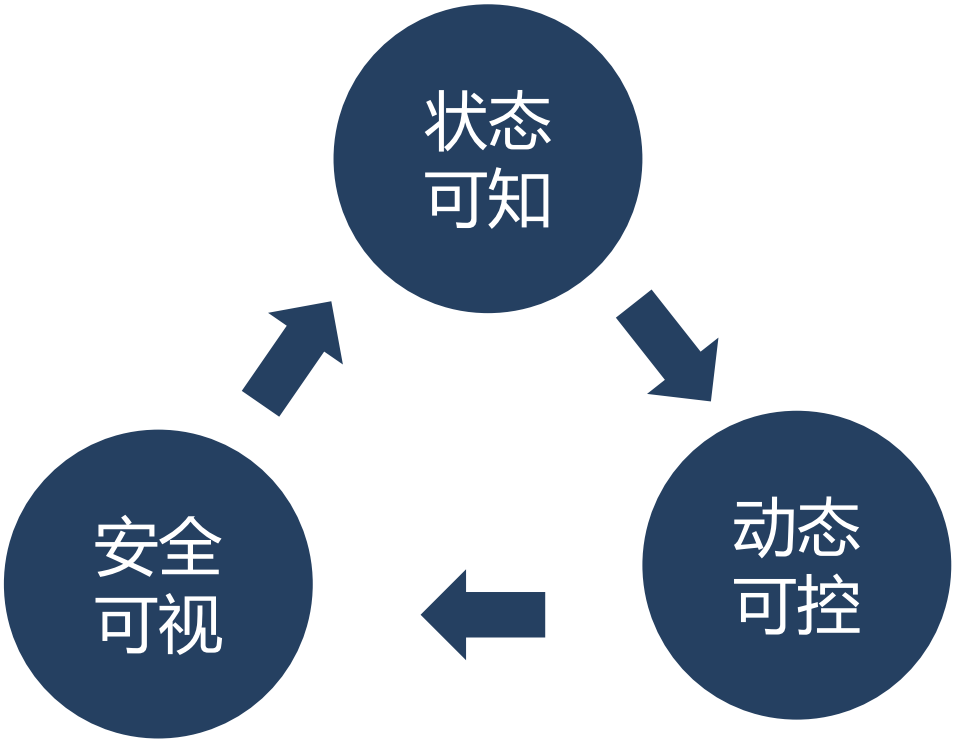
可管–数据使用



通过给数据加上水印能够及时发现数据篡改行为，并且能够追溯数据的流转过程

通过替换、重排、加密、截断、掩码、日期偏移取整等方式，防止敏感数据暴露

部署在网络出口处，对通过网络传输的数据进行内容智能分析和识别，并根据策略对违规内容进行审计、告警及阻断

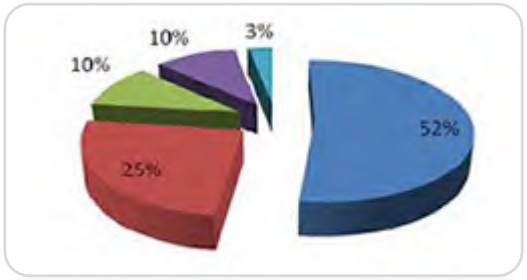


- 可知：
全面采集大数据环境运行过程的各类信息，采取各种分析方式，发现识别各种安全风险、事件

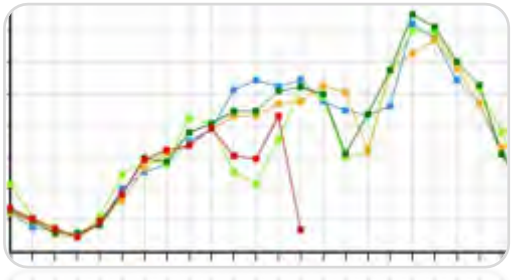
- 可视：
通过图表等形象的展示方式，展示大数据环境的运行状态及安全风险、事件

