MODUL SITEM INFORMASI MANAGEMEN

(MAN 611)

MODUL PERTEMUAN 09

Big Data and Business Intelligence

DISUSUN OLEH

Dr. Fransiskus Adikara, S.Kom, MMSI

UNIVERSITAS ESA UNGGUL

2019

BIG DATA AND BUSINESS INTELLIGENCE

1. Kemampuan Akhir Yang Diharapkan

After reading this session, you will be able to answer the following questions:

1. What are the principal tools and technologies for accessing informa- tion from databases to improve business performance and decision making?
2. Why are information policy, data administration, and data quality assurance essential for managing the firm's data resources?

2. Uraian dan Contoh

2.1. USING DATABASES TO IMPROVE BUSINESS

A. PERFORMANCE AND DECISION MAKING

Businesses use their databases to keep track of basic transactions, such as paying suppliers, processing orders, keeping track of customers, and paying employees. But they also need databases to provide information that will help the company run the business more efficiently, and help managers and employees make better decisions. If a company wants to know which product is the most popular or who is its most profitable customer, the answer lies in the data.

B. THE CHALLENGE OF BIG DATA

Up until about five years ago, most data collected by organizations consisted of transaction data that could easily fit into rows and columns of relational database management systems. Since then, there has been an explosion of data from Web traffic, e-mail messages, and social media content (tweets, status messages), as well as machine-generated data from sensors (used in smart meters, manufacturing sensors, and electrical meters) or from electronic trading systems. These data may be unstructured or semi-structured and thus not suitable for relational database products that organize data in the form of columns and rows. We now use the term big data to describe these datasets with volumes so huge that they are beyond the ability of typical DBMS to capture, store, and analyze.

Big data doesn’t refer to any specific quantity, but usually refers to data in the petabyte and exabyte range—in other words, billions to trillions of records, all from different sources. Big data are produced in much larger quantities and much more rapidly than traditional data. For example, a single jet engine is capable of generating 10 terabytes of data in just 30 minutes, and there are more than 25,000 airline flights each day. Even though “tweets” are limited to 140 characters each, Twitter generates over 8 terabytes of data daily. According to the International Data Center (IDC) technology research firm, data are more than doubling every two years, so the amount of data available to organizations is skyrocketing.

Businesses are interested in big data because they can reveal more patterns and interesting anomalies than smaller data sets, with the potential to provide new insights into customer behavior, weather patterns, financial market activity, or other phenomena. However, to derive business value from these data, organizations need new technologies and tools capable of managing and analyzing non-traditional data along with their traditional enterprise data.

C. BUSINESS INTELLIGENCE INFRASTRUCTURE

Suppose you wanted concise, reliable information about current operations, trends, and changes across the entire company. If you worked in a large company, the data you need might have to be pieced together from separate systems, such as sales, manufacturing, and accounting, and even from external sources, such as demographic or competitor data. Increasingly, you might need to use big data. A contemporary infrastructure for business intelligence has an array of tools for obtaining useful information from all the different types of data used by businesses today, including semi-structured and unstructured big data in vast quantities. These capabilities include data warehouses and data marts, Hadoop, in-memory computing, and analytical platforms.

Data Warehouses and Data Marts

The traditional tool for analyzing corporate data for the past two decades has been the data warehouse. A data warehouse is a database that stores current and historical data of potential interest to decision makers throughout the company. The data originate in many core operational transaction systems, such as systems for sales, customer accounts, and manufacturing, and may include data from Web site transactions. The data warehouse extracts current and historical data from multiple operational systems inside the organization. These data are combined with data from external sources and transformed by correcting inaccurate and incomplete data and restructuring the data for management reporting and analysis before being loaded into the data warehouse.

The data warehouse makes the data available for anyone to access as needed, but it cannot be altered. A data warehouse system also provides a range of ad hoc and standardized query tools, analytical tools, and graphical reporting facilities .

