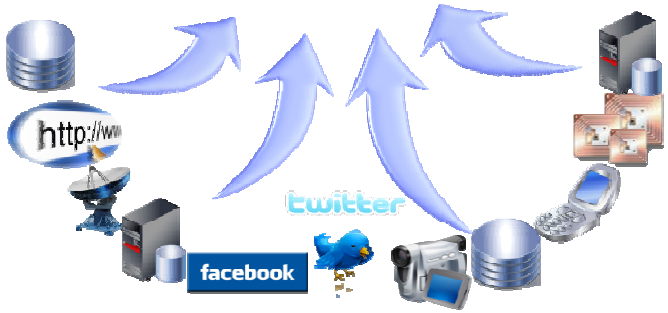
Monitorización de actividades de datos InfoSphere® Guardium® para Big Data

Amal Mashlab
IBM Software Group, Information Management
Responsable de Ventas de Gobierno de la Información para Europa

Cumplir con los constantes cambios de la normativa global y sectorial

Esta gran oportunidad conlleva grandes riesgos para la seguridad

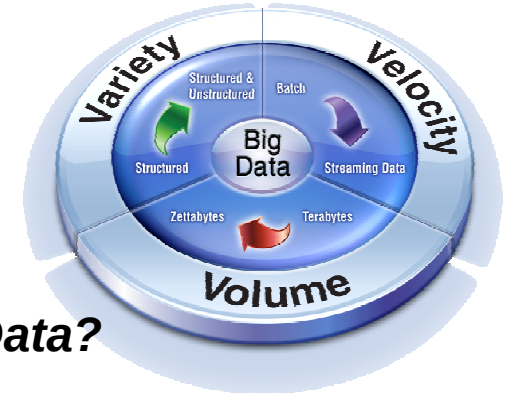
¿Qué es Big Data?



- Toda clase de datos
- Grandes volúmenes
- Conocimiento valioso pero difícil de extraer
- El tiempo suele ser de gran importancia

Oportunidad

Extracción de conocimientos de un enorme volumen, variedad y velocidad de datos a gran velocidad en el momento oportuno y de modo rentable.



¿Qué clase de información se almacena en Big Data?

¿Cómo se consigue la conformidad?

La seguridad es vital en los despliegues empresariales de Big Data

Si no es seguro, será aislado.

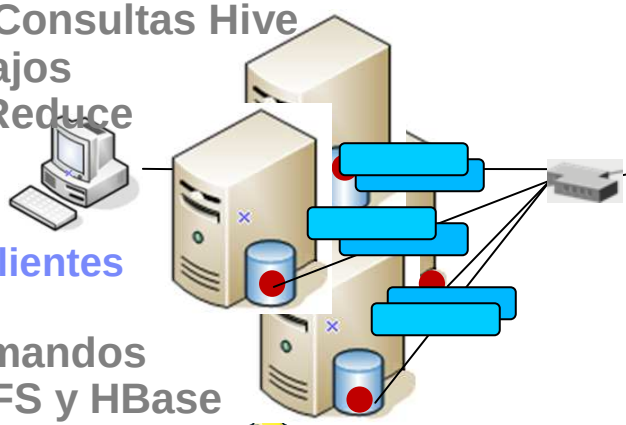
Presentamos Hadoop Activity Monitoring de InfoSphere Guardium

- Impacto mínimo para los recursos de servidores Big Data o la red
- Separación de tareas: datos auditados en un equipo seguro
- Compatibilidad heterogénea (IBM, Cloudera)

Los mensajes relevantes son copiados y enviados al colector

S-TAP InfoSphere Guardium

Consultas Hive
Trabajos MapReduce



Clientes

Comandos HDFS y HBase

hadoop Clúster

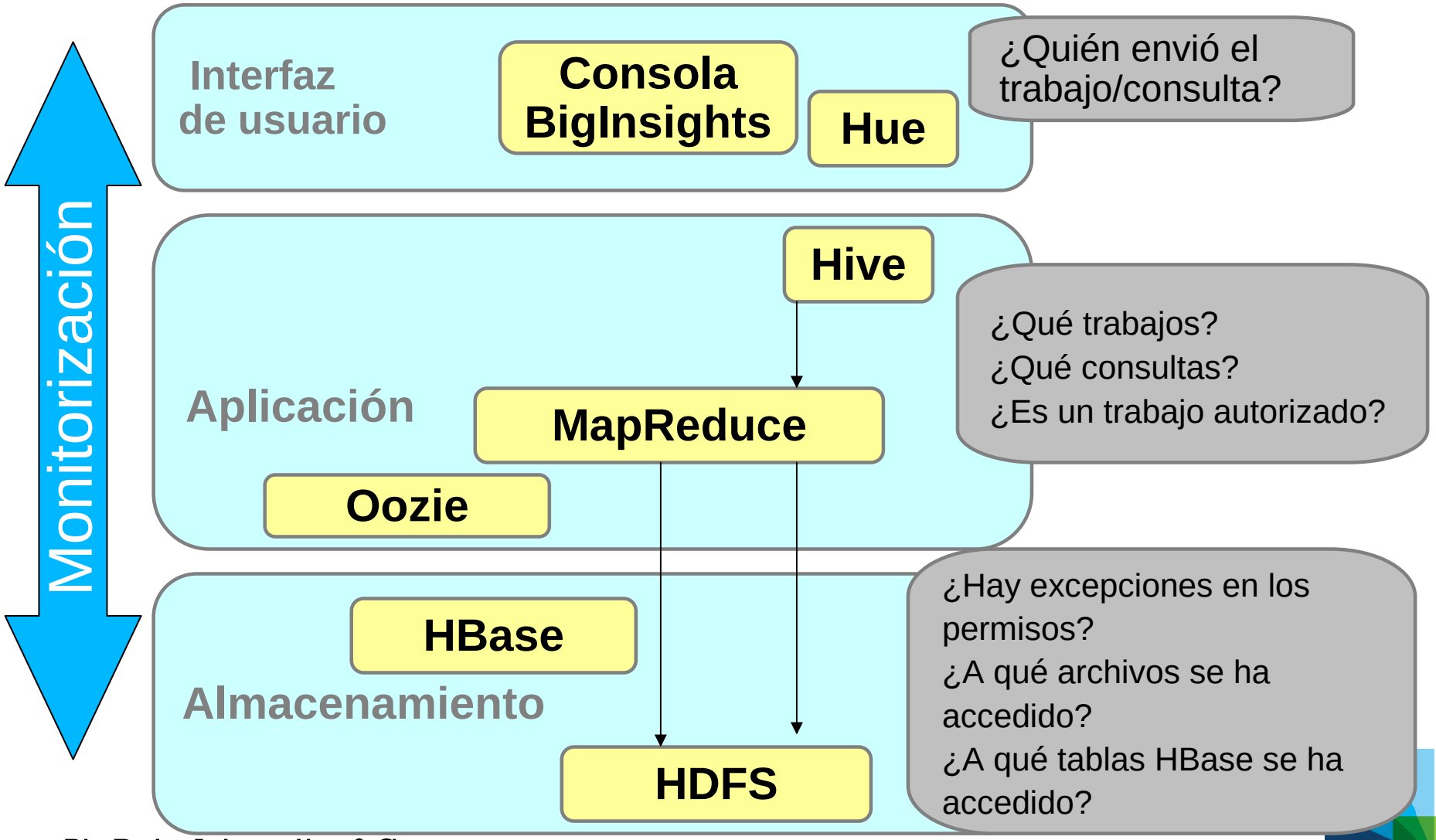
Colector InfoSphere Guardium

¡Alarma de datos confidenciales!

