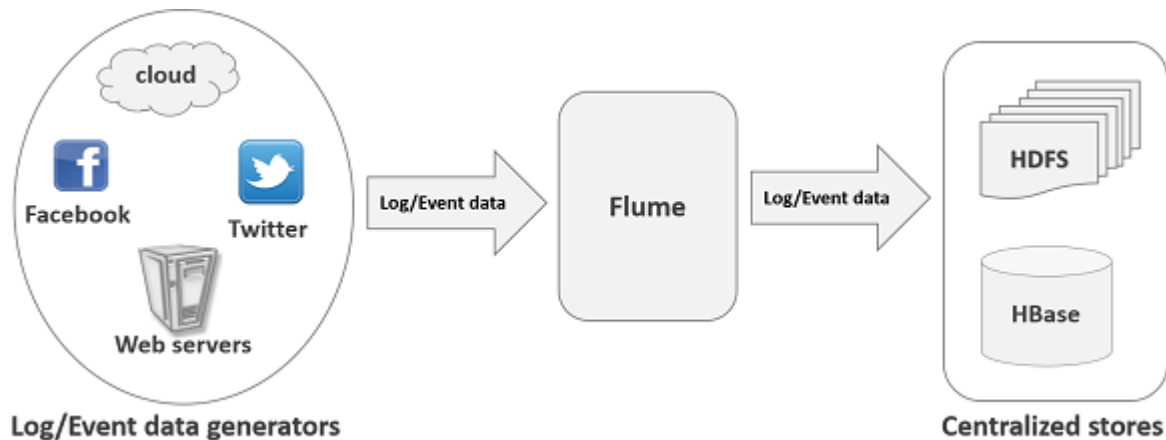


# Flume

# Flume 무엇인가?

- Apache Flume은 다양한 소스의 로그 파일, 이벤트 (etc ...)와 같은 대량의 스트리밍 데이터를 집계하고 수집 및 전송하기 위한 툴 / 서비스 / 데이터 처리 메커니즘임
- Flume은 신뢰성이 높고 분산처리 가능함
- 예를 들면, 웹 서버의 스트리밍 데이터 (로그 데이터)를 HDFS로 저장할 수 있도록 가능하게 하는 경우

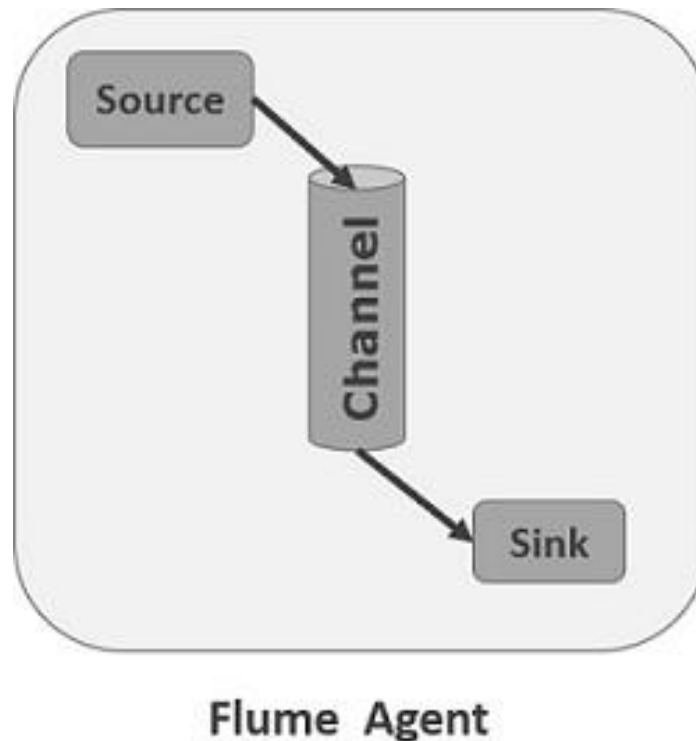


# Flume 응용 분야

- 전자 상거래 웹 애플리케이션에서 특정 지역의 고객 행동을 분석하려고 한다고 가정 했을때, 정의된 로그 데이터를 Hadoop에서 관리하고 분석하려고 하는 경우
- IoT기기나 센서에서 주기적으로 발생하는 로그를 저장하고 분석하기 위한 경우
- 대량의 이벤트성 데이터를 저장해서 처리하는 경우

# Flume 아키텍처

- 에이전트는 Flume의 독립 데몬 프로세스 (JVM)임
- 클라이언트 또는 다른 에이전트로부터 데이터 (이벤트)를 수신하여 다음 대상 (싱크 또는 에이전트)으로 전달함.
- Flume에는 둘 이상의 에이전트가있을 수 있음. 다음 다이어그램은 Flume Agent를 나타냄



# Flume 설정

- 구성 요소의 이름 지정
  - agent\_name.sources = source\_name
  - agent\_name.sinks = sink\_name
  - agent\_name.channels = channel\_name

Sources	Channels	Sinks
<ul style="list-style-type: none"><li>•Avro Source</li><li>•Thrift Source</li><li>•Exec Source</li><li>•JMS Source</li><li>•Spooling Directory Source</li><li>•Twitter 1% firehose Source</li><li>•Kafka Source</li><li>•NetCat Source</li><li>•Sequence Generator Source</li><li>•Syslog Sources</li><li>•Syslog TCP Source</li><li>•Multiport Syslog TCP Source</li><li>•Syslog UDP Source</li><li>•HTTP Source</li><li>•Stress Source</li><li>•Legacy Sources</li><li>•Thrift Legacy Source</li><li>•Custom Source</li><li>•Scribe Source</li></ul>	<ul style="list-style-type: none"><li>•Memory Channel</li><li>•JDBC Channel</li><li>•Kafka Channel</li><li>•File Channel</li><li>•Spillable Memory Channel</li><li>•Pseudo Transaction Channel</li></ul>	<ul style="list-style-type: none"><li>•HDFS Sink</li><li>•Hive Sink</li><li>•Logger Sink</li><li>•Avro Sink</li><li>•Thrift Sink</li><li>•IRC Sink</li><li>•File Roll Sink</li><li>•Null Sink</li><li>•HBaseSink</li><li>•AsynchHBaseSink</li><li>•MorphlineSolrSink</li><li>•ElasticSearchSink</li><li>•Kite Dataset Sink</li><li>•Kafk</li></ul>