

# AISWare Data Infrastructure 产品

## 亚信科技数据基础设施平台产品 V6.1 白皮书

AISWare Data Infrastructure 是亚信科技全力打造的数据基础设施平台产品，数据基础设施平台作为企业数字化驱动和转型能力的基石，为了更好的支持业务，必须要向轻量化，微服务化和支持业务能力灵活构建的方面发展。在企业整体 IT 环境上云的背景下，继续为企业数字化转型提供大数据存储、计算、分析、管控能力，协助客户打造企业级的一体化大数据基础平台。

# 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”）。

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，即使亚信事先获知上述损失可能发生。

亚信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）

亚信科技创立于 1993 年，依托产品、服务、运营和集成能力，为电信运营商及其它大型企业客户提供业务转型及数字化的软件产品及相关服务，致力于成为大型企业数字化转型的赋能者。

根据弗若斯特沙利文的资料，我们是中国电信行业最大的电信软件产品及相关服务供应商，按 2017 年收益计，我们的市场份额为 25.3%。根据同一资料来源，我们也是中国电信行业最大的 BSS 软件产品及相关服务供应商，按 2017 年收益计，我们的市场份额为 50.0%。我们是中国第一代电信软件的供应商，从 20 世纪 90 年代开始与中国移动、中国联通和中国电信长期合作，支撑全国超过十亿用户。与电信运营商的长期合作关系让我们对电信运营商的 IT 及网络环境以及业务运营需求有了深度理解，使我们能够开发出拥有 500 多种任务关键型电信级软件的丰富的产品组合（软件产品主要面向电信运营商，对其业务运营至关重要），包括客户关系管理、计费账务、大数据、物联网及网络智能化产品。截至 2018 年 12 月 31 日，我们有 214 家电信运营商客户，包括中国移动、中国联通和中国电信的总部、省级公司、地市级公司、专业化公司和合营企业。

我们也正在积极拓展在中国非电信企业软件产品及相关服务市场的市场份额。凭借我们在电信软件产品及相关服务市场丰富的行业知识及专长及稳固的领导地位以及全方位、高度专业化的电信级产品图谱，我们相信我们也已经就解决各类企业，尤其是大型企业在业务转型与数字化方面与电信运营商相类似的、最为根本的需求占据了有利地位。截至 2018 年 12 月 31 日，我们有 38 家广电、邮政及金融、电网、汽车等行业的大型企业客户。通过资源、管理、专业知识及技术专长的共享，我们能够同时服务电信和非电信企业市场，凭借协同效应赢取新业务并保持竞争优势。

### 部分企业荣誉资质

ISO 9001 质量管理体系认证	2018 年中国软件业务收入前百家企业前 20 强
ISO 20000IT 服务管理体系认证	
信息系统集成及服务资质（一级）	2018 年中国电子信息行业社会贡献 500 强
CMMI 5 级（能力成熟度模型集成 5 级）认证	2018 年中国电子信息研发创新能力 50 强企业
国家规划布局内重点软件企业	

# 目录

一. 摘要.....	6
二. 缩略语与术语解释.....	8
三. 产品概述.....	9
3.1 趋势与挑战.....	9
3.2 产品定义.....	10
3.3 产品定位.....	10
四. 产品整体架构.....	11
五. 产品功能架构.....	12
5.1 产品整体功能概述.....	12
5.1.1 大数据基础平台功能架构.....	13
5.1.2 集群管控平台功能架构.....	18
5.1.3 实时计算引擎功能架构.....	21
5.1.4 高速查询引擎功能架构.....	23
5.2 基础功能.....	25
5.3 特色功能.....	31
5.3.1 大数据多租户能力模型.....	31
5.3.2 开放集成协议彰显架构优势.....	32
5.3.3 跨中心的异构集群云化.....	32
5.3.4 智能租户资源分配.....	32
5.3.5 分布式列式数据库增强.....	33
5.3.6 构建统一的数据湖能力.....	33
5.3.7 云边集群统一管控.....	34
六. 产品优势特性.....	35
七. 产品价值.....	37
八. 产品差异化优势.....	38
九. 应用场景.....	39
9.1 电信行业湖仓一体.....	39
9.2 多数数据中心的云边协同.....	41
9.3 通用型大规模分布式储算.....	43

十. 产品客户成功故事 .....	45
10.1 大数据平台升级及数据迁移方案 .....	45
十一. 资质与荣誉.....	49
11.1 大数据相关软件著作权.....	49
11.2 大数据相关专利.....	50
11.3 信通院认证 .....	51
11.4 国产化兼容适配.....	51
11.4.1 适配海光服务器 .....	51
11.4.2 适配华为TaiShan200服务器.....	51
11.4.3 适配银河麒麟服务器 .....	52
十二. 联系我们 .....	53

# 一. 摘要

从宏观形势来看，大数据在各行各业发挥了越来越重要的作用，已经逐步上升到国家战略层面。2020年10月29日，中国共产党第十九届中央委员会第五次全体会议审议通过了《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》，明确提出了“十四五”时期经济社会发展指导方针，这为做好未来五年经济社会发展工作指明了方向，提供了可遵循的原则。

数字化转型、智能化升级带来的需求，企业内部有降本增效的需求，企业向原生数字化企业发展，要实现业务数据化必须要规范建设大数据平台。商业生态环境无时无刻不在变化，企业也需要不断调整、扩展业务边界，加强生态合作，为数据集成、查询、实时分析等都带来需求。科技赋能5G、数字孪生、大数据、云计算、人工智能等，加速科技突破。大数据技术的不断突破创新，主要来源于厂商的主动性突破创新，例如湖仓一体、跨域跨源交互式查询等。

近年来大数据技术和应用迅猛发展，技术的不断演进，通过横向扩展，分布式集群部署方式比传统集中式架构性能更优，在数据平台架构云化重构、实时应用支撑、能力开放、智能运维等方面发挥了重要作用，为企业的大数据中心从大数据存储、计算、PaaS能力、运维、开放都有了飞跃式的发展。在此国家政策扶持、技术发展趋势良好的背景下，亚信紧跟开源大数据技术的发展，自研开发的大数据基础平台为电信行业，其它大企业提供完整的大数据平台解决方案，助力企业在大数据中心建设，发挥大数据价值方面保驾护航。

亚信 AISWare Data Infrastructure 是集大数据采集、数据转换、数据计算、数据存储、资源管控、运维一体化完整产品集和解决方案。

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 Infrastructure 产品。

## 二. 缩略语与术语解释

以下对文中提到的缩略语和专业术语，进行了解释说明。

表 2-1 缩略语与术语解释

缩略语或术语	英文全称	解释
CLI	Command-line interface	命令行界面
DIF	Data Infrastructure	数据基础设施平台
HDFS	Hadoop Distributed File System	Hadoop 分布式文件系统
KDC	Key Distribution Center	密钥分发中心
NTP	Network Time Protocol	计算机时间同步化的协议
OSB	Open Service Broker	纳管组件的接入接口标准
RBAC	Role Based Access Control	基于角色的访问控制
SQL	Structured Query Language	结构化查询语言
YARN	Apache Hadoop YARN	Hadoop 资源管理器

## 三. 产品概述

AISWare Data Infrastructure 帮助企业从存储、计算、分析、处理、管控、运维提供一体化的系统平台，立足于开源，提供专业的大数据服务、运维、咨询，协助客户打造企业级的高效安全的大数据平台，并为客户提供大数据平台智能化的集群管控运维、集群洞察工具，提供可视化、实时的分布式流数据开发处理平台。

