

# Big Data On Kubernetes

基于Kubernetes的新一代云原生大数据平台

# CONTENTS / 目录

## 01 / 公司简介

智领数据, 云享未来 / P3

## 02 / 传统大数据平台的挑战

传统大数据平台的挑战 / P5

## 03 / 云原生大数据架构优势

什么是云原生架构 / P6  
云原生大数据架构优势 / P6  
云原生大数据趋势 / P7

## 04 / 智领云Kubernetes Data Platform

KDP架构 / P8  
KDP核心优势 / P10  
KDP核心功能 / P12  
KDP界面展示 / P16

## 05 / 应用场景

传统技术的升级改造 / P20  
高效的集群发布和运维 / P21  
自助式的数字创新 / P22  
提升IT架构资源效率 / P23

## 06 / 性能比对

性能对比 / P24

## 07 / 信创支持

部署环境 / P25  
国产软硬件支持 / P25

## 08 / 企业荣誉

企业荣誉 / P26

## 09 / 我们的服务

我们的服务 / P27

# 01 / 公司简介

## COMPANY PROFILE

### 智领数据 云享未来

智领云科技有限公司（LinkTime Cloud）成立于2016年8月，是国内首个标准化和产品化的云原生大数据核心技术及服务提供商。智领云科技致力于云计算和大数据核心技术的研发和推广，为企业数字化转型提供最佳的产品解决方案和服务。公司研发的BDOS云原生大数据系列产品已在医疗，政务，物联网，教育，金融，能源多个行业落地，并获得了客户的一致好评。

智领云科技在成立之初即获得国内外著名投资机构以及硅谷独角兽公司多位创始人的天使投资；2019年初获得线性资本数千万元pre-A融资；2020年7月获得金沙江联合主投、线性资本跟投的数千万A轮融资。

# 02 / 传统大数据平台的挑战

## CHALLENGES OF TRADITIONAL BIG DATA PLATFORMS

传统大数据平台,一般是指以Hadoop为基础的大数据生态技术上构建的,通过组合一系列的大数据技术组件和工具来实现数据采集、存储、计算和分析、以及数据应用开发的大数据平台。

# 传统大数据平台的挑战

## 传统大数据平台难以实现资源的隔离

多租户环境下的数据开发效率提升, 需要以资源隔离的方式来保证租户之间的计算作业互相不影响, 特别是不能出现某一个或几个租户独占集群资源的情况。但Hadoop系统本身的出发点就不是为了多租户环境而设计的, 其目前的资源隔离实现也不完善。

## Hadoop存算合一的耦合架构决定了它的资源利用率无法提高

在一个Hadoop集群中, 一个节点既是存储节点 (datanode), 也是计算节点。当存储资源不够的时候, 增加节点可以进行存储扩容, 但会造成计算资源的利用率下降; 同样, 当计算资源不够而进行扩容的时候, 存储资源利用率就会下降。

## 传统大数据平台难以集成新的计算和存储技术

Hadoop系统在部署其他组件的时候, 对这些组件与HDFS和Yarn的版本适配有严格要求, 引入一个新的计算和存储组件的难度是非常高的, 往往需要几天甚至是几周的时候。

## Hadoop集群资源无法做到快速的弹性扩容和缩容

Hadoop的节点扩容和缩容流程, 导致扩缩容无法在很快的时间内完成, 数据备份以较小的传输率运行在后台, 往往要持续几个小时以上。

总而言之, 传统大数据平台因为其结构性的缺陷导致了多租户环境下数据开发效率低、集群资源利用率不高、以及集成新技术很复杂等问题, 依靠Hadoop生态技术框架本身的发展是不可能解决这些问题的。

# 03/ 云原生大数据架构趋势

## ARCHITECTURE TRENDS

### 什么是云原生架构 /

并不是运行在云主机上的程序或者容器化的程序就是云原生程序。其架构需要满足以下标准：

- Containerization—容器化
- Dynamic Management—动态管理
- Micro-service—微服务
- Orchestration—统一编排
- Automation—自动化

### 云原生大数据架构优势 /

#### 标准化发布运维

大数据组件的标准化发布运维，简化大数据应用场景的运维复杂度

#### 弹性扩展

充分利用微服务和容器化的优势，统一集群运营，弹性扩展，避免数据孤岛，应用孤岛

#### 节约成本

共享K8s资源管理及运维体系，降低资源需求及管理复杂度

#### 细粒度资源分配

集成端到端开发流程管理，数据即产品，利于数据能力的抽象共享和复用

#### 高效开发

敏捷式协同开发，加速技术创新与落地

#### 极致性能

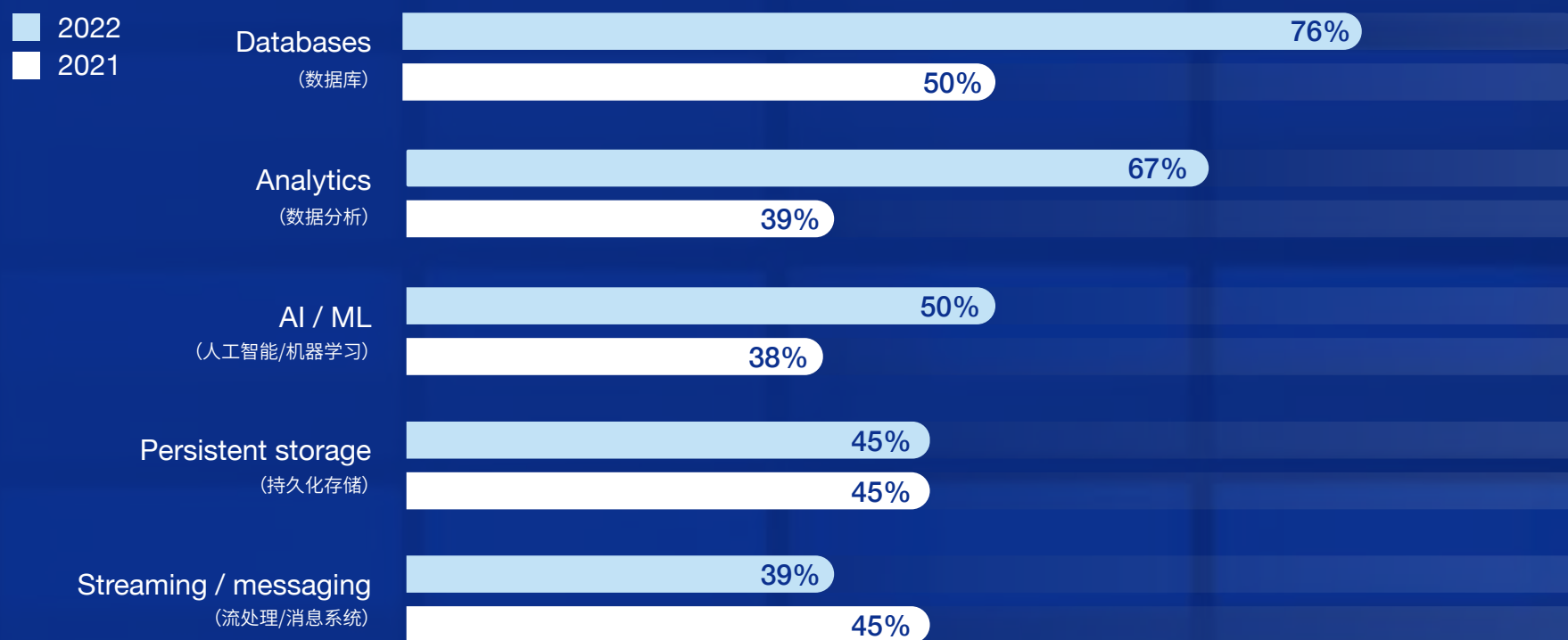
存算分离的架构，易于实现的混合调度，可将系统资源使用率提高到60%-70%



# 云原生大数据趋势

Kubernetes作为云原生时代的“操作系统”，已于2021年3月和5月分别正式支持Spark on K8s、Kafka on K8s，HDFS也有了云原生的对标方案，大数据平台的云原生化已是大势所趋。