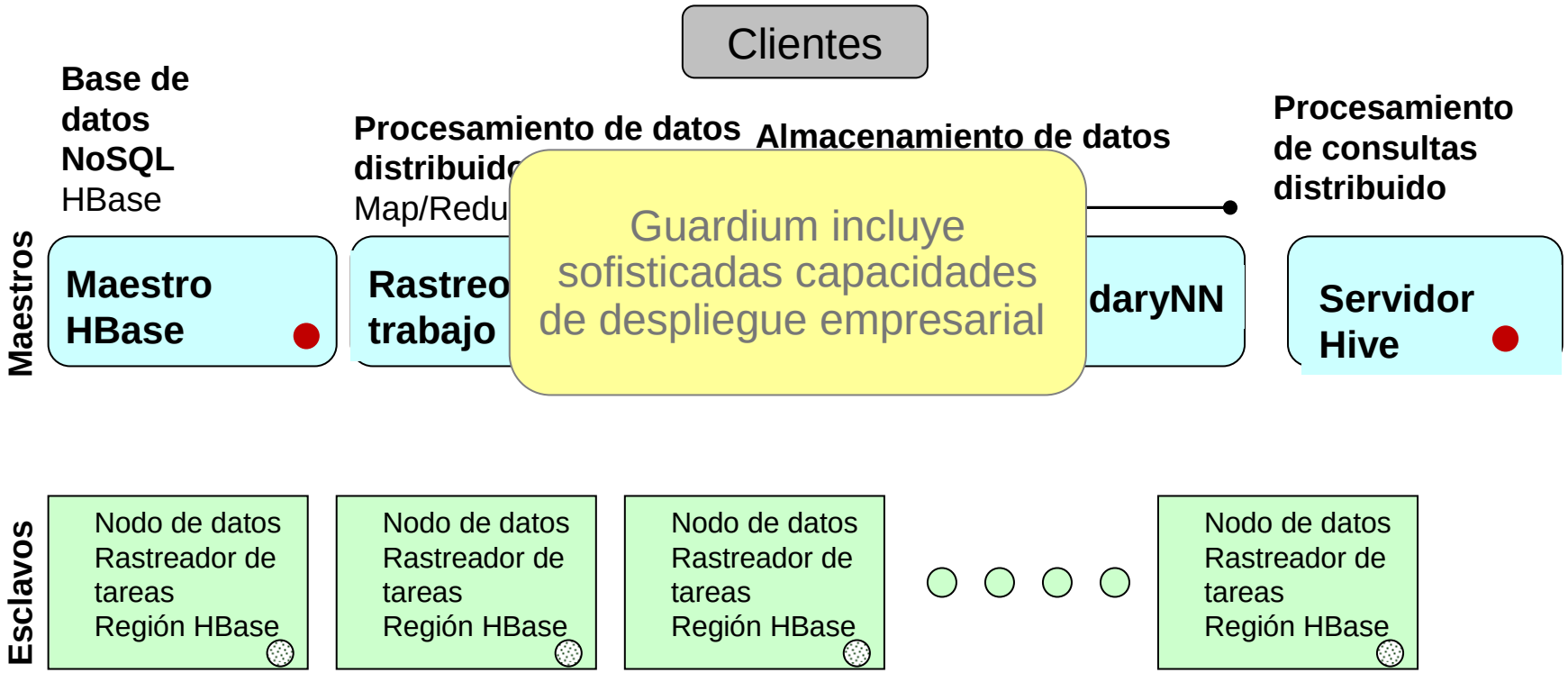
Access Rule Description	Client IP	Server IP	DB User Name	Full SQL String
Access Rule: sensitive files: Alert	9.70.145.118	9.70.145.113	SVORUGA	__WGPB message {struct:1='getFileInfo',struct:2= {struct:1='/user/svoruga/testme'}, struct:3='org.apache.hadoop.hdfs.protocol.ClientProtocol',varint:4=1}
Access Rule: sensitive files: Alert	9.70.145.118	9.70.145.113	SVORUGA	__WGPB message {struct:1='getListing',struct:2= {struct:1='/user/svoruga/testme',struct:2='', varint:3=0},struct:3='org.apache.hadoop.hdfs.protocol.ClientProtocol',varint:4=1}

