Knox

Knox Gateway Proposal

Abstract

Knox Gateway is a system that provides a single point of secure access for Apache Hadoop clusters.

Proposal

The Knox Gateway (“Gateway” or “Knox”) is a system that provides a single point of authentication and access for Apache Hadoop services in a cluster. The goal is to simplify Hadoop security for both users (i.e. who access the cluster data and execute jobs) and operators (i.e. who control access and manage the cluster). The Gateway runs as a server (or cluster of servers) that serve one or more Hadoop clusters.

- Provide perimeter security to make Hadoop security setup easier
- Support authentication and token verification security scenarios
- Deliver users a single cluster end-point that aggregates capabilities for data and jobs
- Enable integration with enterprise and cloud identity management environments

Background

An Apache Hadoop cluster is presented to consumers as a loose collection of independent services. This makes it difficult for users to interact with Hadoop since each service maintains it’s own method of access and security. As well, for operators, configuration and administration of a secure Hadoop cluster is a complex and many Hadoop clusters are insecure as a result.

The goal of the project is to provide coverage for all existing Hadoop ecosystem projects. In addition, the project will be extensible to allow for new and/or proprietary Hadoop components without requiring changes to the gateway source code. The gateway is expected to run in a DMZ environment where it will provide controlled access to these Hadoop services. In this way Hadoop clusters can be protected by a firewall and only limited access provided through the firewall for the gateway. The authentication components of the gateway will be modular and extensible such that it can be integrated with existing security infrastructure.

Rationale

Organizations that are struggling with Hadoop cluster security result in a) running Hadoop without security or b) slowing adoption of Hadoop. The Gateway aims to provide perimeter security that integrates more easily into existing organizations’ security infrastructure. Doing so will simplify security for these organizations and benefit all Hadoop stakeholders (i.e. users and operators). Additionally, making a dedicated perimeter security project part of the Apache Hadoop ecosystem will prevent fragmentation in this area and further increase the value of Hadoop as a data platform.

Current Status

Prototype available, developed by the list of initial committers.

Meritocracy

We desire to build a diverse developer community around Gateway following the Apache Way. We want to make the project open source and will encourage contributors from multiple organizations following the Apache meritocracy model.

Community

We hope to extend the user and developer base in the future and build a solid open source community around Gateway. Apache Hadoop has a large ecosystem of open source projects, each with a strong community of contributors. All project communities in this ecosystem have an opportunity to participate in the advancement of the Gateway project because ultimately, Gateway will enable the security capabilities of their project to be more enterprise friendly.

Core Developers

Gateway is currently being developed by several engineers from Hortonworks - Kevin Minder, Larry McCay, John Speidel, Tom Beerbower and Sumit Mohanty. All the engineers have deep expertise in middleware, security & identity systems and are quite familiar with the Hadoop ecosystem.

Alignment

The ASF is a natural host for Gateway given that it is already the home of Hadoop, Hive, Pig, HBase, Oozie and other emerging big data software projects. Gateway is designed to solve the security challenges familiar to the Hadoop ecosystem family of projects.

Known Risks

Orphaned products & Reliance on Salaried Developers

The core developers plan to work full time on the project. We believe that this project will be of general interest to many Hadoop users and will attract a diverse set of contributors. We intend to demonstrate this by having contributors from several organizations recognized as committers by the time Knox graduates from incubation.

Inexperience with Open Source

All of the core developers are active users and followers of open source. As well, Hortonworks and the affiliated mentors have a strong heritage of success with contributions to Apache Hadoop Projects.

Homogeneous Developers

The current core developers are from Hortonworks, however, we hope to establish a developer community that includes contributors from several corporations.

Reliance on Salaried Developers

Currently, the developers are paid to do work on Gateway. However, once the project has a community built around it, we expect to get committers and developers from outside the current core developers.

Relationships with Other Apache Products

Gateway is going to be used by the users and operators of Hadoop, and the Hadoop ecosystem in general.

A Excessive Fascination with the Apache Brand

Our interest in developing Gateway in Apache project is to follow an established development model, as well since many of the Hadoop ecosystem projects also are part of Apache, Gateway will complement those projects by following the same development and contribution model.

Documentation

There is documentation in Hortonworks' internal repositories. These can be shared upon request and will be transferred into the Apache CM system if this proposal is accepted.

Initial Source

The current initial source can be found in a GitHub repository. https://github.com/hortonworks/knox.git

Source and Intellectual Property Submission Plan

The complete Gateway code is under Apache Software License 2.

External Dependencies

The Gateway dependencies are listed below, separated by Category A and Category B as defined in the Apache Third-Party Licensing Policy. Note: These are the direct dependencies. Indirect dependencies are not included.

