

Best practices for protecting Enterprise Information in BigData & Datawarehouse

Anwar Ali,
Senior Solution Consultant,
Information Management



Big data– a growing phenomenon

20 Petabytes of data every day



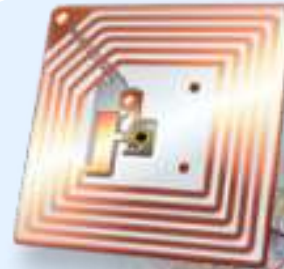
12+ TBs of tweet data every day

25+ TBs of log data every day



76 million smart meters in 2009... 200M by 2014

30 billion RFID tags today (1.3B in 2005)



6 billion mobile phones world wide



100s of millions of GPS enabled devices sold annually

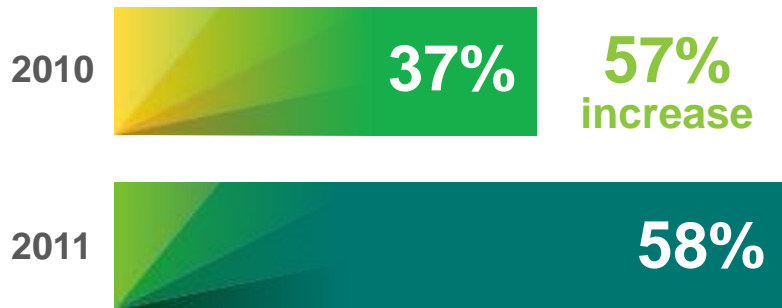


2+ billion people on the Web by end 2011



Analytically sophisticated organizations outperform others

Respondents who say analytics creates a competitive advantage



Organizations achieving a competitive advantage with analytics are

2.2x

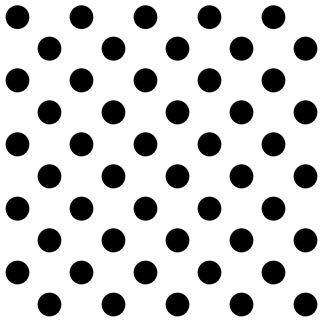
more likely to substantially outperform their industry peers

Source: The New Intelligent Enterprise, a joint *MIT Sloan Management Review* and IBM Institute of Business Value analytics research partnership. Copyright © Massachusetts Institute of Technology 2011

