

**Information Systems for Engineers**  
**Fall 2018**  
**Final Exam**  
**8 February 2019**

Department of Computer Science  
ETH Zürich

**Rules (please read carefully):**

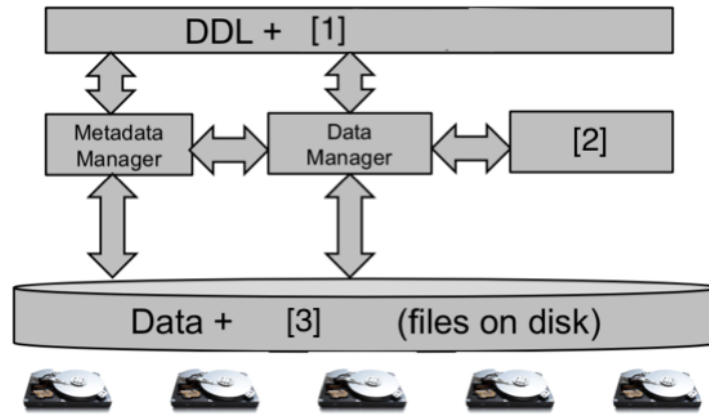
- You have 120 minutes for the exam.
- Please write your name and Legi number on the answer sheet **now** but do not start reading any questions until the proctors tell you to start.
- This is a multiple-choice examination. Most questions come with four possible answers (a, b, c, and d), among which exactly one is correct, **unless specified otherwise**. Picking the correct answer for a question will give you one point.
- Some multiple-choice questions have multiple correct answers. You must mark all the correct answers to get one point, otherwise no points given (i.e., no partial points).
- For multiple-choice questions, please **fill the relevant square completely on the answer sheet** to mark your answers.
- Some questions are open and expect you to write down an integer/letter or a list of integers/letters. The answer must be fully correct to get one point, otherwise no points given (i.e., no partial points).
- For open questions, **write down your answer on the answer sheet**. If your answer includes a list of integers, separate your integers with “comma” (e.g., 0, 1, 2, 3, ...).
- **Any answers or marks on other pages than the answer sheet will be completely ignored even if they are correct.**
- The exam consists of 100 questions. The maximum number of points that can be collected is 100.
- There is no penalty for incorrect answers. We, thus, encourage you to answer every question.
- If you think that you marked a wrong answer and you would like to change it, **cross the wrong answer and fill the one that you think it is the correct answer**. If an option is already crossed, we will consider that answer as “unmarked”. **Thus, if the crossed option is correct, you will not get a point.**
- **Only use black or blue pen. Do not use red pens.**
- The exam consists of 28 pages (including this page). **Let proctors know in advance if you have a missing page.**

*This page left blank intentionally*

## 1. Data Modeling – Basics

1. Which one of the following does **NOT** specify a data structure?
  - a. Trees
  - b. Graphs
  - c. Big Data
  - d. Cubes
  
2. Which one of the following has the correct descending order of the data sizes (from left to right)?
  - a. Tera > Peta > Giga > Mega > Kilo
  - b. Zetta > Giga > Tera > Peta > Kilo
  - c. Exa > Peta > Mega > Giga > Kilo
  - d. Yotta > Tera > Giga > Mega > Kilo
  
3. In a 3-tier architecture (with one top level, middle level, and bottom level), which one of the following is a list of the layers that can be present in a 3-tier architecture?
  - a. User Interface, Business Logic, Persistency
  - b. User Interface, Persistency, Maintenance (e.g., repairing broken disks)
  - c. Business Logic, User Interface, Maintenance
  - d. Project requirements, User Interface, Business Logic
  
4. Which one of the following is currently **NOT** one of the *main* features offered by most of Database Management System?
  - a. To save development costs
  - b. To avoid data loss
  - c. To enforce integrity constraints
  - d. To infer results given a set of data with Machine Learning
  
5. Which of the following storage locations **can** be used as the storage layer of a database? Select **ALL** correct answers (e.g., if both a and b are correct, then mark both).
  - a. User local hard drive
  - b. Server local hard drive
  - c. Shared network drive
  - d. CPU caches
  
6. What determines the validity of an instance against a schema? Select **ALL** correct answers (e.g., if both a and b are correct, then mark both).
  - a. Structure (set of attributes)
  - b. Types
  - c. Size of the disk
  - d. Integrity Constraints
  
7. Which one of the following is correct for SQL?
  - a. SQL is only an imperative language
  - b. SQL is only a declarative language
  - c. SQL is both an imperative and a declarative language
  - d. SQL is not a language

The following figure shows the overall database architecture with some missing parts. **Answer the questions 8,9, and 10 according to the figure.** Note that we do not consider PL/SQL in any of the questions below.



8. Which one of the following is the correct answer for the missing part [1]?
  - a. Transaction interpreter
  - b. Metadata interpreter
  - c. Data manipulation language interpreter
  - d. None of the above
  
9. Which one the following is the correct answer for the missing part [2]?
  - a. Transaction manager
  - b. Data manipulation language manager
  - c. DML manager
  - d. None of the above
  
10. Which one of the following is the correct answer for the missing part [3]?
  - a. Transaction
  - b. Metadata
  - c. Data manipulation language
  - d. None of the above

## 2. The Relational Model

11. Which of the following is **NOT** a well-formed set in a mathematical sense?
  - a.  $\{(1, a), (2, 3)\}$
  - b.  $\{\{1, 2, 3\}, \{1, 1\}\}$
  - c.  $\{a, b, c\}$
  - d.  $\{\}$
  
12. Which of the following is an incorrect statement for a table?
  - a. Order of attributes does not matter
  - b. Order of tuples does not matter
  - c. Extensions cannot be empty
  - d. In the set-based algebra, tuples must be unique
  
13. Which one of the following is a data type usable as a domain in a relational database?
  - a. Row
  - b. Integer
  - c. Attribute
  - d. Relational table
  
14. Which of the following is NOT correct about extensions?
  - a. It contains a set of tuples
  - b. There is always a relation between a set of attributes and a set of extensions
  - c. A set of extensions never contains partial functions from one set to another
  - d. A set of extensions may evolve over time

Consider the following table and answer Questions from 15 to 20 based on the table below.

