Evaluación de la calidad del diseño

Sivana Hamer - sivana.hamer@ucr.ac.cr
Escuela de Ciencias de la Computación

UNIVERSIDAD DE COSTA RICA

¿Cómo determinamos que un diseño es de calidad?

Es importante considerar como objetivo del diseño…

Preservar información

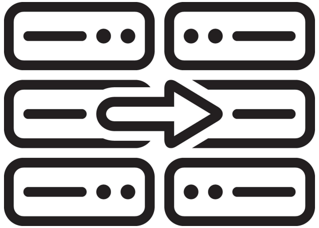
Minima redundancia

Falso en

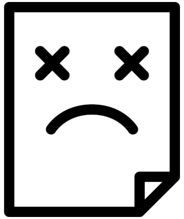
NO SQL

Existen algunas guías que ayudan a determinar informalmente la calidad

Semántica clara



Evitar redundancia



Reducir NULLs

Considerar tuplas falsas

Semántica clara busca que los elementos de la base de datos sean sencillos de entender

STUDENT_SCHOOL_FACULTY

<u>Email</u>	Name	Phone_number	<u>Acronym</u>	Number_students

Guideline #1: While designing the relational schema, the attributes should be easy to understand and explain. Thus, every entity or relationship has a straightforward meaning representing only one concept.

Evitar redundancia buscar reducir espacio y anomalías (inserción, borrado y modificación)

STUDENT_SCHOOL_FACULTY

Student_mail	Student_name	School_acronym	School_phone	Faculty_name
alicia.armando23	Alicia Armando	ECCI	2511-8000	Ingenería
bob.benavidez	Bob Benavidez	ECCI	2511-8000	Ingenería
bob.benavidez	Bob Benavidez	EMat	2511-6551	Ciencias Básicas
carlos.calvo	Carlos Calvo	EAN	2511-9180	Ciencias Sociales
daniela.delgado	Daniela Delgado	EMat	2511-6551	Ciencias Básicas

Guideline #2: Design to avoid insertion, deletion or modification anomalies. If not possible, clearly state the anomalies and update the database correctly.

Al **insertar** hay que asegurarse que se duplique la información correctamente

STUDENT_SCHOOL_FACULTY

Student_mail	Student_name	School_acronym	School_phone	Faculty_name
alicia.armando23	Alicia Armando	ECCI	2511-8000	Ingenería
bob.benavidez	Bob Benavidez	ECCI	2511-8000	Ingenería
bob.benavidez	Bob Benavidez	EMat	2511-6551	Ciencias Básicas
carlos.calvo	Carlos Calvo	EAN	2511-9180	Ciencias Sociales

\+

daniela.delgado	Daniela Delgado	EMat	2511-6551	Ciencias Básicas

Al **borrar** se puede perder todos los datos en la BD

STUDENT_SCHOOL_FACULTY

Student_mail	Student_name	School_acronym	School_phone	Faculty_name
alicia.armando23	Alicia Armando	ECCI	2511-8000	Ingeniería
bob.benavidez	Bob Benavidez	ECCI	2511-8000	Ingeniería
bob.benavidez	Bob Benavidez	EMat	2511-6551	Ciencias Básicas
daniela.delgado	Daniela Delgado	EMat	2511-6551	Ciencias Básicas
		EAN	251[illegible]180	Ciencias Sociales

Al **actualziar** se debe asegurar que se actualizen todos los datos

STUDENT_SCHOOL_FACULTY

Student_mail	Student_name	School_acronym	School_phone	Faculty_name
alicia.armando23	Alicia Armando	ECCI	2511-8000	Ingenería
bob.benavidez	Bob Benavidez	ECCI	2511-8000	Ingenería
bob.benavidez	Bob Benavidez	EMat	2511-6551	Ciencias Básicas
carlos.calvo	Carlos Calvo	EAN	2511-9180	Ciencias Sociales
daniela.delgado	Daniela Delgado	EMat	2511-6551	Ciencias Básicas

Reducir NULLs busca gastar menos espacio

STUDENT

<u>Email</u>	Name	Position_association	Fax_number	Kindergarden_grade	Instagram

Guideline #3: Avoid NULL attributes while possible. If not possible, ensure that they do not apply to the majority of tuples in the relation.

