Hadoop Security

Hadoop Contains Sensitive Data

- All data is security relevant
- Improper usage or breaches of data will cause huge damage to the business
- Hadoop is governed by the same security requirements as any data center platform

Hadoop is Subject to Compliance Adherence

- Organization are often subject to comply with regulations such as HIPPA, PCI-DSS that require protection of personal information
- Adhere to other corporate security policies

Hortonworks Secure Data Lakes (July 2015)

Five pillars of enterprise security

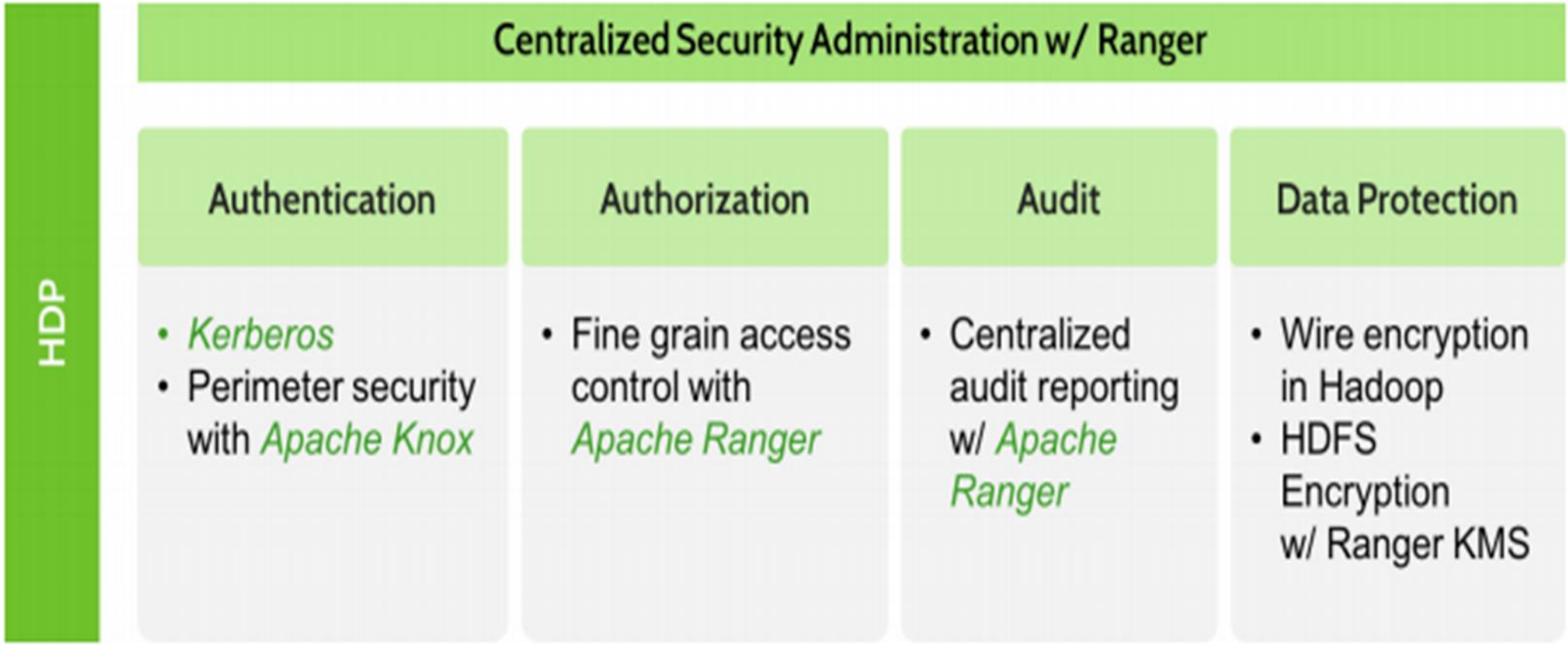
Administration Central management and consistent security	How do I set policy across the entire cluster?
Authentication Authenticate users and systems	Who am I/ prove it?
Authorization Provision access to data	What can I do?
Audit Maintain a record of data access	What did I do?
Data Protection Protect data at rest and in motion	How can I encrypt data at rest and over the wire?

