

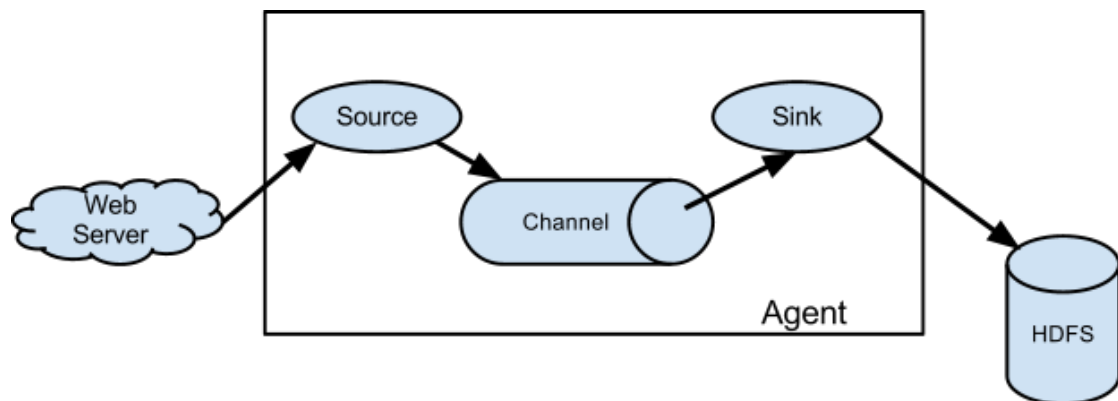
## 第 1 章 Flume 概述

### 1.1 Flume 概念

Flume 是一种分布式，可靠和可用的服务，用于高效收集，聚合和移动大量日志数据。它具有基于流数据流的简单灵活的架构。它具有可靠的可靠性机制和许多故障转移和恢复机制的强大和容错能力。它使用一个简单的可扩展数据模型，允许在线分析应用程序。

Flume 是 Cloudera 公司开发的用于实时收集服务器（apache/nginx 等）日志数据的框架，Flume 很多时候和 storm 以及 spark streaming 等流式处理框架结合使用。

### 1.2 Flume 组成架构



#### Agent

具有字节有效载荷和可选的一组字符串属性的数据流的单元。Flume 代理(Agent)是一个 (JVM) 进程，它承载事件从外部源传递到下一个目标（跳）的组件。是 Flume 数据传输的基本单元，以事件的形式将数据从源头送至目的地。

Agent 就是一个 java 进程，Agent 主要有 3 个部分组成，Source、Channel、Sink。