GLA 数据资产管控平台

技术白皮书

西藏国路安科技股份有限公司

二〇一七年四月

1 背景

大数据时代，数据资产被业界公认为是最宝贵的资产之一，其价值也得到了普遍认同。然而面对如此有价值的资产，其安全性就显得格外重要。近年来数据窃取手段快速演进，数据泄露事件层出不穷，即使信息化建设成熟的政府行业同样难于幸免。因此仅仅网络边界安全防护远远不够，作为存储数据的仓库——数据资产自身的安全壁垒还需加固。

另一方面，据了解，很多部委在信息化建设过程中，多年来由于业务系统不断的增加，使得系统愈加庞大，并且它们之间的关系也越来越复杂。 但是由于前期没有统一规划、人员/厂商变动频繁、技术更新快等原因，使得多数单位的数据资产处于混沌状态，很难说清所有应用系统的数据内容、流转情况、使用情况。有的单位尝试手工进行梳理 ,但是投入大量的人力、物力却只能梳理出一小部分的数据资产。对于数据资产的使用，客户经常面临以下这些情况：

- 数据资产底账不清，客户单位往往不知道现在究竟有哪些数据库？数据库支撑哪些业务系统？数据库中又存了哪些数据？这些数据究竟该归谁来进行管理？

- 对大部委来说，各业务部门每年都会上线众多的应用系统，产生新的数据资产，但这些系统和资产通常没有被统一管理，使得运维部门、安全部门面临困惑。

- 每个系统都有自己的账户管理体系，但是人员调动或者离职时，相应的系统账户、数据库账号并没有及时调整，导致系统中存在大量的死账号。

- 应用系统在大部分情况下是正常运行的，但不能确保每个系统都是在合规使用，之前就发生过某应用系统留有后门，夜里偷传关键数据的情况。应用系统的数据流转无法跟踪。

- 几十年来，建设了众多的应用系统，但这些系统的利用率究竟怎样？哪些数据资产更经常被使用，哪些又长期处于静默状态？

2016 年，国路安参与了某部委的《XX 信息中心数据安全管理系统服务项目》，在项目需求跟踪和梳理过程中，我们了解到数据资产底账梳理的重要性和紧迫性，并与客户一起探讨如何建立有序的数据资产使用流程，提升数据资产管控平台能力。

在项目的实施过程中，数据资产采集探针，采用流量侦听、DPI 技术、协议识别和协议分析技术镜像侦听捕获数据资产动态信息，为资产管理平台提供全量的数据资产信息，并对敏感字段、敏感信息进行标识上报。结合采用主动扫描的方式实现对静态数据的采集和梳理。数据资产管理平台实现资产标识、认领，热度分析、流向展示，账号热度分析，资产的注册登记，资产分级分类等功能。

与此同时，中办、国办 2016 年 7 月发布了《国家信息化发展战略纲要》，纲要中指出：*“建立信息资源基本制度体系。探索建立信息资产权益保护制度，实施分级分类管理，形成重点信息资源全过程管理体系。加强采集管理和标准制定，提高信息资源准确性、可靠性和可用性。依法保护个人隐私、企业商业秘密，确保国家安全。研究制定信息资源跨境流动管理办法。”*

