

Faculty of Science

Course: CSCI 3030U Database Systems and Concepts

Final Exam (50%)

Time limit: 2 hours

Name: Onyedikachi Kalu Student ID: 100557846

1. (4 points)

Inference test over Functional Dependencies (FDs): construct test with two tuples with values 0 and ?. Describe all intermediate steps in your reasoning (points b-e). Assume a set of FDs F = {ABC 🡪 D, ABC 🡪 E, DEF 🡪 GI, ABC 🡪 F, G 🡪 IJ, KLM 🡪 NZ}.

* 1. Define Functional Dependency.

Functional dependency is a set attributes that determine other attributes in the relation

Based on prescribed FDs is it true that:

* 1. AB 🡪 GI? NO BECAUSE WE DON’T KNOW THAT AB - > C, SO WE CANNOT ASSUME THAT AB ->
  2. ABC 🡪 IJ? YES BECAUSE

**ABC: ABC-> D**

**ABC-> E**

**ABC-> F**

**DEF -> GI**

**G -> IJ**

* 1. DEF 🡪 F? YES BECAUSE DEF -> DEFGIJ
  2. DEF 🡪 KLM? NO BECAUSE DEF –> DEFGIJ AND KLM IS NOT IN THERE

Answer:

**ABC: ABC-> D**

**ABC-> E**

**ABC-> F**

**DEF -> GI**

**G -> IJ**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A | B | C | D | E | F | G | I | J | K | L | M | N | Z |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ? | ? | ? | ? | ? |

**AB:** **AB->AB**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A | B | C | D | E | F | G | I | J | K | L | M | N | Z |
| 0 | 0 | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |

**DEF : DEF -> GI**

**G -> IJ**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A | B | C | D | E | F | G | I | J | K | L | M | N | Z |
| ? | ? | ? | 0 | 0 | 0 | 0 | 0 | 0 | ? | ? | ? | ? | ? |

1. (4 Points)

What is the advantage of set-based canonical form for ODs over list-based canonical form?

Set-based canonical form is faster and more efficient

Given the list-based OD [ABC] |-> [DE] provide the mapping of it to set-based form.

[ABC] : [] |-> D , [ABC] : [] |-> E

{}: A~D

{A}: B~D

{D}: A~E

{AD}: B~E

{ABD}: C~E

What is the complexity of the mapping of list-based ODs to set-based ODs?

The complexity is linear

1. (4 points)

Closure test:

* 1. What is the closure of the set of attributes?
  2. Assume a set of FDs F = {AB 🡪 CD, E 🡪 F, GH 🡪 IJ, F 🡪 GH, I 🡪 K, LM 🡪 N, N 🡪 O}. Compute a closure of ABE+. Describe all intermediate steps.
  3. Based on the closure information is true that ABE 🡪 I, ABE 🡪 K, ABE 🡪 KL?

**ABE+ = ABE**

**ABE+ = ABE**

AB - > CD and AB is a subset of ABE+

ABE+ = ABECD

**ABE+ = ABECD**

E -> F and E is a subset of ABE+

ABE+ = ABECDF

**ABE+ = ABECDF**

F -> GH and F is a subset of ABE+

ABE+ = ABECDFGH

**ABE+ = ABECDFGH**

GH -> IJ and GH is a subset of ABE+

ABE+ = ABECDFGHIJ

**ABE+ = ABECDFGHIJ**

I -> K and I is a subset of ABE+

ABE+ = ABECDFGHIJK

**ABE+ = ABECDFGHIJ**

LM -> N but LM is NOT a subset of ABE+

ABE+ = ABECDFGHIJK

**ABE+ = ABECDFGHIJ**

N -> O and N is NOT a subset of ABE+

ABE+ = ABECDFGHIJK

Answer:

ABE 🡪 I : YES because I is in the ABE+

, ABE 🡪 K: YES because K is in the ABE+

, ABE 🡪 KL NO because L is NOT in the ABE+

1. (4 Points)

BCNF:

* 1. What does it mean that relation is in BCNF (provide definition)? BCNF stands for Boyce Codd Normal Form. BCNF is a case where a set functional dependencies implies all other attributes in the relation
  2. Assume set of FDs F = {AB 🡪 CD, AB 🡪 E, AB -> FG} over table T. Is table T in BCNF. Provide explanation. If answer is no, decompose T.