Four dimensions of big data



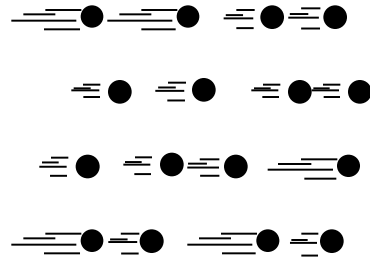
Volume



Data at Rest

Terabytes to exabytes
of existing data to
process

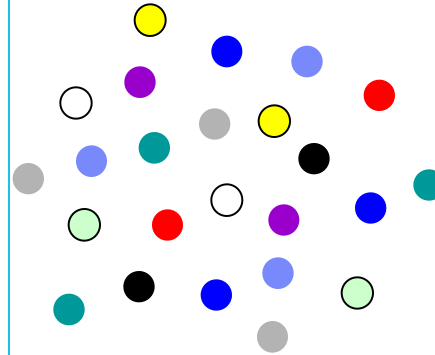
Velocity



Data in Motion

Streaming data,
milliseconds to
seconds to respond

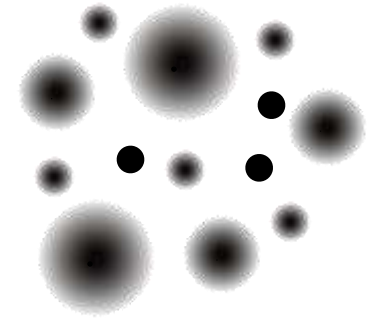
Variety



Data in Many Forms

Structured,
unstructured, text,
multimedia

Veracity*

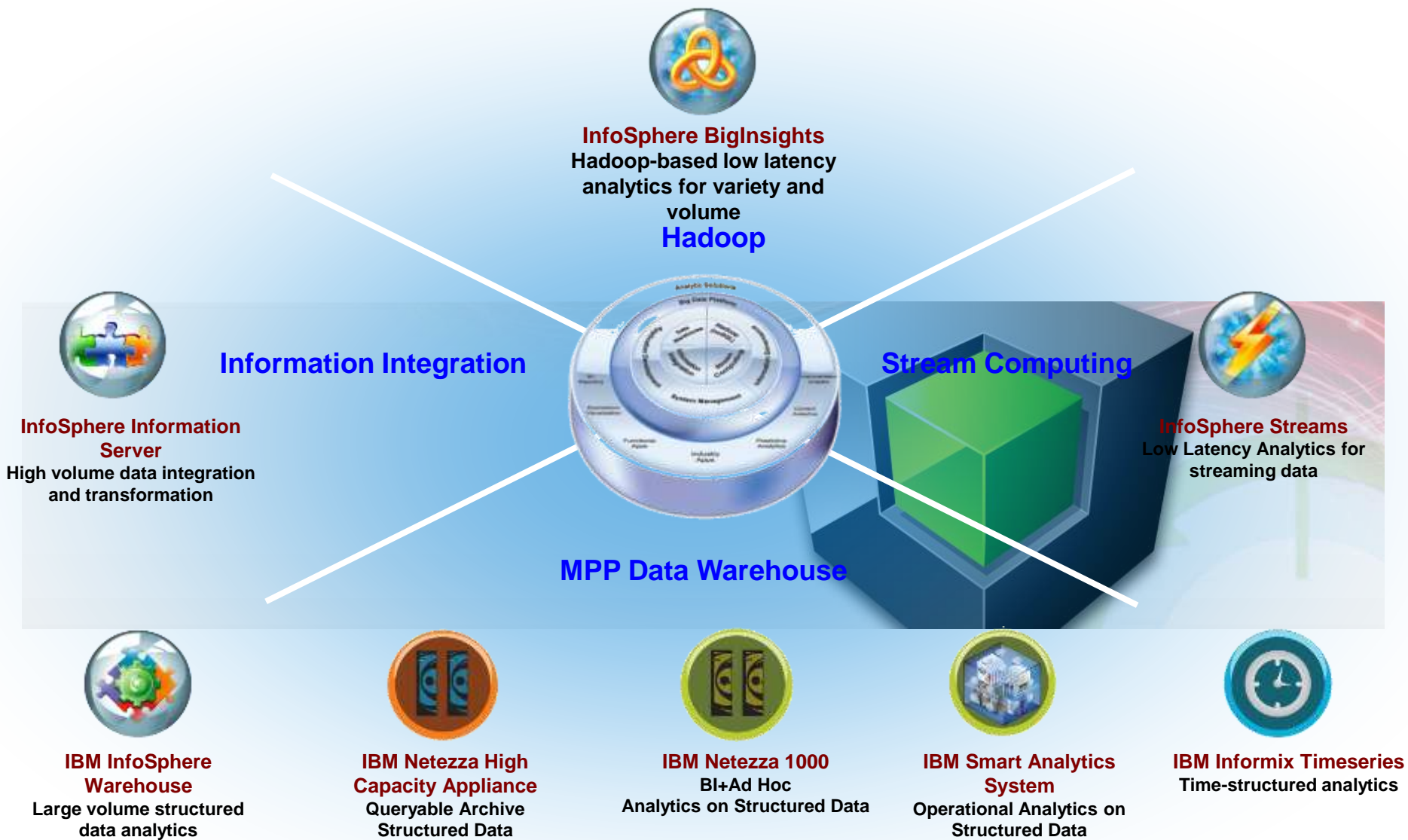


Data in Doubt

Uncertainty due to
data inconsistency
& incompleteness,
ambiguities, latency,
deception, model
approximations

* Truthfulness, accuracy or precision, correctness

The IBM Big Data Platform, Enterprise Big Data Landscape



Security and compliance concerns in big data environments

Big Data Platform

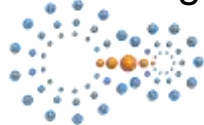
Structured



Unstructured



Streaming



- Who is running specific big data requests?
- What map-reduce jobs are they running?
- Are they trying to download all of the sensitive data for non-authorized purposes?,
- Is there an exceptional number of file permission exceptions?
- Are these jobs part of an authorized program list accessing the data?
- Has some new query application been developed that you were previously unaware existed?

- Massive volume of structured data movement
 - 2.38 TB / Hour load to data warehouse
 - High-volume load to Hadoop file system
- Ingest unstructured data into Hadoop file system
- Integrate streaming data sources



Clients

Hadoop Cluster

Regulations require data protection, no matter where data resides!



Traditional approaches miss the mark

855

security incidents in 2011, compromising **174 M** records

\$5.5M

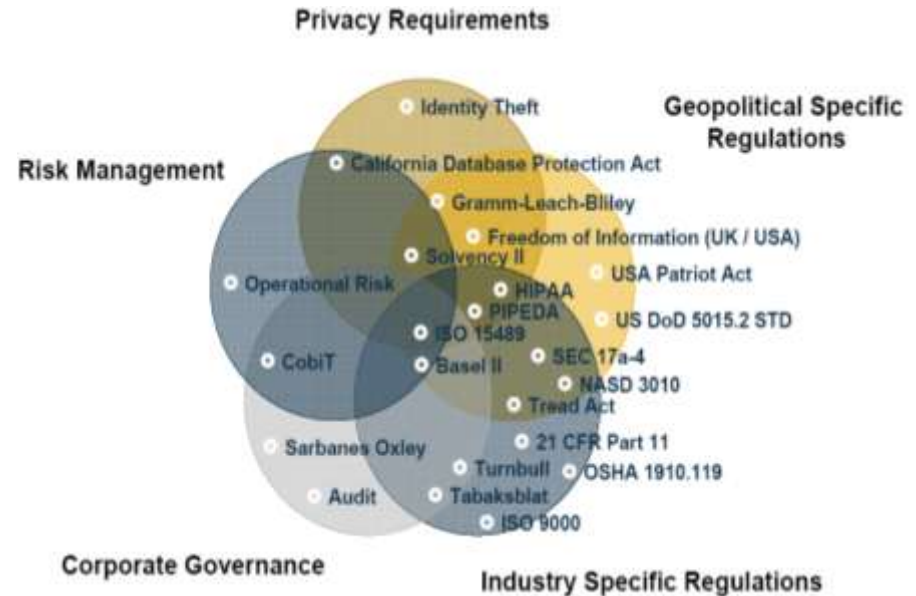
average cost per breach

\$3M

cost of losing customer loyalty (lost business) following a data breach

556M

consumer cybercrime victims in 2011, or 18 adults became a victim every second



\$3.5M

Yearly average cost of compliance

More frequent and damaging data breaches



Attacks impact 650 Israeli government websites, wipes databases, leaks emails and passwords

November 2012: Thousands of email addresses/passwords, hundreds of Web sites, government & privately owned



SQL Injection Infects Millions of Web Pages; Attacks up 69%

September 2012: Attacker takes full control of OS, DB and App



Rogue employee steals over 9,000 drivers licenses records

February 2013: Employee had no "business need" to access data

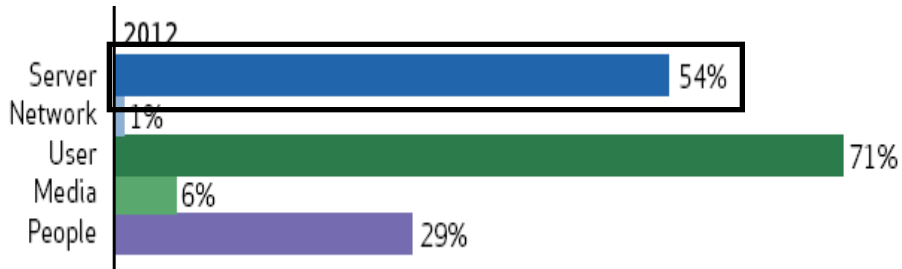


Medical records exposed for nearly 300,000 clients; FTC sues

February 2013: FTC demands accountability for privacy practices

Database servers are a primary source

Real time security is needed –
66% of breaches that remain undiscovered for months or more!



WHY?

