



CNUTCon 2017
全球运维技术大会

苏宁大数据平台运维实践

王志强
苏宁云商·大数据平台技术总监

QCon

全球软件开发大会

10月17-19日 上海 · 宝华万豪酒店

➤ 扫码锁定席位

九折即将结束

团购还享更多优惠，折扣有效期至9月17日
扫描右方二维码即可查看大会信息及购票



如果在使用过程中遇到任何问题，可联系大会主办方，欢迎咨询！

微信: qcon-0410

电话: 010-84782011

ArchSummit

全球架构师峰会 2017

➤ 扫码锁定席位

12月8-9日 北京 · 国际会议中心

七折即将截止立省2040元

使用限时优惠码AS200，
以目前最优惠价格报名ArchSummit
仅限前20名用户，优惠码有效期至9月19日，
扫描右方二维码即可使用



如果在使用过程中遇到任何问题，可联系大会主办方，欢迎咨询！

微信: aschina666

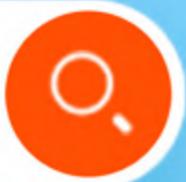
电话: 15201647919



极客搜索

全站干货，一键触达，只为技术

s.geekbang.org



扫描二维码立即体验

有没有一种搜索方式，能整合 InfoQ 中文站、极客邦科技旗下12大微信公众号矩阵的全部资源？

极客搜索，这款针对极客邦科技全站内容资源的轻量级搜索引擎，做到了！

扫描上方二维码，极客搜索！

这里只有 技术领导者

EGO会员第二季招募季正式开启



—— E小欧 ——

报名时间：9月1日-9月15日
扫描添加E小欧，
邀您进入EGO会员预报名群

立即报名



你认为以下哪些关键字对运维最重要？



TABLE OF CONTENTS

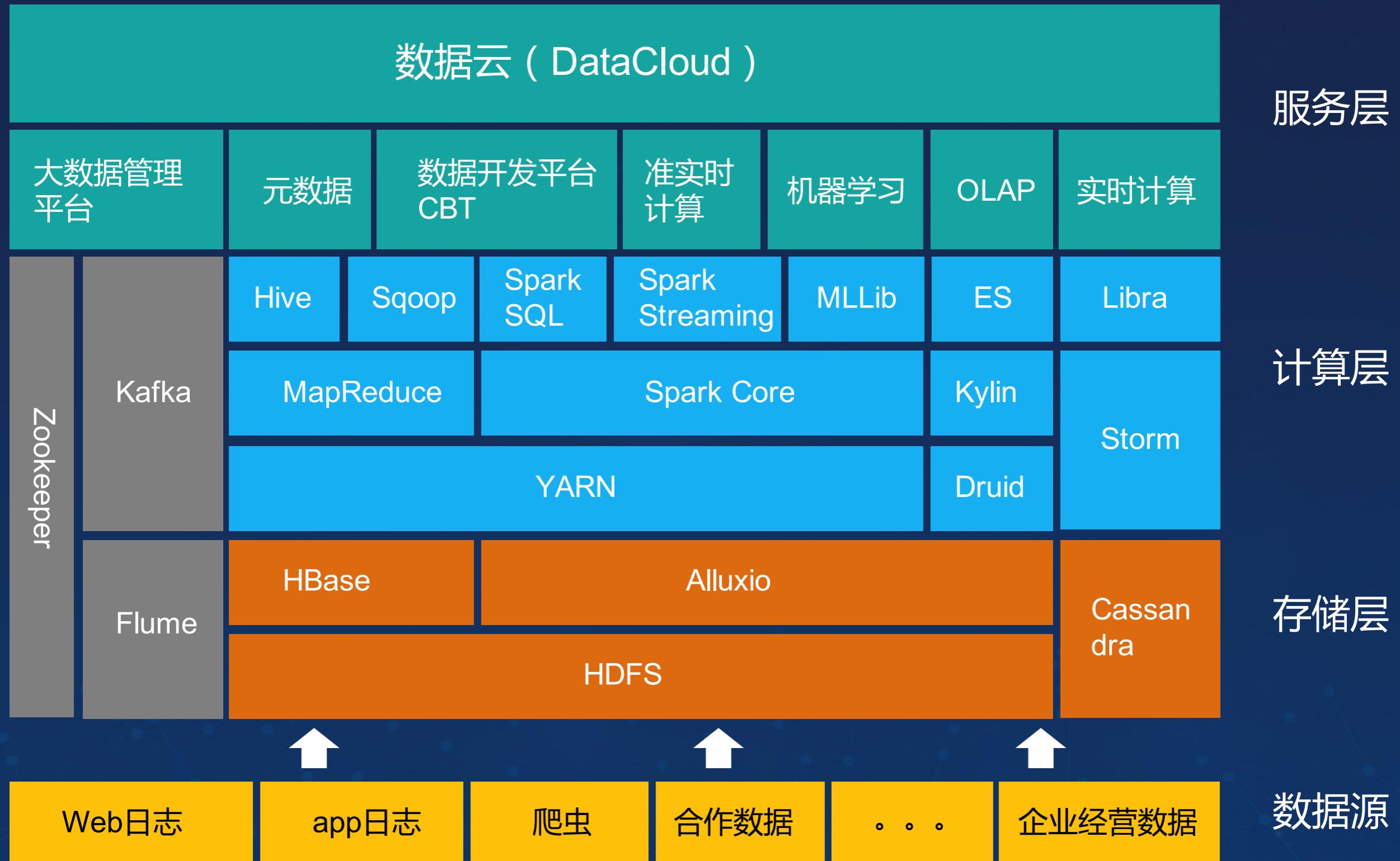
苏宁大数据平台基本介绍

大数据平台运维的痛点及解决方案

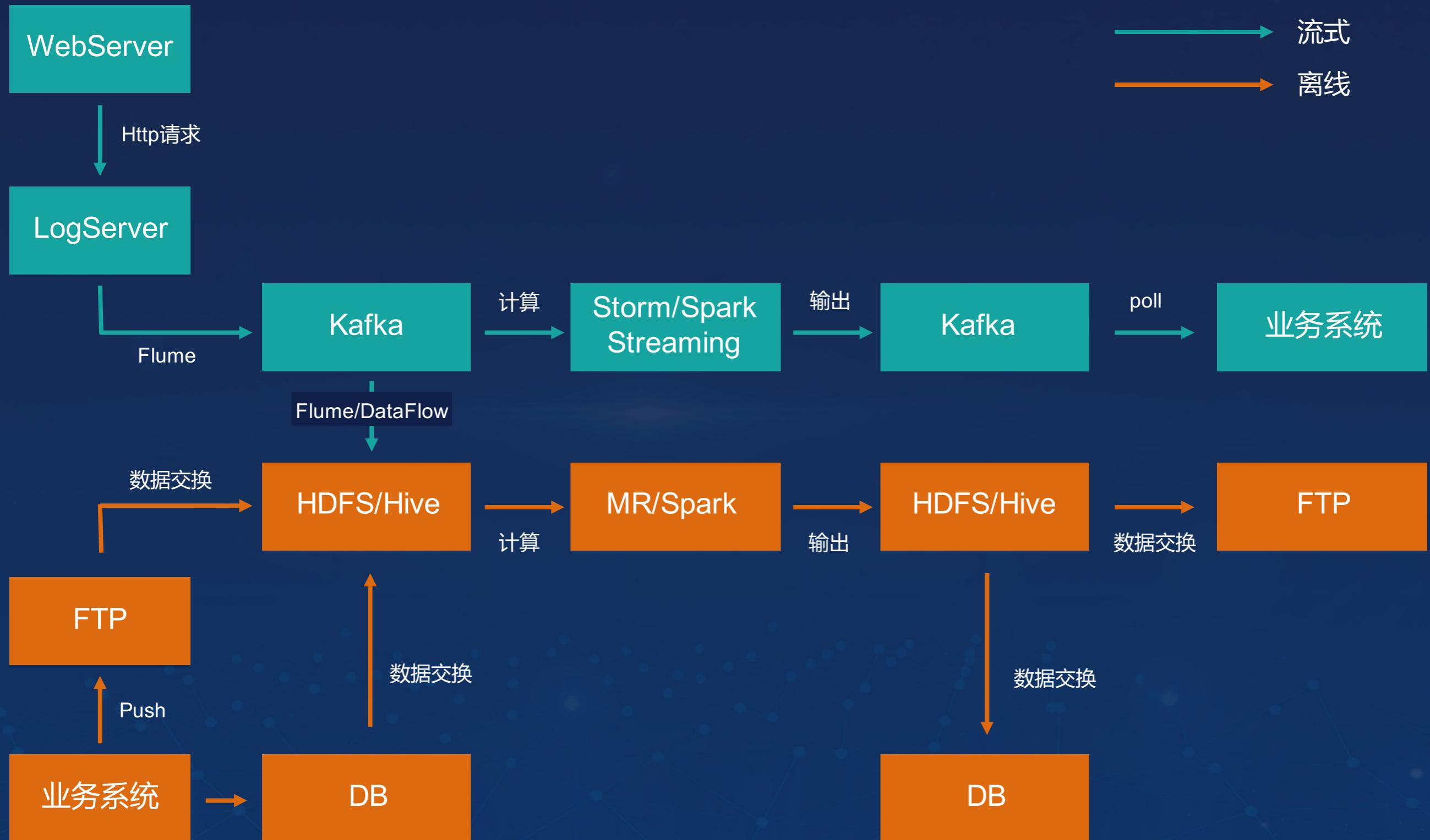
平台优化及增强

DOING & TO DO

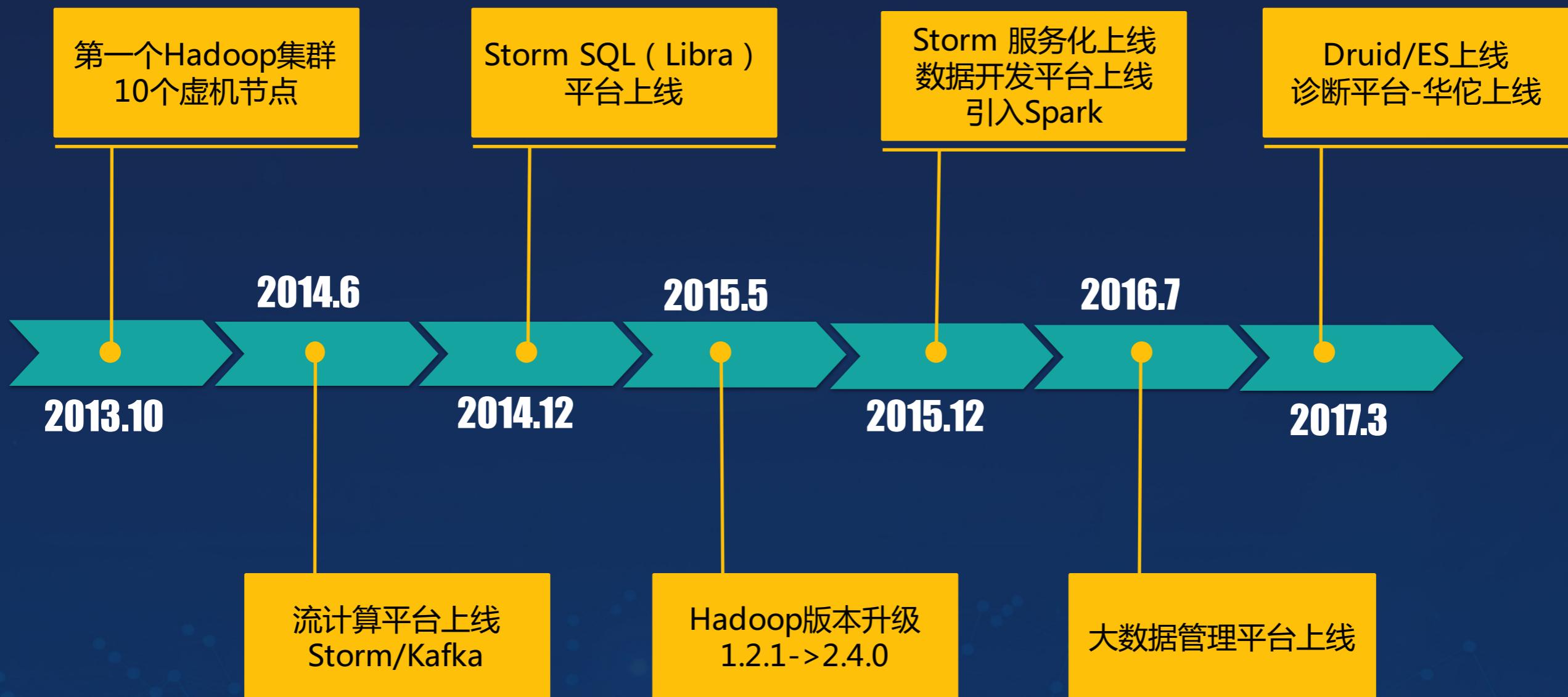
苏宁大数据平台软件栈介绍



典型数据流向



平台发展历程



平台规模

离线计算

1500节点
5万任务流，30万任务/天
处理500TB/天
新增30TB/天
150账户，500+开发人员
Hive:MR:Spark=90:3:7

KV存储(HBase)

160节点，6个集群
50个NameSpace
600+张表
高峰期200万QPS

流式计算

Storm :
1400虚拟机，41个集群
单集群最大174节点
1500流计算任务
Spark Streaming :
93节点，80个任务

OLAP引擎

Druid :
39节点
30个DataSource
Elastic Search :
15节点
4个Index

TABLE OF CONTENTS

苏宁大数据平台基本介绍

大数据平台运维的痛点及解决方案

平台优化及增强

DOING & TO DO

大数据平台运维的痛点

- 痛点1. 部署及运维复杂
- 痛点2. 无资源使用视图
- 痛点3. 任务相互影响，资源隔离性差
- 痛点4. 排查问题耗时长，应用优化门槛高

痛点1. 部署及运维复杂

平台管理员

集群组件辣么多，部署环境辣么多，累成狗。 . .

堡垒机->跳板机->运维机，就为了开个权限。 . .

谁懂Flume运维的苦？

业务开发

我开发用到一个用户行为的数据，在哪个表里？

配置crontab的方式串联任务，伤不起。 . .

不同账户间的任务相互依赖严重，跨账户依赖难实现

解决1. 平台化、自动化

大数据管理平台：主机管理，集群管理自动化

The screenshot shows a web-based management interface for a 'Big Data Management Platform'. The top navigation bar includes links for '帮助' (Help) and a user profile for '王志强'. On the left, a sidebar menu lists several categories: '集群管理' (Cluster Management), '主机' (Host) which is highlighted with a red border, '客户端' (Client), '账户' (Account), '权限' (Permissions), '资源' (Resources), '监控' (Monitoring), '资源展板' (Resource Dashboard), and '任务展板' (Task Dashboard). The main content area is titled '主机管理' (Host Management) and displays a table of host information. The table columns are: 编号 (Number), IP, Host, 主机类型 (Host Type), and 操作 (Operations). The data in the table is as follows:

编号	IP	Host	主机类型	操作
1	10.27.1.138	namenode1-sit.cnsuning.com	cluster	详情 销毁
2	10.27.1.141	namenode2-sit.cnsuning.com	cluster	详情 销毁
3	10.27.1.142	slave01-sit.cnsuning.com	cluster	详情 销毁
4	10.27.1.143	slave02-sit.cnsuning.com	cluster	详情 销毁
5	10.27.0.241	slave03-sit.cnsuning.com	cluster	详情 销毁
6	10.27.0.242	slave04-sit.cnsuning.com	cluster	详情 销毁

A large teal button labeled '主机列表' (Host List) is centered over the table. The bottom right corner of the page features the 'Geekbang InfoQ' logo.

解决1. 平台化、自动化

大数据管理平台：主机管理，集群管理自动化

大数据管理平台

帮助 王志强

集群管理 >

主机

客户端

账户

权限

资源 >

监控 >

资源展板 >

任务展板 >

OLAP >

HBase HA管理 >

HBase数据探查

主机管理 > 详情

主机基本信息

HostName namenode1-sit.cnsuning.com
IP Addr 10.27.1.138
OSVersion

主机资源信息

CPU 24 Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz
内存 264484056
磁盘 3.5T 详情
网络 2000Mb/s

刷新

全部启动 全部停止

主机部署组件信息

HMaster common-hbase-sit 启动 停止
NameNode common-hdfs-sit 启动 停止
JournalNode common-hdfs-sit 启动 停止

主机详情

主机运行组件

组件	实例名	状态	操作
HMaster	common-hbase-sit	启动	停止
NameNode	common-hdfs-sit	启动	停止
JournalNode	common-hdfs-sit	启动	停止

解决1. 平台化、自动化

大数据管理平台：主机管理，集群管理自动化

The screenshot shows the 'Big Data Management Platform' interface. On the left is a dark sidebar with white icons and text: '集群管理' (highlighted with a red box), '集群展板', '长任务进度', '软件部署', '配置管理', '主机', '客户端', '账户', '权限', '资源' (with a right arrow), '监控' (with a right arrow), and '资源展板' (with a right arrow). The main area is titled '集群展板' (Cluster Dashboard). It features tabs for 'HDFS集群' (selected), 'YARN集群', 'HBase集群', and 'ZooKeeper集群'. Below the tabs is a card for 'common-hdfs-sit': 'sit环境HDFS集群', '详情 节点', and a blue button '运行正常'. A green arrow points from a teal box labeled '集群列表' to the 'ZooKeeper集群' tab. In the top right corner, there are '帮助' (Help) and a user profile for '王志强'. A red box highlights the '创建集群' (Create Cluster) button.

