CS465 Databases and Scripting Languages
Midterm
Fall 2017

1. This exam is closed-note, closed book.
2. You may not use any electronic devices.
3. Circle the appropriate answer(s) for the multiple choice/answer questions.
4. The last three pages of the exam gives the relations you should use for many of the exam questions. I recommend ripping these pages off the exam so that you can easily refer to them.
5. You must answer all of the questions. The CS465 version of the exam has 100 points and the CS565 version of the exam has 118 points. If you are an undergraduate please do not answer the CS565 questions as no extra credit will be awarded.
6. Good luck!

Multiple Choice (30 points): Choose the best answer from the following choices. Circle only one answer! If two answers seem like they might work, choose the best of the two answers.

1. In the relational model, the entity integrity constraint refers to the fact that:
    a. All attributes of the foreign key must be null or else reference an existing tuple in another relation
    b. The primary key of an entity must be non-null
    c. If an entity is deleted from a relation, all entities that depend on that entity will also be deleted from the database
    d. The relation representing an entity must have non-null fields for every attribute in the relation

2. Once the read/write arm has moved to the appropriate track of a disk, the time required for the appropriate block of a disk to move under the read/write arm is called:
    a. the transfer rate
    b. rotational latency
    c. the seek time
    d. the random access time

3. What is the formal basis for the DDL of a relational database?
    a. SQL
    b. Relational algebra
    c. Relational calculus
    d. Relational model

4. What is the English-language meaning of the following relational algebra query. I am abbreviating the relations as follows (RC = RaceCourse, CS = CourseSegment):

   $\Pi_{CS.segmentNo,\ distance,\ description}\ (\sigma_{courseName='Neyland\ Dr'}\ RC \bowtie CS)$

   a. Print information about each course segment on the Neyland Dr. race course.
   b. Print information about each course segment named Neyland Dr.
   c. Print information about each course segment whose start or finish line is on Neyland Dr.
   d. Print information about each course segment that has been part of a race named "Neyland Dr".

5. B+ trees have an advantage over extendible hashing for what type of query?
   a. Point query
   b. Range query
   c. Projection query
   d. Join query

6. In the RaceResult relation, am I likely to want to place a secondary index on runnerNo?
   a. No, it is an artificial key without much meaning
   b. No, the relation is too small for indexing
   c. No, a query on this attribute will retrieve a significant fraction of the relation
   d. Yes, it will be frequently used in joins involving the RaceResult relation
   e. Yes, it will be frequently used to develop index-only plans to solve a query
   f. Yes, it will often be used to sort the results of a query

7. Am I likely to want to place an index on the courseNo attribute in the RaceCourse relation?
   a. No, it is an artificial key without much meaning
   b. No, the relation is too small for indexing
   c. No, it involves a long character string
   d. Yes, it is the primary key of the RaceCourse relation and will be frequently used in joins
   e. Yes, it is a foreign key in the RaceCourse relation and will be frequently used in joins
   f. Yes, it will often be used to sort the results of a query

8. If extendible hashing is used for a secondary index, what is the worst case number of disk accesses that will be required to access a record from a relation if none of the accesses can be made from cache (include disk accesses for fetching both index blocks and the record block)
    a. 1
    b. 2
    c. 3
    d. 4
    e. $\log_{M/2}N$

9. courseNo in the Race relation is an example of what type of key?
    a. Primary key
    b. Foreign key
    c. Super key
    d. Alternate key

10. For the Volunteer relation, does the following dependency represent a full functional dependency?

    firstname, lastname, phoneNo → city, street, zipcode

    a. Yes, the determinant uniquely determines the volunteer's address.
    b. No, the determinant does not uniquely determine the volunteer's address and hence the dependency is invalid.
    c. No, phoneNo uniquely determines the volunteer's address and hence firstname and lastname are not required in the determinant.
    d. No, the pair (lastname, phoneNo) is the minimal set of attributes that uniquely determine the volunteer's address and hence firstname is not required in the determinant.
    e. No, a full functional dependency cannot contain all of the attributes in a relation

11. Which of the following queries would require set division?
    a. Find all people who are associated with a race as either a runner or volunteer.
    b. Find all people who have both run in a race and volunteered for a race.
    c. Find all people who have volunteered for all the races named "No Frills 10K"
    d. Find all people who have run a race in the past 40 days

