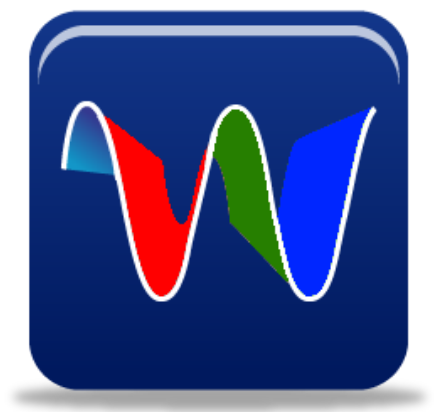
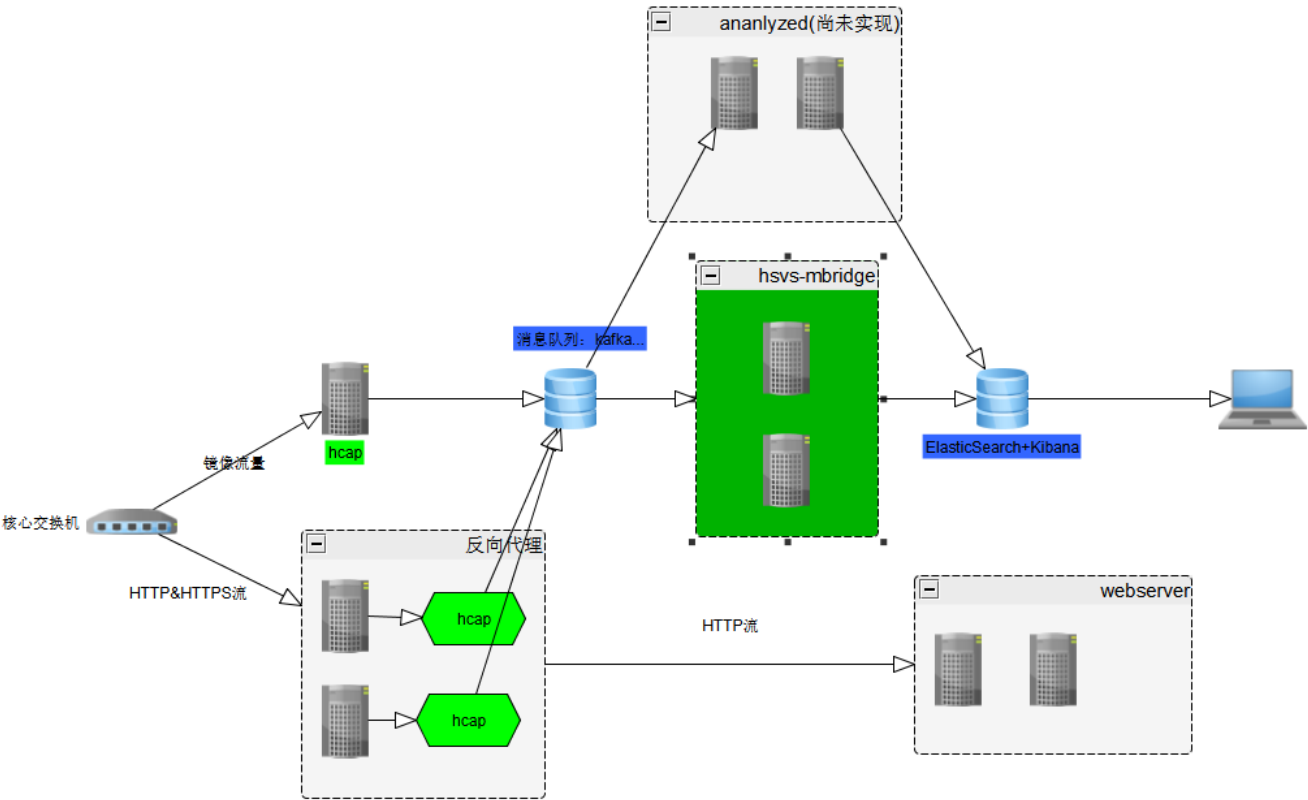


# HSVS-kafka方案



本文档适用于hcap:1.0,hsvs-mbridge:1.1.2 版本，此版本为当前latest分支

## 拓扑图



绿色部分为HSVS提供

蓝色部分需要用户自己安装或者使用企业已有资源

## 安装步骤

### 1. 安装部署ElasticSearch 和 Kibana

版本要求ElasticSearch(version >= 5.0.0)

