CHAPTER 2
DATA GOVERNANCE AND IT ARCHITECTURE SUPPORT LONG-TERM PERFORMANCE

IT at Work

2.1 Data Quality Determines Systems Success and Failure

1. Why was an EIS designed and implemented?
The EIS was designed to provide senior managers with internal and external data and key performance indicators (KPIs) that were relevant to their specific needs.

2. What problems did executives have with the EIS?
The executives found that half of the data generated from their EIS was irrelevant to corporate-level decision making concerning their Strategic Business Units (SBU) and that some of relevant timely data, crucial for decision making, was not available when and how they wanted them.

3. What were the two reasons for those EIS problems?
The application architecture was not designed for customized report generation. In the EIS system, the SBUs were reporting sales and revenue at different timeframes.
In addition, the user interfaces were too complicated to get to the required information; so much so, that the analysts had to first extract KPI-related data and then work on them for producing the information required by the executives.

4. How did the CIO improve the EIS?
The CIO put in place a dedicated team to design and develop a new system with a business-driven architecture instead of a financial reporting-driven architecture. The new system used standardized data formatting across the company, thus eliminating data inconsistencies.

5. What are the benefits of the new IT architecture?
The new system provided reliable KPI reports on inventory turns, cycle times, and profit margins of all SBUs. It was easy to modify reports, eliminating ad hoc analyses. There was a reduction in resources required for maintaining the system. EIS use by executives improved due to relatively reliable underlying data in the system.

6. What are the benefits of data governance?
Data formats are standardized, data inconsistencies eliminated, and therefore high-quality data that employees and business partners can trust and access on demand is provided. Improved data quality can bring benefits such as accurate sales forecasting and order processing, increased sales growth, a better customer experience, and greater customer loyalty and retention.

2.2 EA is Dynamic

No questions.

2.3 Business Continuity with Virtualization

1. What business risks had Liberty Wines faced?
As their business grew, their IT facility could not handle the increased data volume. The systems were slow and required greater maintenance efforts. This meant loss of employee productivity, thus affecting its core business processes such as order processing and inventory management.

By not providing the level of service that customers expect in a fast-paced environment or the ability to accommodate growth, they risked not only loss of future business but also the potential loss of current customers.

2. How does Liberty Wines' IT infrastructure impact its competitive advantage?
By reducing costs (power, air conditioning, and hardware replacement), improving resilience and stability through a backup system, and speeding up business processes by enabling apps to run faster, the bottom line was improved and employees can provide better customer service with improved productivity. Further, future business growth can be accommodated easily and quickly. Each of these contribute to an advantage over competitors by reducing costs, improving service, and providing for future growth.

3. How did server virtualization benefit Liberty Wines and the environment?
Reducing the number of servers from 10 to 4, one being part of a backup system, caused a cut in power use and air conditioning costs of 60 percent, improving the bottom line and reducing the carbon footprint. Applications ran faster through better utilization which, in turn, resulted in better customer service and inventory management. Hardware replacement costs were cut by $69,500 and servers can be added easily and quickly when needed.

2.4 Unilever

No questions.

Review Questions

2.1 Information Management

1. Explain information management.
Information management is the use of IT tools and methods to collect, process, consolidate, store, and secure data from sources that are often fragmented and inconsistent. A modern organization needs to manage a variety of information which goes beyond the structured types like numbers and texts to include semi-structured and unstructured contents such as video and sound. The digital library includes content from social media, texts, photos, videos, music, documents, address books, events, and downloads. Maintaining—updating, expanding, porting—an organization's digital library's contents on a variety of platforms is the task of Information Management. Specifically, Information Management deals with how information is organized, stored, and secured, and the speed and ease with which it is captured, analyzed and reported.

2. Why do organizations still have information deficiency problems?
Over many decades, changes in technology and the information companies require, along with different management teams, changing priorities, and increases or decreases in IT investments as they compete with other demands on an organization's budget, have all contributed. Other common reasons include: data silos (information trapped in departments' databases), data lost or bypassed during transit, poorly designed user interfaces requiring extra effort from users, non-standardized data formats, and fast-moving changes in the type of information desired, particularly unstructured content, requiring expensive investments.

3. What is a data silo?
A data silo is one of the data deficiencies that can be addressed. It refers to the situation where the databases belonging to different functional units (e.g., departments) in an organization are not shared between the units because of a lack of integration. Data silos support a single function and therefore do not support the cross-functional needs of an organization. The lack of sharing and exchange of data between functional units raises issues regarding reliability and currency of data, requiring extensive verification to be trusted. Data silos exist when there is no overall IT architecture to guide IS investments, data coordination, and communication.

4. Explain KPIs and give an example.
KPIs are performance measurements. These measures demonstrate the effectiveness of a business process at achieving organizational goals. KPIs present data in easy-to-comprehend and comparison-ready formats. KPIs help reduce the complex nature of organizational performance to a small number of understandable measures.

Examples of key comparisons are actual vs. budget, actual vs. forecasted, and this year vs. prior years.

5. What three factors are driving collaboration and information sharing?
Forrester (forrester.com) identified three factors driving the trend toward collaboration and information sharing technology. These are:
- Global, mobile workforce (a growing number of employees telecommute)
- Mobility-driven consumerization (cloud-based collaboration solutions are on the rise)
- Principle of any (there is growing need to connect anybody anytime anywhere and on any device)

6. What are the business benefits of information management?
The following four benefits have been identified:
- Improves decision quality (due to timely response using reliable data)
- Improves the accuracy and reliability of management predictions ("what is going to happen" as opposed to financial reporting on "what has happened.")
- Reduces the risk of noncompliance (due to improved compliance with regulation resulting from better information quality and governance), and
- Reduces the time and cost of locating relevant information (due to savings in time and effort through integration and optimization of repositories)

2.2 Enterprise Architecture and Data Governance

1. Explain the relationship between complexity and planning. Give an example.

As enterprise information systems become more complex, the importance of long-range IT planning increases dramatically. Companies cannot simply add storage, new apps, or data analytics on an as needed basis and expect those additions to work with the existing systems. The relationship between complexity and planning is easier to see in physical things such as skyscrapers and transportation systems. If you are constructing a simple cabin in a remote area, you do not need a detailed plan for expansion or to make sure that the cabin fits into its environment. If you are building a simple, single-user, non-distributed system, you would not need a well-thought out growth plan either. Therefore, it is no longer feasible to manage big data, content from mobiles and social networks, and data in the cloud without the well-designed set of plans, or blueprint, provided by EA. The EA guides and controls software add-ons and upgrades, hardware, systems, networks, cloud services, and other digital technology investments.