<b>ID</b> <i>Integer</i>	<b>Name</b> <i>String</i>	<b>Height (cm)</b> <i>Real</i>	<b>Birth Date</b> <i>Date</i>	<b>Early Birth</b> <i>Boolean</i>	<b>Hospital</b> <i>String</i>
0	Cicek	53.2	1993-03-13	true	Hospital A
1	Emma	49.9	1h 16m	false	Hospital A
2	Can	52	1992-04-19	true	Hospital B
3	92	-1	1999-03-21	false	Hospital C
4		49	2018-12-29	false	Hospital
6	David	tall	2000-11-04	true	Any

15. Which of the following is the set of attributes of the table?
- {Integer, String, Real, Date, Boolean, String}
  - {Integer → "ID", String → "Name", Real → "Height (cm)", Date → "Birth Date", Boolean → "Early Birth", String → "Hospital"}
  - {"ID", "Name", "Height (cm)", "Birth Date", "Early Birth", "Hospital"}
  - {"Name", "Height (cm)", "Birth Date", "Early Birth", "Hospital"}
16. Which of the following is a tuple from the table?
- {Integer → "ID", String → "Name", Real → "Height (cm)", Date → "Birth Date", Boolean → "Early Birth", String → "Hospital"}
  - {Integer → 0, String → "Cicek", Real → 53.2, Date → 1993-03-13, Boolean → true, String → "Hospital A"}
  - {ID → 2, Name → "Can", Height (cm) → 52, Birth Date → 1992-04-19, Early Birth → true, Hospital → "Hospital B"}
  - None of the above
17. How many rows violate the schema. Your answer is: \_\_\_\_\_ (Only write a number, not the rows. Write 0 if there is none)
18. Which of the following is a correct statement about the fourth row (ID = 3). Select **ALL** correct answers (e.g., if both a and b are correct, then mark both).
- Fourth row violates the schema because "**Name**" is an integer value
  - Fourth row violates the schema because "**Height (cm)**" cannot be a negative value
  - Fourth row violates the schema because "**Height (cm)**" cannot be an integer value
  - Fourth row does not violate the schema.
19. Which of the following is a correct statement about the fifth row ("ID" = 4). Select **ALL** correct answers (e.g., if both a and b are correct, then mark both).
- Fifth row violates the schema because "**Name**" cannot be an empty string
  - Fifth row violates the schema because "**Hospital**" cannot have "Hospital" value as a string
  - Fifth row violates the schema because "**Height (cm)**" cannot be an integer value
  - Fifth row does not violate the schema.
20. Which of the following is a correct statement about the sixth row ("ID"=6). Select **ALL** correct answers (e.g., if both a and b are correct, then mark both).
- Sixth row violates the schema because "**Height (cm)**" cannot be **tall**
  - Sixth row violates the schema because "**Hospital**" cannot be set to "**Any**"
  - Sixth row violates the schema because "**ID**" should be 5 instead of 6.
  - Sixth row does not violate the schema

### 3. Data Definition with SQL

21. Which of the following is **NOT** a SQL type that describes *exact numbers* (i.e., *Number, exact*)? You may assume that you have an unlimited precision.
- real
  - integer
  - numeric
  - decimal
22. Which of the following is **NOT** a DDL operation that can be performed on a table?
- Create
  - Modify
  - Select
  - Drop
23. Which of the following is optional while creating a table?
- The name of the table
  - The attributes
  - The associated domains
  - Consistency constraints (e.g., NULL or NOT NULL)
24. Which one of the following is a correct statement about primary keys and foreign keys?
- It is not possible to distinguish a foreign key from a primary key
  - Both foreign and primary keys do not allow duplicates
  - A foreign key can be allowed to have NULL values whereas a primary key cannot
  - Both foreign key and primary key uniquely identifies a row in the table
25. Which of the following is **NOT** possible when deleting values from Table A referenced by Table B?
- Cascading the deletion from Table A to Table B for the referenced values
  - Set NULL the referenced values in Table B
  - Insert new value in Table B with default values
  - Restrict the deletion of references from Table A

Consider the following tables and answer the questions from 26 to 30 based on the tables below where the domains are omitted.

**countries**

id	name	population	avg_age
1	Switzerland	8503111	42.23
2	USA	327162500	37.81
3	Turkey	82405728	30.20
4	Italy	59261451	45.92

