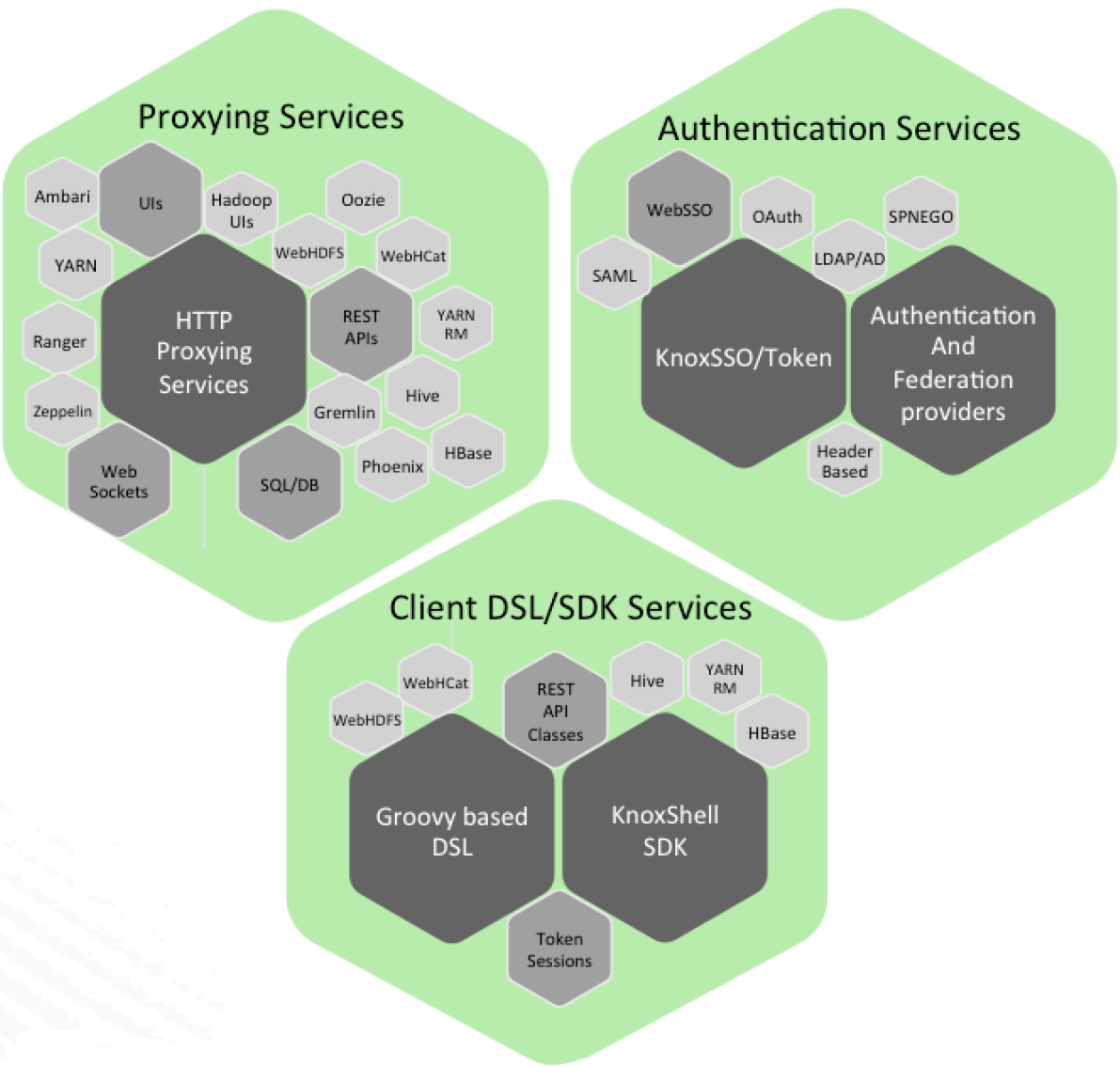
Announcing Apache Knox 1.6.1!

REST API and Application Gateway for the Apache Hadoop Ecosystem

The Apache Knox™ Gateway is an Application Gateway for interacting with the REST APIs and UIs of Apache Hadoop deployments.

The Knox Gateway provides a single access point for all REST and HTTP interactions with Apache Hadoop clusters.

Knox delivers three groups of user facing services:



- **Proxying Services**
  Primary goals of the Apache Knox project is to provide access to Apache Hadoop via proxying of HTTP resources.
- **Authentication Services**
  Authentication for REST API access as well as WebSSO flow for UIs. LDAP/AD, Header based PreAuth, Kerberos, SAML, OAuth are all available options.
- **Client Services**
  Client development can be done with scripting through DSL or using the Knox Shell classes directly as SDK. The KnoxShell interactive scripting environment
  combines the interactive shell of groovy shell with the Knox Shell SDK classes for a interating with data
  from your deployed Hadoop cluster.

