

**UCLLOUD 优刻得**

中国第一家公有云科创板上市公司

股票代码：688158

# USDP 3.1.13

# 大数据平台操作手册

发布日期：2020年02月22日

文档更新：2023年09月14日



下一代云基础设施

筑基数字化底座

## 版权信息

---

版权所有 ©2023 优刻得科技股份有限公司 保留一切权利。

本文档中出现的任何文字叙述、文档格式、图片、方法及过程等内容，除另有特别注明外，其著作权或其它相关权利均属于优刻得科技股份有限公司。非经优刻得科技股份有限公司书面许可，任何单位和个人不得以任何方式和形式对本文档内的任何部分擅自进行摘抄、复制、备份、修改、传播、翻译成其它语言、将其全部或部分用于商业用途。

### 注意

您购买的产品、服务或特性等应受优刻得科技股份有限公司商业合同和条款约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用权利范围之内。除非合同另有约定，优刻得科技股份有限公司对本文档内容不做任何明示或暗示的声明或保证。

## 关于文档

---

优得刻科技股份有限公司在编写本文档时已尽最大努力保证其内容准确可靠，但优得刻科技股份有限公司不对本文本中的遗漏、不准确或错误导致的损失和损害承担责任。

由于产品版本升级或其它原因，本文档内容会不定期更新，除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

## 目录

<b>1 产品简介</b> .....	<b>6</b>
1.1 产品概述 .....	6
1.2 核心优势 .....	6
1.2.1 Web 控制台, 轻松管理大数据集群 .....	7
1.2.2 一键开启或关闭 Kerberos 安全模式, 为大数据服务的安全保驾护航 .....	7
1.2.3 大数据集群平滑升级, 为需求升级提供有力支撑 .....	8
1.2.4 配置角色组管理, 有效提升资源利用率和管理效率 .....	8
1.2.5 UDH 更新及发布, 多版本可灵活选择 .....	8
1.2.6 宿主环境的修复能力, 为运维人员减负 .....	9
1.2.7 轻量便捷的操作 .....	9
1.2.8 全面的生态支持 .....	9
1.3 产品架构 .....	10
1.4 应用场景 .....	11
1.4.1 离线/实时数仓架构场景 .....	11
1.4.2 流批一体架构场景 .....	12
1.4.3 数据湖/湖仓一体架构场景 .....	12
<b>2 名词解释</b> .....	<b>14</b>
<b>3 功能简介</b> .....	<b>15</b>
3.1 USDP 核心功能 .....	15
3.2 UDH 发行版所含大数据生态服务列表 .....	16
<b>4 部署 USDP</b> .....	<b>19</b>
<b>5 访问 USDP 控制台</b> .....	<b>20</b>
5.1.1 在浏览器中访问 USDP 控制台 .....	20
<b>6 多集群管理</b> .....	<b>21</b>
6.1 创建大数据集群 .....	21
6.1.1 集群创建向导 .....	22
6.2 集群切换 .....	26
6.3 非 HADOOP 生态数据集群 .....	26

6.3.1 适用场景.....	26
6.3.2 参考架构图.....	27
6.3.3 资源规划.....	27
6.3.4 示例-创建 Kafka 集群.....	27
<b>7 单集群管理.....</b>	<b>28</b>
7.1 集群管理.....	28
7.1.1 集群级控制.....	28
7.1.2 集群主机管理.....	29
7.1.3 集群大数据服务及服务操作.....	29
7.1.4 集群监控数据概览.....	30
7.2 主机管理.....	30
7.2.1 主机控制.....	31
7.2.2 主机控制及监控.....	31
7.2.3 主机视角-大数据服务实例管理.....	32
7.2.4 为集群扩展主机.....	33
7.3 服务管理.....	37
7.3.1 服务控制及监控.....	38
7.3.2 服务实例管理.....	40
7.3.3 服务配置管理.....	43
7.3.4 服务 WEB UI 访问.....	45
7.3.5 为集群扩展服务-向导.....	46
7.4 告警管理.....	48
7.4.1 告警规则.....	48
7.4.2 活跃告警.....	51
7.4.3 历史告警.....	52
<b>8 用户管理.....</b>	<b>53</b>
<b>9 平台管理.....</b>	<b>54</b>
9.1 UDH 源.....	54
9.2 安全管理.....	54
9.2.1 名词解释.....	55

9.2.2 启用 Kerberos 向导.....	55
9.3 授权管理.....	62
9.3.1 名词解释.....	62
9.3.2 上传或更换授权.....	63
9.4 审核.....	63
9.4.1 操作日志.....	63
9.5 诊断.....	63
9.5.1 事件.....	63
9.6 集群信息说明.....	64
9.6.1 各服务常用端口.....	64
9.6.2 各大数据服务组件 WEB UI 登录账号.....	65
9.6.3 各组件服务部署规则说明.....	66

# 1 产品简介

## 1.1 产品概述

智能大数据平台 USDP (UCloud Smart Data Platform ) 是 UCloud 自主研发的智能化、轻量级的企业级自研大数据基础服务平台，提供大数据集群管理和运维能力，全面兼容开源生态，用户提供一站式、开箱即用的大数据生态服务支持，快速构建大数据的分析处理能力和高效的服务管理与运维能力。凭借其完备得管理功能、丰富的管理工具集、细致的监控体系，可大大简化对大数据基础平台原生且复杂的维护控制工作，提高集群管理效率，提升系统稳定性，降低管理成本，使用户聚焦于数据分析及价值挖掘中，达到降本增效的目的。USDP 多集群部署和集中管理的多个集群能力，可方便快速的管理成百乃至上千台服务器资源在各集群中的使用。该应用程序使安装过程自动化，从而将大数据集群部署时间从数天缩短至数分钟；为您提供集群范围内实时运行的主机和服务的视图；提供单个中央控制台，以在整个群集中实施服务配置管理；结合监控指标采集扩展，将集群及服务状态以及性能表现尽收眼底，为优化集群性能和资源利用率提供强有力的支撑。

## 1.2 核心优势

大数据业务系统作为企业信息系统的重要组成部分，近些年来亦成为信创的关注焦点之一。针对私有化部署场景，UCloud 推出的一站式智能大数据平台 USDP，可灵活构建于 IDC 物理服务器、云 IaaS 虚拟化，依托于自研的 USDP Manager 管理工具，实现对多套大数据集群的管理，并可使用户独享大数据集群。支持开源 Hadoop 全生态，进行集群、服务、监报告警、故障诊断等智能化的运维和管理操作，从而协助用户轻松构建和管理大数据业务分析处理能力。



### 1.2.1 Web 控制台，轻松管理大数据集群

USDP 集中化的管理控制台，在本次版本发布中，增加了集群管理视图功能，该视图使集群管理员对整个集群的服务状态一目了然；并可采取便捷的管理措施调整，保证系统的高可用和稳定性。

在 USDP 中，管理员可通过自动化向导的方式快速取得大数据服务对业务的支持；集中化的管理界面中，企业运营团队可以便捷的控制和调整服务配置和资源分配，以及一键开启/关闭 Kerberos，极大简化配置和管理的复杂性；自动化向导支持快速部署集群、扩展集群主机、给集群添加新的大数据服务，扩展服务实例等操作；结合预制的告警模板和自定义告警，使用户可以清晰掌握集群和集群中所有服务组件的运行状况。

### 1.2.2 一键开启或关闭 Kerberos 安全模式, 为大数据服务的安全保驾护航

数据和服务的安全保障，一直是企业非常重视的问题。USDP 3.0 具备大数据服务安全性保护能力，通过流程化配置，快速开启对集群服务的检测，结合细粒度的权限控制能力，使得大数据集群服务及数据的安全性整体上得到保护。

在 USDP 中，支持了向导化和自动化管理安全模式，通过 Kerberos 的运用，为集群中的用户、服务和主机提供身份认证和授权管理能力，其强大的安全性和跨平台支持特性，确保只有经过身份验证的用户才能访问受保护的资源，为



集群免受未经授权的访问和攻击提供了强有力的保护，帮助集群实现高度的安全性和可靠性。集群中的各个组件和服务（如 HDFS、YARN 等）可以通过 Kerberos 进行认证和授权管理，从而保护数据和应用程序的安全性。助力企业提升管理效率和信息安全水平。

### 1.2.3 大数据集群平滑升级，为需求升级提供有力支撑

集群的持续平滑升级能力，是保障用户基于该平台构建和管理的大数据服务，并获得持续维护和升级的重要支撑，平台及服务的灵活扩展能力，是应对随业务需求不断变化的架构优化的迫切需要。

企业在大数据相关业务方面的不断拓展和深入，数据类型变得更加丰富，数据量级爆发式增长，对数据处理时效的需求不断提高。因此，对于企业前期围绕数据仓库技术构建的大数据平台系统也带来了更高要求，亟需进行系统性升级和技术架构拓展，以满足企业业务持续发展的需要。

USDP 支持集群的持续升级和组件特性等持续拓展。用户可灵活选择大数据平台架构。无论是数仓架构的优化、流式计算的引入，甚至向流批一体架构、湖仓一体等架构演进，通过 USDP 都能获得支撑。

### 1.2.4 配置角色组管理，有效提升资源利用率和管理效率

角色组的功能，是将服务配置按实例角色类型分配给相应的角色组，组中各个角色继承这个组配置，助力大规模分布式服务便捷管理；根据集群主机环境、服务的特殊要求，为不同的主机或服务自定义分配不同的角色组，从而达到资源利用率和管理效率的有效提升。

### 1.2.5 UDH 更新及发布，多版本可灵活选择

UCloud 大数据组件发行套件，集成了丰富的开源项目，使企业可灵活构建一个功能先进的大数据系统；套件提供了强大的自动部署、管理和监控工具，便于用户操作维护大数据集群；套件中包含了更多的补丁和功能特性，为分布式大数据系统提供稳定性和性能保障。本次 UDH v3.0.0 的发布中，使其囊括的数据采集工具、流批计算引擎、调度系统、存储系统等服务，可全面兼容 Hadoop



### 3.3.4 版本生态。

## 1.2.6 宿主环境的修复能力，为运维人员减负

大数据集群宿主环境的修复及初始化能力，大大简化了基础运维在搭建大规模集群前，所需要进行的繁琐的基础设施环境准备工作；自动化修复工具良好的幂等性支持，能有效控制手动运维的出错概率，极大的降低运维人员的工作量和维护成本，保障了系统稳定性、可靠性及安全性。通过工具自动化的能力实现快速部署、配置、升级等复杂操作，使用户更加专注于数据分析业务的推进。

## 1.2.7 轻量便捷的操作

面对众多大数据服务，USDP 试图做到傻瓜式的“一键操作”，即，一键环境检查，一键环境修复，一键部署。省去用户费力排查环境问题，配置时间同步，系统优化等相关工作。

相比于目前已存在的部分运维工具，USDP 支持极其方便的傻瓜式部署方案，即，用户无需提前四处准备离线包，直接启动 USDP 服务后，通过界面的简单交互，在离线环境中即可完成一切部署过程。并且在部署的过程中，有详细的日志可供查看。

## 1.2.8 全面的生态支持

与同类产品不同，USDP 以完全中立，无捆绑的立场支持了超过 23 种服务，并将持续支持更多服务与组件的协同工作。

同时，USDP 产品中的服务完全基于 Apache 版本进行集成，并进行大量兼容性测试与压力测试，确保各个服务稳定运行的同时，还能兼顾 Apache 官方频繁的 BUG 修复。因此，用户也无需承担额外的学习成本。

## 1.3 产品架构



图 1-1 USDP 智能大数据平台产品架构图

当今数字化时代，数据已成为企业成功的关键要素。为达到有效地获取、存储、处理和分析巨量规模大数据的目的，企业需要安全、强大、全面的大数据解决方案。USDP 正以此为定位，提供企业级安全灵活、自主可控、生态丰富、高度兼容、稳定可靠、管理简便、易于维护的基础大数据平台，USDP 解决方案涵盖了数据接入、离线分析、实时计算、交互式查询、数据检索、资源调度、任务调度、元数据管理、数据安全、数据存储等能力，同时 USDP 统一管理门户提供丰富的管理工具及功能，为企业一站式建设大数据解决方案提供有力支撑。

在 USDP 支持的生态服务使用场景方面，可以灵活根据业务需要进行组件搭配组合，支持实现离线分析场景、实时数据计算场景、交互式分析查询场景、海量高性能数据检索查询场景、数据科学及价值挖掘等其他场景。USDP 提供了多种丰富的工具和技术，使大数据平台能够轻松地连接到业务数据源，提供对结构化、半结构化、非结构化等各种类型和格式数据的批量或实时采集接入。

离线分析是大数据平台核心价值之一，通过离线分析任务，企业可以不断地推进数据价值挖掘、洞察业务，从而促进对业务决策的支撑。USDP 提供了全面的批处理框架及计算引擎，帮助企业从数据资产要素中获得更多价值。

行业的飞速发展，带动了企业在信息数据的获取、处理、反馈等方面有了更

加苛刻的时效性要求，实时计算带来的数据洞察与快速决策辅助，正是应对这一要求的核心有效途径。USDP 提供了 Flink、SparkStreaming 等优秀的高性能实时数据处理引擎，满足性能要求的同时，提高业务吞吐量和高并发支撑，使得企业能够立即处理流式数据，快速获得决策和价值变现。

数据价值分析，是一场持久战，数据科学家，数据工程师需要不断地在已沉淀的数据资产要素中探索和挖掘。灵活的交互式查询，统一的 SQL 分析语法可极大的简化分析算法要求，从复杂的算法逻辑开发中解放出更多的时间，有效的优化开发成本。而海量数据价值的极速检索及高并发、高性能访问支持也十分关键，USDP 提供了性能优异的 HBase 分布式数据库，来满足基于海量数据随机读写的高性能场景，支撑业务用户和数据分析师的要求。

USDP 支持基于对象存储的存算分离架构，并且在异构存储平台的基础上，提供软硬件缓存加速方案，使得对数据的读写效率得到大幅提升。而数据湖架构进一步优化了传统数仓架构的诸多难题，流/批增量数据的分钟级入湖效率及流式数据读取，这些存储方案进一步促进企业获得数据安全可靠性和可扩展性的能力，及数据要素管控的便捷性。

USDP 提供的大数据生态技术，已全面兼容国产化软硬件设备运行环境，包括 x86 及 ARM 架构的国产化芯片和操作系统。并能稳定健康的运行。USDP 提供完备的工具体系和管理功能，如自动化部署、多集群管理、安全管控、服务控制、日志在线检索、监控告警和平滑升级等管理服务，不断夯实此类管理功能，将是平台稳定运行和高效维护的保障基础。