http://info.hortonworks.com/rs/549-QAL-086/images/**Security**-White-Paper.pdf

Hortonworks Data Platform 2.3

http://info.hortonworks.com/rs/549-QAL-086/images/**Security**-White-Paper.pdf

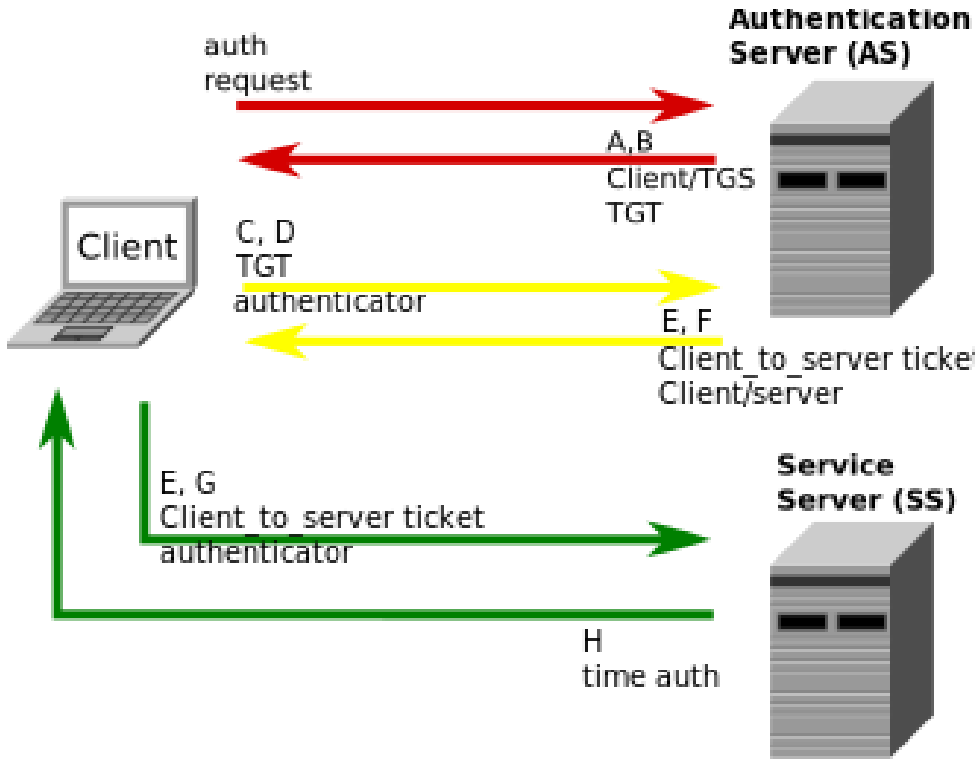
Hortonworks: Roles of Security Components



