ITMD 523 *Advanced Topics in Data Management* HW 3

Student Name **Henry Post** Section **422**

Instructor ***Luke Papademas*** Due Date 08/05

Part	1	2	3	4	**TOTAL**	**Score**
Maximum Points	25 points	25 points	25 points	25 points	**100** points	

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

- OLAPs are concerned with analyzing data: Sum, count, average, stdev, comparisons…
- Less transactions, more complex calculations
- OLAP saves intermediate representations of data (like yearly, quarterly, monthly aggregates)
- OLAP systems store data in multidimensional schemas like Star and Galaxy, with Fact and Dimensional tables joined logically.
- This scheme is used in data mining to get data from OLAP systems.

- OLTP is about data reads and writes, and fast small transactions. Single-row CRUD ops are common.
- In OLTP, effectiveness is measured by the number of transactions per second.
- Very few indexes as compared to OLAP. Why use them?
- Data is highly normalized, low redundancy compared to OLAP.
- Large number of joins compared to OLAP.

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

Student Name **Henry Post** Section **422**

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 global parcel shipping company
- a sports memorabilia firm
- a nationwide college or university
- an office supply superstore
- a talent agency
- a vitamin and health shop

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

- A vitamin and health shop:
  - What's the most high-performing shop?
    - Where is it located?
    - When does it perform well?
  - What shops have the highest rate of stolen or lost goods?
  - What products are not selling well at all?
    - Is it even worth it to keep those products stocked?
    - What products could replace them?

- A nationwide college or university
  - GPA breakdown
    - What students perform the worst?
    - What characteristics do these students share?
    - By year
    - By major
  - Professor Ratings
    - Only on one subject?
    - Only on one class?
    - All classes (might want to fire them...)

Student Name **Henry Post** Section **422**

(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) Replace double-spaces with single-spaces.

```
-- Replace all double-spaces with single spaces.
UPDATE my_table
SET my_field = (REPLACE(my_field, '  ', ' '))
FROM my_table
WHERE (CHARINDEX('  ', my_field) > 0);
```

B) Finding duplicate rows:

```
-- Find duplicate rows
SELECT *, COUNT(*)
FROM table_name
GROUP BY duplicated_column
HAVING COUNT(*) > 1
```

C) Remove rows with a column that has a numeric value:

```
-- Delete rows with a numerical column
DELETE FROM my_table
WHERE (ISNUMERIC(my_column) = 1);
```

D) Delete rows with an alphanumerical column:

```
-- Delete rows with an alphanumerical column
DELETE FROM my_table
WHERE my_column LIKE "%[^a-zA-Z0-9]%";
```

E) asdf

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, CustomerID, ShippingDate, OrderStatus}

OrderDetails {OrderID, ProductID, Quantity, LineNumber}

Products {ProductID, Name, Description, Quantity, UnitPrice}

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

1. See all orders that are not yet delivered that are >3 weeks old.
2. See the products that are the most-sought after.
3. See the customers that order the most amount of products.
4. See a breakdown of the customers that pay the highest amount of money per month for orders.
5. See the least popular products.

(2) (Types of Relationships)

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

A weak relationship in an ERD is one that requires foreign keys in order to be fully identified.

An example of this could be a table that only contains two foreign keys, and that's it.

Student Name **Henry Post** Section **422**

Consider a 'Customer' and 'Account' table. They are not related directly to each other, but a 'CustomerAccount' contains two keys: 'CustomerID' and 'AccountID'. It is used to join ONE customer to ONE account.

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?

A multivalued attribute is a value that can hold more than one atomic piece of information. It is one that can be subdivided.

An example of this is a full name (last+first) or an address (street number, house number, etc.)

(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	$16,000
2	August	$17,280
3	September	$23,660
4	October	$17,493
5	November	$28,805
6	December	$32,470

Student Name **Henry Post** Section **422**

- SUM
+ Could be used to see total sales!

- RANK
+ Order by highest sales

- PARTITION
+ See quadrants of performance

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.

A ternary relationship is one where multiple entities can have a relationship with one object.

An example of this is a “Customer” and “Employee” table where both are allowed to buy items in a “Purchase” table.

(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

Student Name **Henry Post** Section **422**

Then choose one of these categories, such as Shipping Rates or B&N Membership, and discuss how database management plays a role in connecting the Web site with the customers of the company.