**located\_in**

id	name	continent
1	Switzerland	Europe
2	USA	North America
4	Italy	Europe

26. Which one of the following SQL statement generates the table schema above with name **countries**? Assume that we do not run any *ALTER* command after constructing the table.
- CREATE TABLE** countries(id VARCHAR(30), name VARCHAR(30), population BIGINT, avg\_age numeric(4,1));
  - CREATE TABLE** countries(id SMALLINT, name VARCHAR(10), population BIGINT, avg\_age numeric(4,1));
  - CREATE TABLE** countries(id SMALLINT, name VARCHAR(30), population BIGINT, avg\_age numeric(4,2));
  - CREATE TABLE** countries(id SMALLINT, name VARCHAR(30), population BIGINT, avg\_age BIGINT);
27. Assume that **id** in table **located\_in** is a primary key and **name** in table **located\_in** has a *UNIQUE* constraint. Based on this assumption, which of the following statements are possible? Select **ALL** correct answers (e.g., if both a and b are correct, then mark both).
- id** in table **countries** is a primary key
  - (id, name)** in table **countries** is a foreign key that references **(id, name)** in table **located\_in**
  - id** in table **countries** is a foreign key that references **id** in table **located\_in**
  - continent** in table **located\_in** is a primary key
28. Assume that **id** in **countries** is a primary key, **avg\_age** is numeric and has a scale value of 2, and **population** is integer. Based on this fact, which of the following codes cause an error? Select **ALL** correct answers (e.g., if both a and b are correct, then mark both).
- INSERT INTO** countries VALUES(5, 'Germany');
  - INSERT INTO** countries VALUES(6, 'Norway', 5258300, 39.2);
  - INSERT INTO** countries VALUES(7, 'China', 37.01, 1386229115);
  - INSERT INTO** countries VALUES(4, 'Italy', 59261451, 45.92);



29. Assume that **id** is a primary key in table **countries**. We would like to delete a record in table **located\_in** whenever a corresponding record in table **countries** is deleted. Which of the following is a constraint for table **countries** that serves our purpose?
- FOREIGN KEY(id) REFERENCES countries (id) ON DELETE REMOVE**
  - FOREIGN KEY(id) REFERENCES countries (id) ON DELETE CASCADE**
  - FOREIGN KEY(id) REFERENCES countries (id) ON DELETE RESTRICT**
  - FOREIGN KEY(id, name) REFERENCES countries (id, name) ON DELETE REMOVE**
30. We would like to change the **name** of each country located in the **continent** of *Europe* in table **located\_in** in order to change their **name** as *EU*. Which of the following code makes these changes?
- UPDATE located\_in SET name = 'Europe' WHERE continent = 'EU';**
  - UPDATE located\_in SET continent = 'Europe' WHERE name = 'EU';**
  - UPDATE located\_in SET name = 'EU' WHERE continent = 'Europe';**
  - UPDATE located\_in SET continent = 'EU' WHERE name = 'Europe';**

#### 4. Relational Algebra

31. Which of the following is **NOT** a kind of relational operation?
- Binary operations
  - Renaming operations
  - Relvar operations
  - Filter operations
32. Which of the following is **never** a constraint for set operations?
- The attributes must be the same
  - The domains must be the same
  - The number of rows must be the same
  - All of the above are constraints

Consider the following relational database where we specify the database name and attributes of a table below. **Answer the questions from 33 to 40 based on this relational database.** Note that date field is in the form “dd/mm/yyyy”. Underscores determines a set of primary key for the table

Attributes<sub>book</sub> = {bid, title, author, category}  
 Attributes<sub>customer</sub> = {cid, cname, gender}  
 Attributes<sub>sales</sub> = {cid, bid, date, bstorename}  
 Attributes<sub>bookstore</sub> = {name, bcity}

33. Which of the following finds the names of customers who bought at least one book since the beginning of this year?

- $\sigma_{date \geq "01/01/2019"}(sales \bowtie customer)$
- $\pi_{cname}(customer \bowtie (\sigma_{date \geq "01/01/2019"}(sales)))$
- $\pi_{cname}(customer) \bowtie (\sigma_{date \geq "01/01/2019"}(sales))$
- $customer \bowtie_{cid=cid} (\sigma_{date \geq "01/01/2019"}(sales))$

34. Which of the following explains correctly what the following relational algebra does in plain English:  $sales \bowtie_{bstorename=name} (\sigma_{bcity="Zurich"}(bookstore))$

- Finds the name of the bookstores in Zurich that have at least one book in their system
- Finds the titles of all books located in any bookstore with the name Zurich
- Finds all books sold in any bookstore in Zurich
- Finds all bookstores located in Zurich

35. Which of the following finds the names of the customers who bought at least one book from “Zurich Store” in Zurich since the beginning of last year

- $\pi_{cname}(book \bowtie (\sigma_{date \geq "01/01/2018"}(sales) \bowtie (customer \bowtie_{bstorename=name} (\sigma_{bcity="Zurich"}(\sigma_{name="Zurich Store"}(bookstore))))))$
- $\pi_{cname}(customer \bowtie (\sigma_{date \geq "01/01/2018"}(sales) \bowtie (sales \bowtie_{bstorename=name} (\sigma_{bcity="Zurich"}(\sigma_{name="Zurich Store"}(bookstore))))))$
- $\pi_{cname}(customer \bowtie (book \bowtie (\sigma_{bcity="Zurich"}(\sigma_{name="Zurich Store"}(bookstore))))))$
- All of the above

Consider the following operations that are labelled with capital letters from A to F. In the questions below from 36 to 40, give the corresponding letter of the correct operations below that exactly does what being asked.

