



New York Oracle Users Group, Inc.

Bridging the Gap Between Privacy and Data Insight

Ulf Mattsson
CTO, Protegrity

ulf . mattsson [at] protegrity . com



Bridging the Gap Between Privacy and Data Insight



Ulf Mattsson, CTO Protegrity

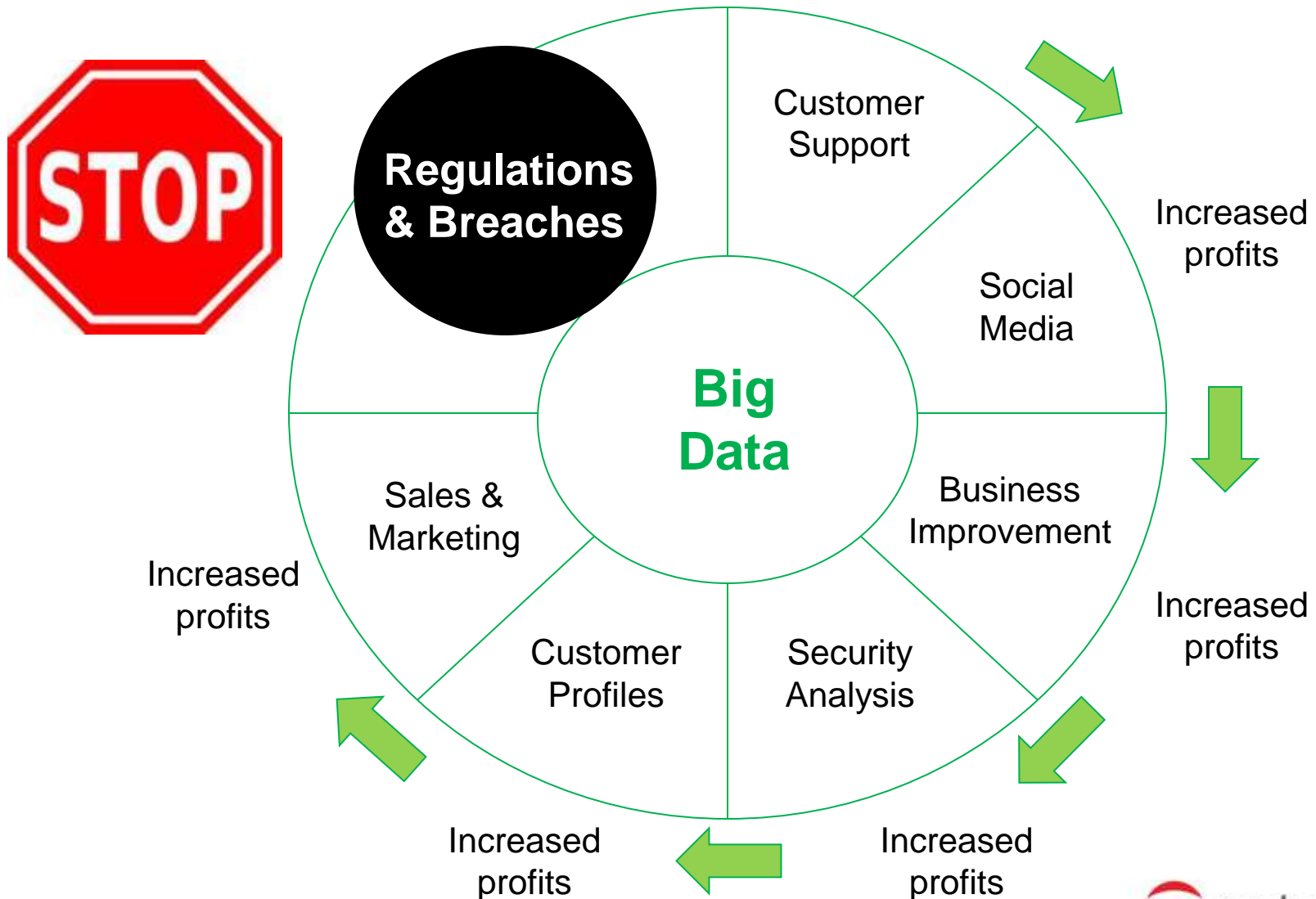
- 20 years with IBM Research & Development & Global Services
- Started Protegrity in 1994 (Data Security)
- Inventor of 20+ patents
 - Encryption, Tokenization & Intrusion Prevention
- Member of
 - PCI Security Standards Council (PCI SSC)
 - American National Standards Institute (ANSI) X9
 - Encryption & Tokenization
 - International Federation for Information Processing
 - IFIP WG 11.3 Data and Application Security
 - NYOUG, ISACA , ISSA and Cloud Security Alliance (CSA)



Agenda

- HIPAA, PCI DSS & Privacy Laws
- Oracle's Big Data Platform
- Big Data and Threats
 - Ways to Hack Big Data
- What's the Problem with Securing Big Data?
 - Hadoop Beyond Kerberos
 - New Protection Techniques
 - Speed of Different Protection Methods
- Risk Adjusted Data Protection
 - De-Identifying Sensitive Data
 - Research Studies
- A Data Protection Methodology
 - Best Practices for Protecting Big Data

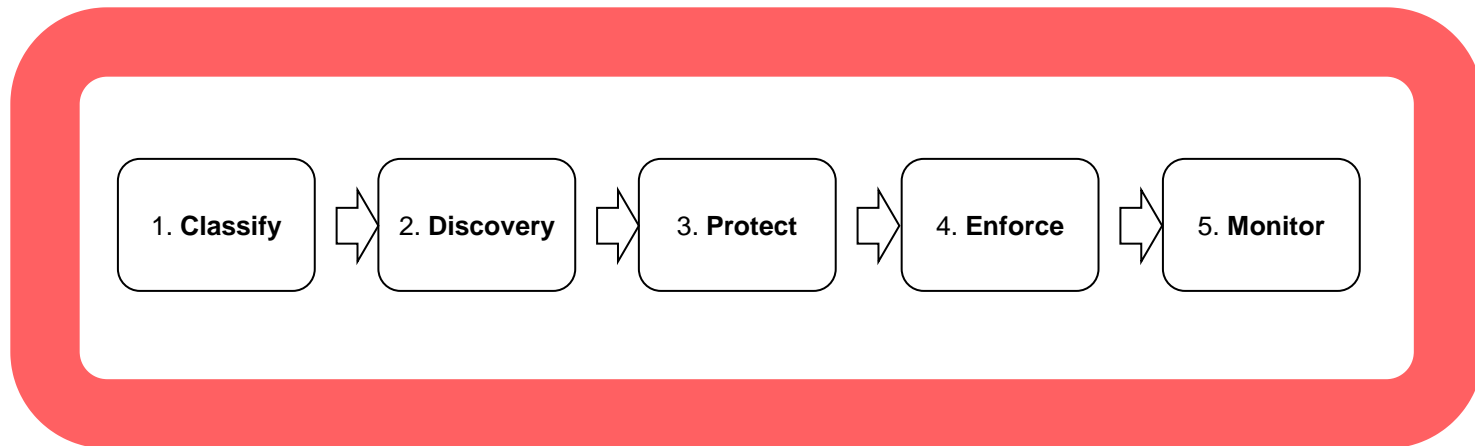
Balancing security and data insight



Balancing security and data insight

- Tug of war between security and data insight
- Big Data is designed for access, not security
- Privacy regulations require de-identification which creates problems with privileged users in an access control security model
- Only way to truly protect data is to provide data-level protection
- Traditional means of security don't offer granular protection that allows for seamless data use

Five Point Data Protection Methodology



1

Classify

Determine what data is sensitive to your organization.

1. Classify

Data is classified as sensitive and must be protected in response to Laws and regulations such as PCI DSS, HIPAA, State Privacy Laws and others.

Companies may also have Intellectual Property and other Company Secrets that must be secured against the competition.

All companies have sensitive data about their Customers and about their Employees.

1. Classify: Examples of Sensitive Data

Sensitive Information	Compliance Regulation / Laws
Credit Card Numbers	PCI DSS
Names	HIPAA, State Privacy Laws
Address	HIPAA, State Privacy Laws
Dates	HIPAA, State Privacy Laws
Phone Numbers	HIPAA, State Privacy Laws
Personal ID Numbers	HIPAA, State Privacy Laws
Personally owned property numbers	HIPAA, State Privacy Laws
Personal Characteristics	HIPAA, State Privacy Laws
Asset Information	HIPAA, State Privacy Laws

Breach notification laws are often the catalyst for a deep audit in companies resulting in large fines.

HIPAA PHI: List of 18 Identifiers

1. Names
2. All geographical subdivisions smaller than a State
3. All elements of dates (except year) related to individual
4. Phone numbers
5. Fax numbers
6. Electronic mail addresses
7. Social Security numbers
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers
13. Device identifiers and serial numbers
14. Web Universal Resource Locators (URLs)
15. Internet Protocol (IP) address numbers
16. Biometric identifiers, including finger prints
17. Full face photographic images
18. Any other unique identifying number

PCI DSS (Payment Card Industry Data Security Standard)

Build and maintain a secure network.	<ol style="list-style-type: none">1. Install and maintain a firewall configuration to protect data2. Do not use vendor-supplied defaults for system passwords and other security parameters
Protect cardholder data.	<ol style="list-style-type: none">3. Protect stored data4. Encrypt transmission of cardholder data and sensitive information across public networks
Maintain a vulnerability management program.	<ol style="list-style-type: none">5. Use and regularly update anti-virus software6. Develop and maintain secure systems and applications
Implement strong access control measures.	<ol style="list-style-type: none">7. Restrict access to data by business need-to-know8. Assign a unique ID to each person with computer access9. Restrict physical access to cardholder data
Regularly monitor and test networks.	<ol style="list-style-type: none">10. Track and monitor all access to network resources and cardholder data11. Regularly test security systems and processes
Maintain an information security policy.	<ol style="list-style-type: none">12. Maintain a policy that addresses information security

