项目验收说明

### 一 项目文档

**1）项目计划书，文档结构如下**

       （1）系统数据流程设计

* + - 集群流程图
    - 集群框架图
    - 集群特点

       （2）具体版本选型

       （3）服务器选型

* + - 机器成本考虑
    - 运维成本考虑

       （4）集群规模

* + - 数据量分析
    - 集群规划

       （5）人员配置

* + - * 离线测试集群搭建用时
      * 离线正式集群搭建用时

       （6）首批指标查询用时

* + - * 常用指标用时
      * 临时指标用时
      * 活动指标用时
      * 特殊指标用时

       （7）Bug 处理用时

* + - * 框架 Bug 处理用时
      * 指标 Bug 处理用时

       注意：人员配置部分要内嵌 Excel，以表格的形式展示所有组员每天 上午、下午、晚上 三个时间段的任务，粒度可以更细，不可以更粗。若有线上同学，与线下同学一并分配任务，没有任何区别对待。

       可依据实际情况对文档结构进行补充，鼓励各位创新。

**2）项目文档**

       环境搭建及建模部分可以不写，但是从 ODS 到 ADS 层的所有建表语句，装载语句及脚本都要在文档中出现，且每层应准备一个对应表格对本层所有表进行简要介绍，其中，ODS 层必须标注采集方式（全量/增量）。如下。

descript

ODS 层表格示例

descript

DWD 层表格示例

descript

DIM 层表格示例

descript

DWS 层表格示例

descript

ADS 层表格示例

       可依据实际情况对表格字段进行补充。

# 2 代码

## 2.1ODS层

### 2.1.1 分类表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| base\_category\_info\_full | 维度 | 全量 | ods\_base\_category\_info\_full | 分类表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_base\_category\_info\_full;
2. **CREATE** EXTERNAL **TABLE** ods\_base\_category\_info\_full
3. (
4. `id`                  STRING COMMENT '编号',
5. `category\_name`       STRING COMMENT '分类名称',
6. `create\_time`         STRING COMMENT '创建时间',
7. `update\_time`         STRING COMMENT '更新时间',
8. `deleted`             STRING COMMENT '是否删除'
9. ) COMMENT '分类表'
10. PARTITIONED **BY** (`dt` STRING)
11. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
12. NULL DEFINED **AS** ''
13. LOCATION '/warehouse/edu/ods/ods\_base\_category\_info\_full/';

装载语句

### 2.1.2省份表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| base\_province\_full | 维度 | 全量 | ods\_base\_province\_full | 省份表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_base\_province\_full;
2. **CREATE** EXTERNAL **TABLE** ods\_base\_province\_full
3. (
4. `id`         STRING COMMENT '编号',
5. `**name**`       STRING COMMENT '省份名称',
6. `region\_id`  STRING COMMENT '地区ID',
7. `area\_code`  STRING COMMENT '行政区位码',
8. `iso\_code`   STRING COMMENT '旧版ISO-3166-2编码，供可视化使用',
9. `iso\_3166\_2` STRING COMMENT '新版IOS-3166-2编码，供可视化使用'
10. ) COMMENT '省份表'
11. PARTITIONED **BY** (`dt` STRING)
12. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
13. NULL DEFINED **AS** ''
14. LOCATION '/warehouse/edu/ods/ods\_base\_province\_full/';

装载语句

### 2.1.3来源表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| base\_source\_full | 维度 | 全量 | ods\_base\_source\_full | 来源表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_base\_source\_full;
2. **CREATE** EXTERNAL **TABLE** ods\_base\_source\_full
3. (
4. `id`           STRING COMMENT '引流来源id',
5. `source\_site`  STRING COMMENT '引流来源名称',
6. `source\_url`   STRING COMMENT '引流来源链接'
7. ) COMMENT '来源表'
8. PARTITIONED **BY** (`dt` STRING)
9. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
10. NULL DEFINED **AS** ''
11. LOCATION '/warehouse/edu/ods/ods\_base\_source\_full/';

装载语句

### 2.1.4科目表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| base\_subject\_info\_full | 维度 | 全量 | ods\_base\_subject\_info\_full | 科目表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_base\_subject\_info\_full;
2. **CREATE** EXTERNAL **TABLE** ods\_base\_subject\_info\_full
3. (
4. `id`              STRING COMMENT '编号',
5. `subject\_name`    STRING COMMENT '科目名称',
6. `category\_id`     STRING COMMENT '分类',
7. `create\_time`     STRING COMMENT '创建时间',
8. `update\_time`     STRING COMMENT '更新时间',
9. `deleted`         STRING COMMENT '是否删除'
10. ) COMMENT '科目表'
11. PARTITIONED **BY** (`dt` STRING)
12. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
13. NULL DEFINED **AS** ''
14. LOCATION '/warehouse/edu/ods/ods\_base\_subject\_info\_full/';

装载 语句

### 2.1.5加购表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| cart\_info\_full | 维度 | 全量 | ods\_cart\_info\_full | 加购表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_cart\_info\_full;
2. **CREATE** EXTERNAL **TABLE** ods\_cart\_info\_full
3. (
4. `id`              STRING COMMENT '编号',
5. `user\_id`         STRING COMMENT '用户id',
6. `course\_id`       STRING COMMENT '课程id',
7. `course\_name`     STRING COMMENT '课程名称 (冗余)',
8. `cart\_price`      **DECIMAL**(16, 2) COMMENT '放入购物车时价格',
9. `img\_url`         **BIGINT** COMMENT '商品图片地址',
10. `session\_id`      STRING COMMENT '会话id',
11. `create\_time`     STRING COMMENT '创建时间',
12. `update\_time`     STRING COMMENT '修改时间',
13. `deleted`         STRING COMMENT '是否删除',
14. `sold`            STRING COMMENT '是否已售'
15. ) COMMENT '加购表'
16. PARTITIONED **BY** (`dt` STRING)
17. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
18. NULL DEFINED **AS** ''
19. LOCATION '/warehouse/edu/ods/ods\_cart\_info\_full/';

装载语句

### 2.1.6章节表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| chapter\_info\_full | 维度 | 全量 | ods\_chapter\_info\_full | 章节表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_chapter\_info\_full;
2. **CREATE** EXTERNAL **TABLE** ods\_chapter\_info\_full
3. (
4. `id`              STRING COMMENT '编号',
5. `chapter\_name`    STRING COMMENT '章节名称',
6. `course\_id`       STRING COMMENT '课程id',
7. `video\_id`        STRING COMMENT '视频id',
8. `publisher\_id`    STRING COMMENT '发布者id',
9. `is\_free`         STRING COMMENT '是否免费',
10. `create\_time`     STRING COMMENT '创建时间',
11. `update\_time`     STRING COMMENT '更新时间',
12. `deleted`         STRING COMMENT '是否删除'
13. ) COMMENT '章节表'
14. PARTITIONED **BY** (`dt` STRING)
15. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
16. NULL DEFINED **AS** ''
17. LOCATION '/warehouse/edu/ods/ods\_chapter\_info\_full/';

装载语句

### 2.1.7章节评价表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| comment\_info\_inc | 事实 | 增量 | ods\_comment\_info\_inc | 章节评价表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_comment\_info\_inc;
2. **CREATE** EXTERNAL **TABLE** ods\_comment\_info\_inc
3. (
4. `type` STRING COMMENT '变动类型(是否删除)',
5. `ts`   **BIGINT** COMMENT '变动时间',
6. `data` STRUCT<id :STRING,user\_id :STRING,chapter\_id :STRING,course\_id :STRING,
7. comment\_txt :STRING,create\_time :STRING,deleted :STRING> COMMENT '数据',
8. `old`  MAP<STRING,STRING> COMMENT '旧值'
9. ) COMMENT '章节评价表'
10. PARTITIONED **BY** (`dt` STRING)
11. ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.JsonSerDe'
12. LOCATION '/warehouse/edu/ods/ods\_comment\_info\_inc/';

装载语句

### 2.1.8课程信息表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| course\_info\_full | 维度 | 全量 | ods\_course\_info\_full | 课程信息表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_course\_info\_full;
2. **CREATE** EXTERNAL **TABLE** ods\_course\_info\_full
3. (
4. `id`                STRING COMMENT '编号',
5. `course\_name`       STRING COMMENT '课程名称',
6. `course\_slogan`     STRING COMMENT '课程标语',
7. `course\_cover\_url`  STRING COMMENT '课程封面',
8. `subject\_id`        STRING COMMENT '学科id',
9. `teacher`           STRING COMMENT '讲师名称',
10. `publisher\_id`      STRING COMMENT '发布者id',
11. `chapter\_num`       **BIGINT** COMMENT '章节数',
12. `origin\_price`      **DECIMAL**(16, 2) COMMENT '价格',
13. `reduce\_amount`     **DECIMAL**(16, 2) COMMENT '优惠金额',
14. `actual\_price`      **DECIMAL**(16, 2) COMMENT '实际价格',
15. `course\_introduce`  STRING COMMENT '课程介绍',
16. `create\_time`       STRING COMMENT '创建时间',
17. `deleted`           STRING COMMENT '是否删除',
18. `update\_time`       STRING COMMENT '更新时间'
20. ) COMMENT '课程信息表'
21. PARTITIONED **BY** (`dt` STRING)
22. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
23. NULL DEFINED **AS** ''
24. LOCATION '/warehouse/edu/ods/ods\_course\_info\_full/';

装载语句

### 2.1.9收藏表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| favor\_info\_inc | 事实 | 增量 | ods\_favor\_info\_inc | 收藏表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_favor\_info\_inc;
2. **CREATE** EXTERNAL **TABLE** ods\_favor\_info\_inc
3. (
4. `type` STRING COMMENT '变动类型',
5. `ts`   **BIGINT** COMMENT '变动时间',
6. `data` STRUCT<id :STRING,course\_id :STRING,user\_id :STRING,create\_time :STRING,update\_time:STRING,deleted :STRING> COMMENT '数据',
7. `old`  MAP<STRING,STRING> COMMENT '旧值'
8. ) COMMENT '收藏表'
9. PARTITIONED **BY** (`dt` STRING)
10. ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.JsonSerDe'
11. LOCATION '/warehouse/edu/ods/ods\_favor\_info\_inc/';

装载语句

### 2.1.10知识点表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| knowledge\_point\_full | 维度 | 全量 | ods\_knowledge\_point\_full | 知识点表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_knowledge\_point\_full;
2. **CREATE** EXTERNAL **TABLE** ods\_knowledge\_point\_full
3. (
4. `id`                STRING COMMENT '编号',
5. `point\_txt`         STRING COMMENT '知识点内容',
6. `point\_level`       STRING COMMENT '知识点级别',
7. `course\_id`         STRING COMMENT '课程id',
8. `chapter\_id`        STRING COMMENT '章节id',
9. `create\_time`       STRING COMMENT '创建时间',
10. `update\_time`       STRING COMMENT '更新时间',
11. `publisher\_id`      STRING COMMENT '发布者id',
12. `deleted`           STRING COMMENT '是否删除'
13. ) COMMENT '知识点表'
14. PARTITIONED **BY** (`dt` STRING)
15. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
16. NULL DEFINED **AS** ''
17. LOCATION '/warehouse/edu/ods/ods\_knowledge\_point\_full/';

装载语句

### 2.1.11订单明细表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| order\_detail\_inc | 维度 | 增量 | ods\_order\_detail\_inc | 订单明细表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_order\_detail\_inc;
2. **CREATE** EXTERNAL **TABLE** ods\_order\_detail\_inc
3. (
4. `type` STRING COMMENT '变动类型',
5. `ts`   **BIGINT** COMMENT '变动时间',
6. `data` STRUCT<id :STRING,course\_id :STRING,course\_name :STRING,order\_id :STRING,user\_id :STRING,origin\_amount
7. :**DECIMAL**(16, 2),coupon\_reduce :**DECIMAL**(16, 2),final\_amount :**DECIMAL**(16, 2),create\_time
8. :STRING,update\_time :STRING> COMMENT '数据',
9. `old`  MAP<STRING,STRING> COMMENT '旧值'
10. ) COMMENT '订单明细表'
11. PARTITIONED **BY** (`dt` STRING)
12. ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.JsonSerDe'
13. LOCATION '/warehouse/edu/ods/ods\_order\_detail\_inc/';

装载语句

### 2.1.12订单表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| order\_info\_inc | 事实 | 增量 | ods\_order\_info\_inc | 订单表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_order\_info\_inc;
2. **CREATE** EXTERNAL **TABLE** ods\_order\_info\_inc
3. (
4. `type` STRING COMMENT '变动类型',
5. `ts`   **BIGINT** COMMENT '变动时间',
6. `data` STRUCT<id :STRING,user\_id :STRING,origin\_amount :**DECIMAL**(16, 2),coupon\_reduce :**DECIMAL**(16, 2),final\_amount :**DECIMAL**(16, 2),order\_status :STRING,out\_trade\_no :STRING,trade\_body
7. :STRING,session\_id :STRING,province\_id :STRING,create\_time :STRING,expire\_time :STRING,update\_time :STRING> COMMENT '数据',
8. `old`  MAP<STRING,STRING> COMMENT '旧值'
9. ) COMMENT '订单表'
10. PARTITIONED **BY** (`dt` STRING)
11. ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.JsonSerDe'
12. LOCATION '/warehouse/edu/ods/ods\_order\_info\_inc/';

装载语句

### 2.1.13支付表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| payment\_info\_inc | 事实 | 增量 | ods\_payment\_info\_inc | 支付表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_payment\_info\_inc;
2. **CREATE** EXTERNAL **TABLE** ods\_payment\_info\_inc
3. (
4. `type` STRING COMMENT '变动类型',
5. `ts`   **BIGINT** COMMENT '变动时间',
6. `data` STRUCT<id :STRING,out\_trade\_no :STRING,order\_id :STRING,alipay\_trade\_no
7. :STRING,total\_amount :**DECIMAL**(16, 2),trade\_body :STRING,payment\_type
8. :STRING,payment\_status :STRING,create\_time :STRING,update\_time
9. :STRING,callback\_content :STRING,callback\_time
10. :STRING> COMMENT '数据',
11. `old`  MAP<STRING,STRING> COMMENT '旧值'
12. ) COMMENT '支付表'
13. PARTITIONED **BY** (`dt` STRING)
14. ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.JsonSerDe'
15. LOCATION '/warehouse/edu/ods/ods\_payment\_info\_inc/';

装载语句

### 2.1.14课程评价表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| review\_info\_inc | 事实 | 增量 | ods\_review\_info\_inc | 课程评价表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_review\_info\_inc;
2. **CREATE** EXTERNAL **TABLE** ods\_review\_info\_inc
3. (
4. `type` STRING COMMENT '变动类型(是否删除)',
5. `ts`   **BIGINT** COMMENT '变动时间',
6. `data` STRUCT<id :STRING,user\_id :STRING,course\_id :STRING,
7. review\_txt :STRING,review\_stars :STRING,create\_time :STRING,deleted :STRING> COMMENT '数据',
8. `old`  MAP<STRING,STRING> COMMENT '旧值'
9. ) COMMENT '课程评价表'
10. PARTITIONED **BY** (`dt` STRING)
11. ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.JsonSerDe'
12. LOCATION '/warehouse/edu/ods/ods\_review\_info\_inc/';

装载语句

### 2.1.15测验表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| test\_exam\_inc | 维度 | 增量 | ods\_test\_exam\_inc | 测验表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_test\_exam\_inc;
2. **CREATE** EXTERNAL **TABLE** ods\_test\_exam\_inc
3. (
4. `type` STRING COMMENT '变动类型(是否删除)',
5. `ts`   **BIGINT** COMMENT '变动时间',
6. `data` STRUCT<id :STRING,paper\_id :STRING,user\_id :STRING,
7. score :**DECIMAL**(16, 2),duration\_sec :**BIGINT**,create\_time :STRING,
8. submit\_time :STRING,update\_time :STRING,deleted :STRING> COMMENT '数据',
9. `old`  MAP<STRING,STRING> COMMENT '旧值'
10. ) COMMENT '测验表'
11. PARTITIONED **BY** (`dt` STRING)
12. ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.JsonSerDe'
13. LOCATION '/warehouse/edu/ods/ods\_test\_exam\_inc/';

装载语句

### 2.1.16测验问题表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| test\_exam\_question\_inc | 维度 | 增量 | ods\_test\_exam\_question\_inc | 测试问题表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_test\_exam\_question\_inc;
2. **CREATE** EXTERNAL **TABLE** ods\_test\_exam\_question\_inc
3. (
4. `type` STRING COMMENT '变动类型(是否删除)',
5. `ts`   **BIGINT** COMMENT '变动时间',
6. `data` STRUCT<id :STRING,exam\_id :STRING,paper\_id :STRING,question\_id :STRING,user\_id :STRING,
7. answer :STRING,is\_correct :STRING,score :**DECIMAL**(16, 2),create\_time :STRING,
8. update\_time :STRING,deleted :STRING> COMMENT '数据',
9. `old`  MAP<STRING,STRING> COMMENT '旧值'
10. ) COMMENT '测验表'
11. PARTITIONED **BY** (`dt` STRING)
12. ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.JsonSerDe'
13. LOCATION '/warehouse/edu/ods/ods\_test\_exam\_question\_inc/';

装载语句

### 2.1.17试卷表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| test\_paper\_full | 维度 | 全量 | ods\_test\_paper\_full | 试卷表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_test\_paper\_full;
2. **CREATE** EXTERNAL **TABLE** ods\_test\_paper\_full
3. (
4. `id`             STRING COMMENT '编号',
5. `paper\_title`    STRING COMMENT '试卷名称',
6. `course\_id`      STRING COMMENT '课程id',
7. `create\_time`    STRING COMMENT '创建时间',
8. `update\_time`    STRING COMMENT '更新时间',
9. `publisher\_id`   STRING COMMENT '发布者id',
10. `deleted`        STRING COMMENT '是否删除'
11. ) COMMENT '试卷表'
12. PARTITIONED **BY** (`dt` STRING)
13. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
14. NULL DEFINED **AS** ''
15. LOCATION '/warehouse/edu/ods/ods\_test\_paper\_full/';

装载语句

### 2.1.18试卷问题表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| test\_paper\_question\_full | 维度 | 全量 | ods\_test\_paper\_question\_full | 试卷问题表 |

建表语句

1. DROP TABLE IF EXISTS ods\_test\_paper\_question\_full;
2. CREATE EXTERNAL TABLE ods\_test\_paper\_question\_full
3. (
4. `id`             STRING COMMENT '编号',
5. `paper\_id`       STRING COMMENT '试卷id',
6. `question\_id`    STRING COMMENT '题目id',
7. `score`          DECIMAL(16, 2) COMMENT '得分',
8. `create\_time`    STRING COMMENT '创建时间',
9. `deleted`        STRING COMMENT '是否删除',
10. `publisher\_id`   STRING COMMENT '发布者id'
11. ) COMMENT '试卷问题表'
12. PARTITIONED BY (`dt` STRING)
13. ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
14. NULL DEFINED AS ''
15. LOCATION '/warehouse/edu/ods/ods\_test\_paper\_question\_full/';

装载语句

### 2.1.19知识点问题表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| test\_point\_question\_full | 维度 | 全量 | ods\_test\_point\_question\_full | 知识点问题表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_test\_point\_question\_full;
2. **CREATE** EXTERNAL **TABLE** ods\_test\_point\_question\_full
3. (
4. `id`             STRING COMMENT '编号',
5. `point\_id`       STRING COMMENT '知识点id',
6. `question\_id`    STRING COMMENT '问题id',
7. `create\_time`    STRING COMMENT '创建时间',
8. `publisher\_id`   STRING COMMENT '发布者id',
9. `deleted`        STRING COMMENT '是否删除'
10. ) COMMENT '知识点问题表'
11. PARTITIONED **BY** (`dt` STRING)
12. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
13. NULL DEFINED **AS** ''
14. LOCATION '/warehouse/edu/ods/ods\_test\_point\_question\_full/';

装载语句

### 2.1.20问题信息表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| test\_question\_info\_full | 维度 | 全量 | ods\_test\_question\_info\_full | 问题信息表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_test\_question\_info\_full;
2. **CREATE** EXTERNAL **TABLE** ods\_test\_question\_info\_full
3. (
4. `id`             STRING COMMENT '编号',
5. `question\_txt`   STRING COMMENT '题目内容',
6. `chapter\_id`     STRING COMMENT '章节id',
7. `course\_id`      STRING COMMENT '课程id',
8. `question\_type`  STRING COMMENT '题目类型',
9. `create\_time`    STRING COMMENT '创建时间',
10. `update\_time`    STRING COMMENT '更新时间',
11. `publisher\_id`   STRING COMMENT '发布者id',
12. `deleted`        STRING COMMENT '是否删除'
13. ) COMMENT '问题信息表'
14. PARTITIONED **BY** (`dt` STRING)
15. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
16. NULL DEFINED **AS** ''
17. LOCATION '/warehouse/edu/ods/ods\_test\_question\_info\_full/';

装载语句

### 2.1.21问题选项表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| test\_question\_option\_full | 维度 | 全量 | ods\_test\_question\_option\_full | 问题选项表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_test\_question\_option\_full;
2. **CREATE** EXTERNAL **TABLE** ods\_test\_question\_option\_full
3. (
4. `id`             STRING COMMENT '编号',
5. `option\_txt`   STRING COMMENT '选项内容',
6. `question\_id`     STRING COMMENT '题目id',
7. `is\_correct`      STRING COMMENT '是否正确',
8. `create\_time`    STRING COMMENT '创建时间',
9. `update\_time`    STRING COMMENT '更新时间',
10. `deleted`        STRING COMMENT '是否删除'
11. ) COMMENT '问题选项表'
12. PARTITIONED **BY** (`dt` STRING)
13. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
14. NULL DEFINED **AS** ''
15. LOCATION '/warehouse/edu/ods/ods\_test\_question\_option\_full/';