Considerar tuplas falsas busca considerar las tuplas que se van a generar al realizar un join

STUDENT

Email	Name	Faculty_name
alicia.armando23	Alicia Armando	Ingenería
bob.benavidez	Bob Benavidez	Ingenería
bob.benavidez	Bob Benavidez	Ciencias Básicas
carlos.calvo	Carlos Calvo	Ciencias Sociales
daniela.delgado	Daniela Delgado	Ciencias Básicas

SCHOOL

School_acronym	School_phone	Faculty_name
ECCI	2511-8000	Ingenería
EMat	2511-6551	Ciencias Básicas
EAN	2511-9180	Ciencias Sociales
ECCC	2511-3600	Ciencias Sociales

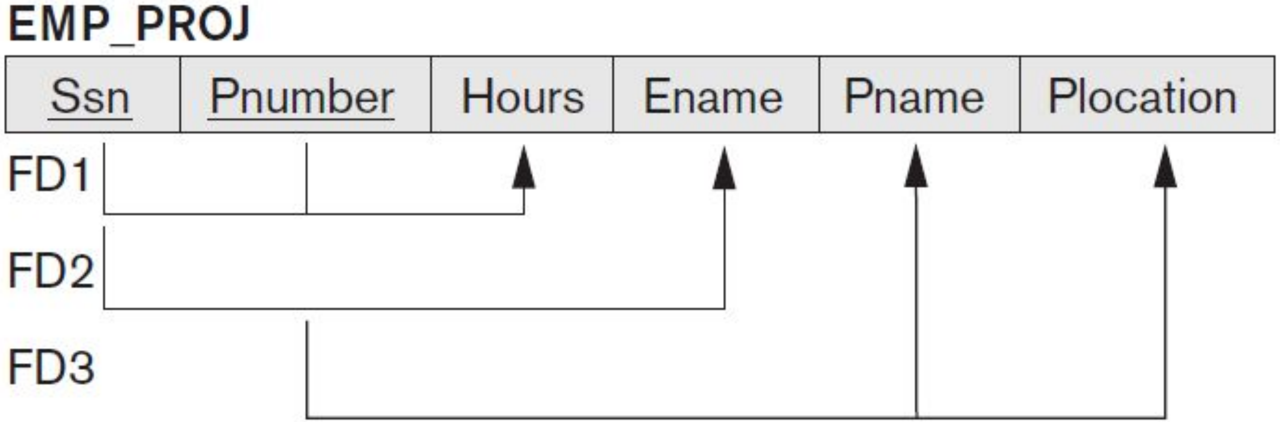
STUDENT_SCHOOL_FACULTY

Student_mail	Student_name	School_acronym	School_phone	Faculty_name
alicia.armando23	Alicia Armando	ECCI	2511-8000	Ingenería
bob.benavidez	Bob Benavidez	ECCI	2511-8000	Ingenería
bob.benavidez	Bob Benavidez	EMat	2511-6551	Ciencias Básicas
carlos.calvo	Carlos Calvo	EAN	2511-9180	Ciencias Sociales
carlos.calvo	Carlos Calvo	ECCC	2511-3600	Ciencias Sociales
daniela.delgado	Daniela Delgado	EMat	2511-6551	Ciencias Básicas

Guideline #4: Design relation schemas so that they can be joined correctly, using the appropriate primary and foreign keys that do not generate false tuples.

Se pueden definir formalmente algunas de estas relaciones con **dependencias funcionales (FD)**

A *functional dependency* is a constraint denoted by $X \rightarrow Y$ for a set of attributes X and Y for a relation R.It can be defined when any two tuples that have $t_1[X] = t_2[X]$ must also have $t_1[Y] = t_2[Y]$.



All candidate keys X can determine the attributes of any attribute Y in R. Therefore, $X \rightarrow R$.

