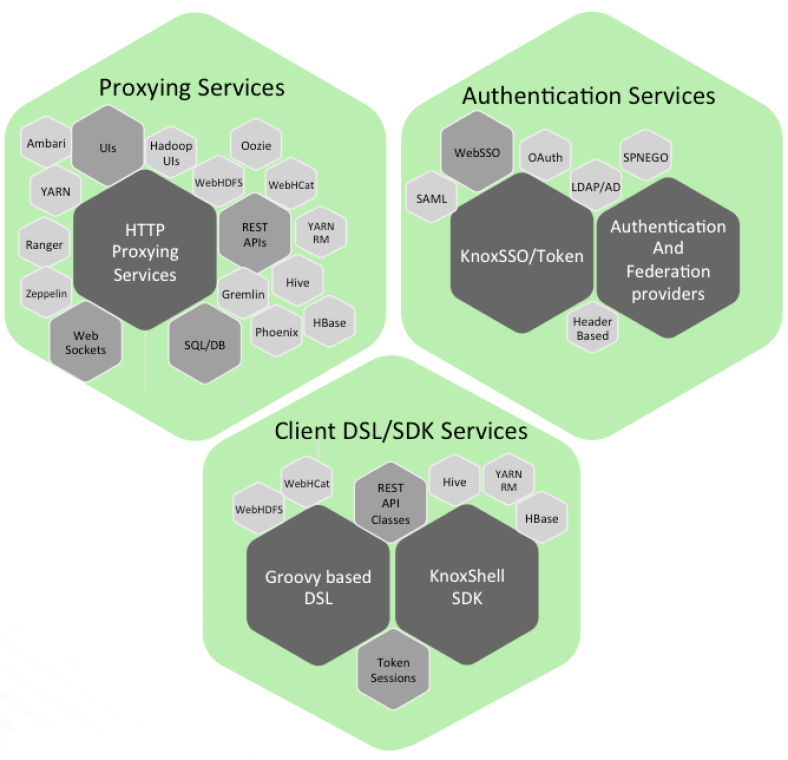
Apache Knox

Apache Knox是一个用于hadoop安全的REST Full API Gateway，为Apache Hadoop集群提供唯一的REST 入口（不只服务与Hadoop，对于Web化的应用，基本上都可以使用Knox做保护）。Knox以类似防火墙的形式挡在Hadoop集群之前，接管所有用于请求，例如Web UI访问，HDFS内容查看、Hive及HBase的数据操作等，相对Kerbeors来说更清爽，但是对内部集群的保护要求高。主要提供的功能如下：



1. Proxing Service

Apache Knox项目的主要目标是提供Hadoop集群的HTTP的代理功能。

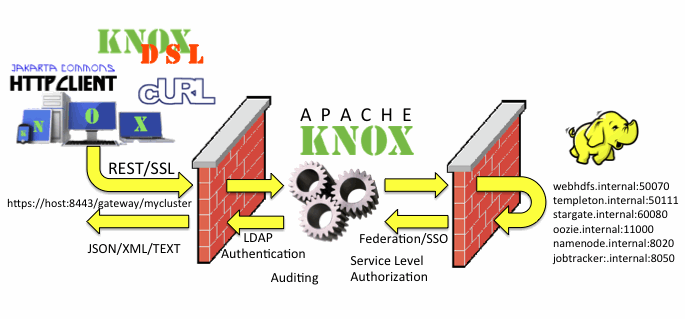
1. Authentication Service

不仅提供WebSSO流程，同时提供REST API的访问认证。可选择LDAP/AD, Header based PreAuth,Kerberors,SAML及Oauth等均是可选项。

