NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
VAALDRIEHOEKKAMPUS

Benodigdhede vir hierdie vraestel/Requirements for this paper:			
Multikeusekaarte/ Multi-choice cards:		Nie-programmeerbare sakrekenaar/ Non-programmable calculator:	
Grafiekpapier/ Graph paper:		Draagbare Rekenaar/ Laptop:	

Oopboek-eksamen/ Open book examination?	NEE/ NO

EKSAMEN/TOETS EXAMINATION/TEST:	**November 1st Opportunity 2017**	KWALIFIKASIE/ QUALIFICATION:	**B.Com, B.Sc, B.Sc in IT**
MODULEKODE/ MODULE CODE:	**ITRW321**	TYDSDUUR/ DURATION:	**3 hours**
MODULEBESKRYWING/ MODULE DESCRIPTION:	**Databases II**	MAKS/ MAX:	**125**
EKSAMINATOR(E)/ EXAMINER(S):	**AR BOTES**	DATUM/ DATE:	**03/11/2017**
		TYD/TIME:	**09:00**
MODERATOR:	**PROF TANYA STOTT**		

INSTRUCTIONS

1. *Answer all questions in the script provided.*
2. *Number questions clearly.*
3. *Write in BLUE or BLACK pen, diagrams in pencil and diagram text in pen.*
4. *Please write neatly; if answers cannot be read it cannot be marked.*
5. *Any many-to-many relationships in logical database diagrams will incur a 3 mark(-3) penalty per instance.*

QUESTION 1

1.

1.1. List and briefly explain the steps performed during the logical design stage. (8)

1.2. List and briefly explain the steps performed during the physical design stage. (6)

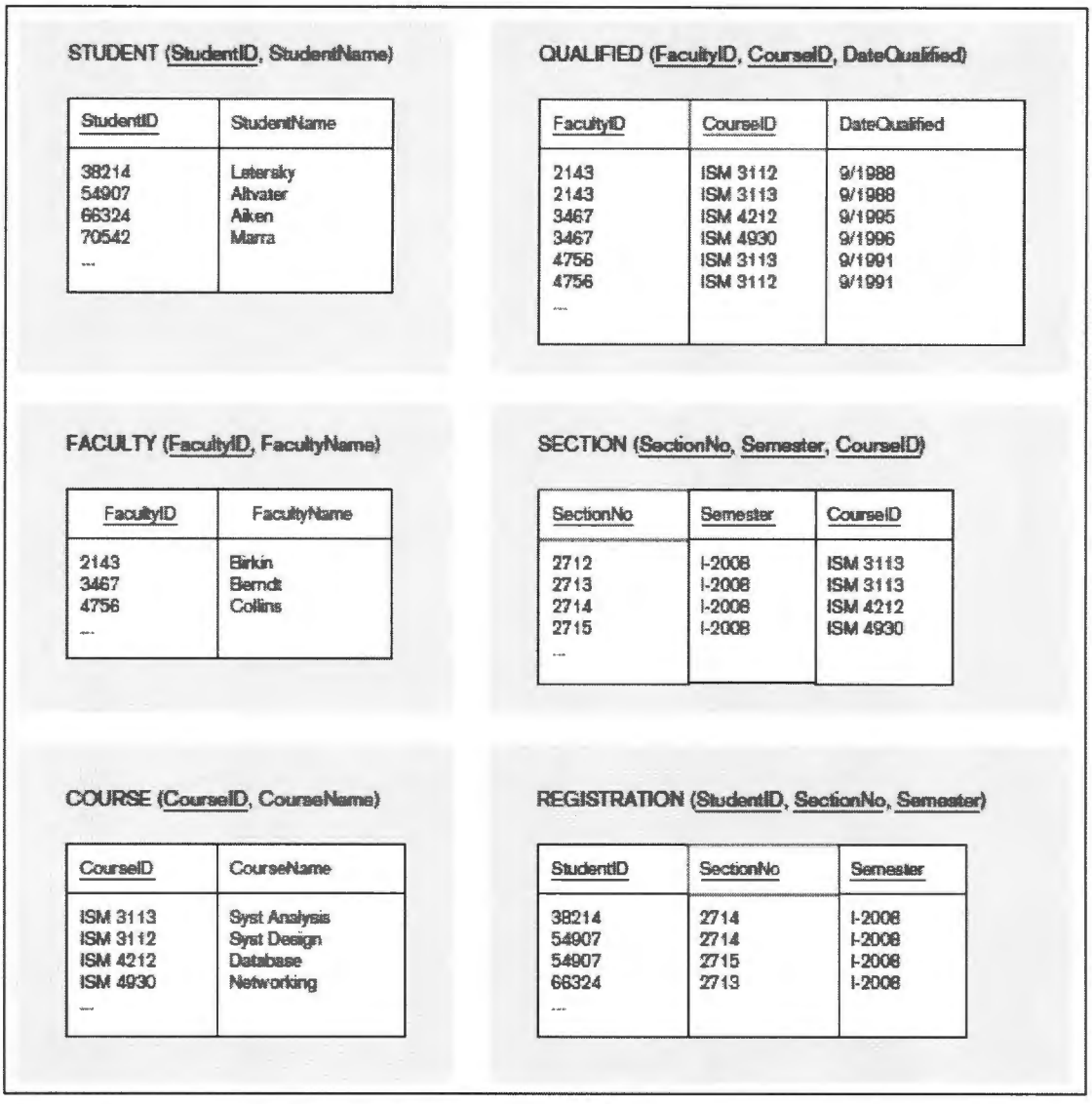
1.3. How do systems analysis and systems development fit into a discussion about information systems? (6)

