|  |  |  |  |
| --- | --- | --- | --- |
| Instructor | ***Luke Papademas*** | Due Date |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Part** | 1 | 2 | 3 | 4 | **TOTAL** | **Score** |
| *Maximum Points* | 25 points | 25 points | 25 points | 25 points | **100**G101010 pointsG |  |

**Textbook Reading Assignment** Thoroughly read Week 1 - 4 course lecture notes.

**Part 1 Concept Check - Advanced Topics in Data Management**

**(1) ( Data Warehousing: OLAP versus OLTP )**

An information system category that prioritizes transaction processing, which deal with operational data is Online Transaction Processing ( OLTP ) . Another type of information system category is Online Analytical Processing ( OLAP ) .

OLAP concentrates on performing analytical processing and OLTP focuses on providing transactional processing.

View the video production at the link shown below ( or similar video ) and list five facts that you learned concerning these two information systems.

[**https://www.youtube.com/watch?v=q5f1xOA9IQQ**](https://www.youtube.com/watch?v=q5f1xOA9IQQ)

|  |
| --- |
| OLAP: OLAP is Primarily used for complex operations such as Aggregations, Sum, Average, Count. In the OLAP the Data is basically stored in the Multidimensional Databases.  Aggregation: In OLAP data need to be aggregated.  Indexes: OLAP has many indexes for the performance optimization.  Joins: OLAP systems are having less joins and they are usually denormalized.  Effectiveness: Time required to execute an query can be considered as an effective measure for the OLAP systems  OLAP Applications are widely used in the data mining techniques. OLAP is having 2 to 3 hours of data latency which represents the time it takes before transferring the data.  OLTP: OLTP is primarily used for large number of online transactions such as Update, Read, Delete.  Indexes: OLTP is having only five indexes as compared to the OLAP.  Aggregation: In OLTP the data is not aggregated as if we compare with the OLAP with aggregated data. |

**(2) ( Data Science and Business Intelligence: OLAP )**

OLAP focuses on multi - dimensional views and analysis of data queries. For example, we can query data from a standpoint of examining company sales by region, year, sales amount and salesperson.

The typical OLAP operations will include: (1) drill - down, (2) dice and slice,   
 (3) rotate or pivot and (4) roll - up or consolidation

Using one or more of the business or organizational entities listed below discuss a multi - dimensional analysis that can be performed on the entities with any or all of the above operations.

**Business / Organizational Entities**

• a custom trophy and award firm • a custom T- Shirt firm

• a nationwide college or university • a musical instrument shop

• an HVAC repair service • an ice cream parlor

You can review the information at the link below, which summaries the OLAP operations.

[**http://www.cs.ccsu.edu/~markov/ccsu\_courses/DataMining-2.html**](http://www.cs.ccsu.edu/~markov/ccsu_courses/DataMining-2.html)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A Custom T-Shirt Firm:  When you consider an example of online firm that takes custom T-shirt Orders through their websites. According to the conceptual model model we can observe the outer view of the data. For Eg: Total number of T-shirts sold in last year.  If we consider total number of the Sales in last quarter, we can use one of the widely know OLAP operation drill-down method.  Drill-Down: The less detailed data is converted in to the more detailed one by moving down the hierarchy.  At the top level we have a data cube named as Total Sales having different subset of attributes.  Total Sales      Sales By Month Age-specification PaymentMethod  (In a quarter) (Most sales from (Most Preferred Payment  Age group) Method)   |  |  | | --- | --- | | Quarter | Sales | | 1 | 13,455 | | 2 | 22,689 | | 3 | 11,000 | | 4 | 25,000 |   By Using Drill-Down approach we go deeper in the conceptual hierarchy and extract out more detailed information.   |  |  |  |  | | --- | --- | --- | --- | | Months | Jan | Feb | Mar | | Sales | 1200 | 5490 | 2500 | | Payment Method | Offline | Offline | Online | |  |  |  |  |   We can provide further analysis and extract more data from Payment Method data cube,   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Months | Jan | Feb | Mar | Apr | | Online |  |  | Zelle | Zelle | | Offline | Cash | DD |  |  |   Similarly we can drill-down through the data cube and uncover more data secret layers allowing us to procure deeper understanding of situation. In terms of Age-Specification we can simply drill-down to specific age group, most commonly used materials and colors. Such analysis will help us to keep track of customers interest and change over inventory and product design according to the customers need.  Inversely, we can repeat the process moving from highly detailed data levels to low detailed data levels.  Roll-Up Operation: It is widely used to get the more general view of the data. We will move from the Months, online and offline table to Payments method table with less details and more summary focus data cube. |
|  |

