Leverage Security in Hadoop Without Sacrificing Usability

Syed Mahmood
Sr. Product Marketing Manager, Hortonworks

Vincent Lam
Director Product Marketing, Protegrity

Security in Open Enterprise Hadoop

protegrity

Hortonworks®

Security Challenges for a Hadoop Data Lake

Central Repository

of critical, sensitive data

Long-term Retention

of data stored for years or decades

Reliable Integration

always secure despite a fluctuating ecosystem

Dynamic Access

permits users to analyze data in new and different ways, always in flux

protegrity

Hortonworks®

Our Comprehensive Approach To Security

Administration	**Centrally manage consistent security**	How do I set policy across the entire cluster?
Authentication	**Prove the identity of systems and users**	Who are you and how can you prove it?
Authorization	**Provide secure access to data**	What can you do once you're authenticated?
Audit	**Maintain a record of data access events**	What did you do and when did you do it?
Data Protection	**Safeguard data at rest and in motion**	How can you encrypt the data?

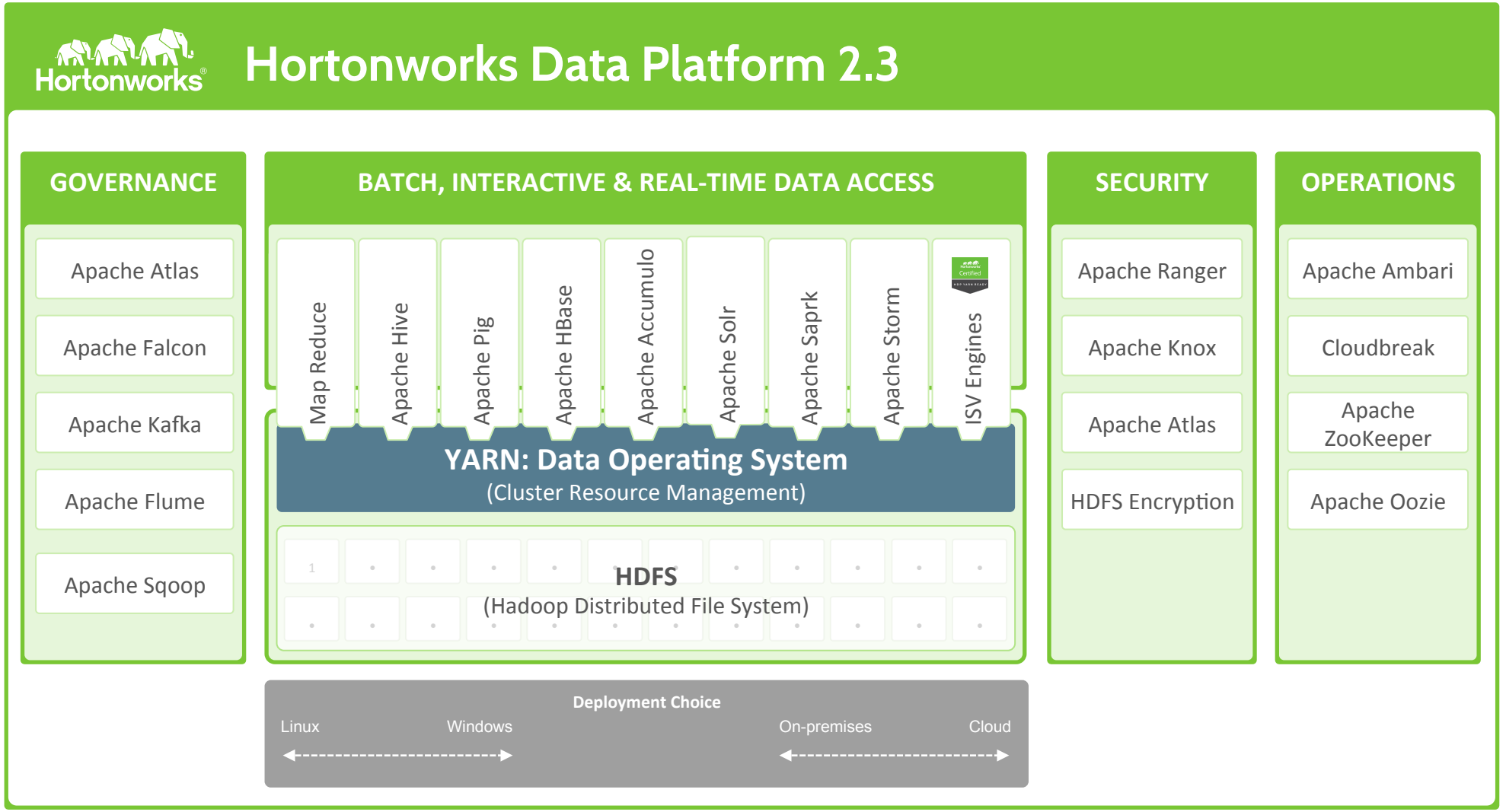
Our Comprehensive Approach To Security

Administration	
Centrally manage consistent security	Single administrative console to set policy across the entire cluster
Authentication	
Prove the identity of systems and users	Integrated with existing Active Directory and LDAP solutions, for perimeter and cluster
Authorization	
Provide secure access to data	Consistent authorization controls across all Apache™ components within HDP
Audit	
Maintain a record of data access events	Consistent, accessible record of data access events across all cluster components
Data Protection	
Safeguard data at rest and in motion	Encrypt data in motion and data at rest and refer to partner encryption solutions

Our Comprehensive Approach To Security

Administration	Centrally manage consistent security	APACHE RANGER
Authentication	Prove the identity of systems and users	KERBEROS & APACHE KNOX
Authorization	Provide secure access to data	APACHE RANGER
Audit	Maintain a record of data access events	APACHE RANGER & APACHE ATLAS
Data Protection	Safeguard data at rest and in motion	HDFS TDE with RANGER KMS

