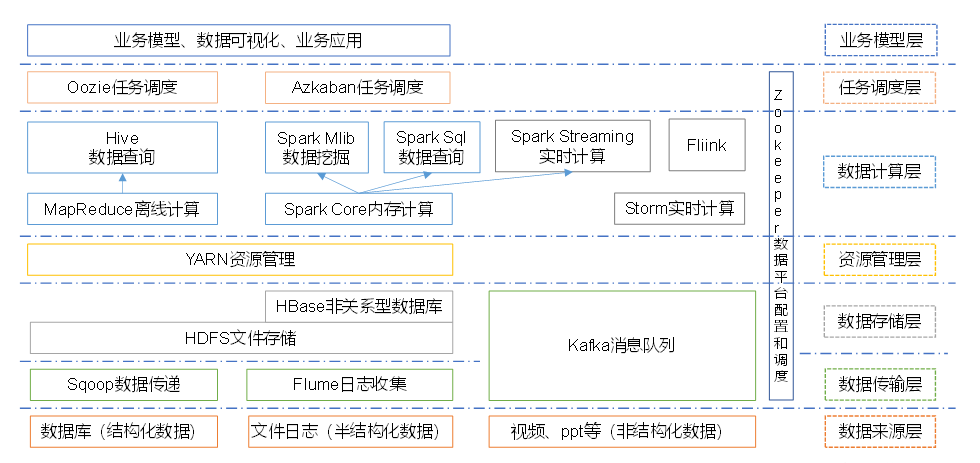
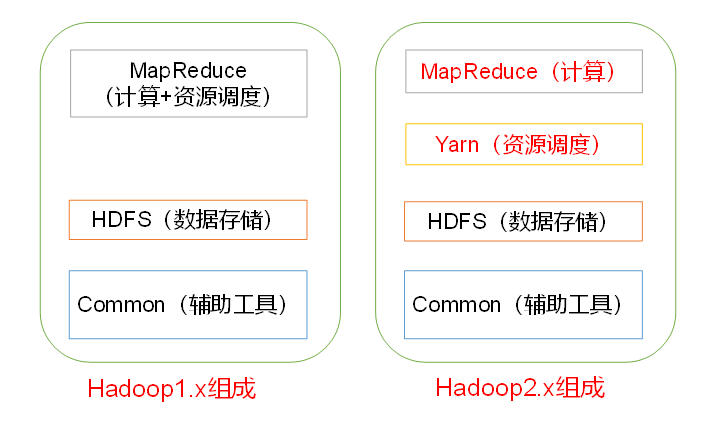
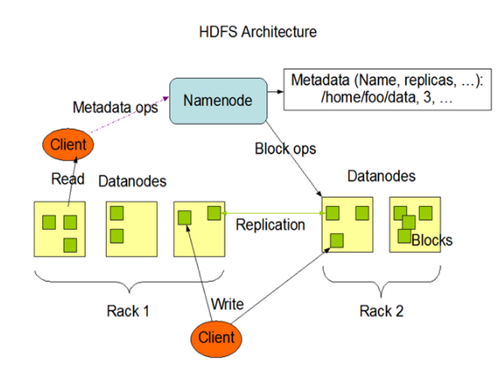
# 大数据技术生态体系