12. Suppose you have the following ER diagram:

    Course -----------------→ Students
    1..* enrolls 1..*

    What is the degree of this relationship?
    a. 1 : 1
    b. 1 : *
    c. * : 1
    d. * : *

13. What is the English-language meaning of the following relational calculus query:

$$\{Ra.raceName, Ra.date \mid Race(Ra) \land Ra.date \geq 2014-01-01 \land$$
$$Ra.date < 2016-01-01 \land (\exists Ru)(\exists Res)(Runner(Ru) \land RaceResult(Res) \land$$
$$Ru.runnerName =' Smiley\ VanderZanden' \land Ru.runnerNo =$$
$$Res.runnerNo \land Res.raceNo = Ra.raceNo)\}$$

a. List the names and dates of all races in which Smiley Vander Zanden has been a volunteer
b. List the names and dates of all races in which Smiley Vander Zanden has been a runner.
c. List the names and dates of all races in which Smiley Vander Zanden has either run or volunteered.
d. List the names and dates of all races in 2014 and 2015 in which Smiley Vander Zanden either ran or volunteered.
e. List the names and dates of all races that Smiley Vander Zanden ran in 2014 and 2015
f. List the names and dates of all races in which Smiley Vander Zanden volunteered in 2014 and 2015.

14. Which of the following relational algebra queries allows us to derive a join between relations R and S using the fundamental relational algebra operations, as opposed to ones derived from the fundamental algebra operations?

a. $\sigma(R-(R-S))$
b. $\sigma(R \cup S-(R \cap S))$
c. $(R\ \sigma\ S) X\ R$
d. $\sigma(R\ X\ S)$

15. Behold the following chasm trap:

Branch $_{1..1} \longrightarrow {}_{1..*}$ Staff $_{0..1} \longrightarrow {}_{0..*}$ PropertyForRent

How can I redraw the above ER diagram to fix this chasm trap?

a. Branch $_{1..1} \longleftarrow {}_{1..*}$ Staff $_{0..1} \longrightarrow {}_{0..*}$ PropertyForRent
b. Branch $_{1..1} \longleftarrow {}_{1..*}$ Staff $_{0..1} \longleftarrow {}_{0..*}$ PropertyForRent
c. Staff $_{1..*} \longleftarrow {}_{1..1}$ Branch $_{1..1} \longrightarrow {}_{1..*}$ PropertyForRent
d. Branch $_{1..1} \longrightarrow {}_{1..*}$ Staff $_{0..1} \longrightarrow {}_{0..*}$ PropertyForRent

1..1 1..*

SQL and Relational Algebra Queries (16 points)

16. Which of the following SQL queries will return the list of volunteers who have been a race director for at least one race?

    a. SELECT firstname, lastname FROM Volunteer
       WHERE job = "race director";
    b. SELECT firstname, lastname FROM RaceStaffing
       WHERE job = "race director";
    c. SELECT firstname, lastname FROM Volunteer v
       WHERE phoneNo IN
       (SELECT phoneNo FROM RaceStaffing WHERE job = "race director");
    d. SELECT firstname, lastname FROM Volunteer v, RaceStaffing rs
       WHERE rs.job = "race director";

17. Which of the following relational algebra queries will list the names of all race courses that are run at a kilometer distance and that are >= 10 kilometers.

    a. $\Pi_{\text{courseName}}(\sigma_{\text{distanceMeasure='km' and distance >= 10}} \text{RaceCourse})$
    b. $\Pi_{\text{description}}(\sigma_{\text{distanceMeasure='km' and distance >= 10}} \text{RaceCourse} \bowtie \text{CourseSegment})$
    c. $\Pi_{\text{courseName}}(\sigma_{\text{distanceMeasure='km' or distance >= 10}} \text{RaceCourse})$
    d. $\sigma_{\text{distanceMeasure='km' and distance >= 10}} \text{RaceCourse}$
    e. $\sigma_{\text{distanceMeasure='km' or distance >= 10}} \text{RaceCourse}$
    f. $\sigma_{\text{distanceMeasure='km' and distance >= 10}} \text{RaceCourse} \bowtie \text{CourseSegment}$