**(3) ( Data Maintenance: Scrubbing Data in an MS Access Database Table )**

You have been assigned the task of scrubbing an MS Access database table, which is used for data maintenance purposes. Explain how you would use MS Access to perform each of these scrubbing tasks.

(a) Locate any fields that contain two spaces and replace the two spaces with one space.

(b) Locate any duplicate records.

(c) Eliminate any records that have a particular field having a number value.

(d) Eliminate any records that have a field having an alpha - numeric value.

(e) Update any records that have any non - printing characters.

You can create your own sample table and then show explain or how the table is scrubbed.

|  |
| --- |
| a> Locate any fields that contain two spaces and replace the two spaces with one space.  There are various ways in which we can effectively scrub the table, the purpose of scrubbing the database is to make database look good. When we see the database in MS Access, we can edit the data in the individual columns by simply clinking in each column field.  Moreover, we can also use the queries to replace double spaces in a particular column by using the replace method.  Consider the following table were indicating columns before replacing two spaces with the one.  To remove the spaces from form a particular column we will use Replace method.  Graphical user interface, application, Word  Description automatically generated  Consider the following Query it selects the names from the column by simply replacing the double space by the single space and presents the result.  Graphical user interface, application, Word  Description automatically generated  Moreover, we can use the update query to permanently replace the values in the column having two spaces with the one space.  Graphical user interface, application, Word  Description automatically generated  Graphical user interface, application, Word  Description automatically generated  After executing the Update and replace query, Column entity with the two spaces has been replaced with the one space.  b> To Find the duplicate records, we will have to use the use the query design tab in the MS Access. For E.g. Following university table each and every record is unique so it will reflect null after duplicate analysis.  Graphical user interface, application  Description automatically generatedGraphical user interface, application  Description automatically generatedGraphical user interface, application  Description automatically generatedGraphical user interface, application  Description automatically generatedGraphical user interface, application  Description automatically generated  As above results show no columns because there are no duplicate columns in the table.  c> Eliminate any records that have a particular field having a number value.  To check the data type of fields we can select the table and can click on View option and then Design View from the drop-down list which in turn will show us all the values and data types related to each column.  d>Similarly, we can check the data type if the values are alphanumeric in the table such as values having the data type short text.  Moreover, In the databases we can simply use LIKE and wildcard characters to accurately trace the certain range of digits or character using % [0-9]% or %[A-Z, A-Z]%.  e> The Non-Printable characters are the ASCII Characters with the range 1 to 31.  The first 32 characters in the ASCII table are unprintable control codes and used to control the peripherals such as printers.  We can use the Asc function in MS Access to get the ascii value of a particular character.  With the help of View we can switch to the design view and identify the type of each column.  Table  Description automatically generated with medium confidence  Consider the list of non-printable ASCII characters as mentioned above.  With the help of above commands we can update the existing table. |

**Part 2 DBMS Concepts - Advanced Topics in Data Management**

**(1) (Entity Relationship Diagrams - ERDs )**

An entity relationship diagram (ERD ) depicts the conceptual database as viewed by end user. An ERD also illustrates the database’s main components:

Entities, Attributes, Relationships

Examine these tables their attributes and their relationships.