Category A Dependencies

- Apache Commons - ASLv2.0
  - commons-io:commons-io#2.4
  - commons-cli:commons-cli#1.2
  - commons-codec:commons-codec#1.7
  - org.apache.commons:commons-digester3#3.2
  - org.apache.commons:commons-vfs2#2.0
- Apache Hadoop - ASLv2.0
  - org.apache.hadoop:hadoop-auth#0.23.3
  - org.apache.hadoop:hadoop-core#1.0.3
- Apache Geronimo - ASLv2.0
  - org.apache.geronimo.components:geronimo-jaspi#2.0.0
  - org.apache.geronimo.specs:geronimo-osgi-locator#1.1
- Apache Shiro - ASLv2.0
  - org.apache.shiro:shiro-web#1.2.1
- ApacheDS - ASLv2.0
  - org.apache.directory.server:apacheds-all#1.5.5
- Log4J - ASLv2.0
  - log4j:log4j#1.2.17
- SL4J - MIT

    - org.slf4j:slf4j-api#1.6.6
    - org.slf4j:slf4j-log4j12#1.6.6
- Guava - ASLv2.0
    - com.google.guava:guava#14.0-rc1
- HttpClient - ASLv2.0
    - org.apache.httpcomponents:httpclient#4.2.1
- Jetty - ASLv2.0
    - org.eclipse.jetty:jetty-server#8.1.7.v20120910
    - org.eclipse.jetty:jetty-servlet#8.1.7.v20120910
    - org.eclipse.jetty:jetty-webapp#8.1.7.v20120910
    - org.eclipse.jetty:jetty-jaspi#8.1.7.v20120910
    - org.eclipse.jetty.aggregate:jetty-all#8.1.7.v20120910
    - org.eclipse.jetty:test-jetty-servlet#8.1.7.v20120910
- JBoss ShrinkWrap - ASLv2.0
    - org.jboss.shrinkwrap:shrinkwrap-api#1.0.1
    - org.jboss.shrinkwrap:shrinkwrap-impl-base#1.0.1
    - org.jboss.shrinkwrap.descriptors:shrinkwrap-descriptors-api-javaee#2.0.0-alpha-4
    - org.jboss.shrinkwrap.descriptors:shrinkwrap-descriptors-impl-javaee#2.0.0-alpha-4

Category A Dependencies (Test)

- EasyMock - ASLv2.0
    - org.easymock:easymock#3.0
- XML Matchers - ASLv2.0
    - org.xmlmatchers:xml-matchers#0.10
- Hamcrest - BSDv3
    - org.hamcrest:hamcrest-api#1.0
    - org.hamcrest:hamcrest-core#1.2.1
    - org.hamcrest:hamcrest-library#1.2.1
- JsonPath - ASLv2.0
    - com.jayway.jsonpath:json-path#0.8.1
    - com.jayway.jsonpath:json-path-assert#0.8.1
- XMLTool - ASLv2.0
    - com.mycila.xmltool:xmltool#3.3
- REST-assured - ASLv2.0
    - com.jayway.restassured:rest-assured#1.6.2

Category B Dependencies

- Jersey - CDDLv1.1 or GPL2wCPE
    - com.sun.jersey:jersey-server#1.14
    - com.sun.jersey:jersey-servlet#1.14
- Jerico - EPLv1.0
    - net.htmlparser.jericho:jericho-html#3.2
- Servlet - CDDLv1.0 or GPLv2
    - javax.servlet:javax.servlet-api#3.0.1
- JUnit - CPLv1.0
    - junit:junit#4.11

Cryptography

The Gateway uses cryptographic software indirectly as a result of having two dependencies: ApacheDS and Apache Shiro. Gateway does not include any special or custom cryptographic technologies.

ApacheDS is an ASF project and has been classified Export Commodity Control Number (ECCN) 5D002.C.1 due to it's dependency on Bouncy Castle. More information on the ApacheDS classification can be found at http://svn.apache.org/repos/asf/directory/apacheds/trunk/installers/README

Apache Shiro is an ASF project and has been classified Export Commodity Control Number (ECCN) 5D002.C.1. More information on the Apache Shiro classification can be found at http://svn.apache.org/repos/asf/shiro/trunk/README

Required Resources

Mailing lists

knox-dev AT incubator DOT apache DOT org knox-commits AT incubator DOT apache DOT org knox-user AT hms incubator apache DOT org knox-private AT incubator DOT apache DOT org

Subversion Directory

https://svn.apache.org/repos/asf/incubator/knox

Issue Tracking

JIRA Knox (KNOX)

Initial Committers

- Kevin Minder (kevin DOT minder AT hortonworks DOT com)
- Larry McCay (lmccay AT hortonworks DOT com)
- John Speidel (jspeidel AT hortonworks DOT com)
- Tom Beerbower (tbeerbower AT hortonworks DOT com)
- Sumit Mohanty (smohanty AT hortonworks DOT com)
- Venkatesh Seetharam (venkatesh AT hortonworks DOT com)

Affiliations

- Kevin Minder (Hortonworks)
- Larry McCay (Hortonworks)
- John Speidel (Hortonworks)
- Tom Beerbower (Hortonworks)
- Sumit Mohanty (Hortonworks)
- Venkatesh Seetharm (Hortonworks)
- Owen O'Malley (Hortonworks)
- Mahadev Konar (Hortonworks)
- Alan Gates (Hortonworks)
- Devaraj Das (Hortwonrks)
- Chris Douglas (Microsoft)
- Chris Mattmann (NASA)
- Tom White (Cloudera)

Sponsors

Champion

Devaraj Das (ddas AT apache DOT org)

Nominated Mentors

- Owen O'Malley (omalley AT apache DOT org)
- Mahadev Konar (mahadev AT apache DOT org)
- Alan Gates (gates AT apache DOT org)
- Devaraj Das (ddas AT apache DOT org)
- Chris Douglas (cdouglas AT apache DOT org)
- Chris Mattmann (chris DOT a DOT mattmann AT jpl DOT nasa DOT gov)
- Tom White (tom DOT e DOT white AT gmail DOT com)

Sponsoring Entity

Incubator PMC