18. Which of the following SQL queries will list the names of all runners who have finished first in a race. Do not print a runner more than once and print the runners in alphabetical order.
    a. SELECT runnerName FROM Runner, RaceResult
       WHERE Runner.runnerNo = RaceResult.runnerNo
       and RaceResult.finishPosition = 1
       GROUP BY runnerName;
    b. SELECT DISTINCT runnerName FROM RaceResult
       WHERE finishPosition = 1
       ORDER BY runnerName;
    c. SELECT DISTINCT runnerName FROM Runner, RaceResult
       WHERE Runner.runnerNo = RaceResult.runnerNo
       and RaceResult.finishPosition = 1
       ORDER BY runnerName;
    d. SELECT DISTINCT runnerName FROM Runner, RaceResult
       HAVING Runner.runnerNo = RaceResult.runnerNo
       and RaceResult.finishPosition = 1
       GROUP BY runnerName;

19. Which of the following queries will find the names of volunteers who have helped in more than 10 races?
    a. SELECT firstname, lastname FROM Volunteer NATURAL JOIN RaceStaffing
       GROUP BY phoneNo, firstname, lastname
       HAVING COUNT(*) > 10;
    b. SELECT firstname, lastname FROM Volunteer NATURAL JOIN RaceStaffing
       WHERE Count(*) > 10;
    c. SELECT firstname, lastname FROM Volunteer NATURAL JOIN RaceStaffing
       ORDER BY phoneNo, firstname, lastname
       HAVING COUNT(*) > 10;
    d. SELECT firstname, lastname FROM Volunteer
       GROUP BY phoneNo, firstname, lastname
       HAVING COUNT(*) > 10;

Multiple Answer (13 points): Choose all answers that are appropriate from the following choices.

20. What are the candidate keys for the RaceResult relation (**2 correct answers**)?
    a. raceNo
    b. runnerNo
    c. time
    d. finishNo
    e. raceNo, runnerNo
    f. raceNo, time
    g. raceNo, finishPosition
    h. runnerNo, time
    i. runnerNo, finishPosition
    j. time, finishPosition

21. Which of the following operations are considered fundamental to the relational algebra, as opposed to a convenience operation that can be derived from the basic operations **(5 correct anwers)**?
    a. Select
    b. Project
    c. Join
    d. Set Difference
    e. Set Union
    f. Set Division
    g. Set Intersection
    h. Cartesian Product

22. Which of the following properties must be observed when using normalization to perform a decomposition of a relation into a set of smaller relations **(2 correct answers)**?
    a. A constraint in the original relation can be enforced by one or more constraints on each of the smaller relations
    b. There must be no duplicated columns in any of the smaller relations (i.e., given any pair of relations from the set of smaller relations, there is no shared column)
    c. Any tuple in the original relation can be re-constructed from corresponding tuples in the smaller relations
    d. Any column that contains a multi-valued attribute in the original relation must be included as a column containing a multi-valued attribute in at least one of the smaller relations.

23. **(CS465-4 points, CS565-6 points) ER Diagrams:** The restaurant database has three entities—restaurants, menu items, and customers. Two of the relationships are 1) restaurants offer menu items, and 2) customers order menu items at restaurants. The ER diagram might be drawn as:

```
         ______________________________/order\___________________________
        |                              \     /                           |
        |                                 |                              | (3)
    Restaurants -----------------------→ Menu Items                  Customers
          1           offer            2
```

The numbers shown above represent missing multiplicities for the relationships. You may assume that:

a. Restaurants offer at least one menu item.
b. A menu item must be offered by at least one restaurant.
c. Restaurants offer at most 30 menu items.
d. Menu items can be offered at an unlimited number of restaurants.
e. Given a restaurant and a customer, that customer may have ordered 0 or more menu items from that restaurant.
f. At any restaurant, zero or more customers has ordered each menu item
g. A customer may have ordered the same menu item at 0 or more restaurants

For each number, enter the letter associated with its multiplicity (numbers 1 and 2 refer to the restaurants offer menu items relationship and number 3 refers to the customers order menu items at restaurants relationship):

a. 1..1
b. 0..1
c. 1..0
d. 1..*
e. *..*
f. 0..*
g. *..0
h. *..1
i. 0..30
j. 1..30
k. 30..0
l. 30..1

1) ______________

2) ______________

3) ______________ **(CS 565 students only)**

Normalization (16 points): The following universal relation shows the orders placed by customers at a restaurant chain. Each line represents a single item ordered by a customer.

