



数梦工场 一体化智能数据平台 使用手册

V1.0.0

声明

版权声明

版权所有 © 杭州数梦工场科技有限公司 2021。保留一切权利。
非经本公司书面许可，任何单位和个人不得擅自摘抄、复制、修改本文档内容的部分或全部，并不得以任何形式传播。

商标声明



为杭州数梦工场科技有限公司的注册商标。本文档中提及的其他公司的商标、产品标识及商品名称，由其所有人各自拥有。

注意

本文档仅作为使用指导，文档内容可能包含技术上不准确或与产品功能及操作不相符的地方，以数梦工场最终解释为准。本文档中的所有陈述、信息和建议等均不构成数梦工场任何明示或暗示的担保、保证。因为产品版本升级或其他原因，数梦工场将不定期对本文档内容进行更新或修正，恕不另行通知。

公司信息

杭州数梦工场科技有限公司
地址：杭州市西湖区云栖小镇山景路
网址：<http://www.dtdream.com>

修订记录

修订记录

文档版本	发布日期	修改说明
V1.0.0	2021-12-25	第一次正式发布。

目录

1 平台功能简介	1
1.1 数据服务链	1
1.2 数据集成	1
1.3 数据开发	1
1.4 数据治理	1
1.5 大数据计算平台	1
2 用户配置	3
2.1 组织管理	3
2.2 用户管理	3
2.3 租户管理	3
2.4 权限管理	3
3 数据集成	4
3.1 数据连接管理	4
3.2 转换管理	4
3.3 作业管理	5
4 数据开发	6
4.1 数据开发流程	6
4.2 离线开发	6
4.2.1 工作流	6
4.2.2 工作节点	7
4.3 数据管理	7
4.3.1 创建本地 Hive 表	7
4.4 空间管理	9
4.5 发布运行	错误!未定义书签。
5 数据治理	11
5.1 元数据采集	11
5.1.1 同步数据连接	11
5.1.2 自定义添加	11
5.2 数据标准	12
5.2.1 创建标准字典表	12
5.2.2 查看数据标准接入情况	12
5.3 数据模型	12
5.4 数据血缘	13
5.5 质量评估	13
6 数据服务链	14
6.1 功能概述	14
6.1.1 数据资源目录	14
6.1.2 通用数据服务	14
6.1.3 级联管理	14

6.2 资源开发	14
6.3 数据汇聚	15
6.4 资源服务发布	错误!未定义书签。
6.5 资源服务订阅	错误!未定义书签。
6.6 级联管理	16
6.7 大盘管理	16
6.7.1 平台页面展示	16
6.7.2 部门页面展示	16
7 大数据计算平台	17
7.1 告警中心	17
7.1.1 主机告警	17
7.1.2 服务告警	17
7.1.3 告警推送设置	18
7.1.4 告警策略设置	19
7.2 操作审计日志	20
7.3 服务中心	20
7.3.1 服务监控	20
7.3.2 服务管理	22
7.4 运维大盘	25
7.4.1 主机概况	26
7.4.2 服务概况	26
7.4.3 集群监控	27
7.5 集群管理	28
7.5.1 主机监控	28
7.5.2 主机管理	29

1 平台功能简介

1.1 数据服务链

数据服务链把多种数据使用场景抽象成数据服务，比如数据汇聚、数据推送、协议转换、统一查询、数据高铁、比对订阅等，可以快速支撑开发和上线多种数据服务，实现数据服务的即插即用，还可以支持水平部门间的横向级联以及垂直部门间的纵向级联，通过横纵向数据服务链联盟，打造全国信息资源“一盘棋”，实现数据资源的资产化、数据资产的服务化、数据服务的价值化。

1.2 数据集成

数据集成提供对企业数据集成处理功能。通过数据集成，用户可以从不同结构的数据源中抽取数据，对数据进行复杂的加工转换，最后将数据加载到各种存储结构中。例如实现从多个异构的数据源（不同数据库、结构化文件等）抽取数据，并加工成统一的数据格式，最后加载到数据仓库中，供商业智能等应用使用。

1.3 数据开发

数据开发是大数据环境下多用户协同进行数据作业程序开发、测试、生产运维需求的大数据集成开发 IDE，在一体化数据平台中处于承上启下的关键位置：对下支持多种大数据平台，向上为各数据产品提供引擎管理、作业管理、统一 workflow 调度等重要基础服务。

数据开发支持多种计算和存储引擎服务，包括华为 FusionInsight、阿里 MaxCompute、数梦 EMR、星河 TDH 等。

数据开发提供全链路智能大数据开发功能，您可以使用 DataStudio，对数据进行传输、转换和集成等操作，从不同的数据存储引入数据，并进行转化和开发，最后将处理好的数据同步至其它数据系统。

1.4 数据治理

数据治理能够提供异构元数据采集、分类能力，支持高效检索；提供数据清洗能力，规整数据内容；提供数据建模加工能力，形成高价值主题专题。同时，数据治理还具备数据标准的制定能力和数据质量的管控能力，为数据治理各环节提供规范支撑和质量保证；具备智能化能力，提升数据治理各环节效率。

1.5 大数据计算平台

针对大规模数据分析场景所设计的商用大数据产品，为用户提供便捷的分析、处理海量数据的手段，使用户可以不必关心分布式计算细节，从而达到简单高效地分析海量数据的目的。

大数据计算平台提供针对 TB/PB 级别数据的离线和实时处理能力，主要应用于日志分析、数据挖掘、商业智能等领域，适用于政府、运营商、教育科研机构、中小企业和公共事业单位等。
大数据计算平台提供存储（HDFS）和计算(YARN)的引擎能力，需要配合大数据产品使用。

2 用户配置

用户管理平台可以提供组织管理、用户管理、租户管理、权限管理等功能。

2.1 组织管理

使用管理员账号登录，用户可以在“组织管理”中，创建组织，可以达到用户的有效隔离。

2.2 用户管理

使用管理员账号登录，用户可以在“用户管理”中，创建用户，用户填写基本信息。

2.3 租户管理

使用管理员账号登录，用户可以在“租户管理”中，创建租户，将组织与租户及应用关联起来，达到资源的有效隔离。

2.4 权限管理

使用前面租户所有人账号登录，用户可以在“用户授权”界面，创建授权，将用户与角色及权限关联起来，达到权限的有限下放与隔离。

3 数据集成

数据集成主要包括以下主要功能模块：

- 数据连接管理，实现 MySQL、Oracle、SQL Server、RDS、ADS、ODPS、Hive、HDFS 等不同类型数据连接的创建及状态检测等功能。
- 转换管理，在数据连接管理的基础上创建数据交互转换。转换管理维护具体的转换任务，用户还可以通过各种转换插件对数据进行处理，最终输出到目的表。
- 作业管理，在转换管理、文件管理等功能的基础上，提供调度执行的服务。调度类型有手工执行，按秒、按分钟、按天、按周、按月、按季度、按年定时执行，实时执行以及通过 API 执行。

3.1 数据连接管理

在数据连接管理页面单击“创建”，进入创建新数据连接页面，如下图所示。

The screenshot shows a web form titled "创建数据连接" (Create Data Connection). The form contains several input fields, each with a red asterisk indicating it is required. The fields are: "数据连接名称" (Data Connection Name), "描述" (Description), "数据连接类型" (Data Connection Type), "服务器主机名/地址" (Server Hostname/Address), "端口" (Port), "数据库名称" (Database Name), "用户名" (Username), and "密码" (Password). Each field has a placeholder text "请输入..." (Please enter...). At the bottom of the form, there are three buttons: "确定" (Confirm), "测试" (Test), and "取消" (Cancel).

