International Journal of Control and Automation
Vol. 13, No.1, (2020), pp. 11 - 21

The Analytical Study on Enterprise/Public Information Protection with Big Data: Challenges and Impact on Data

Mr. Shailesh Gahane, Ms. Mohini Patil, Dr. Onkar Kemkar
Research Scholar, Sandip University Nashik

Abstract:

Today, we have moved on at pace to web and mobile client enabled multitier computing, business process management and service oriented architecture. But still the problem of distributed data remains. If anything it is getting worse, Cloud computing is one of the tools for us, meaning that business application systems now exist outside the enterpriselevel as well as in public area. Multiple data warehouses or service providers have also emerged as a line of business thathas taken hold of business intelligence and analytics. Data warehouse applications have emerged, creating more data stores yet, and for many, master and reference data is still not under control.

The wave of cloud computing is Big Data, which is widespread in the industry by storm. Given the difficulty already upon us in managing data and managing information security in a distributed computing environment, what then is the impact of big data on the enterprise? This is an important question in front of all of us. What is big data, and how much of a challenge does it pose to the already overextended need to enforce enterprise information protection?

In this paper, we have first define what enterprise information security and privacy involves, then looking at necessities that need to be met before introducing big data and identifying what impact this has on those necessities.

Keywords: *Multitier Computing, Business Process Management, Service Oriented Architecture, Cloud Computing, Data Warehouse, Big Data, Enterprise Information Protection, Information Security, Data Transmission, Mobile Application.*

I. Introduction about Enterprise Information/Data Protection:

In the late 1980's and early 1990's, we saw a fundamental shift in computing when companies began to move faraway from centralized systems towards client-server computing and distributed systems. However, what happened wasn't such a lot of distributed systems but more like 'standalone' autonomous systems. Each system was intended to support a particular business purpose and was contained of an application deployed on its servers with its data. Thanks to the doorway of multiple servers within the enterprise produced a replacement problem that the way to manage this new, more large ,and sophisticated environment. This resulted in agent-based distributed systems management software emerging to assist systems administrators in using the facility of the network to manage multiple servers across the organization.

While this helped people manage and monitor multiple systems, one 'sideeffect' of the increase of distributed computing that wasn't well addressed was that the matter of distributed data. The impact of distributed data was significant therein data became difficult to share, keep consistent, and synchronized, especially as users of the various applications began to maintain data in several application-specific databases. The result was that data became far more difficult to manage. It needed to flow between systems to implement business processes. File transfers speeded, and demand grew for data to be combined with supporting cross-functional management reporting and analysis. The

bunch of jobs grew at a high rate moving data between systems, then the age of the info warehouse was born.

The information protection challenge has been expanded in the areas ofrisk prevention andchange management to show the overall complexitythat most enterprises are facing when it comes to undertaking enterpriseinformation protection. It is without hesitation a scary problem and wellworthy of a devoted team to deal with it.