2. Explain Enterprise architecture.

Enterprise architecture (EA) is the way IT systems and processes are structured. EA is an ongoing process of creating, maintaining, and leveraging IT. It helps to solve two critical challenges: where an organization is going and how it will get there. EA helps, or impedes, day-to-day operations and efforts to execute business strategy.

There are two problems that the EA is designed to address:

- IT systems' complexity. IT systems have become unmanageably complex and expensive to maintain.
- Poor business alignment. Organizations find it difficult to keep their increasingly expensive IT systems aligned with business needs.
- EA is the roadmap that is used for controlling the direction of IT investments and it is a significant item in long-range planning. It is the blueprint that guides the build out of overall IT capabilities consisting of four sub-architectures (see question 3). EA defines the vision, standards, and plan that guide the priorities, operations, and management of the IT systems supporting the business.

3. What are the four components of EA?

The four components are:

- Business Architecture (the processes the business uses to meet its goals);
- Application architecture (design of IS applications and their interactions);
- Data architecture (organization and access of enterprise data);
- Technical architecture (the hardware and software infrastructure that supports applications and their interactions)

4. What are the business benefits of EA?

EA cuts IT costs and increases productivity by giving decision makers access to information, insights, and ideas where and when they need them.

EA determines an organization's competitiveness, flexibility, and IT economics for the next decade and beyond. That is, it provides a long-term view of a company's processes, systems, and technologies so that IT investments do not simply fulfill immediate needs.

EA helps align IT capabilities with business strategy—to grow, innovate, and respond to market demands, supported by an IT practice that is 100 percent in accord with business objectives.

EA can reduce the risk of buying or building systems and enterprise apps that are incompatible or unnecessarily expensive to maintain and integrate.

5. How can EA maintain alignment between IT and business strategy?

EA starts with the organization's target–where it is going—not with where it is. Once an organization identifies the strategic direction in which it is heading and the business drivers to which it is responding, this shared vision of the future will dictate changes in business, technical, information, and solutions architectures of the enterprise, assign priorities to those changes, and keep those changes grounded in business value. EA guides and controls software add-ons and upgrades, hardware, systems, networks, cloud services, and other digital technology investments which are aligned with the business strategy.

6. What are the two ways that data are used in an organization?

Data are used in an organization for running the business (transactional or operational use) and for improving the business (analytic use.)

7. What is the function of data governance?

Data governance is the process of creating and agreeing to standards and requirements for the collection, identification, storage, and use of data. The success of every data-driven strategy or marketing effort depends on data governance. Data governance policies must address structured, semi-structured, and unstructured data (discussed in Section 2.3) to ensure that insights can be trusted.

Data governance allows managers to determine where their data originates, who owns them, and who is responsible for what—in order to know they can trust the available data when needed. Data governance is an enterprise-wide project because data cross boundaries and are used by people throughout the enterprise.

8. Why has interest in data governance and MDM increased?

As data sources and volumes continue to increase, so does the need to manage data as a strategic asset in order to extract its full value. Making business data consistent, trusted, and accessible across the enterprise is a critical first step in customer-centric business models. With appropriate data governance and MDM, managers are able to extract maximum value from their data, specifically by making better use of opportunities that are buried within behavioral data. Strong data governance is needed to manage the availability, usability, integrity, and security of the data used throughout the enterprise so that data are of sufficient quality to meet business needs.

9. What role does personal conflict or politics play in the success of data governance?

There may be a culture of distrust between technology and employees in an organization. To overcome this, there must be a genuine commitment to change. Such a commitment must come from senior management. A methodology, such as data governance, cannot solve people problems. It only provides a framework in which such problems can be solved.

2.3 Information Systems: The Basics

1. Contrast data, information, and knowledge.

Data, or raw data, refers to a basic description of products, customers, events, activities, and transactions that are recorded, classified, and stored. Data are the raw material from which information is produced and the quality, reliability and integrity of the data must be maintained for the information to be useful.

Information is data that has been processed, organized, or put into context so that it has meaning and value to the person receiving it.

Knowledge consists of data and/or information that have been processed, organized, and put into context to be meaningful, and to convey understanding, experience, accumulated learning, and expertise as they apply to a current problem or activity.
Define TPS and give an example.

2. Define TPS and give an example.

Transaction processing systems are designed to process specific types of data input from ongoing transactions. TPSs can be manual, as when data are typed into a form on a screen, or automated by using scanners or sensors to capture data.

Organizational data are processed by a TPS--sales orders, payroll, accounting, financial, marketing, purchasing, inventory control, etc. Transactions are either:

- Internal transactions: Transactions that originate from within the organization or that occur within the organization. Examples are payroll, purchases, budget transfers, and payments (in accounting terms, they're referred to as accounts payable).
- External transactions: Transactions that originate from outside the organization, e.g., from customers, suppliers, regulators, distributors, and financing institutions.

TPSs are essential systems. Transactions that do not get captured can result in lost sales, dissatisfied customers, and many other types of data errors having financial impact. For example, if accounting issues a check as payment for an invoice (bill) and that check is cashed, if that transaction is not captured, the amount of cash on the financial statements is overstated, the invoice continues to show as unpaid, and the invoice may be paid a second time. Or if services are provided, but not recorded, the company loses that service revenue.

3. When is batch processing used?

Batch processing is used when there are multiple transactions which can be accumulated and processed at one time. These transactions are not as time sensitive as those that need to be processed in real time. The transactions may be collected for a day, a shift, or over another period of time, and then they are processed. Batch processing often is used to process payroll in a weekly or bi-weekly manner. Batch processing is less costly than real-time processing.