Companies often build enterprise-wide data warehouses, where a central data warehouse serves the entire organization, or they create smaller, decentral- ized warehouses called data marts. A data mart is a subset of a data warehouse in which a summarized or highly focused portion of the organization’s data is placed in a separate database for a specific population of users. For example, a company might develop marketing and sales data marts to deal with cus- tomer information. Bookseller Barnes & Noble used to maintain a series of data marts—one for point-of-sale data in retail stores, another for college bookstore sales, and a third for online sales.

Hadoop

Relational DBMS and data warehouse products are not well-suited for organiz- ing and analyzing big data or data that do not easily fit into columns and rows used in their data models. For handling unstructured and semi-structured data in vast quantities, as well as structured data, organizations are using Hadoop. Hadoop is an open source software framework managed by the Apache Software Foundation that enables distributed parallel processing of huge amounts of data across inexpensive computers. It breaks a big data problem down into sub-problems, distributes them among up to thousands of inexpen- sive computer processing nodes, and then combines the result into a smaller data set that is easier to analyze. You've probably used Hadoop to find the best airfare on the Internet, get directions to a restaurant, do a search on Google, or connect with a friend on Facebook.

Hadoop consists of several key services: the Hadoop Distributed File System (HDFS) for data storage and MapReduce for high-performance parallel data processing. HDFS links together the file systems on the numerous nodes in a Hadoop cluster to turn them into one big file system. Hadoop's MapReduce was inspired by Google's MapReduce system for breaking down processing of huge datasets and assigning work to the various nodes in a cluster. HBase, Hadoop's non-relational database, provides rapid access to the data stored on HDFS and a transactional platform for running high-scale real-time applications.

Hadoop can process large quantities of any kind of data, including structured transactional data, loosely structured data such as Facebook and Twitter feeds, complex data such as Web server log files, and unstructured audio and video data. Hadoop runs on a cluster of inexpensive servers, and processors can be added or removed as needed. Companies use Hadoop for analyzing very large volumes of data as well as for a staging area for unstructured and semi-struc- tured data before they are loaded into a data warehouse. Facebook stores much of its data on its massive Hadoop cluster, which holds an estimated 100 petabytes, about 10,000 times more information than the Library of Congress. Yahoo uses Hadoop to track user behavior so it can modify its home page to fit their interests. Life sciences research firm NextBio uses Hadoop and HBase to process data for pharmaceutical companies conducting genomic research. Top database vendors such as IBM, Hewlett-Packard, Oracle, and Microsoft have their own Hadoop software distributions. Other vendors offer tools for moving data into and out of Hadoop or for analyzing data within Hadoop.

In-Memory Computing

Another way of facilitating big data analysis is to use in-memory computing, which relies primarily on a computer's main memory (RAM) for data storage. (Conventional DBMS use disk storage systems.) Users access data stored in system primary memory, thereby eliminating bottlenecks from retrieving and reading data in a traditional, disk-based database and dramatically shortening query response times. In-memory processing makes it possible for very large sets of data, amounting to the size of a data mart or small data warehouse, to reside entirely in memory. Complex business calculations that used to take hours or days are able to be completed within seconds, and this can even be accomplished on handheld devices.

The previous chapter describes some of the advances in contemporary computer hardware technology that make in-memory processing possible, such as powerful high-speed processors, multicore processing, and falling computer memory prices. These technologies help companies optimize the use of memory and accelerate processing performance while lowering costs.