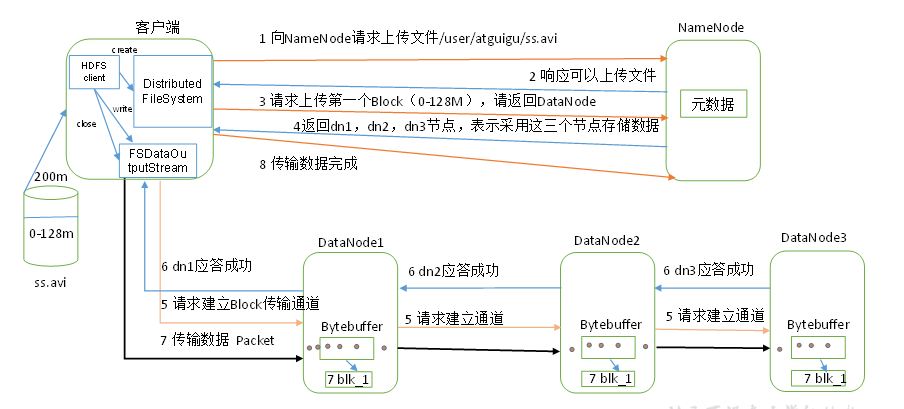
# Hadoop



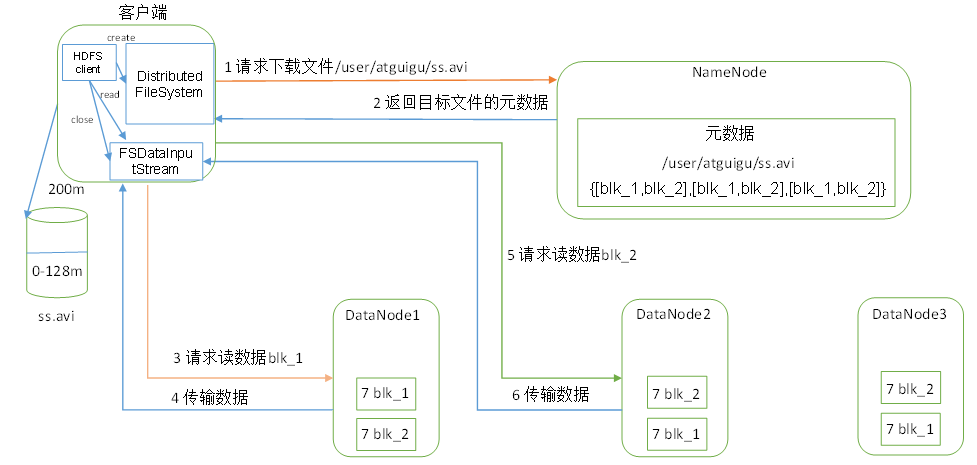
## HDFS



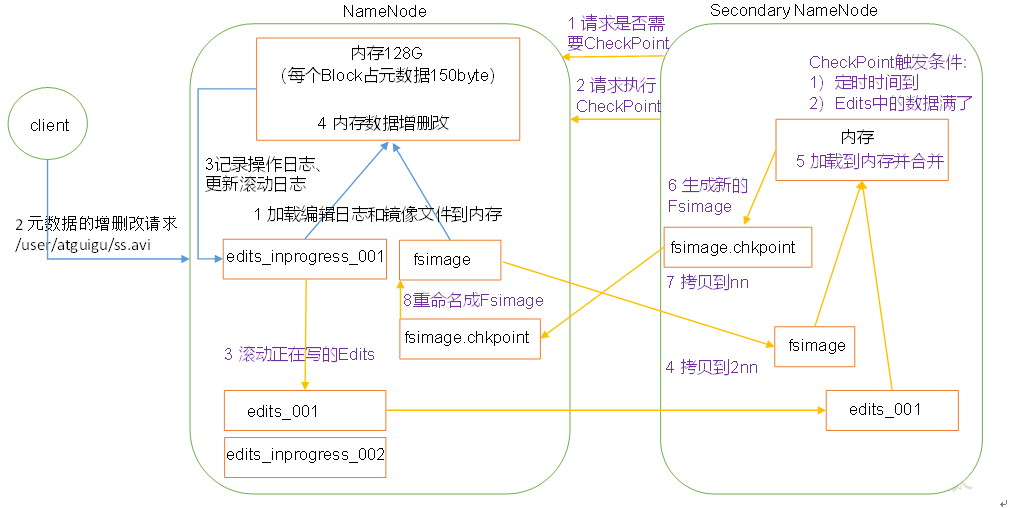
### HDFS的写数据流程



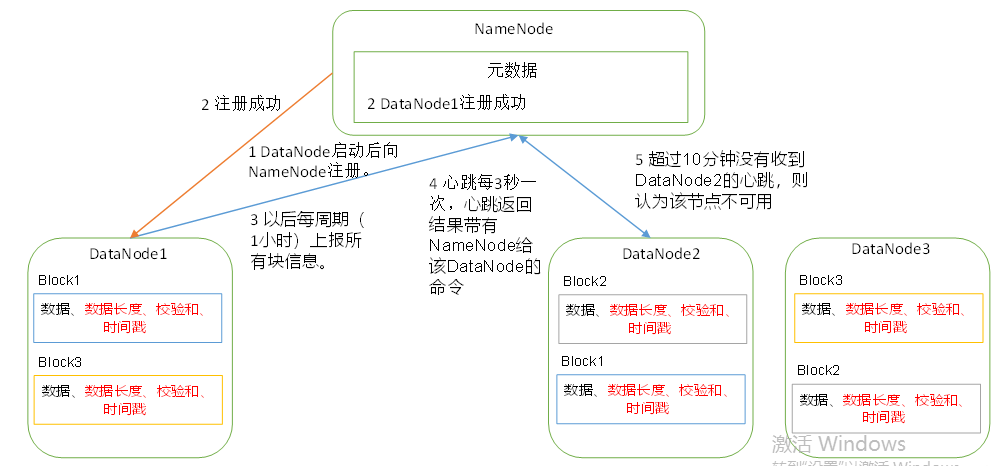
