SLOTS 03

CHAPTER 3

DESIGN THEORY FOR RELATIONAL DATABASES

OBJECTIVES

Understand concepts of:

- ❑ **Functional Dependencies** (Phụ thuộc hàm)
- ❑ **Normalization** (Chuẩn hoá lược đồ CSDL)
- ❑ **Decomposition** (Phân rã lược đồ CSDL)
- ❑ **Multi-valued Dependencies** (Phụ thuộc hàm đa trị)

CONTENT

- ❑Functional Dependencies (SLOT3)
- ❑Rules about FDs (SLOT 4)
- ❑Key & Super-Key (SLOT 4)
- ❑Normal forms (SLOT 5)

I.

FUNCTIONAL DEPENDENCIES

3.1. FUNCTIONAL DEPENDENCIES (FDS)

Functional dependencies (FDs)

A **functional dependency:** constraint between **two sets of attributes** in a relation (describe relationship among attributes in a relation)

X (include $A_1A_2...A_n$): a set of attributes in R

Y (include $B_1B_2...B_m$): another set of attribute in R

3.1. FUNCTIONAL DEPENDENCY (FD)

We write **X→Y** **(read as** X determines Y or Y is functionally dependent on X) if and only if each *X* value is associated with precisely one *Y* value.

Means given a **value** of X, we can find **one and exactly one value** of Y

Examples 1

Student

<u>sId</u>	sName	sDoB
1	John Smith	1999-01-12
2	Remesh Shah	1998-02-28
3	Susan Black	1999-08-10
4	John Smith	1999-01-12
5	John Doe	1998-02-28

Functional dependencies:

sId → {sName, sDob}

sName → sDoB

title	year	length	genre	studioName	star Name
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Hark Hamill
Star Vars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	drama	MGM	Vivien Leigh
Wayne's World	1992	95	comedy	Paramount	Dana Carvey
Wayne'a World	1992	95	comedy	Paramount	Hike Meyers

Easy to see that: the following FD is true

title,year → length, genre, studioName

Exercise: How about the FD

title,year → starName ???

title, year → startName does not hold in Movies1 relation

Identifying Functional Dependencies

Find the FDs in the relation below:

CarReg	*hireDate*	*Make*	*model*	*custNo*	*custName*	*outletNo*	*outletLoc*
MS34 0GD	14/5/03	Ford	Focus	C100	Smith, J	01	Bearsden
MS34 0GD	15/5/03	Ford	Focus	C201	Hen, P	01	Bearsden
NS34 TPR	16/5/03	Nissan	Sunny	C100	Smith, J	01	Bearsden
MH34 BRP	14/5/03	Ford	Ka	C313	Blatt, O	02	Kelvinbridge
MH34 BRP	20/5/03	Ford	Ka	C100	Smith, J	02	Kelvinbridge
MD510PQ	20/5/03	Nissan	Sunny	C295	Pen, T	02	Kelvinbridge

Identifying Functional Dependencies

FD1: carReg -> make, model, outletNo, outletLoc

FD2: custNo -> custName

FD3: outletNo -> outletLoc

FD4: model -> make (only if we assume a model name is unique to a make)

FD5: carReg, hireDate -> make, model, custNo, custName, outletNo, outletLoc

…

KEY & SUPPER KEY

Student

StudentID	Name	Major
1	Alice	Computer Science
2	Bob	Mathematics
3	Charlie	Computer Science
4	David	Physics

KEY of relation

Key

- It has no duplicate values, it has unique values.
- It cannot be NULL.
- It can identify only one tuple (a record) at a time.
- Primary keys are not necessarily to be a single column; more than one column can also be a primary key for a table.

SUPPER KEY

Review key of relation, candidate keys (alternate keys), primary key, Super-key

Super-key: A set of attributes that contains a key is called a *super-key*.

- Every super-key satisfies the first condition of a key: it functionally determines all other attributes of the relation
- If K is a key, L is a super key, then: $K \subseteq L$
- A key is also a **super key**

3.2.

RULES ABOUT FDS

3. RULES ABOUT FDS

1. Armstrong's Axioms (Hệ tiên đề Armstrong)

2. The Closure of Attributes (Bao đóng của tập thuộc tính)

3. Closing Sets of FDs & minimal basic (Bao đóng của tập phụ thuộc hàm & phủ tối thiểu (phụ thuộc hàm tối thiểu))

4. Projecting FDs

5. Anomalies introduction

3.1.
Armstrong's Axioms

3.1. Armstrong's Axioms

A. Fundamental Rules: Let X, Y, Z are sets of attributes

- **Reflexivity (Luật phản xạ):**

  If X is a subset of Y, then Y→X

- **Augmentation (Luật thêm/tăng cường)**

  If X→Y, then XZ→YZ for any Z

- **Transitivity (Luật bắc cầu)**

  If X→Y and Y→Z, then X→Z

3.1. Armstrong's Axioms

B. Additional rules (Các luật mở rộng): Let X, Y, Z, W are sets of attributes

- **Union/Combining (Luật hợp)**:

