渠道分析

数据分析是广告优化的基础能力,广告投放过程中通过数据分析挖掘改进点,可以数倍降低用户获取的成本,同时 量化广告投放的ROI。本项目以某电商的广告渠道投放及生产系统的Demo数据为例(Demo样例数据约5万条), 对渠道数据和生产数据进行采集,将投放的广告数据与生产数据进行深度的关联,得到渠道数据模型,并从广告点 击率、获客成本等维度对数据进行分析。

BDOS Online提供一套数据工作流系统,能够根据时间或数据可用性来运行这些程序和查询,实现任务的编排、调 度、监控。并以DAG拖拉拽的方式,全任务管理、监控作业及其依赖关系,提供数据采集、数据处理及数据导出等 步骤初始化后的定时自动化调度运行,并支持任务排队和插队机制。包括步骤如下:

第一步: URL文件导入,采集行业广告渠道信息和业务数据到HDFS 第二步: HDFS到Hive导入,把数据从HDFS导入到数仓(Hive) 第三步: Hive程序,从广告点击率、获客成本等维度对数据进行分析 第四步: ETL程序,把数据从数仓导入到指定的MySQL库,对结果数据的可视化展示进行数据准备 第五步: Superset,通过Superset对结果数据进行BI可视化展示

步骤介绍

创建机构

创建渠道分析个人/机构项目

备注: 仅机构管理员可创建机构项目

点击首页【创建新项目】,填写项目名称和描述,进行项目创建,创建后会进入数据工程项目界面



添加项目步骤

1. URL文件导入,采集行业广告渠道信息和业务数据到HDFS

创建项目后,添加【数据采集-URL文件导入】步骤至该项目。可点击步骤名称,修改名称为:1-URL文件导入 - 上 传广告、订单、报表数据

	数据工程 数据资产 集成工	具 数据浏览工具 管理 权限配置	Ø admin → □ 用户手册	e beta
当前项目步骤 (1) 《 点击切换步骤详情	数据采集 数据转换 数据	品分析·· 数据质量· 数据管理· 数据服务· 数据应用· BI报表·		
□ 1-URL文件导入 URL文件导入	数据工程 / 个人项目 / 渠道分析-3	杨景体验 URL文件导入 1-URL文件导入 - 上传广告、订单、报表数据 ∠		
•	过程编写			导入
	导入数据源配置 配置需要导入的目标URL和文件	* 目标URL下载		
		若输入的URL设有httpselfnttp前缀,则会默认加上`https://'作为前缀 文件名称		
		请确保填写后缀,例如:.csv,.txt,.xis		
	目标目录设置 选择目标目录	HDFS目录选择		
	And PP and PU and AN	/user/beta 点击测觉 不填表示默认选择当前目录,可输入*/*查看并选择子目录		
		是否选择覆盖同名文件		

点击Tab【步骤README】,并复制为本演示场景提供的Demo数据

数据工程 / 个人项	目 / 渠道分析-场景体验 URL文件导入 1-URL文件导入 - 上传广告、订单、报表数据 🖉
步骤配置 运行	记录 步骤README 教程 留言版
README.md 🖉	
URL文件导入步	5骤 - 上传广告、订单数据
过程编写	
广告渠道线索Demo数	擺web下載路径: http://linktime-public.oss-cn-qingdao.aliyuncs.com/Project_online/public_program/channel-clue-demonew.csv
广告渠道生产系统De	mo订单数据web下载路径: http://linktime-public.oss-cn-qingdao.aliyuncs.com/Project_online/public_program/channel-order-demonew.csv
广告渠道转化Demo数	北Web下载路径: http://linktime-public.oss-cn-gingdao.aliyuncs.com/Project_online/public_program/channel-report-demonew.csv
导入数据源配置	
名称	内容
目标URL下载	输入本演示场景提供的Demo数据链接,多个文件链接需分别进行多次导入
文件名称	可不填写,系统将默认读取链接上的文件名称及后缀:也可自定义文件名称,并填写文件后缀,如 .csv, .txt, .xls(系统当前支持的文件类型)
目标目录设置	
名称	内容
HDFS目录选择	保持默认
填写完毕后点击导。	λ

过程编写

广告渠道线索Demo数据Web下载路径: <u>http://linktime-public.oss-cn-qingdao.aliyuncs.com/Project_online/public</u>_program/channel-clue-demonew.csv

广告渠道生产系统Demo订单数据web下载路径: <u>http://linktime-public.oss-cn-qingdao.aliyuncs.com/Project_onlin</u> e/public_program/channel-order-demonew.csv 广告渠道转化Demo数据web下载路径: <u>http://linktime-public.oss-cn-qingdao.aliyuncs.com/Project_online/public</u>_program/channel-report-demonew.csv

导入数据源配置

名称	内容
目标 URL下 载	输入本演示场景提供的Demo数据链接,多个文件链接需分别进行多次导入
文件名 称	可不填写,系统将默认读取链接上的文件名称及后缀;也可自定义文件名称,并填写文件后缀, 如 .csv, .txt, .xls(系统当前支持的文件类型)