1. Client Service

客户端开发可以通过DSL的脚本或者直接使用作为SDK的Knox Shell

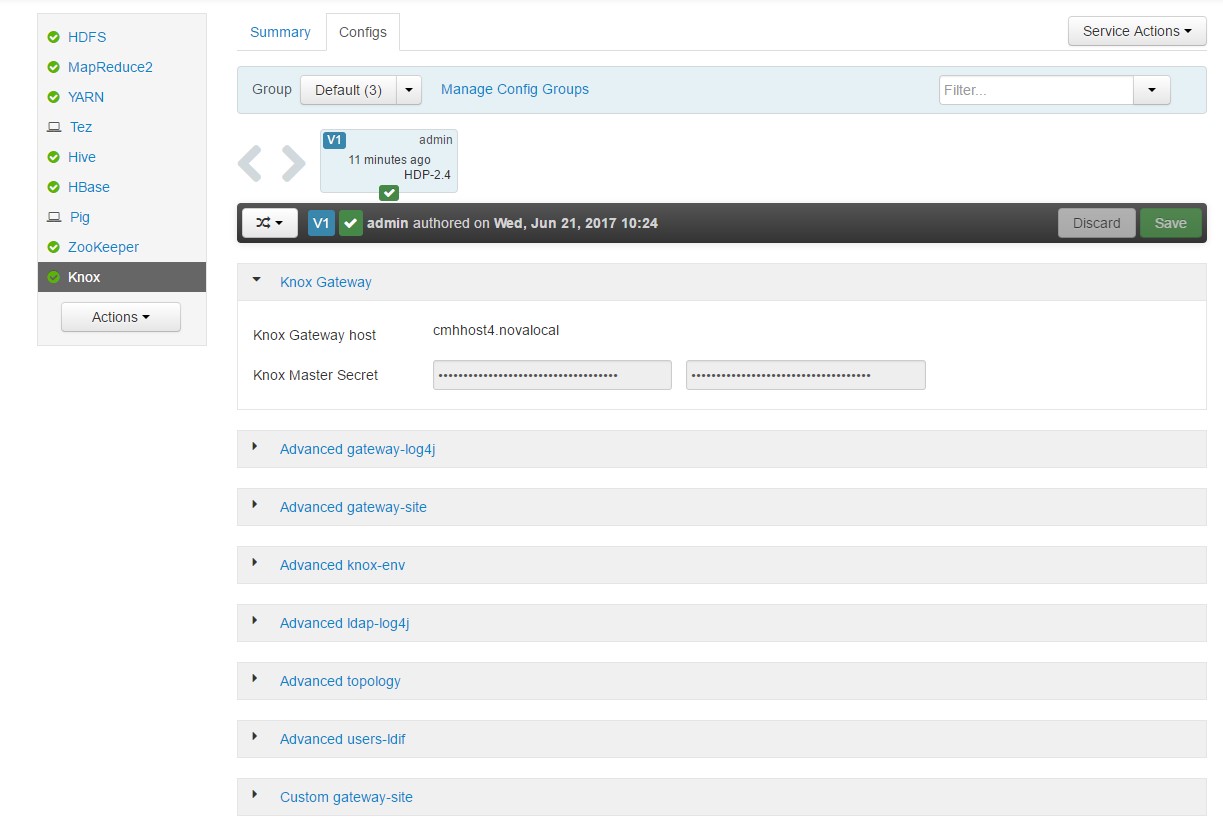
下图是Knox的使用流程图：



Knox的认证使用是通过LDAP来实现的。

# 1.安装和使用

通过Ambari 安装，设置的Kmox Master Secret为：qwe123，安装成功后如下图所示：



查看topology目录：

*cd knox-server/conf/topologies*

可以在目录中查看topologies，如default.topologies

*<service>*

*<role>WEBHDFS</role>*

*<url>http://cmhhost4.novalocal:50070/webhdfs</url>*

*</service>*

*<service>*

*<role>WEBHBASE</role>*

*<url>http://cmhhost6.novalocal:8080</url>*

*</service>*

*<service>*

*<role>HIVE</role>*

*<url>http://cmhhost5.novalocal:10001/cliservice</url>*

*</service>*

*<service>*

*<role>RESOURCEMANAGER</role>*

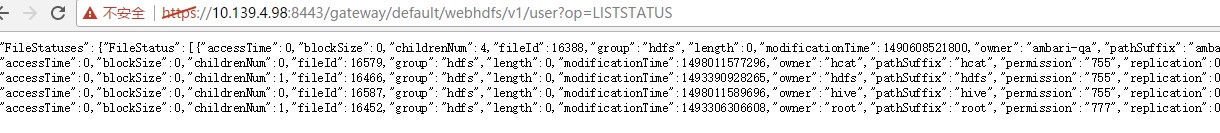
*<url>http://cmhhost5.novalocal:8088/ws</url>*

