

系统使用说明-用户篇

- 1. 使用前提-权限申请说明
- 2. binlog数据接入Hive
 - 2.1 新增接入
 - 2.2 修改接入
- 3. binlog表导出

1. 使用前提-权限申请说明

使用该系统前请确认有如下系统权限：

- 大数据集群生产账号，用于访问数据，申请地址：<http://dpp.intra.xiaojukeji.com/index#/apply/personaccount>

平台地址：<http://bigdata.xiaojukeji.com/alpheus/#/>

2. binlog数据接入Hive

2.1 新增接入

用户只需提交一次申请，系统自动流转以下流程：



1， 首页列表页，查找到需要接入的mysql表，点击操作列的“接入信息”，在弹出框中，点击“新增接入”。

如果没有找到mysql表，需要在元数据系统提交主从库信息。http://bigdata.xiaojukeji.com/ddc_metadata_fe/#/mysqlSearch
审核通过后，一个小时内会同步到该平台。

业务线: 全部mysql库: account_cmysql表: anti_coupon_g_hive表: 请输入表名查询任务列表

库名	表名	topic	用户名	表中文名	业务线	国家	操作
account_c	anti_coupon_g_coupon_bind		account_c_a_r	绑券发券	专快车	中国	详情 编辑 接入信息

接入列表

+ 新增接入

id	名称	Hive库	Hive表	更新时间	状态	操作
暂无数据						

2， 在新增接入页，填写接入信息，如下所示，

注：1) 接入名称默认为mysql表名

2) 任务优先级 默认为T3 （低优先级） 用于说明数据/报表的重要程度 （请用户谨慎选择）

3) 是否有效用于标识ETL流程状态，如果置为无效，ETL流程将停止，不再拉取binlog数据更新到Hive表中。

4) 增量类型 全量类型 解释说明链接 增量&全量数据接入

5) binlog数据所要接入的Hive库，如果不了解，可以咨询身边数仓或BI同学，本部门数据所在Hive库

6) binlog数据所要接入的Hive表名称，系统自动生成，规则为ods_binlog_ + mysql表名 + 更新类型(increment, whole)

7) 主键，默认为表的主键，也可以自定义主键，主要用于数据去重使用

8) 过滤表达式中，用户可以自定义数据过滤规则，如order_id < 1000 && uid > 39764385
(逻辑表达式)，后续版本支持系统内置变量过滤

9) 保留删除是针对问题跟踪场景，可以选择是否将mysql中已删除的记录还原到hive中，如果选择保留，binlog表中以system_rule_etl_delete_flag字段标识是否为删除记录，0-正常记录，1-删除记录

10) 保留重复是为增量更新类型设计，用于标识是否对周期内数据去重

11) 是否回溯数据也是为增量更新类型设计，用于标识是否对增量数据进行回溯，配合增量字段使用

新增接入

* 中文名称 人车关系表 1

* 业务线 专快车

* 接入类型 增量 4

* 主键 id 7

☒ 保留删除 9

☐ 保留重复 10

* 任务优先级 T3(其他) 2

* mysql库 charon

* Hive库 请选择 5

过滤表达式 请输入内容 8

☒ 是否回溯数据 11

增量字段 请选择 12

☒ 是否生效 3

* mysql表 g_driver_vehicle_relation

Hive表 ods_binlog_g_driver_vehicle_relation_increment 6

导出任务列表

mysql表-字段名称	hive表-字段名称	操作
id	id	编辑
gvid	gvid	编辑
guid	guid	编辑
uid	uid	编辑
product	product	编辑
_create_time	a_create_time	编辑
_modify_time	a_modify_time	编辑
_status	a_status	编辑

保存接入

返回首页

问题反馈

3, 在新增接入页，点击”保存接入“，任务提交完毕。


任务列表页面，如下图所示：

ID	工单类型	库名	表名	表中文名	业务线	状态	是否全量	全量状态	查看
281	新增接入	manhattan_riskmanage	rc_invoke_10_1023	风控调用	营销	数据治理	是	全量swan配置	详情 列表
280	新增接入	uemc_noah	n_session_message	会话消息表	客服	血缘关系配置	否		详情 列表
279	新增接入	uemc_noah	n_dialogue_session	会话表	客服	血缘关系配置	否		详情 列表
278	新增接入	users	contacts_10_127	用户通讯录备份	顺风车	血缘关系配置	是	全量swan配置	详情 列表
274	新增接入	upm_data	upm_role_feature_relation	角色权限关联表	安全	血缘关系配置	是	全量swan配置	详情 列表
273	新增接入	upm_data	upm_role_users	角色关联角色组表	安全	血缘关系配置	是	全量swan配置	详情 列表
272	新增接入	upm_data	upm_role_group	角色组表	安全	血缘关系配置	是	全量swan配置	详情 列表
271	新增接入	upm_data	upm_feature_group_role_relation	功能组和角色的关系表	安全	血缘关系配置	是	全量swan配置	详情 列表
270	新增接入	upm_data	upm_feature_group_relation	功能组和功能的关系表	安全	血缘关系配置	是	全量swan配置	详情 列表
269	新增接入	upm_data	upm_feature_group	功能组表	安全	血缘关系配置	是	全量swan配置	详情 列表

