Database Normalization

Dr. Kamal Gulati

Database Normalization

- The main goal of Database Normalization is to restructure the **logical data model** of a database to:
  - Eliminate Redundancy
  - Organize Data Efficiently
  - Reduce the potential for Data Anomalies.

Redundancy

- **Remove Unnecessary data,**
- **Remove Duplicate data,**
- **Remove surplus data and**
- **Remove unneeded data**

Duplication Vs Redundant Data

- Duplicated Data: When an attribute has two or more identical values
- Redundant Data: If you can delete data with a loss of information

Data Efficiently

- Data should be Competently,
- Data should be Resourcefully and
- Data should be Satisfactory.

Data Anomalies

- Data anomalies are inconsistencies in the data stored in a database as a result of an operation such as update, insertion, and/or deletion.
- Such inconsistencies may arise when have a **particular record stored in multiple locations and not all of the copies are updated**.
- We can prevent such anomalies by implementing **7 different level of** normalization called Normal Forms (NF)

Different Type of Normalization

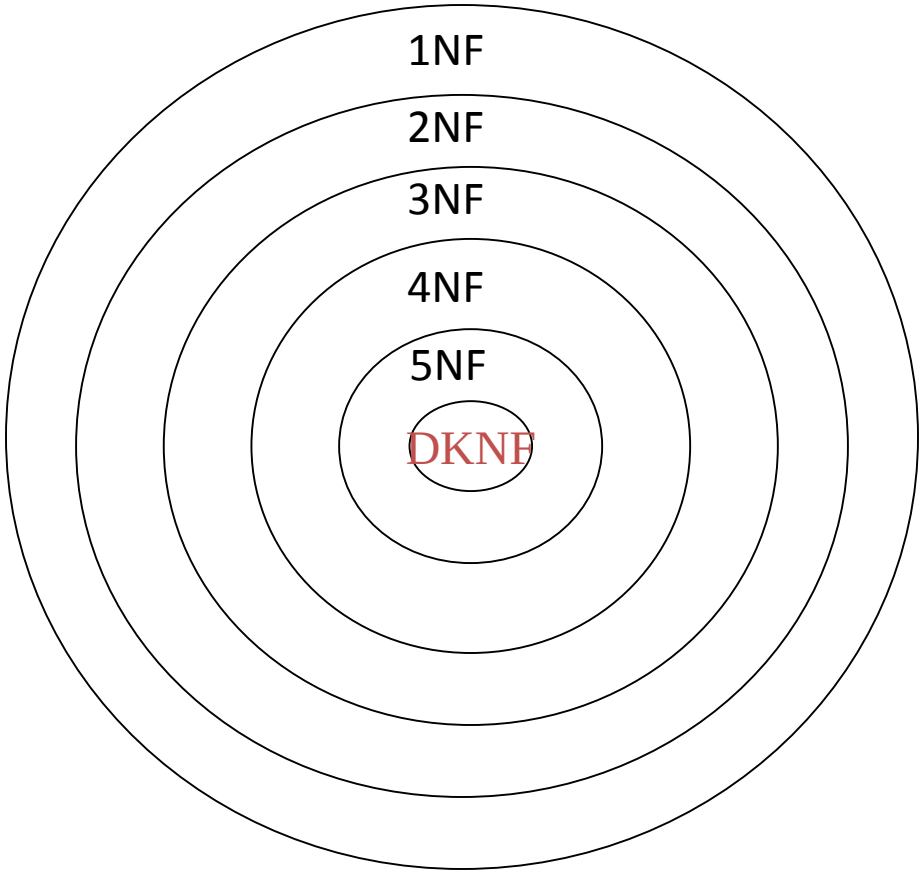
- In database design, we start with one single table, with all possible columns. A lot of redundant data would be present since it's a single table.
- **The process of removing the redundant data, by splitting up the table in a well defined fashion is called normalization.**

Levels of Normalization

- **1. First Normal Form (1NF)**
- **2. Second Normal Form (2NF)**
- **3. Third Normal Form (3NF)**
- **4. Boyce-Codd Normal Form (BCNF)**
- **5. Fourth Normal Form (4NF)**
- **6. Fifth Normal Form (5NF)**
- **7. Domain Key Normal Form (DKNF)**

Most databases should be 3NF or BCNF in order to avoid the database anomalies.

