网易新闻点击量分析-操作指南

数据开发的目的是使用各种工具来分析数据,从数据中产生可指导行动的商业洞见,是从数据到价值的转换过程。 在很多场景下,数据分析人员需要对海量的数据进行快速的分析,性能上要像在传统的数据仓库中运行查询语句一 样,在几秒钟内得到数据分析的结果。

本README将以向导的形式,向大家展示一个典型的端到端大数据分析实例,包括数据采集、数据处理和数据分析 及数据导出这几个步骤:

第一步,数据采集: 在BDOS Online大数据平台,通过爬虫步骤爬取网易网页新闻,并将这些新闻数据存入 HDFS(为方便用户体验,本系统将提供已爬取完成的一个样例数据供用户下载,用户可通过URL文件导入,导入 到系统的HDFS;

第二步,数据处理:在Hive程序步骤中对新闻数据进行清洗并统计;

第三步,数据导出: 再将统计后的数据导入到MySQL数据仓库中;

第四步,BI报表: 使用Superset进行数据探索,将MySQL数据仓库中的统计数据以可视化的方式展示出来,分析 网易新闻的受欢迎程度。(即将推出)

用户只需克隆本项目,通过几次简单的点击,即可完成端到端的数据分析场景。

步骤介绍

1.克隆公共项目

用户登陆 BDOS Online 后,通过项目类型筛选公共项目,选择企业数仓-场景体验进行克隆



点击**克隆**,并自定义项目名称

	工具 管理	● common + □ 用户手册 🕑 prentissdemo
数据工程		列表形式: 回日
ah sin / Hitta		最近打开
Mattern BDOS Online 新手教程→	机器学习	维分 企业数合-场景体 个人项目 : 2021-12-21 09:21:36 prentissdemo
□ 项目类型 公共项目 ◎ □ □ 工業项目 ○ 归档项	个人项目 个人专属空间项目	防傷好 JL測试 (机构项目) : 2021-12-20 14:04:59 beta
企业数仓-场景体验 公共项目	* 项目名称	eta :
渠道分析-场景演示 <u>公共项目</u>	项目描述	ata :
	取測 确定	

进入项目主页

	数据工程 集成工具 数据浏览工具 管理	♥ common → 目用户手册
当前项目步骤 ① 《 拖拽步骤添加到工作流	数据采集。 数据转换。 数据分析。 数据质量。 数据服务。 数据应用。	BI报表 ~
4-数据转换导出 ETL程序 Ⅰ &	数据工程 / 个人项目 企业数仓-场景体验_clone01 ∠	
3-数据统计 Hive程序 ● 品	项目主页 工作流与依赖视图 列表视图 调度实例记录 教程	
2-数据清洗 Hive程序 ● ▲	項目信息 项目名称:企业数仓-场景体验_clone01 ∡	负责人: prentissdemo
目 1-新闻数据导入 数据库采集 ● ▲	创建时间: 2021-12-21 09:48:15 项目简介 🖉	修改时间: 2021-12-21 09:48:15
•	README.md 🗶 🖸	
	本项目为克建项目,克隆项目只保留数据处理步骤,每个步骤内的具体过程设置,请根据实际需要重 数据开发的目的是使用各种工具来分析数据,从数据中产生可指导行动的商业洞见,是从数据到价值 仓库中运行查询语句一样,在几秒钟内得到数据分析的结果。	新进行填入。 的转换过程。在很多场景下,数据分析人员需要对海量的数据进行快速的分析,性能上要像在传统的数据
	本README将以向导的形式,向大家展示一个典型的端到端大数据分析实例,包括数据采集、数据如	理和数据分析及数据导出这几个步骤:
	第一步。数据采集: 在BDOS Online大数据平台,通过爬虫步骤爬取网易网页新闻,并将这些新闻数 件导入,导入到系统的HDFS;	居存入HDFS(为方便用户体验,本系统将提供已爬取完成的一个样例数据供用户下载,用户可通过URL文
	第二步,数据处理:在Hive程序步骤中对新闻数据进行清洗并统计;	
	第三步, 数据导出: 再将统计后的数据导入到MySQL数据仓库中;	
	第四步,BI报表: 使用Superset进行数据探索,将MySQL数据仓库中的统计数据以可视化的方式展示	出来,分析网易新闻的受欢迎程度。(即将推出)
	用户只需克隆本项目,通过几次简单的点击,即可完成端到端的数据分析场景。	