3.2 转换管理

一个完整的转换，必须要有数据源端和数据目的端，即输入插件和输出插件是必选项，中间处理步骤插件可根据实际转换需求选择配置。

此处新建转换流程以数据清洗为例，流程如下所示：

1. 拖拽需要使用的插件至画布：输入、数据清洗和输出插件。

2. 按照输入>清洗>输出的顺序连接各插件。
3. 按照业务需求配置各插件。
4. 保存转换。
5. 运行转换。

3.3 作业管理

一个需要调度执行的任务称为一个作业，在作业管理页面单击“创建”，进入创建作业页面，如下图所示。



4 数据开发

数据集成开发平台 DTSphere DataStudio（以下简称 DataStudio）是杭州数梦工场科技有限公司推出的数据作业开发、生产、一站式开发平台。

DataStudio 支持多种计算引擎，包括 DTSphere EMR（以下简称 EMR）、MaxCompute、FusionInsight、TDH、ZDH 等，提供 SQL、Graph 计算、Spark、机器学习等多种数据处理服务，并支持列存储等多种存储格式。

通过 DataStudio，从不同的数据存储引入数据并进行转化和开发，最后将处理好的数据同步至其他数据系统。

4.1 数据开发流程

主要包括资源管理、创建数据表、配置数据源、工作节点开发、 workflow 开发、发布运行六大部分：



1. **创建数据表**：数据表是进行大数据开发的直接数据来源，数据加工后的结果也储存在数据表中。创建数据表，或使用共享数据表。
2. **数据接入**：根据需要配置数据源，将储存在外部（如 MySQL、Oracle）的数据导入到计算引擎数据表，用于数据开发和清洗。
3. （可选）**资源管理**：添加资源文件、自定义函数供工作节点开发时使用。
4. **工作流开发**：将工作节点配置为工作流，也可使用已发布的工作流或导入工作流，修改、调试工作流，调度工作流及查看工作流日志等。
5. **工作节点开发**：创建、修改、调试工作节点。按需使用已配置数据源的数据表、共享数据表，按需使用资源文件、自定义函数。
6. **发布运行**：配置调度策略，启动调度。

4.2 离线开发

DataStudio 按照业务种类组织相关的不同类型的节点，让您能够更好地以业务进行开发。

4.2.1 工作流

工作流按照业务逻辑将一个或者多个工作节点组织在一起，以 DAG 图的形式展现，描述了一个或者多个节点之间的逻辑依赖关系和运行规则。

支持多种节点类型，包括数据集成转换、Spark SQL、Phoenix SQL、GaussDB SQL 等多种类型节点，通过节点之间的相互依赖，对复杂的数据进行分析处理。

系统支持以目录的形式分类管理工作流与工作节点，将工作流和工作节点以层级结构展示。便于用户

将工作流和工作节点从业务的角度来组织数据开发。

DataStudio 提供多样的工作流调度能力：

- 支持根据时间、依赖关系，进行任务触发的机制。
- 支持分钟、小时、天、周、月、年多种调度周期配置。

4.2.2 工作节点

工作节点是 DataStudio 数据处理和分析过程的最基本单元，每个工作节点对应工作流有向无环图（即 DAG 图）中的一个节点，目前支持的工作节点如下表所示。

工作节点类型	说明
数据集成	
数据集成转换	提供稳定高效、弹性伸缩的数据集成转换服务。
数据集成作业	
数据开发	
Spark SQL	采用的是类似于 SQL 的语法，可以看作是标准 SQL 的子集，但不能因此简单的把 Spark SQL 等价成一个数据库，它在很多方面并不具备数据库的特征，如事务、主键约束、索引等。
MaxCompute SQL	当计算引擎为 MaxCompute 时，离线开发支持 MaxComputeSQL 节点，MaxCompute SQL 采用的是类似于 SQL 的语法。它的语法是标准语法 ANSI SQL92 的一个子集，并有自己的扩展，具体请参见 MaxCompute 产品介绍。
GaussDB SQL	用于用户直接连接到 GaussDB 集群并执行相应 SQL 语句。
Spark	用于用户自定义算法的运行，目前支持 java、Scala、Python 类型算法的运行。
HiveHBase 数据直通	用于 Hive 数据表和 Phoenix 数据表数据的互导。
算法开发	
MLab	用户利用 MLab 算法实验室对数据进行加工处理、数据挖掘、统计分析。

4.3 数据管理

数据管理是对数据表的操作，支持数据表的创建、查看、删除，只能在绑定有计算引擎项目的空间内使用。

4.3.1 创建本地 Hive 表

可通过多种方式来创建 Hive 数据表，包括可视化建表、DDL 建表和 Spark SQL 节点建表三种途径。

所建表的列数较少时，建议采用可视化建表模式，反之建议采用 DDL 模式建表或者 Spark SQL 节点建表。

4.3.1.1 DDL 建 Hive 表

步骤1 左侧导航栏选择“数据管理>Hive 数据表”，进入本地 Hive 表列表。

步骤2 单击右上角“创建”，进入创建数据表页面，在创建数据表页面中对创建模式选择“DDL 模

式”，从默认的可视化建表页面跳转到 DDL 建表页面。



步骤3 编写 SQL 建表语句。

```

/****示例代码****/
create table if not exists table2
(
id string comment'用户 ID',
name string comment'用户名称'
)partitioned by(dt string)
LIFECYCLE 7;
    
```

使用 LIFECYCLE 关键字可管理生命周期（单位为天）。为避免在未备份的情况下删除有用数据，请仅对无需保留的数据表设置生命周期。

4.3.1.2 Spark SQL 节点建表

Spark SQL 节点建表需要先创建 Spark SQL 节点，创建节点后，单击工具栏运行按钮，运行成功后，返回数据表列表，即可看到创建的数据表。

4.3.1.3 可视化建 Hive 表

步骤1 左侧导航栏选择“数据管理>Hive 数据表”，进入本地 Hive 表列表页面。

步骤2 单击右上角“创建”，进入创建数据表页面，默认为可视化建表模式，如下图所示。

创建模式: 可视化模式

* 数据库名称: wjd

* 数据表名称: test001

描述: test1

生命周期: 7

普通列 当列数过多时, 建议使用DOL模式创建, 您已创建2列

列名	类型	长度	小数点	描述	操作
xb	字符串(string)			请输入描述	删除
xm	字符串(string)			请输入描述	删除

添加普通列

分区列 您已创建1列

列名	类型	长度	小数点	描述	操作
data	字符串(string)			请输入描述	删除

添加分区列

保存 重置

4.4 空间管理

可配置以下几项空间属性:

- 是否开启血缘: 开启后运行工作流(包含手动运行和调度运行)将发送 SQL 任务的血缘信息, 记录 SQL 中表的血缘关系。
- 可发布目标空间、可授权目标空间。
- 是否开启周期调度: 未开启调度的空间可进行调度配置但不可执行调度。
- 工作流/工作流 2.0 等待调度超时时间(分钟): 当工作流实例达到理论执行时间, 但依赖条件没有满足, 且等待时间达到设置的超时时间, 则工作流实例处于等待调度超时状态。0 表示不经等待, 立即进入超时状态; -1 表示一直等待, 不进入超时状态。修改后, 需要先停止调度再重新启动调度, 否则调度将按原超时时间计算。
- 失败节点重试次数: 0 表示不重复, 最大重试次数为 10 次。
- 失败节点重试间隔: 重试需等待 30~300 秒的间隔。
- 允许编辑工作流/工作流 2.0 和节点内容: 未启用时, 用户无法使用离线开发/离线开发 2.0 功能。如果开启则无法关闭, 请谨慎操作。
- 引擎设置: 可配置 GaussDB 引擎、达梦引擎、人大金仓引擎以及清洗服务地址。
- 如果空间内需要使用 DataClean 节点, 可配置清洗服务地址。清洗服务地址为 DataRiver 质量节点的地址。
- 设置空间级别变量。

4.5 发布运行

实际应用中, 常常使用脱敏数据进行开发, 使用真实数据进行生产。通常, 真实数据仅对生产人员可见, 对开发、测试人员不可见。为此, DataStudio 的工作空间按照类型分成开发空间、生产空间。

1. 开发空间基于脱敏数据, 由开发人员进行节点、工作流的开发和验证。
2. 最后发布到实际的生产空间, 基于真实的数据来生产调度、运维, 并进行监控告警。

发布运行使得开发空间、生产空间之间能够快速同步, 开发项目可以迅速部署到生产环境。较之工作流的导入导出, 发布运行更方便快捷, 免去网络传输等待时间以及人为失误。

发布运行主要提供以下功能：

- 发布方**租户管理员**可以选择工作流（可选调度信息）、jar 包、自定义函数、共享包等资源，配置发布计划，一键发布到目标空间，在接收方接受前可撤回。发布方**空间所有用户**都可以查看发布包状态。
- 接收方**租户管理员**可以查看、接受或拒绝收到的发布包。接收方**空间所有用户**都可以查看收到的发布包。

需要在创建空间时或发布前，在“空间管理>空间属性”中绑定可发布的目标空间。

5 数据治理

5.1 元数据采集

数据连接根据接入方式不同，分同步连接和自定义添加数据连接两种。

5.1.1 同步数据连接

通过在平台创建引擎和添加项目，将 MaxCompute 和 EMR 计算引擎上添加的项目自动同步至“数据连接”中，以“项目名@引擎名”的形式展现。

5.1.2 自定义添加

在“连接管理”界面，单击“新建”，创建新的数据连接，包括添加 MySQL、Hive2、DRDS、ADS、RDS、Oracle、GaussDB、Kafka、DM、KINGBASE 等连接类型的数据连接。

新建数据连接

* 数据连接类型: 请选择数据源连接类型

* 数据连接名称: 请输入数据连接名称

数据连接描述: 请输入数据连接描述

* 服务器地址: 请输入服务器地址

* 端口号: 请输入端口号

* 数据库名称: 请输入数据库名称

* 同步元数据: 是 否

* 用户名: 请输入用户名

* 密码: 请输入密码

* 配置等价连接信息: 是 否

关闭 测试连通性 确定

5.2 数据标准

5.2.1 创建标准字典表

字典表是定义数据元值域的一种方式，如：在电话区号中，杭州市区号的取值为 0571；在公安性别中，男的取值为 1。

字典名		字典值	
1	男	1	1
2	女	2	2
3	未知	3	0

5.2.2 查看数据标准接入情况

在“总体大盘”界面，查看标准数据元和标准字典表的接入汇总。



5.3 数据模型

逻辑模型是根据业务规则确定的实体的集合，配置逻辑模型主要包括添加实体以及对应的属性。

模型上传
下载模板

模型导入
 PowerDesigner导入

将文件拖拽到此处上传或单击图标选择文件上传

仅支持单个文件上传，支持10MB以内的zip、xls、xlsx文件。

数据模型.zip 5.6M

关闭 导入

5.4 数据血缘

数据血缘，用于记录数据的来源、数据之间的流向、数据之间的关系等信息，并通过可视化图形进行展示。

5.5 质量评估

在“质量评估”界面，从多个维度评估数据表的质量，并根据预设权重，综合计算质量得分。

名称	执行频率	租户名称	状态	最新执行完成时间	执行统计	规则数	最新执行成功/失败数	评估分数	操作
> 全表	每天	1125	● 调度中	2021-12-24 11:03:06	详情	2	2 / 0	70.72	详情 导出
> 数据地图质量评估-20211223172224	手动执行	sk实时清洗租户2	● 已完成	2021-12-23 17:23:05	详情	6	6 / 0	100	详情 导出
> 数据地图质量评估-20211222144219	手动执行	科技和信息化司	● 已完成	2021-12-23 16:46:06	详情	16	16 / 0	98.38	详情 导出
> quality_001_规则检查_1637658519028	手动执行	quality	● 已完成	2021-12-21 17:18:08	详情	3	0 / 3		详情 导出
> 发撒倍德萨	手动执行	测试用大数据组织	● 已完成	2021-12-16 20:08:08	详情	1	1 / 0		详情 导出
> odps-身份证和手机号码	手动执行	delete2	● 已完成	2021-12-16 19:52:08	详情	2	2 / 0	87.13	详情 导出
> 3表评估-非全表	手动执行	quality_autotest	● 已完成	2021-12-16 19:32:15	详情	4	4 / 0	89.17	详情 导出
> 121604	手动执行	quality_autotest	● 已完成	2021-12-16 18:41:14	详情	1	1 / 0	100	详情 导出
> 121603	手动执行	quality_autotest	● 已完成	2021-12-16 18:27:07	详情	1	1 / 0	100	详情 导出
> 121605	手动执行	quality_autotest	● 已完成	2021-12-16 18:25:07	详情	2	2 / 0	62.5	详情 导出

共有591条，每页显示: 10 < 1 2 3 4 5 ... 60 > Go

6 数据服务链

6.1 功能概述

数据服务链提供信息资源目录、通用数据服务、数据级联服务、平台管理、平台安全等主要功能。

6.1.1 数据资源目录

数据资源目录可以帮助用户了解数据代表的业务含义，弥补业务与 IT 之间的隔阂，让数据更容易的被组织内部的各种角色所理解。同时，还可以很容易的获取数据的属性信息，让用户对数据资料有更全面和直观的了解。