- Database servers contain your client's most valuable information
 - Financial records
 - Customer information
 - Credit card and other account records
 - Personally identifiable information
 - Patient records
- High volumes of structured data
- Easy to access

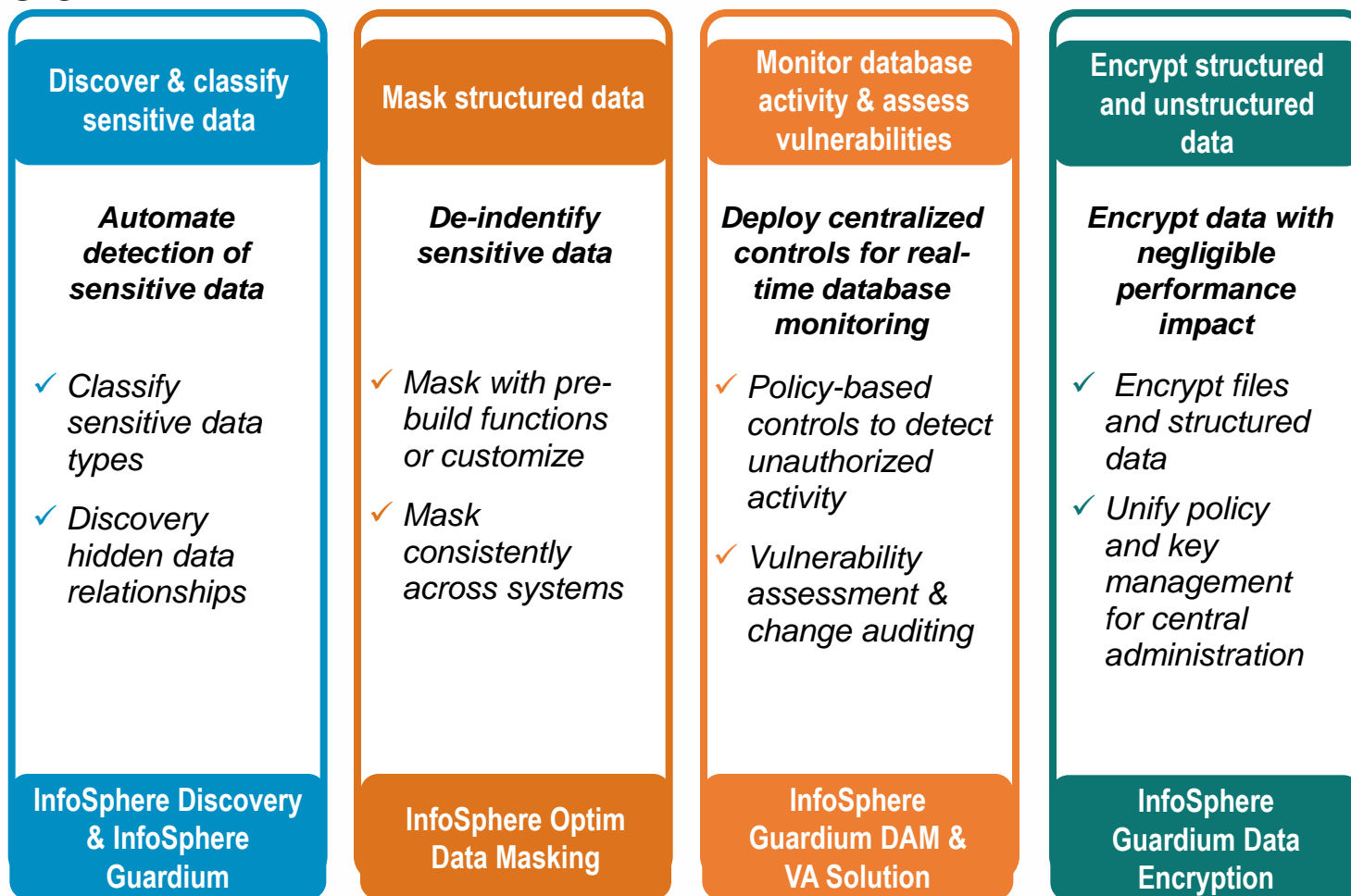
2013 Data Breach Report from Verizon Business RISK Team

<http://www.verizonenterprise.com/DBIR/2013/>



“Go where the money is... and go there often.” - Willie Sutton

InfoSphere provides a complete data protection approach



Satisfy compliance and regulatory mandates

Database Activity Monitoring & Vulnerability Assessment

Addressing the Full Lifecycle of Database Security & Compliance

Real-Time Database Security & Monitoring



The Compliance Mandate – What do you need to monitor?

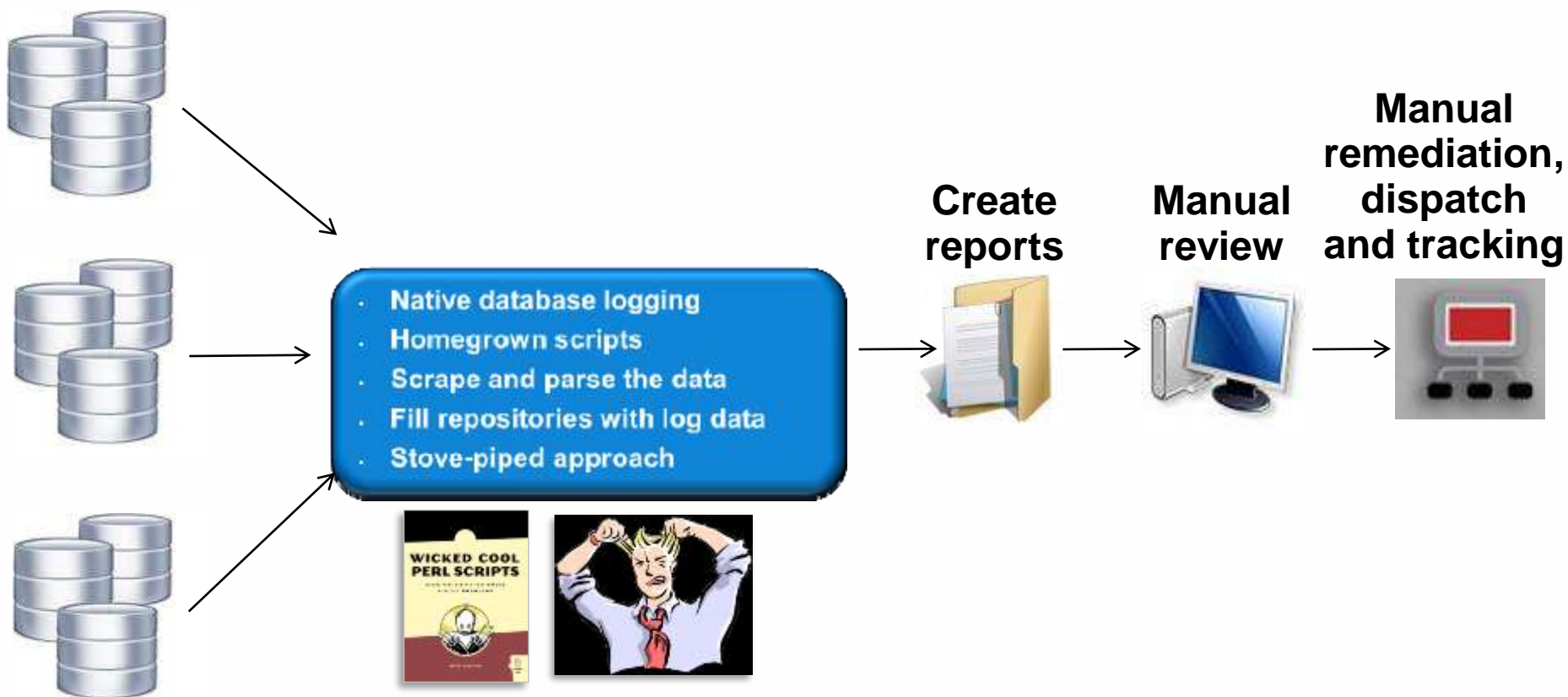
Audit Requirements	COBIT (SOX)	PCI-DSS	ISO 27002	Data Privacy & Protection Laws	NIST SP 800-53 (FISMA)
1. Access to Sensitive Data (Successful/Failed SELECTs)		✓	✓	✓	✓
2. Schema Changes (DDL) (Create/Drop/Alter Tables, etc.)	✓	✓	✓	✓	✓
3. Data Changes (DML) (Insert, Update, Delete)	✓		✓		
4. Security Exceptions (Failed logins, SQL errors, etc.)	✓	✓	✓	✓	✓
5. Accounts, Roles & Permissions (DCL) (GRANT, REVOKE)	✓	✓	✓	✓	✓

