作业：高可用分布式日志收集系统

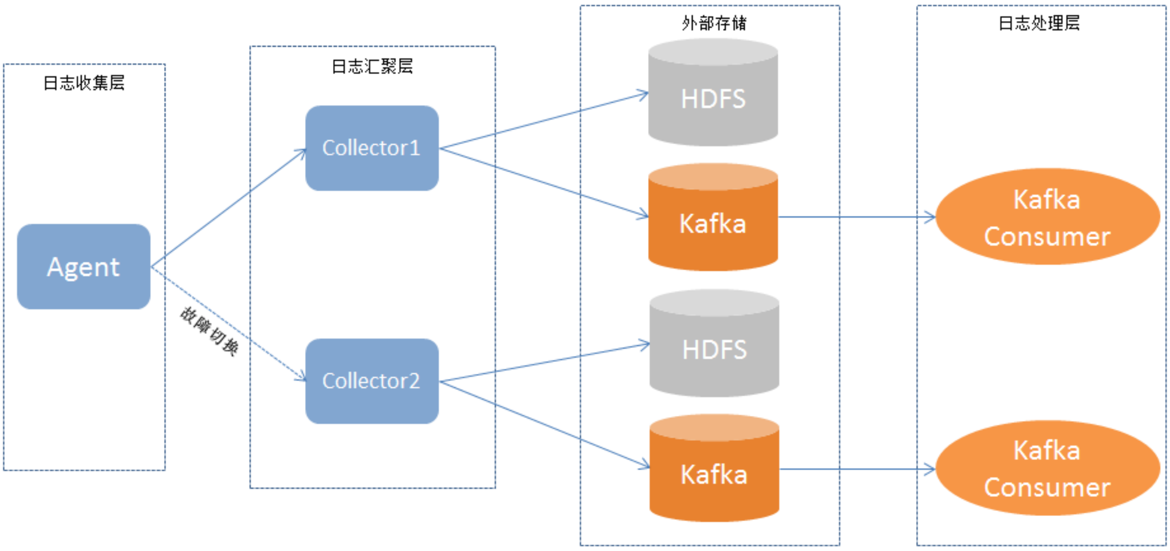
需求分析：

使用分层日志收集架构设计一个高可用的分布式日志收集系统，将收集到的日志数据分别发送到kafka的ad_log主题和HDFS的ad_log目录，目录格式如：ad_log/20190509。然后使用java编写Kafka Consumer来消费Kafka的ad_log主题的数据。

实现思路：

1.创建一个日志收集agent，用于收集服务器本地日志文件中的数据，agent中定义**容错处理器（failover sinkprocessor）**，一个agent对应多个collector，实现任何一个collector挂掉不也不影响系统的日志收集服务。

agent:　taildir source —> channel —> failover sinkprocessor —> avro sink



```
agent.sources = r1
agent.channels = c1
agent.sinks = k1 k2
agent.sources.r1.type = TAILDIR
agent.sources.r1.positionFile = /bigdata/flume/taildir/position/taildir_position.json
agent.sources.r1.filegroups = f1
agent.sources.r1.filegroups.f1 = /bigdata/taildir_log/test1/test.log
agent.sources.r1.channels = c1
agent.channels.c1.type = memory
agent.sinkgroups = g1
agent.sinkgroups.g1.sinks = k1 k2
agent.sinkgroups.g1.processor.type = failover
agent.sinkgroups.g1.processor.priority.k1 = 10
agent.sinkgroups.g1.processor.priority.k2 = 5
agent.sinks.k1.type = avro
agent.sinks.k1.channel = c1
agent.sinks.k1.hostname = 192.168.2.130
agent.sinks.k1.port = 8888
agent.sinks.k2.type = avro
agent.sinks.k2.channel = c1
agent.sinks.k2.hostname = 192.168.2.130
agent.sinks.k2.port = 8889
```

