VINSYS

DELiVERiNGSKiLLS
DRiViNGSUCCESS

www.VINSYS.COM

NORMALIZATION

VINSYS

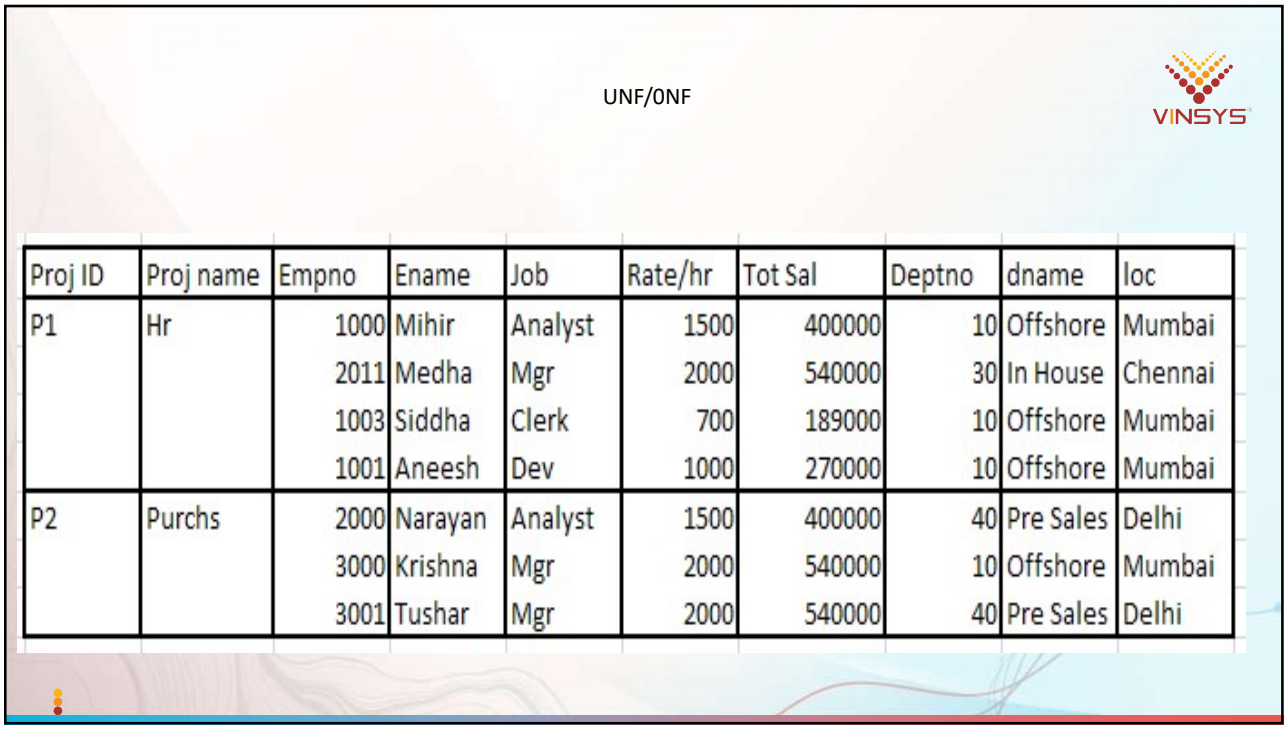
Normalization

- Normalization is the process of efficiently organizing data in a database with two goals in mind
- First goal: eliminate redundant data
  - for example, storing the same data in more than one table
- Second Goal: eliminate Insert,Update,Delete anomalies

Benefits of Normalization

- Less storage space
- Quicker updates
- Less data inconsistency
- Clearer data relationships
- Easier to add data
- Flexible Structure

UNF/0NF



Proj ID	Proj name	Empno	Ename	Job	Rate/hr	Tot Sal	Deptno	dname	loc
P1	Hr	1000	Mihir	Analyst	1500	400000	10	Offshore	Mumbai
		2011	Medha	Mgr	2000	540000	30	In House	Chennai
		1003	Siddha	Clerk	700	189000	10	Offshore	Mumbai
		1001	Aneesh	Dev	1000	270000	10	Offshore	Mumbai
P2	Purchs	2000	Narayan	Analyst	1500	400000	40	Pre Sales	Delhi
		3000	Krishna	Mgr	2000	540000	10	Offshore	Mumbai
		3001	Tushar	Mgr	2000	540000	40	Pre Sales	Delhi

Need for NORMALIZATION Contd..

