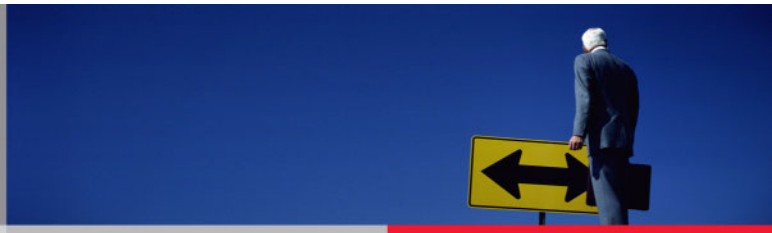
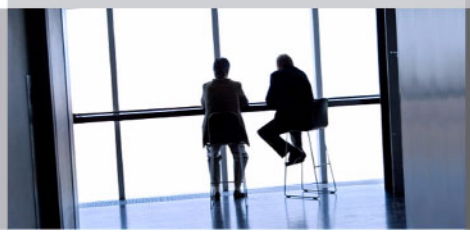
Smarter Analytics: Information Architecture for a New Era of Computing





Redguides

for Business Leaders

Mandy Chessell

- Build lasting value with an integrated information architecture
- Incorporate the benefits of big data into your infrastructure
- Protect, improve, and govern your critical information

Redbooks

Introduction

IBM Global Technology Outlook 2013[1] highlights how recent advances in technology are changing the relationship between an organization and the people it interacts with; whether they are customers, business partners, or employees. It describes the emergence of the contextual enterprise.

A *contextual enterprise* is an organization that dynamically adapts to the changing needs of its individual customers by using information from a wide range of sources. A contextual enterprise has these characteristics:

- Actively engages with external parties such as customers, business partners, and analysts
- Uses information from multiple sources to provide meaningful context to communications with customers
- Drives new business activity as a result of the interaction with external parties
- Uses analysis and correlation of information from both inside and outside of the organization to extract important facts and identify events than need action

A local business adapts to the changing needs of its customers naturally. Its employees are immersed in the community and continuously receive information from a broad range of sources. Some of this information is immediately relevant to the business; other pieces of information might relate to additional opportunities in the future.

Whenever a local event occurs, the local business is present, serving its customers where and when necessary. If its customers need additional goods and services, the local business is aware of that and can make choices about whether or not to extend the range of offerings.

How can a larger organization, serving many communities, provide the same level of service as a local business?

[1] IBM® Global Business Outlook 2013, http://www.zurich.ibm.com/pdf/isl/infoportal/Global_Technology_Outlook_2013.pdf

Some of the answers are determined by the culture of the organization:

- Are employees encouraged to reach out to customers and provide a personalized service?
- Are they able to trigger internal discussions about the untapped needs they see in the market?
- Are they able to record and share appropriate insights about their customers among their colleagues?

Information technology (IT) also has a part to play because it is the means by which an organization scales its operations. IT provides the access points to many channels of communication and, most important, organizes, collates, processes, and retains important information that supports its activities.

This IBM Redguide™ publication focuses on how IT takes advantage of new sources of information to deliver insight to the enterprise, enabling it to act in the context of its wider environment. The guide begins with a view of the types of information that must be included in the context and where this information is from. It also covers how this information can be integrated into the enterprise and the practices that ensure that the management of this information is cost-effective, compliant with regulations, and sustainable.

The big data challenge and opportunity

Every organization collects information about the people and organizations it interacts with, the business activity it performs, and the assets it uses. Often this information is distributed among the organization's applications, creating many partial views of the activity of the business.

The missing information from each application's view can affect the decisions the organization makes. The following scenario illustrates this point. It describes information use in a business website application.

A potential customer connects to the website, browses some pages, and selects the option to register as a customer. The registration process collects essential information about the person and, if the responses are appropriate, the website application accepts this person as a new customer.

Next, this customer selects goods or services through the website. When the additional information necessary to place an order passes the verification process, the order is accepted and processed.

Over time, as the same customer returns to the website and selects additional goods or services, the website application builds up knowledge about the customer's preferences. Special offers and other promotions can be made to the customer by matching the customer's use of the website to historical knowledge of similar customers. The successful matching of offers to each particular customer leads to additional sales and increased customer loyalty.

In this example, the website application is working with the information that its customers have explicitly given through the website. There might be additional information about each customer in other applications that support specific types of customer accounts or the customer's interaction with other of the organization's channels.

Consolidated views of information

Technology such as master data hubs and data warehouses consolidate information from multiple applications into specialized servers to give a holistic perspective of how the organization is operating. Content management systems provide organized collections of documents that can be managed through well-defined processes. Together these solutions provide state-of-the-art management of an organization's information. These solutions enable the organization to support reporting, analytics, multi-channel interaction with customers and partners, and provide demonstrable compliance to necessary regulations for their industry.

By accessing a consolidated view of customer information from a master data hub, the website application is working with the best data about the customer that the organization has. However, this application still has only its internal view of the customer. For example, the application is unaware of purchases made by a customer using other websites. As a result, the organization's website might continue making inappropriate offers to some customers because of incomplete knowledge of their needs. There is increased value in using additional information that fills in the gaps.

Systems of engagement

Website applications are typically passive. Potential customers move to another website if they do not find what they want and the organization with the first website has little understanding of what was missing from its offerings.

In situations where the customer relationship is important, the website might offer the potential customer the opportunity to connect to an employee so they can engage in a conversation about the customer's needs. Some organizations are taking this a step further and have their employees engage directly in the electronic communities that their customers are using.

In both of these examples, the interaction between the customer and the employee is two-way; these systems are sometimes referred to as *systems of engagement*.

