ArgusProposal

Apache Argus Proposal

Abstract

Argus is a framework to enable, monitor and manage comprehensive data security across the Hadoop platform.

The name "Argus" is derived from Argus Panoptes, a 100-eyed giant in Greek mythology, endowed with a role to keep "an eye" open and be an effective watchman at all times.

Background

The vision with Argus is to provide comprehensive security across the Apache Hadoop ecosystem. With the advent of Apache YARN, the Hadoop platform can now support a true data lake architecture. Enterprises can potentially run multiple workloads, in a multi tenant environment. Data security within Hadoop needs to evolve to support multiple use cases for data access, while also providing a framework for central administration of security policies and monitoring of user access.

XA Secure, a Hadoop security focused startup, developed the initial technology behind Argus. XA Secure was acquired by Hortonworks, which now is contributing the technology to the open source community to extend and innovate.

Rationale

Many of the projects in the Hadoop ecosystem have their own authentication, authorization, and auditing components. There are no central administration and auditing capabilities. We are looking to address these enterprises security needs of central administration and comprehensive security through the Argus project. Our initial focus would be around authorization and auditing, the longer term vision would be to tie all aspects around data security within the Hadoop platform.

Proposal Details

The vision of Argus is to enable comprehensive data security across the Hadoop platform. The goal is provide a single user interface or API to manage security policies, monitor user access and policy changes history. The framework would work with individual components in enforcing these policies and in capturing relevant audit information.

Initial Goals

- Donate the Argus source code and documentation to the Apache Software Foundation
- Setup and standardize the open governance of the Argus project
- Build a user and developer community
- Deeper Integration with Hadoop Platform
- Enable integration with Apache Storm, Apache Knox and Apache Falcon for authorization and auditing
- Configurable centralized storage of audit data into HDFS
- Enable framework to be run in both Linux and Windows environments
- Rationalize install procedure, making it easier for enterprises to deploy

Longer Term Goals

In longer term, Argus should provide a comprehensive security framework for Hadoop platform components, covering the following

- Centralized security administration to manage all security related tasks in a central UI
- Fine grained authorization to do a specific action and/or operation with Hadoop component/tool and managed through a central administration tool
- Standardize authorization method across all Hadoop components
- Enhanced support for different authorization methods - Role based access control, attribute based access control etc
- Enable tag based global policies
- Centralize auditing of user access and administrative actions (security related) within all the components of Hadoop

Current Status

Argus' technology is currently being used by enterprises and is under active development.

The key components of Argus are:

- Enterprise Security Administration Portal
- A Java Web Application, designed for administration of security policies from a single location for the entire hadoop cluster (and even multiple hadoop clusters)
- Security Agents
- A light-weight Java Agent, which will be embedded into the hadoop component (e.g. Hive, HBase and Hadoop) as an authorization provider to enforce the security policies and also collect access events/logs.
- User/Group Synchronizer Module

- A standalone daemon which allows the user/group information to be synched from the enterprise user repositories like LDAP/AD to Argus local database. This user/group information in Argus local database will help the security policy administrators
- to define security policies by selecting users/groups from a drop-down box (instead of typing their name/group in a text-box).
- to delegate policy administration to other users/groups
- to restrict view of reports based on the users/groups
- Centralized Audit Logs and Monitoring
- Log events to central data storage/database
- Interactive query of audit events
- Audit administrator actions

The initial version provides ability to

- Define security policies using a central security administration UI.
- Fine grained access control for HDFS (file level) , Hive (column level) and HBase (column level)
- Framework to record access/operational events/logs as part of auditing and view using a central monitoring UI.
- Support delegated policy administration
- Centralized audit monitoring and query capabilities

Meritocracy

We plan to invest in supporting a meritocracy. We will discuss the requirements in an open forum. Several companies have already expressed interest in this project, and we intend to invite additional developers to participate. We will encourage and monitor community participation so that privileges can be extended to those that contribute.

Community

We are happy to report that there are existing Apache committers and corporate users who are closely involved in the project already. We hope to extend the user and developer base further in the future and build a solid open source community around Argus, growing the community and adding committers following the Apache meritocracy model.

Core Developers

The initial technology within Argus was originally built by the team at XA Secure. XA Secure was founded and managed by experienced members with a wide background in enterprise security. Some of the XA's core team have been proposed as core developers for this project. The developer list also include an Apache member and PMC members from several Apache projects (Hadoop, HBase, and Knox). A concern is that all of the core developers are employed by Hortonworks and thus an emphasis will be on increasing the diversity of the developer community.

