

AISWare Data OS

数据中台操作系统 V5.4 白皮书

AISWare Data OS 是为企业提供数据开发治理开放一站式服务的智能平台，平台以元数据为驱动，为企业提供数据的“采、存、算、管、用”智能化管理，打造具备“人”特性的数据中台套件。帮助企业释放数据价值，降低管理成本，规避数据风险。

声明

任何情况下，与本软件产品及其衍生产品、以及与之相关的全部文件（包括本文件及其任何附件中的全部信息）相关的全部知识产权（包括但不限于著作权、商标和专利）以及技术秘密皆属于亚信科技（中国）有限公司（“亚信科技”）。

本文件中的信息是保密的，且仅供用户指定的接收人内部使用。未经亚信科技事先书面同意本文件的任何用户不得对本软件产品和本文件中的信息向任何第三方（包括但不限于用户指定接收人以外的管理人员、员工和关联公司）进行开发、升级、编译、反向编译、集成、销售、披露、出借、许可、转让、出售分发、传播或进行与本软件产品和本文件相关的任何其他处置，也不得使该等第三方以任何形式使用本软件产品和本文件中的信息。

未经亚信科技事先书面允许，不得为任何目的、以任何形式或任何方式对本文件进行复制、修改或分发。本文件的任何用户不得更改、移除或损害本文件所使用的任何商标。

本文件按“原样”提供，就本文件的正确性、准确性、可靠性或其他方面，亚信科技并不保证本文件的使用或使用后果。本文件中的全部信息皆可能在没有任何通知的情形下被进一步修改，亚信科技对本文件中可能出现的任何错误或不准确之处不承担任何责任。

在任何情况下，亚信科技均不对任何因使用本软件产品和本文件中的信息而引起的任何直接损失、间接损失、附带损失、特别损失或惩罚性损害赔偿（包括但不限于获得替代商品或服务、丧失使用权、数据或利润、业务中断），责任或侵权（包括过失或其他侵权）承担任何责任，即使亚信科技事先获知上述损失可能发生。

亚信科技产品可能加载第三方软件。详情请见第三方软件文件中的版权声明。

亚信科技控股有限公司 (股票代码: 01675.HK)

亚信科技是中国领先的软件产品及服务提供商, 拥有丰富的软件产品开发和大型软件工程实施经验。公司深耕市场 30 年, 在 5G、云计算、大数据、人工智能、物联网、数智运营、业务及网络支撑系统等领域具有先进的技术能力和众多成功案例, 客户遍及通信、广电、能源、政务、交通、金融、邮政等行业。

2022 年, 亚信科技完成收购商业决策服务领域的领先企业艾瑞市场咨询股份有限公司 (「艾瑞咨询」), 并整合形成新的“艾瑞数智”品牌。通过此次收购, 亚信科技的核心能力从产品研发、交付服务、数据运营、系统集成延伸至咨询规划、智能决策, 成为领先的数智化全栈能力提供商。

亚信科技始终致力于将 5G、AI、大数据等数智技术赋能至百行千业, 与客户共创数智价值。公司以“产品与服务双领先”为目标, 产品研发围绕数智、云网、IT 及中台产品体系持续聚焦, 实现行业引领, 其中云网产品保持国际引领, 数智产品实现国内领先, 部分国际先进, IT 领域产品处于国内第一阵营。

面向未来, 亚信科技将努力成为最可信赖的数智价值创造者, 并依托数智化全栈能力, 创新客户价值, 助推数字中国。

部分企业资质

能力成熟度模型集成 CMMI5 级认证

信息系统建设和服务能力评估 (CS4 级)

云管理服务能力评估证书卓越级

数字化可信服务 - 研运数字化治理能力认证

1S09001 质量管理体系认证证书

150200001T 服务管理体系认证证书

1S027001 信息安全管理体系统认证证书

企业信用等级 (AAA 级) 证书

信息系统安全集成服务资质 (二级)

信息系统安全开发服务资质 (二级)

部分企业荣誉

连续多年入选中国软件业务收入百强榜单

连续多年入选中国软件和信息服务业竞争力百强企业

中国软件行业最具影响力企业

中国软件和信息服务业最有价值品牌

中国软件和信息服务业最具影响力的行业品牌

中国数字与软件服务最具创新精神企业奖

中国电子信息行业社会贡献 50 强

中国人工智能领航企业

新型智慧城市领军企业

IDC 未来运营领军者

目录

1 摘要	6
2 缩略语与术语解释	7
3 产品概述	10
3.1 趋势与挑战.....	10
3.2 产品定义.....	11
3.3 产品定位.....	11
4 产品功能架构	12
5 产品基础功能	13
5.1 数据开发治理开放一体化.....	13
5.2 统一采集.....	14
5.3 统一调度.....	16
6 产品特色功能	17
6.1 数据管治副驾.....	17
6.2 主动元数据.....	18
6.3 多模态数据管理.....	19
6.4 数据采集增强.....	20
6.5 跨域传输增强.....	21
6.6 实时数据计算增强.....	22
6.7 分布式调度增强.....	23
7 产品差异化优势	25
8 场景解决方案	26
8.1 数据管治副驾.....	26
8.2 数据编织之主动元数据管理.....	27
8.3 数据价值运营.....	27
8.4 主数据管理.....	28
8.5 跨域数据传输加速.....	28
9 产品客户成功故事	30
9.1 云边协同大数据平台.....	30
9.2 运营商集团数据治理.....	31
9.3 金融机构数据治理.....	31
9.4 大数据中心数据资源管理.....	32
9.5 钢铁集团主数据治理.....	33
10 资质与荣誉	34

11 联系我们.....35

AsialInfo Confidential

1 摘要

当今社会已全面迎来数字化时代，物理世界的数字化属性不断挖掘、数字价值深度释放，以大数据、人工智能、区块链、云计算等为代表的新一代信息技术飞速迭代，加快向产业和行业下沉。数字化转型赋能产业、重塑业务模式，成为新旧动能转化，实现经济高质量发展的必由之路。

数据是数字化转型的基础也是关键，拥有数据的企业，都在考虑如何从拥有的数据中，挖掘新的市场机会，规避经营风险，提升企业竞争优势。但是随着万物互联，产生的数据越来越多，结构越来越复杂，如何管理数据，让数据更好的产生价值，服务于各行各业，对我们的系统支撑都带来了挑战。