目标目录设置

名称	内容
HDFS目录选择	保持默认

填写完毕后点击**导入**,并查看运行结果

记录列表					
5件名称	存储路径	URL地址	下载时间	状态	上传用户
hannel-report-demonew.csv	/user/beta	http://linktime-public.oss-cn-qin gdao.aliyuncs.com/Project_onlin e/public_program/channel-report -demonew.csv	2021-11-04 06:34:37	成功	beta
nannel-order-demonew.csv	/user/beta	http://linktime-public.oss-cn-qin gdao.aliyuncs.com/Project_onlin e/public_program/channel-order -demonew.csv	2021-11-04 06:34:28	成功	beta
hannel-clue-demonew.csv	/user/beta	http://linktime-public.oss-cn-qin gdao.aliyuncs.com/Project_onlin e/public_program/channel-clue-	2021–11–04 06:34:01	成功	beta

2. HDFS到Hive导入,把数据从HDFS导入到数仓(Hive)

点击添加【数据转换-HDFS到Hive导入】步骤至该项目。可点击步骤名称,修改名称为:2 - HDFS到Hive导入 - 导入广告、订单数据到 Hive。

2.1 导入线索表: channel_clue_demonew到目标Hive库

HDFS数据源配置

名称	内容
HDFS目录文件选择	输入 / 查看并选择表:channel_clue_demonew

Hive目录表设置

名称	内容
数据库	下拉框选择系统提供的个人Hive数据库,如:user_xxx(xxx为当前登陆用户名)
数据表状态	点击新表需创建后点击 打开向导

保持默认,并点击**下一步**

			● ◆ 件格式								 _			
	/user/beta /ch	nannel-clue-dem	onew.csv	点击浏览	6						G 7 GOULT			
HDFS又忤选择	只能选择当前目录	或者子目录下的文例	件, 可输入"/"查看并	选择文件列表										
空印八百姓	27.0		に見い原始	+0.4=++()_)		司马索林	20219							
于权力闸付	進亏		运来力 陪付	1941719(\n)		りち子付	X515							
☑ 使用第一行作	作为表头 預算	3												
id	advertiser_i d	ad_id	ad_name	province_na me	city_name	name	address	create_time	create_time _detail	clue_type	external_url	app_name	request_id	order_id
315	27057551	21351	gyao	广东省	珠海市	陈丹丹	甘肃省彬市 梁平石家庄 街E座 1033 98	2020/8/5 0:00	2020/8/5 1 1:00	0	https://37. cn/	抖音	9a9356854 1acfdbd300 1ec275e32 8966	100018
316	39850300	21360	nwan	江苏省	南京市	张鑫	河南省龙市 和平关岭路 b座 319251	2020/8/5 0:00	2020/8/5 1 1:00	0	https://ww w.zh.cn/	抖音	221240d2e afd6b3d92f 04cf6cc8e 01de	100023
317	4049746	21360	jingdai	江苏省	苏州市	施建军	湖北省湖州 市蓋山太原 路O座 5094 05	2020/8/5 0:00	2020/8/5 1 1:00	0	https://ud. cn/	抖音	394e891a5 793c8f53a 021742fc15 6982	100064
318	4049746	21362	fangjun	江苏省	苏州市	侯燕	甘肃省秀荣 县梁平南昌 街u座 1844 96	2020/8/5 0:00	2020/8/5 1 1:00	0	https://25. cn/	抖音	7db75dbcc1 1369ce9913 cf3000928 d4e	100075

数据表导入向导

名称	内容
目标表名称	用户可自定义(只能使用大小写字母以及下划线和数字。例如 user_test.table1)
数据格式	选择csv
高级配置	保持默认

修改字段clue_type原始类型boolean型为string型

据表导入向导						
	文件格式				导入数据库	
目标表名称 user	r_beta.channel_clue_demonew					
表名称作	格式为鼓振滚表,只能使用大小写字母以及下划线和数字。例如 user_test	t.table1				
数据格式 ● csv	× 1					
高级配置 💼 🖬) #9/02					
	八 <u>奴</u> 第 用第二行在与主礼					
0 07						
字段 名称中5	R能包含大小写字母, 数字以及下划线					
名和	Id	英型	bigint	315	316	
名称	advertiser_id	类型	bigint	27057551	39850300	
名称	ad_id	类型	bigint	21351	21360	
名称	ad_name	类型	string	gyao	nwan	
名称	province_name	类型	string	广东省	江苏省	
名称	city_name	类型	string	珠海市	南京市	
名称	name	类型	string	陈丹丹	张鑫	
名称	address	类型	string	甘肃省彬市梁平石家庄	河南省龙市和平关岭路	
名称	create_time	类型	string	2020/8/5 0:00	2020/8/5 0:00	
名称	create_time_detail	类型	string	2020/8/5 11:00	2020/8/5 11:00	
名称	clue_type	类型	string	0	0	
名称	external_url	类型	string	https://37.cn/	https://www.zh.cn/	
名称	app_name	类型	string	抖音	抖音	
	request id	供型	string	9a93568541act/dbd3	221240d2eafd6h3d9	