**YES** this is in BCNF because the set of FD’s is {AB} and AB implies everything else in the relation.

* 1. Assume set of FDs F = {AB 🡪 CD, AB 🡪 E, AB -> FG, FG -> IJ, FG -> MNOPQ} over table T’. Is table T’ in BCNF? Provide explanation. If answer is no, decompose T’ with all intermediate steps.

NO it is not in BCNF because the set of FD’s is {AB, FG} AB doesn’t violate BCNF but FG does because FG doesn’t imply AB.

**DECOMPOSE**

TABLE1(AB, CD, E, FG)

TABLE2(FG, IJ, MNOPQ)

TABLE1 is in BCNF because the set of FD’s is {AB} and AB implies everything else

TABLE2 is in BCNF because the set of FD’s is {FG} and FG implies everything else

Answer:

1. (4 Points)



* 1. What is the difference between CASCADE and RESTRICT grant option?

When CASCADE is revoked, anyone that person GRANTED the privilege, no longer has access to it while RESTRICT means the people that were granted the privilege would still have access to it.

* 1. Assume user A executes REVOKE P FROM B CASCADE. Describe how is it going to affect the grant diagram presented above? Does user C has a privilege P after this operation?

In this case, BP and CP would no longer have access to P but the CP at the bottom part would still have access to it

Answer:

1. (4 Points)

What are the limitations of the shortest path approach over graph search? Provide a motivating example over a sample graph with co-authors of the papers.

Propose, how to resolve this issue. Provide an example why the second approach works better.

1. (4 Points)

Airline Frequent Flyer Case Study (building data wareshouse).

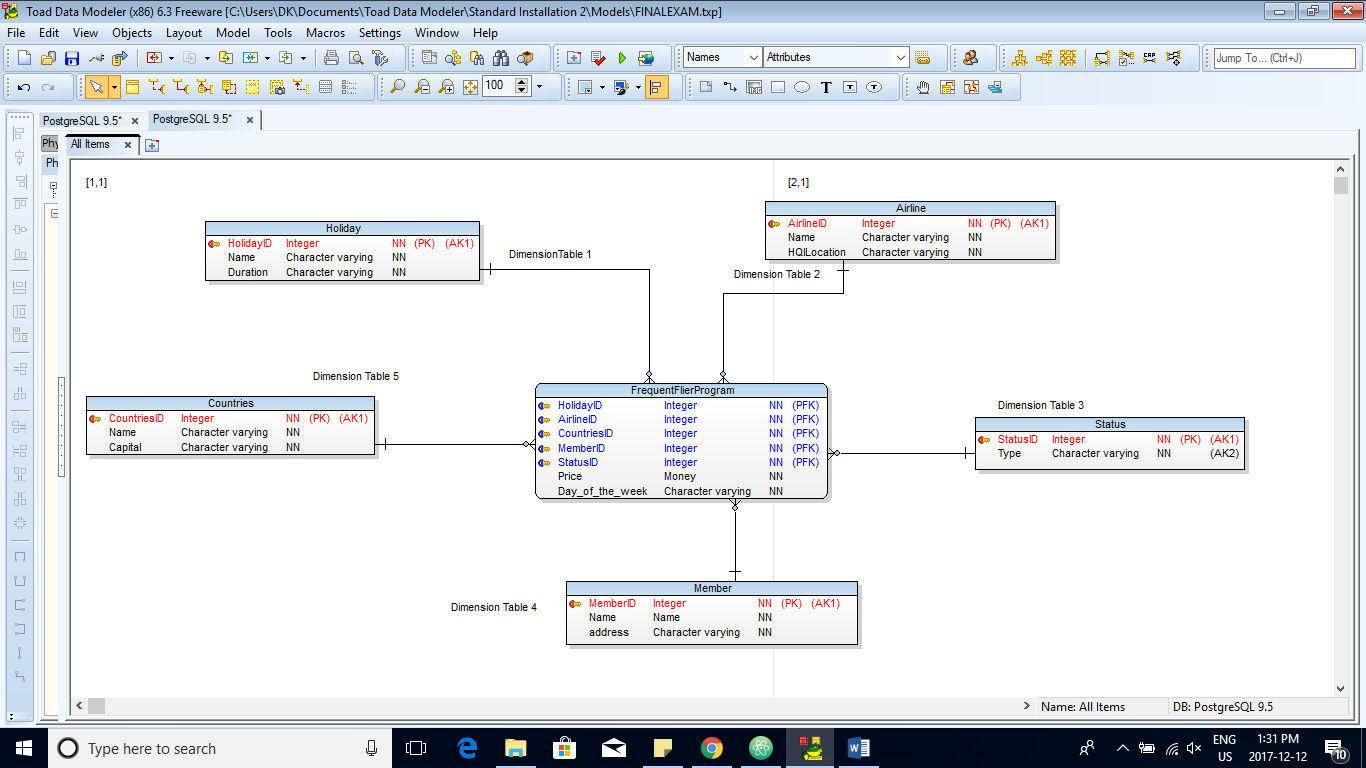
In this case the airline’s marketing department wants to analyze the flight activity of each member of its frequent flyer program. The department is interested in seeing what flights the company’s frequent flyers take, which days they fly and what are the name of the holidays associated with that day, which planes they fly, which countries they fly to, what fare basis they pay, how often they upgrade, and what proportion of these frequent flyers have titanium, platinum, gold, or aluminum status.

