ODS层设计规范

更新时间：2020-07-10 16:06:25

[**本页目录**](javascript:void(0))

* [数据同步及处理规范](https://help.aliyun.com/document_detail/117434.html?spm=a2c4g.11174283.6.916.2e1d590eUM5fev#title-nj1-gai-nn1)
* [命名规范](https://help.aliyun.com/document_detail/117434.html?spm=a2c4g.11174283.6.916.2e1d590eUM5fev#title-zw5-ppa-oau)
* [数据存储及生命周期管理规范](https://help.aliyun.com/document_detail/117434.html?spm=a2c4g.11174283.6.916.2e1d590eUM5fev#title-dk7-68w-rw1)
* [数据质量规范](https://help.aliyun.com/document_detail/117434.html?spm=a2c4g.11174283.6.916.2e1d590eUM5fev#title-uly-r5y-n41)

本文为您介绍ODS层设计规范。

**数据同步及处理规范**

* 数据同步方式的选择

数据同步规范如下。基本规范通过需求形式落地到DataWorks的数据集成，规范落地情况依赖工具的推进节奏：

* + 一个系统的源表只允许同步一次到MaxCompute。
  + 数据集成只同步离线全量数据。增量数据同步可以通过数据传输服务DTS实现。
* 数据加载与处理
  + 数据集成同步全量数据时会直接进入全量表的当日分区。
  + DTS同步数据处理方式如下：
    - 全量初始化：对于每个同步表，全量初始化的数据都会独立存储在MaxCompute的全量基线表中，表名的默认格式为**源表名\_base**。
    - 增量数据同步：增量数据会实时同步到MaxCompute，并存储在增量日志表中，表名的默认格式为**源表名\_log**。每个同步表对应一个增量日志表。在同步增量数据时，DTS会将多条记录合并到一个文件并写入MaxCompute。
  + DTS全量数据合并。DTS根据全量基线表和增量日志表得到某个时刻表的全量数据，并通过MaxCompute SQL合并全量数据。您可以通过DataWorks配置一个全量数据Merge节点，当全量数据完成Merge后，可自动调度后续的计算分析节点。您也可以配置调度周期，进行周期性数据的离线分析。
  + 所有ODS层的表都以统计日期及时间分区表方式存储，数据成本由存储管理和策略控制。
  + 如果源系统新增了字段，您需要重新配置数据集成同步作业。如果目标表的字段在源系统中不存在，数据集成自动填充NULL。

**命名规范**

* 表命名规范

表命名规则：**{层次}{源系统表名}{保留位/delta与否}**。

* + 增量数据：**{project\_name}.s{源系统表名}delta**。
  + 全量数据：**{project\_name}.s{源系统表名}**。
  + ODS ETL过程的临时表：**{project\_name}.tmp{临时表所在过程的输出表}{从0开始的序号}**。
  + 按小时同步的增量表：**{project\_name}.s{源系统表名}{delta}\_{hh}**。
  + 按小时同步的全量表：**{project\_name}.s{源系统表名}{hh}**。
  + 当不同源系统同步到同一个Project下的表命名冲突时，您需要给同步较晚的表名加上源系统的**dbname**以解决冲突。
* 字段命名规范
  + 字段默认使用源系统的字段名。
  + 字段名与MaxCompute关键字冲突时，在源字段名后加上**col**，即**源字段名col**。MaxCompute关键字详情请参见[保留字与关键字](https://help.aliyun.com/document_detail/27872.html" \l "concept-zxk-v5f-vdb" \o "本文为您介绍MaxCompute SQL中的所有保留字与关键字。)。
* 同步任务命名规范
  + 任务名：**{源系统表名}[delta]**。

**说明** 同一Project下异库同名表的任务名为**{源系统表名}{tddl的appname}[\_delta]**。

* + 任务的输出名称，即输出表的名称，需要与数据存储及生命周期管理规范保持一致。详情请参见[数据存储及生命周期管理规范](https://help.aliyun.com/document_detail/117434.html?spm=a2c4g.11174283.6.916.2e1d590eUM5fev#section-jvp-q65-xrn)。

**数据存储及生命周期管理规范**

| **数据表类型** | **存储方式** | **最长存储保留策略** |
| --- | --- | --- |
| ODS流水型全量表 | 按天分区 | * 不可再生情况下，永久保存。 * 日志（数据量非常大，例如一天数据量大于100 GB）数据保留24个月。 * 自主设置是否保留历史月初数据。 * 自主设置是否保留特殊日期数据。 |
| ODS镜像型全量表 | 按天分区 | * 重要的业务表及需要保留历史的表视情况保存。 * ODS全量表的默认生命周期为2天，支持通过ds=max\_pt(tablename)方式访问数据。 |
| ODS增量表 | 按天分区 | * 有对应全量表，最多保留最近14天分区数据。 * 无对应全量表，需要永久保留数据。 |
| ODS ETL过程临时表 | 按天分区 | 最多保留最近7天分区。 |
| DBSync非去重数据 | 按天分区 | 由应用通过中间层保留历史数据，默认ODS层不保留历史数据。 |

**数据质量规范**

* 每个ODS全量表必须配置唯一性字段标识。
* 每个ODS全量表必须有注释。
* 每个ODS全量表必须监控分区空数据。
* 仅有监控要求的ODS表才需要创建数据质量监控规则。您可以通过DataWorks配置数据质量监控规则，详情请参见[配置数据质量监控](https://help.aliyun.com/document_detail/146697.html#task-2357812)。
* 建议对重要表的重要枚举类型字段进行枚举值变化及枚举值分布监控。
* 建议对ODS表的数据量及数据记录数设置周同环比监控，如果周同环比无变化，表示源系统已迁移或下线