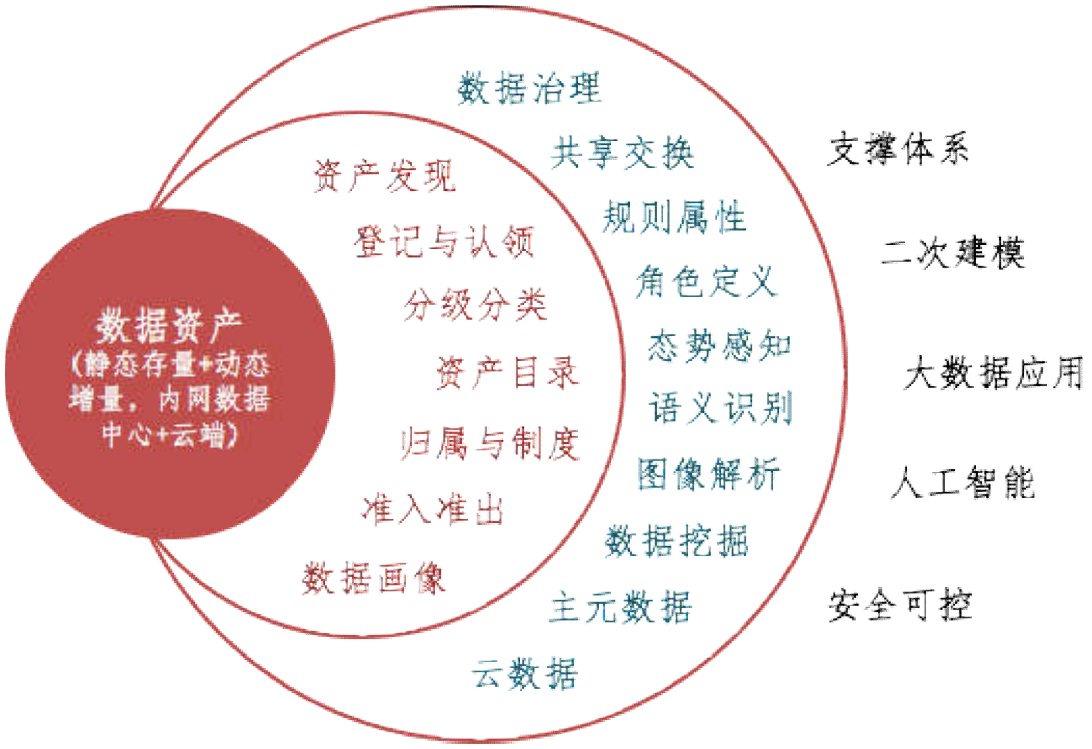
为了更好的满足市场需求，提高市场竞争力，国路安拟立项进一步基于大数据技术开发完善GLA数据资产管控平台，以数据资产采集探针作为基本数据采集方式，逐步形成国路安数据资产安全管控方案。

1.1 政策导向和法律法规

- **2015 年 9 月，国务院印发《促进大数据发展行动纲要》**
    - 大力推动政府部门数据共享。加强顶层设计和统筹规划，明确各部门数据共享的范围边界和使用方式，厘清各部门数据管理及共享的义务和权利。
- **2016 年 7 月，中办、国办发布的《国家信息化发展战略纲要》**
    - 建立信息资源基本制度体系。探索建立信息资产权益保护制度，实施分级分类管理，形成重点信息资源全过程管理体系。加强采集管理和标准制定，提高信息资源准确性、可靠性和可用性。
- 9 月 5 日，国务院出台《政务信息资源共享管理暂行办法》。
    - 《办法》指出，按照“谁经手，谁使用，谁管理，谁负责”的原则，使用部门应根据履行职责需要依法依规使用共享信息。
- **工业和信息化部正式印发了《大数据产业发展规划（2016 - 2020 年）》**
    - “推动数据开放与共享、加强技术产品研发、深化应用创新”三大重点，完善“发展环境和安全保障能力”两个支撑，打造一个“数据、技术、应用与安全协同发展的自主产业生态体系”，提升我国对大数据的“资源掌控、技术支撑和价值挖掘”三大能力。

2 产品概述

数据资产监管平台，作为数据资产安全管理系统的展示平台，旨在利用自学习、数据挖掘、大数据分析等技术将所有探针采集到的数据资产的静态动态信息进行有效梳理，对数据资产进行资产属性注册登记，形成数据资产目录。进而实现数据资产的分级分类，数据资产动态感知等功能。最终给客户提供合理的资产管理建议。协助客户建立数据资产有序管理制度。



2.1 产品设计目标

本产品建设目标是通过各种技术工具与服务结合，完善数据资产属性，建立数据资产目录，对数据资产进行分级分类，并进行数据资产态势感知，能够对整体的数据资产情况做到"说得清"、"看得见"、"管的住"、"能审计"。

具体来说系统能够实现对数据资产的动态管理，完成存量的清理和增量的管理，利用自学习技术，建立行为基线，对敏感资产的违规行为产生告警，并自动化完成数据资产的定期检查，为数据资产干系人建立数据资产安全管理的线上平台。

