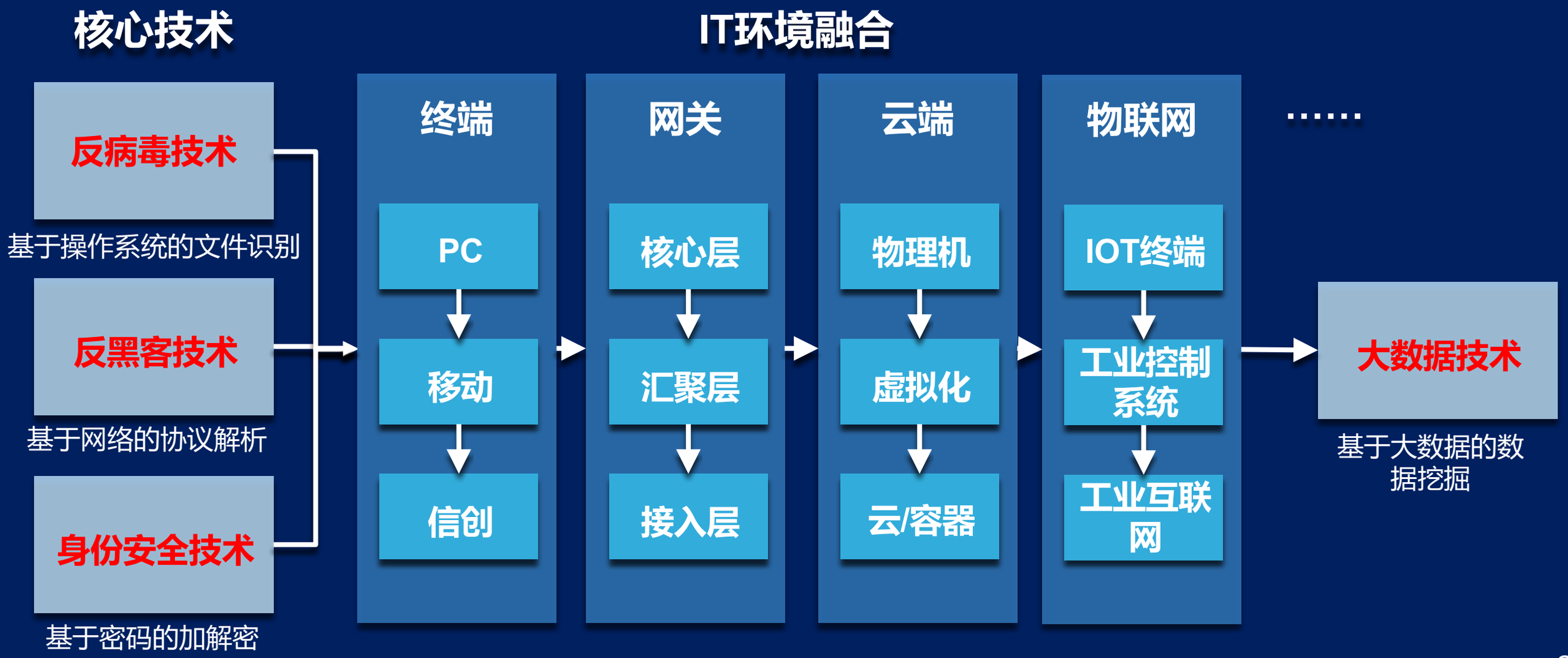
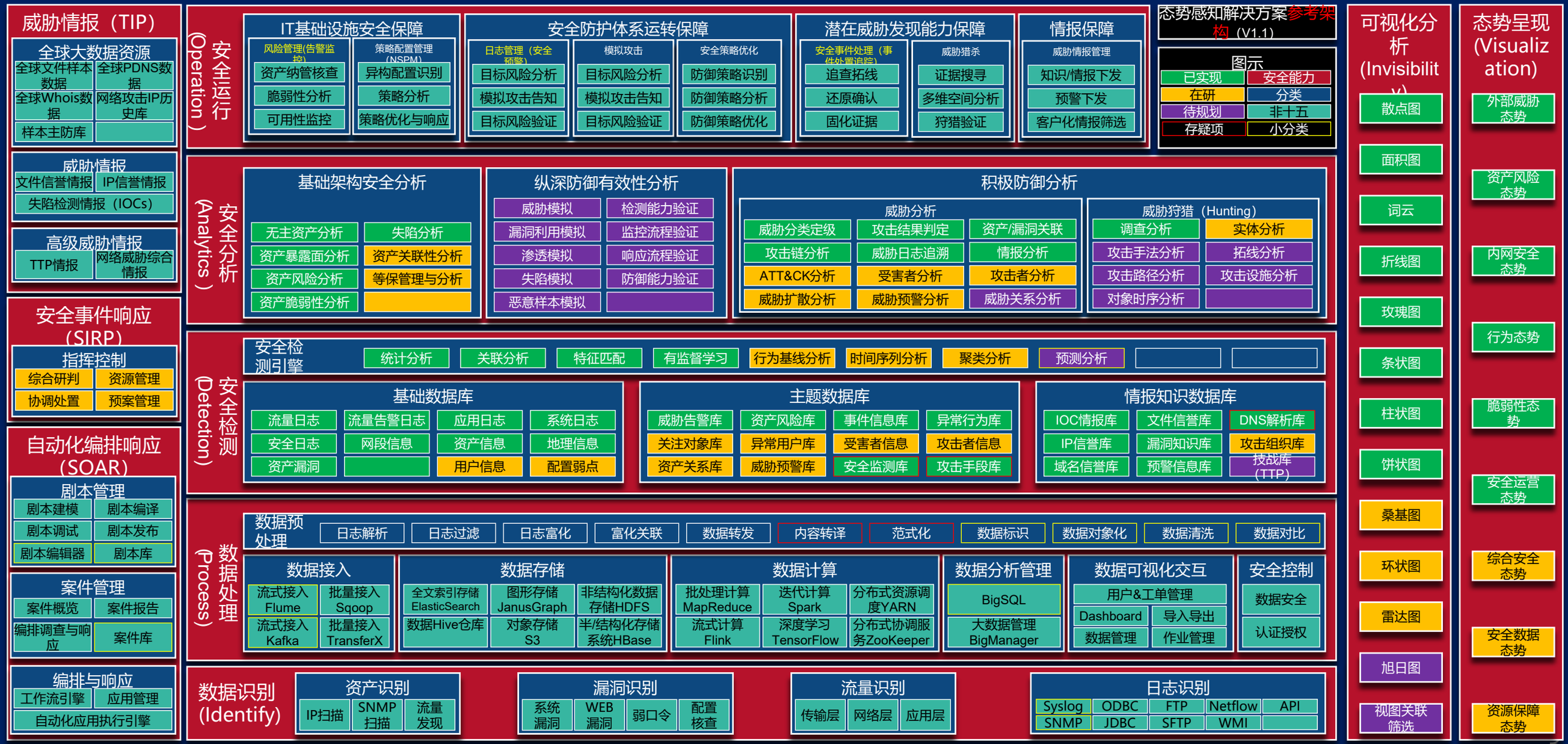
安全数据中台

