基于阿里云搭建实时数据仓库项目

阿里云大学 & 尚硅谷 联合出品

课程目标

- 1) 学习搭建一个**实时数据仓库**, 掌握数据**采集、存储、计算、输出、展示**等整个业务流程。
- 2)整个实时数据仓库系统是在**阿里云架构上搭建**,掌握并学会运用**各个服务组件**,及 各个组件之间如何**联动**。
 - 3)前置知识要求
 - 熟练掌握SQL语法
 - 对Hadoop大数据体系有一定的了解

第1章 课程目录

- 1. 项目需求及架构设计
 - 1.1 项目需求分析
 - 1.2 项目框架
 - 1.2.1 阿里云技术框架
 - 1.2.2 技术选型
 - 1.2.3 系统架构设计
 - 1.2.4 业务流程
 - 1.3 电商表结构

- 2.业务数据准备
- 3.缓冲数据
- 4.同步业务数据
- 5.实时数仓分层
- 6.数据可视化

1.1 项目需求分析

- 1) 实时采集埋点日志数据
- 2)实时采集业务数据库中数据
- 3)对数据进行清洗和处理
- 4)保存数据到分析型数据库
- 5)对结果进行可视化展示



1.2.1 阿里云技术框架

阿里云产品	简介	类比
ECS	弹性服务器	Linux服务器
RDS	关系型数据库	MySql
DataHub	数据总线	Kafka +各种服务接口
实时计算	实时计算	Spark、Flink
DataWorks (Stream Studio)	可 视 化 StreamCompute 的开发管理平台	目前没有
AnalyticDB for MySql	分析型数据库	MySql集群
DataV	可视化数据展示工具	Tableau、Echarts、Kibana

1.2.2 技术选型

数据采集传输:

数据存储:

阿里云框架

DataHub、DTS

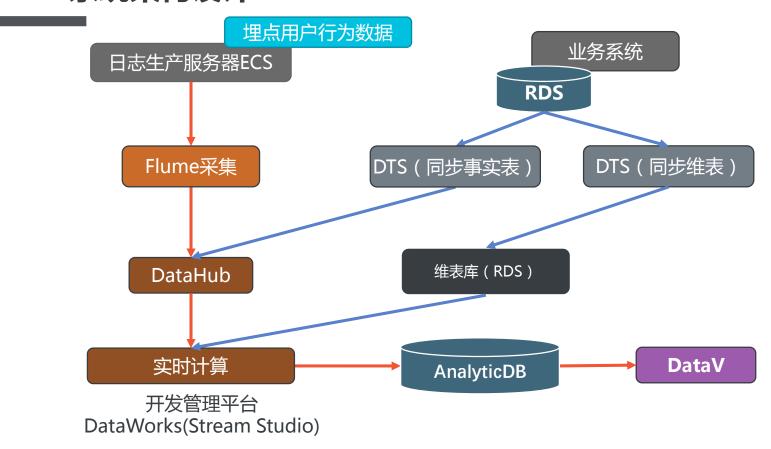
Flume、Kafka、Canal、MaxWell

开源框架

RDS、AnalyticDB MySql、Hadoop、HBase

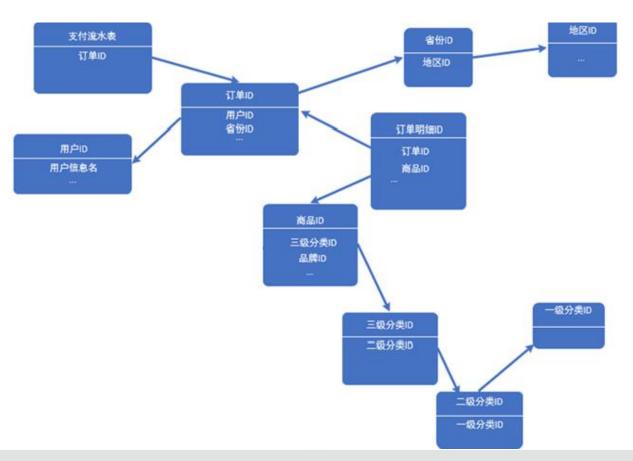
数据可视化: DataV 、QuickBI Tableau、Echarts、Kibana

1.2.3 系统架构设计



1.2.4 业务流程 业务系统 order detail 埋点用户行为数据 order info RDS 日志生产服务器ECS payment info user info sku info base_region Flume采集 base_category1 DTS同步维表 DTS同步事实表 base_category2 base_province base category3 DataHub | ods层 (order_detail和order_info) dwd层(dwd_paid_order_detail) 聚合商品 聚合省份 维表库(RDS) 实时计算 dim sku info dim province ods_order_detail和 ods_order_info 双流join ads_sku_stat dwd_paid_order_detail ads_province_stat 和dim_province dwd_paid_order_detail 存储结果 **DataV** 和dim_sku_info (AnalyticDB) join

1.3 电商表结构



课程说明

第2~6章,采用Word课件授课