### 3.1 趋势与挑战

国家大数据发展战略及政策为大数据发展保驾护航，大数据政策开始向各大行业和各细分应用领域延伸扩展，行业应用成为关注重点。受宏观政策环境，技术进步与数字应用普及渗透等多种因素影响。新基建政策提出加快第五代移动通信、工业互联网、大数据中心等建设。完善宏观经济治理政策指出提升大数据等现代技术手段辅助治理能力。发展战略性新兴产业政策明确推动互联网、大数据、人工智能等同各产业深度融合。

国家加强了对科技基础平台建设的支持，并且随着新型技术的发展，大数据平台迎来新的机遇；同时平台从百家争鸣进入到加剧竞争的阶段，市场对于产品化程度，技术深度，专业化都提出了新的要求。开源社区走向封闭：Cloudera 和 Hortonworks 合并，Hadoop 开源社区活跃度下降。Hadoop3.x 新特性质量，稳定性下降，从开源到商用之间的鸿沟不断增大。

在技术层面，以开源为主导、自主可控要求提升，大数据平台建设从粗放走向精细，从小规模走向大规模，从少量使用走向多租户。对技术上的挑战主要体

现在新技术特性的理解推广和使用，大数据平台规模上的增长，带来运维和性能调优的挑战。国产化，国产生态深度融合的要求逐步提升，带来对核心技术自主可控的强要求。

## 3.2 产品定义

数据基础设施平台（AISWare Data Infrastructure），基于完整的 Hadoop 技术生态，为企业提供多类型数据存储、流批数据计算，跨域跨源查询，实时检索，数据湖管理，多集群管控，租户资源分配以及智能化辅助运维等技术能力，并提供专业的大数据运维、咨询等服务，协助客户打造高效，安全的企业级大数据基础平台。

## 3.3 产品定位

AISWare Data Infrastructure 是亚信推出的数据基础设施平台，作为企业数字化驱动和转型能力的基石，为了更好的支持业务，必须要向轻量化，微服务化和支持业务能力灵活构建的方面发展。在企业整体 IT 环境上云的背景下，继续为企业数字化转型提供大数据存储，计算，分析，管控能力，协助客户打造企业级的高效的，安全的一体化的数据基础设施平台。

## 四. 产品整体架构

亚信大数据域产品集包括：

AISWare Data Discovery 数据探索分析平台

AISWare Knowledge Graph 知识图谱

AISWare Data OS 数据操作系统

AISWare Data Go 数据资产治理系统

AISWare Data Infrastructure 数据基础设施平台

AISWare BigData 产品体系中产品间关系及 AISWare Data Infrastructure 位置如图所示：



图 4-1 亚信大数据域产品集总体架构

## 五. 产品功能架构

本章节重点阐述产品的功能体系架构，对每个功能模块进行了详细的描述。

### 5.1 产品整体功能概述

亚信 AISWare Data Infrastructure 数据基础设施平台整体技术架构如图所示，其中包括大数据基础平台、集群管控、实时计算引擎、高速查询引擎，共四部分。

大数据基础平台是亚信推出的商业版本大数据平台，实现为租户提供统一的资源、运维管理及多种常用的容器化组件，并根据客户需求提供完整的计算、存储及组件使用环境，以支撑客户上层业务运营需求。

集群管控平台以多租户管理核心，面向企业实现大数据集群资源管控，实现大数据平台的租户能力开放管理能力，同时提供大数据集群资源、性能、安全的深度洞察和智能运维能力，全面保障大数据集群的合理部署和持续优化，达到充分利用资源的目的。

实时计算引擎实现多数据源的实时融合计算，提供实时业务所需要的计算场景。提升对于实时流数据的读取、处理和输出能力。

高速查询引擎应对大数据平台跨多种数据源的交互查询要求，屏蔽底层多类型跨数据源查询的复杂性，实现按租户权限为客户提供标准 SQL 的交互式高效查询场景。

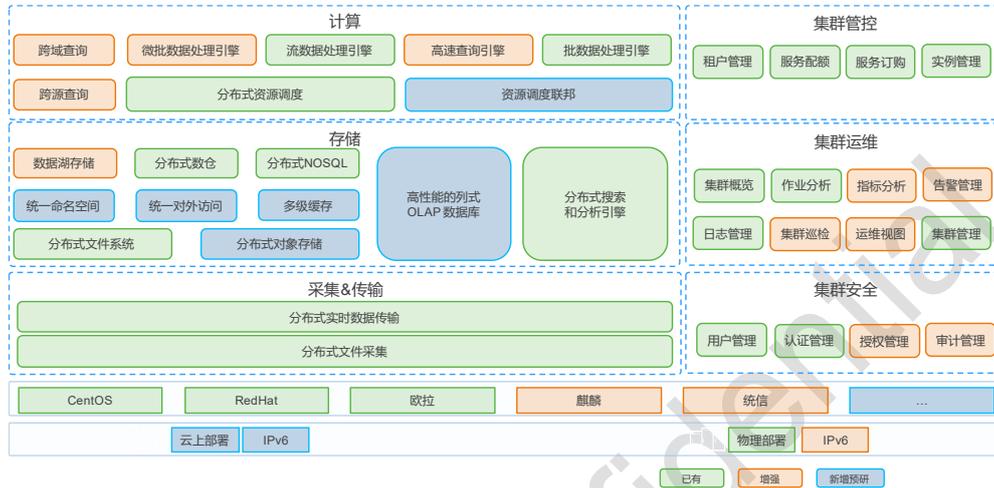


图 5-1 AISWare Data Infrastructure 产品功能架构

### 5.1.1 大数据基础平台功能架构

大数据基础平台具备管控、运维、监控及安全的完备平台体系，提供安全可靠、运维便捷、支持国产化的大数据储算平台。为企业提供统一的多类型数据存储能力，支持数据湖及数据仓库的建设，满足大规模数据存储诉求。满足从离线处理到在线处理（秒级到毫秒级）、交互式查询等数据处理诉求，满足企业业务发展针对数据分析决策，支撑企业数字化能力。

提供一套完整的基于分布式文件系统的海量数据采集、存储、计算处理及运维的综合基础平台，采用分布式文件系统、列存储或混合存储、压缩、延迟加载等技术，只需要较为廉价的硬件设备投入即可提供对海量数据的存储能力。

采用分布式调度和资源管理技术，保证分布式并行运算的安全、高效和可靠；通过列存储引擎，提供列式数据的实时存取，满足实时应用对大数据的读、写能力；采用交互式类 SQL 语句完成分析查询功能，提供快速、精准的数据多维分析功能，免去程序开发的复杂性，降低开发的难度，满足离线分析型应用对大数据的处理要求。

用户可以快速的搭建起自身的企业级数据基础设施平台，并在其上开展自身的数据分析业务。数据基础设施平台作为开放的基础平台产品，为用户提供了强大而丰富的平台能力，如数据存储能力、数据查询能力、数据计算和多维分析能力、索引分析能力等。



图 5-2 大数据基础平台功能架构

- 分布式文件系统

HDFS（Hadoop Distributed File System）是 Hadoop 的分布式文件系统。HDFS 集群主要由 NameNode（管理者）和多个 DataNode（工作者）组成。NameNode 用来管理元数据，DataNode 用来存储真实数据。用户可以通过 NameNode 与文件元数据建立联系或修改文件，并且通过 DataNode 直接访问实际文件的内容。