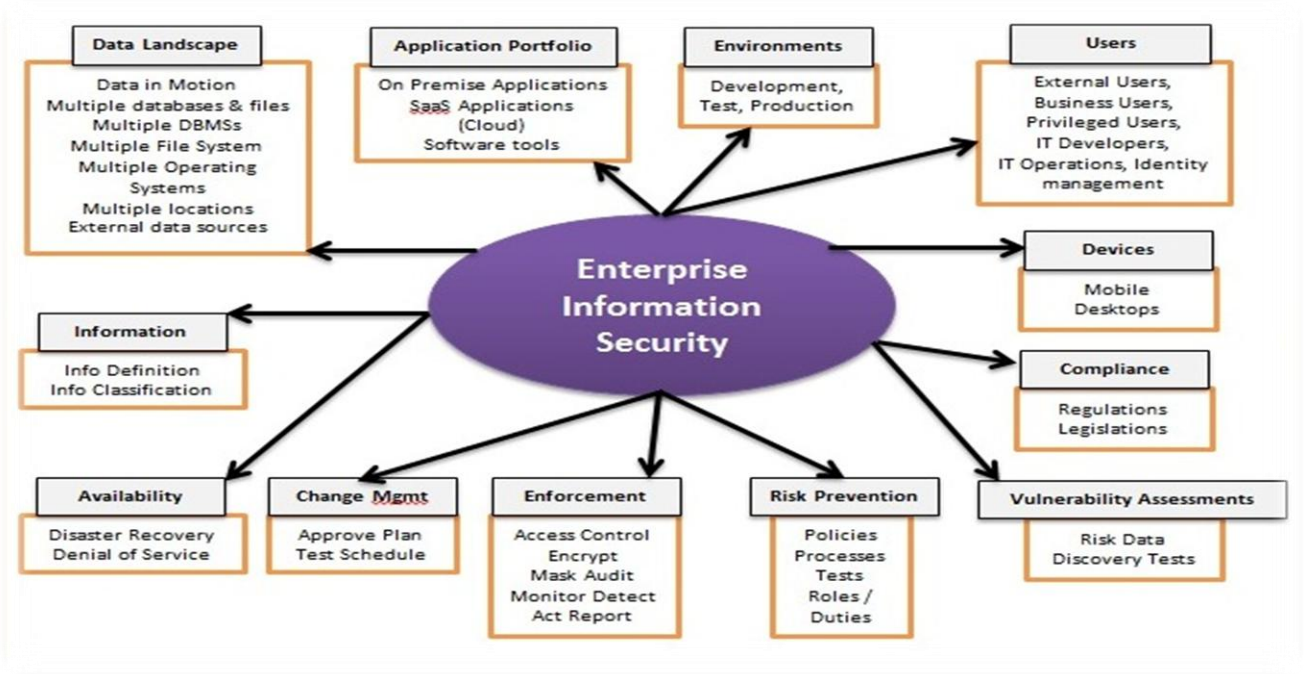


Figure 1: Source: Intelligent Business Strategies

Looking at above Figure 1 in a clockwise direction from the box entitledinformation, tackling the challenge of enterprise information protectioninvolves:

- Classifying information to define what is sensitive and what is not.
- Understanding the current data landscape to define where thatcomplex information is located in order to safeguard it.
- Understanding the present application portfolio and the informationthese applications access and conserve.
- Ensuring enterprise information protection covers multipleenvironments.
- Accounting for altered types of users and the devices they picked toaccess applications andinformation with.
- Upholding compliance regulations and legislation.
- Assessing vulnerability to information security/privacy breaches.
- Describingguidelines, procedures, characters, tests and actions needed toprevent breaches in security.
- Applying those rules to impose information security.
- Managing change so that information security is not compromisedby changes made.
- Maintaining information availability.

In addition to this, imposing enterprise information security and privacy alsocontains integration with other substructure software. In order for enterprise information protection to be implemented, meansdefining what information has to be protected and then locating that data tobe able to secure it. For most organizations, that data is stored indatabases and files scattered across a highly distributed landscape ofmultiple DBMS and file systems that run on a range of operating systemson servers in multiple locations. This contains both on-premise and cloudbased stages. The fact that complex data could be extensively distributedacross this landscape rises the risks of security being bargained.Finding, controlling access to andprotecting sensitive data content in thiskind of environment is a real challenge without the right tools to help you.In terms of users, identity management, user risk classification,authentication, authorization and multi-device access security are all verymuch part of an enterprise information protection initiative to control useraccess. Both desktop and mobile devices need to be considered with theaddedproblem that mobile devices can be easily lost or whipped. Typesof user are also important when it comes to information access and authorization. This would include outdoor users such as Partners, Clientsand Suppliers, Internal Business Users and IT Specialists. SomeIT professionals are privileged in that they have administrative power thatallows them access to potentially any data including sensitive customer,employee and financial data. For this reason, privileged IT professionalsthemselves need to be monitored and duties separated to control what they can and cannot do without authorization.