[20]

QUESTION 2

2. Questions below related to the following case study of the class scheduling which includes 3NF relations along with some sample data shown.

Not shown in these figures are data for an ASSIGNMENT relation, which represents a many-to-many relationship between faculty and sections.



STUDENT (StudentID, StudentName)

StudentID	StudentName
38214	Letersky
54907	Altvater
66324	Aiken
70542	Marra
...	

QUALIFIED (FacultyID, CourseID, DateQualified)

FacultyID	CourseID	DateQualified
2143	ISM 3112	9/1988
2143	ISM 3113	9/1988
3467	ISM 4212	9/1995
3467	ISM 4930	9/1996
4756	ISM 3113	9/1991
4756	ISM 3112	9/1991
...		

FACULTY (FacultyID, FacultyName)

FacultyID	FacultyName
2143	Birkin
3467	Berndt
4756	Collins
...	

SECTION (SectionNo, Semester, CourseID)

SectionNo	Semester	CourseID
2712	I-2008	ISM 3113
2713	I-2008	ISM 3113
2714	I-2008	ISM 4212
2715	I-2008	ISM 4930
...		

COURSE (CourseID, CourseName)

CourseID	CourseName
ISM 3113	Syst Analysis
ISM 3112	Syst Design
ISM 4212	Database
ISM 4930	Networking
...	

REGISTRATION (StudentID, SectionNo, Semester)

StudentID	SectionNo	Semester
38214	2714	I-2008
54907	2714	I-2008
54907	2715	I-2008
66324	2713	I-2008
...		

2.1. Using this database, write the SQL code to represent each of the following transactions. Use BEGIN TRANSACTION, COMMIT, and ROLLBACK to group the SQL statements in logical transactions.

2.1.1. Create both forms of the INSERT command to add a student with a student ID of 65798 and last name Lopez to the Student table. (4)

2.1.2. Now write a command that will remove Lopez from the Student table. (2)

2.1.3. Modify the name of course ISM 4212 from Database to Introduction to Relational Databases. (2)

2.2. Assuming that pessimistic locking with two-phase locking protocol is used, create a chronological list of the locking, unlocking, and data manipulation activities that would occur during the complete processing of the transactions described in 2.1. (7)

2.3. Write the transaction log for the above transactions from question 2.1. (10)

[25]