Levels of Normalization



Each higher level is a subset of the lower level

Brief History / Overview

- Database Normalization was first proposed by **Edgar F. Codd**.
- Codd defined the first three Normal Forms, which we'll look into, of the 7 known Normal Forms.
- In order to do normalization we must know what the requirements are for each of the three Normal Forms that we'll go over.
- One of the key requirements to remember is that Normal Forms are progressive. That is, in order to have 3rd NF we must have 2nd NF and in order to have 2nd NF we must have 1st NF.

First Normal Form (1NF)

Dr. Kamal Gulati

1. First Normal Form (1NF)

- A table is in its first normal form if it contains no repeating attributes or groups of attributes

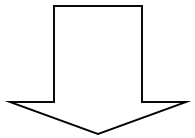
Non-Normalised Table

STUDENT

Number	Name	Classes
001231	William Hartnell	Information Systems, Systems Analysis, Data Communications
001232	Patrick Troughton	Systems Analysis, Data Communications
001233	Jon Pertwee	OO Programming, Systems Analysis, Data Communications
001234	Tom Baker	Systems Analysis, Data Communications

First Normal Form

- To convert data for unnormalised form to 1NF, simply convert any repeated attributes into part of the candidate key
- STUDENT(Number, Name, Classes)



- STUDENT(Number, Name, Classes)

First Normal Form

STUDENT

Number	Name	Classes
001231	William Hartnell	Information Systems
001231	William Hartnell	Systems Analysis
001231	William Hartnell	Data Communications
001232	Patrick Troughton	Systems Analysis
001232	Patrick Troughton	Data Communications
001233	Jon Pertwee	OO Programming
001233	Jon Pertwee	Systems Analysis
001233	Jon Pertwee	Data Communications
001234	Tom Baker	Systems Analysis
001234	Tom Baker	Data Communications

1. First Normal Form (1NF)

- The requirements to satisfy the 1st NF:
  - Each table has a **primary key**: minimal set of attributes which can uniquely identify a record
  - The values in **each column of a table are atomic** (No multi-value attributes allowed).
  - There are **no repeating groups**: two columns do not store similar information in the same table.

- A relation is said to be in first normal form if and only if all underlying domains contain **atomic values only**.
- After 1NF, we can still have redundant data.

Over to you...

RefNo	Name	Address	Status	AccNo
345	C.J. Date	23, High Street	Business	120768, 348973
543	F.D. Rolland	45, The Ash	Domestic	987654
675	D.R. Howe	17, Low Street	Business	745363, 678453, 348973

Table in 1NF

RefNo	Name	Address	Status	AccNo
345	C.J. Date	23, High Street	Business	120768
345	C.J. Date	23, High Street	Business	348973
543	F.D. Rolland	45, The Ash	Domestic	987654
675	D.R. Howe	17, Low Street	Business	745363
675	D.R. Howe	17, Low Street	Business	678453
675	D.R. Howe	17, Low Street	Business	348973

Second Normal Form (2NF)

Dr. Kamal Gulati

2NF

A table is in the second normal form if it's in the first normal form AND no column that is not part of the primary key is dependant only a portion of the primary key.

Functional Dependency

- The concept of functional dependency in central to normalisation and, in particular, strongly related to 2NF.