解决1. 平台化、自动化

大数据管理平台：主机管理，集群管理自动化

大数据管理平台

帮助 王志强

集群管理 >

主机

客户端

账户

权限

资源 >

监控 >

资源展板 >

任务展板 >

OLAP >

HBase HA管理 >

HBase数据探查

集群展板 > common-yarn-sit详情

common-yarn-sit的基本信息

sit环境YARN集群

查看页面(使用window运维机访问) >>

配置信息

集群详细信息

查看详细配置 >>

Version hadoop-2.4.0.5

Zookeepers namenode1-sit.cnsuning.com,namenode2-sit.cnsuning.com,slave01-sit.cnsuning.com

依赖hdfs common-hdfs-sit

ResourceManagers namenode2-sit.cnsuning.com,namenode1-sit.cnsuning.com

NodeManager节点数 10

运行信息

查看节点 >>

运行正常 40% 34

解决1. 平台化、自动化

元数据管理：数据字典，权限申请审批实施自动化

元数据管理 MDM

搜索 all

14082455

概览

项目总表数 678

申请外部表记录数 1

表对外授权记录数 15

占用的存储量 395.34 GB

表信息概况

项目信息完整度分析

项目内热点排名top5

序名	表名	表中文名	表owner	热度
1	test11		徐伟14082455	7
2	test		徐伟14082455	1
3	table_comment		邢虎16050578	1
4	books_parquet		黄伟良14071882	1
5	test_td		郝雷15090284	1

申请中的外项目记录

序名	表名	所属项目	项目管理员	申请人	申请权限周期	申请时间	操作

解决1. 平台化、自动化

元数据管理：数据字典，权限申请审批实施自动化

The screenshot shows a web-based metadata management system. At the top, there's a header bar with the title "元数据管理 MDM" and a search input field containing "dpa_". On the right side of the header is a user profile icon and the number "14082455". Below the header, on the left, is a sidebar with navigation links: "概览", "数据表管理", "权限管理", and "系统管理". The main content area has tabs for "Hive(1000)" and "Kafka(3)", with "Hive(1000)" currently selected. Below the tabs are filters: "项目" set to "ETL系统(ETL)", "类目" (category) with an empty input field, and a search input field also containing "dpa_". There are "查询" (Search) and "重置" (Reset) buttons. The main list displays several data dictionary entries, each with a blue link:

- dpa_br_outer_click_d**
项目：ETL系统(ETL) 所属数据库：bi_dpa 负责人：彭虎(11080108) 最后更新时间：2014-10-08 15:29:39
描述：
类目：
- dpa_ord_bill_tf**
项目：ETL系统(ETL) 所属数据库：bi_dpa 负责人：彭虎(11080108) 最后更新时间：2017-07-18 11:38:39
描述：
类目：
- dpa_ord_bill_tf_d**
项目：ETL系统(ETL) 所属数据库：bi_dpa 负责人：彭虎(11080108) 最后更新时间：2017-05-09 20:40:51
描述：
类目：
- dpa_ord_bill_tf_spark**
项目：ETL系统(ETL) 所属数据库：bi_tmp 负责人：彭虎(11080108) 最后更新时间：2015-12-22 14:48:40
描述：
类目：
- dpa_ord_bill_tf_tmp**
项目：ETL系统(ETL) 所属数据库：bi_dpa 负责人：彭虎(11080108) 最后更新时间：2015-12-25 09:09:19
描述：

A large green rectangular overlay box is centered over the middle of the list, containing the text "数据字典，支持搜索" in white.

解决1. 平台化、自动化

元数据管理：数据字典，权限申请审批实施自动化

The screenshot shows the '元数据管理 MDM' (Metadata Management MDM) interface. On the left, there's a sidebar with navigation items: 概览 (Overview), 数据表管理 (Table Management), 权限管理 (Permission Management), and 系统管理 (System Management). The main area displays a table named 'dpa_ord_bill_tf_d' under the 'Hive' database. The table basic information panel includes fields like 表名 (Table Name: dpa_ord_bill_tf_d), 中文名 (Chinese Name: bi_dpa), 属于数据库 (Database: ETL系统(ETL)), and 项目管理员 (Project Manager: 彭虎 11080108). A red box highlights the '项目管理员' field. The '权限申请' (Permission Application) button at the bottom of this panel is highlighted with a large teal box. A modal dialog titled '权限申请' (Permission Application) is open in the center. It contains fields for '申请的表' (Table applied for: dpa_ord_bill_tf_d), '权限有效期' (Permission Validity Period: 永久 - Permanent), and '申请理由' (Reason for application: 请输入理由,不超过200字 - Please enter the reason, no more than 200 characters). At the bottom of the dialog are '确定' (Confirm) and '取消' (Cancel) buttons. The top right corner of the dialog has a close button (X). The status bar at the bottom of the page says 'Copyright © 2016 苏宁云商. All rights reserved.'