数据中台的过去、现在和未来

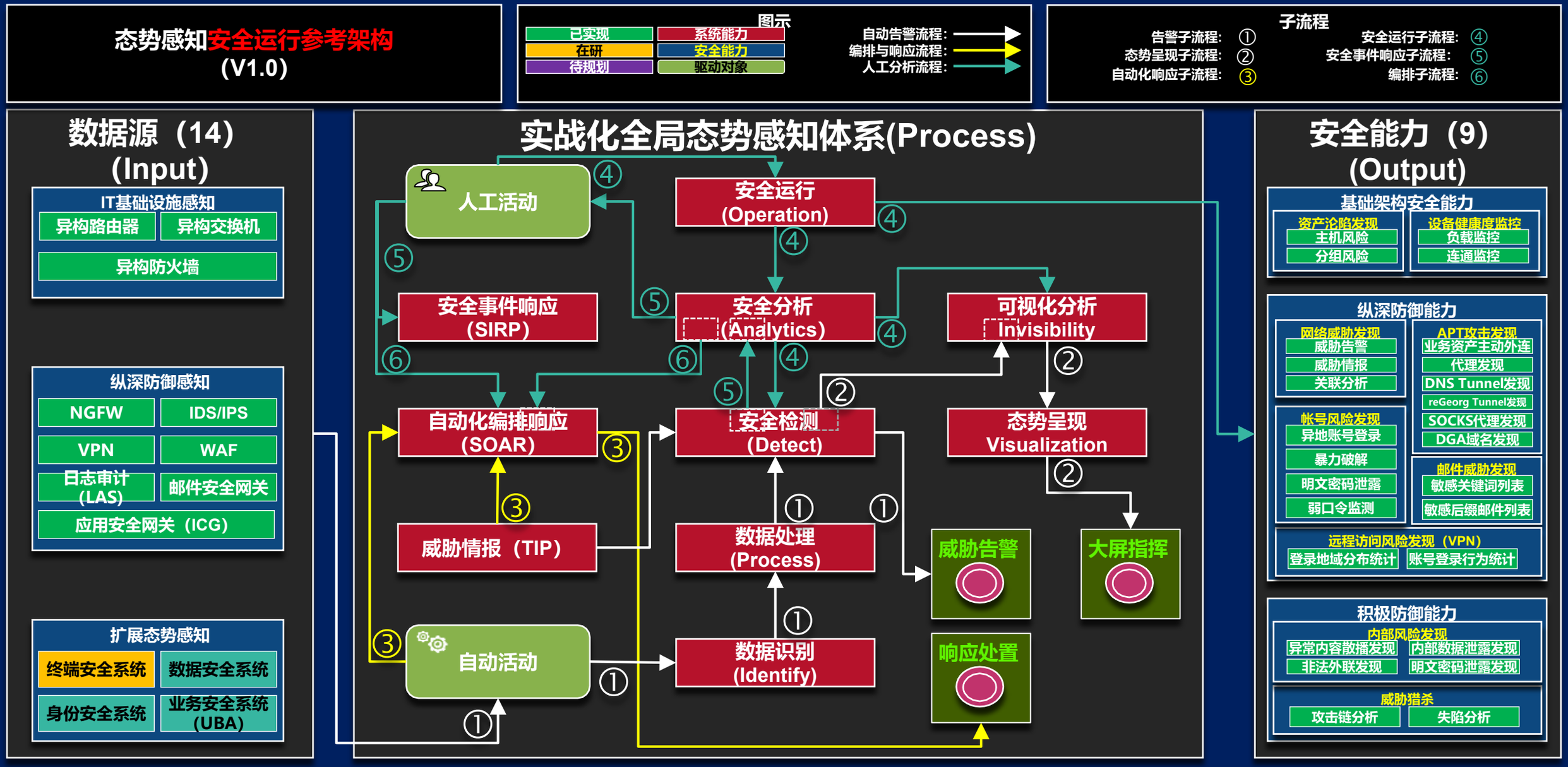
大数据中台的技术本质


核心技术
IT环境融合
反病毒技术
基于操作系统的文件识别
反黑客技术
基于网络的协议解析
身份安全技术
基于密码的加解密
终端
PC
移动
信创
网关
核心层
汇聚层
接入层
云端
物理机
虚拟化
云/容器
物联网
IOT终端
工业控制系统
工业互联网
……
大数据技术
基于大数据的数据挖掘