The conversations with customers that the systems of engagement support are powered by context, which consists of the following items:

- Information about the customers themselves
- A history of the exchanges that have taken place before and with whom
- The current reason for the exchange

The richer the context, the more meaningful the conversation can be. Any system or employee that connects to these electronic communities should be seen as a gateway into your organization, demonstrating the accumulated knowledge and capabilities your organization brings to the world.

Additional sources of information

Increasingly people are comfortable using the Internet for many aspects of their lives such as, connecting with their friends, locating information, shopping, registering opinions, and doing business. Often they are carrying a smartphone that also pinpoints their location. All of this activity, along with the massive number of sensors collecting bits of data from many places is creating an explosion of digital data that can be collected, processed, and used to develop greater insight into the activity in the wider world.

Much of the data related to human collaboration is saved in a text form such as, emails, web pages, documents, conversations using instant messaging, and social media technology. The interesting facts that an organization is looking for are widely scattered throughout this data. Specialized analytical technology is required to step through this text and pick out the interesting facts and events. There are huge volumes of data being produced at a rapid rate, resulting in large quantities accumulated in a short period of time. This data must be processed at sufficient speed for opportunities to be detected and action taken before this information becomes irrelevant as time passes.

Big data technology

New technology designed for handling large volumes of data is growing in popularity. This technology is creating excitement in many organizations because of its ability to scale to handle enormous amounts of data of various types and formats at an affordable price. The following types of technology are available:

- Map-reduce (Hadoop) technology

  Map-reduce technology distributes parallel processing of highly variable data across a large, multi-computer file system.

- Streams technology

  Streams technology manages data created at high speed from Internet feeds and devices.

- Search and discovery technology

  Search and discovery technology finds the correct information from a wide variety of sources.

- Advanced analytics

  Advanced analytics discovers patterns in data that predict what is likely to happen in the near future or indicate an opportunity to create a better outcome.

These technologies enable an organization to pull in and process additional data from third parties or from untapped internal systems. To be useful, the results of this processing must be passed to the systems that are supporting the organization.

Selecting new sources of information

The ability to handle large amounts of diverse information provides the organization with broader insight into its customer's perception of its goods and services, if the information used is relevant. Including unnecessary details can delay, confuse, and mislead analysis in the same way as missing or out-of-date information effects results.

An organization that wants to augment its existing information with new sources of information needs a focused approach to ensure these outcomes:

- New and additional information is relevant to the business requirements
- The best sources for the new information are being used
- The new information is made available to the business systems and people that need to access the new information

Often questions help to identify the information that is needed for the decision process. For example, if customers are no longer purchasing from your organization, and you do not know why, maybe decisions related to customer service should be your initial focus. If there is a new market opportunity, what information is needed to understand where the value is and

how to enter the market? If fraud is suspected, how can it be detected or, better still, prevented? Consider the following general questions:

- What does the decision-maker need to know to understand the current situation?
  - How did this situation occur?
  - Why did the situation occur?
  - Who is affected?
  - What resources, services, and assets are involved?
  - When did this occur?
  - Where did it occur?
  - What is the impact of this situation?
- How does the decision-maker assess how critical the situation is, and after the assessment is made, how does that person decide how to act?
  - Do they need to act?
  - Why do they need to act?
  - How do they implement this action?
  - Who is affected?
  - What resources, services, assets are involved?
  - When do they act?
  - Where do they act?

Some questions are easily answered from general knowledge or information stored in your organization's systems and some of this information is readily available. Other information requires additional processing to extract the facts in a format that is suitable for the decision-maker to use.

Consider these example situations:

- Understanding what agreement was made with a customer during a phone conversation. This event needs a transcript or recording of the conversation so it can be analyzed to extract the important facts and sentiment of the conversation.
- Understanding which web pages are being browsed on the company website. This is accomplished by analyzing the details of the user’s navigation through the website captured in a click-stream log.

Other information might come from third-party sources that have a different perspective on the subject area.

It might be necessary to infer new information from the information you have. For example, if a customer is browsing the support pages of your website, maybe that customer has a problem with one of the products purchased. A call from a support person to verify if the customer needs help can increase the customer’s loyalty.

Analytics can be used to predict near-term events or to optimize the use of resources for that customer. This approach works by detecting patterns of behavior in historical data. For example, predicting the best types of products to offer to a particular customer based on the buying patterns of similar customers might be possible.

Through this investigation, your technical team can build up a model of the information that might improve the operations of your organization. Mapping this information to your organization’s systems identifies three types of information gaps:

- Information that is missing from the organization's systems
- Information that is not of sufficient quality for the decision makers
- Information that is available in a system other than the one used by the decision-makers

Filling these gaps requires a focus on three activities:

- Agreeing on the information strategy for the organization

  This activity describes how to provide the type and quality of information that is needed to support the organization's business strategy. The information strategy must cover three types of work:

  - Exploratory investigation of patterns in data, to understand the potential cause and effect of proposed activity
  - Projects that make changes to IT infrastructure or business processes
  - Everyday business operations that run continuously

- Developing an evolving information landscape

  This activity explains how the IT infrastructure can be transformed over several iterations to deliver the information strategy:

  - Incorporating new sources of information into your organization's systems. Often this involves deploying new systems to store this information.
  - Using techniques such as master data management (MDM) and content management to create consolidated sources of information for operational use.
  - Connecting systems together, so information can flow to the places where it is needed.

- Operating an information governance program

  This activity actively controls and monitors how information is being used and managed.

The information strategy must take a holistic approach to the management of information. It requires a *system of systems* style of architecture where each system has a well-defined role, both in serving its direct users and also being a part of the management and flow of information across the organization.