为此，亚信打造了以数据为核心的 Data OS 产品，从数据治理、数据开发、数据开放、数据安全、数据运营五个方面进行全方位支撑。

本白皮书将从产品概述、技术架构、主要功能、客户价值、产品优势等几个方面阐述亚信 AISWare Data OS 产品。

2 缩略语与术语解释

缩略语与术语解释如表 2-1:

表 2-1 术语解释

缩略语或术语	英文全称	解释
API	Application Programming Interface	应用程序编程接口，一种允许不同软件系统之间进行通信和交互的规范，定义了一组规则和协议，使得不同的软件系统可以互相访问和共享数据。
B/S 架构	Browser/Server 架构	浏览器/服务器模式，是 Web 兴起后的一种网络结构模式，Web 浏览器是客户端最主要的应用软件。这种模式统一了客户端，将系统功能实现的核心部分集中到服务器上，简化了系统的开发、维护和使用。
Token	Token	访问鉴权验证信息。
Flink SQL	Flink SQL	Flink 实时计算为简化计算模型，降低用户使用实时计算门槛而设计的一套符合标准 SQL 语义的开发语言。
Spark SQL	Spark SQL	Spark 计算为简化计算模型，降低用户使用计算门槛而设计的一套符合标准 SQL 语义的开发语言。
Hadoop	Hadoop	是一个由 Apache 基金会所开发的分布式系统基础架构，是为了能够跨分布式文件系统处理和存储大数据而构建的一种开源框架。
HBase	HBase	一个分布式的、可扩展的、高性能的大数据存储系统，基于 Hadoop 生态系统构建。HBase 提供了面向列的存储模型，支持海量数据的随机读写和实时查询，适用于大数据存储和分析等场景。
HDFS	Hadoop Distributed File System	Hadoop 分布式文件存储系统，主要为各类分布式计算框架如 Spark、MapReduce 等提供海量数据存储服务，客户端可通过路径来访问文件。
IDE	Integrated Development	集成开发环境。

缩略语或术语	英文全称	解释
	Environment	
MPP	Massively Parallel Processing	大规模并行处理。
SFTP	Secret File Transfer Protocol	是一种安全的文件传输协议，一种通过网络传输文件的安全方法；它确保使用私有和安全的数据流来安全地传输数据。
DistCp	DistCp	Hadoop 开源的跨域传输方案，是用于大规模集群内部和集群之间拷贝的工具。
Broker	Broker	分布式调度集群的一个处理节点，常驻进程。
DP	Data Processing	产品自研的数据处理执行器架构。
RDB	Relational Database	关系型数据库。
Kafka	Kafka	Kafka 是由 Apache 软件基金会开发的一个开源流处理平台，由 Scala 和 Java 编写。Kafka 是一种高吞吐量的分布式发布订阅消息系统，它可以处理消费者在网站中的所有动作流数据。
Redis	Redis Database	Redis 数据库，一个开源的内存数据结构存储系统，通常用作缓存、消息队列和数据库等场景。Redis 支持多种数据结构，如字符串、哈希表、列表、集合和有序集合等，同时也提供了丰富的命令和功能，如事务、管道和 Lua 脚本等。
Yarn	Apache Hadoop Yarn	资源调度平台，负责为运算程序提供服务器运算资源，相当于一个分布式的操作系统平台，而 MapReduce 等运算程序则相当于运行于操作系统之上的应用程序。
Failover	Failover	故障转移。
Leader	Leader	主节点，调度集群中负责调度任务处理核心功能的常驻进程。

缩略语或术语	英文全称	解释
Checkpoint	Checkpoint	数据库事件。
TCL	Tool Command Language	一种解释执行的脚本语言。

AsialInfo Confidential

3 产品概述

AISWare Data OS 是以数据融合为基础、以智慧运营为衔接、以生产执行为服务重点打造的数据操作系统产品。产品提供深化跨域数据整合并沉淀公共的数据能力，实现数据的分层与水平解耦，同时整合数据处理、治理、安全、运营等多种能力，满足客户精细数据服务、数据开放和智慧运营的要求。

3.1 趋势与挑战

国内政策推进数据要素市场化配置改革和数据资源入表，促进数据合规高效流通。国际社会推动数据安全流动利用，释放数据价值。数据产业市场规模和数据规模持续增长，数据中台产品市场空间巨大。

参考业界最新研究成果，数据中台围绕新兴技术，重点考虑主动元数据驱动架构、数据开发治理智能化、数据价值评估及产品内核增强。

由于运营效率的原因或创新发展的需要，企业对数字化的需求会越来越高，为此也会对现有的数据管理带来挑战。

挑战之一：根据服务实时性。对于一些实时性要求高的场景，比如实时营销、位置服务等。为此需要构建面向实际业务生产提供全流程打通的数据能力，将离线分析能力、实时处理能力进行融合，实现统一可视化编排及监控，并结合数据挖掘能力形成企业级端到端数据能力。

挑战之二：数据资产管理全面性。随着 5G、物联网技术的推进，越来越多领域的的数据汇聚在一起，数据量越来越大，结构越来越复杂，数据服务的对象越来越多，对数据的一致性、准确性有更高的要求，需要具备全域的全流程端到端的数据资产管理支撑能力。

挑战之三：创新带来的变化性。在数字化创新过程中，上层应用的变化越来越快，如何快速的支撑上层对数据服务的需求已经迫在眉睫。为此需要以数据为中心，构建企业级数据中台来支撑前端对数据的需求，实现数据服务的沉淀与共享，快速满足前端业务的变化。

3.2 产品定义

构建“数据整合、能力共享、服务创新”体系的企业级数智中枢。AISWare Data OS 数据中台操作系统涵盖数据集成、数据开发、数据治理、数据安全、数据开放、数据运营等能力，通过屏蔽技术和数据多样性，提供全生命周期、全景式、全流程的操作能力。可面向电信、金融、政府、交通、能源等各个行业，构建“数据整合、能力共享、服务创新”的企业级数智中枢，助力企业数字化转型。

3.3 产品定位

AISWare Data OS 是为企业提供数据开发治理开放一站式服务的智能平台，平台以元数据为驱动，为企业提供数据的“采、存、算、管、用”智能化管理，打造具备“人”特性的数据中台套件。帮助企业释放数据价值，降低管理成本，规避数据风险。



图 3-1 产品定位

4 产品功能架构

亚信 AISWare Data OS 功能架构如图 4-1 所示：



图 4-1 产品功能架构

5 产品基础功能

AISWare Data OS 基础功能如下：

5.1 数据开发治理开放一体化

数据开发治理一体化的功能架构如图 5-1 所示：



图 5-1 数据开发治理开放一体化功能架构

- 数据建模

提供可视化的模型设计能力，从基本信息、字段信息、存储信息、物化信息等维度进行模型设计。支持与数据标准的前向映射。支持用户以多人协作迭代式进行模型设计与交付，提高设计开发的效率和一致性。