- A.  $\pi_{cid,cname}(customer) - \pi_{cid,cname}(customer \bowtie (sales \bowtie (\sigma_{category="Novel"}(book))))$   
 B.  $\sigma_{bid \geq 5}(\gamma_{category, count(bid)}(sales \bowtie_{bstorename=name} (\sigma_{bcity="Zurich"}(bookstore))))$   
 C.  $\sigma_{bcount \geq 5}(\gamma_{cid, count(bid) \rightarrow bcount}(\sigma_{category="Novel"}(book) \bowtie sales))$   
 D.  $\pi_{cid,cname}(customer) - \pi_{cid,cname}(customer \bowtie sales)$   
 E.  $\sigma_{avg \geq 5}(\gamma_{category, AVERAGE(bid) \rightarrow avg}(sales))$   
 F.  $\gamma_{MAX(bcount) \rightarrow maxb}(\gamma_{cid, count(bid) \rightarrow bcount}(\sigma_{date \geq "01/01/2019"}(\sigma_{date \leq "08/02/2019"}(sales))))$

36. Which of the operations above finds the id and name of the customers who never bought a book in Novel category?  
 Your answer is: \_\_\_\_\_ (Please only give a single letter such as 'A')
37. Which of the operations above finds the id of the customers who bought the maximum number of books since the beginning of the year?  
 Your answer is: \_\_\_\_\_ (Please only give a single letter such as 'A')
38. Which of the operations above finds the id and name of the customers who did not buy any book?  
 Your answer is: \_\_\_\_\_ (Please only give a single letter such as 'A')
39. Which of the operations above finds the number of books in each category that are sold at least 5 times and carried by the bookstores in Zurich?  
 Your answer is: \_\_\_\_\_ (Please only give a single letter such as 'A')
40. Which of the operations above finds the id of the customers who bought at least 5 books in "Novel" category?  
 Your answer is: \_\_\_\_\_ (Please only give a single letter such as 'A')

## 5. Queries with SQL

### ➤ Part I

Considering the following schema, and their corresponding data, answer questions **41 to 51**.

```
CREATE TABLE Course(  
    id INT NOT NULL,  
    title VARCHAR(60) NOT NULL,  
    number VARCHAR(30) NOT NULL,  
    PRIMARY KEY (id));  
  
CREATE TABLE Department(  
    id INT NOT NULL,  
    name VARCHAR(60) NOT NULL,  
    PRIMARY KEY (id));  
  
CREATE TABLE Catalogue(  
    course_id INT NOT NULL,  
    department_id INT NOT NULL,  
    semester VARCHAR(30) NOT NULL,  
    lecturer VARCHAR(60) NOT NULL,  
    std_count INT,  
    PRIMARY KEY (course_id, department_id, semester));
```

#### Course

ID	TITLE	NUMBER
1	'Information Systems for Engineers'	'252-0834-00L'
2	'Differential Geometry I'	'401-3531-00L'
3	'Communication in Mathematics'	'401-0000-00L'
4	'Computer Architecture'	'401-0000-00L'
5	'Digital Circuits'	'227-0003-00L'
6	'Information Retrieval'	'252-0341-01L'

#### Department

ID	NAME
1	'Computer Science'
2	'Information Technology and Electrical Engineering'
3	'Mathematics'

## Catalogue

COURSE_ID	DEPARTMENT_ID	SEMESTER	LECTURER	STD_COUNT
1	1	'2018 Autumn'	'Ghislain Fourny'	157
2	3	'2018 Autumn'	'Will Merry'	95
3	3	'2018 Autumn'	'Will Merry'	44
4	1	'2017 Autumn'	'Onur Mutlu'	35
4	1	'2018 Autumn'	'Onur Mutlu'	22
5	2	'2018 Autumn'	'Mathieu Luisier'	40

41. How many rows does the following query return?

```
SELECT department_id FROM Catalogue WHERE semester = '2018 Autumn'  
UNION  
SELECT department_id FROM Catalogue WHERE semester = '2018'
```

- a. 3
- b. 4
- c. 5
- d. 6

42. How many rows does the following query return?

```
SELECT lecturer FROM Catalogue WHERE semester = '2018 Autumn'  
INTERSECT  
SELECT lecturer FROM Catalogue WHERE semester = '2018 Spring' OR department_id = 1
```

- a. 0
- b. 1
- c. 2
- d. 3

43. How many rows does the following query return?

```
SELECT * FROM Catalogue  
RIGHT OUTER JOIN Course ON Catalogue.course_id = Course.id
```

- a. 4
- b. 5
- c. 6
- d. 7

44. How many rows does the following query return?

```
SELECT lecturer FROM Catalogue  
LEFT OUTER JOIN Department ON Catalogue.department_id = Department.id
```

- a. 3
- b. 4
- c. 5
- d. 6

*This page left blank intentionally*

45. How many rows does the following query return?

```
SELECT lecturer FROM Department
LEFT OUTER JOIN Catalogue ON Catalogue.department_id = Department.id
UNION
SELECT lecturer FROM Department
RIGHT OUTER JOIN Catalogue ON Catalogue.department_id = Department.id
```

- a. 3
- b. 4
- c. 6
- d. 12

46. How many rows does the following query return?

```
SELECT AVG(Catalogue.std_count) FROM Catalogue
RIGHT OUTER JOIN Course ON Catalogue.course_id = Course.id
WHERE Course.id < 3
```

- a. 1
- b. 3
- c. 5
- d. 7

47. What are the first two values returned by the following query? First value is: \_\_\_\_\_ and the second value is: \_\_\_\_\_

```
SELECT nmb FROM (
  SELECT (Catalogue.std_count / 10.0) as nmb FROM Catalogue
  INNER JOIN Course ON Catalogue.course_id = Course.id ORDER BY nmb ASC
) as sum_tble WHERE nmb > 2
```

48. What is the average value returned by the following query? Your answer is: \_\_\_\_\_ (Please only write down a single value)

```
SELECT MIN(std_count) FROM (
  (SELECT * FROM Department WHERE name = 'Computer Science') as cs_tbl
  LEFT OUTER JOIN
  (SELECT * FROM Catalogue
  WHERE std_count > 50 OR semester = '2017 Autumn') as ctlg_tbl
  ON cs_tbl.id = ctlg_tbl.department_id
) as join_tbl
```