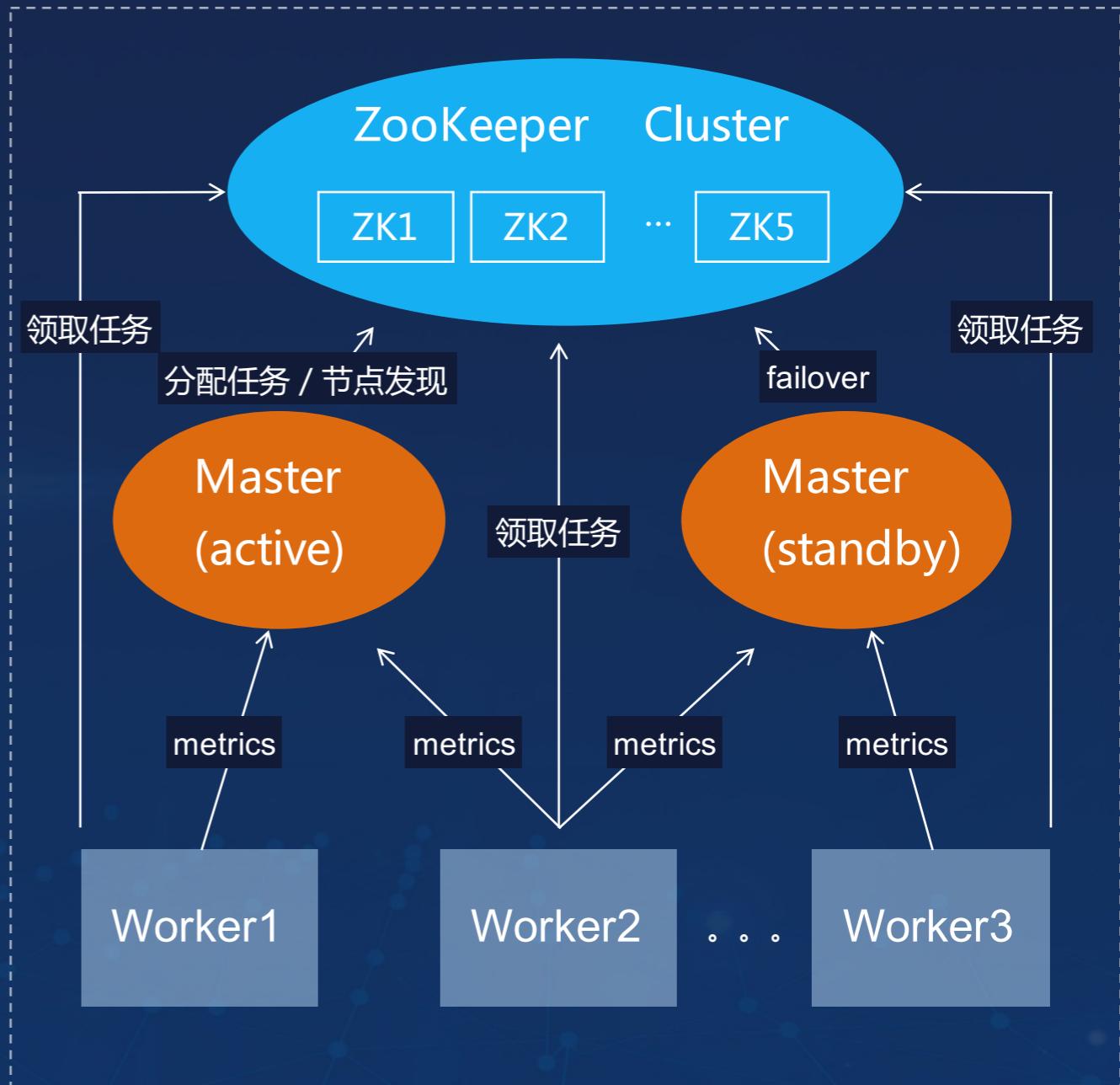
解决1. 平台化、自动化

数据流管理平台：集成Flume，智能扩缩容，插件式

2节点小集群
对接<1000qps业务

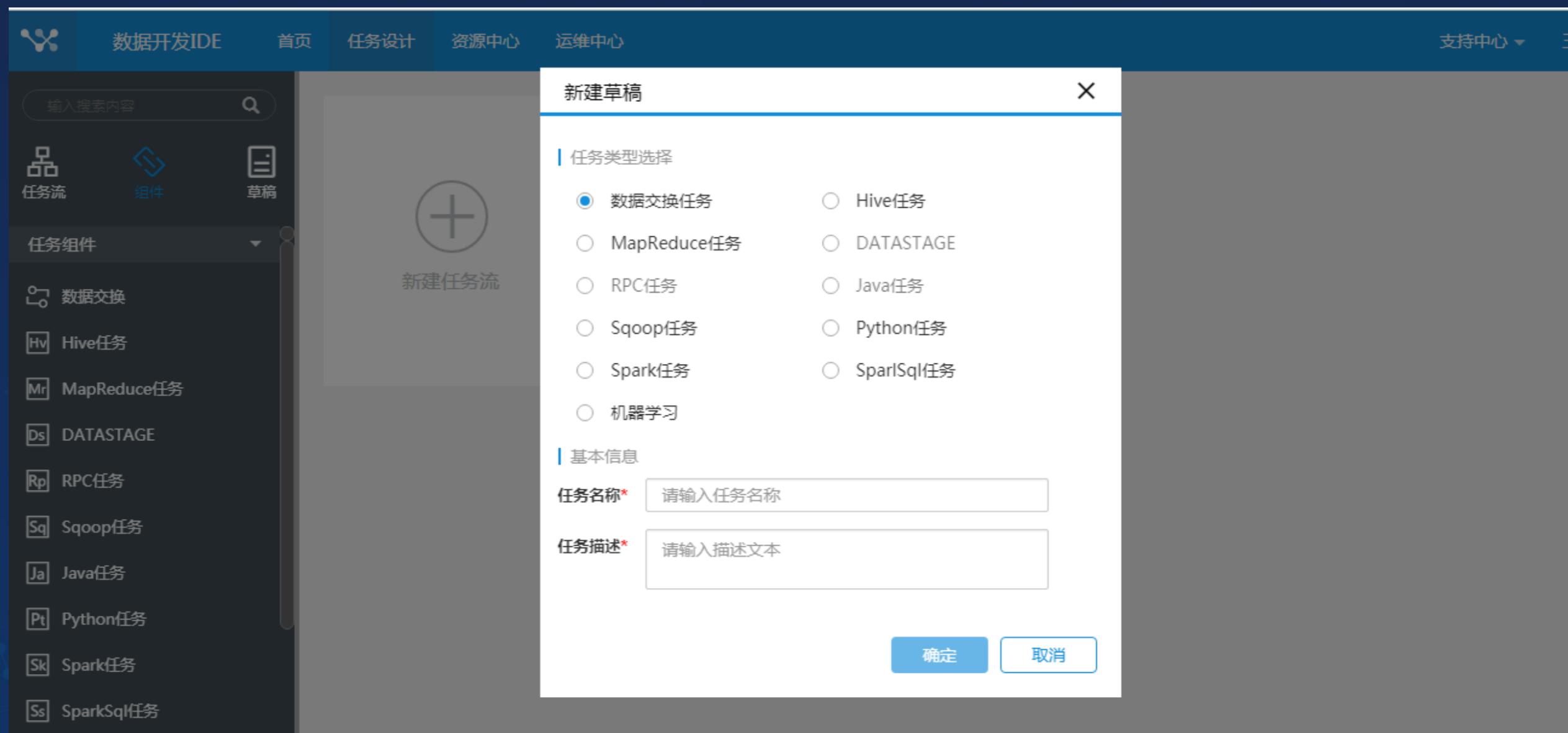
10节点中集群
对接1000-10000qps业务

30节点大集群
对接>10000qps业务



解决1. 平台化、自动化

数据开发平台：支持10种不同的任务类型，支持任务流/任务管理，解决复杂依赖问题，可扩展



解决1. 平台化、自动化

数据开发平台：支持10种不同的任务类型，支持任务流/任务管理，解决复杂依赖问题，可扩展

The screenshot shows the Data Development Platform's task configuration interface. The top navigation bar includes '首页' (Home), '任务设计' (Task Design), '资源中心' (Resource Center), '运维中心' (Operations Center), '支持中心' (Support Center), and '王志强' (Wang Ziqiang). The left sidebar lists recent tasks: '66666677', 'SQOOP_test1', 'sq_job_java', 'sangqiang_mr_con...', 'sangqiang_spark_c...', 'sanqiang_hive_co...', 'sangqiang_datax_c...', 'SPARKSQL_149568...', and 'SPARK_149568316...'. The main area displays a form for configuring a task named 'sangqiang_spark_config'. The form fields include:

- 任务名称: sangqiang_spark_config
- Jar名称: JavaOK.jar
- Main Class: com.suning.java.Test
- Args: --Xms200M
- spark版本: spark-1.5.2

Below the form is a section titled 'spark opts' containing a table with three rows of parameters:

