

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

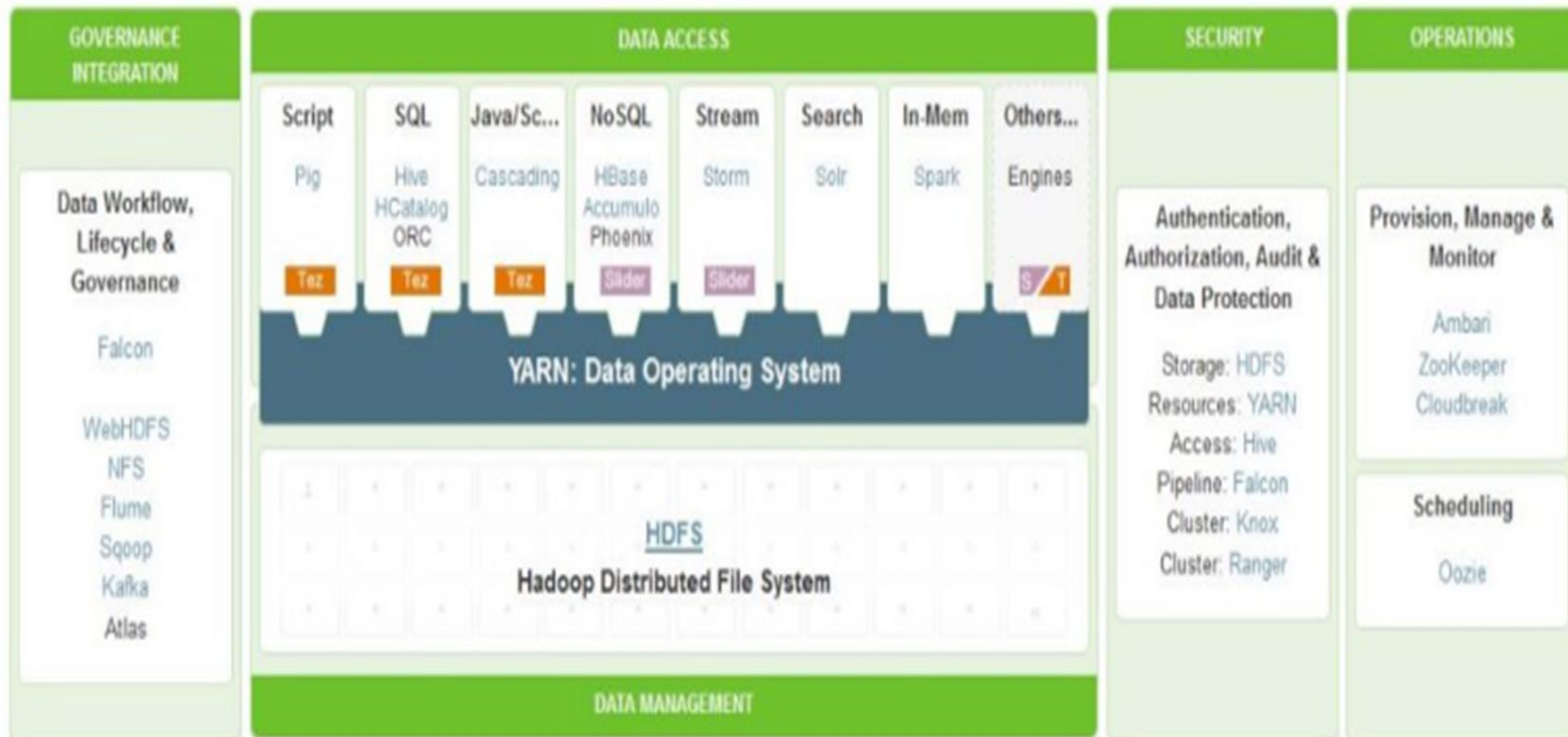
What did I do?

Data Protection

Protect data at rest and in motion

How can I encrypt data at rest and over the wire?