- The previous slide is said to be in 'UN_NORMAL FORM(UNF)' because:-
  - a) There are repeating groups in the slide-
    - For each Project there is multiple employee's & Department's info.
    - This is against RDBMS according to which each relation(Table) must contain an 'atomic' value

Need for NORMALIZATION Contd..

b) Insert,Update,Delete Anomalies:-

- Insert :- If we want to insert a new Dept information/Job Information, we can't do without adding Project compulsorily.
- Update :- If information about Project gets changed/dept's Information gets changed, changes have to be made all over the table causing redundancy . Even if we forget to make changes at one place ,our data is going to be in the inconsistent state
- Delete :- If Project is removed employees as well as Department's Information is going.

  If Employee resigns Or if Department is going,then Project's info gets removed.

Need for NORMALIZATION

- To remove all these problems from the current table we need to Normalize it by separating data into multiple tables & by following different rules of Normalization to avoid previously specified problems.

1NF

Rule :-

- All attribute values are atomic.
- No repeating group, no composite attributes

1NF

Proj ID	Proj name	Empno	Ename	Job	Rate/hr	Tot Sal	Deptno	dname	loc
P1	Hr	1000	Mihir	Analyst	1500	400000	10	Offshore	Mumbai
P1	Hr	2011	Medha	Mgr	2000	540000	30	In House	Chennai
P1	Hr	1003	Siddha	Clerk	700	189000	10	Offshore	Mumbai
P1	Hr	1001	Aneesh	Dev	1000	270000	10	Offshore	Mumbai
P2	Purchs	2000	Narayan	Analyst	1500	400000	40	Pre Sales	Delhi
P2	Purchs	3000	Krishna	Mgr	2000	540000	10	Offshore	Mumbai
P2	Purchs	3001	Tushar	Mgr	2000	540000	40	Pre Sales	Delhi

2 Normal Form(2NF)

- A Relation is in 2NF , only if it is in 1NF, &
- Second normal form (2NF) further addresses the concept of removing duplicative data.
- Here we talk about **'Functional Dependency'**.
- Each value of attribute A is associated with only one value of attribute B.
- All non-prime attributes are fully dependent on the candidate keys. Which is creating relationships between these new tables and their predecessors through the use of foreign keys.

2NF

Proj ID	Proj name	Empno	Ename	Job	Rate/hr	Tot Sal	Deptno	dname	loc
P1	Hr	1000	Mihir	Analyst	1500	400000	10	Offshore	Mumbai
P1	Hr	2011	Medha	Mgr	2000	540000	30	In House	Chennai
P1	Hr	1003	Siddha	Clerk	700	189000	10	Offshore	Mumbai
P1	Hr	1001	Aneesh	Dev	1000	270000	10	Offshore	Mumbai
P2	Purchs	2000	Narayan	Analyst	1500	400000	40	Pre Sales	Delhi
P2	Purchs	3000	Krishna	Mgr	2000	540000	10	Offshore	Mumbai
P2	Purchs	3001	Tushar	Mgr	2000	540000	40	Pre Sales	Delhi

Proj ID	Proj name
P1	Hr
P2	Purchs

Empno	Ename	Job	Rate/hr	Tot Sal	Proj ID	Deptno
1000	Mihir	Analyst	1500	400000	P1	10
2011	Medha	Mgr	2000	540000	P1	30
1003	Siddha	Clerk	700	189000	P1	10
1001	Aneesh	Dev	1000	270000	P1	10
2000	Narayan	Analyst	1500	400000	P2	40
3000	Krishna	Mgr	2000	540000	P2	10
3001	Tushar	Mgr	2000	540000	P2	40

Deptno	dname	loc
10	Offshore	Mumbai
30	In House	Chennai
40	Pre Sales	Delhi

3NF

- No **transitive dependency**
- A relation is said to be in 3NF if it is in 2NF &
- All (non-key) attributes must, and only, be functionally dependent on **Key Attribute**
- A → B and B → C, So A → C

