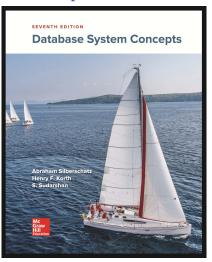
Database System Concepts, 7th Edition Chapter 7: Relational Database Design

Silberschatz, Korth and Sudarshan

April 20, 2025

Database System Concepts



Content has been extracted from Database System Concepts, Seventh Edition, by Silberschatz, Korth and Sudarshan. Mc Graw Hill Education. 2019.

Visit https://db-book.com/.

Plan

Features of Good Relational Design

Functional Dependencies

Decomposition Using Functional Dependencies

Normal Forms

Functional Dependency Theory

Algorithms for Decomposition using Functional Dependencies

Decomposition Using Multivalued Dependencies

More Normal Form

Atomic Domains and First Normal Form

Database-Design Process

Modeling Temporal Data

Features of Good Relational Designs

```
classroom(building, room_number, capacity)
department(dept_name, building, budget)
course(course_id, title, dept_name, credits)
instructor(ID, name, dept_name, salary)
section(course_id, sec_id, semester, year, building, room_number, time_slot_id)
teaches(ID, course_id, sec_id, semester, year)
student(ID, name, dept_name, tot_cred)
takes(ID, course_id, sec_id, semester, year, grade)
advisor(s_ID, i_ID)
time_slot(time_slot_id, day, start_time, end_time)
prereq(course_id, prereq_id)
```

Figure 7.1 Database schema for the university example.

Features of Good Relational Designs

▶ Suppose we combine *instructor* and *department* into *in_dep*, which represents the natural join on the relations *instructor* and *department*.

ID	name	salary	dept_name	building	budget
22222	Einstein	95000	Physics	Watson	70000
12121	Wu	90000	Finance	Painter	120000
32343	El Said	60000	History	Painter	50000
45565	Katz	75000	Comp. Sci.	Taylor	100000
98345	Kim	80000	Elec. Eng.	Taylor	85000
76766	Crick	72000	Biology	Watson	90000
10101	Srinivasan	65000	Comp. Sci.	Taylor	100000
58583	Califieri	62000	History	Painter	50000
83821	Brandt	92000	Comp. Sci.	Taylor	100000
15151	Mozart	40000	Music	Packard	80000
33456	Gold	87000	Physics	Watson	70000
76543	Singh	80000	Finance	Painter	120000

Figure 7.2 The in_dep relation.

- ▶ There is repetition of information
- ▶ Need to use null values (if we add a new department with no instructors)

Decomposition

- ▶ The only way to avoid the repetition-of-information problem in the *in_dep* schema is to decompose it into two schemas −*instructor* and *department* schemas.
- ► Not all decompositions are good. Suppose we decompose: employee(ID, name, street, city, salary) into

```
employee1 (ID, name)
employee2 (name, street, city, salary)
The problem arises when we have two employees with the
same name.
```

▶ The next slide shows how we lose information –we cannot reconstruct the original employee relation– and so, this is a lossy decomposition.

A Lossy Decomposition

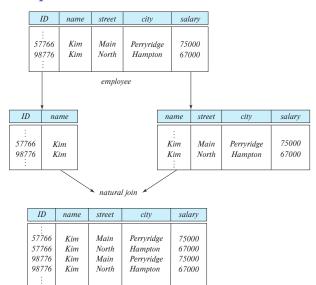


Figure 7.3 Loss of information via a bad decomposition.

Lossless Decomposition

- ▶ Let R be a relation schema and let R_1 and R_2 form a decomposition of R. That is $R = R_1 \bigcup R_2$.
- We say that the decomposition is a **lossless** decomposition if there is no loss of information by replacing R with the two relation schemas $R_1 \cup R_2$.
- ► Formally,

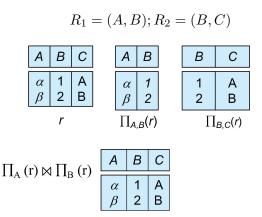
$$\Pi_{R_1}(r)\bowtie\Pi_{R_2}(r)=r$$

▶ And, conversely a decomposition is lossy if

$$r \subset \Pi_{R_1}(r) \bowtie \Pi_{R_2}(r) = r$$

Example of Lossless Decomposition

▶ Decomposition of R = (A, B, C):



Normalization Theory

- ightharpoonup Decide whether a particular relation R is in "good" form.
- ▶ In the case that a relation R is not in "good" form, decompose it into set of relations $\{R_1, R_2, \dots, R_n\}$ such that,
 - ► Each relation is in good form.
 - ▶ The decomposition is a lossless decomposition.
- ▶ Our theory is based on:
 - 1. functional dependencies.
 - 2. multivalued dependencies.