2

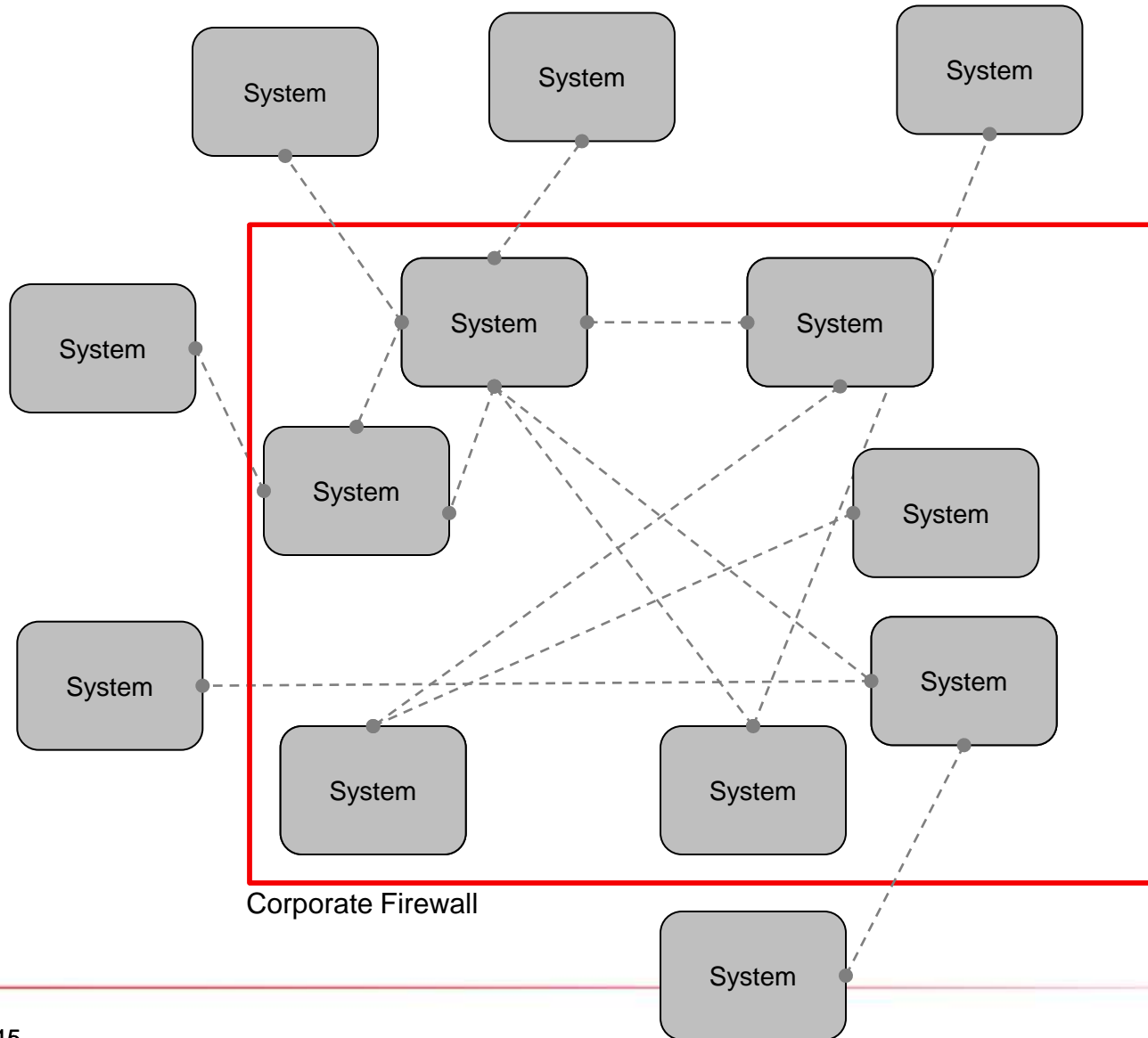
Discovery

Discover where the sensitive data is located, how it flows, who can access it, the performance requirements and other requirements for protection.

2. Discovery

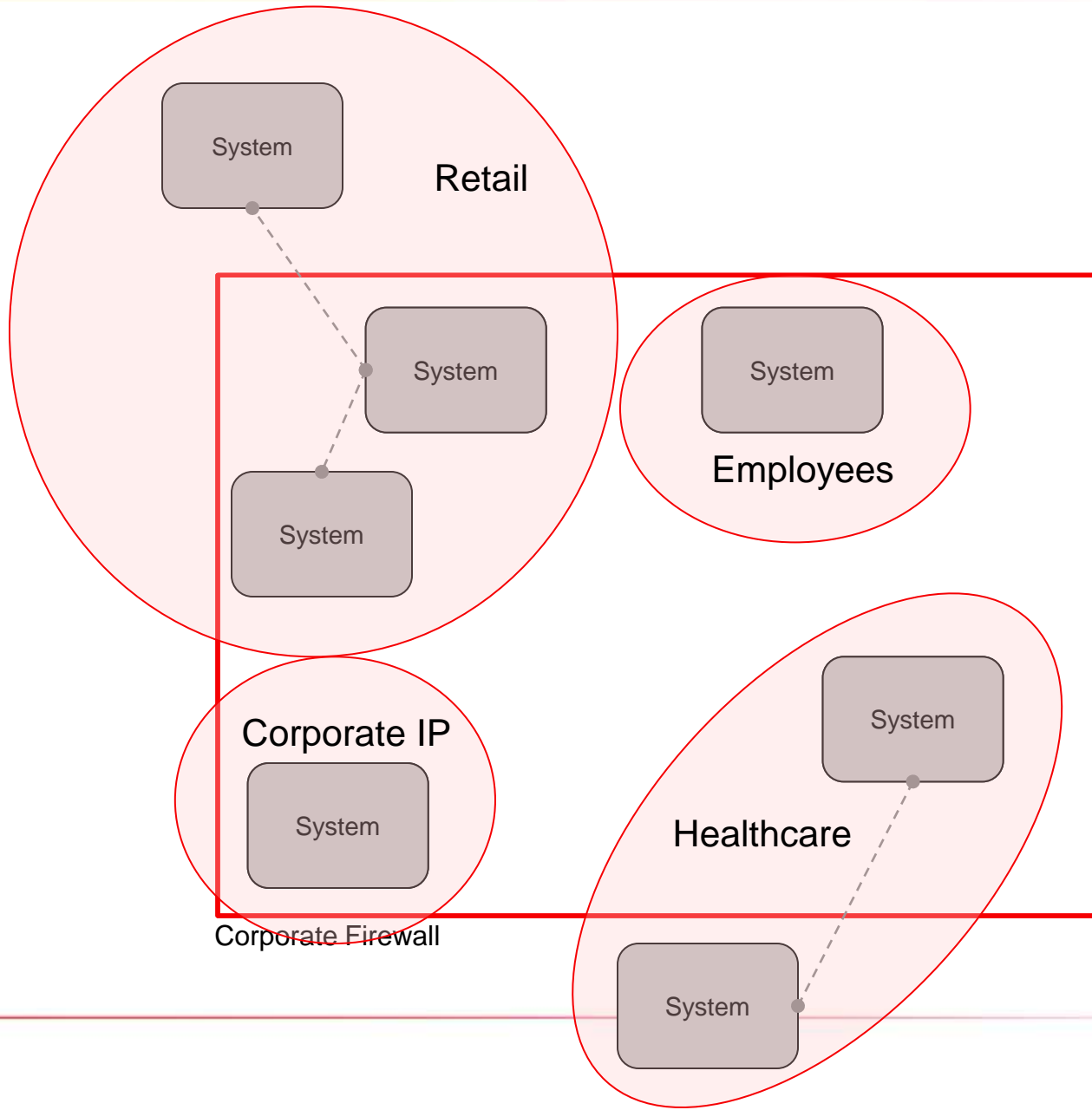
- The Discovery process peers into the enterprise to find sensitive data in preparation for delivering an optimal protection solution.
 - Existing Sensitive Data
 - New Sensitive Data
 - Archived Data

2. Discovery in a large enterprise with many systems

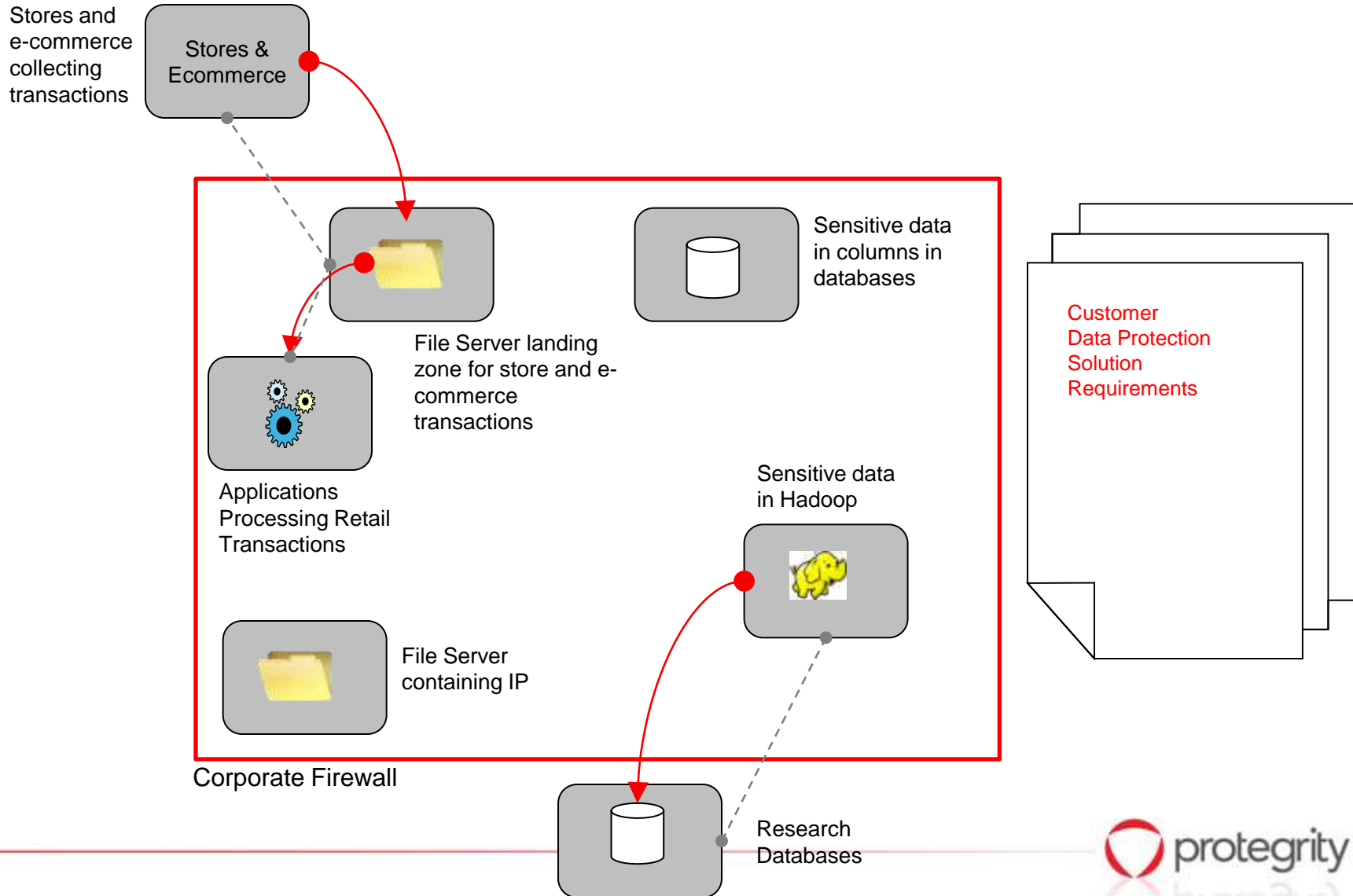


Focus on systems that contain sensitive data

2. Discovery: Determine the context to the Business



2. Discover: Context to the Business and to Security



3

Protect

Protect the sensitive data at rest and in transit.

Google fired engineer for privacy breach

David Barksdale, a Google engineer, was sacked earlier this year for improperly accessing the accounts of several Google users, Google confirms.

by Tom Krazit | September 14, 2010 5:27 PM PDT

Google confirmed on Tuesday that it fired an employee earlier this year for violating its policies on accessing the accounts of its users.

Earlier in the day, [Gawker](#) reported that David Barksdale, an engineer in Google's Seattle offices, used his position as a key engineer evaluating the health of Google's services to break into the Gmail and Google Voice accounts of several children. After parents of the children complained to Google, Gawker said Barksdale--who was not accused of anything with sexual overtones--was dismissed, and Google confirmed that move late Tuesday.