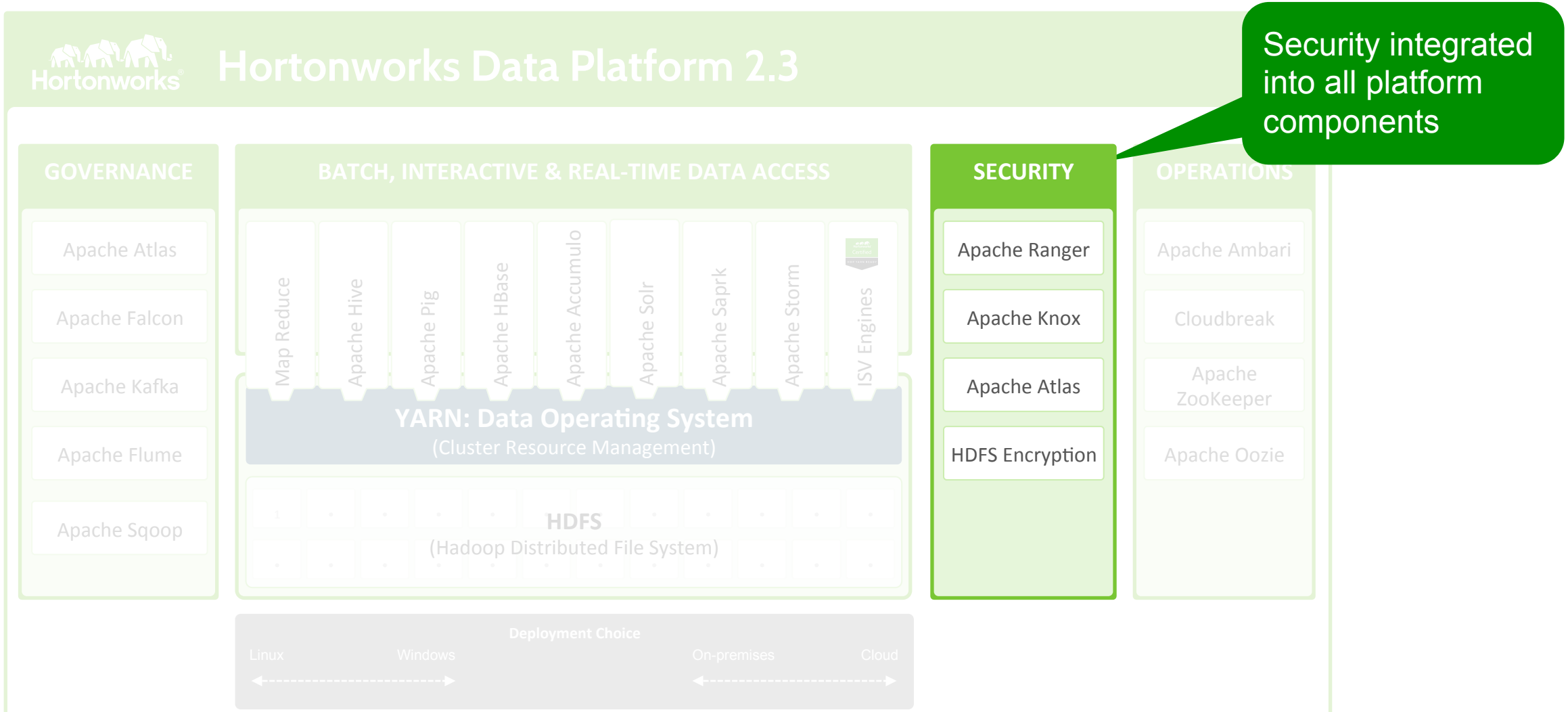
Integrated Platform Security



protegrity

Hortonworks

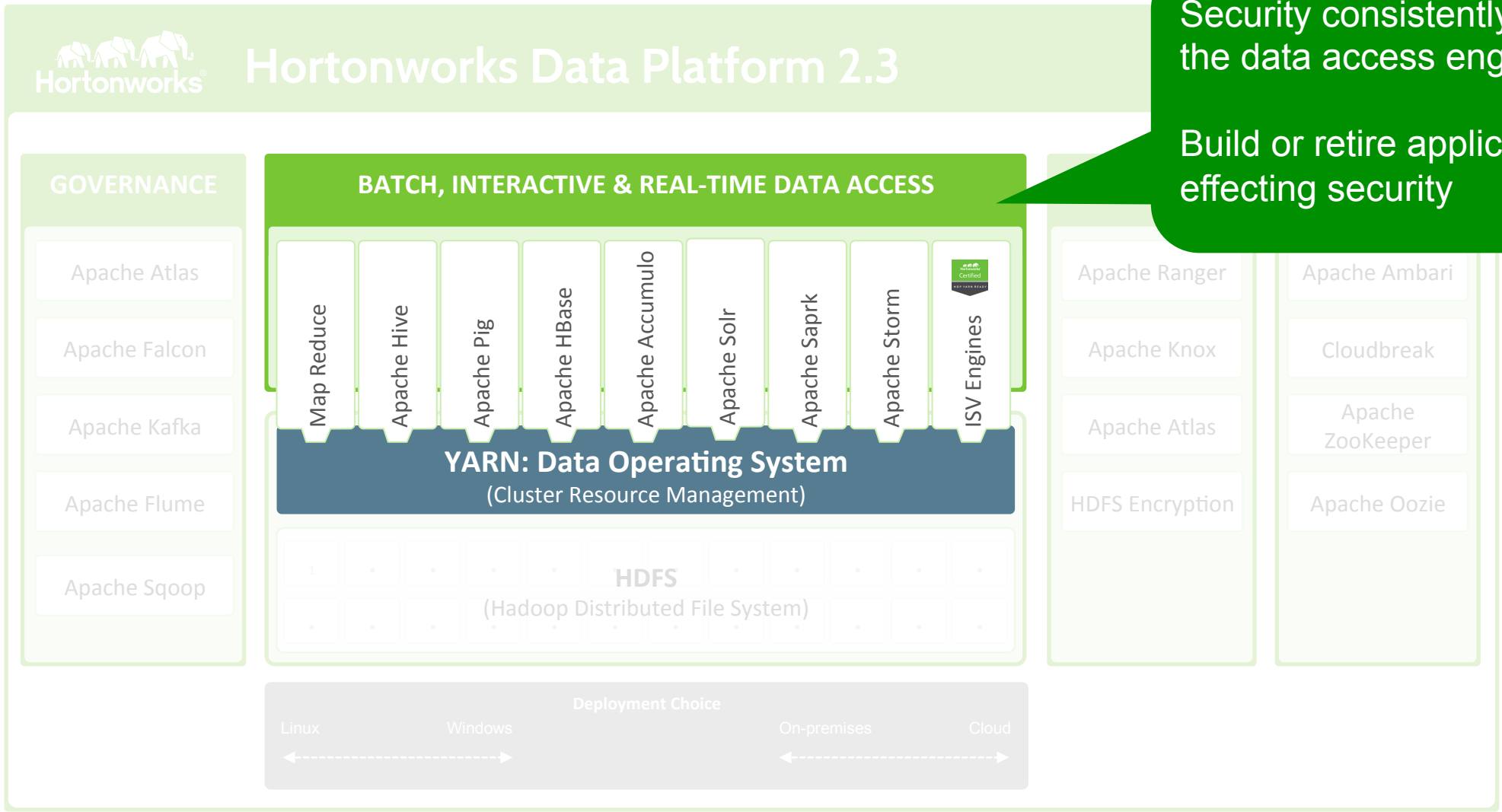
Integrated Platform Security



protegrity

Hortonworks

Integrated Platform Security



Security consistently applied across the data access engines

Build or retire applications without effecting security

protegrity

Hortonworks

Apache Ranger

