

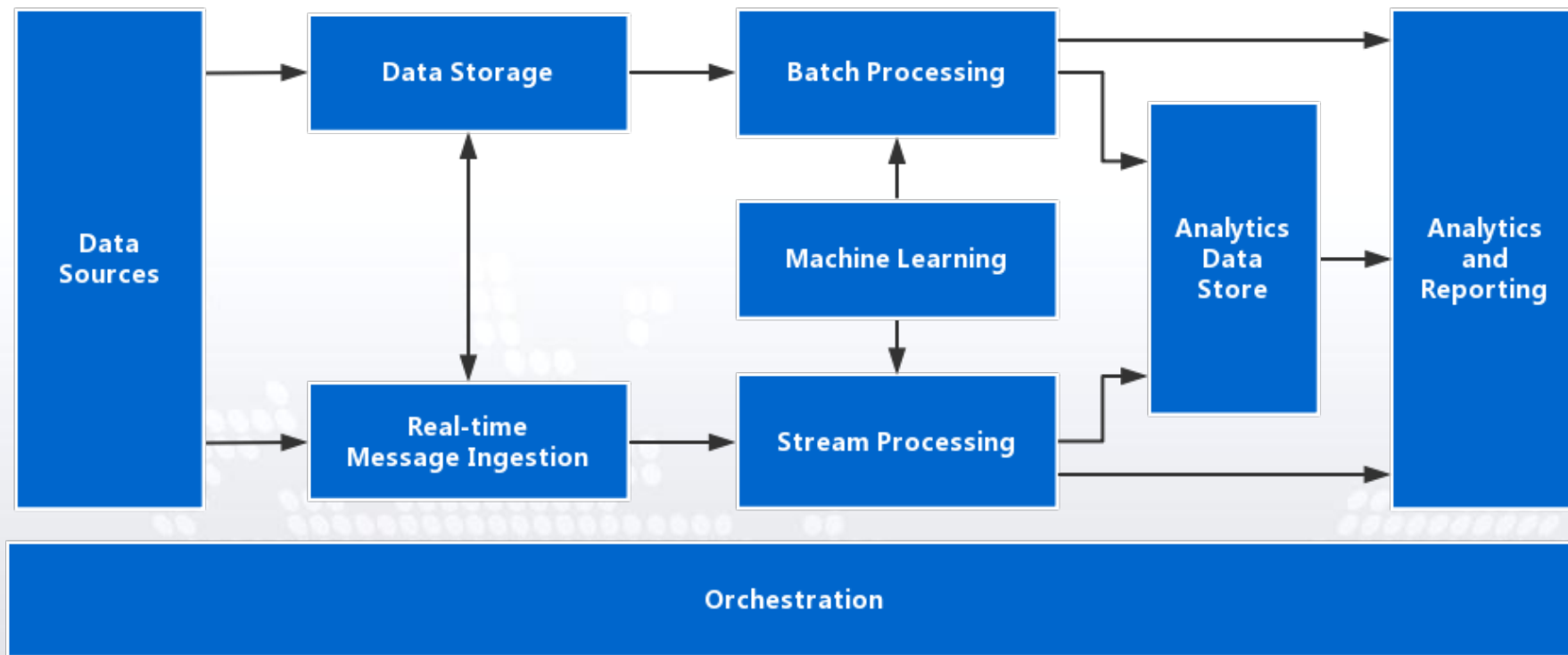
# 2019



## 企业级数据仓库实战



# 典型的数据流如何设计



Data Source：数据源

Data Storage：数据存储

Real-time message ingestion：实时消息接收通道

Batch Processing：批处理

Stream Processing：流处理

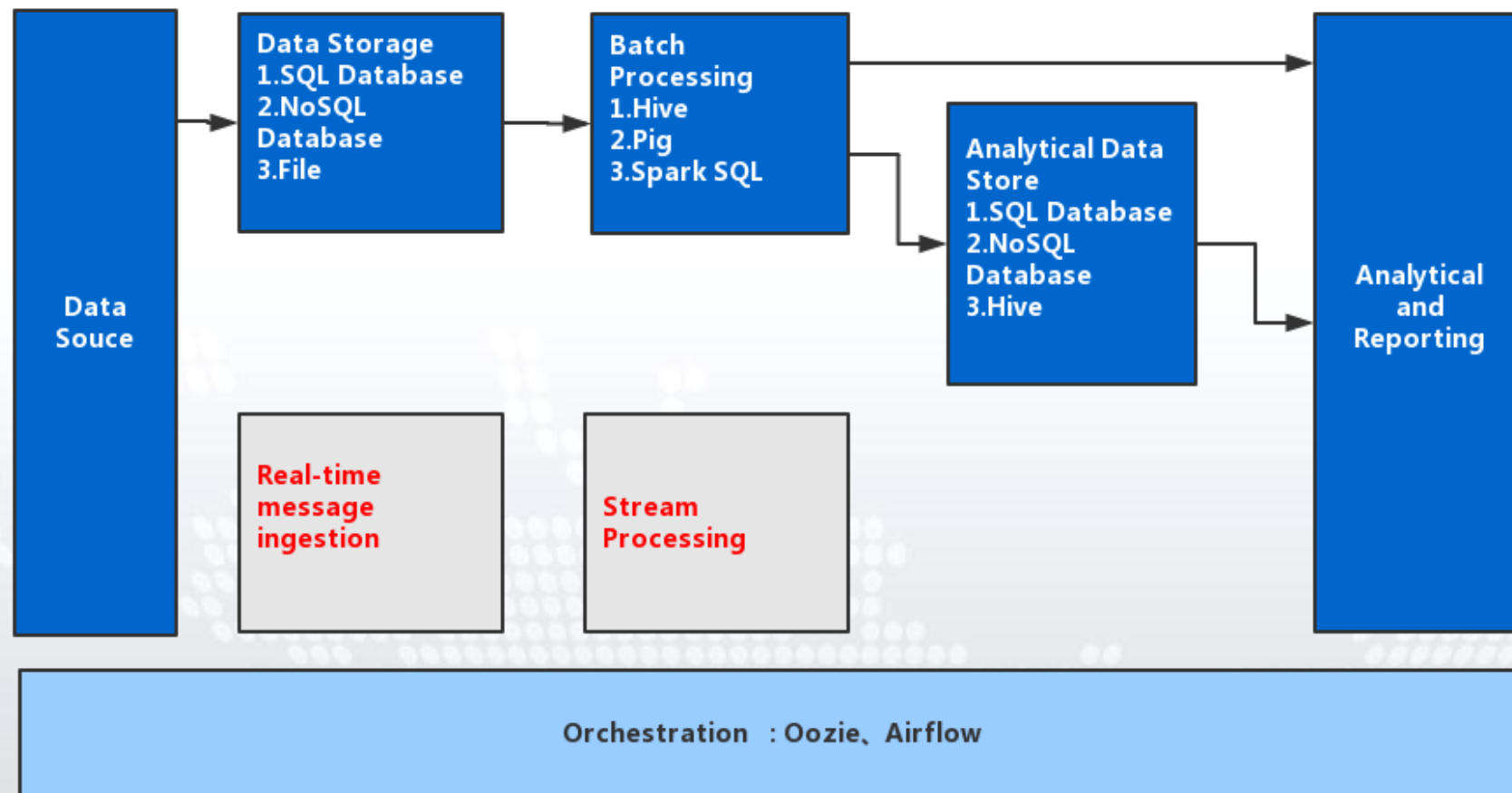