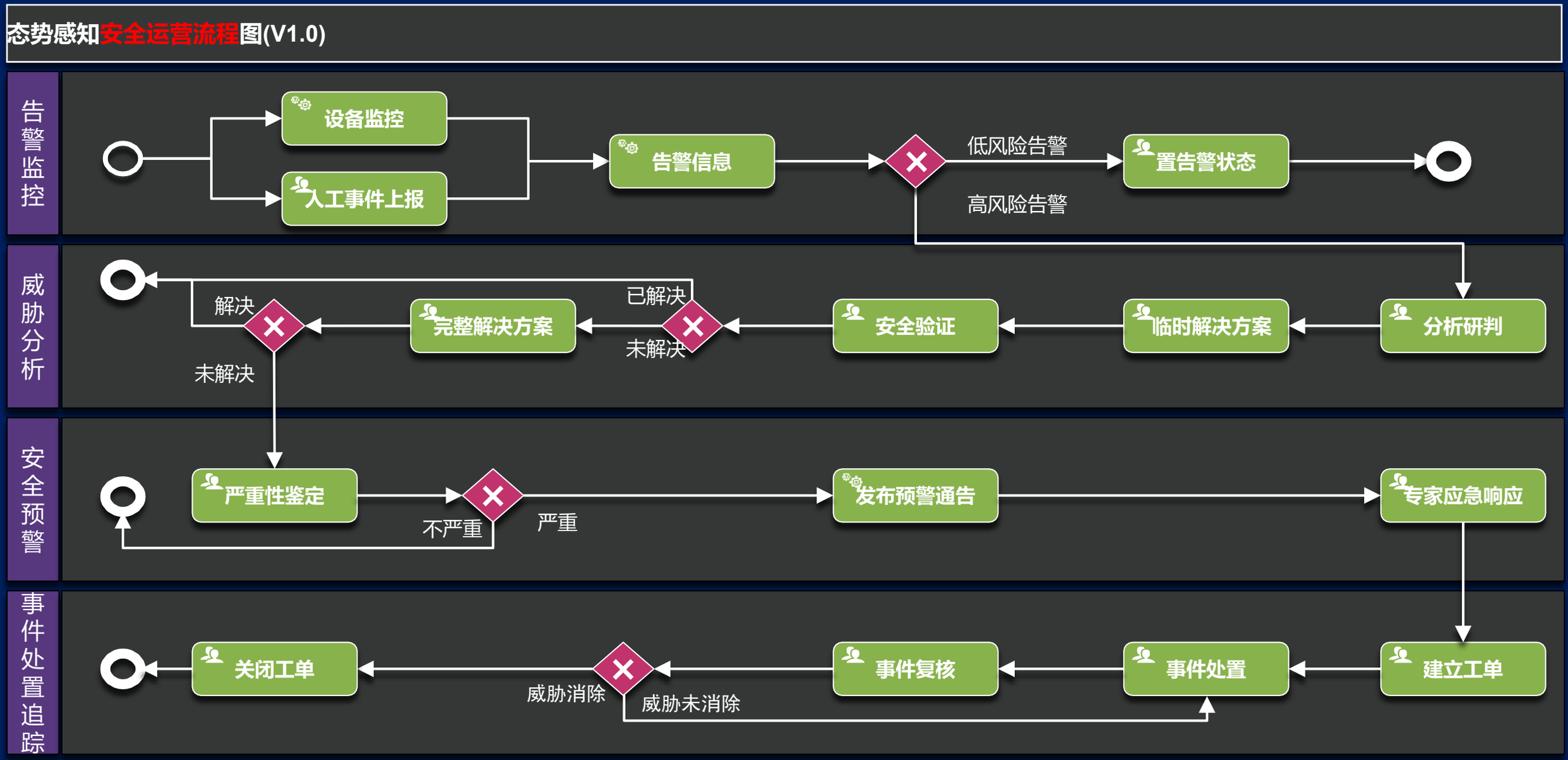
大数据安全的终局


威胁情报（TIP）
全球大数据资源
全球文件样本数据
全球PDNS数据
全球Whois数据
网络攻击IP历史库
样本主防库
威胁情报
文件信誉情报
IP信誉情报
失陷检测情报（IOCs）
高级威胁情报
TTP情报
网络威胁综合情报
安全事件响应（SIRP）
指挥控制
综合研判
资源管理
协调处置
预案管理
自动化编排响应（SOAR）
剧本管理
剧本建模
剧本编译
剧本调试
剧本发布
剧本编辑器
剧本库
案件管理
案件概览
案件报告
编排调查与响应
案件库
编排与响应
工作流引擎
应用管理
自动化应用执行引擎
安全运行（Operation）
IT基础设施安全保障
风险管理(告警监控)
资产纳管核查
脆弱性分析
可用性监控
策略配置管理（NSPM）
异构配置识别
策略分析
策略优化与响应
安全防护体系运转保障
日志管理（安全预警）
目标风险分析
模拟攻击告知
目标风险验证
模拟攻击
目标风险分析
模拟攻击告知
目标风险验证
安全策略优化
防御策略识别
防御策略分析
防御策略优化
潜在威胁发现能力保障
安全事件处理（事件处置追踪）
追查拓线
还原确认
固化证据
威胁猎杀
证据搜寻
多维空间分析
狩猎验证
情报保障
威胁情报管理
知识/情报下发
预警下发
客户化情报筛选
态势感知解决方案参考架构（V1.1）
图示
已实现
安全能力
在研
分类
待规划
非十五
存疑项
小分类
安全分析（Analytics）
基础架构安全分析
无主资产分析
失陷分析
资产暴露面分析
资产关联性分析
资产风险分析
等保管理与分析
资产脆弱性分析
纵深防御有效性分析
威胁模拟
检测能力验证
漏洞利用模拟
监控流程验证
渗透模拟
响应流程验证
失陷模拟
防御能力验证
恶意样本模拟
积极防御分析
威胁分析
威胁分类定级
攻击结果判定
资产/漏洞关联
攻击链分析
威胁日志追溯
情报分析
ATT&CK分析
受害者分析
攻击者分析
威胁扩散分析
威胁预警分析
威胁关系分析
威胁狩猎（Hunting）
调查分析
实体分析
攻击手法分析
拓线分析
攻击路径分析
攻击设施分析
对象时序分析
安全检测（Detection）
安全检测引擎
统计分析
关联分析
特征匹配
有监督学习
行为基线分析
时间序列分析
聚类分析
预测分析
基础数据库
流量日志
流量告警日志
应用日志
系统日志
安全日志
网段信息
资产信息
地理信息
资产漏洞
用户信息
配置弱点
主题数据库
威胁告警库
资产风险库
事件信息库
异常行为库
关注对象库
异常用户库
受害者信息
攻击者信息
资产关系库
威胁预警库
安全监测库
攻击手段库
情报知识数据库
IOC情报库
文件信誉库
DNS解析库
IP信誉库
漏洞知识库
攻击组织库
域名信誉库
预警信息库
战术库（TTP）
数据处理（Process）
数据预处理
日志解析
日志过滤
日志富化
富化关联
数据转发
内容转译
范式化
数据标识
数据对象化
数据清洗
数据对比
数据接入
流式接入 Flume
批量接入 Sqoop
流式接入 Kafka
批量接入 TransferX
数据存储
全文索引存储 ElasticSearch
图形存储 JanusGraph
非结构化数据存储HDFS
数据Hive仓库
对象存储 S3
半/结构化存储系统HBase
数据计算
批处理计算 MapReduce
迭代计算 Spark
分布式资源调度YARN
流式计算 Flink
深度学习 TensorFlow
分布式协调服务ZooKeeper
数据分析管理
BigSQL
大数据管理 BigManager
数据可视化交互
用户&工单管理
Dashboard
导入导出
数据管理
作业管理
安全控制
数据安全
认证授权
数据识别（Identify）
资产识别
IP扫描
SNMP扫描
流量发现
漏洞识别
系统漏洞
WEB漏洞
弱口令
配置核查
流量识别
传输层
网络层
应用层
日志识别
Syslog
ODBC
FTP
Netflow
API
SNMP
JDBC
SFTP
WMI
可视化分析（Invisibility）
散点图
面积图
词云
折线图
玫瑰图
条状图
柱状图
饼状图
桑基图
环状图
雷达图
旭日图
视图关联筛选
态势呈现（Visualization）
外部威胁态势
资产风险态势
内网安全态势
行为态势
脆弱性态势
安全运营态势
综合安全态势
安全数据态势
资源保障态势