Plan

Features of Good Relational Design

Functional Dependencies

Decomposition Using Functional Dependencies

Normal Forms

Functional Dependency Theory

Algorithms for Decomposition using Functional Dependencies

Decomposition Using Multivalued Dependencies

More Normal Form

Atomic Domains and First Normal Form

Database-Design Process

Modeling Temporal Data

Functional Dependencies

- ► There are usually a variety of constraints (rules) on the data in the real world.
- ► For example, some of the constraints that are expected to hold in a university database are:
 - ▶ Students and instructors are uniquely identified by their ID.
 - ► Each student and instructor has only one name.
 - Each instructor and student is (primarily) associated with only one department.
 - Each department has only one value for its budget, and only one associated building.

Functional Dependencies (Cont.)

- ▶ An instance of a relation that satisfies all such real-world constraints is called a **legal instance** of the relation.
- ► A legal instance of a database is one where all the relation instances are legal instances
- ► Constraints on the set of legal relations.
- ▶ Require that the value for a certain set of attributes determines uniquely the value for another set of attributes.
- ▶ A functional dependency is a generalization of the notion of a key.

Functional Dependencies Definition

 \triangleright Let R be a relation schema

$$\alpha \subseteq R$$
 and $\beta \subseteq R$

► The functional dependency

$$\alpha \to \beta$$

holds on R if and only if for any legal relations r(R), whenever any two tuples t_1 and t_2 of r agree on the attributes α , they also agree on the attributes β . That is,

$$t_1[\alpha] = t_2[\alpha] \Rightarrow t_1[\beta] = t_2[\beta]$$

 \triangleright Example: Consider r(A, B) with the following instance of r:

1	4	
1	5	
3	7	

ightharpoonup On this instance, $B \to A$ hold; $A \to B$ does **NOT** hold,

Closure of a Set of Functional Dependencies

- ▶ Given a set F set of functional dependencies, there are certain other functional dependencies that are logically implied by F.
 - ▶ If $A \to B$ and $B \to C$, then we can infer that $A \to C$.
- ► The set of all functional dependencies logically implied by *F* is the **closure** of *F*.
- \blacktriangleright We denote the closure of F by F^+ .

Plan

Features of Good Relational Design

Functional Dependencies

Decomposition Using Functional Dependencies

Normal Forms

Functional Dependency Theory

Algorithms for Decomposition using Functional Dependencies

Decomposition Using Multivalued Dependencies

More Normal Form

Atomic Domains and First Normal Form

Database-Design Process

Modeling Temporal Data

Keys and Functional Dependencies

- ▶ K is a superkey for relation schema R if and only if $K \to R$.
- \blacktriangleright K is a candidate key for R if and only if:
 - ightharpoonup K o R, and
 - ightharpoonup for no $\alpha \subset K$, $\alpha \to R$
- ▶ Functional dependencies allow us to express constraints that cannot be expressed using superkeys. Consider the schema:

in_dep(ID, name, salary, dept_name, building, budget).
We expect these functional dependencies to hold:

$$dept_name \rightarrow building$$

$$ID \rightarrow building$$

but would not expect the following to hold:

 $dept_name \rightarrow salary$

Use of Functional Dependencies

- ▶ We use functional dependencies to:
 - ➤ To test relations to see if they are legal under a given set of functional dependencies.
 - If a relation r is legal under a set F of functional dependencies, we say that r satisfies F.
 - ▶ To specify constraints on the set of legal relations.
 - We say that F holds on R if all legal relations on R satisfy the set of functional dependencies F.
- ▶ Note: A specific instance of a relation schema may satisfy a functional dependency even if the functional dependency does not hold on all legal instances.
 - ► For example, a specific instance of *instructor* may, by chance, satisfy

 $name \rightarrow ID$.