Functional Dependency

- If ‘X’ is a set of attributes within a relation, then we say ‘A’ (an attribute or set of attributes), is functionally dependant on X, if and only if, for every combination of X, there is only one corresponding value of A
- We write this as :

X -> A

Table in 1NF

RefNo	Name	Address	Status	AccNo
345	C.J. Date	23, High Street	Business	120768
345	C.J. Date	23, High Street	Business	348973
543	F.D. Rolland	45, The Ash	Domestic	987654
675	D.R. Howe	17, Low Street	Business	745363
675	D.R. Howe	17, Low Street	Business	678453
675	D.R. Howe	17, Low Street	Business	348973

Functional Dependency

- It is clear that :

  RefNo -> Name, Address, Status

or, most correctly,

AccNo, RefNo -> Name, Address, Status

Second Normal Form

RefNo	AccNo
345	120768
345	348973
543	987654
675	745363
675	678453
675	348973

RefNo	Name	Address	Status
345	C.J. Date	23, High Street	Business
543	F.D. Rolland	45, The Ash	Domestic
675	D.R. Howe	17, Low Street	Business

2. Second Normal Form (2NF)

- All requirements for 1st NF must be met.
- Redundant data across multiple rows of a table must be moved to a separate table.
  - The resulting tables must be related to each other by use of foreign key.

- A relation is said to be in 2NF if and only if it is in 1NF and every non key attribute is fully dependent on the primary key.
- After 2NF, we can still have redundant data.

Over to you...

Supplier#	Part#	City	Quantity
S1	P1	London	1000
S1	P2	London	1500
S1	P3	London	3400
S1	P4	London	2100
S2	P2	Paris	3400
S2	P3	Paris	1000
S4	P1	Nuku alofa	5
S4	P4	Nuku alofa	7

Table in Second Normal Form

Supplier#	City
S1	London
S2	Paris
S4	Nuku alofa


Supplier#	Part#	Quantity
S1	P1	1000
S1	P2	1500
S1	P3	3400
S1	P4	2100
S2	P2	3400
S2	P3	1000
S4	P1	5
S4	P4	7

Third Normal Form (3NF)

Dr. Kamal Gulati

A table is in the third normal form if it is the second normal form and there are no *non-key columns* dependant on *other non-key columns* that could not act as the primary key.

Table in Second Normal Form

Branch Name	Address	Manager No	Acc No	Balance	Type
Rathmines	15 Upr Rathmimes Road	1234	1205	-£123.45	‘D’
Rathmines	15 Upr Rathmimes Road	1234	6784	£67.54	‘C’
Rathmines	15 Upr Rathmimes Road	1234	9843	£43.43	‘C’
Dame St.	1 Dame Street	1101	5422	£34.50	‘C’
Dame St.	1 Dame Street	1101	0998	£666.66	‘D’

Table in Third Normal Form

Acc No 	Balance	Type	Branch Name
1205	-£123.45	‘D’	Rathmines
6784	£67.54	‘C’	Rathmines
9843	£43.43	‘C’	Rathmines
5422	£34.50	‘C’	Dame St.
0998	£666.66	‘D’	Dame St.

Branch Name 	Address	Manager No
Rathmines	15 Upr Rathmimes Road	1234
Dame St.	1 Dame Street	1101

3. Third Normal Form (3NF)

- The requirements to satisfy the 3rd NF:
  - All requirements for 2nd NF must be met.
  - Eliminate fields that do not depend on the primary key;
    - That is, any field that is dependent not only on the primary key but also on another field must be moved to another table.

- A relation is said to be in 3NF, if and only if it is in 2NF and every non key attribute is non-transitively dependent on the primary key.

Boyce-Codd Normal Form (BCNF)

Boyce-Codd Normal Form

- **All attributes in a relation should be dependant upon the key, the whole key and nothing but the key**

Table in Third Normal Form

CourseNo	Lecturer	Time	Room
FT225/2	P. O'Byrne	9.00	121
FT222/1	D. Gordon	10.00	666
DT266/2	K. O'Brien	1.00	343
DT266/1	D. Carroll	11.00	876
FT222/4	K. O'Brien	3.00	343
FT228/3	D. Gordon	4.00	666

On the Above table:

1. **Identify Primary Key**
2. **Identify Composite Key**
3. **Identify Candidate Key**
4. **Which Columns are having repetition of data?**