http://info.hortonworks.com/rs/549-QAL-086/images/**Security**-White-Paper.pdf

Kerberos

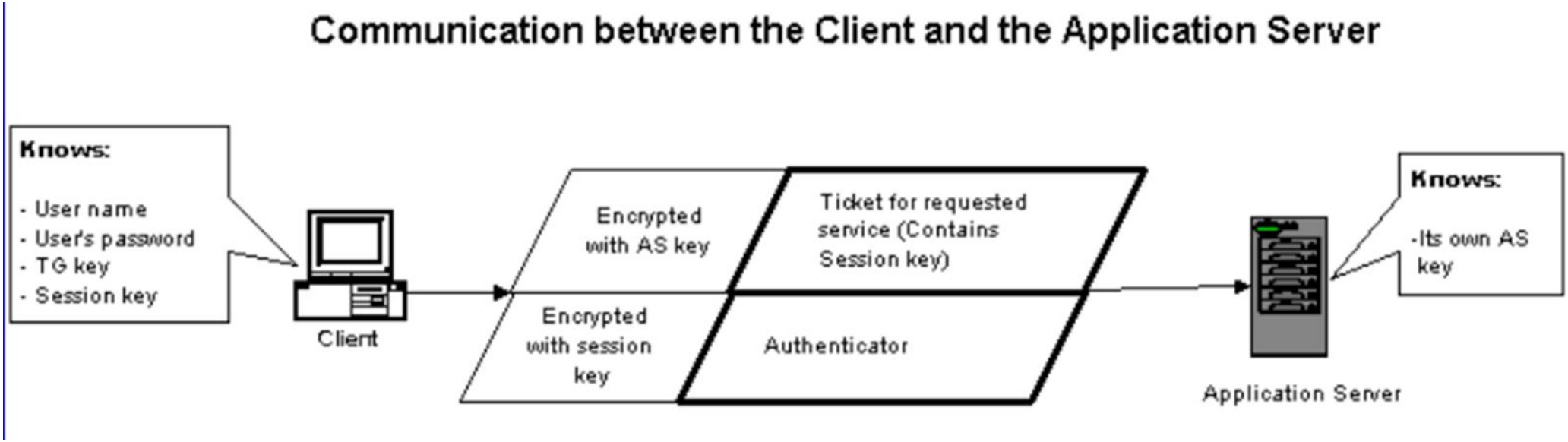
Kerberos



- Client requests an "Auth" server for a "ticket" to access a service
  - Sends: UserID, Service command
  - The Auth server knows all UserIDs passwords, and Service Keys
- Server checks UserID and service and:
  - Creates ticket containing service request encrypted using the service server key
  - Ticket encrypted with users password and returned to client
- Client decrypts the ticket & uses it to make all service requests to the cluster
  - At the cluster, the ticket is decrypted using the service server key
- Each ticket will have an "expiry time", ~8 hours

Published in the late 80's as a network security protocol
https://en.wikipedia.org/wiki/Kerberos_(protocol) http://web.stanford.edu/~torg/kerberos-overview.html