利用数据挖掘、大数据分析等技术将所有探针采集到的数据资产的静态动态信息进行有效梳理，对数据资产进行资产属性注册登记，形成数据资产目录。进而实现数据资产的分级分类，数据资产动态感知等功能。协助客户建立数据资产有序管理制度。

3 产品功能

3.1 资产梳理

可实现的功能包括：资产目录（来源：导入/注册/扫描/侦听）大小资产目录、资产属性登记、资产标识、资产认领等功能。

3.2 建立资产目录

通过主动扫描技术能够精准的感知到所有资产的库体结构信息：库名、表名、表列名称、账号等，建立大小资产目录。

3.3 资产属性注册登记

通过对数据资产的属性注册登记，实现数据资产信息的注册管理。数据资产属性包括：基本属性、业务属性、干系人属性、重要性、安全属性、使用属性等。

3.4 资产分级分类

通过自学习，对数据资产按照资产属性进行不同维度的资产分级分类，为用户提供不同的分级分类依据，对不同级别类型的资产采取不同的安全管理策略。保证核心资产有序的使用。

3.5 资产风险监测

采用大数据技术采集海量的数据资产访问行为数据，采集的数据应确保全面性、准确性和实时性，基于海量的行为访问数据和智能分析算法，将形成数据资产访问行为基线，建立合规的数据资产数据访问模型。

更重要的是，系统还将监测数据资产的各类安全事件，如发生敏感信息的流转、非授权的访问、不合规的异常访问、突发大数据量访问等异常行为，系统将第一时间发出告警。并挖掘潜在风险预测及评估。

3.6 态势感知

结合大数据技术，对数据资产信息进行深度挖掘、关联分析，进而生成数据资产安全态势分析。包括：资产分布、资产备案分析、资产热度分析、敏感资产分析、敏感资产分析、资产风险等。

3.7 策略配置

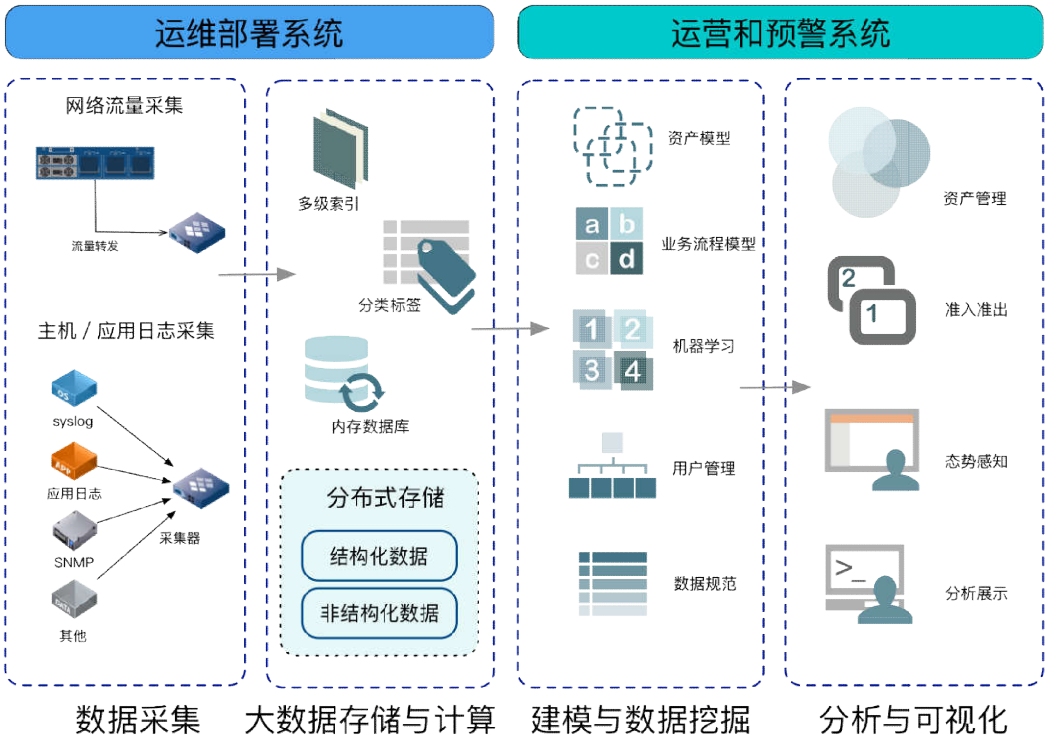
资产监管平台通过策略配置实现对采集探针的管理和控制。如采集信息的范围、采集信息的细粒度、采集信息标识等。