支持 Federation，在集群中将会有多个 NameNode，他们之间相互独立且不需要互相协调，各自分工，管理自己的区域。增加多个备用 NameNode 技术，降低单个备用 NameNode 瘫痪带来的集群管理风险。引入了纠删码技术，与三副本策略相比，提高 50% 以上的存储利用率。

- 分布式资源调度

在 Hadoop 中每个应用程序被表示成一个作业，每个作业又被分成多个任务。JobTracker 是一个后台服务进程，启动之后，会一直监听并接收来自各个 TaskTracker 发送的心跳信息，包括资源使用情况和任务运行情况等信息。它的主要功能是：作业管理、状态监控和任务调度等。

在 YARN 中，有一个全局的资源管理器（ResourceManager）和每个应用程序的应用程序管理器（ApplicationMaster）。ResourceManager 和每个节点（NodeManager）组成了处理数据的框架，ResourceManager 是整个系统资源的最终决策者。每个应用程序的 ApplicationMaster 从 ResourceManager 获得资源，并在 NodeManager 上执行和监控任务。

- 批数据处理引擎

Hive 是建立在 Hadoop 上的数据仓库基础构架。它提供了一系列的工具，可以用来进行数据提取转化加载，可以对存储在 Hadoop 中的数据查询和分析。Hive 定义了简单的类 SQL 查询语言，称为 HQL，它允许熟悉 SQL 的用户查询数据。

Spark 是一个大数据分布式编程框架，它使用函数式编程方式扩展了 MapReduce 模型，从而支持更多计算类型。Spark 不仅实现了 MapReduce 的 map 函数和 reduce 函数及计算模型，还提供更为丰富的算子，如 filter、join、

groupByKey 等。Spark 使用内存缓存来提升性能，缓存同时提升了迭代算法的性能，这使得 Spark 非常适合数据处理任务。

- 实时数据处理

Spark Streaming 是 Spark 核心 API 的一个扩展，可以实现高吞吐量的、具备容错机制的实时流数据处理。Spark Streaming 将接收到的实时数据，按照一定时间间隔，对数据进行拆分，交给 Spark Engine 引擎，最终得到一批批的结果。

Apache Flink 是一个分布式处理引擎，用于对无界和有界数据流进行有状态计算。Flink 可以在所有常见的集群环境中运行。

Kafka 是一个低延迟高吞吐的分布式消息队列，适用于消费离线和在线的消息，用于收集和发送事件和日志数据。Kafka 通过副本来实现消息的可靠存储，同时消息间通过特有机制来确认消息的落地，避免单机故障造成服务中断。同时副本也可以增加扇出带宽，支持更多的下游消费者订阅。

- 分布式列式数据库

HBase 是一个高可靠性、高性能、面向列、可伸缩的分布式存储系统，利用 HBase 技术可在廉价的设备上搭建起大规模结构化存储集群。在 Hadoop 生态系统中，HDFS 为 HBase 提供了高可靠性的底层存储支持，Zookeeper 为 HBase 提供了稳定服务和 failover 机制。

- 数据湖存储

数据湖可以支持多类型存储媒介同时存在于数据湖当中，湖和多仓数据存储架构，实现了跨湖、跨仓统一的元数据管理及协同分析的融合集约化数据存储与管理的解决方案。

Hudi (Hadoop Upsert Delete and Incremental) 是基于湖仓一体的表映射层技术，对多种格式的文件提供基于表的管理能力，如表版本控制、事务支持、表的增删改、表结构的动态调整、多种查询引擎的支撑。

- 主机操作系统网络适配

支持基于 x86 架构的国产 CPU (海光)，支持基于 ARM 架构的国产 CPU (鲲鹏、飞腾)。支持国产操作系统 (麒麟、欧拉、BC\_Linux)。

产品适配 IPv6 网络，现有网络从 IPv4 切换到 IPv6 后，通过调整配置即可正常使用。

- 统一管理与运维

对 Hadoop 平台组件的可视化一键部署，实现集群配置的统一管理与维护，实现对 Hadoop 平台告警及通知。

用户使用已经创建好的平台账号并登录系统，在菜单栏上切换不同的功能模块 (首页、服务、主机、告警、后台操作、退出登录)，成功进入系统后，用户能够监控集群运行的各种指标，如 CPU、内存、网络、磁盘 I/O 等。

在平台首页，可以查看集群主机的运行状况，查看集群服务的版本号及各服务运行的健康情况，对集群全部服务和单个服务进行启、停等操作。在服务的详情页中，可以查看服务的组件运行状况和对应的设置。

提供主机管理的功能，可以查看集群所有主机的信息和安装的组件情况，同时也可以对主机进行管理操作。在主机的详情页面，查看主机服务运行状况和重要指标数据。

### 5.1.2 集群管控平台功能架构

集群管控平台实现了分布式集群 Hadoop 的多集群接入管控能力，同时提供面向租户分配订购集群资源、多集群洞察分析、集群资源和作业的智能化分析能力。实现多集群的安全管理、授权管理和统一认证机制。

集群管控平台在功能架构上主要分为集群的接入单元和管理运维单元，其中管理运维单元中主要提供管理、运维和分析能力，集群的接入单元实现多集群组件的统一接入和监管能力。



图 5-3 集群管控平台功能架构

- 集群资源管理

集群资源管理为整个系统的核心功能，实现大数据平台的资源分配及线上申请两种模式。租户管理员可以根据子租户租户需求分配资源，对于租户资源使用不均衡，可以根据使用动态调整。

在租户资源分配后，租户管理员可根据需求创建服务实例，服务实例的资源配额不能超过该租户的资源额度。租户成员需要提出申请，租户管理员审批通过后，实例变更生效。租户管理员或者成员均可删除该实例。删除实例时要做风险提示，删除后该实例物理删除，需要慎重操作。

- 多租户管理

构建树形多租户模型，逐级适配不同的管理模式。支持租户的生命周期管理和时间到期配置。实现 RBAC（Role Based Access Control，基于角色的访问控制）的方式来实现用户的权限管控。提供集群的用户统一管理、集群组件的权限管理和基于 Kerberos 的认证管理。

其中权限管理进一步实现大数据组件 HBase、HDFS、Hive、Kafka、MapReduce2、Spark 的细粒度权限管控功能，满足细分权限的分配，满足租户对权限控制的需求。对于每个服务实例均支持多条权限策略。支持对多个用户的授权功能。

**表 5-1 细粒度权限**

服务名称	权限控制粒度	权限
HBase	命名空间	支持对 HBase 的命名空间的细粒度权限管理，支持表、列簇、列的授权，权限包括：查询、更新、创建、删除等。

服务名称	权限控制粒度	权限
HDFS	文件目录、文件数	支持文件路径的权限管理，包括文件的读/写/执行。
Hive	数据库、表	支持数据库的细粒度权限管理，实现表、列的授权，权限包括查询、更新、创建、删除、修改。
Kafka	主题	支持 Kafka 实例主题的细粒度权限管理，权限包括：发布、消费、配置、创建、删除等。
MapReduce2	队列	支持队列权限管理，包括：队列管理、队列查询。
Spark	文字内容队列	支持队列权限管理，包括：队列管理、队列查询。

- 集群洞察分析

集群洞察分析可以对集群进行深度分析，以提升集群使用效率。其中包含：作业监控、日志管理、告警管理和运维巡检。

提供租户维度的资源视图，显示按照服务配额参数显示占用情况，并呈现服务实例的实际使用量并显示明细情况。使平台管理员及租户管理员能及时了解集群资源使用情况，对集群资源动态调整，提升集群资源利用率。