- 数据采集

基于多样的数据采集与交换场景，支持 30+ 异构数据源与的采集交换。采用去中心化架构，提高并发性及高可用性。提供按租户方式进行数据交换权限分配，从开发、上线、监控整套流程，提升数据采集的安全性。

- 数据开发

基于 DataOps（Data Operations 数据操作）理念的敏捷开发运维 IDE（Integrated Development Environment 集成开发环境）工具包，完美集成流批一体化开发、运维及治理能力，一套工具全流程、端到端支持数据模型及程序开发、任务调度、任务运维及嵌入式数据治理管控，提升开发及管理效率，保障数据开发质量。

- 任务调度

提供了从任务的生成，依赖判断，任务执行到状态维护的任务全生命周期管理。通过实时心跳检测，对任务的健康度进行监控。并提供了在异常状态下自我恢复的能力。提供模版调度配置，通过模版批量生成子任务，支持批处理与流任务的统一调度和运维。

- 数据开放

提供可视化编排方式进行服务开发，并对服务生产全过程进行管理。面向于数据消费者，采取多种方式提供数据服务，并保障服务调用的高效、稳定和安全。提供可视化数据目录帮助用户查找数据服务，并实现对用户、服务的统一管理、授权与审计。

- 数据标准

通过数据标准（词根标准、维度标准、基础标准、命名标准）的制定、发布、映射和稽核，建立严谨的数据标准管理体系，统一企业对数据的认知和定义，保证数据的一致性和合规性，从而提升企业数据质量。

- 数据质量

通过数据质量规则制定、数据质量定期校验、数据质量异常告警、数据质量问题处理，建立闭环的数据质量管理机制，及时发现并解决数据质量问题，提升数据的完整性、合规性、准确性、唯一性、一致性和及时性，赋能业务精准决策。

- 数据安全

通过数据安全分级分类、敏感数据定义和识别、敏感数据脱敏规则制定等能力，实现对数据中台范围内数据的安全管控。

- 存储周期

通过存储策略配置、存储信息监控告警等能力，提升存储资源的利用率，节省企业的存储成本。

- 主数据

通过主数据建模、集成、质量、维护和开放，提供全面的主数据管理服务，提升主数据质量，降低管理成本，支持业务高效发展。

5.2 统一采集

统一采集的功能架构如图 5-2 所示：



图 5-2 统一采集功能架构

- 采集数据稽核

提供采集前，采集中以及采集后数据稽核，及时发现数据问题，保障后续数据处理链路数据来源可信、可用。

- 采集执行框架

提供采集任务执行相关任务从启动到完成的全生命周期管理和控制，包含：采集任务启停、异常处理；采集任务的异常处理；源端目标端或服务的连接以及数据读取及写入；数据库外清洗、转换等数据处理。

- 采集日志

提供数据采集任务运行相关的状态及信息监控和运维操作能力，支持采集任务实例状态、日志以及依赖和血缘等信息查看。

- 跨域数据传输

当前基于 UDT (UDP-based Data Transfer Protocol 基于 UDP 的数据传输协议) 跨域传输机制，当碎文件多时，存在资源消耗大，传输慢问题，因此提供三种传输方案 (Distcp 、 Socket 、 UDT)，同时借鉴 Yarn 资源容量调度机制，设计基于带宽的容量调度，通过配置任务的最低运行带宽和最大带宽，结合实际集群带宽利用率进行动态调整，实现任务带宽资源隔离，保障任务并发执行效率及最大化带宽资源利用率，提高输出效率。

- 库外数据处理

在数据同步过程中，提供统一的库外数据处理能力，如库外数据清洗、转换等，实现不依赖特定数据库语法以及计算资源的统一数据处理能力。

- 传输资源隔离

提供传输资源分组功能，可实现与租户、项目以及任务等粒度的灵活绑定，实现数据传输资源的隔离，保障传输安全和效率。

5.3 统一调度

统一采集的功能架构如图 5-3 所示：



图 5-3 统一调度功能架构

- 任务调度

支持调度任务创建、管理、外部依赖管理和调度任务实例生成状态检测；提供数据采集、处理任务运行相关的状态及信息监控和运维操作能力；支持任务运行状态核心指标统计；支持任务实例状态、日志以及依赖和血缘等信息查看；支持对任务进行重跑、暂停，补数据等。

- 任务优先级

提供任务运行优先级控制能力，保障调度和执行资源基于优先级进行合理的利用，有效保障高优先级任务的执行。

- 任务依赖

提供丰富的任务依赖策略，支持时间依赖、任务依赖、文件依赖等依赖类型，满足复杂场景海量任务的调度运行。

- 分布式调度执行框架

通过云原生改造、中间件支持 Redis 集群以及任务并发策略增强，提供更稳更好用的任务调度能力。

- 集群监控运维

提供对调度集群服务、中间件以及依赖主机资源的核心指标的监控和运维操作，提升集群服务监控和运维效率，保障集群的稳定运行。

- 调度互联互通

在云边协同调度场景，提供边缘和中心调度节点的互联互通机制，实现统一的调度能力。

6 产品特色功能

AISWare Data OS 产品特色功能包括大数据多租户能力模型、开源及第三方快速集成、处理脚本自动生成与优化诊断等。

6.1 数据管治副驾

数据管治副驾的功能架构如图 6-1 所示：



图 6-1 数据管治副驾功能架构

- 大模型+知识库

通过大模型+本地向量知识库的方式，构建数据管治副驾智能化的基础能力。通过持续注入内外部知识（企业业务知识，企业数据管理制度、数据相关的国家政策和发文，行业监管要求，数据管理领域的白皮书和标准），降低大模型幻觉，提升答案准确性。

- 数据开发智能服务

实现数据开发过程中各场景的智能化（如：智能建模、程序智能编排、代码注释生成、代码智能纠错、任务异常原因分析等），降低数据开发和运维的门槛和成本。

- 数据治理智能服务

实现数据治理各场景的智能化（如：数据标准智能映射、敏感数据智能识别、敏感数据智能纠错、质量阈值智能推荐、质量异常自动定位等），降低数据治理的人工成本。

- 数据开放智能服务

实现数据开放各场景的智能化（如：数据目录智能生成、数据智能上架、语义找表、语义分析等），降低数据开放的人工成本。

6.2 主动元数据

主动元数据的功能架构如图 6-2 所示：



图 6-2 主动元数据功能架构

- 元模型管理

针对不同类型数据对象，提供元模型管理能力，支持定义元模型的元素信息及元模型之间的关联关系。

- 元数据主动采集

接入采集对象数据源，选择合适的采集适配器，配置元数据采集任务，定制采集并更新元数据信息。在结构化元数据采集的基础之上，扩展对非结构化数据（图片、音频、视频、文档）的元数据采集能力。