## 1.4 应用场景

### 1.4.1 离线/实时数仓架构场景

离线数仓将数据从源系统中抽取出来，经过清洗、转换和加载（ETL）等步骤，使数据按照一定的规则组织到数据仓库中，再通过报表等方式对数据进行分析和挖掘。而离线数仓所擅长处理的大规模数据能力，却不可避免的带来一定的数据延迟性，而实时数仓解决了这一问题。

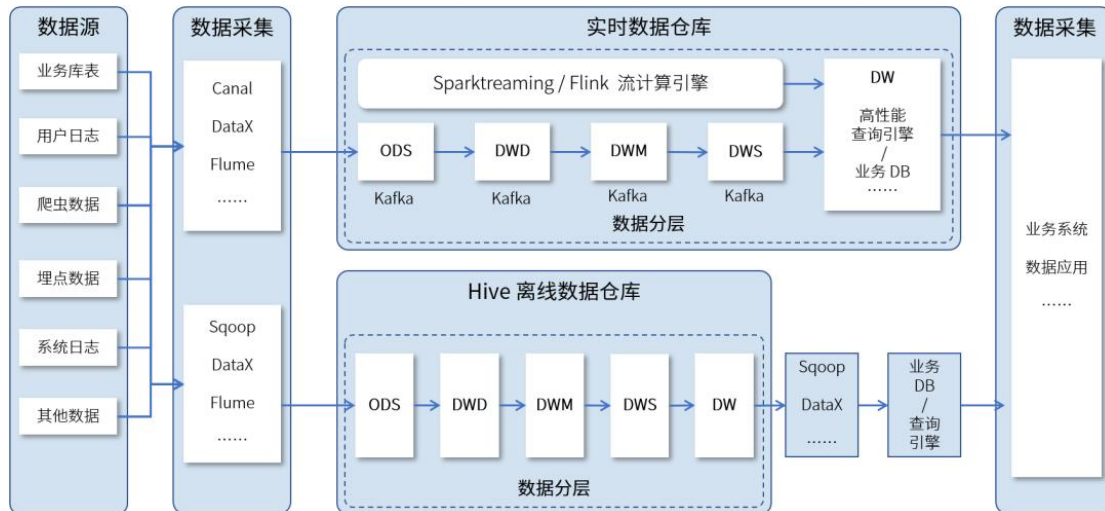


图 1-2 离线/实时数仓架构场景架构图

在实时数仓中，以实时或近乎实时的方式处理数据，将数据通过流式处理引擎（例如 Kafka、Flink 等）实时抽取、清洗、转换和加载到数据仓库中，再通过可视化工具等方式对数据进行实时监控和分析。基于 USDP 中丰富的大数据技术的选择，企业可根据自身的发展需要，灵活搭建并不断优化整合自己的大数据平台架构，构建适宜的数仓架构场景。

### 1.4.2 流批一体架构场景

流批一体架构的思想是将流处理和批处理整合在一起，提供更加全面和高效的实时数据分析能力。如使用 Flink 引擎及 Kafka 消息队列等引擎和工具的组合架构，可以在实时数据流中进行复杂的事件驱动处理，并兼顾批处理任务。

依托于 Flink 引擎提供的丰富的流处理和批处理 API，和强大的状态管理和容错能力，使得流批一体架构可以更加可靠和高效地处理数据。在减少架构的复杂性和维护成本的同时，使企业达到更加灵活地处理数据的目的，并更好地适应业务需求的变化。而 USDP 中提供了丰富的大数据生态服务、工具和框架的支持，能很好帮助企业构建灵活且复杂的大数据处理架构。

### 1.4.3 数据湖/湖仓一体架构场景

在将数据湖和数据仓库的优点深度结合的“湖仓一体”数据架构中，以多种格式的数据统一存储为基础，可很好的避免数据冗余和一致性问题；统一高效的

数据处理和清洗带给数据质量和可用性有力保障；在该架构的支撑下，企业可开展如批处理、流处理、实时处理等多种模式的数据分析方式，借助 SQL 查询分析、OLAP 分析、数据挖掘、机器学习来满足应用的分析需求；通过该架构，更有利于企业实现数据的管理和治理，从而提高数据的可靠性和可信度。

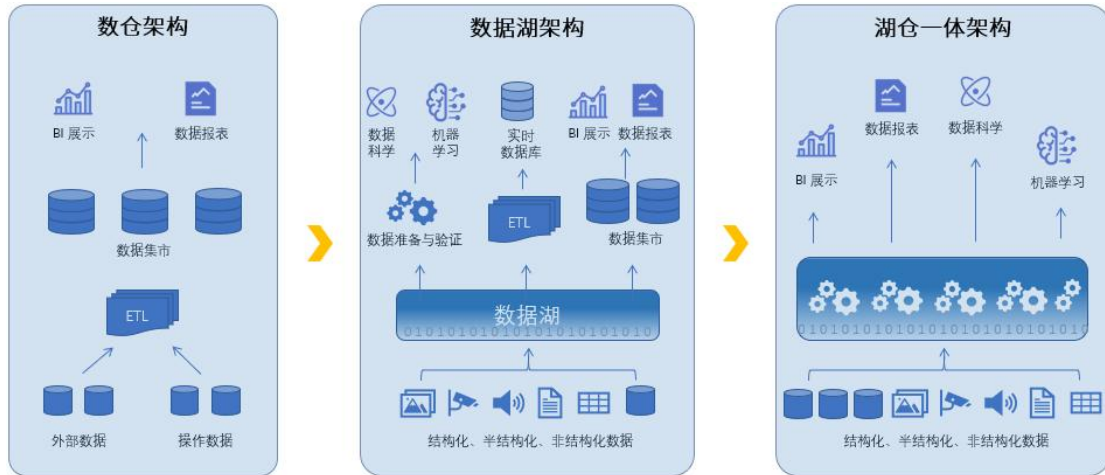


图 1-3 由数据仓库架构向湖仓一体架构演进

## 2 名词解释

**专业版：**通过 USDP 快速构建并管理中小规模大数据基础平台。

**企业版：**通过 USDP 提供稳健的企业级大数据平台的管理特性及安全的控制能力。

**信创版：**通过 USDP 轻松获得技术生态全面兼容国产化软硬件环境的信创大数据平台。

**实例：**大数据组件服务运行的服务实例，例如在集群的某主机中正在运行 DataNode 进程就是 HDFS 服务的一个实例。

**角色配置组：**某一类配置分组，例如 DataNode 配置都归属于默认 DataNode 角色配置组。支持自定义配置项（基于 default role group 默认角色组生成）。

**UDH：**UCloud's Distribution Including Apache Hadoop（UCloud 大数据组件发行版）。

**集群管理视图：**即集群管理首页，该页面可全面操作、控制、监控整个集群的常规且核心的管理功能，提升操作便捷性及管理效率。



## 3 功能简介

### 3.1 USDP 核心功能

- ◆ 统一管理控制台，提供强大的大数据集群管理和监控工具。可便捷管大数据集群，掌握平台运行情况，确保平台及集群的稳定性和可靠性。
- ◆ 多集群管理能力，可配合底层网络策略实现多集群间的跨网络跨地域的访问及隔离；
- ◆ 集群主机管理能力，可线性扩展集群资源、主机状态监控、主机实例管理等；
- ◆ 集群大数据服务组件管理能力，对集群服务全方位的管理和维护、服务和实例的监控；
- ◆ 集群资源池可视化配置管理能力，促进集群计算资源池充分管理和优化，提升资源利用率；
- ◆ 提供对 YARN 集群中所有任务的实时监控，以及任务的资源利用情况和执行状态；
- ◆ 角色化配置管理,Key Value 级别配置同一角色组中的所有实例配置，并支持部分实例自定义扩展配置项；
- ◆ 可视化的服务配置管理、历史版本记录及回滚、实例配置过期待重启提醒等；
- ◆ 一键开启或关闭 Kerberos 安全模式；
- ◆ 一键配置或关闭 LDAP；
- ◆ 中心化管理 UDH 源，按需自动下载并安装所需大数据服务组件；
- ◆ 集群级管理视图，包括服务便捷操作及集群资源监控等；
- ◆ 自定义监控指标扩展，对业务服务实现精细化全面的监控能力；



- ◆ 自定义监控图表，满足对重点服务所关注的指标进行归类展示，便于轻松掌控集群及服务的稳定性、高性能服务表现。
- ◆ 丰富的告警模板套用和灵活的自定义告警设置能力；
- ◆ 服务异常终止时自动拉起；
- ◆ 支持邮件、企业微信、钉钉、飞书等多种告警通知方式；
- ◆ 完善的主机环境修复及初始化工具；
- ◆ 支持持续平滑升级能力，提供管理工具与服务升级、大数据集群升级能力；

## 3.2 UDH 发行版所含大数据生态服务列表

服务	版本	描述
HDFS	3.3.4	Apache Hadoop 分布式文件系统 (HDFS) 是 Hadoop 应用程序使用的主要存储系统。HDFS 创建多个数据块副本并将它们分布在整个群集的计算主机上，以启用可靠且极其快速的计算功能。
YARN	3.3.4	Apache Hadoop MapReduce 2.0 (MRv2) 或 YARN 是支持 MapReduce 应用程序的数据计算框架（需要 HDFS）。
Hive	3.1.3	Apache Hive 数据仓库软件有助于使用 SQL 读取，写入和管理驻留在分布式存储中的大型数据集。
TEZ	0.10.2	Apache TEZ 是支持 DAG 作业的开源计算框架，它可以将多个有依赖的作业转换为一个作业从而大幅提升 DAG 作业的性能。
HBase	2.4.13	Apache HBase 是 Hadoop Database，是一个高可靠性、高性能、面向列、可伸缩的分布式存储系统。
Phoenix	5.1.2	Apache Phoenix 是 HBase 的 SQL 驱动，Phoenix 使得 HBase 支持通过 JDBC 的方式进行访问，并将你的 SQL 查询转成 HBase 的扫描和相应的动作。
Impala	4.1.1	Apache Impala 分布式计算服务。
Flink	1.16.1	分布式计算引擎

服务	版本	描述
Spark	3.2.3	Apache Spark 是一个通用的大数据分析引擎,具有高性能、易用和普遍性等特点。
Hue	4.11.0	Apache Hue 可视化管理服务。
Datax	3.0	数据采集与转储服务
Sqoop	1.4.7	Apache Sqoop 是一种用来将 Apache Hadoop 和结构化数据存储 (如关系数据库) 中的数据相互转移的工具。
Flume	1.11.0	Flume 是 Cloudera 提供的一个高可用的, 高可靠的, 分布式的海量日志采集、聚合和传输的系统, Flume 支持在日志系统中定制各类数据发送方, 用于收集数据; 同时, Flume 提供对数据进行简单处理, 并写到各种数据接受方 (可定制) 的能力。
Kafka	2.8.1	Apache Kafka 是一种高吞吐量的分布式发布订阅消息系统。
Oozie	5.2.1	Oozie 是群集中管理数据处理作业的工作流协调服务。
DolphinScheduler	3.1.4	Apache DolphinScheduler 是一个分布式易扩展的可视化 DAG 工作流任务调度开源系统。
Solr	8.11.2	Apache Solr 基于 Lucene 的流行、高性能的开源企业级搜索平台。
Prometheus	2.38.0	Prometheus 是一个开源的系统监控和报警系统用于拉取大数据平台监控数据。
Atlas	2.3.0	元数据管理服务
Ranger	2.3.0	Apache Ranger 授权服务。
Zookeeper	3.5.10	Apache ZooKeeper 是用于维护和同步配置数据的集中服务。
Canal	1.1.6	Canal 是 MySQL binlog 增量订阅和消费组件
EFAK	3.0.1	是一个监控系统, 监控的 kafka 集群、可视化消费者线程、偏移量、所有者等
Hudi	0.13.0	Apache Hudi 是一种数据湖的存储格式, 在 Hadoop 文件系统之上提供了更新数据和删除数据的能力以及消费变化数据的能力。
Iceberg	1.3.0	Apache Iceberg 是一个开源的大数据表格式和存储库, 它提供了一种可靠的、可扩展的、版本化的数据表存储方案,

服务	版本	描述
		专为大规模数据湖和数据仓库场景设计。
Kylin	4.0.3	Apache Kylin 是一个开源的、分布式的分析型数据仓库，提供 Hadoop/Spark 之上的 SQL 查询接口及多维分析 (OLAP) 能力以支持超大规模数据。
Kyuubi	1.7.1	Apache Kyuubi 一个分布式和多租户网关，用于在 Lakehouse 上提供 Serverless SQL。
Trino	418	Trino 是一个以惊人的速度运行的查询引擎 快速分布式 SQL 查询引擎，用于大数据分析，帮助用户探索数据宇宙。
Zeppelin	0.10.1	Apache Zeppelin 是一个基于 Web 的笔记本，可以启用数据驱动，交互式数据分析和与 SQL，Scala，Python，R 等的协作文档。
ZKUI	2.0	ZKUI 是一个基于 Web 的 Apache ZooKeeper 管理界面，用于简化 ZooKeeper 集群的管理和监控。

表 3-1 3.2 USDP UDH 发行版所含大数据生态服务列表

## 4 部署 USDP

详细参见 USDP 安装手册。

## 5 访问 USDP 控制台

### 5.1.1 在浏览器中访问 USDP 控制台

在浏览器中访问 USDP 控制台：

```
http://<usdp_server_ip>:<usdp_server_port>
```

提示：

- ◆ “usdp\_server\_ip” 为启动 USDP-Server 的节点 IP 地址。如果浏览器所在节点与 USDP 不在同一网络环境，则需要自行搭建 VPN，或通过 USDP 部署节点网络互通的 Windows 机器的浏览器来访问该 IP。
- ◆ “usdp\_server\_port” 为 inventories/hosts.yml 配置文件中 usdp\_server\_port 的取值。
- ◆ 默认用户名和密码为 admin / admin （登录后建议修改默认密码）