"We dismissed David Barksdale for breaking Google's strict internal privacy policies. We carefully control the number of employees who have access to our systems, and we regularly

Big Data and The Insider Threat

Google fired engineer for privacy breach

David Barksdale, Google engineer, improperly accessed users' data, Google confirms.

by Tom Krazit | September 1, 2013

Google confirmed on Tuesday that it fired an engineer this year for violating its privacy policies and exposing its users' data.

Earlier in the day, Gawker reported that a Google engineer evaluating the security of a new feature for the Gmail and Google Voice services had accessed the parents of the children of a former Google employee, David Barksdale—who was not a Google employee. Google confirmed that the engineer had accessed the data.

"We dismissed David Barksdale for violating our privacy policies. We carefully control the number of people who have access to our systems, and we regularly



... earlier this year for violating its privacy policies, Google confirmed.



... was dismissed, and

... privacy policies. We carefully control the number of people who have access to our systems, and we regularly

Privacy Laws (See List in Appendix)

○ 54 International Privacy Laws

○ 30 United States Privacy Laws, including

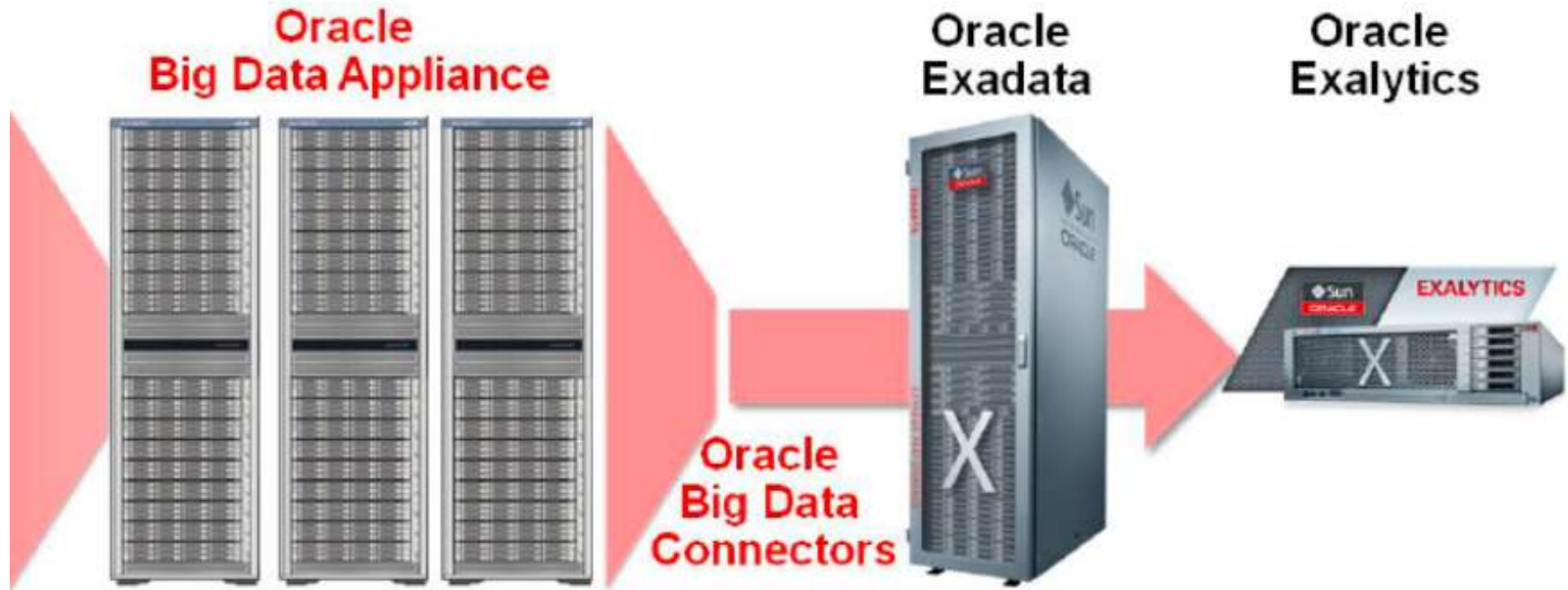
- **Financial Services** - Gramm-Leach-Bliley Act (GLBA), Sarbanes-Oxley Act (SARBOX), USA PATRIOT ACT, PCI Data Security Standard, and the Basel II Accord (EU)
- **Healthcare and Pharmaceuticals** - HIPAA (Health Insurance Portability and Accountability Act of 1996) and FDA 21 CFR Part 11
- **Infrastructure and Energy** - Guidelines for FERC and NERC Cybersecurity Standards, the Chemical Sector Cyber Security Program and Customs-Trade Partnership Against Terrorism (C-TPAT)
- **Federal Government** - Compliance with FISMA and related NSA Guidelines and NIST Standards