安全运行架构

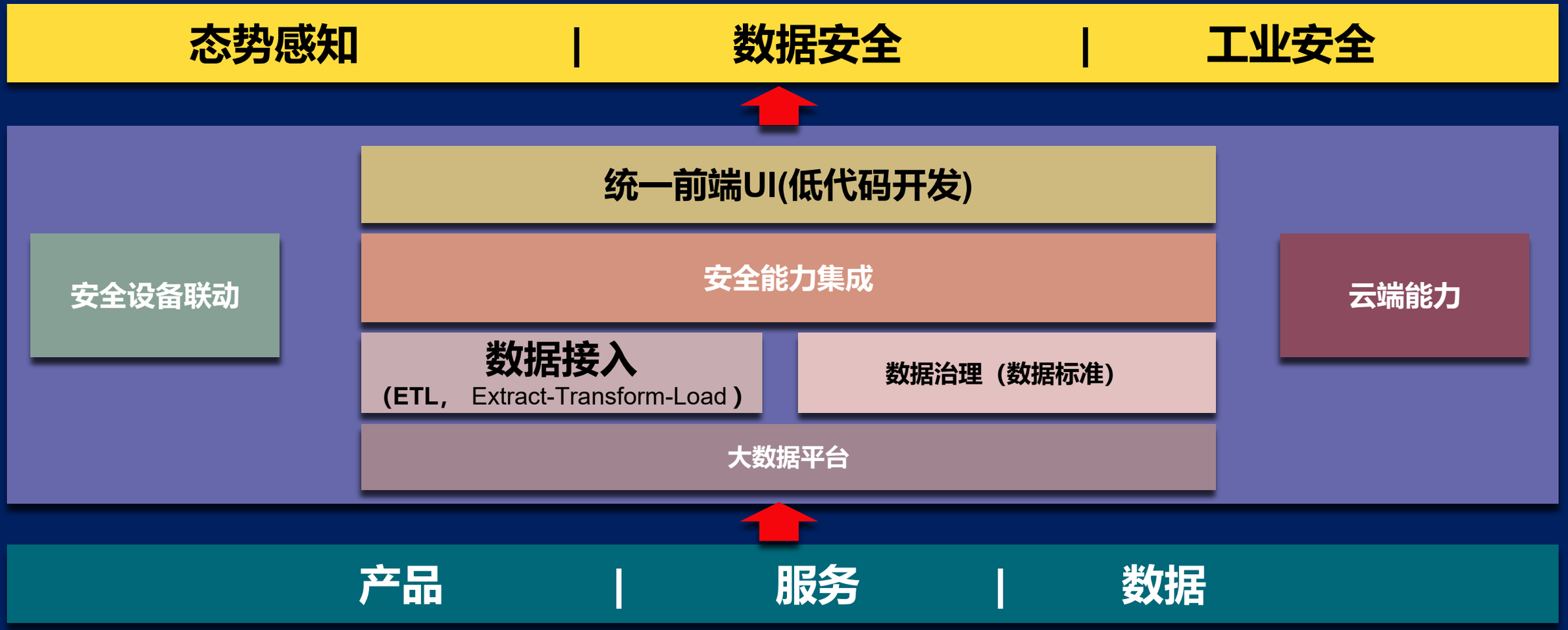

态势感知安全运行参考架构
(V1.0)
图示
已实现
在研
待规划
系统能力
安全能力
驱动对象
自动告警流程:
编排与响应流程:
人工分析流程:
子流程
告警子流程: ①
态势呈现子流程: ②
自动化响应子流程: ③
安全运行子流程: ④
安全事件响应子流程: ⑤
编排子流程: ⑥
数据源（14）
(Input)
IT基础设施感知
异构路由器
异构交换机
异构防火墙
纵深防御感知
NGFW
IDS/IPS
VPN
WAF
日志审计(LAS)
邮件安全网关
应用安全网关（ICG）
扩展态势感知
终端安全系统
数据安全系统
身份安全系统
业务安全系统(UBA)
实战化全局态势感知体系(Process)
人工活动
安全运行(Operation)
安全事件响应(SIRP)
安全分析(Analytics)
可视化分析Invisibility
自动化编排响应(SOAR)
安全检测(Detect)
态势呈现Visualization
威胁情报（TIP）
数据处理(Process)
威胁告警
大屏指挥
自动活动
数据识别(Identify)
响应处置
安全能力（9）