49. What is the value on the last row returned by the following query? Your answer is: \_\_\_\_\_ (Please only write down a single value)

```
SELECT MIN(std_count) FROM (
  SELECT course_id, MIN(std_count) as std_count FROM Catalogue
  WHERE semester = '2018 Autumn'
  GROUP BY course_id
) as A
```

50. How many rows does the following query return?

```
SELECT * FROM Course CROSS JOIN Catalogue
WHERE Catalogue.course_id = Course.id
```

- a. 6
- b. 12
- c. 14
- d. 18

51. How many rows does the following query return?

```
SELECT DISTINCT * FROM Course  
CROSS JOIN Catalogue
```

- a. 7
- b. 12
- c. 18
- d. 36



## ➤ Part II

Considering the following schema and its data, answer questions 52 to 60.

```

CREATE TABLE Author(
    SSN INT NOT NULL,
    name VARCHAR(60) NOT NULL,
    bdate DATE,
    country VARCHAR(60),
    PRIMARY KEY (SSN));

CREATE TABLE Publisher(
    pub_name VARCHAR(60),
    pub_city VARCHAR(60),
    PRIMARY KEY(pub_name));

CREATE TABLE Book(
    ISBN INT NOT NULL,
    title VARCHAR(60) NOT NULL,
    sold_copies BIGINT NOT NULL,
    price INT NOT NULL,
    publisher_name VARCHAR(60),
    PRIMARY KEY (ISBN),
    FOREIGN KEY(publisher_name) REFERENCES Publisher(pub_name));

CREATE TABLE Authorship(
    SSN INT NOT NULL,
    ISBN INT NOT NULL,
    year INT NOT NULL,
    income BIGINT NOT NULL,
    PRIMARY KEY (SSN, ISBN),
    FOREIGN KEY(SSN) REFERENCES Author(SSN),
    FOREIGN KEY(ISBN) REFERENCES Book (ISBN));
    
```

### Author

SSN	NAME	BDATE	COUNTRY
1	'Isaac Asimov'	'1920-01-02'	'Russia'
2	'Ray Bradbury'	'1920-08-22'	'USA'
3	'Ursula Guin'	'1929-10-21'	'USA'
4	'George Orwell'	'1903-06-23'	'UK'

### Publisher

NAME	COUNTRY
'Secker and Warburg'	'London'
'Gnome Press'	'New York City'
'Doubleday'	'New York City'
'Ballantine Books'	'New York City'
'Parnassus Press'	'Berkeley'

## Book

ISBN	title	sold_copies	price	publisher_name
1	'Animal Farm',	30000000	1	'Secker and Warburg'
2	'I, Robot'	25000000	7	'Gnome Press'
3	'Foundation'	35000000	5	'Gnome Press'
4	'The Gods Themselves'	15000000	5	'Doubleday'
5	'Fahrenheit 451'	12000000	3	'Ballantine Books'
6	'The Martian Chronicles'	20000000	5	'Doubleday'
7	'A Wizard of Earthsea'	25000000	7	'Parnassus Press'

## Authorship

SSN	ISBN	year	income
4	1	1945	50000000
1	2	1950	100000000
1	3	1951	120000000
1	4	1972	50000000
2	5	1953	30000000
2	6	1950	40000000
3	7	1968	85000000

Now consider the following query that includes missing parts in it with the numbers labelled with these missing parts. We would like to query the name of oldest author(s) who has obtained a total income greater than 100000000 between 1950 and 1970. **Answer the questions from 52 to 55 based on the query below.**

```
SELECT name FROM Author WHERE bdate = (  
  SELECT [1] FROM Author  
  WHERE SSN IN (  
    SELECT SSN FROM Authorship  
    WHERE year BETWEEN 1950 AND 1970  
    GROUP BY [2]  
    HAVING [3] > 100000000  
  )  
)
```

52. Which of the following fills the gap [1] in the query above so that it returns results corresponding to the problem description?
- MAX (bdate)
  - MIN (bdate)
  - MAX (year)
  - MIN (year)
53. Which of the following fills the gap [2] in the query above so that it returns results corresponding to the problem description?
- bdate
  - year
  - SSN
  - Income
54. Which of the following fills the gap [3] in the query above so that it returns results corresponding to the problem description?
- SUM(income)
  - income
  - MAX(income)
  - AVG(income)
55. Based on the description of the problem above, what is the author returned by the corresponding query?
- Isaac Asimov
  - Ray Bradbury
  - Ursula Guin
  - George Orwell
56. Consider the query below. What are the SSN values returned by this query?

```

SELECT DISTINCT A.SSN, A.name FROM Author A, Authorship AUS
WHERE AUS.SSN = A.SSN AND income > ALL (
    SELECT income FROM Author A, Authorship AUS
    WHERE AUS.SSN = A.SSN AND A.name = 'George Orwell')

```

- 1
  - 1 and 2
  - 1 and 3
  - 1, 2, and 3
57. Consider the same query in Question 56. How many rows does this query return? (Hint: Note that we use **DISTINCT** with **SELECT**)
- 1
  - 2
  - 3
  - 4

58. Consider the query below. What are the names of the authors returned by this query? Select **ALL** correct answers (e.g., if both a and b are correct, then mark both).

```
SELECT DISTINCT A.SSN, A.name
FROM Author A, Authorship AUS, Book B
WHERE AUS.SSN = A.SSN AND AUS.ISBN = B.ISBN AND B.publisher_name =
'Doubleday'
```

- a. Isaac Asimov
- b. Ray Bradbury
- c. Ursula Guin
- d. George Orwell

59. How many rows does the following query return?

```
SELECT DISTINCT AUS.ISBN FROM Authorship AUS, (
  SELECT AUS.ISBN FROM Authorship AUS, Author A
  WHERE AUS.SSN = A.SSN EXCEPT
  SELECT B.ISBN FROM Book B WHERE B.publisher_name = 'Doubleday') AS A
WHERE A.ISBN = AUS.ISBN
```