2. 启动项目步骤

2.1 数据库采集,新闻数据导入

克隆公共项目-企业数仓后,进入步骤1:新闻数据导入

	资源消耗(即将推出) 步骤README 教程	
编与		品 馀仔
	•	
	采集数据源配置 采集目标配置	
	* 数据源 ●	
	public_mysql 🗸	
	没找到数据源? 点击此处创建新数据源	
	*表 0	
	warehouse_source 选择表	
	* 数据范围 ◎	
	◎ 全量	
	○ 増量	
	导出的字段 (列): ◎ 已选/源表: 8/8	
	Id VARCHAR () Item VARCHAR () Itime VARCHAR () Ititle VARCHAR ()	
	+ urts I VARCHAR ③ clicks I VARCHAR ③ create_timestamp I VARCHAR ③	
	update_timestamp VARCHAR ③	
	where ● 验证语法 自定义变量	
	结合自定义变量进行增量时间字段筛选的示例: edit_time>=str_to_date('date_1_day_ago','%Y%m%d') and edit_time <str_to_date('yyyymmdd','%y%m%d') (示例释义: 采集edit_time为过去一天的数据,其中'date_1_day_ago'在自定义时间变量中定义的过去一天,%Y%m%d%H表示指 定etrion的时间格式, 'wwwMMdd'表示当前时间)</str_to_date('yyyymmdd','%y%m%d') 	
	语句中 {{yyyy}}, {{MM}}, {{dd}}, {{H-H}}, {{mm}}, {{ss}}, {{timestamps}}, {{username}} 超定模板, 会被实际调度时间替换, timestamp格式, 例	
	如:1623254400, usemame为系统当前登录用尸名。	

保持默认配置内容并点击下一步

过程编写		□保存
	HDFS Hive	
	✓ 是否需要对数据进行分区 ◎	
	数据分区字段 💿	
	时间格式 YYYY-MM-DD HH:mm:ss	
	1 天 > 之前 >	
	基准是预期调度时间变量值:; 描述: 天之前 ○ 变量计算(以现在为时间基准计算)	
	数据分区存储路径格式 ◎	
	month hour minute second	
	month=MM/hour=HH/minute=mm/second=ss	
	* Hive数据库名称 ◎	
	public_project_data v	
	* Hive表名称 ◎	
	{{username}}_public_project_data_warehouse_source_full	
	* HDFS存储路径 ◎	
	/user/prentissdemo4/ingestion/{{username}}_public_project_data_warehouse_source_full	
	HDFS存储文件名与Hive表名称一致	
	± 2	咨

点击**保存**后,点击**试运行**

查看运行记录

运行模式	手动运行		运行状态	运行状态				查询
运行名称					用户	运行开始时间	运行结束时间	运行状态
nlinejob	p148 s720 p	rentissdemo4 202	1 12 21 11 20	35	prentissdemo4	2021-12-21 11:20:36	2021-12-21 11:21:43	成功

查看导入到公共 Hive 库的数据表

Online 数据工	程 集成工具 数	据浏览工具 管理					□ 用户手册	Prentissdemo4
管理								
数据源管理 数据表管理								
数据源类型 Hive Mysql	★ 数据库 ● public_ SCHEMA DDL ME 默认展示前10条数据	project_data v						
	id	item	time	title	urls	clicks	create_timestamp	update_timestamp
prentissdemo4_public_pr oject_data_warehouse_so	479460	新闻	2021-03-13 06:01:00	货拉拉坠车女孩被	https://news.163.c	10714	2021-03-13 06:01:00	2021-03-13 06:01:00
urce_full	479461	新闻	2021-03-13 06:02:	官方通报女辅警敲	https://news.163.c	6382	2021-03-13 06:02:	2021-03-13 06:02:
	479462	新闻	2021-03-13 06:03:	美将进一步限制供	https://news.163.c	5390	2021-03-13 06:03:	2021-03-13 06:03:
	479463	新闻	2021-03-13 06:04:	两会闭幕后 这3位	https://news.163.c	2079	2021-03-13 06:04:	2021-03-13 06:04:
	479464	新闻	2021-03-13 06:05:	2021年3月12日外交	https://news.163.c	1320	2021-03-13 06:05:	2021-03-13 06:05:
	479465	新闻	2021-03-13 06:06:	俄军收到8辆T-90M	https://war.163.co	1094	2021-03-13 06:06:	2021-03-13 06:06:
	479466	新闻	2021-03-13 06:07:	副局长回应与女辅	https://news.163.c	889	2021-03-13 06:07:	2021-03-13 06:07:
	479467	新闻	2021-03-13 06:08:	缅甸军方向西方服	https://news.163.c	880	2021-03-13 06:08:	2021-03-13 06:08:
	479468	新闻	2021-03-13 06:09:	晨枫:中国涨军费	https://war.163.co	862	2021-03-13 06:09:	2021-03-13 06:09:
	479469	新闻	2021-03-13 06:10:00	首飞刚过一个多月	https://war.163.co	812	2021-03-13 06:10:00	2021-03-13 06:10:00