进入 USDP 管理控制台时，登录界面如下图所示：

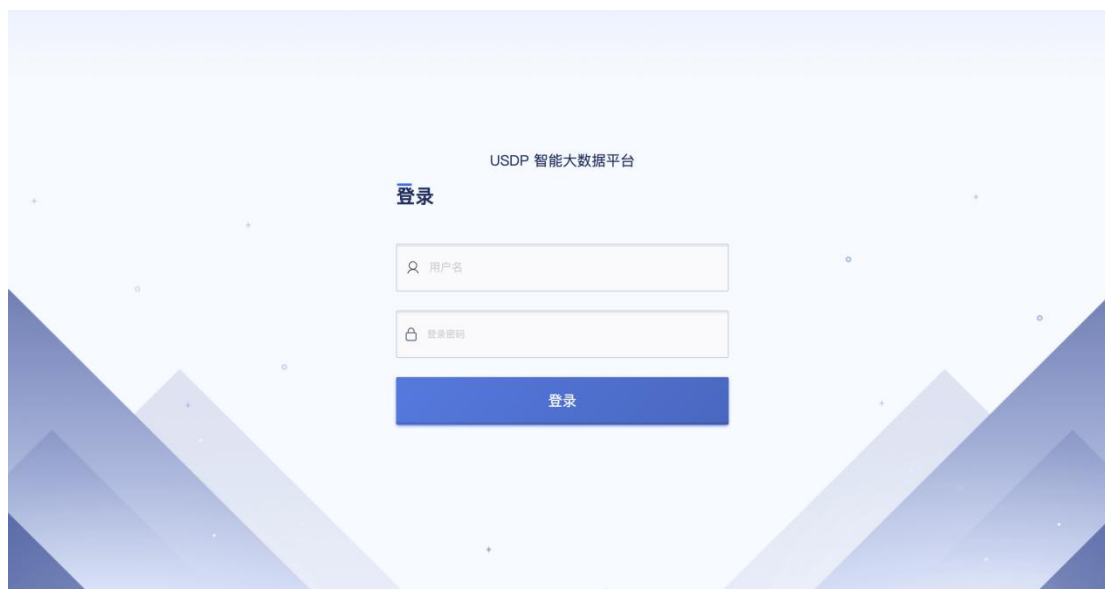


图 5-1 USDP 控制台登陆页面

登录后可在 USDP 控制台中，更改 admin 管理员的密码。

## 6 多集群管理

本章节中，即将介绍通过 USDP 智能大数据平台对于多集群的创建、切换、与管理等相关操作。其中集群为 USDP 支持的并且以服务目标为主灵活创建的多个分布式服务，如围绕 Hadoop 的生态服务集群、Kafka 集群等。

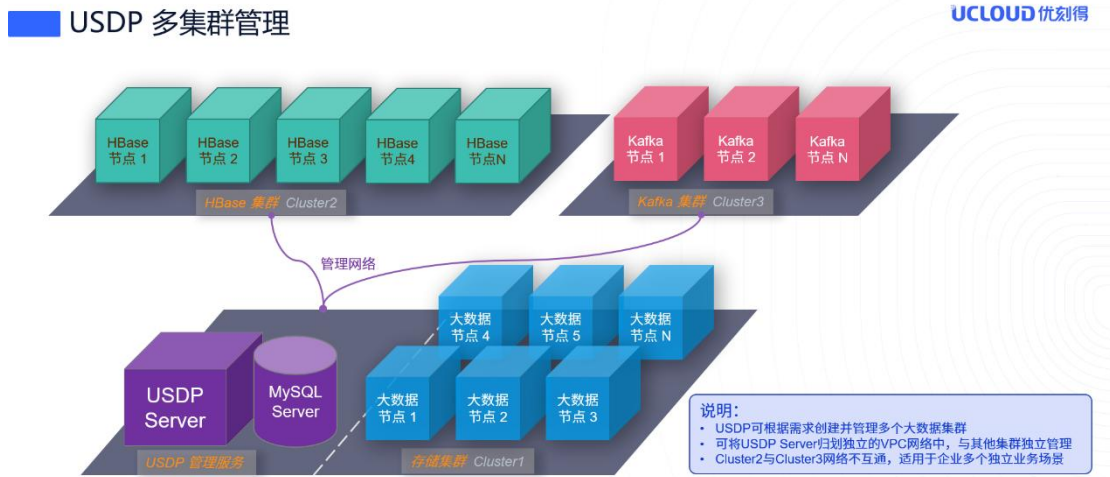


图 6-1 USDP 多集群管理示意图

如图所示，可结合网络 VPC (vlan 等) 技术实现 VPC 间的互通管理，大数据集群 Cluster1 与 Cluster2 可能承载的是不同且不相关的大数据处理业务，因此无法互相访问和共享数据。

基于此类场景，在使用 USDP 管理大数据业务系统时，使用者可根据业务需求，来规划所需的大数据集群数量及单集群的规模和服务；根据业务形态，决策集群间的隔离性；对集群层面进行管理操作。

### 6.1 创建大数据集群

本章节指南，将介绍快速规划并创建一个 USDP 智能大数据集群。为了便于用户快速创建大数据集群，USDP 管理服务已将大数据集群创建过程中的众多服务、配置整合成向导的形式，用户可通过向导中的指引快速搭建出一个全新的大数据集群。

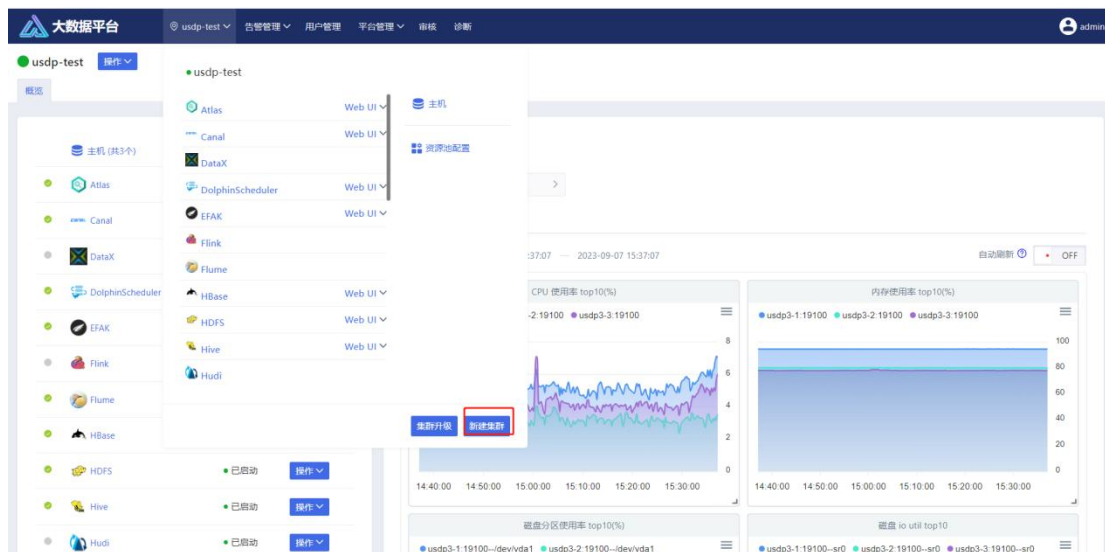


图 6-2 USDP 控制台-集群切换及新建集群入口

登陆 USDP 控制台时，默认会选择集群 ID 号靠前的集群，并显示该集群管理视图页面，此时点击顶部菜单栏“当前集群”菜单，会弹出下拉菜单中展示以创建并管理的集群列表，以及当前集群内已部署使用的大数据服务列表，在下拉菜单底部点击“新建集群”按钮。此时，即可进入创建新集群的向导流程中，后续操作与《创建首个大数据集群》章节的操作步骤一致，可前往参考。

### 6.1.1 集群创建向导

首次安装 USDP，系统会提示“新建集群”，点击“新建集群向导”按钮，即可快速搭建一个全新的大数据集群。

若在正常使用 USDP 过程中，需要创建新的大数据集群时，可点击 USDP Web 控制台顶部的“集群”菜单，在弹出的下拉菜单列表中，点击“新建集群”按钮，即可进入创建向导中。

#### Step1: 选择软件版本

进入向导第一步，此时用户需要定义即将新创建的集群的基本信息，主要包括“集群名称”、集群创建需要依赖的 USDP “软件版本”、以及集群框架类型。

- ◆ 集群名称：输入即将创建的集群的名称，创建成功后该名称不可更改；



- ◆ 软件版本：选择 3.1.4；确认与 udh 包版本一致
- ◆ 集群架构：选择 Hadoop 框架

系统会在下方展示出 V3.1.4 版本所包含的所有可使用服务及服务的简要描述信息。

## Setp2: 指定集群节点

手动输入相应节点主机的完全限定域名 (hostname) 即可，若未曾修改节点的 hostname，则根据 hosts.yml 文件中相应的配置内容进行填写。

为了方便批量录入，USDP 支持用户以表达式的方式输入，规则如下：

- ◆ 可单行输入每一个节点的完全限定域名；
- ◆ 可通过 “[ ]” 辅助输入有数字规律的节点完全限定域名；例如 usdp3-[1-3] 表示包含 “usdp3-1”、“usdp3-2”、“usdp3-3” 共三个节点。

USDP 需要通过 “root” 用户以及 “22” 端口，进行管理所有主机服务器。暂不支持对端口号进行更改。

USDP 需要对指定的主机进行环境检测，例如节点间的网络互通性，待检查完成，将在页面中显示所有节点检测的状态。点击节点列表中“点击查看详情”栏的相应字样，即可查看该主机上的已检测以及正在检测的日志详情。

待所有节点连接成功后，选择要加入集群的节点，点击“下一步”。

## Setp3: 检查节点环境

在该步骤中，USDP 需要检查对应节点是否已经初始化完成，比如：JDK、MySQL、时间服务器等是否已经正确安装及配置等。

待所有检查工作完成后，点击右下角的向导“下一步”按钮继续。

检查过程中，可点击列表“详情”栏的“检查中”等状态按钮查看对应节点的任务详情。

当检查过程中出现失败的情况，首先可尝试点击右侧的“全部重试”按钮，如果仍旧报错，此时可重点查看并参考失败详情及日志说明来定位失败具体原因，必要时登录对应主机进行修复后重试；

## Setp4: 配置 UDH 源

USDP 需要通过 http 协议的源地址，为创建集群的各个主机提供安装包的支持，因为此处 USDP 将自动检测 UDH 源，并完成相应配置更改，所以需要在创建集群前，前往 USDP Web 控制台“系统管理” - “UDH 源”中配置正常的源。此处参见本文档《配置 UDH 源》章节内容进行操作。

## Setp5: 选择服务

UDH 源中支持的大数据生态服务较多，在集群安装时，使用者可根据实际业务需要，灵活选择所需安装的服务和组件。

在选中需要的大数据服务后，点击“下一步”按钮时，USDP 将自动检测使用者已选择的服务所需要与其他服务存在依赖的关系。若所选服务需要依赖没有选中的服务时，可根据页面中提示信息，调整依赖或被依赖的服务后，继续执行向导流程。

## Setp6: 选择实例安装主机

UDH 中支持的大数据生态服务，均由不同的实例构成，安装集群时，使用者可根据业务需求做好服务在集群主机中的分布规划，将不同的服务及实例分别安装到集群的哪些主机上去。

USDP 会根据使用者在上一步中选择的大数据服务，自动计算各服务在当前集群规模中如何分布和部署，以推荐方案的形式选择好分布情况，并呈现给使用者。此时，使用者可在推荐方案的基础上，进行部分调整。调整后，点击向导“下一步”按钮。

## Setp7: 服务配置

到达该步骤，需使用者根据选择的服务及硬件环境进行部分配置项设置，如

某服务的数据存储路径、需要的数据库信息填写等等信息，可点击右上角“显示所有说明”进行参考。

使用者自行配置之前，可参考 USDP 默认根据集群环境、服务配置要求等信息补充在页面中的各配置项信息，如服务管理元数据信息、默认填充 USDP 初始化时已部署的 MySQL 的相关连接信息，使用者仅需要点击“测试连接”按钮即可测试是否正常。

USDP 允许使用者将服务元数据存储的数据库更改为使用者已使用并维护的其他 MySQL 数据库；仅需要在此处配置该库的连接信息并测试连接成功。

如 HDFS DataNode 数据目录的设置中，若集群各主机存在多块数据盘，可根据实际规划及需求，手动点击“添加”可设置 HDFS 多块数据盘的文件路径。注意：数据目录路径需填写绝对路径，如/data1/usdp/hadoop/hdfs/dn，其中 /data1 目录已挂载数据盘，子目录不需要创建。

## Setp8: 配置信息总览

完成前几个向导步骤中的信息配置后，USDP 会将所有选择及配置信息汇总展示。使用者浏览检查无误后，点击向导“开始部署”按钮继续执行。

## Setp9: 安装并启动服务

向导到达此步骤，USDP 根据配置信息自动生成部署任务计划，并开始逐项执行服务安装工作。

当服务安装发生错误时，使用者可以点击查看报错节点“失败”详情，参考详情提示信息，进行手动修复错误操作，或点击“全部重试”按钮，重新进行服务组件部署工作。

在部署过程中，各主机中正在进行的部署进展，均可通过进度条实时展示，便于使用了解执行进展。可在“点击查看详情”栏中的各个链接中，查看当前执行步骤的详细情况。安装任务执行日志，可点击执行成功/执行中的各任务名称进行查看。

## Setp10: 集群创建完成

待所有任务执行完成，主机进度条显示 100% 时，即表示集群已创建完毕，点击“下一步”按钮，便会看到集群创建成功的欢迎页面。此时，退出安装向导，即可进入该集群管理视图页面。

## 6.2 集群切换

当前 USDP 服务管理着至少 2 个以上的集群时，可点击控制台顶部菜单栏“当前集群”菜单，会弹出下拉菜单中展示已创建并可管理的集群列表，选择其中任一集群点击其“集群名称”，即可切换至该集群管理视图页面。

## 6.3 非 Hadoop 生态数据集群

USDP 支持对非依赖 Hadoop 的部分其他数据处理集群的创建与管理，例如 Kafka 集群，Kafka 仅依赖 Zookeeper 进行分布式协调器，因此可以创建独立的 Kafka 集群来提供消息队列服务。

### 6.3.1 适用场景

- ◆ 一套管理服务，即可管理 Hadoop 大数据集群，又可管理其他数据处理集群的场景；
- ◆ 获得统一的集群级、服务级监控服务；
- ◆ 统一的集群服务异常告警服务；

USDP 允许非依赖 Hadoop 相关服务的数据服务与 Hadoop 集群共享节点资源。在实际使用中，建议使用者严格根据不同服务集群的计划使用场景和资源消耗预估，合理规划集群节点及资源分配，使得各大数据发挥出最优的服务性能和表现。