3.8 系统管理

系统管理模块实现探针管理、人员管理、权限管理、角色管理等系统基本配置。

4 产品优势

4.1 基于大数据和机器学习的技术框架

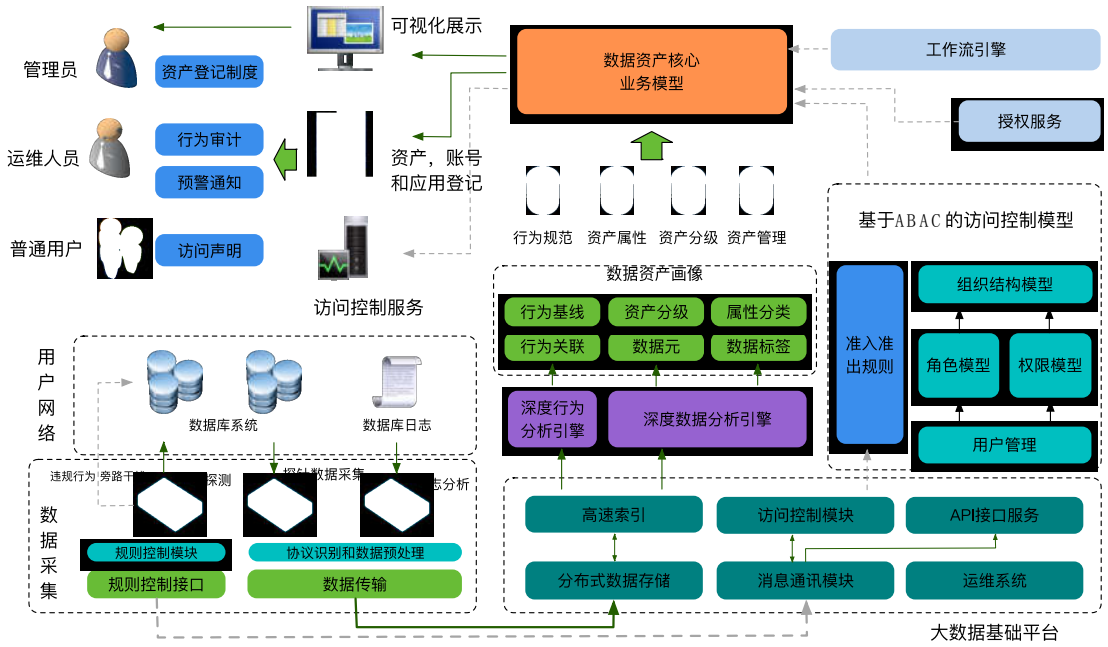


技术架构设计中，遵循采集，存储，分析处理，展示的流程框架。在大数据平台架构中，采用 lambda 架构，支持实时和批处理两种模式的处理。同时系统在建设时，还会兼顾运维部署系统和运营预警系统的建设。满足不同职能部门的业务需求。

4.2 分布式的数据库探针采集技术

1. 分布式数据采集探针，支持结构化与非结构化数据资产的采集
2. 高性能全量数据采集，支持万兆网络流量采集，支持分布式部署
3. 实时监测数据资产的变动、监测数据资产间的数据流向、监测对数据资产的所有访问行为
4. 精准的深度协议识别能力，识别 20 多种主流的数据库协议，支持 SQL 语义分析，细粒度审计每个用户的数据库访问行为

