# 平台功能文档&使用手册

# 为什么选择Data平台

66

- 1. 从数据生产到数据展现的完整服务
- 2. 生成报表只需写hql,不需要写代码
- 3. 原始数据的统一管理
- 4. 灵活的调度
- 5. 方便的维度上卷和下钻
- 6. 多样而又易于配置的图表类型
- 7. 统一的权限管理

## 概览

用户通过 hive sql 的方式做数据预处理,把所有的数据都划分成维度和指标两大类,借助 grouping sets 语法在一个hql下生成多个维度组合下的数据,同时在写入数据时系统把一个项目下同一维度组合下的所有数据关联在一起以此达到数据的快捷查询和分析。

在数据可视化方面,我们提供了图、表、时间框等可视化元素,用户通过可视化点选或者自定义的方式 生成一个报表需要的图表和表格等元素,以此来实现数据的展现。

## 使用入门

作为数据分析师和商业运营人员,data数据分析平台对他们来说是查看、分析数据和提高工作效率的有力武器,它能及时收集记录分析数据信息。

### 第一步:添加项目

- 1. 选择类型、设置项目名称、点击保存项目配置。
- 2. 添加分类,输入中文名和英文名,点击确定保存.
- 3. 添加hql, 经过hql解析

### 第二步:添加报表

- 1. 报表基本信息设置:选择项目、输入报表名称等设置点击确定保存。
- 2. 时间条件设置:选择时间区类型、时间粒度、时间偏移量等设置,点击确定保存。
- 3. 图标区域设置: 图表类型、标题、选取数据源设置, 点击确定保存。
- 4. 表格区域设置:表格类型、表格名称、选择维度指标数据配置,点击确定保存。
- 5. 点击保存配置,完成新的报表配置

### 第三步:添加菜单

1. 选择一级菜单名称、报表, 点击保存后

## 第四步: 查看报表

1. 进入分析平台点开你设置的左侧菜单,就可以展现出你的报表。

# 报表

## 介绍

用户通过图、表、时间工具的可视化配置,展现多图多表格的不同种类型的报表,同时提供用户下载、 收藏、订阅邮件的功能,方便用户随时查看、分析、统计数据,赋予了用户对数据进行分析挖掘的能力。

## 时间选择工具条

用户可按小时、天、月、选择获取相应的数据,同时支持下载、收藏、查看报表创建人。

时间选择框的设置分如下几步:

- 1. 时间区类型:分为区间和单天。
- 2. 默认时间间隔: 开始时间和结束时间相差的天数。
- 3. 默认结束时间偏移量: 结束时间默认当天的前几天,同时支持用户自定义设置偏移量的数据。
- 4. 快捷时间设置: 前天、昨天、最近7天、最近30天。

## 图表

以维度和指标为坐标系构建出曲线图、柱状图、饼图、行政图、漏斗图、曲线图、行政区划图,同时支持任意维度的堆积图、形成多图表混合展现的可视化图表。

#### 支持的图表类型

- 1. 饼图
- 2. 柱状图
- 3. 行政区划图
- 4. 漏斗图
- 5. 曲线图
- 6. 堆积图

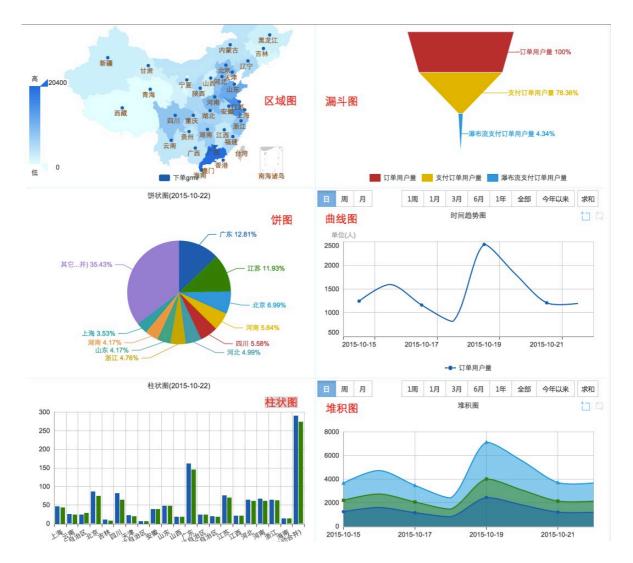
### 特点

- 1. 图表设置有数据事件关联,同时可以排序,设置后事件会在图表相应的数据节点标注显示。
- 2. 图表可编辑、可拖拽、可排序、可下载截图、可邮件订阅。
- 3. 图表带有月、周、3月、6月、1年等快捷按钮对数据进行统计计算。
- 4. 丰富的图表元素。