点击**执行**

2.2 导入线索表: channel_order_demonew到目标Hive库

HDFS数据源配置

名称	内容
HDFS目录文件选择	输入 / 查看并选择表: channel_order_demonew

Hive目录表设置

名称	内容
数据库	下拉框选择系统提供的个人Hive数据库,如:user_xxx(xxx为当前登陆用户名)
数据表状态	点击新表需创建后点击 打开向导

保持默认,并点击**下一步**

参考截图修改字段类型后,点击执行

BDOS Onlin	ie	数据工程	数据资产 集成工具 数据浏览工具	管理	权限配置					🛛 admin 👻	回 用户手册	e beta
当前项目步员	数据表导入向导	₽									×	
点击切换步骤:												100
R HDF			0						•			
1-0			文件格式	~			T 12	7.46	导入数据库			
URL		名称	order_sum_amount	类型	double		128 修改为类型: float	4122.37				
[名称	order_need_pay_amount	类型	double		648.24	3404.86				
l		名称	order_payed_amount	类型	double	~	648.24	3404.86				
		名称	order_pay_way	类型	string		货到付款	货到付款				
		名称	order_activity_discount	类型	bigint		30	20				
		名称	order_activity_id	类型	boolean		1修改为类型: string	0				
		名称	order_activity_name	类型	string	~	双11跨店满减	首购优惠				
		名称	order_activity_type	类型	string		普通	首购				
		名称	order_source	类型	string		微信	微信				
		名称	order_source_type	类型	boolean	-	1修改为类型: string	1				
		名称	order_plan	类型	boolean	~	1	1				
		名称	order_idea	类型	boolean	~	1	1				
		名称	order_keywords	类型	string	~	童装	童装				
		名称	year	类型	bigint		2020	2020				
		名称	month	类型	bigint		8	8				
		名称	day	类型	bigint		5	5				
										E	一步 执行	
		_				_					9613	1

2.3 导入线索表: channel_report_demonew到目标Hive库

HDFS数据源配置

名称	内容
HDFS目录文件选择	输入 / 查看并选择表: channel_report_demonew

Hive目录表设置

名称	内容
数据库	下拉框选择系统提供的个人Hive数据库,如:user_xxx(xxx为当前登陆用户名)
数据表状态	点击新表需创建后点击 打开向导

保持默认,并点击**下一步**

参考截图修改字段类型后,点击执行

Conlin	ne 数据	計程	数据资产 集成工具 数据浏览工具	管理	权限配置		🛛 admin 👻	□ 用户手册	e beta
当前项目步骤	数据表导入向导							×	
a HDF HDR			♥ 文件格式						
1-UI URL:		使用	第一行作为表头						
	字段 名称	东中只能	包含大小写字母,数字以及下划线						
	名利	脉	id	类型	bigint	77057 197378			
	名利	豚	cost	类型	double	修改字段类型为: float 20054			
	名利	脉	ad_show	类型	bigint ~	79914 97342			
	名利	称	convert_cost	类型	double	%修改字段类型为:float010.79			
	名利	除	campaign_id	类型	bigint	7 7			
	名利	除	campaign_name	类型	string	巨奥 巨奥			
	名利	除	click	类型	bigint	79 97			
	名種	除	stat_datetime	类型	string	2020/8/11 11:00 2020/8/21 9:00			
	名利	除	ad_id	类型	bigint	21351 21360			
	名和	称	form	类型	boolean	修改字段类型为: string			
	名利	除	ad_name	类型	string	acao acao			
	名利	称	ad_convert	类型	bigint	5 5			
	名種	除	advertiser_id	类型	bigint	3571811 3571811			
	名種	原	ad_type	类型	string	jrtt jrtt			
							L	一步 执行	

执行完成后,可查看运行记录

乱工程 / 个人项目 / 東進分析场景体验 HDFS動Hive导入 ▲ 集記置 通行記录 歩骤FIEADME 教程 留言版								
记录列表								
表名	HDFS 文件路径	类型	数据源	导入策略	创建时间	导入结果	操作类型	用户
user_beta.channel_report _demonew	/user/beta/channel-rep ort-demonew.csv	hive	user_beta	overwrite	2021-11-04 14:26:30	成功	导入新增的表	beta
user_beta.channel_order_ demonew	/user/beta/channel-ord er-demonew.csv	hive	user_beta	overwrite	2021-11-04 14:22:32	成功	导入新增的表	beta
user_beta.channel_clue_d emonew	/user/beta/channel-clue -demonew.csv	hive	user_beta	overwrite	2021-11-04 14:15:36	成功	导入新增的表	beta
			д	3条 < 1 >	10条/页 ~			