Using Toad Data Modeler provide a **star-schema** data warehouse for airline frequent flyer case study.

Provide description which tables are fact tables and dimension tables.

Which attributes are dimension attributes and dependent attributes in a fact table?

Answer:



FACT TABLE: Frequent Flier Program

DEPENDENT ATTRIBUTES: Price

Day\_of\_the\_week

DIMENSION ATTRIBUTES: CountriesID

MemberID

StatusID

AirlineID

HolidayID

DIMENSION TABLES: Countries

Member

Status

Airline

Holiday

1. (4 Points)

Provide DTD for airline frequent flyer database described in Point 6. In your specification include multiplicity, IDs, IDREFs and all other necessary details.

Answer:

<[!DOCTYPE PROGRAM

<!ELEMENT PROGRAM(FREQUENTFLIER\*) >

<!ELEMENT FREQUENTFLIER(COUNTRIES+, MEMBER+, STATUS+, AIRLINE+, HOLIDAY\*)>

<!ATTLIST FREQUENTFLIER price IDREF #REQUIRED>

<!ATTLIST FREQUENTFLIER Day\_of\_the\_week IDREF #REQUIRED>

<!ELEMENT COUNTRIES EMPTY>

<!ATTLIST COUNTRIES name ID #REQUIRED>

<!ELEMENT MEMBER EMPTY>

<!ATTLIST MEMBER name ID #REQUIRED>

<!ELEMENT HOLIDAY EMPTY>

<!ATTLIST HOLIDAY name ID #REQUIRED>

<!ELEMENT STATUS EMPTY>

<!ATTLIST STATUS name ID #REQUIRED>

<!ELEMENT AIRLINE EMPTY>

<!ATTLIST AIRLINE name ID #REQUIRED>

]>

­

<PROGRAM>

< FREQUENTFLIER price = "1,000" Day\_of\_the\_week = “Monday”>

</ FREQUENTFLIER >

<COUNTRIES name = "Nigeria" >

</COUNTRIES>

<MEMBER name = "Dikachi Kalu" >

</MEMEBER>

<HOLIDAY name = "Christmas" >

</HOLIDAY>

<STATUS name = "Platinum" >

</STATUS>

<AIRLINE name = "KLM" >

</AIRLINE>

</PROGRAM >

1. (4 Points)

Provide an example XML document for the DTD described in Point 7.