2.2 转数据分析--Hive程序步骤:数据清洗

通过本步骤对 Hive 库表数据进行清洗,并存入到新的 Hive 表。进入步骤2:数据清洗,点击进入编辑界面

数据工	程 / 个人项目 / 企业数仓-场景体验 Hive程序 2-数据清洗 ∠
步骤配	置 运行记录 资源消耗(即将推出) 步骤README 教程
过程组	烏写 保存 试运行
编写	万克 ● 高級设置 預览 ● ○●
1 2 3 4 5 6 6 7 7 8 9 9 10 11 12 13 14 15 5 16 17 18 19 20 21 22 22 22 22 22 22 22 22 22 22 22 22	<pre> 创建数据清洗结果数据存放表,设置表的字段名称初类型 CREATE TABLE IF NOT EXISTS public_project_data.{{username}}_warehouse1 (</pre>



```
-- 创建数据清洗结果数据存放表,设置表的字段名称和类型
CREATE TABLE IF NOT EXISTS public_project_data.{{username}}_warehousel (
    `item` string,
    `time` string,
    `title` string,
    `urls` string,
    `clicks` int)
PARTITIONED BY(year string,month string,day string)
```

```
-- 插入分区字段: year、month、day,并从原始表选择字段数据写入数据清洗结果表
INSERT OVERWRITE TABLE public_project_data.{{username}}_warehouse1
PARTITION(year='2021',month='03',day='13')
SELECT item
     , `time`
     , title
     , urls
     , clicks
 FROM (SELECT item
            , `time`
            , title
            , urls
            , clicks
            , row_number() over (PARTITION BY urls, substr(create_timestamp,1,10) ORDER
BY clicks desc) as rn
       FROM public_project_data.
{{username}}_public_project_data_warehouse_source_full_demo
       WHERE create timestamp>='2021-03-13 00:00:00'
      ) t
WHERE t.rn=1
```

```
保持默认,点击保存后,点击试运行
```

```
数据工程 / 个人项目 / 企业数仓-场景体验 Hive程序 3-数据清洗 ∠
步骤配置
           运行记录
                       资源消耗(即将推出)
                                             步骤README
                                                             教程
                                                                                                             🛛 保存
                                                                                                                        ▶ 试运行
过程编写
                                                                                                                    预览 🛛 🕥
  编写主程序 🗢 高级设置
         创建数据清洗结果数据存放表,设置表的字段名称和类型
   2
     CREATE TABLE IF NOT EXISTS user_beta.warehouse1 (
          `item` string,
`time` string,
`title` string,
   3
   4
   5
     `urls` string,
`clicks` int)
PARTITIONED BY(year string,month string,day string)
   6
   8
     STORED AS TEXTFILE;
   9
  10
        插入分区字段: year、month、day, 并从原始表选择字段数据写入数据清洗结果表
  11
  12 INSERT OVERWRITE TABLE user_beta.warehouse1 PARTITION(year='2021',month='03',day='13')
  13 SELECT item
  14
             `time
          ,
          , title
  15
         , urls
  16
  17
            clicks
     FROM (SELECT item
  18
19
```

运行完成后,可以在运行记录中确认运行结果。

查看运行记录

参考截图查看Hive程序运行记录,可点击运行名称查看日志详情

运行模式 手动运行		运行状态	运行状态				查
运行名称				用户	运行开始时间	运行结束时间	运行状态
nlinejob p148 s719 p	rentissdemo4 2021	12 21 11 33	03	prentissdemo4	2021-12-21 11:33:03	2021-12-21 11:33:46	成功

