Normalization

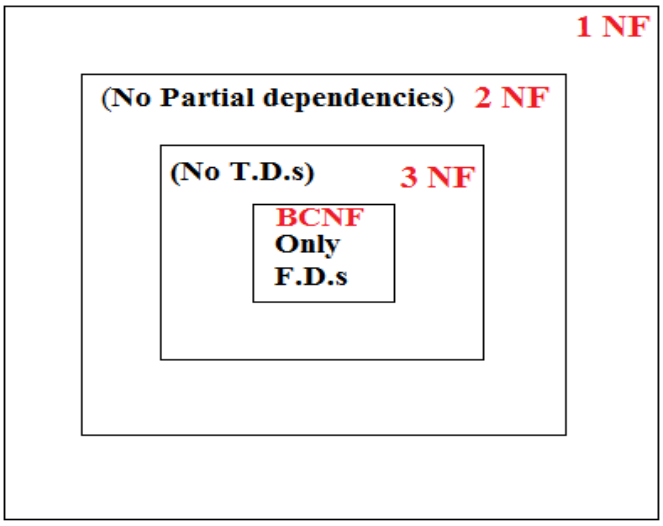
- As one paragraph contains a single idea similarly one table must contain an information about single idea, otherwise we have to repeat one info for other.

Roll no	name	Age	Br_code	Br_name	Br_hod_name
1	A	19	101	Cs	Abc
2	B	18	101	Cs	Abc
3	C	20	101	Cs	Abc
4	D	20	102	Ec	Pqr

Roll no	name	Age	Br_code
1	A	19	101
2	B	18	101
3	C	20	101
4	D	20	102

Br_code	Br_name	Br_hod_name
101	Cs	Abc
102	Ec	Pqr

- **Normalization of data** (Decomposition of Relation) can be considered a process of analyzing the given relation schema to achieve the desirable properties of minimizing redundancy using Decomposition.
- The tool we use for normalization is functional dependencies and candidate keys.
- Functional dependency can be used only to normalize up to BCNF.
- A series of normal form tests that can be carried out on individual relation schemas so that the relational database can be **normalized** to any desired degree.
- 1NF>>>2NF>>3NF>>BCNF



Hierarchy of Normal forms

Q Relational database schema normalization is NOT for: (**NET-AUG-2016)**
(1) reducing the number of joins required to satisfy a query.
(2) eliminating uncontrolled redundancy of data stored in the database.
(3) eliminating number of anomalies that could otherwise occur with inserts and deletes.
(4) ensuring that functional dependencies are enforced.
Ans: a

Q. match the following database terms to their function: **(NET-DEC-2015)**

	List - I		**List - II**
(a)	Normalization	(i)	Enforces match of primary key to foreign key
(b)	Data Dictionary	(ii)	Reduces data redundancy in a database
(c)	Referential Integrity	(iii)	Defines view(s) of the database for particular user(s)
(d)	External Schema	(iv)	Contains metadata describing database structure

Codes :

	(a)	**(b)**	**(c)**	**(d)**
(1)	(iv)	(iii)	(i)	(ii)
(2)	(ii)	(iv)	(i)	(iii)
(3)	(ii)	(iv)	(iii)	(i)
(4)	(iv)	(iii)	(ii)	(i)

Ans. 2

Q Decomposition help in eliminating some of the problems of bad design **(NET-JUNE-2011)**
(A) Redundancy
(B) Inconsistencies
(C) Anomalies
(D) All of the above
Ans: d

FIRST NORMAL FORM

- A Relation table is said to be in first normal form iff each attribute in each cell have single value(atomic). Means a Relation should not contain any multivalued or composite attributes.
- Other implications of first normal form
    - Every row should be unique, that is no two rows should have the same values of all the attributes.
    - There must be a primary key.
    - Every column should have a unique name
    - Order of row and column is irrelevant

Customer			
Customer ID	**First Name**	**Surname**	**Telephone Number**
123	Pooja	Singh	555-861-2025, 192-122-1111
456	San	Zhang	(555) 403-1659 Ext. 53; 182-929-2929
789	John	Doe	555-808-9633

- This table is not in first normal form, as column telephone number, contains multiple value in a single cell.