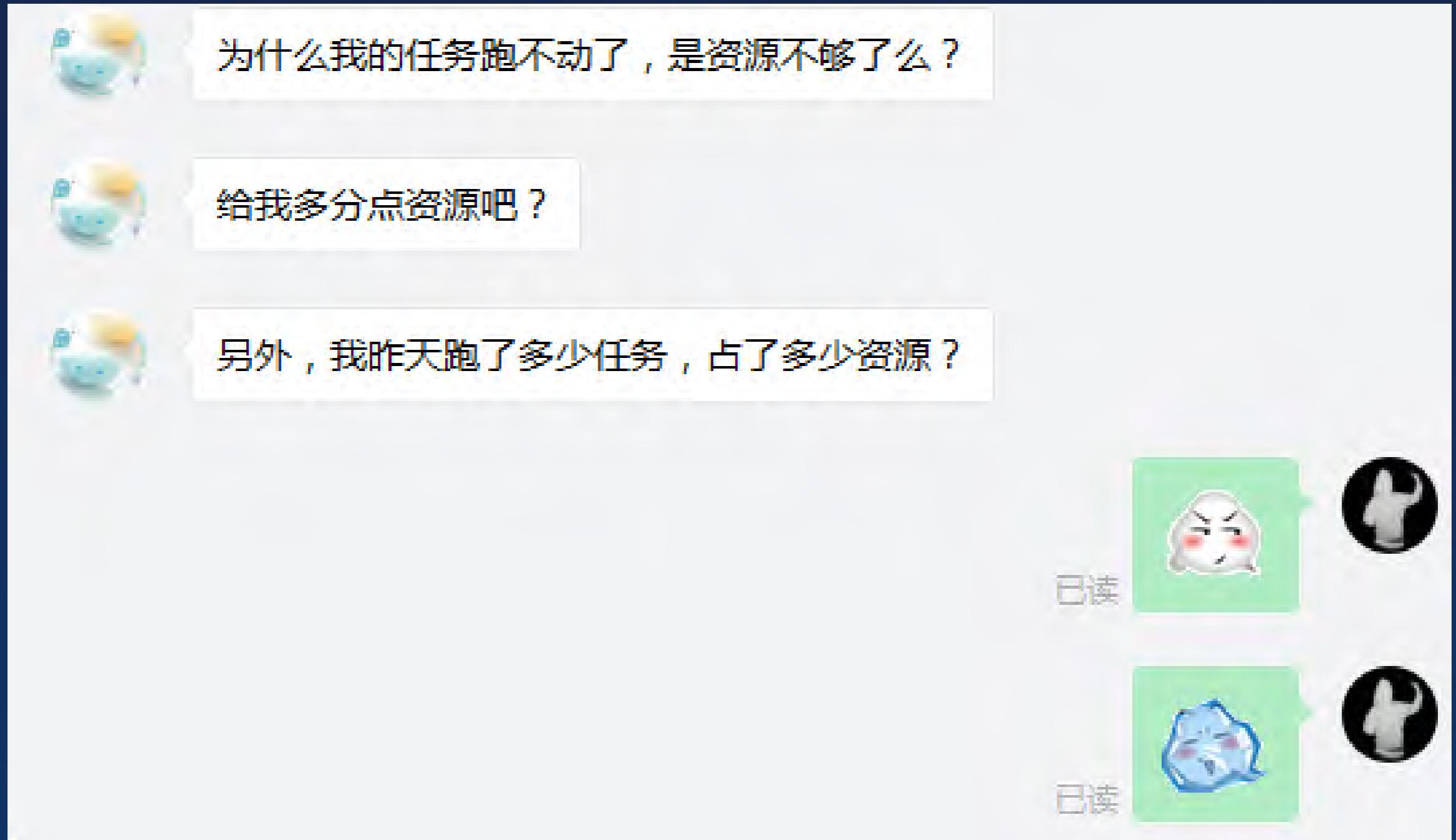
参数名	参数值
	3333
	222
	1111

解决1. 平台化、自动化

数据开发平台：支持10种不同的任务类型，支持任务流/任务管理，解决复杂依赖问题，可扩展

The screenshot displays the Data Development IDE interface. The top navigation bar includes '首页' (Home), '任务设计' (Task Design), '资源中心' (Resource Center), '运维中心' (Operations Center), '支持中心' (Support Center), and '王志强' (Wang Ziqiang). The left sidebar features icons for '任务流' (Task Flow), '组件' (Components), and '草稿' (Drafts), along with a list of saved tasks: '9_onlyonce_6...', 'sync_bi_dpa.t...', 'ISE_HBASE_O...', 'HP_MSA_H', 'test_liuyh', '智能客服离线...', '111', 'HP_MSA_D', 'HP_CCBS_INIT', and '小时任务测试'. A search bar at the top is followed by a toolbar with various icons. The main workspace shows a complex task flow graph with numerous nodes and dependencies. A legend at the bottom identifies node types: '日志查看' (Log View) and '任务流面板' (Task Flow Panel). A blue button at the bottom left says '+ 新建任务流' (New Task Flow).

痛点2. 无资源使用视图



解决2. 资源可视化、人民币化

- ✓ 存储/计算资源计量计费
- ✓ 资源池使用可视化
- ✓ 任务展板

解决2. 资源可视化、人民币化

存储/计算资源计量计费

大数据管理平台 帮助

账户: sousuo 昨天 7天 30天

存储容量统计
总容量 **684.77GB**

过去1日新增 -288.62GB -29.65% ↓
总存储成本 4.01元

计算资源统计
总耗时 **11 ch**

过去1日新增 0ch -0.98% ↑
总计算成本 4.24元

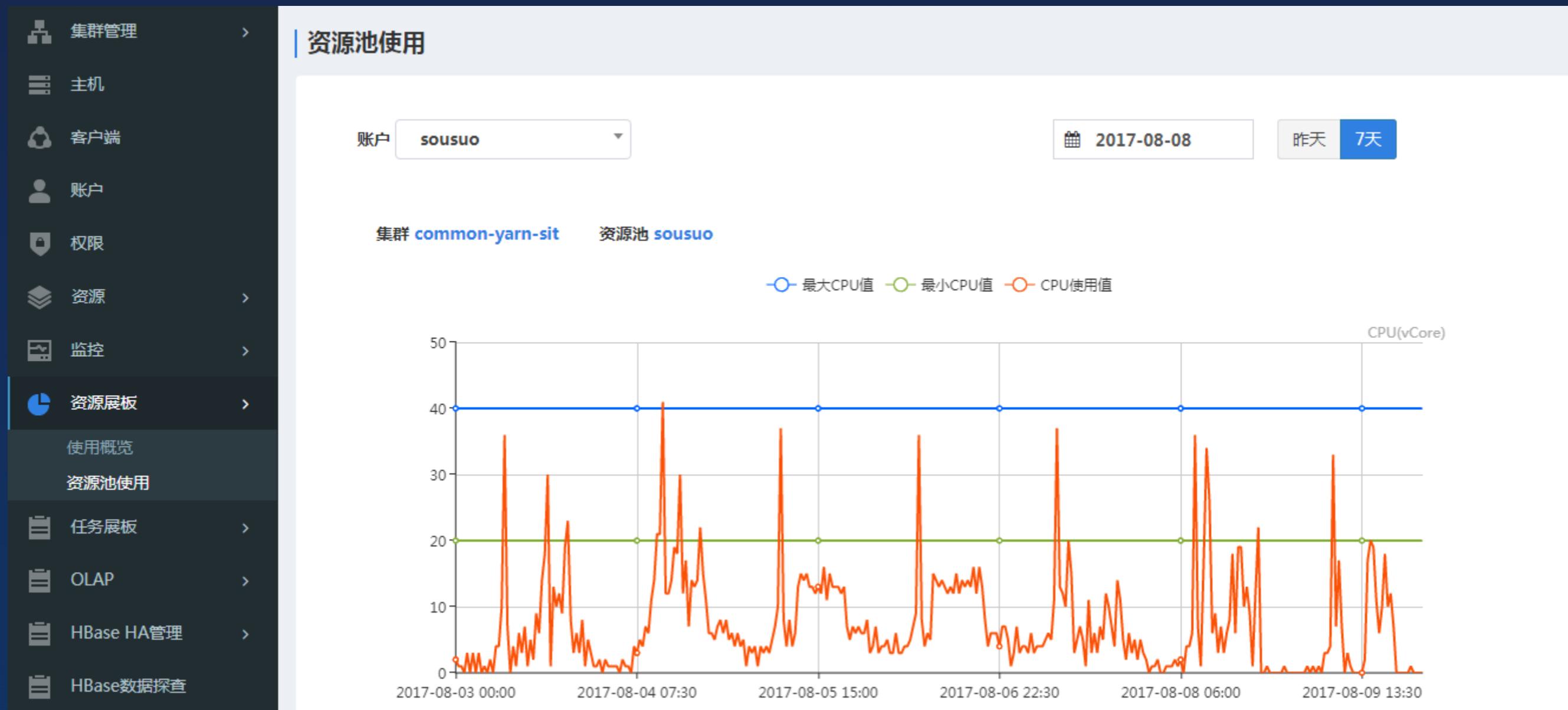
计费方法 总成本 **8.25元**

过去30天存储容量使用折线图

Geekbang & InfoQ

解决2. 资源可视化、人民币化

资源使用可视化



解决2. 资源可视化、人民币化

任务展板

大数据管理平台

帮助 王志强

集群管理 >

主机

客户端

账户

权限

资源 >

监控 >

资源展板 >

任务展板 >

排行榜

任务概览

任务详情

任务概览

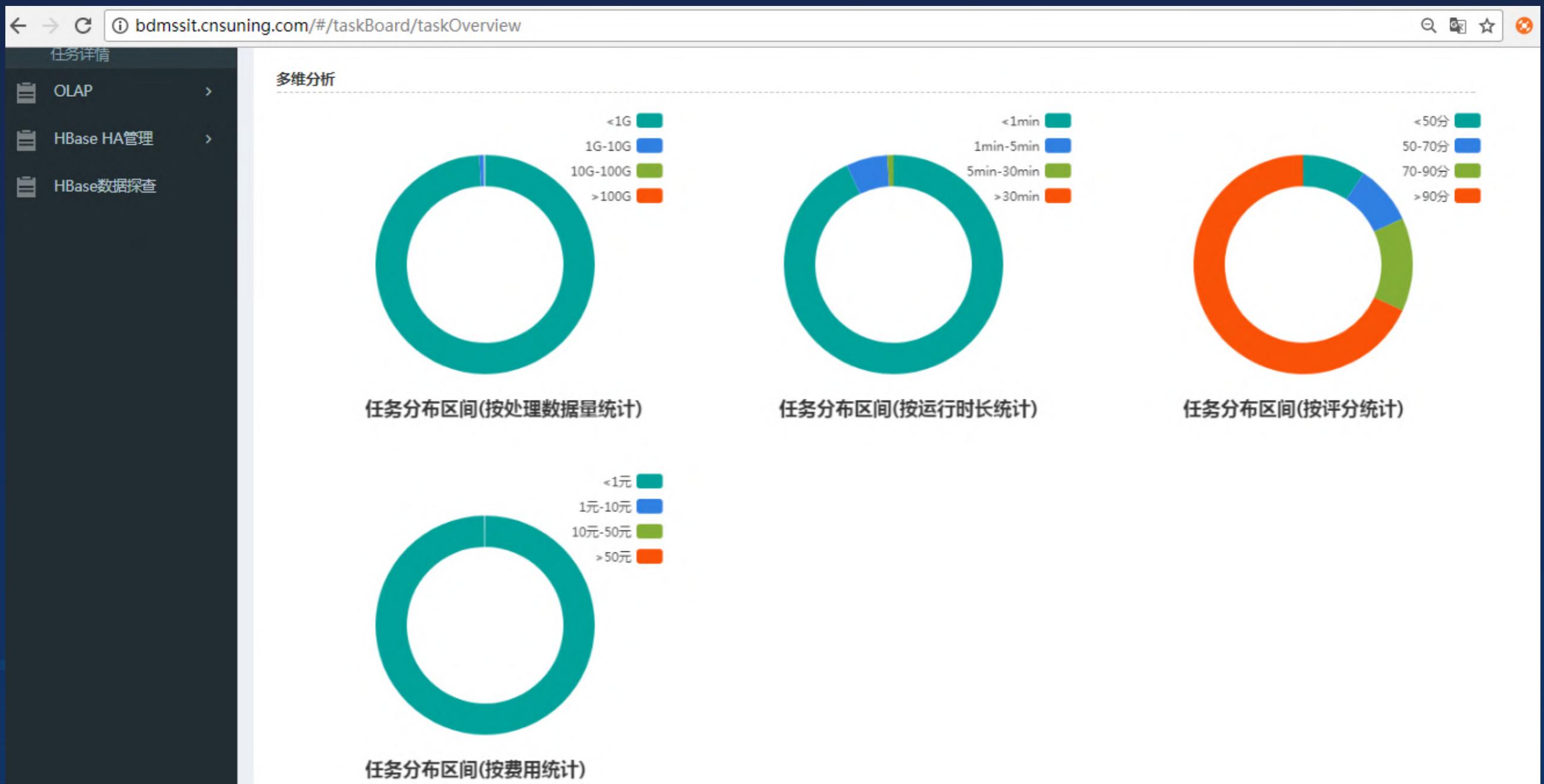
账户 All 集群 common-yarn-sit 组件 All 日期 2017-08-09

