

# 基于Flink的资讯场景实时数仓

本文介绍如何针对资讯聚合类业务场景搭建基于消息队列Kafka版和实时计算Flink的实时数仓。

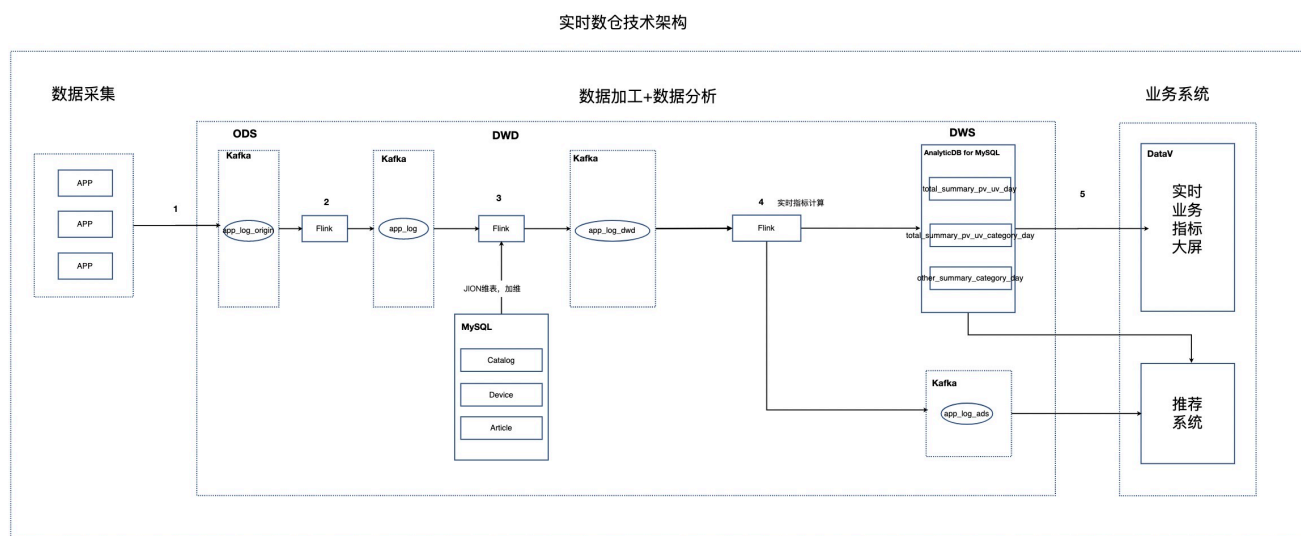
## 场景描述

本文首先介绍什么是实时数仓以及相关技术架构，接着介绍资讯聚合类业务的典型场景及其业务目标，并据此设计了相应的技术架构。然后介绍如何部署基础环境和搭建实时数仓，并介绍业务系统如何使用实时数仓。

## 解决的问题

- 通过消息队列Kafka版和实时计算Flink实现实时ETL和数据流。
- 通过消息队列Kafka版和实时计算Flink实现实时数据分析。
- 通过消息队列Kafka版和实时计算Flink实现事件触发。

## 部署架构图



## 选用的产品

### • 消息队列Kafka版

阿里云基于Apache Kafka构建的高吞吐量、高可扩展性的分布式消息队列服务，广泛用于日志收集、监控数据聚合、流式数据处理、在线和离线分析等，是大数据生态中不可或缺的产品之一，阿里云提供全托管服务，用户无需部署运维，更专业、更可靠、更安全。

更多关于消息队列Kafka版的介绍，参见[消息队列Kafka版产品详情页](#)。

### • 实时计算

实时计算（Alibaba Cloud Realtime Compute, Powered by Ververica）是阿里云提供的基于Apache Flink构建的企业级大数据计算平台。在PB级别的数据集上可以支持亚秒级别的处理延时，赋能用户标准实时数据处理流

程和行业解决方案；支持Datastream API作业开发，提供了批流统一的Flink SQL，简化BI场景下的开发；可与用户已使用的大数据组件无缝对接，更多增值特性助力企业实时化转型。

更多关于实时计算的介绍，参见[实时计算产品详情页](#)。

- **DataV数据可视化**

DataV旨在让更多的人看到数据可视化的魅力，帮助非专业的工程师通过图形化的界面轻松搭建专业水准的可视化应用，满足您会议展览、业务监控、风险预警、地理信息分析等多种业务的展示需求。

更多关于阿里云DataV数据可视化的介绍，参见[DataV数据可视化产品详情页](#)。

- **专有网络VPC**

专有网络VPC帮助您基于阿里云构建出一个隔离的网络环境，并可以自定义IP地址范围、网段、路由表和网关等；此外，也可以通过专线、VPN、GRE等连接方式实现云上VPC与传统IDC的互联，构建混合云业务。

更多关于专有网络VPC的介绍，参见[专有网络VPC产品详情页](#)。

- **云数据库RDS**

阿里云关系型数据库（Relational Database Service，简称RDS）是一种稳定可靠、可弹性伸缩的在线数据库服务。基于阿里云分布式文件系统和SSD盘高性能存储，RDS支持MySQL、SQL Server、PostgreSQL、PPAS（Postgre Plus Advanced Server，高度兼容Oracle数据库）和MariaDB TX引擎，并且提供了容灾、备份、恢复、监控、迁移等方面的全套解决方案，彻底解决数据库运维的烦恼。

更多关于云数据库RDS的介绍，参见[云数据库RDS产品文档](#)。

- **分析型数据库MySQL版**

分析型数据库MySQL版（AnalyticDB for MySQL）是一种高并发低延时的PB级实时数据仓库，全面兼容MySQL协议以及SQL:2003语法标准，可以毫秒级针对万亿级数据进行即时的多维分析透视和业务探索。

更多关于分析型数据库MySQL版的介绍，参见[分析型数据库MySQL版产品详情页](#)。

- **对象存储OSS**

阿里云对象存储服务（Object Storage Service，简称OSS），是阿里云提供的海量、安全、低成本、高可靠的云存储服务。

更多关于对象存储OSS的介绍，参见[对象存储OSS产品详情页](#)。