

Foundation of business intelligence assignment



Important principles of database designing 9 5. Principal tools for accessing information from database to improve business performance and decision making 1 1 6. Reasons of Information policy, data administration and data quality assurance essential for managing a firm's data resources 15 7. Closure 16 8. References 17 1. Introduction Database systems are the information heart of modern enterprises, where they are used for processing business transactions and for understanding and managing the enterprise. Business intelligence is the analysis of data to produce insights useful for managing the enterprise and increasingly, in routine business operations such as intelligent supply chain management. The knowledge of Database Management helps business intelligences to know transaction processing and decision support, how to use data mining technologies to discover the structure, trends, and relationships in the data to produce valuable business insights and effective decision support processes.

This assignment includes the following relevant issues regarding Database and Information Management: 1 . Problems of Managing Data resources in a traditional file environment and how they are solved by data management system . Major capabilities of database management systems (DB'S) and the reason of Relational database becoming so powerful 3. Important principles of database designing 4. Principal tools for accessing information from database to improve business performance and decision making 5.

Reasons of Information policy, data administration and data quality assurance essential for managing a firm's data resources This assignment will be a very helpful as a source of information to get clear concept of what

is database, how DB'S and relational database helps business intelligence to solve business problems, tools of database to improve business reference, importance of Information policy, data administration and data quality assurance for managing a firm. They are solved by data management system

File Organizing Terms and Concepts In a traditional file organizing concept a computer system organizes data in a hierarchy that starts with bits and bytes and progresses to fields, records, files and database. A Bit represents the smallest unit of a data (0 and 1). The widely used code conversion method is known as " American Standard Code of Information Interchange" (ASCII). A Byte is a group of bits. (1 Byte consists of 8 bits) It represents singly character/letter/number or symbol. A field is a group of words or a complete number or contents of Bytes.

A record is a group of related fields/data. A file is a group of records of same types. A database is made up of a group of related files. Problems with the traditional file environment: In most organizations, systems tend to grow independently without a company-wide plan. Accounting, finance, manufacturing, human resources and sales and marketing all develop their own systems and data files. Each application requires its own files and its own computer program to operate. This process leads to multiple master files creation, maintenance and are operated by different divisions and departments.

Day by day it gets difficult to maintain the huge files in different locations and places. The problems arise in this system is mentioned below: a) Data Redundancy and Inconsistency: Duplication of multiple data files meaning same types of files are stored in more than one or many places. It causes the

<https://assignbuster.com/foundation-of-business-intelligence-assignment/>

wastage of storage resources and leads to data inconsistency meaning that the same attributes may have different values. B) Program-Data

Dependence: It refers to the coupling of data stored in files and the specific programs required to update and maintain those files such that changes in programs require changes to the data.) Lack of flexibility: A traditional file system can deliver routine scheduled reports after extensive programming efforts, but it cannot deliver ad-hoc reports to respond to unanticipated information requirements in a timely fashion. D) Poor security: The management of data has little control so it has poor security. Management may have no way of knowing who is accessing or even making changes to the organization's data. E) Lack of data sharing and availability: As the information are in separate parts of the organization, it can't be related to one another and it can't be shared or accessed different parts of the organization.

Solving the problems through Database management system: A database management system (DB'S) is a software that permits an organization to centralize data, manage them efficiently, and provide access to the stored data by application programs. The DB'S acts as an interface between application programs and the physical data files. When the application program calls for a data item, such as gross pay, the DB'S finds this item in the database and presents it to the application program. The DB'S uncouples programs and data, enabling data to stand on their own.

Access and availability of information will be increased and program development and maintenance costs reduced because users and programmers can perform ad-hoc queries of data in the database. The DB'S

<https://assignbuster.com/foundation-of-business-intelligence-assignment/>

enables the organization to centrally manage data, their use and security. The database management software makes the physical database available for different logical views required by users. For example, the Human Resource Database illustrates the benefits of the employee through Name, USN and Health Care information while payroll information are shown through the employee's name, social security number, gross pay and net pay.