数据服务链以数据目录的形式展示部门发布的数据及服务，其他用户可以按分类、部门浏览数据，也支持关键字搜索数据。

6.1.2 通用数据服务

数据服务链对数据服务进行全生命周期的监督管理，以数据服务的形式支撑各种业务场景。数据服务全生命周期包括：服务开发、服务发布、服务订阅、服务调用、质量监控和优化提升。

数据服务链中的数据服务不是固化的，可在具体场景中进行不断生长迭代，不同行业、不同场景、不同用户都可进化出适合自己的数据服务集。

典型服务包括数据汇聚服务、数据推送服务、数据查询服务、数据统计服务、比对订阅服务等。

6.1.3 级联管理

提供级联功能，多个不同的数据服务链产品可以通过级联功能受控级联自己的目录和目录所挂接的服务。有效实现上级数据服务建设者向下级单位发布建设成果、满足加快促进地方信息化建设的要求。

6.2 资源开发

系统支持的资源类型有：库表、接口、文件、实时数据流。

- 库表资源是指数据库中的数据，数据库可以是关系型数据库也可以是非关系型数据库。
- 文件资源是指可通过文件服务器访问的资源，可以是单个文件也可以是文件夹下的多个文件。
- 接口资源是指可通过 API 接口访问的资源。
- 实时数据流是指 Kafka 实时数据资源。

不同的服务可挂接不同的资源，目录添加资源前，需要完成对应类型资源的创建。

6.3 数据汇聚

通过上级集中创建并发布数据汇聚任务的方式，让所有的下级单位收到同样标准的数据汇聚任务流程；下级单位通过按照汇聚任务流程的要求提供本地数据，让自己的数据上报给上级单位；在上报过程中，通过标准项检查，来保证下级汇聚的数据正确性；在上级单位提供数据合并功能，让上级单位直接看到一张从各个地方汇聚而成的汇总表，减轻上级单位的数据合并工作负担。

6.4 资源服务发布

通过目录注册功能发布的数据，必须经过必要的审批流程后才能展示在数据目录中。目录注册流程如下所示。



- 目录编制阶段，用于完成目录基础信息的录入。
- 添加资源属于服务挂接阶段，用于录入该服务挂接的资源。
- 添加服务属于服务挂接阶段，在添加资源后，在资源上添加服务。添加服务完成后，将目录提交部门审核。
- 部门审核阶段，用于审核已提交目录的真实性及准确性，并且发布到中心平台。
- 中心审核阶段，用于审核部门发布的目录是否符合中心平台的规范。
- 中心审批完成后将直接发布到门户，供数据使用人员使用该目录。

上述流程是一个完整的流程，用户可根据自己部门的实际情况对流程进行定制。

6.5 资源服务订阅

资源服务订阅是需求部门获取供应部门发布数据的过程。根据数据资源形态以及资源共享级别的不同，资源服务订阅流程有所不同。资源服务订阅过程有严格的流程管理，

除数据汇聚可提供免申请汇聚（任务下发）的方式进行数据使用外，其他服务都需要数据使用人员通过提交申请进行资源服务订阅。

申请资源服务订阅流程如下：



- 服务申请阶段，用于触发启动交换流程。
- 申请方审批阶段，用于确认是否受理该项申请。
- 提供审批阶段，用于确认是否将资源提供给申请部门。
- 平台运营审批阶段，用于决策是否给申请部门数据使用。
- 执行阶段，该阶段可以使用资源对应的数据。

上述流程是一个完整的流程，用户可根据自己部门的实际情况对流程进行定制。

6.6 级联管理

级联针对的对象是信息资源目录。信息资源目录可以支撑包括库表、文件和接口各种类型的资源，这些资源可以以多种不同的服务方式挂接在信息资源目录上。当信息资源目录被级联，所挂接的服务将和信息资源一道被级联到下级节点。

数据服务链提供横向、纵向的级联功能，可实现目录、服务和标准的级联。

- 横向级联指的是平行部门之间的交互联动，比如省与省之间、市与市之间，打通平行部门之间的资源目录和服务，实现跨省、跨市场景下的资源目录和服务的可见、可申请、可使用。
- 纵向级联指的是垂直部门之间的交互联动，比如部省之间、省市之间，打通垂直部门之间的资源目录和服务，实现部省、省市之间的资源目录和服务的可见、可申请、可使用。

6.7 大盘管理

6.7.1 平台页面展示

平台用户在“资源目录”界面，查看整个平台内所有部门下目录发布及目录资源挂载等相关信息。通过筛选部门，可以查看指定部门下的资源目录统计信息。

页面数据默认 60 秒刷新一次，可在页面右上角单击“设置”调整刷新时长。



6.7.2 部门页面展示

部门用户可查看当前部门下目录发布及目录资源挂载等相关信息。

页面数据默认 60 秒刷新一次，可在页面右上角单击“设置”调整刷新时长。

7 大数据计算平台

7.1 告警中心

告警中心可对主机告警及服务告警进行设置和告警信息管理，支持告警通知。

7.1.1 主机告警

在主机告警页面单击“修改”，进行阈值设置，用于对 CPU、内存、磁盘的告警阈值设置。

CPU阈值	WARNING	50	%	CRITICAL	80	%
内存阈值	WARNING	50	%	CRITICAL	80	%
磁盘阈值	WARNING	50	%	CRITICAL	80	%

[修改](#)

在告警详情页面，查看当前和历史告警信息，用于监控“警告”或“严重”告警，如出现告警，系统将自动发送通知给收件人邮箱。



7.1.2 服务告警

在服务告警页面，打开当前和历史告警的告警开关，显示告警名称、告警状态、最近状态变化时间。单击告警名称后可以查看详细信息。

告警名称	告警状态	最近状态变化时间	告警开关
datanode cpu limit	正常	2017-11-30 20:20:04	已开启
DataNode Web UI	正常	2017-11-29 17:49:11	已开启
NameNode High Availability Health	正常	2017-11-30 19:52:52	已开启
NameNode Directory Status	正常	2017-11-30 19:53:29	已开启
NameNode RPC Latency	正常	2017-11-30 19:53:27	已开启
NameNode Web UI	正常	2017-11-30 19:52:28	已开启
HDFS Upgrade Finalized State	正常	2017-11-29 17:40:03	已开启
DataNode Health Summary	正常	2017-11-30 19:54:27	已开启
HDFS Capacity Utilization	正常	2017-11-30 19:55:26	已开启
Percent DataNodes Available	正常	2017-11-29 17:49:20	已开启

单击任意告警，弹出告警设置界面，单击右上角 按钮，可以在本页修改告警设置（可能包含检查间隔、阈值等）。

datanode socket limit 已开启

描述 This service-level alert will trigger if socket number beyond a certain threshold.