Solution

- An apparent solution is to introduce more columns:

Customer				
Customer ID	**First Name**	**Surname**	**Telephone Number1**	**Telephone Number2**
123	Pooja	Singh	555-861-2025	192-122-1111
456	San	Zhang	(555) 403-1659 Ext. 53	182-929-2929
789	John	Doe	555-808-9633	

- An arbitrary and hence meaningless ordering has been introduced: why is 555-861-2025 put into the Telephone Number1 column rather than the Telephone Number2 column?
- There's no reason why customers could not have more than two telephone numbers, so how many Telephone Number *N* columns should there be?
- It is not possible to search for a telephone number without searching an arbitrary number of columns.
- Adding an extra telephone number may require the table to be reorganized by the addition of a new column rather than just having a new row (tuple) added.

Designs that comply with 1NF

- To bring the model into the first normal form, we split the strings we used to hold our telephone number information into "atomic" (i.e. indivisible) entities: single phone numbers. And we ensure no row contains more than one phone number.

Customer			
Customer ID	**First Name**	**Surname**	**Telephone Number**
123	Pooja	Singh	555-861-2025
123	Pooja	Singh	192-122-1111
456	San	Zhang	182-929-2929
456	San	Zhang	(555) 403-1659 Ext. 53
789	John	Doe	555-808-9633

- Note that the "ID" is no longer unique in this solution with duplicated customers. To uniquely identify a row, we need to use a combination of (ID, Telephone Number). The value of the combination is unique although each column separately contains repeated values. Being able to uniquely identify a row (tuple) is a requirement of 1NF.

An alternative design uses two tables:

Customer Name		
Customer ID	**First Name**	**Surname**
123	Pooja	Singh
456	San	Zhang
789	John	Doe

Customer ID	Telephone Number
123	555-861-2025
123	192-122-1111
456	(555) 403-1659 Ext. 53
456	182-929-2929
789	555-808-9633

- Using **Customer ID** as key, a *one-to-many* relationship exists between the name and the number tables. A row in the "parent" table, **Customer Name**, can be associated with many telephone numbers rows in the "child" table, **Customer Telephone Number**, but each telephone number belongs to one, and only one customer. It is worth noting that this design meets the additional requirements for second and third normal form.

Q Relations produced from E - R Model will always be in _____. **(NET-JULY-2018)**

(1) 1 NF **(2)** 2 NF **(3)** 3 NF **(4)** 4 NF

Ans (1)

- **Prime attribute: -** A attribute is said to be prime if it is part of any of the candidate key
- **Non-Prime attribute: -** A attribute is said to be non-prime if it is not part of any of the candidate key

E.g. R(ABCD)

AB>CD

Here candidate key is AB so, A and B are prime attribute, C and D are non-prime attributes.

- ***PARTIAL DEPENDENCY*-** When a non –prime attribute is dependent only on a part (Proper subset) of candidate key then it is called partial dependency. (PRIME > NON-PRIME)

- ***TOTAL DEPENDENCY*-** When a non –prime attribute is dependent on the entire candidate key then it is called total dependency.

Q R(ABCD) AB>D, A>C

SECOND NORMAL FORM

Relation R is in 2NF if,

- R should be in 1 NF.
- R should not contain any Partial dependency. (that is every non-key column should be fully dependent upon candidate key)S

Example to see advantage of redundancy

Q R(A, B, C) B>C

A	B	C
a	1	X
b	2	Y
a	3	Z
C	3	Z
D	3	Z
E	3	Z

A	B
A	1
B	2
A	3
C	3
D	3
E	3

B	C
1	X
2	Y
3	z

Even if one some subset become null then also, we can find the desired attribute

Q Consider the following relation R(ABCDEF) with the FD set F = (A → B, C → D, E → F).
Find the 2 NF decomposition for R.
a) (ABC) (CD) (EF) (ABCD)
b) (AB)(CD)(EF)
c) (AB) (CDE)(EF) (ACE)
d) (AB)(CD)(EF)(ACE)

- Every table with two attributes will always be in second normal form.