Woman gets Prison Time in Identity Theft

AP By ROXANA HEGEMAN | Associated Press – Mon, Mar 25, 2013

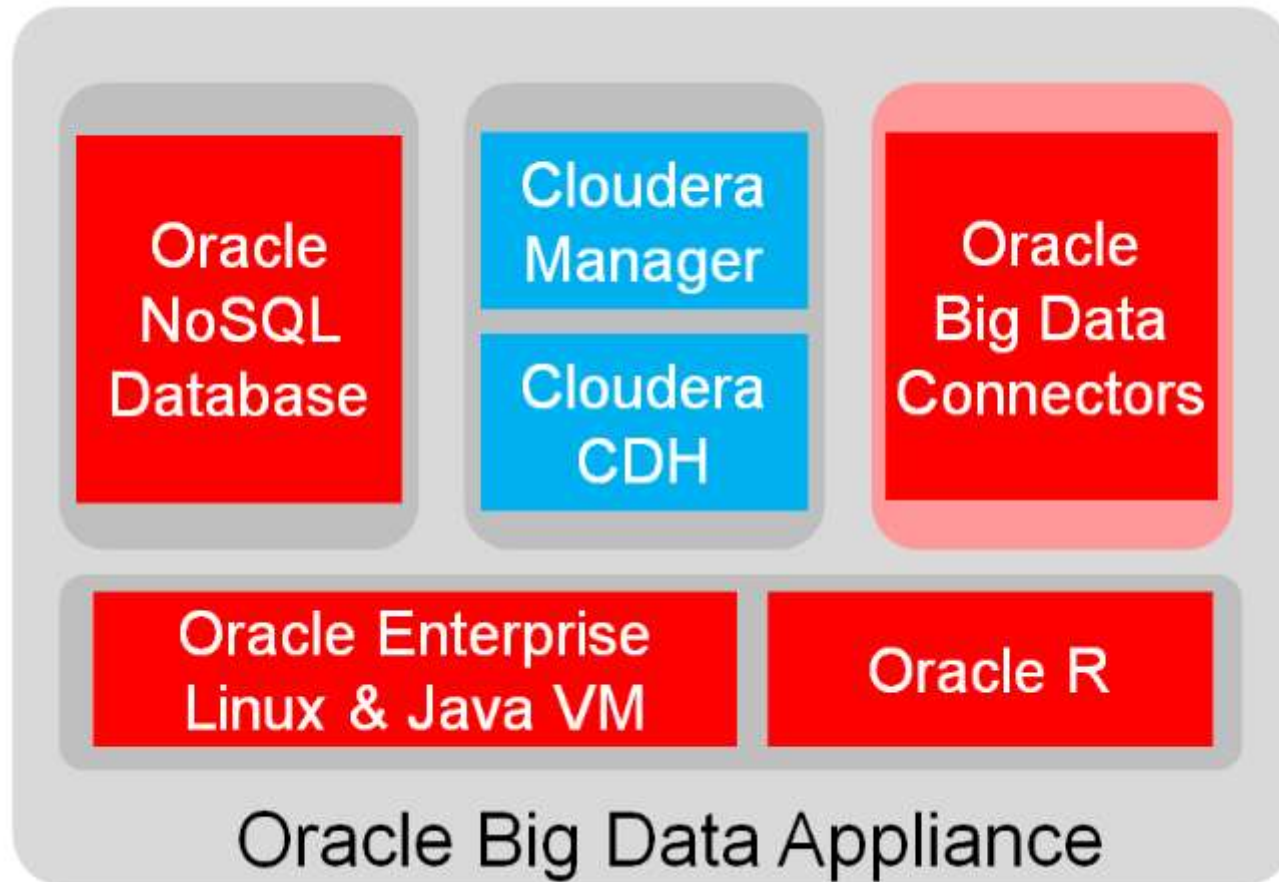


Oracle's Big Data Platform



Source: <http://www.oracle.com/us/technologies/big-data/index.html>

Software on Oracle Big Data Appliance

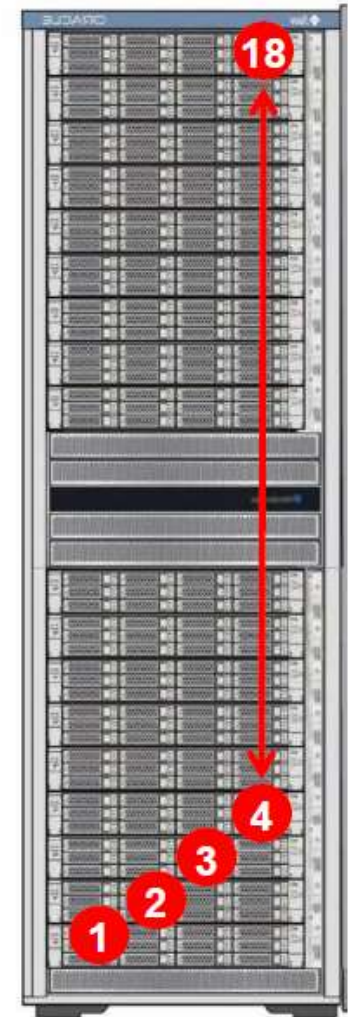


Source: <http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf>

Software Layout - Oracle Big Data Appliance

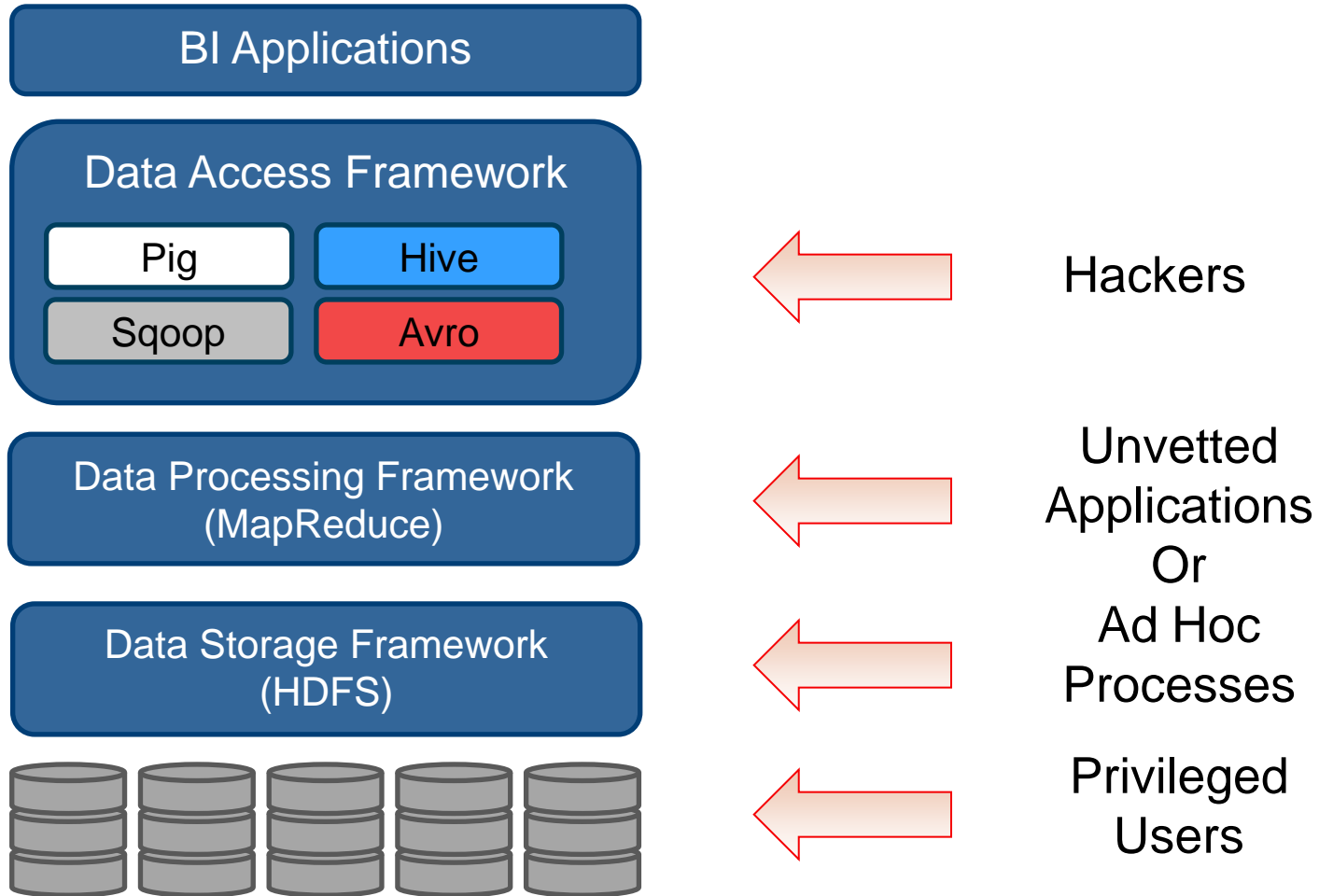
	Node 1	Node 2	Node 3	Node 4-18
Master	NameNode	Secondary NameNode	JobTracker	DataNode
	Balancer	Cloudera Manager	MySQL Master	TaskTracker
	HBase Master	Zookeeper	ODI Agent	HBase Region Server
		NoSQL DB Administration*	Hive Server, Hue	NoSQL DB Storage Node*
Slave	DataNode	Data Node	Data Node	
	NoSQL DB Storage Node*	NoSQL DB Storage Node*	NoSQL DB Storage Node*	
		MySQL DB Slave		

Hadoop Processes
 MySQL
 HBase
 Oracle NoSQL DB

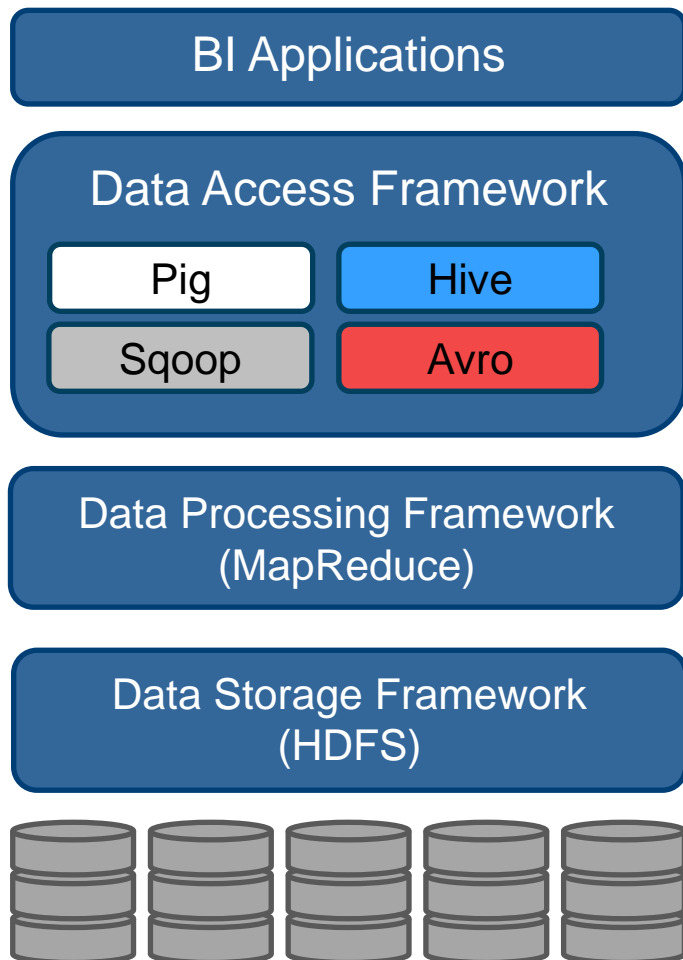


Source: 04_Oracle_Big_Data_Appliance_Deep_Dive.pdf

Many Ways to Hack Big Data



Hadoop - Protection Beyond Kerberos



API enabled **Field level data protection** with **Policy** based access control and Monitoring



API enabled **Field level data protection** with **Policy** based access control and Monitoring



Field level data protection with **Policy** based access control and Monitoring; existing and new data



Volume Encryption with **Policy** based access control of Files and Monitoring

What's the Problem with Securing Big Data?

- Analytics
- Inter-node data movement
- Encryption
 - Data size
 - Data type
 - Performance (SLA)
 - Table scans

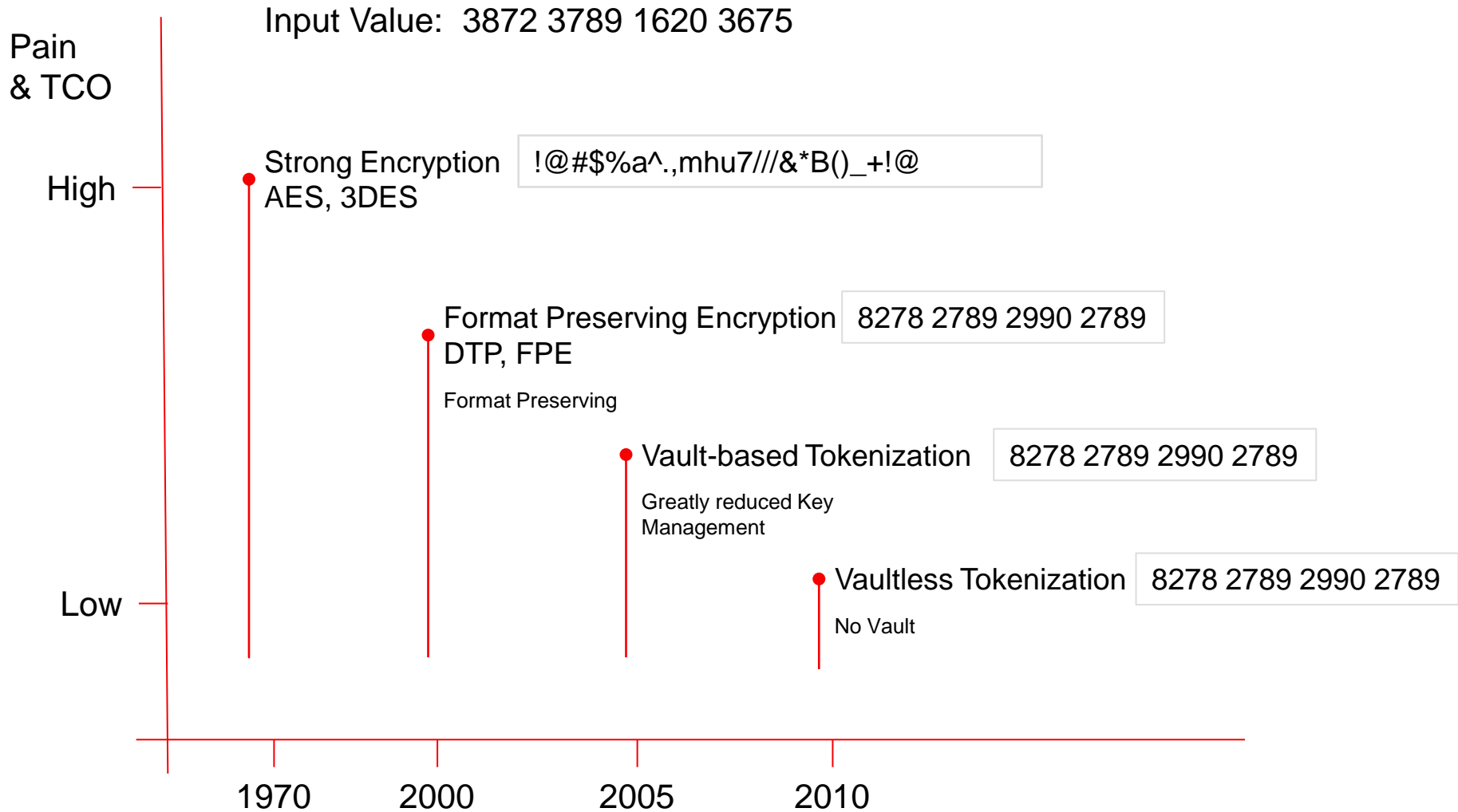
PCI DSS # 3.4

Render PAN* unreadable anywhere it is stored by using any of the following approaches:

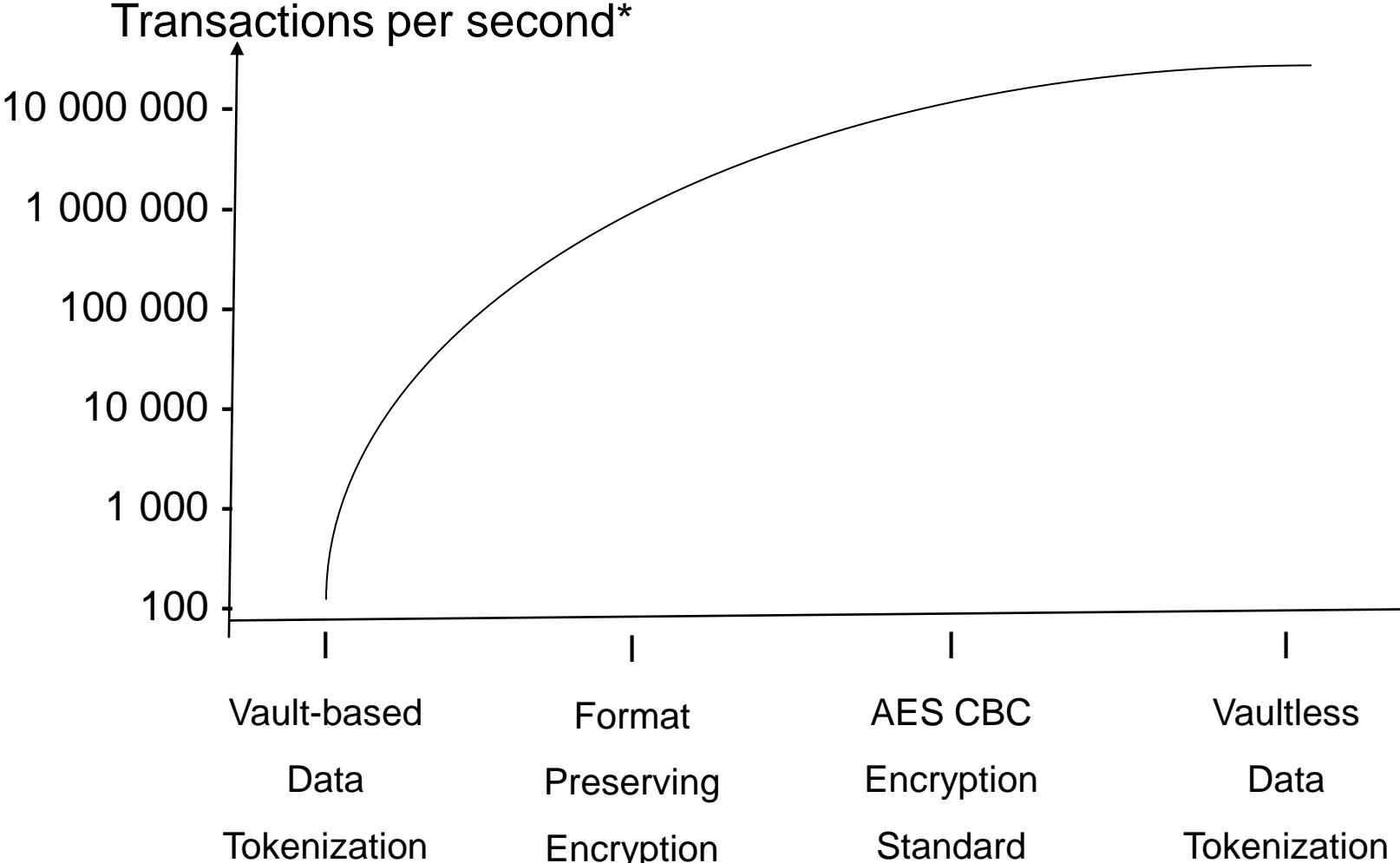
- ❖ **Index tokens** and pads
- ❖ Strong cryptography with associated key-management processes and procedures
- ❖ One-way hashes based on strong cryptography
- ❖ Truncation

* : Primary Account Number (credit card number)

Reduction of Pain with New Protection Techniques



Speed of Different Protection Methods



*: Speed will depend on the configuration



Protection Granularity: Field Protection

Production Systems

Encryption

- Reversible
- Policy Control (authorized / Unauthorized Access)
- Lacks Integration Transparency
- Complex Key Management
- Example !@#%a^.,mhu7///&*B()_+!@

Non-Production Systems

Masking

- Not reversible
- No Policy, Everyone can access the data
- Integrates Transparently
- No Complex Key Management
- Example 0389 3778 3652 0038

Protection Granularity: Field Protection

Production Systems

Encryption

- Reversible
- Policy Control (authorized / Unauthorized Access)
- Lacks Integration Transparency
- Complex Key Management
- Example !@#\$%a^.,mhu7///&*B()_+!@

Vaultless Tokenization / Pseudonymization

- Reversible
- Policy Control (Authorized / Unauthorized Access)
- Integrates Transparently
- No Complex Key Management
- Business Intelligence Credit Card: 0389 3778 3652 0038

Non-Production Systems

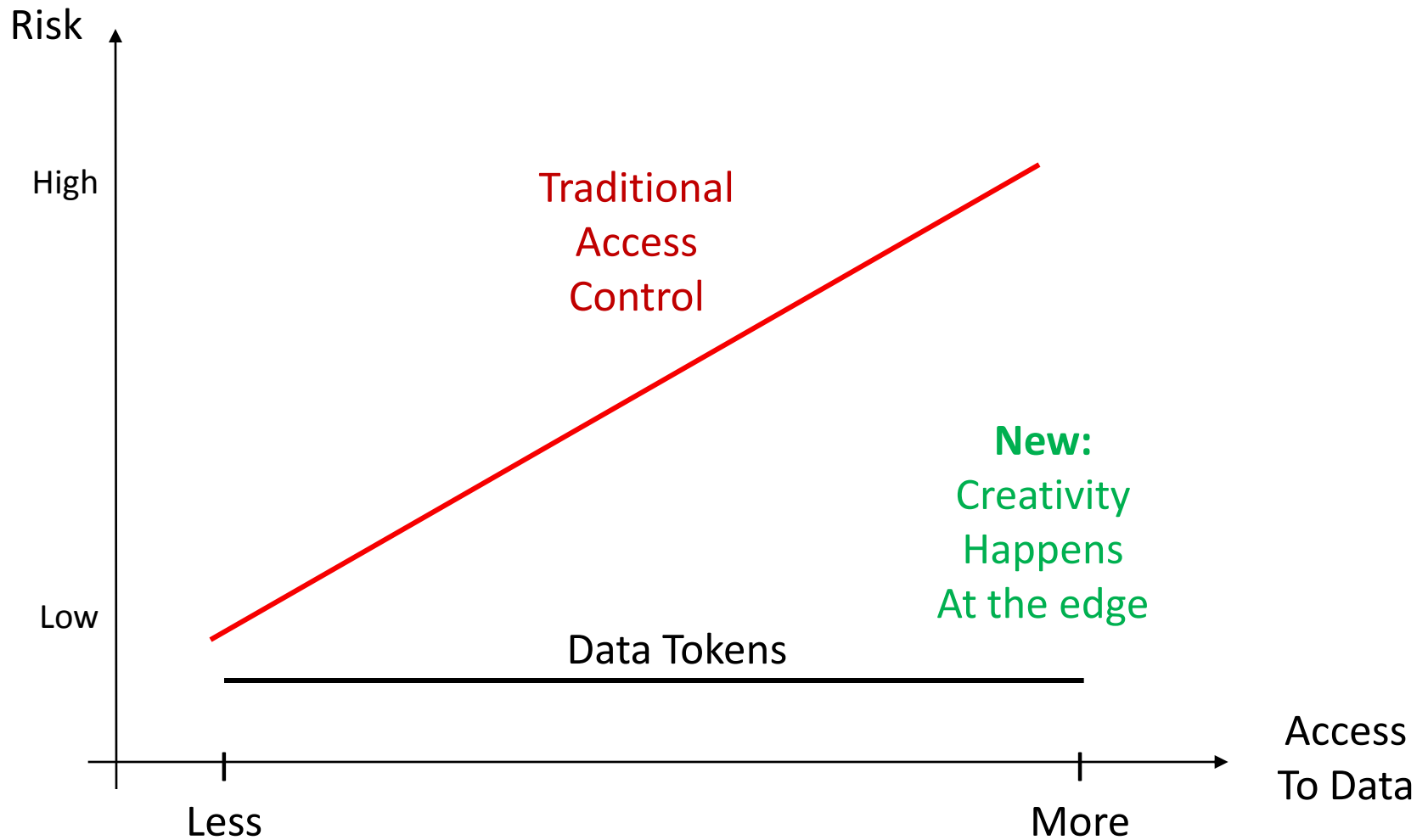
Masking

- Not reversible
- No Policy, Everyone can access the data
- Integrates Transparently
- No Complex Key Management
- Example 0389 3778 3652 0038




“Tokenization Gets Traction”

- Aberdeen has seen a steady increase in enterprise use of tokenization for protecting sensitive data over encryption
- Nearly half of the respondents (47%) are currently using tokenization for something other than cardholder data
- Over the last 12 months, tokenization users had 50% fewer security-related incidents than tokenization non-users

New Data Security = More Creativity



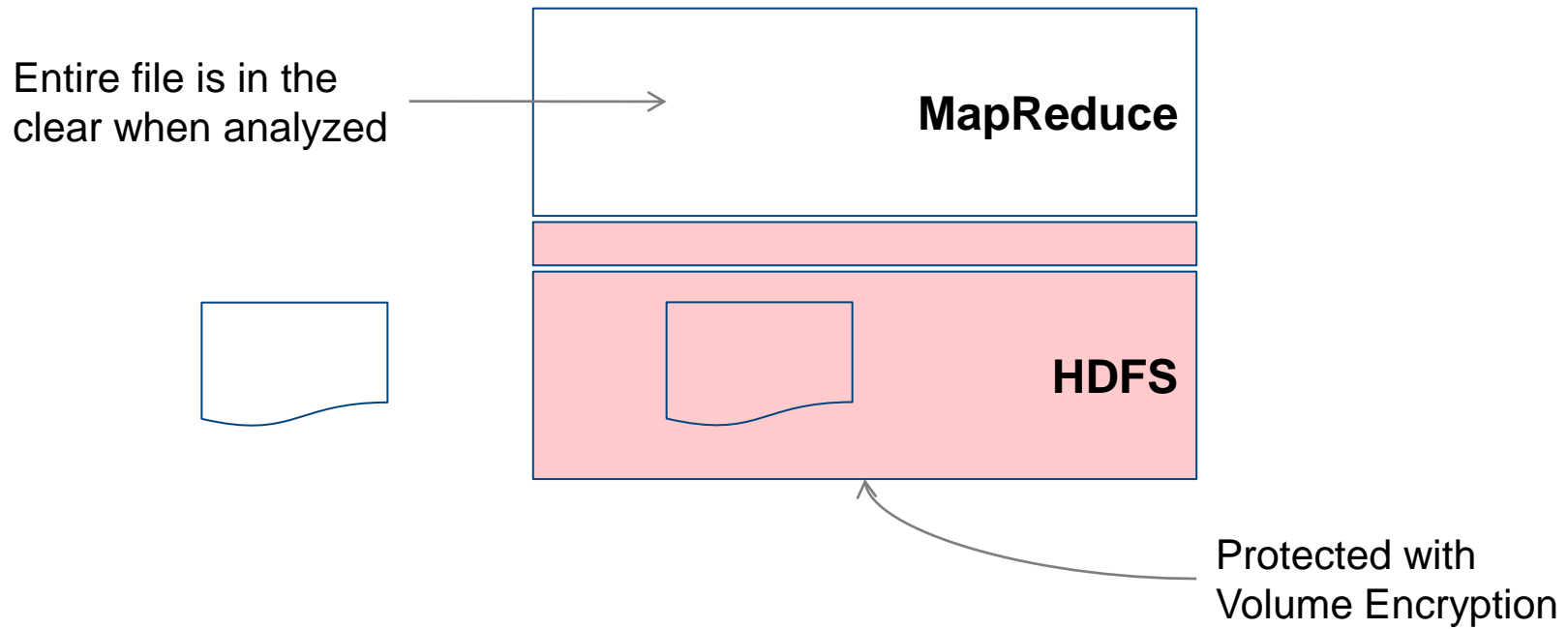
De-Identified Sensitive Data

Field	Real Data	Tokenized / Pseudonymized
Name	Joe Smith	csu wusoj
Address	100 Main Street, Pleasantville, CA	476 srta coetse, cysieondusbak, CA
Date of Birth	12/25/1966	01/02/1966
Telephone	760-278-3389	760-389-2289
E-Mail Address	joe.smith@surferdude.org	eo.e.nwuer@beusorpdqo.org
SSN	076-39-2778	076-28-3390
CC Number	3678 2289 3907 3378	3846 2290 3371 3378
Business URL	www.surferdude.com	www.sheyinctao.com
Fingerprint		Encrypted
Photo		Encrypted
X-Ray		Encrypted
Healthcare Data – Primary Care Data	Dr. visits, prescriptions, hospital stays and discharges, clinical, billing, etc.	Protection methods can be equally applied to the actual healthcare data, but not needed with de-identification