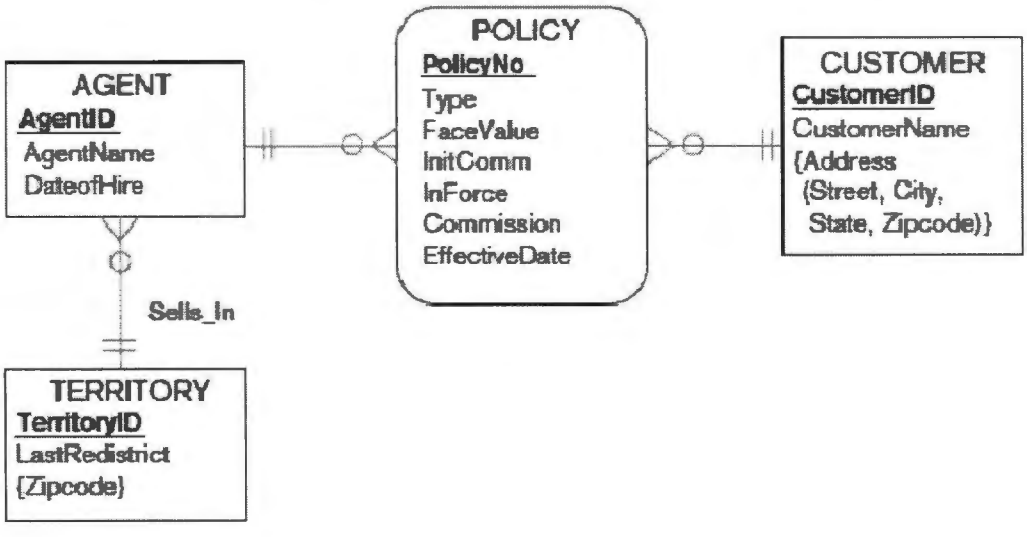
QUESTION 3

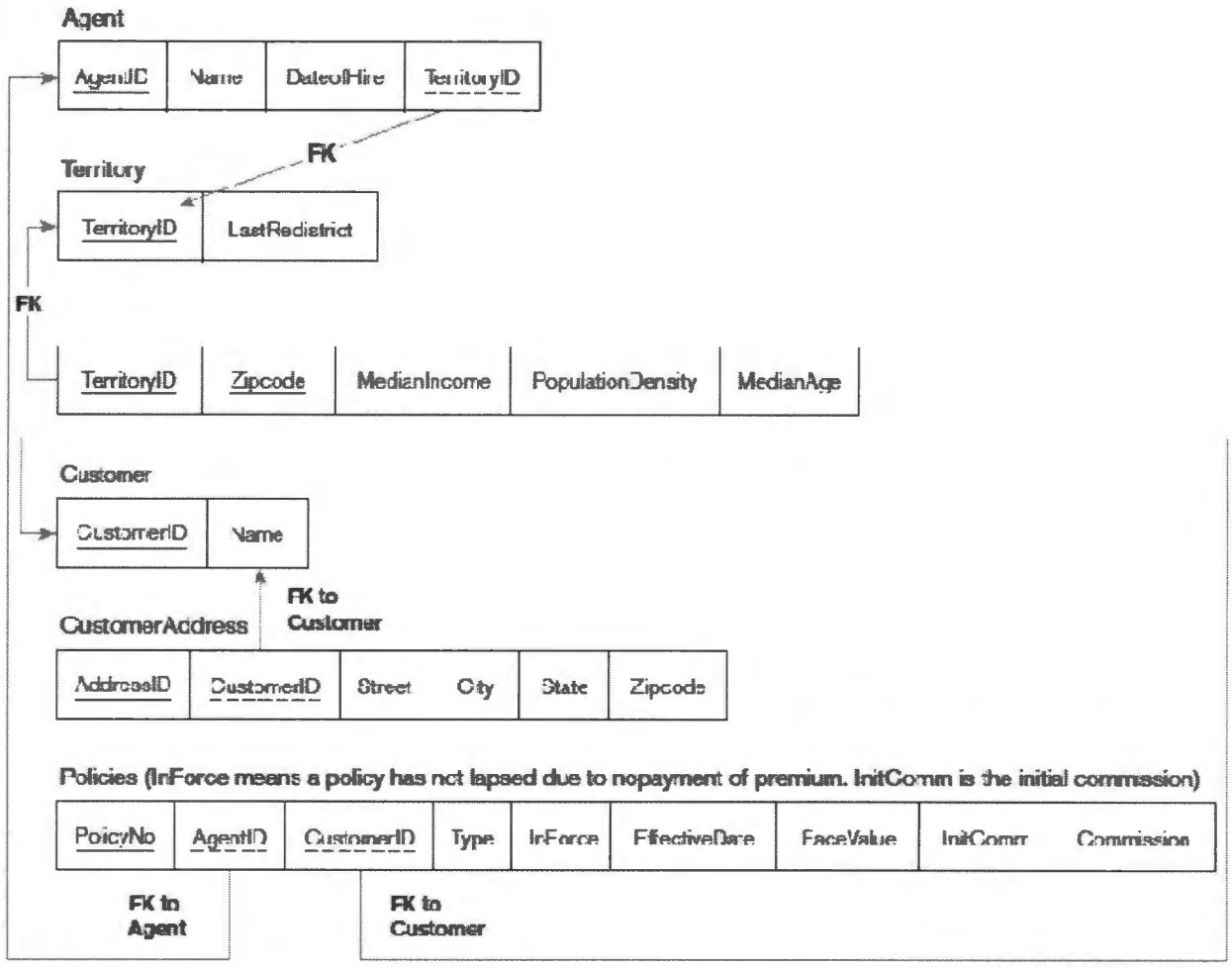
3. Analyse the case study below and answer the questions to follow:

WorkCube Insurance Company, which is primarily involved in the sale of annuity products, would like to design a data mart for its sales and marketing organization. Presently, the OLTP system is a legacy system residing on a Novell network consisting of approximately 600 different flat files. For the purposes of our case study, we can assume that 30 different flat files are going to be used for the data mart. Some of these flat files are transaction files that change constantly. The OLTP system is shut down overnight on Friday evening beginning at 6 P.M. for backup. During that time, the flat files are copied to another server, an extraction process is run, and the extracts are sent via FTP to a UNIX server. A process is run on the UNIX server to load the extracts into Oracle and rebuild the star schema. For the initial loading of the data mart, all information from the 30 files was extracted and loaded.

On a weekly basis, only additions and updates will be included in the extracts.
Although the data contained in the OLTP system are broad, the sales and marketing organization would like to focus on the sales data only. After substantial analysis, the ERD shown in the figure below was developed to describe the data to be used to populate the data mart. Sales and marketing is interested in viewing all sales data by territory, effective date, type of policy, and face value. In addition, the data mart should be able to provide reporting by individual agent on sales as well as commissions earned. Occasionally, the sales territories are revised (i.e., zip codes are added or deleted). The Last Redistrict attribute of the Territory table is used to store the date of the last revision. Some sample queries and reports are listed here:

- Total sales per month, by territory, by type of policy.
- Total sales per quarter by territory, by type of policy.
- Total sales per month by agent, by type of policy.
- Total sales per month by agent, by zip code.
- Total face value of policies, by month of effective date.
- Total face value of policies by month of effective date, by agent.
- Total face value of policies, by quarter of effective date.
- Total number of policies in force, by agent.
- Total number of policies not in force, by agent.
- Total face value of all policies sold by an individual agent.
- Total initial commission paid on all policies to an agent.
- Total initial commission paid on policies sold in a given month by agent.
- Total commissions earned by month, by agent.
- Top-selling agent by territory, by month.





Commissions are paid to an agent upon the initial sale of a policy. The InitComm field of the policy table contains the percentage of the face value paid as an initial commission. The Commission field contains a percentage that is paid each month as long as a policy remains active or in force. Each month, commissions are calculated by computing the sum of the commission on each individual policy that is in force for an agent.

3.1. Create a star schema for this case study. How did you handle the time dimension? (19)

3.2. Would you prefer to normalize (snowflake) the star schema of your answer to question 3.1? If so, how and why? Redesign the star schema to accommodate your recommended changes. (2)

[21]

QUESTION 4

4. The South African Vuvuzela (Pty) Ltd., wants to design a distributed relational database. The company is headquartered in Cape Town and has major operations in Durban, Johannesburg, and Pretoria. The database involved consists of five tables, labelled A, B, C, D, and E, with the following characteristics:

   - Table A consists of 500,000 records and is heavily used in Cape Town and Durban.
   - Table B consists of 100,000 records and is frequently required in all four cities.
   - Table C consists of 800 records and is frequently required in all four cities.
   - Table D consists of 75,000 records. Records 1-30,000 are most frequently used in Durban. Records 30,001-75,000 are most frequently used in Johannesburg.
   - Table E consists of 20,000 records and is used almost exclusively in Cape Town.

   Design a distributed relational database for Vuvuzela (Pty) Ltd. Justify your placement, replication, and partitioning of the tables.

Redraw and use the table template below:

Table/Location	Capetown	Durban	Johannesburg	Pretoria
Table A				
Table B				
Table C				
Table D				
Table E				

[13]

QUESTION 5

5. Draw a diagram to illustrate how the ADO.NET framework integrates from a client applications to the database.

[16]

QUESTION 6

6.