检查间隔 120 分

主机名称	状态	最近状态变化时间	告警内容
at1.dtdream.com	正常	2017-11-08 18:58:08	the current sockets is 1337.0, the thresh...
at2.dtdream.com	警告	2017-11-08 18:58:02	the current sockets is 2252.0, the thresh...
at3.dtdream.com	正常	2017-11-08 18:58:09	the current sockets is 978.0, the threshol...
at4.dtdream.com	正常	2017-11-08 18:58:06	the current sockets is 531.0, the threshol...
at5.dtdream.com	正常	2017-11-08 18:58:09	the current sockets is 472.0, the threshol...

7.1.3 告警推送设置

在告警推送设置页面，开启告警推送开关，设置推送地址，系统会将告警信息推动到 SCBB 的运维中心。

告警推送设置

是否开启告警推送: ON

推送地址:

7.1.4 告警策略设置

在告警通知设置页面，单击右上角“创建告警策略”，按照如下图所示填写相关内容，用于系统发送邮件通知告警。

创建告警策略 ✕

* 告警名称

服务 所有 自定义

* 告警级别

描述

接收邮箱

SMTP服务器

SMTP端口

发送邮箱

身份认证

邮件告警内容如下图所示，告警内容为 HDFS 服务告警，数据节点内存上限，内存使用：73.82%，集群：master，主机：host3.dtdream.com。



7.2 操作审计日志

在审计日志页面，查看 EMR 平台的操作记录，包含对项目空间、应用的操作。

id	操作内容	操作人	操作时间
42	执行管理操作	dtdream	2018-03-08 13:52:25
40	执行管理操作	dtdream	2018-03-06 11:16:40
38	执行管理操作	emradmin	2018-03-02 16:34:57
36	执行自检	emradmin	2018-03-02 16:34:16
35	执行管理操作	nzwatch	2018-02-06 11:27:49
33	停止主机 auth10.master.com 上的所有组件	huangn	2018-01-31 15:23:20
31	停止主机 auth11.master.com 上的所有组件	huangn	2018-01-31 15:17:26
29	停止主机 auth12.master.com 上的所有组件	huangn	2018-01-31 14:50:51
27	重启服务 SPARKDOT	huangn	2018-01-22 11:05:21
25	执行管理操作	poseidon	2018-01-12 17:38:08

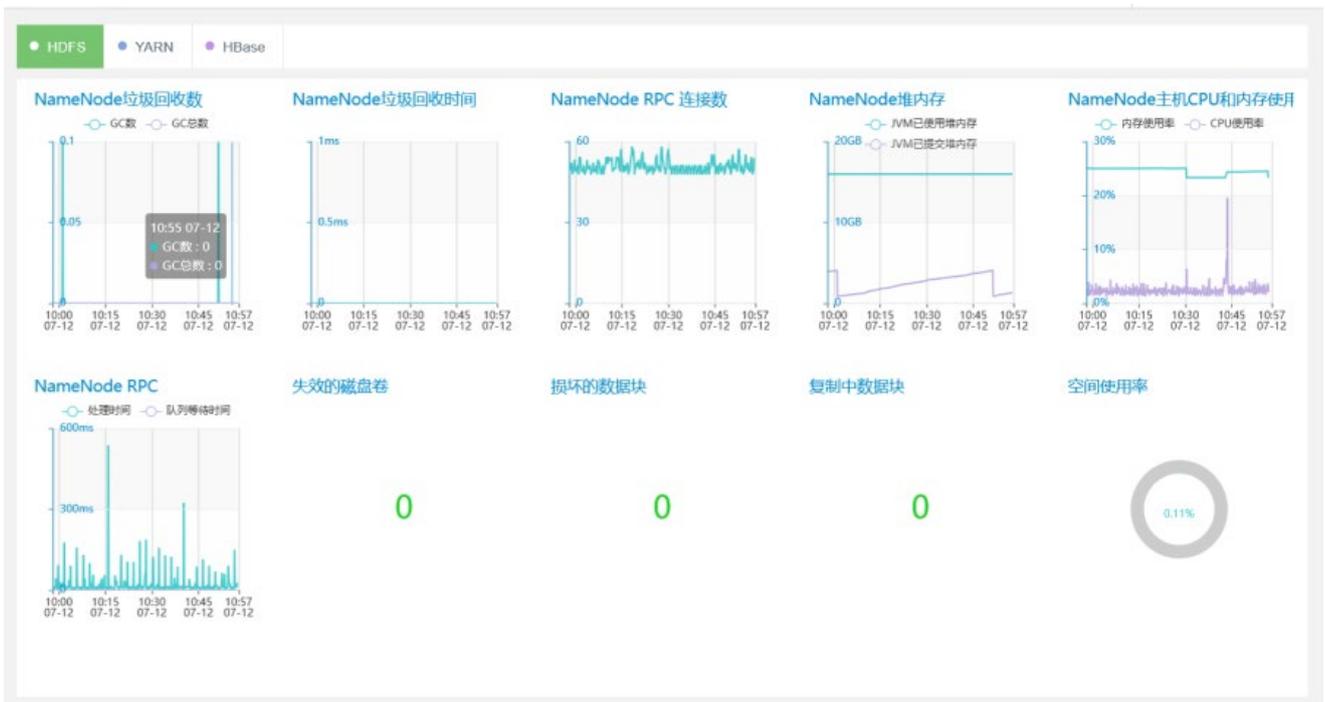
7.3 服务中心

服务中心对 EMR 包含的服务进行集中监控、管理。

7.3.1 服务监控

HDFS

在服务监控页面，选择 HDFS，查看 HDFS 支持的监控项：NameNode 垃圾回收数、NameNode 垃圾回收时间、NameNode RPC 连接数、NameNode 堆内存、NameNode 主机 CPU 和内存使用率、NameNode RPC、失效的磁盘卷、损坏的数据块、复制中数据块、空间使用率。