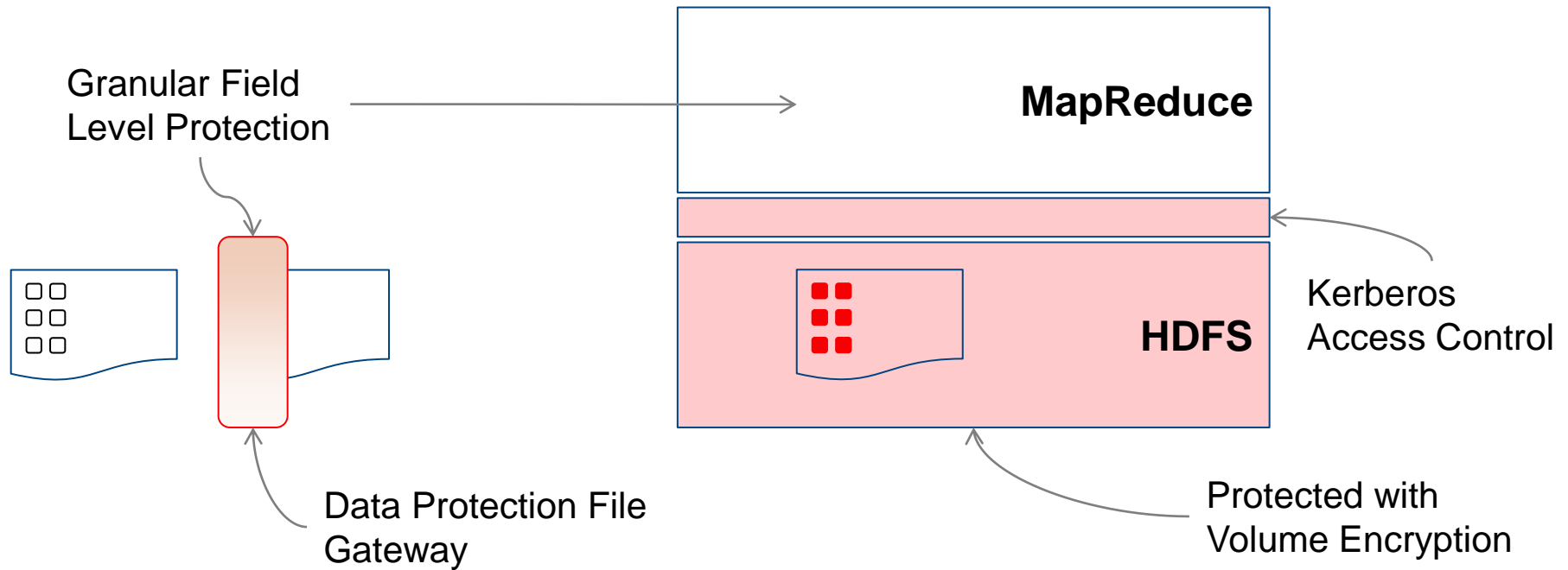
Flexibility in Token Format Controls

Type of Data	Input	Token	Comment
Credit Card	3872 3789 1620 3675	8278 2789 2990 2789	Numeric
Credit Card	3872 3789 1620 3675	3872 3789 2990 3675	Numeric, First 6, Last 4 digits exposed
Credit Card	3872 3789 1620 3675	3872 qN4e 5yPx 3675	Alpha-Numeric, Digits exposed
Account Num	29M2009ID	497HF390D	Alpha-Numeric
Date	10/30/1955	12/25/2034	Date - multiple date formats
E-mail Address	yuri.gagarin@protegrity.com	empo.snaugs@svtiensnni.snk	Alpha Numeric, delimiters in input preserved
SSN	0756722278 or 075-67-2278	287382567 or 287-38-2567	Numeric, delimiters in input
Binary	0x010203	0x123296910112	
Decimal	123.45	9842.56	Non length preserving
Alphanumeric Indicator	5105 1051 0510 5100	8278 2789 299A 2781	Position to place alpha is configurable
Invalid Luhn	5105 1051 0510 5100	8278 2789 2990 2782	Luhn check will fail
Multi-Merchant or Multi-Client ID	3872 3789 1620 3675	ID 1: 8278 2789 2990 2789 ID 2: 9302 8999 2662 6345	This supports delivery of a different token to different merchant or clients based on the same credit card number.

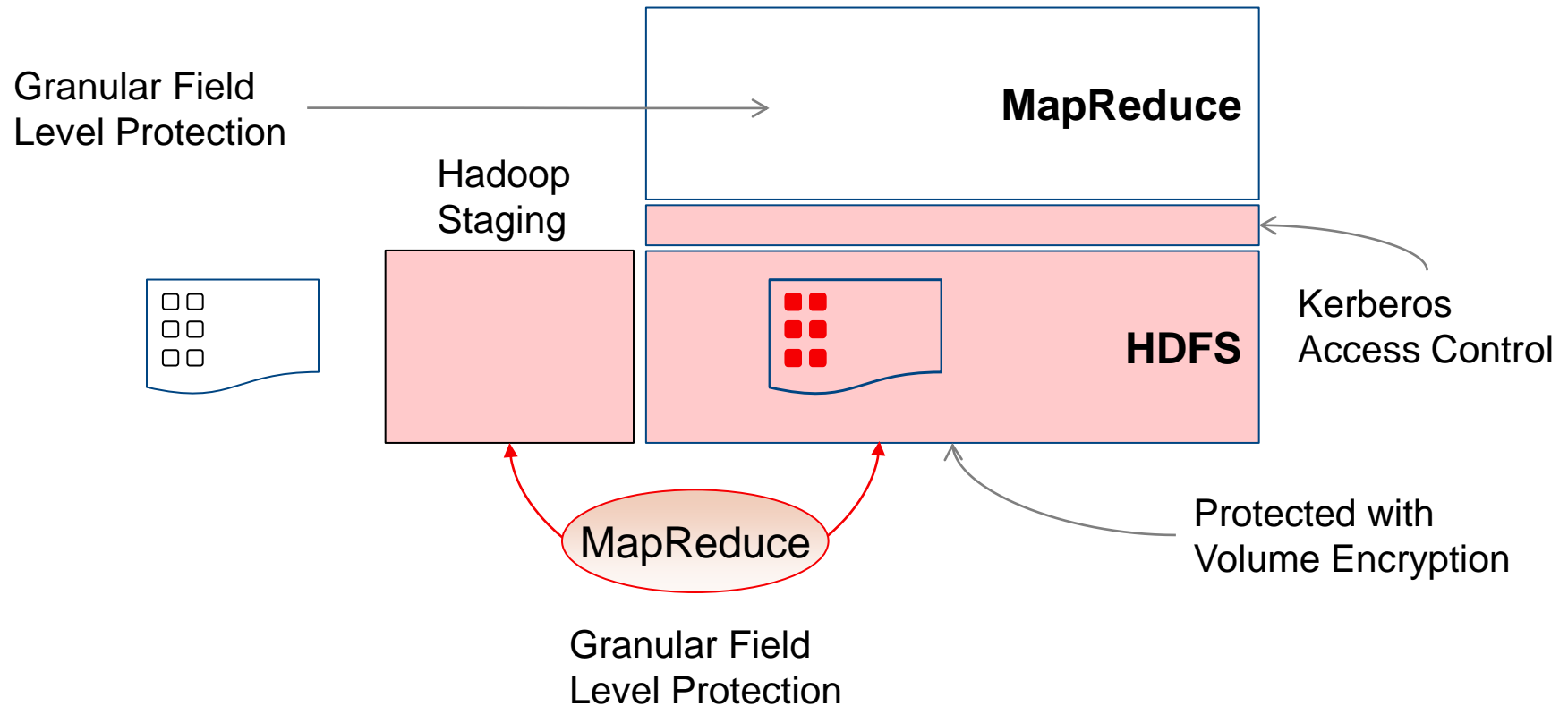
Volume Encryption



Volume Encryption + Gateway Field Protection



Volume Encryption + Internal MapReduce Field Protection



4

Enforce

Policies are used to enforce rules about how sensitive data should be treated in the enterprise.

4. Enforce

The goal of policy enforcement is to;

1. Hide sensitive data from un-authorized users but disclose sensitive data to authorized users.
2. Deliver the minimum information to an individual or a process who needs the information to accomplish a task. Least Privilege by NIST.
3. Collect information about who is attempting to access sensitive data – both authorized and unauthorized.

A Data Security Policy

What

What is the sensitive data that needs to be protected. **Data Element.**

How

How you want to protect and present sensitive data. There are several methods for protecting sensitive data. Encryption, tokenization, monitoring, etc.

Who

Who should have access to sensitive data and who should not. Security access control. **Roles & Members.**

When

When should sensitive data access be granted to those who have access. Day of week, time of day.