- 元数据主动生成

通过 AI（Artificial Intelligence 人工智能）技术，基于表字段的英文名，自动补齐核心元数据（表中文名、表描述、表标签、字段中文名、字段描述等）信息。

- 元数据解析

通过解析器，自动解析 SQL 和 TCL 等脚本，基于血缘探测结果获取元数据之间的关系信息。

- 元数据维护

各数据对象采集后的元数据缺失或不准确，支持人工补充和更新。完善元数据信息后，通过元数据发布/变更流程，正式发布符合标准的元数据，并支持通过版本管理追溯元数据的历史版本信息。

- 端到端全链路血缘

在现有的解析能力的基础上引入 AI 大模型，利用大模型对代码的理解能力，增强对复杂 SQL 的解析能力，提升表级和字段级血缘解析的准确率。

- 元数据图谱

基于元元模型构建元数据图谱，在数据详情中增加数据图谱，展示当前数据对象与其他数据对象之间的关系，提升用户理解数据的效率。

- 元数据质量

针对各类数据对象提供多种维度元数据质量稽核能力，及时发现和处理元数据的质量异常。

- 元数据主动服务

基于元数据，构建多场景的元数据服务能力（元数据查询、血缘查询、影响分析、元数据质量识别等），并通过 API（Application Programming Interface 应用程序编程接口）等方式向周边系统提供查询服务。同时，通过对元数据存储数据库的选型及优化提升查询速度。

6.3 多模态数据管理

通过对多模数据统一管理，解决传统数据分析与 AI 模型之间的“数据搬家”问题，打通数据分析与 AI 模型引擎，实现基于一份数据多模分析，提升数据驱动决策的准确性和可信性，让数据工程师灵活进行模型和特征训练，实现数据与 AI 开发高效无缝互通。

多模态元数据采集：提供非结构化数据的元数据采集服务，通过丰富的元数据，提升模型开发人员对数据的理解力。

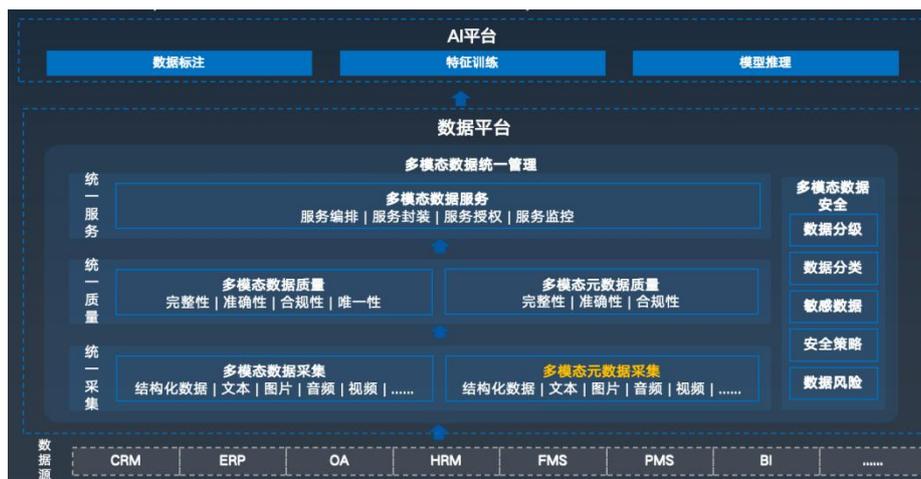


图 6-3 多模态数据管理

6.4 数据采集增强

当前采集执行框架不支持分布式部署、实时采集以及 Checkpoint 等特性，无法有效支撑飞速增长的数据同步需求，通过引入分布式采集执行引擎、库外数据转换等能力，以提高资源利用和执行效率。

- 分布式执行框架
 - 引入新一代超高性能分布式数据集成引擎，实现更细粒度任务分布式调度，预计执行效率提升 20%~30%。
 - 新增支持流批采集开发，简化使用、降低门槛，预计交维和培训时间缩短一半。
- 库外数据转换