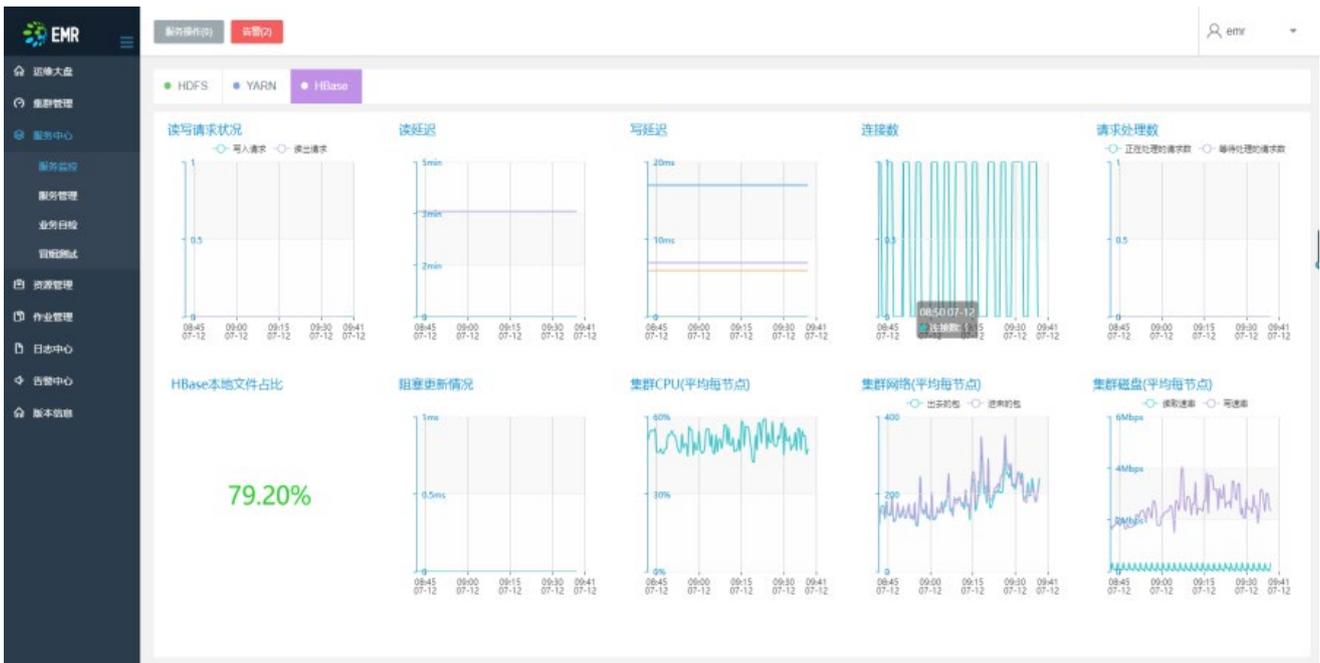
YARN

在服务监控页面，选择 YARN，查看 YARN 支持的监控项：集群内存、内存使用率、CPU 使用率、Container 失败率、正在等待资源的 App 数量、App 失败率、集群磁盘、集群网络、集群 CPU。



HBase

在服务监控页面，选择 HBase，查看 HBase 支持的监控项：读写请求状况、读延迟、写延迟、连接数、请求处理数、HBase 本地文件占比、阻塞更新情况、集群 CPU、集群网络、集群磁盘。



7.3.2 服务管理

在概览页面，主要展示服务的状态和服务的组件状况。

The service overview for HDFS shows the following component status:

- DataNode: 5/5 开启
- HDFS Client: 安装
- JournalNode: 3/3 开启
- Standby NameNode: 开启
- Active NameNode: 开启
- ZKFailoverController: 2/2 开启