装载语句

### 2.1.22用户章节进度表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| user\_chapter\_process\_inc | 事实 | 增量 | ods\_user\_chapter\_process\_inc | 用户章节进度表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_user\_chapter\_process\_inc;
2. **CREATE** EXTERNAL **TABLE** ods\_user\_chapter\_process\_inc
3. (
4. `type` STRING COMMENT '变动类型',
5. `ts`   **BIGINT** COMMENT '变动时间',
6. `data` STRUCT<id :STRING,course\_id :STRING,chapter\_id :STRING,user\_id :STRING,position\_sec
7. :**BIGINT**,create\_time :STRING,update\_time
8. :STRING,deleted :STRING> COMMENT '数据',
9. `old`  MAP<STRING,STRING> COMMENT '旧值'
10. ) COMMENT '用户章节进度表'
11. PARTITIONED **BY** (`dt` STRING)
12. ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.JsonSerDe'
13. LOCATION '/warehouse/edu/ods/ods\_user\_chapter\_process\_inc/';

装载语句

### 2.1.23用户表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| user\_info\_inc | 维度 | 增量 | ods\_user\_info\_inc | 用户表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_user\_info\_inc;
2. **CREATE** EXTERNAL **TABLE** ods\_user\_info\_inc
3. (
4. `type` STRING COMMENT '变动类型',
5. `ts`   **BIGINT** COMMENT '变动时间',
6. `data` STRUCT<id :STRING,login\_name :STRING,nick\_name :STRING,passwd :STRING,**name** :STRING,phone\_num :STRING,email
7. :STRING,head\_img :STRING,user\_level :STRING,birthday :STRING,gender :STRING,create\_time :STRING,operate\_time
8. :STRING,status :STRING,real\_name :STRING> COMMENT '数据',
9. `old`  MAP<STRING,STRING> COMMENT '旧值'
10. ) COMMENT '用户表'
11. PARTITIONED **BY** (`dt` STRING)
12. ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.JsonSerDe'
13. LOCATION '/warehouse/edu/ods/ods\_user\_info\_inc/';

装载语句

### 2.1.24视频表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| video\_info\_full | 维度 | 全量 | ods\_video\_info\_full | 视频表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_video\_info\_full;
2. **CREATE** EXTERNAL **TABLE** ods\_video\_info\_full
3. (
4. `id`              STRING COMMENT '编号',
5. `video\_name`      STRING COMMENT '视频名称',
6. `during\_sec`      **BIGINT** COMMENT '时长',
7. `video\_status`    STRING COMMENT '上传状态',
8. `video\_size`      **BIGINT** COMMENT '视频大小',
9. `video\_url`       STRING COMMENT '存储路径',
10. `video\_source\_id` STRING COMMENT '云端资源编号',
11. `version\_id`      STRING COMMENT '版本号',
12. `chapter\_id`      STRING COMMENT '章节id',
13. `course\_id`       STRING COMMENT '课程id',
14. `publisher\_id`    STRING COMMENT '发布者id',
15. `create\_time`     STRING COMMENT '创建时间',
16. `update\_time`     STRING COMMENT '更新时间',
17. `deleted`         STRING COMMENT '是否删除'
18. ) COMMENT '视频表'
19. PARTITIONED **BY** (`dt` STRING)
20. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
21. NULL DEFINED **AS** ''
22. LOCATION '/warehouse/edu/ods/ods\_video\_info\_full/';

装载语句

### 2.1.25VIP变化表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HDFS表名 | 维度/事实 | 全量/增量 | ODS表名 | 描述 |
| vip\_change\_detail\_inc | 事实 | 增量 | ods\_vip\_change\_detail\_inc | VIP变化表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_vip\_change\_detail\_inc;
2. **CREATE** EXTERNAL **TABLE** ods\_vip\_change\_detail\_inc
3. (
4. `type` STRING COMMENT '变动类型',
5. `ts`   **BIGINT** COMMENT '变动时间',
6. `data` STRUCT<id :STRING,user\_id :STRING,from\_vip :STRING,to\_vip :STRING,create\_time :STRING> COMMENT '数据',
7. `old`  MAP<STRING,STRING> COMMENT '旧值'
8. ) COMMENT 'VIP 变化表'
9. PARTITIONED **BY** (`dt` STRING)
10. ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.JsonSerDe'
11. LOCATION '/warehouse/edu/ods/ods\_vip\_change\_detail\_inc/';

装载语句

## 2.2DIM层

### 2.2.1地区维度表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 维度信息 | 维度表类型 | 同步类型 | 表名 | 描述 |
| 地区维度 | 全量表 |  | dim\_province\_full | 地区维度表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dim\_province\_full;
2. **CREATE** EXTERNAL **TABLE** dim\_province\_full
3. (
4. `id`            STRING COMMENT '编号',
5. `region\_id`     STRING COMMENT '大区id',
6. `province\_name` STRING COMMENT '省份名称',
7. `area\_code`     STRING COMMENT '行政区位码',
8. `iso\_code`      STRING COMMENT '国际编码',
9. `iso\_3166\_2`    STRING COMMENT 'ISO3166 编码'
10. ) COMMENT '地区维度表'
11. PARTITIONED **BY** (`dt` STRING)
12. STORED **AS** ORC
13. LOCATION '/warehouse/edu/dim/dim\_province\_full/'
14. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

首日

1. **insert** overwrite **table** ${APP}.dim\_province\_full partition(dt='$do\_date')
2. **select**
3. id,
4. **name**,
5. region\_id,
6. area\_code,
7. iso\_code,
8. iso\_3166\_2
9. **from** ${APP}.ods\_base\_province\_full
10. **where** dt='$do\_date';

每日

1. **insert** overwrite **table** ${APP}.dim\_province\_full partition(dt='$do\_date')
2. **select**
3. id,
4. region\_id,
5. **name**,
6. area\_code,
7. iso\_code,
8. iso\_3166\_2
9. **from** ${APP}.ods\_base\_province\_full
10. **where** dt='$do\_date';

### 2.2.2用户维度表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 维度信息 | 维度表类型 | 同步类型 | 表名 | 描述 |
| 用户维度 | 拉链表 |  | dim\_user\_zip | 用户维度表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dim\_user\_zip;
2. **CREATE** EXTERNAL **TABLE** dim\_user\_zip
3. (
4. `id`           STRING COMMENT '编号',
5. `login\_name`   STRING COMMENT '用户名称',
6. `nick\_name`    STRING COMMENT '用户昵称',
7. `real\_name`    STRING COMMENT '用户姓名',
8. `phone\_num`    STRING COMMENT '手机号码',
9. `email`        STRING COMMENT '邮箱',
10. `user\_level`   STRING COMMENT '用户等级',
11. `birthday`     STRING COMMENT '用户生日',
12. `gender`       STRING COMMENT '性别',
13. `create\_time`  STRING COMMENT '创建时间',
14. `operate\_time` STRING COMMENT '操作时间',
15. `start\_date`   STRING COMMENT '开始日期',
16. `end\_date`     STRING COMMENT '结束日期'
17. ) COMMENT '用户维度表'
18. PARTITIONED **BY** (`dt` STRING)
19. STORED **AS** ORC
20. LOCATION '/warehouse/edu/dim/dim\_user\_zip/'
21. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

首日

1. **insert** overwrite **table** ${APP}.dim\_user\_zip partition (dt='9999-12-31')
2. **select**
3. data.id,
4. data.login\_name,
5. data.nick\_name,
6. md5(data.real\_name),
7. md5(if(data.phone\_num regexp '^(13[0-9]|14[01456879]|15[0-35-9]|16[2567]|17[0-8]|18[0-9]|19[0-35-9])\\\d{8}$',data.phone\_num,null)),
8. md5(if(data.email regexp '^[a-zA-Z0-9\_-]+@[a-zA-Z0-9\_-]+(\\\.[a-zA-Z0-9\_-]+)+$',data.email,null)),
9. data.user\_level,
10. data.birthday,
11. data.gender,
12. data.create\_time,
13. data.operate\_time,
14. '2022-02-21' start\_date,
15. '9999-12-31' end\_date
16. **from** ${APP}.ods\_user\_info\_inc
17. **where** dt='$do\_date'
18. and type='bootstrap-insert';

每日

1. **insert** overwrite **table** ${APP}.dim\_user\_zip partition(dt)
2. **select**
3. id,
4. login\_name,
5. nick\_name,
6. real\_name,
7. phone\_num,
8. email,
9. user\_level,
10. birthday,
11. gender,
12. create\_time,
13. operate\_time,
14. start\_date,
15. if(rn=2,date\_sub('$do\_date',1),end\_date) end\_date,
16. if(rn=1,'9999-12-31',date\_sub('$do\_date',1)) dt
17. **from**
18. (
19. **select**
20. id,
21. login\_name,
22. nick\_name,
23. real\_name,
24. phone\_num,
25. email,
26. user\_level,
27. birthday,
28. gender,
29. create\_time,
30. operate\_time,
31. start\_date,
32. end\_date,
33. row\_number() over (partition **by** id **order** **by** start\_date **desc**) rn
34. **from**
35. (
36. **select**
37. id,
38. login\_name,
39. nick\_name,
40. real\_name,
41. phone\_num,
42. email,
43. user\_level,
44. birthday,
45. gender,
46. create\_time,
47. operate\_time,
48. start\_date,
49. end\_date
50. **from** ${APP}.dim\_user\_zip
51. **where** dt='9999-12-31'
52. **union**
53. **select**
54. id,
55. login\_name,
56. nick\_name,
57. md5(real\_name) **name**,
58. md5(if(phone\_num regexp '^(13[0-9]|14[01456879]|15[0-35-9]|16[2567]|17[0-8]|18[0-9]|19[0-35-9])\\\d{8}$',phone\_num,null)) phone\_num,
59. md5(if(email regexp '^[a-zA-Z0-9\_-]+@[a-zA-Z0-9\_-]+(\\\.[a-zA-Z0-9\_-]+)+$',email,null)) email,
60. user\_level,
61. birthday,
62. gender,
63. create\_time,
64. operate\_time,
65. '$do\_date' start\_date,
66. '9999-12-31' end\_date
67. **from**
68. (
69. **select**
70. data.id,
71. data.login\_name,
72. data.nick\_name,
73. data.real\_name,
74. data.phone\_num,
75. data.email,
76. data.user\_level,
77. data.birthday,
78. data.gender,
79. data.create\_time,
80. data.operate\_time,
81. row\_number() over (partition **by** data.id **order** **by** ts **desc**) rn
82. **from** ${APP}.ods\_user\_info\_inc
83. **where** dt='$do\_date'
84. )t1
85. **where** rn=1
86. )t2
87. )t3;

### 2.2.3课程维度表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 维度信息 | 维度表类型 | 同步类型 | 表名 | 描述 |
| 课程维度 | 全量表 |  | dim\_course\_info\_full | 课程维度表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dim\_course\_info\_full;
2. **CREATE** EXTERNAL **TABLE** dim\_course\_info\_full
3. (
4. `id`               STRING COMMENT '编号',
5. `course\_name`      STRING COMMENT '课程名称',
6. `subject\_id`       STRING COMMENT '学科id',
7. `teacher`          STRING COMMENT '讲师名称',
8. `chapter\_num`      STRING COMMENT '章节数',
9. `origin\_price`     **DECIMAL**(16, 2) COMMENT '价格',
10. `reduce\_amount`    **DECIMAL**(16, 2) COMMENT '优惠金额',
11. `actual\_price`     **DECIMAL**(16, 2) COMMENT '实际价格',
12. `course\_introduce` STRING COMMENT '课程介绍',
13. `create\_time`      STRING COMMENT '创建时间',
14. `update\_time`      STRING COMMENT '更新时间',
15. `subject\_name`     STRING COMMENT '科目名称',
16. `category\_id`      STRING COMMENT '分类',
17. `category\_name`    STRING COMMENT '分类名称'
18. ) COMMENT '课程维度表'
19. PARTITIONED **BY** (`dt` STRING)
20. STORED **AS** ORC
21. LOCATION '/warehouse/edu/dim/dim\_course\_full/'
22. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

首日

1. **insert** overwrite **table** ${APP}.dim\_course\_info\_full partition (dt = '$do\_date')
2. **select** course.id,
3. course\_name,
4. subject\_id,
5. teacher,
6. chapter\_num,
7. origin\_price,
8. reduce\_amount,
9. actual\_price,
10. course\_introduce,
11. create\_time,
12. update\_time,
13. subject\_name,
14. category\_id,
15. category\_name
16. **from** (
17. **select** id,
18. course\_name,
19. subject\_id,
20. teacher,
21. chapter\_num,
22. origin\_price,
23. reduce\_amount,
24. actual\_price,
25. course\_introduce,
26. create\_time,
27. update\_time
28. **from** ${APP}.ods\_course\_info\_full
29. **where** dt = '$do\_date') course
30. left join
31. (
32. **select** id,
33. subject\_name,
34. category\_id
35. **from** ${APP}.ods\_base\_subject\_info\_full
36. **where** dt = '$do\_date'
37. ) base
38. **on** course.subject\_id = base.id
39. left join
40. (
41. **select** id,
42. category\_name
43. **from** ${APP}.ods\_base\_category\_info\_full
44. **where** dt = '$do\_date'
45. ) category
46. **on** base.category\_id = category.id
47. **where** base.id **is** not null;

每日

1. **insert** overwrite **table** ${APP}.dim\_course\_info\_full partition (dt='$do\_date')
2. **select** course.id,
3. course\_name,
4. subject\_id,
5. teacher,
6. chapter\_num,
7. origin\_price,
8. reduce\_amount,
9. actual\_price,
10. course\_introduce,
11. create\_time,
12. update\_time,
13. subject\_name,
14. category\_id,
15. category\_name
16. **from** (
17. **select** id,
18. course\_name,
19. subject\_id,
20. teacher,
21. chapter\_num,
22. origin\_price,
23. reduce\_amount,
24. actual\_price,
25. course\_introduce,
26. create\_time,
27. update\_time
28. **from** ${APP}.ods\_course\_info\_full
29. **where** dt = '$do\_date') course
30. left join
31. (
32. **select** id,
33. subject\_name,
34. category\_id
35. **from** ${APP}.ods\_base\_subject\_info\_full
36. **where** dt = '$do\_date'
37. ) base
38. **on** course.subject\_id = base.id
39. left join
40. (
41. **select** id,
42. category\_name
43. **from** ${APP}.ods\_base\_category\_info\_full
44. **where** dt = '$do\_date'
45. ) category
46. **on** base.category\_id = category.id
47. **where** base.id **is** not null;

### 2.2.4章节维度表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 维度信息 | 维度表类型 | 同步类型 | 表名 | 描述 |
| 章节维度 | 全量表 |  | dim\_chapter\_full | 章节维度表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dim\_chapter\_full;
2. **CREATE** **TABLE** dim\_chapter\_full
3. (
4. `id`           STRING COMMENT '编号',
5. `chapter\_name` STRING COMMENT '章节名称',
6. `video\_id`     STRING COMMENT '视频id',
7. `is\_free`      STRING COMMENT '是否免费',
8. `video\_name`   STRING COMMENT '视频名称',
9. `create\_time`  STRING COMMENT '创建时间',
10. `update\_time`  STRING COMMENT '更新时间'
11. ) COMMENT '章节维度表'
12. PARTITIONED **BY** (`dt` STRING)
13. STORED **AS** ORC
14. LOCATION '/warehouse/edu/dim/dim\_chapter\_full/'
15. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

首日

1. **insert** overwrite **table** ${APP}.dim\_chapter\_full partition (dt = '$do\_date')
2. **select** chapter.id,
3. chapter\_name,
4. video\_id,
5. is\_free,
6. video\_name,
7. create\_time,
8. update\_time
9. **from** (
10. **select** id,
11. chapter\_name,
12. video\_id,
13. is\_free,
14. create\_time,
15. update\_time
16. **from** ${APP}.ods\_chapter\_info\_full
17. **where** dt = '$do\_date'
18. ) chapter
19. **full** outer join
20. (
21. **select** id,
22. video\_name
23. **from** ${APP}.ods\_video\_info\_full
24. **where** dt = '$do\_date'
25. ) video
26. **on** chapter.id = video.id;

每日

1. **insert** overwrite **table** ${APP}.dim\_chapter\_full partition (dt ='$do\_date')
2. **select**
3. a1.id,
4. chapter\_name,
5. video\_id,
6. is\_free,
7. video\_name,
8. create\_time,
9. update\_time
10. **from** ( **select**
11. id
12. ,chapter\_name
13. ,video\_id
14. ,is\_free
15. ,create\_time
16. ,update\_time
17. **from** ${APP}.ods\_chapter\_info\_full
18. **where** dt='$do\_date'
19. ) a1
20. left join (
21. **select**
22. id,
23. video\_name
24. **from** ${APP}.ods\_video\_info\_full
25. **where** dt='$do\_date'
26. ) a2 **on** a1.video\_id = a2.id;

### 2.2.5问题信息维度表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 维度信息 | 维度表类型 | 同步类型 | 表名 | 描述 |
| 问题维度 | 全量表 |  | dim\_question\_full | 问题信息维度表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dim\_question\_full;
2. **CREATE** EXTERNAL **TABLE** dim\_question\_full
3. (
4. `id`            STRING COMMENT '编号',
5. `question\_txt`  STRING COMMENT '题目内容',
6. `chapter\_id`    STRING COMMENT '章节id',
7. `course\_id`     STRING COMMENT '课程id',
8. `question\_type` STRING COMMENT '题目类型',
9. `create\_time`   STRING COMMENT '创建时间',
10. `update\_time`   STRING COMMENT '更新时间',
11. `option\_txt`    STRING COMMENT '选项内容',
12. `is\_correct`    STRING COMMENT '是否正确',
13. `question\_id`   STRING COMMENT '题目ID'
14. ) COMMENT '问题信息维度表'
15. PARTITIONED **BY** (`dt` STRING)
16. STORED **AS** ORC
17. LOCATION '/warehouse/edu/dim/dim\_question\_full/'
18. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

首日

1. **insert** overwrite **table** ${APP}.dim\_question\_full partition (dt = '$do\_date')
2. **select** question.id,
3. question\_txt,
4. chapter\_id,
5. course\_id,
6. question\_type,
7. create\_time,
8. update\_time,
9. option\_txt,
10. is\_correct,
11. question\_id
12. **from** (
13. **select** id,
14. question\_txt,
15. chapter\_id,
16. course\_id,
17. question\_type,
18. create\_time,
19. update\_time
20. **from** ${APP}.ods\_test\_question\_info\_full
21. **where** dt = '$do\_date'
22. ) question
23. left join
24. (
25. **select**
26. option\_txt,
27. is\_correct,
28. question\_id
29. **from** ${APP}.ods\_test\_question\_option\_full
30. **where** dt = '$do\_date'
31. ) **option**
32. **on** question.id = **option**.question\_id;

每日

1. **insert** overwrite **table** ${APP}.dim\_question\_full partition (dt = '$do\_date')
2. **select** question.id,
3. question\_txt,
4. chapter\_id,
5. course\_id,
6. question\_type,
7. create\_time,
8. update\_time,
9. option\_txt,
10. is\_correct,
11. question\_id
12. **from** (
13. **select** id,
14. question\_txt,
15. chapter\_id,
16. course\_id,
17. question\_type,
18. create\_time,
19. update\_time
20. **from** ${APP}.ods\_test\_question\_info\_full
21. **where** dt = '$do\_date'
22. ) question
23. **full** outer join
24. (
25. **select** id,
26. option\_txt,
27. is\_correct,
28. question\_id
29. **from** ${APP}.ods\_test\_question\_option\_full
30. **where** dt = '$do\_date'
31. ) **option**
32. **on** question.id = **option**.id;

### 2.2.6知识点维度表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 维度信息 | 维度表类型 | 同步类型 | 表名 | 描述 |
| 知识点维度 | 全量表 |  | dim\_test\_point\_question\_full | 知识点维度表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dim\_test\_point\_question\_full;
2. **CREATE** EXTERNAL **TABLE** dim\_test\_point\_question\_full
3. (
4. `id`           STRING COMMENT '编号',
5. `point\_id`     STRING COMMENT '知识点id',
6. `question\_id`  STRING COMMENT '问题id',
7. `publisher\_id` STRING COMMENT '发布者id',
8. `create\_time`  STRING COMMENT '创建时间',
9. `update\_time`  STRING COMMENT '更新时间',
10. `point\_txt`    STRING COMMENT '知识点内容',
11. `point\_level`  STRING COMMENT '知识点级别',
12. `course\_id`    STRING COMMENT '课程id',
13. `chapter\_id`   STRING COMMENT '章节id'
14. ) COMMENT '知识点维度表'
15. PARTITIONED **BY** (`dt` STRING)
16. STORED **AS** ORC
17. LOCATION '/warehouse/edu/dim/dim\_test\_point\_question\_full/'
18. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

首日

1. **insert** overwrite **table** ${APP}.dim\_test\_point\_question\_full partition (dt = '$do\_date')
2. **select**
3. a1.id
4. ,point\_id
5. ,question\_id
6. ,publisher\_id
7. ,create\_time
8. ,update\_time
9. ,point\_txt
10. ,point\_level
11. ,course\_id
12. ,chapter\_id
13. **from** (**select**
14. id,
15. point\_id,
16. question\_id
17. **from** ${APP}.ods\_test\_point\_question\_full) a1
18. right join
19. (
20. **select**
21. id
22. ,publisher\_id
23. ,create\_time
24. ,update\_time
25. ,point\_txt
26. ,point\_level
27. ,course\_id
28. ,chapter\_id
29. **from** ${APP}.ods\_knowledge\_point\_full
30. ) a2
31. **on** a1.point\_id= a2.id;