增强支持库外清洗转换处理，避免非必要入库及计算，预计 ETL 任务配置到执行全链路耗时降低 20%+。

- 重跑机制优化

提供 Checkpoint 机制，优化任务重做流程，减少处理时间。

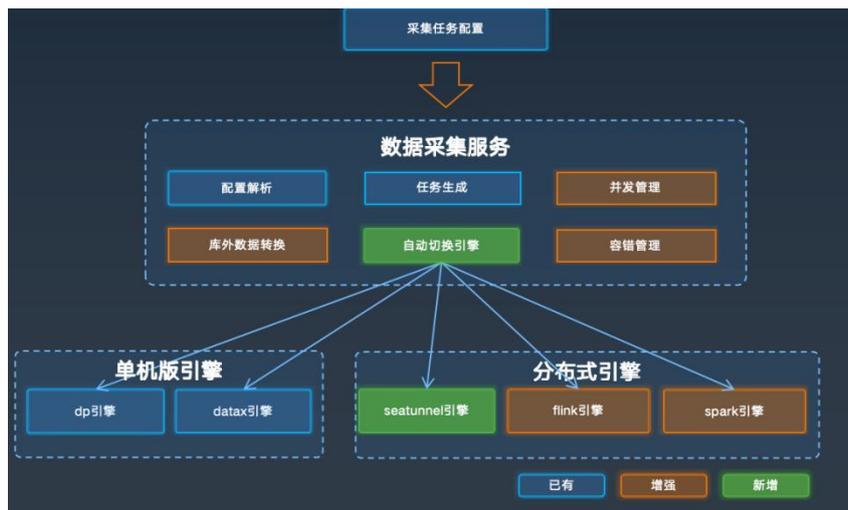


图 6-4 数据采集增强

6.5 跨域传输增强

针对小碎文件和大文件场景，改进开源 Distcp 工具存在的管理和性能及稳定性问题，预估可以节省 50% 以上的 Broker 资源。

- 支持多人并发提交迁移 Job

可以多人随时选择迁移，不需要集群负责人配合执行迁移任务。

- 多 Job 状态管理

多个任务并行执行，平台记录每个 Job 的各个状态。

- 多 Job 进度管理

定时刷新 Job 的进度，通过平台可以看到迁移的进度。

- 大任务智能拆解

对于一些单次搬运数据量比较大的任务，Job 会自动拆解为多个 Job，提高整个搬运的效率。

- 模糊匹配

模糊匹配的功能可以用于多个层级，选择有一定特征的路径一起迁移，减少重复操作。

- 支持文件排序，避免长尾文件

提供灵活文件排序策略，文件合理排序均衡传输，运行过程中实时检测整体运行情况。

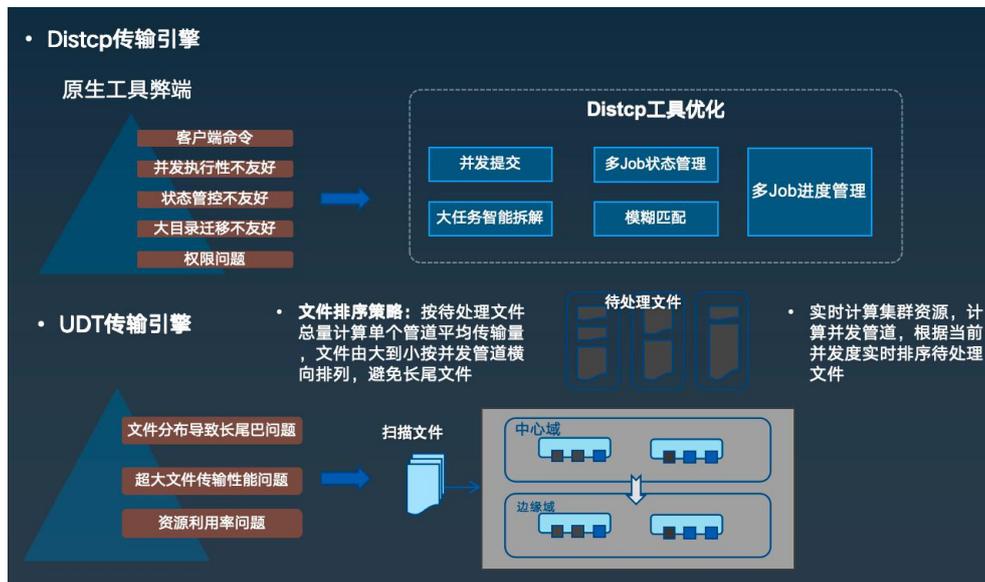


图 6-5 跨域传输增强

6.6 实时数据计算增强

针对多表实时关联计算场景，尤其对历史数据进行关联下，开源流处理框架不可能在内存中存储如此大量的数据，从而无法支持这类场景，因此基于 Flink 引擎自研 StreamSQL 用于简化和增强实时数据计算，提高实时计算效率和稳定性。

- 关联索引表，提升关联效率

通过结合流处理状态和 HBase 索引表，实现二级缓存机制，以提高关联操作的效率和速度。

- 算子下推，减少计算数据量

将关联操作中的条件下推至更低层，减少数据传输和计算量，从而提高处理效率。

- 缓存表关联，提升维表关联性能

针对变化较小的维表，设计缓存表来进行关联操作，以减少维表变化带来的大数据量回撤，提高关联性能。

- 关联查询异步化，减少查询阻塞

利用列分页查询过滤器和异步 IO (Input / Output 输入 / 输出技术) 技术, 实现并发查询, 解决一对多场景下数据膨胀带来的查询时效问题, 提高关联查询效率。

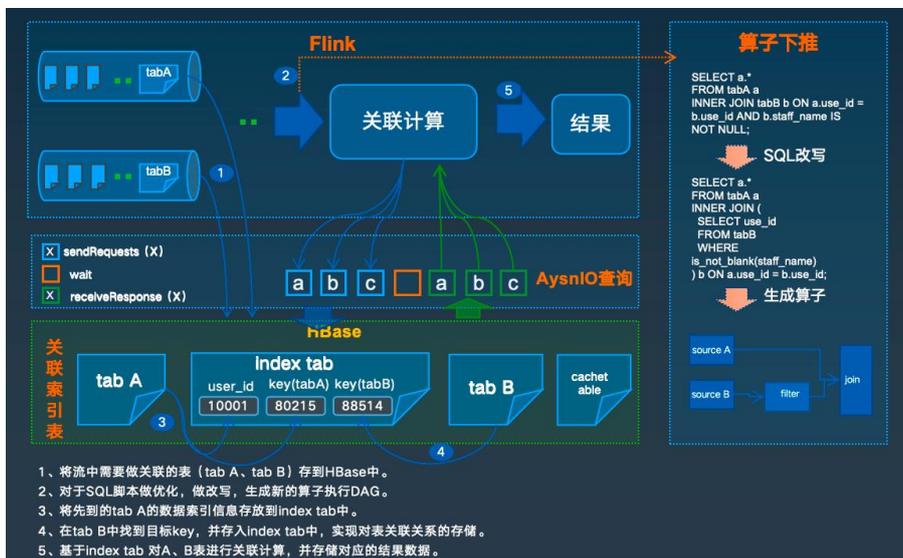


图 6-6 实时数据计算增强

6.7 分布式调度增强

企业数字化运营需求催生了海量的数据任务调度需求, 产品在实际项目中面临部署运维效率、中间件连接数瓶颈以及任务并发需求复杂等挑战, 需进一步升级架构、增强任务和资源调度策略。

- 分布式调度 on K8S (Kubernetes 开源容器编排平台)
 - 提供公平调度、任务拓扑、SLA (Service Level Agreement 服务等级协议) 调度、作业抢占、回填等资源调度策略, 通过队列的资源预留、队列容量管理, 提升多租户的动态资源共享的能力。
 - K8S 统一编排容器, 屏蔽环境及程序的差异, 实现集群节点快速部署与扩缩容, 降低后台部署、运维成本。
- 中间件升级
 - 新增支持 Redis Cluster, 解决哨兵模式的连接数瓶颈, 预计任务并发量提升到 3.5w+。
 - 支持队列数据分区存储, 降低故障情况数据丢失概率。