Trivial Functional Dependencies

- ▶ A functional dependency is **trivial** if it is satisfied by all instances of a relation.
 - **E**xample:
 - ightharpoonup ID, name
 ightarrow ID
 - ightharpoonup name
 ightharpoonup name
 ightharpoonup name
 - ▶ In general, $\alpha \to \beta$ is trivial if $\beta \subseteq \alpha$.

Lossless Decomposition

- ▶ We can use functional dependencies to show when certain decomposition are lossless.
- ▶ For the case of $R = (R_1, R_2)$, we require that for all possible relations r on schema R

$$r = \Pi_{R_1}(r) \bowtie \Pi R_2(r)$$

▶ A decomposition of R into R_1 and R_2 is lossless decomposition if at least one of the following dependencies is in F^+ :

$$R_1 \bigcap R_2 \to R_1$$
$$R_1 \bigcap R_2 \to R_2$$

▶ The above functional dependencies are a sufficient condition for lossless join decomposition; the dependencies are a necessary condition only if all constraints are functional dependencies.

Example

- R = (A, B, C) $F = (A \to B, B \to C)$
- ► $R_1 = (A, B)$; $R_2 = (B, C)$ Lossless decomposition: $R_1 \cap R_2 = \{B\}$ and $B \to BC$
- ► $R_1 = (A, B)$; $R_2 = (A, C)$ Lossless decomposition: $R_1 \cap R_2 = \{A\}$ and $A \to AC$
- ▶ Note: $B \to BC$ is a shorthand notation for $B \to \{B, C\}$.

Dependency Preservation

- ► Testing functional dependency constraints each time the database is updated can be costly,
- ▶ It is useful to design the database in a way that constraints can be tested efficiently.
- ▶ If testing a functional dependency can be done by considering just one relation, then the cost of testing this constraint is low.
- ▶ When decomposing a relation it is possible that it is no longer possible to do the testing without having to perform a Cartesian Produced.
- ► A decomposition that makes it computationally hard to enforce functional dependency is said to be NOT dependency preserving.

Dependency Preservation Example

► Consider a schema:

dept_advisor(s_ID, i_ID, department_name)

▶ With function dependencies:

$$i_ID \rightarrow dept_name \\ s_ID, dept_name \rightarrow i_ID$$

- ▶ In the above design we are forced to repeat the department name once for each time an instructor participates in a dept_advisor relationship.
- ► To fix this, we need to decompose dept_advisor.
- ▶ Any decomposition will not include all the attributes in:

$$s_ID, dept_name \rightarrow i_ID$$

▶ Thus, the composition NOT be dependency preserving.

Plan

Features of Good Relational Design

Functional Dependencies

Decomposition Using Functional Dependencies

Normal Forms

Functional Dependency Theory

Algorithms for Decomposition using Functional Dependencies

Decomposition Using Multivalued Dependencies

More Normal Form

Atomic Domains and First Normal Form

Database-Design Process

Modeling Temporal Data

Boyce-Codd Normal Form

▶ A relation schema R is in BCNF with respect to a set F of functional dependencies if for all functional dependencies in F^+ of the form:

$$\alpha \to \beta$$

where $\alpha \subseteq R$ and $\beta \subseteq R$, at least one of the following holds:

- $ightharpoonup \alpha \to \beta$ is trivial (i.e., $\beta \subseteq \alpha$)
- $ightharpoonup \alpha$ is a superkey for R.

Boyce-Codd Normal Form (Cont.)

▶ Example schema that is not in BCNF:

```
in_dep (ID, name, salary, dept_name, building, budget)
because:
```

- $ightharpoonup dept_name o building, budget$
 - ▶ holds on in_dep
 - ▶ but...
- dept_name is not a superkey.
- ► When decompose in_dept into instructor and department:
 - ▶ instructor is in BCNF.
 - department is in BCNF.