A new information landscape

Figure 1 on page 7 shows an architectural approach to augmenting existing systems with new information to support new information requirements that are needed by the contextual enterprise. This information architectural approach has three goals:

- Combine the best information from across the organization for use both by operational and analytical systems.
- Augment this consolidated information with relevant facts and event triggers by analyzing previously untapped sources of information.
- Help to more easily locate information and deliver it to where the business most needs it.

In Figure 1, most of the new information is entering on the left side of the picture and is being consumed by the information interaction solutions shown on the right side. In the middle of the figure is the processing that interprets, transforms, consolidates, and moves information so it is consumable by the information interaction solutions.

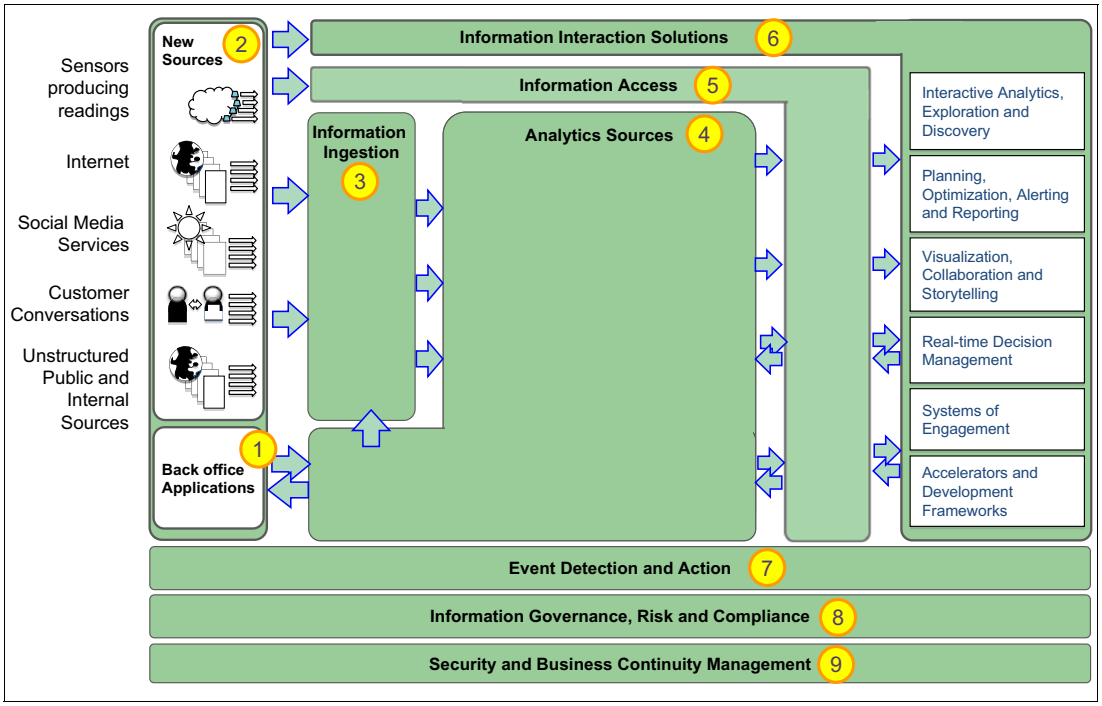


Figure 1 A new information landscape

The arrows that point right-to-left in Figure 1 represent feedback and additional information that is being generated by the information interaction solutions on the right.

The subsystems shown in Figure 1 are as follows:

- Back office applications (item 1)

  Back office applications execute and process the business transactions. These systems include the order processing, billing, marketing, product development, and sales types of systems. In general they operate independently using their own data, but they can exchange information with the operational systems in the analytics sources.

- New sources (item 2)

  New sources are systems that supply information to augment the information generated by the back office applications. These sources might come from third parties, or might be internal sources, such as logs, emails, and other text-based sources that have been too expensive to process in the past.

- Information ingestion (item 3)

  Information ingestion collects together raw information from the back office applications and new sources for validation, correlation, cleansing, and transformation.

- Analytic sources (item 4)

  Analytics sources provide the information for different types of analytics processing.

  - Some of this analytics processing occurs inside of the systems hosting the analytics sources.
  - Some analytics processing occurs in the provisioning engines as information is moved between the analytics sources.
  - Some analytics processing occurs in the information interaction systems.

  The analytics sources include shared operational systems such as master data hubs, reference data hubs, activity data hubs, and content management hubs. The analytics sources also include systems such as data warehouses, map-reduce (Hadoop), files, databases, and data marts that host historical information harvested from many sources.

- Information access (item 5)

  Information access enables the information interaction solutions to locate and consume information without needing to understand exactly where it is physically stored and maintained. This capability is part of an approach called information virtualization, which is described in more detail in “Information virtualization” on page 13.

- Information interaction (item 6)

  Information interaction solutions include the systems of engagement and advanced analytics capability. They predominantly work with information drawn from the analytics sources. Information interaction solutions can produce some information that is fed back into the analytics sources.

- Event detection and action (item 7)

  Event detection and action triggers business processes and other activity when significant events are detected in the applications, analytic sources, and information interaction solutions.

- Information governance, risk and compliance (item 8)

  Information governance, risk, and compliance provide the capability to ensure information is appropriately managed throughout its lifecycle with the level of quality and protection that is consistent with the information's sensitivity and use.

- Security and business continuity (item 9)

  Security and business continuity management ensures all systems have the security and availability appropriate to their importance to the business.

Figure 2 on page 9 shows the subsystems from Figure 1 on page 7 overlaid with information zones, which are represented by dotted-line boxes. An information zone defines a collection of systems where information is used and managed in a specific way. The zones overlap with one another when the same information is being used for multiple purposes. This approach is necessary when there is a high volume of information making it uneconomical for each team to have its own private copy of the information. This architecture ensures the availability of suitable information for all of the teams in the organization, with maximum reuse and flexibility to support new use cases.

