



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

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

阿里云框架

- 数据采集传输： DataHub、DTS
- 数据存储： RDS、AnalyticDB
- 数据计算： 实时计算
- 数据可视化： DataV、QuickBI

开源框架

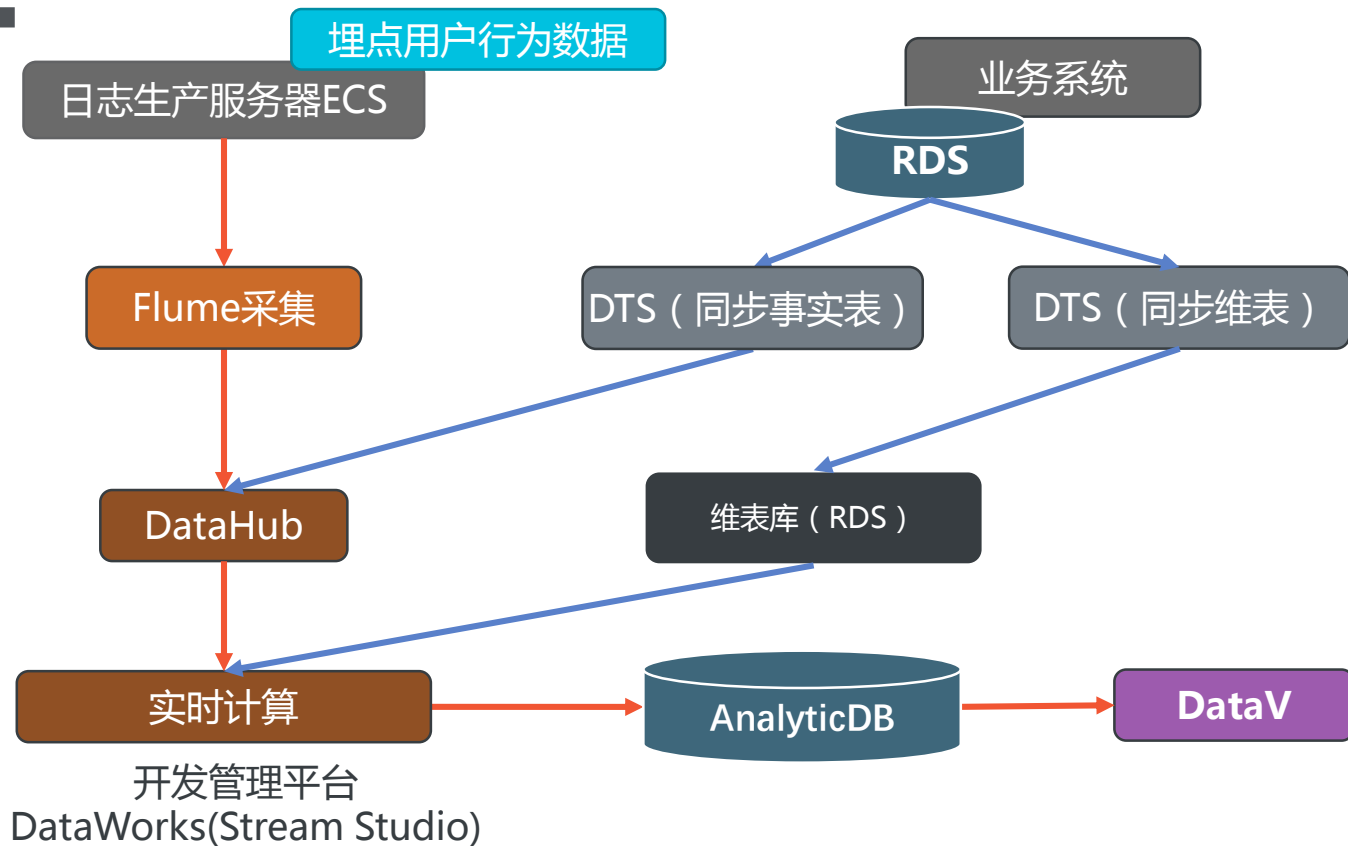
Flume、Kafka、Canal、MaxWell

MySQL、Hadoop、HBase

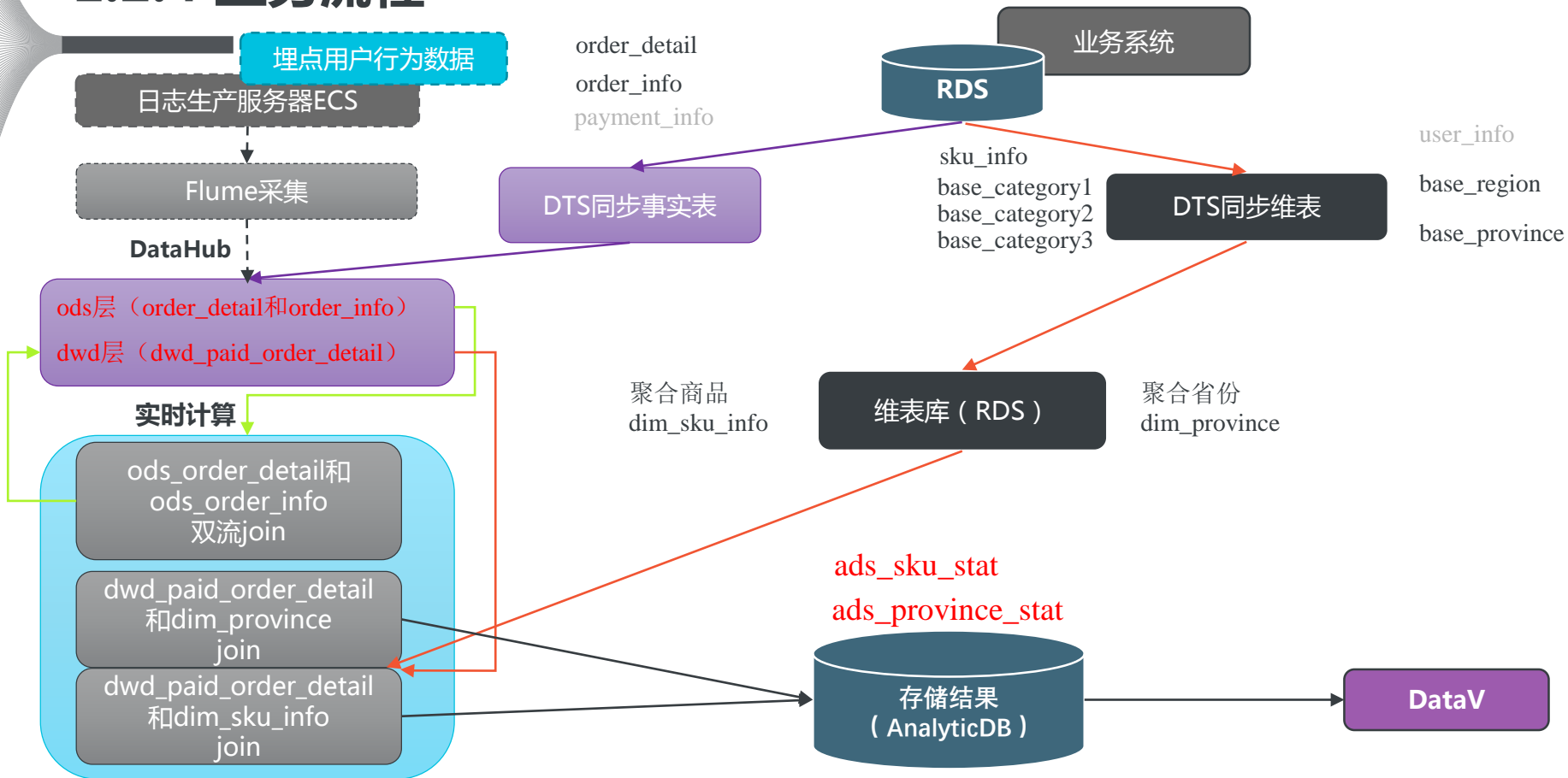