Decomposing a Schema into BCNF

- ▶ Let R be a schema that is not in BCNF. Let $\alpha \to \beta$ be the Fuctional Dependency that causes a violation of BCNF.
- \blacktriangleright We decompose R into:
 - $\triangleright (\alpha \bigcup \beta)$
 - $ightharpoonup (R (\beta \alpha))$
- ► In our example of in_dep,
 - $ightharpoonup \alpha = dept_name$
 - $\triangleright \beta = building, budget$

and in_dep is replaced by

- $(\alpha \bigcup \beta) = (dept_name, building, budget)$
- $(R (\beta \alpha)) = (ID, name, salary, dept_name)$

BCNF and Dependency Preservation

- ▶ It is not always possible to achieve both BCNF and dependency preservation.
- ► Consider a schema:

▶ With function dependencies:

$$i_ID \rightarrow \quad dept_name \\ s_ID, dept_name \rightarrow \quad i_ID$$

- dept_advisor is not in BCNF:
 - $ightharpoonup i_ID$ is not a superkey.
- ► Any decomposition of dept_advisor will not include all the attributes in:

$$s_ID, dept_name \rightarrow i_ID$$

▶ Thus, the composition is NOT be dependency preserving.

Third Normal Form

► A relation schema R is in **third normal form (3NF)** if for all:

$$\alpha \to \beta$$
 in F^+

at least one of the following holds:

- $ightharpoonup \alpha \to \beta$ is trivial (i.e., $\beta \in \alpha$).
- $ightharpoonup \alpha$ is a superkey for R.
- Each attribute A in $\beta \alpha$ is contained in a candidate key¹ for R.
- ▶ If a relation is in BCNF it is in 3NF (since in BCNF one of the first two conditions above must hold).
- ► Third condition is a minimal relaxation of BCNF to ensure dependency preservation (will see why later).

¹NOTE: each attribute may be in a different candidate key.

3NF Example

► Consider a schema:

dept_advisor(s_ID, i_ID, dept_name)

▶ With function dependencies:

$$i_ID \rightarrow dept_name$$

$$s_ID, dept_name \rightarrow i_ID$$

- ▶ Two candidate keys = $\{s_ID, dept_name\}, \{s_ID, i_ID\}$
- ▶ We have seen before that dept_advisor is not in BCNF.
- \triangleright R, however, is in 3NF:
 - $ightharpoonup s_ID$, $dept_name$ is a superkey,
 - ▶ $i_ID \rightarrow dept_name$ and i_ID is NOT a superkey, but:
 - $ightharpoonup dept_name i_ID = dept_name$ and
 - ▶ dept_name is contained in a candidate key.

Redundancy in 3NF

- \triangleright Consider the schema R below, which is in 3NF:
 - ightharpoonup R = (J, K, L),
 - $ightharpoonup F = \{JK \to L, L \to K\},$
 - ▶ And an instance table:

J	L	K
j_1	l_1	k_1
j_2	l_1	k_1
j_3	l_1	k_1
null	l_2	k_2

▶ What is wrong with the table?

Redundancy in 3NF

- \triangleright Consider the schema R below, which is in 3NF:
 - ightharpoonup R = (J, K, L),
 - $ightharpoonup F = \{JK \to L, L \to K\},$
 - ▶ And an instance table:

J	L	K
j_1	l_1	k_1
j_2	l_1	k_1
j_3	l_1	k_1
null	l_2	k_2

- ▶ What is wrong with the table?
 - Repetition of information,
 - Need to use null values (e.g., to represent the relationship l_2 , k_2 where there is no corresponding value for J)

Comparison of BCNF and 3NF

- ▶ Advantages to 3NF over BCNF. It is always possible to obtain a 3NF design without sacrificing losslessness or dependency preservation.
- ▶ Disadvantages to 3NF.
 - ▶ We may have to use null values to represent some of the possible meaningful relationships among data items.
 - ▶ There is the problem of repetition of information.

Goals of Normalization

- ightharpoonup Let R be a relation scheme with a set F of functional dependencies.
- ightharpoonup Decide whether a relation scheme R is in "good" form.
- ▶ In the case that a relation scheme R is not in "good" form, decompose it into a set of relation scheme $\{R_1, R_2, \dots, R_n\}$ such that:
 - ► Each relation scheme is in good form,
 - ▶ The decomposition is a lossless decomposition,
 - Preferably, the decomposition should be dependency preserving.

How good is BCNF?

- ▶ There are database schemas in BCNF that do not seem to be sufficiently normalized...
- ► Consider a relation:

inst_info(ID, child_name, phone)

where an instructor may have more than one phone and can have multiple children.

► Instance of inst_info

ID	$\mathbf{child}_{-}\mathbf{name}$	phone
99999	David	512-555-1234
99999	David	512 - 555 - 4321
99999	William	512-555-1234
99999	William	512 - 555 - 4321

How good is BCNF? (Cont.)

