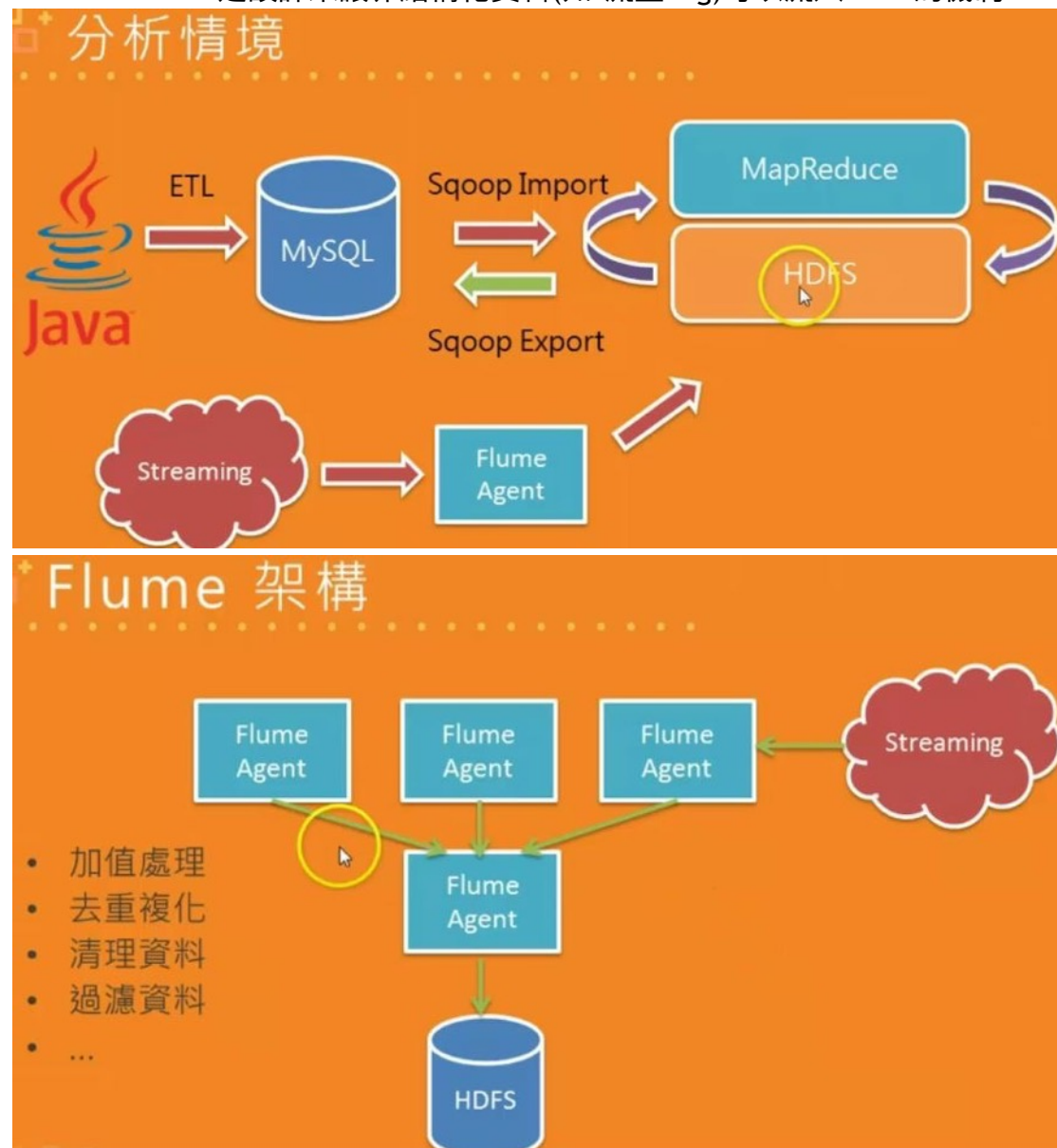


Flume Setup

1. 簡介：

- Flume是設計來讓非結構化資料(如. 流量Log)可以流入HDFS的機制

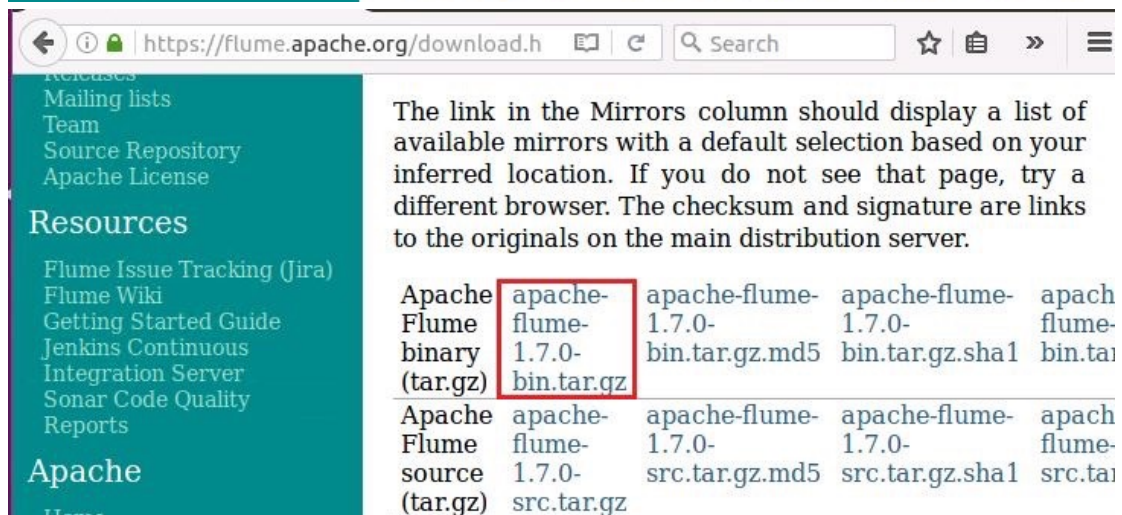


2. install Flume

- 訪問Flume網站([網址](#))



The screenshot shows the Apache Flume homepage. The header features the Apache Flume logo on the left and the text "Apache Flume" in large green letters on the right. Below the logo is a navigation menu with links: "How to Get Involved", "Download" (highlighted with a red box), "Documentation", "Releases", "Mailing lists", "Team", "Source Repository", and "Apache License". The main content area has a heading "Welcome to Apache Flume" and a paragraph describing Flume as a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data.



The screenshot shows the Apache Flume download page. The left sidebar contains a navigation menu with links: "Releases", "Mailing lists", "Team", "Source Repository", "Apache License", "Resources", "Flume Issue Tracking (Jira)", "Flume Wiki", "Getting Started Guide", "Jenkins Continuous Integration Server", "Sonar Code Quality Reports", and "Apache". The main content area has a heading "The link in the Mirrors column should display a list of available mirrors with a default selection based on your inferred location. If you do not see that page, try a different browser. The checksum and signature are links to the originals on the main distribution server." Below this is a table of mirrors.