Order(chairNo, tableNo, restaurantNo, restaurantAddress, menuItemNo, menuItemDescription, menuItemPrice, quantityOrdered, date, time, creditCardNo, cardHolderName)

Make the following assumptions:

1. This relation has the following functional dependencies:
   chairNo, tableNo, restaurantNo, date, time → creditCardNo
   chairNo, tableNo, restaurantNo, date, time, menuItemNo → quantityOrdered
   creditCardNo → cardHolderName
   menuItemNo → menuItemDescription, menuItemPrice
   restaurantNo → restaurantAddress

2. The entities are restaurants, menu items, and cardholders.
3. The primary key is (chairNo, tableNo, restaurantNo, date, time, menuItemNo)

Answer the following questions.

24. This table is susceptible to various types of anomalies. Circle all of the following anomalies that apply **(3 correct answers)**:
    a. If you delete the last order containing a menu item, you lose information about the menu item
    b. If you delete the last order containing a chair number you lose information about that chair number
    c. If you change the quantity ordered, then you must update every instance of that quantity ordered or you will have an inconsistency
    d. If you change a menu item's description, you must change every instance of that description or you will have an inconsistency
    e. If you change a restaurant number you must change every instance of that restaurant number or you will have an inconsistency
    f. You cannot insert a new restaurant unless it already has an order associated with it
    g. You cannot insert a new table unless it has an associated restaurant
    h. You cannot insert a new order unless it has a chair, table, and restaurant number associated with it.

25. Next to each of the following functional dependencies, indicate whether they are:
    a. Used to convert the relation from 1st to 2nd normal form
    b. Used to convert the relation from 2nd to 3rd normal form
    c. Not used in the decomposition because the dependency represents a candidate or primary key

    Please circle the correct letter next to each dependency

    a b c chairNo, tableNo, restaurantNo, date, time → creditCardNo
    a b c chairNo, tableNo, restaurantNo, date, time, menuItemNo → quantityOrdered
    a b c creditCardNo → cardHolderName
    a b c menuItemNo → menuItemDescription, menuItemPrice
    a b c restaurantNo → restaurantAddress

26. **Queries (CS 565 students only):** Write the following queries in the requested language:

    a. (**SQL- 8 points)** Print the name of each runner and their average time in 10K races.

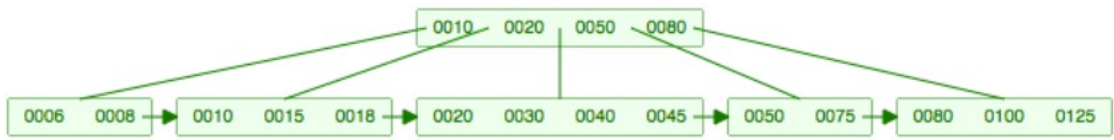
    b. **(Relational Calculus-8 points)** Write a relational calculus query that lists all the races run by Winnie the Pooh, including the race name, date, Winnie's time, and Winnie's finish position.

27. **ER Diagrams (13 points):** Draw an ER diagram for the race database. Only draw the entities and relationships. Do not worry about multiplicities. Make sure you label the relationships and where appropriate, draw arrows indicating the directionality of the relationship.

    Hint: One or more of the relations are not entities but relationships.

28. **B+ Trees (8 points)** Show what the following B+ tree looks like after 42 is inserted into it.

- Assume that M = 5 and L = 4.
- When you split a node and one of the two nodes must contain an odd number of keys, put the odd number of keys in the leftmost node.



The following information should be used in answering the questions on this exam. You are given the following relations about a races database that is maintained by a running club.