De un estado r no se pueden definir una FD, pero sí asumir que **puede existir** una FD

A	B	C	D
a1	b1	c1	d1
a1	b2	c2	d2
a2	b2	c2	d3
a3	b3	c4	d3

puede $B \rightarrow C$

X	Y
b1	c1
b2	c2
b3	c4

A	B	C	D
a1	b1	c1	d1
a1	b2	c2	d2
a2	b2	c2	d3
a3	b3	c4	d3

no puede $A \rightarrow B$

X | Y
a1 | b1
a1 | b2

A	B	C	D
a1	b1	c1	d1
a1	b2	c2	d2
a2	b2	c2	d3
a3	b3	c4	d3

¿Puede B→ A?

A) Si

B) No

A	B	C	D
a1	b1	c1	d1
a1	b2	c2	d2
a2	b2	c2	d3
a3	b3	c4	d3

¿Puede $\{A, B\} \rightarrow D$?

A) Si

B) No

A	B	C	D
a1	b1	c1	d1
a1	b2	c2	d2
a2	b2	c2	d3
a3	b3	c4	d3

¿Puede C → B?

A) Si

B) No

Nota: Que puede B → C no implica puede C → B

A	B	C	D
a1	b1	c1	d1
a1	b2	c2	d2
a2	b2	c2	d3
a3	b3	c4	d3

¿Puede D → C?

A) Si

B) No

Normalización fue usar primero para revisar *compliance* con una forma normal

Information Retrieval P. BAXENDALE, Editor

A Relational Model of Data for Large Shared Data Banks

E. F. Codd
IBM Research Laboratory, San Jose, California

Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). A prompting service which supplies such information is not a satisfactory solution. Activities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed. Changes in data representation will often be needed as a result of changes in query, update, and report traffic and natural growth in the types of stored information.

Existing noninferential, formatted data systems provide users with tree-structured files or slightly more general network models of the data. In Section 1, inadequacies of these models are discussed. A model based on *n*-ary relations, a normal form for data base relations, and the concept of a universal data sublanguage are introduced. In Section 2, certain operations on relations (other than logical inference) are discussed and applied to the problems of redundancy and consistency in the user's model.

KEY WORDS AND PHRASES: data bank, data base, data structure, data organization, hierarchies of data, networks of data, relations, derivability, redundancy, consistency, composition, join, retrieval language, predicate calculus, security, data integrity

CR CATEGORIES: 3.70, 3.73, 3.75, 4.20, 4.22, 4.29

The relational view (or model) of data described in Section 1 appears to be superior in several respects to the graph or network model [3, 4] presently in vogue for non-inferential systems. It provides a means of describing data with its natural structure only—that is, without superimposing any additional structure for machine representation purposes. Accordingly, it provides a basis for a high level data language which will yield maximal independence between programs on the one hand and machine representation and organization of data on the other.

A further advantage of the relational view is that it forms a sound basis for treating derivability, redundancy, and consistency of relations—these are discussed in Section 2. The network model, on the other hand, has spawned a number of confusions, not the least of which is mistaking the derivation of connections for the derivation of relations (see remarks in Section 2 on the "connection trap").