- a. 2
- b. 4
- c. 5
- d. 7

60. Consider the query below. How many rows does the following query return? Your answer is: \_\_\_\_\_ (Please only write an integer value)

```
SELECT pub_name FROM Publisher
WHERE pub_name NOT IN
(SELECT publisher_name FROM Book WHERE publisher_name IS NOT NULL)
```

## 6. Database Design Theory

61. Which of the following is *always* the correct definition of the first normal form?

- a. A primary key must be a set of a single attribute of a table
- b. All the attributes must be atomic (indivisible)
- c. There must be at least two tables in a relational database design
- d. The tuples must be ordered with respect to an attribute in a table

62. Which of the following is *always* the correct definition of the Boyce-Codd normal form?

- a. Superkeys may not appear on the left-hand-side of a non-trivial functional dependency.
- b. Superkeys may not appear on the right-hand-side of a non-trivial functional dependency.
- c. Only superkeys may appear on the right-hand-side of a non-trivial functional dependency.
- d. Only superkeys may appear on the left-hand-side of a non-trivial functional dependency.

63. Which of the following is *always* a correct statement given a specific universe?

- a. If we remove an attribute from the set of attributes of a superkey, the resulting set is still a superkey
- b. If we remove an attribute from the set of attributes of a candidate key, the resulting set is still a candidate key
- c. A candidate key is a superkey
- d. A superkey is a candidate key

Consider the following table and **answer Questions 64, 65, 66, and 67 according to the table below**. We assume that the considered universe only contains this table (no undesired superkeys).

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
a	b	c	d
e	b	f	h
h	i	c	d
l	m	n	o
o	b	c	d

64. How many superkeys does the table above have?

- a. 1
- b. 4
- c. 8
- d. 9

65. Which of the following is a non-trivial functional dependency for the table above?

- a.  $B \rightarrow C, D$
- b.  $B \rightarrow A$
- c.  $C \rightarrow D$
- d.  $D \rightarrow A$

66. Which of the following functional dependencies **must** be a part of a minimal basis?

- a.  $A \rightarrow B$
- b.  $A, B \rightarrow C, D$
- c.  $D \rightarrow A$
- d.  $B, C, D \rightarrow A$

67. How many attributes does the candidate key include?

- a. 1
- b. 2
- c. 3
- d. 4

Consider each of the following sets of functional dependencies for relation schema  $R = (A, B, C, D, E)$ . We denote the closure of an attribute  $A$  as  $\{A\}^+ = S$ , where  $S$  is the set of attributes captured under the closure. **Answer the Questions 68, 69, 70, and 71 based on the functional dependencies below.**

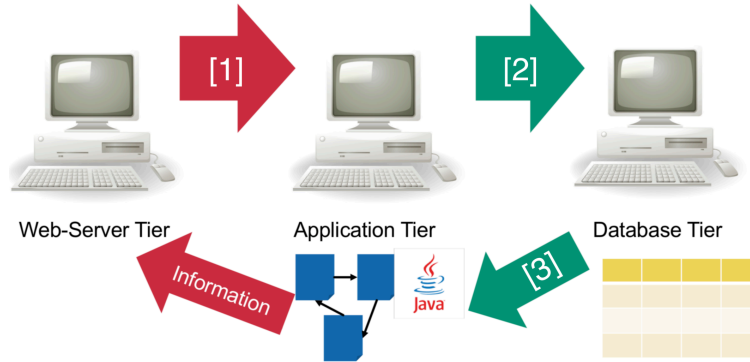
$A \rightarrow B, C$   
 $C, D \rightarrow E$   
 $B \rightarrow D$   
 $E \rightarrow A$

68. Which of the following sets is  $\{A\}^+$ ?
- $\{B, C\}$
  - $\{A, B, C\}$
  - $\{A, B, C, D\}$
  - $\{A, B, C, D, E\}$
69. Which of the following sets is  $\{B, D\}^+$ ?
- $\{B, D\}$
  - $\{B, D, E\}$
  - $\{A, B, D, E\}$
  - $\{A, B, C, D, E\}$
70. How many candidate keys does  $R$  have?
- 0
  - 1
  - 3
  - 4
71. Assume that we change the functional dependency " $C, D \rightarrow E$ " to " $C, D \rightarrow B$ ". How many candidate keys does  $R$  have now?
- 1
  - 2
  - 3
  - 4
72. Which statement is not part of the impossibility triangle regarding the decomposition of a table into the Boyce-Codd normal form?
- No table has a superkey
  - All tables are in Boyce-Codd normal form
  - The decomposition is lossless
  - The decomposition preserves functional dependencies
73. Consider the relational schema  $R(A, B, C, D, E, F)$ , and the functional dependencies  $(A, B \rightarrow D)$   $(A, D \rightarrow E)$ ,  $(E, F \rightarrow C)$ ,  $(A, B \rightarrow F)$ . Which of the following is a correct statement about the decomposition of  $R$  into  $R_1(A, E, F)$  and  $R_2(B, C, D, E, F)$ ?
- It is a lossless join, but the dependencies are not preserved
  - It is a lossless join and the dependencies are preserved
  - It is not a lossless join and the dependencies are not preserved
  - It is not a lossless join but the dependencies are preserved

