

**Flume Architecture**

A **Flume agent** is a **JVM** process which has 3 components -**Flume Source, Flume Channel** and **Flume Sink**- through which events propagate after initiated at an external source.

1. In the above diagram, the events generated by external source (WebServer) are consumed by Flume Data Source. The external source sends events to Flume source in a format that is recognized by the target source.
2. Flume Source receives an event and stores it into one or more channels. The channel acts as a store which keeps the event until it is consumed by the flume sink. This channel may use a local file system in order to store these events.
3. Flume sink removes the event from a channel and stores it into an external repository like e.g., HDFS. There could be multiple flume agents, in which case flume sink forwards the event to the flume source of next flume agent in the flow.

Flume has a flexible design based upon streaming data flows. It is fault tolerant and robust with multiple failovers and recovery mechanisms.

* Flume carries data between sources and sinks. This gathering of data can either be scheduled or event-driven. Flume has its own query processing engine which makes it easy to transform each new batch of data before it is moved to the intended sink.
* Possible **Flume sinks** include **HDFS** and **HBase**. Flume can also be used to transport event data including but not limited to network traffic data, data generated by social media websites and email messages.