0	成功	
	数据已经成功导入channel_report_demonew	

3. Hive程序,从广告点击率、获客成本等维度对数据进行分析

点击添加【数据分析-Hive程序】步骤至该项目。可点击步骤名称,修改名称为:3 - Hive程序- 数据计算。

编写主程序

备注:个人项目,请用实际登陆名替换beta,实际自定义的三个输入表名替换此次的三个表名。机构项目,请使用org_< 实际机构名称>替换user_beta

输出表: user_beta.table_flow_output

线索输入表: user_beta.channel_clue_demonew

订单输入表: user_beta.channel_order_demonew

```
CREATE TABLE if not exists user beta.table flow output (
`id` int COMMENT '唯一ID',
`ad type` string COMMENT '渠道类型',
`cost` float COMMENT '展现数据-总花费',
`ad show` int COMMENT '展现数据-展示数',
`convert_cost` float COMMENT '转化数据-转化成本',
`advertiser_id` int COMMENT '广告主ID',
`campaign name` string COMMENT '广告组name',
`click` int COMMENT '展现数据-点击数',
`stat_datetime` string COMMENT '数据起始时间',
`ad id` int COMMENT '计划id',
`form` int COMMENT '落地页转化数据-表单提交',
`ad_name` string COMMENT '计划name',
`ad_convert` int COMMENT '转化数据-转化数',
`ad_order` int COMMENT '订单数',
`ad amount` float COMMENT '订单总金额'
);
set hive.exec.dynamic.partition=true;
set hive.exec.dynamic.partition.mode=nonstrict;
set hive.exec.max.dynamic.partitions.pernode=3000;
set hive.exec.max.dynamic.partitions=3000;
set hive.exec.max.created.files=3000;
WITH report clue table AS (
SELECT report.*, clue.order id FROM user beta.channel report demonew report
LEFT JOIN
(SELECT ad id, order id from user beta.channel clue demonew) clue
ON report.ad_id = clue.ad_id
),
report_clue_order_table AS (
SELECT rc.*, orders.order payed amount
FROM report clue table rc
LEFT JOIN
(SELECT * FROM user_beta.channel_order_demonew) orders
ON rc.order_id = orders.order_id
)
insert overwrite table user beta.table flow output
select
jr.id,
jr.ad_type,
jr.cost,
jr.ad_show,
jr.convert_cost,
```

```
jr.advertiser_id,
jr.campaign_name,
jr.click,
jr.stat_datetime,
jr.ad_id,
jr.form,
jr.ad_name,
jr.ad_convert,
rco.ad_order,
rco.ad_amount
FROM user_beta.channel_report_demonew jr
LEFT JOIN (
    SELECT ad_id, count(1) as ad_order, sum(order_payed_amount) as ad_amount
    FROM report_clue_order_table where order_id is not null and order_payed_amount > 0
GROUP BY ad_id
) rco
ON jr.ad_id = rco.ad_id
```

点击**保存**后,点击**试运行**,查看运行日志

过程编写	日保存・试道
编写主程序 高级设置	預览
1 GREATE TABLE if not exists user_beta.table_flow_output { 2 'id 'int COMMENT '通一口', 'ad_tops string COMMENT '通道要要'', 'ad_show int COMMENT '通道要要要', 'ad_show int COMMENT '通道要要要', 'ad_show int COMMENT '消遣我要要", 'ad_show int COMMENT '消遣我要要", 'ad_show int COMMENT '消遣我要要", 'ad_show int COMMENT '消害我要要", 'ad_show int COMMENT '消害我要", 'ad_show int COMMENT '消害我要要", 'ad_show int COMMENT '消害我要", 'ad_show int COMMENT '消害我要", 'ad_show int COMMENT '消害我要", 'ad_show int COMMENT '消害我要", 'ad_show int COMMENT 'jim ad_show int 'int 'int 'int 'int 'int 'int 'int	<pre>1 CREATE TABLE if not exists user_beta.table_flow_output (2 'id' int COMENT '请定是=10', 3 'del.yeg' string COMENT '请注决是', 4 'cost float COMENT '请注决是', 5 'del.moin int COMENT '请定法是=4花, 6 'convert_cost float COMENT '诉定法是=4花, 7 'devertiser_id' int COMENT '诉定法是=4花, 9 'click' int COMENT '请定法是=4花, 9 'click' int COMENT '请法法, 9 'click' int COMENT '请法法, 10 'del.not' 'iff.'del.not', 11 'del.not', 12 'del.not', 13 'del.not', 14 'del.not', 15 'del.not', 15 'del.not', 16 'del.not', 17 'ABLE, 18 'del.not', 16 'del.not', 17 'ABLE, 18 'del.not', 17 'del.not', 18 'del.not', 19 'del.not', 19 'del.not', 10 'del.not', 10 'del.not', 11 'del.not', 11 'del.not', 12 'del.not', 13 'del.not', 14 'del.not', 15 'del.not', 16 'del.not', 17 'ABLE, 18 'del.not', 17 'del.not', 18 'del.not', 19 'del.not', 19 'del.not', 10 'del.not', 10 'del.not', 10 'del.not', 10 'del.not', 11 'del.not', 11 'del.not', 12 'del.not', 13 'del.not', 14 'del.not', 15 'del.not', 15 'del.not', 16 'del.not', 17 'ABLE, 18 'del.not', 17 'del.not', 16 'del.not', 17 'del.not', 17 'del.not', 17 'del.not', 17 'del.not', 18 'del.not', 19 'del.not', 19 'del.not', 10 'del.not', 11 'del.not', 11 'del.not', 12 'del.not', 13 'del.not', 14 'del.not', 15 'del.not', 15 'del.not', 16 'del.not', 17 'del.not', 17 'del.not', 17 'del.not', 17 'del.not', 17 'del.not', 18 'del.not', 18 'del.not', 18 'del.not', 18 'del.not', 18 'del.not', 18 'del.not', 19 'del.not', 10 'del.not',</pre>
战运行日志	
※ 运行日志 11 停止运行任务 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