Informes y alertas InfoSphere Guardium

Monitorizar quién, qué y cómo

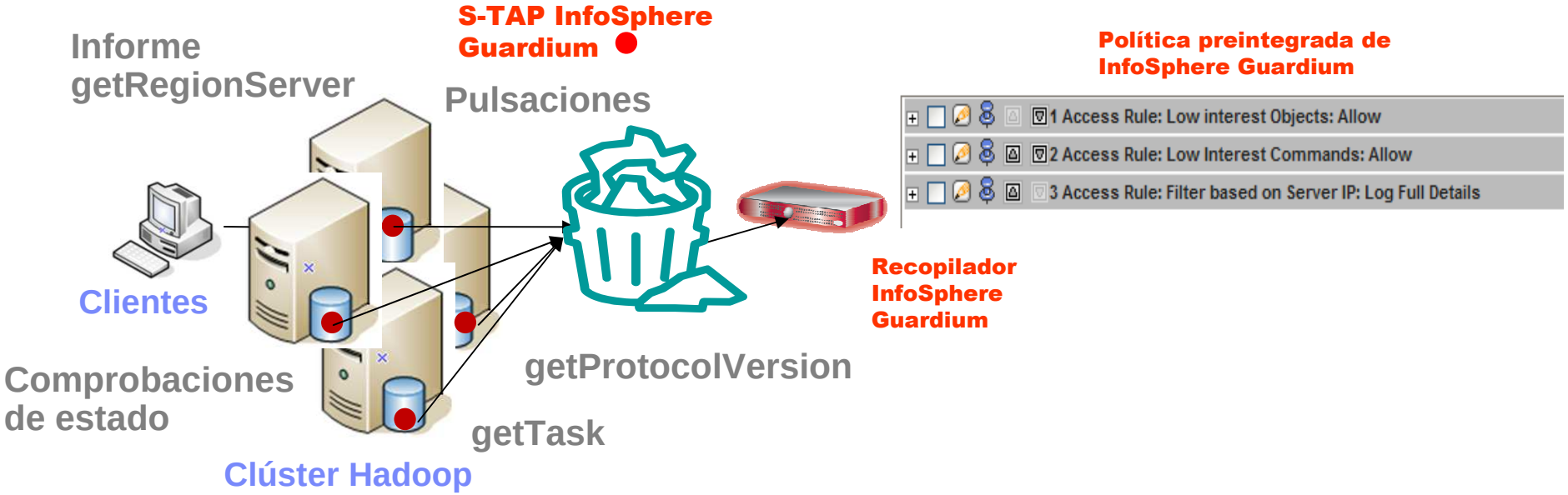


Ubicación de S-TAP en el clúster



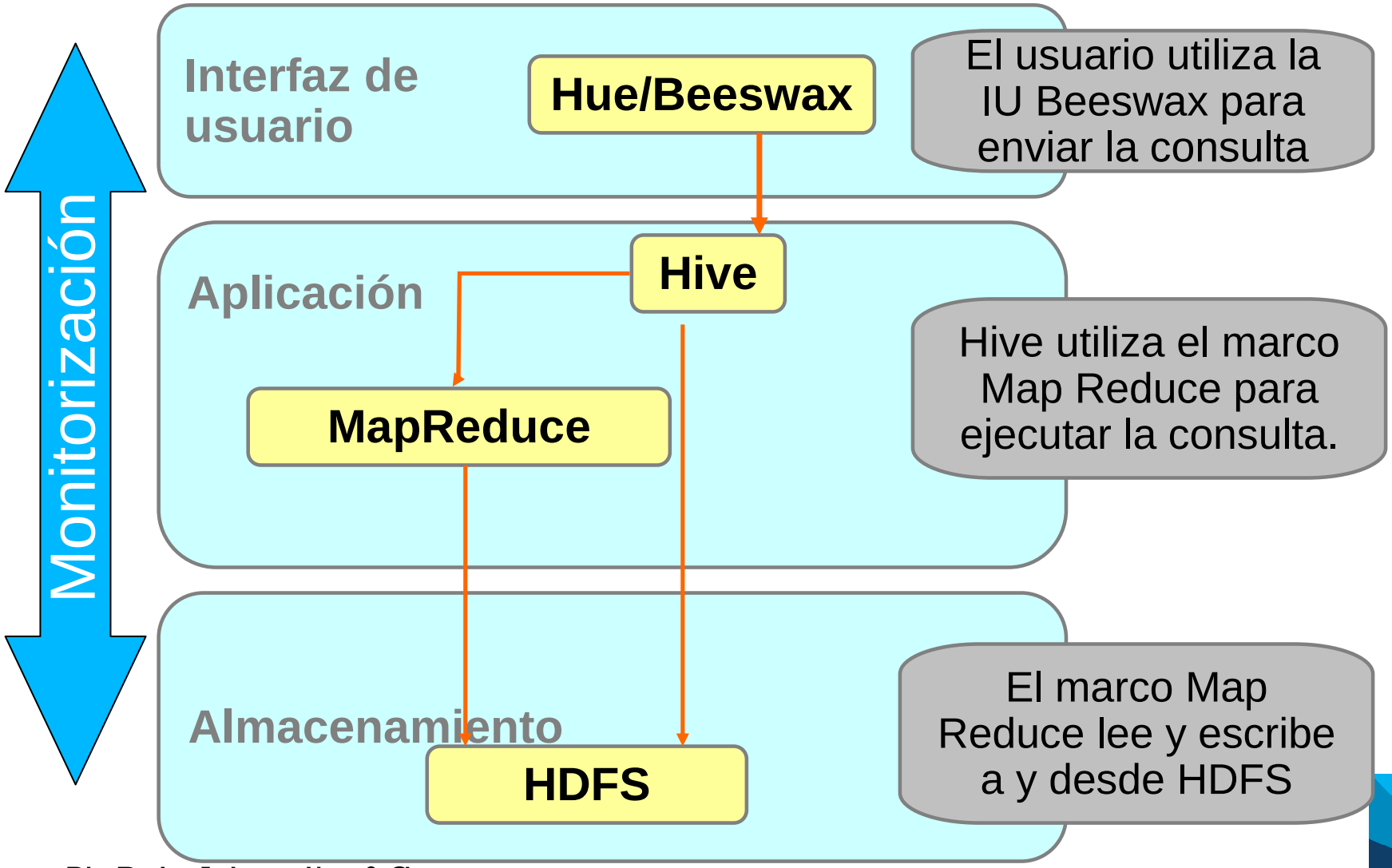
- S-TAP
- S-TAP opcional requerido únicamente para monitorizar comandos Put HBase
- Instalación centralizada y actualización de S-TAP
- Proceso, integración y scripts de automatización

Control basado en políticas para cumplir requisitos de auditoría detallados



- Política en Hadoop para filtrar el tráfico innecesario y reducir el ruido
- Existen otras políticas de auditoría predefinidas para ayudar a cumplir los requisitos de auditoría de PCI, SOX, etc.

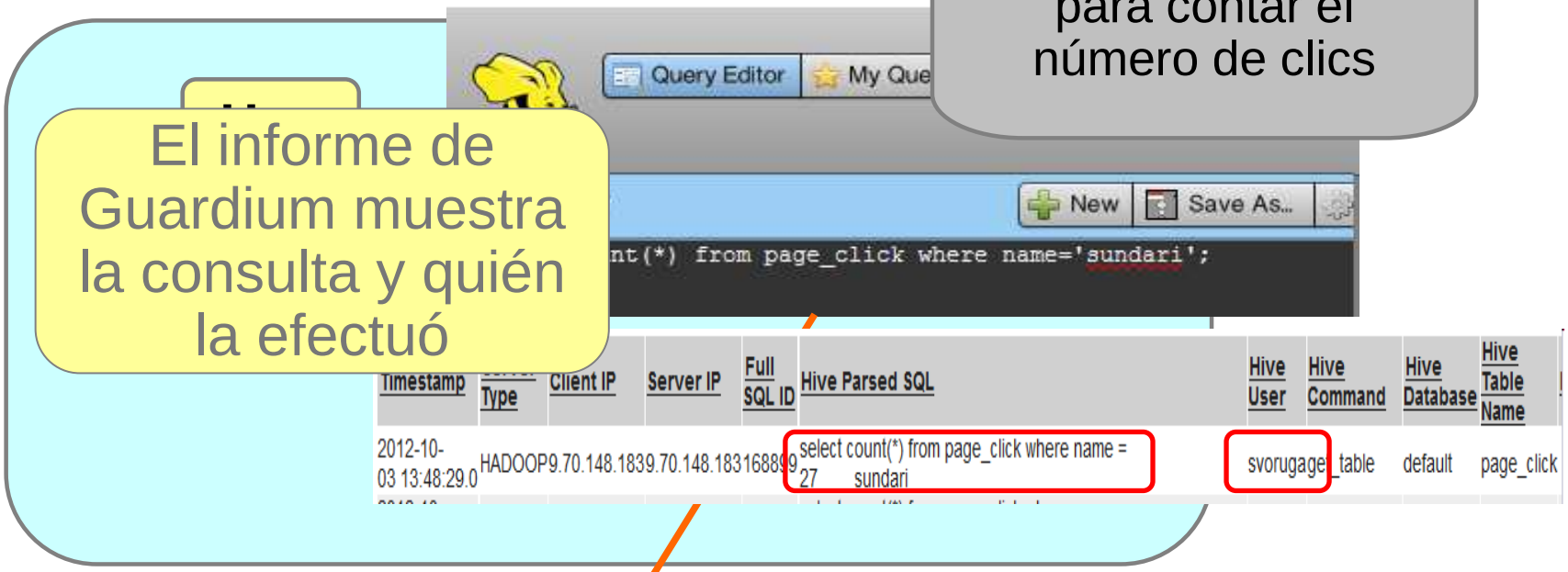
Ejemplo con Cloudera: Monitorización de una consulta Hive



