

数仓命名规范

- 一、数仓分层
- 二、数据库命名规则
- 三、数仓表名命名规范
 - 1、ods层表
 - 2、sdd层表
 - 3、dim层表
 - 4、dwd层表
 - 5、dws层表
 - 6、ads层表
 - 7、临时表
 - 8、中间表
- 四、建表规范
 - 1、建表模板
 - 2、字段映射类型
- 五、任务名命名规范

一、数仓分层

层级	描述
ods	对接源数据，将原始数据几乎无处理的存放在数据仓库系统，结构上与源系统基本保持一致。非结构化的数据进行结构化处理。
sdd	数据标准层。根据国家以及行业标准制定的一层，用于兼容垂直行业内的数据差异，规范数仓模型。
dim	维表层。建立一致数据分析维表，可以降低数据计算口径和算法不同意风险。
dwd	明细事实表层。以业务过程作为建模驱动，基于每个具体的业务过程特点，构建最细粒度的明细层事

	实表，某些重要维度属性字段适当冗余，即宽表化处理。
dws	公共汇总层。以分析的主题对象作为建模驱动，基于上层的应用和产品的指标需求，构建公共汇总层。为上层提供公共指标，建立汇总宽表、明细事实表。
ads	数据应用层。存放数据产品个性化的统计指标数据，个性化指标加工。定制化复杂性指标，基于应用的数据组装，宽表集市。趋势指标。

二、数据库命名规则

数据库名按照项目来划分，一个项目只保留一个数据库名，不同层级之间得表通过前缀来区分，命名规则：项目。

例：贵燃项目：

数据库名
guiran

三、数仓表名命名规范

单分区增量全量标识简写

标识	简写
增量	i
全量	f

常用更新周期简写

更新周期	简写
小时	h
日	d
周	w
月	m

年	y
---	---

常用统计时间范围简写

中文名	英文名
最近1天	1d
最近3天	3d
最近1周	1w
最近2周	2w
最近1月	1m
最近2月	2m
最近3月	3m
最近6月	6m
180天以前	bh
自然周	cw
自然月	cm
自然季度	cq
截止当日	td
年初截止当日	sd
零点截止当前	tt
财年	fy
最近1小时	1h
准实时	ts
未来1周	f1w
未来4周	f4w

常用行业标准简写

业务域、数据域、业务过程、业务标准、数据集市等 需要梳理好后更新

1、ods层表

数据库名.层级 数据库名 源表名 _ 更新周期+单分区增量全量标识

- 1 例：应收应付流水表
- 2 按天全量数据：dbname.ods_project_accruals_record_df
- 3 按天增量数据：dbname.ods_project_accruals_record_di

2、sdd层表

数据库名.层级 业务域/pub 业务标准 _ 更新周期+单分区增量全量标识

- 1 等业务标准梳理好后补充
- 2 例：应收应付流水表
- 3 按天全量数据：dbname.sdd_pub_业务标准_df
- 4 按天增量数据：dbname.sdd_pub_业务标准_di

3、dim层表

数据库名.层级 业务域/pub 维度定义[自定义命名] 更新周期+单分区增量全量标识

- 1 例：时间维度表
- 2 dbname.dim_pub_date_df

4、dwd层表

数据库名.层级 业务域/pub 数据域 业务过程[自定义命名] _ 更新周期+单分区增量全量标识

- 1 例如：应收应付明细表
- 2 dbname.dwd_project_trade_accruals_record_detail_df

5、dws层表

数据名.层级 业务域/pub 数据域 数据粒度[自定义命名] 统计时间范围[更新周期]+[_单分区增量全量标识]

- 1 例如：淘宝卖家交易订单最近1天汇总表
- 2 `dbname.dws_tb_trd_slr_ord_1d`

6、ads层表

数据名.层级 业务域/pub 数据集市[自定义命名] 统计时间范围[更新周期]+[_单分区增量全量标识]

- 1 例如：应收应付汇总表
- 2 `dbname.ads_project_trade_accruals_record_1d`

7、临时表

临时表命名在各层级的基础上加上tmp{n}_前缀标识，n是同一个表有多个临时表时，用数字区分，并且需要及时删除

- 1 例如：淘宝卖家交易订单汇总临时表
- 2 `tmp1_dws_tb_trd_slr_ord_1d`
- 3 `tmp2_dws_tb_trd_slr_ord_1d`

8、中间表

中间表命名在各层级的基础上加上mid{n}_前缀标识，并且需要及时删除数据

- 1 例如：淘宝卖家订单汇总中间表
- 2 `mid1_dws_tb_trd_slr_ord_1d`
- 3 `mid2_dws_tb_trd_slr_ord_1d`

四、建表规范

1、建表模板

只需要指定分区字段就行。特殊的表，根据实际情况自行修改。

```
1 drop table if exists [dwd].[table_name];
2 create table if not exists [dwd].[table_name](
3     col1 type [comment]
4     ,col2 type [comment]
5     ,.....
6 ) [comment]
7 partitioned by (pcol type [comment])
```

2、字段映射类型

从源数据库到ods层全都转换成string格式，在sdd层根据源表数据类型在按照下面对应关系转换

业务库类型(mysql等)	数仓字段类型 (hive)
整形 (small int,int,long,bigint等)	bigint
浮点型 (decimal, float, double等)	double
时间 (date,datetime,timestamp等)	timestamp
字符串 (char,varchar,string等)	string
布尔型 (bool,boolean等)	boolean

五、任务名命名规范

job_表名。一个 job任务，只能有一个输出表。

```
1 例如：和工程交易相关的ods层表的调度任务
2      job_ods_project_trade_df
```