预览数据表

用户从【管理-个人-数据表管理】菜单进入,通过数据源类型和数据库筛选,可预览上一个步骤创建的数据表数据

	数据工程 集成工具	数据浏览工具 管理					□ 用户手册	9 prentissdemo4
管理								
↑								
数据源管理 数据表管理								
数据源类型 O Hive O M	fysql *数据库 10 pt	ublic_project_data 🗸 🗸						
prentissdemo4_warehouse	Q SCHEMA DDL 默认展示前10条数据	METADATA EXAMPLE						
T NIXERAUG-CK	item	time	title	urls	clicks	year	month	day
prentissdemo4_warehou e1	us 新闻	2021-03-24 13:32:	中国式养娃智商税	http://caozhi.news	994	2021	03	13
	新闻	2021-03-18 11:06:00	年轻人最爱的城市	http://data.163.co	950	2021	03	13
	房产	2021-03-13 15:59:00	保利·花海湾 现已开	http://gz.house.16	1	2021	03	13
	房产	2021-03-13 15:52:00	城投珠江·天河壹品	http://gz.house.16	1	2021	03	13
	房产	2021-03-21 02:20:	越秀·天悦江湾 已开	http://gz.house.16	1	2021	03	13
	房产	2021-03-13 16:01:00	世茂天越 已开盘 单	http://gz.house.16	1	2021	03	13
	房产	2021-04-04 21:33:	珠江国际创业中心	http://gz.house.16	1	2021	03	13
	房产	2021-03-13 15:44:00	城投珠江·天河壹品	http://gz.house.16	1	2021	03	13
	房产	2021-04-02 16:56:	湾区金融城 待售 户	http://gz.house.16	1	2021	03	13
	房产	2021-03-18 20:36:	广州地铁82亿豪夺	http://gz.house.16	1	2021	03	13

2.3 数据分析--Hive程序步骤:数据统计

通过本步骤对数据进行统计分析,并把输出写入新的 Hive 表。进入步骤3:数据统计,点击进入编辑界面

数据工	工程 / 个人项目 / 企业数仓-场景体验 Hve程序 3-数据统计 ∠	
步骤配	配置 运行记录 资源消耗(即将推出) 步骤README 教程	
过程纲	呈编写	□保存 试运行
编写	皐写主程序 ◆ 高级设置	预览 🛛 🚺
1 2 3 3 4 5 6 7 8 9 100 111 122 131 14 15 16 17 18 19 200 21 221 223 224	<pre>1 創建数据就计结构表,设置表的字段名称和类型 2 CREATE TABLE IF NOT EXISTS public_project_data.{{username}}_warehouse2 (</pre>	

```
编写主程序
```

```
-- 创建数据统计结构表,设置表的字段名称和类型
CREATE TABLE IF NOT EXISTS public_project_data.{{username}}_warehouse2 (
   `item` string,
   `time` string,
   `title` string,
   `urls` string,
   `clicks` int,
   `last_clicks` int,
   `new_clicks` int)
PARTITIONED BY(year string, month string, day string)
STORED AS TEXTFILE;
-- 插入分区字段: year、month、day,并从原始表选择字段,对数据进行时间筛选,点击数量进行汇总统计,并把
结构写入数据统计结果表
INSERT OVERWRITE TABLE public project data.{{username}} warehouse2
PARTITION(year='2021',month='03',day='13')
SELECT item
     , `time`
     , title
     , urls
     , clicks
     , lag(clicks,1,0) over(PARTITION BY urls ORDER BY year,month,day) as last_clicks
     , clicks - lag(clicks,1,0) over(PARTITION BY urls ORDER BY year, month, day) as
new_clicks
FROM public_project_data.{{username}}_warehouse1
WHERE CONCAT(year,'-',month,'-',day,' 00:00:00')>='2021-03-13 00:00:00'
AND CONCAT(year, '-', month, '-', day, ' 00:00:00') <= '2021-04-01 00:00:00'
```

保持默认,点击**保存**后,点击**试运行**

查看运行记录

运行完成后,可以在运行记录中确认运行结果。

运行模式	手动运行		运行状态	运行状态				查询
运行名称					用户	运行开始时间	运行结束时间	运行状态
nlinejob i	o148 s718 prentiss	sdemo4 2021	12 21 11 37	49	prentissdemo4	2021-12-21 11:37:50	2021-12-21 11:38:28	成功