(Output)
基础架构安全能力
资产沦陷发现
主机风险
分组风险
设备健康度监控
负载监控
连通监控
纵深防御能力
网络威胁发现
威胁告警
威胁情报
关联分析
APT攻击发现
业务资产主动外连
代理发现
DNS Tunnel发现
reGeorg Tunnel发现
SOCKS代理发现
DGA域名发现
帐号风险发现
异地账号登录
暴力破解
明文密码泄露
弱口令监测
邮件威胁发现
敏感关键词列表
敏感后缀邮件列表
远程访问风险发现（VPN）
登录地域分布统计
账号登录行为统计
积极防御能力
内部风险发现
异常内容散播发现
内部数据泄露发现
非法外联发现
明文密码泄露发现
威胁猎杀
攻击链分析
失陷分析

安全运行流程

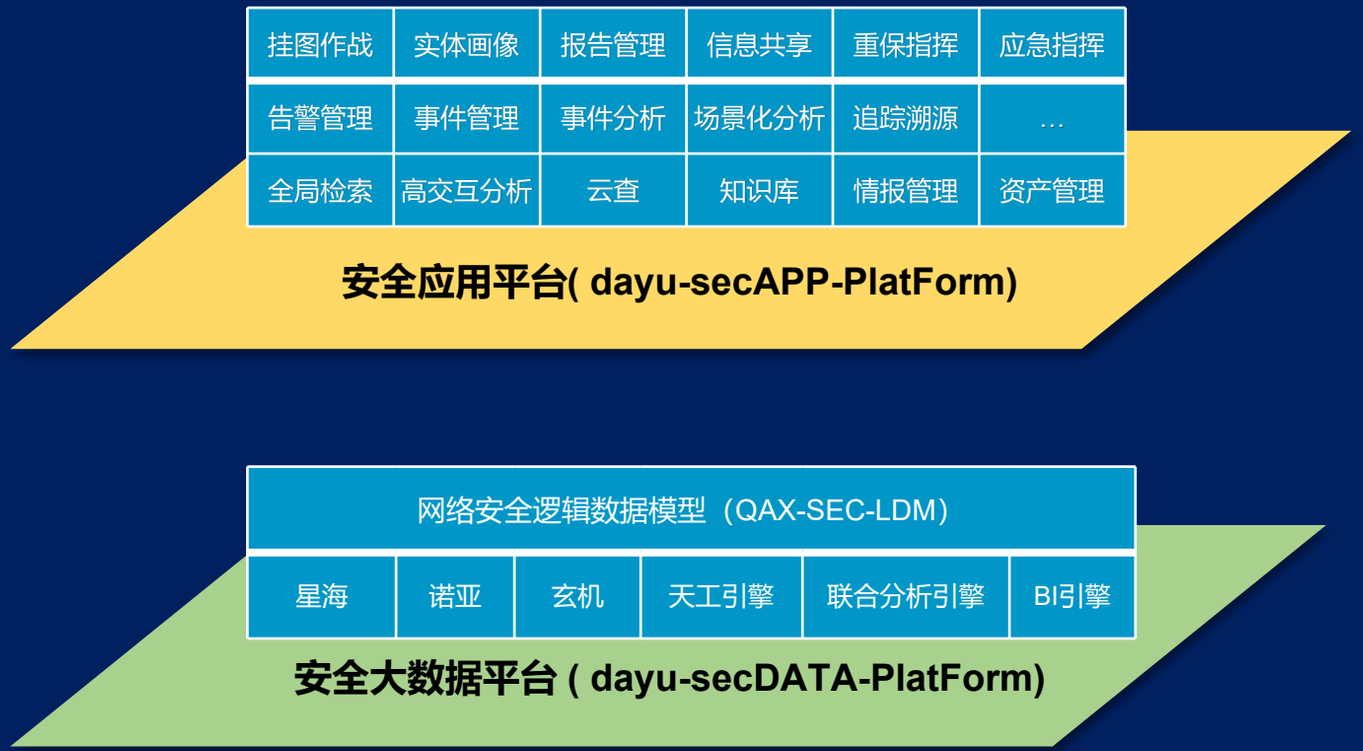

态势感知安全运营流程图(V1.0)
告警监控
设备监控
人工事件上报
告警信息
低风险告警
高风险告警
置告警状态
威胁分析
分析研判
临时解决方案
安全验证
已解决
未解决
完整解决方案
解决
未解决
安全预警
严重性鉴定
不严重
严重
发布预警通告
专家应急响应
事件处置追踪
建立工单
事件处置
事件复核
威胁消除
威胁未消除
关闭工单