4. When are real-time processing capabilities needed?

Online transaction processing (OLTP), or real-time processing, is used when a system must be updated as each transaction occurs. The input device or website for entering transactions must be directly linked to the transaction processing system (TPS). This type of entry is used for more

time sensitive data, such as reservation systems in which the user must know how many seats or rooms are available.

5. Explain why TPSs need to process incoming data before they are stored.
Processing improves data quality, which is important because reports and decisions are only as good as the data they are based on. As data is collected or captured, it is validated to detect and correct obvious errors and omissions.

Data errors detected later may be difficult to correct or time-consuming. You can better understand the difficulty of detecting and correcting errors by considering identity theft. Victims of identity theft face enormous challenges and frustration trying to correct data about them.

6. Define MIS and DSS and give an example of each.
General purpose reporting systems are referred to as management information systems (MIS). Their objective is to provide reports to managers for tracking operations, monitoring, and control.

MIS is used by middle managers in functional areas and provides routine information for planning, organizing, and controlling operations. Types of reports include:
Periodic: reports created to run according to a pre-set schedule, such as daily, weekly, and quarterly.

Exception: reports generated only when something is outside the norm, either higher or lower than expected. An example might be increased sales in a hardware store prior to a hurricane. Ad hoc, or on demand, reports are unplanned reports generated as needed.

Decision support systems (DSS) are interactive applications that support decision making. Configurations of a DSS range from relatively simple applications that support a single user to complex enterprise-wide systems. A DSS can support the analysis and solution of a specific problem, to evaluate a strategic opportunity, or to support ongoing operations. These systems support unstructured and semi-structured decisions, such as whether to make-or-buy-or-outsource products, or what new products to develop and introduce into existing markets. Decision support systems are used by decision makers and managers to combine models and data to solve semi-structured and unstructured problems with user involvement.

To provide such support, DSSs have certain characteristics to support the decision maker and the decision making process.

Three defining characteristics of DSSs are:
• an easy-to-use interactive interface
• models that enable sensitivity analysis, what if analysis, goal seeking, and risk analysis
• data from multiple sources - internal and external sources plus data added by the decision maker who may have insights relevant to the decision situation.

Having models is what distinguishes DSS from MIS. Some models are developed by end users through an interactive and iterative process. Decision makers can manipulate models to conduct

experiments and sensitivity analyses, such as what-if, and goal-seeking. What-if analysis refers to changing assumptions or data in the model to see the impacts of the changes on the outcome. For example, if sales forecasts are based on a 5 percent increase in customer demand, a what if analysis would replace the 5 percent with higher and/or lower demand estimates to determine what would happen to sales if the demands were different. With goal seeking, the decision maker has a specific outcome in mind and needs to figure out how that outcome could be achieved and whether it's feasible to achieve that desired outcome. A DSS also can estimate the risk of alternative strategies or actions.

California Pizza Kitchen (CPK) uses a DSS to support inventory decisions. CPK has 77 restaurants located in various states in the U.S. Maintaining inventory of all restaurants at optimal levels was challenging and time-consuming. A DSS has made it easy for the managers to keep records updated and make decisions. Many CPK restaurants increased sales by 5 percent after implementing a DSS.

7. Why are databases inappropriate for doing data analysis?
Databases are used for recording and processing transactions. Due to the number of transactions, the data in the databases are constantly in a state of change making it difficult to use for complex decision making.

2.4 Data Centers, Cloud Computing, and Virtualization

1. What is a data center?
A data center consists of a large number of network servers (Figure 2.13) used for the storage, processing, management, distribution, and archiving of data, systems, Web traffic, services, and enterprise applications. Data center also refers to the building or facility that houses the servers and equipment.

2. Describe cloud computing.
Cloud computing is the general term for infrastructures that use the Internet and private networks to access, share, and deliver computing resources.

3. What is the difference between data centers and cloud computing?
A main difference between a cloud and data center is that a cloud is an off-premise form of computing that stores data on the Internet. In contrast, a data center refers to on-premises hardware and equipment that store data within an organization's local network. Cloud services are outsourced to a third-party cloud provider who manages the updates, security, and ongoing maintenance. Data centers are typically run by an in-house IT department.

A data center is owned by the company. Since only the company owns the infrastructure, a data center is more suitable for organizations that run many different types of applications and have complex workloads. A data center, like a factory, has limited capacity. Once it is built, the amount of storage and the workload the center can handle does not change without purchasing and installing more equipment.

A data center is physically connected to a local network, which makes it easier to restrict access to apps and information by only authorized, company-approved people and equipment. However,

the cloud is accessible by anyone with the proper credentials and Internet connection. This accessibility arrangement increases exposure to company data at many more entry and exit points.

Cloud computing is the delivery of computing and storage resources as a service to end-users over a network. With cloud computing, shared resources (such as hard drives for storage) and software apps are provided to computers and other devices on-demand, like a public utility. That is, it's similar to electricity - a utility that companies have available to them on-demand and pay for it based on usage. Cloud systems are scalable. That is, they can be adjusted to meet changes in business needs.

A drawback of the cloud is control because a third party manages it. Companies do not have as much control as they do with a data center.

4. What are the benefits of cloud computing?
Answers may vary.

Many IT infrastructures are extremely expensive to manage and too complex to easily adapt. Because cloud computing resources are scalable "on demand", this increases IT agility and responsiveness. In a business world where first movers gain the advantage, IT responsiveness and agility provide a competitive edge. Access to data in the cloud is possible via any device that can access the Internet, allowing users to be more responsive and productive.

Cloud services are outsourced to a third-party cloud provider who manages the updates, security, and ongoing maintenance, including backups and disaster recovery, relieving this burden from the business. The business saves the costs of increased staff, power consumption, and disposal of discontinued hardware. Additionally, cloud services significantly reduce IT costs and complexity through improved workload optimization and service delivery.

5. How can cloud computing solve the problem of managing software licenses?
Cloud computing makes it more affordable for companies to use services that in the past would have been packaged as software and required buying, installing and maintaining on any number of individual machines. A major type of service available via the cloud is called software as a service, or SaaS.