Exploring 2NF

Proj ID	Proj name	Empno	Ename	Job	Rate/hr	Tot Sal	Deptno	dname	loc
P1	Hr	1000	Mihir	Analyst	1500	400000	10	Offshore	Mumbai
P1	Hr	2011	Medha	Mgr	2000	540000	30	In House	Chennai
P1	Hr	1003	Siddha	Clerk	700	189000	10	Offshore	Mumbai
P1	Hr	1001	Aneesh	Dev	1000	270000	10	Offshore	Mumbai
P2	Purchs	2000	Narayan	Analyst	1500	400000	40	Pre Sales	Delhi
P2	Purchs	3000	Krishna	Mgr	2000	540000	10	Offshore	Mumbai
P2	Purchs	3001	Tushar	Mgr	2000	540000	40	Pre Sales	Delhi

Proj ID	Proj name
P1	Hr
P2	Purchs

Empno	Ename	Job	Rate/hr	Tot Sal	Proj ID	Deptno
1000	Mihir	Analyst	1500	400000	P1	10
2011	Medha	Mgr	2000	540000	P1	30
1003	Siddha	Clerk	700	189000	P1	10
1001	Aneesh	Dev	1000	270000	P1	10
2000	Narayan	Analyst	1500	400000	P2	40
3000	Krishna	Mgr	2000	540000	P2	10
3001	Tushar	Mgr	2000	540000	P2	40

Deptno	dname	loc
10	Offshore	Mumbai
30	In House	Chennai
40	Pre Sales	Delhi

3NF

Proj ID	Proj name
P1	Hr
P2	Purchs

Empno	Ename	Job	Tot Sal	Proj ID	Deptno
1000	Mihir	Analyst	400000	P1	10
2011	Medha	Mgr	540000	P1	30
1003	Siddha	Clerk	189000	P1	10
1001	Aneesh	Dev	270000	P1	10
2000	Narayan	Analyst	400000	P2	40
3000	Krishna	Mgr	540000	P2	10
3001	Tushar	Mgr	540000	P2	40

Deptno	dname	loc
10	Offshore	Mumbai
30	In House	Chennai
40	Pre Sales	Delhi

Job	Rate/hr
Analyst	1500
Mgr	2000
Clerk	700
Dev	1000

Effects after NORMALIZATION

1. Same data got separated into number of tables as & when with every step
   (1NF,2NF etc.) followed of normalization.
2. There is no redundancy or minimal redundancy.
3. Insert/Update/Delete Anomalies are removed
   Inserting new data without affecting records.
   Updation of records doesn't cause redundancy
   Deletion of records independently is possible now

Example with first three forms

Suppose we have this Invoice Table

Invoice Table Violate's Normalization Form 1

	Customer Information											
Invoice#	Cust#	Name	Addr	Quant1	Part1	Amt1	Quant2	Part2	Amt2	Quant3	Part3	Amt3
1001	43	Jones	121 1st	200	Screw	2.00	300	Nut	2.25	100	Washr	0.75
1002	55	Smith	222 2nd	1	Motor	52.00	5	Brace	44.44			
1003	43	Jones	121 1st	10	Saw	121.00						

First Normal Form: No repeating groups.

- The above table violates 1NF because it has columns for the first, second, and third line item.

- Solution: you make a separate line item table, with it's own key, in this case the combination of invoice number and line number

Table now in 1NF

Complies with Normalization Form 1, Violate's Normalization Form 2

Invoice table

Invoice#
1001
1002
1003

Line item table

		Customer Information					
Invoice#	LIne#	Cust#	Name	Address	Quant1	Part1	Amt1
1001	1	43	Jones	121 1st	200	Screw	2.00
1001	2	43	Jones	121 1st	300	Nut	2.25
1001	3	43	Jones	121 1st	100	Washr	0.75
1002	1	55	Smith	222 2nd	1	Motor	52.00
1002	2	55	Smith	222 2nd	10	Saw	121.00
1003	1	43	Jones	121 1st	5	Brace	44.44

Second Normal Form:
Each column must depend on the *entire* primary key.