数据工程 / 个人项目 / 渠道分析-场景体验 Hive程序- 数据计算 ∠									
步骤配置 上下游数据 运行记录 资源消耗(即将推出) 步骤README 教程	留言版								
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_p662_s4072_beta_2021_11_04_14_38_09	beta	2021-11-04 14:38:10	2021-11-04 14:43:24	成功					
	共1条 < 1	> 10条/页 ~							

查看输出表

点击【集成工具-Hue】进入Hue主界面

Conline 数据工程 数据资产 集成工具 数据浏览工具	管理 初闭配置 0 co	mmon - 四用户手册 🕒 beta
集成工具 用户可选择以下精选的数据探索工具 机构工具 个人工具		
Superset Superset	Hue MLflow Hwe是用于访问BDOS HDFS的Web应用程序交互的界面,可以使用Hue中的应用程序交互的界面,可以使用Hue中的应用程序访问BDOS HDFS,使用表并运行Hve宣询。 MLflow	。 台,可以追踪实验,重现实验,共享和部
II 6 0 C	g I I I I I I I I I I I I I I I I I I I	•

选择【Notebook-Editor-Hive】进入

HUE	Notebo	ok 👻	Q Se	arch saved documents	🌡 beta	
┋ 41 ■	Edito	r 🕨	🗎 Notebook	Add a description		0
	(1) + Ə o	1 Exam	 ♥ Hive ♥ MySQL 	Lename, or press CTRL + space	Database user_beta ♥ Type text ♥ � ?	< ²