2021年—2022年，在Kubernetes上运行的不同类型的工作负载的占比及趋势



# 04 / 智领云 Kubernetes Data Platform

## PRODUCT INTRODUCTION

Kubernetes Data Platform 是智领云Kubernetes大数据平台(以下简称“KDP”)

### KDP简介

KDP是市场上首个可完全在Kubernetes上部署的容器化云原生大数据平台,深度整合了云原生架构的优势,将大数据组件及数据应用纳入Kubernetes管理体系,标准化系统管理,提升系统运行效率,降低运维成本,消除应用孤岛及数据孤岛,解决传统Hadoop大数据平台在部署,运维,运行效率上由于架构限制带来的难点。

### 主要功能

支持K8s集群环境下生产级别的运行稳定性及高可用。

实现基于OpenID协议和开源软件Keycloak的身份认证和访问控制方案。

支持统一的调度机制以及精细化的作业调度策略

支持基于Kerberos协议的安全认证和基于Apache Ranger的统一授权管理框架。

优化计算引擎性能,包括实现Spark 和 Hive On Spark 的Locality优化等。

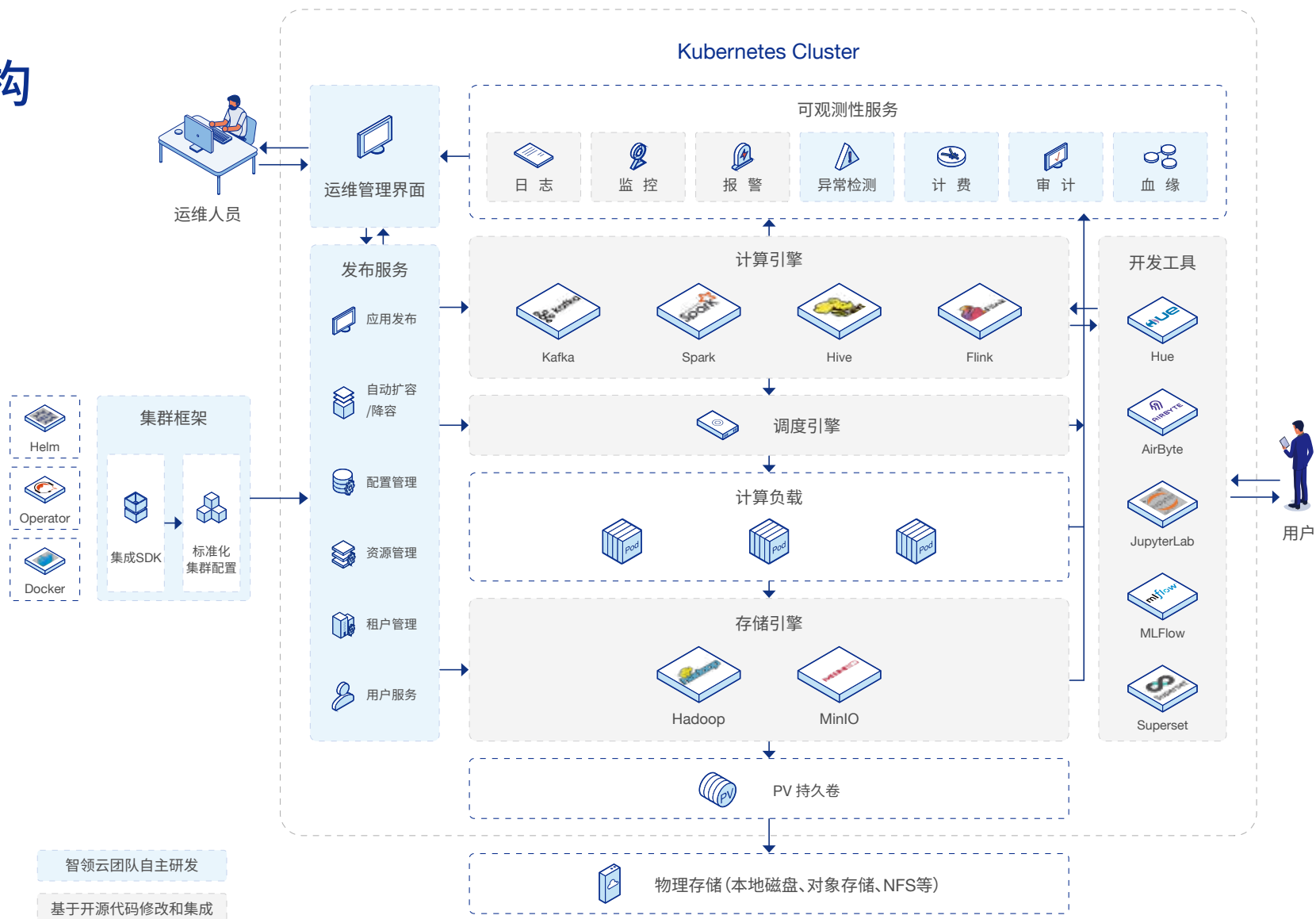
支持以标准化(Operator或者Helm chart)的集群部署方式进行适配。

支持可观测性服务的适配,包括监控、报警、日志、审计、计费等等。

实现与BDOS云原生大数据集成基座的适配,实现可视化的发布、配置修改、运维等。



# KDP架构



KDP核心优势 /

## 提升集群资源使用率

### 共享资源池

KDP以K8s技术为基座, 将整个集群资源作为一个可以共享的资源池, 实现实时分析作业和离线分析作业的统一资源调度机制, 并进行精细化调度。

### 公有云部署

通过云服务提供商提供的接口, 实现对计算和存储资源的动态扩缩容, 灵活应对短时间出现的性能瓶颈问题。

### 实时/离线作业混部

同一个KDP节点上可同时运行实时分析作业和离线分析作业, 通过作业混部能把集群资源使用率从传统大数据平台的30%左右, 提升到60%以上。

### 私有云部署

通过虚拟化系统提供的接口也能实现一定程度的资源动态扩缩容, 并可把备用的一部分虚机在不同的系统中进行共享, 提高资源复用率。

KDP核心优势 /

## 简化大数据组件集成及配置安装流程

### 标准化的配置管理流程

采用统一的K8s文件配置方式, 对大数据组件的镜像文件、软件依赖、部署步骤、K8s网络和存储的对接、计算和存储资源的配额、监控指标的输出、日志文件的采集等内容进行描述, 极大简化大数据组件与K8s集群的集成。

### 自动化的安装流程

基于标准化集成和配置流程, 安装程序会通过这些配置文件来自动执行大数据组件到K8s集群的部署、更新、运维、升级等操作, 降低大数据平台部署的复杂性, 减少大数据平台运维的成本。

KDP核心优势 /

## 提高计算作业的运行效率

### 提供灵活调度策略

采用Volcano作为一级调度和K8s作为二级调度来实现调度机制的精细化管理。基于Volcano的多实例部署，分别管理不同命名空间或者不同节点组的作业调度，实现大规模作业调度的低延时。

### 支持集群的弹性扩容

在集群的计算资源出现不足的情况下，利用K8s的弹性扩容技术来从容应对计算作业的性能瓶颈，实现计算资源及集群资源的动态扩容，以提供更好的计算性能。

KDP核心优势 /

## 降低运维成本并提升运维效率

### 更轻量、更少资源消耗的日志系统

KDP通过为大数据组件的pod部署一个sidecar容器来实现日志采集，所有日志都是以流文件的形式推送和存储在Loki日志系统中，方便运维人员通过统一界面管理、下载与查询。同时，Loki支持以对象存储的方式存储日志，借助K8s存算分离的技术，进一步降低日志系统的整体资源消耗成本。

### 标准化大数据组件运维操作

通过K8s标准的Operator操作方式，完成大数据组件的部署、升级、扩容、备份等操作，相比去掌握各种不同的大数据组件的运行原理和运维操作，其学习成本要低很多。同时，KDP将一些基本的大数据组件运维操作集成到运维管理界面，通过统一界面，进一步提升运维人员执行运维操作的效率。