IT designers who build systems and IT operations staff who manage andrun those systems also need to be taken into description. IT designers oftenwork with constructionof data during development andtesting. Thereforeinformation has to be protected in development, testing and productionenvironments.Also information constructed from accessing complex information may itselfcontain subsets of that complex information.

Enterprise/Public information security has to be universal, covering all bases toavoid information risks that might breach regulation, cause non-compliancewith regulations or adversely impact the organization's ability to meet itsown business purposes. It involves being able to classify and locate sensitive data, assess the vulnerability of the organization to potentialbreaches in security, implement prevention measures to avoid putting dataat risk, monitor events that may signal a problem and respond in a timelymanner to minimize the impact of these events when they occur. The Accesscontrol, sensitive and confidential data masking and network encryption / decryption are central to it. Companies need to identify what the information risks are and whatcontrols are in place to secure and protect information to reduce theserisks. Thesecontrols may be in the form of access approval processes,data masking and encryption policies,auditing, backup policies, retention policies and other checks and balances. If a violation occurs, then adamage limitation process is needed to manage losses and managechanges to procedures to avoid the same thing happening again. Alsothere need to be process in place to re-test security when changes aremade so that risk exposure is not increased as a result of the changes.

II. Requirements for Enterprise Information Protection:

Having defined what enterprise information protection is and looked at where organizations are in terms of tackling this challenge, the nextquestion to ask is "What are the requirements for enterprise informationprotection?".

These necessities need to cover the complete range ofattentions shown Figure 1 to prevent safety breaches. It is worth adding additionalrequirements to that list to cover everything shown in Figure 1 in thispaper that is not covered in the aforementioned paper. These supplementary requirementsare listed under the box headlines shown in Figure 1 for convenience.

III. Information Classification Requirements:

I. Structured Information

- It should be possible to define common data definitions forall master data (e.g. customer, product, asset, site, supplieretc.), reference data (code sets), transaction types,hierarchies and metrics in a business glossary and thenclassify what data is sensitive.
- It should be possible to define and attach policies toindividual data item definitionsand/or complete data entitiesto control data privacy and access security for master data,reference data, transaction data, relationship data(hierarchies) and metrics.

II. Unstructured Data

- It should be possible to define standard document andcontent types for the organization to describe what adocument/image/rich media file is e.g. for documents thedocument types could include a supplier contract, amarketing brochure, a customer contract, an equipmentmaintenance manual etc.
- It should be possible to define an enterprise classification forthe organization to define what is the image/document/richmedia file is about e.g. a booklet is about an insurance product, a conservation of manual is about a specific make andmodel of asset (equipment).
- It should be possible to define and attach policies toindividual document and contentitems to control privacy andaccess security associated with specific document/contenttypes about a specific topic e.g. to secure access tocontracts associated with a specific customer, financialreports associated with a business unit, or annual reviewdocuments associated with all employees.

IV. Data Landscape Requirements:

It should be promising to secure and protect data in motion eventhough it has not yet been stored.

User Requirements:

It should be possible to centrally manage the identity of individualusers or federate identity management so that a single view of allusers accessing applications, information and software tools insideand outside the organization can be seen to prevent creationduplicate users and so that authentication andauthorization can bemanaged centrally.

Device Requirements:

I. Device Security

- It should be possible to centrally manage mobile devicesecurity by making configuration outlines containing devicesecurity guidelines, APN settings, VPN configuration information, Wi-Fisettings, email account settings and licenses that permit mobile smart phones and tabletsto work with your enterprise and public systems.
- It should be potential to enforce device confirmation tosecure access to a mobile device.
- It should be possible to configure memory limits on mobiledevices to limit the amount of data they are allowed tohold on the device.