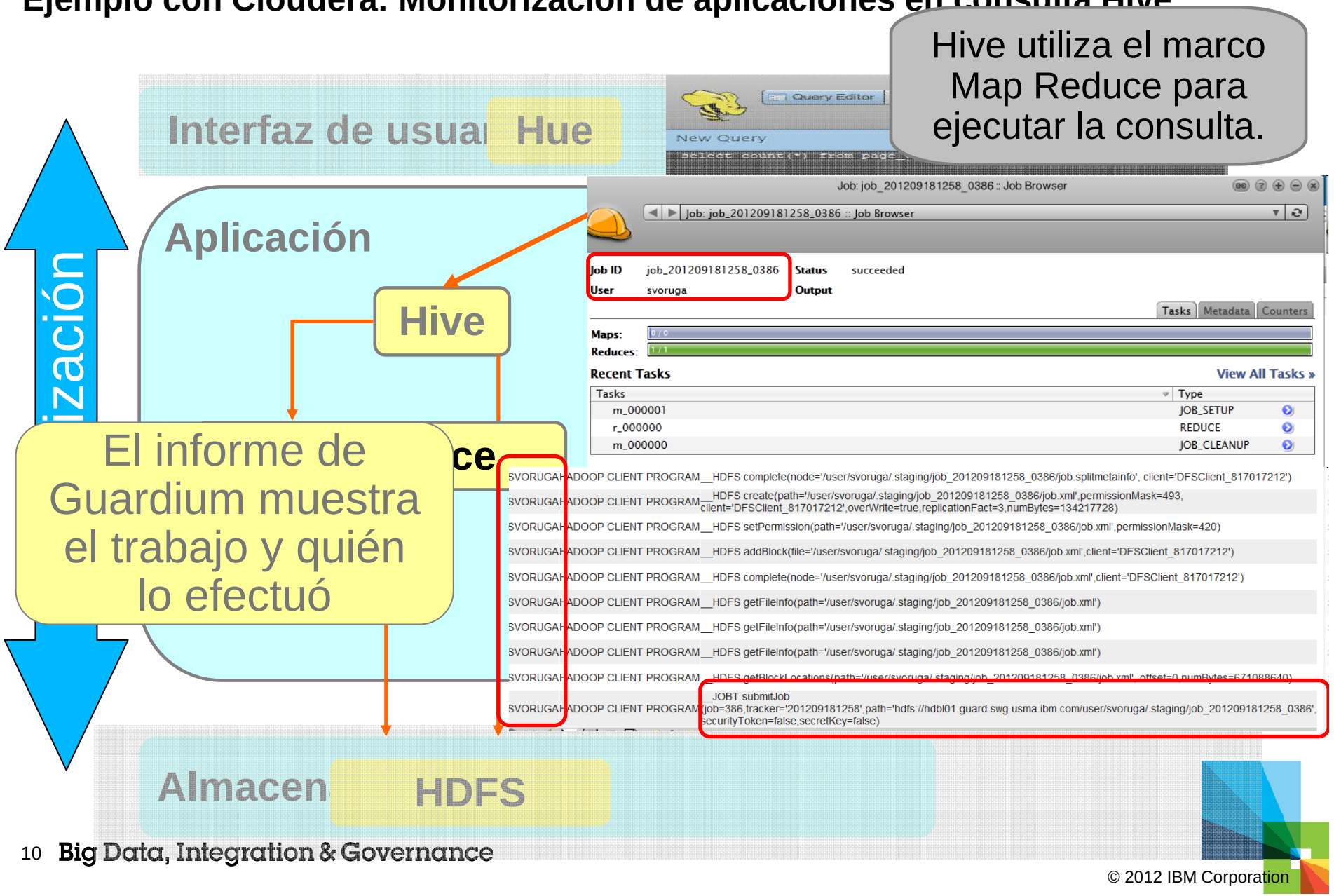
Ejemplo con Cloudera: Monitorización de la interfaz de usuario en consulta Hive