### 新建/编辑

图表区域设置											×
图表标题*:	秒杀产	딞									
图表类型*:	饼图										
数据处理	Top1	.0	٧	自定义数	字						
图表显示设置*:	一行	两个	٧								
数据源*:	选择	数据源									
	类型	列显示名称	列key	列说	明	数据过滤	3	计算值		隐藏选择 ■全选	图例名称 (隐藏选择)
	维度	一级类目	复制	cata	alog19		¥	-			
	指标	拍下订单数	复制	Ð			▼	-	1	<b>D</b> -	0-
	表林	格颜色渐变   □									
	<b>着加</b>	行									确定取消

## 效果



## 表格

表格的组件可配置主、副表,以及特定的表格类型,支持不同维度、指标、自定义指标数据的过滤、计算、百分比、即时过滤、搜索,排序、添加外链等多样化的设置,同时可以自定义报表实现跨项目维度 和指标的统计、加工、整合,以及数据的添加查看对比趋势的功能。

### 多表格类型

表格类型	表格描述
主表	展示的是概览数据
副表	展示的是细节数据,比如天级,小时级,周级细节数据
普通表格	从数据源选取所需的指标和维度,展示相应的数据,是最常用的表格类型
对比表格	选择指定的指标展示出该指标的当日值、前一日值、当日相比前一日变化率、 上周同日值、当日相比上周同日变化率、周平均值,方便进行对比分析
聚合表格	相比较普通报表多了一个该指标数值可以求最大值、求最小值、求和、求平均值、 计数的功能
自定义表格	用户自定义sql,通过 <mark>\$TABLE(项目英文名,维度组合)</mark> 函数来确定维度、 指标的存储表,同时可以配置跨项目的报表,支持直接查询Mysql业务库。

## sql限制条件(自定义表格)

- 1. 支持配置data平台现有项目的数据和直接查询业务库mysql数据
- 2. 查询data平台现有项目的数据表名用函数 \$TABLE(项目英文名,维度组合) 的形式替换
- 3. 查询业务库mysql数据表名用函数 \$DB(库名.表名) 的形式替换
- 4. 项目管理的配置查询功能里获取sql的指标组合,项目英文名 维度组合且指标组合请使用sql字段组合里的名称
- 5. 表中的时间列统一用 cdate 字段
- 6. sql必须添加开始时间和结束时间条件限制,开始时间用 \$START(0),结束时间 \$END(0),即:每个sql必须加上 where cdate>='\$START(0)' and cdate<='\$END(0)'

## sql例子:

```
select
    a.cdate1 cdate,
    a.area_id,
    a.gm∨,
    b.gmv_year
from
(
    select
        cdate cdate1,
        partner_area_id as area_id,
        kpi_report_run_kpi_month_report_gmv gmv
        $TABLE(kpi_runing_report,partner_area_id,mayjor,mouths)
    where
        cdate>='$START(0)' and cdate<='$END(0)' )a</pre>
    join
        (
            select
                cdate cdate2,
                partner_area_id,
                kpi_uqa_all_kpi_qua_all_d_gmv_year gmv_year
                $TABLE(kpi_runing_report,partner_area_id)
            where
                cdate>='$START(0)' and cdate<='$END(0)' )b</pre>
on a.area_id=b.partner_area_id and a.cdate1=b.cdate2
```

### 数据源设置

数据设置	设置描述
类型	维度: 展示的是维度数据;指标:展示的指标数据,可进行数据过滤、百分比等功能; udc: 用户可自行计算指标与指标间的率值,通过点击 添加一行 按钮出现udc一行; 列名: 自定义表格,用户通过输入sql获取的字段列统称为列名,同时字段列可自行配置所需的数据
列显示 名称	表格的表头显示的列名称
列key	英文key,点击复制按钮可复制key,方便udc计算值使用此功能
列说明	当前列的详细注释说明,udc列可自定义
数据过滤	相当于sql的搜索条件,支持 =、like、not like、start with、end with、in、not in、<=、>=、<、>、!=、 正则 等条件
计算值	仅udc列可使用; 100*general.posterposter.paid_gmv/sum(general.posterposter.paid_gmv)
百分比	表格该列数据显示%
搜索	模糊查询功能勾选:表格显示上面则会出现模糊搜索的输入查询框,方便数据的模糊查询; 精确搜索:即输入内容完全匹配