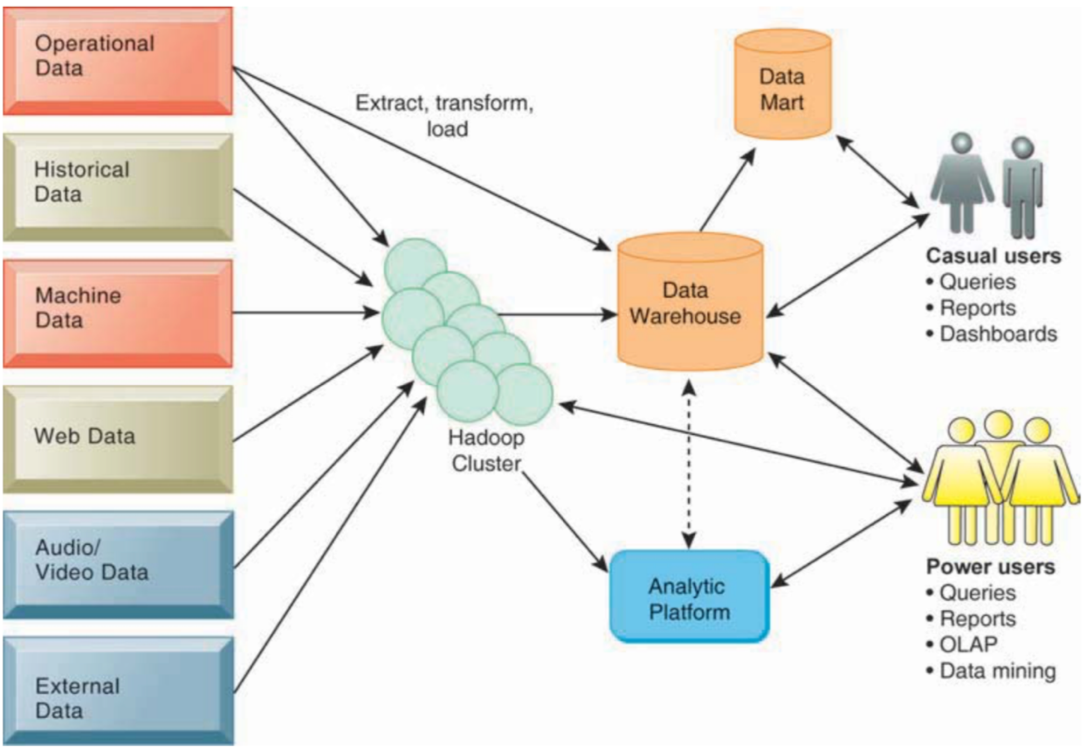
Leading commercial products for in-memory computing include SAP's High Performance Analytics Appliance (HANA) and Oracle Exalytics. Each provides a set of integrated software components, including in-memory database software and specialized analytics software, that run on hardware optimized for in-memory computing work.

Centrica, a gas and electric utility, uses HANA to quickly capture and analyze the vast amounts of data generated by smart meters. The company is able to analyze usage every 15 minutes, giving it a much clearer picture of usage by neighborhood, home size, type of business served, or building type. HANA also helps Centrica show its customers their energy usage patterns in real-time using online and mobile tools.

Analytic Platforms

Commercial database vendors have developed specialized high-speed analytic platforms using both relational and non-relational technology that are optimized for analyzing large datasets. These analytic platforms, such as IBM Netezza and Oracle Exadata, feature preconfigured hardware-software systems that are specifically designed for query processing and analytics. For example, IBM Netezza features tightly integrated database, server, and storage compo- nents that handle complex analytic queries 10 to 100 times faster than tradi- tional systems. Analytic platforms also include in-memory systems and NoSQL non-relational database management systems.

FIGURE 6.12 COMPONENTS OF A DATA WAREHOUSE



A contemporary business intelligence infrastructure features capabilities and tools to manage and analyze large quantities and different types of data from multiple sources. Easy-to-use query and reporting tools for casual business users and more sophisticated analytical toolsets for power users are included.

Figure 6.12 illustrates a contemporary business intelligence infrastructure using the technologies we have just described. Current and historical data are extracted from multiple operational systems along with Web data, machine- generated data, unstructured audio/visual data, and data from external sources that's been restructured and reorganized for reporting and analysis. Hadoop clusters pre-process big data for use in the data warehouse, data marts, or an analytic platform, or for direct querying by power users. Outputs include reports and dashboards as well as query results. Chapter 12 discusses the various types of BI users and BI reporting in greater detail.

2.2. ANALYTICAL TOOLS: RELATIONSHIPS, PATTERNS, TRENDS

Once data have been captured and organized using the business intelligence technologies we have just described, they are available for further analysis using software for database querying and reporting, multidimensional data analysis (OLAP), and data mining. This section will introduce you to these tools, with more detail about business intelligence analytics and applications in Chapter 12.