*</service>*

1）访问webhdfs，链接如下所示：

*https://{gateway-host}:{gateway-port}/{gateway-path}/{cluster-name}/webhdfs*

查看到的结果如下所示：



在输入链接时需要输入用户名和密码：guest : guest-passwd

2） 访问hbase，链接如下<http://10.139.4.106:8080/status/cluster>，通过knox访问如下：

[*https://10.139.4.98:8443/gateway/default/hbase/status/cluster*](https://10.139.4.98:8443/gateway/default/hbase/status/cluster)



1. Admin URL

Knox Admin API的URL如下所示：

|  |  |
| --- | --- |
| *Gateway* | *https://{gateway-host}:{gateway-port}/{gateway-path}/admin/api/v1* |

执行如下：

*curl -u admin:admin-password -i -k https://10.139.4.98:8443/gateway/admin/api/v1/version*

*<?xml version="1.0" encoding="UTF-8"?>*

*<ServerVersion>*

*<version>0.6.0.2.4.3.0-227</version>*

*<hash>fce038086ebdb2f76ed3149de8822b4c5bab6701</hash>*

*</ServerVersion>*

获取所有的Topology如下所示：

*[root@cmhhost6 ~]# curl -u admin:admin-password -i -k https://10.139.4.98:8443/gateway/admin/api/v1/topologies*

*<?xml version="1.0" encoding="UTF-8"?>*

*<topologies>*

*<topology>*

*<name>admin</name>*

*<timestamp>1471269880000</timestamp>*

*<uri>https://10.139.4.98:8443/gateway/admin</uri>*

*<href>https://10.139.4.98:8443/gateway/admin/api/v1/topologies/admin</href>*

*</topology>*

*<topology>*

*<name>default</name>*

*<timestamp>1498012390000</timestamp>*

*<uri>https://10.139.4.98:8443/gateway/default</uri>*

*<href>https://10.139.4.98:8443/gateway/admin/api/v1/topologies/default</href>*

*</topology>*

*</topologies>*

获取具体的toplogies信息如下：

*Example cURL Request [http://knox.apache.org/books/knox-0-12-0/markbook-section-link.png](http://knox.apache.org/books/knox-0-12-0/user-guide.html#Example+cURL+Request)*

*curl -u admin:admin-password -i -k -H Accept:application/json https://{gateway-host}:{gateway-port}/{gateway-path}/admin/api/v1/topologies/{topology-name}*

# 2.Apache Knox 概括

部署Knox的Hadoop集群至少部署WebHDFS,WebHCat(i.e. Templeton)及Oozier，通过配置也可以通过Knox Gateway来访问HBase/Stargate或者Hive。

1. Gateway不依赖Hadoop Cluster，可以作为独立的服务
2. 运行在集群中的服务都必须通过gateway来访问，该服务的hostname或者ip仅对gateway来说是可访问的，对用户来说是透明的

## 2.1 Apache Knox目录结构

可以通过解压zip/archive 文件来安装Knox，下表是各目录：

|  |  |
| --- | --- |
| 目录 | 描述 |
| conf/ | 存放gateway全局的配置文件，非cluster的配置 |
| data/ | 包含安全和topology specific的artifacts在运行时读写的数据 |
| conf/topologies | Hadoop Clustes描述的topology 文件，用于部署集群代理 |
| data/security | 保存master的秘钥和keystore目录 |
| data/security/keystores | 保存Gateway标识keystore和credential stores，用于部署cluster topology |
| data/services | 目前支持的服务behavior定义 |
| bin/ | 包括执行脚本，客户端和Servers的batch及Jars文件 |
| data/deployments | 对定义的Hadoop Clusters进行访问保护的集群topologies |
| lib | Gateway的所有组件的JARS |
| dep | Gateway依赖的所有组件JARS |
| ext | 用户提供的扩展gateway功能的JARS |
| pids | 运行的ldap和gateway servers的进程id |
| samples | 用于学习gateway功能的样例 |
| templates | 用于复制及定制的默认的配置文件 |
| README | Apache Knox Gateway的基本信息 |
| ISSUES | 已知问题的基本信息 |
| CHANGES | 版本之间的变化信息 |
| LICENSE | 版权声明 |
| NOTICE | Notice文件 |

