A flume is a channel that directs water from a source to some other location where water is needed. As its clever name implies, Apache® Flume™ was created to allow you to flow data from a source into your [Hadoop®](https://www.ibm.com/analytics/us/en/technology/hadoop/) environment.

In **Flume**, the entities you work with are called sources, decorators, and sinks. A source can be any data source, and Flume has many predefined source adapters. A sink is the target of a specific operation (and in Flume, among other paradigms that use this term, the sink of one operation can be the source for the next downstream operation). A decorator is an operation on the stream that can transform the stream in some manner, which could be to compress or uncompress data, modify data by adding or removing pieces of information, and more.

Flume Example

A number of predefined source adapters are built into Flume. For example, some adapters allow the flow of anything coming off a TCP port to enter the flow, or anything coming to standard input (stdin). A number of text file source adapters give you the granular control to grab a specific file and feed it into a data flow or even take the tail of a file and continuously feed the flow with whatever new data is written to that file. The latter is very useful for feeding diagnostic or web logs into a data flow, since they are constantly being appended to, and the TAIL operator will continuously grab the latest entries from the file and put them into the flow. A number of other predefined source adapters, as well as a command exit, allow you to use any executable command to feed the flow of data.

Types of Sink

Basic:

**This sink can be a text file, the console display, a simple HDFS path, or a null bucket where the data is simply deleted.**

**Agent Tier Event:**

**This is used when you want the sink to be the input source for another operation. When you use these sinks, Flume will also ensure the integrity of the flow by sending back acknowledgments that data has actually arrived at the sink.**

#### Collector Tier Event

#### *This is where you would land a flow (or possibly multiple flows joined together) into an HDFS-formatted file system.*