Answer:

<xs:complextype>

<xs:sequence>

<xs:element

name = “Country”

type = “xs:string”

minOccurs = “0” maxOccurs = “1”

>

<xs:element

name = “Member”

type = “xs:string”

minOccurs = “0” maxOccurs = “1”

>

<xs:element

name = “Airline”

type = “xs:string”

minOccurs = “0” maxOccurs = “1”

>

<xs:element

name = “Status”

type = “xs:string”

minOccurs = “1” maxOccurs = “1”

>

<xs:element

name = “Holiday”

type = “xs:string”

minOccurs = “0” maxOccurs = “1”

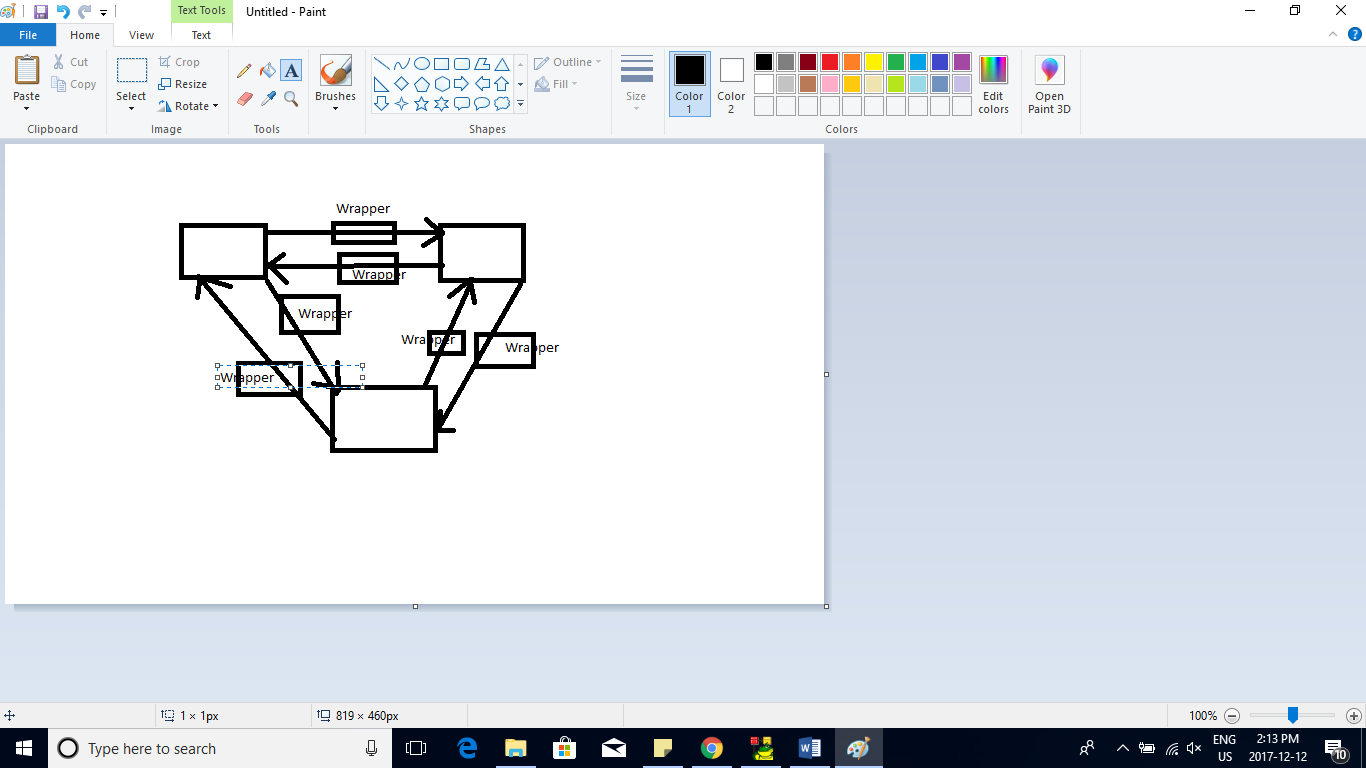
>

</xs:sequence>

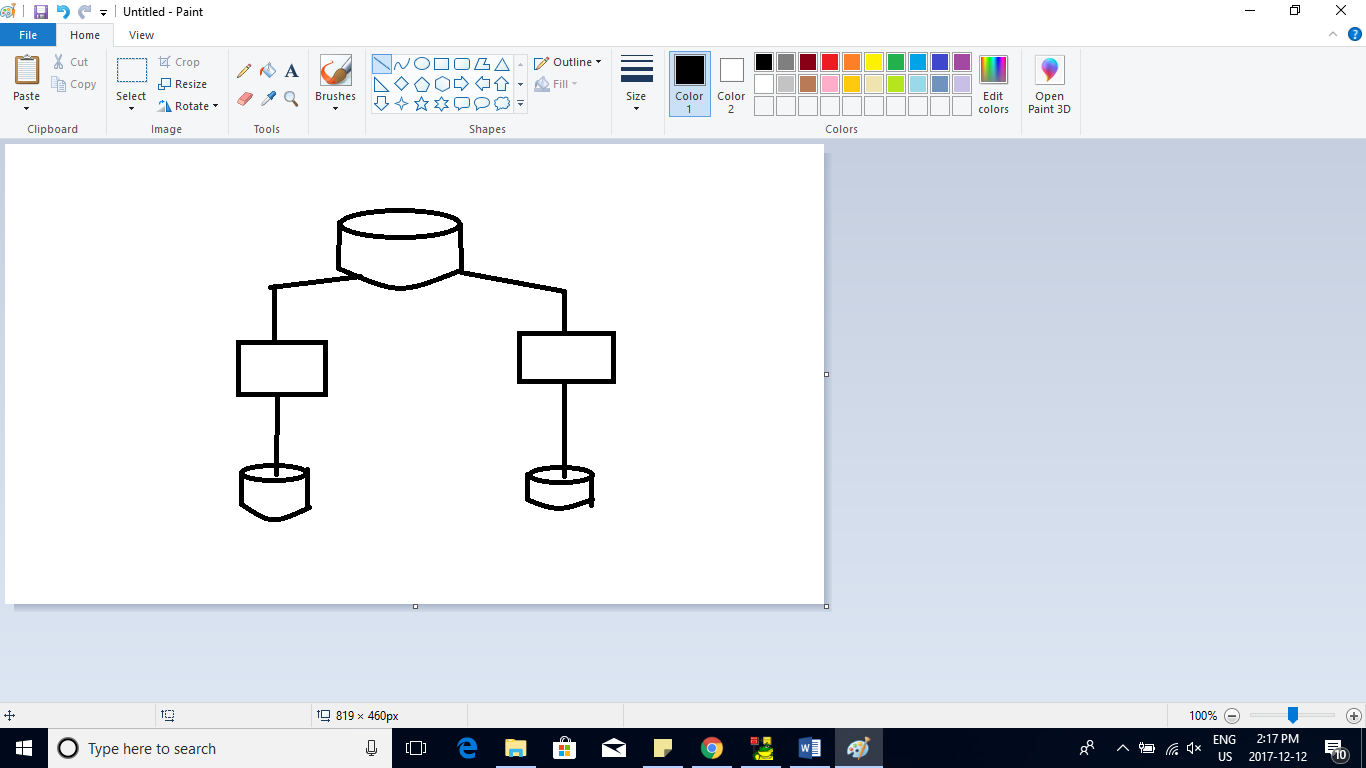
</xs:complextype>

1. (4 Points)
   1. Describe the difference between three different data integration architectures: Federations, Data Warehouse, and Mediator. (Support your answer by diagrams.)