每日

1. **insert** overwrite **table** ${APP}.dim\_test\_point\_question\_full partition (dt = '$do\_date')
2. **select**
3. a1.id
4. ,point\_id
5. ,question\_id
6. ,publisher\_id
7. ,create\_time
8. ,update\_time
9. ,point\_txt
10. ,point\_level
11. ,course\_id
12. ,chapter\_id
13. **from** (**select**
14. id,
15. point\_id,
16. question\_id
17. **from** ${APP}.ods\_test\_point\_question\_full) a1
18. right join
19. (
20. **select**
21. id
22. ,publisher\_id
23. ,create\_time
24. ,update\_time
25. ,point\_txt
26. ,point\_level
27. ,course\_id
28. ,chapter\_id
29. **from** ${APP}.ods\_knowledge\_point\_full
30. ) a2
31. **on** a1.point\_id= a2.id;

### 2.2.7试卷维度表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 维度信息 | 维度表类型 | 同步类型 | 表名 | 描述 |
| 试卷维度 | 全量表 |  | dim\_paper\_full | 试卷维度表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dim\_paper\_full;
2. **CREATE** EXTERNAL **TABLE** dim\_paper\_full
3. (
4. `id`           STRING COMMENT '编号',
5. `paper\_title`  STRING COMMENT '试卷名称',
6. `course\_id`    STRING COMMENT '课程id',
7. `create\_time`  STRING COMMENT '创建时间',
8. `update\_time`  STRING COMMENT '更新时间',
9. `publisher\_id` STRING COMMENT '发布者id'
10. ) COMMENT '试卷维度表'
11. PARTITIONED **BY** (`dt` STRING)
12. STORED **AS** ORC
13. LOCATION '/warehouse/edu/dim/dim\_paper\_full/'
14. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

首日

1. **insert** overwrite **table** ${APP}.dim\_paper\_full partition (dt = '$do\_date')
2. **select** id,
3. paper\_title,
4. course\_id,
5. create\_time,
6. update\_time,
7. publisher\_id
8. **from** ${APP}.ods\_test\_paper\_full
9. **where** dt = '$do\_date';

每日

1. **insert** overwrite **table** ${APP}.dim\_paper\_full partition (dt = '$do\_date')
2. **select** id,
3. paper\_title,
4. course\_id,
5. create\_time,
6. update\_time,
7. publisher\_id
8. **from** ${APP}.ods\_test\_paper\_full
9. **where** dt = '$do\_date';

### 2.2.8时间维度表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 维度信息 | 维度表类型 | 同步类型 | 表名 | 描述 |
| 时间维度 | 全量表 |  | dim\_date | 时间维度表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dim\_date;
2. **CREATE** EXTERNAL **TABLE** dim\_date
3. (
4. `date\_id`    STRING COMMENT '日期ID',
5. `week\_id`    STRING COMMENT '周ID,一年中的第几周',
6. `week\_day`   STRING COMMENT '周几',
7. `day`        STRING COMMENT '每月的第几天',
8. `month`      STRING COMMENT '一年中的第几月',
9. `quarter`    STRING COMMENT '一年中的第几季度',
10. `year`       STRING COMMENT '年份',
11. `is\_workday` STRING COMMENT '是否是工作日',
12. `holiday\_id` STRING COMMENT '节假日'
13. ) COMMENT '时间维度表'
14. STORED **AS** ORC
15. LOCATION '/warehouse/edu/dim/dim\_date/'
16. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

1. **insert** overwrite **table** ${APP}.dim\_date
2. **select** \*
3. **from** ${APP}.tmp\_dim\_date\_info;

### 2.2.9时间维度表-临时表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 维度信息 | 维度表类型 | 同步类型 | 表名 | 描述 |
| 时间维度 |  |  | tmp\_dim\_date\_info | 时间临时表 |

建表语句

1. **DROP** **TABLE** IF EXISTS tmp\_dim\_date\_info;
2. **CREATE** EXTERNAL **TABLE** tmp\_dim\_date\_info
3. (
4. `date\_id`    STRING COMMENT '日',
5. `week\_id`    STRING COMMENT '周ID',
6. `week\_day`   STRING COMMENT '周几',
7. `day`        STRING COMMENT '每月的第几天',
8. `month`      STRING COMMENT '第几月',
9. `quarter`    STRING COMMENT '第几季度',
10. `year`       STRING COMMENT '年',
11. `is\_workday` STRING COMMENT '是否是工作日',
12. `holiday\_id` STRING COMMENT '节假日'
13. ) COMMENT '时间维度表'
14. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
15. LOCATION '/warehouse/edu/tmp/tmp\_dim\_date\_info/';

## 2.3DWD层

### 2.3.1互动域章节评论事务事实表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 业务过程 | 粒度 | 表名 | 描述 | 同步类型 |
|  | 增量表 | dwd\_interaction\_comment\_add\_inc | 互动域章节评论事务事实表 | 增量 |

建表语句

1. **DROP** **TABLE** IF EXISTS dwd\_interaction\_comment\_add\_inc;
2. **CREATE** EXTERNAL **TABLE** dwd\_interaction\_comment\_add\_inc
3. (
4. `id`          STRING COMMENT '编号',
5. `user\_id`     STRING COMMENT '用户id',
6. `chapter\_id`  STRING COMMENT '章节id',
7. `course\_id`   STRING COMMENT '课程id',
8. `date\_id`     STRING COMMENT '评论日期id',
9. `create\_time` STRING COMMENT '评论时间'
11. ) COMMENT '互动域章节评论事务事实表'
12. PARTITIONED **BY** (`dt` STRING)
13. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
14. STORED **AS** ORC
15. LOCATION '/warehouse/edu/dwd/dwd\_interaction\_comment\_add\_inc/'
16. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

每日

1. **insert** overwrite **table** ${APP}.dwd\_interaction\_comment\_add\_inc partition (dt)
2. **select** data.id,
3. data.user\_id,
4. data.chapter\_id,
5. data.course\_id,
6. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
7. data.create\_time,
8. date\_format(data.create\_time, 'yyyy-MM-dd')
9. **from** ${APP}.ods\_comment\_info\_inc
10. **where** dt = '$do\_date';

首日

1. **insert** overwrite **table** ${APP}.dwd\_interaction\_comment\_add\_inc partition (dt='$do\_date')
2. **select** data.id,
3. data.user\_id,
4. data.chapter\_id,
5. data.course\_id,
6. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
7. data.create\_time
8. **from** ${APP}.ods\_comment\_info\_inc
9. **where** dt = '$do\_date'
10. and type = 'insert';

### 2.3.2互动域课程评价事务事实表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 业务过程 | 粒度 | 表名 | 描述 | 同步类型 |
|  | 增量表 | dwd\_interaction\_review\_add\_inc | 互动域课程评价事务事实表 | 增量 |

建表语句

1. **DROP** **TABLE** IF EXISTS dwd\_interaction\_review\_add\_inc;
2. **CREATE** EXTERNAL **TABLE** dwd\_interaction\_review\_add\_inc
3. (
4. `id`           STRING COMMENT '编号',
5. `user\_id`      STRING COMMENT '用户id',
6. `course\_id`    STRING COMMENT '课程id',
7. `review\_stars` STRING COMMENT '评价，几星好评',
8. `date\_id`      STRING COMMENT '评价日期id',
9. `create\_time`  STRING COMMENT '评价时间'
11. ) COMMENT '互动域课程评价事务事实表'
12. PARTITIONED **BY** (`dt` STRING)
13. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
14. STORED **AS** ORC
15. LOCATION '/warehouse/edu/dwd/dwd\_interaction\_review\_add\_inc/'
16. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

每日

1. **insert** overwrite **table** ${APP}.dwd\_interaction\_review\_add\_inc partition (dt)
2. **select** data.id,
3. data.user\_id,
4. data.course\_id,
5. data.review\_stars,
6. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
7. data.create\_time,
8. dt
9. **from** ${APP}.ods\_review\_info\_inc
10. **where** dt = '$do\_date';

首日

1. **insert** overwrite **table** ${APP}.dwd\_interaction\_review\_add\_inc partition (dt='$do\_date')
2. **select** data.id,
3. data.user\_id,
4. data.course\_id,
5. data.review\_stars,
6. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
7. data.create\_time
8. **from** ${APP}.ods\_review\_info\_inc
9. **where** dt = '$do\_date'
10. and type = 'insert';

### 2.3.3互动域收藏事务事实表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 业务过程 | 粒度 | 表名 | 描述 | 同步类型 |
|  | 增量表 | dwd\_interaction\_favor\_add\_inc | 互动域收藏事务事实表 | 增量 |

建表语句

1. **DROP** **TABLE** IF EXISTS dwd\_interaction\_favor\_add\_inc;
2. **CREATE** EXTERNAL **TABLE** dwd\_interaction\_favor\_add\_inc
3. (
4. `id`          STRING COMMENT '编号',
5. `course\_id`   STRING COMMENT '课程id',
6. `user\_id`     STRING COMMENT '用户id',
7. `date\_id`     STRING COMMENT '收藏日期id',
8. `create\_time` STRING COMMENT '收藏时间'
10. ) COMMENT '互动域收藏事务事实表'
11. PARTITIONED **BY** (`dt` STRING)
12. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
13. STORED **AS** ORC
14. LOCATION '/warehouse/edu/dwd/dwd\_interaction\_favor\_add\_inc/'
15. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

每日

1. **insert** overwrite **table** ${APP}.dwd\_interaction\_favor\_add\_inc partition (dt)
2. **select** data.id,
3. data.course\_id,
4. data.user\_id,
5. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
6. data.create\_time,
7. dt
8. **from** ${APP}.ods\_favor\_info\_inc;

首日

1. **insert** overwrite **table** ${APP}.dwd\_interaction\_favor\_add\_inc partition (dt='$do\_date')
2. **select** data.id,
3. data.course\_id,
4. data.user\_id,
5. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
6. data.create\_time
7. **from** ${APP}.ods\_favor\_info\_inc
8. **where** type='insert';

### 2.3.4交易域加购事务事实表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 业务过程 | 粒度 | 表名 | 描述 | 同步类型 |
|  | 增量表 | dwd\_trade\_cart\_add\_inc | 交易域加购事务事实表 | 增量 |

建表语句

1. **DROP** **TABLE** IF EXISTS dwd\_trade\_cart\_add\_inc;
2. **CREATE** EXTERNAL **TABLE** dwd\_trade\_cart\_add\_inc
3. (
4. `id`          STRING COMMENT '编号',
5. `user\_id`     STRING COMMENT '用户id',
6. `course\_id`   STRING COMMENT '课程id',
7. `cart\_price`  **DECIMAL**(10, 2) COMMENT '放入购物车时价格',
8. `session\_id`  STRING COMMENT '会话id',
9. `date\_id`     STRING COMMENT '加购日期id',
10. `create\_time` STRING COMMENT '加购时间'
11. ) COMMENT '交易域章加购事务事实表'
12. PARTITIONED **BY** (`dt` STRING)
13. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
14. STORED **AS** ORC
15. LOCATION '/warehouse/edu/dwd/dwd\_trade\_cart\_add\_inc/'
16. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

每日

1. **insert** overwrite **table** ${APP}.dwd\_trade\_cart\_add\_inc partition (dt)
2. **select** id,
3. user\_id,
4. course\_id,
5. cart\_price,
6. session\_id,
7. date\_format(create\_time, 'yyyy-MM-dd') date\_id,
8. create\_time,
9. dt
10. **from** ${APP}.ods\_cart\_info\_full
11. **where** dt = '$do\_date';

首日

1. **insert** overwrite **table** ${APP}.dwd\_trade\_cart\_add\_inc partition (dt='$do\_date')
2. **select** data.id,
3. data.user\_id,
4. data.course\_id,
5. data.cart\_price,
6. data.session\_id,
7. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
8. data.create\_time
9. **from** ${APP}.ods\_cart\_info\_inc
10. **where** dt = '$do\_date'
11. and type = 'insert';

### 2.3.5交易域下单事务事实表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 业务过程 | 粒度 | 表名 | 描述 | 同步类型 |
|  | 增量表 | dwd\_trade\_order\_detail\_inc | 交易域下单事务事实表 | 增量 |

建表语句

1. **DROP** **TABLE** IF EXISTS dwd\_trade\_order\_detail\_inc;
2. **CREATE** EXTERNAL **TABLE** dwd\_trade\_order\_detail\_inc
3. (
4. `id`            STRING COMMENT '编号',
5. `course\_id`     STRING COMMENT '课程id',
6. `order\_id`      STRING COMMENT '订单id',
7. `user\_id`       STRING COMMENT '用户id',
8. `session\_id`    STRING COMMENT '会话 id',
9. `province\_id`   STRING COMMENT '地区id',
10. `date\_id`       STRING COMMENT '下单日期id',
11. `create\_time`   STRING COMMENT '下单时间',
12. `origin\_amount` **DECIMAL**(16, 2) COMMENT '原始金额'
13. ) COMMENT '交易域下单明细事务事实表'
14. PARTITIONED **BY** (`dt` STRING)
15. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
16. STORED **AS** ORC
17. LOCATION '/warehouse/edu/dwd/dwd\_trade\_order\_detail\_inc/'
18. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

每日

1. **insert** overwrite **table** ${APP}.dwd\_trade\_order\_detail\_inc partition (dt)
2. **select** b.id,
3. course\_id,
4. order\_id,
5. b.user\_id,
6. session\_id,
7. province\_id,
8. date\_id,
9. create\_time,
10. nvl(origin\_amount, 0) origin\_amount,
11. dt
12. **from** (
13. **select** data.id,
14. data.user\_id,
15. data.session\_id,
16. data.province\_id
17. **from** ${APP}.ods\_order\_info\_inc
18. ) a
19. right join (
20. **select** data.id,
21. data.course\_id,
22. data.course\_name,
23. data.order\_id,
24. data.user\_id,
25. data.origin\_amount,
26. data.coupon\_reduce,
27. data.final\_amount,
28. data.create\_time,
29. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
30. data.update\_time,
31. date\_format(data.create\_time, 'yyyy-MM-dd') dt
32. **from** ${APP}.ods\_order\_detail\_inc
33. ) b
34. **on** a.user\_id = b.user\_id;

首日

1. **insert** overwrite **table** ${APP}.dwd\_trade\_order\_detail\_inc partition (dt='$do\_date')
2. **select** b.id,
3. course\_id,
4. order\_id,
5. b.user\_id,
6. session\_id,
7. province\_id,
8. date\_id,
9. create\_time,
10. nvl(origin\_amount, 0) origin\_amount
11. **from** (
12. **select** data.id,
13. data.user\_id,
14. data.session\_id,
15. data.province\_id
16. **from** ${APP}.ods\_order\_info\_inc
17. **where** type='insert'
18. ) a
19. right join (
20. **select** data.id,
21. data.course\_id,
22. data.course\_name,
23. data.order\_id,
24. data.user\_id,
25. data.origin\_amount,
26. data.coupon\_reduce,
27. data.final\_amount,
28. data.create\_time,
29. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
30. data.update\_time,
31. date\_format(data.create\_time, 'yyyy-MM-dd') dt
32. **from** ${APP}.ods\_order\_detail\_inc
33. **where** type='insert' and dt='$do\_date'
34. ) b
35. **on** a.user\_id = b.user\_id;

### 2.3.6交易域支付成功事务事实表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 业务过程 | 粒度 | 表名 | 描述 | 同步类型 |
|  | 增量表 | dwd\_trade\_pay\_detail\_suc\_inc | 交易域支付成功事务事实表 | 增量 |

建表语句

1. **DROP** **TABLE** IF EXISTS dwd\_trade\_pay\_detail\_suc\_inc;
2. **CREATE** EXTERNAL **TABLE** dwd\_trade\_pay\_detail\_suc\_inc
3. (
4. `id`                STRING COMMENT '编号',
5. `order\_id`          STRING COMMENT '订单id',
6. `user\_id`           STRING COMMENT '用户id',
7. `course\_id`         STRING COMMENT '课程id',
8. `payment\_type\_code` STRING COMMENT '支付类型编码',
9. `date\_id`           STRING COMMENT '支付日期id',
10. `callback\_time`     STRING COMMENT '支付成功时间',
11. `total\_amount`      **DECIMAL**(16, 2) COMMENT '支付金额'
12. ) COMMENT '交易域支付成功事务事实表'
13. PARTITIONED **BY** (`dt` STRING)
14. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
15. STORED **AS** ORC
16. LOCATION '/warehouse/edu/dwd/dwd\_trade\_pay\_detail\_suc\_inc/'
17. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

每日

1. **insert** overwrite **table** ${APP}.dwd\_trade\_pay\_detail\_suc\_inc partition (dt)
2. **select** id
3. , pay.order\_id
4. , user\_id
5. , course\_id
6. , payment\_type\_code
7. , date\_format(callback\_time,'yyyy-MM-dd') date\_id
8. , callback\_time
9. , total\_amount
10. , date\_format(callback\_time,'yyyy-MM-dd') dt
11. **from** (
12. **select** data.id,
13. data.course\_id,
14. data.user\_id,
15. data.order\_id,
16. data.final\_amount total\_amount
17. **from** ${APP}.ods\_order\_detail\_inc
18. **where**  dt = '$do\_date'
19. ) oi
20. join
21. (
22. **select** data.order\_id,
23. data.payment\_type payment\_type\_code,
24. data.callback\_time
25. **from** ${APP}.ods\_payment\_info\_inc
26. **where**  dt = '$do\_date'
27. and data.payment\_type='1102'
28. ) pay
29. **on** oi.order\_id = pay.order\_id;

首日

1. **insert** overwrite **table** ${APP}.dwd\_trade\_pay\_detail\_suc\_inc partition (dt='$do\_date')
2. **select** id
3. , pay.order\_id
4. , user\_id
5. , course\_id
6. , payment\_type\_code
7. , date\_format(callback\_time,'yyyy-MM-dd') date\_id
8. , callback\_time
9. , total\_amount
10. **from** (
11. **select** data.id,
12. data.course\_id,
13. data.user\_id,
14. data.order\_id,
15. data.final\_amount total\_amount
16. **from** ${APP}.ods\_order\_detail\_inc
17. **where**  dt = '$do\_date' and type='insert'
18. ) oi
19. join
20. (
21. **select** data.order\_id,
22. data.payment\_type payment\_type\_code,
23. data.callback\_time
24. **from** ${APP}.ods\_payment\_info\_inc
25. **where**  dt = '$do\_date' and type='insert'
26. and data.payment\_type='1102'
27. ) pay
28. **on** oi.order\_id = pay.order\_id;

### 2.3.7考试域答卷事务事实表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 业务过程 | 粒度 | 表名 | 描述 | 同步类型 |
|  | 增量表 | dwd\_exam\_paper\_exam\_inc | 考试域答卷事务事实表 | 增量 |

建表语句

1. **DROP** **TABLE** IF EXISTS dwd\_exam\_paper\_exam\_inc;
2. **CREATE** EXTERNAL **TABLE** dwd\_exam\_paper\_exam\_inc
3. (
4. `id`           STRING COMMENT '编号',
5. `user\_id`      STRING COMMENT '考生id',
6. `paper\_id`     STRING COMMENT '考卷id',
7. `date\_id`      STRING COMMENT '考试日期id',
8. `create\_time`  STRING COMMENT '创建时间',
9. `submit\_time`  STRING COMMENT '结束考试的时间',
10. `score`        **DECIMAL**(16, 2) COMMENT '考试分数',
11. `duration\_sec` STRING COMMENT '考试时长'
12. ) COMMENT '考试域答卷事务事实表'
13. PARTITIONED **BY** (`dt` STRING)
14. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
15. STORED **AS** ORC
16. LOCATION '/warehouse/edu/dwd/dwd\_exam\_paper\_exam\_inc/'
17. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

每日

1. **insert** overwrite **table** ${APP}.dwd\_exam\_paper\_exam\_inc partition (dt)
2. **select** data.id,
3. data.user\_id,
4. data.paper\_id,
5. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
6. data.create\_time,
7. data.submit\_time,
8. data.score,
9. data.duration\_sec,
10. dt
11. **from** ${APP}.ods\_test\_exam\_inc
12. **where** dt = '$do\_date';

首日

1. **insert** overwrite **table** ${APP}.dwd\_exam\_paper\_exam\_inc partition (dt='$do\_date')
2. **select** data.id,
3. data.user\_id,
4. data.paper\_id,
5. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
6. data.create\_time,
7. data.submit\_time,
8. data.score,
9. data.duration\_sec
10. **from** ${APP}.ods\_test\_exam\_inc
11. **where** dt = '$do\_date' and type = 'insert';

### 2.3.8考试域答题事务事实表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 业务过程 | 粒度 | 表名 | 描述 | 同步类型 |
|  | 增量表 | dwd\_exam\_question\_exam\_inc | 考试域答题事务事实表 | 增量 |