74. Consider the relational schema  $R(X, Y, Z)$  and the functional dependencies  $(X \rightarrow Y)$  and  $(Y \rightarrow Z)$ . Which of the following is a correct statement about the decomposition of  $R$  into  $R_1(X, Y)$  and  $R_2(X, Z)$ ?
- It is a lossless join, but the dependencies are not preserved
  - It is a lossless join and the dependencies are preserved
  - It is not a lossless join and the dependencies are not preserved
  - It is not a lossless join but the dependencies are preserved
75. Consider the relational schema  $R(X, Y, Z)$  and the functional dependencies  $(X \rightarrow Y)$  and  $(Y \rightarrow Z)$ . Which of the following is a correct statement about the decomposition of  $R$  into  $R_1(X, Y)$  and  $R_2(Y, Z)$ ?
- It is a lossless join, but the dependencies are not preserved
  - It is a lossless join and the dependencies are preserved
  - It is not a lossless join and the dependencies are not preserved
  - It is not a lossless join but the dependencies are preserved
76. Given a relation  $R(A, B, C, D, E)$  with the set of functional dependencies  $(A \rightarrow D)$ ,  $(B \rightarrow C)$ ,  $(C, D \rightarrow E)$ ,  $(A \rightarrow B)$ ,  $(A \rightarrow C)$ , and  $(E \rightarrow B)$ , how many dependencies violate BCNF?
- 1
  - 2
  - 3
  - 4
77. Consider the same relation  $R$  and the functional dependencies given in the previous question (Question 76). Using the functional dependency  $B \rightarrow C$ , we would like to decompose  $R$  so that the decomposition will be in BCNF. Which of the following is a correct decomposition of  $R$  into BCNF relations?
- $R_1(B, C)$  and  $R_2(A, D, E)$
  - $R_1(B, C)$  and  $R_2(A, B, D, E)$
  - $R_1(B, C)$ ,  $R_2(B, E)$ , and  $R_3(A, D, E)$
  - $R_1(B, C)$ ,  $R_2(B, D)$ , and  $R_3(A, B, E)$

## 7. The Ecosystem around SQL

The following figure shows the flow of information in a 3-tier stack that can be used in a web-server with a database in the backend. However, some parts of the figure are missing as indicated by numbers [1], [2], and [3]. **Answer the questions 78, 79, and 80 according to the figure below.**

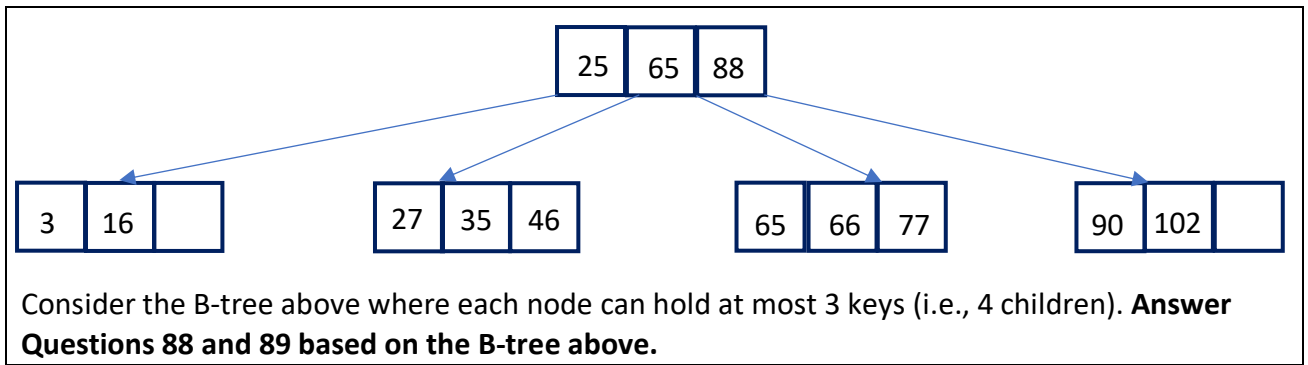


78. Which one of the following is the correct description of the interaction happening between the Web-Server Tier and the Application Tier (see [1])?
- Request needed data with SQL
  - Request needed information
  - Send over the data output by the query
  - Resolve functional dependencies
79. Which one of the following the correct description of the interaction happening between the Application Tier and the Database Tier (see [2])?
- Request needed data with SQL
  - Request needed information
  - Send over the data output by the query
  - Resolve functional dependencies
80. Which one of the following is the correct description of the interaction happening between the Database Tier and the Application Tier (see [3])?
- Request needed data with SQL
  - Request needed information
  - Send over the data output by the query
  - Resolve functional dependencies
81. Which of the following is a correct statement about the following schedule?  
 $w_1(B); r_1(A); r_2(B); r_1(B); w_1(A); w_2(B); r_2(A); w_2(A);$
- The schedule is already serial
  - The schedule is not serializable because there is a circular dependency  $T_1 \rightarrow T_2$  and  $T_2 \rightarrow T_1$
  - The schedule is serializable because there is a non-circular dependency  $T_1 \rightarrow T_2$
  - The schedule is serializable because there is a non-circular dependency  $T_2 \rightarrow T_1$
82. Which of the following is a correct statement about the following schedule?  
 $r_2(A); r_1(B); w_2(B); r_1(C); w_3(D); r_1(A); w_1(A); w_3(A);$
- The schedule is not serializable because there is a circular dependency  $T_3 \rightarrow T_1 \rightarrow T_3$
  - The schedule is not serializable because there is a circular dependency  $T_1 \rightarrow T_2 \rightarrow T_1$
  - The schedule is serializable because there is a non-circular dependency  $T_1 \rightarrow T_2$
  - The schedule is serializable because there is a non-circular dependency  $T_2 \rightarrow T_3$