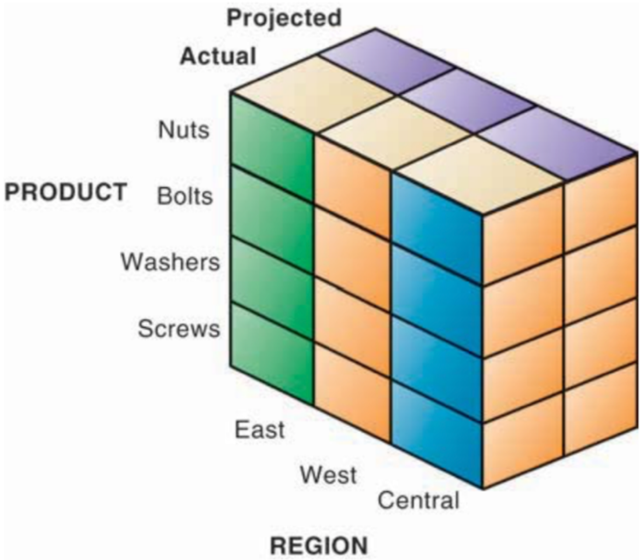
Online Analytical Processing (OLAP)

Suppose your company sells four different products—nuts, bolts, washers, and screws—in the East, West, and Central regions. If you wanted to ask a fairly straightforward question, such as how many washers sold during the past quarter, you could easily find the answer by querying your sales database. But what if you wanted to know how many washers sold in each of your sales regions and compare actual results with projected sales?

To obtain the answer, you would need online analytical processing (OLAP). OLAP supports multidimensional data analysis, enabling users to view the same data in different ways using multiple dimensions. Each aspect of information— product, pricing, cost, region, or time period—represents a different dimension. So, a product manager could use a multidimensional data analysis tool to learn how many washers were sold in the East in June, how that compares with the previous month and the previous June, and how it compares with the sales forecast. OLAP enables users to obtain online answers to ad hoc questions such as these in a fairly rapid amount of time, even when the data are stored in very large databases, such as sales figures for multiple years.

Figure 6.13 shows a multidimensional model that could be created to represent products, regions, actual sales, and projected sales. A matrix of actual sales can be stacked on top of a matrix of projected sales to form a cube with six faces. If you rotate the cube 90 degrees one way, the face showing will be product versus actual and projected sales. If you rotate the cube 90 degrees again, you will see region versus actual and projected sales. If you rotate 180 degrees from the original view, you will see projected sales and product versus region. Cubes can be nested within cubes to build complex views of data. A company would use either a specialized multidimensional database or a tool that creates multidimen- sional views of data in relational databases.

FIGURE 6.13 MULTIDIMENSIONAL DATA MODEL



This view shows product versus region. If you rotate the cube 90 degrees, the face that will show is product versus actual and projected sales. If you rotate the cube 90 degrees again, you will see region versus actual and projected sales. Other views are possible.

Data Mining

Traditional database queries answer such questions as, “How many units of product number 403 were shipped in February 2013?” OLAP, or multidi- mensional analysis, supports much more complex requests for information, such as, “Compare sales of product 403 relative to plan by quarter and sales region for the past two years.” With OLAP and query-oriented data analysis, users need to have a good idea about the information for which they are looking.

Data mining is more discovery-driven. Data mining provides insights into corporate data that cannot be obtained with OLAP by finding hidden patterns and relationships in large databases and inferring rules from them to predict future behavior. The patterns and rules are used to guide decision making and forecast mining include associations, sequences, classifications, clusters, and forecasts.