Q If a relation is in 2NF then: **(NET-JUNE-2008)**
(A) every candidate key is a primary key
(B) every non-prime attribute is fully functionally dependent on each relation key
(C) every attribute is functionally independent
(D) every relational key is a primary key
Ans: B

TRANSITIVE DEPENDENCY

A functional dependency from non-Prime attribute to non-Prime attribute is called transitive
E.g.- R(A, B, C, D) with A as a candidate key
A->B
B->C [transitive dependency]
C->D [transitive dependency]

THIRD NORMAL FORM

Let R be the relational schema, it is said to be in 3 NF

- R should be in 2NF
- It must not contain any transitive dependency

THIRD NORMAL FORM DIRECT DEFINATION-

A relational schema R is said to be 3 NF if every functional *dependency in R from α-->β, either α is super key or β is the prime attribute*

A	B	C
A	1	P
B	2	Q
C	2	Q
D	2	Q
E	3	R
F	3	R
G	4	S

A	B
A	1
B	2
C	2
D	2
E	3
F	3
G	4

B	C
1	P
2	Q
3	R
4	S

Q Which normal form is considered as adequate for usual database design? **(NET-JUNE-2013)**
(A) 2NF **(B)** 3NF **(C)** 4NF **(D)** 5NF
Ans: b

Q Third normal form is based on the concept of _____. **(NET-DEC-2012)**
(A) Closure Dependency **(B)** Transitive Dependency
(C) Normal Dependency **(D)** Functional Dependency
Ans: b

Q If a relation is in 2NF and 3NF forms then: **(NET-DEC-2007)**
(A) no non-prime attribute is functionally dependent on other non-prime attributes
(B) no non-prime attribute is functionally dependent on prime attributes
(C) all attributes are functionally independent
(D) prime attribute is functionally independent of all non-prime attributes
Ans: a

BCNF (BOYCE CODD NORMAL FORM)

A relational schema R is said to be BCNF if every functional *dependency in R from α-->β*

- *α must be a super key*

E.g.- R (A, B, C, D)

{

AB->C [No violation of 2NF,3NF, BCNF]

C->D [No violation of 2NF,3NF, BCNF]

D->A [violation of BCNF, D not a candidate/super key]

} *Candidate key= {AB}, {DB}, {CB}*

R(A, B, C) AB>C, C>B

A	B	C
A	B	B
B	B	C
B	A	D
A	A	E
C	C	B
D	C	B
E	C	B
F	C	B

A	B
A	B
B	B
B	A
A	A
C	C
D	C
e	C
f	c

C	B
B	B
C	B
D	A
E	A

Q (NET-JULY-2019)

In relational databases, if relation R is in BCNF, then which of the following is true about relation R?

1. R is in 4NF
2. R is not in 1NF
3. R is in 2NF and not in 3NF
4. R is in 2NF and 3NF

Ans: 4

Some important note points on Normalization:

- If a relation R does not contain any non- trivial dependency, then R Is in BCNF.
- A Relation with two attributes is always in BCNF.
- A relation schema R consist of only simple candidate key then, R is always in 2NF but may or may not be in 3NF or BCNF.
- A Relation schema R consist of only prime attributes then R is always in 3NF, but may or may not be in BCNF.
- A relation schema R in 3NF and with only simple candidate keys, then R surely in BCNF.

Q R(ABCDEF) (A, BC, DEF) (BCNF) →
A>BCDEF
BC>ADEF
DEF>ABC

……

Q R(ABC)(AB, BC)(3 NF)
AB>C
C>A

Q R(ABCD)(AD, BD, CD)(3 NF)
A>B
B>C
C>A

Q R(ABCD)(AB, BD)(3 NF)
AB>CD
D>A

Q R(ABCDE)(ACD, BCD, CDE)(3 NF)
A>B
BC>E
DE>A

Q R(ABCD)(AB, AD, BC, CD)(3 NF)
AB>CD
C>A
D>B

Q R(ABCDE)(AB, BC, BD)(3 NF)
AB>CD
D>A
BC>DE

Q R(ABCDE)(BC, CD)(3 NF)
BC>ADE
D>B

Q R(ABCDEF)(C, D, AB, BE, BF)(3 NF)
AB>C
C>D
D>BE
E>F
F>A

Q R(WXYZ)(Y, XW, XZ)(3 NF)
Z>W
Y>XZ
XW>Y

Q R(ABCDE)(A, E, BC, CD)(3 NF)
A>BC
CD>E
B>D
E>A

Q (ABCDE)(ACD, BCD, CDE)(3 NF)
A>B
BC>E
DE>A

……..