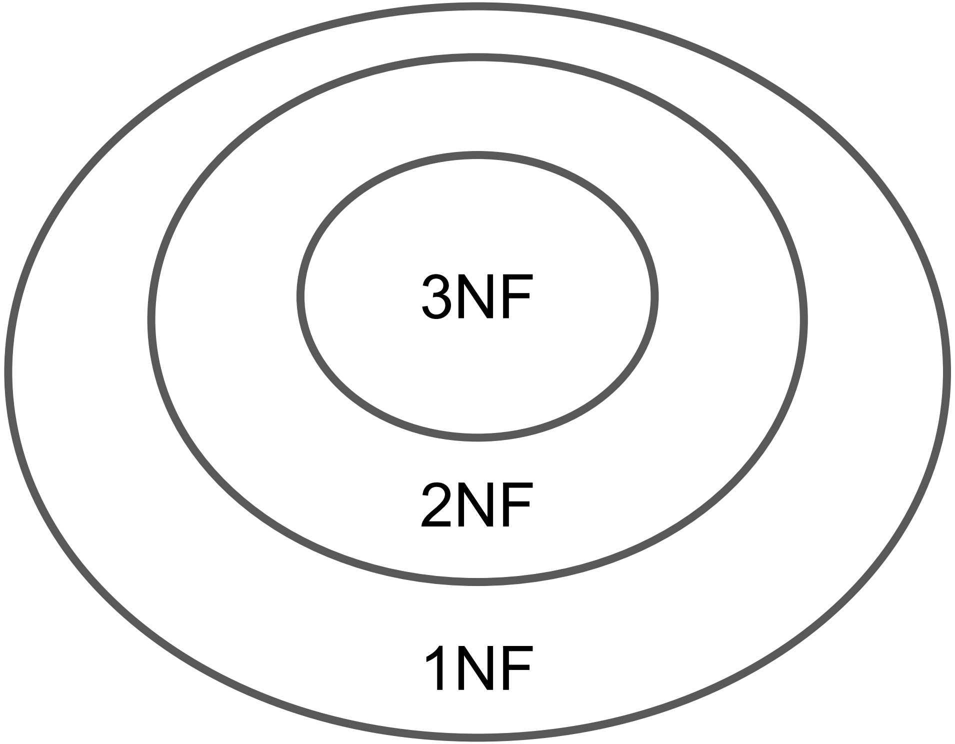
Finally, the relational view permits a clearer evaluation of the scope and logical limitations of present formatted data systems, and also the relative merits (from a logical standpoint) of competing representations of data withir single system. Examples of this clearer perspective cited in various parts of this paper. Implementations systems to support the relational model are not discuss

1.2. Data Dependencies in Present Systems

The provision of data description tables in recently veloped information systems represents a major adva toward the goal of data independence [5, 6, 7]. Such tab facilitate changing certain characteristics of the data rep sentation stored in a data bank. However, the variety data representation characteristics which can be chang *without logically impairing some application programs* still quite limited. Further, the model of data with whi users interact is still cluttered with representational prop-

Vamos a ver tres formas normales

Las formas normales de mayores requieren de haber cumplido con los niveles menores



1NF se cumple si solo tiene atributos atómicos y relaciones no anidadas

No 1NF

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

1NF con redundancia

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocation
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

Se quita la redundancia moviendo el atributo en una nueva relación

1NF con redundancia

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocation
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