II. Mobile Application Security

- It should be possible to enforce mobile applicationauthentication via user ID and password so that users haveto log in to applications that provide access to sensitive and confidential datao It should be possible to enforce role-based access toapplication functionality from any device so that the user isonly authorized to use specific request functionality.
- It should be possible to allow access to specific applicationsfrom a mobile device for a set period after which access tothose applications automatically expires.

III. Data Transmission Security

- It should be possible to encrypt sensitive data flowing over apublic or private network to a desktop or mobile device orencrypt the data in the data store and decrypt beforesending only if a user is authorized to see it.

IV. Device Information security

- It should be possible to clear mobile device caches ofapplication specific information immediately an applicationcloses or after a defined period.
- It should be possible for a user to control what subset(s) ofinformation other users or user groups are allowed to seewhen sharing that information from a mobile device.
- It should be possible to enforce role-based access toapplication and information services from any device so thatthe user is only authorized to see specific information.
- It should be possible to allow access to specific informationfrom a mobile device for a set period after which access tothat information automatically expires.

V. Security Monitoring and Audit

- It should be possible to log all mobile security damages andadvices for auditing and writing purposes.

Change Management Requirements:

• It should be conceivable to define methods that require approval whenvariations are made to:

- Information classifications
- Security and privacy policies
- Schema
- Data replication and synchronization
- Access and manipulation privileges
- Application functionality

- It should be possible to run susceptibility testing before and afterchanges are made to ensure that safety risk exposure has notbeen improved as a result of the changes made. It should also bepossible to runsusceptibility testing to look for databasesusceptibilities when changes have not been made.
- It should be possible to co-ordinate controlled reversal of changes toinformation classification, schema, privileges and policies toprevious versions of privileges if changes result in security breaches.

Description About Big Data:

Big data can be considered into two areas:

- Big Data Transaction Processing
- Big Data Analytics

Big Data Transaction Processing is about extreme versions of transactions that can update data in NoSQL DBMSs, relational DBMSs or file systems. Obviously, relational DBMS is used because the so-called ACID attributes are mistaken in most NoSQL DBMSs. This is only a problem if the loss of operation is objectionable. Banking Deposit Big Data Analytics is about advanced analytics on traditional unstructured and multi-structured data. It is the term for new types of workloads and basic technologies needed to solve business problems that we could not previously support due to technical limitations, prohibitive costs, or both. So Big Data Analytics is not just about data size. Data complexity (variation of data types) and analytical complexity are important if the data size is correct. Big Data Analytics is about the data volume, the speed of data (the rate at which the data is produced) and the analytical workload associated with some combination of data variation, including complex analytics and complex data types. This is also the case for structured and multi-structured data.

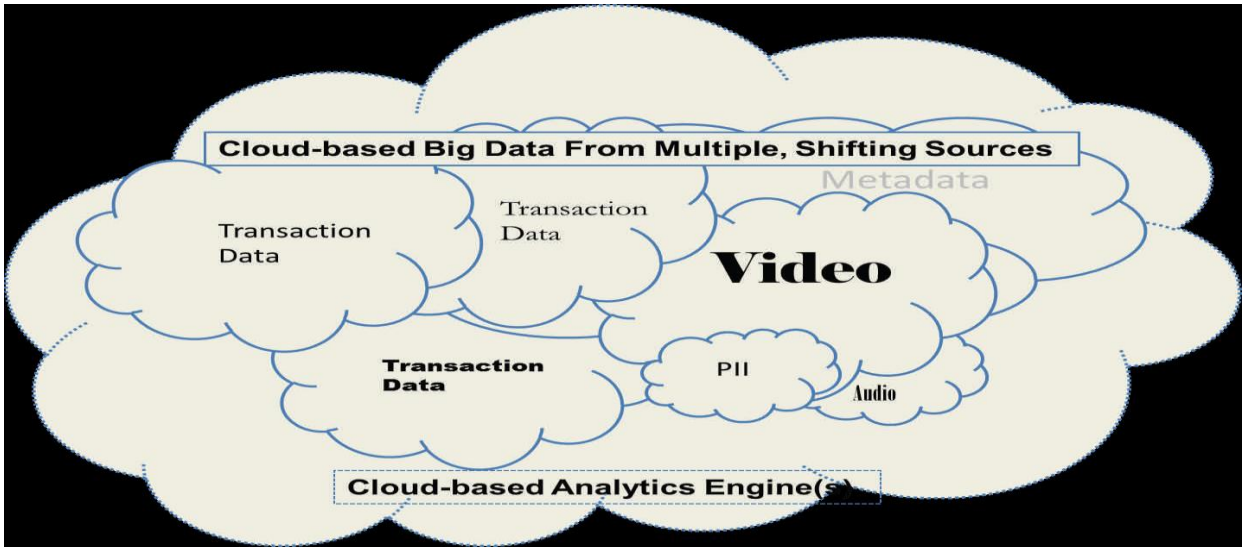
New Big Data Analytical Workload:

With the emergence of new data sources and the need to simultaneously analyze traditional unstructured content from live data streams to large amounts of structured content, many businesses are now realizing that in the analytical era, using a single enterprise data warehouse cannot be solved because of the spectrum of workload. Beyond the old data warehouse there are new big data analytics tasks. Here are:

1. Analysis of data in motion
2. Exploratory analysis of un-modeled multi-structured data
3. Analytical processing of ETL and analytical processing of un-modeled data to enrich data in a data warehouse or analytical tool
4. Analysis of the relationship in the graph
5. Complex analysis of structured data
6. Loading and re-processing of stored data

The Impact of Big Data on Information Protection:

Now that we have introduced big data and understood what it is, the nextquestion is "What is the impact of big data on information protection?"



The impact of big data on the organization:

కొత్త New sources of information.

Tion Data in motion as well as additional data at rest.

Most analytical data stores (some of these data stores may be in the cloud) in a more analytical environment.

- Big data platform wide storage eg. Analytical RDBMS Column Data Store, Hadoop Distributed File System (HDFS) or NoSQL Graph Database.
- Analyze new analytical work.
- Sandbox for data scientists to conduct experimental analysis.
- Tenders for accessing and analyzing new tools and big data.
- More complex information management in a big data environment
- Aly Supply data to many analytical data stores.

Analytical big data move data between analytical systems and data warehouses.

In spite of all this impact on data security, the data landscape is now more comprehensive, as every big data analytic platform has a different way of storing unskilled data in this rapidly evolving area.

Protecting and protecting data from new sources on different big data platforms is essential. This includes particularly large contract data and e-commerce logs. The abundance of structured and multi-structured data brought to a big data store for analysis is of interest to cyber criminals. Data sources such as location sensor data, customer data from smart phones, on-line transaction data,

user interaction data and web logins may be an example. Security around big data is an issue.