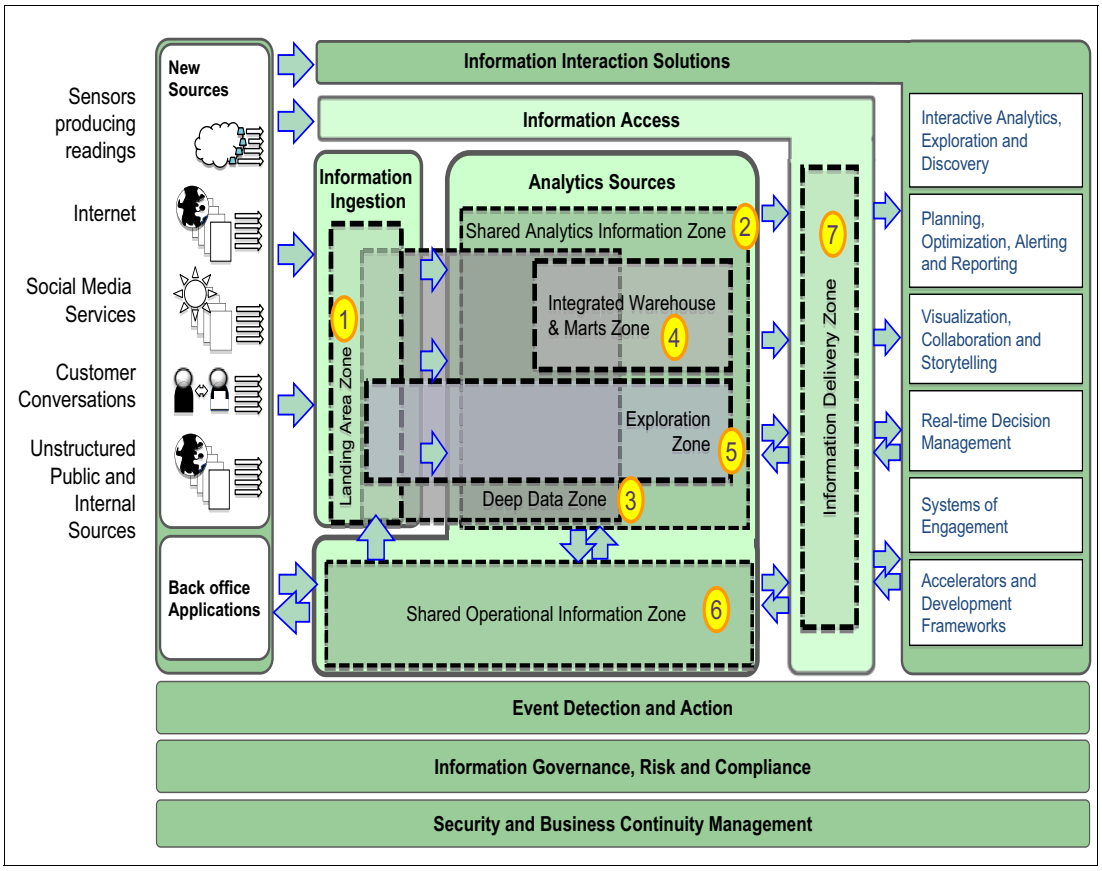


Figure 2 Information zones in the new information landscape

The information zones shown in Figure 2 are defined as follows:

- Landing area zone (item 1)

  Landing area zone manages raw data just received from the applications and the new sources. This data has had minimal verification and reformatting performed on it

- Shared analytics information zone (item 2)

  Shared analytics information zone contains information that has been harvested for reporting and analytics.

- Deep data zone (item 3)

  Deep data zone contains detailed information that is used by analytics to create new insight and summaries for the business. This data is kept for some time after the analytics processing has completed to enable detailed investigation of the original facts if the analytics processing discovers unexpected values.

- Integrated warehouse and marts zone (item 4)

  Integrated warehouse and marts zone contains consolidated and summarized historical information that is managed for reporting and analytics.

- Exploration zone (item 5)

  Exploration zone provides the data that is used for exploratory analytics. Exploratory analytics uses a wide variety of raw data and managed information.

- Shared operational information zone (item 6)

  Shared operational information zone has systems that contain consolidated operational information that is being shared by multiple systems. This zone includes the master data hubs, content hubs, reference data hubs, and activity data hubs.

- Information delivery zone (item 7)

  The information delivery zone contains information that has been prepared for use by the information interaction solutions. The information delivery zone typically contains read only information that is regularly refreshed to support the needs of the systems using it. It provides some of the authoritative information sources that are used in information virtualization where the original source of information is not suitable for direct access.

The systems within these information zones are shown in Figure 3.