Envío de la consulta para contar el número de clics

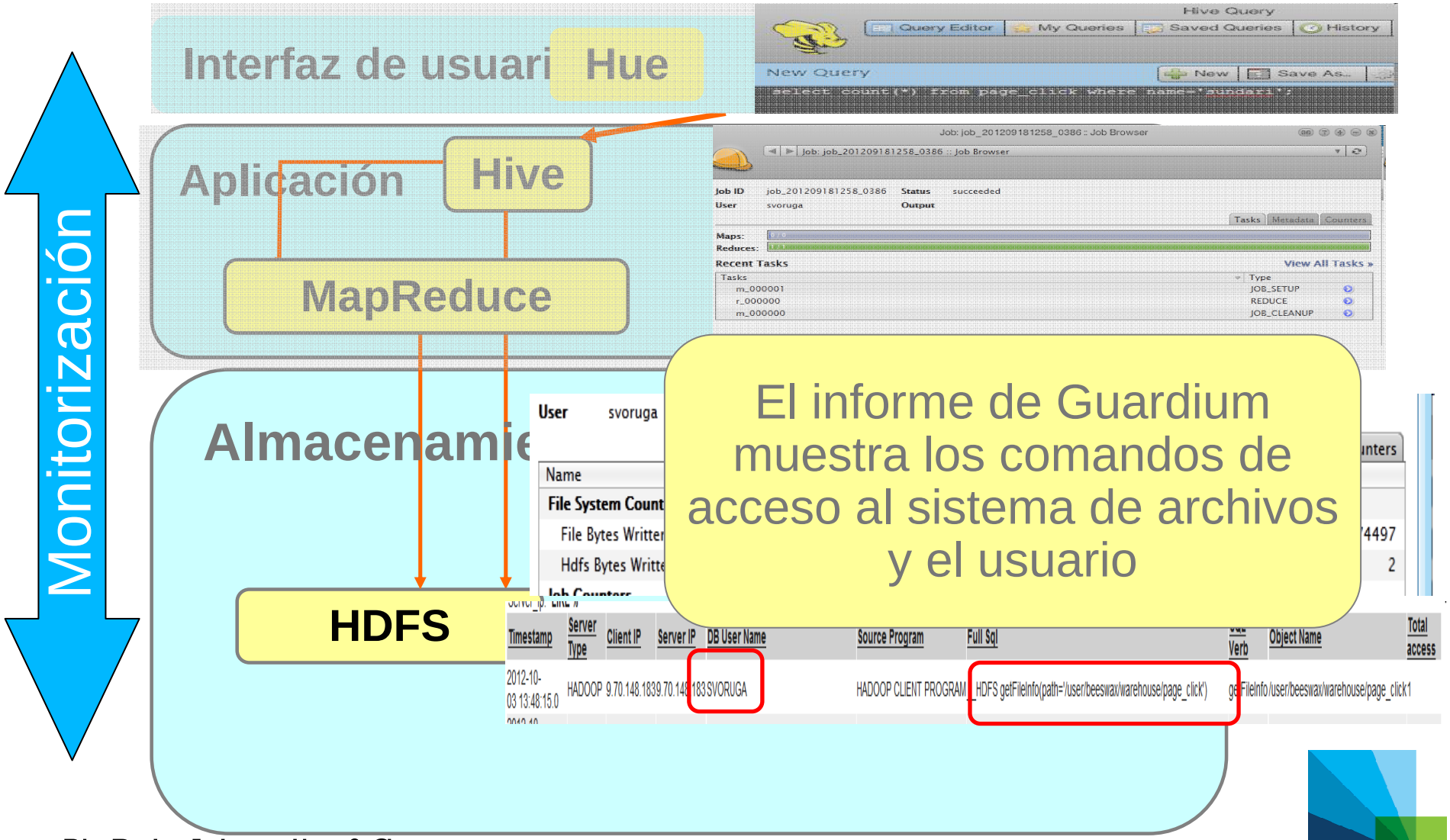
El informe de Guardium muestra la consulta y quién la efectuó



Ejemplo con Cloudera: Monitorización de aplicaciones en consulta Hive



Ejemplo con Cloudera: Monitorización del almacenamiento en consulta Hive



Ejemplo con IBM InfoSphere BigInsights: Monitorización

Ejecución del trabajo de análisis de la tarjeta de crédito

Monitorización

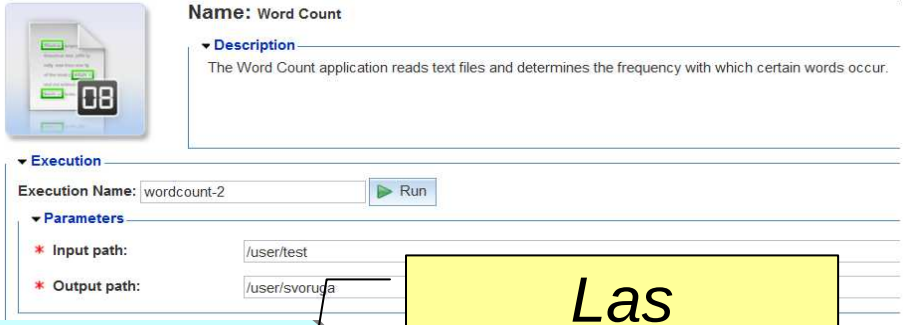
Consola BigInsights

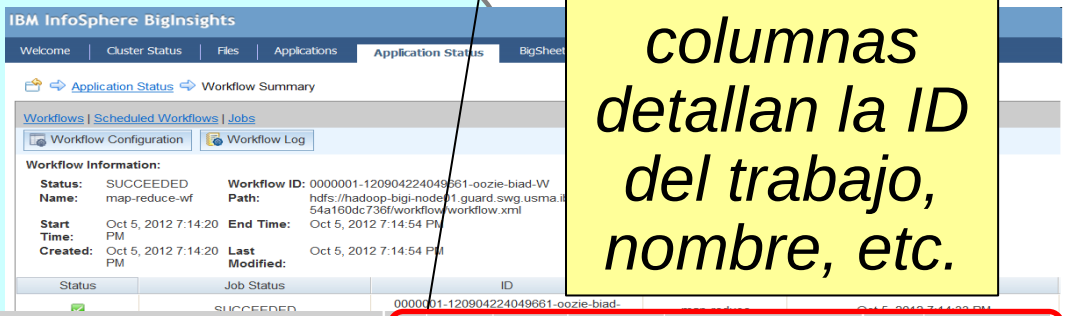
Mensaje detallado con información de permisos y del grupo

Oozie

Almacen

HDFS

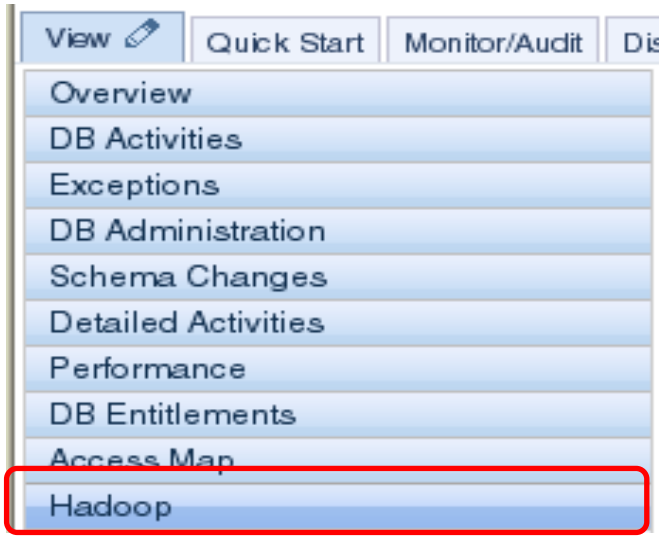


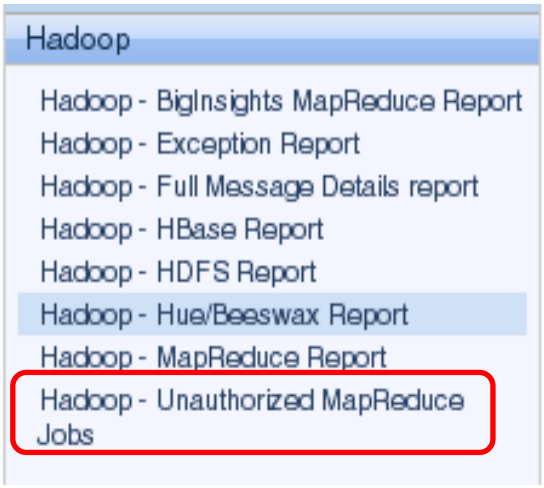


Las columnas detallan la ID del trabajo, nombre, etc.