Complies with Normalization Form 2, Violate's Normalization Form 3

Invoice table

	Customer Information		
Invoice#	Cust#	Name	Address
1001	43	Jones	121 1st
1002	55	Smith	222 2nd
1003	43	Jones	121 1st

Line item table

Invoice#	LIne#	Quant1	Part1	Amt1
1001	1	200	Screw	2.00
1001	2	300	Nut	2.25
1001	3	100	Washr	0.75
1002	1	1	Motor	52.00
1002	2	10	Saw	121.00
1003	1	5	Brace	44.44

customer address could go in the invoice table (see above), but this would cause data redundancy if several invoices were for the same customer. It would also cause an update nightmare when the customer changes his address, and would require extensive programming

Third Normal Form:

Each column must depend on *directly* on the primary key.

Complies with Normalization Form 3

Invoice table

Invoice#	Cust#
1001	43
1002	55
1003	43

Customer table

Cust#	Name	Address
43	Jones	121 1st
55	Smith	222 2nd

Line item table

Invoice#	LIne#	Quant1	Part1	Amt1
1001	1	200	Screw	2.00
1001	2	300	Nut	2.25
1001	3	100	Washr	0.75
1002	1	1	Motor	52.00
1002	2	10	Saw	121.00
1003	1	5	Brace	44.44

Assignments

OrderNum	OrderDate	PartNum	NumOrdered
21608	10/20/2003	AT94	11
21610	10/20/2003	DR93 DW11	1 1
21613	10/21/2003	KL62	4
21614	10/21/2003	KT03	2
21617	10/23/2003	BV06 CD52	2 4
21619	10/23/2003	DR93	1
21623	10/23/2003	KV29	2

1NF

Order_ID	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	Product_ID	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2004	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2004	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2004	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2004	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
1007	10/25/2004	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3

2NF

Order

Order_ID	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	Product_ID	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2004	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2004	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2004	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2004	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
1007	10/25/2004	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3

Order

Order_ID	Order_Date	Cust_ID	Cust_Name	Cust_Address
1006	10/24/2004	2	Value Furniture	Plano, TX
1007	10/25/2004	6	Furniture Gallery	Boulder, CO

Product

Prod_ID	Prod_Desc	Prod_Finish	Unit_Price
4	Entertainment Center	Natural Maple	650.00
5	Writer's Desk	Cherry	325.00
7	Dining Table	Natural Ash	800.00
11	4-Dr Dresser	Oak	500.00

OrderItem

Ord_ID	**Prod_ID**	Ord_Quan
1006	7	2
1006	5	2
1006	4	1
1007	11	4
1007	4	3

Transforming to 3NF

Product

Prod_ID	Prod_Desc	Prod_Finish	Unit_Price
4	Entertainment Center	Natural Maple	650.00
5	Writer's Desk	Cherry	325.00
7	Dining Table	Natural Ash	800.00
11	4-Dr Dresser	Oak	500.00

OrderItem

Ord_ID	**Prod_ID**	Ord_Quan
1006	7	2
1006	5	2
1006	4	1
1007	11	4
1007	4	3

- Move the attributes involved in transitive dependency to another relation

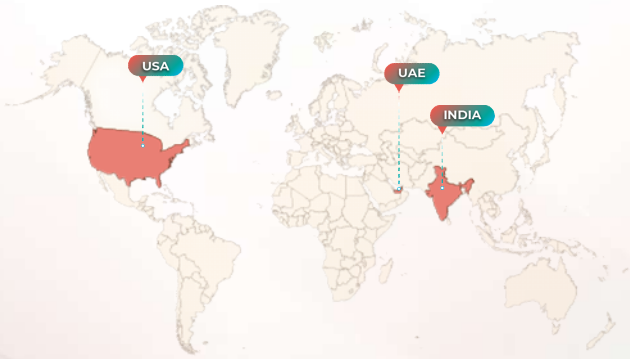
Order

Order_ID	Order_Date	Cust_ID
1006	10/24/2004	2
1007	10/25/2004	6
1008	11/1/2004	2

Customer

Cust_ID	Cust_Name	Cust_Address
2	Value Furniture	Plano, TX
6	Furniture Gallery	Boulder, CO

Physical Presence



www.vinsys.com

enquiry@vinsys.com

Shivaji Niketan, Tejas Society, Behind Kothrud Bus Stand, Near Mantri Park, Kothrud, Pune - 411029.

304, City Tower 2, Near Crown Plaza, Sheikh Zayed Road. Dubai, UAE.
PO.Box - 213279

132 West 31st Street, First Floor, New York, 10001, USA