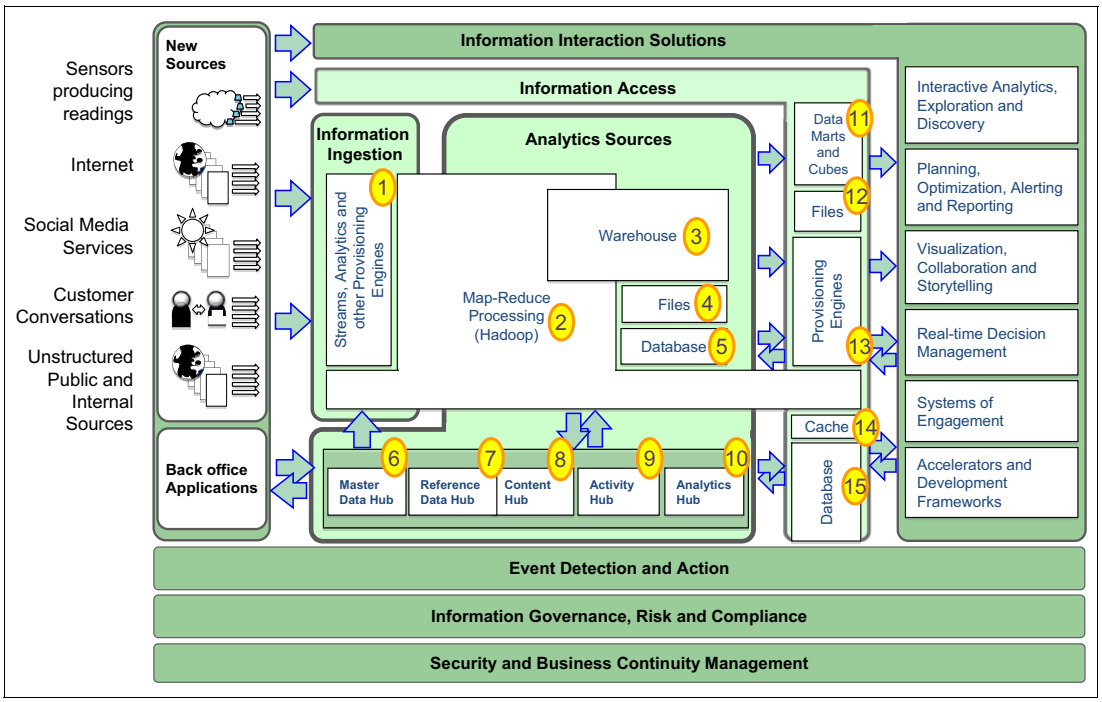


Figure 3 Systems supporting the new information landscape

The systems shown in Figure 3 are described as follows:

- Streams, analytics and other provisioning engines (item 1)

  Streams, analytics and other provisioning engines take raw data and perform initial verification, correlation, consolidation and transformation to make the information consistent, simplifying subsequent processing. Streams provisioning engine is particularly important when receiving information from high speed feeds that is arriving too quickly to store. Streams engine is able to extract important facts to pass on to the other systems.

- Map-reduce processing (Hadoop) (item 2)

  Map-reduce processing (Hadoop) provides a flexible storage system that can hold data in many formats. Schemas and other forms of annotations can be mapped onto the data after it is stored, allowing it to be used for multiple purposes.

- Warehouses (item 3)

  Warehouses are still needed to provide efficient access to consolidated and reconciled information for analytics, reports, and dashboards.

- Files (item 4)

  Files are used for many purposes particularly for moving large amounts of information.

- Databases (item 5)

  Databases offer structured storage for real-time access.

- Master data hubs (item 6)

  Master data hubs consolidate and manage master data such as customer, supplier, product, account, and asset, as part of a master data management (MDM) program.

- Reference data hubs (item 7)

  Reference data hubs manage code tables and hierarchies of values used to transform and correlate information from different sources.

- Content hubs (item 8)

  Content hubs manage documents and other media such as image, video, and audio files that must be controlled and managed through formal processes.

- Activity data hubs (item 9)

  Activity data hubs manage consolidated information about recent activity, including analytical decisions that are related to the entities in the master data and content hubs. This type of data is normally managed by applications. However, because the organization is processing information about activity from outside of the scope of its applications, a new type of hub is needed to manage the dynamic nature of this type of information.

- Analytics hubs (item 10)

  Analytics hubs appear among the operational hubs to create additional insight in real time. These hubs are supporting advanced analytics such as predictive analytics and optimization, along with business rules.

- Data marts and cubes (item 11)

  Data marts and cubes still provide subsets of information specifically formatted for a particular team or style of processing. They are read-only copies of information that are regularly refreshed from the analytical sources.

- Files (item 12)

  Files are still used as input to certain analytics processing particularly data mining.

- Provisioning engines (item 13)

  Provisioning engines are either serving up information in real-time to the information interaction solutions, or creating and maintaining specialized stores of information in the information delivery zone that will be used by the information interaction solutions

- Cache (item 14)

  Cache represents in-memory stores providing fast access to read-only information for the information interaction solutions.

- Databases (item 15)

  Databases (in this context) can provide real-time storage for information created by the information interaction solutions.

Collectively these systems supply authoritative information to the information interaction solutions.

Authoritative sources are maintained using a deterministic and governed information supply chain that synchronizes the authoritative sources with other copies of the information. Each

system shown in Figure 3 on page 10 is responsible for implementing part of the information supply chain.

Figure 4 illustrates the types of processing, or capabilities, that are supported by the systems in the information supply chain. Many of these capabilities are used in multiple places in the information supply chain, because each system maintains the information it is responsible for and makes this information available to systems downstream in the information supply chain.