Because applications are hosted by vendors and provided on demand, rather than via physical installations or seat licenses (a key characteristic of cloud computing), applications are accessed online through a Web browser instead of stored on a computer. Companies pay only for the computing resources or services they use. Vendors handle the upgrades and companies do not purchase or manage software licenses. They simply pay for the number of concurrent users.

6. What is an SLA? Why are SLAs important?
An SLA is a negotiated agreement between a company and service provider that can be a legally binding contract or an informal contract.

An SLA serves "as a means of formally documenting the service(s), performance expectations, responsibilities, and limits between cloud service providers and their users. A typical SLA describes levels of service using various attributes such as: availability, serviceability, performance, operations, billing, and penalties associated with violations of such attributes." (Cloud Standards Customer Council, 2012, pp. 5–6.)

7. What factors should be considered when selecting a cloud vendor or provider?
From Table 2.5:

Factors	Examples of Questions to Be Addressed
Delays	What are the estimated server delays and network delays?
Workloads	What is the volume of data and processing that can be handled during a specific amount of time?
Costs	What are the costs associated with workloads across multiple cloud computing platforms?
Security	How are data and networks secured against attacks? Are data encrypted and how strong is the encryption? What are network security practices?
Disaster recovery and business continuity	How is service outage defined? What level of redundancy is in place to minimize outages, including backup services in different geographical regions? If a natural disaster or outage occurs, how will cloud services be continued?
Technical expertise	Does the vendor have expertise in your industry or and understanding business processes? Does the vendor understand what you need to do and have the technical expertise to fulfill those obligations?
Insurance in case of failure	Does the vendor provide cloud insurance to mitigate user losses in case of service failure or damage? This is a new and important concept.
Third-party audit, or an unbiased assessment of the ability to rely	Can the vendor show objective proof with an audit that it can

on the service provided by the vendor	live up to the promises it is making?

8. When are private clouds used instead of public clouds?
Companies or government agencies set up their own private clouds when they need stronger security and control for regulated industries and critical data.

9. Explain three issues that need to be addressed when moving to cloud computing or services.
Issues that need to be addressed when moving to public cloud computing or services include:

Infrastructure issues – Cloud computing runs on a shared infrastructure so there is less customization for a company's specific requirements. The network and WAN (wide area network) become more critical in the IT infrastructure. Network bandwidth is also an issue as enough is needed to support the increase in network traffic. With cloud computing, it may be more difficult to get to the root of performance problems, like the unplanned outages that occurred with Google's Gmail and Workday's human resources apps. The trade-off is cost vs. control.

Disruption issues – There is a risk of disrupting operations or customers in the process of moving operations to the cloud.

Management issues – Putting part of the IT architecture or workload into the cloud requires different management approaches, different IT skills, and knowing how to manage vendor relationships and contracts.

(The astute student may also describe the following:
Strategic issues such as deciding which workloads to export to the cloud; which set of standards to follow for cloud computing; how to resolve privacy and security issues; and how departments or business units will get new IT resources.)

10. How does a virtual machine (VM) function?
A virtual machine (VM) is a software layer that runs its own Operating System (OS) and apps as if it were a physical computer. A VM behaves exactly like a physical computer and contains its own virtual (software based) CPU, RAM, hard drive and Network Interface Card. An OS cannot tell the difference between a VM and a physical machine, nor can apps or other computers on a network tell the difference. (See Fig 2.13 for details)

11. Explain virtualization.
Virtualization is a concept that has several meanings in IT and therefore several definitions. The major type of virtualization is hardware virtualization, which remains popular and widely used. Virtualization is often key part of an enterprise's disaster recovery plan. In general, virtualization separates business applications and data from hardware resources. This separation allows companies to pool hardware resources—rather than to dedicate servers to applications—and assign those resources to applications as needed. The major types of virtualization are the following:

- Storage virtualization is the pooling of physical storage from multiple network storage devices into what appears to be a single storage device that is managed from a central console.
- Network virtualization combines the available resources in a network by splitting the network load into manageable parts, each of which can be assigned (or reassigned) to a particular server on the network.
- Hardware virtualization is the use of software to emulate hardware or a total computer environment other than the one the software is actually running in. It allows a piece of hardware to run multiple operating system images at once. This kind of software is sometimes known as a virtual machine.
- Virtualization increases the flexibility of IT assets, allowing companies to consolidate IT infrastructure, reduce maintenance and administration costs, and prepare for strategic IT initiatives. Virtualization is not primarily about cost-cutting, which is tactical reason. More importantly, for strategic reasons, virtualization is used because it enables flexible sourcing, and cloud computing.

12. What are the characteristics and benefits of virtualization?
Memory-intensive: VMs need a huge amount of RAM (random access memory, or primary memory) because of their massive processing requirements.

Energy-efficient: VMs minimize energy consumed running and cooling servers in the data center— representing up to a 95 percent reduction in energy use per server.

Scalability and load balancing: Virtualization provides load balancing to handle the demand for requests to the site. The VMware infrastructure automatically distributes the load across a cluster of physical servers to ensure the maximum performance of all running VMs.

13. When is load balancing important?
When a big event happens, such as the Super Bowl, millions of people hit a Web site at the same time. Virtualization provides load balancing to handle the demand for requests to the site. Load balancing is key to solving many of today's IT challenges.

2.5 Virtualization and VM (Virtual Machines)

1. What is SaaS?
Any software that is provided on demand is referred to as software as a service, or SaaS.
SaaS is a widely used model in which software is available to users as needed. Specifically, in SaaS, a service provider hosts the application at its data center and customers access it via a standard Web browser. Other terms for SaaS are on-demand computing and hosted services. The idea is basically the same: Instead of buying and installing expensive packaged enterprise applications, users can access software apps over a network, with an Internet browser being the only necessity. A SaaS provider licenses an application to customers either on-demand, through a subscription, based on usage (pay-as-you-go), or increasingly at no cost when the opportunity exists to generate revenue from advertisements or through other methods.