## KDP核心功能 /

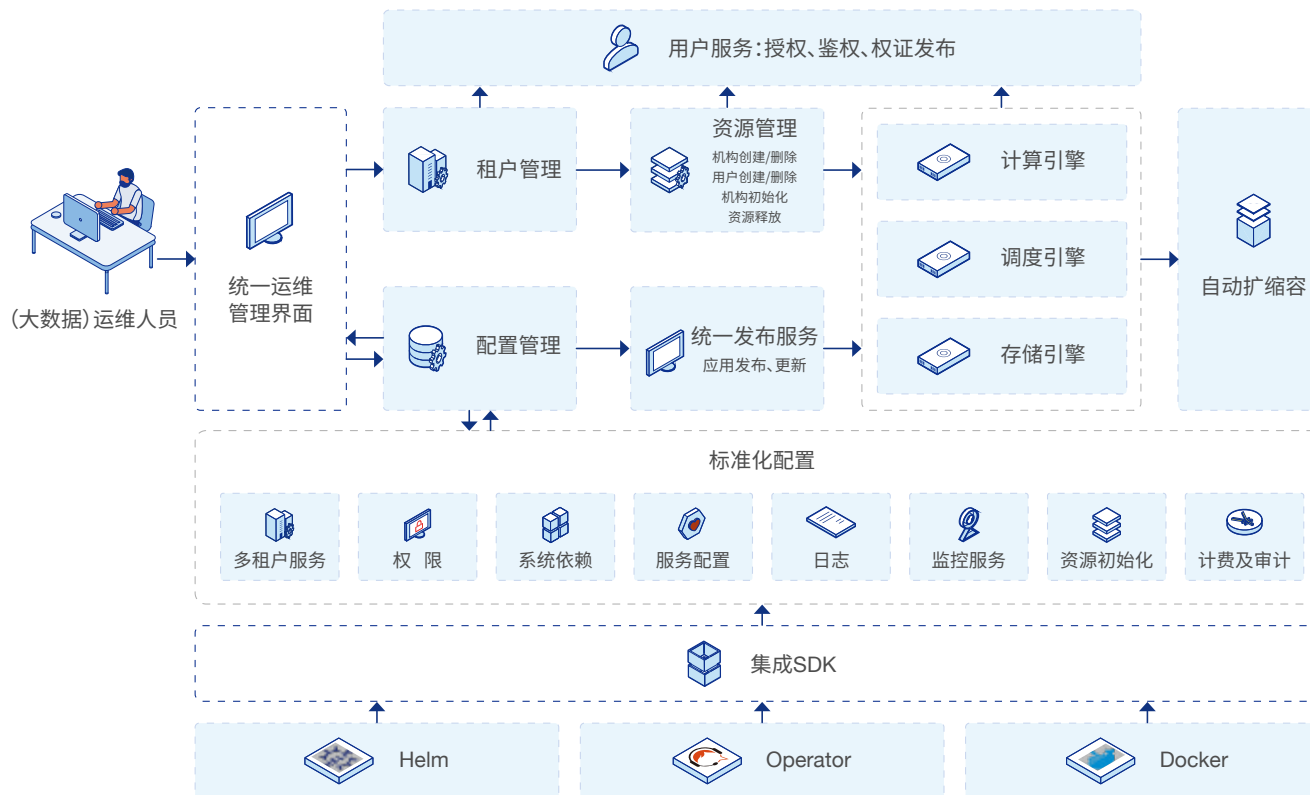
# 标准集成框架与发布服务

### 标准集成流程

允许各种大数据组件以标准化配置文件的形式与统一系统服务对接,形成标准化部署,运行和运维流程。在K8s配置的基础上提供封装,简化大数据组件的配置流程,标准化组件与系统服务及其它组件之间的对接机制。

### 统一发布服务、更新、运维和升级操作

实现大数据组件从配置文件到K8s集群的统一发布、更新、运维与升级。支持大数据类型的负载(如大量即起即停的批处理任务),可与大数据租户体系、权限管理、用户管理、资源管理进行对接和集成。



KDP核心功能 /

## 计算引擎和存储引擎的云原生改造与集成

### 计算引擎云原生改造与集成

业界首个基于Spark计算引擎在K8s上运行Hive SQL的产品实现;将Spark和Flink纳入调度引擎,实现统一的资源调度管理,支持云原生的精细化调度引擎Volcano,作为K8s之上的一级调度系统.

### 存储引擎的云原生改造与集成

实现HDFS基于PV的云原生存储模式及datanode (在硬件环境支持的情况下) 的弹性扩容;通过声明PVC,以PV的方式来实现存储引擎背后的物理存储;支持Hive分别在MinIO和HDFS中同时进行冷热数据存储。



KDP核心功能 /

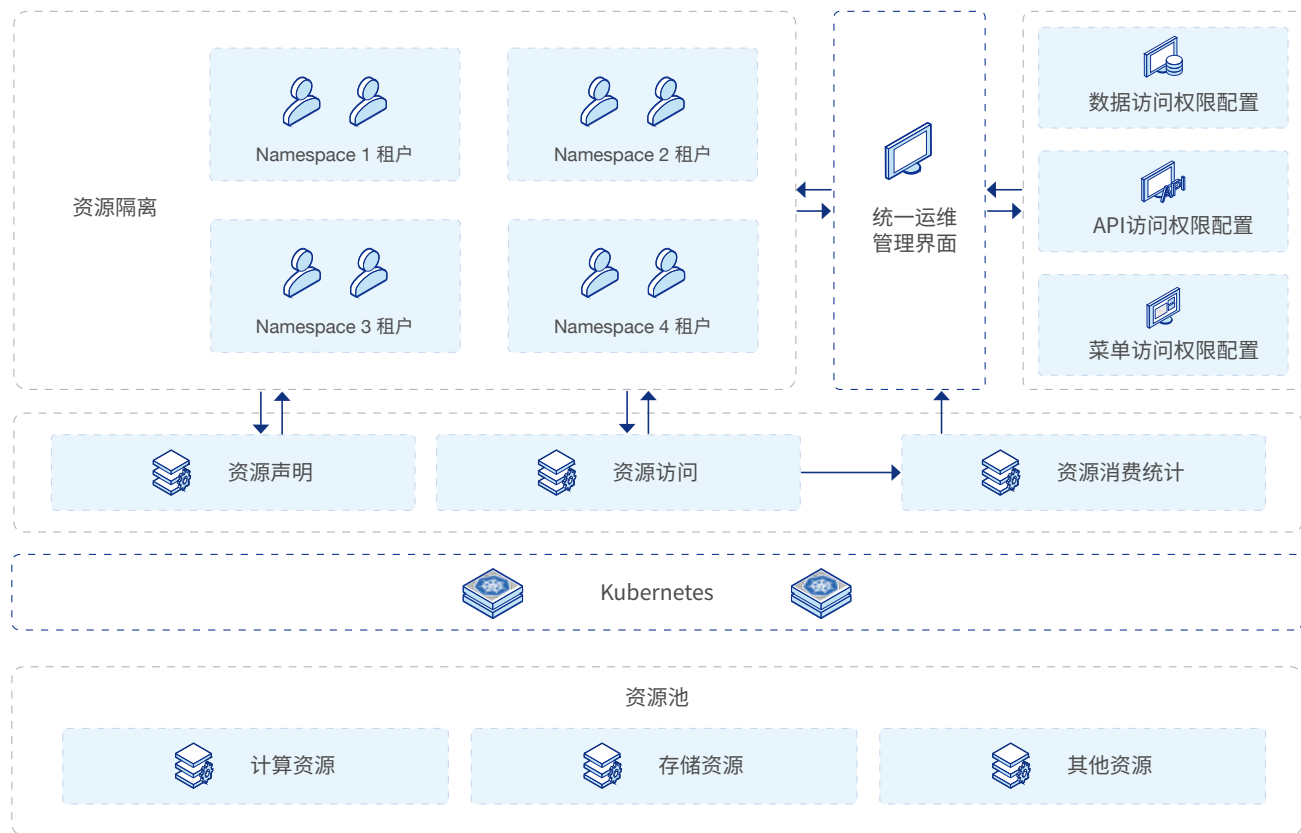
## 云原生的多租户和安全管理

### 云原生的多租户管理

每个新用户都会创建单独的用户账号和对应的Kerberos keytab, 并加入相应的安全组。每个安全组都有自己独立的K8s命名空间, 并独立发布计算作业, 每个命名空间都有对应资源配额, 实现多租户的用户管理和资源隔离。