  if $X \rightarrow Y$ AND $X \rightarrow Z$ then $X \rightarrow YZ$

- **Decomposition/Splitting (Luật tách):**

  if $X \rightarrow YZ$, then $X \rightarrow Y$ and $X \rightarrow Z$

3.1. Armstrong's Axioms

B. Additional rules (Các luật mở rộng): Let X, Y, Z, W are sets of attributes

- Pseudotransitivity (Luật giả bắc cầu):

  If X → Y and WY → Z then WX →Z

C. Trivial FDs (Các phụ thuộc hàm tầm thường): right side is a subset of left side

- Ex: FLD →FD

3.1. Armstrong's Axioms

- A set of FD's *S* **follows (suy dẫn)** from a set of FD's *T* if every relation instance that satisfies all the FD's in *T* also satisfies all the FD's in *S*
- Two sets of FD's *S* and *T* are **equivalent** if and only if *S* **follows** from *T*, and *T* **follows** *S*

3.2.
The Closure of Attributes

3.2. The Closure of Attributes

- The closure of a set of attributes $\{A_1, A_2, \ldots, A_n\}$ (bao đóng của tập thuộc tính)under FD's in S *(denoted* $\{A_1, A_2, \ldots, A_n\}^+)_F$ is the set of attributes B such that every relation that satisfies all the FD's in set S also satisfies $A_1A_2\ldots A_n \rightarrow B$
- That is, $A_1A_2\ldots A_n \rightarrow B$ follows from the FD's of S
- $A_1, A_2, \ldots, A_n \in \{A_1, A_2, \ldots, A_n\}+$, because $A_1A_2\ldots A_n \rightarrow A_i$ is trivial

3.2. The Closure of Attributes

- **Algorithm 3.7: Closure of a set of attributes**

  Input: A set of attributes $\{A_1, A_2, \ldots, A_n\}$ and a set of FD's S

  Output: The closure $\{A_1, A_2, \ldots, A_n\}^+$

  1. If necessary, split the FD's of S, so each FD in S have singleton right side
  2. Let X be a set of attributes that will become the closure. Initialize X to be $\{A_1, A_2, \ldots, A_n\}$
  3. Repeatedly search for some FD: $B_1B_2 \ldots B_m \rightarrow C$, such that B_1, B_2, …, B_m are in X, but C is not
     a) If such C is found, add to X, and repeat the search
     b) If such C is not found, no more attributes can be added to X
  4. The set X is the correct value of $\{A_1, A_2, \ldots, A_n\}$+

Compute the Closure of Attributes

Example 3.8: Let us consider a relation with attributes A, B, C, D, E, and F. Suppose that this relation has the FD's $AB \rightarrow C$, $BC \rightarrow AD$, $D \rightarrow E$, and $CF \rightarrow B$. What is the closure of $\{A, B\}$, that is, $\{A, B\}^+$?

First, split $BC \rightarrow AD$ into $BC \rightarrow A$ and $BC \rightarrow D$. Then, start with $X = \{A, B\}$. First, notice that both attributes on the left side of FD $AB \rightarrow C$ are in X, so we may add the attribute C, which is on the right side of that FD. Thus, after one iteration of Step 3, X becomes $\{A, B, C\}$.

Next, we see that the left sides of $BC \rightarrow A$ and $BC \rightarrow D$ are now contained in X, so we may add to X the attributes A and D. A is already there, but D is not, so X next becomes $\{A, B, C, D\}$. At this point, we may use the FD $D \rightarrow E$ to add E to X, which is now $\{A, B, C, D, E\}$. No more changes to X are possible. In particular, the FD $CF \rightarrow B$ can not be used, because its left side never becomes contained in X. Thus, $\{A, B\}^+ = \{A, B, C, D, E\}$. $\square$

Compute the Closure of Attributes

EX1:

R(A, B, C, D)

S = {A→B, B→C, C→D, D→A}

Compute A^+_S , B^+_S , BD^+_S.

What are some the keys of R?

Compute the Closure of Attributes

Solution:

R(A, B, C, D)

F = {A→B, B→C, C→D, D→A}

A^+_F = ABCD = R^+ (a key of R)

B^+_S = BCDA = R^+ (a key of R)

BD^+_S = BDAC = R^+ (a supperkey of R)

Compute the Closure of Attributes

EX2:

Q(ABCDEGH),

$F = \{B \rightarrow A,\ DA \rightarrow CE,\ \ D \rightarrow H,\ GH \rightarrow C, AC \rightarrow D\ \}$

Compute A^+_F ,ABC^+_F $DB^+_{F,}$ BG^+_F

What are some the keys of R?