Q R(ABCDE)(AE)(2 NF)
A>B
B>E
C>D

……

Q R(ABCDE)(ac)(1NF)
A>B

B>E

C>D

Q R(ABCDE)(ab)(1NF)

AB>C

B>D

D>E

Q R(ABCDE)(AB)(1NF)

AB>C

B>D

D>E

Q R(ABCD)(AB)(1 NF)

AB>C

B>D

Q R(ABCDEF)(BF)(1 NF)

AB>C

C>D

B>AE

Q R(ABCDEFGHIJ)(AB)(1 NF)

AB>C

A>DE

B>F

F>GH

D>IJ

Q R(ABCDEFGHIJ)(ABD) (1NF)

AB>C

AD>GH

BD>EF

A>I

H>J

Q R(ABCDE)(CE)(1 NF)
CE>D
D>B
C>A

Q R(ABCDEFGH)(AE)(1 NF)
A>BC
ABE>CDGH
C>GD
D>G
E>F

Q R(ABCDEF)(ABD, BCD)(1 NF)
AB>C
DC>AE
E>F

Q R(VWXYZ)(VW, XW)(1NF)
Z>Y
Y>Z
X>YV
VW>X

Q R(ABCDEF)(ABC, ACD)(1 NF)
ABC>D
ABD>E
CD>F
CDF>B
BF>D

Q R(ABCDE)(ABD)(1 NF)
BD>E
A>C

Q Consider the following four relational schemas. For each schema, all non-trivial functional dependencies are listed. The underlined attributes are the respective primary keys. (GATE-

2020) (2 Marks)

Schema I: Registration (rollno, courses)
Field ‘courses’ is a set-valued attribute containing the set of courses a student has registered for.
Non-trivial functional dependency
rollno → courses

Schema II: Registration (rollno, coursid, email)
Non-trivial functional dependencies:
rollno, courseid → email
email → rollno

Schema III: Registration (rollno, courseid, marks, grade)
Non-trivial functional dependencies:
rollno, courseid, → marks, grade
marks → grade

Schema IV: Registration (rollno, courseid, credit)
Non-trivial functional dependencies:
rollno, courseid → credit
courseid → credit

Which one of the relational schemas above is in 3NF but not in BCNF?
(a) Schema 1 **(b)** Schema 2 **(c)** Schema 3 **(d)** Schema 4
Ans: b

Q Which one of the following statements if FALSE? **(GATE- 2017) (1 Marks)**
a) Any relation with two attributes is in BCNF
b) A relation in which every key has only one attribute is in 2NF
c) A prime attribute can be transitively dependent on a key in a 3NF relation
d) A prime attribute can be transitively dependent on a key in a BCNF relation
ANSWER D

Q A database of research articles in a journal uses the following schema. **(GATE- 2016) (2 Marks)**

(VOLUME, NUMBER, STARTPAGE, ENDPAGE, TITLE, YEAR, PRICE)
The primary key is (VOLUME, NUMBER, STARTPAGE, ENDPAGE) and the following functional dependencies exist in the schema.

(VOLUME, NUMBER, STARTPAGE, ENDPAGE) → TITLE
(VOLUME, NUMBER) → YEAR
(VOLUME, NUMBER, STARTPAGE, ENDPAGE) → PRICE

The database is redesigned to use the following schemas.

(VOLUME, NUMBER, STARTPAGE, ENDPAGE, TITLE, PRICE)
(VOLUME, NUMBER, YEAR) Which is the weakest normal form that the new database satisfies, but the old one does not

(1) 1NF **(2)** 2NF **(3)** 3NF **(4)** BCNF

Ans: 2

Q If every non-key attribute is functionally dependent on the primary key, then the relation is in _____. **(NET-NOV-2017)**

(1) First normal form
(2) Second normal form
(3) Third normal form
(4) Fourth normal form

Ans: (2)

Q In RDBMS, different classes of relations are created using _____ technique to prevent modification anomalies. **(NET-NOV-2017)**

(1) Functional Dependencies
(2) Data integrity
(3) Referential integrity
(4) Normal Forms

Ans: 4

Q For a database relation R(a, b, c, d) where the domains of a, b, c and d include only atomic values, and only the following functional dependencies and those that can be inferred from them hold : a → c b → d The relation is in _____.**(NET-JULY-2017)**

(1) First normal form but not in second normal form
(2) Second normal form but not in third normal form
(3) Third normal form
(4) BCNF

Ans: a

Q For a database relation R(A, B, C, D) where the domains of A, B, C and D include only atomic values, only the following functional dependencies and those that can be inferred from them are : A → C B → D The relation R is in _____. **(NET-JAN-2017)**

(1) First normal form but not in second normal form.