### 云原生的安全管理

实现系统所有集成大数据工具的单点登录、Kerberos安全认证及Ranger授权管理。





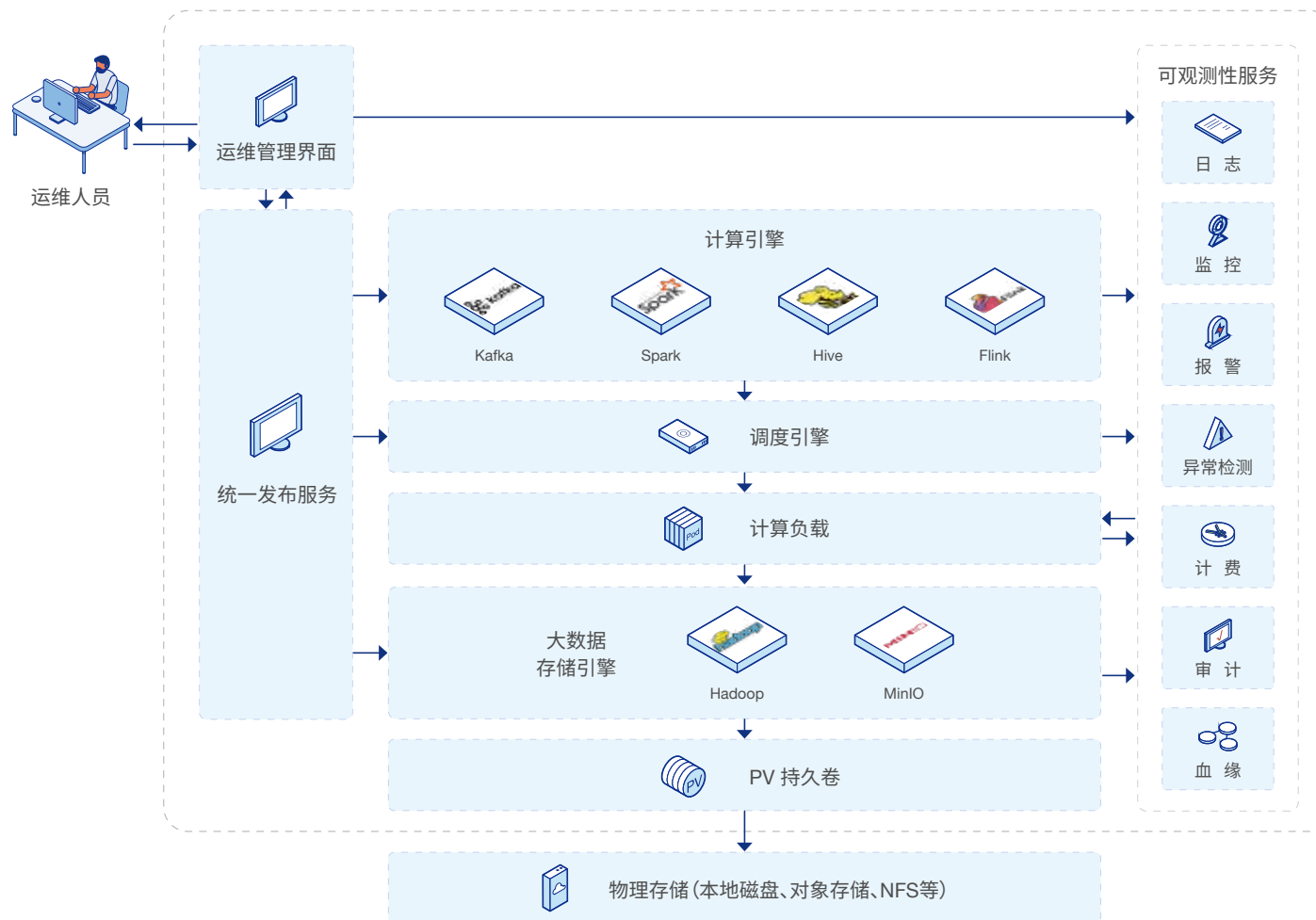
# KDP核心功能 / 可观测性服务

## 日志聚合

提供大数据组件及其执行的 Workload 的日志,对性能及稳定性的指标进行监控和报警,并通过统一界面进行日志管理。

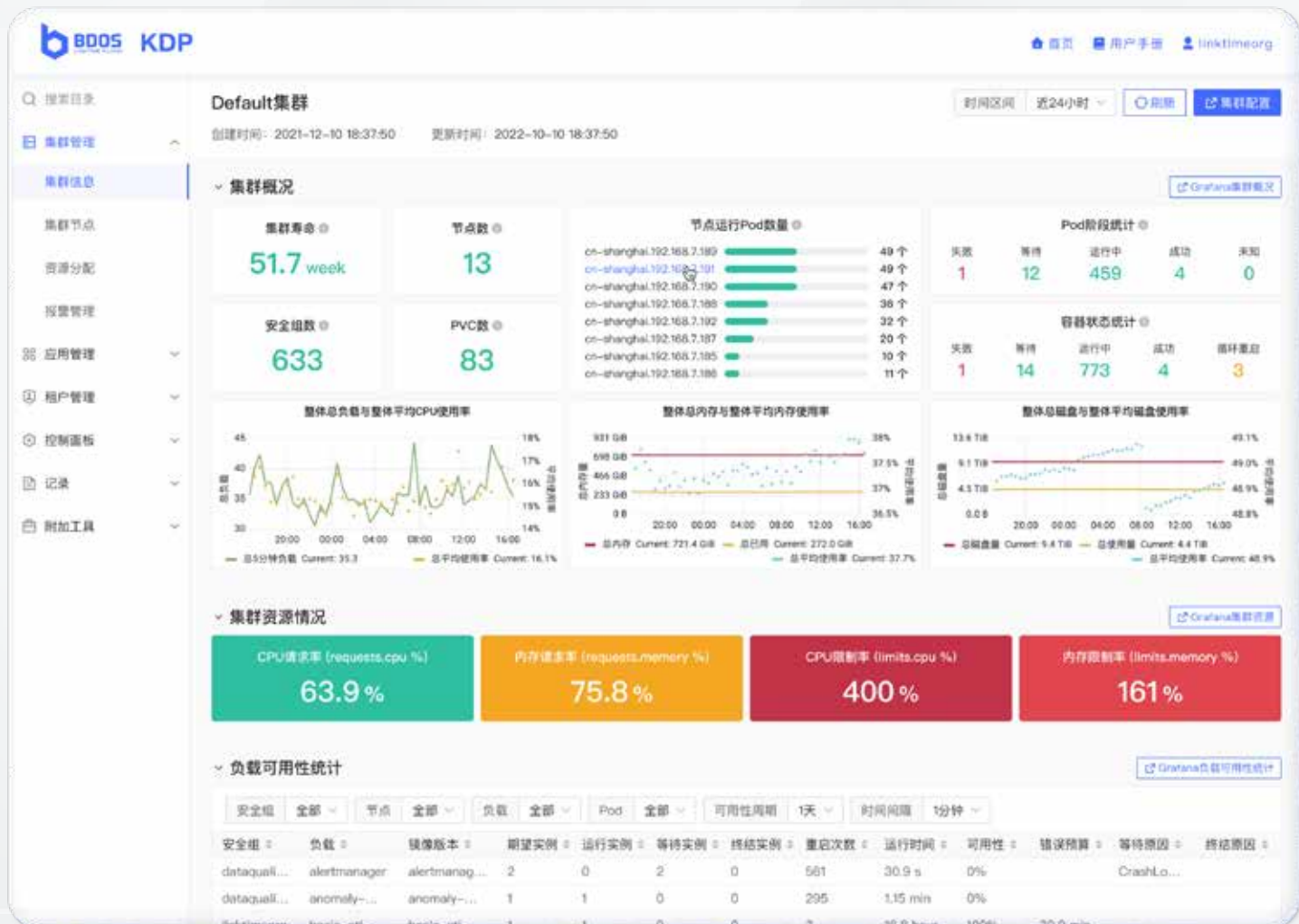
## 计费与审计

收集大数据计算引擎和存储引擎的资源使用数据,对其进行整理、存储和分析,提供HTTP API,让用户能清晰地看到各安全组、各用户、各计算作业等的资源使用情况的统计,并通过运维管理界面进行查询和审计。