### HDFS的读数据流程



### NameNode工作机制

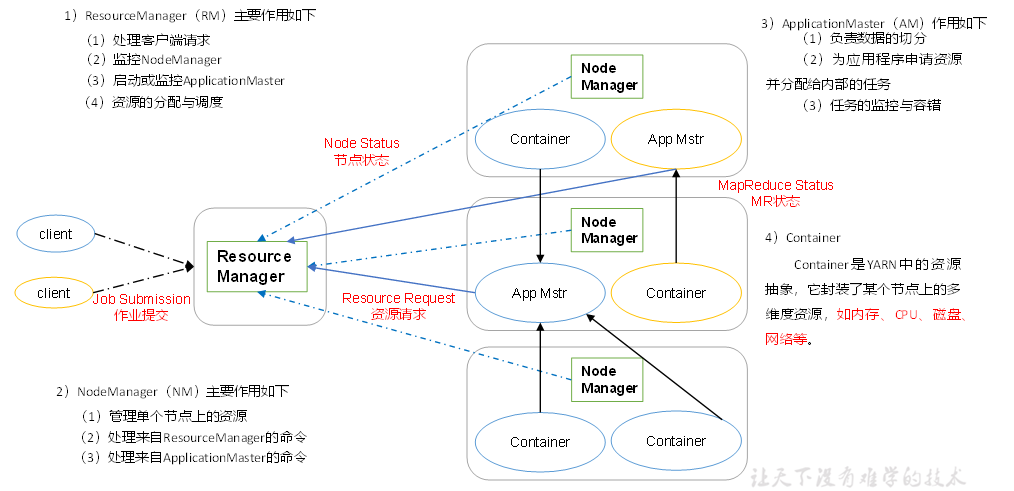


### DataNode工作机制

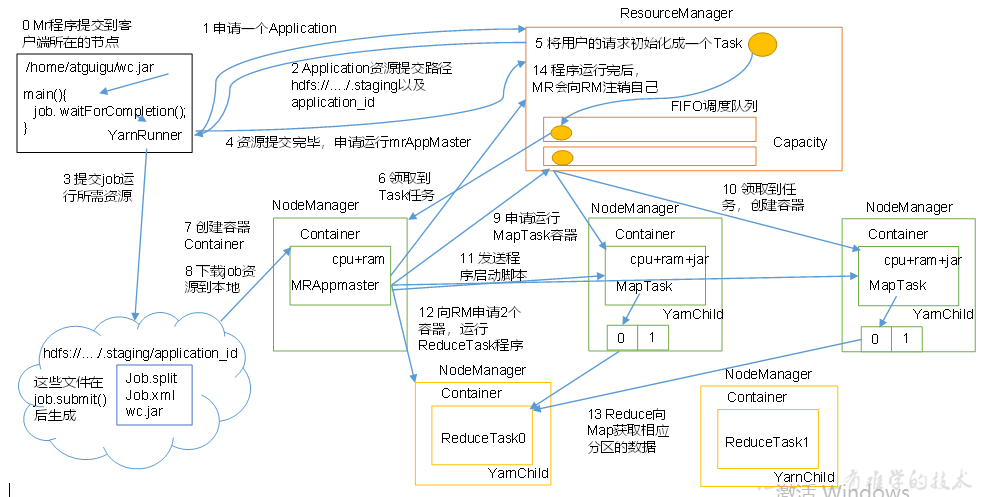


## YARN

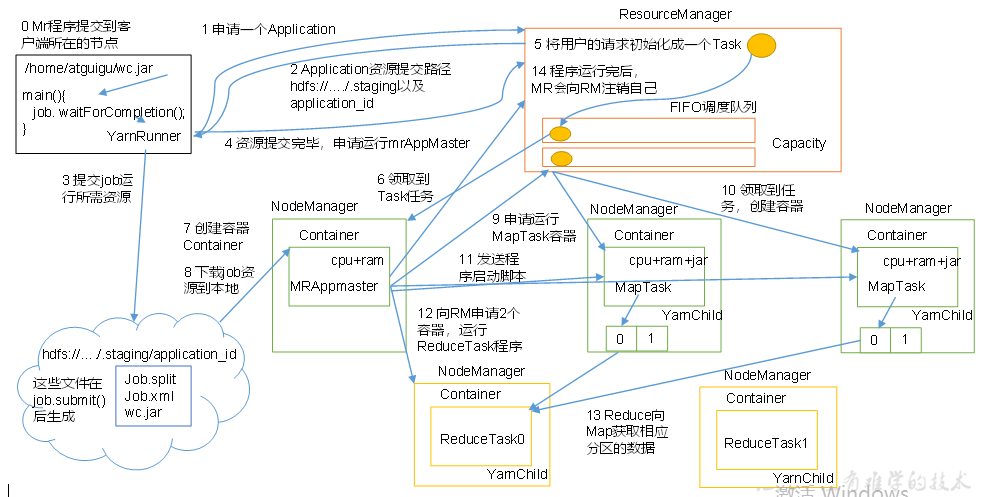
### YARN基本架构



### YARN工作机制

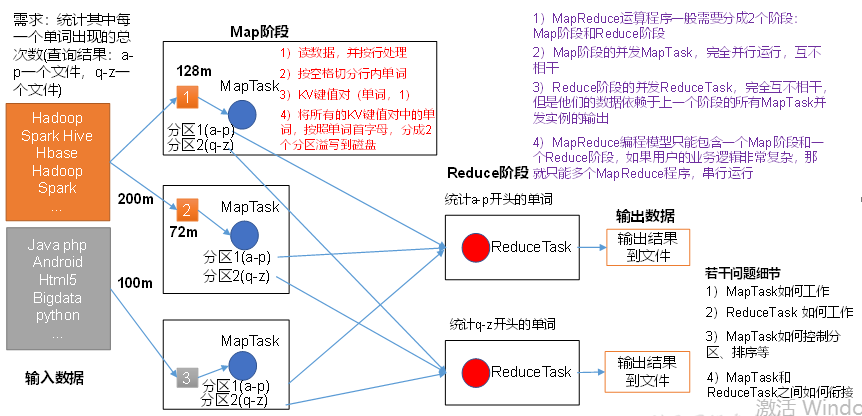