总任务数: 10753 成功任务数: 10747 失败任务数: 1 主动杀死任务数: 5

总耗费	处理数据总量	平均处理数据量	平均运行时间	MR/SPARK任务占比
222元	1.69TB	165MB	31s	10753 / 0

解决2. 资源可视化、人民币化

任务展板



解决2. 资源可视化、人民币化

任务展板

大数据管理平台

帮助 王志强

集群管理 >

主机

客户端

账户

权限

资源 >

监控 >

资源展板 >

任务展板 > **排行榜**

排行榜

组件 All

2017-08-09

任务总量Top10

排名	账户	任务数
1	csi	3347
2	bigdata	2232
3	erp	1274
4	lbi	1086
5	bi	779
6	srdss	361
7	scpdm	240
8	smbm	205
9	sopdm	191
10	pos	155

处理数据量Top10

排名	账户	处理数据量
1	spider	706.78GB
2	srdss	271.88GB
3	ztbd	210.14GB
4	bigdata	165.41GB
5	pcms	148.15GB
6	mobdss	118.66GB
7	sousuo	29.57GB
8	pos	27.91GB
9	scpdm	15.15GB
10	lbi	12.13GB

任务评分Top10 (由高到低)

排名	账户	评分
1	lbi	100

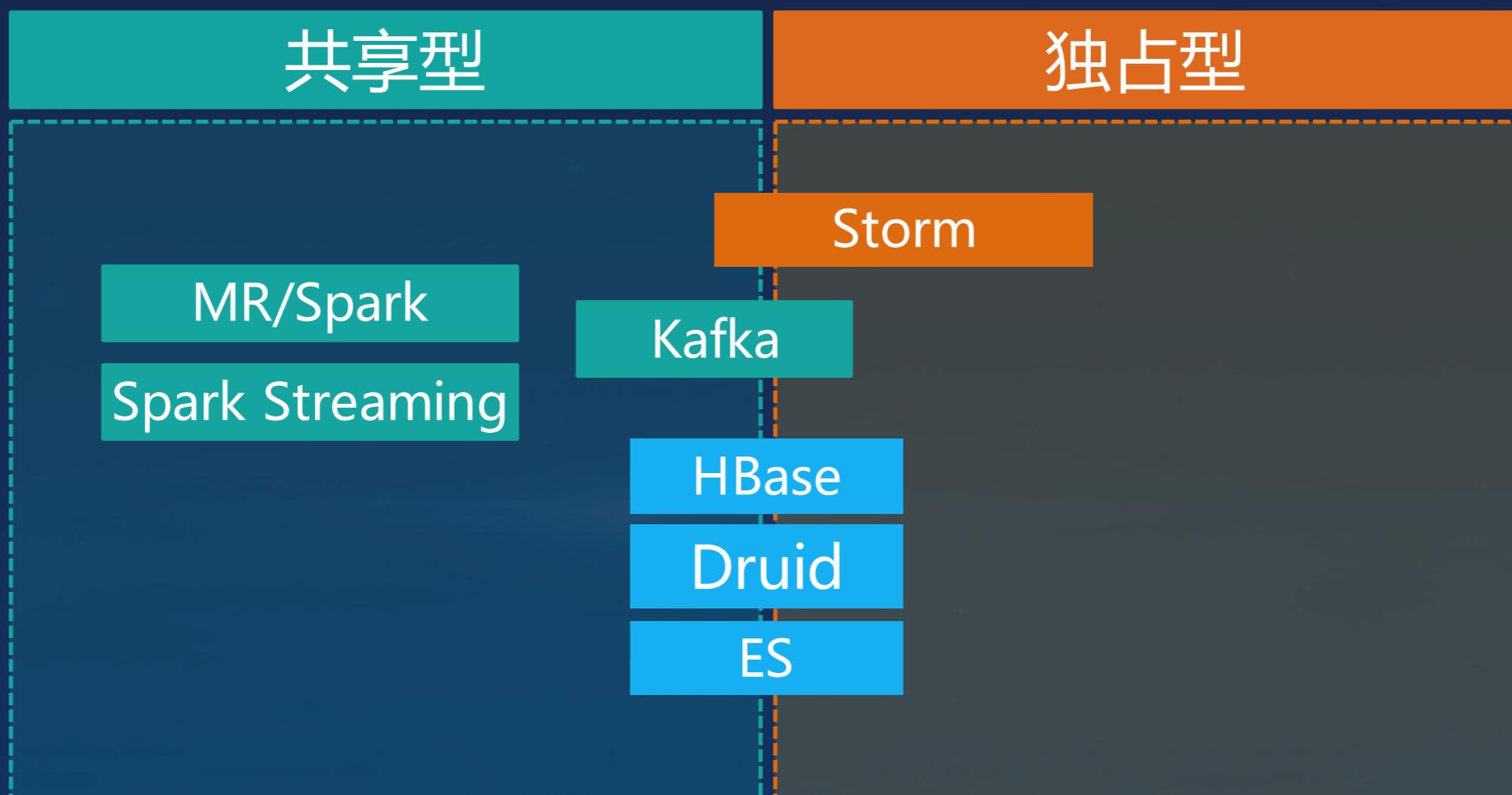
任务评分Top10 (由低到高)

排名	账户	评分
1	lbi	49

痛点3. 任务互相影响，资源隔离性差

我的任务为什么突然变慢了？
是不是别人影响我了？

解决3. 差异化服务、物理隔离



资源浪费：离线、流本来就是错峰的

解决3. 差异化服务、物理隔离

共享型-YARN

FairScheduler：每个账户一个Pool，配置最大最小资源；最小资源是能保证的资源，最大资源是上限；

优先级: backport from 2.8；改进为7个优先级，每提升1个优先级，获得分配权重增加1倍；

降低长GC的影响：尽快失败，合理设置 GCTimeLimit=90 GCHeapFree Limit=10 ParallelGCThreads=8；

Cgroup

开启抢占

解决3. 差异化服务、物理隔离

共享型-Kafka



避免大集群：对ZK的压力，对controller的压力

逻辑分区：按照不同的业务场景分配，避免不同保障级别业务的相互影响

解决3. 差异化服务、物理隔离

独占型-Storm



独立的物理区域

物理CPU与虚机CPU 1:1 亲和性绑定

CPU ST从40%+降低至1%

痛点4. 排查问题耗时长，应用优化门槛高

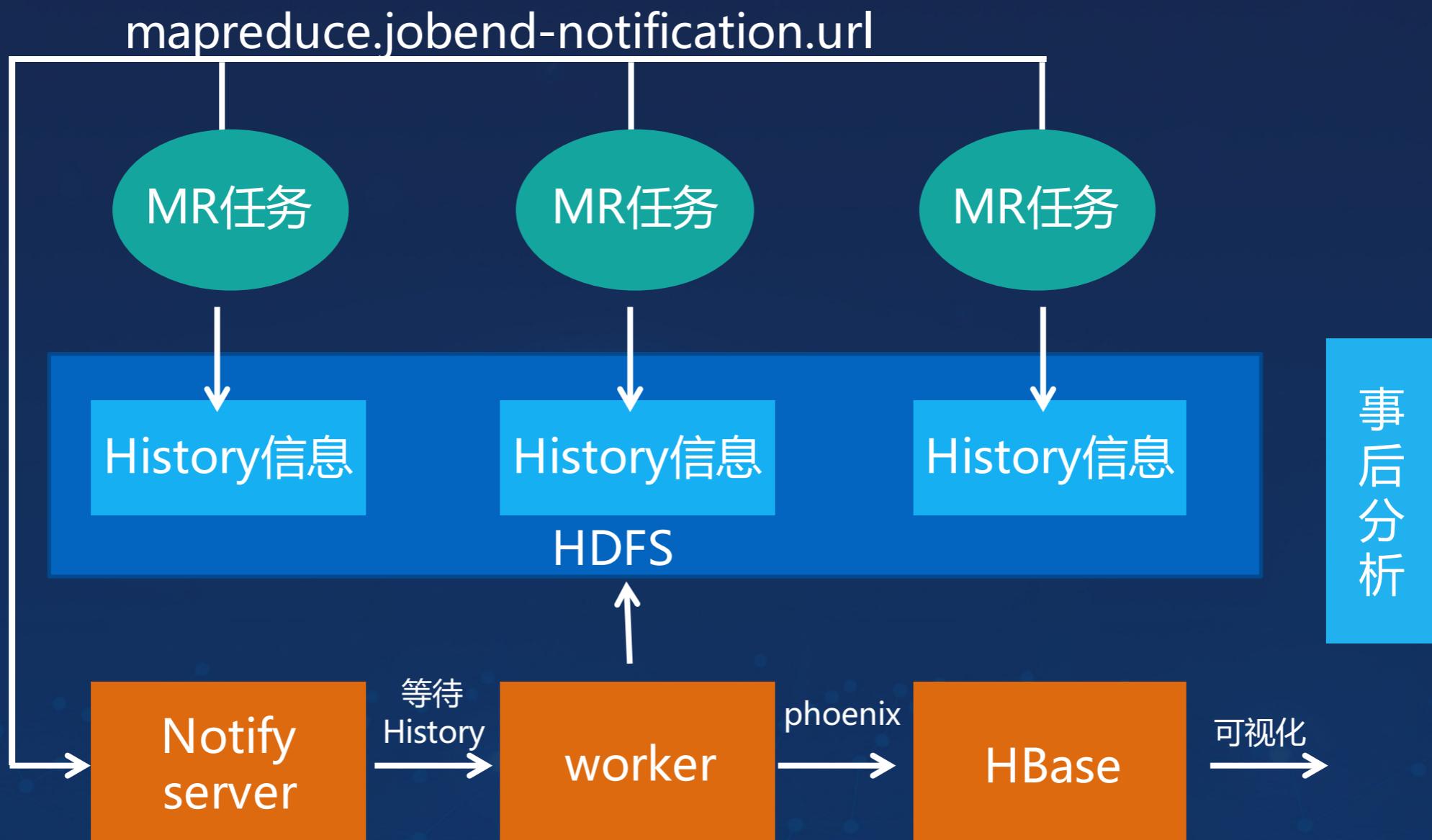


用户不清楚任务为什么失败

用户不清楚该如何优化任务

解决4. 智能诊断、优化建议

离线计算平台-MR任务分析平台



解决4. 智能诊断、优化建议

离线计算平台-MR任务分析平台

任务排行榜			
评分最低Top20			
排名	账户	jobID	评分
1	bi	job_1482378778761_19995963	44
2	aps	job_1482378778761_19986972	44
3	cloudytrace	job_1482378778761_19882779	44
4	aps	job_1482378778761_20003396	44
5	serrmsg	job_1482378778761_20040120	44
6	serrmsg	job_1482378778761_19987516	44
7	serrmsg	job_1482378778761_19987516	44
8	lbi	job_1482378778761_19888904	44
9	lbi	job_1482378778761_19888904	44
10	birec	job_1482378778761_20032814	44