建表语句

1. **DROP** **TABLE** IF EXISTS dwd\_exam\_question\_exam\_inc;
2. **CREATE** EXTERNAL **TABLE** dwd\_exam\_question\_exam\_inc
3. (
4. `id`          STRING COMMENT '编号',
5. `exam\_id`     STRING COMMENT '考试id',
6. `paper\_id`    STRING COMMENT '试卷id',
7. `question\_id` STRING COMMENT '问题id',
8. `date\_id`     STRING COMMENT '考试日期id',
9. `create\_time` STRING COMMENT '创建时间',
10. `score`       **DECIMAL**(16, 2) COMMENT '本题得分'
11. ) COMMENT '考试域答题事务事实表'
12. PARTITIONED **BY** (`dt` STRING)
13. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
14. STORED **AS** ORC
15. LOCATION '/warehouse/edu/dwd/dwd\_exam\_question\_exam\_inc/'
16. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

每日

1. **insert** overwrite **table** ${APP}.dwd\_exam\_question\_exam\_inc partition(dt)
2. **select** data.id,
3. data.exam\_id,
4. data.paper\_id,
5. data.question\_id,
6. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
7. data.create\_time,
8. data.is\_correct,
9. data.score,
10. dt
11. **from** ${APP}.ods\_test\_exam\_question\_inc
12. **where** dt = '$do\_date';

首日

1. **insert** overwrite **table** ${APP}.dwd\_exam\_question\_exam\_inc partition (dt='$do\_date')
2. **select** data.id,
3. data.exam\_id,
4. data.paper\_id,
5. data.question\_id,
6. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
7. data.create\_time,
8. data.is\_correct,
9. data.score
10. **from** ${APP}.ods\_test\_exam\_question\_inc
11. **where** dt = '$do\_date' and type = 'insert';

### 2.3.9播放域用户章节观看事务事实表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 业务过程 | 粒度 | 表名 | 描述 | 同步类型 |
|  | 增量表 | dwd\_play\_user\_chapter\_process\_inc | 播放域用户章节观看事务事实表 | 增量 |

建表语句

1. **DROP** **TABLE** IF EXISTS dwd\_play\_user\_chapter\_process\_inc;
2. **CREATE** EXTERNAL **TABLE** dwd\_play\_user\_chapter\_process\_inc
3. (
4. `id`           STRING COMMENT '编号',
5. `user\_id`      STRING COMMENT '用户id',
6. `course\_id`    STRING COMMENT '课程id',
7. `chapter\_id`   STRING COMMENT '章节id',
8. `date\_id`      STRING COMMENT '观看日期id',
9. `create\_time`  STRING COMMENT '创建时间',
10. `position\_sec` STRING COMMENT '章节观看进度'
11. ) COMMENT '播放域用户章节观看事务事实表'
12. PARTITIONED **BY** (`dt` STRING)
13. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
14. STORED **AS** ORC
15. LOCATION '/warehouse/edu/dwd/dwd\_play\_user\_chapter\_process\_inc/'
16. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

每日

1. **insert** overwrite **table** ${APP}.dwd\_play\_user\_chapter\_process\_inc partition (dt)
2. **select** data.id,
3. data.user\_id,
4. data.course\_id,
5. data.chapter\_id,
6. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
7. data.create\_time,
8. data.position\_sec,
9. date\_format(data.create\_time, 'yyyy-MM-dd')
10. **from** ${APP}.ods\_user\_chapter\_process\_inc
11. **where** dt = '$do\_date';

首日

1. **insert** overwrite **table** ${APP}.dwd\_play\_user\_chapter\_process\_inc partition (dt='$do\_date')
2. **select** data.id,
3. data.user\_id,
4. data.course\_id,
5. data.chapter\_id,
6. date\_format(data.create\_time, 'yyyy-MM-dd') date\_id,
7. data.create\_time,
8. data.position\_sec
9. **from** ${APP}.ods\_user\_chapter\_process\_inc
10. **where** dt = '$do\_date'
11. and type = 'insert';

### 2.3.10播放域视频播放事务事实表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 业务过程 | 粒度 | 表名 | 描述 | 同步类型 |
|  | 增量表 | dwd\_play\_video\_play\_inc | 播放域视频播放事务事实表 | 增量 |

建表语句

1. **DROP** **TABLE** IF EXISTS dwd\_play\_video\_play\_inc;
2. **CREATE** **TABLE** dwd\_play\_video\_play\_inc
3. (
4. `user\_id`      STRING COMMENT '用户id',
5. `course\_id`    STRING COMMENT '课程id',
6. `chapter\_id`   STRING COMMENT '章节id',
7. `video\_id`     STRING COMMENT '视频id',
8. `play\_sec`     STRING COMMENT '播放时长',
9. `position\_sec` STRING COMMENT '播放进度',
10. `data\_id`      STRING COMMENT '播放日期'
11. ) COMMENT '播放域视频播放事务事实表'
12. PARTITIONED **BY** (`dt` STRING)
13. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
14. STORED **AS** ORC
15. LOCATION '/warehouse/edu/dwd/dwd\_play\_video\_play\_inc/'
16. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

每日

1. **insert** overwrite **table** ${APP}.dwd\_play\_video\_play\_inc partition (dt)
2. **select** video\_id
3. , t1.user\_id
4. , t1.chapter\_id
5. , course\_id
6. , video\_time
7. , date\_format(from\_utc\_timestamp(ts, 'GMT+8'), 'yyyy-MM-dd') watch\_data
8. , watch\_progress
9. , play\_duration
10. , dt
11. **from** (
12. (**select** video\_id
13. , user\_id
14. , course\_id
15. , chapter\_id
16. , video\_time
17. , watch\_progress
18. , play\_duration
19. , ts
20. , dt
21. **from** (
22. (**select** dt,
23. id,--视屏编号
24. chapter\_id, --章节id
25. course\_id, --课程id
26. during\_sec video\_time --视屏总时长
27. **from** ${APP}.ods\_video\_info\_full
28. **where** dt = '$do\_date') ovi
29. right join
30. (**select** appvideo.play\_sec     play\_duration,--播放时长
31. appvideo.position\_sec watch\_progress, --播放进度
32. appvideo.video\_id,--视频id
33. common.uid            user\_id,
34. ts
35. **from** ${APP}.ods\_log\_inc
36. **where** dt = '$do\_date'
37. and appvideo.play\_sec **is** not null
38. and err.error\_code **is** null) oli
39. **on** ovi.id = oli.video\_id
40. )
41. ) t1);

首日

1. **insert** overwrite **table** ${APP}.dwd\_play\_video\_play\_inc partition (dt='$do\_date')
2. **select** video\_id
3. , t1.user\_id
4. , t1.chapter\_id
5. , course\_id
6. , video\_time
7. , date\_format(from\_utc\_timestamp(ts, 'GMT+8'), 'yyyy-MM-dd') watch\_data
8. , watch\_progress
9. , play\_duration
10. **from** (
11. (**select** video\_id
12. , user\_id
13. , course\_id
14. , chapter\_id
15. , video\_time
16. , watch\_progress
17. , play\_duration
18. , ts
19. **from** (
20. (**select**
21. id,--视屏编号
22. chapter\_id, --章节id
23. course\_id, --课程id
24. during\_sec video\_time --视屏总时长
25. **from** ${APP}.ods\_video\_info\_full
26. **where** dt = '$do\_date') ovi
27. right join
28. (**select** appvideo.play\_sec     play\_duration,--播放时长
29. appvideo.position\_sec watch\_progress, --播放进度
30. appvideo.video\_id,--视频id
31. common.uid            user\_id,
32. ts
33. **from** ${APP}.ods\_log\_inc
34. **where** dt = '$do\_date'
35. and appvideo.play\_sec **is** not null
36. and err.error\_code **is** null) oli
37. **on** ovi.id = oli.video\_id
38. )
39. ) t1);

### 2.3.11用户域用户登录事务事实表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 业务过程 | 粒度 | 表名 | 描述 | 同步类型 |
|  | 增量表 | dwd\_user\_login\_inc | 用户域用户登录事务事实表 | 增量 |

建表语句

1. **DROP** **TABLE** IF EXISTS dwd\_user\_login\_inc;
2. **CREATE** EXTERNAL **TABLE** dwd\_user\_login\_inc
3. (
4. `user\_id`        STRING COMMENT '用户ID',
5. `date\_id`        STRING COMMENT '日期ID',
6. `login\_time`     STRING COMMENT '登录时间',
7. `channel`        STRING COMMENT '应用下载渠道',
8. `province\_id`    STRING COMMENT '省份id',
9. `version\_code`   STRING COMMENT '应用版本',
10. `mid\_id`         STRING COMMENT '设备id',
11. `brand`          STRING COMMENT '设备品牌',
12. `model`          STRING COMMENT '设备型号',
13. `operate\_system` STRING COMMENT '设备操作系统'
14. ) COMMENT '用户域用户登录事务事实表'
15. PARTITIONED **BY** (`dt` STRING)
16. STORED **AS** ORC
17. LOCATION '/warehouse/edu/dwd/dwd\_user\_login\_inc/'
18. TBLPROPERTIES ("orc.compress" = "snappy");

装载语句

每日

1. **insert** overwrite **table** ${APP}.dwd\_user\_login\_inc partition (dt)
2. **select** user\_id
3. , date\_format(from\_utc\_timestamp(ts, 'GMT+8'), 'yyyy-MM-dd')          date\_id
4. , date\_format(from\_utc\_timestamp(ts, 'GMT+8'), 'yyyy-MM-dd HH:mm:ss') login\_time
5. , channel
6. , province\_id
7. , version\_code
8. , mid\_id
9. , brand
10. , model
11. , operate\_system
12. , date\_format(from\_utc\_timestamp(ts, 'GMT+8'), 'yyyy-MM-dd')
13. **from** (**select** user\_id,
14. channel,
15. province\_id,
16. version\_code,
17. mid\_id,
18. brand,
19. model,
20. operate\_system,
21. ts
22. **from** (**select** user\_id,
23. channel,
24. province\_id,
25. version\_code,
26. mid\_id,
27. brand,
28. model,
29. operate\_system,
30. ts,
31. row\_number() over (partition **by** session\_id **order** **by** ts) rn
32. **from** (**select** user\_id,
33. channel,
34. province\_id,
35. version\_code,
36. mid\_id,
37. brand,
38. model,
39. operate\_system,
40. ts,
41. concat(mid\_id, '-',
42. last\_value(session\_start\_point, **true**)
43. over (partition **by** mid\_id **order** **by** ts)) session\_id
44. **from** (**select** common.uid                              user\_id,
45. common.ch                               channel,
46. common.ar                               province\_id,
47. common.vc                               version\_code,
48. common.mid                              mid\_id,
49. common.os                               operate\_system,
50. common.md                               model,
51. common.ba                               brand,
52. ts,
53. if(page.last\_page\_id **is** null, ts, null) session\_start\_point
54. **from** ${APP}.ods\_log\_inc
55. **where** dt = '$do\_date'
56. and page **is** not null) t1) t2
57. **where** user\_id **is** not null) t3
58. **where** rn = 1) t4;

首日

1. **insert** overwrite **table** ${APP}.dwd\_user\_login\_inc partition (dt='$do\_date')
2. **select** user\_id
3. , date\_format(from\_utc\_timestamp(ts, 'GMT+8'), 'yyyy-MM-dd')          date\_id
4. , date\_format(from\_utc\_timestamp(ts, 'GMT+8'), 'yyyy-MM-dd HH:mm:ss') login\_time
5. , channel
6. , province\_id
7. , version\_code
8. , mid\_id
9. , brand
10. , model
11. , operate\_system
12. **from** (**select** user\_id,
13. channel,
14. province\_id,
15. version\_code,
16. mid\_id,
17. brand,
18. model,
19. operate\_system,
20. ts
21. **from** (**select** user\_id,
22. channel,
23. province\_id,
24. version\_code,
25. mid\_id,
26. brand,
27. model,
28. operate\_system,
29. ts,
30. row\_number() over (partition **by** session\_id **order** **by** ts) rn
31. **from** (**select** user\_id,
32. channel,
33. province\_id,
34. version\_code,
35. mid\_id,
36. brand,
37. model,
38. operate\_system,
39. ts,
40. concat(mid\_id, '-',
41. last\_value(session\_start\_point, **true**)
42. over (partition **by** mid\_id **order** **by** ts)) session\_id
43. **from** (**select** common.uid                              user\_id,
44. common.ch                               channel,
45. common.ar                               province\_id,
46. common.vc                               version\_code,
47. common.mid                              mid\_id,
48. common.os                               operate\_system,
49. common.md                               model,
50. common.ba                               brand,
51. ts,
52. if(page.last\_page\_id **is** null, ts, null) session\_start\_point
53. **from** ${APP}.ods\_log\_inc
54. **where** dt = '$do\_date'
55. and page **is** not null) t1) t2
56. **where** user\_id **is** not null) t3
57. **where** rn = 1) t4;

### 2.3.12用户域用户注册事务事实表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 业务过程 | 粒度 | 表名 | 描述 | 同步类型 |
|  | 增量表 | dwd\_user\_register\_inc | 用户域用户注册事务事实表 | 增量 |

建表语句

1. **DROP** **TABLE** IF EXISTS dwd\_user\_register\_inc;
2. **CREATE** EXTERNAL **TABLE** dwd\_user\_register\_inc
3. (
4. `user\_id`     STRING COMMENT '用户ID',
5. `date\_id`     STRING COMMENT '日期ID',
6. `create\_time` STRING COMMENT '注册时间'
7. ) COMMENT '用户域用户注册事务事实表'
8. PARTITIONED **BY** (`dt` STRING)
9. STORED **AS** ORC
10. LOCATION '/warehouse/edu/dwd/dwd\_user\_register\_inc/'
11. TBLPROPERTIES ("orc.compress" = "snappy");

装载语句

每日

1. **insert** overwrite **table** ${APP}.dwd\_user\_register\_inc partition (dt)
2. **select** ui.user\_id,
3. date\_format(create\_time, 'yyyy-MM-dd') date\_id,
4. create\_time,
5. date\_format(create\_time, 'yyyy-MM-dd')
7. **from** (**select** data.id user\_id,
8. data.create\_time
9. **from** ${APP}.ods\_user\_info\_inc
10. **where** dt = '$do\_date'
11. and type = 'insert') ui
12. left join
13. (**select** common.ar  province\_id,
14. common.uid user\_id
15. **from** ${APP}.ods\_log\_inc
16. **where** dt = '$do\_date'
17. and common.uid **is** not null) log
18. **on** ui.user\_id = log.user\_id;

首日

1. **insert** overwrite **table** ${APP}.dwd\_user\_register\_inc partition (dt='$do\_date')
2. **select** ui.user\_id,
3. date\_format(create\_time, 'yyyy-MM-dd') date\_id,
4. create\_time
5. **from** (**select** data.id user\_id,
6. data.create\_time
7. **from** ${APP}.ods\_user\_info\_inc
8. **where** dt = '$do\_date'
9. and type = 'insert') ui
10. left join
11. (**select** common.ar  province\_id,
12. common.uid user\_id
13. **from** ${APP}.ods\_log\_inc
14. **where** dt = '$do\_date'
15. and common.uid **is** not null) log
16. **on** ui.user\_id = log.user\_id;

### 2.3.13流量域页面浏览

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 业务过程 | 粒度 | 表名 | 描述 | 同步类型 |
|  | 增量表 | dwd\_traffic\_page\_view\_inc | 流量域页面浏览 | 增量 |

建表语句

1. **DROP** **TABLE** IF EXISTS dwd\_traffic\_page\_view\_inc;
2. **CREATE** EXTERNAL **TABLE** dwd\_traffic\_page\_view\_inc
3. (
4. `province\_name`  STRING COMMENT '省份',
5. `brand`          STRING COMMENT '手机品牌',
6. `channel`        STRING COMMENT '渠道',
7. `is\_new`         STRING COMMENT '是否首次启动',
8. `model`          STRING COMMENT '手机型号',
9. `mid\_id`         STRING COMMENT '设备id',
10. `operate\_system` STRING COMMENT '操作系统',
11. `user\_id`        STRING COMMENT '会员id',
12. `version\_code`   STRING COMMENT 'app版本号',
13. `page\_item`      STRING COMMENT '目标id ',
14. `page\_item\_type` STRING COMMENT '目标类型',
15. `last\_page\_id`   STRING COMMENT '上页类型',
16. `page\_id`        STRING COMMENT '页面ID ',
17. `source\_type`    STRING COMMENT '目标类型',
18. `date\_id`        STRING COMMENT '日期id',
19. `view\_time`      STRING COMMENT '跳入时间',
20. `session\_id`     STRING COMMENT '所属会话id',
21. `during\_time`    **BIGINT** COMMENT '持续时间毫秒'
22. ) COMMENT '流量域页面浏览日志表'
23. PARTITIONED **BY** (`dt` STRING)
24. STORED **AS** ORC
25. LOCATION '/warehouse/edu/dwd/dwd\_traffic\_page\_view\_inc'
26. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

每日

1. **insert** overwrite **table** ${APP}.dwd\_traffic\_page\_view\_inc partition (dt)
2. **select** province\_id,
3. brand,
4. channel,
5. is\_new,
6. model,
7. mid\_id,
8. operate\_system,
9. user\_id,
10. version\_code,
11. page\_item,
12. page\_item\_type,
13. last\_page\_id,
14. page\_id,
15. source\_type,
16. date\_format(from\_utc\_timestamp(ts, 'GMT+8'), 'yyyy-MM-dd')                                        date\_id,
17. date\_format(from\_utc\_timestamp(ts, 'GMT+8'), 'yyyy-MM-dd HH:mm:ss')                               view\_time,
18. concat(mid\_id, '-', last\_value(session\_start\_point, **true**) over (partition **by** mid\_id **order** **by** ts)) session\_id,
19. during\_time,
20. date\_format(from\_utc\_timestamp(ts, 'GMT+8'), 'yyyy-MM-dd')                                        date\_id

23. **from** (**select** common.ar                               province\_id,
24. common.ba                               brand,
25. common.ch                               channel,
26. common.is\_new                           is\_new,
27. common.md                               model,
28. common.mid                              mid\_id,
29. common.os                               operate\_system,
30. common.uid                              user\_id,
31. common.vc                               version\_code,
32. common.sc                               source\_type,
33. page.during\_time,
34. page.item                               page\_item,
35. page.item\_type                          page\_item\_type,
36. page.last\_page\_id,
37. page.page\_id,
38. ts,
39. if(page.last\_page\_id **is** null, ts, null) session\_start\_point
40. **from** ${APP}.ods\_log\_inc
41. **where** dt = '$do\_date'
42. and page **is** not null) log;

首日

1. **insert** overwrite **table** ${APP}.dwd\_traffic\_page\_view\_inc partition (dt='$do\_date')
2. **select** province\_id,
3. brand,
4. channel,
5. is\_new,
6. model,
7. mid\_id,
8. operate\_system,
9. user\_id,
10. version\_code,
11. page\_item,
12. page\_item\_type,
13. last\_page\_id,
14. page\_id,
15. source\_type,
16. date\_format(from\_utc\_timestamp(ts, 'GMT+8'), 'yyyy-MM-dd')                                        date\_id,
17. date\_format(from\_utc\_timestamp(ts, 'GMT+8'), 'yyyy-MM-dd HH:mm:ss')                               view\_time,
18. concat(mid\_id, '-', last\_value(session\_start\_point, **true**) over (partition **by** mid\_id **order** **by** ts)) session\_id,
19. during\_time

22. **from** (**select** common.ar                               province\_id,
23. common.ba                               brand,
24. common.ch                               channel,
25. common.is\_new                           is\_new,
26. common.md                               model,
27. common.mid                              mid\_id,
28. common.os                               operate\_system,
29. common.uid                              user\_id,
30. common.vc                               version\_code,
31. common.sc                               source\_type,
32. page.during\_time,
33. page.item                               page\_item,
34. page.item\_type                          page\_item\_type,
35. page.last\_page\_id,
36. page.page\_id,
37. ts,
38. if(page.last\_page\_id **is** null, ts, null) session\_start\_point
39. **from** ${APP}.ods\_log\_inc
40. **where** dt = '$do\_date'
41. and page **is** not null) log;

## 2.4DWS层

### 2.4.1最近1日汇总表

#### 2.4.1.1交易域用户课程粒度订单最近1日汇总表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| 1天 | 用户课程 | 汇总最近一日的用户课程粒度的订单 | dws\_trade\_user\_course\_order\_1d | 交易域用户课程粒度订单最近1日汇总表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_trade\_user\_course\_order\_1d;
2. **CREATE** EXTERNAL **TABLE** dws\_trade\_user\_course\_order\_1d(
3. `user\_id`             STRING COMMENT '用户id',
4. `course\_id`             STRING COMMENT '课程id',
5. `course\_name`  STRING COMMENT '课程名称',
6. `order\_count\_1d`           **BIGINT** COMMENT '最近1日下单数',
7. `order\_user\_num\_1d`              **BIGINT** COMMENT '最近1日下单人数',
8. `order\_total\_amount\_1d` **DECIMAL**(16, 2) COMMENT '最近1日下单最终金额'
9. )COMMENT '交易域用户课程粒度订单最近1日汇总事实表'
10. PARTITIONED **BY** (`dt` STRING)
11. STORED **AS** ORC
12. LOCATION '/warehouse/edu/dws/dws\_trade\_user\_course\_order\_1d'
13. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

1. **set** hive.**exec**.**dynamic**.partition.mode=nonstrict;
2. **insert** overwrite **table** dws\_trade\_user\_course\_order\_1d partition(dt)
3. **select**
4. user\_id,
5. course\_id ,
6. course\_name ,
7. order\_count\_1d ,
8. order\_user\_num\_1d ,
9. order\_total\_amount\_1d,
10. dt
11. **from**
12. (
13. **select**
14. dt,
15. user\_id,
16. course\_id ,
17. count(\*)  order\_count\_1d ,
18. count(**distinct**(user\_id)) order\_user\_num\_1d ,
19. sum(origin\_amount) order\_total\_amount\_1d
20. **from** dwd\_trade\_order\_detail\_inc
21. **GROUP** **BY** dt,user\_id,course\_id
22. )od
23. left join
24. (
25. **select**
26. id,
27. course\_name
28. **from** dim\_course\_info\_full
29. **GROUP** **BY** id,course\_name
30. )cour
31. **on** od.course\_id=cour.id;