- Associations are occurrences linked to a single event. For instance, a study of supermarket purchasing patterns might reveal that, when corn chips are purchased, a cola drink is purchased 65 percent of the time, but when there is a promotion, cola is purchased 85 percent of the time. This informa- tion helps managers make better decisions because they have learned the profitability of a promotion.
- In sequences, events are linked over time. We might find, for example, that if a house is purchased, a new refrigerator will be purchased within two weeks 65 percent of the time, and an oven will be bought within one month of the home purchase 45 percent of the time.
- Classification recognizes patterns that describe the group to which an
- item belongs by examining existing items that have been classified and by inferring a set of rules. For example, businesses such as credit card or

telephone companies worry about the loss of steady customers. Classification helps discover the characteristics of customers who are likely to leave and can provide a model to help managers predict who those customers are so that the managers can devise special campaigns to retain such customers.

- Clustering works in a manner similar to classification when no groups have yet been defined. A data mining tool can discover different groupings within data, such as finding affinity groups for bank cards or partitioning a database into groups of customers based on demographics and types of personal investments.

Although these applications involve predictions, forecasting uses predictions in a different way. It uses a series of existing values to forecast what other values will be. For example, forecasting might find patterns in data to help managers estimate the future value of continuous variables, such as sales figures.

These systems perform high-level analyses of patterns or trends, but they can also drill down to provide more detail when needed. There are data mining applications for all the functional areas of business, and for government and scientific work. One popular use for data mining is to provide detailed analyses of patterns in customer data for one-to-one marketing campaigns or for identifying profitable customers.

Caesars Entertainment, formerly known as Harrah's Entertainment, is the largest gaming company in the world. It continually analyzes data about its customers gathered when people play its slot machines or use its casinos and hotels. The corporate marketing department uses this information to build a detailed gambling profile, based on a particular customer's ongoing value to the company. For instance, data mining lets Caesars know the favorite gaming experience of a regular customer at one of its riverboat casinos, along with that person's preferences for room accommodations, restaurants, and entertain- ment. This information guides management decisions about how to cultivate the most profitable customers, encourage those customers to spend more, and attract more customers with high revenue-generating potential. Business intel- ligence improved Caesars's profits so much that it became the centerpiece of the firm's business strategy.

Text Mining and Web Mining

However, unstructured data, most in the form of text files, is believed to account for over 80 percent of useful organizational information and is one of the major sources of big data that firms want to analyze. E-mail, memos, call center transcripts, survey responses, legal cases, patent descriptions, and service reports are all valuable for finding patterns and trends that will help employees make better business decisions. Text mining tools are now available to help businesses analyze these data. These tools are able to extract key elements from unstructured big data sets, discover patterns and relation- ships, and summarize the information.

Businesses might turn to text mining to analyze transcripts of calls to customer service centers to identify major service and repair issues or to measure customer sentiment about their company. Sentiment analysis software is able to mine text comments in an e-mail message, blog, social media conversation, or survey form to detect favorable and unfavorable opinions about specific subjects.

For example, the discount broker Charles Schwab uses Attensity Analyze software to analyze hundreds of thousands of its customer interactions each month. The software analyzes Schwab's customer service notes, e-mails, survey responses, and online discussions to discover signs of dissatisfac- tion that might cause a customer to stop using the company's services. Attensity is able to automatically identify the various "voices" customers use to express their feedback (such as a positive, negative, or conditional voice) to pinpoint a person's intent to buy, intent to leave, or reaction to a specific product or marketing message. Schwab uses this information to take corrective actions such as stepping up direct broker communication with the customer and trying to quickly resolve the problems that are making the customer unhappy.

The Web is another rich source of unstructured big data for revealing patterns, trends, and insights into customer behavior. The discovery and analysis of useful patterns and information from the World Wide Web is called Web mining. Businesses might turn to Web mining to help them understand customer behavior, evaluate the effectiveness of a particular Web site, or quantify the success of a marketing campaign. For instance, marketers use the Google Trends and Google Insights for Search services, which track the popularity of various words and phrases used in Google search queries, to learn what people are interested in and what they are interested in buying.

