

# COS系统的使用规范

Author: koray

Version: 1.0 (持续更新完善)

Date: 2020-11-20

## 1. COS目录命名规范

### 命名格式

- **Cos桶名+公共目录名(项目名\_公共功能)+数据层级名+留存时间标识+表名**

示例：cosn://emr-1254463213/lazada\_dw/ods/ods\_mysql\_station/

### 1. Cos桶命名

- 大数据平台的Cos桶名称：**emr**
- 爬虫数据的Cos桶名称：**spider\_log**

### 2. 公共目录名

#### 项目名

- 目前项目有 **lazada** 、 **shopee**

项目	示例
lazada	cosn://emr/lazada_dw/
shopee	cosn://emr/shopee_dw/

#### 公共功能

功能	标识	示例
hive数仓(data warehouse)	<b>dw</b>	cosn://emr/lazada_dw/
数据导出的缓存层	<b>bigdata_export</b>	cosn://emr/bigdata_export/

### 3. 数据层级名

数仓层级	说明	示例
ods层	数据贴源层，映射cos的原始数据	cosn://emr/lazada_dw/ods/
dwd层	数据明细层，对ods层进行ETL处理从而得出基础表	cosn://emr/lazada_dw/dwd/
app层	数据应用层，聚合分析dwd层的数据，供后续业务使用	cosn://emr/lazada_dw/app/

### 4. 留存时间标识

暂时不启用

留存时间	标识
长期保存(不删除)	long_term
保留1年	one_year
保留近半年(180天)	half_year
保留近30天	thirty_days
保留近7天	seven_days
保留一天	one_day
脚本中设置的维护数据任务目录	others

### 5. 表名

- 本文档要求表名与数据存储路径最后一个名字一致，原始日志表(ods)除外。
  - Cos存储路径：cosn://emr-1254463213/lazada\_dw/ods/ods\_mysql\_station/
  - 表名为：ods\_mysql\_station
- 表的命名规范详见大数据开发规范文档

## 2.日志服务的COS投递设置

基本配置

配置项	参数	例子
COS存储桶	<b>COS桶名</b>	spider-log-1254463213
目录前缀	<b>项目_主题_csv</b>	lzd_item_csv/
分区格式	<b>%Y-%m-%d/%H/主题</b>	%Y-%m-%d/%H/item (hive表分区字段格式yyyy-MM-dd)
路径	<b>{COS存储桶}{目录前缀}{分区格式}_{random}_{index}.{type}</b>	cosn://spider-log-1254463213.cos.ap-shanghai.myqcloud.com/ <b>lzd_item_csv/2020-11-09/10/item</b> _2f25fd1d-b34d-4de6-b3bf-e8a754a85fb1_000
压缩	无	
投递文件大小	100M	
投递间隔时间	5min	

高级设置

配置项	参数
投递格式	<b>csv</b>
键值索引	<b>content</b>
分隔符	<b>制表符</b>
转义符	<b>空</b>