即时过滤	从获取的表格数据中随时查询过滤; <b>多选过滤条件</b> :可设置多选过滤条件; <b>过滤条件的数据</b> :可为空,可输入数据,当设置数据后表头会显示select下拉搜索框; <b>默认搜索值</b> :数据查询框的默认值
外链	表格此列会有超链接显示; 功能: 1.支持平台内部报表跳转。 2.支持外站跳转,外链式报表钻取。通过报表之间获取跳转来实现数据的上卷或者下钻。 数据格式: 变量替换规则为 \${keyname} ,则将keyname替换为该行该keyname的值;例如: 一阶一级渠道的外链配置为 /visual/index/1819? first_source=\${first_source} , 则该报表一阶一级渠道的每个值可以跳转到 /visual/index/1819?first_source=WAP页
图片显示	输入图片的url
是否固 定	表格的左侧固定,右侧可以左右滑动的功能
默认排序	相当于sql的order by功能; 不排序: 默认值; 升序: order by asc 降序: order by desc
隐藏选 择	设置隐藏选择后,此列不显示在表格里
宽度(像 素)	表格列的宽度; 默认值100
设置默 认页码	表格数据的翻页功能
用户自 定义列 是否想 对占比	此功能只用于 用户自定义列udc有效

## 对比表格的特点

对比设 置值	设置描述
中文名 称	支持: 当日值、前一日值、当日相比前一日变化率、上周同日值、当日相比上周周日变化率、周平均值
显示	勾选:表格显示此列,反之则不显示

## 聚合表格的特点

指标 设置描述 指标 支持 最大值、最小值、求和、求平均数、计数的

## 新建表格

- 1. 输入表格名称:
- 2. 选择表格类型,表格类型的不同如上介绍;
- 3. 选择维度, 获取相应的指标后, 继续选择指标确定;
- 4. 数据过滤、百分比、搜素、即时过滤、外链、翻页、排序、表格颜色渐变等进行设置保存,同时 维度和指标之间可以拖拽排序;
- 5. 表格展示的数据;





#### 展示表格

#### 时间选择工具条:

可根据时间选择, 获取相应的图表数据。

### 自定义显示:

表格表头的列选择是否显示隐藏的功能,同时下载表格时也会隐藏对应列,方便下载查看该表格数据。 *模糊查询*:

用户输入或者选择,模糊匹配查询相对应的数。

### 数据提醒功能:

时间选择区域内的时间内、如果数据未生成、便会展示数据未生成的提示。

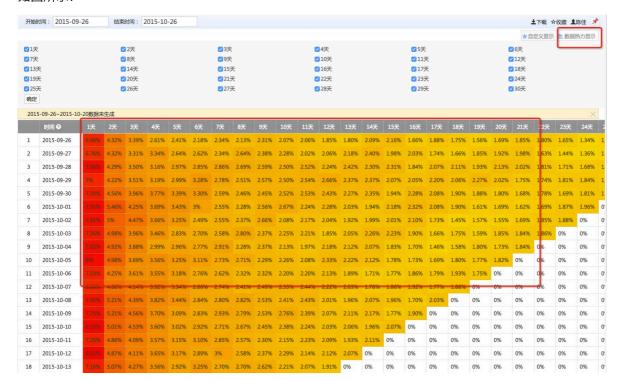
### 如图所示:



数据热力显示功能:

数据设置颜色渐变后,数据会根据大小的顺序由红色逐渐变黄色的渐变,更加直观展示数据的变化趋势。

#### 如图所示:



# 菜单

## 菜单展示

为方便用户的统一查找与管理,所有的菜单展示在页面的左侧。

## 设置菜单

一级菜单:一般是固定的,如果有需要可以邮件申请添加。

二级菜单:由用户自由设置,可随意发挥,设置后属于一级菜单下的子菜单。

报表: 选择需要在二级菜单下的报表,可多选,自由发挥。

外链: 用户想要其他平台的报表连接到数据平台,可启用此功能,格式(名称:外链url)。保存后即

刻生效。

## 邮件订阅

用户订阅邮件后,以邮件的形式接收查看该报表的数据。

邮件设置:邮件标题默认报表的名称,收件人为公司的邮箱。

## 项目

在使用平台前你需要新建分析项目,并进行数据和操作的相关配置。这一环节的重点是要在你提交的 HQL中体现分析项目中所关心的维度和指标。

## 配置参数

### • 平台时间

调度时间:hql任务被调起的时间。eg:天级别任务调度时间为每天上午4点。

运行时间: hql任务的时间。eg:该hql跑 2015-11-11 的数据即运行时间为 2015-11-11。

依赖时间: 默认为运行时间, 可以通过依赖表的 起始时间/终止时间 配置。

数据展现时间:查看数据时的时间 打开配置项的话可以在hql自定义。

#### • 调度周期

平台调度周期分为: 小时, 天, 周, 月, 每天定点, 每月定天, 每周定天

### • 数据运行时间(即函数替换基准时间)和调度时间偏移量

hql任务运行时间为调度时间减去该偏移量。

周期	格式	解释
天/ 周/ 月	-1day	默认偏移量为 -1day 。如您需要调度时间( <mark>2015-11-11</mark> )hql任务运行提前2天( <mark>2015-10-31</mark> )数据,那么该偏移量可以配置为 -2day
小时	-1hour	默认偏移量为 -1hour 。如您需要调度时间(2015-11-11 12:00:00)hql任务运行提前2小时(2015-11-11 10:00:00)数据,那么该偏移量可以配置为 -2hour

### • 回溯次数

回溯次数控制hql任务回跑天/小时数。如对于天级别任务回溯次数参数配置为2,偏移配置为-1day,那么 2015-11-11 调度任务时会调度起 2015-11-10,2015-11-09,2015-10-08 三天任务。

### • 自定义数据展现时间

如需hql任务自定义展现时间选中该选项即可,如果选中该选项hql自定义展现时间维度名必须为 cdate 且作为第一个维度, grouping sets 必须是 cdate 维度和其他维度的组合。

#### • 参数配置

设置hql参数和使用udf, transform

### • 维度/指标字段类型设置

目前支持 字符串(100字符)、字符串(200字符)、字符串(1024字符)、数字4种字段类型。对于维度指标设置类型后不可更改,同一个项目中相同的维度只能有一种类型。

#### • 依赖表依赖时间设置

通过依赖表的 起始时间/终止时间 参数设置检测依赖时间,设置依赖时间格式为 \$DATE(0)/\$DATE(0), \$HOUR(0)/\$HOUR(0) 。如运行时间为 2015-11-11 的任务需要检测依赖

表 2015-11-10 和 2015-11-11 数据是否ready,那么该参数可配置为 \$DATE(-1)/\$DATE(0)

## 创建项目

- 添加一个项目并完成数据配置和操作配置。在项目中您应该尽可能详尽地填写中英文名和项目解释。添加的项目可能会有审批流程。
- 添加分类,这里提供给用户对HQL进行分类管理的功能,区分不同的HQL查询类别,比如"交易相关"和"商品相关"
- 添加Hive sql并完成Hql解析。请将您分析项目中关心的维度,指标等分析量体现在Hive SQL中,点击"解析"后系统会根据HQL的结构解析出您关注的统计量。



### 输入HQL

## hql

```
SELECT
     first source,
     first_sub_source,
     client_device,
     client version,
     GROUPING ID.
     count(distinct device_id) AS uv,
     count(*) AS py
   FROM
                                                                          输入 hql
   (
     SELECT
       r1_first_source as first_source,
       r1_second_source as first_sub_source,
       client device,
       client_version,
       device id
```

进行HQL的解析,注意各个量的cn\_name,explain等属性是可以点击修改的(维度name相同但cn\_name 不同,在报表配置数据源时维度名显示最后一个中文名)



## 解析信息

## metrics 指标

	cn_name	е
uv	uv	F
pv	pv	F

# dimensions 维度

name	cn_name
nid_first	nid_first
nid	nid
eid	eid
poster_name	poster_name
e_nid	e_nid
client_device	client_device
client_version	client_version

## 操作配置

• 项目执行日期时间段、多选执行sql列表

## 操作配置(点击操作配置的确定按钮之后,保存按钮才会生效)



• 注意一定要点击确定后 点击保存设置方可保存此项目



• 在项目中添加报表的操作在下面的报表项目中说明。

编辑一个已经生成的项目,只需点击编辑按钮,即可进入和创建项目相似的界面。相关操作可参考创建项目。



## 查询项目

项目相关的查询功能主要体现在查询项目,运行状况查询,报表查询三个方面。

1. 查询项目 用户只需在查询框中键入项目名称,即可完成对项目的模糊查询



1. 运行详情查询

点击运行详情按钮,可以显示一个项目运行的各种状态。同时,点击新增启动,可以启动HQL,而点击日志,可以查到HQL的运行日志(如MR的执行日志等)。



### 1. 项目下报表查询

报表查询有两种方式, 一种是直接查询, 一种是多维查询。



点击查询可以看到项目中的所有报表。



多维查询允许用户通过指标和维度查询到报表中的部分数据,例如在"主APP指标"项目中选择APP用户访问量>1000000的记录。



多维查询的结果就是由用户自定义指标和维度组成的报表同时在页面右上方有重定制报表的按钮: 添加列和添加图标.添加的方法和报表管理中的操作时一致的.将在报表管理中详细说明。