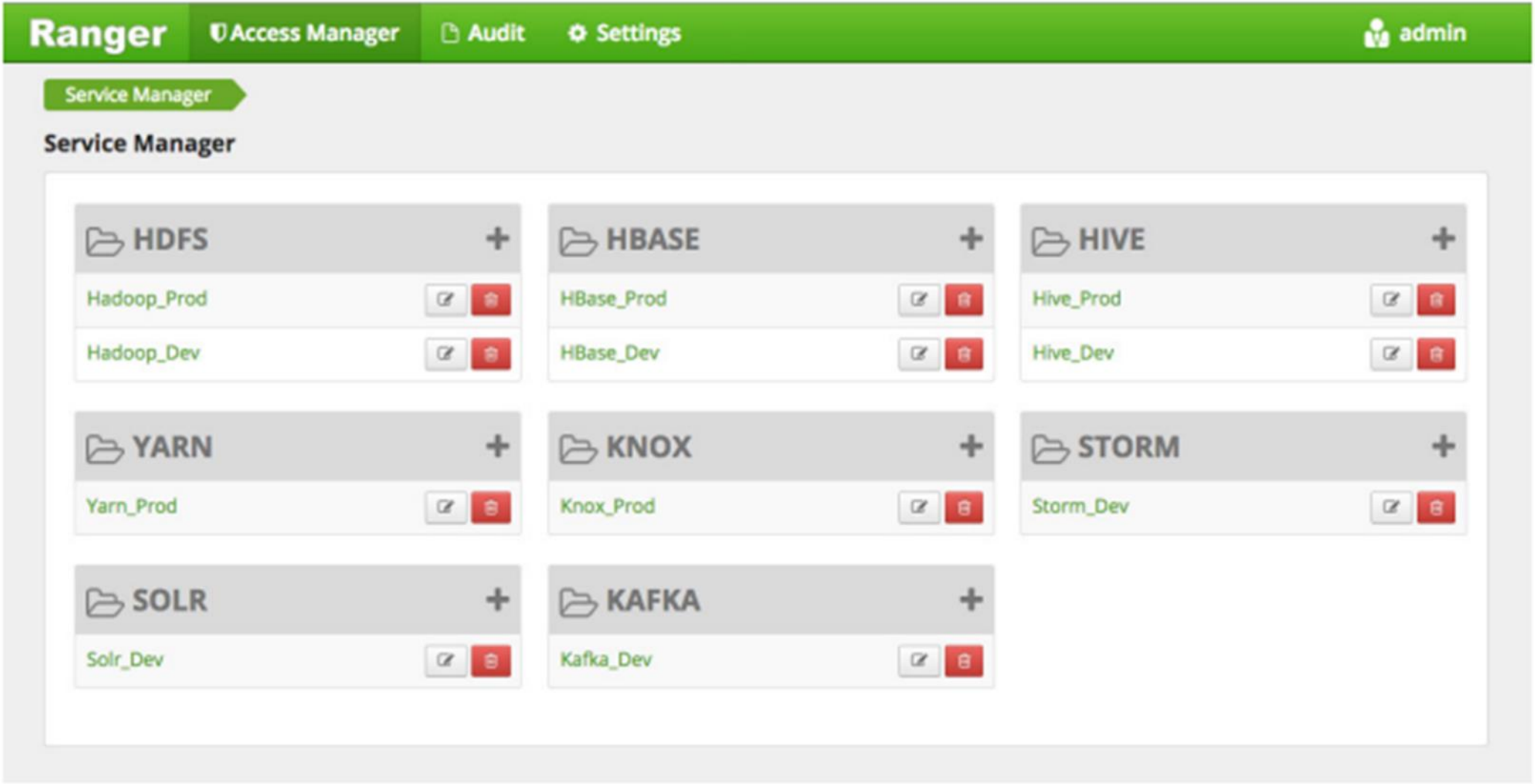
Kerberos



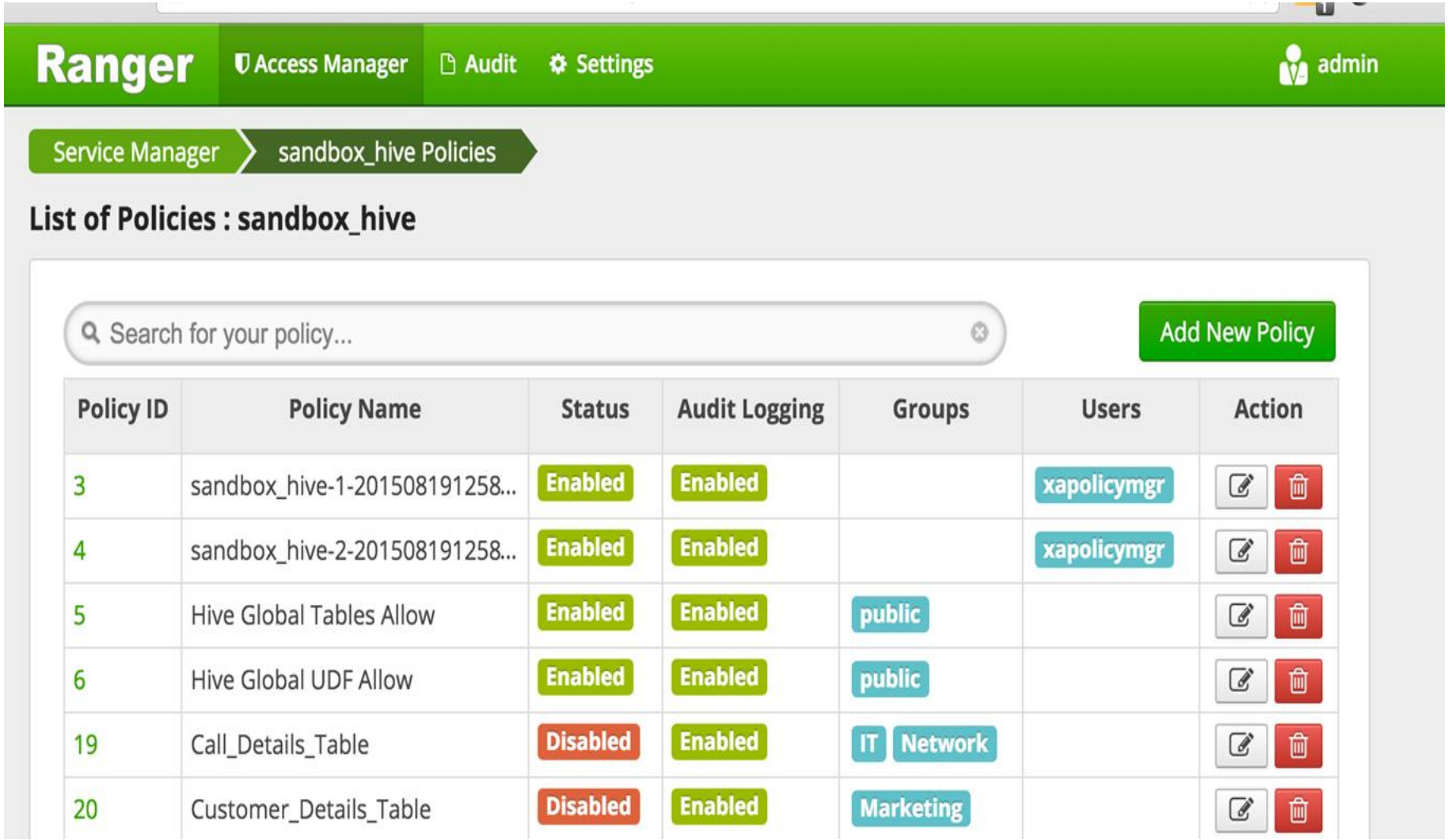
Published in the late 80's as a network security protocol
https://en.wikipedia.org/wiki/Kerberos_(protocol) http://web.stanford.edu/~torg/kerberos-overview.html