Hortonworks Data Platform 2.3



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

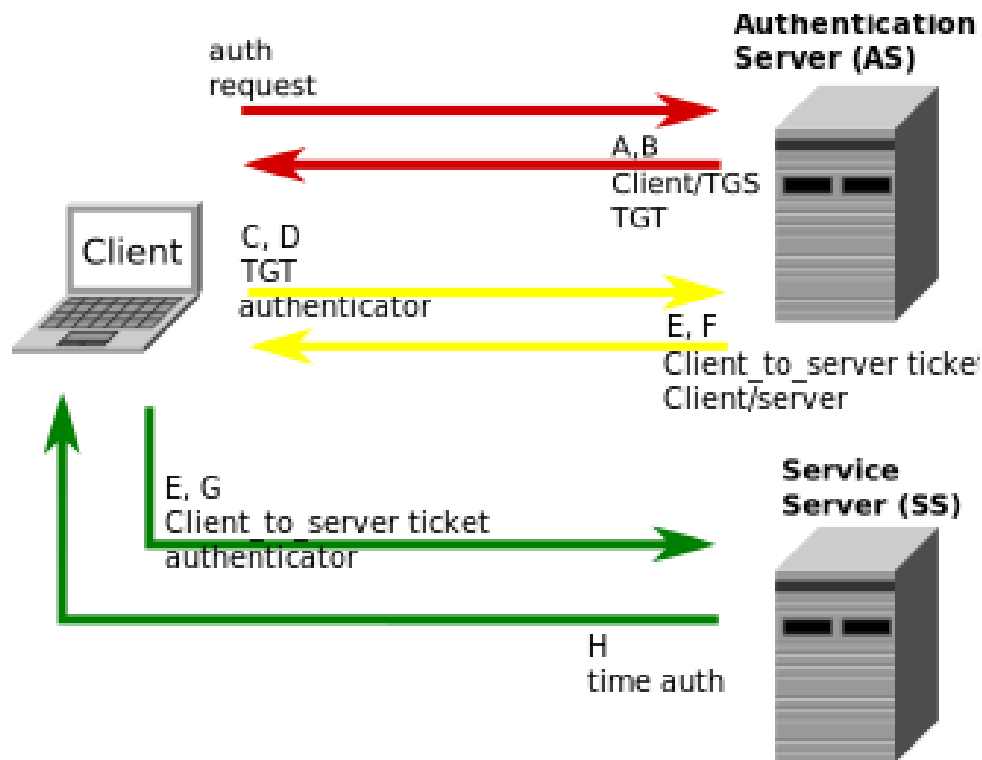
Hortonworks: Roles of Security Components

HDP	Centralized Security Administration w/ Ranger			
	Authentication	Authorization	Audit	Data Protection
	<ul style="list-style-type: none">• <i>Kerberos</i>• Perimeter security with <i>Apache Knox</i>	<ul style="list-style-type: none">• Fine grain access control with <i>Apache Ranger</i>	<ul style="list-style-type: none">• Centralized audit reporting w/ <i>Apache Ranger</i>	<ul style="list-style-type: none">• Wire encryption in Hadoop• HDFS Encryption w/ Ranger KMS