DDL = Data Definition Language (aka schema changes)

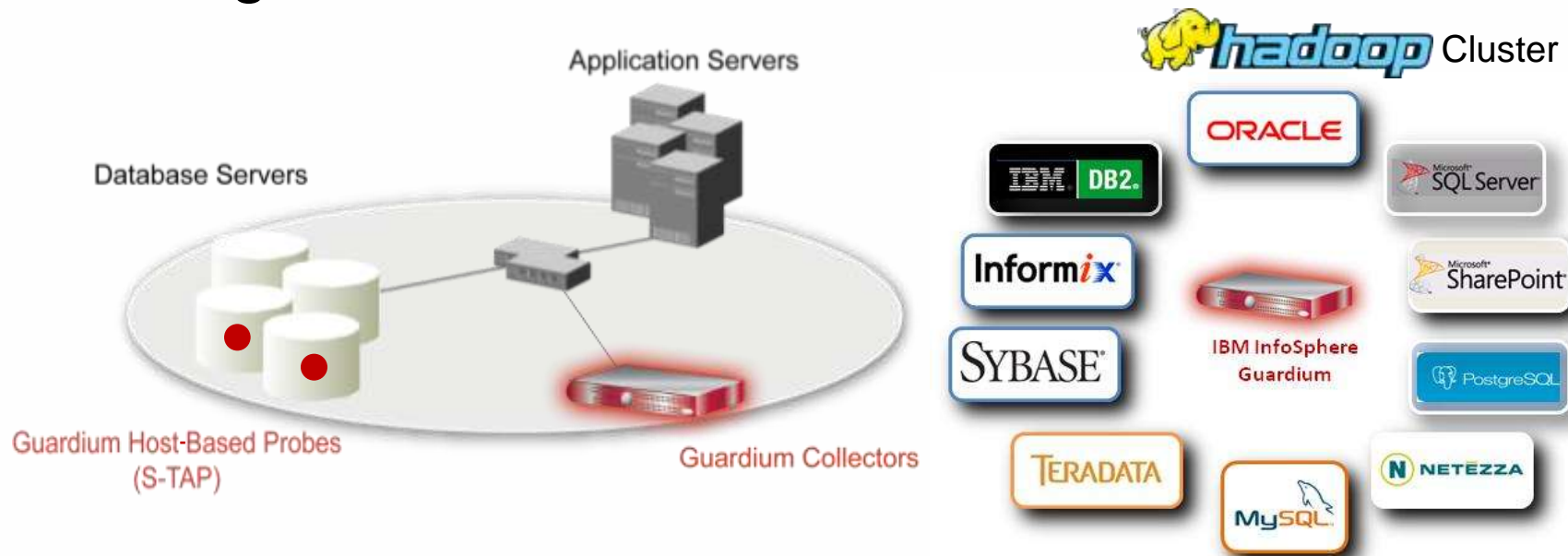
DML = Data Manipulation Language (data value changes)

DCL = Data Control Language

What Database Audit Tools are Enterprises Using Today?



Non-Invasive, Real-Time Database Security & Monitoring



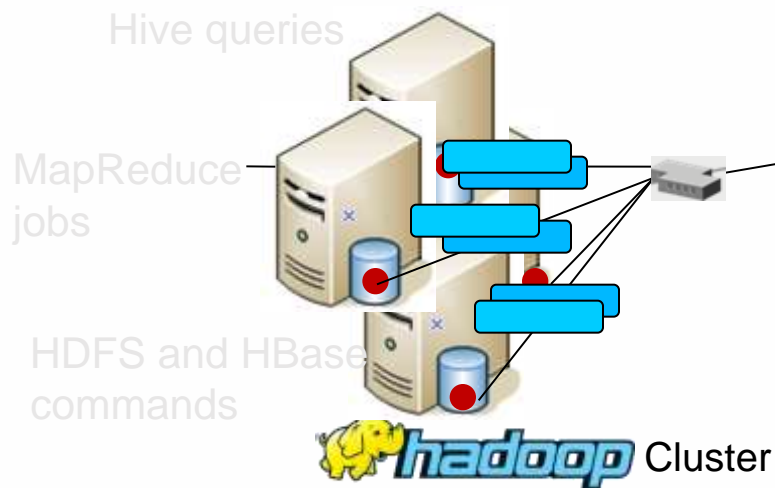
- Continuously monitors all database activities (including local access by superusers)
- Heterogeneous, cross-DBMS solution
- Does not rely on native DBMS logs
- Minimal performance impact (2-3%)
- No DBMS or application changes
- Supports Separation of Duties
- Activity logs can't be erased by attackers or DBAs
- Automated compliance reporting, sign-offs & escalations (SOX, PCI, NIST, etc.)
- Granular, real-time policies & auditing
 - *Who, what, when, where, how*

Guardium monitors data activities on BigInsights

Relevant messages are copied and sent to collector

InfoSphere Guardium S-TAP ●

- Minimal impact to big data server resources or network
- Separation of duties – audited data on secure appliance
- Heterogeneous support (IBM, Cloudera)