- ► There are no non-trivial functional dependencies and therefore the relation is in BCNF.
- ► Insertion anomalies i.e., if we add a phone 981-992-3443 to 99999, we need to add two tuples:

(99999, David, 981-992-3443) (99999, William, 981-992-3443)

Higher Normal Forms

- ► It is better to decompose inst_info into:
 - ▶ inst_child

ID	$\operatorname{child}_{-\operatorname{name}}$
99999	David
99999	William

▶ inst_phone

ID	phone
99999	512-555-1234
99999	512-555-4321

► This suggests the need for higher normal forms, such as Fourth Normal Form (4NF), which we shall see later

Features of Good Relational Design

Functional Dependencies

Decomposition Using Functional Dependencies

Normal Forms

Functional Dependency Theory

Algorithms for Decomposition using Functional Dependencies

Decomposition Using Multivalued Dependencies

More Normal Form

Atomic Domains and First Normal Form

Database-Design Process

Functional-Dependency Theory Roadmap

- ▶ We now consider the formal theory that tells us which functional dependencies are implied logically by a given set of functional dependencies.
- ▶ We then develop algorithms to generate lossless decompositions into BCNF and 3NF.
- ▶ We then develop algorithms to test if a decomposition is dependency-preserving.

Closure of a Set of Functional Dependencies

- ▶ Given a set F set of functional dependencies, there are certain other functional dependencies that are logically implied by F.
 - ▶ If $A \to B$ and $B \to C$, then we can infer that $A \to C$
 - ▶ etc.
- ightharpoonup The set of all functional dependencies logically implied by F is the closure of F.
- \blacktriangleright We denote the closure of F by F^+ .

Closure of a Set of Functional Dependencies

- ▶ We can compute F^+ , the closure of F, by repeatedly applying **Armstrong's Axioms**:
 - ▶ **Reflexive rule**: if $\beta \subseteq \alpha$, then $\alpha \to \beta$,
 - ▶ Augmentation rule: if $\alpha \to \beta$, then $\gamma \alpha \to \gamma \beta$,
 - **► Transitivity rule**: if $\alpha \to \beta$, and $\beta \to \gamma$, then $\alpha \to \gamma$.
- ► These rules are:
 - sound generate only functional dependencies that actually hold, and
 - **complete** generate all functional dependencies that hold.

Example of F^+

$$ightharpoonup R = (A, B, C, G, H, I)$$

$$F = \{A \rightarrow B$$

$$A \rightarrow C$$

$$CG \rightarrow H$$

$$CG \rightarrow I$$

$$B \rightarrow H\}$$

- \triangleright Some members of F^+
 - $ightharpoonup A \to H$ by transitivity from $A \to B$ and $B \to H$.
 - ▶ $AG \to I$ by augmenting $A \to C$ with G, to get $AG \to CG$ and then transitivity with $CG \to I$.
 - ▶ $CG \to HI$ by augmenting $CG \to I$ to infer $CG \to CGI$, and augmenting of $CG \to H$ to infer $CGI \to HI$, and then transitivity.

Closure of a Set of Functional Dependencies (Cont.)

- ► Additional rules:
 - ▶ Union rule: If $\alpha \to \beta$ holds and $\alpha \to \gamma$ holds, then $\alpha \to \beta \gamma$ holds.
 - **Decomposition rule**: If $\alpha \to \beta \gamma$ holds, then $\alpha \to \beta$ holds and $\alpha \to \gamma$ holds.
 - ▶ Pseudotransitivity rule: If $\alpha \to \beta$ holds and $\gamma\beta \to \delta$ holds, then $\alpha\gamma \to \delta$ holds.
- ▶ The above rules can be inferred from Armstrong's axioms.

Procedure for Computing F^+

▶ To compute² the closure of a set of functional dependencies F:

```
F^+ = F apply the reflexivity rule /* Generates all trivial dependencies */
repeat

for each functional dependency f in F^+
  apply the augmentation rule on f
  add the resulting functional dependencies to F^+

for each pair of functional dependencies f_1 and f_2 in F^+
  if f_1 and f_2 can be combined using transitivity
  add the resulting functional dependency to F^+
until F^+ does not change any further
```

Figure 7.7 A procedure to compute F^+ .

²NOTE: We shall see an alternative procedure for this task later.