Apache Flume binary (tar.gz)	apache-flume-1.7.0-bin.tar.gz	apache-flume-1.7.0-bin.tar.gz.md5	apache-flume-1.7.0-bin.tar.gz.sha1	apach flume- bin.ta
Apache Flume source (tar.gz)	apache-flume-1.7.0-src.tar.gz	apache-flume-1.7.0-src.tar.gz.md5	apache-flume-1.7.0-src.tar.gz.sha1	apach flume- src.ta

HTTP

<http://apache.stu.edu.tw/flume/1.7.0/apache-flume-1.7.0-bin.tar.gz>

<http://ftp.mirror.tw/pub/apache/flume/1.7.0/apache-flume-1.7.0-bin.tar.gz>

<http://ftp.tc.edu.tw/pub/Apache/flume/1.7.0/apache-flume-1.7.0-bin.tar.gz>

<http://ftp.twaren.net/Unix/Web/apache/flume/1.7.0/apache-flume-1.7.0-bin.tar.gz>

- 請開啟HadoopMaster主機的終端
- 下載Flume，輸入 `wget http://apache.stu.edu.tw/flume/1.7.0/apache-flume-1.7.0-bin.tar.gz`

```

hduser@hadoopmaster: ~
hduser@hadoopmaster:~$ wget http://apache.stu.edu.tw/flume/1.7.0/apache-flume-1.7.0-bin.tar.gz
--2016-12-26 10:33:49-- http://apache.stu.edu.tw/flume/1.7.0/apache-flume-1.7.0-bin.tar.gz
Resolving apache.stu.edu.tw (apache.stu.edu.tw)... 120.119.118.1, 2001:e10:c41:eee::1
Connecting to apache.stu.edu.tw (apache.stu.edu.tw)|120.119.118.1|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 55711670 (53M) [application/x-gzip]
Saving to: 'apache-flume-1.7.0-bin.tar.gz'

ume-1.7.0-bin.tar.g  2%[          ]  1.29M  278KB/s  eta 3m 59s