3.3.
Closing Sets of FDs
& minimal basic

3.3. Closing Sets of FDs

- Suppose a set of FD's *S*, any set of FD's T equivalent to *S* is said to be a ***basis*** for *S*.
  Then we say T is a ***basis*** for *S*
- Just work with only FD's that have *singleton right sides*
- A ***minimal basis (minimal cover)*** for FD's *S* is a ***basis B*** that satisfies three conditions:
  - All the FD's in *B* have **singleton right sides**
  - If any FD is removed from *B*, the result is no longer a basis
  - If for any FD in *B* we remove one or more attributes from the left side, the result is no longer a basis

Steps to Find Minimal Cover

1. Split the right-hand attributes of all FDs.

Example: A->XY => A->X, A->Y

2. Remove Redundant Attributes:

For each FD, check if any attribute can be removed without changing the implied FD. If so, remove it.

3. Remove Redundant FDs:

For each FD, check if it can be implied by other FDs in the set. If so, remove it.

3.3. Closing Sets of FDs

Example:

R(A,B,C,D)

F = {AB→CD, B→C, C→ D}

Find minimal basis:

Step 1: F1 = {AB→C, AB→D, B→C, C→ D}

Step 2: F2 = {B→D, B→C, C→ D}

Step 3: F3 = {B→C, C→ D}

EXERCISE

Exercise 3.2.1: Consider a relation with schema $R(A, B, C, D)$ and FD's $AB \rightarrow C$, $C \rightarrow D$, and $D \rightarrow A$.

a) What are all the nontrivial FD's that follow from the given FD's? You should restrict yourself to FD's with single attributes on the right side.

b) What are all the keys of R?

c) What are all the superkeys for R that are not keys?

EXERCISE

Exercise 3.2.2: Repeat Exercise 3.2.1 for the following schemas and sets of FD's:

i) $S(A, B, C, D)$ with FD's $A \rightarrow B$, $B \rightarrow C$, and $B \rightarrow D$.

ii) $T(A, B, C, D)$ with FD's $AB \rightarrow C$, $BC \rightarrow D$, $CD \rightarrow A$, and $AD \rightarrow B$.

iii) $U(A, B, C, D)$ with FD's $A \rightarrow B$, $B \rightarrow C$, $C \rightarrow D$, and $D \rightarrow A$.

EXERCISE

! **Exercise 3.2.10:** Suppose we have relation $R(A, B, C, D, E)$, with some set of FD's, and we wish to project those FD's onto relation $S(A, B, C)$. Give the FD's that hold in S if the FD's for R are:

a) $AB \to DE$, $C \to E$, $D \to C$, and $E \to A$.

b) $A \to D$, $BD \to E$, $AC \to E$, and $DE \to B$.

c) $AB \to D$, $AC \to E$, $BC \to D$, $D \to A$, and $E \to B$.

d) $A \to B$, $B \to C$, $C \to D$, $D \to E$, and $E \to A$.

In each case, it is sufficient to give a minimal basis for the full set of FD's of S.

3.4.
Projecting FDs

What happens to …

- … a set of FD's S of R when we project R on some attributes?
- That is, suppose a relation R with set of FD's S, and $R_1 = \pi_L(R)$. What FD's hold in R_1?

3.4. Projecting FDs

- To find a FDs of projection, we
  - Follow from *S*, and
  - Involve only attributes of R_1

3.4. Projecting FDs

Algorithm 3.12: Projecting a Set of FD's

- **Input**: R, $R_1=\pi_L(R)$, *S* a set of FD's that hold in R
- **Output**: the set of FD's that hold in R_1
- **Method:**
  - T is the set of FD's that hold in R_1. Initially, T is empty
  - For each set of attributes X of R_1, compute $X+$. Add to *T* all non-trivial FD's $X \rightarrow A$ such that *A* is both in $X+$ and an attribute of R_1
  - Construct a minimal basis from *T*

3.4. Projecting FDs

Algorithm 3.12: Projecting a Set of FD's

- **Compute a minimal basis from *T***
  - If there is an FD *F* in *T* that follows from other FD's in *T*, then remove *F* from *T*
  - Let Y → B is a FD in *T* , with at least two attributes in *Y*, and let *Z* is *Y* with one of its attributes removed:
    - If *Z* → *B* follows from the other FD's in T (including *Y* → *B*), then replace *Y* → *B* by *Z* → *B*
  - Repeat the above steps in all possible ways until no more changes to *T* can be made

3.4. Projecting FDs

Two notations

- (1) Closing the empty set and the set of all attributes cannot yield a nontrivial FD.
- (2) If we have already knew that the closure of some set *X* is all attributes, then we cannot discover any new FD's by closing supersets of *X*

3.4. Projecting FDs

Example:

R (A,B,C,D)

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

$R1 = \pi_{A,C,D}(R)$.

