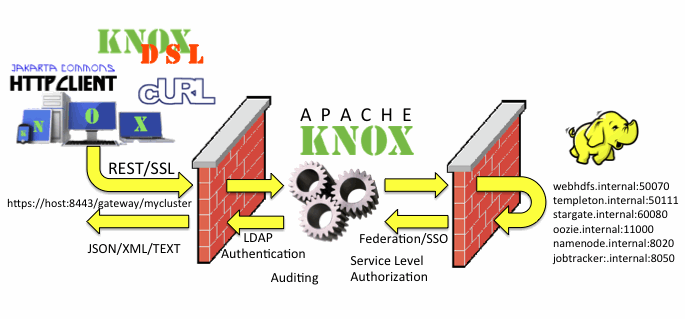
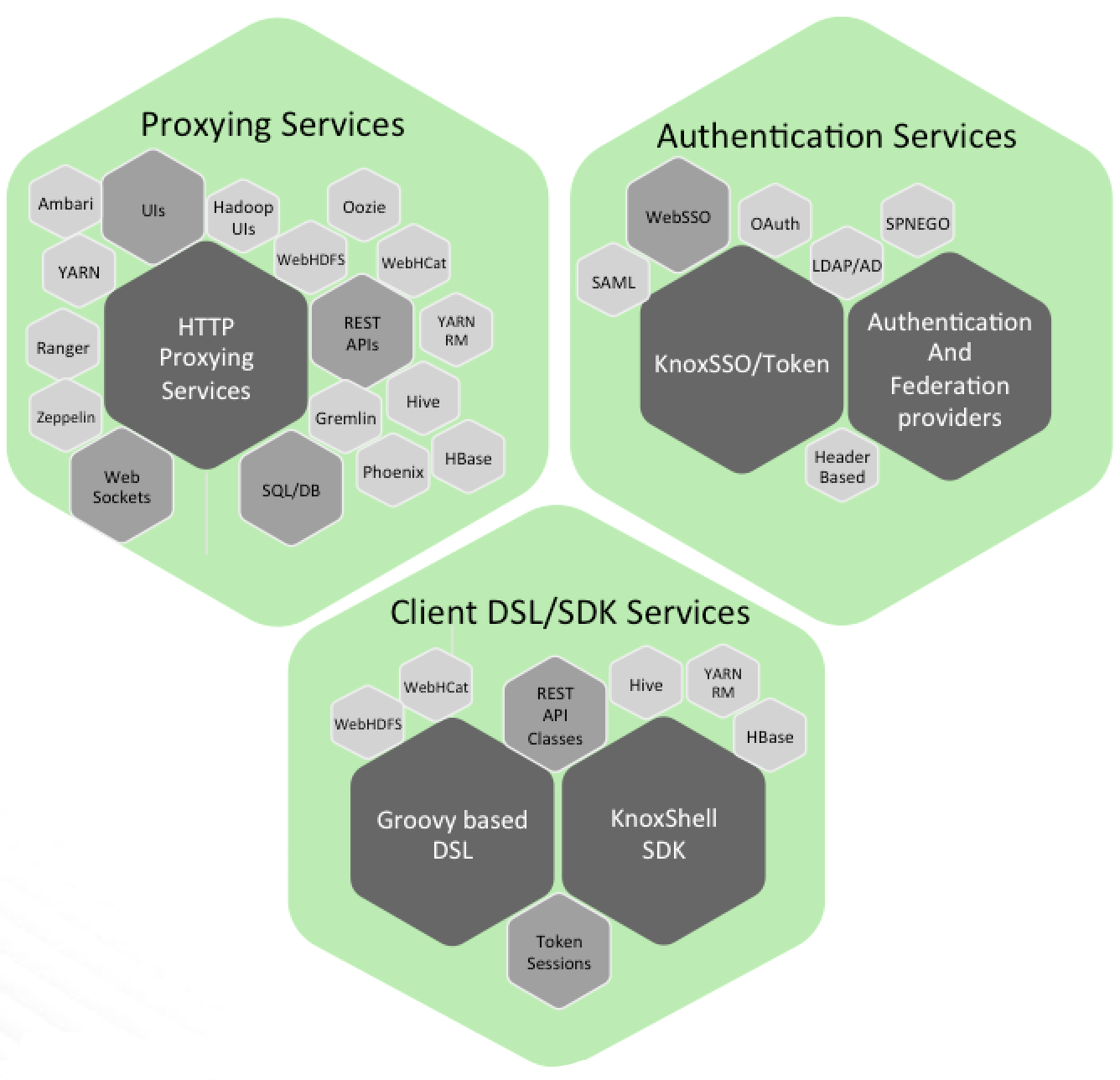
# Apache Knox概述