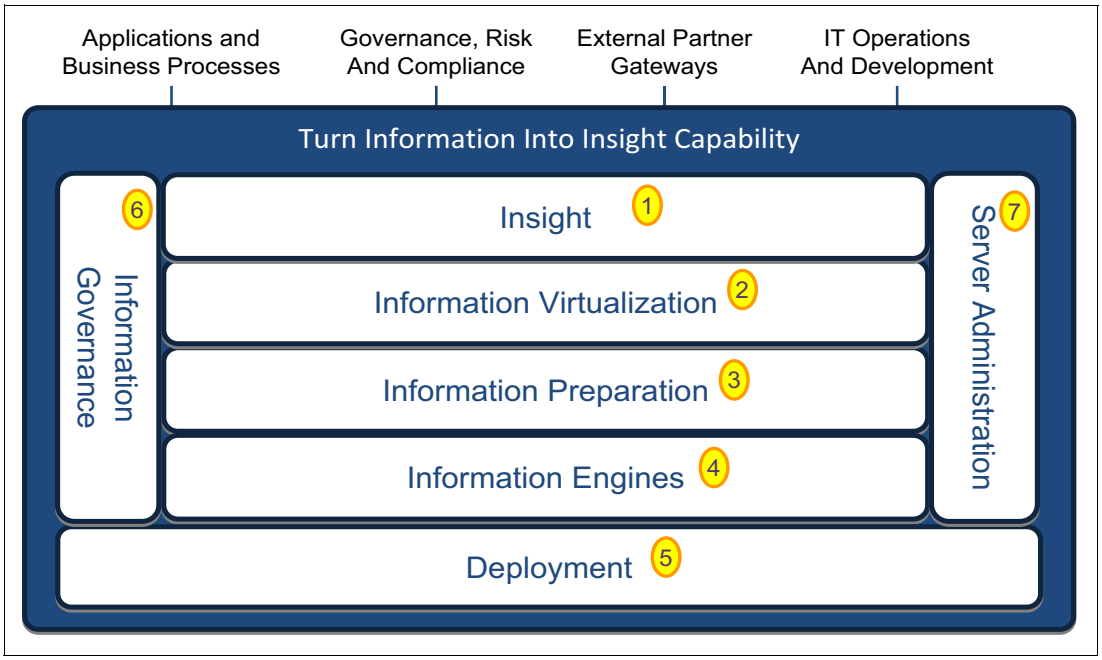


Figure 4 Turn information into insight capability

The capabilities shown in Figure 4 are as follows:

- Insight (item 1)

  Insight is the processing that creates new information by analyzing existing information. Analytics processing occurs at many points along the information supply chain and includes these capabilities:

  - Understanding an information source and categorizing its content based on detected patterns of values
  - Summarizing information
  - Matching and merging duplicated information from different sources
  - Correlating and linking related information values together

- Information virtualization (item 2)

  Information virtualization provides views over stored information that has been adapted to suit the needs of systems and processes that use the information. For more detail about information virtualization, see “Information virtualization” on page 13.

- Information preparation (item 3)

  Information preparation is the function that cleanses, enriches, transforms, consolidates, and links information so it is high quality and easy to use. This capability is also referred to as *information curation*. The information engines automate much of this activity, which is augmented with specific checks and decisions made by human subject matter experts.

- Information engines (item 4)

  Information engines are the specialized servers that support the analytic sources. They include databases and data warehouses; provisioning engines for moving data, engines for performing analytics, hubs for storing operational data. Information engines also support search and exploratory servers, enabling people to find the information they need.

- Deployment (item 5)

  Deployment describes the infrastructure platform on which the information engines run. Deployment can consist of servers, appliances, or cloud-based infrastructure.

- Information governance (item 6)

  Information governance defines the people, processes, and technology that ensure information is appropriately managed to support the organization's use of it. See also “Active information governance” on page 14.

- Server administration (item 7)

  Server administration supports standard facilities, such as startup, shutdown, configure, security, backup, and restore, necessary to manage servers from an IT operations perspective.

Information virtualization

Simplifying the interaction between systems increases the chance that shared information is used and used properly across the organization. The organization must control how the information is accessed to ensure it is available when it is needed and at an appropriate level of performance. The approach that provides these services is referred to as *information virtualization*.

Information virtualization enables information to be stored in various locations but managed, viewed, and used as a single collection. It maximizes the sharing of information, while ensuring individual teams receive information that is customized to their needs.

Information virtualization has two layers as illustrated in Figure 5.

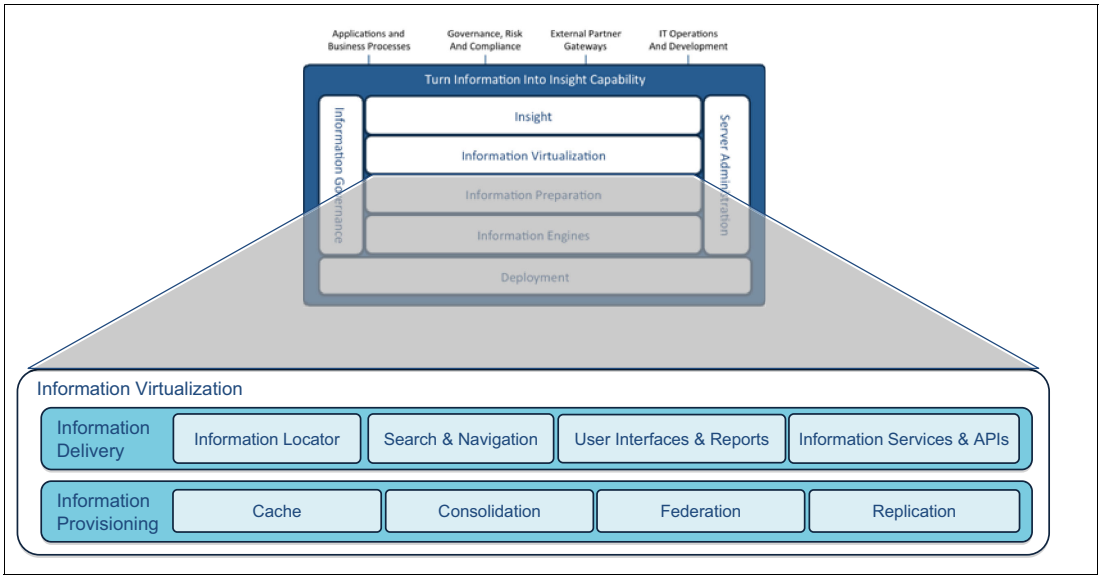


Figure 5 Information virtualization capabilities

The information virtualization capabilities shown in Figure 5 on page 13 are defined as follows:

- Information delivery

  Consumer focused access points include user interfaces, services, and application program interfaces (APIs) that provide information to both systems and people. This information can be the business information itself or descriptions and statistics about the business information, which is called *metadata*. Metadata is used to locate the correct information to use.