Find the FD's of R1?

3.4. Projecting FDs

Step1: **Compute the closure of the singleton set**

$A^+ = \{A,B,C,D\}$, and B is not in R1, then new FD's $A \rightarrow C$, $A \rightarrow D$

$C^+=\{C,D\}$, then new FD's $C \rightarrow D$

$D^+=\{D\}$, no new FD's

3.4. Projecting FDs

Step2: Compute the closure of the doubleton set

Since {A}+ include all attributes, no care any more for supersets of {A}

{C,D}+={C,D}, no new FD's holds in R1

- Finally, there are three FD's A→C, A→D, C→D hold in R1
- A→D is transitive from A→C, and C→D
- So, minimal basis is {A→C, C→D}

3.5.
Anomalies introduction

3.5. Anomalies introduction

- Careless selection of **a relational database schema** can lead to **redundancy** and **related anomalies**

So, in this session we shall tackle the problems of relational database designing

- Problems such as **redundancy** that occur when we try to cram too much into a single relation are called ***anomalies.***

3.5. Anomalies introduction

title	*year*	*length*	*genre*	*studioName*	*starName*
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With The Wind	1939	231	drama	MGM	Vivien Leigh
Wayne's World	1992	95	comedy	Paramount	Dana Carvey
Wayne's World	1992	95	comedy	Paramount	Mike Meyers

Redundancy?
Update Anomalies?
Deletion Anomalies

3.5. Anomalies introduction

The principal kinds of anomalies that we encounter are:

- ***Redundancy***: information maybe repeated unnecessarily in several tuples (**exp**: the length and genre)
- ***Update Anomalies***: We may change information in one tuple but leave the same information unchanged in another. (**exp**: if we found that *Star Wars* is 125 minutes long, we may change the length in the first tuple but not in the second and third tuples)
- ***Deletion Anomalies***: If a set of values becomes empty, we may lose other information as a side effect. (**exp**: if we delete "Fox" from the set of studios, then we have no more studios for the movie "Star Wars"

REVIEW 01

Example 3.1 : Let us consider the relation

```
Movies1(title, year, length, genre, studioName, starName)
```

Example 3.5 : In Example 3.1 the set of FD's:

$$\texttt{title year} \rightarrow \texttt{length}$$
$$\texttt{title year} \rightarrow \texttt{genre}$$
$$\texttt{title year} \rightarrow \texttt{studioName}$$

is equivalent to the single FD:

$$\texttt{title year} \rightarrow \texttt{length genre studioName}$$

that we asserted there. □

REVIEW 02

Example 3.8: Let us consider a relation with attributes A, B, C, D, E, and F. Suppose that this relation has the FD's $AB \to C$, $BC \to AD$, $D \to E$, and $CF \to B$. What is the closure of $\{A, B\}$, that is, $\{A, B\}^+$?

Example 3.9: Consider the relation and FD's of Example 3.8. Suppose we wish to test whether $AB \to D$ follows from these FD's. We compute $\{A, B\}^+$, which is $\{A, B, C, D, E\}$, as we saw in that example. Since D is a member of the closure, we conclude that $AB \to D$ does follow.

On the other hand, consider the FD $D \to A$. To test whether this FD follows from the given FD's, first compute $\{D\}^+$. To do so, we start with $X = \{D\}$. We can use the FD $D \to E$ to add E to the set X. However, then we are stuck. We cannot find any other FD whose left side is contained in $X = \{D, E\}$, so $\{D\}^+ = \{D, E\}$. Since A is not a member of $\{D, E\}$, we conclude that $D \to A$ does not follow. □

REVIEW 03

Example 3.11: Consider a relation $R(A, B, C)$ such that each attribute functionally determines the other two attributes. The full set of derived FD's thus includes six FD's with one attribute on the left and one on the right; $A \to B$, $A \to C$, $B \to A$, $B \to C$, $C \to A$, and $C \to B$. It also includes the three nontrivial FD's with two attributes on the left: $AB \to C$, $AC \to B$, and $BC \to A$. There are also FD's with more than one attribute on the right, such as $A \to BC$, and trivial FD's such as $A \to A$.

Relation R and its FD's have several minimal bases. One is

$$\{A \to B,\ B \to A,\ B \to C,\ C \to B\}$$

Another is $\{A \to B,\ B \to C,\ C \to A\}$. There are several other minimal bases for R, and we leave their discovery as an exercise. $\square$