Web mining looks for patterns in data through content mining, structure mining, and usage mining. Web content mining is the process of extracting knowledge from the content of Web pages, which may include text, image, audio, and video data. Web structure mining examines data related to the structure of a particular Web site. For example, links pointing to a document indicate the popularity of the document, while links coming out of a docu- ment indicate the richness or perhaps the variety of topics covered in the document. Web usage mining examines user interaction data recorded by a Web server whenever requests for a Web site's resources are received. The usage data records the user's behavior when the user browses or makes trans- actions on the Web site and collects the data in a server log. Analyzing such data can help companies determine the value of particular customers, cross marketing strategies across products, and the effectiveness of promotional campaigns.

The Interactive Session on Technology describes organizations' experiences as they use the analytical tools and business intelligence technologies we have described to grapple with "big data" challenges.

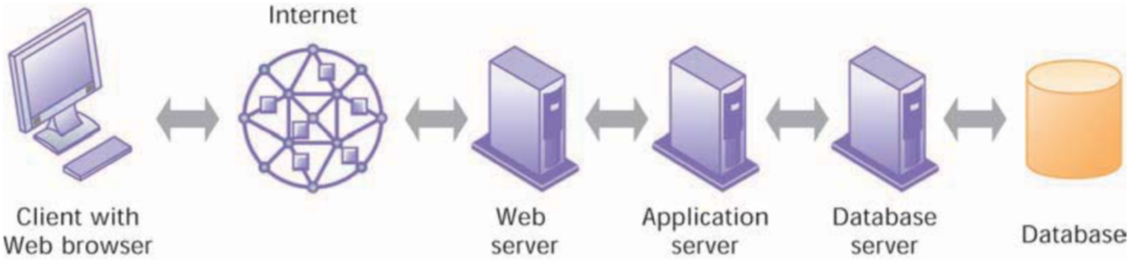
2.3. DATABASES AND THE WEB

Have you ever tried to use the Web to place an order or view a product catalog? If so, you were probably using a Web site linked to an internal corporate database. Many companies now use the Web to make some of the informa- tion in their internal databases available to customers and business partners.

Suppose, for example, a customer with a Web browser wants to search an online retailer's database for pricing information. Figure 6.14 illustrates how that customer might access the retailer's internal database over the Web. The user accesses the retailer's Web site over the Internet using Web browser software on his or her client PC. The user's Web browser software requests data from the organization's database, using HTML commands to communi- cate with the Web server.

Because many back-end databases cannot interpret commands written in HTML, the Web server passes these requests for data to software that translates HTML commands into SQL so the commands can be processed by the DBMS working with the database. In a client/server environment, the

FIGURE 6.14 LINKING INTERNAL DATABASES TO THE WEB



Users access an organization's internal database through the Web using their desktop PCs and Web browser software.

DBMS resides on a dedicated computer called a database server. The DBMS receives the SQL requests and provides the required data. Middleware trans- fers information from the organization's internal database back to the Web server for delivery in the form of a Web page to the user.

Figure 6.14 shows that the middleware working between the Web server and the DBMS is an application server running on its own dedicated computer (see Chapter 5). The application server software handles all application operations, including transaction processing and data access, between browser-based computers and a company's back-end business applications or databases. The application server takes requests from the Web server, runs the business logic to process transactions based on those requests, and provides connectivity to the organization's back-end systems or databases. Alternatively, the software for handling these operations could be a custom program or a CGI script. A CGI script is a compact program using the Common Gateway Interface (CGI) specification for processing data on a Web server.

There are a number of advantages to using the Web to access an organiza- tion's internal databases. First, Web browser software is much easier to use than proprietary query tools. Second, the Web interface requires few or no changes to the internal database. It costs much less to add a Web interface in front of a legacy system than to redesign and rebuild the system to improve user access.

Accessing corporate databases through the Web is creating new efficiencies, opportunities, and business models. ThomasNet.com provides an up-to-date online directory of more than 650,000 suppliers of industrial products, such as chemicals, metals, plastics, rubber, and automotive equipment. Formerly called Thomas Register, the company used to send out huge paper catalogs with this information. Now it provides this information to users online via its Web site and has become a smaller, leaner company.