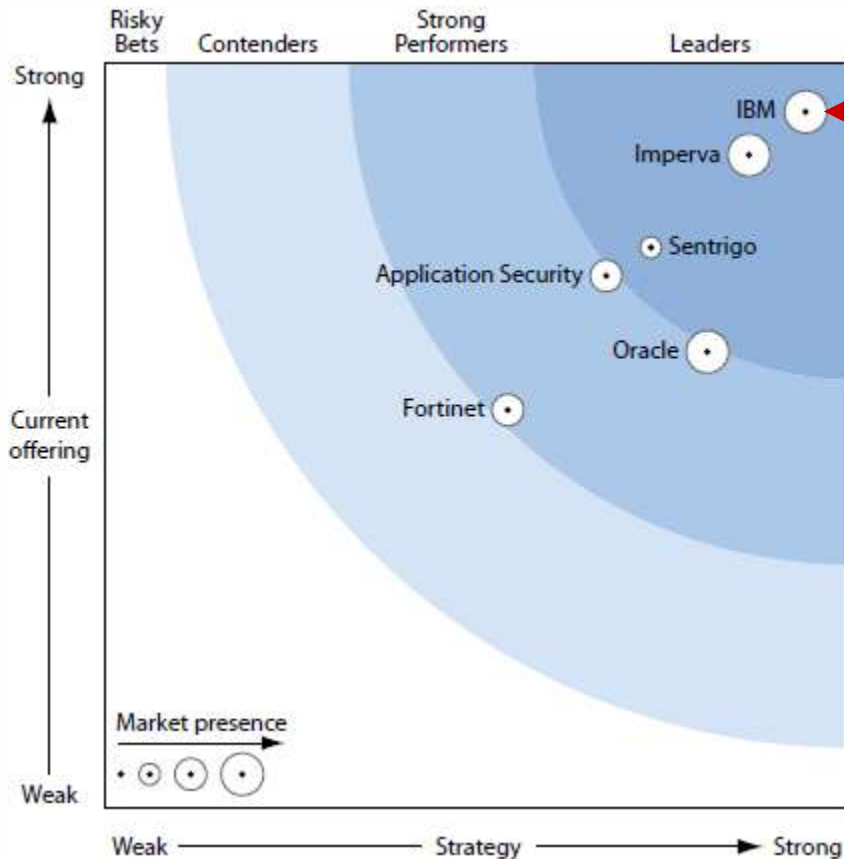
Sensitive data alert!

<u>Access Rule Description</u>	<u>Client IP</u>	<u>Server IP</u>	<u>DB User Name</u>	<u>Full SQL String</u>
Access Rule: sensitive files: Alert9.70.145.1189.70.145.113SVORUGA				__WGPB message {struct:1='getFilefo',struct:2=
				struct:3='org.apache.hadoop.hdfs.protocol.ClientProtocol',varint:4=1}
Access Rule: sensitive files: Alert9.70.145.1189.70.145.113SVORUGA				__WGPB message {struct:1='getListing',struct:2=
				varint:3=0},struct:3='org.apache.hadoop.hdfs.protocol.ClientProtocol',varint:4=1}

InfoSphere Guardium reporting and alerting

Highest Overall Score for Current Offering, Strategy, & Market Presence

FORRESTER



- “Guardium continues to demonstrate its **leadership** in supporting **very large heterogeneous environments**, delivering **high performance and scalability**, simplifying administration, and performing **real-time database protection**.”
- “IBM continues to **focus on innovation** and extending the Guardium product to integrate with other IBM products.”
- **#1 score in all 3 Top Categories and all 17 subcategories** along with perfect scores for Audit Policies; Auditing Repository; Corporate Strategy; Installed Base; Services; and International Presence.
- “Guardium offers **support for almost any of the features that one might find** in an auditing and real-time protection solution.”
- “Guardium offers **strong support for database-access auditing, application auditing, policy management, auditing repository, and real-time protection**.”
- “Guardium has been **deployed across many large enterprises and hundreds of mission-critical databases**.”
- “IBM offers **comprehensive professional services to help customers with complex environments** as well as those who need assistance **implementing database security across their enterprise**.”

The Forrester Wave is copyrighted by Forrester Research, Inc. Forrester and Forrester Wave are trademarks of Forrester Research, Inc. The Forrester Wave is a graphical representation of Forrester’s call on a market and is plotted using a detailed spreadsheet with exposed scores, weightings, and comments. Forrester does not endorse any vendor, product, or service depicted in the Forrester Wave. Information is based on best available resources. Opinions reflect judgment at the time and are subject to change.

Source: “The Forrester Wave™: Database Auditing And Real-Time Protection

Validated by Industry Experts



"Dominance in this space"
#1 Scores for Current Offering,
Architecture & Product Strategy



**"Guardium is ahead of the pack
and gaining
speed."**



*2007 Editor's Choice Award
in "Auditing and
Compliance"*



**"Most Powerful Compliance
Regulations Tools ... Ever"**



"Top of DBEP Class"
"Practically every feature you'll
need to lock down sensitive data."



"Enterprise-class data security
product that should be on every
organization's radar."

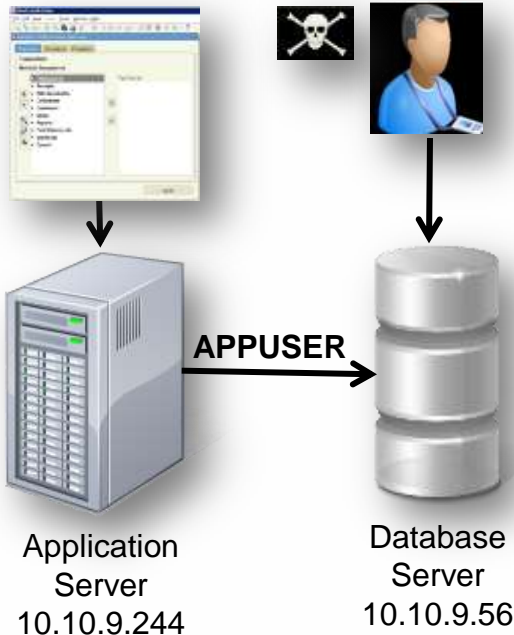


*"5-Star Ratings: Easy
installation, sophisticated
reporting, strong policy-based
security."*



Solution Illustration

Granular Policies with Detective & Preventive Controls



Rule #1 Description: non-App Source AppUser Connection

Category: Security Classification: Breach Severity: MED

Hot Server IP [] / [] and/or Group: Production Servers

Hot Client IP [] / [] and/or Group: Authorized Client IPs

Hot Client MAC [] Hot. Protocol [] and/or Group []

Hot DB Name []

Hot DB User: APPUSER

Field Name [] Object: INVENTORY Command: DROP TABLE

Min. Ct. 0 Reset Interval (minutes) 0

Continue to next Rule Rec. Vals.