输入查询语句,查看输出表数据

HUe	Notel	ook 👻	Q Search saved docum	ents					41
2	9	🖗 Hive 💿 🗛	dd a name Add a description						i
Suser_beta ables ow_output	(1) + 🕽 o	1 select * from user_b	eta.table_flow_output limit 10				0.64s Database user_beta ▼	Type text 👻 🍄	?
		INFO : Compiling com INFO : Semantic Ana INFO : Returning Hi FieldSchema(name thab pe:float, commentinul a(name table_floa-out	mand(queryId=root_20211104154501 lyiis Completed te_flow_output.cost, type:float, l), fialdSchema(name:table_flow, gut.click, type:int, comment_table_flow.	_baa0eB4f-69de-42f5-aa78-c ieldSchema(name:table_flov comment:null), FieldSchema output.advertiser_id, type 1), FieldSchema(name:table)	 126cdcd596c2): select * from v_output.id, type:int, comme (name:table_flow_output.ad_ :int, comment.null), FieldS = flow_output.atat_datetime,	user_beta.table_flow_output lin nt:null), FieldSchema(name:table show, type:int, comment:null), Fi hema(name:table_flow_outputs_flow type:string, comment:null), Fie	iit 10 _flow_output.ad_type, type:strin ieldSchema(name:table_flow_outpu paigo_name, type:string, comment JüSchema(name:table_flow_output.	x hg, comment:null), tr.convert_cost, ty :null), FieldSchem ad_dd, type:int, c	
		omment:null), FieldS: d_convert, type:int, es:null) INFO : Completed cor INFO : Executing cor INFO : Concurrency r INFO : Completed exec INFO : OK	<pre>chema(name:table_flow_output.form comment:null), FieldSchema(name: apiling command(queryId=root_2021 node is disabled, not creating a mmand(queryId=root_20211104154501 couting command(queryId=root_2021</pre>	, type:int, comment:null), table_flow_output.ad_order 1104154501_baa0e84f-69de-4 lock manager _baa0e84f-69de-42f5-aa78-c 1104154501_baa0e84f-69de-4	<pre>FieldSchema(name:table_to) r, type:int, comment:null), 1 42f5-aa78-d26cdcd596c2); Tim 4265cdcd596c2): select * from 42f5-aa78-d26cdcd596c2); Tim</pre>	<pre>w_output.ad_name, type:string, c FieldSchema(name:table_flow_outp a taken: 0.217 seconds user_beta.table_flow_output lin b taken: 0.0 seconds</pre>	comment:null), FieldSchema(name:t uut.ad_amount, type:float, commen nit 10	:able_flow_output.a ht:null)], properti	
		omment:null, FieldS d_convert, type:int, es:null) INF0 : Completed con INF0 : Completed con INF0 : Completed exc INF0 : Completed exc INF0 : OK Query History Q (2) table_flow_o	comment:null), FieldScheme(name: npiling comment/null), FieldScheme(name: npiling command(queryId=root_2021 node is disabled, not creating a mund(queryId=root_2021104154601 ecuting command(queryId=root_2021 Saved Queries Q. Query Buil utputid table_flow_output.ad_typ	<pre>, type:int, comment:null), table_flow_output.ad_order 10415459_baa0e84f-60de-4 lock manager _baa0e84f-69de-42ff-aa78-c 1104154501_baa0e84f-69de-4 der Results(10) Q_x^a a table_flow_output.cost</pre>	<pre>FieldSchema(name:table_tio r, type:int.comment:null). 12f5-aa78-d26cdcd596c2); Tim 12f6cdcd596c2): select + from 12f5-aa78-d26cdcd596c2); Tim 12f5-aa78-d26cdcd596c2); Tim</pre>	<pre>woutput.ad_name, type:string, c ieldSchema(neme:table_flow_output taken: 0.217 seconds user_beta.table_flow_output lin taken: 0.0 seconds table_flow_output.convert_cost</pre>	comment:null), FieldSchema(name:t uut.ad_amount, type:float, commen nit 10 table_flow_output.advertiser_id	table_flow_output.a at:null)], properti table_flow_outpu	
		omment:null, FieldS d_convert, type:int, es:null) INF0 : Completed cor INF0 : Completed cor INF0 : Completed ex INF0 : Completed ex INF0 : OK Query History Q (2) table_flow_o	hema(name:table_flow.output.form comment:null), FieldSchema(name: npiling command(queryId=root_2021 node is disabled, not creating a mand(queryId=root_2021104154691 Saved Queries Q. Query Bui utput.id table_flow_output.ad_typ jrtt i==	<pre>, type:int, comment:null); table_flow_output.ad_order 110415450_baa0e84f-60de-4 lock manager _baa0e84f-69de-42f5-aa78-c 110415450_baa0e84f-69de-4 ider Results (10) Q_x^a a table_flow_output.cost 6943.16 17865 2</pre>	<pre>FieldSchema(name:table_tio r, type:int.comment:null). 12f5-aa78-d26cdcd596c2); Tim 12f6cdcd596c2): select * from 12f5-aa78-d26cdcd596c2); Tim table_flow_output.ad_show 33313 390.000</pre>	<pre>w_output.ad_name, type:string, c ieldSchema(neme:table_flow_output table_flow_output) e taken: 0.217 seconds user_beta.table_flow_output lin t table_flow_output.convert_cost 6943.16 9041.42</pre>	comment:null), FieldSchema(name:t uut.ad_amount, type:float, commen nit 10 table_flow_output.advertiser_id 3571811 74662032	table_flow_output.a ti:null)], properti 合政电子 exer	
		d_convert, type:int, es:null) INFO : Completed con INFO : Completed con INFO : Completed con INFO : Completed ext INFO : OK Query History Q ☐ table_flow_o 1 721152 2 45271 2 45271	<pre>chema(name:table_flow.output.form comment:null), FieldSchema(name: ppling comment(queryId=root_2021 mannd(queryId=root_2021104154501 ceuting command(queryId=root_2021 Saved Queries Q. Query Bui utput.id table_flow_output.ad_typ jrtt int</pre>	, type:int, comment:null), table_flow_output.ad_order 1104154501_baa0e84f-69de-4 lock manager _baa0e84f-69de-42f5-aa78-c 1104154501_baa0e84f-69de-4 ider Results (10) Q e [*] a table_flow_output.cost 6943.16 1785.2 1864.9	<pre>FieldSchema(name:table_tio ; type:int_comment:null) 12f5-aa78-d26cdcd596c2); Tim 226cdcd596c2); select + from 12f5-aa78-d26cdcd596c2); Tim table_flow_output.ad_show 33313 89408 91770</pre>	<pre>w_output.ad_name, type:string. c ieldSchema(neme:table_flow_output user_beta.table_flow_output lin table_flow_output.convert_cost 6943.16 8942.62 413.46</pre>	comment:null), FieldSchema(name:t uut.ad_amount, type:float, commen nit 10 table_flow_output.advertiser_id 3571811 74563913 74563913	table_flow_output.a tt:null], properti table_flow_outpu 合取电子 创纪 上来	
		onsent:null, FieldS d_compett, type:int, es:null) INFO : Completed con INFO : Concurrency: INFO : Executing con INFO : Completed exe INFO : OK Query History Q I INFO : OK INFO : OK INFO : OK INFO : OK	<pre>chema(name:table_flow.output.form comment:null), FieldSchema(name: ppling comment(queryId=root_2821 mand(queryId=root_2821184154501 command(queryId=root_2821184154501 saved Queries Q Query Bui utput.id table_flow_output.ad_typ jrtt jrtt int</pre>	, type:int, comment:null), table_flow_output.ad_order 1104154591_baa@e84f-69de-4 lock manager _baa@e84f-69de-42f5-aa78-c 1104154501_baa@e84f-69de-4 ider Results (10) Q <i>c</i> [*] a table_flow_output.cost 6943.16 17885.2 18869.8 1024 4	<pre>FletGachema(name:table_tion ; type:int.comment:null) it2f5-aa78-d2ccdcd59ec2): Tim d2ccdcd59ec2): select + from iz2f5-aa78-d2ccdcd59ec2); Tim table_flow_output.ad_show 33313 89408 91779 50724</pre>	<pre>woutput.ad_name, type:string. c ieldSchema(neme:table_flow_output e taken: 0.217 seconds user_beta.table_flow_output line table_flow_output.convert_cost 6943.16 8942.62 4717.46 1026.4</pre>	comment:null), FieldSchema(name:t uut.ad_amount, type:float, commen at 10 table_flow_output.advertiser_id 3571811 74563913 74563913 74563913	table_flow_output.a tit:null]], properti 合联电子 创忆 七章	