Overview

The Knox API Gateway is designed as a reverse proxy with consideration for pluggability in the areas of policy enforcement, through providers and the backend services for which it proxies requests.

Policy enforcement ranges from authentication/federation, authorization, audit, dispatch, hostmapping and content rewrite rules. Policy is enforced through a chain of providers that are defined within the topology deployment descriptor for each Apache Hadoop cluster gated by Knox. The cluster definition is also defined within the topology deployment descriptor and provides the Knox Gateway with the layout of the cluster for purposes of routing and translation between user facing URLs and cluster internals.

Each Apache Hadoop cluster that is protected by Knox has its set of REST APIs represented by a single cluster specific application context path. This allows the Knox Gateway to both protect multiple clusters and present the REST API consumer with a single endpoint for access to all of the services required, across the multiple clusters.

Simply by writing a topology deployment descriptor to the topologies directory of the Knox installation, a new Apache Hadoop cluster definition is processed, the policy enforcement providers are configured and the application context path is made available for use by API consumers.

While there are a number of benefits for unsecured Apache Hadoop clusters, the Knox Gateway also complements the kerberos secured cluster quite nicely.

Coupled with proper network isolation of a Kerberos secured Apache Hadoop cluster, the Knox Gateway provides the enterprise with a solution that:

- Integrates well with enterprise identity management solutions
- Protects the details of the cluster deployment (hosts and ports are hidden from endusers)
- Simplifies the number of services that clients need to interact with

Supported Apache Hadoop Services

The following Apache Hadoop ecosystem services have integrations with the Knox Gateway:

Ambari
Cloudera Manager
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)
Apache Flink
Kafka REST Proxy

Supported Apache Hadoop ecosystem UIs

Name Node UI
Job History UI
Yarn UI
Apache Oozie UI
Apache HBase UI
Apache Spark UI
Apache Ambari UI
Apache Impala
Apache Ranger Admin Console
Apache Zeppelin
Apache NiFi
Hue
Livy

Configuring Support for new services and UIs

Apache Knox provides a configuration driven method of adding new routing services.
This enables for new Apache Hadoop REST APIs to come on board very quickly and easily. It also enables users and developers to add support for custom REST APIs to the Knox gateway as well.
This capability was added in release 0.6.0 and furthers the Knox commitment to extensibility and integration.

Home Page

Knox provides a conenient Home Page that may be used as the front door to your deployment and the resources that you have published for access through Apache Knox. This is a nice alternative to having to distribute a link to the administrative interface in order to get Quick Links.

Authentication

Providers with the role of authentication are responsible for collecting credentials presented by the API consumer, validating them and communicating the successful or failed authentication to the client or the rest of the provider chain.

Out of the box, the Knox Gateway provides the Shiro authentication provider. This is a provider that leverages the Apache Shiro project for authenticating BASIC credentials against an LDAP user store. There is support for OpenLDAP, ApacheDS and Microsoft Active Directory.

Federation/SSO

For customers that require credentials to be presented to a limited set of trusted entities within the enterprise, the Knox Gateway may be configured to federate the authenticated identity from an external authentication event. This is done through providers with the role of federation. The set of out-of-the-box federation providers include:

KnoxSSO Default Form-based IDP -

The default configuration of KnoxSSO provides a form-based authentication mechanism that leverages the Shiro authentication to authenticate against LDAP/AD with credentials collected from a form-based challenge.

Pac4J -

The pac4j provider adds numerous authentication and federation capabilities including: SAML, CAS, OpenID Connect, Google, Twitter, etc.

HeaderPreAuth -

A simple mechanism for propagating the identity through HTTP Headers that specify the username and group for the authenticated user. This has been built with vendor usecases such as SiteMinder and IBM Tivoli Access Manager.

KnoxSSO

The KnoxSSO service is an integration service that provides a normalized SSO token for representing the authenticated user. This token is generally used for WebSSO capabilities for participating UIs and their consumption of the Apache Hadoop REST APIs.
KnoxSSO abstracts the actual identity provider integration away from participating applications so that they only need to be aware of the KnoxSSO cookie. The token is presented by the browser as a cookie and applications that are participating in the KnoxSSO integration are able to cryptographically validate the presented token and remain agnostic to the underlying SSO integration.

Authorization

The authorization role is used by providers that make access decisions for the requested resources based on the effective user identity context. This identity context is determined by the authentication provider and the identity assertion provider mapping rules. Evaluation of the identity context's user and group principals against a set of access policies is done by the authorization provider in order to determine whether access should be granted to the effective user for the requested resource.

Out of the box, the Knox Gateway provides an ACL based authorization provider that evaluates rules that comprise of username, groups and ip addresses. These ACLs are bound to and protect resources at the service level.
That is, they protect access to the Apache Hadoop services themselves based on user, group and remote ip address.

Audit

The ability to determine what actions were taken by whom during some period of time is provided by the auditing capabilities of the Knox Gateway. The facility is built on an extension of the Log4j framework and may be extended by replacing the out of the box implementation with another.