Kerberos



Kerberos



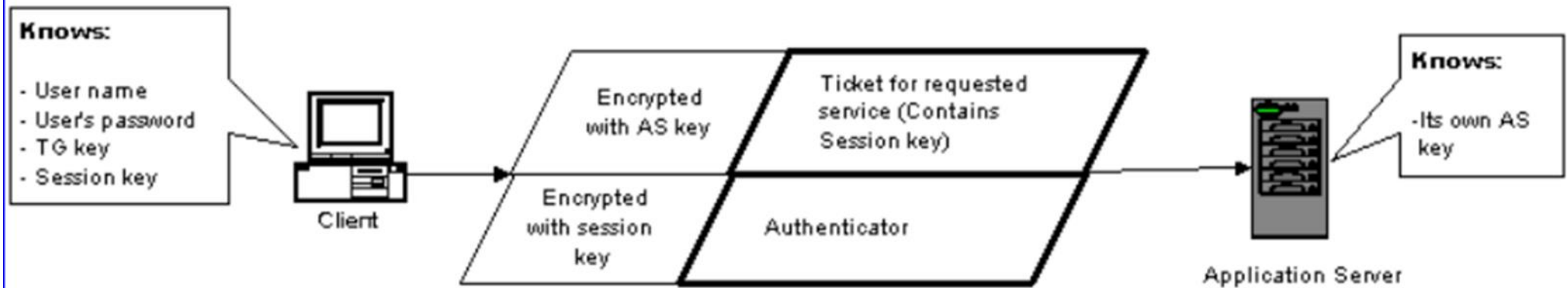
- Client requests an "Auth" server for a "ticket" to access a service
 - Sends: UserID, Service command
 - The Auth server knows all UserIDs passwords, and Service Keys
- Server checks UserID and service and:
 - Creates ticket containing service request encrypted using the service server key
 - Ticket 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_\(protocol\)](https://en.wikipedia.org/wiki/Kerberos_(protocol)) <http://web.stanford.edu/~torg/kerberos-overview.html>

Kerberos

Communication between the Client and the Application Server



Published in the late 80's as a network security protocol

[https://en.wikipedia.org/wiki/Kerberos_\(protocol\)](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

The screenshot displays the Apache Ranger web interface. At the top is a green navigation bar with the 'Ranger' logo, links for 'Access Manager', 'Audit', and 'Settings', and a user profile for 'admin'. Below this, a 'Service Manager' tab is selected. The main content area, titled 'Service Manager', contains a grid of service configuration cards. Each card represents a service category with a folder icon, a plus sign for expansion, and a list of instances. Each instance has a green checkmark icon and a red stop icon.

Service	Instance	Status
HDFS	Hadoop_Prod	Active
	Hadoop_Dev	Active
HBASE	HBase_Prod	Active
	HBase_Dev	Active
HIVE	Hive_Prod	Active
	Hive_Dev	Active
YARN	Yarn_Prod	Active
KNOX	Knox_Prod	Active
STORM	Storm_Dev	Active
SOLR	Solr_Dev	Active
KAFKA	Kafka_Dev	Active