Comprehensive security for Enterprise Hadoop

Ranger Centralizes Security for Deep Visibility

Apache Ranger

Centralized Platform

Consistently define, administer and manage security policies

Define a policy once and apply it to all the applicable components across the stack

Fine-grained Definitions

Administer security for:

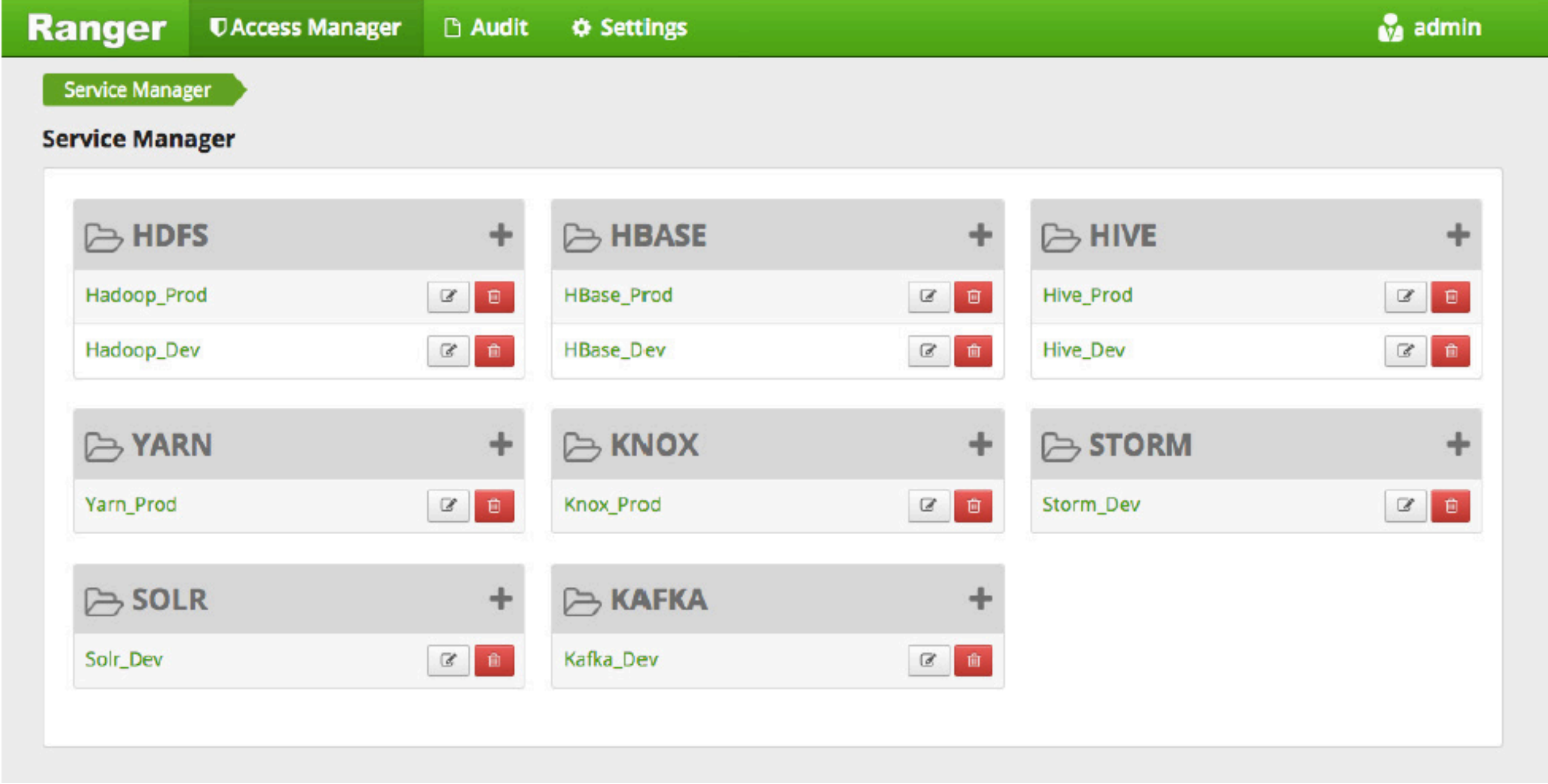
- Database
- Table
- Column
- LDAP Groups
- Specific Users