预览数据表

用户从【管理-个人-数据表管理】菜单进入,通过数据源类型和数据库筛选,可预览上一个步骤创建的数据表数据

Solutione 数据	工程 集成工具	数据浏览工具	管理						□ 用户手册	Prentissdemo4
管理 个人 效振波管理 <u>数据表管理</u>										
数据源类型	* 数据库 ③ F SCHEMA DDL 默认展示前10条数据	METADATA	AMPLE							
	item	time	title	urls	clicks	last_clicks	new_clicks	year	month	day
e2	新闻	2021-03-24 1	中国式养娃智	http://caozhi	994	0	994	2021	03	13
	新闻	2021-03-18 1	年轻人最爱的	http://data.16	950	0	950	2021	03	13
	房产	2021-03-13 1	保利·花海湾	http://gz.hou	1	0	1	2021	03	13
	房产	2021-03-13 1	城投珠江·天河	http://gz.hou	1	0	1	2021	03	13
	房产	2021-03-21 0	越秀·天悦江湾	http://gz.hou	1	0	1	2021	03	13
	房产	2021-03-13 1	世茂天越 已开	http://gz.hou	1	0	1	2021	03	13
	房产	2021-04-04	珠江国际创业	http://gz.hou	1	0	1	2021	03	13
	房产	2021-03-13 1	城投珠江·天河	http://gz.hou	1	0	1	2021	03	13
	房产	2021-04-02 1	湾区金融城 待	http://gz.hou	1	0	1	2021	03	13
	房产	2021-03-18 2	广州地铁82亿	http://gz.hou	1	0	1	2021	03	13

2.4 数据转换--ETL程序:数据转换导出

通过本步骤,对结果数据进行转换,从 Hive 库转换到目标 MySQL 库。

写			□保存 ▶
•	0		Ø
输入源配置	输出源配置	字段转换映射	数据查看
	* 数据源类型 ◎		
	Hive		
	* 数据库 ①		
	public_project_data		
	*表 ①		
	warehouse2		
	Query 🔍		
	select * from public_project_data.{{username}}_wa	rehouse2 查询字段	
	♥ 高级设置		
	参数配置 配置输入源		

保持默认并点击**下一步**

◆	•	<u></u>	
输入源配置	And a local sector and		
	输出源配查	字段转换映射	数据查看
	* 数据源类型 🕕		
	MySQL		
	* 数据源 ③		
	public_mysql		
	方式选择 💿 🔿 选择已有表 🔹 创建新表 💿		
	* 新表名		
	{{username}}_warehouse_mysql		
	MySQL表的名称不能以下划线开头,仅支持大/小写字母、数字、下划线及其组合 设置主键列 ◎	Ĩ	
	不选则默认使用第一列作为主键		
	* 建表语句 ① 点击获取建表语句		
	create table if not exists {{username}}_warehouse_mysql	(`item` 🗐	
	<pre>varchar(200),`time` varchar(200),`title` varchar(200),`u</pre>	urls` varchar(200),`clicks`	
	<pre>int(11),`last_clicks` int(11),`new_clicks` int(11) , PRI</pre>	(MARY KEY (`item`))	
	ENGINE=InnoDB DEFAULT CHARSET=utf8		
	* 数据变更 ◎		
	Update		

注:请**点击获取建表语句**

保持默认,并点击**下一步**

查看字段映射

选择需要检查的字段并进行匹配校验

编写								□ 保存 🕨 试
	O					•		0
	输入源配置		输出源配置		字顧	段转换映射	数据	居查看
前匹配信息	』:输入字段个数:7,输出:	字段个数: 7, 已匹配1	个数:7			参照输入表匹配	字段名匹配(精准)	→ 开始匹配
序号 🜖	输入表字段	类型	匹配状态 💿		序号 🕤	输出表字段	类型	匹配状态 💿
添加一行	+				添加一	行 +		
1	item	string	• 成功	••	1	item	varchar(200)	• 成功
2	time	string	- 成功	••	2	time	varchar(200)	- 成功
3	title	string	• 成功	••	3	title	varchar(200)	• 成功
4	urls	string	• 成功	••	4	urls	varchar(200)	• 成功
5	clicks	int	• 成功	••	5	clicks	int(11)	• 成功
6	last_clicks	int	• 成功	••	6	last_clicks	int(11)	• 成功
7	new clicks	int	- 成功	••	7	new_clicks	int(11)	- 成功