- **Primary Key**:
  - 1. CourseNo

- **Composite Key:**
  - 1. Lecturer + Time
  - 2. Time + Room

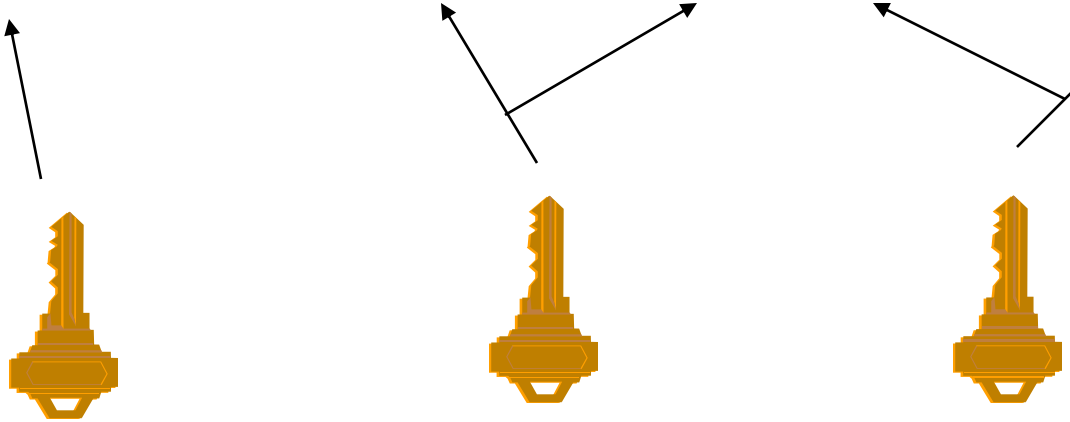
- **Candidate Key:**
  - 1. CourseNo
  - 2. Time

Columns are having repetition of data

- Lecturer
- Room

Table in Third Normal Form

CourseNo	Lecturer	Time	Room
FT225/2	P. O'Byrne	9.00	121
FT222/1	D. Gordon	10.00	666
DT266/2	K. O'Brien	1.00	343
DT266/1	D. Carroll	11.00	876
FT222/4	K. O'Brien	3.00	343
FT228/3	D. Gordon	4.00	666



Redundancy in 3NF

- The combination of ROOM, TIME is unique to each tuple, no room is used twice at the same time (thus it is in 3NF).

- But, we know there is a redundancy in that ROOM depends LECTURER, therefore, we split the table...

Tables in BCNF

CourseNo	Lecturer	Time
FT225/2	P. O'Byrne	9.00
FT222/1	D. Gordon	10.00
DT266/2	K. O'Brien	1.00
DT266/1	D. Carroll	11.00
FT222/4	K. O'Brien	3.00
FT228/3	D. Gordon	4.00

Lecturer	Room
P. O'Byrne	121
D. Gordon	666
K. O'Brien	343
D. Carroll	876

Difference between BCNF and 3NF

- Most relations in 3NF are also in BCNF, the only time this may not be true is when there is **more than one candidate key for a relation and at least one of is composite**.

Fourth Normal Form

Fourth normal form (or 4NF) requires that there be no non-trivial multivalued dependencies of attribute sets on something other than a superset of a candidate key. A table is said to be in 4NF if and only if it is in the BCNF and multivalued dependencies are functional dependencies. The 4NF removes unwanted data structures: multivalued dependencies.

Fifth Normal Form

Fifth normal form (5NF and also PJ/NF) requires that there are no non-trivial join dependencies that do not follow from the key constraints. A table is said to be in the 5NF if and only if it is in 4NF and every join dependency in it is implied by the candidate keys.

Fifth normal form is satisfied when all tables are broken into as many tables as possible in order to avoid redundancy. Once it is in fifth normal form it cannot be broken into smaller relations without changing the facts or the meaning.

Domain/Key Normal Form

Domain/key normal form (or DKNF) requires that the database contains no constraints other than domain constraints and key constraints.

The relation is in DKNF when there can be no insertion or deletion anomalies in the database.

Conclusion for Normalization

- We have seen how Database Normalization can decrease redundancy, increase efficiency and reduce anomalies by implementing three of seven different levels of normalization called Normal Forms.
- The first three NF's are usually sufficient for most small to medium size applications.