4.3 完整的数据资产管理方案



1. 资产梳理采用定期扫描/人工表单报送/网络侦听等方式，从而实现对静默资产、活跃资产以及核心资产等多种类数据资产梳理。

2. 建立资产目录通过资产梳理建立从应用系统到库表结构的自上而下的大、小资产目录结构。实现资产的统一管理 :资产标识、资产认领、核心资产划分、敏感资产标识等。

3. 资产属性注册登记通过对数据资产的属性注册登记 ,实现数据资产完整信息的注册管理。数据资产属性包括：基本属性、业务属性、干系人属性、重要性、安全属性、使用属性等。

4. 资产分级分类通过自学习 ,对数据资产按照资产属性进行不同维度的资产分级分类，为用户提供不同的分级分类依据，对不同级别类型的资产采取不同的安全管理策略。保证核心资产有序的使用。

5. 资产风险监测采用大数据技术采集海量的数据资产访问行为数据 ,采集的数据应确保全面性、准确性和实时性，基于海量的行为访问数据和智能分析算法，将形成数据资产访问行为基线，建立合规的数据资产数据访问模型。

6. 态势感知结合大数据技术，对数据资产信息进行深度挖掘、关联分析，进而生成数据资产安全态势分析。包括：资产分布、资产备案分析、资产热度分析、敏感资产分析、资产风险分析等。

5 产品型号

5.1 探针设备型号

探针型号	说明
GLA-DAP-PROBE-200	对于 200M 流量以下单探针网络，体现综合能力
GLA-DAP-PROBE-500	对于 500M 流量以下单探针网络，体现综合能力
GLA-DAP-PROBE-1000	针对千兆网络，主打数据存储能力。适用于中性流量网络和有较长时间存储需求的用户
GLA-DAP-PROBE-10000	针对万兆网络，主打运算和数据挖掘能力。适用于大流量网络，用户除了存储还希望对数据进行深入挖掘的需求

5.2 大数据集群服务器

大数据集群可以支持由常用的服务器组合成大数据集群，下面是大数据集群中的不同服务器角色和配置推荐，

服务器角色	用途	配置说明
高密度节点	适合各种类型的服务部署	2U / 256G 内存 / 32T 硬盘
存储节点	适合作为分布式存储节点	2U / 48G 内存 / 48-156T 硬盘
计算节点	适合并行计算用途	2U / 256G 内存 / 1T 硬盘

通用服务节点	适合各类业务应用系统，运维系统的部署	2U / 48G 内存 / 4T 硬盘

5.3 超融合一体机

超融合一体机采用高密度集成的方式，将普通服务器集成到一台集成设备中，节约了物理设备空间

2U 四节点架构（标准版）

4U 八节点架构（增强版）

7U 十节点架构（高端版）

5.4 典型应用场景的设备选型推荐

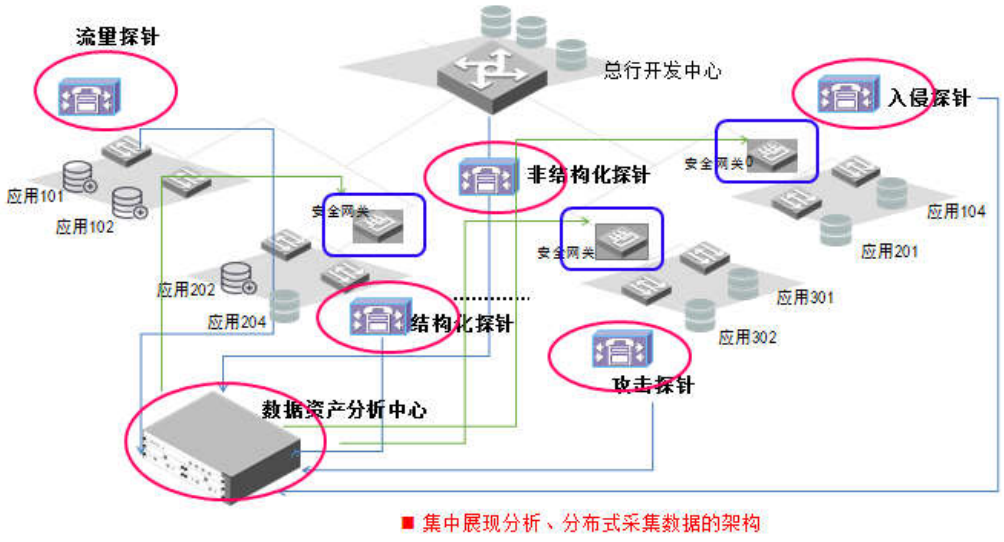
	低端版	标准版	增强版	高端版
适应网络场景	200Mb以下网络的数据采集	500Mb以下网络的数据采集	千兆以下网络的数据采集	万兆以下网络的数据采集
适应的用户场景	POC，Demo，简单网络采集	小型网络采集	中型网络采集	大型网络采集
探针部署模式	单探针部署	单一网络中单探针／多探针部署	单一网络中单探针／多探针部署，多探针组合模式部署	多重网络下多探针组合部署
大数据平台节点数	3-6节点	6-10节点	15-30节点	30节点以上
部署方式	1. 一体机方案 1. 一台2U4节点超融合一体机 2. 机架式方案 1. 通用节点1台 2. 存储节点3-5台	1. 一体机方案 1. 一台4U8节点超融合一体机 2. 机架式方案 1. 通用节点 1台 2. 高密度节点2台 3. 存储节点3-7 台	1. 一体机方案 1. 4U8节点/7U10节点 2. 机架式方案 1. 通用节点 2台 2. 高密度节点N台 3. 存储节点N台	1. 根据场景重新计算