平台定义

- 平台集**安全开发运行框架**、**安全大数据平台**、**安全分析应用**平台三位一体的安全中台。整合集团安全产品、服务、数据，为我司新一代基于实战攻防的态势感知（监管、运营、攻防）以及数据安全、工业安全等产线和重点项目，提供平台级支撑。



平台的应用架构



安全应用平台：

提供资产、告警、事件、情报、检索为核心应用，支撑业务快速产品化。

- **态势感知**
- **数据安全**
- **工控安全**

安全大数据平台：

基于安全数据模型，提供数据接入、治理及工具集，为数据驱动的安全业务开发提供支撑

安全运行框架：

可扩展、高可用、高性能和安全性的服务运行框架

平台的安全能力

资产识别

主要集中在于对资产及其脆弱性（漏洞）的识别。

- 对设备、系统、软件、网站、服务等资产类型的统一管理
- 自动发现资产（扫描+流量发现）的统一纳管、运营
- 对接漏洞日志，关联相关资产形成资产漏洞档案
- 通过cpe和cve匹配，分析资产潜在漏洞，分析漏洞影响范围

威胁检测

对内、外部威胁的检测能力

- 各类设备、平台告警的穿透、过滤
- 针对流式数据进行实时日志关联分析（网络攻击类）
- 针对全量数据的离线分析
- 对海量告警进行归并
- 关联告警、日志自动发现高价值安全事件
- UEBA（基线模型）
- 文件动静态检测
- 重点目标的布控
  如：全流量采集监控

威胁分析

针对告警、事件的分析能力及针对特定安全场景的分析能力

- 高性能、跨数据源的全局统一检索
- 高交互分析能力（类Splunk）
- 告警、事件的基本展示、分析
- 针对不同告警分类的个性化展示
- 关键证据采集分析（PCAP包等）
- 根据攻击链、钻石模型对网络威胁的结构化分析
- 场景化分析
- 多人协同分析
- 溯源能力
- ATT&CK映射

风险评分和优先排序

针对告警、事件的威胁评分；对资产、单位、行业、区域的安全评分；对攻击者的评分；

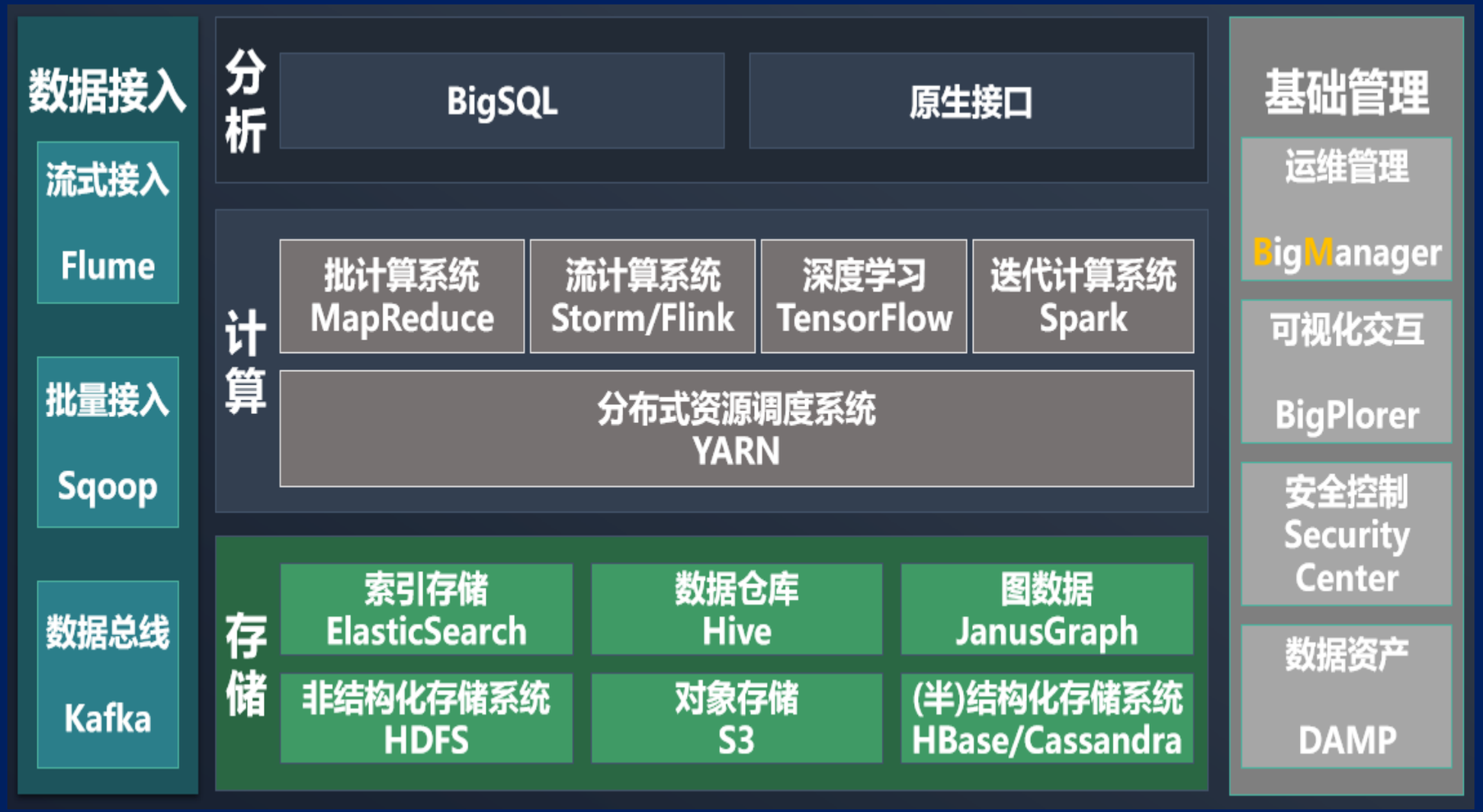
- 告警威胁评分
- 事件威胁评分
- 资产安全指数
  （设备、系统、单位、行业、区域）
- 攻击者威胁指数