#### 2.4.1.2交易域用户粒度订单最近1日汇总表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| 1天 | 用户 | 汇总最近一日的用户粒度的订单 | dws\_trade\_user\_order\_1d | 交易域用户粒度订单最近1日汇总表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_trade\_user\_order\_1d;
2. **CREATE** EXTERNAL **TABLE** dws\_trade\_user\_order\_1d
3. (
4. `user\_id`                   STRING COMMENT '用户id',
5. `order\_count\_1d`            **BIGINT** COMMENT '最近1日下单次数',
6. `order\_total\_amount\_1d`     **DECIMAL**(16, 2) COMMENT '最近1日下单金额'
7. ) COMMENT '交易域用户粒度订单最近1日汇总事实表'
8. PARTITIONED **BY** (`dt` STRING)
9. STORED **AS** ORC
10. LOCATION '/warehouse/edu/dws/dws\_trade\_user\_order\_1d';

装载语句

1. **insert** overwrite **table** dws\_trade\_user\_order\_1d partition(dt)
2. **select**
3. user\_id,
4. count(**distinct**(order\_id)),
5. sum(origin\_amount),
6. dt
7. **from** dwd\_trade\_order\_detail\_inc
8. **group** **by** user\_id,dt;

#### 2.4.1.3考试域用户试卷粒度考试最近1日汇总表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| 1天 | 用户 | 汇总最近一日的用户的考试信息 | dws\_exam\_user\_paper\_exam\_1d | 考试域用户试卷粒度考试最近1日汇总表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_exam\_user\_paper\_exam\_1d;
2. **CREATE** EXTERNAL **TABLE** dws\_exam\_user\_paper\_exam\_1d
3. (
4. `user\_id`   STRING COMMENT '用户id',
5. `paper\_id` STRING COMMENT '试卷id',
6. `paper\_title` STRING COMMENT '试卷名称',
7. `course\_id` STRING COMMENT '课程id',
8. `course\_name` STRING COMMENT '课程名称',
9. `score\_1d` **DECIMAL**(10, 2) COMMENT '最近1日试卷分数',
10. `duration\_sec\_1d` **BIGINT** COMMENT '最近1日考试时长'
11. ) COMMENT '考试域用户试卷粒度最近1日汇总表'
12. PARTITIONED **BY**(`dt` STRING)
13. STORED **AS** ORC
14. LOCATION '/warehouse/edu/dws/dws\_exam\_user\_paper\_exam\_1d/'
15. TBLPROPERTIES('orc.compress' = 'snappy');

装载语句

1. **insert** overwrite **table** dws\_exam\_user\_paper\_exam\_1d partition(dt)
2. **select**
3. epe.user\_id,
4. epe.paper\_id,
5. tp.paper\_title,
6. tp.course\_id,
7. ci.course\_name,
8. epe.score\_1d,
9. epe.duration\_sec\_1d,
10. dt
11. **from**
12. (
13. **select**
14. dt,
15. user\_id,
16. paper\_id,
17. sum(score) score\_1d,
18. sum(duration\_sec) duration\_sec\_1d
19. **from** dwd\_exam\_paper\_exam\_inc
20. **where** dt='2022-02-21'
21. **group** **by** user\_id,paper\_id
22. ) epe
23. left join
24. (
25. **select**
26. id,
27. paper\_title,
28. course\_id
29. **from** dim\_paper\_full
30. **where** dt='2022-02-21'
31. ) tp
32. **on** epe.paper\_id=tp.id
33. left join
34. (
35. **select**
36. id,
37. course\_name
38. **from** dim\_course\_info\_full
39. **where** dt='2022-02-21'
40. ) ci
41. **on** tp.course\_id=ci.id;

#### 2.4.1.4交易域用户粒度支付最近1日汇总表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| 1天 | 用户 | 汇总最近一日的用户粒度的支付信息 | dws\_trade\_user\_payment\_1d | 交易域用户粒度支付最近1日汇总表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_trade\_user\_payment\_1d;
2. **CREATE** EXTERNAL **TABLE** dws\_trade\_user\_payment\_1d
3. (
4. `user\_id`           STRING COMMENT '用户id',
5. `payment\_count\_1d`  **BIGINT** COMMENT '最近1日支付次数',
6. `payment\_amount\_1d` **DECIMAL**(16, 2) COMMENT '最近1日支付金额'
7. ) COMMENT '交易域用户粒度支付最近1日汇总事实表'
8. PARTITIONED **BY** (`dt` STRING)
9. STORED **AS** ORC
10. LOCATION '/warehouse/edu/dws/dws\_trade\_user\_payment\_1d'
11. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

1. **insert** overwrite **table** dws\_trade\_user\_payment\_1d partition(dt)
2. **select**
3. user\_id,
4. count(**distinct**(order\_id)),
5. sum(payment\_amount),
6. dt
7. **from** dwd\_trade\_pay\_detail\_suc\_inc
8. **group** **by** user\_id,dt;

#### 2.4.1.5交易域用户粒度加购最近1日汇总表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| 1天 | 用户 | 汇总最近一日的用户粒度的订单加购信息 | dws\_trade\_user\_cart\_add\_1d | 交易域用户粒度加购最近1日汇总表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_trade\_user\_cart\_add\_1d;
2. **CREATE** EXTERNAL **TABLE** dws\_trade\_user\_cart\_add\_1d
3. (
4. `user\_id`           STRING COMMENT '用户id',
5. `cart\_add\_count\_1d` **BIGINT** COMMENT '最近1日加购次数'
6. ) COMMENT '交易域用户粒度加购最近1日汇总事实表'
7. PARTITIONED **BY** (`dt` STRING)
8. STORED **AS** ORC
9. LOCATION '/warehouse/edu/dws/dws\_trade\_user\_cart\_add\_1d'
10. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

1. **insert** overwrite **table** dws\_trade\_user\_cart\_add\_1d partition(dt)
2. **select**
3. user\_id,
4. count(\*),
5. dt
6. **from** dwd\_trade\_cart\_add\_inc
7. **group** **by** user\_id,dt;

#### 2.4.1.6播放域用户视频粒度观看一日汇总表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| 1天 | 用户 | 汇总最近一日的用户观看视频的信息 | dws\_user\_video\_play\_1d | 播放域用户视频粒度观看一日汇总表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_user\_video\_play\_1d;
2. **CREATE** EXTERNAL **TABLE** dws\_user\_video\_play\_1d
3. (
4. `user\_id`                STRING COMMENT '用户id',
5. `course\_id`              STRING COMMENT '课程id',
6. `course\_name`            STRING COMMENT '课程名称',
7. `chapter\_id`             STRING COMMENT '章节id',
8. `chapter\_name`              STRING COMMENT '章节名称',
9. `video\_id`               STRING COMMENT '视频id',
10. `during\_video\_time\_1d`   **BIGINT** COMMENT '最近一日视频播放时长',
11. `view\_video\_count\_1d`    **BIGINT** COMMENT '最近一日视频播放次数',
12. `view\_video\_progress\_1d` **DECIMAL**(16, 2)COMMENT '最近一日视频播放进度'
13. )COMMENT '--播放域用户视频粒度观看一日汇总表'
14. PARTITIONED **BY** (`dt` STRING)
15. STORED **AS** ORC
16. LOCATION '/warehouse/edu/dws/dws\_user\_video\_play\_1d'
17. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

1. **insert** overwrite **table** dws\_user\_video\_play\_1d partition(dt='2022-02-21')
2. **select**
3. t1.user\_id,
4. t1.course\_id,
5. t2.course\_name,
6. t1.chapter\_id,
7. t3.chapter\_name,
8. t1.video\_id,
9. during\_video\_time\_1d,
10. view\_video\_count\_1d,
11. view\_video\_progress\_1d
12. **from**
13. (**select**
14. user\_id,
15. course\_id,
16. chapter\_id,
17. video\_id,
18. sum(play\_sec) during\_video\_time\_1d,
19. count(\*)  view\_video\_count\_1d,
20. **max**(position\_sec) view\_video\_progress\_1d
21. **from**
22. dwd\_play\_video\_play\_inc
23. **group** **by** dt,user\_id, video\_id,course\_id,chapter\_id) t1
24. left join(
25. **select**
26. id,
27. course\_name
28. **from**
29. dim\_course\_info\_full
30. )t2
31. **on** t1.course\_id=t2.id
32. left join (
33. **select**
34. id,
35. chapter\_name
36. **from**
37. dim\_chapter\_full
38. ) t3
39. **on** t1.chapter\_id=t3.id;

#### 2.4.1.7流量域会话粒度页面浏览最近1日汇总表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| 1天 | 会话 | 汇总最近一日每个会话浏览页面的信息 | dws\_traffic\_session\_page\_view\_1d | 流量域会话粒度页面浏览最近1日汇总表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_traffic\_session\_page\_view\_1d;
2. **CREATE** EXTERNAL **TABLE** dws\_traffic\_session\_page\_view\_1d
3. (
4. `session\_id`     STRING COMMENT '会话id',
5. `mid\_id`         string comment '设备id',
6. `brand`          string comment '手机品牌',
7. `model`          string comment '手机型号',
8. `operate\_system` string comment '操作系统',
9. `version\_code`   string comment 'app版本号',
10. `channel`        string comment '渠道',
11. `source\_type`    string comment '来源类型',
12. `final\_amount`    string comment '总最终金额',
13. `during\_time\_1d` **BIGINT** COMMENT '最近1日访问时长',
14. `page\_count\_1d`  **BIGINT** COMMENT '最近1日访问页面数'
15. ) COMMENT '流量域会话粒度页面浏览最近1日汇总表'
16. PARTITIONED **BY** (`dt` STRING)
17. STORED **AS** ORC
18. LOCATION '/warehouse/edu/dws/dws\_traffic\_session\_page\_view\_1d'
19. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

1. **insert** overwrite **table** dws\_traffic\_session\_page\_view\_1d partition(dt='2022-02-21')
2. **select**
3. session\_id,
4. mid\_id,
5. brand,
6. model,
7. operate\_system,
8. version\_code,
9. channel,
10. a.source\_type,
11. sum(b1.final\_amount),
12. sum(during\_time),
13. count(\*)
14. **from**
15. (
16. **select**
17. user\_id,
18. mid\_id,
19. brand,
20. model,
21. operate\_system,
22. version\_code,
23. channel,
24. source\_type,
25. during\_time
26. **from** dwd\_traffic\_page\_view\_inc
27. **where** dt='2022-02-21'
28. )a
29. left join
30. (
31. **select**
32. source\_site source\_type
33. **from** ods\_base\_source\_full
34. **where** dt='2022-02-21'
35. )b
36. **on** a.source\_type=b.source\_type
37. left join
38. (
39. **select**
40. data.user\_id     ,
41. data.session\_id ,
42. data.final\_amount   final\_amount
43. **from** ods\_order\_info\_inc
44. **where** dt='2022-02-21'
45. )b1
46. **on** a.user\_id=b1.user\_id
47. **where** session\_id **is** not null
48. **group** **by** session\_id,mid\_id,brand,model,operate\_system,version\_code,channel,a.source\_type;

#### 2.4.1.8流量域访客页面粒度页面浏览最近1日汇总事实表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| 1天 | 用户 | 汇总最近一日的访客粒度的页面浏览信息 | dws\_traffic\_page\_visitor\_page\_view\_1d | 流量域访客页面粒度页面浏览最近1日汇总事实表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_traffic\_page\_visitor\_page\_view\_1d;
2. **CREATE** EXTERNAL **TABLE** dws\_traffic\_page\_visitor\_page\_view\_1d
3. (
4. `mid\_id`         STRING COMMENT '访客id',
5. `brand`          string comment '手机品牌',
6. `model`          string comment '手机型号',
7. `operate\_system` string comment '操作系统',
8. `page\_id`        STRING COMMENT '页面id',
9. `during\_time\_1d` **BIGINT** COMMENT '最近1日浏览时长',
10. `view\_count\_1d`  **BIGINT** COMMENT '最近1日访问次数'
11. ) COMMENT '流量域访客页面粒度页面浏览最近1日汇总事实表'
12. PARTITIONED **BY** (`dt` STRING)
13. STORED **AS** ORC
14. LOCATION '/warehouse/edu/dws/dws\_traffic\_page\_visitor\_page\_view\_1d'
15. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

1. **insert** overwrite **table** dws\_traffic\_page\_visitor\_page\_view\_1d partition(dt='2022-02-21')
2. **select**
3. mid\_id,
4. brand,
5. model,
6. operate\_system,
7. page\_id,
8. sum(during\_time),
9. count(\*)
10. **from** dwd\_traffic\_page\_view\_inc
11. **where** dt='2022-02-21'
12. **group** **by** mid\_id,brand,model,operate\_system,page\_id;

### 2.4.2最近n日汇总表

#### 2.4.2.1交易域用户课程粒度订单最近n日汇总表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| n天 | 用户 | 汇总最近n日的用户购买课程的订单信息 | dws\_trade\_user\_course\_order\_nd | 交易域用户课程粒度订单最近n日汇总表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_trade\_user\_course\_order\_nd;
2. **CREATE** EXTERNAL **TABLE** dws\_trade\_user\_course\_order\_nd
3. (
4. `user\_id`               STRING COMMENT '用户id',
5. `course\_id`             STRING COMMENT '课程id',
6. `course\_name`  STRING COMMENT '课程名称',
7. `order\_count\_7d`           **BIGINT** COMMENT '最近7日下单数',
8. `order\_user\_num\_7d`              **BIGINT** COMMENT '最近7日下单人数',
9. `order\_total\_amount\_7d` **DECIMAL**(16, 2) COMMENT '最近7日下单最终金额',
10. `order\_count\_30d`           **BIGINT** COMMENT '最近30日下单数',
11. `order\_user\_num\_30d`              **BIGINT** COMMENT '最近30日下单人数',
12. `order\_total\_amount\_30d` **DECIMAL**(16, 2) COMMENT '最近30日下单最终金额'
13. ) COMMENT '交易域用户课程粒度订单最近n日汇总事实表'
14. PARTITIONED **BY** (`dt` STRING)
15. STORED **AS** ORC
16. LOCATION '/warehouse/edu/dws/dws\_trade\_user\_course\_order\_nd'
17. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

1. **insert** overwrite **table** dws\_trade\_user\_course\_order\_nd partition(dt)
2. **select**
3. dt,
4. user\_id ,
5. course\_id ,
6. course\_name ,
7. sum(if(dt>=date\_add('2022-02-21',-6),order\_count\_1d,0)),
8. sum(if(dt>=date\_add('2022-02-21',-6),order\_user\_num\_1d,0)),
9. sum(if(dt>=date\_add('2022-02-21',-6),order\_total\_amount\_1d,0)),
10. sum(order\_count\_1d),
11. sum(order\_user\_num\_1d),
12. sum(order\_total\_amount\_1d)
13. **from** dws\_trade\_user\_course\_order\_1d
14. **where** dt>=date\_add('2022-02-21',-29)
15. **group** **by**  dt,user\_id , course\_id , course\_name;

#### 2.4.2.2交易域用户粒度订单最近n日汇总表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| n天 | 用户 | 汇总最近n日的用户的订单信息 | dws\_trade\_user\_order\_nd | 交易域用户粒度订单最近n日汇总表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_trade\_user\_order\_nd;
2. **CREATE** EXTERNAL **TABLE** dws\_trade\_user\_order\_nd
3. (
4. `user\_id`                        STRING COMMENT '用户id',
5. `order\_count\_7d`             STRING COMMENT '最近7日下单次数',
6. `order\_total\_amount\_7d`      **DECIMAL**(16, 2) COMMENT '最近7日下单金额',
7. `order\_count\_30d`            **BIGINT** COMMENT '最近30日下单次数',
8. `order\_total\_amount\_30d`     **DECIMAL**(16, 2) COMMENT '最近30日下单金额'
9. ) COMMENT '交易域用户粒度订单最近n日汇总事实表'
10. PARTITIONED **BY** (`dt` STRING)
11. STORED **AS** ORC
12. LOCATION '/warehouse/gmall/edu/dws\_trade\_user\_order\_nd'
13. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

1. **insert** overwrite **table** dws\_trade\_user\_order\_nd partition(dt='2022-02-21')
2. **select**
3. user\_id,
4. sum(if(dt>=date\_add('2022-02-21',-6),order\_count\_1d,0)),
5. sum(if(dt>=date\_add('2022-02-21',-6),order\_total\_amount\_1d,0)),
6. sum(order\_count\_1d),
7. sum(order\_total\_amount\_1d)
8. **from** dws\_trade\_user\_order\_1d
9. **where** dt>=date\_add('2022-02-21',-29)
10. **group** **by** user\_id,dt;

#### 2.4.2.3考试域用户试卷粒度考试最近n日汇总表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| n天 | 用户试卷 | 汇总最近n日的用户做过的试卷的信息 | dws\_exam\_user\_paper\_exam\_nd | 考试域用户试卷粒度考试最近n日汇总表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_exam\_user\_paper\_exam\_nd;
2. **CREATE** EXTERNAL **TABLE** dws\_exam\_user\_paper\_exam\_nd
3. (
4. `user\_id`   STRING COMMENT '用户id',
5. `paper\_id` STRING COMMENT '试卷id',
6. `paper\_title` STRING COMMENT '试卷名称',
7. `course\_id` STRING COMMENT '课程id',
8. `course\_name` STRING COMMENT '课程名称',
9. `score\_7d` **DECIMAL**(10, 2) COMMENT '最近7日试卷分数',
10. `duration\_sec\_7d` **BIGINT** COMMENT '最近7日考试时长',
11. `score\_30d` **DECIMAL**(10, 2) COMMENT '最近30日试卷分数',
12. `duration\_sec\_30d` **BIGINT** COMMENT '最近30日考试时长'
13. ) COMMENT '考试域用户试卷粒度最近n日汇总表'
14. PARTITIONED **BY**(`dt` STRING)
15. STORED **AS** ORC
16. LOCATION '/warehouse/edu/dws/dws\_exam\_user\_paper\_exam\_nd/'
17. TBLPROPERTIES('orc.compress' = 'snappy');

装载语句

1. **insert** overwrite **table** dws\_exam\_user\_paper\_exam\_nd partition(dt='2022-02-21')
2. **select**
3. user\_id,
4. paper\_id,
5. paper\_title,
6. course\_id,
7. course\_name,
8. sum(`if`(dt>=date\_sub('2022-02-21', 6), score\_1d, 0)) score\_7d,
9. sum(`if`(dt>date\_sub('2022-02-21', 6), duration\_sec\_1d, 0)) duration\_sec\_7d,
10. sum(score\_1d) score\_30d,
11. sum(duration\_sec\_1d) duration\_sec\_30d
12. **from** dws\_exam\_user\_paper\_exam\_1d
13. **where** dt>=date\_sub('2022-02-21', 29)
14. **group** **by** user\_id,paper\_id,paper\_title;

#### 2.4.2.4播放域用户视频粒度观看n日汇总表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| n天 | 用户视频 | 汇总最近n日的用户观看视频的信息 | dws\_user\_video\_play\_nd | 播放域用户视频粒度观看N日汇总表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_user\_video\_play\_nd;
2. **CREATE** EXTERNAL **TABLE** dws\_user\_video\_play\_nd
3. (   `user\_id`                        STRING COMMENT '用户id',
4. `course\_id`                      STRING COMMENT '课程id',
5. `course\_name`                    STRING COMMENT '课程名称',
6. `chapter\_id`                     STRING COMMENT '章节id',
7. `chapter\_name`                   STRING COMMENT '章节名称',
8. `video\_id`                       STRING COMMENT '视频id',
9. `during\_video\_time\_7d`           **BIGINT** COMMENT  '最近7日视频播放时长',
10. `view\_video\_count\_7d`            **BIGINT** COMMENT  '最近7日视频播放次数',
11. `view\_video\_progress\_7d`         **DECIMAL**(16, 2) COMMENT  '最近7日视频播放进度',
12. `during\_video\_time\_30d`          **BIGINT** COMMENT '最近30日视频播放时长',
13. `view\_video\_count\_30d`           **BIGINT** COMMENT  '最近30日视频播放次数',
14. `view\_video\_progress\_30d`        **DECIMAL**(16, 2) COMMENT '最近30日视频播放进度'
15. ) COMMENT '播放域用户视频粒度观看N日汇总表'
16. PARTITIONED **BY** (`dt` STRING)
17. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
18. STORED **AS** ORC
19. LOCATION '/warehouse/edu/dws/dws\_user\_video\_play\_nd/'
20. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

1. **insert** overwrite **table** dws\_user\_video\_play\_nd partition(dt='2022-02-21')
2. **select**
3. user\_id,
4. course\_id,
5. course\_name,
6. chapter\_id,
7. chapter\_name,
8. video\_id,
9. sum(if(dt>=date\_sub('2022-02-21',6),during\_video\_time\_1d,0)),
10. sum(if(dt>=date\_sub('2022-02-21',6),view\_video\_count\_1d,0)),
11. **max**(view\_video\_progress\_1d),
12. sum(if(dt>=date\_sub('2022-02-21',29),during\_video\_time\_1d,0)),
13. sum(if(dt>=date\_sub('2022-02-21',29),view\_video\_count\_1d,0)),
14. **max**(view\_video\_progress\_1d)
15. **from** dws\_user\_video\_play\_1d
16. **where** dt>=date\_sub('2022-02-21',29)
17. **group** **by** user\_id,course\_id,course\_name,chapter\_id,chapter\_name,video\_id;