6 产品部署

GLA 数据资产管控平台的部署模式为旁路模式。

6.1 旁路模式

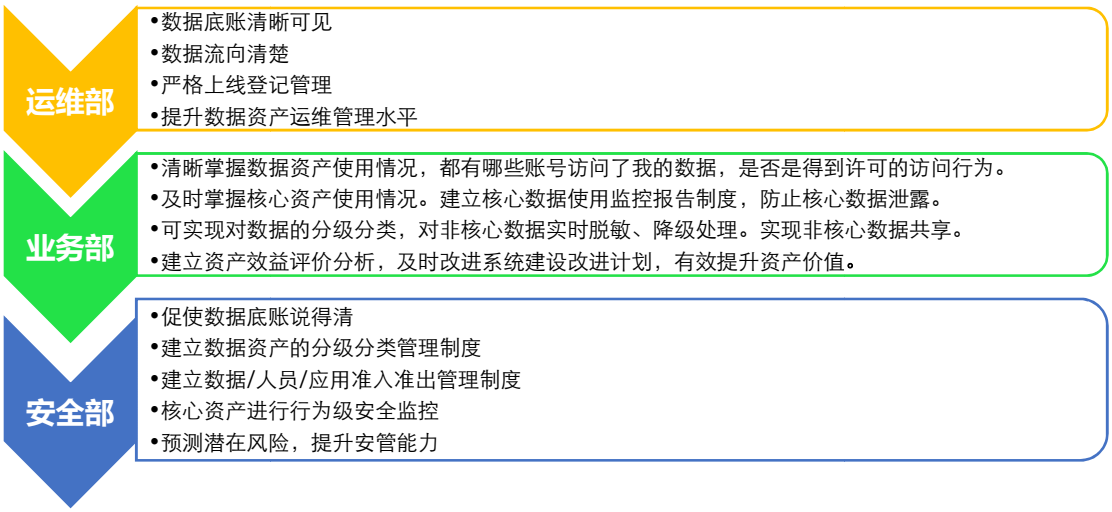
GLA 数据资产管控平台不串联到网络环境中，而是旁挂在交换机设备上，通过流量镜像方式对网络中的流量进行检测和告警。其部署方式与 IDS 相似。部署结构图如下图所示。



7 用户价值

掌握三大基础信息	建立两个制度	实施一个审计
• 数据资产信息 • 人员账号信息 • 应用账号信息	• 核心数据资产行为许可制度 • 核心资产用户行为声明制度	• 核心数据资产行为级审计

7.1 解决的用户问题



8 服务支持

1. 产品部署实施
2. 定制化功能与界面开发
3. 技术分析，数据分析支持，业务建模支持
4. 产品运维