处理数据量最多Top20			
排名	账户	jobID	处理数据量
1	sousuo	job_1482378778761_20032979	2.57TB
2	sousuo	job_1482378778761_20039489	2.57TB
3	srdss	job_1482378778761_20002279	2.57TB
4	spider	job_1482378778761_19879587	2.44TB
5	spider	job_1482378778761_19875138	2.39TB
6	safe	job_1482378778761_19953908	1.99TB
7		job_1482378778761_19950502	1.99TB
8		job_1482378778761_19961139	1.99TB
9	srdss	job_1482378778761_19996728	1.46TB
10	osmos	job_1482378778761_19879911	1.41TB

不同维度的排行榜

总CPU时间最长Top20			
排名	账户	jobID	耗时
1	bi	job_1482378778761_19896644	3442184
2	bi	job_1482378778761_19903213	2259819
3	erp	job_1482378778761_20004500	2040751
4	erp	job_1482378778761_20010608	1533820

Task数排名Top20			
排名	账户	jobID	Task数
1	sopdm	job_1482378778761_19987327	22908
2	aps	job_1482378778761_19979424	20925
3	aps	job_1482378778761_19977850	20391
4	ztbd	job_1482378778761_19916949	17344

解决4. 智能诊断、优化建议

离线计算平台-MR任务分析平台

任务详情 > job_1482378778761_19896644 输入jobID 🔍

基本信息

账户名称	bi
集群名称	common_yarn_prd3
任务名称	tsor_ous_vbeln_plan_unfnsh_d(20170408)
任务状态	SUCCEEDED
提交时间	2017-04-09 02:39:25
结束时间	2017-04-09 03:55:21
总CPU时间	3442184s
处理数据量	216.47GB
map/reduce	504 / 775

任务统计

map平均执行时间	132s
reduce平均执行时间	21s
平均map处理数据量	455.17MB
map输出数据量	352.30GB
map输出条数	8,625,703,184
map spill条数	17,266,181,015
reduce spill条数	8,625,703,184
最长map执行时间	4456s
最长reduce执行时间	4394s

评分详情 (评分方法)

平均map处理时间得分	20
平均map处理数据量得分	20
平均reduce处理数据量得分	20
spill数据量得分	20
长尾任务得分	16

96分

任务运行信息

任务评分依据

The screenshot displays a web-based interface for monitoring and analyzing MapReduce tasks. At the top, there's a header with '任务详情' and a search bar for 'jobID'. Below the header, there are three main sections: '基本信息' (Basic Information), '任务统计' (Task Statistics), and '评分详情 (评分方法)' (Scoring Details). The '基本信息' section shows various task parameters like account name, cluster name, and execution times. The '任务统计' section provides detailed metrics for map and reduce operations, including execution times, data processed, and spill counts. The '评分详情' section lists the scoring criteria and their corresponding scores. Two green arrows originate from teal-colored callout boxes labeled '任务运行信息' and '任务评分依据', both pointing to specific data tables within the interface.

解决4. 智能诊断、优化建议

离线计算平台-MR任务分析平台

| 任务详情 > job_1499408566046_8795319

基本信息

账户名称	bjredbaby
集群名称	common_yarn_prd3
任务名称	create table tmp_zq_zs_order_06 stored ...t(Stage-1)
任务状态	FAILED
提交时间	2017-08-10 17:14:18
结束时间	2017-08-10 17:15:37
总CPU时间	1.25min
处理数据量	0MB
map/reduce	1 / 4

失败

失败原因

失败原因 数字格式化异常
建议解决方案 ["检查数字格式化"]

失败诊断信息

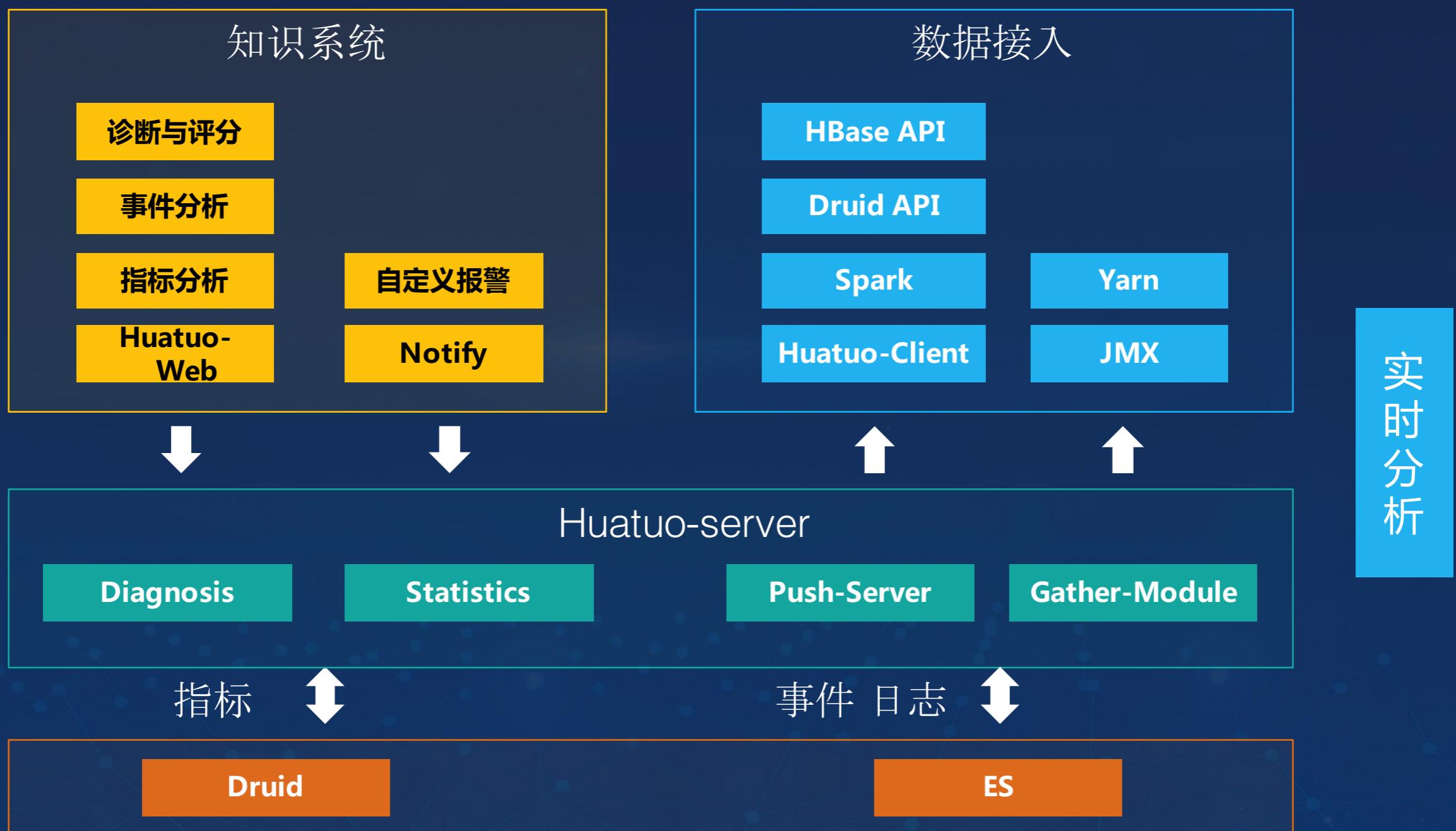
←

异常日志信息

```
Log Type: stderr Log Length: 243 log4j:WARN No appenders could be found for logger (org.apache.hadoop.metrics2.impl.MetricsSystemImpl). log4j:WARN
```

解决4. 智能诊断、优化建议

离线计算平台-华佗智能诊断平台(spark)



解决4. 智能诊断、优化建议

离线计算平台-华佗智能诊断平台(spark)

资源角度	性能角度	故障角度
<ul style="list-style-type: none">•宿主机器状态分析•HDFS资源使用分析	<ul style="list-style-type: none">•Task耗时链分析•长尾Task分析	<ul style="list-style-type: none">•Shuffle数据倾斜•HDFS Commit阻塞
<ul style="list-style-type: none">•Driver/Executor进程状态分析•资源利用率分析	<ul style="list-style-type: none">•任务调度Overhead分析•Reduce并发度分析•JDBC并发度分析Kafka读并发度分析	<ul style="list-style-type: none">•Executor丢失•Spill事件分析•高维Parquet写性能诊断•RDD Size Estimator耗时诊断•任务事件流
<ul style="list-style-type: none">•Cache 利用率分析•Shuffle内存利用率分析		