- 任务并发策略增强

新增支持按照任务的周期、类型、重要程度等维度进行执行时间范围的“黑白名单”策略设定（如非关键任务在指定时间段内运行），提升调度 SLA 服务等级。

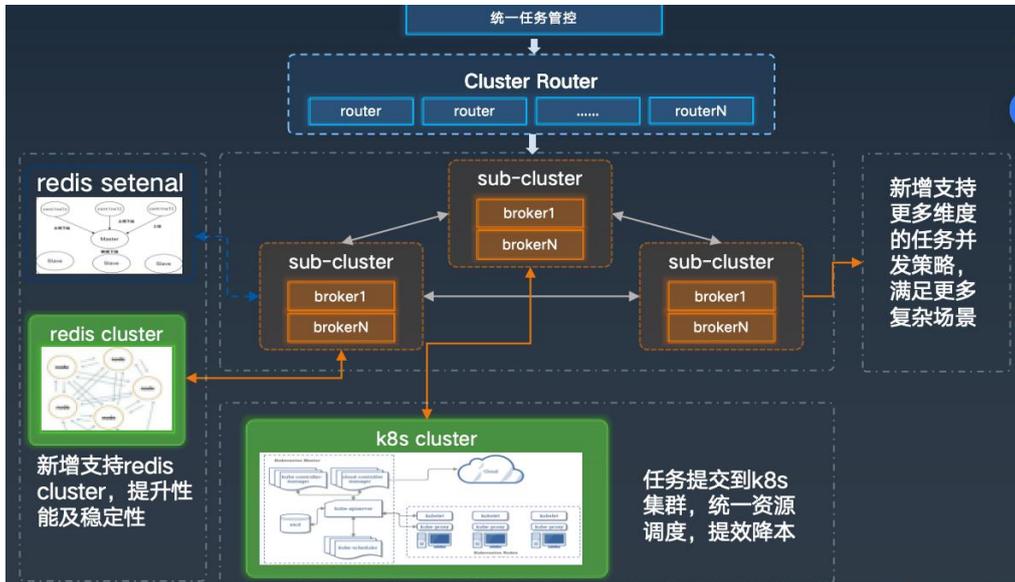


图 6-7 分布式调度增强

7 产品差异化优势

- 多源异构数据采集

支持 30+异构数据库的数据高效采集。

- 一体化开发治理

把数据治理嵌入到数据开发流程中，通过事前治理提升效率。

- 高效调度传输

支持千万级/日任务调度、3.5W+并发调度、TB~EB 级数据高效同步。

- 大模型驱动

实现基于大模型的数据智能开发和治理，减少人工重复操作。

- 多维度数据评估

提供多维度的数据量化评估服务，实现数据从资源向资产的转化。

- 全面信创支持

全面支持各类国产芯片、操作系统、中间件、数据库。

8 场景解决方案

本章主要介绍 AISWare Data OS 不同的应用场景。

8.1 数据管治副驾

通过一站式数据管理副驾,实现数据开发和治理多个场景的智能化和自动化,降低人工成本,提升管理效率。

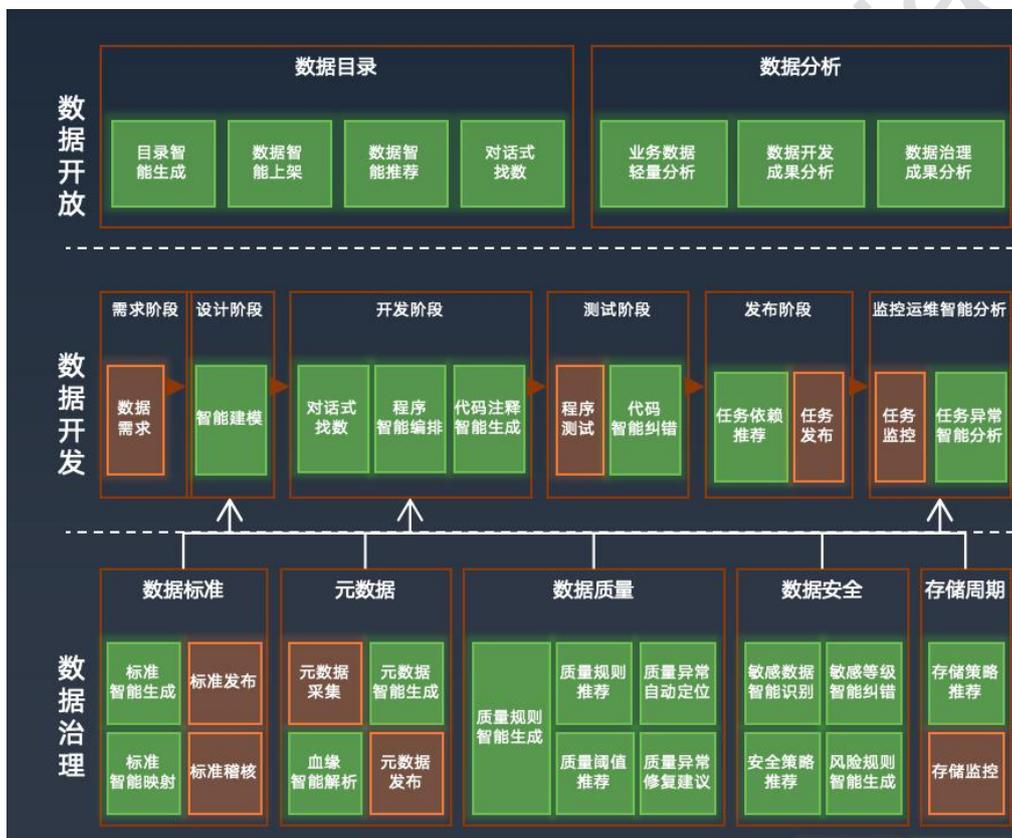


图 8-1 数据管治副驾解决方案



图 8-2 数据管治副驾界面

8.2 数据编织之主动元数据管理

基于 AISWare Data OS 产品，为企业客户构建主动元数据管理能力，实现以元数据为驱动，主动改进数据管理的工具和活动。



图 8-3 数据编织之主动元数据管理解决方案

8.3 数据价值运营

基于 AISWare Data OS 产品，为企业客户提供丰富完善的数据价值评估体系，构建全面评估机制及运营机制。