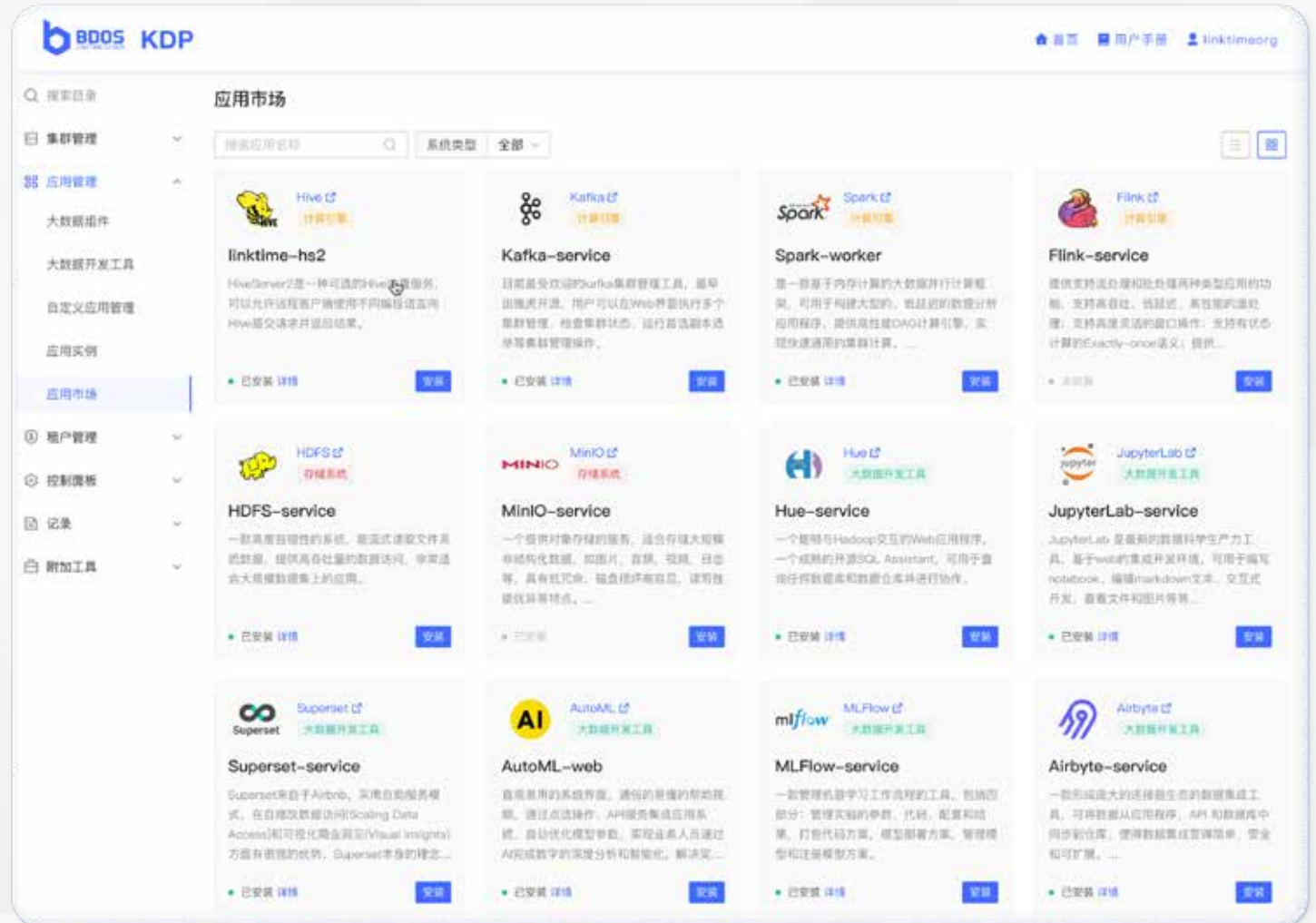
# 集群管理/ Cluster Management

统一界面进行集群精细化管理



# 应用管理/ Application Management

应用一键安装、部署与运维



# APPLICATION MANAGEMENT

# 资源管理/ Resource Management

细粒度资源计算与统计

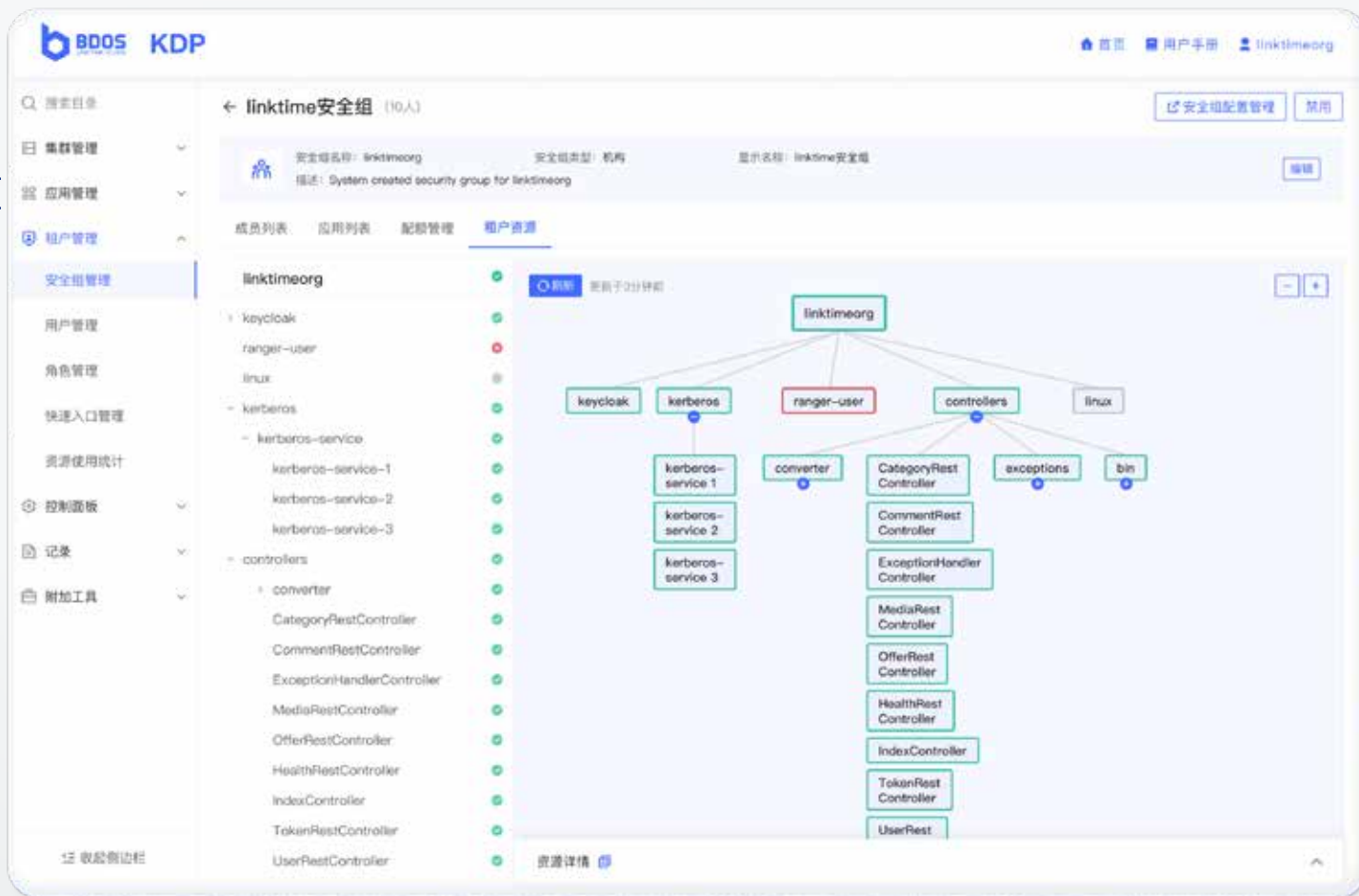
The screenshot displays the '安全组管理' (Security Group Management) page in the LinkTimeCloud KDP interface. The page features a search bar, filter options for type and status, and a table listing various security groups. Each group entry includes its name, member count, type (e.g., '系统创建的安全组'), description, status (e.g., '已激活'), and resource usage metrics for CPU, memory, and disk. Action buttons for editing, configuration, and disabling are provided for each group.