### 6.3.2 参考架构图

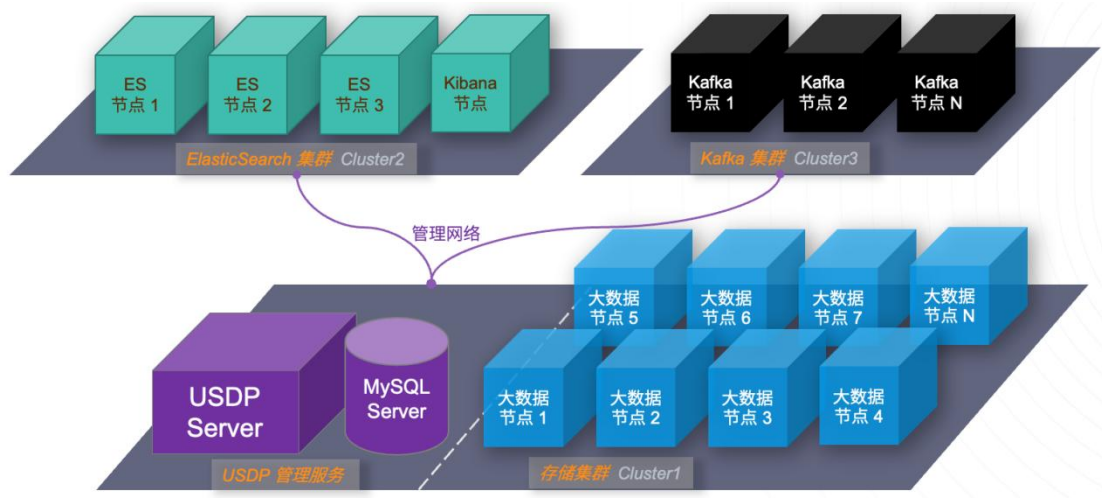


图 6-3 USDP 管理多个不同类型的大数据业务集群

### 6.3.3 资源规划

可参考本文档 《部署规划 – 多集群多节点规划》 的方案完成本次规划。

注意：多个集群无法共用同一台主机节点，不可重复使用！请合理规划节点用途；

### 6.3.4 示例-创建 Kafka 集群

登陆 USDP 控制台时，默认会选择集群 ID 号靠前的集群，并显示该集群管理视图页面，此时点击顶部菜单栏“当前集群”菜单，会弹出下拉菜单中展示以创建并管理的集群列表，以及当前集群内已部署使用的大数据服务列表，在下拉菜单底部点击“新建集群”按钮。

进入创建新集群的向导流程中后，请注意在“集群框架”表单处，选择“Hadoop 框架”，并继续向导流程。

接下来，请注意在向导“选择服务”页面处，可仅选择 Kafka 相关服务即可，此处可严格参照使用者对这部分规划。其他向导内的操作步骤，可参考本文档《创建首个大数据集群》章节指引，流程基本一致，此处不在赘述。

## 7 单集群管理

在本章节中，将介绍如何通过 USDP 智能大数据平台管理单个大数据集群，其中将涉及到对集群的节点管理、集群内运行的大数据生态服务的管理、服务的状态及分布管理、服务的配置管理、服务原生 Web UI 的集成、服务的监控管理等。

### 7.1 集群管理

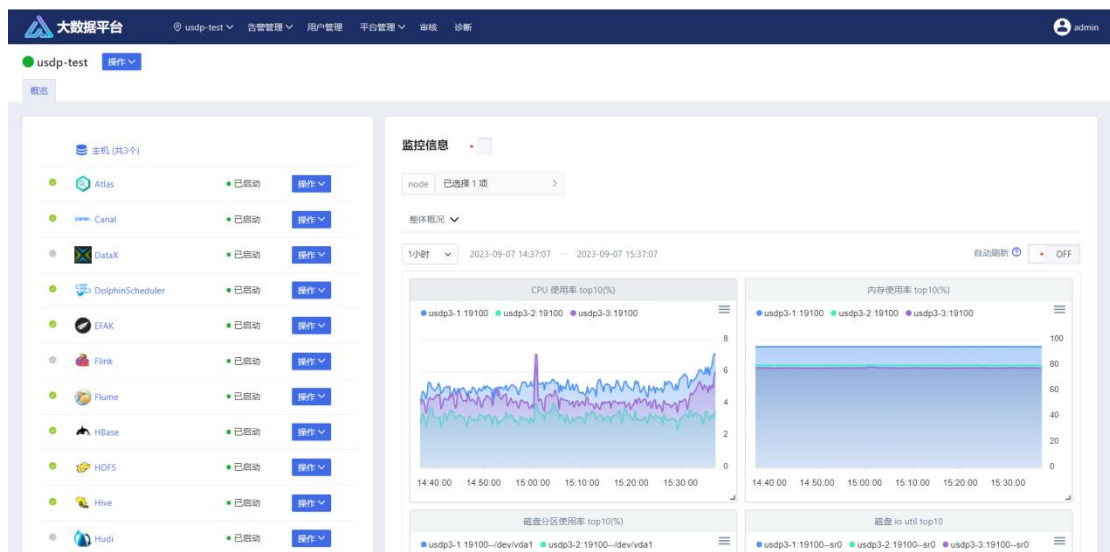


图 7-1 USDP 服务集群管理视图 – 集群管理视图

当用户在 USDP 控制台操作并切换至某一集群时，首先将进入该集群管理视图页面，在该页面中提供的信息及操作，主要包括集群级控制、集群主机管理、集群大数据服务预览、集群监控数据概览四方面的管理内容。

#### 7.1.1 集群级控制

在集群管理视图的顶部，主要展示了集群的工作状态（绿色圆点表示运行中）、集群名称、对集群的便捷操作等。

集群工作状态有：

状态	说明
----	----

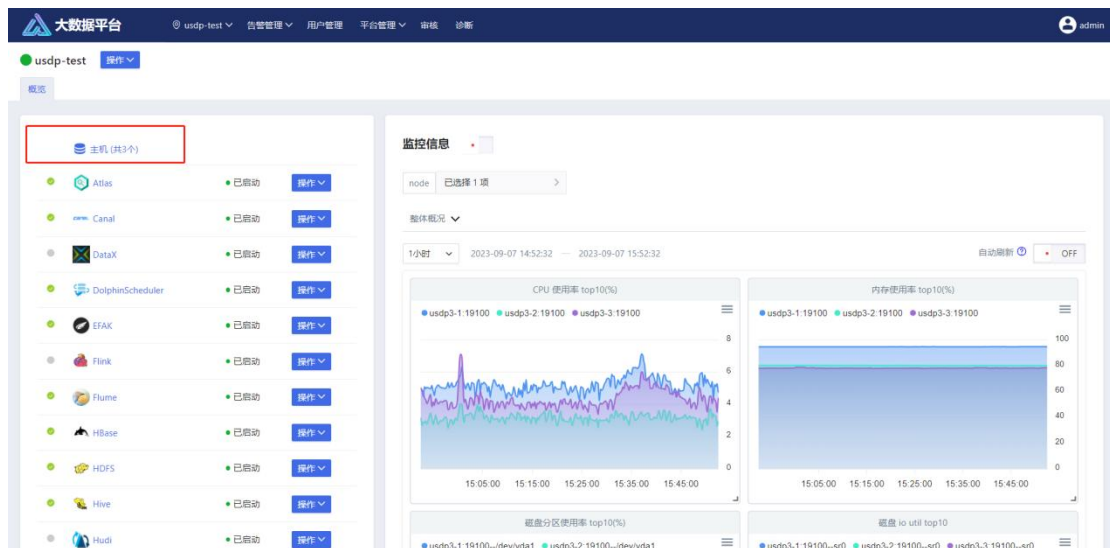


状态	说明
<span style="color: green;">●</span>	运行中
<span style="color: orange;">●</span>	集群新建中
<span style="color: yellow;">●</span>	集群部署中
<span style="color: blue;">●</span>	集群维护中
<span style="color: blue;">●</span>	集群变更中

点击集群管理视图中，集群名称右侧的下拉菜单，即可找到对集群便捷操作的功能，包括添加主机、启用 Kerberos、添加服务、设置 LDAP、解除 LDAP 等。

### 7.1.2 集群主机管理

在集群管理视图“概况”选项卡页面左侧，统计展示了当前集群所使用并管理的所有主机节点的数量，点击此处即可进入集群主机统一管理页面。关于对主机的控制操作，在后面的章节中进行具体介绍。



### 7.1.3 集群大数据服务及服务操作

在集群管理视图“概况”选项卡页面左侧，还展示了当前集群中已部署并处于工作状态的所有大数据生态服务的列表，并逐一展示了各服务的状态、服务 Logo 及名称、操作状态、服务级便捷操作按钮等。



服务工作状态有：

状态	说明
	可用状态
	部分可用状态
	不可用状态
	客户端实例不适用状态

对大数据服务的便捷操作主要有，服务的“启动”、“停止”、“重启”、“滚动启动”、“滚动重启”、“删除”等操作，这些将在后续章节中具体介绍。

### 7.1.4 集群监控数据概览

在集群管理视图“概况”选项卡页面右侧，展示了当前集群整体的监控数据，如各节点 CPU 使用率情况、内存使用率、磁盘使用率等情况，鼠标移至相应的图标上时即可查看到对应时间刻度各主机节点的该指标具体情况，辅助用户判断和参考。

监控图表支持手动设置时间粒度并画图，包括最近的 1 小时、3 小时、6 小时、12 小时、1 天、3 天、7 天、自定义时间粒度；支持图表自动刷新展示。

## 7.2 主机管理

在集群管理视图服务列表上方点击“主机”按钮，即会进入该集群的主机管理界面。



图 7-2 集群主机管理页面

在该页面中，将集中对当前集群所管理的所有主机进行统一管理，包括对主机层面的操作、主机的资源使用情况概览及详细监控、大数据服务的分布查看，以及大数据服务组件的管理操作等。

主机管理页面中，可以查看到各台主机的关键信息，便于日常维护和使用，含主机名、主机的状态、主机的 IP 地址；还可查看各主机的资源情况，含主机 CPU 的核数、CPU 使用情况、主机的内存使用情况、主机的磁盘使用情况。

## 7.2.1 主机控制

对于集群主机的控制层面，支持主机的扩展添加、删除操作。

删除时，为避免误操作，系统会自动判断待删除主机上是否还有运行大数据实例，并提醒使用者，需要删除主机上的所有实例后才可以删除该主机。

## 7.2.2 主机控制及监控

除了在主机管理页面中可以查看到主机的部分关键信息外，点击各“主机名”可进一步查看该主机的详细信息及监控信息。

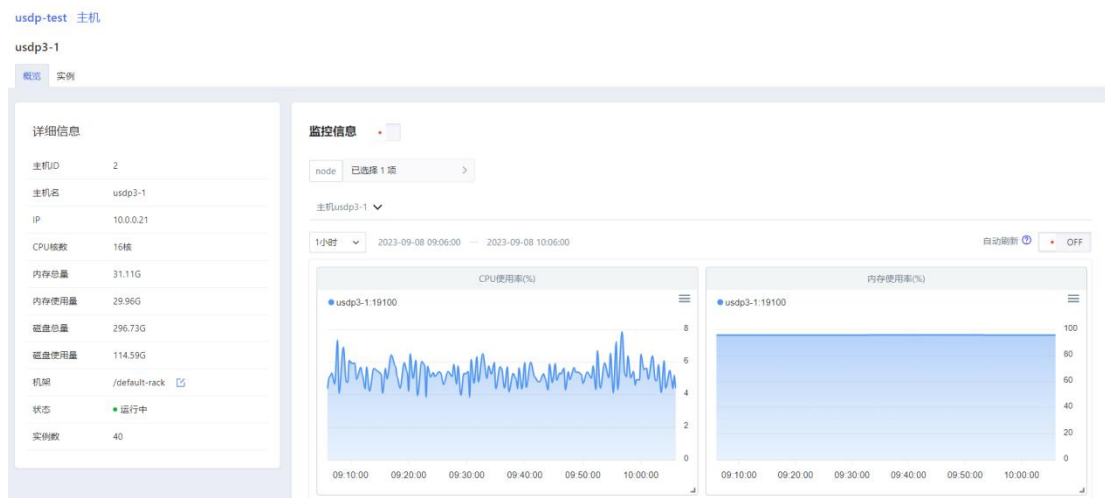


图 7-3 主机详情页面

在该页面中，左侧显示了该主机的详细信息，右侧显示该主机的监控指标信息；

监控图表支持手动设置时间粒度并画图，包括最近的 1 小时、3 小时、6 小

时、12 小时、1 天、3 天、7 天的；支持自定义时间粒度；支持图表自动刷新展示。

### 7.2.3 主机视角-大数据服务实例管理

主机管理页面中，可概览性的查看到各主机上正在运行的大数据服务“实例数量”。点击实例统计数字，可进一步查看运行在该主机中的大数据服务实例详情。

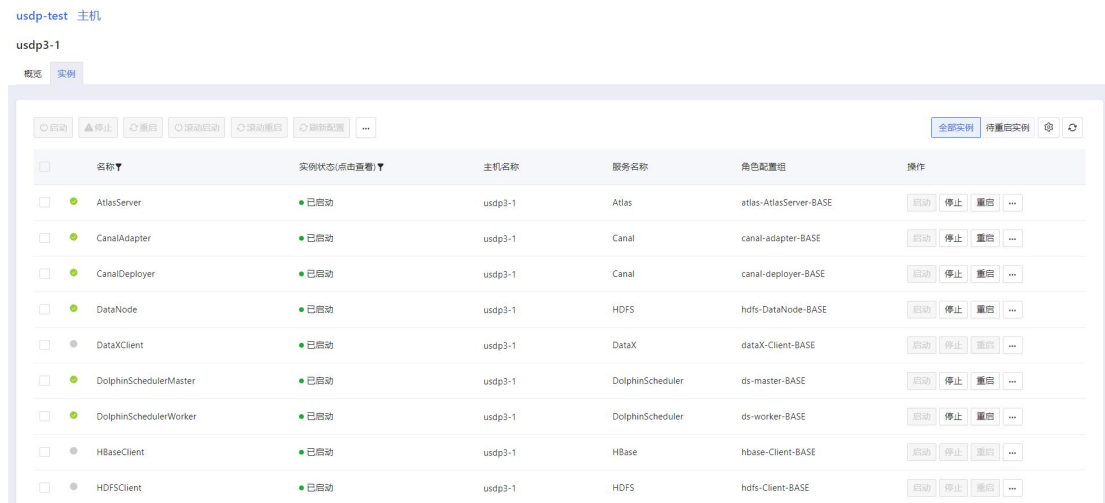


图 7-4 主机中运行的大数据服务实例列表

主机“实例”详情页面中，可清楚查看到，运行在当前主机中的实例列表，如上图所示，可查看到实例的运行“状态”、“名称”、“实例的启动状态”、实例所在的“主机名”、实例归属的“服务名称”、实例所属的“角色配置组”。

支持使用者对该主机上运行的大数据服务实例进行逐一/批量管理操作，如上图所示，可逐一/批量的进行“启动”、“停止”、“重启”、“滚动启动”、“滚动重启”、“删除”等操作。