Closure of Attribute Sets

- ▶ Given a set of attributes α , define the closure of α under F (denoted by α^+) as the set of attributes that are functionally determined by α under F.
- Algorithm to compute α^+ , the closure of α under F:

```
 \begin{array}{l} \textit{result} := \alpha; \\ \textbf{repeat} \\ \textbf{for each} \ \text{functional dependency} \ \beta \rightarrow \gamma \ \textbf{in} \ F \ \textbf{do} \\ \textbf{begin} \\ \textbf{if} \ \beta \subseteq \textit{result} \ \textbf{then} \ \textit{result} := \textit{result} \cup \gamma; \\ \textbf{end} \\ \textbf{until} \ (\textit{result} \ \text{does not change}) \\ \end{array}
```

Figure 7.8 An algorithm to compute α^+ , the closure of α under F.

Example of Attribute Set Closure

$$R = (A, B, C, G, H, I)$$

$$F = \{A \rightarrow B \\ A \rightarrow C \}$$

$$\begin{array}{c} CG \to H \\ CG \to I \\ B \to H \} \end{array}$$

- \triangleright $(AG)^+$
 - 1. result = AG.
 - 2. result = $ABCG(A \rightarrow C \text{ and } A \rightarrow B)$.
 - 3. result = $ABCGH(CG \rightarrow H \text{ and } CG \subseteq AGBC)$.
 - 4. result = $ABCGHI(CG \rightarrow I \text{ and } CG \subseteq AGBCH)$.
- ightharpoonup Is AG a candidate key?
 - 1. Is AG a super key?
 - 1.1 Does $AG \to R$? == Is $R \supseteq (AG)^+$.
 - 2. Is any subset of AG a superkey?
 - 2.1 Does $A \to R$? == Is $R \supseteq (A)^+$.
 - 2.2 Does $G \to R$? == Is $R \supseteq (G)^+$.
 - 2.3 In general: check for each subset of size n-1.

Features of Good Relational Design

Functional Dependencies

Decomposition Using Functional Dependencies

Normal Forms

Functional Dependency Theory

Algorithms for Decomposition using Functional Dependencies

Decomposition Using Multivalued Dependencies

More Normal Form

Atomic Domains and First Normal Form

Database-Design Process

Features of Good Relational Design

Functional Dependencies

Decomposition Using Functional Dependencies

Normal Forms

Functional Dependency Theory

Algorithms for Decomposition using Functional Dependencies

Decomposition Using Multivalued Dependencies

More Normal Form

Atomic Domains and First Normal Form

Database-Design Process

Features of Good Relational Design

Functional Dependencies

Decomposition Using Functional Dependencies

Normal Forms

Functional Dependency Theory

Algorithms for Decomposition using Functional Dependencies

Decomposition Using Multivalued Dependencies

More Normal Form

Atomic Domains and First Normal Form

Database-Design Process

Features of Good Relational Design

Functional Dependencies

Decomposition Using Functional Dependencies

Normal Forms

Functional Dependency Theory

Algorithms for Decomposition using Functional Dependencies

Decomposition Using Multivalued Dependencies

More Normal Form

Atomic Domains and First Normal Form

Database-Design Process

Features of Good Relational Design

Functional Dependencies

Decomposition Using Functional Dependencies

Normal Forms

Functional Dependency Theory

Algorithms for Decomposition using Functional Dependencies

Decomposition Using Multivalued Dependencies

More Normal Form

Atomic Domains and First Normal Form

Database-Design Process

Features of Good Relational Design

Functional Dependencies

Decomposition Using Functional Dependencies

Normal Forms

Functional Dependency Theory

Algorithms for Decomposition using Functional Dependencies

Decomposition Using Multivalued Dependencies

More Normal Form

Atomic Domains and First Normal Form

Database-Design Process

End of Chapter 7.

TDT5FTOTTC

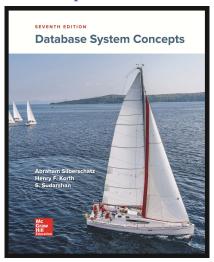


Top 5 Fundamental Takeaways

Top 5 Fundamental Takeaways

5

Database System Concepts



Content has been extracted from Database System Concepts, Seventh Edition, by Silberschatz, Korth and Sudarshan. Mc Graw Hill Education. 2019.
Visit https://db-book.com/.