Other companies have created entirely new businesses based on access to large databases through the Web. One is the social networking service Facebook, which helps users stay connected with each other and meet new people. Facebook features “profiles” with information on more than 950 million active users with information about themselves, including interests, friends, photos, and groups with which they are affiliated. Facebook maintains a massive database to house and manage all of this content.

There are also many Web-enabled databases in the public sector to help consumers and citizens access helpful information. The Interactive Session on Organizations describes one of these databases, which has generated controversy over its methods for providing consumer product safety data.

2.4. MANAGING DATA RESOURCES

Setting up a database is only a start. In order to make sure that the data for your business remain accurate, reliable, and readily available to those who need it, your business will need special policies and procedures for data management.

A. ESTABLISHING AN INFORMATION POLICY

Every business, large and small, needs an information policy. Your firm’s data are an important resource, and you don’t want people doing whatever they want with them. You need to have rules on how the data are to be organized and maintained, and who is allowed to view the data or change them.

An information policy specifies the organization’s rules for sharing, disseminating, acquiring, standardizing, classifying, and inventorying information. Information policy lays out specific procedures and account- abilities, identifying which users and organizational units can share information, where information can be distributed, and who is responsi- ble for updating and maintaining the information. For example, a typical information policy would specify that only selected members of the payroll and human resources department would have the right to change and view sensitive employee data, such as an employee’s salary or social security number, and that these departments are responsible for making sure that such employee data are accurate.

If you are in a small business, the information policy would be established and implemented by the owners or managers. In a large organization, managing and planning for information as a corporate resource often requires a formal data administration function. Data administration is responsible for the specific policies and procedures through which data can be managed as an organizational resource. These responsibilities include developing information policy, planning for data, overseeing logical database design and data dictionary development, and monitoring how information systems specialists and end-user groups use data.

You may hear the term data governance used to describe many of these activities. Promoted by IBM, data governance deals with the policies and processes for managing the availability, usability, integrity, and security of the data employed in an enterprise, with special emphasis on promoting privacy, security, data quality, and compliance with government regulations.

A large organization will also have a database design and management group within the corporate information systems division that is responsible for defining and organizing the structure and content of the database, and maintaining the database. In close cooperation with users, the design group establishes the physical database, the logical relations among elements, and the access rules and security procedures. The functions it performs are called database administration.

B. ENSURING DATA QUALITY

A well-designed database and information policy will go a long way toward ensuring that the business has the information it needs. However, additional steps must be taken to ensure that the data in organizational databases are accurate and remain reliable.

What would happen if a customer's telephone number or account balance were incorrect? What would be the impact if the database had the wrong price for the product you sold or your sales system and inventory system showed different prices for the same product? Data that are inaccurate, untimely, or inconsistent with other sources of information lead to incorrect decisions, product recalls, and financial losses. Gartner Inc. reported that more than 25 percent of the critical data in large Fortune 1000 companies' databases is inaccurate or incom- plete, including bad product codes and product descriptions, faulty inventory descriptions, erroneous financial data, incorrect supplier information, and incorrect employee data. A Sirius Decisions study on "The Impact of Bad Data on Demand Creation" found that 10 to 25 percent of customer and prospect records contain critical data errors. Correcting these errors at their source and following best practices for promoting data quality increased the productivity of the sales process and generated a 66 percent increase in revenue.

Some of these data quality problems are caused by redundant and inconsistent data produced by multiple systems feeding a data warehouse. For example, the sales ordering system and the inventory management system might both maintain data on the organization's products. However, the sales ordering system might use the term Item Number and the inventory system might call the same attribute Product Number. The sales, inventory, or manufacturing sys- tems of a clothing retailer might use different codes to represent values for an attribute. One system might represent clothing size as "extra large," whereas the other system might use the code "XL" for the same purpose. During the design process for the warehouse database, data describing entities, such as a customer, product, or order, should be named and defined consistently for all business areas using the database.

Think of all the times you've received several pieces of the same direct mail advertising on the same day. This is very likely the result of having your name maintained multiple times in a database. Your name may have been misspelled or you used your middle initial on one occasion and not on another or the information was initially entered onto a paper form and not scanned properly into the system. Because of these inconsistencies, the database would treat you as different people! We often receive redundant mail addressed to Laudon, Lavdon, Lauden, or Landon.