### YARN作业提交流程

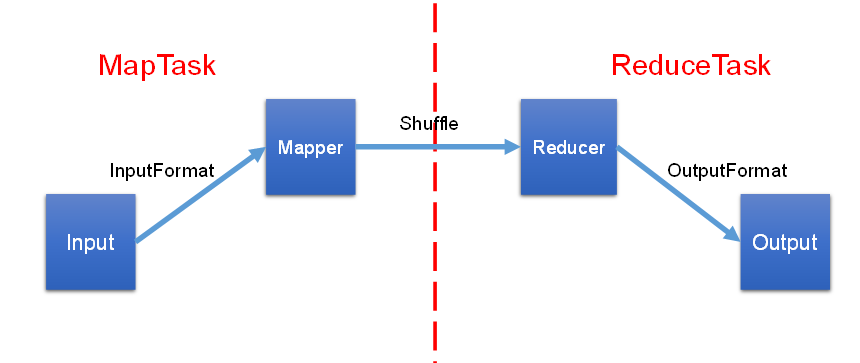


## 1.3 MapReduce

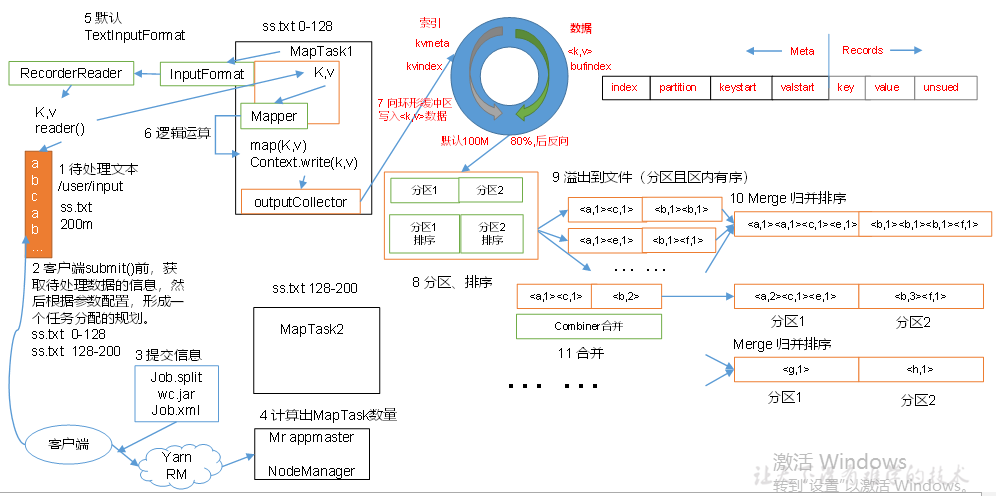
### 1.3.1 MapReduce核心思想

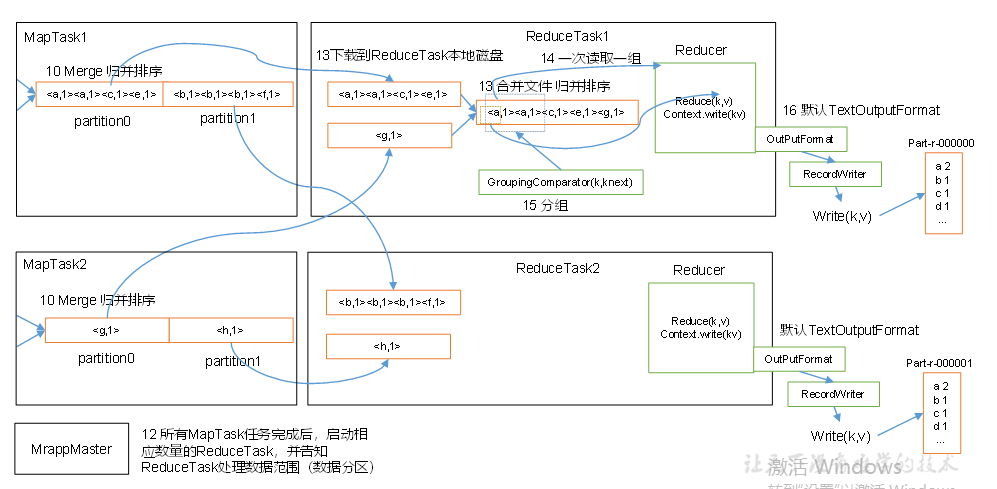
MapReduce数据流

### 1.3.2 MapReduce数据流

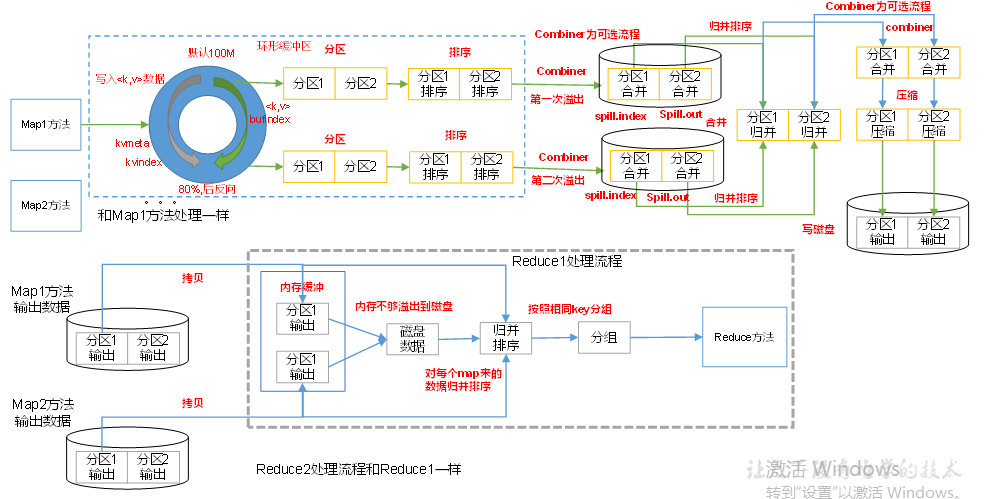