Action: ALERT PER MATCH

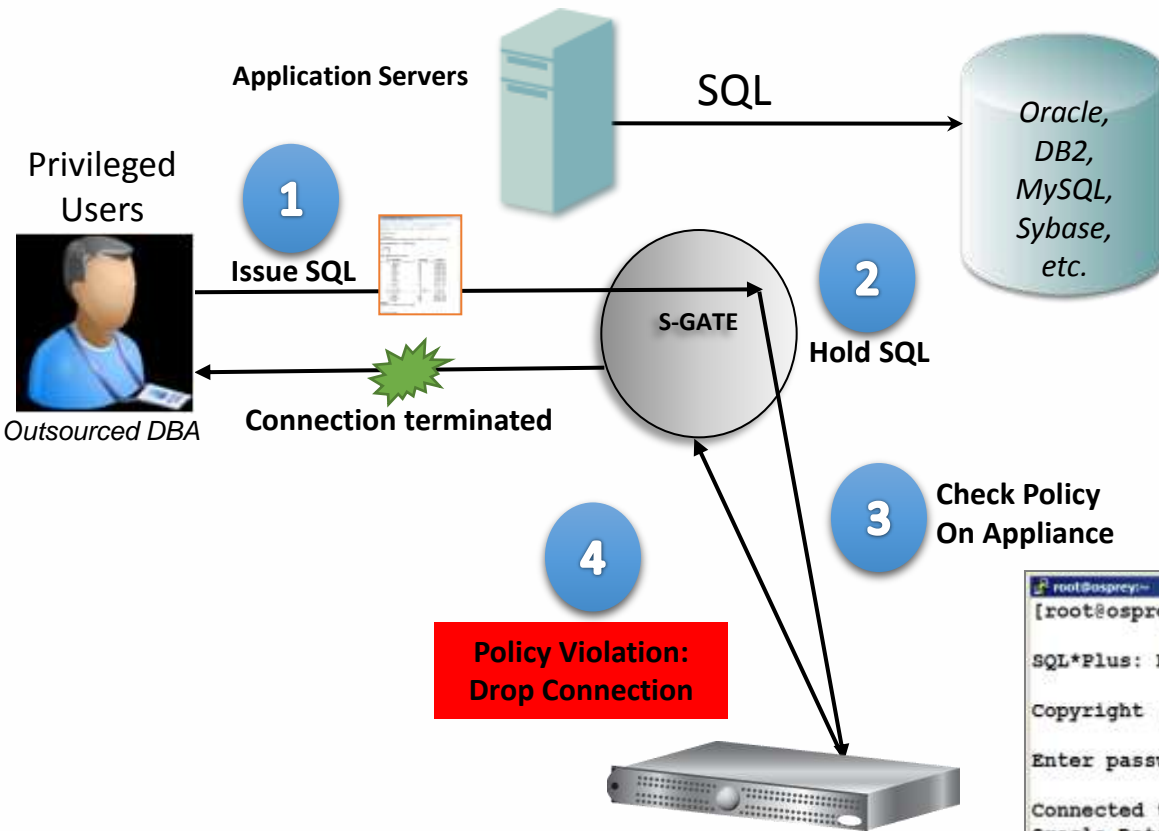
Notification: Notification Type MAIL Mail User marc_gamache@guardium.com

Alert Options:
 ALERT DAILY
 ALERT ONCE PER SESSION
 ALERT PER MATCH
 ALERT PER TIME GRANULARITY
 ALLOW
 IGNORE RESPONSES PER SESSION
 IGNORE SESSION
 IGNORE SQL PER SESSION
 LOG FULL DETAILS
 LOG FULL DETAILS PER SESSION
 LOG FULL DETAILS WITH VALUES
 LOG FULL DETAILS WITH VALUES PER SESSION
 LOG MASKED DETAILS
 LOG ONLY
 RESET
 S-GATE ATTACH
 S-GATE DETACH
 S-GATE TERMINATE
 S-TAP TERMINATE
 SKIP LOGGING

From: GuardiumAlert@guardium.com Sent: Wed 4/15/2009 8:00 AM
 To: Marc Gamache
 Cc:
 Subject: (c1) SQLGUARD ALERT

Subject: (c1) SQLGUARD ALERT Alert based on rule ID non-App Source AppUser Connection
 Category: security Classification: Breach Severity: MED
 Rule # 20267 [non-App Source AppUser Connection]
 Request Info: [Session start: 2009-04-15 06:59:03 Server Type: ORACLE Client IP 192.168.20.160 ServerIP: 172.16.2.152 Client PORT: 11787 Server Port: 1521 Net Protocol: TCP DB Protocol: INS DB Protocol Version: 3.8 DB User: APPUSER
 Application User Name
 Source Program: JDBC THIN CLIENT Authorization Code: 1 Request Type: SQL_LANG Last Error:
 SQL: select * from EmployeeTable

Cross-DBMS, Data-Level Access Control (S-GATE)



- ✓ Cross-DBMS policies
- ✓ Block privileged user actions
- ✓ No database changes
- ✓ No application changes
- ✓ Without risk of inline appliances that can interfere with application traffic

```
root@osprey:~# sqlplus system
[root@osprey ~]# sqlplus system

SQL*Plus: Release 10.2.0.1.0 - Production on Tue May 27 01:13:32 20
Copyright (c) 1982, 2005, Oracle. All rights reserved.

Enter password:

Connected to:
Oracle Database 10g Express Edition Release 10.2.0.1.0 - Production

SQL> select * from creditcard;
select * from creditcard
*
ERROR at line 1:
ORA-03113: end-of-file on communication channel

Session Terminated

SQL>
```

Data Encryption


What is driving data encryption technology?

- Businesses or organizations that retain sensitive data and/or must meet data governance compliance
 - Credit card data
 - PCI - Payment Card Industry Data Security Standard
 - All that process credit card transactions
 - Health data
 - HIPPA - Health Insurance Portability and Accountability Act
 - Health care providers and organizations who retain and/or transmit health related information including financial transactions
 - Financial data
 - Sarbanes-Oxley - Sarbanes-Oxley Act
 - Public companies, public accounting firms, and firms providing auditing services
 - Notification Laws in most states
- Minimal disruption while meeting data security needs and compliance

Why Encrypt Data? Industry Regulations

Payment Card Industry (PCI) Requirements...