Full Sql	DB User Name	Source Program	Application User	BI Job ID	BI User Name	BI Job Name
create source=/user/biadmin/oozie-biad/0000000-121005223302938-oozie-biad-W/wordcount--map-reduce/output/_SUCCESS, permissions=biadmin:supergroup:rw-r--r--	BIADMIN		:PU=biadmin		biadmin	wordcount
create source=/user1/aa/_logs/history/job_201210052231_0002_1349490980576_biadmin_oozie%3Aaction%3AT%3Dmap-reduce%3AW%3Dmap-reduce-wf%3AA%3Dwordco, permissions=biadmin:supergroup:rw-r--r--	BIADMIN		:PU=biadmin	job_201210052231_0002	biadmin	job_201210052231_0002_1349490980576_biadmin_oozie%3Aaction%3AT%3Dmap-reduce%3AW%3Dmap-reduce-wf%3AA%3Dwordco,
create source=/user/biadmin/oozie-biad/0000000-121005223302938-oozie-biad-W/wordcount--map-reduce/newId.properties, permissions=biadmin:supergroup:rw-r--r--	BIADMIN		:PU=biadmin		biadmin	wordcount
create source=/user/biadmin/oozie-biad/0000000-121005223302938-oozie-biad-W/wordcount--map-reduce/0000000-121005223302938-oozie-biad-W@wordcount@0, permissions=biadmin:supergroup:rw-r--r--	BIADMIN		:PU=biadmin		biadmin	wordcount

Informe: Identificación de trabajos no autorizados





Informes predefinidos y personalizables

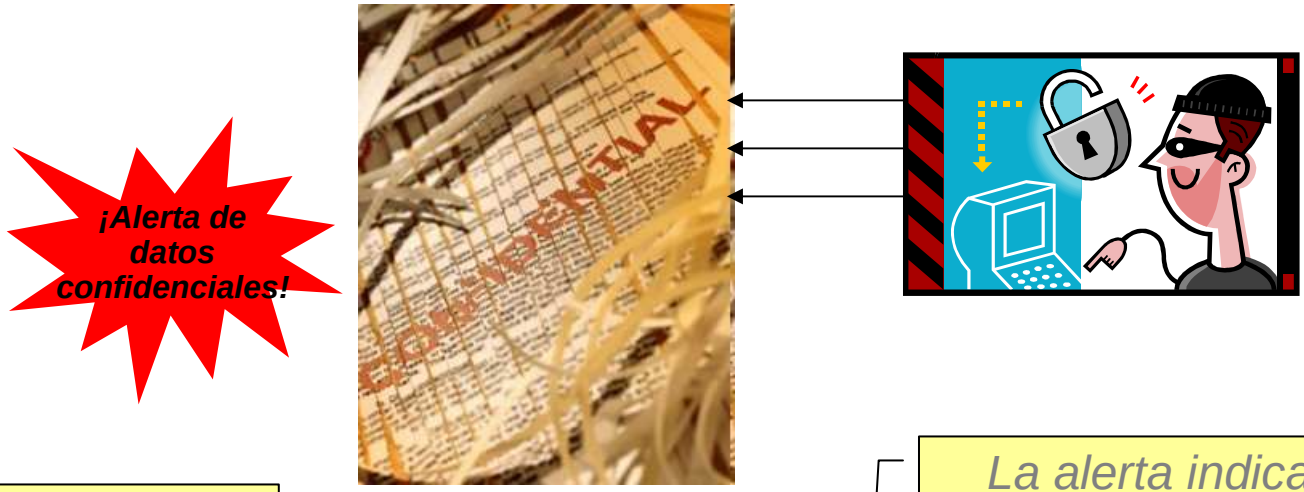
¡Programa no autorizado!

Hadoop - Unauthorized MapReduce Jobs

Start Date: 2012-09-05 11:06:20 End Date: 2012-09-06 14:06:20
Aliases: OFF Sql: LIKE %
UserName: LIKE %authusr% name: LIKE %

Timestamp	Server Type	Client IP	Server IP	Full Sql	OS User	DB User Name	MapReduce User	MapReduce Name	MapReduce Job	Source Program
2012-09-05 14:01:26.0	HADOOP	9.70.146.211	9.70.146.211	__HDFS complete (node='/user/data/output_unauth/_logs/history/0.0.0.0_1345143729831_job_201208161502_0018_authusr_hack', client='DFSClient_-881434930')		AUTHUSR	authusr	hack	ob_201208161502_0018	HADOOP CLIENT PROGRAM
2012-09-05 14:01:10.0	HADOOP	9.70.146.211	9.70.146.211	__HDFS create (path='/user/data/output_unauth/_logs/history/0.0.0.0_1345143729831_job_201208161502_0018_authusr_hack', permissionMask=493,client='DFSClient_-881434930',overWrite=true,replicationFact=1, numBytes=67108864)		AUTHUSR	authusr	hack	ob_201208161502_0018	HADOOP CLIENT PROGRAM