### 2.4.3历史至今汇总表

#### 2.4.3.1交易域用户粒度订单历史至今汇总表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| 历史至今 | 用户 | 汇总历史至今每个用户的所有订单信息 | dws\_trade\_user\_order\_td | 交易域用户粒度订单历史至今汇总表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_trade\_user\_order\_td;
2. **CREATE** EXTERNAL **TABLE** dws\_trade\_user\_order\_td
3. (
4. `user\_id`                   STRING COMMENT '用户id',
5. `order\_date\_first`          STRING COMMENT '首次下单日期',
6. `order\_date\_last`           STRING COMMENT '末次下单日期',
7. `order\_count\_td`            **BIGINT** COMMENT '下单次数',
8. `total\_amount\_td`           **DECIMAL**(16, 2) COMMENT '总金额'
9. ) COMMENT '交易域用户粒度订单历史至今汇总事实表'
10. PARTITIONED **BY** (`dt` STRING)
11. STORED **AS** ORC
12. LOCATION '/warehouse/edu/dws/dws\_trade\_user\_order\_td'
13. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

1. **insert** overwrite **table** dws\_trade\_user\_order\_td partition(dt='2022-02-21')
2. **select**
3. user\_id,
4. **min**(dt) create\_time,
5. **max**(dt) create\_time,
6. sum(order\_count\_1d) origin\_amount,
7. sum(order\_total\_amount\_1d) total\_amount
8. **from** dws\_trade\_user\_order\_1d
9. **group** **by** user\_id;

#### 2.4.3.2用户域用户粒度登录历史至今汇总表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 统计周期 | 统计粒度 | 业务过程 | 表名 | 描述 |
| 历史至今 | 用户 | 汇总历史至今每个用户登录的历史信息 | dws\_user\_user\_login\_td | 用户域用户粒度登录历史至今汇总表 |

建表语句

1. **DROP** **TABLE** IF EXISTS dws\_user\_user\_login\_td;
2. **CREATE** EXTERNAL **TABLE** dws\_user\_user\_login\_td
3. (
4. `user\_id`         STRING COMMENT '用户id',
5. `login\_date\_last` STRING COMMENT '末次登录日期',
6. `login\_count\_td`  **BIGINT** COMMENT '累计登录次数'
7. ) COMMENT '用户域用户粒度登录历史至今汇总事实表'
8. PARTITIONED **BY** (`dt` STRING)
9. STORED **AS** ORC
10. LOCATION '/warehouse/edu/dws/dws\_user\_user\_login\_td'
11. TBLPROPERTIES ('orc.compress' = 'snappy');

装载语句

1. **insert** overwrite **table** dws\_user\_user\_login\_td partition(dt='2022-02-21')
2. **select**
3. u.id,
4. nvl(login\_date\_last,date\_format(create\_time,'yyyy-MM-dd')),
5. nvl(login\_count\_td,1)
6. **from**
7. (
8. **select**
9. id,
10. create\_time
11. **from** dim\_user\_zip
12. **where** dt='9999-12-31'
13. )u
14. left join
15. (
16. **select**
17. user\_id,
18. **max**(dt) login\_date\_last,
19. count(\*) login\_count\_td
20. **from** dwd\_user\_login\_inc
21. **group** **by** user\_id
22. )l
23. **on** u.id=l.user\_id;

## 2.5ADS层

### 2.5.1流量主题

#### 2.5.1.1各来源流量统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 流量 | 来源 | 访客数  会话平均停留时长  会话平均浏览页面数  会话总数  跳出率 | ads\_traffic\_stats\_by\_source | 各来源流量统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_traffic\_stats\_by\_source;
2. **CREATE** EXTERNAL **TABLE** ads\_traffic\_stats\_by\_source
3. (
4. `dt`               STRING COMMENT '统计日期',
5. `recent\_days`      **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `source`           STRING COMMENT '来源',
7. `uv\_count`         **BIGINT** COMMENT '访客人数',
8. `avg\_duration\_sec` **BIGINT** COMMENT '会话平均停留时长，单位为秒',
9. `avg\_page\_count`   **BIGINT** COMMENT '会话平均浏览页面数',
10. `sv\_count`         **BIGINT** COMMENT '会话数',
11. `bounce\_rate`      **DECIMAL**(16, 2) COMMENT '跳出率'
12. ) COMMENT '各来源流量统计'
13. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
14. LOCATION '/warehouse/edu/ads/ads\_traffic\_stats\_by\_source/';

装载语句

1. **insert** overwrite **table** ads\_traffic\_stats\_by\_source
2. **select** \* **from** ads\_traffic\_stats\_by\_source
3. **union**
4. **select**
5. '2022-02-21' dt,
6. recent\_days,
7. source\_type source,
8. cast(count(**distinct**(mid\_id)) **as** **bigint**) uv\_count,
9. cast(avg(during\_time\_1d)/1000 **as** **bigint**) avg\_duration\_sec,
10. cast(avg(page\_count\_1d) **as** **bigint**) avg\_page\_count,
11. cast(count(\*) **as** **bigint**) sv\_count,
12. cast(sum(if(page\_count\_1d=1,1,0))/count(\*) **as** **decimal**(16,2)) bounce\_rate
13. **from** dws\_traffic\_session\_page\_view\_1d lateral **view** explode(array(1,7,30)) tmp **as** recent\_days
14. **where** dt>=date\_add('2022-02-21',-recent\_days+1)
15. **group** **by** recent\_days,source\_type;

#### 2.5.1.2路径分析

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 流量 | source-target | 页面浏览数 | ads\_page\_path | 路径分析 |

建表数据

1. **DROP** **TABLE** IF EXISTS ads\_page\_path;
2. **CREATE** EXTERNAL **TABLE** ads\_page\_path
3. (
4. `dt`          STRING COMMENT '统计日期',
5. `source`      STRING COMMENT '跳转起始页面ID',
6. `target`      STRING COMMENT '跳转终到页面ID',
7. `path\_count`  **BIGINT** COMMENT '跳转次数'
8. --     `页面浏览数`    BIGINT COMMENT '页面浏览数'
9. ) COMMENT '页面浏览路径分析'
10. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
11. LOCATION '/warehouse/edu/ads/ads\_page\_path/';

装载数据

1. **insert** overwrite **table** ads\_page\_path
2. **select** \* **from** ads\_page\_path
3. **union**
4. **select**
5. '2022-02-21' dt,
6. source,
7. nvl(target,'null'),
8. count(\*) path\_count
9. **from**
10. (
11. **select**
12. concat('step-',rn,':',page\_id) source,
13. concat('step-',rn+1,':',next\_page\_id) target
14. **from**
15. (
16. **select**
17. page\_id,
18. lead(page\_id,1,null) over(partition **by** session\_id **order** **by** view\_time) next\_page\_id,
19. row\_number() over (partition **by** session\_id **order** **by** view\_time) rn
20. **from** dwd\_traffic\_page\_view\_inc
21. **where** dt='2022-02-21'
22. )t1
23. )t2
24. **group** **by** source,target;

#### 2.5.1.3各来源下单统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 流量 | 来源 | 销售额  转化率 | ads\_order\_stats\_by\_source | 各来源下单统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_order\_stats\_by\_source;
2. **CREATE** EXTERNAL **TABLE** ads\_order\_stats\_by\_source
3. (
4. `dt`                      STRING COMMENT '统计日期',
5. `recent\_days`             **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `source`                  STRING COMMENT '来源',
7. `order\_amount`            **BIGINT** COMMENT '销售额',
8. `Conversion\_rate`         **BIGINT** COMMENT '转化率'
9. ) COMMENT '各来源下单统计'
10. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
11. LOCATION '/warehouse/edu/ads/ads\_order\_stats\_by\_source/';

装载语句

1. **insert** overwrite **table** ads\_order\_stats\_by\_source
2. **select** \* **from** ads\_order\_stats\_by\_source
3. **union**
4. **select**
5. '2022-02-21' dt,
6. recent\_days,
7. source\_type source,
8. total\_amount,
10. **from** dwd\_trade\_pay\_detail\_suc\_inc lateral **view** explode(array(1,7,30)) tmp **as** recent\_days
11. **where** dt>=date\_add('2022-02-21',-recent\_days+1)
12. **group** **by** recent\_days,source\_type;

### 2.5.2用户主题

#### 2.5.2.1用户变动统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 用户 |  | 流失用户数  回流用户数 | ads\_user\_change | 用户变动统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_user\_change;
2. **CREATE** EXTERNAL **TABLE** ads\_user\_change
3. (
4. `dt`               STRING COMMENT '统计日期',
5. `user\_churn\_count` **BIGINT** COMMENT '流失用户数',
6. `user\_back\_count`  **BIGINT** COMMENT '回流用户数'
7. ) COMMENT '用户变动统计'
8. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
9. LOCATION '/warehouse/edu/ads/ads\_user\_change/';

装载语句

1. **insert** overwrite **table** ads\_user\_change
2. **select** \* **from** ads\_user\_change
3. **union**
4. **select**
5. churn.dt,
6. user\_churn\_count,
7. user\_back\_count
8. **from**
9. (
10. **select**
11. '2022-02-21' dt,
12. count(\*) user\_churn\_count
13. **from** dws\_user\_user\_login\_td
14. **where** dt='2022-02-21'
15. and login\_date\_last=date\_add('2022-02-21',-7)
16. )churn
17. join
18. (
19. **select**
20. '2022-02-21' dt,
21. count(\*) user\_back\_count
22. **from**
23. (
24. **select**
25. user\_id,
26. login\_date\_last
27. **from** dws\_user\_user\_login\_td
28. **where** dt='2022-02-21'
29. )t1
30. join
31. (
32. **select**
33. user\_id,
34. login\_date\_last login\_date\_previous
35. **from** dws\_user\_user\_login\_td
36. **where** dt=date\_add('2022-02-21',-1)
37. )t2
38. **on** t1.user\_id=t2.user\_id
39. **where** datediff(login\_date\_last,login\_date\_previous)>=8
40. )back
41. **on** churn.dt=back.dt;

#### 2.5.2.2用户留存率

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 用户 |  | 留存率 | ads\_user\_retention | 用户留存率 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_user\_retention;
2. **CREATE** EXTERNAL **TABLE** ads\_user\_retention
3. (
4. `dt`              STRING COMMENT '统计日期',
5. `create\_date`     STRING COMMENT '用户新增日期',
6. `retention\_day`   **INT** COMMENT '截至当前日期留存天数',
7. `retention\_count` **BIGINT** COMMENT '留存用户数量',
8. `new\_user\_count`  **BIGINT** COMMENT '新增用户数量',
9. `retention\_rate`  **DECIMAL**(16, 2) COMMENT '留存率'
10. ) COMMENT '用户留存率'
11. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
12. LOCATION '/warehouse/edu/ads/ads\_user\_retention/';

装载语句

1. **insert** overwrite **table** ads\_user\_retention
2. **select** \* **from** ads\_user\_retention
3. **union**
4. **select**
5. '2022-02-21' dt,
6. login\_date\_first create\_date,
7. datediff('2022-02-21',login\_date\_first) retention\_day,
8. sum(if(login\_date\_last='2022-02-21',1,0)) retention\_count,
9. count(\*) new\_user\_count,
10. cast(sum(if(login\_date\_last='2022-02-21',1,0))/count(\*)\*100 **as** **decimal**(16,2)) retention\_rate
11. **from**
12. (
13. **select**
14. user\_id,
15. date\_id login\_date\_first
16. **from** dwd\_user\_register\_inc
17. **where** dt>=date\_add('2022-02-21',-7)
18. and dt<'2022-02-21'
19. )t1
20. join
21. (
22. **select**
23. user\_id,
24. login\_date\_last
25. **from** dws\_user\_user\_login\_td
26. **where** dt='2022-02-21'
27. )t2
28. **on** t1.user\_id=t2.user\_id
29. **group** **by** login\_date\_first;

#### 2.5.2.3用户新增活跃统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 用户 |  | 新增用户数  活跃用户数 | ads\_user\_stats | 用户新增活跃统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_user\_stats;
2. **CREATE** EXTERNAL **TABLE** ads\_user\_stats
3. (
4. `dt`                STRING COMMENT '统计日期',
5. `recent\_days`       **BIGINT** COMMENT '最近n日,1:最近1日,7:最近7日,30:最近30日',
6. `new\_user\_count`    **BIGINT** COMMENT '新增用户数',
7. `active\_user\_count` **BIGINT** COMMENT '活跃用户数'
8. ) COMMENT '用户新增活跃统计'
9. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
10. LOCATION '/warehouse/edu/ads/ads\_user\_stats/';

装载语句

1. **insert** overwrite **table** ads\_user\_stats
2. **select** \* **from** ads\_user\_stats
3. **union**
4. **select**
5. '2022-02-21' dt,
6. t1.recent\_days,
7. new\_user\_count,
8. active\_user\_count
9. **from**
10. (
11. **select**
12. recent\_days,
13. sum(if(login\_date\_last>=date\_add('2022-02-21',-recent\_days+1),1,0)) new\_user\_count
14. **from** dws\_user\_user\_login\_td lateral **view** explode(array(1,7,30)) tmp **as** recent\_days
15. **where** dt='2022-02-21'
16. **group** **by** recent\_days
17. )t1
18. join
19. (
20. **select**
21. recent\_days,
22. sum(if(date\_id>=date\_add('2022-02-21',-recent\_days+1),1,0)) active\_user\_count
23. **from** dwd\_user\_register\_inc lateral **view** explode(array(1,7,30)) tmp **as** recent\_days
24. **group** **by** recent\_days
25. )t2
26. **on** t1.recent\_days=t2.recent\_days;

#### 2.5.2.4用户行为漏斗分析

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 用户 |  | 首页浏览人数  课程详情页浏览人数  加购人数  下单人数  支付人数 | ads\_user\_action | 用户行为漏斗分析 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_user\_action;
2. **CREATE** EXTERNAL **TABLE** ads\_user\_action
3. (
4. `dt`                STRING COMMENT '统计日期',
5. `home\_count`        **BIGINT** COMMENT '浏览首页人数',
6. `good\_detail\_count` **BIGINT** COMMENT '浏览商品详情页人数',
7. `cart\_count`        **BIGINT** COMMENT '加入购物车人数',
8. `order\_count`       **BIGINT** COMMENT '下单人数',
9. `payment\_count`     **BIGINT** COMMENT '支付人数'
10. ) COMMENT '漏斗分析'
11. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
12. LOCATION '/warehouse/edu/ads/ads\_user\_action/';

装载语句

1. **insert** overwrite **table** ads\_user\_action
2. **select** \* **from** ads\_user\_action
3. **union**
4. **select**
5. '2022-02-21' dt,
6. home\_count,
7. good\_detail\_count,
8. cart\_count,
9. order\_count,
10. payment\_count
11. **from**
12. (
13. **select**
14. 1 recent\_days,
15. sum(if(page\_id='home',1,0)) home\_count,
16. sum(if(page\_id='good\_detail',1,0)) good\_detail\_count
17. **from** dws\_traffic\_page\_visitor\_page\_view\_1d
18. **where** dt='2022-02-21'
19. and page\_id in ('home','good\_detail')
20. )page
21. join
22. (
23. **select**
24. 1 recent\_days,
25. count(\*) cart\_count
26. **from** dws\_trade\_user\_cart\_add\_1d
27. **where** dt='2022-02-21'
28. )cart
29. **on** page.recent\_days=cart.recent\_days
30. join
31. (
32. **select**
33. 1 recent\_days,
34. count(\*) order\_count
35. **from** dws\_trade\_user\_order\_1d
36. **where** dt='2022-02-21'
37. )ord
38. **on** page.recent\_days=ord.recent\_days
39. join
40. (
41. **select**
42. 1 recent\_days,
43. count(\*) payment\_count
44. **from** dws\_trade\_user\_payment\_1d
45. **where** dt='2022-02-21'
46. )pay
47. **on** page.recent\_days=pay.recent\_days;

#### 2.5.2.5新增交易用户统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 用户 |  | 新增下单人数  新增支付人数 | ads\_new\_order\_user\_stats | 新增交易用户统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_new\_order\_user\_stats;
2. **CREATE** EXTERNAL **TABLE** ads\_new\_order\_user\_stats
3. (
4. `dt`                        STRING COMMENT '统计日期',
5. `recent\_days`               **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `new\_order\_user\_count`      **BIGINT** COMMENT '新增下单人数',
7. `new\_payment\_user\_count`    **BIGINT** COMMENT '新增支付人数'
8. ) COMMENT '新增交易用户统计'
9. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
10. LOCATION '/warehouse/edu/ads/ads\_new\_order\_user\_stats/';

装载语句

1. **insert** overwrite **table** ads\_new\_order\_user\_stats
2. **select** \* **from** ads\_new\_order\_user\_stats
3. **union**
4. **select**
5. '2022-02-21' dt,
6. recent\_days,
7. sum(if(order\_date\_first>=date\_add('2022-02-21',-recent\_days+1),1,0)) new\_order\_user\_count,
8. sum() new\_pay\_user\_count
9. **from** dws\_trade\_user\_order\_td lateral **view** explode(array(1,7,30)) tmp **as** recent\_days
10. **where** dt='2022-02-21'
11. **group** **by** recent\_days;

#### 2.5.2.6各年龄段下单用户数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 用户 | 年龄段 | 下单人数 | ads\_agegroup\_order\_user\_count | 各年龄段下单用户数 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_agegroup\_order\_user\_count;
2. **CREATE** EXTERNAL **TABLE** ads\_agegroup\_order\_user\_count
3. (
4. `dt`                   STRING COMMENT '统计日期',
5. `recent\_days`          **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `age\_group`            STRING COMMENT '年龄段',
7. `order\_user\_count`     **BIGINT** COMMENT '下单人数'
8. ) COMMENT '各年龄段下单用户数'
9. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
10. LOCATION '/warehouse/edu/ads/ads\_agegroup\_order\_user\_count/';

装载语句

1. **insert** overwrite **table** ${APP}.ads\_agegroup\_order\_user\_count
2. **select** \* **from** ${APP}.ads\_agegroup\_order\_user\_count
3. **union**
4. **select**
5. '$do\_date' dt,
6. recent\_days,
7. sum(order\_user\_count) user\_count,
8. age\_group
9. **from**(
10. **select** t2.recent\_days,
11. t2.user\_id,
12. order\_user\_count,
13. age\_group
14. **from** (**select** 1                       recent\_days,
15. user\_id,
16. count(**distinct** user\_id) order\_user\_count
17. **from** ${APP}.dws\_trade\_user\_order\_1d
18. **where** dt='$do\_date'
19. **group** **by** user\_id
20. **union** all
21. **select** recent\_days, user\_id, count(**distinct** (if(order\_count\_nd > 0, user\_id, null))) order\_user\_count
22. **from** (**select** recent\_days,
23. user\_id,
24. case recent\_days
25. **when** 7 **then** order\_count\_7d
26. **when** 30 **then** order\_count\_30d
27. **end** order\_count\_nd
28. **from** ${APP}.dws\_trade\_user\_order\_nd lateral **view** explode(array(7, 30)) tmp **as** recent\_days
29. **where** dt='$do\_date') t1
30. **group** **by** user\_id, recent\_days) t2
31. **inner** join
32. (**select** recent\_days,
33. user\_id,
34. age\_group,
35. count(**distinct** user\_id)
36. **from** (
37. **select** id user\_id,
38. recent\_days,
39. case
40. **when** (year(`**current\_date**`()) - year(birthday)) > 50 **then** '50岁以上'
41. **when** (year(`**current\_date**`()) - year(birthday)) > 40 **then** '40岁-50岁'
42. **when** (year(`**current\_date**`()) - year(birthday)) > 30 **then** '31岁-40岁'
43. **when** (year(`**current\_date**`()) - year(birthday)) > 20 **then** '21岁-30岁'
44. **when** (year(`**current\_date**`()) - year(birthday)) > 0 **then** '20岁以下'
45. **end** age\_group
46. **from**  ${APP}.dim\_user\_zip
47. lateral **view** explode(array(1, 7, 30)) tmp **as** recent\_days
48. **where** dt='$do\_date'
49. ) t3
50. **group** **by** recent\_days, age\_group, user\_id) t4
51. **on** t4.user\_id=t2.user\_id
52. )tt
53. **where** dt='$do\_date'
54. **group** **by** recent\_days, age\_group,dt;

### 2.5.3课程主题指标

#### 2.5.3.1各分类课程交易统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 课程 | 分类 | 下单数  下单人数  下单金额 | ads\_order\_stats\_by\_cate | 各分类课程交易统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_order\_stats\_by\_cate;
2. **CREATE** EXTERNAL **TABLE** ads\_order\_stats\_by\_cate
3. (
4. `dt`                      STRING COMMENT '统计日期',
5. `recent\_days`             **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `category\_id`             STRING COMMENT '分类id',
7. `category\_name`           STRING COMMENT '分类名称',
8. `order\_count`             **BIGINT** COMMENT '下单数',
9. `order\_user\_count`        **BIGINT** COMMENT '下单人数',
10. `order\_amount`            **BIGINT** COMMENT '下单金额'
11. ) COMMENT '各分类课程交易统计'
12. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
13. LOCATION '/warehouse/edu/ads/ads\_order\_stats\_by\_cate/';