1	Install and maintain a firewall	7	Restrict access to data by business need-to-know
2	Do not use vendor-supplied defaults for passwords. Develop configuration standards	8	Assign a unique ID to each person with computer access
3	Protect stored data <i>Encrypt cardholder number</i>	9	Restrict physical access to cardholder data
4	Encrypt transmission of cardholder data across public networks	10	Track and monitor all access to network resources and cardholder data
5	Use and regularly update anti-virus software	11	Systems should be tested to ensure security is maintained over time and through changes
6	Develop and maintain secure systems and applications	12	Maintain an information security policy

 IBM Database Encryption Expert can help

The Data Threats – Data at Rest & Data in Transit

- Online – internal threats
 - Attackers breaking through perimeter security
 - Privileged user abuse
 - Data replicates to many locations
- Offline – theft and loss
 - Backups typically written to portable media
 - Often stored offsite for long periods



-
- Onwire – internal and external threats
 - Hackers and sniffers picking data off the network



What is IBM Database Encryption Expert?

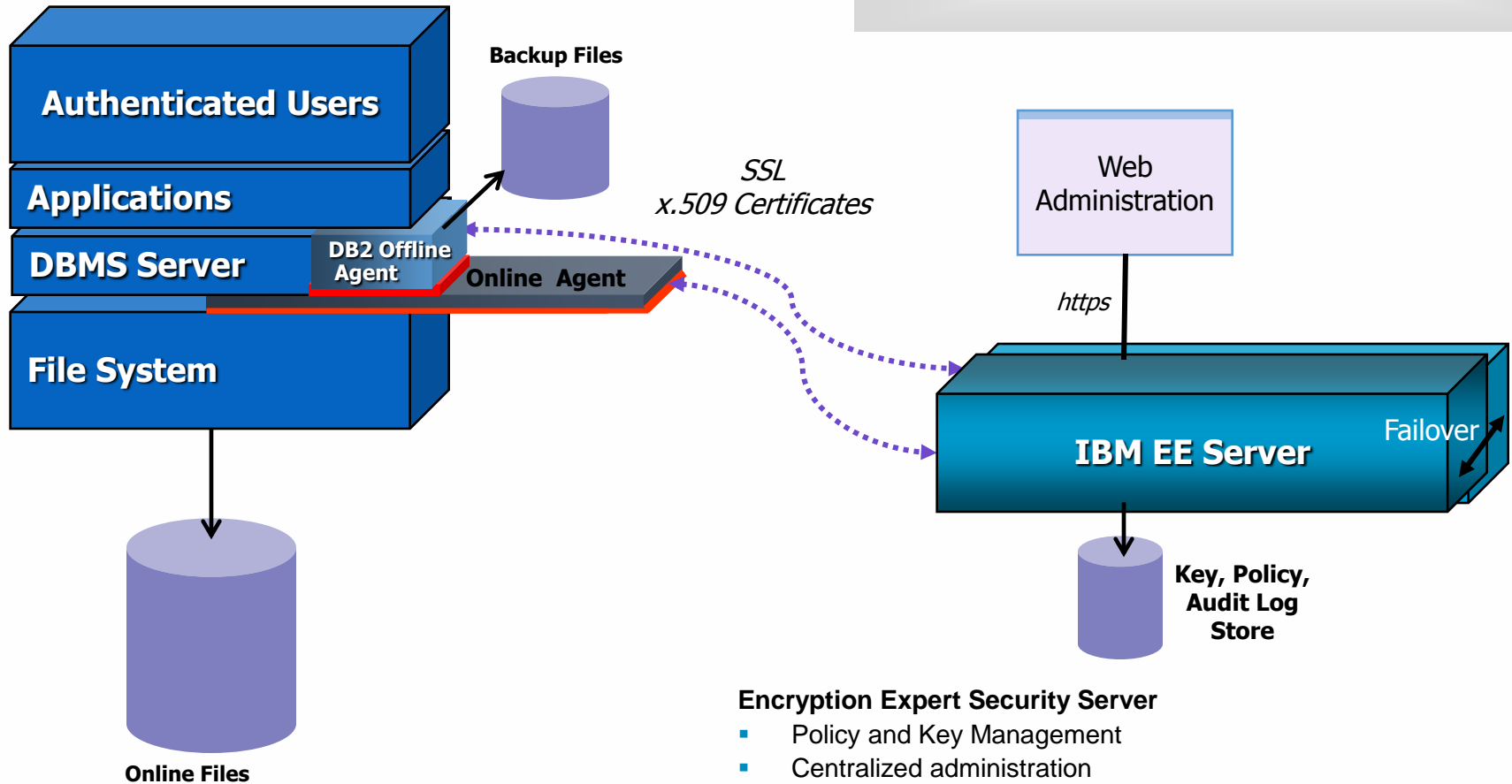
- Data protection for your database environments
 - High performance encryption, access control and auditing
 - Data privacy for both online and backup environments
 - Unified policy and key management for centralized administration across multiple data servers
- Transparency to users, databases, applications, storage
 - No coding or changes to existing IT infrastructure
 - Protect data in any storage environment
 - User access to data same as before
- Centralized administration
 - Policy and Key management
 - Audit logs
 - High Availability



Encryption Expert Architecture

Components:

- EE Security Server
- EE Secure Offline Agent
- EE Secure File System Online Agent



Encryption Expert Security Server

- Policy and Key Management
- Centralized administration
- Separation of duties

Policy Rules

- **WHO** is attempting to access protected data?
 - Configure one or more users, groups, or applications users may invoke who can access protected data
- **WHAT** data is being accessed?
 - Configure a mix of files and directories
- **WHEN** is the data being accessed?
 - Configure a range of hours and days of the week for authorized access
- **HOW** is the data being accessed?
 - Configure allowable file system operations allowed to access the data
e.g. read, write, delete, rename, etc.
- **EFFECT**: Permit; Deny; Apply Key; Audit



Data Masking

Vulnerable non-production environments at risk

Most ignore security in non-production environments



70%

of organizations surveyed use live customer data in non-production environments (testing, Q/A, development)

Database Trends and Applications. *Ensuring Protection for Sensitive Test Data*

\$194

per record
cost of a data breach

The Ponemon Institute. *2012 Cost of Data Breach Study*

50%

of organizations surveyed have no way of knowing if data used in test was compromised

The Ponemon Institute. *The Insecurity of Test Data: The Unseen Crisis*

52%

of surveyed organizations
outsource development

The Ponemon Institute. *The Insecurity of Test Data: The Unseen Crisis*

What is data masking?



- **Definition**

Method for creating a **structurally similar but inauthentic** version of an organization's data. The **purpose is to protect the actual data** while having a functional substitute for occasions when the real data is not required.

- **Requirement**

Effective data masking requires data to be altered in a way that the **actual values cannot be determined** or reengineered, **functional appearance is maintained**.