单击服务操作，支持服务的启动、关闭、重启操作。

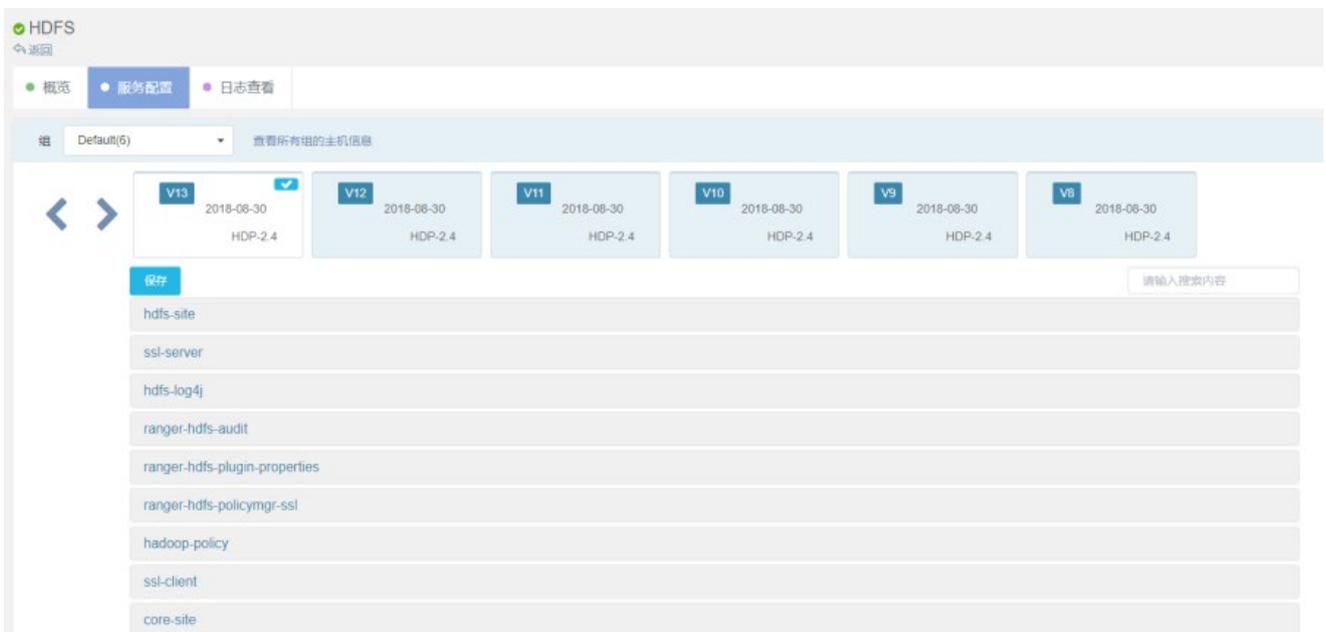
服务操作

- 服务启动
- 服务关闭
- 服务重启

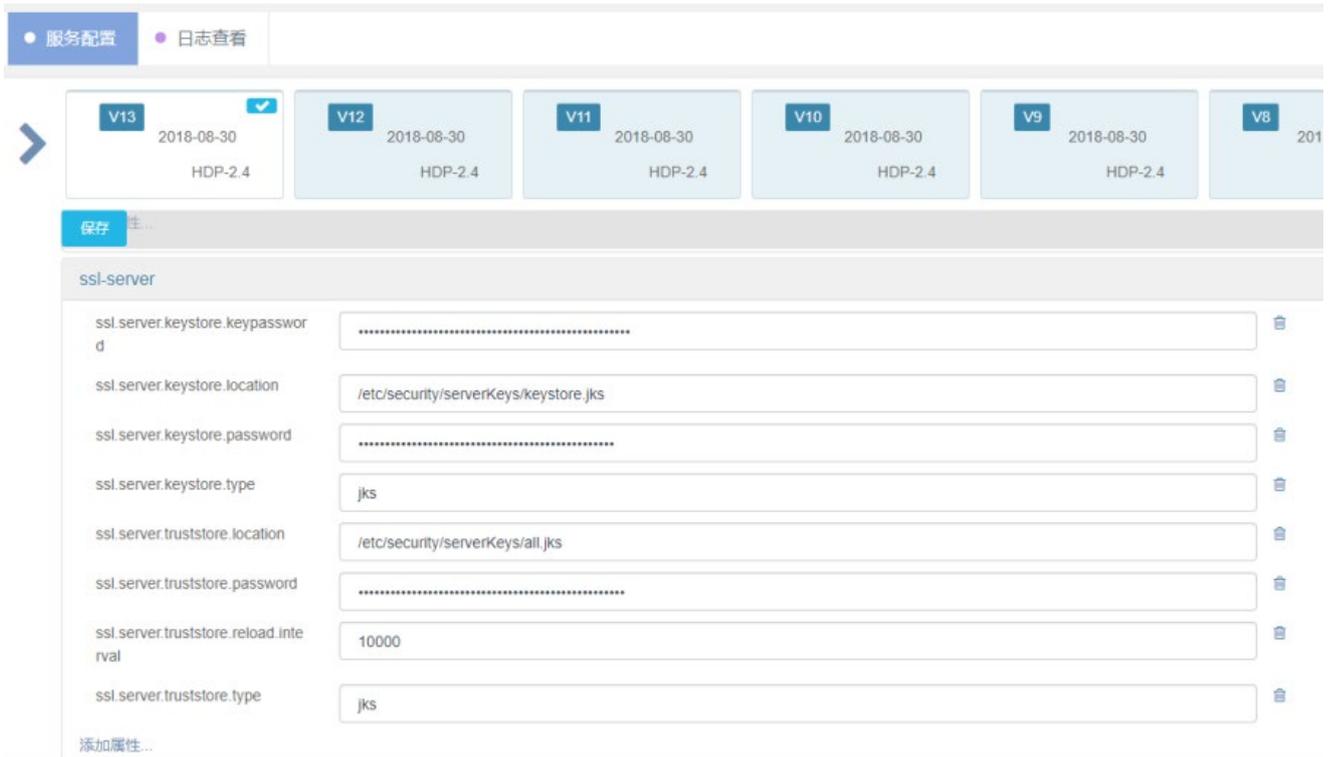
单击左上角“服务操作”，查看服务操作记录。



在服务配置页面，展示所有的配置版本，每个版本相当于一个快照，每次修改配置后，修改的配置会保存为新的版本。



每个服务的配置按照具体的配置文件为单位组织，方便查找。



单击“添加属性”，修改配置文件，用于在已有的配置文件中添加配置项。



通过组下拉列表可筛选该组下的服务配置，用于将集群中的主机进行分组，每一组应用不同的配置，来达到不同主机可以配置不同参数的效果。



单击“查看所有组的主机信息”，可以查看该服务的不同服务分组下的所有主机。

查看 HDFS 组主机信息



您可以查看HDFS 不同组的的主机

Default(0)
datamax11:hdfs_yarn(2)
datamax11:controller_1(1)
datamax11:hdfs_hbase(1)
datamax11:controller_0(1)
datamax11:controller_2(1)

确定

7.4 运维大盘

在运维大盘页面，显示 EMR 节点的大盘信息，包括集群概况、主机概括、服务概况、内存使用率、磁盘使用率、CPU 使用率等，其中集群 CPU、集群内存、集群系统负载、集群网络信息均可以以曲线图展示。



7.4.1 主机概况

选择主机概况，显示每个主机的状态。

主机名	IP地址
node1.dtdream.com 12	10.156.23.210
node2.dtdream.com 16	10.156.23.211
node3.dtdream.com 13	10.156.23.212
node4.dtdream.com 5	10.156.23.213
node5.dtdream.com	10.156.23.214
node6.dtdream.com	10.156.23.215

7.4.2 服务概况

选择服务概况，显示每个服务的状态。单击服务名称可进入详情页面查看详细信息。

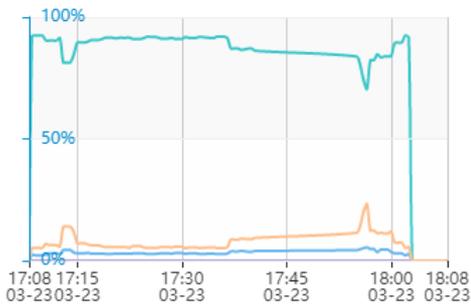
服务名称	服务类型	服务状态
Ambari Metrics	Ambari	正常
Basic Components	Pnd	异常
Databridge	Pnd	正常
Datatrain	Pnd	正常
Flume	Ambari	正常
HDFS	Ambari	正常
MapReduce2	Ambari	正常
Ranger	Ambari	正常

7.4.3 集群监控

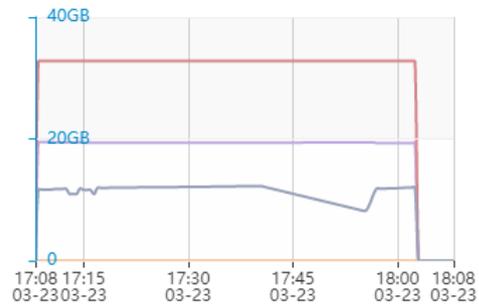
在集群监控页面，以曲线图的形式实时（每隔 15 分钟刷新一次）展示每日集群的 CPU、内存、负载、网络信息。

监控项

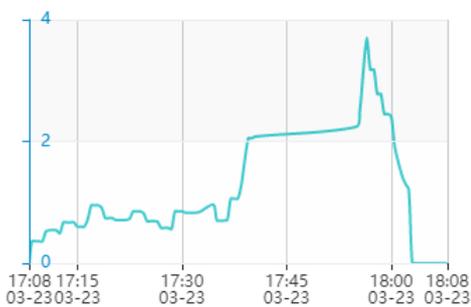
集群CPU(平均每节点)



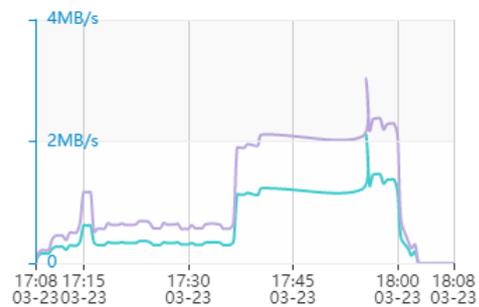
集群内存(平均每节点)



集群系统负载(平均每节点)



集群网络(平均每节点)



