Monitoring Flume

This page will document the metrics monitoring systems built into Flume. This interface is under development and subject to change but is also open for changes to make it more useful.

Nodes

Currently you can monitor metrics in flume nodes by polling particular URLs on the node that provide data via a json interface. There are links for the physical node, and for each logical node. Please file issues if you would like to have finer granularity.

Physical Nodes

Let's say you have a node called foo. You could view all metrics in tabular form by visiting http://foo:35862. To view the data in json format, you can visit http://foo:35862/node/reports.

You would get something similar to this (note, all json snippets have been formatted for human readability):

```
{
  "jvmInfo": {
    "mem.heap.committed": 62783488,
    "mem.heap.init": 64936576,
    "mem.heap.max": 1005518848,
    "mem.heap.used": 10032176,
    "mem.other.committed": 25624576,
    "mem.other.init": 24313856,
    "mem.other.max": 136314880,
    "mem.other.used": 24227704,
    "name": "pn-grimlock.jvm-Info",
    "rt.vmname": "Java HotSpot(TM) 64-Bit Server VM",
    "rt.vmvendor": "Sun Microsystems Inc.",
    "rt.vmversion": "19.1-b02"
  },
  "logicalnodes": {
    "grimlock": "http://localhost:35862/node/reports/grimlock"
  },
  "sysInfo": {
    "hostname": "grimlock",
    "name": "pn-grimlock.system-info",
    "os.arch": "amd64",
    "os.cpus": 1,
    "os.load": 0.11,
    "os.name": "Linux",
    "os.version": "2.6.35-30-generic"
  }
}
```

This page summarizes the information about that particular physical node, the machine it is on and some process jvm process information, and provides links to metrics for each specific logical node.

Logical Nodes

Each logical node has a URL that can be visited to get information just for that particular node.

The generic pattern is http://<node>:35862/node/reports/<logical node>, but for this specific example the url is http://localhost:35862/node/reports/grimlock.
In this example we can see that the source is null, the sink is null, and that this logical node is currently IDLE. This is generally the state a node will be in if it does not have a configuration on the master.

Let's see the see what this metrics json looks like for an agent node and for a collector node.

Collectors

Here's an example of a node setup to be a collector. This is very verbose and provides all metrics for all of the stages of the flume pipeline.
Each stage of a flume sink pipeline is laid out in "dot" notation.

The most interesting fields to follow are likely:

- "source.CollectorSource.number of bytes" - number of bytes accepted by source
- "source.CollectorSource.number of events" - number of events accepted by source
- "sink.GunzipDecorator.UnbatchingDecorator.AckChecksumChecker.InsistentAppend.appendAttempts" - number of events sink attempted to write to subsink including retries (TODO: check)
- "sink.GunzipDecorator.UnbatchingDecorator.AckChecksumChecker.InsistentAppend.appendGiveups" - number of events that cause an irrecoverable error when attempting to write
- "sink.GunzipDecorator.UnbatchingDecorator.AckChecksumChecker.InsistentAppend.appendRequests" - number of events sink was asked to write (TODO: check)
- "sink.Collector.GunzipDecorator.UnbatchingDecorator.AckChecksumChecker.InsistentAppend.appendRetries" - number of times any event has been retried.

**Agents**

BE mode

DFO mode

E2E mode

Failover Chains

**Masters**