Apache Ranger™ is a framework to enable, monitor and manage comprehensive data security across the Hadoop ecosystem.

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

The vision with Ranger 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...

Apache Ranger – Frequently Asked Questions

General

What does Apache Ranger offer for Apache Hadoop and related components?

> Apache Ranger offers a centralized security framework to manage fine grained access control over Hadoop and related components (Apache Hive, HBase etc.). Using the Apache Ranger administration console, users can easily manage policies around accessing a resource (file, folder, database, table, column etc) for a particular set of users and/or groups, and enforce the policies within Hadoop. They also can enable audit tracking and policy analytics for deeper control of the environment. Apache Ranger also provides ability to delegate administration of certain data to other group owners, with an aim of decentralizing data ownership

What projects does Apache Ranger support today

> Apache Ranger supports fine grained authorization and auditing for following Apache projects:

- Apache Hadoop

- Apache Hive

- Apache HBase

- Apache Storm

- Apache Knox

- Apache Solr

- Apache Kafka

- YARN

How does it work over Hadoop and related components

Apache Ranger at the core has a centralized web application, which consists of the policy administration, audit and reporting modules. Authorized users will be able to manage their security policies using the web tool or using REST APIs. These security policies are enforced within Hadoop ecosystem using lightweight Ranger Java plugins, which run as part of the same process as the namenode (HDFS), Hive2Server(Hive), HBase server (Hbase), Nimbus server (Storm) and Knox server (Knox) respectively. Thus there is no additional OS level process to manage.

Is there a single point of failure?

No, Apache Ranger is not a Single Point of Failure. Apache Ranger's plugins run within the same process as the component, e.g. NameNode for HDFS. These agents pull the policy-changes using REST API at a configured regular interval (e.g.: 30 second). The plugin is able to function even if the policy server is temporarily down and will provide the authorization enforcement. Also, the policy manager web application can be hosted on a HA infrastructure. (with multiple apache server, multiple tomcat servers and a standby database server w/o replication setup).

Apache Hadoop

How does Apache Ranger provide authorization in Apache Hadoop?

Apache Ranger provides a plugin for Apache Hadoop, specifically for the NameNode as part of the authorization method. The Apache Ranger plugin is in the path of the user request and is able to make a decision on whether the user request shoud be authorized. The plugin also collects access request details required for auditing

Apache Ranger will enforce the security policies available in the policy database. Users can create a security policy for a specific set of resources (one or more folders and/or files) and assign specific set of permissions (e.g: read, write, execute) to a specific set of users and/or groups. The security policies are stored in the policy manager and are independent from native permissions.

Does Apache Ranger emulated permissions at the unix level for Apache Hadoop?

No, Apache Ranger enforces authorization based on policies entered in the policy administration tool and does not emulate the permissions at the unix level. Apache Ranger does provide a default feature to validate access using native hadoop file-level permissions if the Ranger policies do not cover the requested access

Does the Apache Ranger plugin need to be implemented in each datanode ?

No, the Apache Ranger plugin for Hadoop is only needed in the NameNode.

Apache Hive

How does Apache Ranger provide authorization in Apache Hive?

The Apache Ranger plugin is enabled in Hiveserver2 as part of the authorization

How does Apache Ranger authorization compare to SQL standard authorization?

Apache Hive currently provides two methods of authorization, Storage based authorization and SQL standard authorization, which was introduced in Hive 13. SQL standard authorization provides grant/revoke functionality at database, table level. The commands would be familiar to a DBA admin. Apache Ranger provides a centralized authorization interface for Hive and provides more granular access control at column level through

the Hive plugin. Ranger also provides ability to use wildcard in resource names within the policy.

Apache Knox

How does Apache Ranger provide authorization in Apache Knox? Apache Knox currently provides a service level authorization for users/groups. These acls are stored locally in a file. Apache Ranger has built a plugin for Knox to enable administration of these policies through central UI/REST APIs as well as detailed auditing of Knox user access.

Apache Kafka

How does Apache Ranger provide authorization in Apache Kafka? Security was introduced in Apache Kafka 0.9. Apache Ranger can manage the Kafka ACLs per topic. Users can use Ranger to control who can write to a topic or read from a topic. In addition to providing policies by users and groups, Apache Ranger also supports IP address based permissions to publish or subscribe.

Apache Solr

How does Apache Ranger provide authorization in Apache Solr? Similar to Apache Kafka, security in Apache Solr was introduced recently by the community. Through Apache Ranger, users can build policies for users/groups to query a particular collections in Solr. Efforts are underway in Solr community to provide more granular index level permissions.

YARN

How does Apache Ranger provide authorization in YARN? YARN is widely used in the Hadoop ecosystem as resource management layer for applications. Adminstrators can use YARN to setup queues with a certain capacity and applications

can be given permissions to write to a certain queue. Using Apache Ranger, administrators can manage the policies for who can write to a particular queue