安全组名称	人数	类型	描述	状态	CPU	内存	磁盘	操作			
linktime安全组 linktimeorg	10人	系统	System created security group.	已激活	3.54	20%	4.30G	10%	20M	90%	编辑 配额管理 禁用
演示安全组 linktimedemoorg	8人	系统	System created security group.	未激活 <a href="#">已激活</a>	0	0	0	0	0	0	编辑 配额管理 禁用
demo安全组 demoorg	8人	系统	System created security group.	已激活	7.95	50%	24.54GB	50%	80M	50%	编辑 配额管理 禁用
测试组1 testorg1	8人	其他	System created security group.	已禁用	-	-	-	-	-	-	编辑 配额管理 启用
测试组2 testorg2	8人	系统	System created security group.	激活中 <a href="#">已激活</a>	0	0	0	0	0	0	编辑 配额管理 禁用
测试组3 testorg3	8人	其他	System created security group.	已激活	3.2	12.18%	4.34GB	10%	20M	90%	编辑 配额管理 禁用
测试组4 testorg4	8人	系统	System created security group.	已激活	1.8	11.06%	4.34GB	10%	20M	90%	编辑 配额管理 禁用
测试组5 testorg5	8人	系统	System created security group.	未激活 <a href="#">已激活</a>	0	0	0	0	0	0	编辑 配额管理 禁用
测试组6 testorg6	8人	系统	System created security group.	已激活	1.8	11.06%	4.34GB	10%	20M	90%	编辑 配额管理 禁用
测试组7 testorg7	8人	系统	System created security group.	已激活	1.8	11.06%	4.34GB	10%	20M	90%	编辑 配额管理 禁用