(2) Both in first normal form as well as in second normal form.
(3) Second normal form but not in third normal form.
(4) Both in second normal form as well as in third normal form.
Ans: a

Q Consider a relation R (A, B, C, D, E, F, G, H), where each attribute is atomic, and following functional dependencies exist. **(NET-NOV-2017)**
CH → G
A → BC
B → CFH
E → A
F → EG
The relation R is _____ .
(1) in 1NF but not in 2NF
(2) in 2NF but not in 3NF
(3) in 3NF but not in BCNF
(4) in BCNF
Ans: 1

Q Which of the following statements is false? **(NET-DEC-2014)**
(A) Any relation with two attributes is in BCNF.
(B) A relation in which every key has only one attribute is in 2NF.
(C) A prime attribute can be transitively dependent on a key in 3NF relation.
(D) A prime attribute can be transitively dependent on a key in BCNF relation.
Ans: d

Q The best normal form of relation scheme R(A, B, C, D) along with the set of functional dependencies F = {AB → C, AB → D, C → A, D → B} is **(NET-DEC-2014)**
(A) Boyce-Codd Normal form
(B) Third Normal form
(C) Second Normal form
(D) First Normal form
Ans: b

Q Given the following two statements:
S1: Every table with two single-valued attributes is in 1NF, 2NF, 3NF and BCNF.
S2: AB->C, D->E, E->C is a minimal cover for the set of functional dependencies AB->C, D->E, AB->E, E->C.
Which one of the following is CORRECT? **(GATE- 2014) (1 Marks)**
A) S1 is TRUE and S2 is FALSE.
B) Both S1 and S2 are TRUE.
C) S1 is FALSE and S2 is TRUE.
D) Both S1 and S2 are FALSE.
Ans: a

Q Which of the following is TRUE? **(GATE- 2012) (1 Marks)**

a) Every relation in 3NF is also in BCNF

b) A relation R is in 3NF if every non-prime attribute of R is fully functionally dependent on every key of R

c) Every relation in BCNF is also in 3NF

d) No relation can be in both BCNF and 3NF

Ans: c

Q For a database relation R(a, b, c, d) where the domains of a, b, c, d include only the atomic values. The functional dependency a → c, b → d holds in the following relation **(NET-JUNE-2013)**

(A) In 1NF not in 2NF
(B) In 2NF not in 3NF
(C) In 3NF
(D) In 1NF

Ans: a

Q Which of the following is true? **(NET-DEC-2012)**

(A) A relation in BCNF is always in 3NF.
(B) A relation in 3NF is always in BCNF.
(C) BCNF and 3NF are same.
(D) A relation in BCNF is not in 3NF.

Ans: a

Q A function that has no partial functional dependencies is in ……. form. **(NET-DEC-2009)**

(A) 3 NF **(B)** 2 NF **(C)** 4 NF **(D)** BCNF

Ans: b

Q Which of the following is true? **(NET-JUNE-2008)**

(A) A relation in 3NF is always in BCNF
(B) A relation in BCNF is always in 3NF
(C) BCNF and 3NF are totally different
(D) A relation in BCNF is in 2NF but not in 3NF

Ans: b

Q Relation R with an associated set of functional dependencies, F is decomposed into BCNF. The redundancy (arising out of functional dependencies) in the resulting set relations is. **(GATE- 2002) (1 Marks)**

a) Zero

b) More than zero but less than that of an equivalent 3NF decomposition

c) Proportional to the size of F+
d) Indeterminate
Ans: a

Relation R has eight attributes ABCDEFGH. Fields of R contain only atomic values.