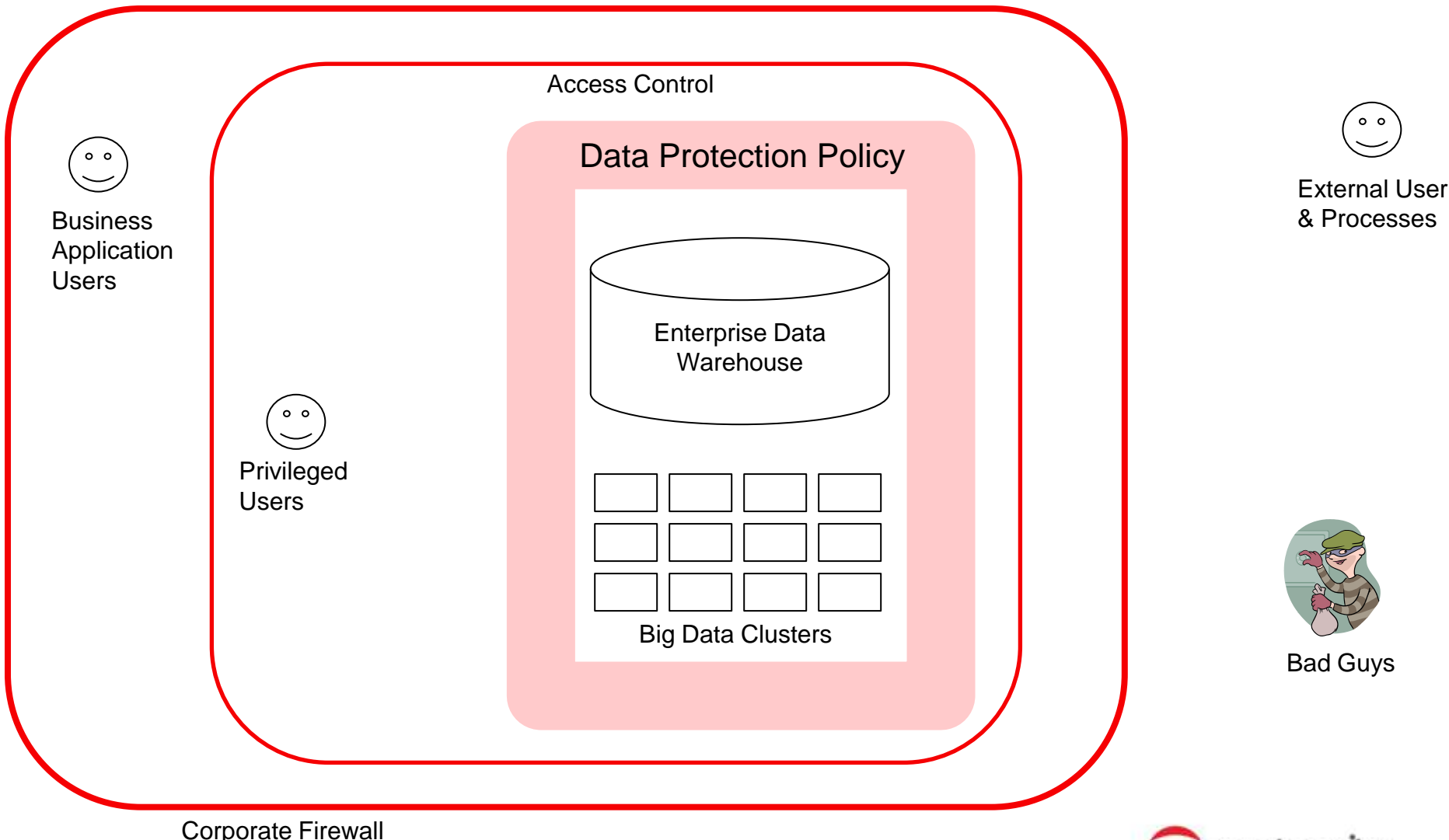
Where

Where is the sensitive data stored? This will be where the policy is enforced. At the protector.

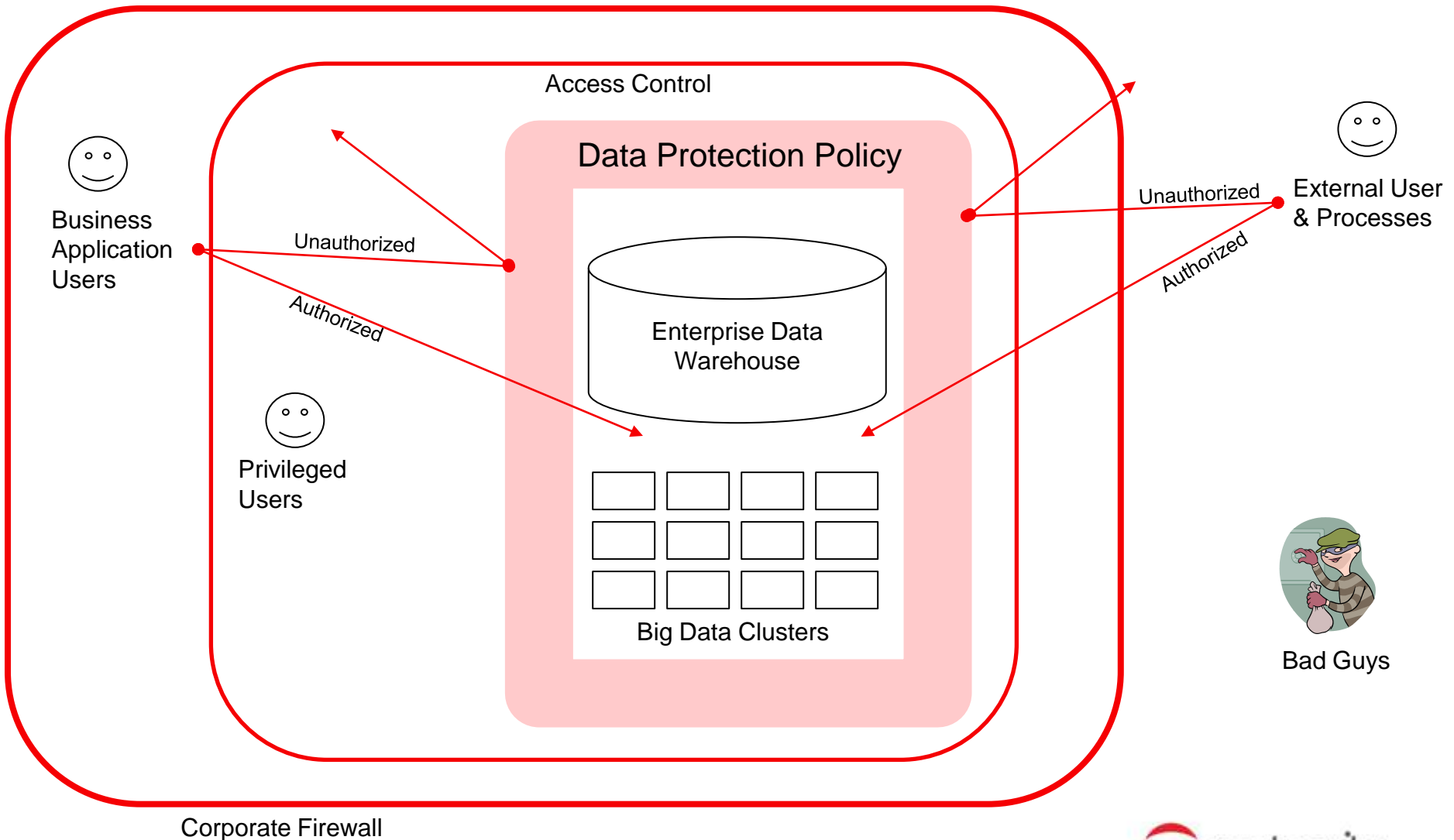
Audit

Audit authorized or un-authorized access to sensitive data. Optional audit of protect/unprotect.

