

# FLUME

11 June 2020 20:20



## To install flume:

```
sudo yum install flume-ng
```

## Features:

To ingest structured data from sources like RDBMS, we use Sqoop. It transfers data from RDBMS to HDFS or HBASE. However, to ingest data from unstructured sources like live source, kafka, log files etc. we use flume.

>>> Flume is used for transfer of **unstructured** or **streaming data** from any place to the HDFS or HBase.

When we run a Sqoop command, it builds a connection with the database, then the data gets transferred from the

>>> Flume is an **always-on non-stop** data ingestion application.

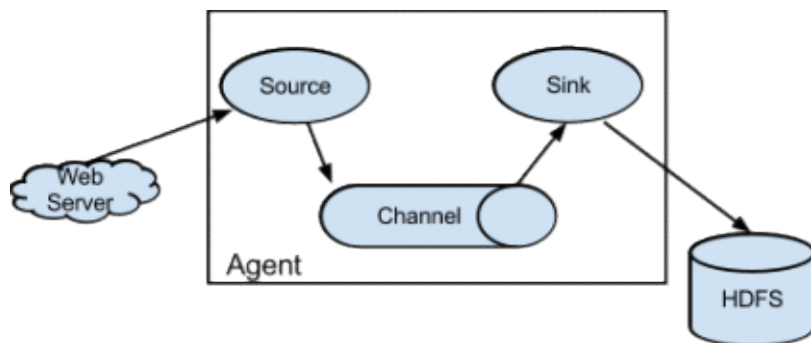
Sqoop is not driven by events

>>> Flume is completely **event driven**

>>> **Whenever we create a flume app, we firstly define a flume agent. Flume agent has 3 main components:**

1. **SOURCE**
2. **SINK**
3. **CHANNEL(memory mainly).**

We require a channel as many a times our reading capacity is more and writing capacity is less, thus we prefer storing the data in a buffer memory so that it can be ingested when it's time comes.



>>> Thus, to use flume, we do a following:

1. Create a flume app
2. Describe configurations of Source
3. Describe configurations of Sink
4. Describe configurations of Channels
5. Binding the source and sink to the channel

>>> For a flume application, we create a **.conf** file.

>>> We can create any number of flume agents at the same time and can fetch data from different sources to our HDFS only