Class Exercise

- Do the Normalization (Up to maximum levels) for the following database of the children in a class, and the pets they have.

Name	Age	Pet Type & Pet Name
Heather	10	Dog - Rex Cat - Thomas
Rachel	10	Cat - Fluff
Jimmy	11	Dog - Kimba
Mike	10	Cat - Thomas

First Normal Form

- **Student_Pet Table**

Name	Age	Pet Type	Pet Name
Heather	10	Dog	Rex
Heather	10	Cat	Thomas
Rachel	10	Cat	Fluff
Jimmy	11	Dog	Kimba
Mike	10	Cat	Thomas

Further Normalizations …..

Second Normal Form

- Pets Table

Pet Type	Pet Name	Name
Dog	Rex	Heather
Cat	Thomas	Heather
Cat	Fluff	Rachel
Dog	Kimba	Jimmy
Cat	Thomas	Mike

- Students Table

Name	Age
Heather	10
Rachel	10
Jimmy	11
Mike	10

Third Normal Form

- Pets Table

P_ID	Pet Type	Pet Name	S_ID
P1	Dog	Rex	S1
P2	Cat	Thomas	S1
P3	Cat	Fluff	S2
P4	Dog	Kimba	S3
P5	Cat	Thomas	S4

- Students Table

S_ID	Name	Age
S1	Heather	10
S2	Rachel	10
S3	Jimmy	11
S4	Mike	10

Fourth Normal Form

- Pets Table

P_ID	Pet Type	Pet Name
P1	Dog	Rex
P2	Cat	Thomas
P3	Cat	Fluff
P4	Dog	Kimba

- Students Table

S_ID	Name	Age
S1	Heather	10
S2	Rachel	10
S3	Jimmy	11
S4	Mike	10

- Students_Pets Table

P_ID	S_ID
P1	S1
P2	S1
P3	S2
P4	S3
P2	S4

Dr. Kamal Gulati

(Hope you are able to understand Database Normalization)

For More Questions / Queries Feel Free to Contact me.

Dr. Kamal Gulati

Dr. Kamal Gulati

Associate Professor |
University Quality Support Head
Mentoring Programme Coordinator &
Exam Superintendent |

[Ph. D., M.Sc. (Computer Science), M.C.A., M.B.A]

Professional Certifications:

- **Certified Microsoft Innovative Educator**
- **Data Science 101 Certification from *Big Data University***
- **R Language 101 Certification from *Big Data University***
- **SQL Certification from *SOLOLEARN.com***
- **Certified IBM Big Data 101 from *Big Data University***
- **R Program & Python Certified from *DataCamp***
- **Wiley Certified Big Data Analyst [WCBDA]**
- **Certification on DBMS from *IIT Mumbai***
- **Certified Cisco Certified Network Associate [CCNA]**
- **Certified Microsoft Certified Professional [MCP]**
- **Certified Brainbench in *(MS Access, MS Project, MySQL 5.7 Administration, Computer Fundamentals, Advanced Ms. Excel & Windows OS)***
- **Real-time Advertising Fundamentals** Certified from RTA Academy

Profile of Dr. Kamal Gulati

- **Worked as Visiting Professor with Stratford University, USA for six months from Jan'2016 to June'2016.**

- **Also worked at Bahrain University in Kingdom of Bahrain Sr. I.T. Faculty (Computer Science Department) for Period of 2 Years.**

- **Have rich experience in the field of teaching and research in Computer Science and Information Technology for almost 15+ years in Academia.**

- **Having experience of working with both private and public institutions and universities as the lecturer and self-instruction material writer for Information Technology courses.**

- **Had number of research papers published in national and international journals and conference proceedings in IEEE and Scopus Index.**

- **Also chaired various National and International Conferences of repute and associated with various International Journals as Editorial Board Member for International and National, Academic Adviser and Research Paper Reviewer.**

- **My current area of interest: Big Data Analytics, R Software, Internet & Web Technology, IT Project Management, Decision Support System, Business Analytics, Management Information System, Database Management System, Data Networking, R Software and Advanced Excel with Visual Basic Macros.**

- **Country Visited: USA, Canada, UAE, Bahrain, Oman (Mostly for Teaching and Research Purpose)**

Profile Contd….