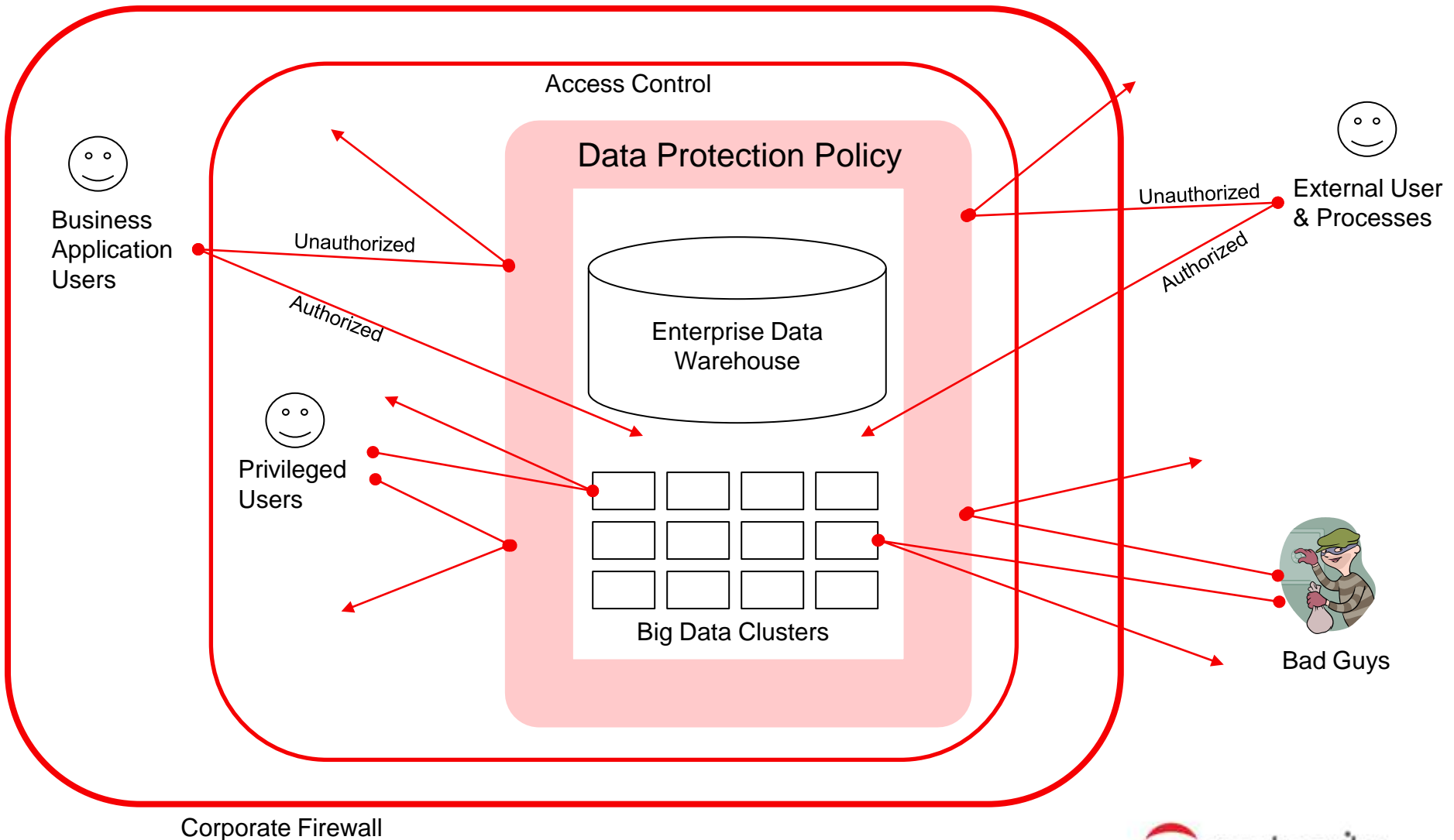
4. Enforce



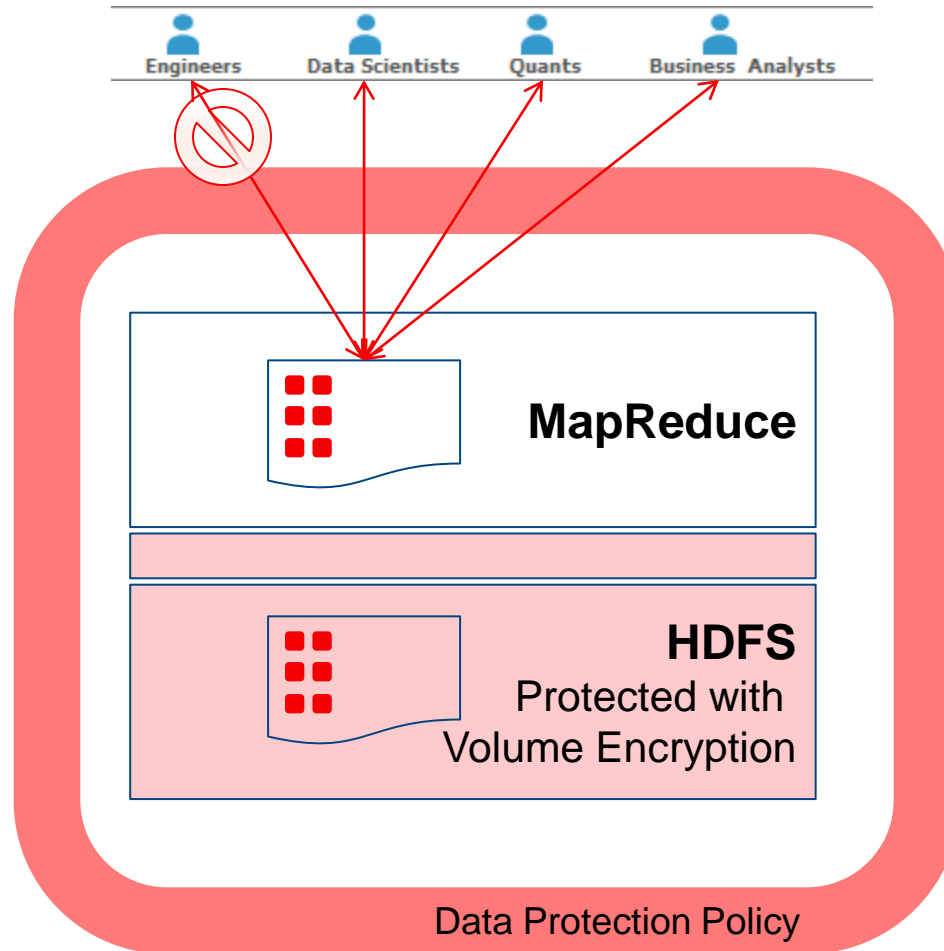
4. Enforce



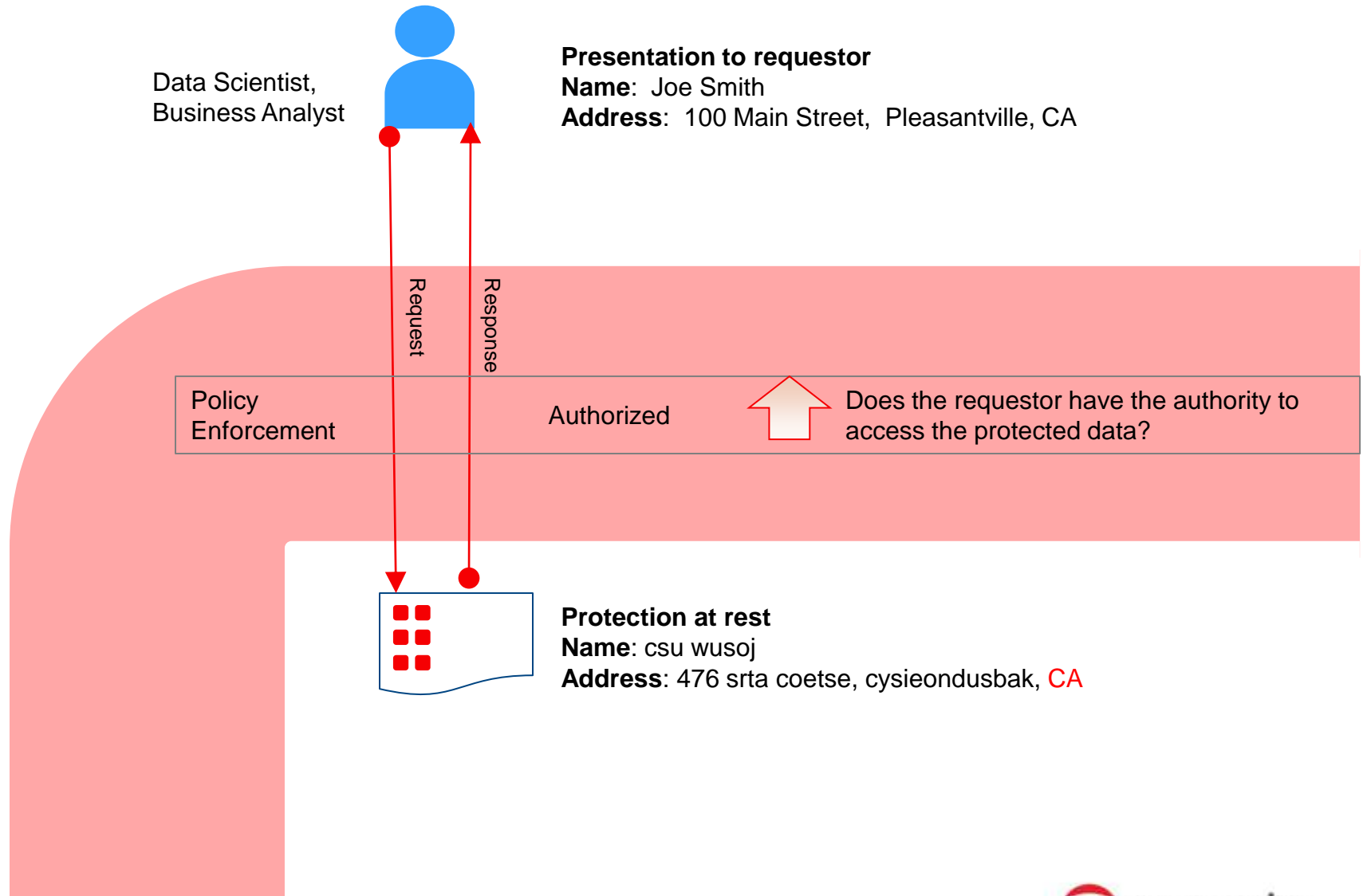
4. Enforce



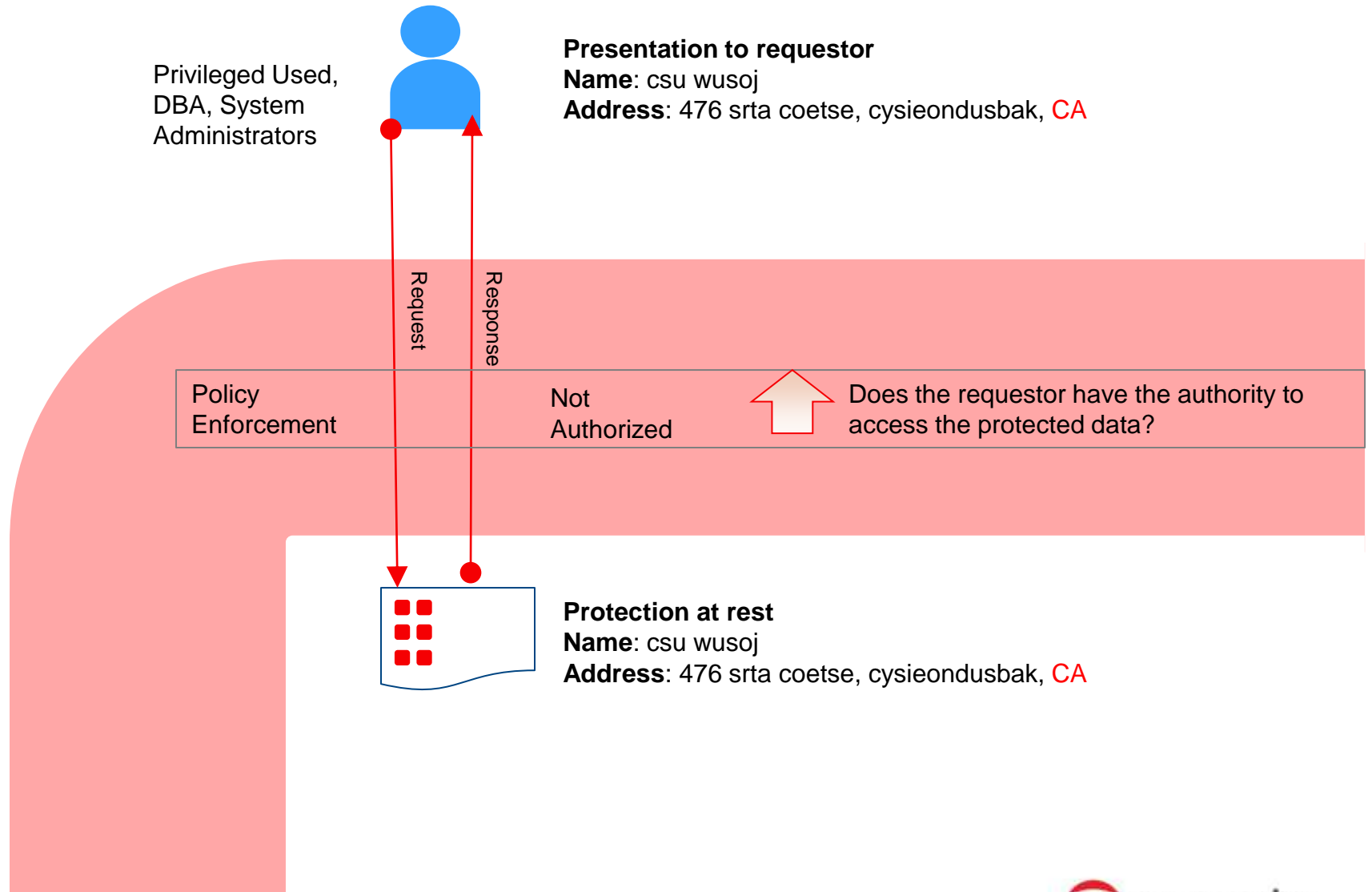
Volume Encryption + Field Protection + Policy Enforcement



4. Authorized User Example



4. Un-Authorized User Example



5

Monitor

A critically important part of a security solution is the ongoing monitoring of any activity on sensitive data.

5. Monitor

Policy enforcement collects information in the form of audit logs about any activity on sensitive data.

Monitoring enables security personnel to gain insights on what's going on with your sensitive data.

It enables the understanding of what's normal and what's not normal activity on sensitive data.

Best Practices for Protecting Big Data

- **Start Early** – Don' wait until you have terabytes of sensitive data in Hadoop before starting your Big Data protection program.
- **Granular protection** in addition to access control and volume protection.
- Future proof your protection. **Select the optimal protection** for today and for the future.
- **Enterprise coverage** to ensure nothing is left vulnerable.
- **Protection against insider threat** is more important today than ever before. Can only achieve this through granular data protection techniques.
- You can protect highly sensitive data while in a way that is mostly **transparent to the analysis process**.
- **Policy based protection** provides a shield to your sensitive data while recording all events on that data.

How Protegrity Can Help

1

We can help you **Classify** the sensitive data that needs to be secured in your enterprise.

2

We can help you **Discover** where the sensitive data sits in your environment and design the optimal security solution.

3

We can help you **Protect** your sensitive data with a flexible set of protectors and protection methods.

4

We can help you **Enforce** policies that will enable business functions while preventing sensitive data from getting in the wrong hands.

5

We can help you **Monitor** activity on sensitive data to gain insights on abnormal behaviors.

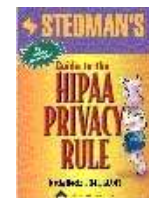
Introduction to Protegrity

○ Proven enterprise data security software and innovation leader

- Sole focus on the protection of data

○ Growth driven by risk management and compliance

- PCI (Payment Card Industry)
- PII (Personally Identifiable Information)
- PHI (Protected Health Information) – HIPAA
- State and Foreign Privacy Laws, Breach Notification Laws



○ Successful across many key industries





Please contact me for more information

Ulf . Mattsson [at] protegrity . Com

Info@protegrity.com



www.protegrity.com