- Information provisioning

  Authoritative information sources are made available to the access points using the most appropriate provisioning mechanisms. The provisioning mechanisms can be the following items:

  – Caching provides read-only, in-memory local copies of data that provide fast access for situations where the same query is issued repeatedly on slowly changing information.

  – Federation is real-time extraction and merging of information from a selection of sources providing the most current (but not necessarily most consistent) information.

  – Consolidation makes use of a pre-constructed source of information fed from multiple sources. This approach provides consistent and complete local information (although it might not be the latest available).

  – Replication is an exact local copy of remotely sourced information, which provides locally stored, read-only access to information that is maintained elsewhere.

The choice of provisioning method is determined by the needs of the systems and processes that use the information.

Active information governance

When systems are connected together, there is an increased obligation to manage and govern the way information is used in the organization. For instance bad information added in one system has the opportunity to contaminate other systems and has a greater impact than ever before.

Information governance (which is part of overall information management shown initially in Figure 4 on page 12) consists of the capabilities shown in Figure 6 on page 15

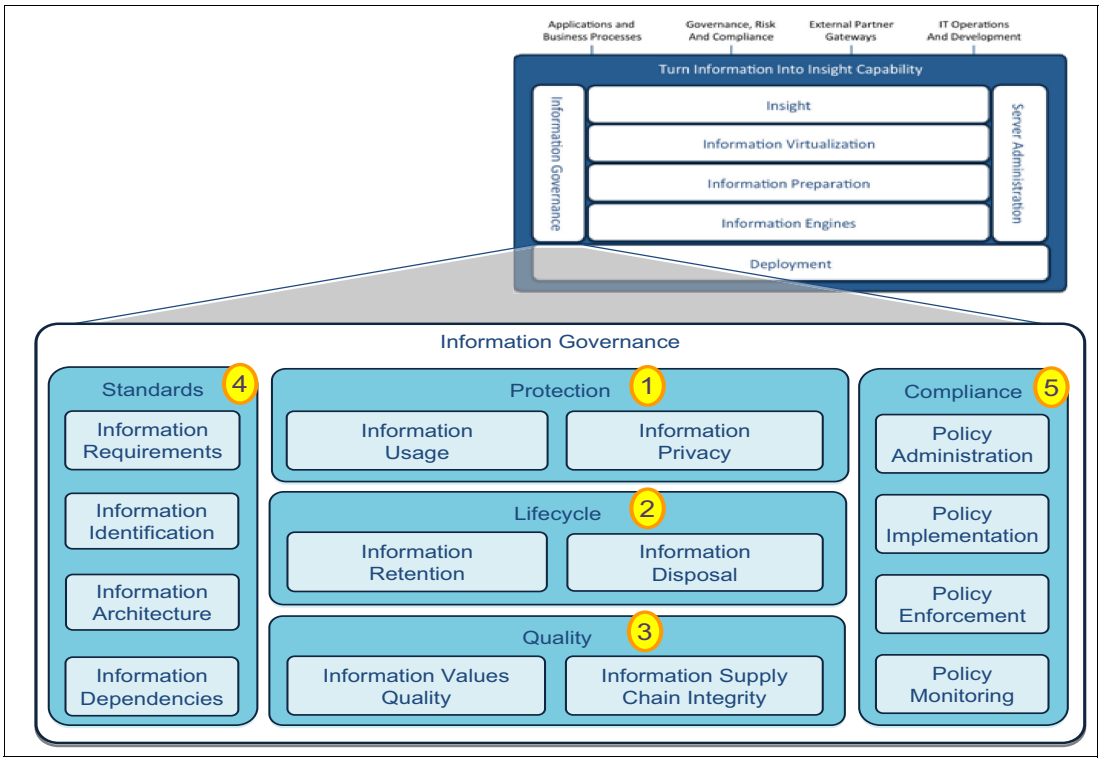


Figure 6 Information governance capabilities

Three main disciplines are in information governance (located in the middle of Figure 6):

- Protection (item 1)

  Protection ensures information is used in appropriate ways, and an individual's privacy is respected.

- Lifecycle (item 2)

  Lifecycle ensures information is appropriately managed throughout its life, so it is retained for as long as it is needed and removed when it is no longer useful.

- Quality (item 3)

  Quality actively improves the accuracy, timeliness, completeness, and precision of information to the level that is appropriate for the consuming people and systems.

These disciplines are supported by (as shown in Figure 6) the following capabilities:

- Standards (item 4)

  Standards define the information strategy and the inventory of information definitions and systems that support them.

- Compliance (item 5)

  Compliance capabilities support the definition of policies, the execution of rules that implement these policies, and the measurement of activities to determine the level of compliance being achieved.

As with the information strategy, information governance should start small and focus on the most important information, and then expand out as it demonstrates its worth. It must evolve with the business, be responsive and accountable, while seeking to communicate and educate people in the appropriate management of information. Most important, it needs senior stakeholders and visible consequences for those who ignore the requirements.

Making strategic change

Although change to the IT infrastructure is required, improving the way information is managed, owned and governed requires a cultural change in the organization. There are many theories, books, and courses about strategic changes in an organization. Although the specifics vary, some common themes are often present, including these common themes:

- Explanations covering the reasons why change is necessary are repeatedly given to everyone affected by the change, backed up by more facts as they reveal themselves.
- The change is openly sponsored at the senior management of the organization.
- New approaches are piloted using small but significant pilots before a change is rolled out to the organization.
- People must be given the correct skills to enable the change.
- Success is measured and celebrated, with care to reward those who have met the challenge.