Deep Visibility

Administrators have complete visibility into the security administration process

protegrity

Hortonworks®


Ranger
Access Manager
Audit
Settings
admin
Service Manager
Service Manager
HDFS
Hadoop_Prod
Hadoop_Dev
HBASE
HBase_Prod
HBase_Dev
HIVE
Hive_Prod
Hive_Dev
YARN
Yarn_Prod
KNOX
Knox_Prod
STORM
Storm_Dev
SOLR
Solr_Dev
KAFKA
Kafka_Dev

Edit Policy

Policy Details :

Policy ID: 18

Policy Name *: Call_Details_Table — enabled

Hive Database *: xademo — include

table *: call_detail_records — include

Hive Column *: phone_number — exclude

Description:

Audit Logging: YES

User and Group Permissions :

Permissions

Select Group	Select User	Permissions	Delegate Admin
developer	Select User	select	

+

Service Manager > sandbox_hive Policies

List of Policies : sandbox_hive

Search for your policy...

Add New Policy

Policy ID	Policy Name	Status	Audit Logging	Groups	Users	Action
3	sandbox_hive-1-20150529142947	Enabled	Enabled	--	xapolicymgr	
4	Hive Global Tables Allow	Disabled	Enabled	public	--	
5	Hive Global UDF Allow	Disabled	Enabled	public	--	
18	Call_Details_Table	Enabled	Enabled	developer	--	
19	Customer_Details_Table	Disabled	Enabled	Marketing	--	
20	Hive Demo Table Loader	Enabled	Enabled	--	hive	
21	Hive Demo UDF Loader	Enabled	Enabled	--	hive	
29	admin policy	Enabled	Enabled	--	admin	

Apache Knox

A single point of secure access for Hadoop clusters

Apache Knox Provides API Security

Single Access Point

Kerberos encapsulation

REST API hierarchy

Consolidated API calls

Multi-cluster support

Central Controls

Eliminates SSH “edge node”

Central API management

Central audit control

Service level authorization

Integrated with Existing Systems

SSO integration – Siteminder and OAM

LDAP and Active Directory integration

Enhanced Security Capabilities in HDP 2.3

NEW

Administration

Centrally manage consistent security

- Authorization and audit for Solr, Kafka and YARN
- Support for custom plugins via Ranger and Knox

Authentication

Prove the identity of systems and users

- Bi-directional SSL between clients and servers
- LDAP data caching

Authorization

Provide secure access to data

- Auth for Kafka, Solr and multi-tenant YARN queues
- Hooks for dynamic policy rules (e.g. by geo-location)

Audit

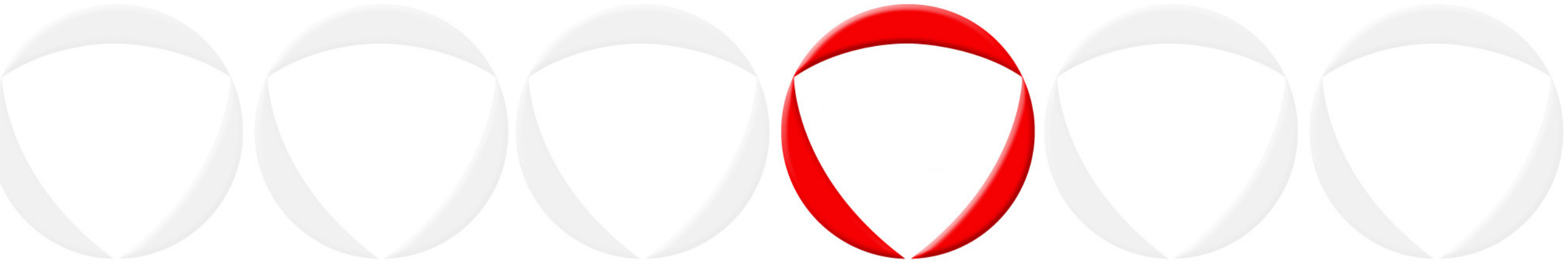
Maintain a record of data access events

- Audit storage in Solr
- Audit optimization for high volume applications

Data Protection

Safeguard data at rest and in motion

- HDFS transparent encryption for data at rest
- Robust, highly available key management store



Leverage Security in Hadoop Without Sacrificing Usability

Vincent Lam

Director, Product Marketing

protegrity

protecting your data.
protecting your business.

protegrity

Hortonworks®

Data Security for the Enterprise

- Fortune 1000 customers
- Global Presence across every industry
- Helping customers increase customer trust while eliminating risk