2. Describe the cloud computing stack.
The cloud computing stack consists of the following three categories:

- SaaS apps are designed for end-users.
- PaaS is a set of tools and services that make coding and deploying these apps faster and more efficient.
- IaaS consists of hardware and software that power computing resources— servers, storage, operating systems, and networks.

See Figure 2.19 for a graphical representation.

3. What is PaaS?
PaaS provides a standard unified platform for app development, testing, and deployment, thus benefiting software development. This computing platform allows the creation of Web applications quickly and easily without the complexity of buying and maintaining the underlying infrastructure. Without PaaS, the cost of developing some apps would be prohibitive. The trend is for PaaS to be combined with IaaS.

4. What is IaaS?
Infrastructure as a service (IaaS) is a way of delivering cloud computing infrastructure as an on-demand service. Rather than purchasing servers, software, data center space, or networks, companies instead buy all computing resources as a fully outsourced service.

5. Why is DaaS growing in popularity?
The DaaS model is growing in popularity as data become more complex, difficult, and expensive to maintain.

Data as a service (DaaS) enables data to be shared among clouds, systems, apps, and so on regardless of the data source or where they are stored. DaaS makes it easier for data architects to select data from different pools, filter out sensitive data, and make the remaining data available on-demand. A key benefit of DaaS is the elimination of the risks and burdens of data management to a third-party cloud provider.

6. How might companies risk violating regulation or compliance requirements with cloud services?
Companies are frequently adopting software, platform, infrastructure, data management and starting to embrace mobility as a service and big data as a service because they typically no longer have to worry about the costs of buying, maintaining, or updating their own data servers. Regulations mandate that confidential data be protected regardless of whether the data are on-premises on in the cloud. Therefore, a company's legal department needs to get involved in these IT decisions. Put simply, moving to cloud services is not simply an IT decision because the stakes around legal and compliance issues are very high.

Critical Thinking Questions

1. Why is a strong market position or good profit performance only temporary?
Markets do not stand still and the basis of competition is changing at a faster pace. By the time a company's financial performance starts tapering off; it might be too late to start building new

market-relevant capabilities. Effective ways to thrive over the long term are to launch new business models and strategies or devise new ways to outperform competitors.

2. Explain the difference between customer-centric and product-centric business models.
Customer-centric business models strive to create the best solution or experience for the customer. In contrast, product-centric models are internally focused on creating the best product. The ability to connect with customers, anticipate their needs, and provide flawless customer service is essential to loyalty and long-term growth because social and mobile technology is forcing businesses to offer excellent customer experiences (CX) across every available touchpoint, including chat, video, mobile apps, and alerts (Figure 2.2).

3. Assume you had:
   a. **A tall ladder with a sticker that listed a weight allowance only 5 pounds more than you weighed. You know the manufacturer and model number.**
   b. **Perishable food with an expiration date 2 days into the future.**
   c. **A checking account balance that indicated you had sufficient funds to cover the balance due on an account.**

In all three cases, you cannot trust the data to be exactly correct. The data could be incorrect by about 20 percent. How might you find the correct data for each instance? Which data might not be possible to verify? How does dirty data impact your decision making?
Answers will vary.

Since you have the manufacturer and model number for the ladder, you could search the Internet for contact information of the manufacturer and check with them on the accuracy of the information on the sticker.

Because perishable food packaged with an expiration date is packaged by a manufacturer, not the retailer, you could not get this information from the retailer. You may or may not even know the manufacturer, depending on the packaging. If known, you might contact the manufacturer as to the accuracy of their data, much as with the ladder, although there may be additional data needed, such as a lot number, which typically would be on the label. However, by the time you might be able to contact the manufacturer, the expiration date may have passed.

The balance of a checking account at any point in time may not represent the amount of spendable money remaining in the account. Any outstanding checks, for example, need to be included in the determination. Being aware of any required minimums also will help prevent additional charges. Being aware of all charges, passed or pending, by balancing your account to regular statements, remembering any required minimums, and keeping track of charges and the current balance between statements will allow you to determine your amount of truly spendable money.

In the case of the perishable food, the data might not be possible to verify.

Dirty data may cause one to feel a need to mistrust what is presented. This might cause you to not purchase such a ladder or the food in the first place. If already purchased, then you might not use the ladder (unless you lost weight!) or might throw away the perishable food rather than consume it. You may feel compelled to carefully balance your checking account and track all charges – a good thing! If, after correctly balancing and tracking charges, you determine that the bank's data is not trustworthy, you likely would change banks quickly.

4. If business data are scattered throughout the enterprise and not synched until the end of the month, how does that impact day-to-day decision making and planning?
Answers will vary.

In such a case, day-to-day decision making and planning cannot rely on this business data, except, perhaps, just after the end-of-month synch. This would mean that business decisions and planning based upon such data would have to be made on an untimely basis or made with considerable risk. This would be extremely risky for critical business decisions and future plans.

5. Assume a bank's data are stored in silos based on financial product—checking accounts, saving accounts, mortgages, and auto loans, and so on. What problems do these data silos create for the bank's managers?
Such silos are unable to share or exchange data, and they cannot consistently be updated. When data are inconsistent across multiple enterprise applications, data quality cannot (and should not) be trusted without extensive verification. Data silos support a single function, and as a result, do not support an organization's cross-functional needs.

With data silos, bank managers waste a lot of time and effort to include data from these silos in their business analysis. There is a greater chance of leaving out routinely messy but important data in their analysis.

6. Why do managers and workers still struggle to find information that they need to make decisions or take action despite advances in digital technology? That is, what causes data deficiencies?
Some reasons for data deficiencies are: data silos; lost or bypassed data (not captured or captured but not in sufficient detail); poorly designed interfaces which increase the risk of errors from misunderstanding the data or ignoring them; non-standardized data formats, where comparisons cannot be made without conversion of the data; and the inability to hit moving targets (ISs cannot respond to rapidly changing business needs.)

7. According to a Tech CEO Council Report, Fortune 500 companies waste $480 billion every year on inefficient business processes. What factors cause such huge waste? How can this waste be reduced?
Reasons for this waste include poor and complex internal processes, data silos and siloed global applications, and applications not aligned to business operations.

Reduction in such waste can be made by business process improvement. Companies need to be able to analyze and adapt their business processes quickly, efficiently, and without disruption.