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

Ranger Access Manager Audit Settings admin

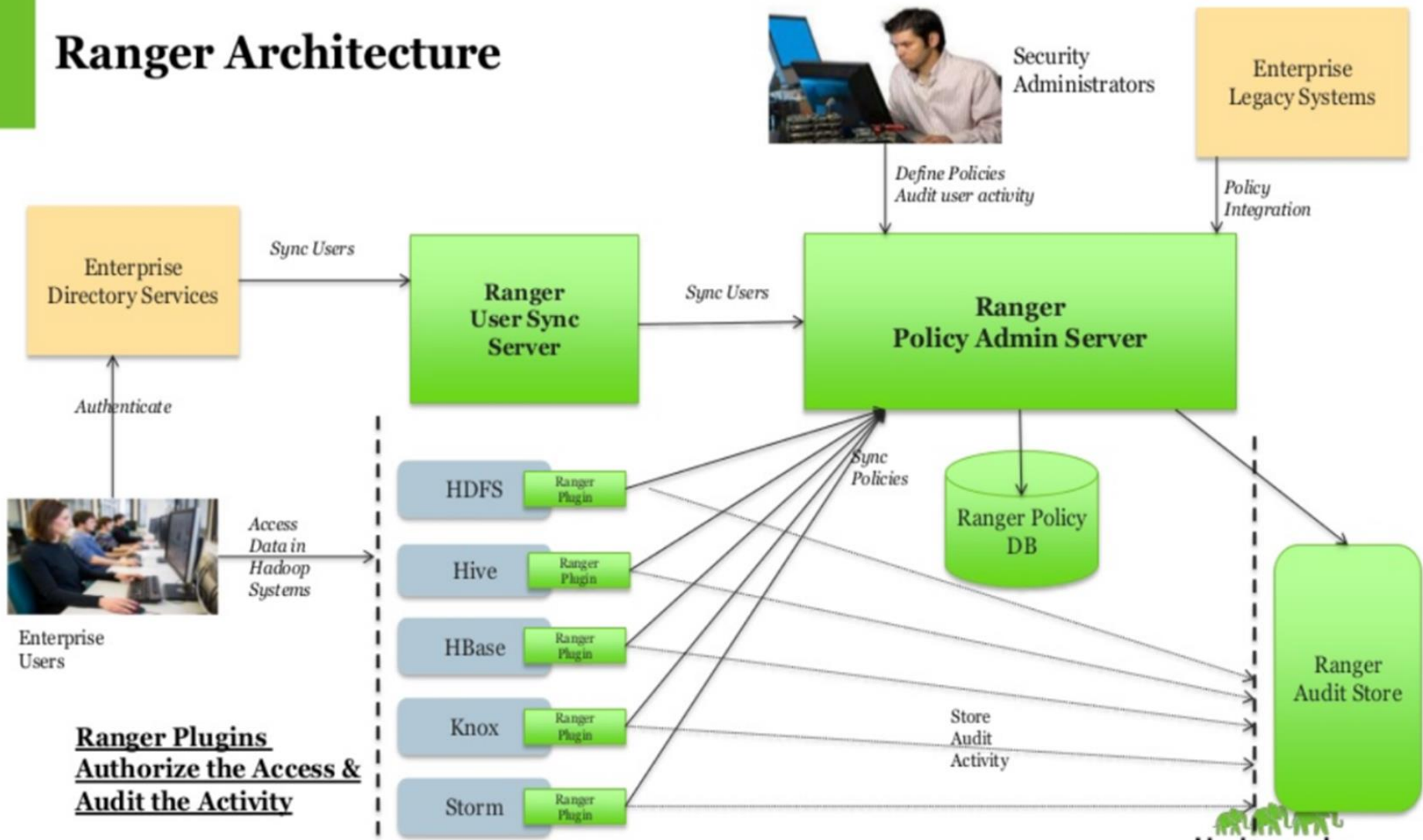
Service Manager > sandbox_hive Policies

List of Policies : sandbox_hive

Add New Policy

Policy ID	Policy Name	Status	Audit Logging	Groups	Users	Action
3	sandbox_hive-1-201508191258...	Enabled	Enabled		xapolicymgr	
4	sandbox_hive-2-201508191258...	Enabled	Enabled		xapolicymgr	
5	Hive Global Tables Allow	Enabled	Enabled	public		
6	Hive Global UDF Allow	Enabled	Enabled	public		
19	Call_Details_Table	Disabled	Enabled	IT Network		
20	Customer_Details_Table	Disabled	Enabled	Marketing		