Apache Ranger

Apache Ranger: “Single Pane of Glass” to the Administrator



Apache Ranger: “Single Pane of Glass” to the Administrator



Ranger Architecture

Security Administrators

Enterprise Legacy Systems

Define Policies
Audit user activity

Policy Integration

Enterprise Directory Services

Sync Users

Ranger User Sync Server

Sync Users

Ranger Policy Admin Server

Authenticate

Enterprise Users

Access Data in Hadoop Systems

HDFS — Ranger Plugin

Hive — Ranger Plugin

HBase — Ranger Plugin

Knox — Ranger Plugin

Storm — Ranger Plugin

Sync Policies

Ranger Policy DB

Ranger Audit Store

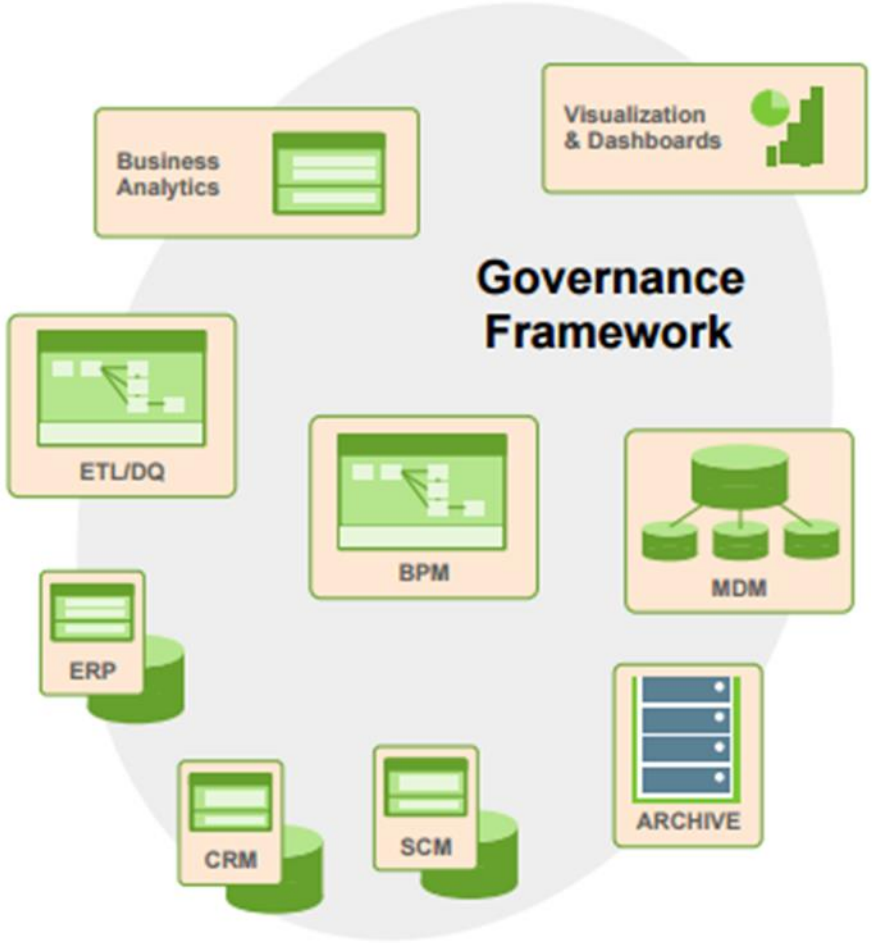
Store Audit Activity

Ranger Plugins Authorize the Access & Audit the Activity

		APACHE RANGER
Platform-wide coverage across Hadoop stack	•	Coverage across HDFS, YARN, Hive, HBase, Storm, Knox, Solr and Kafka
Fine grain authorization	•	Authorize security policies for a database, table and column or a file as well as LDAP based groups or individual user
Provide hooks for dynamic policy-based authorization	•	Specify dynamic conditions in service definitions Flexibility to define unique conditions by service (HDFS, Hive etc.)
Built on pluggable service-based model	•	Custom plugins can be created for any data store