- 告警管理

统一告警管理作为对系统运行异常时的一种通知机制，需要在系统异常时，能够及时有效通知管理员以便其进行介入，及时处理、解决问题。

告警场景千差万别，所以需要针对不同场景分别定义不同的通知策略以及通知内容，提高管理员的运维效率。

#### ➤ 运维巡检

保障集群正常运行是集群管理人员和运维人员的主要工作内容，运维内容包括主机、服务、组件、数据库等服务的维护。在平台本身附带的一些监控和管理能力之外，运维管理员需要更加专业、全面的运维工具以辅助其高效完成运维工作。

针对传统的线下巡检方式，不仅消耗人力成本，同时受限于集群规模，而且耗时很长，不仅效率低，也无法标准化巡检内容。所以提供一套自动运维巡检服务，解决以上所有问题。

#### ● 集群接入

基于 Open Service Broker（纳管组件的接入接口标准，以下简称：OSB）标准并且实现服务的自动接入、自动注册，并支持订购界面的自动生成，实现云边、异构、不同版本的大数据平台资源管控。并规范统一的接入标准，满足物理多租及逻辑多租组件的接入管理。

### 5.1.3 实时计算引擎功能架构

实时计算引擎为企业级用户提供统一的分布式流数据计算的平台，能够满足实时数据的接入、计算、订阅，能够全方位满足实时类业务场景的诉求。

实时数据服务：依据实时业务场景的特点，提供个性化数据订阅和数据推送能力。

实时数据处理：主要对流式数据进行业务逻辑运算，包括：字段计算、多流合并、维度汇总和复杂事件处理。



图 5-4 实时计算引擎功能架构

- 流字段计算

通过高速分布式缓存 Redis Cluster 完成上网类、位置类、订单缴费类等流数据计算，运算速度快，高并发，高吞吐。

- 实时汇总计算

实时的数据汇总计算，满足多指标多维度并行计算，支持 count、sum 等。

- 多流合并计算

提供多种流数据合并的计算能力，例如：位置流和内容流的实时 join 场景，完全基于 Flink、Spark 内存机制，提供时间窗口设定机制，规避时序性、延迟性。

- 实时数据服务

提供数据订阅和数据推送功能，为用户提供便捷的实时数据服务。

- 实时数据订阅

- 表字段级别的细粒度订阅。
- 提供订阅周期：随机、每天、每周、每月和每批次。

- 实时数据发布。
- 统一化的流数据定义、共享、订阅的全周期管理。

### 5.1.4 高速查询引擎功能架构

基于开源分布式 SQL 查询引擎 Presto 构建产品。支持从多种数据源获取数据，一条标准 SQL 查询可以将多个数据源的数据进行合并查询分析，屏蔽底层多类型组件语言的差异性，实现跨数据源数据分析。

实现跨数据源的大规模的交互式查询分析，主要异地多集群的融合查询。支持标准 SQL 查询，对于不同的数据源基于标准 SQL 即可查询，通过标准 SQL 可以实现两个数据源的融合查询。并实现集群的多租户管理，提供多类型的对外访问接口。

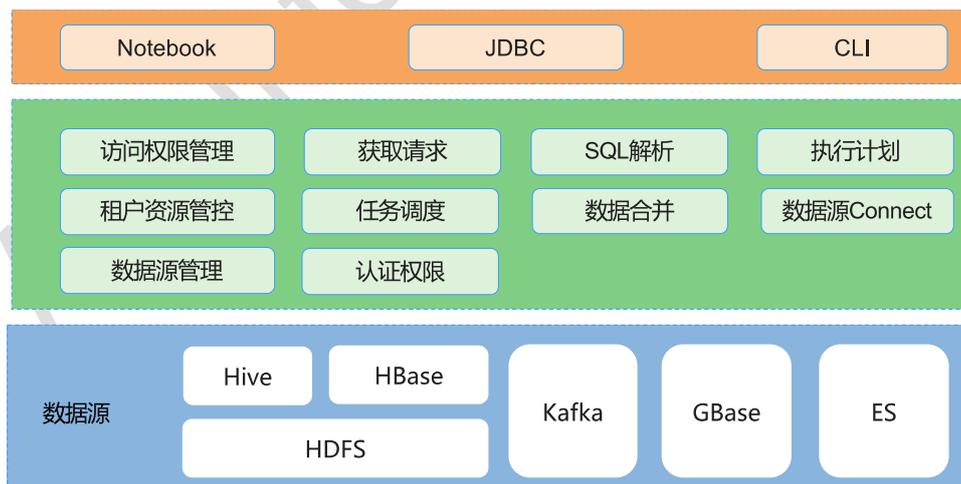


图 5-5 高速查询引擎功能架构

- SQL 解析功能

支持标准 SQL 的语义语法分析和执行，提供生成对应的 SQL 执行计划，并对其分段和调度功能，分配到对应的执行节点，与数据源连接，最终获得执行结果，根据分段执行的接口进行合并处理。

- 权限管理

用户权限管理，实现与权限组件的用户同步，并支持用户的变更及删除操作。

数据源权限管理，支持数据源的权限、Schema 的权限及表权限的管理，并支持安全策略的管理，实现策略的创建、变更及删除。

- 数据源支持

支持 Elasticsearch、GBase、Hive、Kafka、MySQL、PostgreSQL 等常用数据源。

支持湖存储数据源的配置和数据查询，同时满足与其他数据源进行关联查询。

- 访问方式支持

提供 JDBC、CLI（Command-line interface 命令行界面）及 Notebook 三种访问方式。

Notebook 访问可以支持 SQL 语句的输入及执行，并显示执行的记录，对于小数据量可支持查看及预览，大数据量的可以提供下载功能。

- 多租户资源隔离

提供基于 YARN 模式下租户统一的资源管控能力，支持租户间高速查询服务的计算资源共享能力，从而实现多租户模式下，租户资源、权限、任务的隔离。

## 5.2 基础功能

以下是数据基础设施平台产品的基础功能列表：

**表 5-1 基础功能**

功能点	功能点描述
分布式数据存储	使用 Apache 社区的 HDFS，支持分布式数据存储的各种操作，包括： <ul style="list-style-type: none"> <li>• 分布式数据存储，数据一致性；</li> <li>• 分布式文件访问，文件读取；</li> <li>• 分布式文件系统高可用；</li> <li>• 数据均衡、存储策略、压缩存储。</li> </ul>
分布式资源管理	使用 Apache 社区的 YARN，支持分布式资源管理，包括： <ul style="list-style-type: none"> <li>• 集群资源管理；</li> <li>• 集群资源调度；</li> <li>• YARN 服务框架；</li> <li>• 应用优化调度与处理。</li> </ul>
集成 Hive	使用 Apache 社区的 Hive，提供数据仓库 Hive 能力，包括： <ul style="list-style-type: none"> <li>• 支持 DDL、DML、DQL 操作；</li> <li>• Hive Warehouse Connector 支持 Spark；</li> <li>• 元数据、权限管理和配置。</li> </ul>