Alignment

The initial committers strongly believe that a unified security portal for Apache Hadoop, Hive, and HBase will gain broad adoption as an open source, community driven project. Our hope is that the Apache Falcon, Apache Storm, Apache Knox, and other communities will find tremendous value in Argus and will adopt it en masse.

Known Risks

Orphaned Products

The initial code behind Argus is under active development and is being actively used by several enterprises. It is not expected to be orphaned.

Inexperience with Open Source

Many of the core developers have long-standing experience in open source, Dili Aramugam, Kevin Minder and Larry McCay are committers on the Apache Knox project. Sanjay Radia and Owen O'Malley are PMC members on several Apache projects. We have several mentors that will work with the inexperienced committers on building a thriving developer community.

Homogeneous Developers

The current core developers are all from Hortonworks. However, we expect to establish a thriving developer community that includes users of Argus and developers of other Hadoop components.

Reliance on Salaried Developers

Currently, all of the developers are paid to work on Argus. A key goal for the incubation process will be to broaden the developer base.

Relationships with Other Apache Products

The biggest risk is fast rate of growth of new features within the Hadoop ecosystem and security standards not being applied during the initial development of these new products. We believe an active engagement from the Hadoop community would significantly aid adoption of common security framework across the ecosystem and will help in establishing cross component standards.

As mentioned in the Alignment section, Argus is closely integrated with Hadoop, Hive and HBase in a numerous ways. We look forward to collaborating with those communities, as well as other Apache communities.

There is some overlap between the goals of Argus and Apache Sentry. Apache encourages disjoint teams to form independent projects, even when those projects overlap in scope. Additionally, we feel that the distinct code bases, development teams, and different approaches to the problem should be represented by different projects. This will provide better choices for users to choose from.

An Excessive Fascination with the Apache Brand

While we respect the reputation of the Apache brand and have no doubts that it will attract contributors and users, our interest is primarily to give Argus a solid home as an open source project with a broad developer base and to encourage adoption by the related ASF projects and foster innovation around security

Documentation

http://hortonworks.com/blog/hortonworks-acquires-xasecure-to-provide-comprehensive-security-for-enterprise-hadoop/

Initial Source

We will make the initial source available as a patch.

Source and IP Submission Plan

- All source will be moved to Apache Infrastructure
- All outstanding issues in our in-house JIRA infrastructure will be replicated into the Apache JIRA system.
- We will be acquiring a twitter handle for project Argus (eg: @apacheargus)

External Dependencies

Argus has no external dependencies except for some Java libraries that are considered ASF-compatible (JUnit, SLF4J, …) and Apache artifacts : Hadoop, Log4J and the transient dependencies of all these artifacts.

Cryptography

Argus does not incorporate encryption currently.

Required Resources

Mailing Lists:

- argus-dev
- argus-user
- argus-commits
- argus-private

Infrastructure:

- Git repository
- JIRA Argus
- Gerrit for reviewing patches

The existing code includes local host integration tests, so we would like a Jenkins instance to run them whenever a new patch is submitted.

Initial Committers

- Balaji Ganesan (bganesan at hortonworks.com)
- Dilli Arumugam (darumugam at hortonworks.com)
- Don Bosco Durai (bdurai at hortonworks.com)
- Kevin Minder (kminder at apache.org)
- Larry McCay (lmccay at apache.org)
- Madhanmohan Neethiraj (mneethiraj at hortonworks.com)
- Owen O'Malley (omalley at apache.org)
- Ramesh Mani (rmani at hortonworks.com)
- Sanjay Radia (sradia at apache.org)
- Selvamohan Neethiraj (sneethiraj at hortonworks.com)

Affiliations

- Balaji Ganesan - Hortonworks
- Dilli Arumugam - Hortonworks
- Don Bosco Durai - Hortonworks
- Kevin Minder - Hortonworks
- Larry McCay - Hortonworks
- Madhanmohan Neethiraj - Hortonworks
- Owen O'Malley - Hortonworks
- Ramesh Mani - Hortonworks
- Sanjay Radia - Hortonworks
- Selvamohan Neethiraj - Hortonworks

Sponsors

Champion:

- Owen O'Malley (omalley at apache.org) - Hortonworks

Nominated Mentors:

- Alan Gates - Hortonworks
- Daniel Gruno - Quokka IvS
- Devaraj Das - Hortonworks
- Jakob Homan - LinkedIn
- Owen O'Malley - Hortonworks

Sponsoring Entity

Incubator PMC