7.5 集群管理

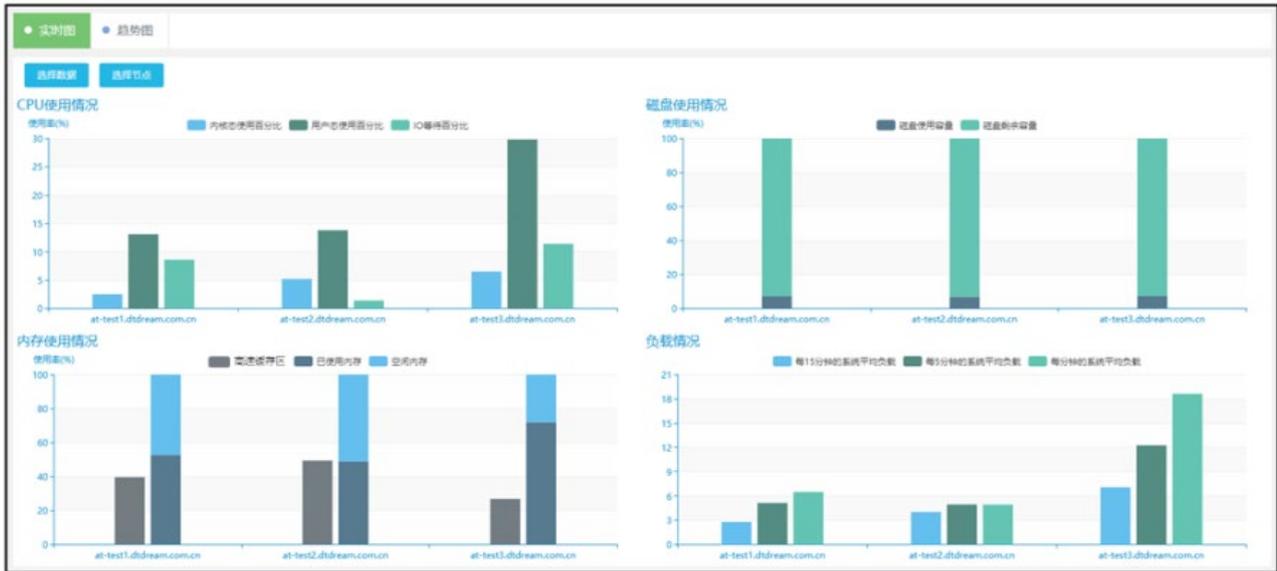
集群管理对集群上所有主机进行集中监控、管理。

7.5.1 主机监控

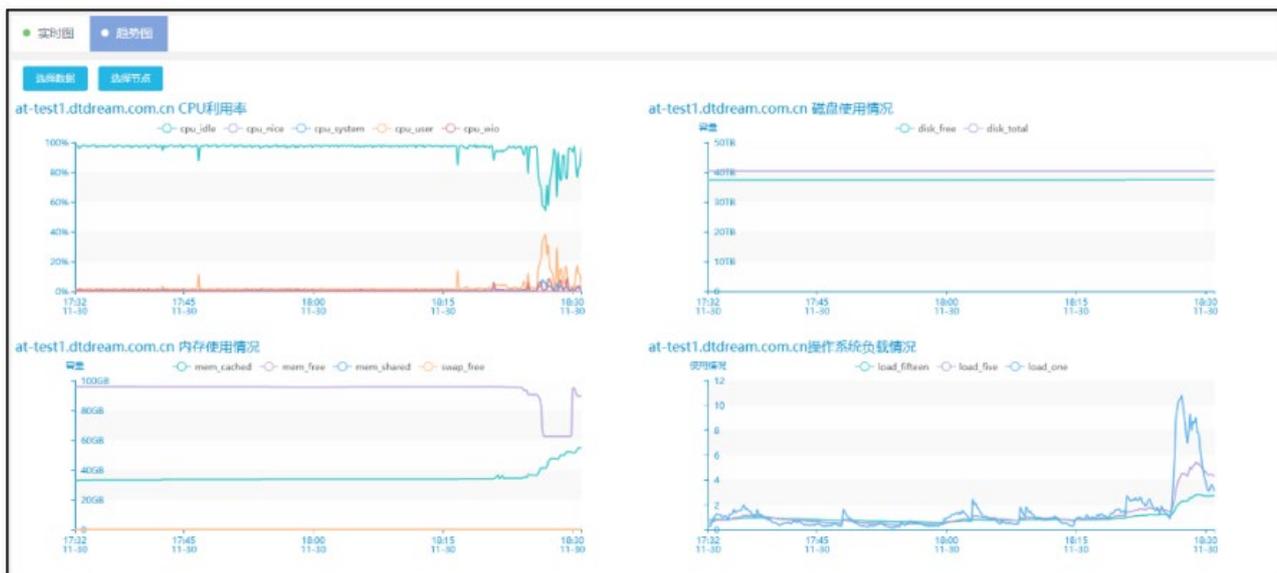
在主机监控页面，包括实时图、趋势图两部分。均可通过单击左上角的“选择数据”选择展示的监控图监控数据，通过单击左上角的“选择节点”选择展示的主机节点。



选择实时图页面，实时展示集群主机的 CPU、磁盘、内存、负载、网络、进程情况。

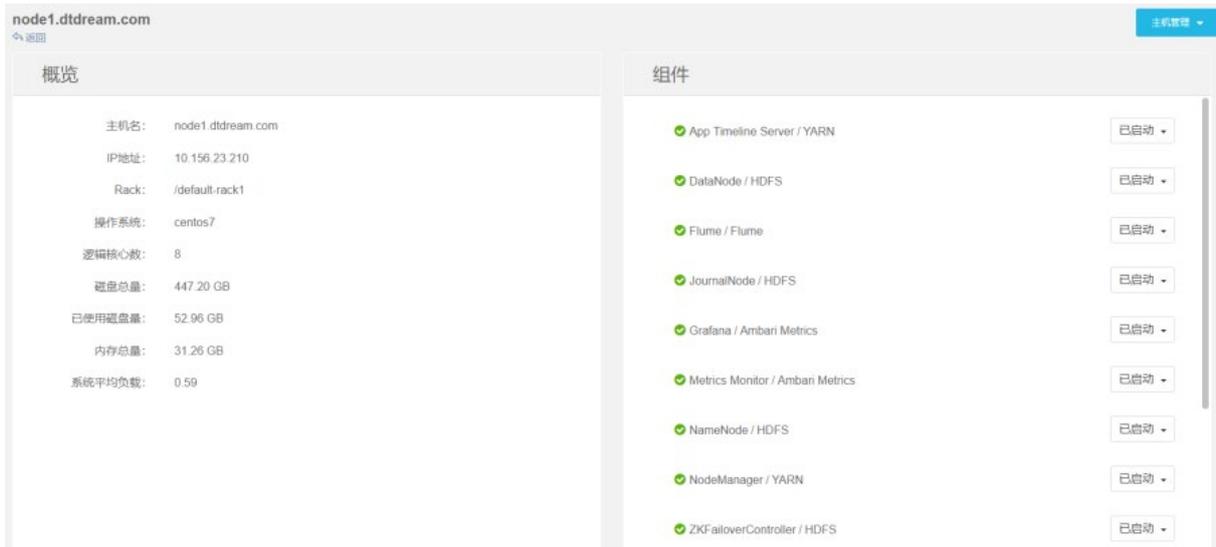


选择趋势图页面，以折线图的形式展示一个主机的 CPU、磁盘、内存、负载、网络、进程在过去一段时间的情况。



7.5.2 主机管理

在主机管理页面，单击主机列表中的主机名进入主机详情页面，查看主机的 IP 地址、Rack、操作系统、逻辑核心数、磁盘总量、已使用磁盘量、内存总量、平均负载以及该主机上安装的组件及组件状态。



在每个主机组所在行，支持启动、停止、重启该主机上的组件。