### 1.3.3 MapReduce工作流程

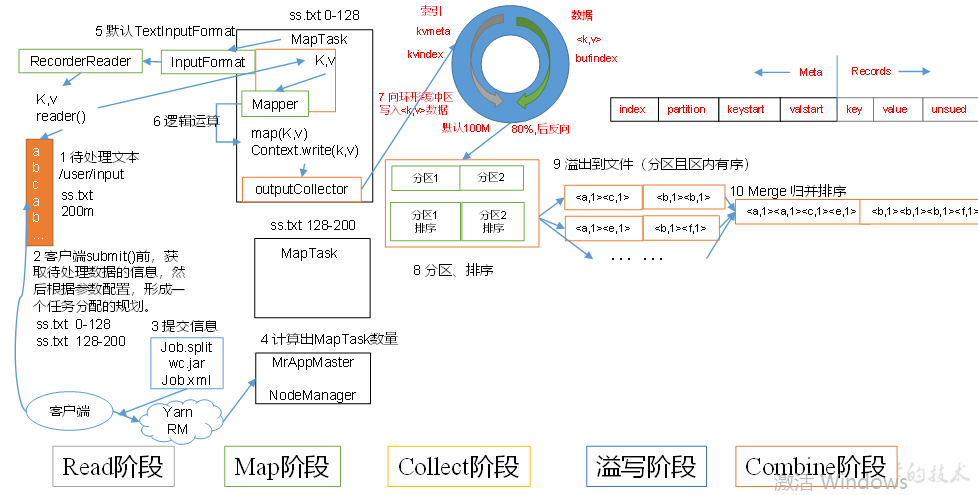




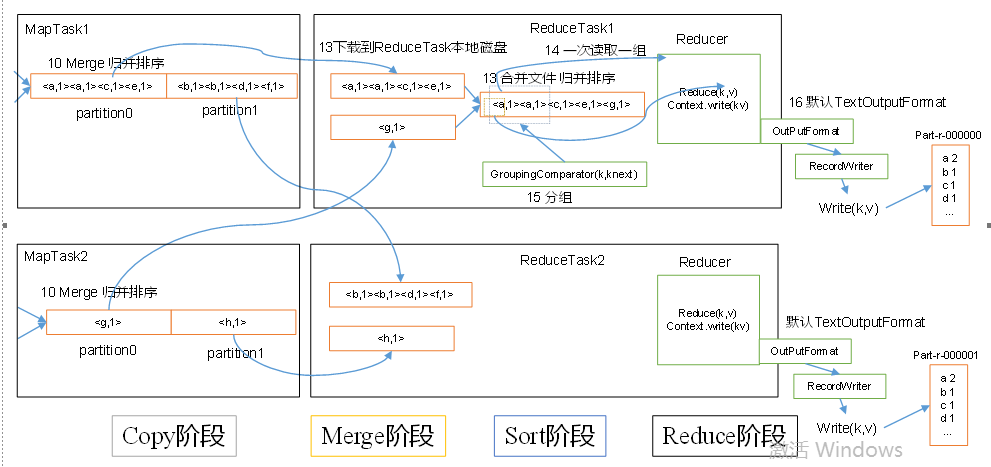
### 1.3.4 Shuffle机制



### 1.3.5 MapTask工作机制

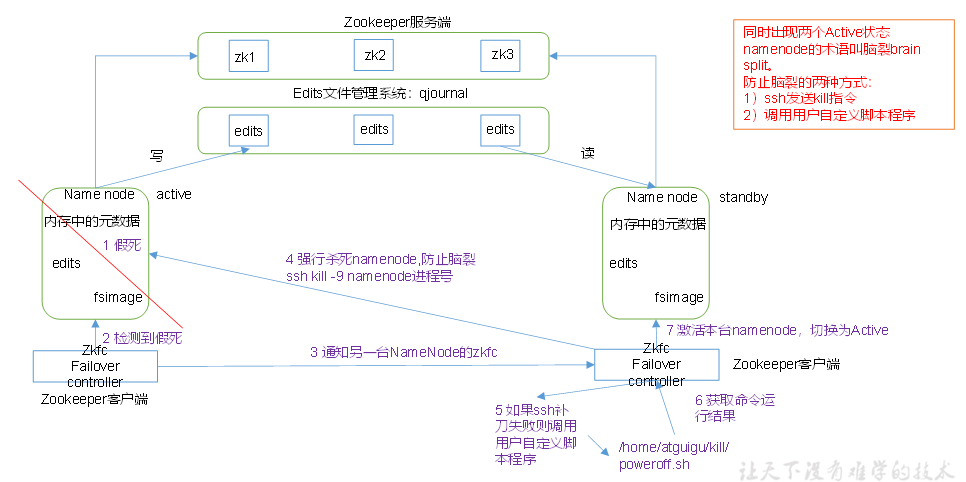