**Orders** {OrderID, OrderDate, Amount, CustomerID, ShippingDate, OrderStatus}

**OrderDetails** {OrderID, ProductID, Quantity, UnitPrice, Discount}

**Products** {ProductID, ProdName, Description, Quantity, UnitPrice, Location}

Without writing any SQL code statements, discuss five separate queries that would be useful to ascertain information regarding this database schema.

|  |
| --- |
| Entity-Relationship Diagram Helps us to identify different components in the databases and their relationships. E-R Diagrams primarily focuses on the Entities their attributes and relationship between them.  For the Given tables, Order table is having the general information about the customers, orders, and shipping date. Whereas products table is having the ID, Description, Unit Price, a Particular Location, and name of the product.   1. One of the important query we can generate is about the getting a orders and customers that have order value over certain amount. we can implement this by querying the order table itself. 2. Next by joining Order Details and the Products table we can procure the information about discount offered on a specific product and can also retrieve the description of that product. 3. By Joining all the three tables with Inner Join we can get the authentic information about Shippingdate, OrderID, CustomerID, ProductDescription and Quantity details. To get detailed summary of the order. 4. Moreover, we can identify list of products for the OrderID Table having the most lucrative discount. The products with higher discount will have higher sales as compared to the other products. 5. From the Products table we can track down the exact location of the product and joining them with order details and order we can combine this data with shipping date, allowing all the product tracking information in one single table. |
|  |

**(2) ( Types of Relationships )**

When designing an ERD, what is meant by a weak relationship? Provide an example of such a relationship.

|  |
| --- |
| Weak Entity Relationship is where primary key of the related entity does not contain the primary key of the parent entity. It is an entity that cannot be uniquely identified by its attribute alone. On the other hand strong entity relationship is where primary key of the related entity will be present inside the primary component of the parent entity.  For E.g.: Carbon and its isotope, Carbon Isotopes only exists due to the carbon. On the other hand, fuel, and the bike. Fuel can exist with or without the bike, making it as a strong relationship.  Moreover, Dependents Employee and Company also follows weak entity relationship.  If you closely focus on the Employees. |

**Part 3 Data Models / Analytics - Advanced Topics in Data Management**

**(1) (Data Modeling )**

Explain multivalued attributes with the help of examples. How are multivalued attributes indicated in the Chen Entity Relationship Model?

|  |
| --- |
| Multivalued Attribute: As its name suggest Multivalued Attribute is an Attribute which holds more than one value. It is usually represented by the double ovels in the E-R Diagram.  In Chen Entity model, Multivalued Attribute is primarily denoted by double line connecting to that attribute. Consider an example of Bike as shown below,  Where Bike is having different attributes including the multivalued attributes.  Diagram  Description automatically generated  Color can be stated as a Multivalued attribute for the Bike and connected with the double lines. |

**(2) ( Data Analytics / Predictive Analytics: Aggregate Functions )**

In SQL database systems such as Oracle and DB2 , aggregate functions, such as averages and other statistical measures, can be used for analytical purposes. The following data illustrates company sales for the second half of the year. List five generic aggregate functions that can be used with the following data. Comment on their usefulness in describing the data.

|  |  |  |
| --- | --- | --- |
| **Table: tblSales** | |  |
|  |  |  |
| index | month | sales |
|  |  |  |
| 1 | July | $19,000 |
| 2 | August | $17,280 |
| 3 | September | $23,660 |
| 4 | October | $17,473 |
| 5 | November | $28,805 |
| 6 | December | $32,470 |

|  |
| --- |
| Five Generic aggregate functions are given as below,  SUM: Sum function will help us get the total sale amount over the duration of the six months starting from July.  COUNT: Count function could effectively help us to count the months occurring within the sales range.  MIN: Min function can be effectively used to calculate the Min sales by considering the sales column and returning the corresponding month.  MAX: Like MIN, MAX can be used to get the month having Maximum number of sales.  AVG: AVG function will help us to calculate the moving average that in turns in future will help us with analytics. moving average can smooth the curve and represent volatility in the stock index. Similarly, in terms of sales with average we can forecast future trends in the sale.  STDDEV: Standard Deviation function is used to calculate the spread of a data from its mean. it is also useful I am demining the volatility of the data. As higher the Standard Deviation higher will be the volatility. |