功能点	功能点描述
集成 HBase	使用 Apache 社区的 HBase，支持列式数据库，包括： <ul style="list-style-type: none"> <li>• HMaster 管理，异步 RPC 机制；</li> <li>• RegionServer 管理、负载均衡，逻辑分组；</li> <li>• 单节点支持部署多个 RegionServer。</li> </ul>
集成 Phoenix	提供 SQL 层访问 HBase 数据，包括： <ul style="list-style-type: none"> <li>• 数据查询操作界面、命令；</li> <li>• 数据查询操作转换；</li> <li>• 列编码与日志。</li> </ul>
集成 Spark	使用 Apache 社区的 Spark，支持 Spark 的各种操作，包括： <ul style="list-style-type: none"> <li>• 缓存持久化；</li> <li>• 元数据检查点、元数据管理；</li> <li>• 数据检查点，数据批处理，时间片管理；</li> <li>• 流输出、监控、监听接口；</li> <li>• 集成 Kyuubi，提供统一的多租户 JDBC 接口。</li> </ul>
集成 Kafka	使用 Apache 社区的 Kafka，提供分布式消息系统，包含： <ul style="list-style-type: none"> <li>• 分布式架构，并支持逻辑分区；</li> <li>• 数据多副本；</li> <li>• 数据生产和消费；</li> <li>• 消息传送机制；</li> <li>• 增强 Kafka 的监控能力。</li> </ul>
集成 Flink	使用 Apache 社区的 Flink，提供的核心功能包括： <ul style="list-style-type: none"> <li>• 提供流式计算和批量计算，支持多流 join 和数据转换；</li> <li>• 时间和计数窗口功能，提供多种窗口类型；</li> </ul>

功能点	功能点描述
	<ul style="list-style-type: none"> <li>• 提供多种内置计算函数，自定义函数接口；</li> <li>• 复杂事件处理；</li> <li>• 支持有状态计算的 Exactly-once 语义。</li> </ul>
集成 Elasticsearch	使用 Apache 社区的 Elasticsearch，提供分布式索引系统，包括： <ul style="list-style-type: none"> <li>• 支持高可用，自动节点恢复，自动数据平衡；</li> <li>• 提供多种分析器、分词器、筛选器及扩充选项转化原始数据；</li> <li>• 提供全文检索，结构化检索，数据分析；</li> <li>• 与 Hadoop 作业交互，支持快照及恢复，支持小文件合并。</li> </ul>
集成 Hudi	使用 Apache 社区的 Hudi，支持的核心功能包括： <ul style="list-style-type: none"> <li>• 基于事务的记录行级增加、修改和删除；</li> <li>• 写时复制模式和读时合并模式同时支持；</li> <li>• 小文件自动控制机制；</li> <li>• 数据版本和历史版本回溯支持；</li> <li>• 多端并发写入控制；</li> <li>• 增量视图、快照视图的读取能力。</li> </ul>
底层基础组件能力	提供底层基础能力，包括： <ul style="list-style-type: none"> <li>• 集成 Ranger，支持认证、授权、审计等；</li> <li>• 组件支持操作系统的 Kerberos 认证；</li> <li>• 集成 NTP 服务；</li> <li>• 集群升级、部署优化。</li> </ul>
集群负载分析	<ul style="list-style-type: none"> <li>• 集群 CPU、内存、存储的监控；</li> <li>• 集群作业的监控；</li> <li>• 集群负载监测分析。</li> </ul>

功能点	功能点描述
多租户管理	租户模型管理，及支持租户的入驻，包括： <ul style="list-style-type: none"> <li>• 租户的生命周期管理，包括创建、变更及删除；</li> <li>• 树形多租户模型，租户到期时间配置；</li> <li>• 租户成员管理，支持三种角色。</li> </ul>
服务资源配额	面向租户按需分配资源，包括： <ul style="list-style-type: none"> <li>• 资源配额或申请；</li> <li>• 资源配额变更；</li> <li>• 资源配额删除。</li> </ul>
服务实例	租户根据应用需求创建所需实例，包括： <ul style="list-style-type: none"> <li>• 服务实例创建或申请；</li> <li>• 服务实例变更；</li> <li>• 服务实例删除。</li> </ul>
细粒度权限管控	支持大数据组件的细粒度权限管控，包含： <ul style="list-style-type: none"> <li>• HDFS 细粒度权限；</li> <li>• Hive 细粒度权限；</li> <li>• HBase 细粒度权限；</li> <li>• Spark 细粒度权限；</li> <li>• Kafka 细粒度权限。</li> </ul>
大数据平台多租户管理	实现大数据平台组件的多租户功能： <ul style="list-style-type: none"> <li>• HDFS 基于 Namespace 实现多租户管理；</li> <li>• Hive 基于 Database 实现多租户管理；</li> <li>• 基于 YARN queue 实现多租户管理；</li> <li>• HBase 基于 Namespace 实现多租户管理；</li> <li>• Spark 基于 YARN 实现多租户管理；</li> </ul>

功能点	功能点描述
	<ul style="list-style-type: none"> <li>• Kafka 的 Topic 多租户管理。</li> </ul>
服务及工具接入管理	开放的服务接入框架，支持组件及工具的接入，包括： <ul style="list-style-type: none"> <li>• 服务接入、注册和变更；</li> <li>• 服务下线和日志。</li> </ul>
运营支撑	支持租户分析及实际用量提醒，包括： <ul style="list-style-type: none"> <li>• 租户级基础统计、服务实际用量分析，图表展示；</li> <li>• 对服务实例的实际使用量提醒，自定义阈值；</li> <li>• 组件的计量计费能力，提供定价、订购及账单。</li> </ul>
集群可视化	<ul style="list-style-type: none"> <li>• 支持指标多样化监控，如 CPU /内存/网络/磁盘等；</li> <li>• 支持查看集群主机和服务的健康状况，版本号展示；</li> <li>• 支持全服务进程的启停等操作；</li> <li>• 支持查看服务的组件运行状况和设置。</li> </ul>
大数据平台多 KDC 统一纳管	<ul style="list-style-type: none"> <li>• 统一入口提供账号的可视化管理界面；</li> <li>• 支持接入多个 KDC 服务，展示不同集群与 KDC 的映射关系；</li> <li>• 多 KDC 的 Principal 的统一管理，账号与 Principal 映射关系；</li> <li>• 实现 keytab 的创建、下载、分享等功能。</li> </ul>
基础组件能力增强	<ul style="list-style-type: none"> <li>• 支持麒麟操作系统；</li> <li>• Spark 认证和权限控制，满足多租户应用场景；</li> <li>• 统一管理集群主机的时钟同步，保证时钟一致；</li> <li>• 支持 IPv6，并提供 IPv6 环境下的集群升级。</li> </ul>

功能点	功能点描述
高速查询引擎功能增强	<ul style="list-style-type: none"> <li>• 支持 GBase 数据源的配置和接入，满足基本的查询与使用；</li> <li>• 支持 On YARN 模式的资源共享；</li> <li>• 提供分布式跨域查询能力，整合多域异构数据源，实现统一的对外接口。</li> </ul>
云边管控功能增强	<ul style="list-style-type: none"> <li>• 统一告警，提供基于静态阈值、智能阈值的告警管理能力；</li> <li>• 统一巡检，支持 Hadoop 集群主机及组件的巡检能力，支持日常巡检集群日常维护、健康状态，发现集群潜在风险。</li> </ul>

## 5.3 特色功能

本章节总共阐述了七大特色功能。

### 5.3.1 大数据多租户能力模型

大数据平台包括组件众多，面向租户资源分配不统一，创建操作比较繁琐，需要功能通过客户端、管理界面、Ranger 来实现，随之租户数量的增多，此管理方式成为租户运营的障碍。