威胁响应

通过对高位（机构、人员）、低位（设备）对象的协调联动，实现不同层级的威胁响应

- 安全设备的联动能力（海龙）
- 响应编排

大数据平台（Hadoop）:大数据基础设施



结构化数据：

- 关系型数据库

半结构化数据：

- 列表数据

非结构数据：

- 文件
- 音/视频
- 图片

批量数据：

- 非实时、大规模、成批量

大数据平台组件

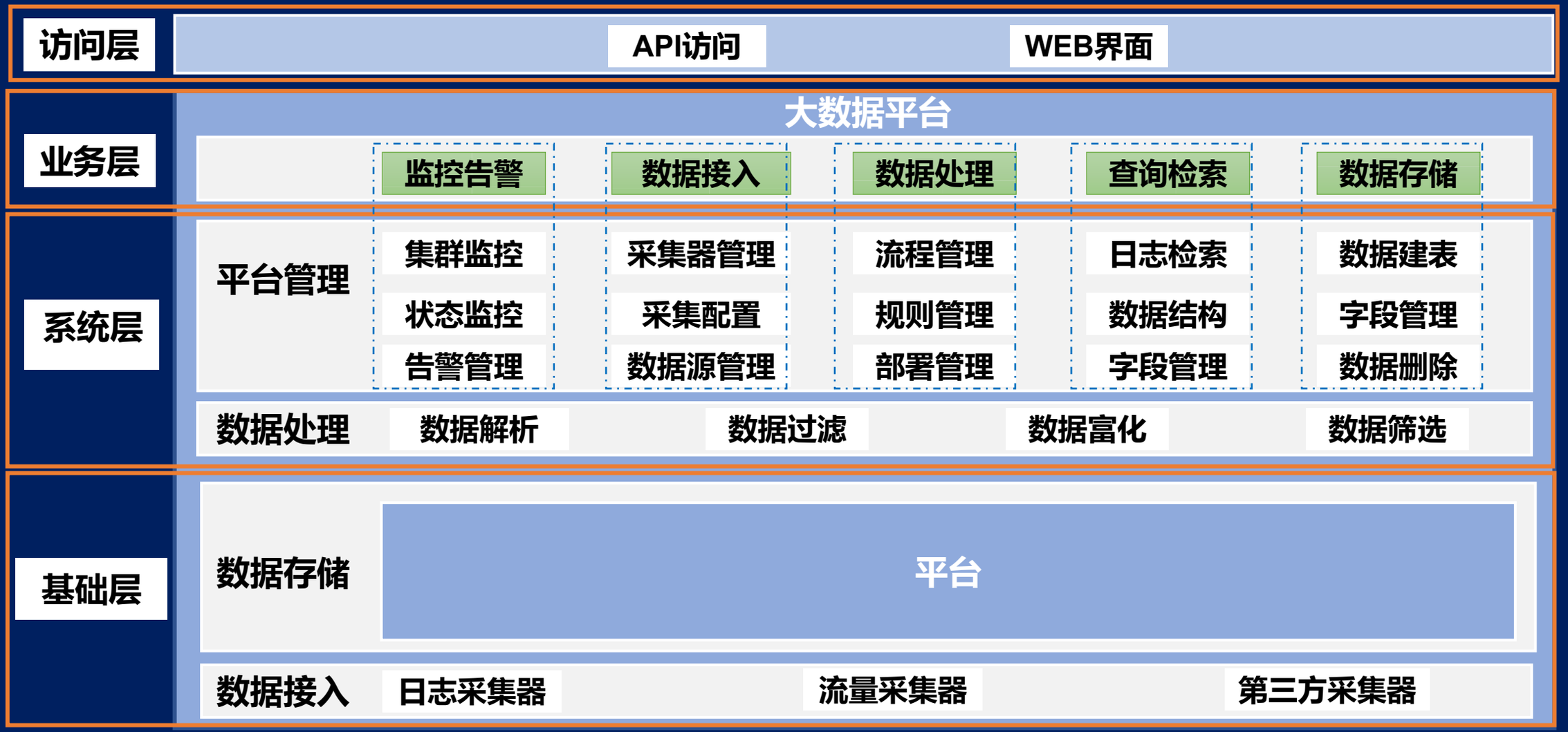
序号	分类	开源组件	描述
1	数据接入	Kafka	流式数据接入的消息队列系统
2		Flume	流式数据接入的分布式的海量日志采集、聚合和传输的系统
3		Sqoop	批量数据接入的开源组数据转存系统，能够实现关系型数据库与大数据数据库之间的数据双向存储
4	数据存储	ElasticSearch	基于Lucene上开发的数据检索搜索服务
5		Hive	结构化的数据仓库
6		图数据库	基于JanusGraph的深度优化图计算组件，应用于图数据库、图计算
7		S3	对象存储，实现非结构化数据加载， 用于大小文件的统一存储
8		HDFS	非结构化存储，文件形数据，能够满足大文件的存储，如日志、视频、图片、音频等文件
9		Hbase	半结构化存储，列存储，以表格形式进行数据存储
10	数据计算	MapReduce	超大规模的数据时将计算分摊到不同节点中进行计算
11		Flink	流式计算，如金融计算、社交网络计算
12		Spark	迭代计算，逐次求解，用于推荐引擎、多维度报表
13		Yarn	分布式资源调度系统，能够针对MapReduce、Spark、Storm等进行资源调度
14		ZooKeeper	分布式协调服务负责整体协调大数据平台的多种组件内部自身的资源调度情况
15	数据分析	BigSQL	主要针对非结构化数据、半结构化数据设计而成，在针对这些数据的挖掘方面能够有很好的效果
16	可视化交互	BigPlorer	进行数据分析与处理，数据提取，数据消费等全方位的产品服务
17	运维管理	BigManager	大数据平台的统一部署、统一调度、统一运维、统一参数配置等

