Authentication for Hadoop HTTP web-consoles

Table of contents

1 Introduction

This document describes how to configure Hadoop HTTP web-consoles to require user authentication.

By default Hadoop HTTP web-consoles (JobTracker, NameNode, TaskTrackers and DataNodes) allow access without any form of authentication.

Similarly to Hadoop RPC, Hadoop HTTP web-consoles can be configured to require Kerberos authentication using HTTP SPNEGO protocol (supported by browsers like Firefox and Internet Explorer).

In addition, Hadoop HTTP web-consoles support the equivalent of Hadoop's Pseudo/ Simple authentication. If this option is enabled, user must specify their user name in the first browser interaction using the `user.name` query string parameter. For example: `http:// localhost:50030/jobtracker.jsp?user.name=babu`.

If a custom authentication mechanism is required for the HTTP web-consoles, it is possible to implement a plugin to support the alternate authentication mechanism (refer to Hadoop Alfredo for details on writing an `AuthenticatorHandler`).

The next section describes how to configure Hadoop HTTP web-consoles to require user authentication.

2 Configuration

The following properties should be in the `core-site.xml` of all the nodes in the cluster.

`hadoop.http.filter.initializers`: add to this property the `org.apache.hadoop.security.AuthenticationFilterInitializer` initializer class.

`hadoop.http.authentication.type`: Defines authentication used for the HTTP web-consoles. The supported values are: `simple | kerberos | #AUTHENTICATION_HANDLER_CLASSNAME#`. The dfeault value is `simple`.

`hadoop.http.authentication.token.validity`: Indicates how long (in seconds) an authentication token is valid before it has to be renewed. The default value is `36000`.

`hadoop.http.authentication.signature.secret.file`: The signature secret file for signing the authentication tokens. If not set a random secret is generated at startup time. The same secret should be used for all nodes in the cluster, JobTracker, NameNode, DataNode and TastTracker. The default value is `${user.home}/hadoop-http-auth-signature-secret`. IMPORTANT: This file should be readable only by the Unix user running the daemons.

`hadoop.http.authentication.cookie.domain`: The domain to use for the HTTP cookie that stores the authentication token. In order to authentiation to work correctly across all nodes in the cluster the domain must be correctly set. There is no default value, the HTTP cookie will not have a domain working only with the hostname issuing the HTTP cookie.

IMPORTANT: when using IP addresses, browsers ignore cookies with domain settings. For this setting to work properly all nodes in the cluster must be configured to generate URLs with hostname.domain names on it.

`hadoop.http.authentication.simple.anonymous.allowed`: Indicates if anonymous requests are allowed when using 'simple' authentication. The default value is `true`

`hadoop.http.authentication.kerberos.principal`: Indicates the Kerberos principal to be used for HTTP endpoint when using 'kerberos' authentication. The principal short name must be `HTTP` per Kerberos HTTP SPNEGO specification. The default value is `HTTP/localhost@$LOCALHOST`.

`hadoop.http.authentication.kerberos.keytab`: Location of the keytab file with the credentials for the Kerberos principal used for the HTTP endpoint. The default value is `${user.home}/hadoop.keytab.i`