## 2.2 支持的服务

下表是所支持的服务：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Service | Version | Non-Secure | Secure | HA |
| WebHDFS | 2.4.0 | y | y | y |
| WebHCat/Templeton | 0.13.0 | y | y | y |
| Oozie | 4.0.0 | y | y | y |
| HBase | 0.98.0 | y | y | y |
| Hive(via WebHcat) | 0.13.0 | y | y | y |
| Hive(via JDBC/ODBC) | 0.13.0 | y | y | y |
| YARN RM | 2.5.0 | y | y | n |
| Storm | 0.9.3 | y | n | n |

## 2.3 Gateway Details

该小结用于描述Knox Gateway，包括以下内容：

1. gateway和Hadoop Clusters中的多个服务之间的URLs映射
2. 如果通过gateway-site.xml来配置gateway和集群相关的topology文件
3. 如果配置不同的policy enforcement provider，例如authentication,authorization,auditing和hostmapping等

### 2.3.1 URL Mapping

该功能比较类似于反向代理，Gateway向外提供的URLs与Hadoop Cluster的URL形成映射关系。

1. 默认的Topology URLs

为了和Hadoop Java Client和已有的Cli Tools兼容，Knox Gateway提供Default Topology的功能。当topology文件和配置的default topology name相符合时，一个特定的URLs映射被安装到gateway中。

这个配置在gateway-site.xml中配置的参数为default.app.topology.name，默认值为sandbox。因此当部署sandbox.xml topology时，下面的Example URLs和Hadoop Cluster的URLs保持一致：

*https://{gateway-host}:{gateway-port}/webhdfs*

*https://{gateway-host}:{gateway-port}/{gateway-path}/{cluster-name}/webhdfs*

1. Fully Qualified URLs

WebHDFS,WebHCat,Oozie和HBase的映射例子如下，在gateway-site.xml中配置，cluster topology的描述文件为{GATEWAY\_HOME}/conf/topologies/{cluster-name}.xml。例子上的服务端口为默认端口，根据实际配置进行改变：

1. WebHDFS

Gateway:https://{gateway-host}:{gateway-port}/{gateway-path}/{cluster-name}/webhdfs

Cluster:http://{webhdfs-host}:50070/webhdfs

1. WebHcat(Templeton)

Gateway: https://{gateway-host}:{gateway-port}/{gateway-path}/{cluster-name}/templeton

Cluster:http://{webhcat-host}:50111/templeton}

1. Oozie

Gateway: https://{gateway-host}:{gateway-port}/{gateway-path}/{cluster-name}/oozie

Cluster:http://{oozie-host}:11000/oozie}

1. HBase

Gateway: https://{gateway-host}:{gateway-port}/{gateway-path}/{cluster-name}/hbase

Cluster: http://{hbase-host}:8080

1. Hive JDBC

Gateway: jdbc:hive2://{gateway-host}:{gateway-port}/;ssl=true;sslTrustStore={gateway-trust-store-path};trustStorePassword={gateway-trust-store-password};transportMode=http;httpPath={gateway-path}/{cluster-name}/hive

Cluster: http://{hive-host}:10001/cliservice

其中：{gateway-host},{gateway-port}和{gateway-path}通过gateway configuration提供，在{GATEWAY\_HOME}/conf/gateway-site.xml，{cluster-name}从cluster topology descriptor，例如从{gateway\_home}/deployements/{cluster-name}.xml中衍化而来。

{webhdfs-host},{webhcat-host},{oozie-host},{hbase-host}和{hive-host}在集群topology descriptor中配置，例如{GATEWAY\_HOME}/conf/topologies/{cluster-name}.xml

注意：50070, 50111, 11000, 8080 及10001 这些端口是 WebHDFS, WebHCat, Oozie, HBase 和Hive的默认端口，这些值在实际的部署时是不同的。

HBase Rest API默认使用的端口是8080，这个端口经常和其他服务冲突，可以在启动的时候指定，例如使用60080端口。

### 2.3.2 Configuration

参考链接：

https://hortonworks.com/hadoop-tutorial/securing-hadoop-infrastructure-apache-knox/

http://www.datastart.cn/tech/2016/06/17/knox.html