大数据 Hive、HBase、HDFS、Kafka、MapReduce2、Spark 组件的多租户模型，通过资源管理能力面向租户开放。

表 5-1 租户配额

服务名称	租户模型	配额参数
Hive	基于 Database 实现多租户管理	数据库大小。
HBase	基于 Namespace 实现多租户管理	最大表数量、region 数。
HDFS	基于 Namespace 实现多租户管理	最大文件数、存储量。
Kafka	实现 Topic 的多租户管理	Topic 分区大小，分区数。
MapReduce2	基于 YARN queue 实现多租户	内存大小。
Spark	基于 YARN queue 实现多租户	内存大小。
Presto	基于 YARN queue 实现多租户	内存大小。

### 5.3.2 开放集成协议彰显架构优势

采用行业通用的 Open Service Broker（纳管组件的接入接口标准，以下简称：OSB）协议实现大数据平台组件的纳管，基于该标准协议集群管控平台可以实现对于不同厂商、不同版本的接入，并且可以支持云边集群资源的统一管控。

提供外部接入的代理接口，实现自动接入，并且在集群管控中自动生成订购界面，待接入组件只需要提供基于 OSB 协议的各类接口定义即可，无需其他开发工作。通用的 OSB 协议，不但可以接入大数据平台组件，也可以实现其他第三方组件及工具接入。

### 5.3.3 跨中心的异构集群云化

大数据云化平台管理，主要解决云化平台上的基于多中心，多云平台上的多集群管理和运营。未来的企业云架构会是一个物理云、虚拟云和容器云混搭的架构，并且基于 5G 的云边架构，形成云端和边缘短的分级云平台。云化大数据平台适应这种基础设施的变化，也会统一整合跨中心、多平台、多集群模式下的大数据平台能力。

### 5.3.4 智能租户资源分配

集群资源利用打分：大数据集群上的作业运行在不同的队列上，会出现队列忙、闲不均衡的情况，属于资源分配不合理，如果能对它们整体合理调度，可以提升集群的资源利用率。该功能综合集群的资源使用率、任务完成率、等待率进行建模,最后的给集群资源利用打分。

AI 模型计算资源分配额度：按照流计算、高优先级、一般队列进行分类，参考资源使用情况、作业等待情况等指标进行优化。该功能需要收集前期作业的运行情况，通过人工智能的算法，不断学习迭代，给出最优化的租户资源分配方案，供管理员选择。

### 5.3.5 分布式列式数据库增强

采用 HBase 构建分布式列式数据库，实现大规模数据可靠的分布式读写。全面兼容开源 Apache 社区高版本的 HBase 组件，并进行大量的兼容性测试、代码适配和安装脚本配置文件增补完善，提供简便的一键式安装包。同时支持适配 x86、ARM 架构的 CPU 单独部署或混合部署方案，支持 CentOS 等主流操作系统。支持单节点部署多 RegionServer 模式，提高机器资源利用效率。具有分布式线性扩展能力，可提供海量级数据存储能力。

### 5.3.6 构建统一的数据湖能力

数据湖是基于 Hudi 组件来构建的统一数据存储池，可对接多种数据输入方式，存储大规模的结构化、半结构化、非结构化数据。它与 Hadoop 生态无缝对接，采用 Flink、Spark 等直接进行数据处理和分析，打破孤岛，洞察业务价值。支持多种数据类型和入湖模式，并提供数据统一清洗。提供企业级数据权限管理，用户可以针对库、表、字段分别设置权限。通过集成 Hive Metastore，提供对外开放服务，便于集成。因此一份数据支持多种分析，是数据湖最大的特点。

- 集群管理集成 Hudi 组件，统一安装、配置及管理。
- 批处理和流处理，支持 Update/Delete 记录，更高效处理数据。

### 5.3.7 云边集群统一管控

大数据的集群管控平台由多集群资源管控向云边协同演进，重点实现云边集群的资源管控、云边集群协同及云边任务调度系统能力。云边集群统一集中运营，实现大数据资源的按需分配，支撑租户需求。云边集群资源统一调度，建立统一的资源调度中心。云边集群跨源查询，提供融合查询能力，支持跨集群的数据查询。

- 多资源池管理，服务实例统一创建、修改、删除。
- 资源池视图、实例批量管理、与资源池间一键同步。

## 六. 产品优势特性

数据基础设施平台产品具备以下优势特性：

- 自主可控支持国产，兼容社区标准

核心技术自研兼容社区标准，支持 x86 和 ARM 架构，适配海光、鲲鹏、飞腾等国产 CPU，兼容欧拉、麒麟、统信等国产操作系统。

- 湖仓一体存储，构建多存储数据湖仓

支持结构化、非结构化及半结构化数据的存储，并支持湖仓一体化存储，并使用纠删码技术节省存储空间，提升存储效率。

- 专业计算引擎支持，助力大数据应用

提供批量查询、实时查询、交互式、点查、全文检索等不同类型的计算任务。提供标准 SQL 查询，支持多租户、安全管理等能力。

- 集群资源管控，支持云边协同管理

以“多租户为核心”实现平台资源管控及运营体系，“按需分批，即分即用”，提供用户、认证、授权及审计管理，保障集群健康稳定。

- 集群洞察分析，保障集群稳定监控

提供大数据集群性能分析、巡检、告警、日志检索、安全管理功能，助力集群健康稳定。

- 专业服务团队，专家运维服务

专业研发及运维团队提供专家服务，为客户提供安装、管控、运维及优化一体化的大数据平台解决方案及高级技术支持服务。

## 七. 产品价值

数据基础设施平台产品具备以下产品价值：

- 多类数据，融合存储

为企业提供统一的多类型海量数据存储能力，支持数据湖及数据仓库，满足数据存储诉求。

- 灵活分析，绽放价值

满足分钟级、秒级到毫秒级的数据处理诉求，提供企业发展所需的数据分析决策。

- 完备体系，健康安全

基于开源自主可控，具备管控、运维、监控及安全的完备平台体系，提供了安全可靠的大数据储算平台。

- 运营管控，效率提升

具备集群资源面向租户的分配及运营能力，提升平台的资源使用效率。

## 八. 产品差异化优势

本章节以表格形式，详细阐述了实时计算引擎、高速查询引擎、大数据基础平台、集群管控平台的各自优势。

表 8-1 产品差异化优势

名称	差异化优势
实时计算引擎	<ul style="list-style-type: none"> <li>• 自主自研实时数据处理引擎，支持多数据源融合计算。</li> <li>• 支持 Flink 及 Spark 架构，支持租户隔离需求。</li> </ul>
高速查询引擎	<ul style="list-style-type: none"> <li>• 自主研发，屏蔽底层多类型跨数据源查询的复杂性。</li> <li>• 基于 YARN 部署，可与 Hadoop 集成提升整体资源利用率。</li> <li>• 支持接入 GBase 数据库，并提供数据查询能力。</li> <li>• 提供跨域查询能力，实现统一的对外接口。</li> </ul>
大数据基础平台	<ul style="list-style-type: none"> <li>• 基于 HDFS 提供分布式文件存储，并提供纠删码等特性。</li> <li>• 对麒麟操作系统的支持。</li> <li>• Spark 认证和权限，满足多租户场景。</li> <li>• NTP 服务的统一管控，保证时钟一致。</li> <li>• 基础组件支持 IPv6，同时提供 IPv4 向 IPv6 升级方案。</li> </ul>
集群管控平台	<ul style="list-style-type: none"> <li>• 多集群多租户管理，提供洞察分析、自动运维和日志分析能力。</li> <li>• 多集群的统一用户管理、认证、授权及审计管理。</li> <li>• 提供基于静态阈值、智能阈值的统一告警管理能力。</li> <li>• 提供 Hadoop 集群主机及组件的统一巡检能力，提前发现集群潜在风险并及时处理。</li> </ul>