装载语句

1. **insert** overwrite **table** ads\_order\_stats\_by\_cate
2. **select** dt
3. , recent\_days
4. , category\_id
5. , category\_name
6. , order\_count
7. , order\_user\_count
8. , order\_amount
9. **from** (
10. **select** dt,
11. category\_id,
12. category\_name,
13. sum(course\_id)             order\_count,
14. sum(order\_user\_num\_1d)     order\_user\_count,
15. sum(order\_total\_amount\_1d) order\_amount
16. **from** (
17. **select** course\_id,
18. order\_count\_1d,
19. order\_user\_num\_1d,
20. order\_total\_amount\_1d,
21. dt
22. **from** dws\_trade\_user\_course\_order\_1d
23. ) a1
24. left join (
25. **select** id, --课程id
26. category\_id,
27. category\_name
28. **from** dim\_course\_info\_full --课程维度表
30. ) b1 **on** a1.course\_id = b1.id
31. **group** **by** category\_id, category\_name, course\_id, dt
32. ) a2 lateral **view** explode(`array`(1, 7, 30)) tmp **as** recent\_days
34. **where** dt = "2022-02-21"
35. and dt >= date\_sub("2022-02-21", recent\_days - 1);

#### 2.5.3.2各学科课程交易统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 课程 | 学科 | 下单数  下单人数  下单金额 | ads\_order\_stats\_by\_sub | 各学科课程交易统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_order\_stats\_by\_sub;
2. **CREATE** EXTERNAL **TABLE** ads\_order\_stats\_by\_sub
3. (
4. `dt`                      STRING COMMENT '统计日期',
5. `recent\_days`             **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `subject\_id`              STRING COMMENT '学科id',
7. `subject\_name`            STRING COMMENT '学科名称',
8. `order\_count`             **BIGINT** COMMENT '下单数',
9. `order\_user\_count`        **BIGINT** COMMENT '下单人数',
10. `order\_amount`            **BIGINT** COMMENT '下单金额'
11. ) COMMENT '各分类学科交易统计'
12. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
13. LOCATION '/warehouse/edu/ads/ads\_order\_stats\_by\_sub/';

装载语句

1. **insert** overwrite **table** ads\_order\_stats\_by\_sub
2. **select** dt
3. , recent\_days
4. , subject\_id
5. , subject\_name
6. , order\_count
7. , order\_user\_count
8. , order\_amount
9. **from** (
10. **select** dt,
11. subject\_id, --学科id
12. subject\_name,
13. sum(course\_id)             order\_count,
14. sum(order\_user\_num\_1d)     order\_user\_count,
15. sum(order\_total\_amount\_1d) order\_amount
16. **from** (
17. **select** course\_id, --课程id
18. order\_count\_1d,
19. order\_user\_num\_1d,
20. order\_total\_amount\_1d,
21. dt
22. **from** dws\_trade\_user\_course\_order\_1d
23. ) a1
24. left join (
25. **select** subject\_id, --学科id
26. subject\_name,
27. kc.id       --课程id
28. **from** (
29. **select** id,          --课程id
30. subject\_id,  --学科id
31. category\_id, --分类id
32. category\_name
33. **from** dim\_course\_info\_full --课程维度表
34. ) kc
35. left join (
36. **select** id,         --科目id
37. subject\_name,
38. category\_id --分类id
39. **from** ods\_base\_subject\_info\_full
40. ) km
41. **on** kc.category\_id = km.category\_id
42. ) b1 **on** a1.course\_id = b1.id
43. **group** **by** subject\_id, subject\_name, course\_id, dt
44. ) a2 lateral **view** explode(`array`(1, 7, 30)) tmp **as** recent\_days
45. **where** dt = "2022-02-21"
46. and dt >= date\_sub("2022-02-21", recent\_days - 1);

#### 2.5.3.3各课程交易统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 课程 | 课程 | 下单数  下单人数  下单金额 | ads\_order\_stats\_by\_course | 各课程交易统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_order\_stats\_by\_course;
2. **CREATE** EXTERNAL **TABLE** ads\_order\_stats\_by\_course
3. (
4. `dt`                      STRING COMMENT '统计日期',
5. `recent\_days`             **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `course\_id`               STRING COMMENT '课程id',
7. `course\_name`             STRING COMMENT '课程名称',
8. `order\_count`             **BIGINT** COMMENT '下单数',
9. `order\_user\_count`        **BIGINT** COMMENT '下单人数',
10. `order\_amount`            **BIGINT** COMMENT '下单金额'
11. ) COMMENT '各分类学科交易统计'
12. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
13. LOCATION '/warehouse/edu/ads/ads\_order\_stats\_by\_course/';

装载语句

1. **insert** overwrite **table** ads\_order\_stats\_by\_course
2. **select** dt
3. , recent\_days
4. , course\_id
5. , course\_name
6. , order\_count
7. , order\_user\_count
8. , order\_amount
9. **from** (
10. **select** dt,
11. course\_id,
12. course\_name,
13. sum(course\_id)             order\_count,
14. sum(order\_user\_num\_1d)     order\_user\_count,
15. sum(order\_total\_amount\_1d) order\_amount
16. **from** dws\_trade\_user\_course\_order\_1d
17. **group** **by** course\_id, course\_name, dt
18. ) a2 lateral **view** explode(`array`(1, 7, 30)) tmp **as** recent\_days
20. **where** dt = "2022-02-21"
21. and dt >= date\_sub("2022-02-21", recent\_days - 1);

#### 2.5.3.4各课程评价统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 课程 | 课程 | 用户平均评分  评价用户数  好评率 | ads\_review\_stats | 各课程评价统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_review\_stats;
2. **CREATE** EXTERNAL **TABLE** ads\_review\_stats
3. (
4. `dt`                      STRING COMMENT '统计日期',
5. `recent\_days`             **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `review\_id`               STRING COMMENT '评价id',
7. `review\_txt`              STRING COMMENT '评价内容',
8. `review\_stars`            **BIGINT** COMMENT '评价',
9. `review\_stars\_avg`        **BIGINT** COMMENT '用户平均评分',
10. `review\_user\_count`       **BIGINT** COMMENT '评价用户数',
11. `favourite\_rate`          **BIGINT** COMMENT '好评率'      --给出5星评价的用户占总评价用户的比率
12. ) COMMENT '各分类学科交易统计'
13. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
14. LOCATION '/warehouse/edu/ads/ads\_review\_stats/';

装载语句

#### 2.5.3.5各分类课程试听留存统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 课程 | 分类 | 试听人数  试听留存率 | ads\_audition\_stats\_by\_cate | 各分类课程试听留存统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_audition\_stats\_by\_cate;
2. **CREATE** EXTERNAL **TABLE** ads\_audition\_stats\_by\_cate
3. (
4. `dt`                      STRING COMMENT '统计日期',
5. `recent\_days`             **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `category\_id`             STRING COMMENT '分类id',
7. `category\_name`           STRING COMMENT '分类名称',
8. `audition\_user\_count`     **BIGINT** COMMENT '试听人数',         --首次播放视频的人数
9. `audition\_retention`      **BIGINT** COMMENT '试听留存率'        --下单人次/试听人次
10. ) COMMENT '各分类课程试听留存统计'
11. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
12. LOCATION '/warehouse/edu/ads/ads\_audition\_stats\_by\_cate/';

装载语句

#### 2.5.3.6各学科试听留存统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 课程 | 学科 | 试听人数  试听留存率 | ads\_audition\_stats\_by\_subject | 各学科试听留存统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_audition\_stats\_by\_subject;
2. **CREATE** EXTERNAL **TABLE** ads\_audition\_stats\_by\_subject
3. (
4. `dt`                      STRING COMMENT '统计日期',
5. `recent\_days`             **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `subject\_id`              STRING COMMENT '学科id',
7. `subject\_name`            STRING COMMENT '学科名称',
8. `audition\_user\_count`     **BIGINT** COMMENT '试听人数',         --首次播放视频的人数
9. `audition\_retention`      **BIGINT** COMMENT '试听留存率'        --下单人次/试听人次
10. ) COMMENT '各分类课程试听留存统计'
11. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
12. LOCATION '/warehouse/edu/ads/ads\_audition\_stats\_by\_subject/';

装载语句

#### 2.5.3.7各课程试听留存统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 课程 | 课程 | 试听人数  试听留存率 | ads\_audition\_stats\_by\_course | 各课程试听留存统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_audition\_stats\_by\_course;
2. **CREATE** EXTERNAL **TABLE** ads\_audition\_stats\_by\_course
3. (
4. `dt`                      STRING COMMENT '统计日期',
5. `recent\_days`             **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `course\_id`               STRING COMMENT '课程id',
7. `course\_name`             STRING COMMENT '课程名称',
8. `audition\_user\_count`     **BIGINT** COMMENT '试听人数',         --首次播放视频的人数
9. `audition\_retention`      **BIGINT** COMMENT '试听留存率'        --下单人次/试听人次
10. ) COMMENT '各分类课程试听留存统计'
11. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
12. LOCATION '/warehouse/edu/ads/ads\_audition\_stats\_by\_course/';

装载语句

### 2.5.4交易主题指标

#### 2.5.4.1交易综合统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 交易 |  | 下单总额GMV  下单数  下单人数 | ads\_order\_stats | 交易综合统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_order\_stats;
2. **CREATE** EXTERNAL **TABLE** ads\_order\_stats
3. (
4. `dt`                 STRING COMMENT '统计日期',
5. `recent\_days`        **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `order\_count`        **BIGINT** COMMENT '下单数',
7. `order\_user\_count`   **BIGINT** COMMENT '下单人数',
8. `order\_total\_amount` **DECIMAL**(16, 2) COMMENT '下单总额'
9. ) COMMENT '交易综合统计'
10. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
11. LOCATION '/warehouse/edu/ads/ads\_order\_stats/';

装载语句

1. **insert** overwrite **table** ads\_order\_stats
2. **select** \* **from** ads\_order\_stats
3. **union**
4. **select** dt,
5. recent\_days,
6. count(order\_id)    order\_count,
7. count(user\_id)     order\_user\_count,
8. sum(origin\_amount) order\_total\_amount
9. **from** dwd\_trade\_order\_detail\_inc lateral **view** explode(`array`(1, 7, 30)) tmp **as** recent\_days
10. **group** **by** dt, order\_id
11. **having** dt = "2022-02-21"
12. and dt >= date\_sub("2022-02-21", recent\_days - 1);

#### 2.5.4.2各省份交易统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 交易 | 省份 | 下单数  下单人数  下单金额 | ads\_order\_by\_province | 各省份交易统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_order\_by\_province;
2. **CREATE** EXTERNAL **TABLE** ads\_order\_by\_province
3. (
4. `dt`                 STRING COMMENT '统计日期',
5. `recent\_days`        **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `province\_id`        STRING COMMENT '省份ID',
7. `province\_name`      STRING COMMENT '省份名称',
8. `area\_code`          STRING COMMENT '地区编码',
9. `iso\_code`           STRING COMMENT '国际标准地区编码',
10. `iso\_code\_3166\_2`    STRING COMMENT '国际标准地区编码',
11. `order\_count`        **BIGINT** COMMENT '下单数',
12. `order\_user\_count`   **BIGINT** COMMENT '下单人数',
13. `order\_total\_amount` **DECIMAL**(16, 2) COMMENT '下单金额'
14. ) COMMENT '各省份交易统计'
15. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
16. LOCATION '/warehouse/edu/ads/ads\_order\_by\_province/';

装载语句

1. **insert** overwrite **table** ads\_order\_by\_province
2. **select** \* **from** ads\_order\_by\_province
3. **union**
4. **select** dt
5. , recent\_days
6. , province\_id
7. , province\_name
8. , area\_code
9. , iso\_code
10. , iso\_code\_3166\_2
11. , order\_count
12. , order\_user\_count
13. , order\_total\_amount
14. **from** (
15. **select** dt,
16. a2.province\_id,
17. **name**       province\_name,
18. area\_code,
19. iso\_code,
20. iso\_3166\_2 iso\_code\_3166\_2,
21. order\_count,
22. order\_user\_count,
23. order\_total\_amount
24. **from** (
25. **select** a1.user\_id,
26. order\_count,
27. order\_total\_amount,
28. order\_user\_count,
29. province\_id --省份id
30. **from** (
31. **select** \*
32. **from** (
33. **select** user\_id,
34. order\_count\_1d        order\_count,
35. order\_total\_amount\_1d order\_total\_amount
36. **from** dws\_trade\_user\_order\_1d
37. **where** dt = "2022-02-21"
38. ) aa
39. join
40. (
41. **select** count(user\_id) order\_user\_count
42. **from** dws\_trade\_user\_order\_1d
43. **where** dt = "2022-02-21"
44. ) bb **on** 1 = 1
45. ) a1
46. left join (
47. **select** data.user\_id,
48. data.province\_id
49. **from** ods\_order\_info\_inc
50. **where** dt = "2022-02-21"
51. ) b1 **on** a1.user\_id = b1.user\_id
52. ) a2
53. left join (
54. **select** id
55. , **name**
56. , region\_id
57. , area\_code
58. , iso\_code
59. , iso\_3166\_2
60. , dt
61. **from** ods\_base\_province\_full
62. **where** dt = "2022-02-21"
63. ) b2
64. **on** a2.province\_id = b2.id
65. ) a3 lateral **view** explode(`array`(1, 7, 30)) tmp **as** recent\_days
66. **where** dt = "2022-02-21"
67. and dt >= date\_sub("2022-02-21", recent\_days - 1);

### 2.5.5考试主题指标

#### 2.5.5.1各试卷相关指标统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 考试 | 试卷 | 平均分  平均时长  用户数 | ads\_test\_paper\_statistics | 各试卷相关指标统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_test\_paper\_statistics;
2. **CREATE** EXTERNAL **TABLE** ads\_test\_paper\_statistics
3. (
4. `dt`                     STRING COMMENT '统计日期',
5. `recent\_days`            **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `test\_paper\_id`           STRING COMMENT '试卷id',
7. `test\_paper\_title`        STRING COMMENT '试卷名称',
8. `succ\_avg\_score`         **BIGINT** COMMENT '完成平均成绩',
9. `succ\_avg\_time`          **BIGINT** COMMENT '完成平均时长',
10. `succ\_user\_count`        **BIGINT** COMMENT '完成用户数'
11. ) COMMENT '各试卷相关指标统计'
12. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
13. LOCATION '/warehouse/edu/ads/ads\_test\_paper\_statistics/';

装载语句

1. **insert** overwrite **table** ads\_test\_paper\_statistics
2. **select** \* **from** ads\_test\_paper\_statistics
3. **union**
4. **select**
5. '2022-02-21' dt,
6. recent\_days,
7. test\_paper\_id,
8. test\_paper\_title,
9. cast(succ\_avg\_score **as** **decimal**(16,2)) succ\_avg\_score,
10. cast(succ\_avg\_time **as** **decimal**(16,2)) succ\_avg\_time,
11. succ\_user\_count
12. **from**
13. (
14. **select**
15. 1                        recent\_days,
16. paper\_id                 test\_paper\_id,
17. paper\_title              test\_paper\_title,
18. avg(score\_1d)            succ\_avg\_score,
19. avg(duration\_sec\_1d)     succ\_avg\_time,
20. count(**distinct**(user\_id)) succ\_user\_count
21. **from** dws\_exam\_user\_paper\_exam\_1d
22. **where** dt='2022-02-21'
23. **group** **by** paper\_id,paper\_title
24. **union** all
25. **select**
26. recent\_days,
27. paper\_id               test\_paper\_id,
28. paper\_title            test\_paper\_title,
29. avg(paper\_score)       succ\_avg\_score,
30. avg(paper\_duration)    succ\_avg\_time,
31. count(**distinct**(if(paper\_score>0,user\_id,null))) succ\_user\_count
32. **from**
33. (
34. **select**
35. recent\_days,
36. user\_id,
37. paper\_id,
38. paper\_title,
39. case recent\_days
40. **when** 7 **then** score\_7d
41. **when** 30 **then** score\_30d
42. **end** paper\_score,
43. case recent\_days
44. **when** 7 **then** duration\_sec\_7d
45. **when** 30 **then** duration\_sec\_30d
46. **end** paper\_duration
47. **from** dws\_exam\_user\_paper\_exam\_nd lateral **view** explode(array(7,30)) tmp **as** recent\_days
48. **where** dt='2022-02-21'
49. )t1
50. **group** **by** recent\_days,paper\_id,paper\_title
51. )t2;

#### 2.5.5.2各课程考试相关指标统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 考试 | 课程 | 平均分  平均时长  用户数 | ads\_course\_test\_statistics | 各课程考试相关指标统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_course\_test\_statistics;
2. **CREATE** EXTERNAL **TABLE** ads\_course\_test\_statistics
3. (
4. `dt`                     STRING COMMENT '统计日期',
5. `recent\_days`            **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `course\_id`              STRING COMMENT '课程id',
7. `course\_name`            STRING COMMENT '课程名称',
8. `succ\_avg\_score`         **BIGINT** COMMENT '完成平均成绩',
9. `succ\_avg\_time`          **BIGINT** COMMENT '完成平均时长',
10. `succ\_user\_count`        **BIGINT** COMMENT '完成用户数'
11. ) COMMENT '各课程考试相关指标统计'
12. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
13. LOCATION '/warehouse/edu/ads/ads\_course\_test\_statistics/';

装载语句

1. **insert** overwrite **table** ads\_course\_test\_statistics
2. **select** \* **from** ads\_course\_test\_statistics
3. **union**
4. **select**
5. '2022-02-21' dt,
6. recent\_days,
7. course\_id,
8. course\_name,
9. cast(succ\_avg\_score **as** **decimal**(16,2))  succ\_avg\_score,
10. cast(succ\_avg\_time **as** **decimal**(16,2))   succ\_avg\_time,
11. succ\_user\_count
12. **from**
13. (
14. **select**
15. 1                        recent\_days,
16. course\_id  ,
17. course\_name ,
18. avg(score\_1d)            succ\_avg\_score,
19. avg(duration\_sec\_1d)     succ\_avg\_time,
20. count(**distinct**(user\_id)) succ\_user\_count
21. **from** dws\_exam\_user\_paper\_exam\_1d
22. **where** dt='2022-02-21'
23. **group** **by** course\_id,course\_name
24. **union** all
25. **select**
26. recent\_days,
27. course\_id            ,
28. course\_name            ,
29. avg(paper\_score)       succ\_avg\_score,
30. avg(paper\_duration)    succ\_avg\_time,
31. count(**distinct**(if(paper\_score>0,user\_id,null))) succ\_user\_count
32. **from**
33. (
34. **select**
35. recent\_days,
36. user\_id,
37. course\_id,
38. course\_name,
39. case recent\_days
40. **when** 7 **then** score\_7d
41. **when** 30 **then** score\_30d
42. **end** paper\_score,
43. case recent\_days
44. **when** 7 **then** duration\_sec\_7d
45. **when** 30 **then** duration\_sec\_30d
46. **end** paper\_duration
47. **from** dws\_exam\_user\_paper\_exam\_nd lateral **view** explode(array(7,30)) tmp **as** recent\_days
48. **where** dt='2022-02-21'
49. )t1
50. **group** **by** recent\_days,course\_id,course\_name
51. )t2;

#### 2.5.5.3各试卷分数分布统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 考试 | 试卷-分数段 | 用户数 | ads\_paper\_garde\_section\_statistics | 各试卷分数分布统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_paper\_garde\_section\_statistics;
2. **CREATE** EXTERNAL **TABLE** ads\_paper\_garde\_section\_statistics
3. (
4. `dt`                     STRING COMMENT '统计日期',
5. `recent\_days`            **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `paper\_id`               STRING COMMENT '试卷id',
7. `paper\_title`            STRING COMMENT '试卷名称',
8. `garde\_section`          STRING COMMENT '分数段',
9. `user\_count`             **BIGINT** COMMENT '用户数'
10. ) COMMENT '各试卷分数分布统计'
11. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
12. LOCATION '/warehouse/edu/ads/ads\_paper\_garde\_section\_statistics/';

装载语句

1. **insert** overwrite **table** ads\_paper\_garde\_section\_statistics
2. **select** \* **from** ads\_paper\_garde\_section\_statistics
3. **union**
4. **select**
5. recent\_days,
6. paper\_id,
7. paper\_title,
8. garde\_section,
9. count(**distinct**(user\_id))   user\_count
10. **from**
11. (
13. **select**
14. 1 recent\_days,
15. paper\_id,
16. paper\_title,
17. user\_id,
18. case
19. **when** score\_1d<60 **then** '0-60'
20. **when** score\_1d>=60 and score\_1d<70 **then** '60-70'
21. **when** score\_1d>=70 and score\_1d<80 **then** '70-80'
22. **when** score\_1d>=80 and score\_1d<90 **then** '80-90'
23. **when** score\_1d>=90 **then** '90-100'
24. **end** garde\_section
25. **from** dws\_exam\_user\_paper\_exam\_1d
26. **where** dt='2022-02-21'
27. **union** all
28. **select**
29. 7       recent\_days,
30. paper\_id,
31. paper\_title,
32. user\_id,
33. case
34. **when** score\_7d<60 **then** '0-60'
35. **when** score\_7d>=60 and score\_7d<70 **then** '60-70'
36. **when** score\_7d>=70 and score\_7d<80 **then** '70-80'
37. **when** score\_7d>=80 and score\_7d<90 **then** '80-90'
38. **when** score\_7d>=90 **then** '90-100'
39. **end** garde\_section
40. **from** dws\_exam\_user\_paper\_exam\_nd
41. **where** dt='2022-02-21'
42. **union** all
43. **select**
44. 30       recent\_days,
45. paper\_id,
46. paper\_title,
47. user\_id,
48. case
49. **when** score\_30d<60 **then** '0-60'
50. **when** score\_30d>=60 and score\_30d<70 **then** '60-70'
51. **when** score\_30d>=70 and score\_30d<80 **then** '70-80'
52. **when** score\_30d>=80 and score\_30d<90 **then** '80-90'
53. **when** score\_30d>=90 **then** '90-100'
54. **end** garde\_section
55. **from** dws\_exam\_user\_paper\_exam\_nd
56. **where** dt='2022-02-21'
57. )garde\_sec
58. **group** **by** recent\_days, paper\_id, paper\_title, garde\_section;