The data for all these views are stored in a single database, where they can be easily managed by the organization. DB'S is basically four types and they are: a) Relational b) Hierarchical c) Network d) Object Oriented/Hybrid

a) Relational : Relational databases represent data as two-dimensional tables. Tables may be referred to as files. Student ID SSL. NO. Contact Dept. No. 121 0191 201 131 02 0171 301 1421 03 0187 401

b) Hierarchical: A hierarchical database model is a data model in which the data is organized into a tree-like structure. It represents information in a one to many relationship, e. G. Arena may have more than one child but each child can only have one parent. Students Fees Database

c) Network: The network model is a database model conceived as a flexible way of representing objects and their relationships. Its distinguishing feature is that the schema, viewed as a graph in which object types are nodes and relationship types are arcs, is not restricted to being a hierarchy. It links all the tables such as in an admission database it will include Student database and Course database.

) Object-oriented : This database is the mixture of relational, hierarchical and network database. 3.

Major capabilities of database management systems (DB'S) and the reason of Relational database becoming so powerful Major capabilities of DB'S include a data definition capability and a data manipulation language. The data definition capability specifies the structure and content of the database. The data dictionary is an automated or manual file that stores information about the data in the database including names, definitions, formats and description of data elements. The data manipulation language such as Structured Query Language (SQL) is a specialized language for accessing and manipulating the data in the database.

Relational database: Relational database represents data as two-dimensional tables (called relations). Tables may be referred to as files. Each table contains data on an entity and its attributes. Microsoft Access is a relational DB'S for desktop systems, whereas DB, Oracle database, Microsoft SQL Server are relational DB'S for large mainframes and midrange computers. The relational database is the primary method for organizing and maintaining data today in information systems because it is so flexible and accessible. It organizes contains data about an entity and its attributes.

Each row represents a record, and each column represents an attribute or field. Each table also contains a key field to uniquely identify each record for retrieval or manipulation. Relational database tables can be combined easily to deliver data required by users, provided that any two or more tables share a common data element. A DB'S includes capabilities and tools for organizing, managing and accessing the data in the database. The most important are its data definition language, data dictionary and data manipulation language.

These capabilities made the relational database so powerful. The Data definition capability is to specify the structure of the content of the database. It is used to create database tables and to define the characteristics of the fields in each table. A data dictionary is an automated or manual file that stores definitions of data elements and their characteristics. Microsoft Access has a rudimentary data dictionary capability that displays information about name, description, size, type, format and other properties of each field in a table.

Query and Reporting is another tool that is used for accessing and manipulating information in database which is used to add, change, delete and retrieve the data in the database. Report generating is another capability that made relational database so popular and powerful. It can display the data of interest in a more structured and polished format than would be possible just by querying. Relational database also has capabilities for developing desktop system applications. These include tools for creating data entry screens, reports and developing the logic for processing transactions. 4.

Important principles of database designing 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. The key to understanding the database design process lies in understanding the way a relational database management system, such as Microsoft Access, stores data. To efficiently and accurately provide you with information, Microsoft Access needs to have the facts about different subjects stored in separate tables.

For example, we might have one table that stores only facts about employees, and another that stores only facts about sales. When we use our data, we then combine and present facts in employees and facts about sales. When we design a database, we first break down the information we want to keep as separate subjects, and then we tell Microsoft Access how the subjects are related to each other so that Microsoft Access can bring the right information together when we need it. The database requires both a conceptual design and a physical design.

The conceptual or logical design of a database is an abstract model of the database from a business perspective, whereas the physical design shows how the database is actually arranged on direct access storage access device. Steps in Designing a Database Step One: To determine the purpose of the database. This will help us decide which facts you want Microsoft Access to store. Step Two: Determine the tables we need. Once we have a

clear purpose for the database, we can divide our information into separate subjects, such as “ Employees” or “ Orders. ” Each subject will be a table in the database.

Step Three: Determine the fields we need. Decide what information we want to keep in each table. Each category of information in a table is called a field and is displayed as a column in the table. For example, one field in an Employees table could be Last Name; another could be Hire Date. Step Four: Determine the relationships. Look at each table and decide how the data in one table is related to the data in other tables. Add fields to tables or create new tables to clarify the relationships, as necessary. Step Five: Refine the design. Analyze design for errors. Then create the tables and add a few records of sample data.

Then we will see whether we are getting our desired results from the tables or not. We have to make adjustments to the design as needed. 5. Principal tools for accessing information from database to improve business performance and decision making Business use their databases to keep track of basic transactions, such as paying suppliers, processing orders, keeping track of customers and paying employees. But 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 customers, the answer lies in the data.