4. ETL程序,把数据从数仓导入到指定的MySQL库

点击添加【数据转换-ETL程序】步骤至该项目。可点击步骤名称,修改名称为:4 - ETL程序- 导出作业用于BI展 示。

过程编写-输入源配置

名称	内容
数据源类型	下拉框选择Hive
数据库	选择个人/机构数据库
表	选择Hive程序步骤创建的输出表

点击**下一步**

过程编写-输出源配置

内容	名称
数据源类型	下拉框选择MySQL
数据源	选择个人/机构数据库(点击查看如何配置 ?
方式选择	创建新表
新表名	用户自定义
设置主键列	保持默认
建表语句	点击获取建表语句
数据变更	选择 Update

点击**下一步**

选择字段匹配标准,进行输入源字段和输出源字段的映射匹配,点击下一步,对数据进行预览

点击**保存**后,点击**试运行**

编写														□保存 ▶
	40	0			40 H 100 P 80	a .			2 F0 F5 10 10 1				• 助探查委	
展示2条		真实数据	· 0 25		THE LAST OFFICE ARE				TRACING				MIST.	
(始数据:														
d	ad_type	cost	ad_show	convert_cost	advertiser_id	campaign_name	click	stat_datetime	ad_id	form	ad_name	ad_convert	ad_order	ad_amount
21152	jrtt	6943.16	33313	6943.16	3571811	合联电子	33	2020/9/23 4:00	213	0	caona	1		
916221	jrtt	17885.2	89408	8942.62	74563913	UKZ	89	2020/12/13 7:00	21300	0	chao47	2		
换后数据:														
d	ad_type	cost	ad_show	convert_cost	advertiser_id	campaign_name	click	stat_datetime	ad_id	form	ad_name	ad_convert	ad_order	ad_amount
21152	jrtt	6943.16	33313	6943.16	3571811	合联电子	33	2020/9/23 4:00	213	0	caona	1		
916221	jrtt	17885.2	89408	8942.62	74563913	创亿	89	2020/12/13 7:00	21300	0	chao47	2		
行日志							1-9							
运行日志	II 停止运行任务	ocute request cpu: (0.50, memory: 1024.00, «	current worker's max	cpu is: 2.00, max me	mory is: 4096.00, max	x remaining cpu is: 2	.00, max remaining m	emory is: 4096.00,	execution nums is	: 0 and total instance	is: 1		
021-11-04 15:	13:09 worker has g	et this execution, read	y to run it											
wnload prog 21–11–04 15: 21–11–04 15: 21–11–04 15:	ram successfully 13:09,172 - main: 89 13:09,197 - main: 89	- NFO - current etl - NFO - ####### h	env is:production ive_metastore_jdbc_url:j	dbc:mysql://master.i	nnodb-cluster.14lb.thi	sdcos.directory:6446/	hive_db?characterE	ncoding=utf8&useSSL	=false,hive_metaste	ore_user:root				
2021-11-04 15: 121-11-04 15: R(Fil	13:13 2021-11-04 19 13:11,856 - main: 89 DW FORMAT DELIMI ELDS TERMINATED	5:13:11,723 - main: 89 - INFO - hive query:I TED BY ' '	- INFO - ######Start t	to parse input hive wi ECTORY '/tmp/etl/a	ith custom sql. aa0a8f163d3e11ecbd8	d0242ac110016'								

5. 工作流与依赖视图

从项目列表视图切换至工作流与依赖视图

	数据工程 数据资产 集成工具 数据消览工具 管理 权限配置	🟮 admin 👻	□ 用户手册	e beta
当前项目步骤 ()) 《 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。 数据留理。 数据服务。 数据应用。 BI报表。			
3 - Hive程序- 数据 Hive程序 &	数据工程 / 个人项目 課道分析-场景体验 ♂			
	工作流与依赖视图 列表视图 调度实例记录 项目主页 數程			
2 - HDFS到Hive导 HDFS到Hive导入	图例: →→ 第次线 ◎ →→ 手动依赖 ◎ (▲ 未创建)		0	启动
 1-URL文作号入 URL文作号入 	©			