值得注意的是，在对某个实例进行“删除”操作时，为避免误操作，需使该实例处于“停止”状态时，才可执行“删除”操作。

## 7.2.4 为集群扩展主机

为达到集群资源可分布式线性扩展目的，使用者可通过 USDP 提供的简单的环境修复及初始化工具，以及可视化的向导化方式，便捷的为集群批量扩展新的主机节点。

### 7.2.4.1 主机环境修复与初始化

#### 为待添加的主机配置 hostname

该部分，可参考本文档[《为主机配置主机名（所有节点）》](#)章节的内容进行操作。

**注意：**主机被添加进集群并且部署服务后，不允许修改主机名（服务配置中的主机名无法同步修改，将会造成大数据服务出现异常）

#### 为待添加的主机配置 ssh key

该部分，可参考本文档[《配置节点 ssh key（所有节点）》](#)章节的内容进行操作。

**注意：**需使用 `ssh-copy-id` 命令为新添加的主机节点的 `id_rsa.pub` 向所要添加进的集群内其他所有节点免密，并对 `USDP-server` 主机节点免密。

#### 将待添加主机的信息添加至 usdp-server 主机的 hosts.yml 文件

USDP 提供的修复及初始化工具，支持分布式环境的幂等性，因此，新增节点配置至 `hosts.yml` 中并执行时，保证了对原有集群以及管理服务节点的不受到影响。

找到 `usdp-server` 所在主机，编辑该主机的 `/data/usdp-srv/env-prepare/inventories/hosts.yml` 文件，在“`work_node`”中追加待添加主机的 `hosts`（IP 地址）和 `Node_name`（主机名，若此前并未配置主机名，则需要在此处配置该节点主机名），注意保持格式无误。

```
all:
  vars:
    # 如果是信创环境，操作系统是 Kylin V10，则需要打开下面的注释，指定 python 版本为
python3
    #ansible_python_interpreter: /usr/bin/python3

    # 如果下面指定的 3 个 ntp_server 不能访问互联网，请设置 with_internet_ntp: false
    # 并且打开 internal_ntp_servers 的注释设置至少一个内部可访问的 ntp server;如内网
ntp 及互联网 ntp 均无法提供，请保持默认 with_internet_ntp: true
    with_internet_ntp: true
    #internal_ntp_servers:
    # - 10.0.1.11
    # - 10.0.1.12

    # 设置时区，缺省设置为 Asia/Shanghai
    #timezone: Asia/Shanghai

    # SSH root 密码，按实际情况修改，如果已经配置了 ssh key 免密登录，则该变量无效
ansible_password: ssh_password_of_root
    # MySQL root 密码，按实际情况修改
mysql_root_password: xxx_mysql_root_password
    # usdp server 监听端口，按实际情况修改
usdp_server_port: 80
    #访问 MySQL 地址，填写实际 vip
mysql_vip: 10.0.0.24
    #MySQL 访问 domain，保持默认即可
mysql_vip_domain: mysql.vip.local

#####
# Groups
#####
# ntp server 选 3 个节点
ntp_server:
  hosts:
    10.0.0.21:
    10.0.0.22:
    10.0.0.23:

# usdp server
usdp_server:
  hosts:
    10.0.0.21:

# work node(s)
# 根据需要放开下面的 NODE_NAME 变量注释，并分别修改成需要的 hostname 值，来设置
```

```
主机的 hostname
# 如果不放开 NODE_NAME 变量, 且 hostname != localhost , 则不设置 hostname , 保留
当前主机名
# 如果不放开 NODE_NAME 变量, 且 hostname == localhost , 则设置 hostname 为 IP 字
符串 (替换 IP 中的 '.' 为 '-') , 例如: 10-76-198-19
#
# 主机名(NODE_NAME)-命名规则:
# 1、主机名只允许包含 ascii 字符里的数字 0-9、字母 a-zA-Z、连字符-、其他都不允许。例
如, 不允许出现其他标点符号, 不允许空格, 不允许下划线, 不允许中文字符
# 2、主机名的开头和结尾字符不允许是连字符
# 3、主机名命名不允许出现 “数字-数字” 这种模式
#
# 主机被添加进集群并且部署服务后, 不允许修改主机名
#
work_node:
  hosts:
    10.0.0.21:
      NODE_NAME: usdp3-1
    10.0.0.22:
      NODE_NAME: usdp3-2
    10.0.0.23:
      NODE_NAME: usdp3-3
    10.0.0.25:
      NODE_NAME: usdp3-4
```

**主机名 (NODE\_NAME) -命名规则** (若手动已为各节点设置主机名, 此处可以不必再配置! ) :

1. 主机名只允许包含 **ascii** 字符里的数字 **0-9**、字母 **a-zA-Z**、连字符-、其他都不允许。例如, 不允许出现其他标点符号, 不允许空格, 不允许下划线, 不允许中文字符。
2. 主机名的开头和结尾字符不允许是连字符。
3. 主机名命名不允许出现 “数字-数字” 这种模式

**注意:** 主机被添加进集群并且部署服务后, 不允许修改主机名 (服务配置中的主机名无法同步修改)

## 执行修复和部署

在 **usdp-server** 主机上, 执行下方命令。

```
cd /data/usdp-srv/env-prepare
make prepare
```

待执行完成，待新增节点即已完成初始化及环境修复，接下来，便可通过 USDP Web 控制台进一步可视化的操作集群主机扩展任务。

### 7.2.4.2 主机扩展-向导

在 USDP 控制台中，将当前集群切换为待添加主机的集群。

在集群管理视图的顶部集群便捷“操作”菜单中，选择“添加主机”，进入添加服务向导。或在主机管理页面顶部点击“添加主机”按钮，便会进入扩展主机的向导。

**注意：**添加主机时，若待添加的主机是首次添加至当前 USDP 集群，则再初始化新主机的操作完成后，需要重新采集集群信息并更换申请新的 license。

#### Step1: 指定集群节点

手动输入待添加主机节点的完全限定域名 (hostname) 即可，若无节点的 hostname 修改，根据 hosts.yml 文件填写已设置的 hostname。

为了方便批量录入，USDP 支持用户以表达式的方式输入，规则如下：

- ◆ 可单行输入每一个节点的完全限定域名；
- ◆ 可通过“[]”辅助输入有数字规律的节点完全限定域名；例如 usdp3-[1-3] 表示包含“usdp3-1”、“usdp3-2”、“usdp3-3”共三个节点。

USDP 需要通过“root”用户以及“22”端口，管理所有主机服务器。暂不支持对端口号进行更改。

USDP 需要对指定添加的主机进行环境检测，例如节点间的网络互通性，待检查完成，将在页面中显示所有节点检测的状态。点击节点列表中“点击查看详情”栏的相应字样，即可查看该主机上的已检测以及正在检测的日志详情。



## Setp2: 检查节点环境

在该步骤中，USDP 需要检查对应节点是否已经初始化完成，比如：JDK、MySQL、时间服务器等是否已经正确安装及配置等。

待所有检查工作完成后，点击右下角的向导“下一步”按钮继续。

检查过程中，可点击列表“详情”栏的“检查中”等状态按钮查看对应节点的任务详情。

当检查过程中出现失败的情况，首先可尝试点击右侧的“全部重试”按钮，如果仍旧报错，此时可重点查看并参考失败详情及日志说明来定位失败具体原因，必要时登录对应主机进行修复后重试；

若碰到未经授权的报错提示时，可检查 USDP “授权管理”中的主机授权的授权数量是否充足，不足时，可扩展更新 License 后再次操作本向导流程。

## Setp3: 配置 UDH 源

USDP 将自动检测到当前环境已配置的 UDH 源，并为待添加的主机配置源信息。待页面显示“进度”为 100%时，即可进入下一步操作。

## Setp4: 集群主机扩展完成

完成上述步骤的执行后，即已成功为集群添加新主机。

一般情况下，新添加的主机节点，使用者会为其规划用作存储或计算服务的水平扩展，若需要稍后再操作，可点击“稍后安装”按钮，退出扩展主机向导。

若需要快速为新节点添加服务或服务实例，可点击“立即安装”按钮，即会进入服务添加向导。此处，可参考本文档《为集群扩展服务-向导》章节的内容继续操作，在此不做赘述。

## 7.3 服务管理

智能大数据平台 USDP 为集群及集群内各大数据服务提供了统一的管理方

式，譬如在集群管理视图页面，就能直观的查看到该集群中已部署使用的各大数据服务的基本健康状态、服务的启停状态，以及针对单个服务的便捷操作；若需要查看各个服务详细的信息，服务的监控信息，服务所需实例的分布情况，管理服务配置，服务原生 Web UI 访问等，辅助使用者更好的管理和使用大数据服务。

本章节，就以 HDFS 服务为例来说明 USDP 提供的服务管理能力，仅供使用者参考。其他的大数据服务管理方式与此类似。

### 7.3.1 服务控制及监控

在浏览器中打开 USDP Web 控制台并登陆，切换需要管理和操控的集群为当前集群。此时，在集群管理视图页面左侧，找到“HDFS”并点击该服务名称，即可跳转进入 HDFS 的详情页。

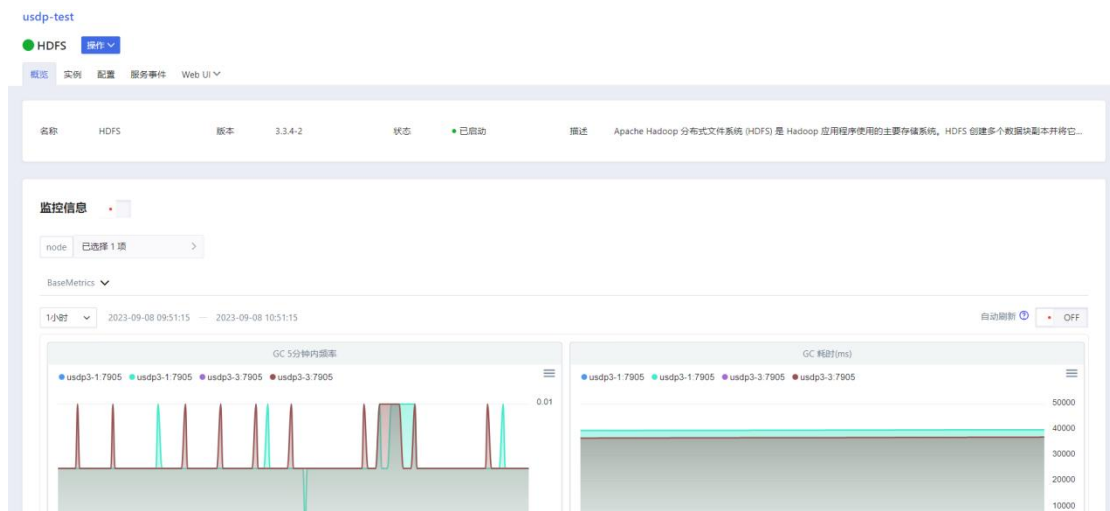


图 7-5 大数据服务详情页

#### 7.3.1.1 服务级控制

如上图所示的部分，会展示当前服务 HDFS 的状态（绿色圆点表示服务运行状态），服务名称，针对该服务的操作等。

关于服务的运行状态，主要有：

状态	说明
----	----

状态	说明
	已选择
	新增服务部署中
	已有服务增量部署实例中
	启动中
	重启中
	服务已部署，并且服务已启动
	服务已部署，但存在已选择的新增实例尚未执行 / 正在执行部署
	停止中
	实例不可用，已停止或者实例处于异常状态
	服务已部署，client 类实例
	移除中

表 7-1 USDP 大数据服务运行状态说明表

目前支持对服务进行“启动”、“停止”、“重启”、“滚动启动”、“滚动重启”、删除等操作。其中，“滚动启动”即为串行启动多个实例；“滚动重启”即为串行重启多个实例。

### 7.3.1.2 状态详情

HDFS 服务详情状态页的顶部位置，展示了 HDFS 服务的相关详细信息，如“服务名称”、“版本号”、服务的运行“状态”、服务“描述”等信息。

### 7.3.1.3 监控详情

HDFS 服务详情状态页面的下方“监控数据”模块中，为使用者展示了关于 HDFS 的部分核心监控指标信息，如与 GC、堆内存、NameNode、DateNode、JournlNode、ZKFC 等等相关的监控图表信息。

监控图表支持手动设置时间粒度并画图，包括最近的 1 小时、3 小时、6 小

时、12 小时、1 天、3 天、7 天的；支持自定义时间粒度；支持图表自动刷新展示。

### 7.3.2 服务实例管理

在服务详情-“实例”页面中，可以看到该服务相关的所有实例的分布情况，对于各个实例，关键信息展示有实例的“监控状态（图标）”、实例本身的“名称”、实例的“实例状态”、实例所在主机的“主机名称”、实例所属的“服务名称”、实例所属的“角色配置组”。

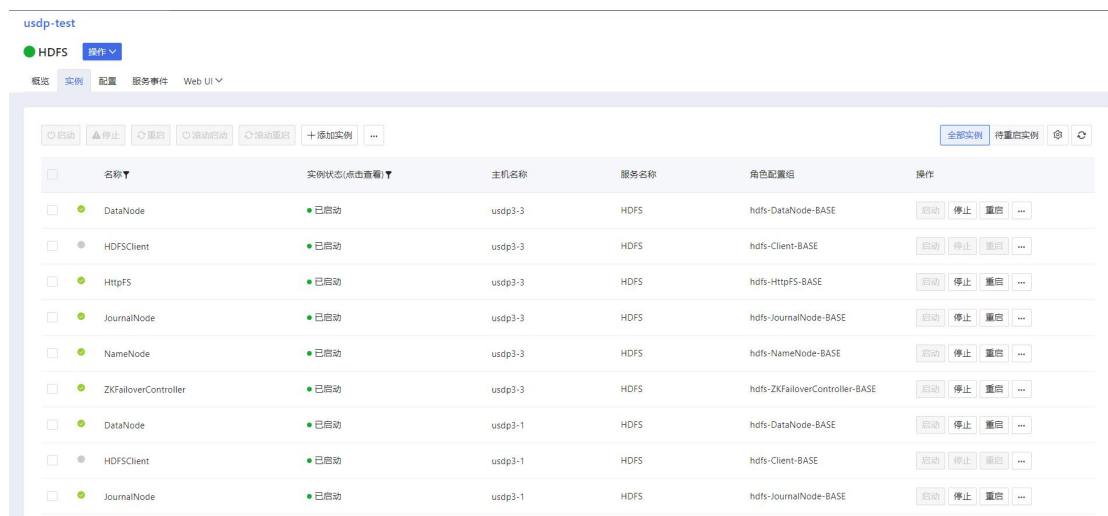


图 7-6 大数据服务实例详情页面

实例状态:

状态	说明
	可用状态
	不可用状态
	未知状态或 client 类实例

表 7-2 USDP 大数据服务实例运行状态说明表

服务“实例”，支持对该服务中已运行的实例进行逐一/批量管理操作，如上图所示，可逐一/批量的进行“启动”、“停止”、“重启”、“滚动启动”、“滚动重启”、“删除”、“添加实例”等操作。

需要注意的是，在对某个实例进行“删除”操作时，为避免误操作，需使该实例处于“停止”状态时，才可执行“删除”操作。

### 7.3.2.1 实例添加-向导

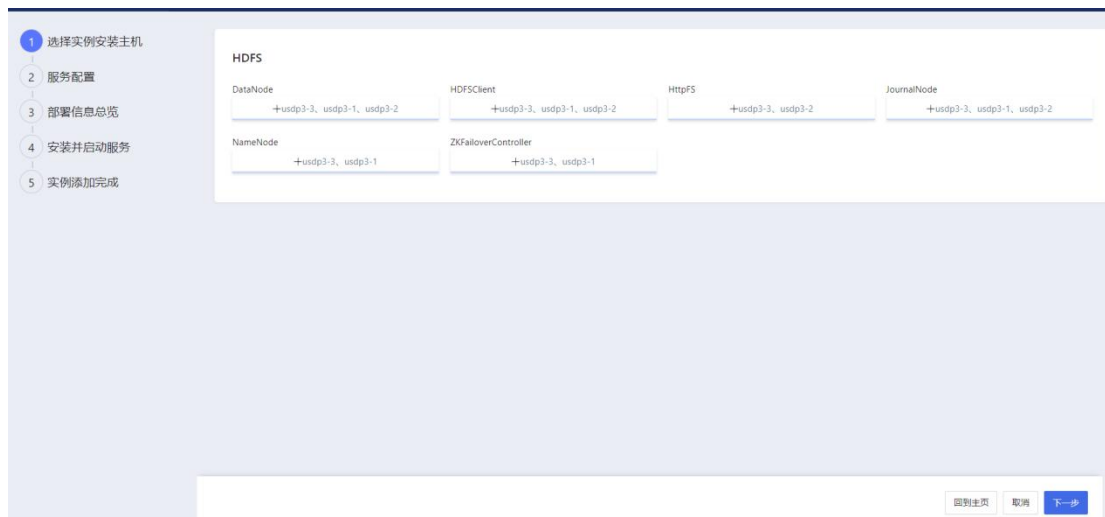
在服务详情-“实例”页面中顶部的实例操作区域，找到“添加实例”按钮，即可进入添加向导。

本章节将以 HDFS 为示例，介绍如何为当前集群的 HDFS 服务扩展 DataNode 实例，其他服务实例的添加操作方式同理。

#### Step1: 选择实例安装节点

UDH 提供的大数据生态服务均由不同的实例构成，安装集群时，使用者可根据业务需求做好服务在集群主机中的分布规划，将不同的服务及实例分别安装到分布式集群的哪些主机上运行。

USDP 会以实例组的方式展示当前服务的各个实例在分布式集群中的分布情况并呈现给使用者。此时，使用者可在点击 DataNode 实例组按钮，为 DataNode 实例组扩展新实例节点。



在选择主机的页面中，会充分展示当前集群中各节点上运行的各服务实例分布情况，需要在集群中寻找并选中符合规划预期，且未运行待添加的实例的主机

节点。

对于 DataNode 实例而言，扩展只能扩展还未部署 DataNode 实例的主机上。此时，选中可部署 DataNode 实例的主机，点击“确定”按钮，并执行向导“下一步”。

#### 提示：

◆ 在为待添加的服务实例选择安装节点时，因在其使用过程中需依赖安装节点的计算、存储和网络资源，因此，建议使用者提前考虑该服务实例的业务场景和可能的资源消耗，

**注意：**请根据当前集群所管理的各个主机硬件配置情况、该主机上已部署服务的使用情况和性能情况，合理规划，以免该服务的运行干扰到当前节点本身提供的服务性能。

## Setp2: 服务配置

进入到服务配置页面时，需使用者自行对所选的实例进行部分配置项设置，如数据存储路径、需要的数据库信息填写等信息；部分实例不涉及此步配置。

如对 DataNode 数据存储目录的设置中，若所选择的主机存在多块数据盘，可根据实际规划及需求，手动“添加”HDFS 数据存储至主机的多块数据盘的文件路径。

## Setp3: 部署信息总览

完成前几个向导步骤中的信息配置后，USDP 会将当前“集群基本信息”、以及“服务基本信息”概览呈现给使用者进行校验。

使用者浏览检查无误后，点击向导“开始部署”按钮继续执行。

## Setp4: 安装并启动服务

在此步骤中，USDP 将根据前几步的配置信息生成部署任务计划，并开始逐项执行服务安装工作，并展示任务执行的进展情况。若需了解执行进展细节。

可在“点击查看详情” 栏中的各个链接中，查看当前执行步骤的详细情况。

当服务安装发生错误时，使用者可以点击查看报错节点“失败”详情，参考详情提示信息，进行手动修复操作，或点击“全部重试”按钮，重新进行服务组件部署工作。

待所有任务执行完成，主机进度条显示 100% 时，即表示已成功为当前服务完成实例添加。

### Step5: 实例添加完成

退出实例添加向导后，即可在服务实例管理页面中查看新实例的分布和运行情况了，如需刷新配置或重启实例，按照提示进行操作。

### 7.3.3 服务配置管理

在智能大数据平台 USDP v3.x 中，USDP 提供所有大数据生态技术服务，使用者均需通过 USDP Web 控制台进行服务配置管理。

**注意：**在 USDP v3.x 中，我们**不希望使用者自行登录集群服务器手动修改 USDP 提供的大数据生态服务组件的配置文件！**如若被自行更改，这将导致手动更改的内容不会被记录至该服务的“历史记录”中，更无法回滚生效；同时，手动操作配置内容可能会被管理服务覆盖，导致丢失的情况。**建议使用者充分使用 USDP Web 控制台中提供的服务配置管理功能**，使用和操作方式可参考本节内容。

以 HDFS 的配置管理为例进行说明，进入服务详情-“配置”页面中，可以为当前大数据服务进行配置的管理操作。



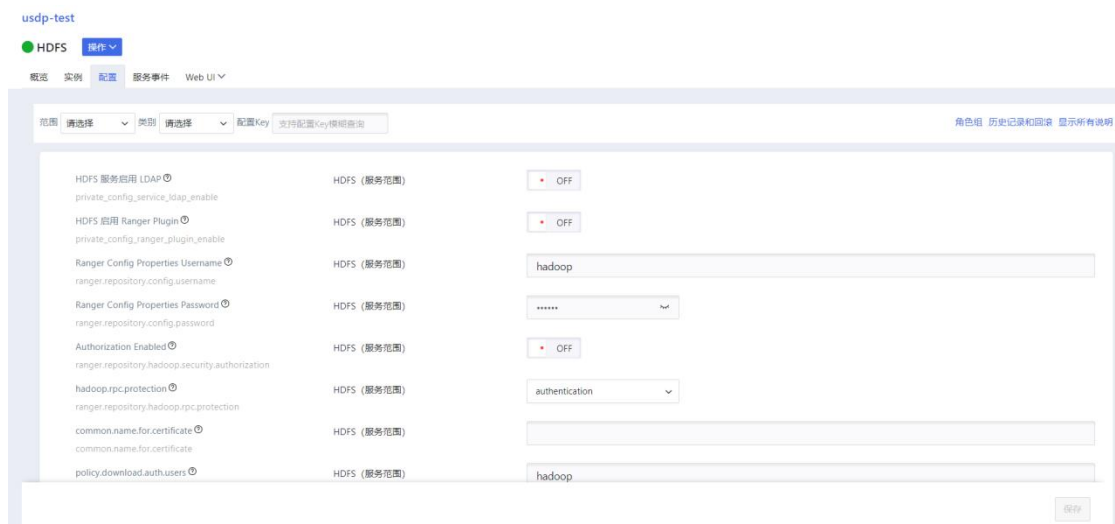


图 7-7 服务配置管理页面

进入 HDFS 的配置页后，即可查看所有与当前服务 HDFS 相关的配置详情，使用者可翻页查找需要调整的配置项进行修改。

因部分大数据服务配置复杂，且配置项较多，USDP 控制台支持实例和配置类别过滤的策略，辅助使用者快速查找到需要调整的配置项，配置可以按照“范围”、“类别”两个维度进行过滤筛选配置项，同时也可通过 Key 进行查询。

### 7.3.3.1 高级配置

当使用者需要调整的配置项并未在配置页面中找到，此时可通过类别筛选“扩展”来查找对应配置项，以源文件的格式输入配置项值“key-value”。如：追加 HDFS 配置“dfs.namenode.http-bind-host”到 hdfs-site.xml 配置文件中，编辑配置项“hdfs\_hdfs\_site\_xml\_extend”，点击添加按钮，输入相应 key-value，或者以 XML 格式输入：

```
<property><name>dfs.namenode.http-bind-host</name><value>0.0.0.0</value></property>
```

使用高级配置代码段进行配置时，支持“以 XML 格式查看”或“视图编辑器”两种方式设置配置信息；

配置修改保存成功后，USDP 将自动检测该更改是否需要当前服务 / 当

前实例，以及依赖关联的 服务 / 实例 进行重启，若需要，将会在各服务实例页面进行显示提醒。如在“实例状态”列中，追加显示“过期配置：需要重启”字样。重启服务实例使其配置生效，则改动的配置项将运用至此服务中所有同角色组中实例配置（客户端配置除外）中。

### 7.3.3.2 配置历史和回滚

USDP 支持服务级别的配置“历史记录和回滚”操作，每一次更改某服务的配置，则会被记录至该服务的“配置”管理页面“历史记录和回滚”中。

点击历史配置列表中某条记录，可在弹出的“版本详细信息”对话框中，查看该记录版本的更改明细（含更改前和更改后的配置项的值变化）；此时，使用者可根据需求判断，是否将该历史配置“回滚”至该版本更改前的历史版本状态；确需要回滚操作，请点击对话框中“还原配置更改”按钮即可；

此时，历史配置列表中产生了一条“已还原”的配置更改历史记录；

### 7.3.3.3 重启生效

配置历史回滚后，USDP 将自动检测该更改是否需要当前服务 / 当前实例，以及依赖关联的 服务 / 实例 进行重启，若需要，将会在各服务实例页面进行显示提醒。如在“实例状态”列中，追加显示“过期配置：需要重启”字样。重启服务实例使其配置生效，则改动的配置项将运用至此服务中所有同角色组中实例配置（客户端配置除外）中。

## 7.3.4 服务 WEB UI 访问

为便于使用者对集群提供的各个大数据服务的高效便捷使用，USDP 已主动集成并支持了各个服务默认自带的所有 Web UI 的快捷打开方式，浏览器所在客户端需提前做好主机名解析，也可手动替换为节点 IP 地址。

更多服务 Web UI 账号，可在[《各大数据服务组件 WEB UI 登录账号》](#)章节中查得。

以 HDFS 的 Web UI 为例，点击 HDFS 服务详情页面中“Web UI”选项卡时，会在弹出的下拉菜单中展开 HDFS 相关的原生 Web 页面链接。点击“NameNode Web UI(usdp3-3)”，会自动在浏览器中打开新的标签页，并显示 usdp3-3 节点上的 HDFS 原生的 dfshealth 管理页面。

以 Ranger 的 Web UI 为例，点击 Ranger 服务详情页面中“Web UI”选项卡时，会在弹出的下拉菜单中展开 Ranger 相关的原生 Web 页面链接，点击“RangerAdmin Web UI (usdp3-2)”进入 Ranger 管理页面，输入登录口令，即可管理 Ranger 服务。

注意：若集群使用内网 IP 进行搭建，部分服务 Web UIs 即会只监听内网请求；还有一部分服务是监听所有 IP 的请求，（不区分内外网络）。以下服务的部分 Web UI 默认监听具体 IP，需要用相应 IP 进行访问：

服务名称	默认密码
HDFS	50070
YARN	8088、8188、19888

表 7-3 USDP 服务端口监听访问来源限制的服务端口列表

更多服务端口，请查看[《各服务常用端口》](#)章节的说明。

### 7.3.5 为集群扩展服务-向导

在使用 USDP 管理大数据业务集群时，使用者可根据需求灵活管理集群中需要的各类大数据服务及相关组件。本节将介绍如何通过 USDP 来为集群添加部署大数据服务，并管理该服务对应的组件。

开始之前，使用者需要先在 USDP 控制台中，切换当前集群为待添加服务或组件的集群。

当前集群管理视图的操作菜单中，选择“添加服务”，进入添加服务向导。

本节将以 Sqoop (一个轻量级的数据采集与转储服务, 主要以 Client 的方式运行于集群的某些主机节点上) 为示例, 介绍如何为当前集群添加 Sqoop 服务, 其他服务添加的操作方式同理。

### Setp1: 选择服务

在本步骤中, 需要选择当前集群中未部署的大数据生态服务。浏览 USDP 支持的服务列表, 找到未部署使用的 Sqoop 服务。选中 Sqoop 左侧的单选框, 点击右下角的“下一步”按钮。

### Setp2: 选择组件安装主机

在“选择组件安装主机”的向导页中, 为 SqoopClient 选择计划安装的节点。点击带“+”的按钮后, USDP 会以对话框的形式, 展示当前集群所管理的主机, 查找到计划部署新增加服务的主机后, 选中主机名左侧的复选框, 选择好后点击向导“下一步”按钮。

在为待添加的服务选择安装主机时, 因在其使用过程中需依赖待安装主机的计算、存储和网络资源, 因此, 建议使用者提前考虑该服务的业务场景和可能的资源消耗, 请根据当前集群所管理的各个主机硬件配置情况、该主机上已部署服务的使用情况和性能情况, 合理规划, 以免该服务的运行干扰到当前主机本身提供的服务性能。

### Setp3: 服务配置

在该步骤中, 需要对服务做一些必要的初始配置。待配置信息设置完成, 点击向导“下一步”按钮。

本案例中所添加的 Sqoop 服务, 在该步骤中无需设置配置信息, 因此可以直接进行下一步操作。

### Setp4: 部署信息总览

在本步骤中 USDP 管理服务将检测出, 已被当前集群管理的主机, 并展示

当前添加的服务已选择部署的主机信息，经浏览确认无误后，点击“开始部署”按钮。

## Step5: 安装并启动服务

进入本步骤是，USDP 将生成部署计划，并开始已在选择的主机上执行部署新增服务的进度。待安装进度完成后，点击“完成”按钮，退出服务或组件安装向导。此时，使用者可以开始使用 Sqoop 服务。

