业务数据：登陆，下单，支付等过程

用户行为数据：页面浏览下的点击，停留，评论，点赞，收藏等。

用户行为数据

业务数据

ODS

DWD

DWS

DWT

ADS

Data Warehouse

Flume

Sqoop

1. 设计模块

1.用户行为数据采集平台

2.业务数据采集平台

3.数据仓库维度建模

4.分析设备，会员，商品，地区，活动等电商核心主题

5.即席查询

6.集群状态监控

7.质量监控

集群服务配置

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | n1 | n2 | n3 |
| hdfs | NameNode | √ |  |  |
| DataNode | √ | √ | √ |
| SecondaryNameNode |  | √ |  |
| yarn | NodeManager | √ | √ | √ |
| ResourceManager |  |  | √ |
| Zookeeper | Zookeeper Server | √ | √ | √ |
| Flume（采集日志） | Flume | √ | √ |  |
| Kafka | Kafka | √ | √ | √ |
| Flume（消费kafka） | Flume |  |  | √ |
| Hive | Hive | √ |  |  |
| Mysql | Mysql | √ |  |  |
| Sqoop | Sqoop | √ |  |  |
| Azkaban | AzkabanWebserver | √ |  |  |
| AzkabanExecutorServer | √ |  |  |
| Spark |  | √ |  |  |
| Kylin |  | √ |  |  |
| Hbase | HMaster | √ |  |  |
| HRegionServer | √ | √ | √ |
| Superset |  | √ |  |  |
| Atlas |  | √ |  |  |
| Solar | Jar | √ |  |  |
| 服务数总计 |  | 18 | 7 | 7 |