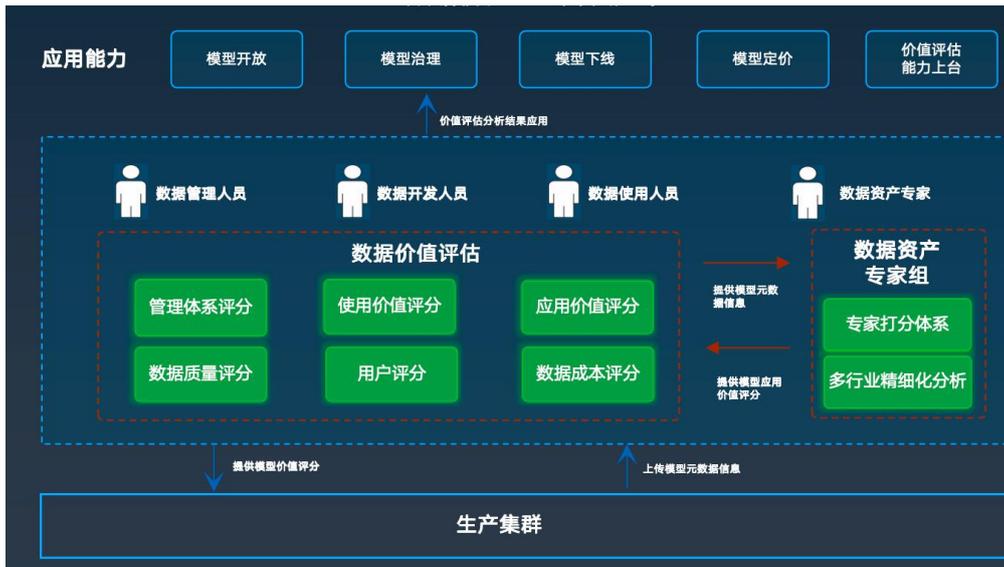


图 8-4 数据价值运营解决方案

8.4 主数据管理

基于 AISWare Data OS 产品，为企业客户基于“业务主数据”打通业务链条，构建主数据采集、运行、管理的主数据体系。

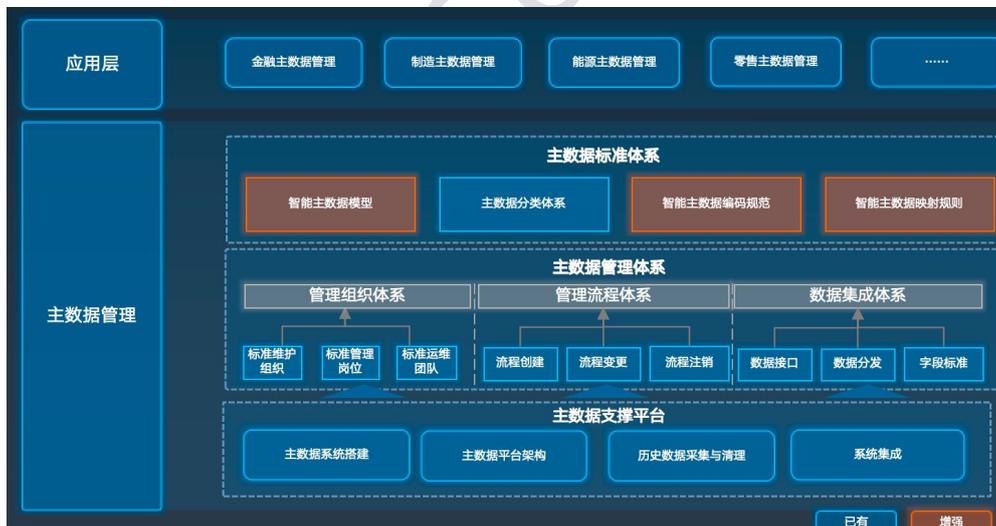


图 8-5 主数据管理解决方案

8.5 跨域数据传输加速

通过综合利用 UDT、实时传输以及 Distcp 等传输技术，探索实现自动基于文件特性匹配最佳传输引擎的跨域传输方案，相较于仅 UDT 传输，取得了整体的最佳传输效率。

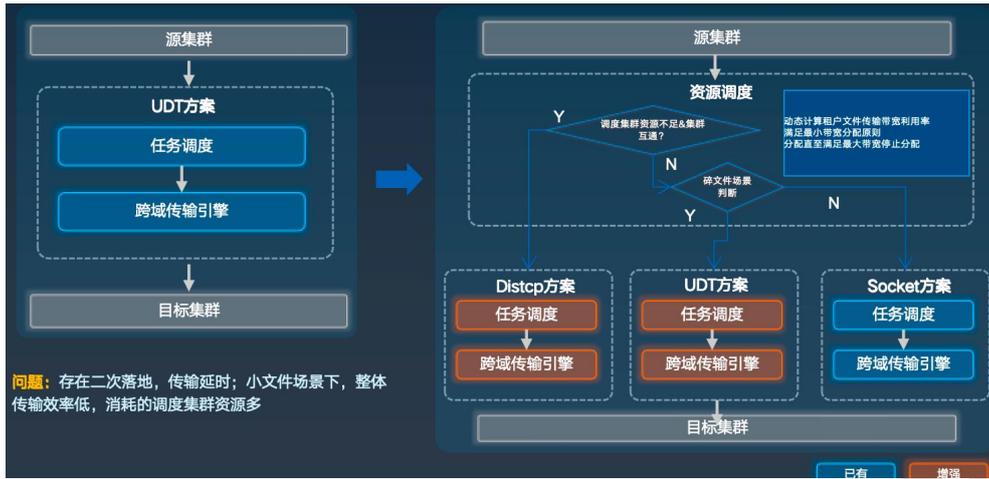


图 8-6 跨域传输加速解决方案

9 产品客户成功故事

以下介绍产品客户成功故事。

9.1 云边协同大数据平台

应用场景

基于亚信科技 AISWare Data OS 产品，搭建“逻辑集中、物理分布”的统一云边协同计算平台。采用分布式协同新型计算架构搭建统一技术底座，实现统一数据管理、统一开发调度、统一开放及统一运营能力。全网类、个性类数据全量引入，支撑全网应用及省专业务创新。

应用效果

- 纳管总部 14 个、省公司 31 个大数据系统元数据；制定标准 20000+个，模型复用率 38%，血缘解析覆盖率达到 80%。
- 入驻集团及专业公司 70 个，已入驻 24 省；构建标准化工具 56 个。
- 支撑 O 域数据贴源接入处理；支撑位置洞察 2 省、内容洞察 1 省数据割接；整合数据服务，构建服务网关。新增交互查询、消息订阅服务，快速支撑租户临时统计类需求。
- 实现对 9 个区域中心计算、任务、服务的统一运营能力；全域运营管理横向拉通，建立平台和运维运营人员的统一管理机制。