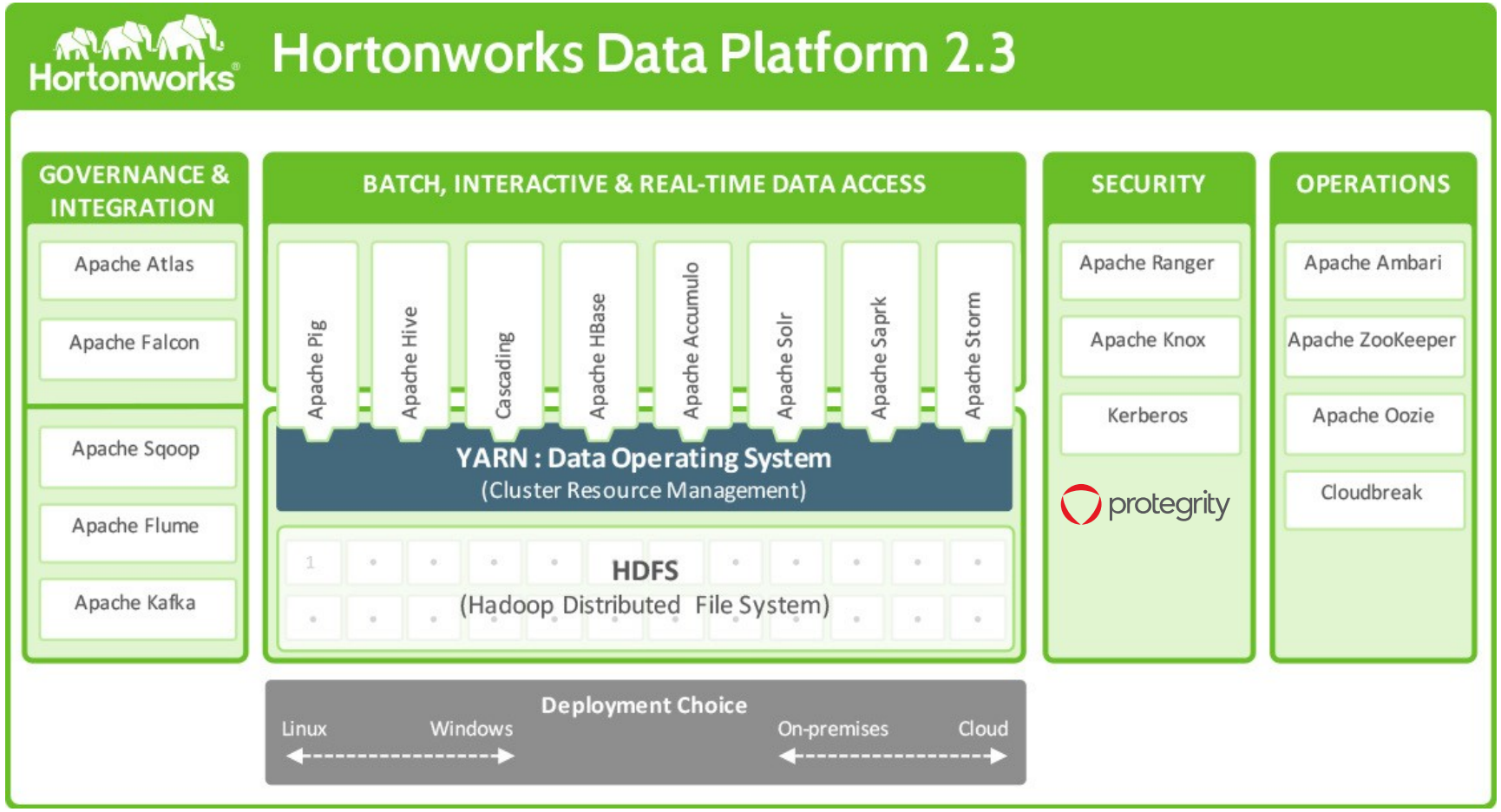
Scalable, data-centric protection of data at-rest, in-transit and in-use, helping businesses secure sensitive or regulated information while maintaining data usability

- Vaultless Tokenization & Encryption
- Centralized Policy, Access Control, Key Management & Reporting
- Heterogeneous, Cross-platform Data Protection
- Rapid Data Security Innovation

protegrity

Hortonworks®

Built to Enhance and Complement the HDP Ecosystem



protegrity

Hortonworks

Improving Security...

Photo: www.mirror.uk.co

While Maintaining Usability!