Ranger Architecture



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

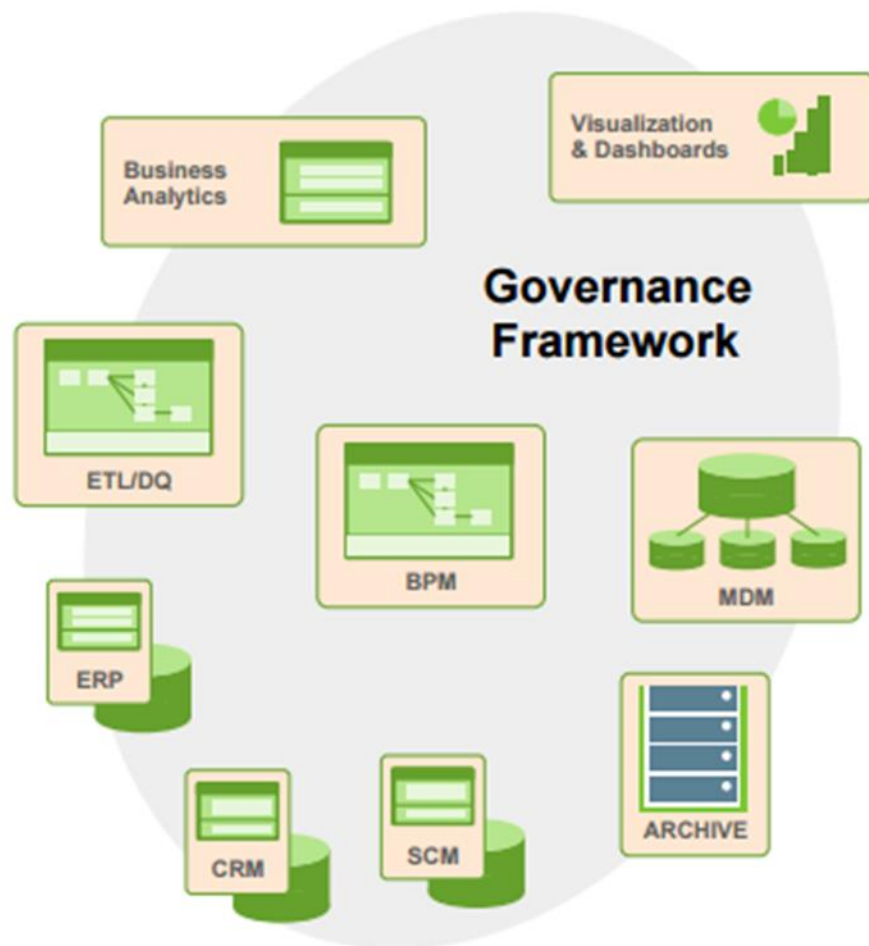
- Specify dynamic conditions in service definitions
Flexibility to define unique conditions by service (HDFS, Hive etc.)

Built on pluggable service-based model

- Custom plugins can be created for any data store

Apache Atlas

Apache Atlas: Enterprise Data Governance Goals

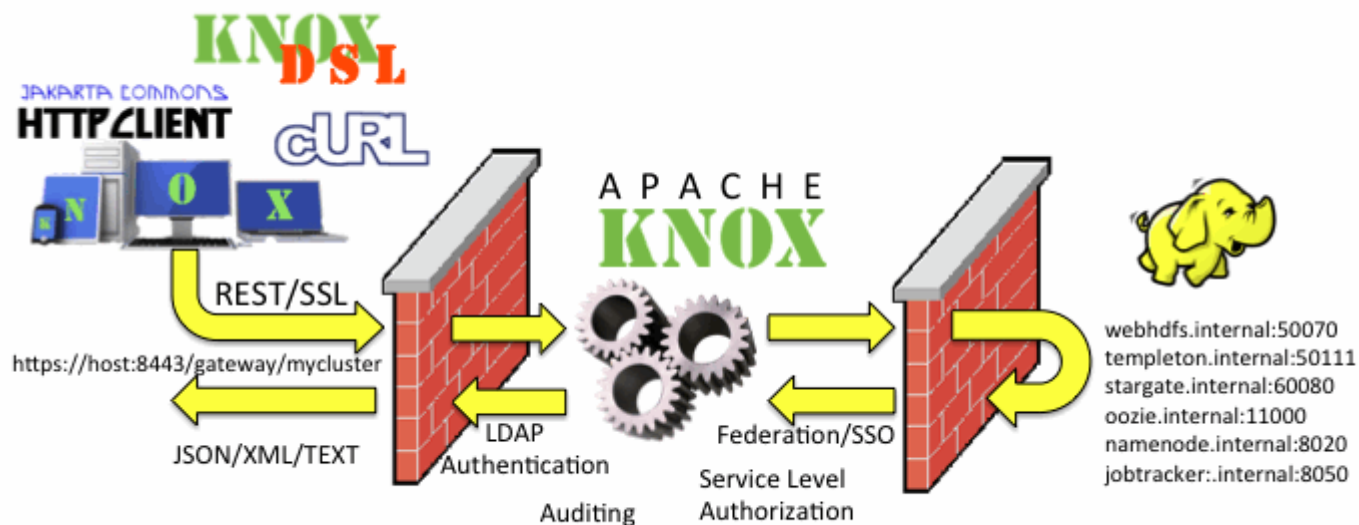


GOAL: Provide a common approach to data governance across all systems and data within the organization

- **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 must 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/>