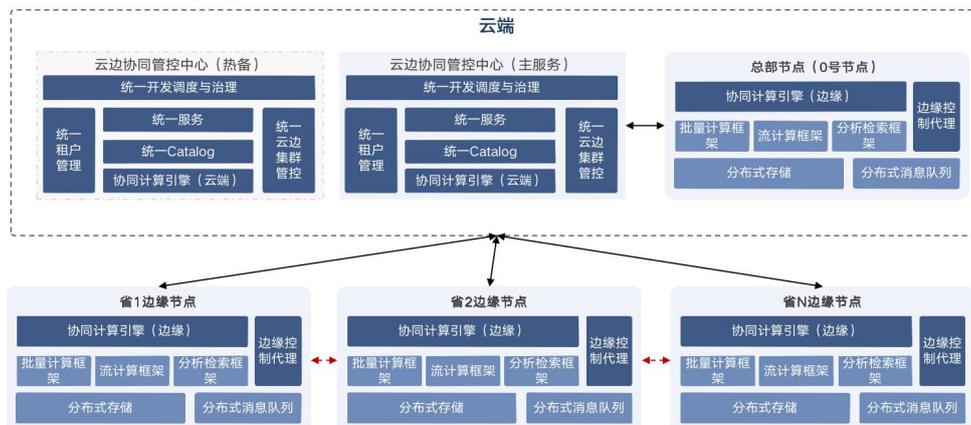


图 9-1 云边协同大数据平台案例

9.2 运营商集团数据治理

应用场景

基于 AISWare Data Go 数据资产管理平台，通过顶层设计打造自顶向下的全网数据治理体系，对接 31 省系统，形成元数据汇聚、过程管控、全面运营、开放服务等能力。以元数据仓库建设为核心，实现全网数据资产管控，完善补充运营、服务能力，为集团、省分、专业公司的数据治理工作，提供可信的、标准的、智能的数据治理能力支撑。

应用效果

- 针对 31 省元数据进行汇聚和稽核，提升了各省元数据的质量。
- 元数据可视化，实现集团+省份 20W+数据资产的可查可见。
- 通过模型优化治理，模型复用度由 31%提升到 38%。
- 进一步完善升级数据治理规范 2.0，完成“1+15”本数据规范分册的编写，以此为基础开展数据质量、数据生命周期、敏感数据管理等专项治理工作。



图 9-2 运营商集团全网数据治理平台案例

9.3 金融机构数据治理

应用场景

在管理层面，基于亚信多年的数据治理实施经验，帮助客户建立了三层式的数据治理组织，制定了多项数据管理规范。在工具层面，基于 AISWare Data Go

数据资产管理平台，以元数据、数据标准、数据质量为主线，为客户构建统一数据治理平台，以盘点数据资产、明确数据责任人、统一数据标准、提升数据质量。

应用效果

- 在元数据、数据标准、数据质量方面，取得一定成效，满足监管要求，保障业务正常开展。
- 采集元数据 35000+，识别数据资产 2000+，开放数据资产：2000+，数据完整性 99.5%，数据合规性 99.6%，数据准确性：99.7%。



图 9-3 金融机构数据治理平台案例

9.4 大数据中心数据资源管理

应用场景

通过顶层规划和系统性设计，拟搭建 9 个子系统。亚信作为核心承建厂商，基于亚信的数据治理体系方法和经验，基于 Data OS 产品能力提供数据开发、治理及共享子系统能力。

应用效果

- 数据开发子系统构建 10000+任务，200 万+调度；数据治理子系统构建 1000+数据标准，200+开发规范；数据共享子系统构建 1000+开放接口，6000+共享模型。
- 为 10+委办、100 亿+条数据提供清洗服务；出具 100+质量报告、10+数据分析报告；构建 100+数据融合任务。
- 支撑健康码应用，完成 4.6 亿+疫情数据，9300 万+共享 2100 万+防疫数据上报。



图 9-4 大数据中心数据资源管理案例

9.5 钢铁集团主数据治理

应用场景

为客户构建“标准体系+管理体系+平台工具”的解决方案，从集团打造面向各子分公司的企业级主数据管理体系，除了提供平台建设，还为客户提供多年经验积累的主数据标准体系内容和管理体系内容。

应用效果

- 制定主数据标准 2000+个，集成主数据 17.56 亿条，开发主数据接口 33 个，制定规范流程 9 个，接入系统 19 个，治理活动覆盖 13 个子公司。
- 主数据完整性从 57.4%提升至 98.3%，主数据准确性从 44.1%提升至 97.8%，主数据一致性从 31.6%提升至 95.4%，主数据开放耗时从 2 天降低至 1 小时。

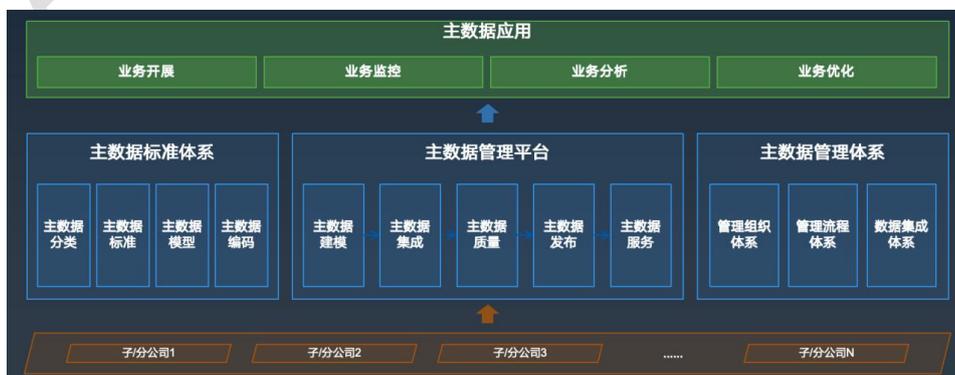


图 9-5 钢铁集团主数据治理案例

10 资质与荣誉

服务电信运营商包括集团公司、省分公司、专业公司的 20 余个客户，市场份额超过 50%，是中国移动集中化大数据中台的主要承建商。同时服务于金融、政府、交通等 40 多家不同行业的客户，累计为 300+项目提供大规模数据产品及服务。

- 落地点 40+家客户
 - 20 个运营商市场，其中移动 14 个、联通 3 个、电信 3 个；
 - 2 个运营商子公司；
 - 20+个垂直行业客户。
- 多项荣誉和奖项



图 10-1 荣誉与奖项

11 联系我们

亚信科技（中国）有限公司

地址：北京市海淀区中关村软件园二期西北旺东路 10 号院东区亚信大厦

邮编：100193

传真：010-82166699

电话：010-82166688

Email：5G@asiainfo.com

网址：www.asiainfo.com



Thank you

依托数智化全栈能力，创新客户价值，助推数字中国。



亚信科技（中国）有限公司保留所有权利