Alarma: Acceso no autorizado a datos confidenciales



La alerta muestra quién ha intentado acceder a los datos confidenciales

La alerta indica el directorio al que se ha accedido

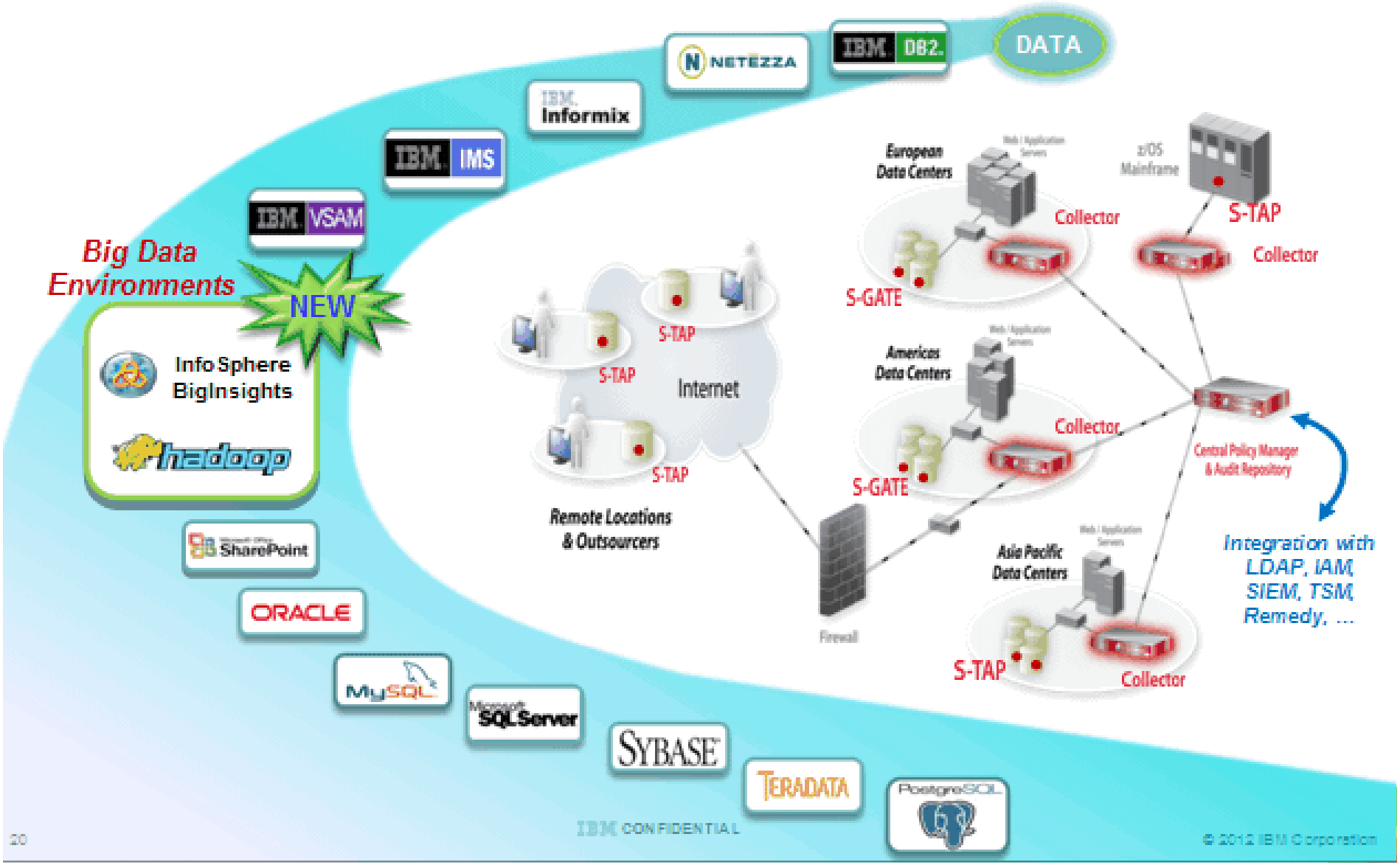
System View | Admini | Monitor | Tap Monitor | Incident Management | My New Reports

Policy Violations / Incid

Start Date: **2012-08-29 14:01:55** End Date: **2012-09-06 14:01:55**
Aliases: **OFF**

Violation Log Id	Timestamp	Category Name	Access Rule Description	Client IP	Server IP	DB User Name	Full SQL String
8	2012-09-05 14:06:28.0		Access Rule: sensitive files: Alert	9.70.146.211	9.70.146.21	UNAUTHUSR	HDFS getFileInfo(path='/user/data/customer.data')
9	2012-09-05 14:06:28.0		Access Rule: sensitive files: Alert	9.70.146.211	9.70.146.21	UNAUTHUSR	HDFS getBlockLocations(path='/user/data/customer.data',offset=0,numBytes=671088640)

Resumen: Monitorización de actividad Hadoop en InfoSphere Guardium



Objetivos finales de la seguridad en entornos Big Data

1

Prevenga las fugas de datos

Prevenga la revelación o la filtración de datos confidenciales

2

Garantice la integridad de los datos confidenciales

Prevenga cambios no autorizados en datos, estructuras de bases de datos, archivos de configuración y registros

3

Reduzca el coste de la conformidad

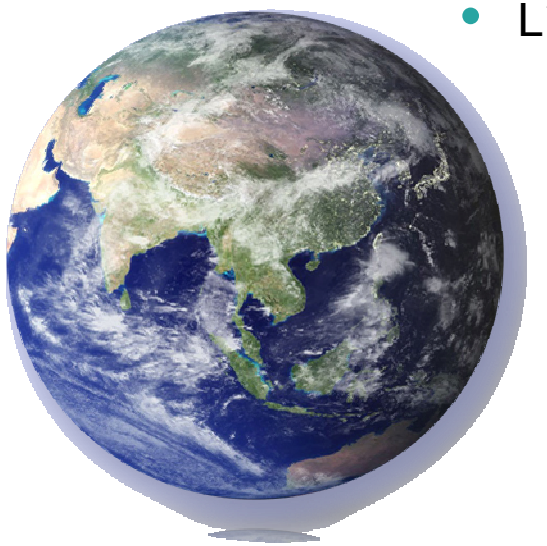
Automatice y centralice los controles

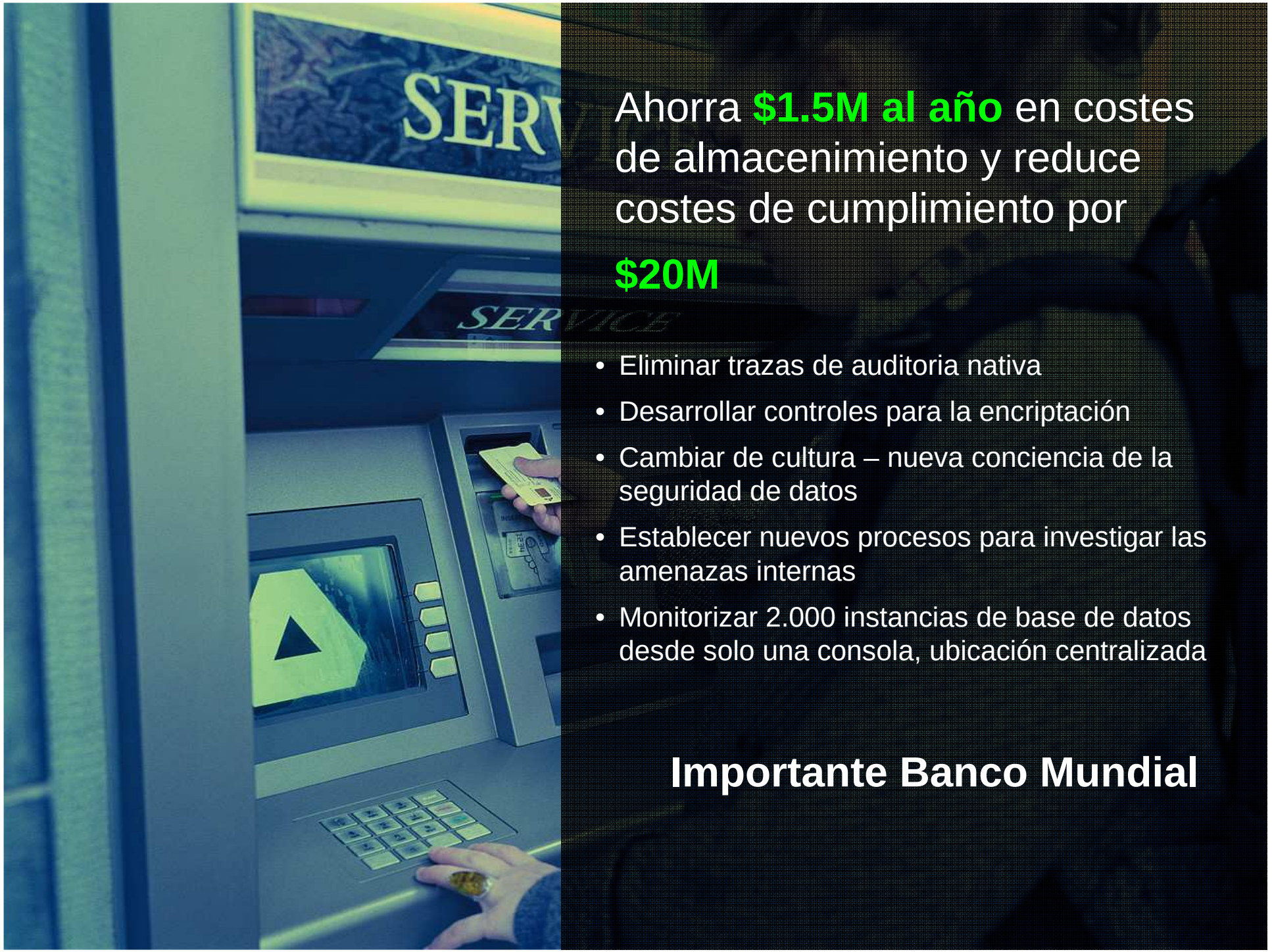
- En distintas normativas, como PCI DSS, normas sobre privacidad de datos, HIPAA/HITECH, etc.
- En entornos heterogéneos, como bases de datos, aplicaciones, data warehouses y plataformas Big Data como Hadoop