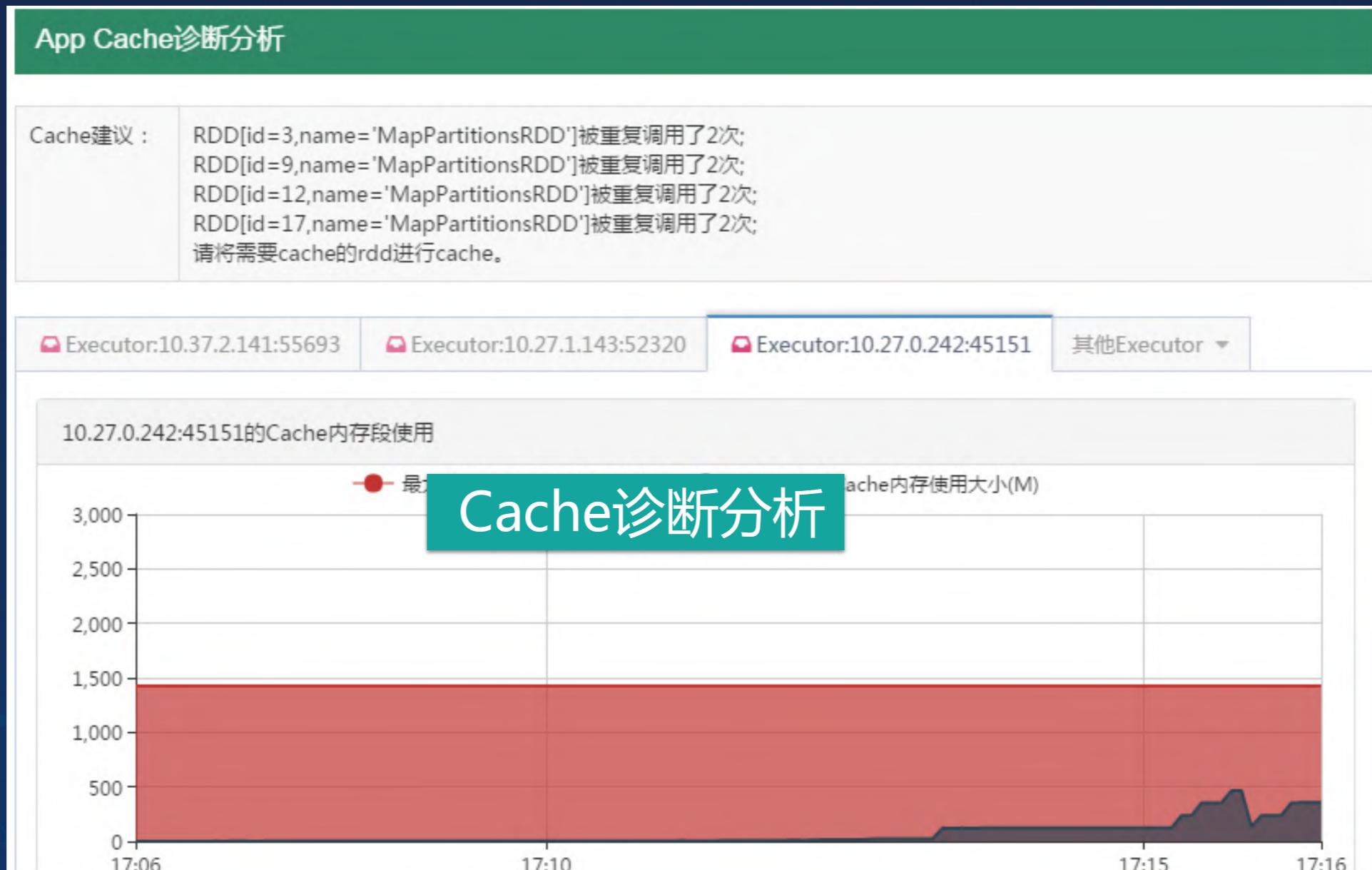
解决4. 智能诊断、优化建议

离线计算平台-华佗智能诊断平台(spark)



解决4. 智能诊断、优化建议

离线计算平台-华佗智能诊断平台(spark)



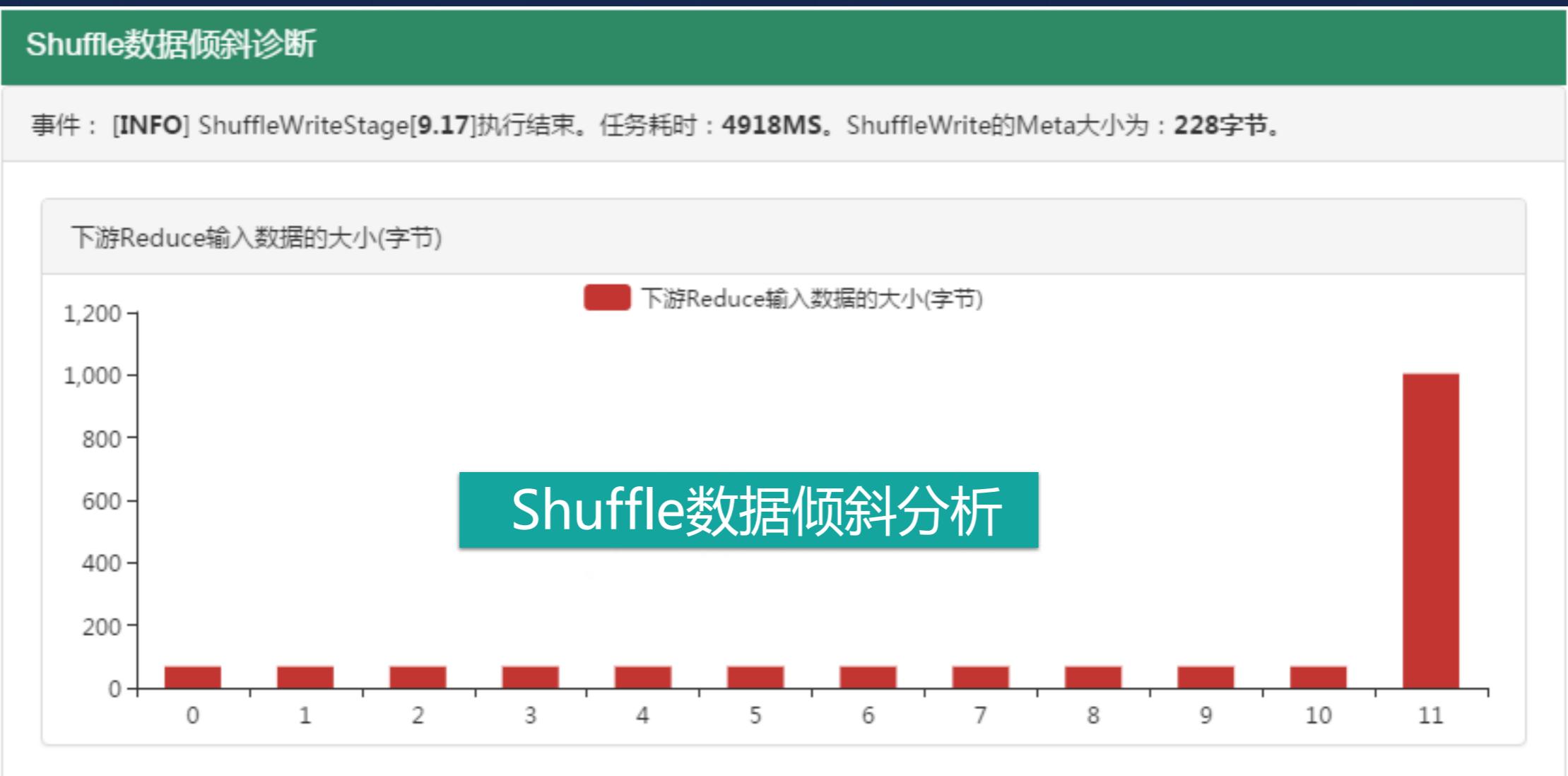
解决4. 智能诊断、优化建议

离线计算平台-华佗智能诊断平台(spark)



解决4. 智能诊断、优化建议

离线计算平台-华佗智能诊断平台(spark)