点击下一步

查看样例数据

	Ø-					•
	输入源配置	输出源配置		字段转换映射		数据查看
展示2条	✓ 真实数据	→ 0 查看				
台数据:						
tem	time	title	urls	clicks	last_clicks	new_clicks
新闻	2021-03-24 13:32:00	中国式养娃智商税 忽悠瘸 了三代人	http://caozhi.news.163.c om/21/0312/15/G4TAO NJU000181TI.html	994	0	994
新闻	2021-03-18 11:06:00	年轻人最爱的城市 别的没 有 快乐管够	http://data.163.com/21/ 0312/19/G4TP1R5E0001 81IU.html	950	0	950
换后数据:						
tem	time	title	urls	clicks	last_clicks	new_clicks
新闻	2021-03-24 13:32:00	中国式养娃智商税 忽悠瘸 了三代人	http://caozhi.news.163.c om/21/0312/15/G4TAO NJU000181TI.html	994	0	994
新闻	2021-03-18 11:06:00	年轻人最爱的城市 别的没 有 快乐管够	http://data.163.com/21/ 0312/19/G4TP1R5E0001 81IU.html	950	0	950

点击**保存**后,点击**试运行**

查看运行结果

可点击运行名称进入,查看运行详情

数据工程 / 个ノ	居工程 / 个人项目 / G 企业数仓-场景体验_模板 数据转换程序 4-数据转换导出 ∠								
步骤配置	运行记录	资源消耗(即将推)	出) 步骤	README	教程				
运行模式	手动运行		运行状态	运行状态				查询	
运行名称					用户	运行开始时间	运行结束时间	运行状态	
onlinejob_p	p148_s717_pr	entissdemo4_2021_	12_21_11_46_	.37	prentissdemo4	2021-12-21 11:46:	37 2021–12–21 11:47:35	成功	

2.5 工作流调度

调整完成后,在项目的【工作流与依赖试图】界面,将左侧的项目步骤中,逐个拖入试图中,并建立如下图所示的依赖。

数据工程 / 个人项目 企业数仓-场景体验_模板 🖉

项目主页	工作流与依赖视图	列表视图	调度实例记录	教程					
0, 0 0	100 200 夏位 图例: 设置流水线调度	300 → 流水线 ●	400 → 手动依頼 ① (● 未	500 (创建)	600	700	800 900	1000	1100 1200 1300 C P 启动
200	8	¹ 据库采集 -新闻数据导入	: 🍇	Hive程序 2-数据清洗	"> 😪	Hive程序 3-数据统计	' → @	ETL程序 :: 4-数据转换导出	
300									
400									
200									
600									

设置调度时间

点击**设置流水线调度**

设置流水线调度			×	
设置说明: 用户可通过对调度周期、调度时间等的设 定,对整条流水线配置调度属性,从而进行 自助、定时按依赖关系顺序调度运行。 带调度属性的滤水结启动后,可通过挂起流 水线并修改调度周期和调度时间等设置,实 现流水线调度属性的修改。	 调度周期: 请输入Cron表达式,不填表示只 请输入Cron表达式,不填表示只运行- * 调度开始时间 ● 2021-12-21 * 调度结束时间 ● 	运行一次 -次 ⓒ 00:00:00	1100 1200	
	 2021-12-22 失败重试次数 ● 3 	 ○ 00:00:00 重试间隔(秒) 60 取消 	S TROLL	

点击**确认**后点击保存,再启动流水线

查看流水线运行状态

数据工程 项目主页	i / ↑/	、项目 企业数仓-场景体验_ 工作流与依赖视图 列表视图	模板 ∠ 调度实例记录	教程				
选择	择步骤	请选择	状态 请选择		开始日期 ⑤ 00:00:00	-	O:00:00 O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O O	重置 查询
批量操作		新执行(补跑)					排序方式	默认 (更新时间) 🛛 🗸 🤟
	ID	调度时间	更新时间	执行时间	步骤名称	状态	耗时(秒)	操作
	138	2021-12-21 11:50:44	2021-12-21 11:50:44		4-数据转换导出	等待上下游依赖		重新执行
	137	2021-12-21 11:50:44	2021-12-21 11:51:54		3-数据统计	资源排队中		重新执行
	136	2021-12-21 11:50:44	2021-12-21 11:51:54	2021-12-21 11:51:17	7 2-数据清洗	成功	37	重新执行
	135	2021-12-21 11:50:44	2021-12-21 11:51:14	2021-12-21 11:50:4	16 1-新闻数据导入	成功	27	重新执行
				共4条 〈	1 > 10条/页 ~			