# 收藏

用户收藏报表后,在左侧最顶端便可看到收藏的报表。

收藏设置: 时间框右侧点击收藏即可, 再次点击即可用取消收藏, 方便而实用的功能。

## timeline

记录项目所有重要事件发生的时间历程。

添加设置: 用户可以任意添加修改编辑。

# 数据映射

对报表中的数值进行解释,替换设置:选择项目、维度、指标,勾选是否替换原始值

数据格式-数值:替换原始数值

# 开发者中心

为开发者提供数据接口API

## 简单流程

第一步:申请应用

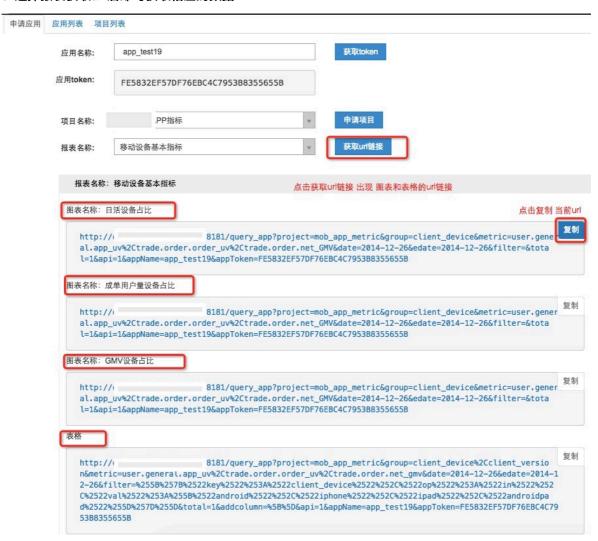
- 1. 输入应用名称后,点击获取token,显示token信息
- 2. 项目名称, 点击申请项目(注:同一应用名称下,项目不能重复申请)

申请应用	应用列表 项目列表	Ę.
	应用名称:	app_test17 1.输入应用名称 获取token
	应用token:	01D55E0364A9F0CE45D35938E7650A7C 2.点击获取token 此处显示获取的token
	项目名称:	msq_metric 3. 选择项目名称 然后点击 申请项目 中请项目

第二步:添加项目



#### : 选择报表获取url后即可获取相应的数据



返回数据信息

# 离线查询工具

为了满足从特定hive表中获取一批特定list的值的需求data平台增加离线查询小工具功能,该离线查询工具在各个层面和报表一样复用目前的权限、菜单等功能。 分析师通过新建配置查询小工具和配置查看该工具的权限(和配置报表一样), 具有使用该工具权限的运营按照工具注释填写查询内容list和该任务查询数据结果邮件接收人(查询数据结果作为邮件附件形式发送)提交任务即可

## 简单流程

### 第一步:新建离线查询工具

- 1. 分析师通过常用工具菜单》离线查询工具进入新建离线查询工具界面,根据提示填写离线查询工具名称、注释、hql。
- 2. 解析hql获取查询字段列名称并填写对应中文名,发送查询结果邮件时如果查询列有中文注释会替换为中文
- 3. 确定保存工具后可以通过菜单管理工具》报表管理查看/编辑离线查询工具

hql	hql格式为格式为 <mark>select [x as ] a,b,c,d,e,f,g from ( 子查询) x</mark> ,子查 询中支持时间函数 <mark>\$DATE</mark> , <mark>\$START</mark> , <mark>\$END</mark> , 支持 <mark>\$vars</mark> (hql中必须包 含 <mark>\$vars</mark> 变量)
\$vars	运营人员根据工具注释填写的查询内容在运行查询任务时会替换该变量
\$DATA	\$DATE 函数替换时间为用户提交任务时间 \$START 和 \$END 函数替换时间为用户选择的 开始时间和结束时间, \$START 和 \$END 具体用法和 \$DATE 用法一致

## 第二步: 查看(提交) 离线查询任务

- 1. 运营人员通过查看该离线工具,根据工具注释填写查询内容
- 2. 根据提示按照正确格式填写数据接收人,任务查询结果为邮件发送给所有邮件接收人
- 3. 提交任务, 任务提交成功后操作界面会跳转到工具查询进度界面

# 多维查询

项目在不用配置报表的情况下,只需要选择多个维度多个指标即可查询获取到相应的数据,同时以表格的形式展示,而且支持下载、增添多图元素的功能。

入口:项目管理=》多维度查询

方式:选择维度和相应的指标点击查询即可展现



.....