8. Explain why organizations need to implement enterprise architecture (EA) and data governance.
Proper EA and data governance can help organizations to improve in both financial and non-financial performance. Success of EA and data governance is measured in financial terms of profitability and return on investment (ROI), and in the nonfinancial terms of improved customer satisfaction, faster speed to market, and lower employee turnover.

9. What two problems can EA solve?
There are two problems that the EA is designed to address:
- IT systems' complexity. IT systems have become unmanageably complex and expensive to maintain.
- Poor business alignment. Organizations find it difficult to keep their increasingly expensive IT systems aligned with business needs.

10. Name two industries that depend on data governance to comply with regulations or reporting requirements. Given an example of each.
Answers may vary, but should include two of the following three listed in the text.
- Food industry. In the food industry, data governance is required to comply with food safety regulations. Food manufacturers and retailers have sophisticated control systems in place so that if a contaminated food product, such as spinach or peanut butter, is detected, they are able to trace the problem back to a particular processing plant or even the farm at the start of the food chain.
- Financial services industry. In the financial services sector, strict reporting requirements of the Dodd–Frank Wall Street Reform and Consumer Protection Act of 2010 are leading to greater use of data governance. The Dodd–Frank Act regulates Wall Street practices by enforcing transparency and accountability in an effort to prevent another significant financial crisis like the one that occurred in 2008.
- Health-care industry. Data are health care's most valuable asset. Hospitals have mountains of electronic patient information. New health-care accountability and reporting obligations require data governance models for transparency to defend against fraud and to protect patients' information.

11. Why is it important for data to be standardized? Give an example of unstandardized data.
When users are presented with data in inconsistent or nonstandardized formats, errors increase. Attempts to compare or analyze data are more difficult and take more time. For example, if the Northeast division reports weekly gross sales revenues per product line and the Southwest division reports monthly net sales per product, you cannot compare their performance without converting the data to a common format.

12. Why are TPSs critical systems?
TPSs are critical systems because transactions that do not get captured can result in lost sales, dissatisfied customers, and many other types of data errors. For example, if accounting issues a check as payment for an invoice (bill), and that transaction is not captured, the amount of cash on the financial statements is overstated and the invoice may be paid a second time. Or if services are provided, but not recorded, the company loses that service revenue.

13. Explain what is meant by data volatility. How does it affect the use of databases for data analysis?
Data volatility means that data in databases are constantly in use or in the process of being updated. In order to effectively analyze data, the data must be extracted, transformed and loaded (ETL) from the operational databases into a data warehouse where the operational data are better formatted for analysis.

14. Discuss why the cloud acts as the great IT delivery frontier.
Cloud is still a relatively new and evolving business model, providing many benefits along with a number of risks. It is a new, flexible, and efficient way to deliver IT services, thus a compellingly great delivery mechanism, especially to an increasingly mobile society with people accessing data and applications using any type of mobile device from anywhere at any time.

The relative newness, the benefits along with the risks, and the evolving of the business model make it a frontier, as frontiers always have associated risks and unknown factors, yet promise great rewards.

15. What are the immediate benefits of cloud computing?
The cloud provides a dynamic infrastructure which helps companies become more agile and responsive while significantly reducing IT costs and complexity.

Cloud computing provides: applications for end-users (SaaS); tools for creating and deploying applications more quickly and efficiently (PaaS); computing resources such as servers, storage, operating systems, and networks; and enables the sharing of data on a selective basis (DaaS.) Cloud systems are scalable so that resources easily can be added as the company need grows. The cloud has greatly expanded the options for enterprise IT infrastructures because any device that accesses the Internet can access, share, and deliver data. Finally, managing software and their licenses involves deploying, provisioning, and updating them—all of which are time-consuming and expensive. Cloud computing overcomes these problems.

16. What are the functions of data centers?
Data centers are used for the storage, processing, management, distribution, and archiving of data, systems, Web traffic, services, and enterprise applications. Data center also refers to the building or facility that houses the servers and equipment.
Since only the company owns the infrastructure, a data center is more suitable for organizations that run many different types of applications and have complex workloads.

17. What factors need to be considered when selecting a cloud vendor?
From Table 2.5:

Factors	Examples of Questions to Be Addressed
Delays	What are the estimated server delays and network delays?
Workloads	What is the volume of data and processing that can be handled during a specific amount of

	time?
Costs	What are the costs associated with workloads across multiple cloud computing platforms?
Security	How are data and networks secured against attacks? Are data encrypted and how strong is the encryption? What are network security practices?
Disaster recovery and business continuity	How is service outage defined? What level of redundancy is in place to minimize outages, including backup services in different geographical regions? If a natural disaster or outage occurs, how will cloud services be continued?
Technical expertise	Does the vendor have expertise in your industry or and understanding business processes? Does the vendor understand what you need to do and have the technical expertise to fulfill those obligations?
Insurance in case of failure	Does the vendor provide cloud insurance to mitigate user losses in case of service failure or damage? This is a new and important concept.
Third-party audit, or an unbiased assessment of the ability to rely on the service provided by the vendor	Can the vendor show objective proof with an audit that it can live up to the promises it is making?

18. What protection does an effective SLA provide?
An SLA is "a means of formally documenting the service(s), performance expectations, responsibilities, and limits between cloud service providers and their users. A typical SLA describes levels of service using various attributes such as: availability, serviceability, performance, operations, billing, and penalties associated with violations of such attributes." (Cloud Standards Customer Council, 2012, pp. 5–6.)

19. Why is an SLA a legal document?
According to the text, "An SLA is a negotiated agreement between a company and service provider that can be a legally binding contract or an informal contract."

Students may not know that a contract, even if informal, is still legally binding if there is an exchange for value, such as money for services. Unless written (and perhaps notarized), it may be unenforceable.

20. How can virtualization reduce IT costs while improving performance?
VMs minimize energy consumed running and cooling servers in the data center—up to a 95 percent reduction in energy use per server. The VMware infrastructure automatically distributes the load across a cluster of physical servers to ensure the maximum performance of all running VMs. Virtualization consolidates servers, which reduces the cost of servers, makes more efficient use of data center space, and reduces energy consumption. All of these factors reduce the total cost of ownership (TCO). Over a three-year lifecycle, a VM costs approximately 75 percent less to operate than a physical server.