3. 可视化展示(仅限企业账号)

备注:如需申请企业账号,请联系

用户通过导航【集成工具-机构工具进入】,点击Superset的进入工具图标,跳转至Superset主界面

	数 据工	型 集成工具 数	据浏览工具 管理 配置						0 admin 👻	回用户手册 😕 beta
集成工具 用F 机构工具	□可选择以下精选的数据探: 个人工具	款工具								
b Süperset	Superset Superset来自于Airbni Access)和可视化商业 念是赋予每个人进行说	b, 采用自助服务模式, 在 洞见(Visual Insights)方面 问数据及数据分析的权力	 通行中 :自缩放数据访问(Scaling Data 有很强的优势,Superset本身的理 ,以便做出明智的数据决策。 	b Hue Hue 是用于访问B 序访问BDOS HD	DOS HDFS的Web应用程序交互的界面 FS,使用表并运行Hive查询。	● 遠行中 面,可以使用Hue中的应用程	Jupyter Jup	ipyterLab pyterLab 是最新的数据科约 tebook、编辑markdown文	*生产力工具,基于webj 本、交互式开发,查看5	• 运行中 的集成开发环境,可用于编写 2件和图片等等
i	0	O	2		0	12	0	ō	O	ß
※ 2	Superset	数据 ▼ 图	周表 看板	h atsaki					+ •	设置 - ■ -
安入3/白	\$X1/d/+	数据集							JULE ALIT	
所有者:	All → 製	上传CSV文件	∔ All →	类型: All - C	と授索					
	名称 🗄	上传Excel	÷	模式 🗧	已修改 🛓	修改人 🗦		所有者	操作	
▦	oiutput01	Physica	al public_project_da	ta public_project_c	data 11分之前	beta bdos		BB		
					« 1 »					
					1-1 总计 1					

通过Superset导航【数据-数据集】进入,点击**+数据集**

3.1 添加数据集

名称	内容	描述
数据源	public_project_data	选择系统默认提供的MySQL公共数据源
模式	public_project_data	选择系统默认的Schema模式
表	下拉框选择	选择需要进行可视化展示的目标表

Superset 数据 · 图表 看板 S	SQL 工具箱 -		+- 设置 - 🔳 -
数据 数据库 数据集 已保存查询	历史查询		批量选择 + 数据集
所有者: All ~ 数据库: All ~ 模式: All ~	类型: All - C	 添加数据集 × 	
名称:	类型 来源 🗄	从公共MySQL数据源: public_project_data中, 操作 选择目标表进行可视化展示	
oiutput01	Physical public_	数据库: mysql public_project_data ~	
		模式	
		Schema: public_project_data 👻	
		选择表 11 INPUBLIC_PROJECT_DATA	
		ā	
<		选择表或输入表名 👻	
		取消新增	

点击**新增**

3.2 添加图表

通过菜单导航至【图表】界面,点击**+图表**

👀 Superset	数据 - 图表 看板	SQL 工具箱 -				+• 设置• 🔳•
图表						批量选择 + 图表
■ 所有者:)	所有 → 创建人:所有 →	可视化类型:所有 ~ 数据集:所有 ~ 收	藏:所有 → Q 搜索			
图表:	可视化类型	数据集	修改人	最后修改 -	创建人	操作

名称	内容	描述
选择数据源	下拉框选择	选择添加的目标数据集
表	点击选择图表类型	选择需要进行展示的图表类型

点击**创建新图表**





点击**执行**后,点击保存

3.3 添加看板

通过菜单导航至【看板】界面,点击**+看板**

Superset 数据。 图表 看板 SQL 工具箱。							🕂 • 设置 • 🔳 •
看板							批量选择 + 看板
■ 所有者:所有 → 创建人:所有 → 状态:所有 → 收藏:所有 →	Q. 搜索						
标题:	修改人	状态 🗄	已修改 🗧	创建人	所有者	操作	
😭 [untitled dashboard]	beta bdos	草稿	4天之前	beta bdos	BB		
< 1 » 1.1 64+1							
		1 1 /2/11					

点击编辑图标进入看板编辑界面,选择【图表】,把模板图表手动拖拽添加至左侧看板画布



点击**保存**