Orchestration：调度

Analytical data store：分析数据存储

Analytics and reporting：分析与报告



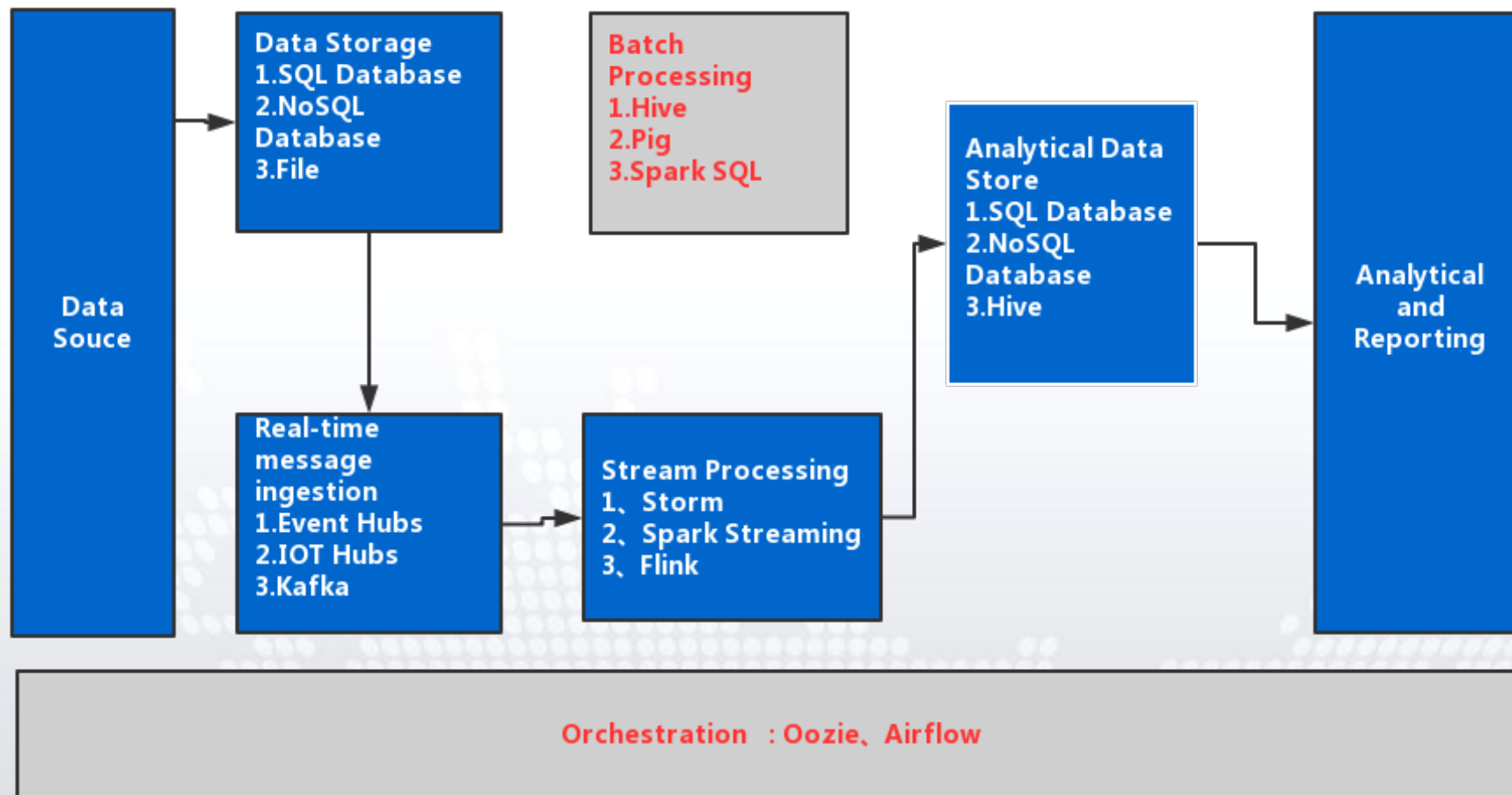
## 批处理



数据存储：一般主要分为两种，日志和表，日志主要是指用户接口日志，用户访问日志，表主要是指业务库表，这些业务库可能是关系型数据库如Mysql/Oracle/MSSQL，也可能是非关系型数据如MongoDB/Redis/Hbase等

批处理：主要使用Hadoop Hive、Pig、Spark SQL等进行处理

分析数据存储：一般可以存在Hive表中，或者关系型和非关系型数据库中



消息接收通道：一般主要为各类型的IOT通道、Kafka等

流处理：目前主流的开源解决方案是Storm、Spark Streaming、Flink等

分析数据存储：一般存储在HBase上，也可以存在Hive上



THANK YOU FOR YOUR GUIDANCE.

谢谢