手动拖拽步骤,并建立依赖关系

注:实线为有实际依赖关系的流水线调度步骤,虚线为逻辑依赖关系,不参与调度

100 0, 0 夏位 医	200 图例:	300 线	400 ● 未创建	50	0	600	700		800													
0,0 XW E	图例: ──) 流水	:线 → 手动依赖	6 (• 未创建)						000	90	0	1000		1100	1	200	1	300	1400		1500	0
	8199):	38 0) 于动化和	● 未回建																	~ · ·	0.1	
•																			(0.1		
•																						
	UF	11文件导入 :		HD)FS到Hive导入			-	Hive程序		: .			FTI 程	席	:						
	Ð,		·····>	5				>	0 115-	ID NY			-> EI		TT IN							
	-	URLX1+ GA		2	- HDFSEJHW	e			3 - HIVE	对王门"				4 - E	111程序-	***						

设置调度时间,启动调度后,实现关联的步骤会按顺序按时运行

BDOS Online 当前項目步骤 () く	数据工程 数据资产 集 	5成工具 数据浏览工具 管理 权环 数据分析 数据质量 数据管理	限配置 数据服务 数据应用 BI报3	N		0 admin 🗸	□ 用户手册	e beta
1014日 Windowsza France	数据工程 / 个人项目 渠道行 工作流与依赖视图 列表:	设置流水线调度			×			
2 - HDFS到Hw号 rcDSBHw号A 9 - URL文作号A URL文作号A		设置说明: 用户可通过对调度周期、调度时间等的设 定,对整条流水线截置调度履管,从而进行 自助、定时按线能关系顺序调度运行。 带调度履性的点水线自动后,可通过挂起流 水线并构态调度周期和调度时间等设置,实 现流水线调度属性的修改。	 環境人口の表达式,不須表示只适行 情報入口の表达式,不須表示只适行一次 * 環境开始封闭 ● 2021-11-05 * 環境結束封闭 ● 2021-11-06 失敗重试次数 ● 	一次 ○ 00:00:00 ◎ 00:00:00 ■ 近向隔(秒)	1200 : 1/ 7	1300	1400	1500
	000		3	60 10399	翻认			

点击**保存**,并**启动**

查看调度实例记录

数据工程 / 个人: 工作流与依赖视图	项目 渠道分析-场景体验 到 列表视图 调度实	. ▲ 例记录 项目主页	教程					
选择步骤	请选择 く	状态 请选择	~ 目 刑	合日期 ③ 00:00:00	- 目结束日期	© 00:00:00		重置 查询
批量操作 重調	新执行(补跑)						排序方式	默认 (更新时间) 🛛 🗸 🗸
D ID	调度时间	更新时间	执行时间	步骤名称	状态	耗时(秒)		操作
7104	2021-11-05 11:45:06	2021-11-05 11:45:07	2021-11-05 11:45:07	3 – Hive程序– 数据计算	运行中			重新执行 停止
				共1条 〈 1 〉	10条/页 🗸			

6. Superset, 对结果数据进行BI可视化展示

点击【集成工具-机构工具-Superset】,点击进入工具。进入Superset界面后,选择【数据源-数据库】后点击+, 添加数据库

添加数据库

名称	内容
数据库	bdos_private< <i>当前登陆用户名</i> >db,如:bdos_private_beta_db
SQLAlchemy URL	SQLAlchemy URI: mysql:// <bdos online当前登陆用户名="">:<bdos online当前登陆密码<br="">>@master.innodb-cluster.l4lb.thisdcos.directory:6446/bdos_private<<i>当前登陆用户名</i>>db? charset=utf8</bdos></bdos>

其他选项可保持默认后,点击**保存**

添加数据表

返回Superset界面后,选择【数据源-数据表】后点击+,添加数据表

👀 Superset 👒 安全 🗸 管理 🗸 S	数据源 🖌 🔟 图表 🐠 看板 🚨 SQL	工具箱 🖌			+ 新建	 4 ~
数据表 SOL LAB 援索 Q	 数据库 型数据表 土传CSV文件 环视图 不等于 ▼ 			▼ 过滤列	jæ ● -	
	表	数据库	修改人	修改于	I	
名称	内容					
数据库	下拉框选择刚添加的	数据库				
表名	可手动输入ETL步骤的	的MySQL输出表名				

其他选项可保持默认后,点击**保存**

返回Superset界面后,选择【图表】后点击+,

💙 Superset	📽 安全 🗸	管理 🗸	数据源 ✔	山 图表	3 看板	▲ SQL工具箱 🗸							+ 新建	-	~	4 ∨
		图表								— T 过	滤列表	€				

配置数据源

名称	内容
数据源	下拉框选择添加的数据表
选择可视化类型	选择 Bar Chart

配置图表

名称	内容
时间范围	选择 未过滤
指标	COUNT(convert_cost)
序列	ad_name

其他选项可保持默认后,点击**运行**