F = {CH→G, A→BC, B→CFH, E→A, F→EG} is a set of functional dependencies (FDs) so that F^+ is exactly the set of FDs that hold for R.

The relation R is
(1) in 1NF, but not in 2NF
(2) in 2NF, but not in 3NF
(3) in 3NF, but not in BCNF
(4) in BCNF
Ans: 1

Q Consider the relation schema R (A B C D) with following FD set
F = {A → BC, C → D}; The relation R is in _____
a) 1 NF **b)** 2 NF **c)** 3 NF **d)** BCNF
Ans b

Q Consider the relation R (ABCDE) with the FD set F = {A → CE, B → D, AE → D}. Identify the highest normal form satisfied by the relation R.
a) 1 NF **b)** 2 NF **c)** 3 NF **d)** BCNF

Q Consider the following relational schema:
Suppliers (Sid: integer, sname: string, city: string, street: string)
Parts (pid: integer, pname: string, color: string)
Catalog (sid: integer, pid: integer, cost: real)

Assume that, in the supplier's relation above, each supplier and each street within a city has a unique name, and (sname, city) forms a candidate key. No other functional dependencies are implied other than those implied by primary and candidate keys. Which one of the following is TRUE about the above schema? **(GATE- 2009) (1 Marks)**
a) The schema is in BCNF
b) The schema is in 3NF but not in BCNF
c) The schema is in 2NF but not in 3NF
d) The schema is not in 2NF
Ans: a

Q Consider the following relational schemes for a library database
Book (Title, Author, Catalog_no, Publisher, Year, Price)
Collection (Title, Author, Catalog_no)
with in the following functional dependencies:
I. Title Author --> Catalog_no
II. Catalog_no --> Title Author Publisher Year
III. Publisher Title Year --> Price
Assume {Author, Title} is the key for both schemes. Which of the following statements is true? **(GATE- 2008) (1 Marks)** **(NET-JUNE-2014)**
(A) Both Book and Collection are in BCNF
(B) Both Book and Collection are in 3NF only
(C) Book is in 2NF and Collection is in 3NF
(D) Both Book and Collection are in 2NF only
Answer: (C)

Q The relation scheme Student Performance (name, courseNo, rollNo, grade) has the following functional dependencies:
name, courseNo → grade
rollNo, courseNo → grade
name → rollNo
rollNo → name
The highest normal form of this relation scheme is **(GATE- 2004) (1 Marks)**
a) 2 NF **b) 3 NF** **c) BCNF** **d) 4 NF**
Ans: b

Q Consider the following functional dependencies in a database **(GATE- 2003) (1 Marks)**
Data_of_Birth → Age
Age → Eligibility
Name → Roll_number
Roll_number → Name
Course_number → Course_name
Course_number → Instructor
(Roll_number, Course_number) → Grade
The relation (Roll_number, Name, Date_of_birth, Age) is:
(A) In second normal form but not in third normal form
(B) In third normal form but not in BCNF
(C) In BCNF

(D) None of the above

Answer: (D)

Q R (W, X, Y, Z) F.D. {Z → W, Y → XZ, XW → Y} the normal form of R is _____

Q The Relation Vendor Order (V_no, V_ord_no, V_name, Qty_sup, unit_price) is in 2NF because **(NET-JUNE-2015)**

(1) Non_key attribute V_name is dependent on V_no which is part of composite key

(2) Non_key attribute V_name is dependent on Qty_sup

(3) Key attribute Qty_sup is dependent on primary_key unit price

(4) Key attribute V_ord_no is dependent on primary_key unit price

Ans. 1

"The Relation Vendor Order (V_no, V_ord_no, V_name, Qty_sup, unit_price) is in 2NF because: Non_key attribute V_name is dependent on V_no which is part of composite key."

Q Given the following two statements:

S1: Every table with two single-valued attributes is in 1NF, 2NF, 3NF and BCNF.

S2: AB->C, D->E, E->C is a minimal cover for the set of functional dependencies AB->C, D->E, AB->E, E->C.

Which one of the following is CORRECT? (GATE – 2014) (2 Marks)

(a) S1 is TRUE and S2 is FALSE

(b) Both S1 and S2 are TRUE

(c) S1 is FALSE and S2 is TRUE

(d) Both S1 and S2 are FALSE

Ans: a