Runner(runnerNo, runnerName, city, street, zipcode, dateOfBirth)
RaceCourse(courseNo, courseName, raceDistance, distanceMeasure, city, startLine, finishLine)
CourseSegment(courseNo, segmentNo, segmentDistance, description)
Race(raceNo, courseNo, raceName, date)
RaceResult(raceNo, runnerNo, time, finishPosition)
Volunteer(lastname, firstname, city, street, zipcode, phoneNo)
RaceStaffing(raceNo, phoneNo, job)

- The RaceCourse relation contains all the race courses on which we run our races. There are about 50 race courses. The distanceMeasure is either 'km' (kilometers) or 'miles'. As an example, a race might be a 5 km race or a 5 mile race. Both the courseNo and courseName uniquely identify the course.
- The CourseSegment relation contains a description of the individual segments that comprise each race course. SegmentNo reflects the order in which the segments are run. Hence a segment number of 1 means it's the first segment which is run on the course, a segment number of 2 means it's the second segment which is run on the course, etc. Hence segment numbers will not be unique in the CourseSegment relation.
- The Race relation contains all the races that we have run. A race may be run on multiple years and over different courses. For example the relations below show that the NoFrills race was run on a Maryville course in 2011 and a Knoxville course in 2012 and 2013.
- The Runner and RaceResult relations should be self-explanatory. The "finishPosition" attribute in RaceResult is the position in which the runner finished (e.g., 1, 2, 3, etc.). There can be no ties for position in a race. That is for a given race, finish positions are unique. However, two runners can finish with the same time so times will not be unique.
- The Volunteer relation keeps track of all volunteers who help out with races. The phoneNo (phone number) for each volunteer is unique. In other words, phoneNo uniquely determines a volunteer. Although it may appear that (firstname, lastname) uniquely determine a volunteer, you may **not** assume that that is the case.
- The RaceStaffing relation tells us which volunteers worked at each race and what job they performed. Volunteers perform different jobs at different races but for any specific race, they perform only one job.
- A course may have multiple different races run on it. For example, in the sample relations below, both the AutumnFest and NoFrills races are held on course C01. There is no limit to the number of races run on a course. A course must have hosted at least one race to be in the database.
- Runners may run in multiple races and must have run in at least one race to be in the database.

Sample entries in the tables might be:

Runner

runnerNo	runnerName	city	street	zipcode	dateOfBirth
1	Smiley	Knoxville	Craghead	37920	11-28-05
2	Winnie The Pooh	Crossville	Ironwood	38412	5-25-73

RaceCourse

courseNo	**courseName**	**distance**	**distanceMeasure**	**city**	**startLine**	**finishLine**
C01	Neyland	5	miles	Knoxville	5500 Neyland Dr	Vet School
C02	Knoxville Downtown	10	km	Knoxville	200 Gay St.	Civic Coliseum
C03	Maryville Greenway	5	miles	Maryville	21 Greves Ln	85 W Main

CourseSegment

courseNo	**segmentNo**	**distance**	**description**
C01	1	1.3km	East on Neyland Dr
C01	2	0.3km	North on Philip Fulmer WayWay
C01	3	0.5km	East on Volunteer Dr
C02	1	0.8km	South on Gay St

Race

raceNo	**courseNo**	**raceName**	**date**
R01	C01	NoFrills	1-3-12
R02	C02	Expo10K	5-25-12
R03	C01	NoFrills	1-4-13
R04	C02	Expo10K	5-26-13
R05	C01	AutumnFest	10-12-13
R06	C02	NoFrills	1-2-11

RaceResult

raceNo	**runnerNo**	**time**	**finishPosition**
R01	1	0:38:50	58
R01	5	0:25:14	3
R02	1	0:48:15	11
R02	8	0:32:13	1

Volunteer

lastname	**firstname**	**city**	**street**	**zipcode**	**phoneNo**
Vander Zanden	Brad	Knoxville	1678 Cardiff Rd.	37920	441-4838
Richards	King	Morristown	45 Oak Dr.	37950	394-3948
Charles	Nancy	Maryville	88 Maple Ln	38393	129-9485
Seddrik	Carol	Knoxville	4800 Cardiff Rd	37920	394-3946

RaceStaffing

raceNo	phoneNo	Job
R01	441-4838	Timekeeper
R01	129-9485	Starter
R01	394-3946	Race director
R02	394-3946	Course director
R02	441-4838	Race Director