Spark、Flink

Tableau、Echarts、Kibana

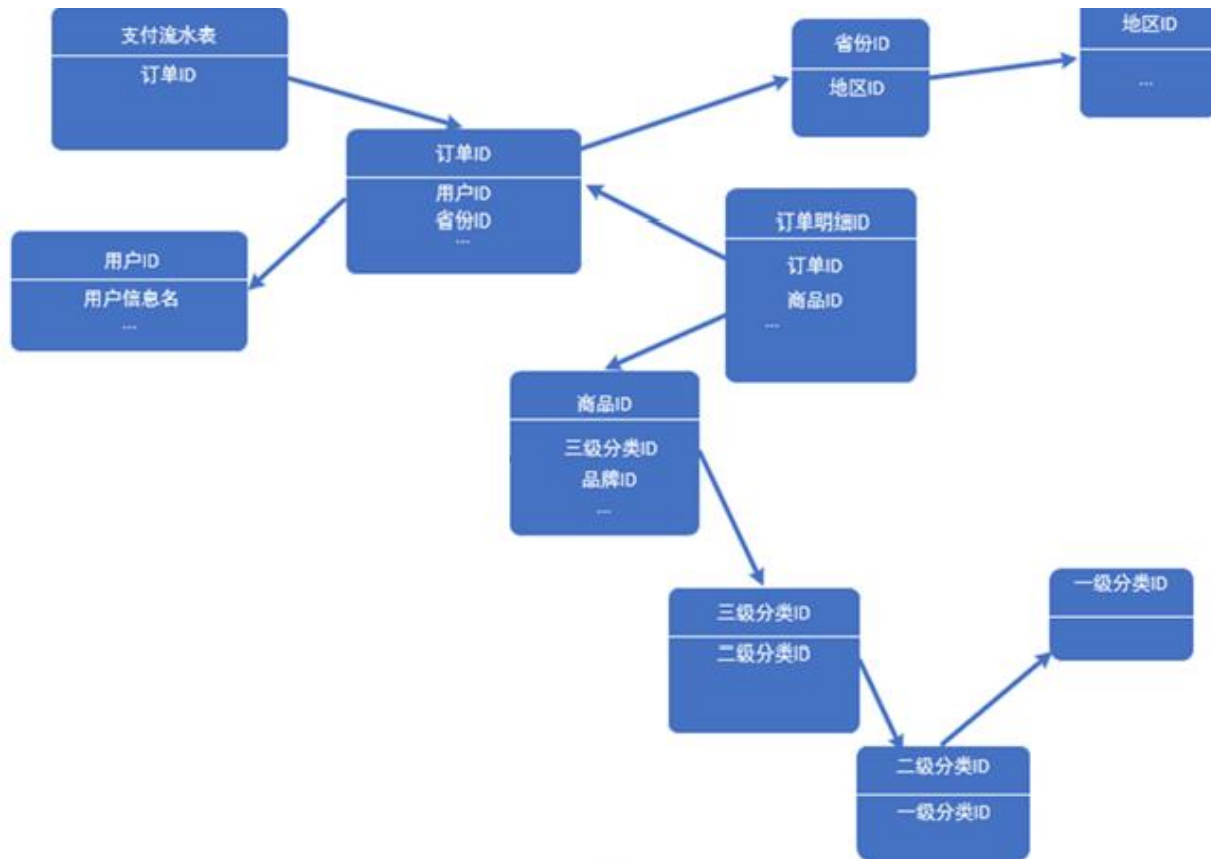
1.2.3 系统架构设计



1.2.4 业务流程



1.3 电商表结构



课程说明

第2~6章，采用Word课件授课