6.1. Evaluate the query below and answer the questions to follow:

```
SELECT EMP.EMP_NUM, EMP_FNAME, EMP_LNAME, EMP_EMAIL, TOTAL
FROM LGEMPLOYEE EMP JOIN
        (SELECT EMPLOYEE_ID, SUM(LINE_QTY) AS TOTAL
         FROM LGINVOICE I JOIN LGLINE L ON I.INV_NUM = L.INV_NUM
              JOIN LGPRODUCT P ON L.PROD_SKU = P.PROD_SKU
              JOIN LGBRAND B ON B.BRAND_ID = P.BRAND_ID
         WHERE BRAND_NAME = 'BINDER PRIME'
                    AND INV_DATE BETWEEN '01-NOV-15' AND '06-DEC-15'
         GROUP BY EMPLOYEE_ID) SUB
    ON EMP.EMP_NUM = SUB.EMPLOYEE_ID
WHERE TOTAL = (SELECT MAX(TOTAL)
                    FROM (SELECT EMPLOYEE_ID, SUM(LINE_QTY) AS TOTAL
                    FROM LGINVOICE I JOIN LGLINE L ON I.INV_NUM = L.INV_NUM
                         JOIN LGPRODUCT P ON L.PROD_SKU = P.PROD_SKU
                         JOIN LGBRAND B ON B.BRAND_ID = P.BRAND_ID
                    WHERE BRAND_NAME = 'BINDER PRIME'
                         AND INV_DATE BETWEEN '01-NOV-15' AND '06-DEC-15'
                    GROUP BY EMPLOYEE_ID));
```

6.1.1. What is the likely data sparsity of the EMP_EMAIL columns? (1)

6.1.2. Should you create an index on LINE_QTY? Why or why not? (2)

6.1.3. What type of database I/O operations will likely be used by the query? (2)

6.1.4. Assuming that there are no table statistics, what type of optimization will the DBMS use? (1)

6.1.5. What type of database I/O operations will likely be used by the query? (3)

6.2. Evaluate the query below with a sample of the data and answer the questions to follow:

```
SELECT      V_CODE, V_NAME, V_CONTACT, V_STATE
FROM  VENDOR
WHERE       V_STATE =' PA'
ORDER BY    V_NAME;
```

State	Number of Vendors	State	Number of Vendors
AK	15	MS	47
AL	55	NC	358
AZ	100	NH	25
CA	3244	NJ	645
CO	345	NV	16
FL	995	OH	821
GA	75	OK	62
HI	68	PA	425
IL	89	RI	12
IN	12	SC	65
KS	19	SD	74
KY	45	TN	113
LA	29	TX	589
MD	208	UT	36
MI	745	VA	375
MO	35	WA	258

6.2.1. Assume that 15,000 vendors are distributed as shown in the Table below. What percentage of rows will be returned by the query? (2)

6.2.2. What type of I/O database operations would be most likely to be used to execute that query? (Assuming and index on V_STATE were created) (2)

6.2.3. Assume that you have 15,000 different products stored in the PRODUCT table and that you are writing a Web-based interface to list all products with a quantity on hand (P_QOH) that is less than or equal to the minimum quantity, P_MIN. What optimizer hint would you use to ensure that your query returns the result set to the Web interface in the least time possible? Write the SQL code. (2)

[15]

QUESTION 7

7. Study the XML data below and convert it to the XML Schema document (XSD).

```
<?xml version="1.0" encoding="UTF-8"?>

<shiporder orderid="889923"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="shiporder.xsd">
  <orderperson>John Smith</orderperson>
  <shipto>
    <name>Ola Nordmann</name>
    <address>Langgt 23</address>
    <city>4000 Stavanger</city>
    <country>Norway</country>
  </shipto>
  <item>
    <title>Empire Burlesque</title>
    <note>Special Edition</note>
    <quantity>1</quantity>
    <price>10.90</price>
  </item>
  <item>
    <title>Hide your heart</title>
    <quantity>1</quantity>
    <price>9.90</price>
  </item>
</shiporder>
```

[10]

QUESTION 8

8. Name and describe the anomaly happening in the transaction table below:

TIME	TRANSACTION	STEP	STORED VALUE
1	T1	Read PROD_QOH	35
2	T1	PROD_QOH = 35 + 100	
3	T1	Write PROD_QOH	135
4	T1	*****ROLLBACK *****	35
5	T2	Read PROD_QOH	35
6	T2	PROD_QOH = 35 – 30	
7	T2	Write PROD_QOH	5

[5]

TOTAL: 125