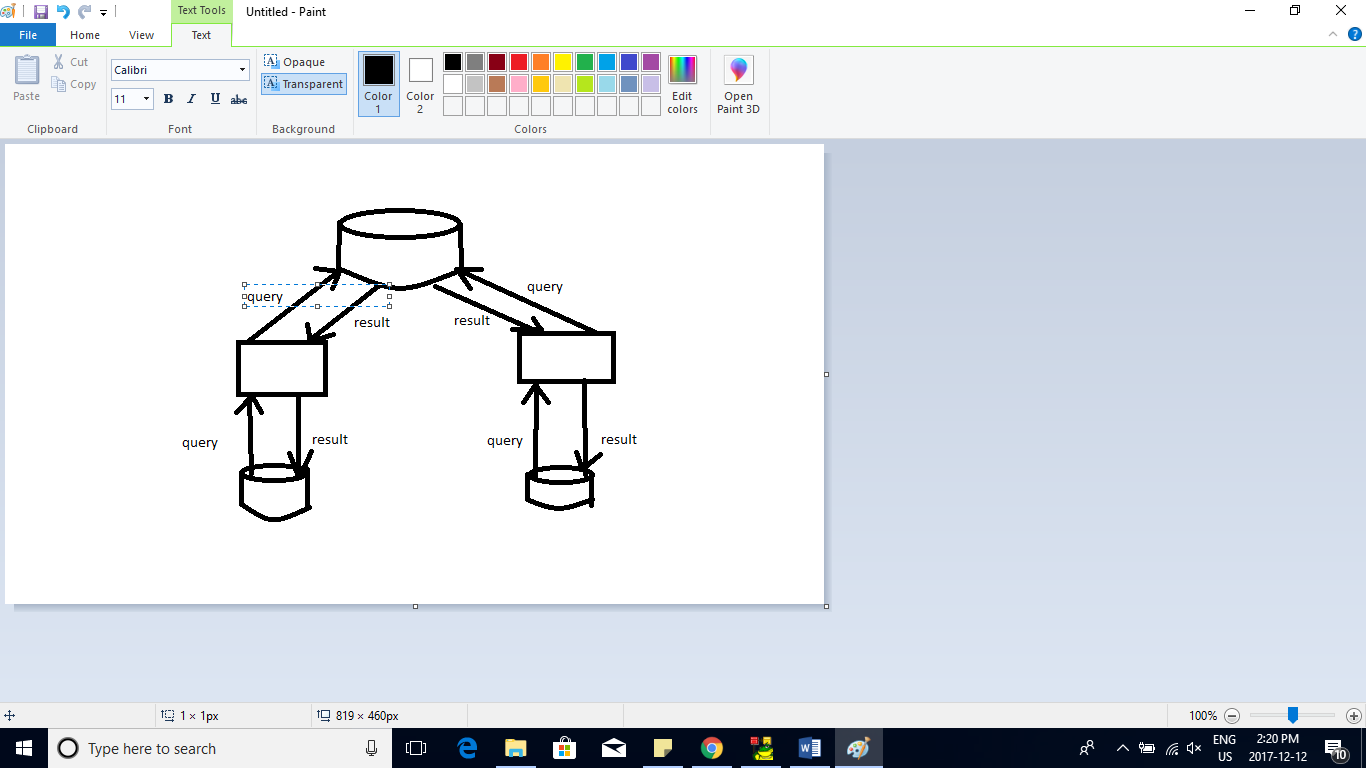
Federations involve all the relations talking to each other with the help of wrappers



Data warehouse involves the going from a local schema to a global schema the copied to a central DB



Mediator is a virtual Data Warehouse that uses queries and results



* 1. What is the difference between OLTP and OLAP systems?

OLTP stands for Online transactions processing while OLAP stands for Online Application processing. OLTP handles all the transactions, anything that involves interacting with the user while OLAP handles the application like what to do when a customer asks clicks on checkout

* 1. Compare superkeys and keys.

A **superkey** is the set of functional dependencies that imply all the attributes in a relation while a **key** is any functional dependency that is not a superkey itself

Answer:

Once completed upload the .doc file on Blackboard and check with person invigilating that your copy is actually available online.

Good luck!