提交时若表格信息不完整，则会弹出“表格信息不完整是否前往补全”的提示信息，若点击确定则会跳转到相应的表格编辑页面，若点击取消则不会跳转页面 如下图所示：

nctual	punctual	编辑
ferrer	referrer	编辑
_region	ride_region	编辑
ate_time	a_create_time	编辑
ify_time	a_modify_time	编辑
status	a_status	编辑
r_name	driver_name	编辑
r_level		编辑
entry_id		编辑
r_grade		编辑
nt_grade		编辑
n_ability	listen_ability	编辑
_valid_date	serial_no_valid_date	编辑
cation_id	qualification_id	编辑
n_id_valid_date	qualification_id_valid_date	编辑
e_town	home_town	编辑
nt_method	payment_method	编辑
direct	is_direct	编辑
channel	sub_channel	编辑

提示

 表信息不完整，请完善表后再次提交，是否前往完善？

取消

确定

4、任务处理完毕后，会有钉钉 + 邮件提醒，访问接入的hive表请到d++平台申请访问权限，链接：<http://dpp.intra.xiaojukeji.com/index#/apply/dataAuth>

2.2 修改接入

1、首页列表页，查找需要修改的接入(可以按照mysql库表查询，也可以按照hive表查询)，点击操作列的“接入信息”，在弹出框中，点击“编辑”。

业务线: 全部
 mysql库: 全部
 mysql表: 请选择
 hive表: ods_binlog_g_car_increment
 查询
 任务列表

接入列表
 新增接入

id	名称	Hive库	Hive表	更新时间	状态	操作
129	车辆信息表	gulstream_ods	ods_binlog_g_car_increment	2017-07-17 18:46:16	有效	详情 编辑

2、在编辑接入页面，修改相关信息。目前支持修改的字段如下图所示，在确认信息后，点击提交接入。

- 注：1) 接入名称
- 任务优先级，
 - 是否生效，如果mysql表已下线或者业务不再实现，可以修改该字段，置为无效
 - 主键
 - 过滤表达式，
 - 保留重复
 - 是否回溯数据
 - 增量字段
 - 如果mysql表中新增了字段，并希望同步修改到Hive表中，可以在此添加字段
 - 如果接入类型为全量时 保留删除也可修改

编辑接入

* 中文名称ods_g_passenger_increases_temp1

* 任务优先级T3(其他)2

☒ 是否生效3

* 业务线专快车

* mysql库gulfstream

* mysql表g_passenger

* 接入类型增量

* Hive库gulfstream_ods

* Hive表ods_g_passenger_increases_temp

* 主键passenger_id4

过滤表达式请输入内容5

导出任务列表

☒ 保留删除10

☐ 保留重复6

☒ 是否回溯数据7

增量字段请选择8

mysql表-字段名称	hive表-字段名称	操作
passenger_id	passenger_id	编辑
area	area	编辑
phone	phone	编辑
role	role	编辑
origin_id	origin_id	编辑
channel	channel	编辑
fast_car_count_call	fast_car_count_call	编辑
count_call	count_call	编辑
count_complete	count_complete	编辑
count_bad	count_bad	编辑

3. 提交接入后进入任务列表页。

3. binlog表导出

目的是将hive中接入的小时级别的binlog数据，导出到其他hive表中(可以是已存在的hive表，不存在的则由alpheus系统自动创建)，根据业务需要，指定数据起止时间及周期，便可以得到天级别类型的导出数据，适应多样化的数据分析需求。

入口都是在接入页面（新增接入页面，或者编辑接入页面），以下以编辑接入为例

1. 首页列表页，查找到需要修改的接入(可以按照mysql库表查询，也可以按照hive表查询)，点击操作列的“接入信息”，在弹出框中，点击“编辑”，进入编辑接入页面