These steps need ongoing focus. Senior management must keep the goals of the initiatives in the forefront of people's minds as they make decisions about how the organization operates.

Maintaining the momentum

As with all change, demonstrating business value is key. Every iteration of the information architecture must deliver value on two levels.

- The IT infrastructure must move closer to the target architecture. This approach simplifies future change and reduces the impact of poor quality information.
- The new capability must provide a tangible business benefit that enables the organization to take advantage of new opportunities.

The information governance program must be agile and business-centric, thus ensuring the business sees the benefit and feels empowered to influence the information governance program's direction.

Finally, senior management in the organization must maintain an active interest in the rollout of the information strategy. As the business strategy evolves, the information strategy evolves to meet it, keeping it vital and relevant to the key stakeholders in the organization.

Conclusion

Making good decisions requires information that is relevant, timely, and accurate. In today's world, there is no lack of data. We are deluged with facts, opinions, and data from many sources. The challenge is to identify what information to rely on and how to correlate the various facts and opinions from many disconnected sources in real time.

To thrive in a globally competitive market where customer expectations are continuously rising, your decision makers need support from strong information management and analytics capabilities. Information management capabilities collect, correct, and correlate data from many sources into coherent collections of information. They provision information for use by business processes, analytics, reporting, dashboards, and other tools used by decision-makers.

Information governance underpins effective information management through policies, business processes, and metrics that control how people within the organization work with information. Together information management and information governance create an information platform where the business can develop and grow towards new markets and opportunities.

This guide described how to use information management technology and information governance to collect, manage, and target the correct information to your decision-makers and the tools that serve them. It described the increased volume and importance that information plays in your organization and how it is important to integrate all the systems so data can be shared.

New capabilities emerging from the new era of computing are not in themselves revolutionary. The shift from application-centric operation to an information and analytics operation will occur through a series of evolutionary steps, each delivering their own value along the way.

IBM Global Business Services® has people skilled in planning for strategic business change. They can guide you through the business transformation required to become an information and analytics-centric organization. Their skills include running workshops to help you formulate your information strategy, building your roadmap, developing new capabilities, and establishing an information governance program.

IBM Software Group provides a wide range of software products that implement key information management capabilities (shown in Table 1).

Table 1 IBM information management solutions

Capability	Product or product family	Web location
Master data hub	IBM InfoSphere® Master Data Management	http://www.ibm.com/software/products/us/en/infosphere-master-data-management/
Reference data hub	IBM InfoSphere Reference Data Management Hub	http://www.ibm.com/software/data/infosphere/mdm/reference-data-management-hub/
Activity data hub	IBM InfoSphere Master Data Management Custom Domain Hub	http://www.ibm.com/software/products/us/en/infomastdatamanacustdomahub/
Content Hub	IBM FileNet® Content Manager	http://www.ibm.com/software/products/us/en/filecontmana/
Information Governance Policy and Standards, Data Movement	IBM InfoSphere Information Server	http://www.ibm.com/software/data/integration/info_server/
Map-Reduce (Hadoop)	IBM InfoSphere BigInsights™	http://www.ibm.com/software/data/infosphere/biginsights/
Streams or Streaming Engine	IBM InfoSphere Streams	http://www.ibm.com/software/products/us/en/infosphere-streams/
Analytics	IBM Business analytics	http://www.ibm.com/software/products/us/en/category/SWQ00
Data exploration and search	IBM InfoSphere Data Explorer	http://www.ibm.com/software/data/infosphere/data-explorer/

IBM Systems and Technology Group delivers workload-optimized IT infrastructure to support your organization's operations. IBM Global Technology Services® group provides hosting environments for cloud-based services and outsourced IT operations.

IBM has the technical and business expertise to help you transform your IT infrastructure to become an information-centric organization.

Other resources for more information

For additional information about this topic, see the following resources:

- The IBM Big Data Hub

  http://www.ibmbigdatahub.com

- Mandy Chessell, Harald Smith, *Patterns of Information Management*, Pearson plc, 2013, ISBN-13: 978-0-13-315550-1

Authors

This guide was produced by a specialist working with the International Technical Support Organization (ITSO).

Mandy Chessell is an IBM Distinguished Engineer and Master Inventor and is currently the Chief Architect of IBM InfoSphere Solutions, working in the Chief Technology Officer (CTO) Office of SWG Information Management. She has expertise in designing information supply chains for information intensive solutions. Mandy joined IBM in 1987 and has held roles for developing new features for various IBM products such as IBM CICS®, IBM TxSeries, Encina, Component Broker and IBM WebSphere® Application Server. She is a Fellow of the Royal Academy of Engineering (FREng), a Chartered Engineer (CEng), and a Fellow of the British Computer Society (FBCS).

Thanks to the following people for their contributions to this project:

LindaMay Patterson
International Technical Support Organization, Rochester Center

Tim Vincent
IBM Fellow, Vice President and CTO for Information Management

Bill O'Connell
IBM Distinguished Engineer, CTO Data Warehousing

Gaurav Rao
IBM Strategic Initiatives Lead

Dan Wolfson
IBM Distinguished Engineer, Chief Architect and CTO, InfoSphere

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new IBM Redbooks® publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document, REDP-5012-00, was created or updated on July 31, 2013.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at
http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

BigInsights™
CICS®
FileNet®
Global Business Services®
Global Technology Services®
IBM®
InfoSphere®
Redbooks®
Redguide™
Redbooks (logo) ®
WebSphere®

The following terms are trademarks of other companies:

Other company, product, or service names may be trademarks or service marks of others.