protegrity

Hortonworks®

Protegrity Enhances the HDP Pillars without Sacrificing Usability

Administration

- Separation of Duties
- Centralized for Hadoop + Enterprise

Authorization

- Enhanced Policy Conditions
- Roles Across Hadoop + Enterprise

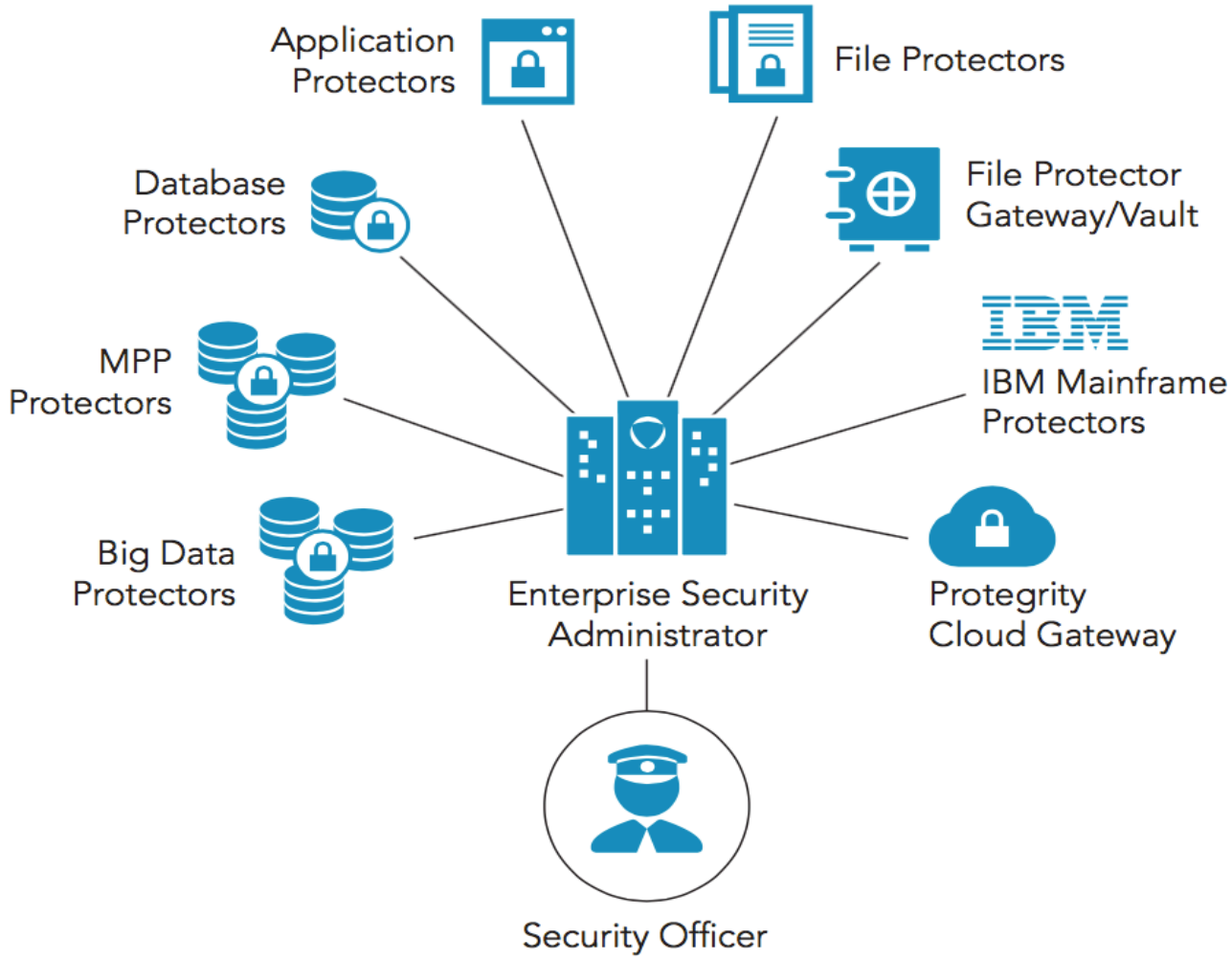
Data Protection

- Fine Grained Protection
- Tokenization
- Encryption
- Masking
- Record Level Protection

protegrity

Hortonworks®

Central Management, Policy Deployment – Hadoop + Beyond



protegrity

Hortonworks®

THE ONLY WAY TO SECURE SENSITIVE DATA IS TO PROTECT THE DATA ITSELF

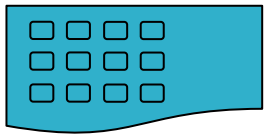
AT REST, IN TRANSIT, IN USE

Granularity of Protecting Sensitive Data

Coarse Grained Protection
(File/Volume)

- Methods: File or Volume encryption
- “All or nothing” approach
- Does NOT secure file contents in use
- OS File System Encryption
- HDFS Encryption
- Secures data at rest and in transit

Fine Grained Protection
(Data/Field)



- At the individual field level
- Fine Grained Protection Methods:
  - Vaultless Tokenization
  - Encryption
  - Format Preserving Encryption
  - Masking/Data Obfuscation
- Data is protected wherever it goes (even In-Use)
- Business intelligence analytics capability retained
  (80%-90% of analytics performed on data in protected form)

protegrity

Hortonworks®

Different Ways to Provide Data-Centric Protection

Encryption

Data converted to binary Ciphertext using mathematical algorithm. Can be one-way (Hash) or reversible (Symmetric/Asymmetric).

Tokenization

Real data is replaced with randomly generated characters of same data type.

Masking

Obfuscates data either statically or dynamically – not reversible

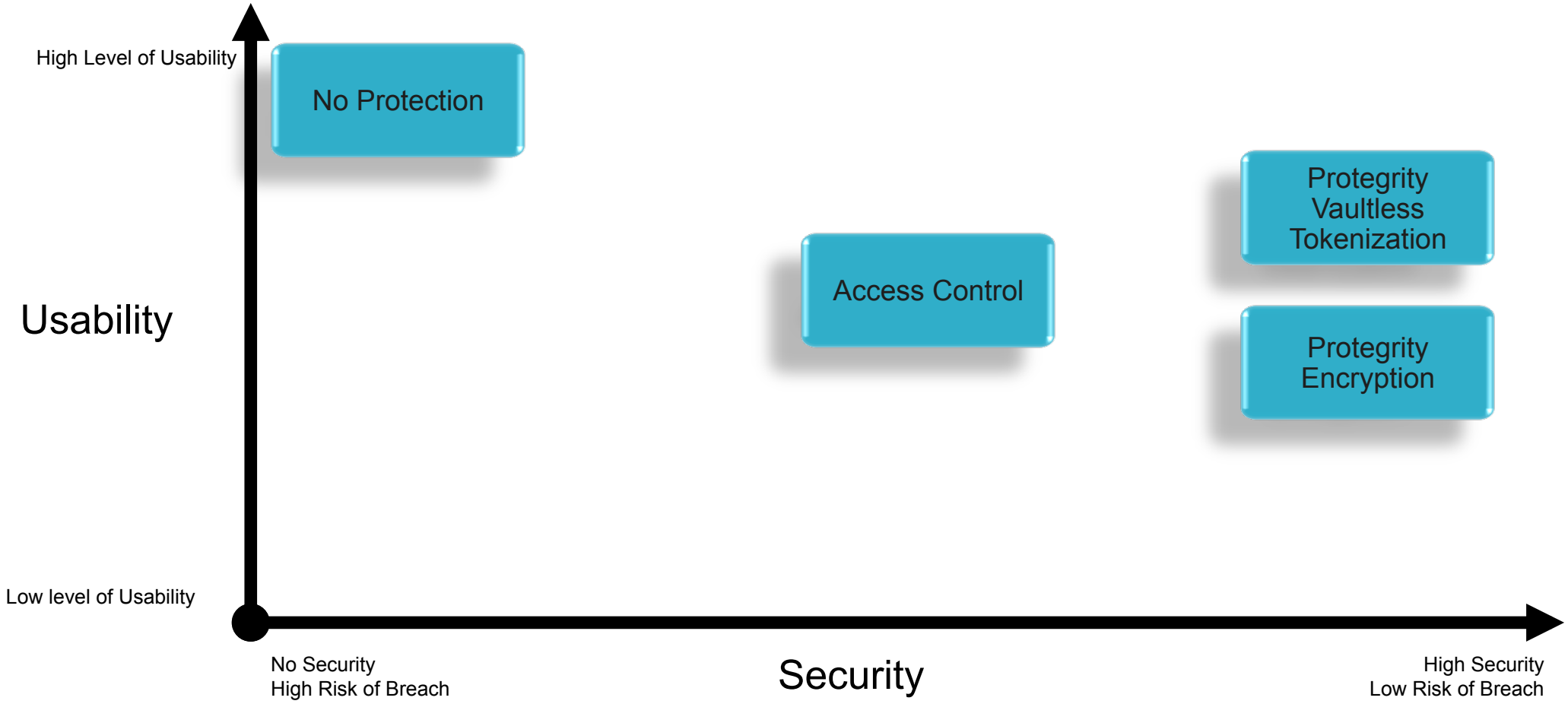
De-Identification or “Anonymization”

Enough data fields are “protected” to sufficiently de-identify or anonymize records

Format Preserving Encryption (FPE or DTP)

Some benefits of both encryption & tokenization at a significant performance cost

Fine Grained Security and Data Usability



protegrity

Hortonworks®

Did You Know?