- **Other Terms Used**

Obfuscation, scrambling, data de-identification

- **Commonly masked data types**

Name, address, telephone, SSN/national identity number, credit card number

- **Methods**

- Static Masking: *Extracts rows from production databases, obfuscating data values that ultimately **get stored in the columns in the test databases***
- Dynamic Masking: *Masks specific data elements **on the fly** without touching applications or physical production data store*

IBM InfoSphere Optim Data Masking Solution



De-identify sensitive information with realistic *but fictional* data



Personal identifiable information is masked with realistic but fictional data

Requirements

- Protect confidential data used in test, training & development systems
- Mask data on screen in applications
- Implement proven data masking techniques
- Support compliance with privacy regulations
- Solution supports custom & packaged ERP applications

Benefits

- Protect sensitive information from misuse and fraud
- Prevent data breaches and associated fines
- Achieve better information governance

Contextually accurate masked data facilitates business processes



Satisfy Privacy regulations

- String literal values
- Character substrings & concatenation
- Random or sequential numbers

Reduce risk of data breaches

- Arithmetic expressions
- Lookup values
- Business data types (CCN, NID)

Maintain value of test data

- Generic mask
- Dates
- User defined

Patient Information			
Patient No.	123456	SSN	333-22-4444
Name	Erica Schafer		
Address	12 Murray Court		
City	Austin	State	TX Zip 78704

Data is masked with contextually correct data to preserve integrity of data

Personal Info Table		
PersNbr	FirstName	LastName
10000	Jeanne	Renoir
10001	Claude	Monet
10002	Pablo	Picasso
	⋮	

Referential integrity is maintained with key propagation

Event Table		
PersNbr	FstNEvtOwn	LstNEvtOwn
10002	Pablo	Picasso
10002	Pablo	Picasso

Gartner Magic Quadrant - Data Masking



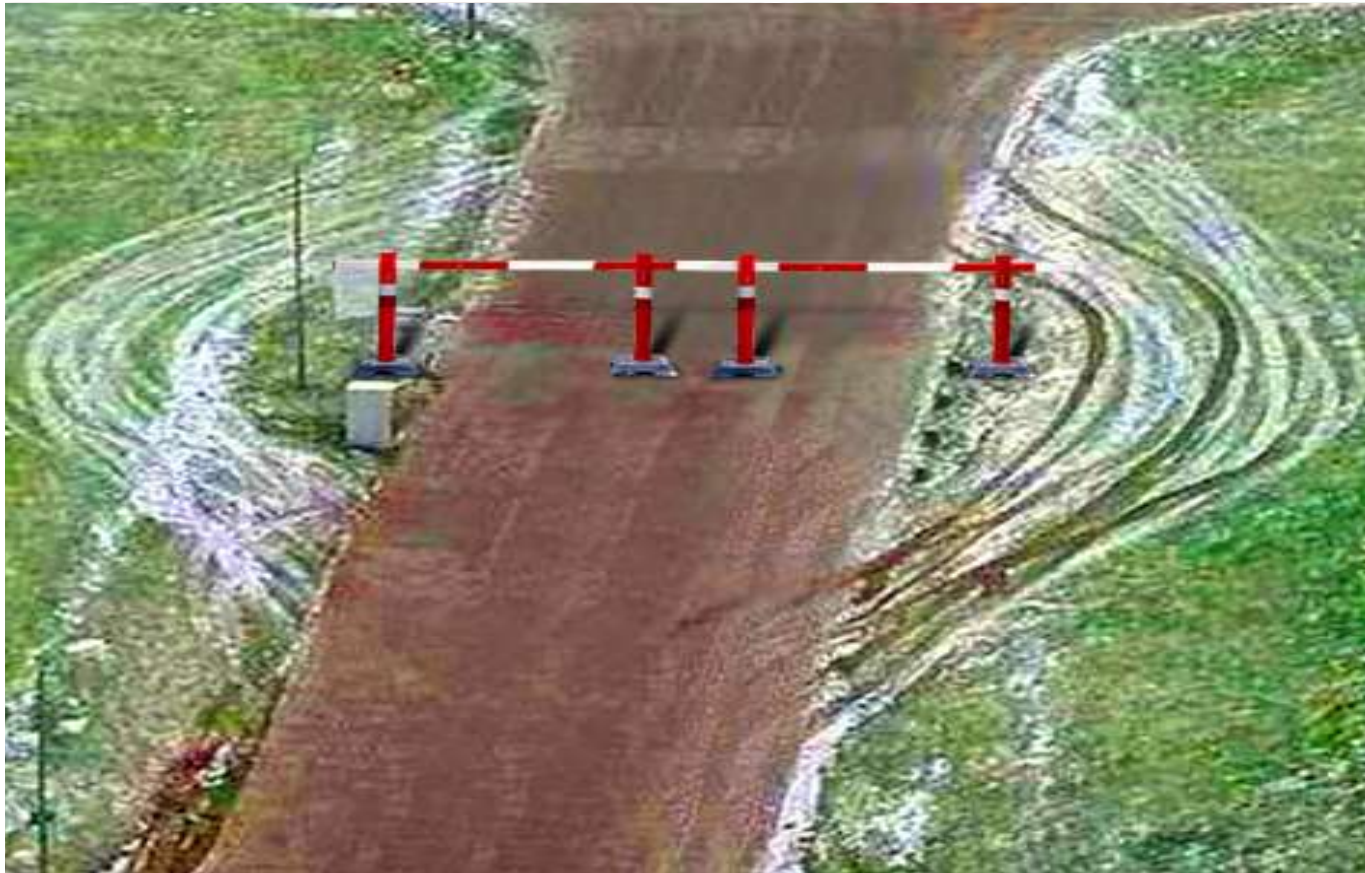
Gartner MQ for Data Masking Technology



The Magic Quadrant is copyrighted 2012 by Gartner, Inc. and is reused with permission. The Magic Quadrant is a graphical representation of a marketplace at and for a specific time period. It depicts Gartner's analysis of how certain vendors measure against criteria for that marketplace, as defined by Gartner. Gartner does not endorse any vendor, product or service depicted in the Magic Quadrant, and does not advise technology users to select only those vendors placed in the "Leaders" quadrant. The Magic Quadrant is intended solely as a research tool, and is not meant to be a specific guide to action. Gartner disclaims all warranties, express or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

Are there ways around your security policies?

Requirements for data security and compliance



The Choice of Market Leaders

JPMORGAN CHASE & CO.	HSBC 	 GE Money Bank	ICBC 	
 BNP PARIBAS			Allianz 	
 Nestlé	 vodafone			 SAMSUNG ENGINEERING
	ING 		 SOCIETE GENERALE Corporate & Investment Banking	
 Rosenberg An AXA Investment Managers Company		 Bank of Tokyo-Mitsubishi UFJ		

The Choice of Government

Questions

Innovate2013
The IBM Technical Summit