```

- 執行解壓縮，輸入 `tar -zxvf apache-flume-1.7.0-bin.tar.gz`

```

hduser@hadoopmaster: ~
hduser@hadoopmaster:~$ tar -zxvf apache-flume-1.7.0-bin.tar.gz
apache-flume-1.7.0-bin/lib/flume-ng-configuration-1.7.0.jar
apache-flume-1.7.0-bin/lib/slf4j-api-1.6.1.jar
apache-flume-1.7.0-bin/lib/slf4j-log4j12-1.6.1.jar
apache-flume-1.7.0-bin/lib/log4j-1.2.17.jar
apache-flume-1.7.0-bin/lib/guava-11.0.2.jar
apache-flume-1.7.0-bin/lib/jsr305-1.3.9.jar
apache-flume-1.7.0-bin/lib/flume-ng-sdk-1.7.0.jar

```

- 移動Sqoop資料夾，輸入 `sudo mv apache-flume-1.7.0-bin /usr/local/flume`

```

hduser@hadoopmaster: ~
hduser@hadoopmaster:~$ sudo mv apache-flume-1.7.0-bin /usr/local/flume
[sudo] password for hduser:
hduser@hadoopmaster:~$

```

- 修改`~/.bashrc`，輸入 `sudo gedit ~/.bashrc`
- 輸入下列指令

```

export FLUME_HOME=/usr/local/flume/
export FLUME_CONF_DIR=/usr/local/flume/conf
export PATH=$PATH:$FLUME_HOME/bin

```

```

*.bashrc
export PATH=$PATH:$SPARK_HOME/bin
#SPARK Variables
#SQOOP Variables
export SQOOP_HOME=/usr/local/sqoop
export PATH=$PATH:$SQOOP_HOME/bin
#SQOOP Variables
#FLUME variables
export FLUME_HOME=/usr/local/flume/
export FLUME_CONF_DIR=/usr/local/flume/conf
export PATH=$PATH:$FLUME_HOME/bin
#FLUME variables

```

- 執行更新過的`~/.bashrc`檔，輸入 `source ~/.bashrc`

```

hduser@hadoopmaster:~$ source ~/.bashrc
hduser@hadoopmaster:~$

```

- 此時已可以執行Flume，輸入 `flume-ng`可看到提示


```
hduser@hadoopmaster: ~  
hduser@hadoopmaster:~$ flume-ng  
Error: Unknown or unspecified command ''  
  
Usage: /usr/local/flume/bin/flume-ng <command> [options]...  
  
commands:  
  help          display this help text  
  agent         run a Flume agent  
  avro-client    run an avro Flume client  
  version       show Flume version info  
  
global options:  
  --conf, -c <conf>    use configs in <conf> directory  
  --classpath, -C <cp> append to the classpath  
  --dryrun, -d         do not actually start Flume, just print the command  
  --plugins-path <dirs> colon-separated list of plugins.d directories. See the  
he                    plugins.d section in the user guide for more details
```

3. 資料測試—產生Log檔並匯入HDFS

- 請將附檔放到/usr/local/flume中，含
 - ✓ Log-generator.py→隨機產生Log記錄，寫到/tmp/log-generator.log中
 - ✓ agent1.properties→Flume Agent1，隨時將Log檔的最新資料讀入
 - ✓ agent2.properties→Flume Agent2，由Agent1讀入資料後寫入HDFS中
 - ✓ agent1的角色可以為多個，agent2做為資料的收集者，並寫入

HDFS，所以在此用兩個agent

```
hduser@hadoopmaster:~$ cd /usr/local/flume
hduser@hadoopmaster:/usr/local/flume$ ls
agent1.properties  CHANGELOG  doap_Flume.rdf  LICENSE  README.md
agent2.properties  conf       docs           log-generator.py  RELEASE-NOTES
bin               DEVNOTES  lib           NOTICE      tools
hduser@hadoopmaster:/usr/local/flume$
```