## 8. Views and Indices

83. Which of the following is a correct statement?
- Lookup in a hash index is linear. Lookup in a B-tree index is logarithmic.
  - Lookup in a hash index is constant. Lookup in a B-tree index is logarithmic.
  - Lookup in a hash index is constant. Lookup in a B-tree index is linear.
  - Lookup in a hash index is logarithmic. Lookup in a B-tree index is linear.
84. A hash index was built on the field "date". Which query will be faster thanks to this index?
- SELECT \* FROM events WHERE date = '2019-01-31'**
  - SELECT \* FROM events WHERE date > '2019-01-31'**
  - SELECT name, date FROM events**
  - SELECT SUM(attendance) FROM events GROUP BY date**
85. Can a view be updated?
- Yes, always
  - No, never
  - Yes, under certain conditions.
  - Updating a view is not semantically meaningful.
86. Indices accelerate query execution. Bob designed a database containing all his sales for his shop and decided to index every single field in all his tables in order to speed up his database. Is this a sensible idea?
- Yes, indices are cheap and should be used everywhere. All SQL vendors actually index most fields automatically.
  - No, because transactions that update the database will be very slow because of the updates made to the indices.
  - Yes, it will make all query and update requests to the database significantly faster, but Bob must buy enough storage space to store the indices.
  - No, because while the insertion of new sales will be accelerated by the index, point queries will become slower due to the index overhead.
87. Can a query that selects based on a field value be executed without an index?
- No, an index is needed to be able to select on a field's value.
  - No, which is why all SQL vendors create indices on the fly when a selection query is sent to the database.
  - Yes, but then a full scan of the table is needed, which is slower than an index lookup.
  - Yes, but in this case the table must be transposed in order to map the selection query as a projection for efficiency.



88. Which of the following statement is correct about the height of the tree (i.e., total number of edges from top to leaf nodes and a tree with only root node has height 0) when an entry with key value 67 is inserted?
- Height of the tree does not change and no further node is created (i.e., height stays as 1)
  - Height of the tree does not change and a new node is created at the top level (i.e., height stays as 1)
  - Height of the tree changes from 1 to 2
  - Height of the tree changes from 1 to 3
89. Which of the following is a correct statement about the nodes in the top level when an entry with key value 67 is inserted?
- There are two nodes in the top level
  - There is a single node in the top level and it only includes the value of 67
  - There is a single node in the top level and it includes the values of 25, 65, 88
  - There is a single node in the top level and it includes the values of 25, 65

## 9. Cubes and OLAP

90. Which of the following is **NOT** a correct comparison between OLTP and OLAP
- OLTP consists of detailed individual records whereas OLAP includes historical summarized consolidated data
  - OLTP usually deals with small sets of records whereas OLAP is usually used for an analysis over big chunks
  - OLTP is usually slower (slow interactive) whereas OLAP is usually fully interactive (fast)
  - There are usually more write requests than reads requests for OLTP whereas there are usually more reads than write requests for OLAP.

Consider the following definitions for OLAP cube operations and **answer the Questions 91, 92, 93, and 94 using these definitions.**

**Definition A:** Displays data at a higher level of detail by replacing an aggregated value with the actual values that it aggregates, sometimes still retaining and displaying the aggregated value as a total. This is typically done by adding more dimensions into the grid mentioned in definition B. The view thus goes from more summarized to more detailed and the grid gets bigger.

**Definition B:** Takes a subset of cube with fewer dimensions, by constraining some dimensions to a single value.

**Definition C:** Cross-tabulates and organizes the cells of a cube onto a user-friendly grid, which is usually two-dimensional. The affected dimensions, also often two of them, are assigned to other rows or columns.

**Definition D:** Displays data at a more abstract level of detail by replacing entire lists of values, corresponding to the members of a dimension, with aggregated values, for example a sum. This is typically done by remove dimensions from the grid mentioned in definition B. The view thus goes from more detailed to more summarized and the grid gets smaller.

91. Which of the definitions above is the definition for Slicing?
- Definition A
  - Definition B
  - Definition C
  - Definition D
92. Which of the definitions above is the definition for Dicing?
- Definition A
  - Definition B
  - Definition C
  - Definition D
93. Which of the definitions above is the definition for Drilling down?
- Definition A
  - Definition B
  - Definition C
  - Definition D
94. Which of the definitions above is the definition for Roll up?
- Definition A
  - Definition B
  - Definition C
  - Definition D
95. Which of the following clause or clauses are used by Dicing operation?
- Only **SELECT**
  - Only **GROUP BY**
  - Both **SELECT** and **GROUP BY**
  - Neither **SELECT** nor **GROUP BY**

## 10. Database Architecture

96. Which of the following is the correct order of the memory devices from faster to slower (from left to right)
- CPU Cache -> Disk -> Memory
  - Memory -> Disk -> Cache
  - CPU Cache -> Memory -> Disk
  - Memory -> Cache -> Disk
97. Which of the following memory types are considered as volatile memory types (i.e., data is lost when the electricity is cut)? Select **ALL** correct answers (e.g., if both a and b are correct, then mark both).
- CPU Cache
  - RAM
  - Disk
  - DVDs
98. Which of the following memory device is the most expensive one per storage unit?
- CPU Cache
  - RAM
  - Disk
  - DVDs
99. Which of the following are the steps in disk optimizations? Select **ALL** correct answers (e.g., if both a and b are correct, then mark both).
- Prefetching blocks into memory before they are used
  - Spreading data over smaller multiple disks
  - Placing the blocks accessed together in the same cylinder
  - Removing the data from the disk once it is accessed
100. What is data independence?
- The absence of non-trivial functional dependencies in a table
  - The clean separation of the logical layer from the physical layer
  - The ability to import and export data from and to any SQL database
  - This is a short name given to the day of January 30, 1970, as a tribute to Edgar Codd's seeding publication.