业务数据

日志数据

日志数据

Flume采集

Flume采集

Flume消费

Hadoop集群存储

Kafka

Kafka

Kafka

Sqoop同步

Zookeeper

集群状态监控报警

定时调度Azkaban

ODS

DWD

DWS

DWT

ADS

Sqoop

MySQL

质量管理，Python+shell

多维分析kylin

Hbase

可视化

SKU = Stock Keeping Unit（库存量基本单位）。现在已经被引申为产品统一编号的简称，每种产品均对应有唯一的SKU号。

SPU（Standard Product Unit）：是商品信息聚合的最小单位，是一组可复用、易检索的标准化信息集合。

举例：例如：iPhoneX手机就是SPU。一台银色、128G内存的、支持联通网络的iPhoneX，就是SKU。

业务表

用户表

订单表

优惠卷领用表

省份表

地区表

订单状态流水表

支付表

订单明细表

退款表

退单表

评价表

订单明细优惠卷关联表

订单明细活动关联表

活动规则表

优惠卷信息表

用户地址表

字典表

收藏表

购物车表

商品信息表

后台管理表

商品信息表

SPU信息表

三级分类表

二级分类表

一级分类表

品牌表

优惠卷范围表

优惠卷信息表

活动商品关联表

活动信息表

活动规则表

SPU销售属性表

SPU销售属性值表

SKU销售属性表

SKU平台属性表

平台属性值表

平台属性表

1. 日志数据（为json格式）

1.页面日志·page（记录一个页面的用户访问情况，包括访问时间，停留时间，页面路径等信息）

商品详情页面

页面id

上页id

页面对象类型

页面对象id

页面来源类型

停留时间

跳入时间

商品详情页

商品列表

商品

1001

搜索结果

30000（毫秒）

13位时间戳

2.事件日志actions（记录应用内一个具体的操作行为，操作类型，操作对象，操作对象描述）

点击

动作id

动作目标类型

动作目标id

动作时间

领劵

优惠卷

优惠卷id

13位时间戳

3.曝光日志displays（记录页面所曝光的内容，类似于一些推荐弹窗或者搜索的推荐广告商品，曝光对象，曝光类型）

曝光

曝光类型

商品

搜索结果

曝光顺序

曝光对象id

曝光对象类型

1001

11

4.启动日志start（记录应用的启动信息）

启动

启动加载时间

开屏广告id

广告播放总时间

用户跳过广告时间点

启动入口

启动时间

图标

500

1002

3000

1200

13位时间戳

5.错误日志err（错误编码，错误信息）

Error\_code 错误码

Msg 错误信息

1. 建表语句（ods层业务数据与日志数据）

|  |
| --- |
| 1. *--ods* drop table if exists gmall.ods\_log; CREATE EXTERNAL TABLE gmall.ods\_log (`line` string) PARTITIONED BY (`dt` string) *-- 按照时间创建分区* STORED AS *-- 指定存储方式，读数据采用LzoTextInputFormat；* INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_log' *-- 指定数据在hdfs上的存储位置* ;  *--装载数据* load data inpath '/origin\_data/gmall/log/topic\_log/2022-01-14' into table ods\_log partition(dt='2022-01-14');  *--建立索引 --hadoop jar /opt/app/hadoop-3.1.3/share/hadoop/common/hadoop-lzo-0.4.20.jar com.hadoop.compression.lzo.DistributedLzoIndexer /warehouse/gmall/ods/ods\_log/dt=2022-01-14  --业务数据* DROP TABLE IF EXISTS gmall.ods\_activity\_info; CREATE EXTERNAL TABLE gmall.ods\_activity\_info(  `id` STRING COMMENT '编号',  `activity\_name` STRING COMMENT '活动名称',  `activity\_type` STRING COMMENT '活动类型',  `start\_time` STRING COMMENT '开始时间',  `end\_time` STRING COMMENT '结束时间',  `create\_time` STRING COMMENT '创建时间' ) COMMENT '活动信息表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_activity\_info/';  DROP TABLE IF EXISTS gmall.ods\_activity\_rule; CREATE EXTERNAL TABLE gmall.ods\_activity\_rule(  `id` STRING COMMENT '编号',  `activity\_id` STRING COMMENT '活动ID',  `activity\_type` STRING COMMENT '活动类型',  `condition\_amount` DECIMAL(16,2) COMMENT '满减金额',  `condition\_num` BIGINT COMMENT '满减件数',  `benefit\_amount` DECIMAL(16,2) COMMENT '优惠金额',  `benefit\_discount` DECIMAL(16,2) COMMENT '优惠折扣',  `benefit\_level` STRING COMMENT '优惠级别' ) COMMENT '活动规则表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_activity\_rule/';  DROP TABLE IF EXISTS gmall.ods\_base\_category1; CREATE EXTERNAL TABLE gmall.ods\_base\_category1(  `id` STRING COMMENT 'id',  `name` STRING COMMENT '名称' ) COMMENT '商品一级分类表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_base\_category1/';  DROP TABLE IF EXISTS gmall.ods\_base\_category2; CREATE EXTERNAL TABLE gmall.ods\_base\_category2(  `id` STRING COMMENT ' id',  `name` STRING COMMENT '名称',  `category1\_id` STRING COMMENT '一级品类id' ) COMMENT '商品二级分类表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_base\_category2/';  DROP TABLE IF EXISTS gmall.ods\_base\_category3; CREATE EXTERNAL TABLE gmall.ods\_base\_category3(  `id` STRING COMMENT ' id',  `name` STRING COMMENT '名称',  `category2\_id` STRING COMMENT '二级品类id' ) COMMENT '商品三级分类表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_base\_category3/';  DROP TABLE IF EXISTS gmall.ods\_base\_dic; CREATE EXTERNAL TABLE gmall.ods\_base\_dic(  `dic\_code` STRING COMMENT '编号',  `dic\_name` STRING COMMENT '编码名称',  `parent\_code` STRING COMMENT '父编码',  `create\_time` STRING COMMENT '创建日期',  `operate\_time` STRING COMMENT '操作日期' ) COMMENT '编码字典表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_base\_dic/';  DROP TABLE IF EXISTS gmall.ods\_base\_province; CREATE EXTERNAL TABLE gmall.ods\_base\_province (  `id` STRING COMMENT '编号',  `name` STRING COMMENT '省份名称',  `region\_id` STRING COMMENT '地区ID',  `area\_code` STRING COMMENT '地区编码',  `iso\_code` STRING COMMENT 'ISO-3166编码，供可视化使用',  `iso\_3166\_2` STRING COMMENT 'IOS-3166-2编码，供可视化使用' ) COMMENT '省份表' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_base\_province/';  DROP TABLE IF EXISTS gmall.ods\_base\_region; CREATE EXTERNAL TABLE gmall.ods\_base\_region (  `id` STRING COMMENT '编号',  `region\_name` STRING COMMENT '地区名称' ) COMMENT '地区表' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_base\_region/';  DROP TABLE IF EXISTS gmall.ods\_base\_trademark; CREATE EXTERNAL TABLE gmall.ods\_base\_trademark (  `id` STRING COMMENT '编号',  `tm\_name` STRING COMMENT '品牌名称' ) COMMENT '品牌表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_base\_trademark/';  DROP TABLE IF EXISTS gmall.ods\_cart\_info; CREATE EXTERNAL TABLE gmall.ods\_cart\_info(  `id` STRING COMMENT '编号',  `user\_id` STRING COMMENT '用户id',  `sku\_id` STRING COMMENT 'skuid',  `cart\_price` DECIMAL(16,2) COMMENT '放入购物车时价格',  `sku\_num` BIGINT COMMENT '数量',  `sku\_name` STRING COMMENT 'sku名称 (冗余)',  `create\_time` STRING COMMENT '创建时间',  `operate\_time` STRING COMMENT '修改时间',  `is\_ordered` STRING COMMENT '是否已经下单',  `order\_time` STRING COMMENT '下单时间',  `source\_type` STRING COMMENT '来源类型',  `source\_id` STRING COMMENT '来源编号' ) COMMENT '加购表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_cart\_info/';  DROP TABLE IF EXISTS gmall.ods\_comment\_info; CREATE EXTERNAL TABLE gmall.ods\_comment\_info(  `id` STRING COMMENT '编号',  `user\_id` STRING COMMENT '用户ID',  `sku\_id` STRING COMMENT '商品sku',  `spu\_id` STRING COMMENT '商品spu',  `order\_id` STRING COMMENT '订单ID',  `appraise` STRING COMMENT '评价',  `create\_time` STRING COMMENT '评价时间' ) COMMENT '商品评论表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_comment\_info/';  DROP TABLE IF EXISTS gmall.ods\_coupon\_info; CREATE EXTERNAL TABLE gmall.ods\_coupon\_info(  `id` STRING COMMENT '购物券编号',  `coupon\_name` STRING COMMENT '购物券名称',  `coupon\_type` STRING COMMENT '购物券类型 1 现金券 2 折扣券 3 满减券 4 满件打折券',  `condition\_amount` DECIMAL(16,2) COMMENT '满额数',  `condition\_num` BIGINT COMMENT '满件数',  `activity\_id` STRING COMMENT '活动编号',  `benefit\_amount` DECIMAL(16,2) COMMENT '减金额',  `benefit\_discount` DECIMAL(16,2) COMMENT '折扣',  `create\_time` STRING COMMENT '创建时间',  `range\_type` STRING COMMENT '范围类型 1、商品 2、品类 3、品牌',  `limit\_num` BIGINT COMMENT '最多领用次数',  `taken\_count` BIGINT COMMENT '已领用次数',  `start\_time` STRING COMMENT '开始领取时间',  `end\_time` STRING COMMENT '结束领取时间',  `operate\_time` STRING COMMENT '修改时间',  `expire\_time` STRING COMMENT '过期时间' ) COMMENT '优惠券表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_coupon\_info/';  DROP TABLE IF EXISTS gmall.ods\_coupon\_use; CREATE EXTERNAL TABLE gmall.ods\_coupon\_use(  `id` STRING COMMENT '编号',  `coupon\_id` STRING COMMENT '优惠券ID',  `user\_id` STRING COMMENT 'skuid',  `order\_id` STRING COMMENT 'spuid',  `coupon\_status` STRING COMMENT '优惠券状态',  `get\_time` STRING COMMENT '领取时间',  `using\_time` STRING COMMENT '使用时间(下单)',  `used\_time` STRING COMMENT '使用时间(支付)',  `expire\_time` STRING COMMENT '过期时间' ) COMMENT '优惠券领用表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_coupon\_use/';  DROP TABLE IF EXISTS gmall.ods\_favor\_info; CREATE EXTERNAL TABLE gmall.ods\_favor\_info(  `id` STRING COMMENT '编号',  `user\_id` STRING COMMENT '用户id',  `sku\_id` STRING COMMENT 'skuid',  `spu\_id` STRING COMMENT 'spuid',  `is\_cancel` STRING COMMENT '是否取消',  `create\_time` STRING COMMENT '收藏时间',  `cancel\_time` STRING COMMENT '取消时间' ) COMMENT '商品收藏表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_favor\_info/';  DROP TABLE IF EXISTS gmall.ods\_order\_detail; CREATE EXTERNAL TABLE gmall.ods\_order\_detail(  `id` STRING COMMENT '编号',  `order\_id` STRING COMMENT '订单号',  `sku\_id` STRING COMMENT '商品id',  `sku\_name` STRING COMMENT '商品名称',  `order\_price` DECIMAL(16,2) COMMENT '商品价格',  `sku\_num` BIGINT COMMENT '商品数量',  `create\_time` STRING COMMENT '创建时间',  `source\_type` STRING COMMENT '来源类型',  `source\_id` STRING COMMENT '来源编号',  `split\_final\_amount` DECIMAL(16,2) COMMENT '分摊最终金额',  `split\_activity\_amount` DECIMAL(16,2) COMMENT '分摊活动优惠',  `split\_coupon\_amount` DECIMAL(16,2) COMMENT '分摊优惠券优惠' ) COMMENT '订单详情表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_order\_detail/';  DROP TABLE IF EXISTS gmall.ods\_order\_detail\_activity; CREATE EXTERNAL TABLE gmall.ods\_order\_detail\_activity(  `id` STRING COMMENT '编号',  `order\_id` STRING COMMENT '订单号',  `order\_detail\_id` STRING COMMENT '订单明细id',  `activity\_id` STRING COMMENT '活动id',  `activity\_rule\_id` STRING COMMENT '活动规则id',  `sku\_id` BIGINT COMMENT '商品id',  `create\_time` STRING COMMENT '创建时间' ) COMMENT '订单详情活动关联表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_order\_detail\_activity/';  DROP TABLE IF EXISTS gmall.ods\_order\_detail\_coupon; CREATE EXTERNAL TABLE gmall.ods\_order\_detail\_coupon(  `id` STRING COMMENT '编号',  `order\_id` STRING COMMENT '订单号',  `order\_detail\_id` STRING COMMENT '订单明细id',  `coupon\_id` STRING COMMENT '优惠券id',  `coupon\_use\_id` STRING COMMENT '优惠券领用记录id',  `sku\_id` STRING COMMENT '商品id',  `create\_time` STRING COMMENT '创建时间' ) COMMENT '订单详情活动关联表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_order\_detail\_coupon/';  DROP TABLE IF EXISTS gmall.ods\_order\_info; CREATE EXTERNAL TABLE gmall.ods\_order\_info (  `id` STRING COMMENT '订单号',  `final\_amount` DECIMAL(16,2) COMMENT '订单最终金额',  `order\_status` STRING COMMENT '订单状态',  `user\_id` STRING COMMENT '用户id',  `payment\_way` STRING COMMENT '支付方式',  `delivery\_address` STRING COMMENT '送货地址',  `out\_trade\_no` STRING COMMENT '支付流水号',  `create\_time` STRING COMMENT '创建时间',  `operate\_time` STRING COMMENT '操作时间',  `expire\_time` STRING COMMENT '过期时间',  `tracking\_no` STRING COMMENT '物流单编号',  `province\_id` STRING COMMENT '省份ID',  `activity\_reduce\_amount` DECIMAL(16,2) COMMENT '活动减免金额',  `coupon\_reduce\_amount` DECIMAL(16,2) COMMENT '优惠券减免金额',  `original\_amount` DECIMAL(16,2) COMMENT '订单原价金额',  `feight\_fee` DECIMAL(16,2) COMMENT '运费',  `feight\_fee\_reduce` DECIMAL(16,2) COMMENT '运费减免' ) COMMENT '订单表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_order\_info/';  DROP TABLE IF EXISTS gmall.ods\_order\_refund\_info; CREATE EXTERNAL TABLE gmall.ods\_order\_refund\_info(  `id` STRING COMMENT '编号',  `user\_id` STRING COMMENT '用户ID',  `order\_id` STRING COMMENT '订单ID',  `sku\_id` STRING COMMENT '商品ID',  `refund\_type` STRING COMMENT '退单类型',  `refund\_num` BIGINT COMMENT '退单件数',  `refund\_amount` DECIMAL(16,2) COMMENT '退单金额',  `refund\_reason\_type` STRING COMMENT '退单原因类型',  `refund\_status` STRING COMMENT '退单状态',*--退单状态应包含买家申请、卖家审核、卖家收货、退款完成等状态。