- 將伺服器HadoopMaster+HadoopSlave1+HadoopSlave2開啟
- 執行Hadoop，於HadoopMaster終端輸入start-all.sh

```
hduser@hadoopmaster:~$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [hadoopmaster]
hadoopmaster: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-
namenode-hadoopmaster.out
hadoopslave2: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-
datanode-hadoopslave2.out
hadoopslave1: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-
datanode-hadoopslave1.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hd
user-secondarynamenode-hadoopmaster.out
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resource
manager-hadoopmaster.out
hadoopslave1: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduse
r-nodemanager-hadoopslave1.out
hadoopslave2: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduse
r-nodemanager-hadoopslave2.out
hduser@hadoopmaster:~$ jps
2929 SecondaryNameNode
2726 NameNode
3127 ResourceManager
3416 Jps
hduser@hadoopmaster:~$
```

- 切換到flume資料夾，輸入cd /usr/local/flume
- 啟動flume agent2，等待flume agent1餵資料，輸入 flume-ng agent --conf-file agent2.properties --name agent2

```
hduser@hadoopmaster:/usr/local/flume$ flume-ng agent --conf-file agent2.properti
es --name agent2
Warning: No configuration directory set! Use --conf <dir> to override.
Info: Including Hadoop libraries found via (/usr/local/hadoop/bin/hadoop) for HD
FS access
Info: Including Hive libraries found via () for Hive access
+ exec /usr/lib/jvm/java-8-openjdk-amd64/bin/java -Xmx20m -cp '/usr/local/flume/
/lib/*:/usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/hadoop/common/lib/*:
/usr/local/hadoop/share/hadoop/common/*:/usr/local/hadoop/share/hadoop/hdfs:/usr
/local/hadoop/share/hadoop/hdfs/lib/*:/usr/local/hadoop/share/hadoop/hdfs/*:/usr
/local/hadoop/share/hadoop/yarn/lib/*:/usr/local/hadoop/share/hadoop/yarn/*:/usr
/local/hadoop/share/hadoop/mapreduce/lib/*:/usr/local/hadoop/share/hadoop/mapred
uce/*:/usr/local/hadoop/contrib/capacity-scheduler/*.jar:/lib/*' -Djava.library.
path=/usr/local/hadoop/lib/native: org.apache.flume.node.Application --conf-fil
e agent2.properties --name agent2
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/flume/lib/slf4j-log4j12-1.6.1.jar!/
org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4
j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
16/12/26 15:03:20 INFO node.PollingPropertiesFileConfigurationProvider: Configur
ation provider starting
16/12/26 15:03:20 INFO node.PollingPropertiesFileConfigurationProvider: Reloadin
```



```

16/12/26 15:03:22 INFO instrumentation.MonitoredCounterGroup: Monitored counter
group for type: SOURCE, name: agent2-src: Successfully registered new MBean.
16/12/26 15:03:22 INFO instrumentation.MonitoredCounterGroup: Component type: SO
URCE, name: agent2-src started
16/12/26 15:03:22 INFO source.AvroSource: Avro source agent2-src started.

```

- 開啟另一終端，一樣切換到flume資料夾，啟動log-generator.py來產生log，輸入python /usr/local/flume/log-generator.py

```

hduser@hadoopmaster: /usr/local/flume$ python log-generator.py
HTTP 204 www.largitdata.com /contact.html 62.211.194.164
HTTP 102 www.ieee.org /about.html 132.180.147.120
HTTP 404 www.largitdata.com /about.html 194.55.102.82
HTTP 200 www.largitdata.com /contact.html 236.189.254.179
HTTP 101 www.ieee.org /service.html 63.151.103.175
HTTP 100 www.largitdata.com /service.html 136.163.21.149
HTTP 304 www.yahoo.com /index.html 39.31.132.57
HTTP 404 www.ieee.org /about.html 204.238.72.130
HTTP 400 www.ieee.org /index.html 80.41.223.182
HTTP 402 www.ieee.org /service.html 17.186.253.95
HTTP 201 www.ieee.org /index.html 34.191.248.54
HTTP 403 www.largitdata.com /about.html 52.205.106.45
HTTP 304 www.yahoo.com /index.html 212.178.113.187
HTTP 304 www.largitdata.com /contact.html 187.125.61.74
HTTP 201 www.ieee.org /about.html 179.4.203.63
HTTP 102 www.yahoo.com /contact.html 23.173.195.101
HTTP 200 www.yahoo.com /contact.html 28.64.72.155
HTTP 304 www.largitdata.com /contact.html 25.115.131.124