## 7.4 告警管理

在使用 USDP 构建并管理大数据平台过程中，当发生或者可能将要发生某种异常状况时，使用者更希望能非常及时的、随时随地的，全面的感知到平台以及集群的运行情况，这对于维护并提供一套稳定的大数据服务至关重要。

USDP 提供了非常丰富的告警规则和指标，使用者可直接导入使用，或自行添加告警规则，来完成对已使用大数据服务状态的监测。USDP 支持用户自定义告警规则，可以灵活的根据需求编写 PromQL。

本章节将介绍如何使用 USDP，为使用者维护的大数据集群进行告警的管理工作。

### 7.4.1 告警规则



图 7-8 集群告警规则管理页面

USDP Web 控制台集群管理视图顶部的菜单栏点击“告警管理”，在下拉菜单中选择“告警规则”，即可进入上图所示的告警规则管理页面。

使用者可通过该页面中查看到所有已对当前集群配置的告警规则；支持翻页、按规则名称检索等辅助浏览的功能；支持按告警的级别，进行分类查找。

### 7.4.1.1 导入告警规则

点击告警规则页面顶部的“导入”按钮，可以打开“导入规则”对话框。

通过该功能，可将 USDP 提供的诸多告警规则复制到当前集群。在导入时，使用者需选择所需要的告警规则进行导入，勾选规则名称左侧的复选框，点击对话框右下角的“确认”按钮，即可完成导入动作。

导入完成后，使用者可根据实际告警需求进行适当调整，如告警频率、告警生效周期、通知媒介等，若无需修改，可直接对导入的告警规则开启“启用”生效。

当点击告警规则列表中的某条规则名称时，可以查看到该规则的配置详情。对告警规则而言，均包括“基本配置”、“生效配置”、“通知配置”三个配置部分。

“基本配置”模块，主要配置了告警规则的触发条件。其中核心配置项为“PromQL”、“执行频率”、“持续时长”三项。允许用户自定义 PromQL，需依赖的指标项，可查找相关 PromQL 资料进行配置。

### 7.4.1.2 添加告警规则

在当前集群的“告警规则”管理页面中，点击左上角的“添加规则”按钮，即可进入告警规则添加页面。

添加告警规则页面由三部组成：

- ◆ 基本配置。在该模块中，需要为待添加的告警规则填写“规则标题”、“告警级别”、“PromQL”、“执行频率”、“持续时长”等关键信息；
- ◆ 生效配置。在该模块中，需要为待添加的告警规则填写是否“立即启用”、“生效时间”周期。

- ◆ 通知配置。在该模块中，需要为待添加的告警规则填写“通知媒介”、“告警接收组”、“启用恢复通知”、“留观时长”、“重复发送频率”、“最大发送次数”等关键信息。

#### 告警规则项说明：

规则项	说明
规则标题	告警规则在规则管理页面显示的名称
规则备注	可为该告警添加自定义的使用用途等说明
告警级别	默认有 3 个级别，可使用该级别为告警规则进行归类，便于查找及管理
PromQL	可自定义对监控指标进行触发条件的设置，遵循 PromQL 语法规则即可，所需要的监控指标，可在 Prometheus-Graph-Metrics Explorer 中查找
执行频率	默认每隔 15 秒，按 PromQL 执行告警查询，查得数据即表示当次有监控数据触发了规则
持续时长	通常持续时长大于等于执行频率，在持续时长内按照执行频率多次执行 PromQL 查询，每次都触发才生成告警；如果持续时长置为 0，表示只要有一次 PromQL 查询触发阈值，就生成告警
立即启用	默认“启用”，即确认添加本规则后，同时将该规则在当前集群中生效
生效时间	默认按 7*24 小时不间断对本规则进行长期生效
通知媒介	可配置按“邮件”、“钉钉”、“企业微信”、“飞书”、“mm bot”等方式进行告警信息接收
告警接收组	配置并按“平台用户组”进行告警信息推送
启用恢复通知	默认“启用”，即该告警状态已恢复时也会发送恢复通知
留观时长	告警恢复的观察时长，在未达到最大时长时间内若无新告警产生，则发送告警恢复通知。该时长应大于“执行频率”；
重复发送频率	默认间隔 60 分钟；如果告警持续未恢复，再间隔时间之后重复提醒告警接收组的成员
最大发送次数	如果值为 0，则不做最大发送次数的限制

### 7.4.1.3 克隆告警规则

当前集群若已添加过告警规则，为了更方便快捷的添加新的告警规则，使用者可选择从已有的告警规则基础上进行克隆，在克隆出的告警规则中根据需求做



出配置调整后，点击确认即可完成克隆。

克隆告警规则时，告警规则编辑页面的操作，可参考本手册《[添加告警规则](#)》章节的说明进行操作。

#### 7.4.1.4 告警规则管理

随着集群在业务中的使用，渐渐地，大数据服务和集群主机规模也会随之增加，此时，需要进行监听的服务状态也会越来越多。在 USDP 控制台中，提供了诸多便于使用者管理告警规则的一些功能。

告警规则管理页面中，支持逐条/批量的操作控制已添加的告警规则，如是否“启用”、“禁用”规则。

告警规则管理页面左上角处，支持按照告警规则的级别（3 个级别）进行过滤查找规则；或在右上角的检索框中按告警规则标题进行模糊查找。

若某项/某些告警规则已不再使用，可在控制台中逐条/批量的对不再使用的告警规则进行“删除”操作。

### 7.4.2 活跃告警



图 7-9 集群活跃告警管理页面

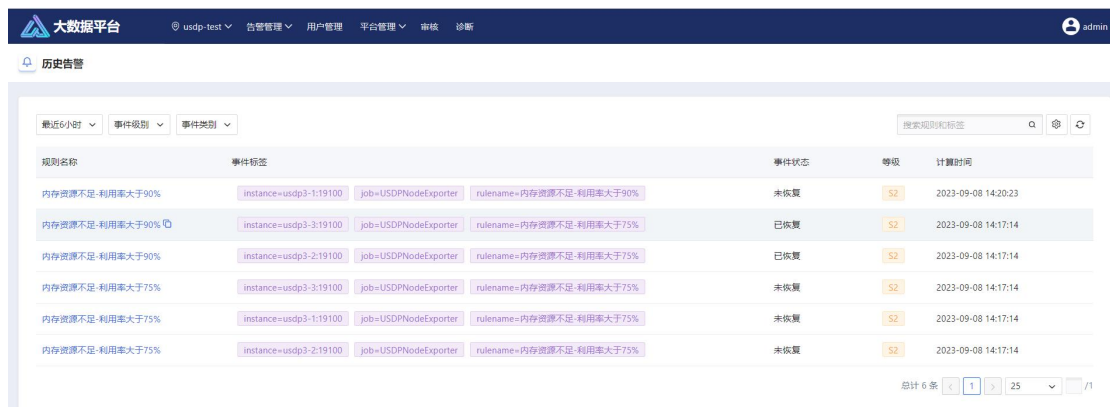
对于集群已配置并已“启用”的告警规则，会根据规则中的 PromQL 条件按执行频率执行监听，当被监听的服务产生了某种异常情况，并且该异常有告警规则可以匹配到，即会触发告警规则，产生一条告警信息。所有未处理的告警，即会进入到活跃告警管理页面中来。

### 7.4.2.1 活跃告警查找

活跃告警管理页面中，支持按照时间段、事件级别辅助使用者过滤查找，还可按照规则名称和标签进行模糊查询。

时间段支持“最近 6 小时”、“最近 12 小时”、“最近 1 天”、“最近 2 天”、“最近 3 天”、“最近 7 天”、“最近 14 天”、“最近 30 天”、“最近 60 天”、“最近 90 天”供使用者选择。

### 7.4.3 历史告警



若使用者对当前集群所有已产生的告警信息进行处理，解除告警状态了，或系统自动恢复正常状态后，相关的告警信息将从“活跃告警”列表中转存至“历史告警”管理页面中，便于使用者对集群及服务的历史状况进行查阅。

#### 7.4.3.1 历史告警查找

历史告警管理页面中，支持按照时间段、事件等级、事件类别几个维度辅助使用者过滤查找，还可按照规则名称和标签进行模糊查询。

时间段支持“最近 6 小时”、“最近 12 小时”、“最近 1 天”、“最近 2 天”、“最近 3 天”、“最近 7 天”、“最近 14 天”、“最近 30 天”、“最近 60 天”、“最近 90 天”供使用者选择。

事件类别包括“Triggered”（已触发的告警）、“Recovered”（已恢复的告警）两种状态辅助过滤查找。

## 8 用户管理



图 8-1 USDP 智能大数据平台用户管理

用户管理页面中，将对 USDP 平台所有的使用用户进行统一管理，当前版本中暂时仅支持管理员用户访问和控制平台。

## 9 平台管理

### 9.1 UDH 源



USDP 使用 http 协议源服务来管理系统需要的安装包。通过该方式，实现对集群层面，以及整个 USDP 平台的版本进行控制。

当 USDP 发布新的 UDH 版本时，使用者便可向 USDP 官方获取新版本的源地址或离线源包，并配置至仓库地址栏，离线源包需解压至部署节点对应目录，之后便可自主选择对某一集群或整个 USDP 平台，进行版本升级。

### 9.2 安全管理

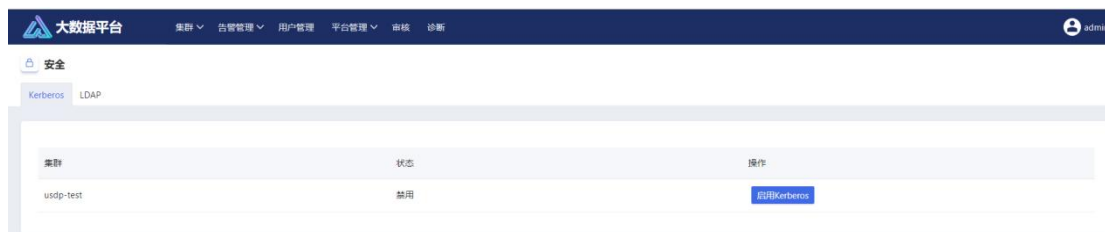


图 9-1 USDP 集群安全管理

USDP 支持通过控制台快速为单个集群进行安全状态管理，如“启用 Kerberos”、“禁用 Kerberos”。

使用 USDP 为集群启用并管理 Kerberos 安全认证前，需要使用者自行部署搭建 KDC 服务并确保其正常运行中，KDC 能正常允许更新票证。启用 kerberos 安全认证时，需要额外搭建 kerberos KDC 服务，并在客户端安装 Kerberos 相关的依赖包。

目前 Kerberos 支持的组件有：Zookeeper、HDFS、YARN、Hive、HBase、Spark、Impala、Hue、Kafka、Ranger。

## 9.2.1 名词解释

- ◆ Kerberos: 是一种网络认证协议，其设计目标是通过密钥系统为客户机 / 服务器应用程序提供强大的认证服务。
- ◆ KDC (Key Distribution Center) : 是一个网络服务，提供 ticket 和临时会话密钥
- ◆ Principal: 认证的主体 例如: `hadoop/kunlun01@UCLLOUD.CN`
- ◆ realm name: 包含 KDC 和许多客户端的 Kerberos 网络, 类似于域, 俗称为领域。例如: UCLLOUD.CN
- ◆ keytab: keytab 是包含 principals 和加密 principal key 的文件

## 9.2.2 启用 Kerberos 向导

当要为某一集群启用 Kerberos 时，在安全管理页面的集群列表中，找到待启用 Kerberos 的集群右侧，点击“启用 Kerberos”按钮，进入启用向导。

在 Kerberos 开启配置向导执行中，该集群中的所有服务将作为向导的一部分重启，以便 Kerberos 能正常开启成功。再此之前，需要使用者确保待开启 Kerberos 的集群中的服务是处于可重启的状态，以免重启服务对业务带来影响。

此向导将指导使用者完成启用 Kerberos 的配置。集群中的所有服务将作为向导的一部分重启。

### Step1: 启用前提信息确认

在该步骤中，需要使用者在开启 Kerberos 前，明确其列出的待确认项目，使用者确认无误后，勾选各项左侧的复选框，再进入向导下一步。

USDP 不提供创建新 KDC，因此，在使用该向导为集群配置并开启 Kerberos 时，需要使用者自行提供和维护一套可用的 KDC；确保该 KDC 允许

可更新的票证；并创建适当的账户。

## Setp2: KDC 信息

配置 KDC 信息，请正确填写相关信息，其中支持同时创建多个 Domain Name。

配置项	配置值	操作
Kerberos 安全领域 (default_realm)	UCLLOUD.CN	
KDC 服务器地址 (kdc_host)	usdp01	
KDC Admin Server Host (admin_server)	usdp01	
Domain Name(s) (domain_names)	UCLLOUD.CN	删除
	HADOOP.COM	删除

图 9-2 启用 Kerberos 向导流程步骤 2

- ◆ Kerberos 安全领域: Kerberos 安全所用领域。
- ◆ KDC 服务器地址: KDC 服务器地址
- ◆ KDC Admin Server Host : KDC 服务器地址,可包含端口号
- ◆ Domain Name(s) : Kerberos 领域列表, 第一个为默认安全领域

## Setp3: KRB5 配置





## Setp6: 集群启用

执行至此步骤，USDP 将生成配置和安装任务进程，开始同步 `keb5.conf` 文件, 并未支持 Kerberos 的发数据服务进行配置启用 `kerberos` 及相关的操作, 之后重启服务使其配置生效。