解决4. 智能诊断、优化建议

流计算平台-健康状态监测系统

进程重启

代码异常

OutOfMemory

宿主机/虚机宕机

Kafka消息堆积

数据量暴增

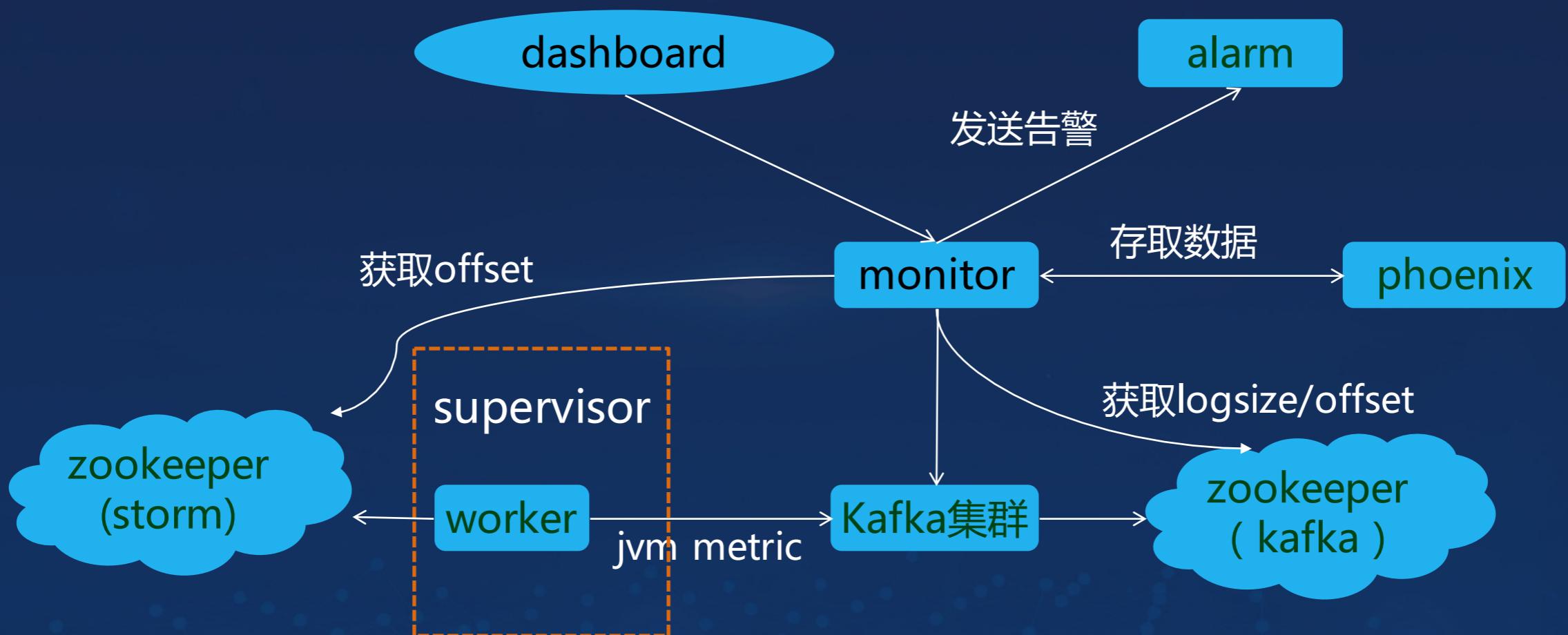
FGC/Capacity

并发不够

任务异常停止

解决4. 智能诊断、优化建议

流计算平台-健康状态监测系统



解决4. 智能诊断、优化建议

流计算平台-健康状态监测系统

ppcsHomePageB健康状态

2017年08月10日15时04分~2017年08月10日22时59分

1、Topology基本信息

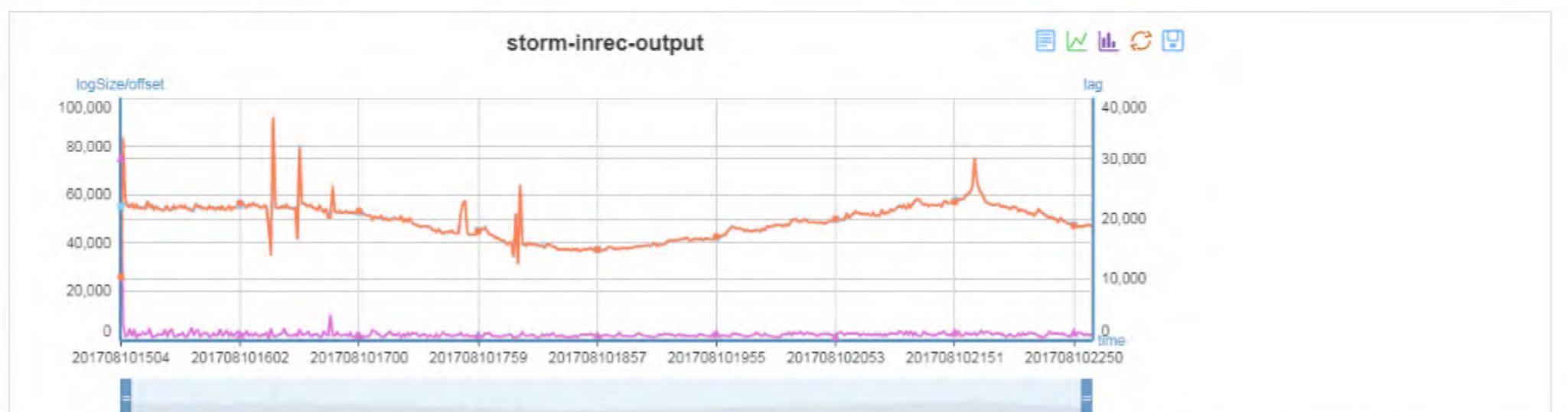
启动时长	Num workers	Num executors	Num tasks	spout.pending	message.timeout(s)	集群ID
7h54m52s	140	449	449	1000	360	ppcs-v-storm

2、Topology配置诊断

- topology的bolt并发设置合理，无需调整。此次启动也无worker发生迁移

3、数据消费视图（2017年08月10日15时04分~2017年08月10日22时59分）

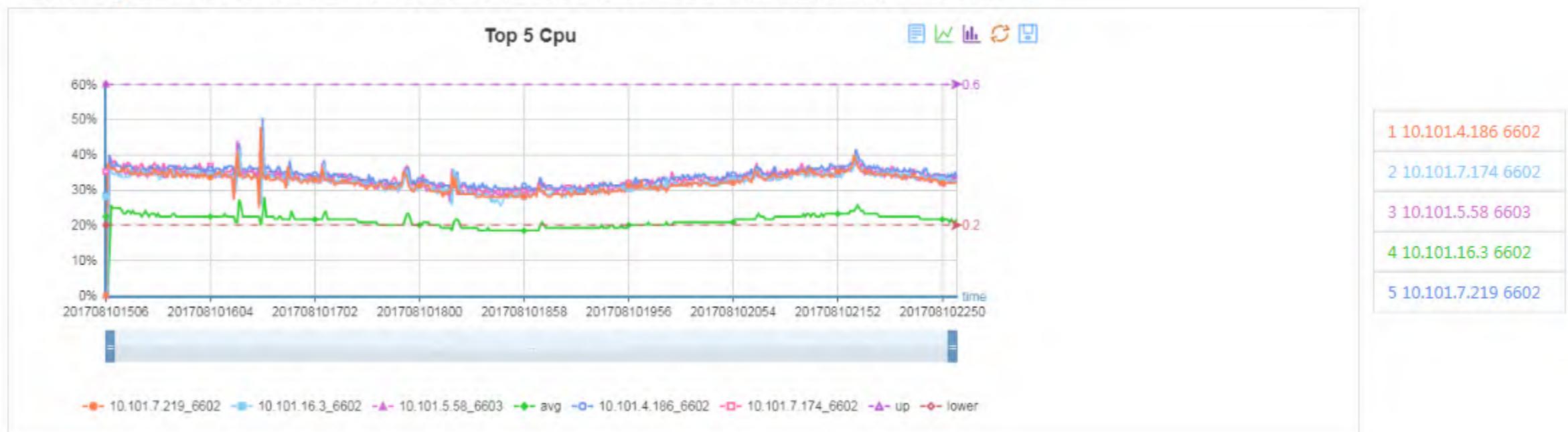
topic名 : storm-inrec-output ▾



解决4. 智能诊断、优化建议

流计算平台-健康状态监测系统

4. Topology运行CPU (2017年08月10日15时04分 ~ 2017年08月10日22时59分)

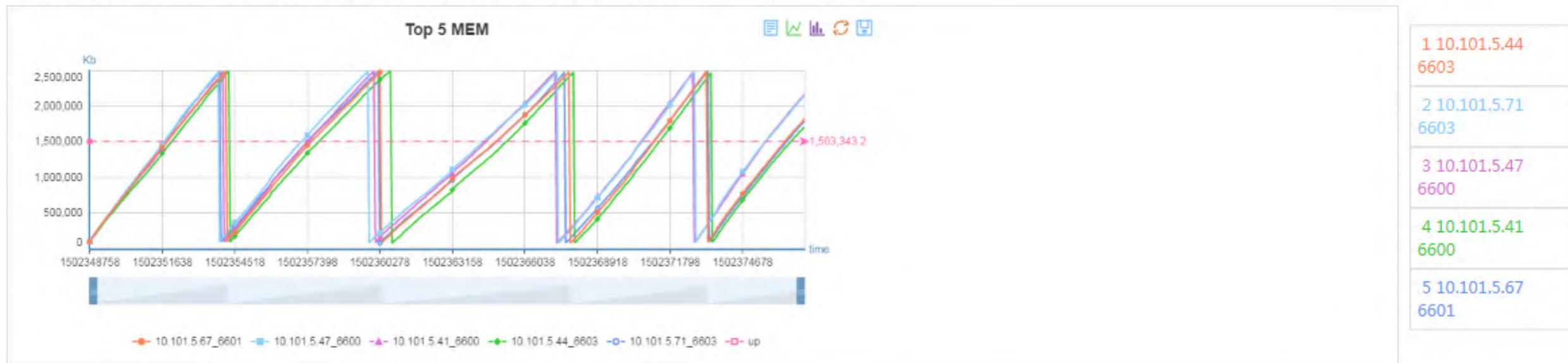


- worker:10.101.7.219_6602的cpu使用率远远高于平均使用率。
- worker:10.101.5.58_6603的cpu使用率远远高于平均使用率。
- worker:10.101.16.3_6602的cpu使用率远远高于平均使用率。
- worker:10.101.4.186_6602的cpu使用率远远高于平均使用率。
- worker:10.101.7.174_6602的cpu使用率远远高于平均使用率。

解决4. 智能诊断、优化建议

流计算平台-健康状态监测系统

5、Topology运行MEM (2017年8月10日15时4分 ~ 2017年8月10日22时59分)



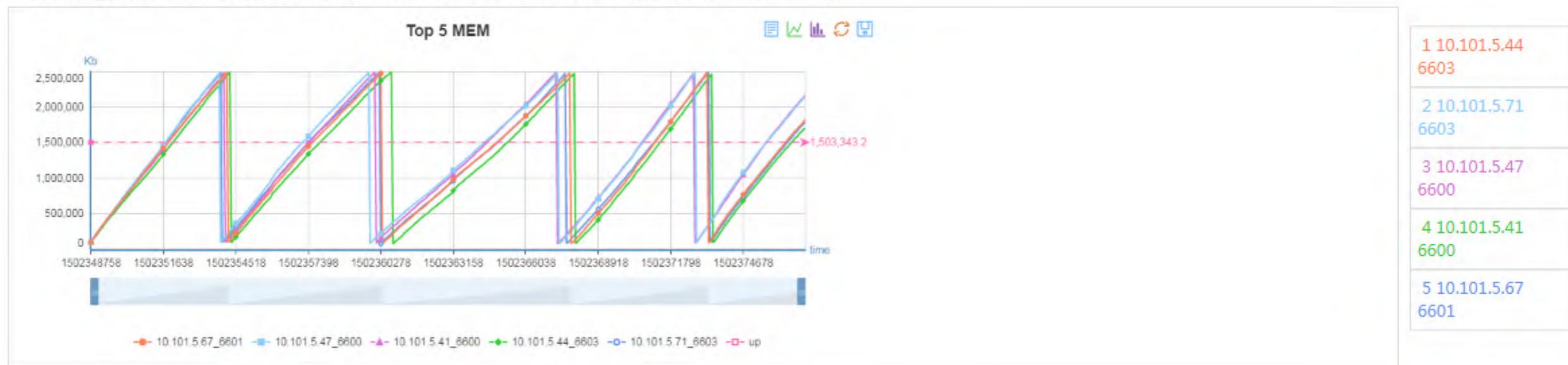
6、Topology运行Capacity (2017年08月10日15时09分 ~ 2017年08月10日22时59分)



解决4. 智能诊断、优化建议

流计算平台-健康状态监测系统

5、Topology运行MEM (2017年8月10日15时4分 ~ 2017年8月10日22时59分)



6、Topology运行Capacity (2017年08月10日15时09分 ~ 2017年08月10日22时59分)



TABLE OF CONTENTS

苏宁大数据平台基本介绍

大数据平台运维的痛点及解决方案

平台优化及增强

DOING & TO DO

平台优化及增强-稳定

- Hive metaserver 连接数过高的问题

修改bonecp的配置 : maxConnectionsPerPartition=1

- Spark Streaming & Druid System CPU过高的问题

设置vm.zone_reclaim_mode=0

- 透明大页导致System CPU过高的问题

echo never >/sys/kernel/mm/transparent_hugepage/defrag

平台优化及增强-安全

- 账户/权限体系：每个系统一个账户，不允许跨账户写。
- Hive metaserver 密码加密

```
<property>
  <name>javax.jdo.option.ConnectionPassword</name>
  <value>OL2vkw+gDtQ=</value>
</property>
```
- 基于User/IP的访问控制策略：RPC层面控制，白名单
- skipTrash禁用：防止误删数据

平台优化及增强-扩展性

结合HDFS的瓶颈问题逐步优化

- 程序优化，扫全表：Hive慎用unix_timestamp方法
- 小文件合并
- YARN日志降低副本至1
- YARN日志单独放在另一个集群
- Federation + Alluxio 实现统一命名空间

TABLE OF CONTENTS

苏宁大数据平台基本介绍

大数据平台运维的痛点及解决方案

平台优化及增强

DOING & TO DO

DOING & TO DO

- ✓ Flink推广
- ✓ OLAP平台建设
- ✓ 流计算消息回溯
- ✓ 多活&灾备
- ✓ 资源统一管理（明年）

你认为以下哪些关键字对运维最重要？



个人号



公众号



THANKS!

智 能 时 代 的 新 运 维

CNUTCon 2017