You Can Analyze Protected Data

De-identification or Anonymization of data allows analytics on protected data across a variety of applications and analytics platforms

About 1%-3% Data is Sensitive

In-Database, Column-Level Encryption or Tokenization are typically applied to these 1% to 3% of the data

Most Analysts Don’t Need The Sensitive Data

80% to 90% of analytics can be performed on “protected” data

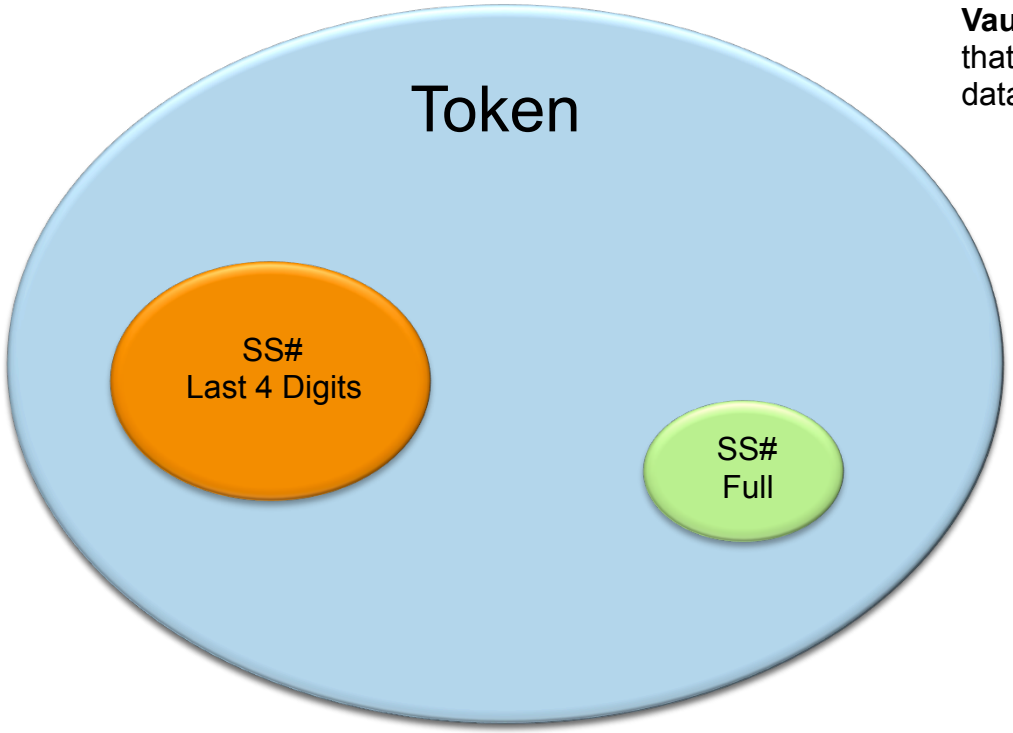
Tokenization Means a Lot Less Overhead

Only 10% to 20% of queries and reports accessing 1% to 3% of the data will introduce any additional de-protection “work”

Data-Centric Security Reduces Exposure and Risk

Vaultless Tokenization is a form of data protection that converts sensitive data into fake data. The real data can be retrieved by authorized users.

Population of users who can perform their job function with only the last 4 digits of the SSN



Population of users who need access to the full SSN to perform their job function