Apache Knox™网关是一个应用程序网关，用于与Apache Hadoop部署的REST API和UI进行交互。

Knox网关为与Hadoop集群的所有REST和HTTP交互提供了一个访问点。

Knox提供三组面向用户的服务：



**代理服务**  
Apache Knox项目的主要目标是通过HTTP资源代理提供对Apache Hadoop的访问。

**身份验证服务**  
用于REST API访问的身份验证以及用于UI的WebSSO流。LDAP / AD，基于标题的PreAuth，Kerberos，SAML，OAuth都是可用的选项。

**客户端服务**  
客户端开发可以通过DSL脚本编写，也可以直接将Knox Shell类用作SDK。

## 总览

Knox API网关被设计为反向代理，考虑到  
策略实施领域中通过其代理请求的提供程序和后端服务的可插入性。

策略实施范围包括身份验证/联合身份验证，授权，审计，调度，主机映射  
和内容重写规则。通过在拓扑  
部署描述符中为Knox门控的每个Apache Hadoop集群定义的供应商链来实施策略。集群定义也在  
拓扑部署描述符中定义，并为Knox网关提供  
集群布局，以便在面向用户的URL和集群内部之间进行路由和转换。

受Knox保护的每个Apache Hadoop集群都有一组REST API，这些API均由单个集群特定的  
应用程序上下文路径表示。这使Knox网关既可以保护多个群集，又可以  
为REST API使用者提供一个端点，以便跨  
多个群集访问所需的所有服务。

只需将拓扑部署描述符写入Knox安装的拓扑目录，即可  
处理新的Apache Hadoop集群定义，配置策略实施提供程序，并使应用程序  
上下文路径可供API使用者使用。

尽管不安全的Apache Hadoop集群有许多好处，  
但Knox网关也很好地补充了kerberos安全集群。

结合Kerberos保护的Apache Hadoop群集的适当网络隔离，  
Knox网关为企业提供了以下解决方案：

与企业身份管理解决方案完美集成

保护群集部署的详细信息（主机和端口对最终用户隐藏）

简化客户端与之交互所需的服务数量

## 支持的Apache Hadoop服务

以下Apache Hadoop生态系统服务已与Knox网关集成：

* Ambari
* WebHDFS (HDFS)
* Yarn RM
* Stargate (Apache HBase)
* Apache Oozie
* Apache Hive/JDBC
* Apache Hive WebHCat (Templeton)
* Apache Storm
* Apache Tinkerpop – Gremlin
* Apache Avatica/Phoenix
* Apache SOLR
* Apache Livy (Spark REST Service)
* Kafka REST Proxy

## 支持的Apache Hadoop生态系统UI

* NameNode UI
* Job History UI
* Yarn UI
* Apache Oozie UI
* Apache HBase UI
* Apache Spark UI
* Apache Ambari UI
* Apache Ranger Admin Console
* Apache Zeppelin
* Apache NiFi
* Hue
* Livy

## 配置对新服务和UI的支持

Apache Knox提供了一种配置驱动的方法来添加新的路由服务。这使新的Apache Hadoop REST API可以非常快速，轻松地加入。它还使用户和开发人员也可以将对自定义REST API的支持添加到Knox网关。

该功能已在0.6.0版中添加，并进一步增强了Knox对可扩展性和集成的承诺。

## 认证方式

具有身份验证角色的提供者负责收集API  
使用者提供的凭据，对其进行验证，并将成功或失败的身份验证传达给客户端或  
其他提供者链。

开箱即用，Knox网关提供Shiro身份验证提供程序。这是一个利用  
Apache Shiro项目针对LDAP用户存储对BASIC凭据进行身份验证的提供程序。支持  
OpenLDAP，ApacheDS和Microsoft Active Directory。