## 九. 应用场景

本章节介绍了产品典型的应用使用场景，包含：湖仓一体、多数据中心的云边协同、大规模分布式储算能力场景。

### 9.1 电信行业湖仓一体

电信行业数据分析的业务场景越来越复杂如实时场景化营销，准实时快速决策，B/O/M 多域融合分析等，对海量数据分析的规模，关联性，时效性提出更高要求，因此需要数据湖和数仓从技术和数据分析两个方向走向融合，构建湖仓一体架构来满足需求。

湖仓一体解决方案，实现多类型存储技术架构并存的湖中有仓体系，数据湖具备事务性数仓能力，发挥其高效的跨湖、仓的多领域数据融合分析能力，驱动客户数据的内在价值挖掘。

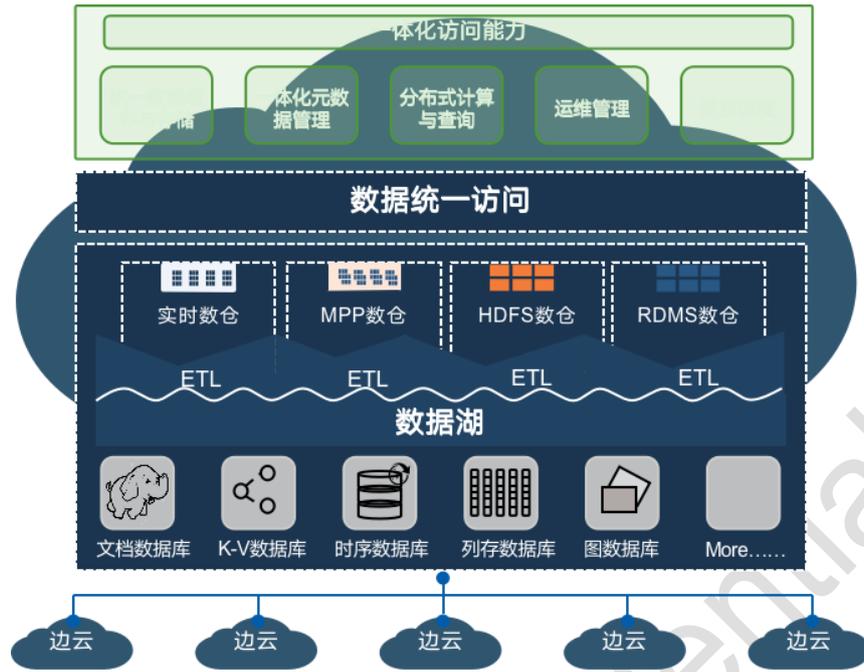


图 9-1 电信行业湖仓一体

- 高时效性能力

利用数据湖内的数仓事务性能力，在湖内数据分析时，实现数仓所达到的一致性、准确性、时效性。

- 全量数据低成本存储能力

在保障全量数据都能够存储的前提下，可大大降低海量数据的存储成本，实现存储资源弹性扩容。

- 跨湖仓的复杂查询能力

屏蔽底层跨湖、仓复杂查询条件，对外提供统一的标准 SQL 的关联查询。

- 全面的湖仓运维管理能力

提供对湖和仓统一的自动化安装部署、监控告警及故障快速定位、安全管理等可视化服务，全面兼容社区开源组件。

## 9.2 多数据中心的云边协同

国家的“东数西算”战略，推动相关大数据 IT 领域，基础软件，信息通信，土木建设，绿色能源等相关产业的大力发展。而大数据 IT 领域和基础软件领域中云边协同架构技术，非常符合国家战略和市场经济发展的技术要求。

云边协同是打造物理分散，逻辑统一的全网大数据平台的总体架构。通过在云端进行全网资源的统一调度协同，将更多简单的计算，实时性高的计算任务向边缘侧倾斜，并与数据的云边协同集成，支撑实时性要求高的典型云边协同应用场景。



图 9-2 多数据中心的云边协同

- 云边协同 PaaS 平台（多租户管理）

提供一致性的多租户体系，负责租户自身管理，租户相关的资源、工具的分配、回收、监控等管理。

- 云边协同大数据中心

负责云端复杂计算任务的执行，高价值数据的集中存储。统一的资源调度中心，负责云端和边缘端资源的协同分配，边缘侧资源的监控，管理及分析。

- 边缘端

接收云端任务，执行边缘侧简单计算任务、实时计算任务和临时数据存储。反馈计算执行结果，向云端反馈相应监控数据。

## 9.3 通用型大规模分布式储算

亚信的大数据分布式储算是一个通过大量实践，实现集中管控、便捷运维、稳定高效的商业化解解决方案。为客户实时数据、海量离线数据的分析场景提供计算、存储的资源服务，使客户利用最小的资源投入，获取最大化的回报。

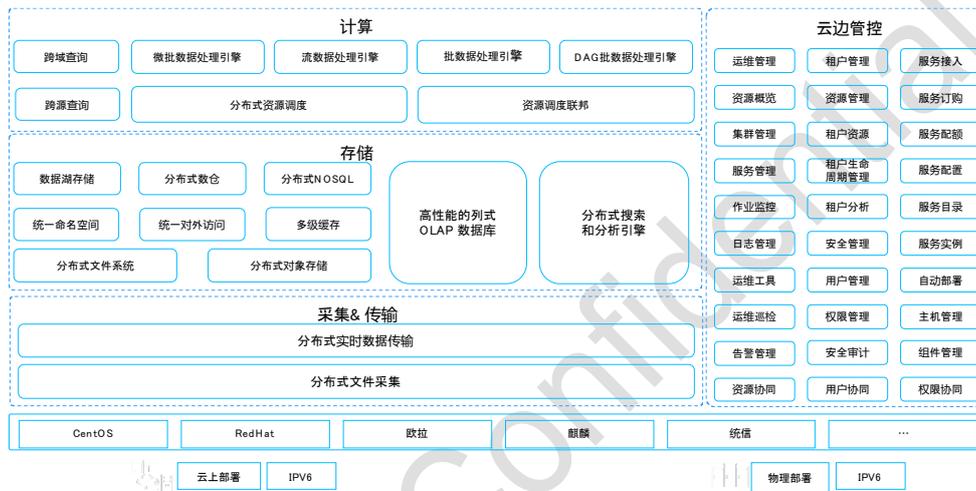


图 9-3 分布式储算能力解决方案

- 低成本，高性能

使用新技术、新架构及不断的迭代优化，为客户提供低成本、高性能，稳定安全的数据技术底座的产品。

- 通用型平台

大数据技术平台具备通用性，运维便捷，通过多类计算引擎、多类存储引擎，丰富工具的自由组合，为多行业打造适用的技术解决方案。

- 生态环境

以数据基础设施平台为基础，实现国产化操作系统和服务器支持，第三方工具接入，打造合作技术生态，共同为数据中台、数据探索分析，BI 分析等提供技术服务。

## 十. 产品客户成功故事

本章节详细阐述数据基础设施平台产品的成功案例。

### 10.1 大数据平台升级及数据迁移方案

随着移动省份数据量的增长，原有平台集群的不断扩容发展，原有的管理模式受到挑战，原有的数据存储技术，多类型 CPU 的环境支撑都需要新组件的升级来进行性能提升和系统安全性的提升。因此需要进行平台的版本升级，老集群数据迁移到新的集群并统一管理，来支撑移动业务的发展。