### 1.3.6 ReduceTask工作机制

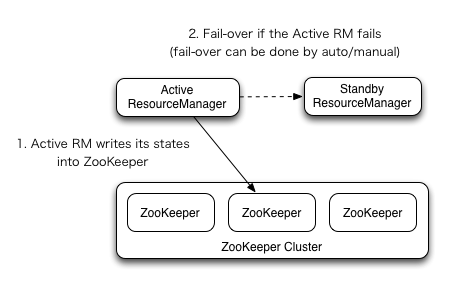


# HA

## 2.1 HDFS-HA自动故障转移工作机制

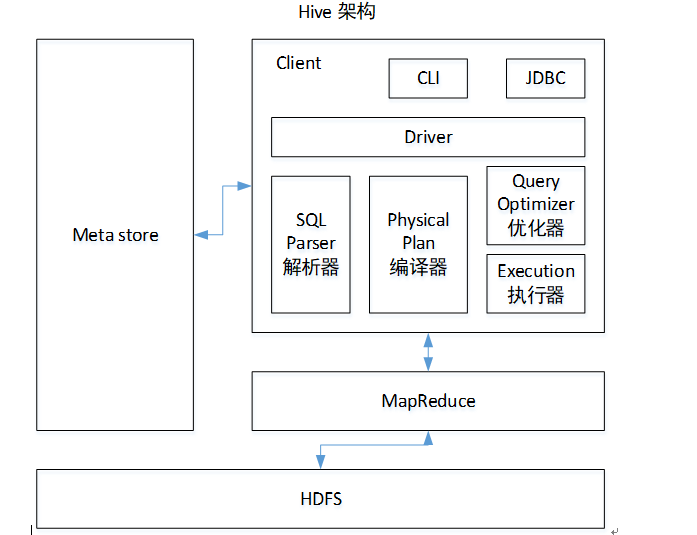


## 2.2 YARN-HA自动故障转移机制



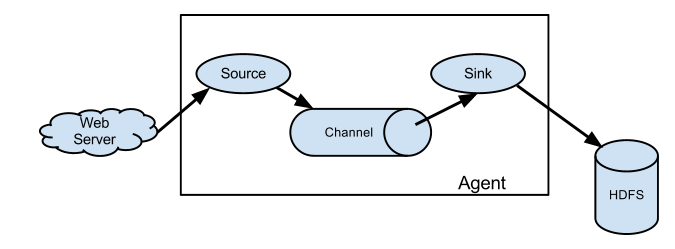