# 用户文档

下载地址： <http://knox.apache.org/books/knox-1-3-0/user-guide.html>

## 介绍

Apache Knox网关是一个为集群中的Apache Hadoop服务提供单点身份验证和访问的系统。目标是简化用户（即访问群集数据并执行作业的用户）和操作员（即控制访问并管理群集的操作员）的Hadoop安全性。网关作为服务器（或服务器集群）运行，可提供对一个或多个Hadoop集群的集中访问。通常，网关的目标如下：

* 为Hadoop REST API提供外围安全性，使Hadoop安全性更易于设置和使用

在外围提供身份验证和令牌验证

启用与企业和云身份管理系统的身份验证集成

在外围提供服务级别授权

* 公开单个URL层次结构，该层次结构聚合Hadoop集群的REST API

限制访问Hadoop集群所需的网络端点（因此限制防火墙漏洞）

隐藏潜在攻击者的内部Hadoop群集拓扑

## 快速开始

以下是使Apache Knox在Hadoop集群上启动并运行的步骤：

1. 验证系统要求
2. 使用Hadoop下载虚拟机​​（VM）
3. 下载Apache Knox网关
4. 使用Hadoop启动虚拟机
5. 安装Knox
6. 启动嵌入在Knox中的LDAP
7. 启动Knox网关
8. 用Knox做Hadoop

## Apache Knox详细信息

### Knox目录布局

| **目录** | **目的** |
| --- | --- |
| conf / | 包含全局应用到网关的配置文件（即不是特定于集群的）。 |
| data/ | 包含特定于安全性和拓扑的工件，这些工件需要在运行时进行读/写访问 |
| conf /topologies/ | 包含表示Hadoop集群的拓扑文件，网关用于部署集群代理 |
| data/security/ | 包含持久的主密钥和密钥库目录 |
| data/security/keystores/ | 包含网关和每个已部署群集拓扑的网关标识密钥库和凭据存储 |
| data/services | 包含当前支持的服务的服务行为定义。 |
| bin/ | 包含客户端和服务器的可执行Shell脚本，批处理文件和JAR。 |
| data/deployments/ | 包含用于保护对特定Hadoop群集的访问的已部署群集拓扑。 |
| lib / | 包含组成网关的所有组件的JAR。 |
| dep/ | 包含网关依赖的所有组件的JAR。 |
| ext / | 可以放置用户提供的扩展JAR的目录，以扩展网关的功能。 |
| pids / | 包含用于运行LDAP和网关服务器的进程ID |
| samples/ | 包含许多样本，可用于探索网关的功能。 |
| templates/ | 包含可以复制和定制的默认配置文件。 |
| README | 提供有关Apache Knox网关的基本信息。 |
| ISSUES | 描述重要的已知问题。 |
| CHANGES | 枚举版本之间的更改。 |
| LICENSE | 记录提供此软件所依据的许可证。 |
| NOTICE | 为包含的依赖项记录所需的归因通知。 |

### Knox支持服务

下表列举了经过测试可与Knox Gateway配合使用的各种Hadoop服务的版本。

| **服务** | **版本** | **不安全** | **安全** | **HA** |
| --- | --- | --- | --- | --- |
| WebHDFS | 2.4.0 | ÿ | ÿ | ÿ |
| WebHCat /Templeton | 0.13.0 | ÿ | ÿ | ÿ |
| Oozie | 4.0.0 | ÿ | ÿ | ÿ |
| Hbase | 0.98.0 | ÿ | ÿ | ÿ |
| Hive（via WebHCat） | 0.13.0 | ÿ | ÿ | ÿ |
| Hive（via JDBC / ODBC） | 0.13.0 | ÿ | ÿ | ÿ |
| Yarn ResourceManager | 2.5.0 | ÿ | ÿ | ñ |
| Kafka（via REST Proxy） | 0.10.0 | ÿ | ÿ | ÿ |
| Storm | 0.9.3 | ÿ | ñ | ñ |
| Solr | 5.5+和6+ | ÿ | ÿ | ÿ |

### 更多例子 [http://knox.apache.org/books/knox-1-3-0/markbook-section-link.png](http://knox.apache.org/books/knox-1-3-0/user-guide.html#More+Examples)

这些示例提供了有关如何通过Apache Knox网关访问各种Apache Hadoop服务的更多详细信息。