**Part 4 Data Design Concepts - Advanced Topics in Data Management**

**(1) ( Database Models )**

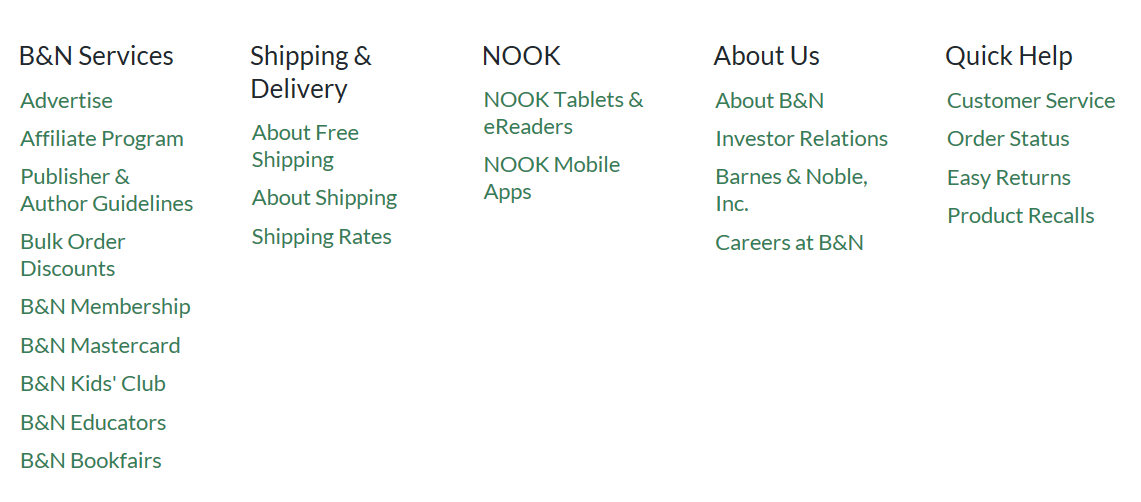
What is a ternary relationship? Provide some business rules examples that specify the need for a ternary or higher - order relationship.

|  |
| --- |
| Inside a relationship when we see three entities are equally participating then such relationship can be called as a ternary relationship.  For Example: Consider the entity relationship diagram having three different components Faculty, Class and Student, these components are connected by the operations.  Availability of the class time depends on the faculty time, Students time, and it also depends on the classroom status as it must not be occupied by some other classes.  Consider the following Data Model representing ternary relationship between three tables.  Graphical user interface  Description automatically generated  Similarly,  Consider product price is dependent on customer, the item and the vendor selling the item itself and can be represented as given below,  Chart  Description automatically generated |

**(2) ( Data Models and eCommerce Web Databases )**

Visit the home page of the Barnes and Noble Web site and examine especially the links towards the bottom portion of the page.

[**https://www.barnesandnoble.com**](https://www.barnesandnoble.com)



Then choose one of these categories, such as Careers at B&N or B&N Kid’s Club and discuss how database management plays a role in connecting the Web site with the customers of the company.

|  |
| --- |
| The Database plays really important role when it comes to the Barnes and Noble and connecting all the books form the databases with the customers. The Barnes and Noble website is connected to the backend system as it directly pull the data from the backend server and convert data stored in terms of object in the visually appealing list representation and present it in front of customers.  Consider the fiction group present on the Barnes and Noble website it helps individuals to understand new book releases and bestselling authors and all the individual book related data from the book price, publication, version and much more.  Customer interested in the Fiction genre could easily navigate through the thousands of e-books with just one click of a filter button. Rather than physically sorting thousands of shelves in the local library. A user can find the desired book with the help of filter function and the database within seconds.  One of the most important advantage of the databases is scalability, with increasing stores and widely expanding business, there is always requirement of more inventory space, with the help of cloud and database we can accomplish that and can provide exceptional service to the customers.  Moreover, if we analyze the about us section we get more important information about Barnes and Nobles, Career section is also embedded in the about us page.  Quick Help Section used to launch complaint and provide FAQ’s. Its is having a feedback form, directly connected to the database, each request generated by the customer allows to assign unique request ID in the database request table.  The new random ticket number is generated from the database and sent to the customer for request process tracking. Request table consist of various components including date, reqid, date, customerinfo, and other details.  Databases increases overall efficiency of an application and act as a backbone for online transaction processing. |