# Hive

## 3.1 Hive架构原理图

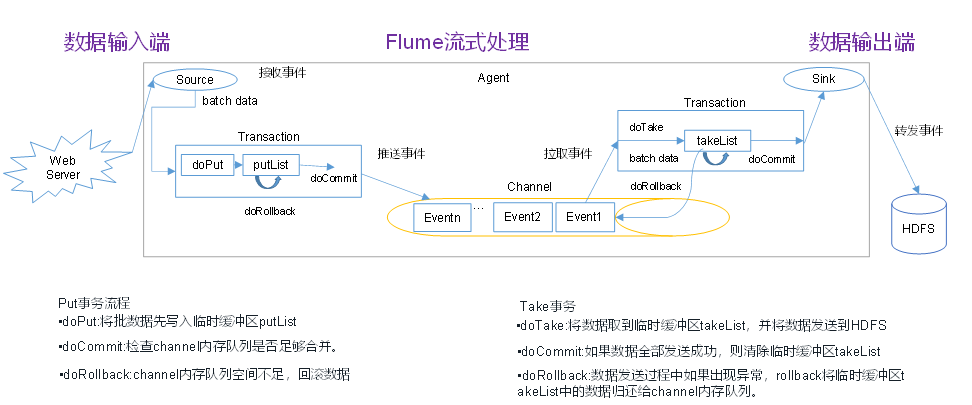


# Flume

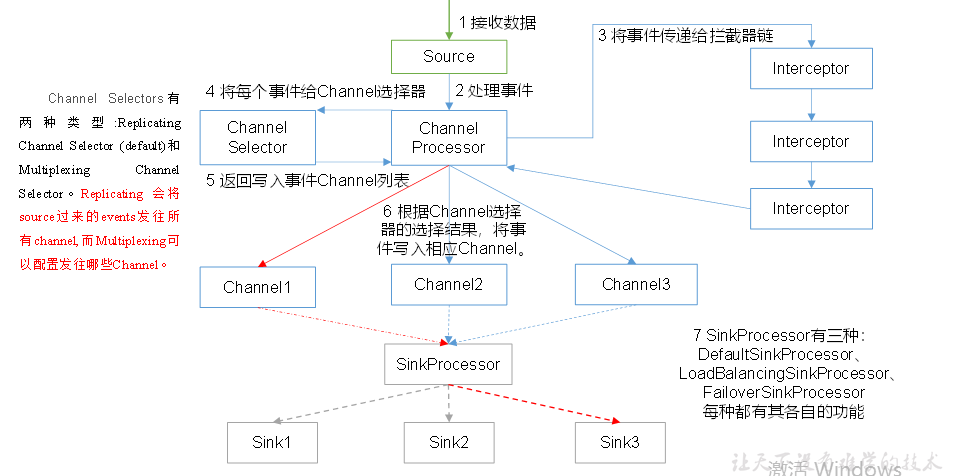
## 4.1 Flume架构



## 4.2 Flume事务



## 4.3 Flume Agent 内部原理

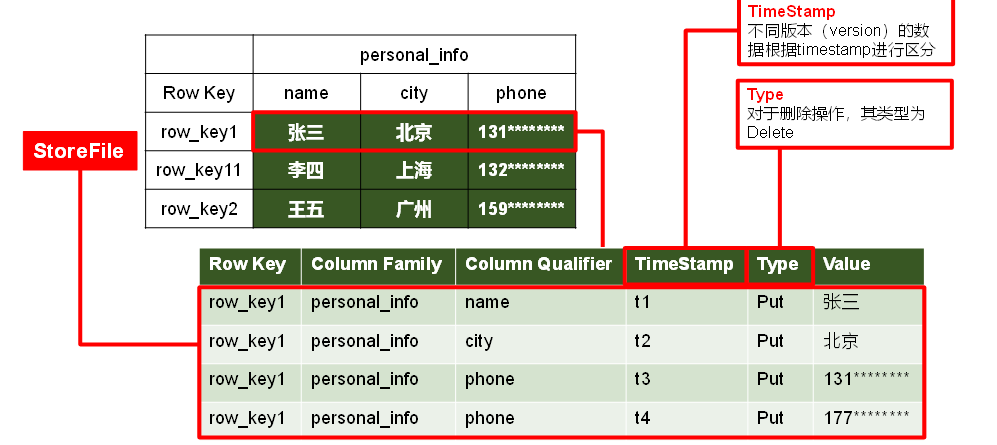


# Hbase

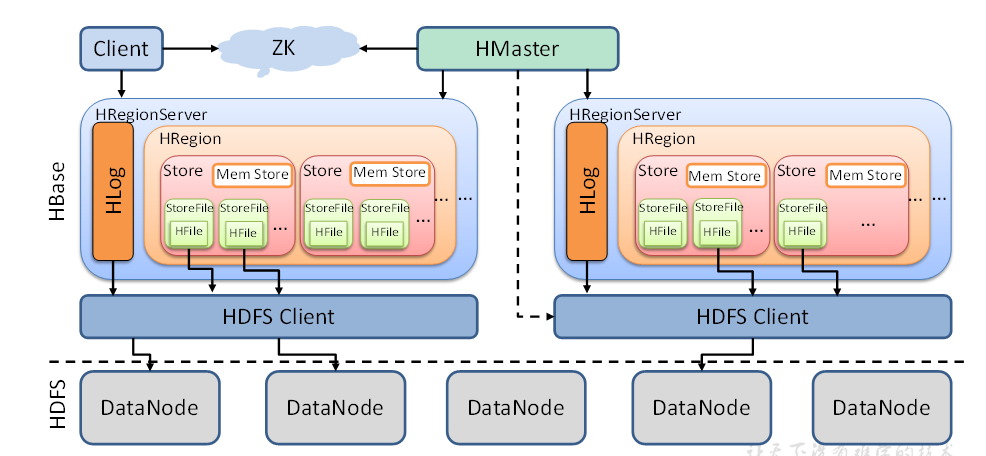