2.创建两个collector，每个collector从日志收集层的**agent接收日志（avro source）**。collector中定义**Replicating Channel Selector**，将相同的数据分别发送到HDFS和Kafka中。

collector: avro source —> replicating channel selector —> channel1,channel2 —> hdfs sink,kafkasink

```
collector1.sources = r1
collector1.channels = c1 c2
collector1.sinks = k1 k2
collector1.sources.r1.type = avro
collector1.sources.r1.bind = 192.168.2.130
collector1.sources.r1.port = 8888
collector1.sources.r1.threads= 3
collector1.sources.r1.interceptors = i1
collector1.sources.r1.interceptors.i1.type = timestamp
collector1.sources.r1.selector.type = replicating

collector1.sources.r1.channels = c1 c2
collector1.channels.c1.type = memory
collector1.channels.c2.type = memory

collector1.sinks.k1.channel = c1
collector1.sinks.k1.type = org.apache.flume.sink.kafka.KafkaSink
collector1.sinks.k1.kafka.topic = ad_log
collector1.sinks.k1.kafka.bootstrap.servers = 192.168.2.130:9092,192.168.2.131:9092
collector1.sinks.k1.kafka.flumeBatchSize = 10
collector1.sinks.k1.kafka.producer.acks = 1

collector1.sinks.k2.channel = c2
collector1.sinks.k2.type = hdfs
collector1.sinks.k2.hdfs.path = /bigdata/flume/ad_log/%Y%m%d/%H%M
collector1.sinks.k2.hdfs.filePrefix = hdfssink-
collector1.sinks.k2.hdfs.fileSuffix = .log
collector1.sinks.k2.hdfs.fileType = DataStream
collector1.sinks.k2.hdfs.writeFormat = Text
```

```
collector1.sinks.k2.hdfs.round = true
collector1.sinks.k2.hdfs.roundValue = 2
collector1.sinks.k2.hdfs.roundUnit = minute
collector1.sinks.k2.hdfs.rollInterval = 30
collector1.sinks.k2.hdfs.rollSize = 0
collector1.sinks.k1.hdfs.rollCount = 0

#collecor2

collector2.sources = r1
collector2.channels = c1 c2
collector2.sinks = k1 k2
collector2.sources.r1.type = avro
collector2.sources.r1.bind = 192.168.2.130
collector2.sources.r1.port = 8889
collector2.sources.r1.threads= 3
collector2.sources.r1.interceptors = i1
collector2.sources.r1.interceptors.i1.type = timestamp
collector2.sources.r1.selector.type = replicating
collector2.sources.r1.channels = c1 c2
collector2.channels.c1.type = memory
collector2.channels.c2.type = memory

collector2.sinks.k1.channel = c1
collector2.sinks.k1.type = org.apache.flume.sink.kafka.KafkaSink
collector2.sinks.k1.kafka.topic = flumeselector
collector2.sinks.k1.kafka.bootstrap.servers = 192.168.2.130:9092,192.168.2.131:9092
collector2.sinks.k1.kafka.flumeBatchSize = 10
collector2.sinks.k1.kafka.producer.acks = 1

collector2.sinks.k2.channel = c2
collector2.sinks.k2.type = hdfs
collector2.sinks.k2.hdfs.path = /data/flume/%Y%m%d/%H%M
collector2.sinks.k2.hdfs.filePrefix = hdfssink-
collector2.sinks.k2.hdfs.fileSuffix = .log
collector2.sinks.k2.hdfs.fileType = DataStream
collector2.sinks.k2.hdfs.writeFormat = Text
collector2.sinks.k2.hdfs.round = true
collector2.sinks.k2.hdfs.roundValue = 2
collector2.sinks.k2.hdfs.roundUnit = minute
collector2.sinks.k2.hdfs.rollInterval = 30
collector2.sinks.k2.hdfs.rollSize = 0
collector2.sinks.k1.hdfs.rollCount = 0
```

```
bin/flume-ng agent --conf conf --conf-file homework/conf/collector1.conf --name
collector1 -Dflume.root.logger=INFO,console
bin/flume-ng agent --conf conf --conf-file homework/conf/collector2.conf --name
collector1 -Dflume.root.logger=INFO,console
bin/flume-ng agent --conf conf --conf-file homework/conf/agent.conf --name agent -
Dflume.root.logger=INFO,console
```