# RESOURCE MANAGEMENT

# 多租户管理/ Multi-tenant Management

机构用户资源初始化全局管理

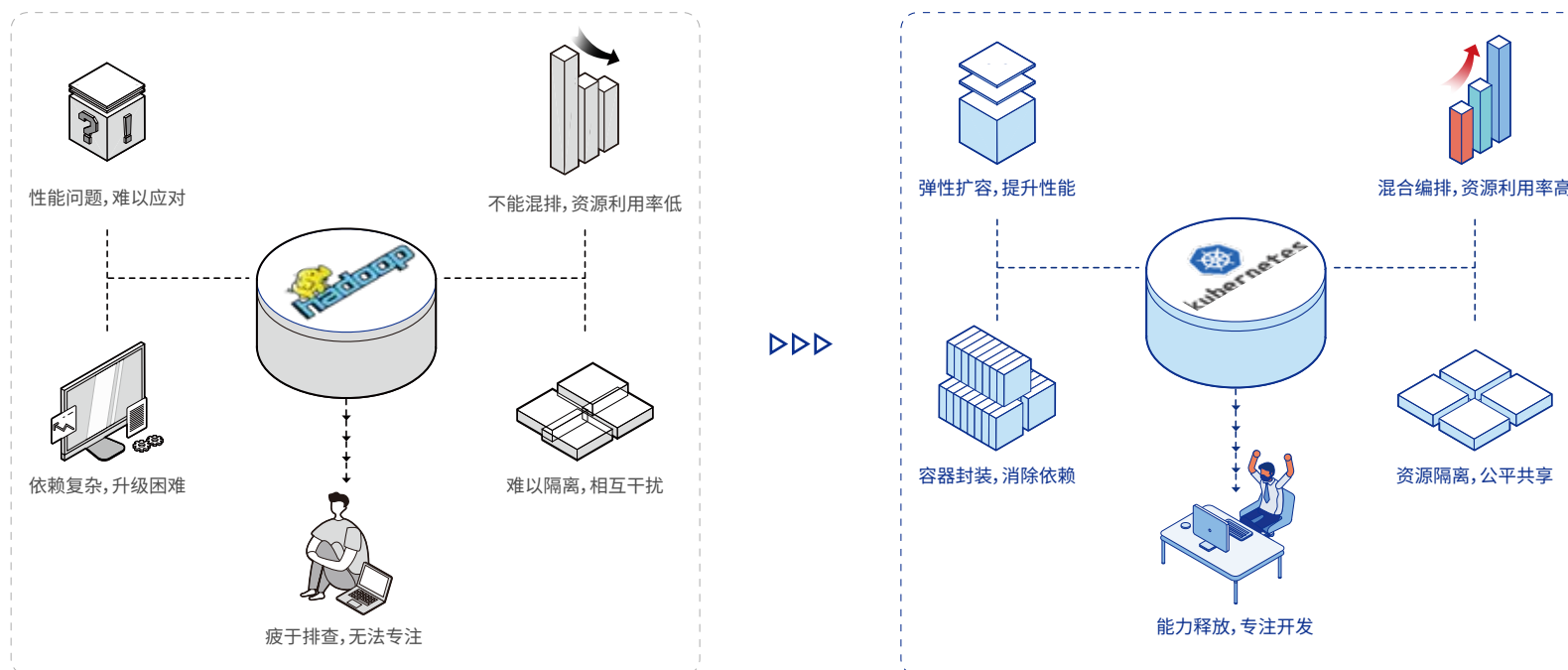


# MULTI-TENANT MANAGEMENT

# 05/ 应用场景 CLASSIC CASES

## 传统技术的升级改造

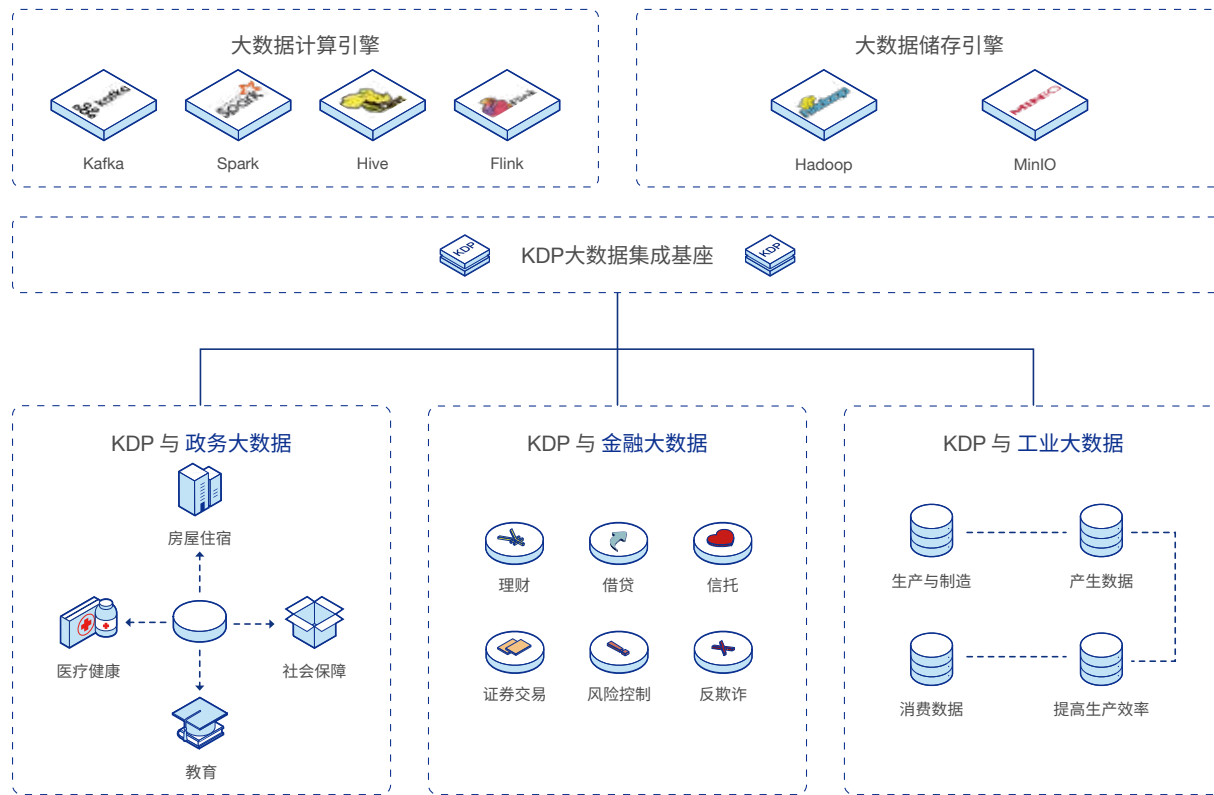
传统大数据平台因为技术扩展迭代流程比较慢,不能及时解决运维中碰到的性能瓶颈,大数据组件之间存在软件包依赖复杂、组件升级困难、新的组件集成耗时费力等问题。传统大数据平台的技术团队面对运维压力而难以专注于业务开发和数据价值的发现,迁移到KDP后,可以显著提升运维效率,降低运维成本,解放技术团队的生产力。





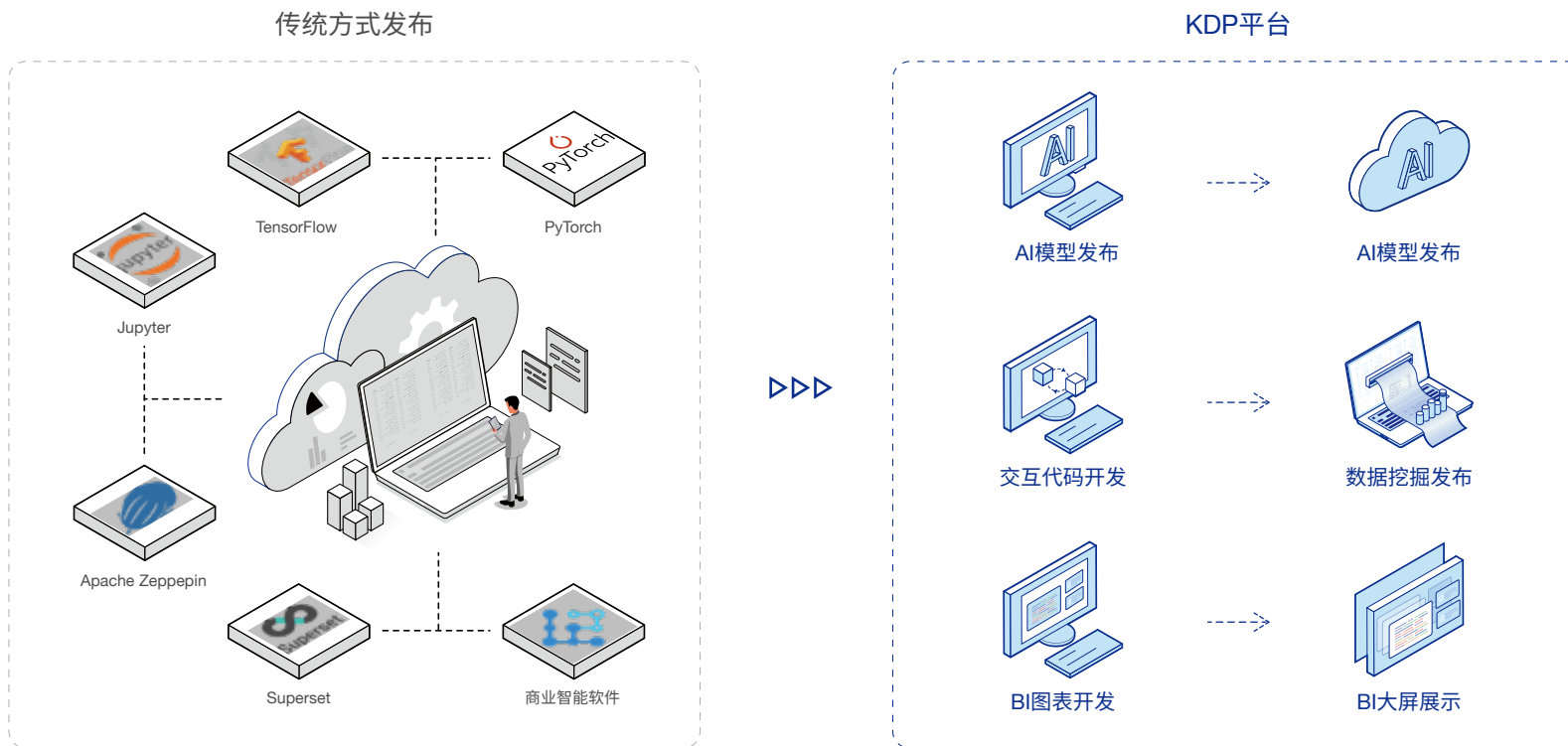
## 高效的集群发布和运维

有的企业作为技术提供方要为多个内部或外部的机构进行大数据集群的部署和实施,但传统大数据平台的软件部署、组件互相适配、计算引擎调优等方案相对复杂,手工部署的步骤多,导致集群部署周期长,项目实施成本很高,运维流程复杂,运维人员能力要求高。在这种场景下,采用KDP云原生大数据平台,可以大幅度提升实施项目的部署效率,降低项目实施运维人力和时间成本。