- **Technical Program Committee for International Conference on Data, Engineering and Applications 2017 (IDEA-2k17) which would be on October 28-29, 2017 at Bhopal. http://www.ideaconference.in**
- **Advisory Board Committee Member for International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) which would be on 1-2 August 2017 at SKR Engineering College, Poonamallee, Tamil Nadu, India. http://ecds.org.in**
- **Advisory Committee Member for International Conference on Innovative Research in Engineering and Science which would be on 16-17 June 2017 at Asian Institute of Technology Conference Center Thailand. http://www.iresconf.org**
- **Advisory Committee Member for International Conference on Cloud Computing and Internet of everything which held on 10-11 Feb'2016 at Delhi-NCR. http://www.ccioet.org**
- **Technical Committee member for InCITe-2016 (International Conference on Information Technology) Theme - Internet of Things: Connect Your Worlds, IT Summit, Amity University 2016 which held on 06-07 Oct, 2016. http://www.amity.edu/incite2016**
- **Technical Speaker for Global perspective on IT business "The Changing Scenario" – Big Data on International Students Conference New Delhi (ISCND) which held on 14-15 Oct, 2016 http://iscnd.com**
- **Advisory Committee Member for International Conference on Sustainable Computing Techniques in Engineering, Science and Management which has held on 09-10 Sep'2016 at Delhi-NCR. http://www.scesm.org**
- **Technical Program Committee Member for Program Committee Member for International Conference on Recent Trends IN ICT, Jaipur, India, Aug 18-19, 2017 http://rtict.org**
- **Program Committee Member for International Conference on Recent Advancement in Computer and Communication Bhopal, India, (IC-RAC-2017) May 26-27, 2017 http://www.icrac.co.in**

Profile Contd….

- **Editorial Board member for the following International Journals:**
  - **International Journal of Computer Science and Innovation http://www.infinitysciences.org**
  - **International Journal of Latest Research in Engineering and Technology http://www.ijlret.com**
  - **International Journal of Latest Trends in Engineering and Technology http://www.ijltet.org**
  - **International Journal of Application or Innovation in Engineering & Management http://www.ijaiem.org**
  - **International Journal for Management http://www.ijm-apm.com**
  - **The International Journal of Emerging Engineering and Embedded Systems http://www.ijeees.org**
  - **Conference Info http://conferenceinfo.org/tpc.php**

- **Expert Speaker for Program "Insurance Beyond Doubt" Presented by Oriental Insurance Co Ltd. https://www.youtube.com/watch?v=GrvJkN_Zn3Q**

BOOK, CHAPTER, and CASE STUDY Published

- Published Book on **“A Study of Changing Trends in E-CRM of Indian Insurance Industry”** Published by LAP Lambert Academic Publishing, one of the top researchers and renowned scientists of Germany with ISBN: 3330009543, 9783330009547. The Book available at Amazon.com.

- Published Real Case Study on **“IoT Security Considerations for Higher Education”** published on Business Cases - RENVOI 2017 BOOK (The Case Centre, UK) with ISBN: 978-1-4828-8840-9, Page 63-70. The Book available at the various online website: Amazon, AbeBooks, Chegg, Barnes & Noble.

- Published Chapter on **"Role of eWorld in Insurance Innovation"** Published by Insurance Institute of India (III), 60 Years Diamond Jubilee, Compendium, Nov 2016 – (Magazine) – One of the premium Insurance Institute of India.

Contact Details

- Email : ***drkamalgulati@gmail.com***
- Website : ***http://mybigdataanalytics.com***
- LinkedIn : **https://in.linkedin.com/in/drkamalgulati**
- **Twitter : @DrKamalGulati**
- **Skype : kamalgulati7**
- **Phone: +91-9718361928**