Simplifique los procesos de revisión para auditoría

Líderes en diversos sectores – Referencias de Clientes

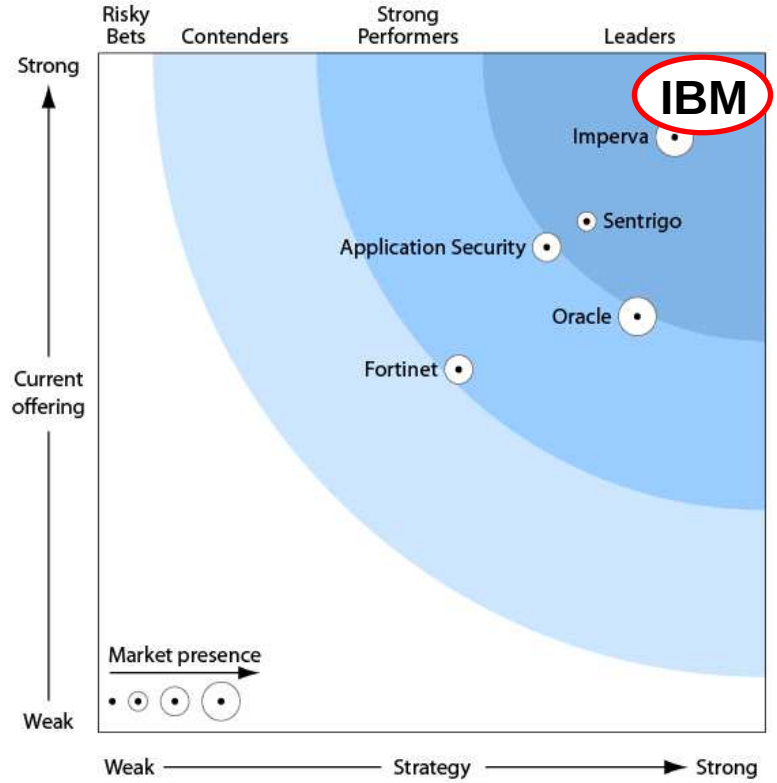
- 9 de los 10 Bancos globales
- 2 de las 3 empresas de Retail globales
- 5 de las 6 aseguradoras globales
- 3 de las empresas de refrescos globales
- La marca más reconocida de PCs
- 9 de las 10 empresas de telco
- Empresas de Gobierno
- Las 3 automotrices
- #1 dedicada a Seguridad
- Proveedores de Energía
- Proveedores de Salud
- Media & entretenimiento
- Líneas aéreas internacionales



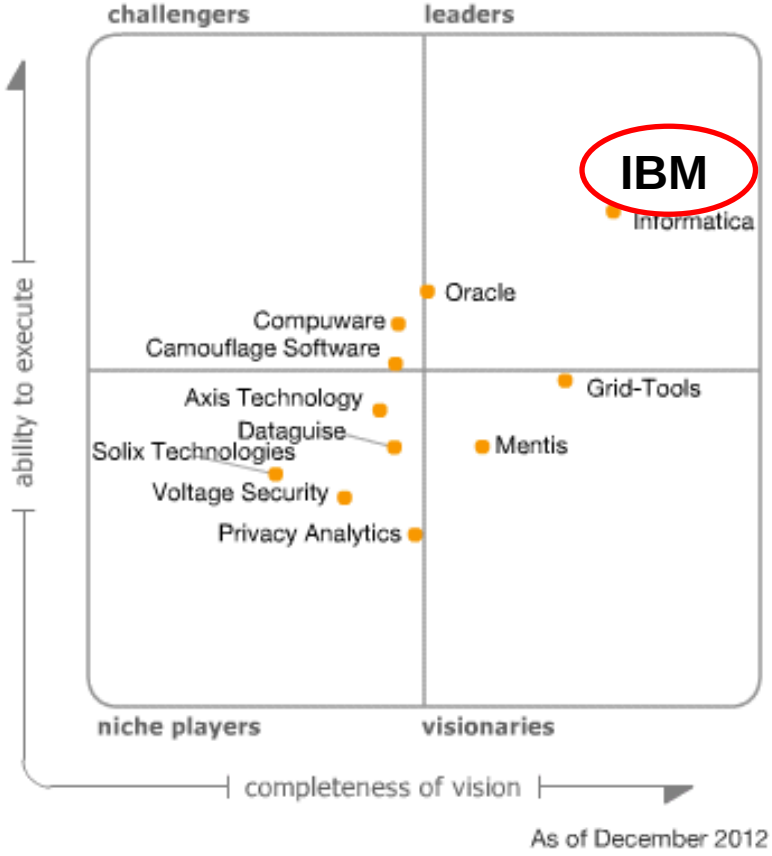

Ahorra $1.5M al año en costes de almacenimiento y reduce costes de cumplimiento por
$20M
• Eliminar trazas de auditoria nativa
• Desarrollar controles para la encriptación
• Cambiar de cultura – nueva conciencia de la seguridad de datos
• Establecer nuevos procesos para investigar las amenazas internas
• Monitorizar 2.000 instancias de base de datos desde solo una consola, ubicación centralizada
Importante Banco Mundial

InfoSphere Guardium e InfoSphere Optim son nº1

Forrester Wave™: Database Auditing And Real-Time Protection, Q2 '11



Gartner MQ for Data Masking Technology




PIENSE
BIG

Muchas Gracias

Big data, integración y gobierno: www.ibm.com/software/es/info/rte/bdig/

@ibmanalytics_es