Online and Interactive Exercises

21. When selecting a cloud vendor to host your enterprise data and apps, you need to evaluate the service level agreement (SLA).
    a. **Research the SLAs of two cloud vendors, such as Rackspace, Amazon, or Google.**
        Answers will vary.
    b. **For the vendors you selected, what are the SLAs' uptime percent? Expect them to be 99.9 percent or less.**
        Answers will vary.
    c. **Does each vendor count both scheduled downtime and planned downtime toward the SLA uptime percent?**
        Answers will vary.
    d. **Compare the SLAs in terms of two other criteria.**
        Answers will vary.
    e. **Decide which SLA is better based on your comparisons.**
        Answers will vary.
    f. **Report your results and explain your decision.**
        Answers will vary.

22. Many organizations initiate data governance programs because of pressing compliance issues that impact data usage. Organizations may need data governance to be in compliance with one or more regulations, such as the Gramm–Leach Bliley Act (GLB), HIPAA, Foreign Corrupt Practices Act (FCPA), Sarbanes Oxley Act, and several state and federal privacy laws.
    a. **Research and select two U.S. regulations or privacy laws.**
    b. **Describe how data governance would help an enterprise comply with these regulations or laws.**
Answers will vary.

23. Visit eWeek.com Cloud Computing Solutions Center for news and reviews at eweek.com/c/s/Cloud-Computing/. Select one of the articles listed under Latest Cloud Computing News. Prepare an executive summary of the article.

Answers will vary.

24. Visit Rackspace at rackspace.com/ and review the company's three types of cloud products. Describe each of those cloud solutions.
Answers will vary.

25. Visit oracle.com. Describe the types of virtualization services offered by Oracle. Oracle provides several virtualization products and solutions including:
Server virtualization provides over 100 pre-built templates to provide fully integrated enterprise management. Oracle's server virtualization products support x86 and SPARC architectures and a variety of workloads such as Linux, Windows, and Oracle Solaris. In addition to solutions that are hypervisor-based, Oracle also offers virtualization built in to hardware and Oracle operating systems.
Data center virtualization reduces the complexity of enterprise management providing centralized control of the entire computing environment. This allows you to keep a central view of hardware, virtual machines and applications—including cloud-based services.

Desktop virtualization provides secure, anywhere access to Oracle Applications and other enterprise software and enables virtualized operating systems on standard client devices. It can be used to securely access nearly any application over the network from standard PCs, Macs, and iPad. It can also be used to run multiple operating systems on laptops and desktop PCs for development and testing.

Network virtualization enables distributed enterprises to deliver trusted connectivity services with less complexity, in significantly less time, and at greatly reduced cost.

26. Visit YouTube.com and search for two videos on virtualization. For each video, report what you learned. Specify the complete URL, video title, who uploaded the video and the number of views.
Answers will vary.

Apply IT Concepts to Business Decisions

27. Financial services firms experience large fluctuations in business volumes because of the cyclical nature of financial markets. These fluctuations are often caused by crises— such as the subprime mortgage problems, the discovery of major fraud, or a slowdown in the economy. These fluctuations require that executives and IT leaders have the ability to cut spending levels in market downturns and quickly scale up when business volumes rise again. Research SaaS solutions and vendors for the financial services sector. Would investment in SaaS help such firms align their IT capacity with their business needs and also cut IT costs? Explain your answer.
Answers will vary but all should address how solutions would help cut IT costs and align IT capacity with business needs.

28. Despite multimillion-dollar investments, many IT organizations cannot respond quickly to evolving business needs. Also, they cannot adapt to large-scale shifts like mergers,

sudden drops in sales, or new product introductions. Can cloud computing help organizations improve their responsiveness and get better control of their IT costs? Explain your answer.

Yes. Improving responsiveness to changes in the business environment and controlling IT costs are some of the advantages of cloud computing. Cloud computing is scalable, providing "convenient, on-demand network access to a shared pool of configuration computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction" (NIST, 2012)." This enables increased responsiveness. Reduction of IT costs are achieved through reduced utility costs, reduced hardware costs (purchase, maintenance, operations), and a possible reduction in application and development costs, along with possibly reduced personnel costs in these areas.

29. Describe the relationship between enterprise architecture and organizational performance.

In order to leverage IT capability for competitive advantage and growth, a guide is needed. EA is that guide.

Every enterprise has a core set of information systems and business processes that execute the transactions that keep it in business. One of the problems that EA is designed to address is poor business alignment of expensive IT systems with business needs. Having the right architecture in place cuts IT costs significantly and increases productivity by giving decision makers access to information, insights, and ideas where and when they need them. The success of EA (and of data governance) is measured in financial terms of profitability and return on investment (ROI), and in nonfinancial terms of improved customer satisfaction, faster speed to market, and lower employee turnover. Thus, EA can improve an organization's performance, helping it to operate more efficiently and effectively.

30. Identify four KPIs for a major airline (e.g., American, United, Delta) or an automobile manufacturer (e.g., GM, Ford, BMW). Which KPI would be the easiest to present to managers on an online dashboard? Explain why.

Answers could vary. Make sure the KPIs are related to goals such as cost reduction, and profit optimization. Airlines: Trends on percentage of seat-occupancy and time lost in breakdown-maintenance; Auto Manufacturer: sales trend by car model, and trends in recall.

Business Case

Case 1 – Opening Case: Detoxing Dirty Data with Data Governance at Intel Security

1. What is the difference between customer-centric and product-centric business models?
Customer-centric business models strive to create the best solution or experience for the customer. In contrast, product-centric models are internally focused on creating the best product.

2. Explain the business challenges caused by Intel Security's dirty data.

Duplicate customer records and incomplete customer data were harming sales. The company could not effectively cross-sell (sell complimentary products or services) or up-sell (sell more expensive models or features). Opportunities to get customers to renew their software licenses—and keep them loyal—were being lost. Data errors degraded sales forecasts and caused order processing mistakes. Time was wasted trying to find, validate, and correct customer records and manually reconcile month-end sales and calculate sales commissions.