此处未涉及到，故该表按增量处理* `create\_time` STRING COMMENT '退单时间' ) COMMENT '退单表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_order\_refund\_info/';  DROP TABLE IF EXISTS gmall.ods\_order\_status\_log; CREATE EXTERNAL TABLE gmall.ods\_order\_status\_log (  `id` STRING COMMENT '编号',  `order\_id` STRING COMMENT '订单ID',  `order\_status` STRING COMMENT '订单状态',  `operate\_time` STRING COMMENT '修改时间' ) COMMENT '订单状态表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_order\_status\_log/';  DROP TABLE IF EXISTS gmall.ods\_payment\_info; CREATE EXTERNAL TABLE gmall.ods\_payment\_info(  `id` STRING COMMENT '编号',  `out\_trade\_no` STRING COMMENT '对外业务编号',  `order\_id` STRING COMMENT '订单编号',  `user\_id` STRING COMMENT '用户编号',  `payment\_type` STRING COMMENT '支付类型',  `trade\_no` STRING COMMENT '交易编号',  `payment\_amount` DECIMAL(16,2) COMMENT '支付金额',  `subject` STRING COMMENT '交易内容',  `payment\_status` STRING COMMENT '支付状态',  `create\_time` STRING COMMENT '创建时间',  `callback\_time` STRING COMMENT '回调时间' ) COMMENT '支付流水表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_payment\_info/';  DROP TABLE IF EXISTS gmall.ods\_refund\_payment; CREATE EXTERNAL TABLE gmall.ods\_refund\_payment(  `id` STRING COMMENT '编号',  `out\_trade\_no` STRING COMMENT '对外业务编号',  `order\_id` STRING COMMENT '订单编号',  `sku\_id` STRING COMMENT 'SKU编号',  `payment\_type` STRING COMMENT '支付类型',  `trade\_no` STRING COMMENT '交易编号',  `refund\_amount` DECIMAL(16,2) COMMENT '支付金额',  `subject` STRING COMMENT '交易内容',  `refund\_status` STRING COMMENT '支付状态',  `create\_time` STRING COMMENT '创建时间',  `callback\_time` STRING COMMENT '回调时间' ) COMMENT '支付流水表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_refund\_payment/';  DROP TABLE IF EXISTS gmall.ods\_sku\_attr\_value; CREATE EXTERNAL TABLE gmall.ods\_sku\_attr\_value(  `id` STRING COMMENT '编号',  `attr\_id` STRING COMMENT '平台属性ID',  `value\_id` STRING COMMENT '平台属性值ID',  `sku\_id` STRING COMMENT '商品ID',  `attr\_name` STRING COMMENT '平台属性名称',  `value\_name` STRING COMMENT '平台属性值名称' ) COMMENT 'sku平台属性表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_sku\_attr\_value/';  DROP TABLE IF EXISTS gmall.ods\_sku\_info; CREATE EXTERNAL TABLE gmall.ods\_sku\_info(  `id` STRING COMMENT 'skuId',  `spu\_id` STRING COMMENT 'spuid',  `price` DECIMAL(16,2) COMMENT '价格',  `sku\_name` STRING COMMENT '商品名称',  `sku\_desc` STRING COMMENT '商品描述',  `weight` DECIMAL(16,2) COMMENT '重量',  `tm\_id` STRING COMMENT '品牌id',  `category3\_id` STRING COMMENT '品类id',  `is\_sale` STRING COMMENT '是否在售',  `create\_time` STRING COMMENT '创建时间' ) COMMENT 'SKU商品表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_sku\_info/';  DROP TABLE IF EXISTS gmall.ods\_sku\_sale\_attr\_value; CREATE EXTERNAL TABLE gmall.ods\_sku\_sale\_attr\_value(  `id` STRING COMMENT '编号',  `sku\_id` STRING COMMENT 'sku\_id',  `spu\_id` STRING COMMENT 'spu\_id',  `sale\_attr\_value\_id` STRING COMMENT '销售属性值id',  `sale\_attr\_id` STRING COMMENT '销售属性id',  `sale\_attr\_name` STRING COMMENT '销售属性名称',  `sale\_attr\_value\_name` STRING COMMENT '销售属性值名称' ) COMMENT 'sku销售属性名称' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_sku\_sale\_attr\_value/';  DROP TABLE IF EXISTS gmall.ods\_spu\_info; CREATE EXTERNAL TABLE gmall.ods\_spu\_info(  `id` STRING COMMENT 'spuid',  `spu\_name` STRING COMMENT 'spu名称',  `category3\_id` STRING COMMENT '品类id',  `tm\_id` STRING COMMENT '品牌id' ) COMMENT 'SPU商品表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_spu\_info/';  DROP TABLE IF EXISTS gmall.ods\_user\_info; CREATE EXTERNAL TABLE gmall.ods\_user\_info(  `id` STRING COMMENT '用户id',  `login\_name` STRING COMMENT '用户名称',  `nick\_name` STRING COMMENT '用户昵称',  `name` STRING COMMENT '用户姓名',  `phone\_num` STRING COMMENT '手机号码',  `email` STRING COMMENT '邮箱',  `user\_level` STRING COMMENT '用户等级',  `birthday` STRING COMMENT '生日',  `gender` STRING COMMENT '性别',  `create\_time` STRING COMMENT '创建时间',  `operate\_time` STRING COMMENT '操作时间' ) COMMENT '用户表' PARTITIONED BY (`dt` STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat' LOCATION '/warehouse/gmall/ods/ods\_user\_info/'; |