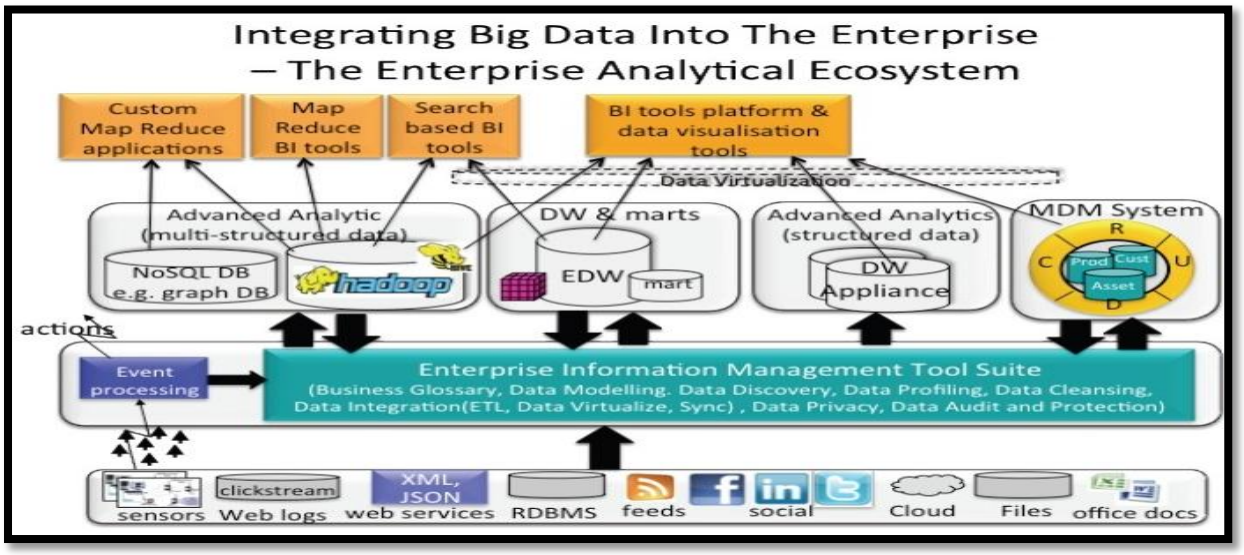


Figure 2:Source: Intelligent Business Strategies

Big Data improves data speeds and stores new file-based critical data for the data landscape, making security more complex. Fig. 2 shows that master data can be supplied to Big Data platforms to analyze big data from different angles. This data can be provided in a variety of formats depending on the big data platform it is loading. For example, data loaded into Hadoop is provided in files to be assigned to the Hadoop cluster. Complex data can also be sourced from older data warehouses in these environments to assist with large data batch analyzes. A new analytical workload is the analysis of data-in-motion along with new applications running on platforms such as Hadoop and NoSQL Database. New tools are reading this data in one or more analytical data stores. Therefore, you need to send control data environments with big data analytics tools and applications, as well as access to data on new Big Data platforms. A new type of consumer is also born: Data Scientist. Data scientists are highly skilled energy users who need a secure environment where they can detect un-modeled multi-structured data and perform complex analysis on large amounts of unstructured data.

V. Using Big Data Analytics for Security Analysis:

Finally, there is another side to this. Big Data analytics may be able tosupportto fight the security concerns by being used to noticecybercrime. Analyzing data in motion to identify fraud is one example of this. Alsostudy of access activity to see what users contact what data and whathave they done to that data.

- New security requirements / recommendations to protect information in big data environments:
- Based on what has been learned, the introduction of big data in the organization demands that the requirements be defined before enterprise information security is extended. To protect data in big data environments, we need to move beyond the built-in data in existing transaction processing systems and data warehouses. Note that you can define requirements that focus on information security for big data environments and information

security from big data environments. The former is concerned with preserving information in big data environments, while using real-time big data analytics to create insights that help protect information in big data and outdated environments.

- The following requirements relate to protecting information in big data environments, and should be included with those already recognized:
- இயக்கatrisk can analyze big data in the data movement and categorize data streams with sensitive data and find out where the important data resides in the data stream.
- விவரிக்க Describing or classifying which files are loaded on a large data platform (egHadoop HDFS) contains complex data and it is reassuring to know where the data lives on a large data base.
- Sensitive Sensitive Structured Data and Multiple Structured Big Data Define and implement policies that encrypt motion and sensitive structured data and multi-structured big data.
- எந்தவொருAny big data analytics data store can define and implement policies that transform sensitive structured data and multi-structured big data into movement and sensitive structured data and multi-structured big data.
- ும்போதுWhen moving this data between big data and traditional data stores during data processing it should be possible to encrypt and edit structured sensitive data and multi-structured data.
- கோப்பு File-based big data can control the connection to all important and important data files stored in legitimate data stores.
- In large data environments it is possible to monitor and record all data management activities related to important data streams and sensitive data files.
- Big data can control who is allowed to create big data analytics sandbox on top of legitimate sites. EgHadoop HDFS and / or Analytical DBMS.
- Sensitive can control which analytics applications have access to important data streams, and which apps can and when which streams are accessed.
- Map Which map minimization analytics applications can restrict access to important data files in Hadoop and other NoSQL data stores, and report which apps are available and when.
- Hadoop ensures that any software tools in other NoSQL data stores have access to sensitive and confidential data, and report on which tools and when any data is retrieved.
- Structured sensitivity and sensitive data generated from text analytics in Hadoop can be encrypted.
- If inserted into search codes built on big data, structured sensitivity data generated from text analytics in Hadoup should be kept secure and secure.
- Sensitive search-based tools that access sensitive and sensitive big data should not display this data to illegal users.
- Direct can connect software tools and big data logical requests to the Active Directory and / or LDAP based directories.
- The use of strategies for sensitive data in Hadoop is promising and should include guidelines for whether the data is stored in HDFS, HBase.