## 5.1 逻辑结构



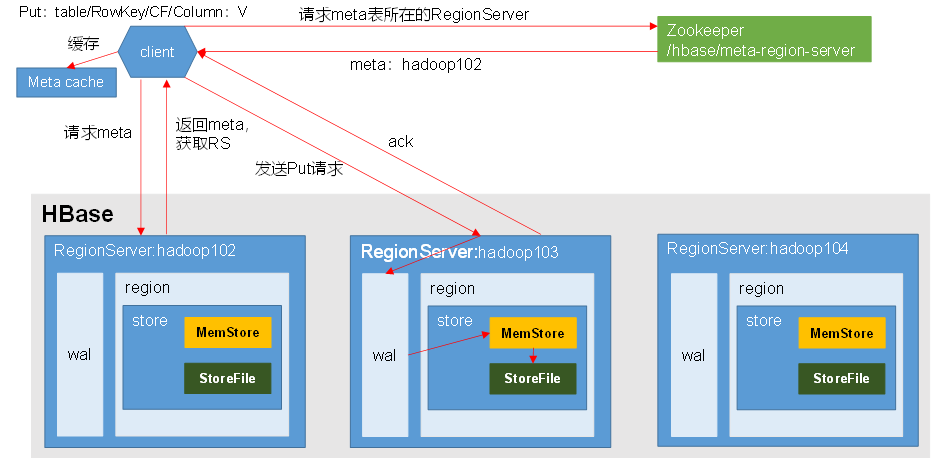
## 5.2 物理存储结构



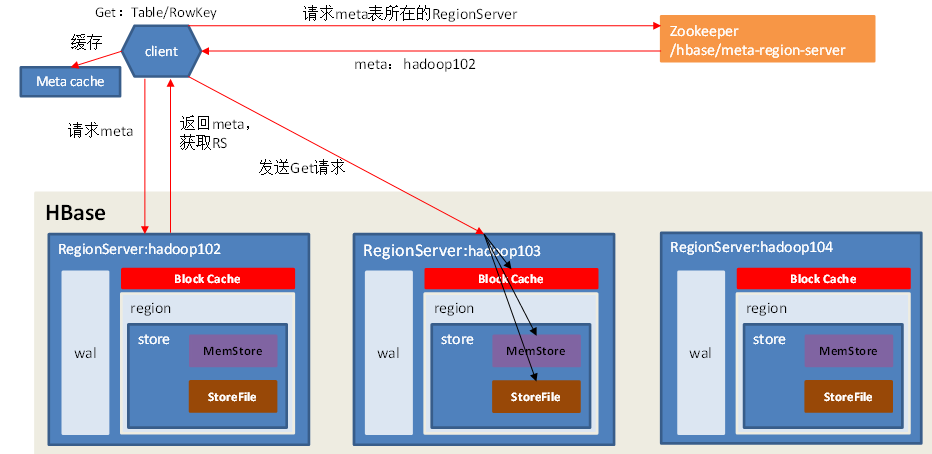
## 5.3 架构原理



## 5.4 写流程



## 5.5 读流程



# Azkaban

# Zookeeper

# Kafka

# 9. Scala

# 10. Spark

# 11. Flink