Apache Atlas

Apache Atlas: Enterprise Data Governance Goals

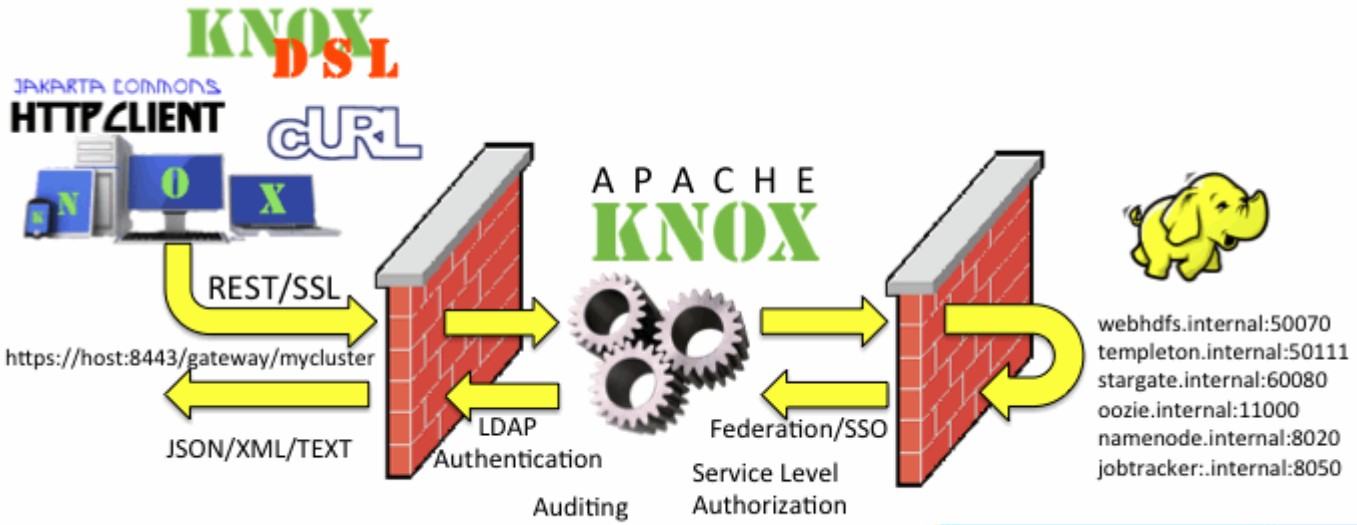


GOAL: Provide a common approach to data governance across all systems and data within the organization

- **Transparent**
  Governance standards & protocols must be clearly defined and available to all
- **Reproducible**
  Recreate the relevant data landscape at a point in time
- **Auditable**
  All relevant events and assets but be traceable with appropriate historical lineage
- **Consistent**
  Compliance practices must be consistent

Apache Knox

Apache Knox:

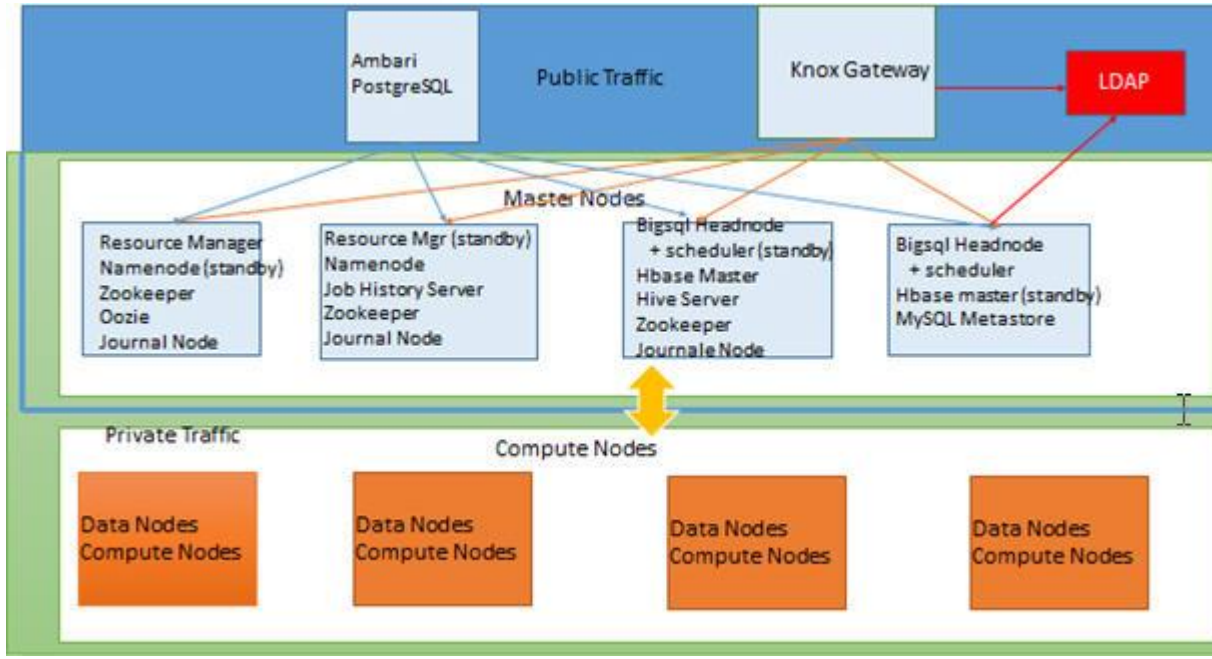


Hides hosts and ports, cluster deployment can be hidden from public

Integrated with enterprise identity management solutions

Creates single point of access for all REST based services

Uses SSL and is excellent when combined with Kerberos



http://www.ibm.com/support/knowledgecenter/SSPT3X_4.1.0/com.ibm.swg.im.infosphere.biginsights.admin.doc/doc/knox_overview.html
https://knox.apache.org/