#### 2.5.5.4各题目正确率统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 考试 | 题目 | 正确率 | ads\_topic\_accuracy\_statistics | 各题目正确率统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_topic\_accuracy\_statistics;
2. **CREATE** EXTERNAL **TABLE** ads\_topic\_accuracy\_statistics
3. (
4. `dt`                            STRING COMMENT '统计日期',
5. `recent\_days`                   **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `question\_id`                   STRING COMMENT '题目id',
7. `question\_title`                STRING COMMENT '题目内容',
8. `question\_accuracy\_rate`        **BIGINT** COMMENT '题目正确率'
9. ) COMMENT '各题目正确率统计'
10. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
11. LOCATION '/warehouse/edu/ads/ads\_topic\_accuracy\_statistics/';

装载语句

1. **insert** overwrite **table** ads\_topic\_accuracy\_statistics
2. **select** \* **from** ads\_topic\_accuracy\_statistics
3. **union**
4. **select**
5. d1.recent\_days,
6. d1.question\_id,
7. e1.question\_txt,
8. cast(succ\_count/all\_count **as** **decimal**(16,2))  succ\_rate
9. **from**
10. (
11. **select**
12. a1.recent\_days   recent\_days,
13. a1.question\_id   question\_id,
14. succ\_count,
15. all\_count
16. **from**
17. (
18. **select**
19. recent\_days,
20. question\_id,
21. is\_correct,
22. if(is\_correct='1',sum(if(date\_id>=date\_add('2022-02-21',-recent\_days+1) ,1,0)),0)  succ\_count
23. **from** dwd\_exam\_question\_exam\_inc lateral **view** explode(array(1,7,30)) tmp **as** recent\_days
24. **where** dt='2022-02-21' and is\_correct='1'
25. **group** **by** recent\_days, question\_id,is\_correct
26. )a1
27. join
28. (
29. **select**
30. recent\_days,
31. question\_id,
32. sum(if(date\_id>=date\_add('2022-02-21',-recent\_days+1),1,0))  all\_count
33. **from** dwd\_exam\_question\_exam\_inc lateral **view** explode(array(1,7,30)) tmp **as** recent\_days
34. **where** dt='2022-02-21'
35. **group** **by** recent\_days, question\_id
36. )c1
37. **on** a1.question\_id=c1.question\_id
38. **group** **by** a1.recent\_days,a1.question\_id,succ\_count,all\_count
39. )d1
40. left join
41. (
42. **select**
43. question\_id,
44. question\_txt
45. **from** dim\_question\_full
46. **where** dt='2022-02-21'
47. **group** **by** question\_id, question\_txt
48. )e1
49. **on** d1.question\_id=e1.question\_id;

### 2.5.6播放主题指标

#### 2.5.6.1各章节视频播放情况统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 播放 | 章节 | 视频播放次数  人均观看时长  观看人数 | ads\_chapter\_video\_play\_statistics | 各章节视频播放情况统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_chapter\_video\_play\_statistics;
2. **CREATE** EXTERNAL **TABLE** ads\_chapter\_video\_play\_statistics
3. (
4. `dt`                            STRING COMMENT '统计日期',
5. `recent\_days`                   **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `chapter\_id`                    STRING COMMENT '章节id',
7. `chapter\_title`                 STRING COMMENT '章节名称',
8. `video\_play\_times`              **BIGINT** COMMENT '视频播放次数',
9. `avg\_watch\_time`                **BIGINT** COMMENT '人均观看时长',
10. `watch\_count`                   **BIGINT** COMMENT '观看人数'
11. ) COMMENT '各章节视频播放情况统计'
12. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
13. LOCATION '/warehouse/edu/ads/ads\_chapter\_video\_play\_statistics/';

装载语句

1. **insert** overwrite **table** ads\_chapter\_video\_play\_statistics
2. **select** \* **from** ads\_chapter\_video\_play\_statistics
3. **union**
4. **select**
5. recent\_day,
6. chapter\_id,
7. chapter\_title,
8. video\_play\_times,
9. avg\_watch\_time,
10. watch\_count
11. **from** (
12. **select**
13. 1                         recent\_day,
14. chapter\_id,
15. chapter\_name              chapter\_title,
16. sum(view\_video\_count\_1d)  video\_play\_times,
17. sum(during\_video\_time\_1d)/count(**distinct** (user\_id)) avg\_watch\_time,
18. count(**distinct** (user\_id)) watch\_count
19. **from** dws\_user\_video\_play\_1d
20. **where** dt = '2022-02-21'
21. **group** **by** chapter\_id, chapter\_name
22. **union** all
23. **select**
24. recent\_days,
25. chapter\_id,
26. chapter\_name,
27. sum(video\_play\_times),
28. sum(avg\_watch\_time)/count(**distinct**(if(watch\_count>0,user\_id,0))),
29. count(**distinct**(if(watch\_count>0,user\_id,0)))
30. **from**
31. (
32. **select**
33. recent\_days,
34. user\_id,
35. chapter\_id,
36. chapter\_name,
37. case recent\_days
38. **when** 7 **then** view\_video\_count\_7d
39. **when** 30 **then** view\_video\_count\_30d
40. **end** video\_play\_times,
41. case recent\_days
42. **when** 7 **then** during\_video\_time\_7d
43. **when** 30 **then** during\_video\_time\_30d
44. **end** avg\_watch\_time,
45. case recent\_days
46. **when** 7 **then** view\_video\_progress\_7d
47. **when** 30 **then** view\_video\_progress\_30d
48. **end** watch\_count
49. **from** dws\_user\_video\_play\_nd lateral **view** explode(array(7,30)) tmp **as** recent\_days
50. **where** dt='2020-06-14'
51. )b1
52. **group** **by** recent\_days, chapter\_id, chapter\_name
53. )a1;

#### 2.5.6.2各课程视频播放情况统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 播放 | 课程 | 视频播放次数  人均观看时长  观看人数 | ads\_course\_video\_play\_statistics | 各课程视频播放情况统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_course\_video\_play\_statistics;
2. **CREATE** EXTERNAL **TABLE** ads\_course\_video\_play\_statistics
3. (
4. `dt`                            STRING COMMENT '统计日期',
5. `recent\_days`                   **BIGINT** COMMENT '最近天数,1:最近1天,7:最近7天,30:最近30天',
6. `chapter\_id`                    STRING COMMENT '课程id',
7. `chapter\_title`                 STRING COMMENT '课程名称',
8. `video\_play\_times`              **BIGINT** COMMENT '视频播放次数',
9. `avg\_watch\_time`                **BIGINT** COMMENT '人均观看时长',
10. `watch\_count`                   **BIGINT** COMMENT '观看人数'
11. ) COMMENT '各课程视频播放情况统计'
12. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
13. LOCATION '/warehouse/edu/ads/ads\_course\_video\_play\_statistics/';

装载语句

1. **insert** overwrite **table** ads\_course\_video\_play\_statistics
2. **select** \* **from** ads\_course\_video\_play\_statistics
3. **union**
4. **select**
5. recent\_day,
6. course\_id,
7. course\_name,
8. video\_play\_times,
9. avg\_watch\_time,
10. watch\_count
11. **from**
12. (
13. **select**
14. 1                         recent\_day,
15. course\_id                  course\_id,
16. course\_name              ,
17. sum(view\_video\_count\_1d)  video\_play\_times,
18. sum(during\_video\_time\_1d)/count(**distinct**(user\_id)) avg\_watch\_time,
19. count(**distinct** (user\_id)) watch\_count
20. **from** dws\_user\_video\_play\_1d
21. **where** dt = '2022-02-21'
22. **group** **by** course\_id, course\_name
23. **union**
24. **select**
25. recent\_days,
26. course\_id,
27. course\_name,
28. sum(video\_play\_times),
29. sum(avg\_watch\_time)/count(**distinct**(if(watch\_count>0,user\_id,0))),
30. count(**distinct**(if(watch\_count>0,user\_id,0)))
31. **from**
32. (
33. **select**
34. recent\_days,
35. user\_id,
36. course\_id,
37. course\_name,
38. case recent\_days
39. **when** 7 **then** view\_video\_count\_7d
40. **when** 30 **then** view\_video\_count\_30d
41. **end** video\_play\_times,
42. case recent\_days
43. **when** 7 **then** during\_video\_time\_7d
44. **when** 30 **then** during\_video\_time\_30d
45. **end** avg\_watch\_time,
46. case recent\_days
47. **when** 7 **then** view\_video\_progress\_7d
48. **when** 30 **then** view\_video\_progress\_30d
49. **end** watch\_count
50. **from** dws\_user\_video\_play\_nd lateral **view** explode(array(7,30)) tmp **as** recent\_days
51. **where** dt='2020-06-14'
52. )b1
53. **group** **by** recent\_days, course\_id, course\_name
54. )a1;

### 2.5.7完课主题指标

#### 2.5.7.1各课程完课人数统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 完课 | 课程 | 完课人数 | ads\_course\_over\_user\_statistics | 各课程完课人数统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_course\_over\_user\_statistics;
2. **CREATE** EXTERNAL **TABLE** ads\_course\_over\_user\_statistics
3. (
4. `dt`            STRING COMMENT '统计日期',
5. `recent\_days`   **BIGINT** COMMENT '最近天数',
6. `course\_id`     STRING COMMENT '课程id',
7. `course\_name`   STRING COMMENT '课程名称',
8. `course\_end\_uc` **BIGINT** COMMENT '完课人数'
9. ) COMMENT '最近1/7/30天完课人数统计'
10. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
11. LOCATION '/warehouse/edu/ads/ads\_course\_over\_user\_statistics/';

装载语句

1. **insert** overwrite **table** ads\_course\_over\_user\_statistics
2. **select** \* **from** ads\_course\_over\_user\_statistics
3. **union**
4. **select** '2022-02-21' `dt`,
5. 1            `recent\_days`,
6. course\_id,
7. course\_name,
8. sum(`if`(flag\_course = 0, 1, 0))
9. **from** (**select** course\_id,
10. course\_name,
11. user\_id,
12. sum(flag\_video) over (partition **by** course\_id,user\_id) flag\_course
13. **from** (
14. **select** user\_id,
15. course\_id,
16. course\_name,
17. `if`(during\_video\_time\_1d > 900 \* 0.9 and view\_video\_progress\_1d > 900 \* 0.9, 0, 1) flag\_video
18. **from** dws\_user\_video\_play\_1d
19. **where** dt = '2022-02-21'
20. ) t1
21. ) t2
22. **group** **by** course\_id,
23. course\_name
24. **union** all
25. **select** '2022-02-21' `dt`,
26. recent\_days,
27. course\_id,
28. course\_name,
29. sum(`if`(flag\_course = 0, 1, 0))
30. **from** (**select** `recent\_days`,
31. course\_id,
32. course\_name,
33. sum(flag\_video) over (partition **by** course\_id,user\_id) flag\_course
34. **from** (**select** `recent\_days`,
35. course\_id,
36. course\_name,
37. user\_id,
38. `if`(play\_sum > 900 \* 0.9 and pos\_max > 900 \* 0.9, 0, 1) flag\_video
39. **from** (
40. **select** recent\_days,
41. course\_id,
42. course\_name,
43. user\_id,
44. case recent\_days
45. **when** 7 **then** during\_video\_time\_7d
46. **when** 30 **then** during\_video\_time\_30d
47. **end** play\_sum ,
48. case recent\_days
49. **when** 7 **then** view\_video\_progress\_7d
50. **when** 30 **then** view\_video\_progress\_30d
52. **end** pos\_max
53. **from** dws\_user\_video\_play\_nd
54. lateral **view** explode(`array`(7, 30)) tmp **as** recent\_days
55. **where** dt = '2022-02-21'
56. ) t3
57. ) t4
58. ) t5
59. **group** **by** recent\_days,
60. course\_id,
61. course\_name;

#### 2.5.7.2完课综合指标

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 播放 |  | 总完课人数  总完课人次 | ads\_all\_course\_over\_user\_statistics | 完课综合指标 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_all\_course\_over\_user\_statistics;
2. **CREATE** EXTERNAL **TABLE** ads\_all\_course\_over\_user\_statistics
3. (
4. `dt`            STRING COMMENT '统计日期',
5. `recent\_days`   **BIGINT** COMMENT '最近天数',
6. `course\_end\_uc` **BIGINT** COMMENT '完课人数',
7. `course\_end\_pv` **BIGINT** COMMENT '完课人次'
8. ) COMMENT '最近1/7/30天完课综合指标统计'
9. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
10. LOCATION '/warehouse/edu/ads/ads\_all\_course\_over\_user\_statistics/';

装载语句

1. **insert** overwrite **table** ads\_all\_course\_over\_user\_statistics
2. **select** \* **from** ads\_all\_course\_over\_user\_statistics
3. **union**
4. **select** '2022-02-21' `dt`,
5. 1            `recent\_days`,
6. count(\*),
7. count(**distinct** user\_id)
8. **from** (**select** course\_id,
9. course\_name,
10. user\_id,
11. sum(flag\_video) over (partition **by** course\_id,user\_id) flag\_course
12. **from** (
13. **select** user\_id,
14. course\_id,
15. course\_name,
16. `if`(during\_video\_time\_1d > 900 \* 0.9 and view\_video\_progress\_1d > 900 \* 0.9, 0, 1) flag\_video
17. **from** dws\_user\_video\_play\_1d
18. **where** dt = '2022-02-21'
19. ) t1
20. ) t2
21. **where** flag\_course=0
22. **union** all
23. **select** '2022-02-21' `dt`,
24. recent\_days,
25. count(\*),
26. count(**distinct** user\_id)
27. **from** (**select** `recent\_days`,
28. course\_id,
29. course\_name,
30. user\_id,
31. sum(flag\_video) over (partition **by** course\_id,user\_id) flag\_course
32. **from** (**select** `recent\_days`,
33. course\_id,
34. course\_name,
35. user\_id,
36. `if`(play\_sum > 900 \* 0.9 and pos\_max > 900 \* 0.9, 0, 1) flag\_video
37. **from** (
38. **select** recent\_days,
39. course\_id,
40. course\_name,
41. user\_id,
42. case recent\_days
43. **when** 7 **then** during\_video\_time\_7d
44. **when** 30 **then** during\_video\_time\_30d
45. **end** play\_sum,
46. case recent\_days
47. **when** 7 **then** during\_video\_time\_7d
48. **when** 30 **then** during\_video\_time\_30d
49. **end** pos\_max
50. **from** dws\_user\_video\_play\_nd
51. lateral **view** explode(`array`(7, 30)) tmp **as** recent\_days
52. **where** dt = '2022-02-21'
53. ) t3
54. ) t4
55. ) t5
56. **where** flag\_course=0
57. **group** **by** recent\_days;

#### 2.5.7.3各课程人均完成章节视频数统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 子主题 | 指标需求 | 表名 | 描述 |
| 播放 | 课程 | 人均完成章节数 | ads\_course\_avg\_user\_chapter\_finished\_statistics | 各课程人均完成章节视频数统计 |

建表语句

1. **DROP** **TABLE** IF EXISTS ads\_course\_avg\_user\_chapter\_finished\_statistics;
2. **CREATE** EXTERNAL **TABLE** ads\_course\_avg\_user\_chapter\_finished\_statistics
3. (
4. `dt`            STRING COMMENT '统计日期',
5. `recent\_days`   **BIGINT** COMMENT '最近天数',
6. `course\_id`     STRING COMMENT '课程id',
7. `course\_name`   STRING COMMENT '课程名称',
8. `chapter\_avg` **DECIMAL**(16, 2) COMMENT '人均完成章节数'
9. ) COMMENT '最近1/7/30天各课程人均完成章节视频数统计'
10. ROW FORMAT DELIMITED FIELDS TERMINATED **BY** '\t'
11. LOCATION '/warehouse/edu/ads/ads\_course\_avg\_user\_chapter\_finished\_statistics/';

装载语句

1. **insert** overwrite **table** ads\_course\_avg\_user\_chapter\_finished\_statistics
2. **select** \* **from** ads\_course\_avg\_user\_chapter\_finished\_statistics
3. **union**
4. **select** '2022-02-21' `dt`,
5. 1            `recent\_days`,
6. course\_id,
7. course\_name,
8. cast(sum(flag\_video) / count(**distinct** user\_id) **as** **decimal**(16, 2))
9. **from** (
10. **select** user\_id,
11. chapter\_id,
12. course\_id,
13. course\_name,
14. `if`(during\_video\_time\_1d > 900 \* 0.9 and view\_video\_progress\_1d > 900 \* 0.9, 1, 0) flag\_video
15. **from** dws\_user\_video\_play\_1d
16. **where** dt = '2022-02-21'
17. ) t1
18. **group** **by** course\_id,
19. course\_name
20. **union** all
21. **select** '2022-02-21' `dt`,
22. recent\_days,
23. course\_id,
24. course\_name,
25. cast(sum(flag\_video) / count(**distinct** user\_id) **as** **decimal**(16, 2))
26. **from** (**select** `recent\_days`,
27. course\_id,
28. course\_name,
29. user\_id,
30. `if`(play\_sum > 900 \* 0.9 and pos\_max > 900 \* 0.9, 1, 0) flag\_video
31. **from** (
32. **select** recent\_days,
33. course\_id,
34. course\_name,
35. user\_id,
36. case recent\_days
37. **when** 7 **then** during\_video\_time\_7d
38. **when** 30 **then** during\_video\_time\_30d
39. **end** play\_sum,
40. case recent\_days
41. **when** 7 **then** view\_video\_progress\_7d
42. **when** 30 **then** view\_video\_progress\_30d
43. **end** pos\_max
44. **from** dws\_user\_video\_play\_nd
45. lateral **view** explode(`array`(7, 30)) tmp **as** recent\_days
46. **where** dt = '2022-02-21'
47. ) t3
48. ) t4
49. **group** **by** recent\_days,
50. course\_id,
51. course\_name;

## 2.6日志表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 导入路径 | 结构 | 全量/增量 | ODS表名 | 描述 |
| /origin\_data/edu/log/topic\_log/\* | 生成的三种日志信息采集到一张日志表中 | 全量 | ods\_log\_inc | 日志表 |

建表语句

1. **DROP** **TABLE** IF EXISTS ods\_log\_inc;
2. **CREATE** EXTERNAL **TABLE** ods\_log\_inc
3. (
4. `common`   STRUCT<ar :STRING,ba :STRING,ch :STRING,is\_new :STRING,md :STRING,mid :STRING,os :STRING,sc :STRING,
5. sid :STRING,uid :STRING,vc :STRING> COMMENT '公共信息',
6. `page`     STRUCT<during\_time :STRING,item :STRING,item\_type :STRING,last\_page\_id :STRING,page\_id
7. :STRING,source\_type :STRING> COMMENT '页面信息',
8. `actions`  ARRAY<STRUCT<action\_id:STRING,item:STRING,item\_type:STRING,ts:**BIGINT**>> COMMENT '动作信息',
9. `displays` ARRAY<STRUCT<display\_type :STRING,item :STRING,item\_type :STRING,`**order**` :STRING,pos\_id
10. :STRING>> COMMENT '曝光信息',
11. `start`    STRUCT<entry :STRING,first\_open :**BIGINT**,loading\_time :**BIGINT**,open\_ad\_id :**BIGINT**,open\_ad\_ms :**BIGINT**,open\_ad\_skip\_ms
12. :**BIGINT**> COMMENT '启动信息',
13. `appVideo` STRUCT<play\_sec :**BIGINT**,position\_sec :**BIGINT**,video\_id :STRING> COMMENT '视频信息',
14. `err`      STRUCT<error\_code:**BIGINT**,msg:STRING> COMMENT '错误信息',
15. `ts`       **BIGINT**  COMMENT '时间戳'
16. ) COMMENT '日志信息表'
17. PARTITIONED **BY** (`dt` STRING)
18. ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.JsonSerDe'
19. LOCATION '/warehouse/edu/ods/ods\_log\_inc/';

装载语句

1. **load** data inpath '/origin\_data/edu/log/topic\_log/2022-02-21' **into** **table** ods\_log\_inc partition(dt='2022-02-21');

# 3 脚本

## 3.1调度脚本

### 3.1.1flume脚本

 

### 3.1.2kafka启动脚本，zookeeper启动脚本，Maxwell启动脚本

### 

### 3.1.3

## 3.2各层初始化脚本

### 3.2.1 Mysql\_to\_hdfs

 

### 3.2.2 Hdfs\_to\_ods

 

### 3.2.3 Ods\_to\_dim

 

### 3.2.4 Ods\_to\_dwd

 

### 3.2.5 Dwd\_to\_dws

 

### 3.2.6 Dws\_1d\_to\_dws\_nd

### 

### 3.2.7 Dws\_1d\_to\_dws\_td

 

# 4 git 仓库地址

       如果是gitee 仓库需要申请开源，否则除项目组成员外无法访问。

https://github.com/MythLei/Online-education-data-warehouse

# 5 截图

       所有指标在 Superset 的可视化展示截图，每个主题域的图表汇总到一起截一张图。

       DS 全流程调度 DAG 图所有结点全部跑通（结点上方有绿色对钩）的截图，如下。