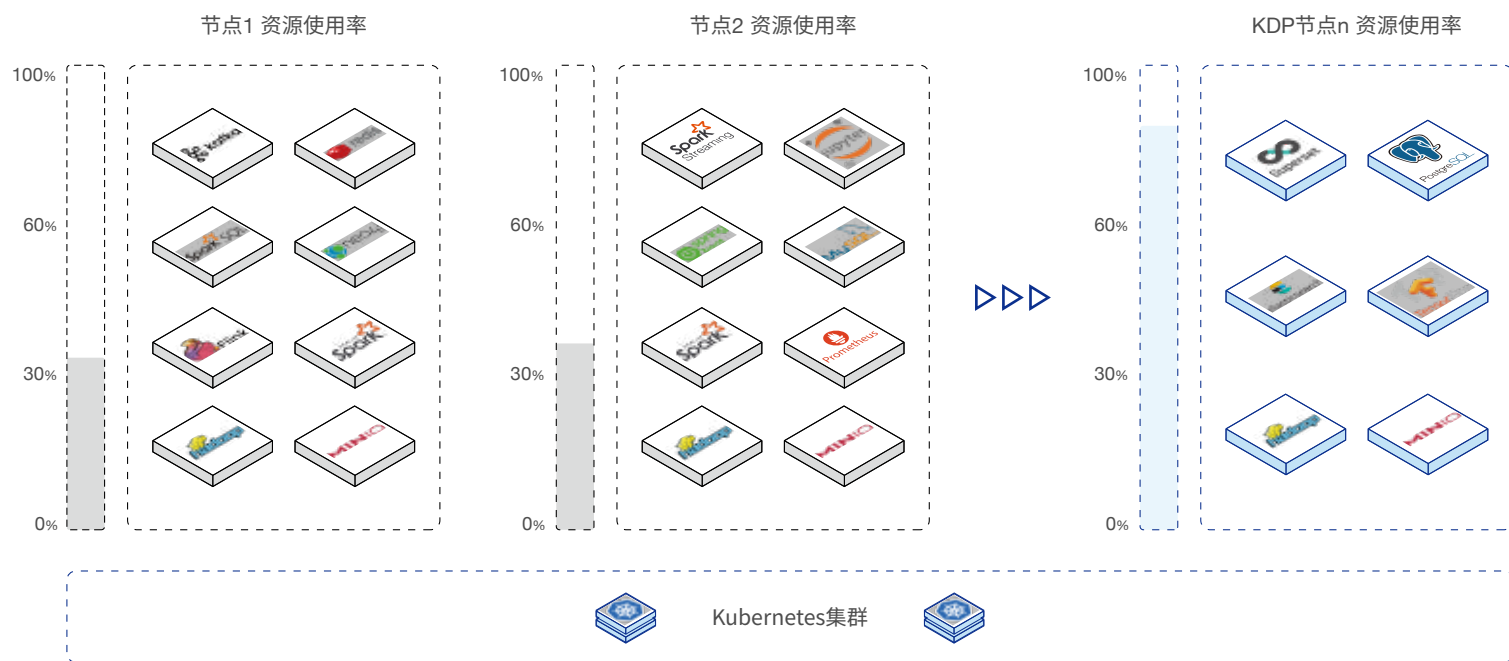
# 数字化创新

有的企业需要有多个大数据集群服务不同的业务部门，业务部门的数据科学家希望能自助式地尝试新的云原生人工智能机器学习工具。很显然，传统大数据平台满足不了这种自助式需要，企业可以通过KDP平台部署提升多平台管理效率，提供数据分析和人工智能开发工具的自助式发布，降低整体资源消耗的成本，加速数据价值的创造过程。



## 提升IT架构资源效率

有的企业在生产环境中运行多种类型的数据应用、不同类型的存储引擎、实时和批处理的计算作业。在传统大数据平台环境下，一般都是采用独立的虚拟机集群来部署这样的生产环境，导致资源使用率很低。采用了KDP以后，企业可以利用作业混排、存算分离和精细化调度等平台特性来提升整体资源使用效率，降低IT架构的投入成本。



# 06 / 性能对比 PERFORMANCE COMPARISON

## 部署效率

时间(少):小时级别系统应用部署

依赖复杂度(低):基于集成框架的标准化配置管理,解决版本兼容、软件包依赖及分布式网络配置等难题

操作难度(低):界面化操作,自动化部署,全程可观测

## 资源效率

资源利用率(高):从30%提升到60%及以上

资源统计颗粒度(细):集群总资源、各个namespace、应用负载、计算作业等的资源消耗统计与聚合

资源管理范围(广):对集群、节点、各个namespace、应用实例等的资源配额进行自助式管理

资源共享程度(高):形成一个可共享的资源池,计算、存储资源可动态调整

## 运维效率

运维人数(少):一个后台工程师只需掌握基本的K8s技术即可进行大数据组件运维

运维门槛(低):统一界面操作及自动化运维

问题定位与排查(快):全局可观测性服务,自动监控与告警,统一日志管理

	传统平台	KDP平台
部署效率	 <p>天/级别部署时间</p>	 <p>小时/级别部署时间</p>
资源效率	 <p>30%资源利用率</p>	 <p>60%及以上资源利用率</p>
运维效率	 <p>运维门槛和运维成本高 自动化运维程度低</p>	 <p>统一界面自动化运维 高度自动化</p>

# 07 / 信创支持

## 部署环境

### 公有云部署

KDP支持在不同的公有云上进行部署,已完成与阿里云容器服务、腾讯云TKE服务、华为云CCE服务、亚马逊EKS服务的适配,包括支持在托管K8s服务集群上部署、支持公有云平台的云原生存储方案、以及对接其他云原生计算或存储组件等。

### 私有云部署

KDP支持私有云部署,支持在用户提供的K8s集群环境上进行部署,已完成与CentOS 7.9操作系统、社区版K8s和OKD等不同环境上的部署适配。

## 国产软硬件支持

KDP支持在所有社区版K8s适配的国产软硬件上进行部署;支持基于飞腾和鲲鹏芯片的国产Arm架构CPU服务器的适配。

目前KDP已完成了与中标和银河麒麟操作系统,以及华为、浪潮、新华三等多家国内服务器的软硬件适配。

# 08 / 企业荣誉 ENTERPRISE HONOR

- **2022**

  - 08月/ 2022大数据产业生态大会, 荣获“中国大数据企业50强”、“大数据投资价值实力企业奖”、“2022数字赋能先锋企业30强”、BDOS云原生K8s大数据平台荣获“大数据领域最佳产品奖”
  - 07月/ 入选《互联网周刊》、德本咨询、eNet研究院联合发布的《2022大数据独角兽企业排行榜》、《2022数据中台TOP50》
  - 03月/ 智领云BDOS入选《开源GitOps产业联盟生态图景1.0版》
  - 02月/ 智领云入选《2022中国数据智能产业图谱1.0版》
- **2021**

  - 12月/ 智领云《云原生数据中台:架构、方法论与实践》图书成功入选“财资中国2021年度好书”, 智领云CTO宋文欣, 荣誉入选首批“开源GitOps产业联盟技术委员会”成员名单
  - 06月/ 智领云成为腾讯云原生加速器首期入选企业, 智领云联合创始人兼CTO宋文欣荣登腾讯云原生加速器首期成员名单
  - 02月/ 通过CMMI 3 认定
- **2020**

  - 12月/ 获东湖高新区2020年科技“小巨人”认定, 计世传媒2020数字创变者大会, 斩获年度中国优秀数字化解决方案大奖, 通过国家高新技术企业认定
- **2019**

  - 05月/ 智领云获线性资本千万级人民币Pre A轮融资, 为企业搭建标准数据中台
- **2018**

  - 10月/ 智领云BDOS大数据操作系统荣获“2018年大数据行业创新产品奖”

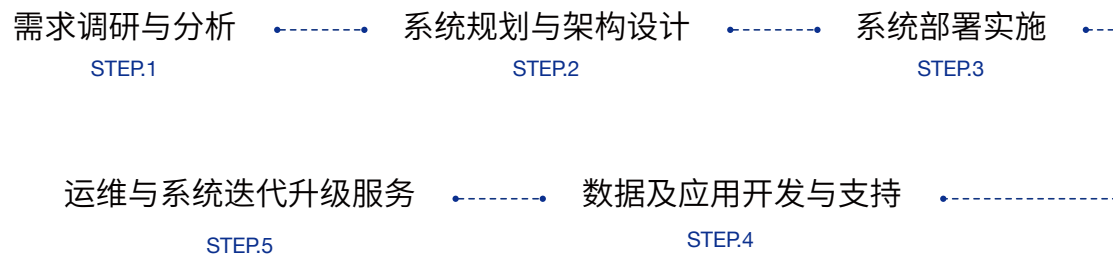


## 09 / 我们的服务

### OUR SERVICES

- 企业级k8s大数据平台建设实施与运维服务
- 企业传统大数据平台云原生改造与升级服务
- 企业数字化转型解决方案咨询与实施
- 大数据、机器学习、AI应用开发、培训和咨询

### 服务流程



# THANKS



电 话 400-8040-019  
邮 箱 info@linktime.cloud  
网 址 www.linktimecloud.com  
地 址 武汉·中国光谷