If a database is properly designed and enterprise-wide data standards estab- lished, duplicate or inconsistent data elements should be minimal. Most data quality problems, however, such as misspelled names, transposed numbers, or incorrect or missing codes, stem from errors during data input. The incidence of such errors is rising as companies move their businesses to the Web and allow customers and suppliers to enter data into their Web sites that directly update internal systems.

Before a new database is in place, organizations need to identify and correct their faulty data and establish better routines for editing data once their data- base is in operation. Analysis of data quality often begins with a data quality audit, which is a structured survey of the accuracy and level of completeness of the data in an information system. Data quality audits can be performed by surveying entire data files, surveying samples from data files, or surveying end users for their perceptions of data quality.

Data cleansing, also known as data scrubbing, consists of activities for detecting and correcting data in a database that are incorrect, incomplete, improperly formatted, or redundant. Data cleansing not only corrects errors but also enforces consistency among different sets of data that originated in separate information systems. Specialized data-cleansing software is available to automatically survey data files, correct errors in the data, and integrate the data in a consistent company-wide format.

Data quality problems are not just business problems. They also pose serious problems for individuals, affecting their financial condition and even their jobs. For example, inaccurate or outdated data about consumers' credit histories maintained by credit bureaus can prevent creditworthy individuals from obtaining loans or lower their chances of finding or keeping a job.

3. Latihan dan Jawaban

1) What are some important database design principles?

   Designing a database requires both a logical design and a physical design. The logical design models the database from a business perspective. The organization's data model should reflect its key business processes and decision-making requirements. The process of creating small, stable, flexible, and adaptive data structures from complex groups of data when designing a relational database is termed normalization. A well-designed relational database will not have many-to-many relationships, and all attributes for a specific entity will only apply to that entity. It will try to enforce referential integrity rules to ensure that relationships between coupled tables remain consistent. An entity-relationship diagram graphically depicts the relationship between entities (tables) in a relational database.

2) What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?

   Contemporary data management technology has an array of tools for obtaining useful information from all the different types of data used by businesses today, including semi-structured and unstruc- tured big data in vast quantities. These capabilities include data warehouses and data marts, Hadoop, in-memory computing, and analytical platforms. OLAP represents relationships among data as a multidimensional structure, which can be visualized as cubes of data and cubes within cubes of data, enabling more sophisticated data analysis. Data mining analyzes large pools of data, including the contents of data warehouses, to find patterns and rules that can be used to predict future behavior and guide decision making. Text mining tools help businesses analyze large unstructured data sets consist- ing of text. Web mining tools focus on analysis of useful patterns and information from the World Wide Web, examining the structure of Web sites and activities of Web site users as well as the contents of Web pages. Conventional databases can be linked via middleware to the Web or a Web interface to facilitate user access to an organization's internal data.

3) Why are information policy, data administration, and data quality assurance essential for managing the firm's data resources?

Developing a database environment requires policies and procedures for managing organizational data as well as a good data model and database technology. A formal information policy governs the maintenance, distribution, and use of information in the organization. In large corporations, a formal data administration function is responsible for information policy, as well as for data planning, data dictionary development, and monitoring data usage in the firm.

Data that are inaccurate, incomplete, or inconsistent create serious operational and financial problems for businesses because they may create inaccuracies in product pricing, customer accounts, and inventory data, and lead to inaccurate decisions about the actions that should be taken by the firm. Firms must take special steps to make sure they have a high level of data quality. These include using enterprise-wide data standards, databases designed to minimize inconsistent and redundant data, data quality audits, and data cleansing soft- ware.

3. Daftar Pustaka

1. Management Information Systems, Managing Digital Firm, 11th Ed, Kenneth C. Laudon, Jane. P. Laudon. (L&L)
2. Management Information Systems With Misource 2007, 8th Ed James A. O'brien, And George Marakas
3. Managing Information Technology 5th Edition Martin, Brown, Dehayes