1NF sin redundancia

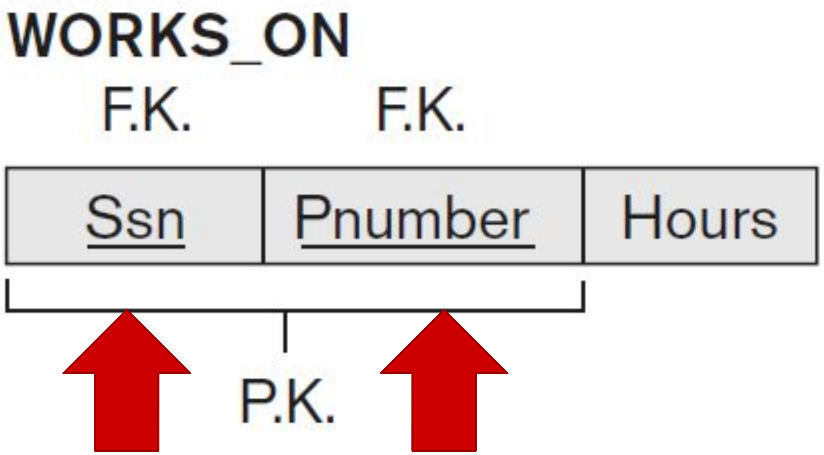
DEPARTMENT

Dname	Dnumber	Dmgr_ssn
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

DEPT_LOCATIONS

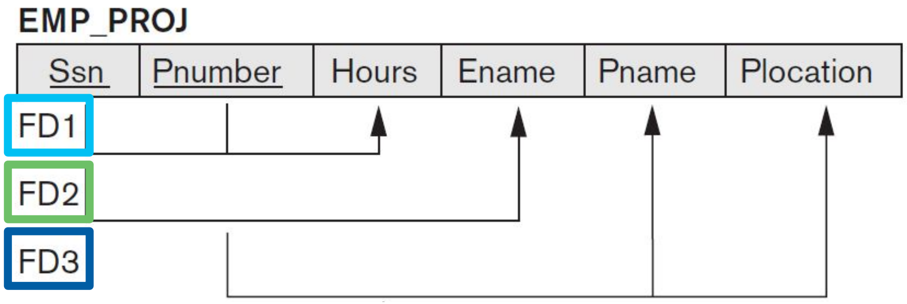
Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

Los **atributos primos** son los atributos que forman parte de una llave candidata

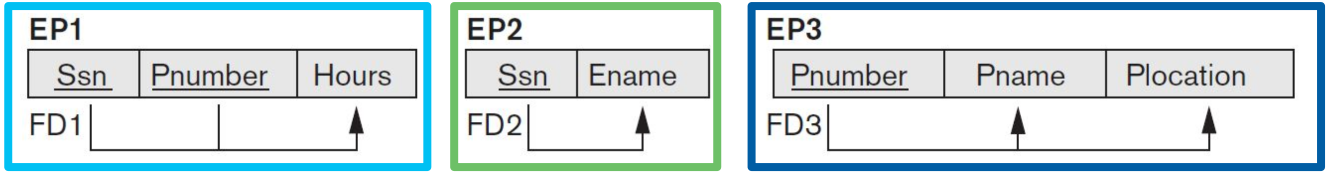


2NF se cumple si todo atributo no primo en R tiene un FD al atributo de la llave de R, por lo que se crea una tabla por FD

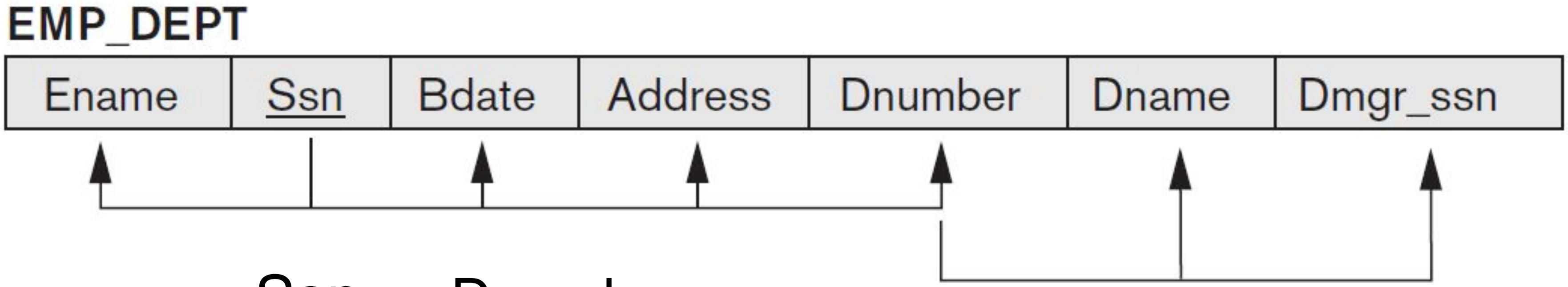
1NF



2NF



Una dependencia transitiva X → Z existe si hay un FD de X → Y y Y → Z, donde Y no es atributo primario