```

- 再開啟另一終端，一樣切換到flume資料夾，啟動flume agent1，由log檔接收資料，輸入flume-ng agent --conf-file agent1.properties --name agent1

```

hduser@hadoopmaster: /usr/local/flume
hduser@hadoopmaster:~$ cd /usr/local/flume
hduser@hadoopmaster: /usr/local/flume$ flume-ng agent --conf-file agent1.properties --name agent1
Warning: No configuration directory set! Use --conf <dir> to override.
Info: Including Hadoop libraries found via (/usr/local/hadoop/bin/hadoop) for HD
FS access
... 11 more
16/12/26 15:05:54 INFO source.ExecSource: Command [tail -F /tmp/log-generator.log] exited with 143
16/12/26 15:05:57 INFO sink.AbstractRpcSink: Rpc sink agent1-sink: Building RpcClient with hostname: localhost, port: 5566
16/12/26 15:05:57 INFO sink.AvroSink: Attempting to create Avro Rpc client.
16/12/26 15:05:57 WARN api.NettyAvroRpcClient: Using default maxIOWorkers

```

- 都正確執行後，可以開啟另一終端，到hdfs中查看資料接收，輸入hadoop fs -ls /flume

```

hduser@hadoopmaster: ~
hduser@hadoopmaster:~$ hadoop fs -ls /flume
Found 200 items
-rw-r--r--  2 hduser supergroup      819 2016-12-26 15:05 /flume/flume.log.14
82735945708
-rw-r--r--  2 hduser supergroup      815 2016-12-26 15:05 /flume/flume.log.14
82735945709
-rw-r--r--  2 hduser supergroup      816 2016-12-26 15:05 /flume/flume.log.14
82735945710
-rw-r--r--  2 hduser supergroup      831 2016-12-26 15:05 /flume/flume.log.14
82735945711
-rw-r--r--  2 hduser supergroup      825 2016-12-26 15:05 /flume/flume.log.14
82735945712
-rw-r--r--  2 hduser supergroup      820 2016-12-26 15:05 /flume/flume.log.14
82735945713
-rw-r--r--  2 hduser supergroup      820 2016-12-26 15:05 /flume/flume.log.14
82735945714
-rw-r--r--  2 hduser supergroup      828 2016-12-26 15:05 /flume/flume.log.14
82735945715

```

- head 看前面十行的內容

```
hduser@hadoopmaster:~$ hadoop fs -cat /flume/flume.log.1482735945907 | head
SEQ[REDACTED]org.apache.hadoop.io.LongWritable"org.apache.hadoop.io.BytesWritable[REDACTED]9zQ[REDACTED]
b 5, [REDACTED]9[REDACTED]5HTTP 402 www.largitdata.com /index.html 89.77.250.158[REDACTED]9[REDACTED]2HTTP
203 www.yahoo.com /contact.html 4.106.146.199[REDACTED]9[REDACTED]3HTTP 400 www.yahoo.com /ser
vice.html 112.254.217.14[REDACTED]9[REDACTED]1HTTP 100 www.ieee.org /contact.html 60.55.234.182
[REDACTED]9[REDACTED]0HTTP 304 www.ieee.org /index.html 14.175.231.243[REDACTED]9[REDACTED]2HTTP 305 www.yahoo
.com /contact.html 174.33.14.188[REDACTED]9[REDACTED]3HTTP 305 www.yahoo.com /contact.html 125.
228.112.86[REDACTED]9[REDACTED]4HTTP 400 www.yahoo.com /index.html 24.95.75.21[REDACTED]9[REDACTED]1HTTP 401 w
ww.yahoo.com /index.html 240.23.105.156[REDACTED]9[REDACTED]8HTTP 402 www.largitdata.com /servi
ce.html 227.142.150.83[REDACTED]9[REDACTED]9zQ[REDACTED]b 5, [REDACTED]3[REDACTED]hduser@hadoopmaster:~$
```