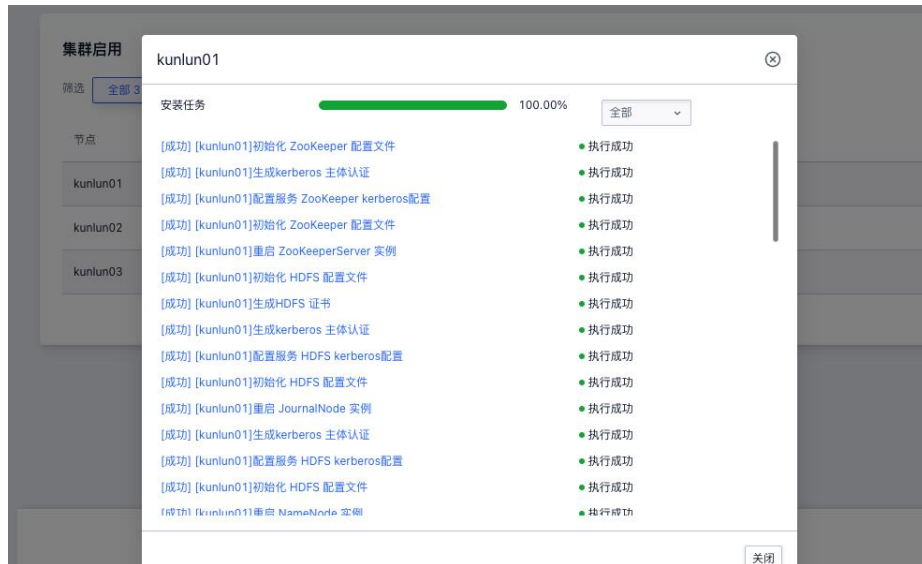


图 9-5 启用 Kerberos 向导流程步骤 6 – 执行详情查看

如上图所示，可以点击查看每个集群主机上的执行详情。

## Setp7: Kerberos 开启完成

至此，则表示已为所选集群成功开启 `kerberos` 认证。

## Setp8 (可选) : 验证 Kerberos

### ◆ 确认 Kerberos 配置信息

```
[root@kunlun01 tmp] ll /etc/krb5.conf
-rwxr-xr-x 1 root root 1031 12月 7 11:09 /etc/krb5.conf
[root@kunlun01 tmp] ll /data/usdp/keytab/hadoop.keytab
-rw-r--r-- 1 root root 1658 12月 2 11:58 /data/usdp/keytab/hadoop.keytab
```

### ◆ 节点机器上初始化 Kerberos 认证

```
[root@kunlun01 usdp] kinit -kt /data/usdp/keytab/hadoop.keytab
hadoop/kunlun01@UCLLOUD.CN
```

```
[root@kunlun01 tmp] ll /tmp/krb5cc_0
-rw----- 1 root root 750 12月 7 11:41 /tmp/krb5cc_0
```

### ◆ 组件 kerberos 确认

#### 未获得 Kerberos 认证时操作 HDFS

```
[root@kunlun02 ~] hadoop fs -ls /
2022-12-07 13:36:04,615 WARN org.apache.hadoop.ipc.Client: Exception encountered while
connecting to the server
org.apache.hadoop.security.AccessControlException: Client cannot authenticate via:[TOKEN,
KERBEROS]
    at org.apache.hadoop.security.SaslRpcClient.selectSaslClient(SaslRpcClient.java:179)
    at org.apache.hadoop.security.SaslRpcClient.saslConnect(SaslRpcClient.java:392)
    at org.apache.hadoop.ipc.Client$Connection.setupSaslConnection(Client.java:623)
    at org.apache.hadoop.ipc.Client$Connection.access$2300(Client.java:414)
    at org.apache.hadoop.ipc.Client$Connection$2.run(Client.java:843)
    at org.apache.hadoop.ipc.Client$Connection$2.run(Client.java:839)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
                                                                    at
org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1878)
    at org.apache.hadoop.ipc.Client$Connection.setupIOStreams(Client.java:839)
    at org.apache.hadoop.ipc.Client$Connection.access$3800(Client.java:414)
    at org.apache.hadoop.ipc.Client.getConnection(Client.java:1677)
    at org.apache.hadoop.ipc.Client.call(Client.java:1502)
    at org.apache.hadoop.ipc.Client.call(Client.java:1455)
                                                                    at
org.apache.hadoop.ipc.ProtobufRpcEngine2$Invoker.invoke(ProtobufRpcEngine2.java:242)
                                                                    at
org.apache.hadoop.ipc.ProtobufRpcEngine2$Invoker.invoke(ProtobufRpcEngine2.java:129)
    at com.sun.proxy.$Proxy9.getFileInfo(Unknown Source)
                                                                    at
org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.getFileInfo(Client
NamenodeProtocolTranslatorPB.java:965)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
                                                                    at
sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
                                                                    at
org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.jav
a:422)
                                                                    at
org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeMethod(RetryInvocationHandl
```

```

er.java:165)
                                                    at
org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invoke(RetryInvocationHandler.java:
157)
                                                    at
org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeOnce(RetryInvocationHandler.
java:95)
                                                    at
org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:359)
  at com.sun.proxy.$Proxy10.getFileInfo(Unknown Source)
  at org.apache.hadoop.hdfs.DFSCClient.getFileInfo(DFSCClient.java:1739)
                                                    at
org.apache.hadoop.hdfs.DistributedFileSystem$29.doCall(DistributedFileSystem.java:1753)
                                                    at
org.apache.hadoop.hdfs.DistributedFileSystem$29.doCall(DistributedFileSystem.java:1750)
  at org.apache.hadoop.fs.FileSystemLinkResolver.resolve(FileSystemLinkResolver.java:81)
                                                    at
org.apache.hadoop.hdfs.DistributedFileSystem.getFileStatus(DistributedFileSystem.java:176
5)
  at org.apache.hadoop.fs.Globber.getFileStatus(Globber.java:115)
  at org.apache.hadoop.fs.Globber.doGlob(Globber.java:362)
  at org.apache.hadoop.fs.Globber.glob(Globber.java:202)
  at org.apache.hadoop.fs.FileSystem.globStatus(FileSystem.java:2124)
  at org.apache.hadoop.fs.shell.PathData.expandAsGlob(PathData.java:343)
  at org.apache.hadoop.fs.shell.Command.expandArgument(Command.java:252)
  at org.apache.hadoop.fs.shell.Command.expandArguments(Command.java:235)
  at org.apache.hadoop.fs.shell.FsCommand.processRawArguments(FsCommand.java:105)
  at org.apache.hadoop.fs.shell.Command.run(Command.java:179)
  at org.apache.hadoop.fs.FsShell.run(FsShell.java:327)
  at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:81)
  at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:95)
  at org.apache.hadoop.fs.FsShell.main(FsShell.java:390)
ls: DestHost:destPort kunlun01:8020 , LocalHost:localPort kunlun02/10.76.196.113:0. Failed
on local exception: java.io.IOException: org.apache.hadoop.security.AccessControlException:
Client cannot authenticate via:[TOKEN, KERBEROS]

```

### 已取得认证时操作 HDFS

```

[root@kunlun02 ~] kinit -kt /data/usdp/keytab/hadoop.keytab hadoop/kunlun02@UCLLOUD.CN
[root@kunlun02 ~] hadoop fs -ls /
Found 4 items
drwxr-xr-x  - root  hadoop      0 2022-12-07 10:54 /kunlun123
-rw-r--r--  3 hdfs  hadoop  69406415 2022-12-07 10:54 /tez.tar.gz
drwxrwx---  - mapred hadoop      0 2022-12-07 10:55 /tmp
drwxr-xr-x  - hive  hadoop      0 2022-12-07 11:44 /user

```

```
[root@kunlun02 ~]
```

### 未获得认证时操作 Hive

```
[root@kunlun01 usdp] hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in
[jar:file:/usr/lib/hbase/lib/client-facing-thirdparty/slf4j-reload4j-1.7.33.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in
[jar:file:/usr/lib/tez/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in
[jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in
[jar:file:/usr/lib/hadoop/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Reload4jLoggerFactory]
Hive Session ID = 7a098aa6-7916-4265-a214-1a9e7898da67

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: true
Exception in thread "main" java.lang.RuntimeException: java.io.IOException: DestHost:destPort kunlun01:8020 , LocalHost:localPort kunlun01/10.76.196.112:0. Failed on local exception: java.io.IOException: org.apache.hadoop.security.AccessControlException: Client cannot authenticate via:[TOKEN, KERBEROS]
    at org.apache.hadoop.hive.ql.session.SessionState.start(SessionState.java:651)
    at org.apache.hadoop.hive.ql.session.SessionState.beginStart(SessionState.java:591)
    at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:747)
    at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:683)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
Caused by: java.io.IOException: DestHost:destPort kunlun01:8020 , LocalHost:localPort kunlun01/10.76.196.112:0. Failed on local exception: java.io.IOException: org.apache.hadoop.security.AccessControlException: Client cannot authenticate via:[TOKEN, KERBEROS]
```

### 已取得认证时操作 Hive

```
[root@kunlun01 tmp] hive
SLF4J: Class path contains multiple SLF4J bindings.
```

```

SLF4J: Found binding in
[jar:file:/usr/lib/hbase/lib/client-facing-thirdparty/slf4j-reload4j-1.7.33.jar!/org/slf4j/impl/StaticLo
ggerBinder.class]
SLF4J: Found binding in
[jar:file:/usr/lib/tez/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in
[jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in
[jar:file:/usr/lib/hadoop/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Reload4jLoggerFactory]
Hive Session ID = 022765f7-d0cf-47ea-9588-2ccb47e74337

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async:
true
Hive Session ID = 05bc29a6-37fa-43dc-8a1c-0bc4d61f12b1
hive> show databases;
OK
default
test
Time taken: 0.876 seconds, Fetched: 2 row(s)
hive> use test;
OK
Time taken: 0.093 seconds
hive> show tables;
OK
employee
Time taken: 0.092 seconds, Fetched: 1 row(s)

```

## 9.3 授权管理

智能大数据平台 USDP，目前已支持完善的“授权许可”相关管理功能，授权许可分为“基础许可”、“拓展许可”、“服务许可”三类。本章节将介绍使用者在正常使用过程中有关“授权许可”的相关问题和处理方法。

### 9.3.1 名词解释

- ◆ 基础许可：平台基础许可，授权后可使用平台基础功能
- ◆ 拓展许可：平台拓展许可，授权后可使用平台增值功能
- ◆ 服务许可：平台服务许可，为购买产品的客户打造的维护保障服务解

决方案

### 9.3.2 上传或更换授权

当未上传基础授权证书或基础授权证书过期时，使用者登录 USDP 控制台时，系统将检查授权是否有效，若无效，则跳转至“授权管理”页面，提示使用者上传新的授权许可证书。成功上传新的许可证书后，方可正常使用和操作 USDP Web 控制台。

在授权管理页面中，点击“信息采集”按钮，此时将下载一份文件，请联系客户经理并提供该文件来获取授权证书。

授权证书获取后，请点击“授权管理”页面的“上传授权”按钮进行上传许可证书文件并完成导入，USDP 系统验证证书文件有效后，使用者可正常进行控制台的其他操作。

导入基础许可证书时的状态，显示已授权字样。此时点击右侧的“查看授权细节”按钮，可查看到该授权的详细信息。

## 9.4 审核

### 9.4.1 操作日志

记录所有用户对当前 USDP 平台进行的控制机操作及变更行为，支持通过“集群 ID”、“API 名称”、“API 接口”、“操作类型”、“请求结果”、“操作者”、“源 IP 地址”进行筛选，默认展示 2 小时内记录，可自由选择“1 小时”、“2 小时”、“6 小时”、“12 小时”、“1 天”、“7 天”、“30 天”、“自定义”，对于筛选结果可使用“下载 CSV”功能导出。

## 9.5 诊断

### 9.5.1 事件

事件类型包括：运维审计类、服务健康检查类

平台定时对服务及实例的健康状态进行检查,对于异常服务发生及恢复时间进行记录;监听平台控制操作过程中存在的一些异常事件,如重启服务及实例失败的情况进行时间告警;

支持通过(事件、实例、服务、集群、主机)ID、事件内容、严重程度、事件类型等进行筛选,也可根据时间段进行筛选,结果支持导出 CSV 格式。

事件处理分为人工干预处理和自动处理恢复,无论那种,当异常行为达到接触事件告警条件时,即可解除告警。

## 9.6 集群信息说明

### 9.6.1 各服务常用端口

大数据集群中各服务常用端口参考如下:

服务名称	端口号
Atlas	22000
Canal	8089
DolphinScheduler	12345
EFAK	18048
HBase	16010
HDFS	50070
Hive	10002
Hue	8888
Impala	25000、25010、25020
Kylin	7070
Oozie	11000
Prometheus	9090
Ranger	6080
Solr	8983



服务名称	端口号
Atlas	22000
Canal	8089
DolphinScheduler	12345
EFAK	18048
HBase	16010
HDFS	50070
Spark	18082
TEZ	9999
Trino	18080
YARN	8088、8188、19888
ZKUI	19090
Zeppelin	28080

## 9.6.2 各大数据服务组件 WEB UI 登录账号

大数据集群中各服务 WebUI 默认账号密码如下：

服务名称	账号	默认密码
Atlas	admin	admin
Canal	admin	123456
DolphinScheduler	admin	dolphinscheduler123
EFAK	admin	123456
HUE	首次登录时创建	首次登录时创建
Ranger	admin	USDAdmin123
KYLIN	admin	KYLIN
ZKUI	admin	admin

其他大数据服务原生 Web UI 不需要登陆即可打开。

### 9.6.3 各组件服务部署规则说明

通过 USDP 所安装的大数据集群中，各类服务与组件的部署路径，遵循本节介绍的规则。

#### 9.6.3.1 组件服务安装目录

默认规则：

```
/usr/lib/${服务名称}
```

以 USDP V3.1.13 的集群 Hive 服务安装为例：

```
/usr/lib/hive
```

#### 9.6.3.2 组件服务配置文件目录

默认规则：

```
/etc/${USDP 服务名称}/conf/${原服务配置文件相对路径}
```

以 USDP V3.1.13 版本的集群 Hive 服务配置文件为例，其 hive-site.xml 相关文件在如下目录中：

```
/etc/hive/conf/hive-site.xml
```

**注意：**在 USDP v3.x 中，我们**不希望**使用者**自行登录集群服务器手动修改 USDP 提供的大数据生态服务组件的配置文件！**如若被自行更改，这将导致手动更改的内容不会被记录至该服务的“历史记录”中，更无法回滚生效；同时，手动操作配置内容可能会被管理服务覆盖，导致丢失的情况。**建议使用者充分使用 USDP Web 控制台中提供的服务配置管理功能**，使用和操作方式可参考本本 [《服务配置管理》](#) 章节说明进行操作。

#### 9.6.3.3 组件服务默认数据存储路径

在部署服务时，默认的数据存储路径可以自定义修改，但官方推荐使用默认数据存储路径，方便后期运维，本例以默认路径为例进行说明，部分数据存放于

hdfs 中，已实际配置为准。

默认规则：

```
/data/usdp/${服务名称}
```

以 USDP V3.1.13 版本的集群 Solr 服务的数据存储为例：

```
/data/usdp/solr
```

### 9.6.3.4 组件服务默认日志存储路径

默认规则：

```
/data/log/${服务名称}/
```

以 USDP V3.1.13 版本的集群 HBase 服务日志为例：

```
/data/log/hbase/hbase-hbase-master-$(hostname).log
```

当开发者需要基于某些服务、组件进行开发作业时，可以进入到相应目录进行脚本、API 等操作。如有特殊需要，也可以自行添加环境变量及 classpath。