3. What is the function of data governance?
Data governance is the control of enterprise data through formal policies and procedures. A goal of data governance is to provide employees and business partners with high-quality data they trust and can access on demand.

4. Describe the function of master data?
Master data are the business-critical information on customers, products, accounts, and other things that is needed for operations and business transactions.

5. Why is it important to keep data synchronized across disparate systems?
Data stored in disparate systems across an enterprise and not kept synchronized often result in data which is duplicated or in conflict, incorrect, incomplete, non-standardized, outdated or invalid, and unusable. These are all characteristics of dirty data.

6. Why did Intel Security need master data management (MDM)?
Answer

7. How did MDM and data governance enable the company to achieve its vision?
Intel Security management recognized that it needed to implement a best-practices customer-centric business model if it was to become the fastest-growing dedicated security company in the world.
MDM would link and synchronize all critical data from those disparate systems into one file, a master file, which provided a common point of reference. Data governance provided control of enterprise data through formal policies and procedures. Data governance and MDM manage the availability, usability, integrity, and security of the data used throughout the enterprise. Data governance and MDM have improved the quality of Intel Security's customer data, which was essential for its customer-centric business model.

8. What benefits did the company achieve as a result of implementing data governance and MDM?
With high-quality data, the company is able to identify up-sell and cross-sell sales opportunities, increasing revenue. Best practices for customer data management improved customer experiences which translated into better customer retention and acquisition.

Case 2 – Business Case: Data Chaos Creates Risk

1. What might happen when each line of business, division, and department develops its own IT apps?

When each line of a business develops its own IT applications, it often is done without a thorough analysis of its relationship with other departmental or divisional systems. This arrangement may lead to the business having IT groups which specifically manage a particular type of application suite or data silo for a particular department or division.

2. What are the consequences of poorly managed apps?
When apps are not well managed, they can generate terabytes of irrelevant data, causing the business to drown in such data. This data chaos could lead to errors. In an effort to manage excessive and massive amounts of data, there is increased risk of relevant information being lost (missing) or inaccurate — that is, faulty or dirty data. Another risk is a data breach.

3. What two risks are posed by data chaos? Explain why.
One risk is faulty, or dirty, data. Because apps are not well managed they can generate terabytes of irrelevant data, causing a business to drown in such data. With excessive and massive amounts of data, there is increased risk of relevant information being lost (missing) or inaccurate. Another risk is data breaches. With large amounts of data spread over various data silos, the security for each not well managed, the opportunities for data breeches are increased.

4. What are the functions of data governance in the healthcare sector?
Answers may vary.

The healthcare industry must comply with regulations or reporting requirements, defend against fraud, and protect patients' information.

Data governance in healthcare helps to reduce or eliminate faulty data and data chaos. Inaccessible or unavailable in separate clinical data silos and valuable "messy" data are routinely left out. Most health-care organizations are drowning in data, yet they cannot get reliable, actionable insights from these data.

Good data governance helps to prevent data breaches.

5. Why is it important to have executives involved in data governance projects?
Data governance is the control of enterprise data through formal policies and procedures. Data governance is an enterprise-wide project because data cross boundaries and are used by people throughout the enterprise. Proper governance reduces risks for the organization.
Executives must be involved in order to determine who owns which data and who can access and/or update that data, particularly data which crosses departmental boundaries. Also, to obtain compliance to data policies and procedures, there must be support from executives who have oversight across the organization.

6. List and explain the costs of data failure.
Costs of data failure include:

Rework - if data are incorrect, someone has to go to the original source and re-enter the correct data; loss of business: not having correct data for customers, prospective customers, services, or products can cause an inability to be customer-centric and therefore cause a loss of business; safety errors (for patients, in healthcare; for the public, in engineering); malpractice lawsuits (in various fields, such as healthcare, legal, or engineering); and a delay in receiving payments: when billing or payment data are not available, or inaccurate, delays can occur before correct payment is received.

7. Calculate the CIDDA over time:
   Q1: G = 40%, M = 50%, TS = 20%
CIDDA = G x M x TS = 40% x 50% x 20% = 4%
   Q2: G = 50%, M = 55%, TS = 30%
CIDDA = G x M x TS = 50% x 55% x 30% = 8.25%
   Q3: G = 60%, M = 60%, TS = 40%
CIDDA = G x M x TS = 60% x 60% x 40% = 14.4%
   Q4: G = 60%, M = 70%, TS = 45%
CIDDA = G x M x TS = 60% x 70% x 45% = 18.9%

8. Why are data the most valuable asset in health care?
Answers may vary.

Due to health-care accountability and reporting obligations, and the huge amount of personal data in electronic form, data becomes very valuable in healthcare, not only for serving the patients' needs, but also for keeping the organization operational and protected from lawsuits. "Beginning in 2012, hospitals have been penalized for high re-admission rates with cuts to the payments they receive from the government (Miliard, 2011)." Managers need to be able to "access an integrated view of relevant clinical and operational information to drive more informed decision making. For example, by predicting which patients might be readmitted, we can reduce costly and preventable readmissions, decrease mortality rates, and ultimately improve the quality of life for our patients" (Miliard, 2011).

Case 3 – Video Case: Cloud Computing: Three Case Studies

1. Explain the value or benefits of each organization's cloud investment.
The three case studies are about cloud computing usage: SaaS (Software as a Service), IaaS (Infrastructure as a Service), PaaS (Platform as a Service).

In the first one, students at a University access Google Apps e-mail (instead of a university e-mail server) from their university portal via a link. Since it is branded, they see their college logo. The value to the university is that there is no need to have an e-mail server of their own.
In the second case, an organization leverages cloud computing through Terremark for managing their virtual server (infrastructure) requirements which may be changing dynamically. This optimizes the cost of infrastructure; there is no need to pay for peak-level resource requirements all of the time.

In the third one, the connection is from the cloud (a Facebook App) to a company's API (a service offered to its customers) on its platform, done in order to attract more customers.