2. 在编辑接入页面，点击导出任务列表（一个binlog表支持导入多个其他表中），如下所示：

编辑接入

* 中文名称车辆信息表

* 任务优先级T3(其他)

☒ 是否生效

* 业务线专快车

* mysql库gulfstream

* mysql表g_car

* 接入类型增量

* Hive库gulfstream_ods

* Hive表ods_binlog_g_car_increment

* 主键car_id

过滤表达式请输入内容

导出任务列表

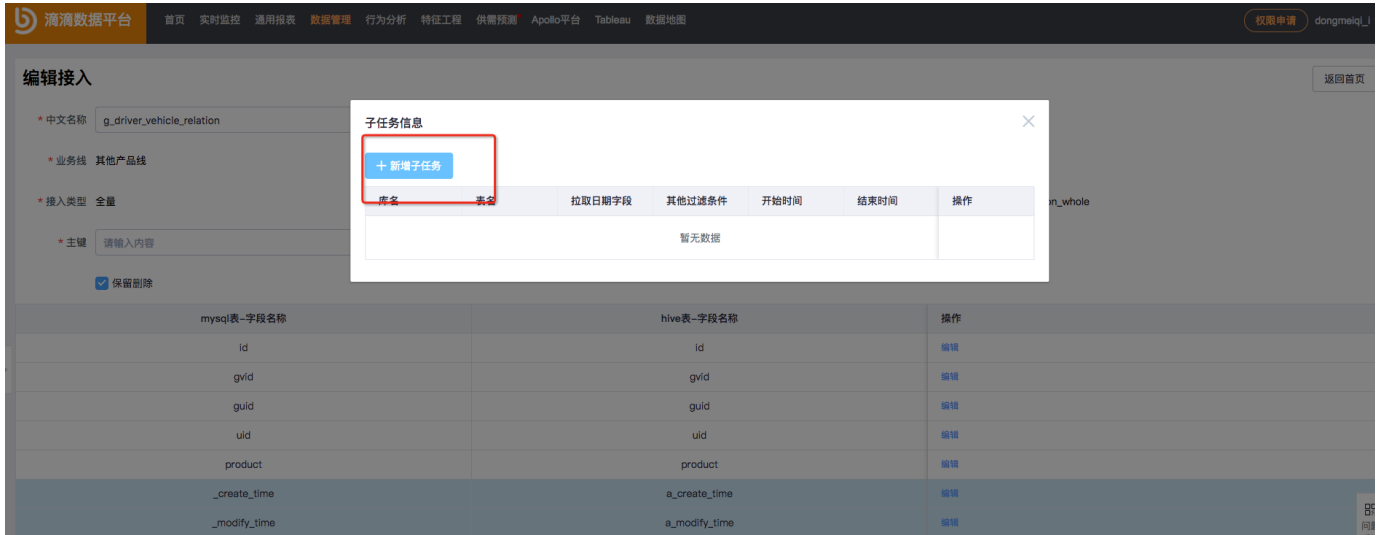
☒ 保留删除

☐ 保留重复

☒ 是否回溯数据

增量字段请选择

mysql表-字段名称	hive表-字段名称	操作
car_id	car_id	编辑
company_id	company_id	编辑
plate_no	plate_no	编辑
level	level	编辑



3. 点击新增子任务，填写任务信息如下所示：

- 注：1) 库名，需要导出的目标Hive库
- 2) 表名，需要导出的目标Hive表
- 3) 开始时间，导出binlog数据的开始分区
- 4) 结束时间，导出binlog数据的结束分区
- 5) 拉取字段（或过滤条件），可以为多个日期字段，逗号分隔，增量类型的binlog数据有效，用于提取满足某个条件的数据，与下方的开始结束时间配合使用（如图中所示，
- 从binlog表前一天的0到23点的分区内，提取a_create_time在前一天00：00：00到23：59：59的记录，去重取最新记录后写入gulfstream_ods.g_car_band），
- 多个日期字段时，每个字段之间用or连接，时间周期超过一天时，按照第一个字段动态分区
- 6) 其他过滤条件，导出任务过滤条件（仅限 sql表达式）
- 7) 如果导出表存在有binlog表中未对应的字段，以null表示

新增子任务信息

库名: abtest 1

表名: apollo_log 2

开始时间: 1 3 天前 0 时

结束时间: 1 4 天前 23 时

拉取日期字段: 多个字段请用#分割 5

其他过滤条件: 如:status = 3 and city_id = 1 6

源表-字段名称	目标表-字段名称
null	testkey
null	key_count
null	unique_key_count

7

确定

4. 信息填写完毕后保存，进入任务列表页。