Flume mit PHP

Wir richten in diesem Beispiel einen Thrift Client ein, der seine Events über Flume ins Hadoop Filesystem (HDFS) schreibt.,

Wir gehen an dieser Stelle davon aus, dass Flume, Thrift und die Thrift PHP Extension installiert sind.
Ansonsten sind die die Verweise um das Nachzuholen:

- [wiki]Apache Thrift Installation[/wiki]
- [wiki]Thrift PHP Extension installieren[/wiki]
- [wiki]Cloudera Flume Installation[/wiki]

== Allgemeine Thrift Klassen ==
Wir benötigen die allgemeinen PHP Klassen von Thrift. Dazu kopieren wir die Bibliotheken aus der Thrift Installation in einen Ordner in unserem Webroot.
Den Code der Extension benötigen wir nicht mehr und löschen ihn.

Quellcode

```
1. mkdir -p /var/www/thrift-example
2. cd ~/thrift-0.5.0
3. cp -r lib/php/src/ /var/www/thrift-example/thrift
4. rm -rf /var/www/thrift-example/thrift/ext
```

== Flume Schema ==
Wir benötigen den Aufbau eines Flume Events im Thrift Format. Dazu kopieren wir die Thrift Definition aus der Flume Installation in unser Webroot.

Quellcode

```
1. cd ~/flume
2. cp src/thrift/flume.thrift /var/www/thrift-example/
```

Das Flume Schema sieht wie folgt aus:

Quellcode

```
1. typedef i64 Timestamp
2. enum Priority {
3. FATAL = 0,
4. ERROR = 1,
5. WARN = 2,
6. INFO = 3,
7. DEBUG = 4,
8. TRACE = 5
9. }
10. enum EventStatus {
11. ACK = 0,
12. COMMITED = 1,
13. ERR = 2
14. }
15. struct ThriftFlumeEvent {
16. 1: Timestamp timestamp,
17. 2: Priority priority,
```

```
3: binary body,
4: i64 nanos,
5: string host,
6: map<string,binary> fields
}
service ThriftFlumeEventServer {
oneway void append( 1:ThriftFlumeEvent evt ),
void close(),
}
```

Alles anzeigen

== Flume Thrift Klassen ==
Nun erzeugen wir uns die PHP Klassencode für den Flume Service und kopieren ihn per Konvention in den packages Ordner innerhalb von Thrift.

Quellcode

```
cd /var/www/thrift-example/
thrift --gen php flume.thrift
mv gen-php/ thrift/packages
```

== Flume Konfiguration ==
Über den Flume Master können wir dem Flume Node mitteilen auf welchem Port und mit welchem Protokoll er lauschen soll.
Dazu betreten wir den Flume Master mit der URL: localhost:35871/

Über "config" aktualisieren wir den Node dann wie folgt.
Wir richten unsere Ordnerstruktur nach Datum aus. Für den Dateiprefix verwenden wir außerdem eine dynamische Platzhaltervariable, die in den "fields" des Flume Events übertragen werden kann.

Quellcode

```
Source: rpcSource(1464)
Sink: collectorSink("hdfs://localhost/tmp/flume/%Y-%m-%d/%H00/","%{cat}")
```

In der Flume Master Overview könnt ihr im Anschluss sehen, falls es zu Fehlern gekommen ist.

== Anwendungscode ==
In unserem Beispiel wollen wir die Events als Remore Procedure Call (RPC) an Flume senden. Wir agieren damit als Sender.
Als Anwendungsfall wollen wir die Anzahl an Seitenaufrufen speichern. Unsere Nachricht sieht dabei wie folgt aus, wir trennen die website und die Session ID mit einem Semikolon.

Quellcode

```
website;sessionid
```

Als Protokoll nehmen wir "TBinaryProtocolAccelerated" weil es die Thrift PHP Extension nutzt und performanter ist. Alternativ lässt sich auch das "TBinaryProtocol" nutzen.

Quellcode

```
<?php
$GLOBALS['THRIFT_ROOT'] = './thrift';
```

```
include_once $GLOBALS['THRIFT_ROOT'] . '/Thrift.php';
include_once $GLOBALS['THRIFT_ROOT'] . '/transport/TSocket.php';
include_once $GLOBALS['THRIFT_ROOT'] . '/protocol/TBinaryProtocol.php';
include_once $GLOBALS['THRIFT_ROOT'] . '/packages/flume/ThriftFlumeEventServer.php';
$transport = new TSocket($host = 'localhost', $port = '1464');
$protocol = new TBinaryProtocolAccelerated($transport);
$client = new ThriftFlumeEventServerClient($protocol, $protocol);
$transport->open();
$timestamp = intval(microtime(true) * 1000);
// log message
$client->append(new ThriftFlumeEvent(array(
'priority' => Priority::INFO,
'timestamp' => $timestamp,
'host' => 'server1',
'body' => 'http://www.easy-coding.de/;100',
'fields' => array(
'cat' => 'torben'
)
)));
$transport->close();
```

Alles anzeigen

Ihr könnt euch den gesamten Quelltext hier herunterladen: easy-coding.de/Attachment/1150/

Hier ein Screenshot aus dem Hue Webinterface:
easy-coding.de/Attachment/1122/