Ssn → Dnumber

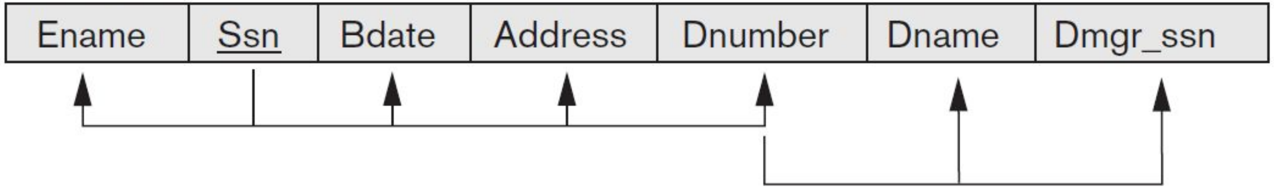
Dnumber → Dmgr_ssn

Dependencia transitiva: Ssn → Dmgr_ssn con Dnumber

3NF sucede cuando no hay dependencias transitivas para atributos no primos

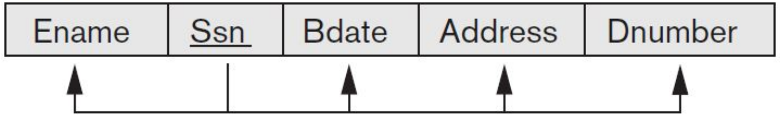
2NF

EMP_DEPT



3NF

ED1



ED2

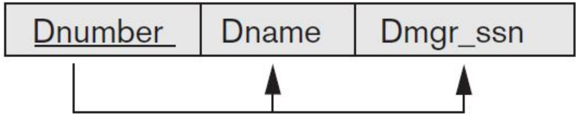
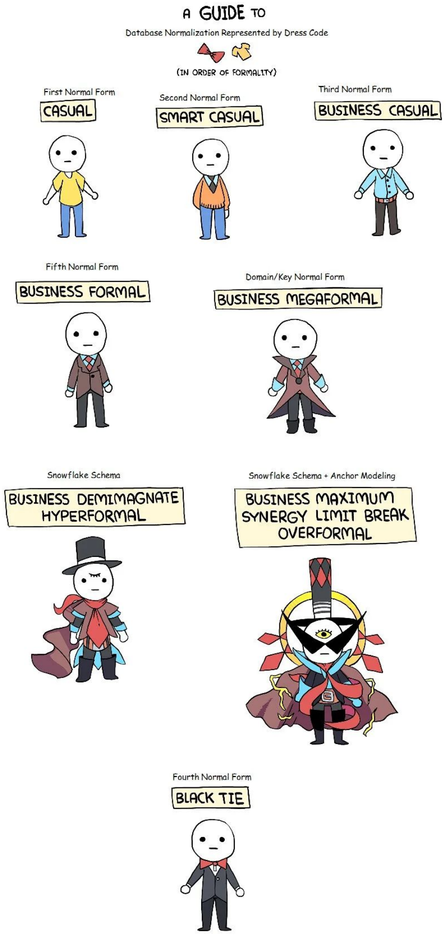


Table 14.1 Summary of Normal Forms Based on Primary Keys and Corresponding Normalization

Normal Form	Test	Remedy (Normalization)
First (1NF)	Relation should have no multivalued attributes or nested relations.	Form new relations for each multivalued attribute or nested relation.
Second (2NF)	For relations where primary key contains multiple attributes, no nonkey attribute should be functionally dependent on a part of the primary key.	Decompose and set up a new relation for each partial key with its dependent attribute(s). Make sure to keep a relation with the original primary key and any attributes that are fully functionally dependent on it.
Third (3NF)	Relation should not have a nonkey attribute functionally determined by another nonkey attribute (or by a set of nonkey attributes). That is, there should be no transitive dependency of a nonkey attribute on the primary key.	Decompose and set up a relation that includes the nonkey attribute(s) that functionally determine(s) other nonkey attribute(s).


A GUIDE TO
Database Normalization Represented by Dress Code
(IN ORDER OF FORMALITY)
First Normal Form
CASUAL
Second Normal Form
SMART CASUAL
Third Normal Form
BUSINESS CASUAL
Fifth Normal Form
BUSINESS FORMAL
Domain/Key Normal Form
BUSINESS MEGAFORMAL
Snowflake Schema
BUSINESS DEMIMAGNATE HYPERFORMAL
Snowflake Schema + Anchor Modeling
BUSINESS MAXIMUM SYNERGY LIMIT BREAK OVERFORMAL
Fourth Normal Form
BLACK TIE

Referencias

- R. Elmasri and S. Navathe, Fundamentals of database systems, 7th ed. Pearson, 2016, chapters 14 and 15.