|  |
| --- |
|  |
| Flume – Network Packet Capturing |
| Implementation and Workflow document |
|  |
| **Abdul Raseeth** |
| **3/21/2013** |

|  |
| --- |
| Detailed implementation document for Apache Flume based Network packet capturing tool |

Contents

[Introduction 2](#_Toc351717702)

[Why Flume? 2](#_Toc351717703)

[Flume Architecture 2](#_Toc351717704)

[Source 2](#_Toc351717705)

[Channel 2](#_Toc351717706)

[Sink 2](#_Toc351717707)

[Source Code 3](#_Toc351717708)

[Building from source 4](#_Toc351717709)

[Installing Flume 4](#_Toc351717710)

[Changing Java bin permission for socket operation 4](#_Toc351717711)

[Running Agent in Flume 4](#_Toc351717712)

[Sample HDFS sink properties file 4](#_Toc351717713)

# Introduction

The objective is to analyze the packets across the network. Custom Flume agents are implemented by extending the Flume abstract source. Jpcap library is used to capture the packets which are handled by Flume sink as events and been stored in HDFS.

# Why Flume?

Apache Flume is a distributed, reliable, and available system for efficiently collecting, aggregating and moving large amounts of log data from many different sources to a centralized data store.

# Flume Architecture

Flume has three major components

* Source
* Channel
* Sink

## Source

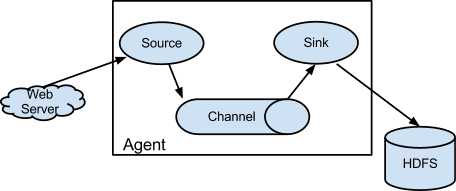
Source can be an external entity like a web server from which an Event raises, An event is a unit of data flows through a flume agent.

## Channel

Channel is a temporary storage place where the events from the sources are stored.

## Sink

Sink is a process which has the responsibility to deliver events from the channel to the terminal repository like HDFS or next agent.



# Source Code

1. Need to import net.sourceforge.jpcap library from <http://jpcap.sourceforge.net/>

import net.sourceforge.jpcap.capture.\*;

import net.sourceforge.jpcap.net.\*;

1. Implementing source by extending AbstractSource class

public class PacCap extends AbstractSource implements Configurable, EventDrivenSource

1. Override Start and Stop of the super class for initializing PacketCapture

@Override

public void start() {

PacketCapture pcap = new PacketCapture();

// Begin capture

pcap.open("eth0", true);

}

@Override

public void stop() {

super.stop();

}

1. Network packets can be listened through PacketListener

public void packetArrived(Packet packet) {

TCPPacket tcppacket = ((TCPPacket) packet);

String captureStr = new String(tcppacket.getTCPData());

getChannelProcessor().processEvent(captureStr);

}

Here the packets are captured as TCP packets and pushed into the channel which maybe a HDFS filesystem. The destination can be configured in the setting file.

agent1.sinks.sink1.type = hdfs

agent1.sinks.sink1.hdfs.path = hdfs://server:54310/user/

# Building from source

1)      You can download the source from [https://github.com/smartek/Network\_Packet\_Reader/tree/master/trunk/src/PacCap](https://by2prd0710.outlook.com/owa/redir.aspx?C=OdG-pPAB_kWTGzYYZWrEkWDP26xe-s8Icmd1B6FGVk3T4Q4dNGOq5K72WEGLbSIu_C2zgvClitc.&URL=https%3a%2f%2fgithub.com%2fsmartek%2fNetwork_Packet_Reader%2ftree%2fmaster%2ftrunk%2fsrc%2fPacCap)

2)      Load the source into any IDE Netbeans / Eclipse, or you can simple install mvn in your linux terminal and build it by ***mvn install***

3)      You can find the binary of the agents as jar at [**https://github.com/smartek/Network\_Packet\_Reader/tree/master/trunk/src/PacCap/target/PacCap-1.0-SNAPSHOT-jar-with-dependencies.jar**](https://by2prd0710.outlook.com/owa/redir.aspx?C=OdG-pPAB_kWTGzYYZWrEkWDP26xe-s8Icmd1B6FGVk3T4Q4dNGOq5K72WEGLbSIu_C2zgvClitc.&URL=https%3a%2f%2fgithub.com%2fsmartek%2fNetwork_Packet_Reader%2ftree%2fmaster%2ftrunk%2fsrc%2fPacCap%2ftarget%2fPacCap-1.0-SNAPSHOT-jar-with-dependencies.jar)

# Installing Flume

1)      Download flume source from [http://www.apache.org/dyn/closer.cgi/flume/1.3.1/apache-flume-1.3.1-src.tar.gz](https://by2prd0710.outlook.com/owa/redir.aspx?C=OdG-pPAB_kWTGzYYZWrEkWDP26xe-s8Icmd1B6FGVk3T4Q4dNGOq5K72WEGLbSIu_C2zgvClitc.&URL=http%3a%2f%2fwww.apache.org%2fdyn%2fcloser.cgi%2fflume%2f1.3.1%2fapache-flume-1.3.1-src.tar.gz)

2)      Installation is not required, We can directly invoke **bin/flume-ng**

Changing Java bin permission for socket operation  
  
        We need to setuid permission for the java bin to read socket as root user  
  
        1) run sudo chmod -R **4755** /usr/lib/jvm/java-7-oracle/bin/  
  
        you can also roll back above command (If needed) using sudo chmod -R **755** /usr/lib/jvm/java-7-oracle/bin/

# Running Agent in Flume

1) Copy the created jar to flume/jar folder

2) Copy the attached config file (hdfs-sink.properties) to flume/Conf folder

4) Change to hduser using su hduser

3) From flume directory run bin/flume-ng agent -C jar/PacCap-1.0-SNAPSHOT-jar-with-dependencies.jar -n agent1 -c conf -f conf/hdfs-sink.properties

# Sample HDFS sink properties file

################################################

# Name the components on this agent

################################################

agent1.sources = source1

agent1.sinks = sink1

agent1.channels = channel1

################################################

# Describe Source

################################################

# Source Twitter

agent1.sources.source1.type = PacCap

agent1.sources.source1.track = iphone

agent1.sources.source1.consumerKey = wqbWQTZwgBfVRXHJ6jImgw

agent1.sources.source1.consumerSecret = q942QsSN2Blp9kqn1suZumkpv6OzKtO9qlJNUSCJw

agent1.sources.source1.accessToken = 627133469-d1SRK1BtKOseh5mOdwlKDizvacRaEieJTKQiNZgE

agent1.sources.source1.accessTokenSecret = eatfrkdNmaJ7PHtrpCxIk3fKlsGhuMCvPM1QSPa8c

################################################

# Describe Sink

################################################

# Sink HDFS

agent1.sinks.sink1.type = hdfs

agent1.sinks.sink1.hdfs.path = hdfs://192.168.10.25:54310/user/paccap

################################################

# Describe Channel

################################################

# Channel Memory

agent1.channels.channel1.type = memory

agent1.channels.channel1.capacity = 1000

agent1.channels.channel1.transactionCapactiy = 100

################################################

# Bind the source and sink to the channel

################################################

agent1.sources.source1.channels = channel1

agent1.sinks.sink1.channel = channel1