- It should be possible to control access to external table functions inRDBMSs that access and manipulate big data in other big datastores outside of the RDBMS itself.
- It should be possible to report on which users and whichapplications accessed and/or manipulated sensitive data in a BigData platform via external table functions on an RDBMS, and when.
- It should be possible encrypt and/or redact sensitive big data prior toproviding that data to an application, RDBMS (via external tablefunction), software tool or user not authorized to see it if that data isnot already protected.
- It should be possible to control access to sensitive data in DataWarehouses and other data stores in an extended analyticalecosystem from MapReduce applications running on Hadoop.
- It should be possible to display sensitive and confidential data file counts and filesizes in filebased Big Data platforms.
- It should be possible to detect and block unauthorized access tosensitive data streams and sensitive data files in big data analyticalenvironments.
- It should be possible to detect and block unauthorized access tosensitive data in Data Warehouses, MDM systems and other datastores from applications running on Hadoop and/or other Big Dataplatforms.
- It should be possible to secure, protect and audit all activity onsensitive big data irrespective of whether that data resides on premiseor in the cloud.

.It is possible to protect, protect and audit all operations on sensitive large data, regardless of whether it is in testing, development or production environments with respect to the use of analytics to help protect information. The following requirements must be included.

. Big data platforms and analytics can be used to collect, monitor and analyze security information to help resolve advanced security and risk use cases.

VI. Conclusion:

The advent of big data at the company and public level has a big impact on corporate information / data security. The distribution of data and the difficulty of protecting and protecting that data have resulted in the emergence of new data stores. What we have seen is that big business is forced to define new requirements when it comes to both data management and enterprise information / data security. Protecting critical data is now difficult as companies create new legitimate workloads that may move between different transaction and analytics data stores.

In this study, we studied how complex, yet complex, enterprise information security is. IBM has recognized that transactional data contains significant amounts of sensitive data, which in turn protects key databases in the IBM system, and introduces technologies to monitor, audit, and protect that data as it moves across data bases and data stores. In a distributed multicultural environment.

There are still many things to do to spread big data on the ground, and for those who are introducing big data into a distributed cloud computing environment, information security is high on the agenda. IBM is already favored as a key supplier of software to tackle enterprise information / data security.

References:

[1] PWC 2012 Global State of Information Security, http://www.pwc.com/gx/en/information-security-survey/giss.jhtml
[2] "Information Governance: Audit and Protection on the IBM System z Platform", Ferguson, October 2011
[3] Multi-structured data can be semi-structured like email or XML or unstructured data like text and video
[4] For more information on these technologies please refer to the paper "Architecting a Big Data Platform for Analytics", Ferguson, September 2012
[5] www.ibm.com/software/data/bigdata
[6] www.cisco.com/en/us/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html
[7] www.nextbillion.net/blog/the-age-of-big-data
[8] jana.com/research/sample/,www.nextbillion.net/blog/2011/10/05/reaching-the-next-billion-through - mobile
[9] ess.santafe.edu/publications.html