数据管理（ETL抽取）

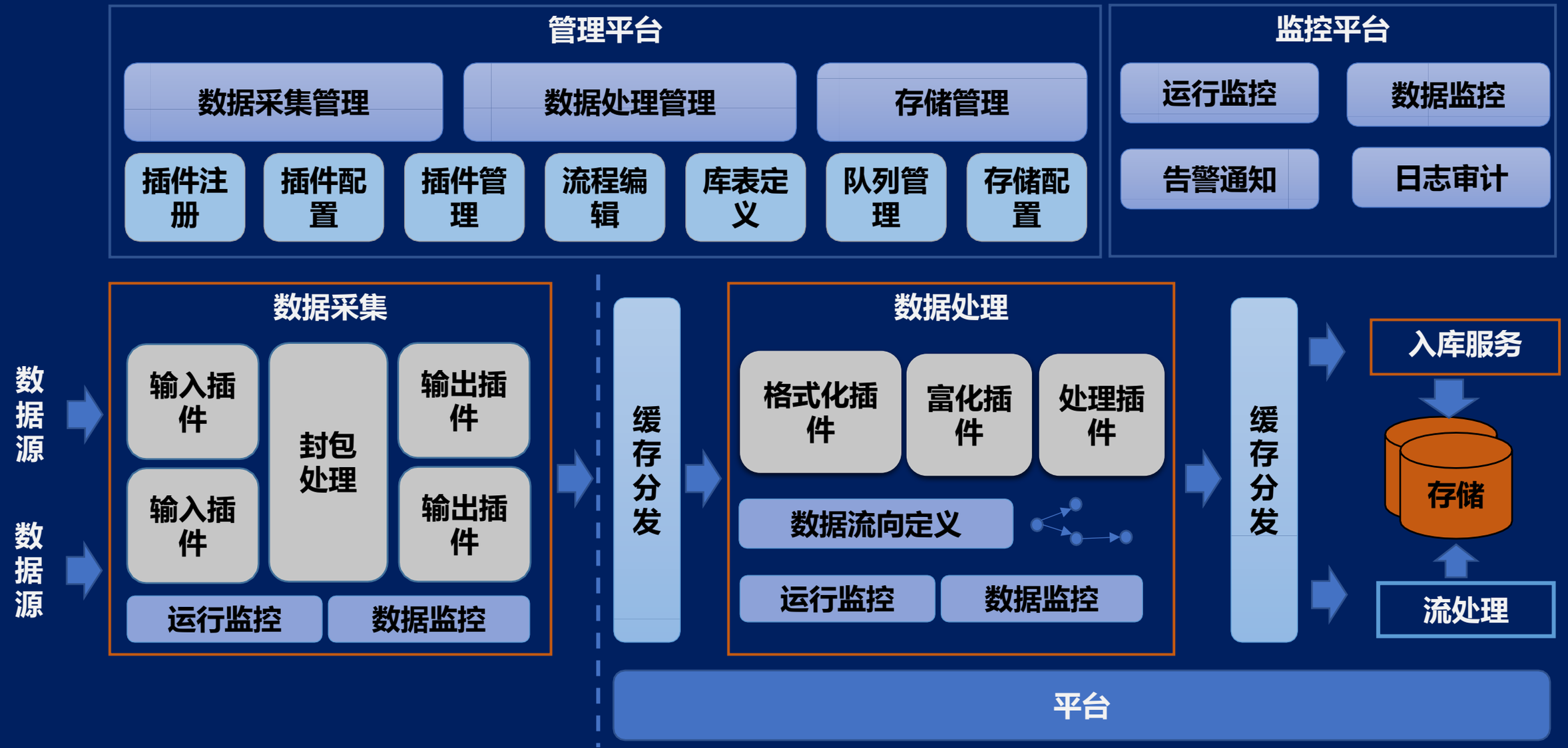
大数据操作平台

- 平台专注于简化大数据应用开发中的流式数据接入、格式转换，富化、实时处理、异常日志，告警等领域的大数据业务系统支撑平台。
- 该平台向安全产品提供流式数据收集、处理、存储、计算、告警等通用型便捷、高效、易用的大数据业务系统支撑环境，可以实现大数据技术与业务解耦、业务无感知的底层技术迭代，提供配置式大数据业务开发环境，高效利用大数据资源的抽象方法。
- 今年已完成49种设备类型，179种设备型号，510种日志类型标准化处理规则506。

平台能力架构


访问层
API访问
WEB界面
大数据平台
业务层
监控告警
数据接入
数据处理
查询检索
数据存储
系统层
平台管理
集群监控
状态监控
告警管理
采集器管理
采集配置
数据源管理
流程管理
规则管理
部署管理
日志检索
数据结构
字段管理
数据建表
字段管理
数据删除
数据处理
数据解析
数据过滤
数据富化
数据筛选
基础层
数据存储
平台
数据接入
日志采集器
流量采集器
第三方采集器

平台技术架构


管理平台
数据采集管理
数据处理管理
存储管理
插件注册
插件配置
插件管理
流程编辑
库表定义
队列管理
存储配置
监控平台
运行监控
数据监控
告警通知
日志审计
数据源
数据源
数据采集
输入插件
输入插件
封包处理
输出插件
输出插件
运行监控
数据监控
缓存分发
数据处理
格式化插件
富化插件
处理插件
数据流向定义
运行监控
数据监控
缓存分发
入库服务
存储
流处理
平台