Improve Your Security Posture

protegrity

Hortonworks®

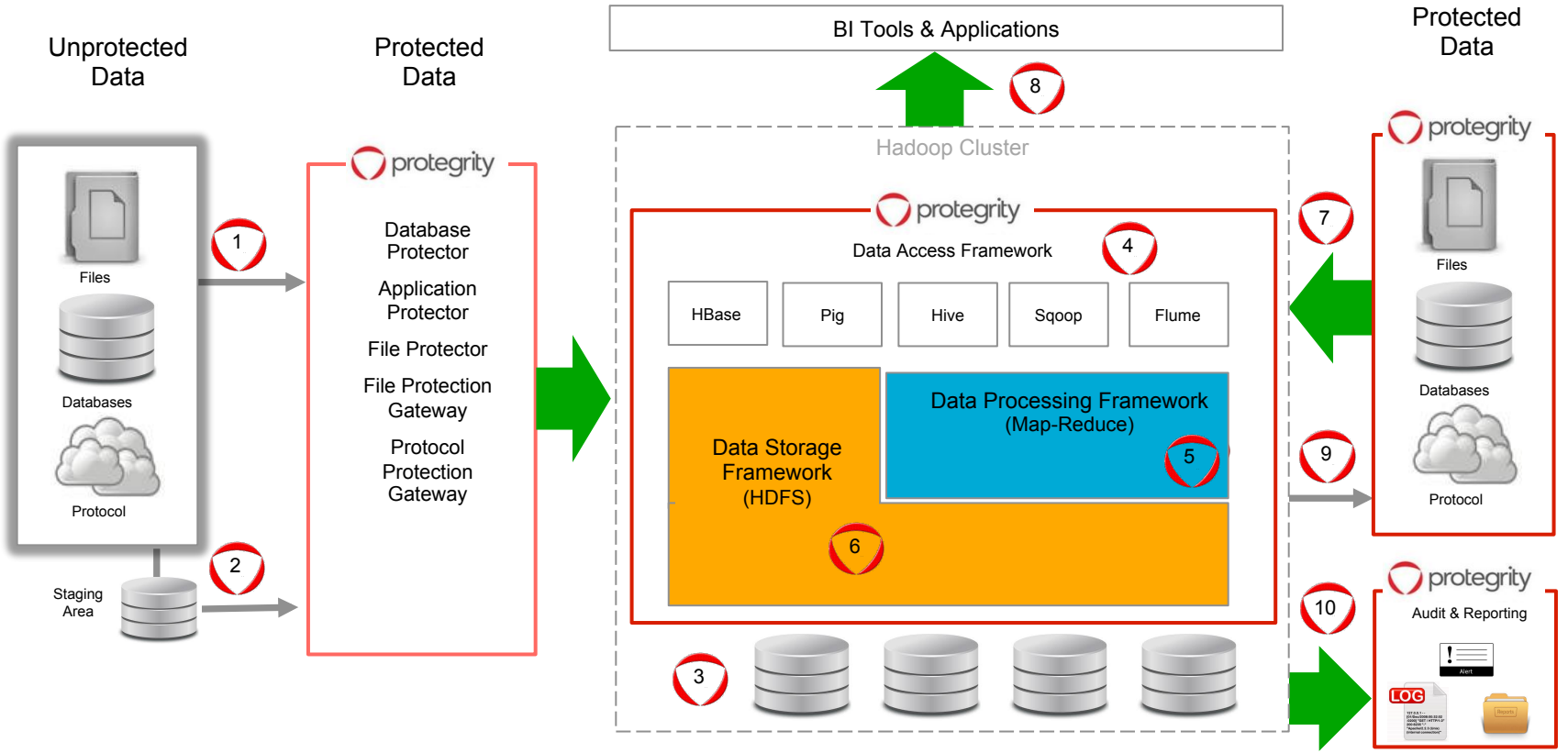
Fine Grained Protection

Identifier	Clear	Protected	Authorized Role 1 * Can see most data in the clear	Authorized Role 2 * Can see limited data in the clear
Name	Joe Smith	csu wusoj	Joe Smith	Joe Smith
Address	100 Main Street, Pleasantville, CA	476 srta coetse, cysieondusbak, HA	100 Main Street, Pleasantville, CA	“No Access”
Date of Birth	12/25/1966	01/02/1966	12/25/1966	01/02/1966
Social Security Number	076-39-2778	478-39-8920	076-39-2778	xxxxx2278
Credit Card Number	3678 2289 3907 3378	3846 2290 3371 3890	xxxx xxxx xxxx 3378	3846 2290 3371 3890
E-mail Address	joe.smith@surferdude.org	eoe.nwuer@beusorpdqo.aku	joe.smith@surferdude.org	joe.smith@surferdude.org
Telephone Number	760-278-3389	998-389-2289	760-278-3389	998-389-2289

protegrity

Hortonworks®

Protection Points Across HDP



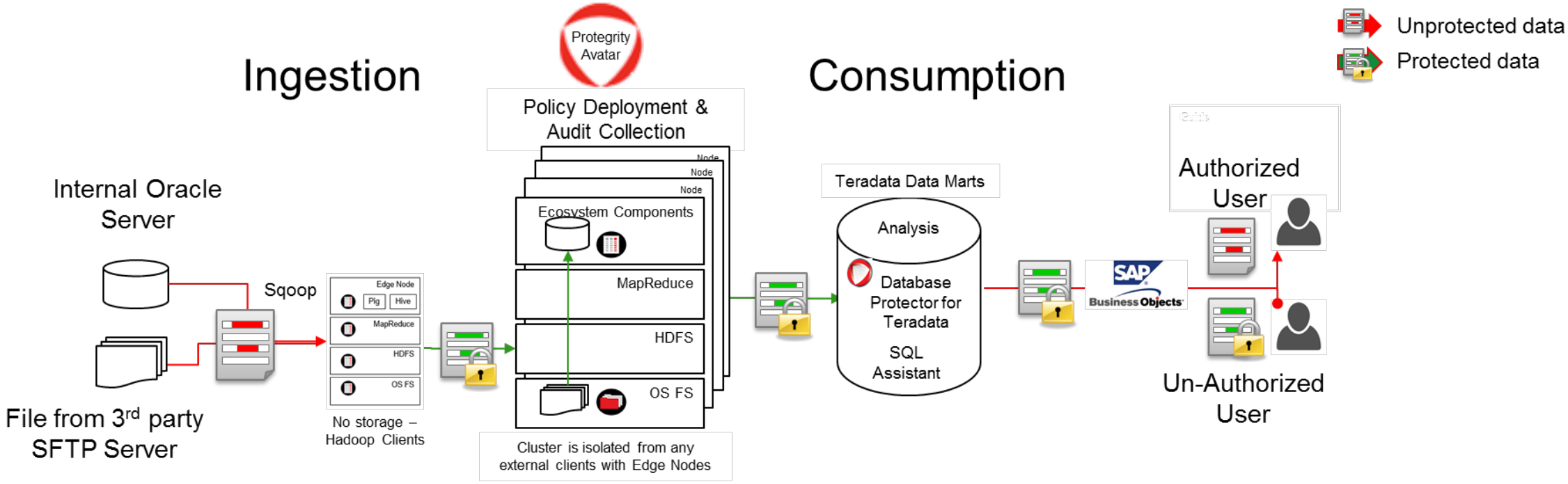
protegrity

Hortonworks®

Use Case: Large Telecommunications Company

Large Telco using Teradata with Hadoop for customer data analytics.

- Understand the consumer and usage patterns to provide best-in-class customer service without compromising privacy of the individual
- Consistent protection across heterogeneous platforms, supporting analytics in Business Objects and Teradata ad-hoc queries



protegrity

Hortonworks®

Thank You

Try Out Protegrity Avatar for Hortonworks

http://www.protegrity.com/products-services/protegrity-avatar-for-hortonworks

Visit Hortonworks (booth 409) and Protegrity (booth 333) at Strata

New York – Sep 29th – Oct 1st

Questions?

Vincent <Vincent.Lam@Protegrity.com>

Syed <smahmood@hortonworks.com>