四，脚本

1.集群文件分发脚本

|  |
| --- |
| if [ $# -lt 1 ]  then  echo Not Enough Arguement!  exit;  fi  #2. 遍历集群所有机器  for host in n1 n2 n3  do  echo ==================== $host ====================  #3. 遍历所有目录，挨个发送  for file in $@  do  #4. 判断文件是否存在  if [ -e $file ]  then  #5. 获取父目录  pdir=$(cd -P $(dirname $file); pwd)  #6. 获取当前文件的名称  fname=$(basename $file)  ssh $host "mkdir -p $pdir"  rsync -av $pdir/$fname $host:$pdir  else  echo $file does not exists!  fi  done  done |

2.集群服务查看脚本

|  |
| --- |
| #! /bin/bash    for i in n1 n2 n3  do  echo --------- $i ----------  ssh $i "$\*"  done |

3.zookeeper群起群关脚本

|  |
| --- |
| #!/bin/bash  case $1 in  "start")  for i in n1 ne2 n3  do  echo --------- zookeeper $i 启动-----------  ssh $i "bash /opt/app/zookeeper/bin/zkServer.sh start"  done  ;;  "stop")  for i in n1 n2 n3  do  echo --------- zookeeper $i 停止-----------  ssh $i "bash /opt/app/zookeeper/bin/zkServer.sh stop"  done  ;;  "status")  for i in n1 n2 n3  do  echo --------- zookeeper $i 状态-----------  ssh $i "bahs /opt/app/zookeeper/bin/zkServer.sh status"  done  ;;  esac |

4.kafka群起群关脚本

|  |
| --- |
| #!/bin/bash  case $1 in  "start")  for i in n1 n2 n3  do  echo "\*\*\*\*\*\*\*\*\*\*\*\*\*\*$i\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"  ssh $i '/opt/app/kafka/bin/kafka-server-start.sh -daemon /opt/app/kafka/config/server.properties'  done  ;;  "stop")  for i in n1 n2 n3  do  echo "\*\*\*\*\*\*\*\*\*\*\*\*\*\*$i\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"  ssh $i '/opt/app/kafka/bin/kafka-server-stop.sh'  done  ;;  esac |

5.集群状态监控启动关闭脚本

|  |
| --- |
| #!/bin/bash  case $1 in  "start"){  echo '----- 启动 prometheus -----'  nohup /opt/app/prometheus/prometheus --web.enable-admin-api --config.file=/opt/app/prometheus/prometheus.yml > /opt/app/prometheus/prometheus.log 2>&1 &  echo '----- 启动 pushgateway -----'  nohup /opt/app/pushgateway/pushgateway --web.listen-address :9091 > /opt/app/pushgateway/pushgateway.log 2>&1 &  echo '----- 启动 grafana -----'  nohup /opt/app/grafana/bin/grafana-server --homepath /opt/app/grafana web > /opt/app/grafana/grafana.log 2>&1 &  for host in n1 n2 n3  do  echo '----- 启动node\_export $host -----'  ssh $host "nohup /opt/app/node\_exporter/node\_exporter > /opt/app/node\_exporter/node\_exporter.log 2>&1 &"  done  };;  "stop"){  echo '----- 停止 grafana -----'  pgrep -f grafana | xargs kill  echo '----- 停止 pushgateway -----'  pgrep -f pushgateway | xargs kill  echo '----- 停止 prometheus -----'  pgrep -f prometheus | xargs kill  echo '----- 停止 node\_export -----'  for host in n1 n2 n3  do  echo '----- 停止node\_export $host -----'  ssh $host "pgrep -f grafana | xargs kill"  done  };;  esac |