基础架构升级到产品 V5.3 版本，满足租户支撑，及支持 x86 和 ARM 架构主机混合模式。新版本的性能和稳定性大幅增强。

客户的痛点在于，大数据平台承载了多部门的租户应用，进而导致：

- 当前 Hadoop 集群资源利用率已经满负荷运转（忙时 CPU 超 85.9%、内存超 73%、存储超 60%）。
- 原有 Hadoop 的核心组件处理性能较低。
- Hadoop 集群规模受到原有架构限制，无法突破集群数量上限。

因此需要通过版本升级来满足以下需求：

- 核心组件优化，降低存储成本、提升计算效能。
- 架构升级实现国产化支撑。
- Hadoop 核心管理架构升级，拓展集群服务器管理规模。

所涉及到的集群规模如下：

- 涉及升级 Hadoop 集群规模约 666 台服务器。
- 可调度资源约是 Vcore 23956 个，存储约 60PB 的资源。
- 实现分布式文件存储、批量计算、内存计算。

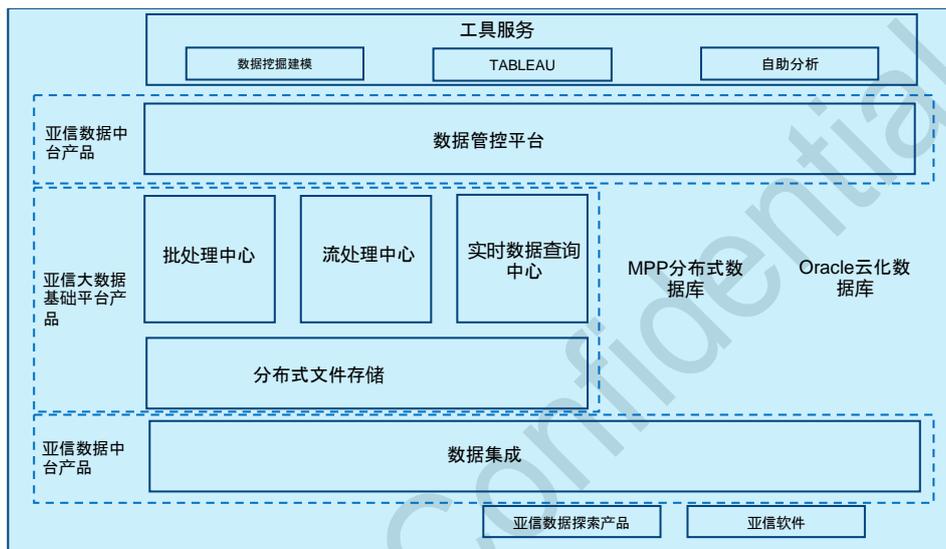


图 10-1 大数据中心架构图

大数据存储计算平台从 V4.2 版本升级到 V5.3 版本，核心组件升级，新增支持流批一体的计算组件，以支撑上层租户的应用。



图 10-2 大数据中心升级及迁移技术方案

- 国产化支持

不仅支持国产化 ARM 架构下部署，还提供基于 ARM 架构和 x86 架构的混合部署模式。

- 存算分离架构

存算分离架构，提升平台可扩展性和资源利用效率。

- HDFS 存储升级

引入纠删码存储管理，对长时间留存数据改为纠删码存储，节省存储资源。

- 资源管理 YARN 升级

从按资源比例分配升级为按实际 CPU、内存数量分配，对租户的资源控制更精准，同时大大增加租户任务并发数管理及租户资源使用监控。

- 核心组件性能升级

提升批处理集群数据处理效率。

引入 Flink 组件，支撑实时业务批流一体化数据处理。

- 提升集群横向扩展

升级集群管理节点能力，提供多 NameNode 联邦扩展，提升 Hadoop 集群横向扩展能力。

# 十一. 资质与荣誉

该产品获得了六项大数据的软件著作权，获得八项大数据专利，并获得了信通院的认证。在国产化兼容方面，也分别获得了海光、华为、银河麒麟的兼容性证书。

## 11.1 大数据相关软件著作权

总共获得六项软件著作权，详情见下表：

表 11-1 大数据相关软件著作权

软件名称	证书号	登记号	日期
亚信智能大数据平台[简称：AISW BDP] V1.0	软 著 登 字 第 5845923 号	2020SR0967227	2020 年 08 月 21 日
亚信多源异构数据处理平台[简称：AISW DPP] V1.0	软 著 登 字 第 5844951 号	2020SR0966255	2020 年 08 月 21 日
亚信数据服务平台[简称：AISW DSOP] V1.0	软 著 登 字 第 5844972 号	2020SR0966276	2020 年 08 月 21 日
亚信大数据分布式存储系统[简称：AISW BDDSS] V1.0	软 著 登 字 第 6300040 号	2020SR1500040	2020 年 09 月 09 日
亚信大数据实时处理平台[简称：AISW BDRPP] V1.0	软 著 登 字 第 5935040 号	2020SR1056344	2020 年 09 月 07 日
亚信大数据基础平台[简称：Data Platform] V5.3	软 著 登 字 第 7375413 号	2021SR0652787	2021 年 05 月 10 日

## 11.2 大数据相关专利

总共获得八项大数据领域的专利，详情见下表：

表 11-1 大数据相关专利

专利名称	备注
一种基于深度学习的大数据集群资源智能优化方法	专利号：201811653945.8
一种面向实时计算的动态窗口调整算法	专利号：201911266309.4
一种抽取关联数据图特征的图数据自动化构建系统	专利号：201911327964.6
一种面向通信领域的海量数据多级索引智能构建系统	专利号：201811640232.8
一种面向通信领域海量图数据查找超级节点的方法	受理中
一种基于数据血缘关系的冷热数据转换的方法	受理中
一种基于动态分类的异构资源调度方案	受理中
一种日志处理方法、装置、电子设备及计算机可读存储介质	受理中

## 11.3 信通院认证

2021 年 6 月 24 日，荣获中国信通院颁发的《分布式批处理平台基础能力大规模专项测试证书》，证书编号：IY8-2021-20048。

经检验，亚信大数据基础平台 V5.3，于中国信通院开展的第 12 批大数据产品能力测评中，在 100 节点集群规模下完成测试，在运维管理、多租户支持、高可用、安全、兼容性、扩展性、数据处理、易用性等方面，均满足分布式批处理平台基础能力测试要求。

## 11.4 国产化兼容适配

本章节介绍了产品对于国产化服务器的兼容适配所取得的技术成果。

### 11.4.1 适配海光服务器

2021 年，亚信科技和中科可控（中科可控信息产业有限公司）进行联合测试，经测试亚信数据基础设施平台完全兼容海光处理器开发的高性能 H 系列服务器，获的《产品兼容性互认证》证书。

### 11.4.2 适配华为 TaiShan200 服务器

2019 年 10 月 25 日，数据基础设施平台完成与华为 TaiShan 200 服务器的兼容性测试，获得 HUAWEI COMPATIBLE 证书及相关认证徽标的使用权。

### 11.4.3 适配银河麒麟服务器

2022 年 04 月 18 日，数据基础设施平台完成与银河麒麟高级服务器操作系统多个版本（飞腾版、鲲鹏版、兆芯版、海光版、AMD64 版）的兼容性测试，并且达到通用兼容性要求及性能、可靠性要求，满足用户的关键性应用需求，获得麒麟软件有限公司的《麒麟软件 NeoCertify 认证》。

## 十二. 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](http://www.asiainfo.com)



# Thank you

亚信科技依托产品、服务、运营、集成能力助力企业数字化，持续创造新价值。

亚信科技（中国）有限公司保留所有权利