There are tools for accessing information from database to improve business performance and decision making. Data Warehouses: 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 website transactions. The above figure shows how a warehouse works. The data warehouse makes the data available to anyone to access as needed, but it can't be altered.

It also provides a range of ad hoc and standardized query tools, analytical tools and graphical reporting facilities. Many firms use intranet portals to make the data warehouse information widely available throughout the firm. Data Marts: Companies often build enterprise-wide data warehouses where a central data warehouse serves the entire organization, or they create smaller decentralized 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 customer information. Principal tools for business intelligence include software for database querying and reporting, tools for multi-dimensional data analysis and tools for data mining. Online Analytical Processing (OLAP) : OLAP (online analytical processing) is computer processing that enables a user to easily and selectively extract and view data from different points of view.

For example, a user can request that data be analyzed to display a spreadsheet showing all of a company's beach ball products sold in Florida in

the month of July, compare revenue figures with those for the same products in September, and then see a comparison of other product sales in Florida in the same time period. To facilitate this kind of analysis, OLAP data is stored in a multidimensional database. Data mining : Data mining is more discovery driven.

Data mining provides insights into corporate data that can't 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 the effect of those decisions. The types of information obtainable from data mining include associations, sequences, classifications, clusters and forecasts. Data mining is a powerful new technology with great potential to help companies focus on the most important in the data they have collected about the behavior of what reports can't effectively reveal.

Although data mining is still in its infancy, companies in a wide range of industries - including retail, finance, health care, manufacturing transportation, and aerospace - are already using data mining tools and techniques to take advantage of historical data. By using pattern recognition technologies and statistical and mathematical techniques to sift through warehoused information, data mining helps analysts recognize significant facts, relationships, trends, patterns, exceptions and anomalies that might otherwise go unnoticed.

For businesses, data mining is used to discover patterns and relationships in the data in order to help make better business decisions. Data mining can

help spot sales trends, develop smarter marketing campaigns, and accurately predict customer loyalty. Specific uses of data mining include:

Market segmentation - Identify the common characteristics of customers who buy the same products from your company. Customer churn - Predict which customers are likely to leave your company and go to a competitor.

Fraud detection - Identify which transactions are most likely to be fraudulent. Direct marketing - Identify which prospects should be included in a mailing list to obtain the highest response rate. Interactive marketing - Predict what each individual accessing a Web site is most likely interested in seeing. Market basket analysis - Understand what products or services are commonly purchased together; e. G. , beer and diapers. Trend analysis - Reveal the difference between a typical customer this month and last.

Data mining technology can generate new business opportunities by:

Automated prediction of trends and behaviors: Data mining automates the process of finding predictive information in a large database. Questions that traditionally required extensive hands-on analysis can now be directly answered from the data. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings.

Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events. Automated discovery of previously unknown patterns: Data mining tools sweep through databases and identify previously hidden

patterns. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. Other pattern discovery problems include detecting fraudulent credit card transactions and identifying anomalous data that could represent data entry keying errors. . Reasons of Information policy, data administration and data quality assurance essential for managing a firm's data resources Developing a database environment requires policies and procedures for managing information policy governs the maintenance, distribution and use of information in he 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, database designed to minimize inconsistent and redundant data, data quality audits and data cleansing software.

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 are accurate and remain reliable. If a database is properly designed and enterprise-wide data standards established, duplicate or

<https://assignbuster.com/foundation-of-business-intelligence-assignment/>

inconsistent data elements should be minimal. Most data quality robbers, however, such as misspelled names, transposed numbers or incorrect or missing codes, stem from errors during data input.

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 database 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 Data cleansing is 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 correct errors but also enforces consistency among different sets of data that originated 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 cleansing is sometimes compared to data purging, where old or useless data will be deleted from a data set.

Although data cleansing can involve deleting old, incomplete or duplicated data, data cleansing is different from data purging in that data purging usually focuses on clearing space for new data, whereas data cleansing focuses on maximizing the accuracy of data in a system. A data cleansing method may use parsing or other methods to get rid of syntax errors,

typographical errors or fragments of records. Careful analysis of a data set can show how merging multiple sets led to duplication, in which case data cleansing may be used to fix the problem.