建议优先使用企业已有的ElasticSearch资源，这里给出一个elasticsearch集群搭建博客教程

hsvs-mbridge 启动后会自动创建模板，无需手动创建索引和模板

## 2. 安装部署kafka

版本要求**kafka(version >= 0.8)**

建议优先使用企业已有的**kafka**资源，如果您需要自己搭建，请参考<http://kafka.apache.org/quickstart>

创建**kafka**的topic `http_session`

```
bin/kafka-topics.sh --create --topic http_session --zookeeper <YOUR_ZK_HOST>:2181 --replication-factor 4 --partitions 4
```

--replication-factor 是设定副本数，这个决定故障时容错的能力和读取数据时的横向扩容速度；  
--partitions 设定分区数，决定写入时的最大能力

查看已经存创建的topic `bin/kafka-topics.sh --list --zookeeper <YOUR_ZK_HOST>:2181`

## 3. 安装部署hsvs-mbridge

本程序是数据预处理模块，会将处理后的数据写入ElasticSearch。本程序部署数量跟第2部中设定的**kafka**数量有关。需被整除，即**hsvs-mbridge**的部署数量乘以N(N为正整数)等于**kafka**创建的分区数。

- 拉取镜像, 如果您内网的机器无法联网，可以使用docker save/load的方式导入到您的机器上

```
docker pull hsvs-mbridge
```

- 下载配置文件模板，修改成您实际环境中的配置

```
wget labs.shellpub.com/hsvs/mbridge.yml
```

- 配置文件说明
    - `kafka_config` 字段配置与**kafka**相关属性
- ```
kafka_zookeeper: "填写您所用kafka集群的zookeeper地址"
```
- `es_config` 字段配置与**elasticsearch**有关属性
- ```
es_host: "填写ElasticMaster Node的IP地址"
```
- `source`: 字段设定数据源，使用**kafka**时，此处应该填写"**kafka**"

我们假定您的配置文件存储路径位于/opt/mbridge.yml

- 启动mbridge

```
docker run --net=host --rm -e ELASTICSEARCH_URL=http://<YOUR_ES_MASTER_HOST>:9200/ -v /opt/mbridge.yml:/hsvs/mt
```

此处必须指定ELASTICSEARCH\_URL传递ElasticSearch地址给hsvs-mbridge，来创建template模板

## 4. 安装部署hcap

本程序是http流还原模块，应当不属于镜像流量或者http反向代理机上

拉取镜像

```
docker pull sort/hcap
```

下载配置文件模板, 修改成您实际环境中的配置

```
wget labs.shellpub.com/hsvs/hcap.ini
```

## 配置文件关键配置说明

```
[MAIN]
NIC_name = eth0 # 网卡名称
promisc = 0 # 是否开启混杂模式，镜像流量模式必须，反向代理模式下无须开启
output_target = kafka # 写入目标，使用kafaka时此处必须配置为kafka

[FILTER]
include_domains = * # 过滤域名，支持通配语法
#exclude_domains = # 排除域名

[OUTPUT_KAFKA]
brokers= #必须参数，此处填写您kafka的broker地址
version=0.11.0.0 #必须参数，要兼容的kafka通信协议版本
topic=http_session #kafka topic
```

运行hcap，我们假定您的配置文件路径位于/opt/hcap.ini

```
docker run --net=host --privileged -d -v /opt/hcap.ini:/hsvs/hcaplite/hcap.ini:ro --rm sort/hcap -Bb -i eth0
```

参数说明：

- Bb 获取请求题和响应体
- i 非必须，指定您所需抓包网卡名称，本参数会覆盖配置文件中的Nic\_Name字段

## 5. 打开kibana创建kibana 索引

启动时间大约为1分钟后，系统中抓取到记录后打开kibana  
<http://<你的KIBANA主机IP>:5601>  
 点击左侧导航栏->Management->Index Patterns->Create Index Pattern  
 在Index name or pattern中输入  
 logstash-http-session-\*  
 在下面`Time Filter field name`中选择@timestamp  
 点击Create完成创建，下面两个[DEPRECATED]选项不要勾选！  
 点击左侧导航栏Discover即可查看抓取到的流量

### Configure an index pattern

In order to use Kibana you must configure at least one index pattern. Index patterns are used to filter data against. They are also used to configure fields.

#### Index name or pattern

Patterns allow you to define dynamic index names using \* as a wildcard. Example: logstash-http-session-\*

#### Time Filter field name ⓘ refresh fields

☐ Expand index pattern when searching [DEPRECATED]

With this option selected, searches against any time-based index pattern that contains a currently selected time range.

Searching against the index pattern `logstash-*` will actually query Elasticsearch for the specified time range. With recent changes to Elasticsearch, this option should no longer be necessary and will be deprecated.

☐ Use event times to create index names [DEPRECATED]