- 可控：
针对发现的安全风险和事件，通过各类型配置与各个安全手段进行联动，实现策略的动态下发和及时的安全管控

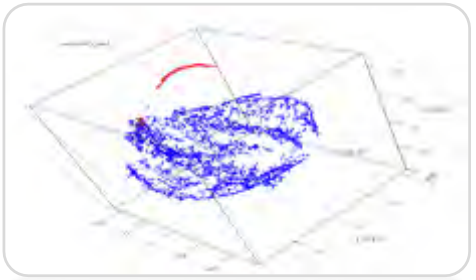

上报
上报
分析层
实时比对
大数据分析
溯源分析
采集层
状态数据
交易数据
环境数据
产生
存储
使用和交互
销毁
数据生命周期



安全事件统计



安全趋势预测

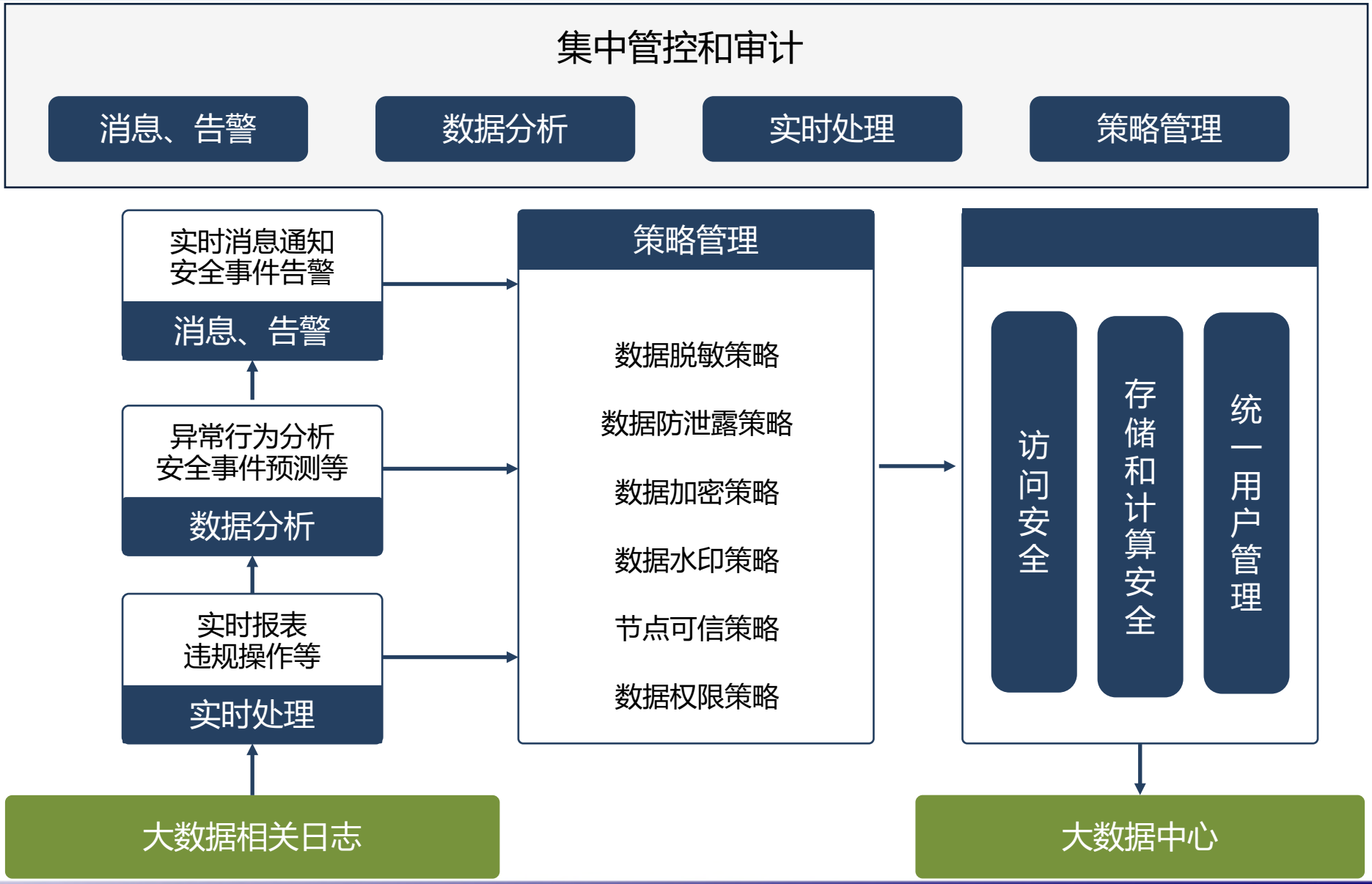


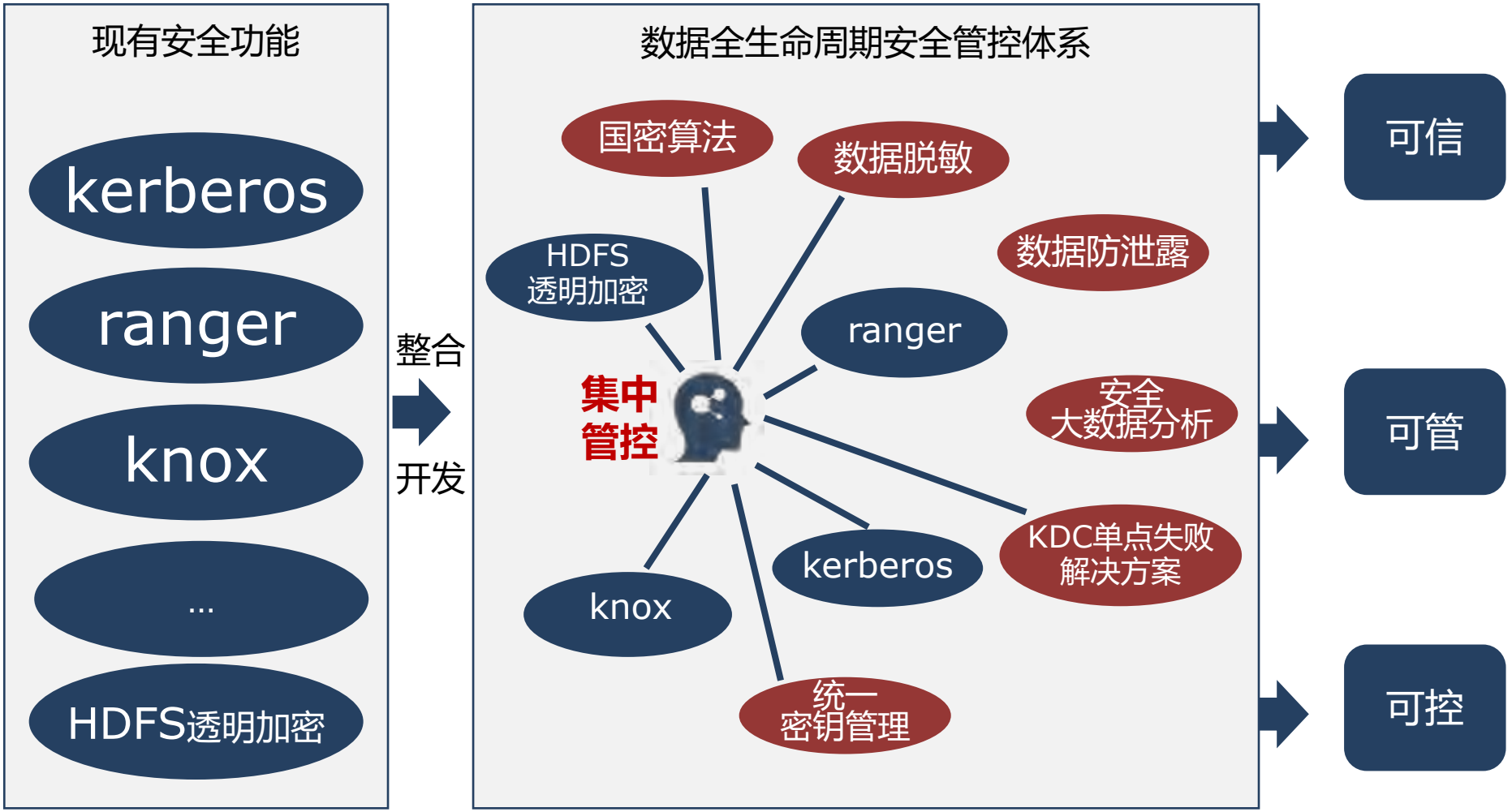
异常行为分析

数据分布及使用情况

安全事件预警

可控–动态管控


集中管控和审计
消息、告警
数据分析
实时处理
策略管理
实时消息通知
安全事件告警
消息、告警
异常行为分析
安全事件预测等
数据分析
实时报表
违规操作等
实时处理
大数据相关日志
策略管理
数据脱敏策略
数据防泄露策略
数据加密策略
数据水印策略
节点可信策略
数据权限策略
访问安全
存储和计算安全
统一用户管理
大数据中心


现有安全功能
kerberos
ranger
knox
...
HDFS透明加密
整合
开发
数据全生命周期安全管控体系
国密算法
数据脱敏
HDFS
透明加密
数据防泄露
ranger
集中
管控
安全
大数据分析
KDC单点失败
解决方案
kerberos
knox
统一
密钥管理
可信
可管
可控

Thanks !

明朝万达
wondersoft.cn

BDTC 2016中国大数据技术大会
Big Data Technology Conference 2016