SCHEMA REFINEMENT & NORMAL FORMS

Topics

- Functional Dependencies (FDs)
- Normalization using FDs
  - First Normal Form (1NF)
  - Second Normal Form (2NF)
  - Third Normal Form (3NF)
  - Fourth and Fifth Normal Form(4NF & 5NF)
  - Boyce-Codd Normal Form (BCNF)
- Lossless and dependency preserving decomposition

Redundancy

- Same information at many places in the DB
- Problems:
  - Wastage of Space
  - Anomalies
    - Update Anomaly
    - Insert Anomaly
    - Delete Anomaly
- Normalization is used for “minimizing” redundancy

Anomalies

Consider the relation:

EMP_PROJ (Emp#, Proj#, Ename, Pname, No_hours)

- **Update Anomaly:** Changing the name of project number P1 from "Billing" to "Customer-Accounting" may cause this update to be made for all 100 employees working on project P1
- **Insert Anomaly:** Cannot insert a project unless an employee is assigned to it

  Inversely - Cannot insert an employee unless he/she is assigned to a project.
- **Delete Anomaly:** When a project is deleted, it will result in deleting all the employees who work on that project. Alternately, if an employee is the sole employee on a project, deleting that employee would result in deleting the corresponding project

Solution

Decompose the relation:

EMP_PROJ (Emp#, Proj#, Ename, Pname, No_hours)

Into the following smaller relations:

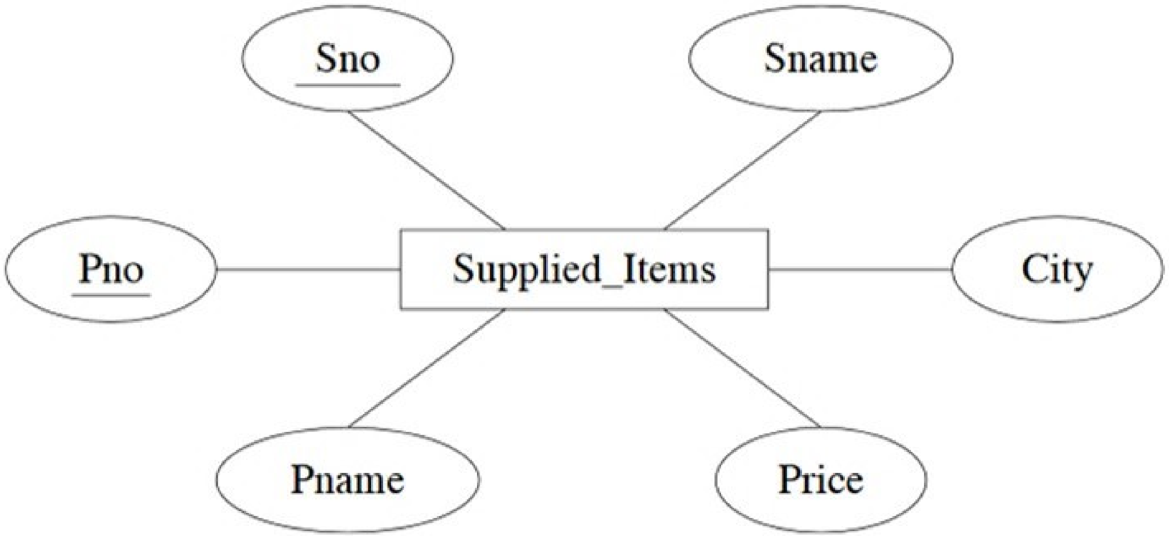
EMP (Emp#, Ename)

PROJ (Proj#, Pname)

EMP_PROJ (Emp#, Proj#, No_hours)

- **What happened to update anomalies?**
- **We need to find out the basis for decomposing a relation to get rid of update anomalies**

Database Design



Supplied_Items

Sno	Sname	City	Pno	Pname	Price
S1	Magna	Ajax	P1	Bolt	0.50
S1	Magna	Ajax	P2	Nut	0.25
S1	Magna	Ajax	P3	Screw	0.30
S2	Budd	Hull	P3	Screw	0.40

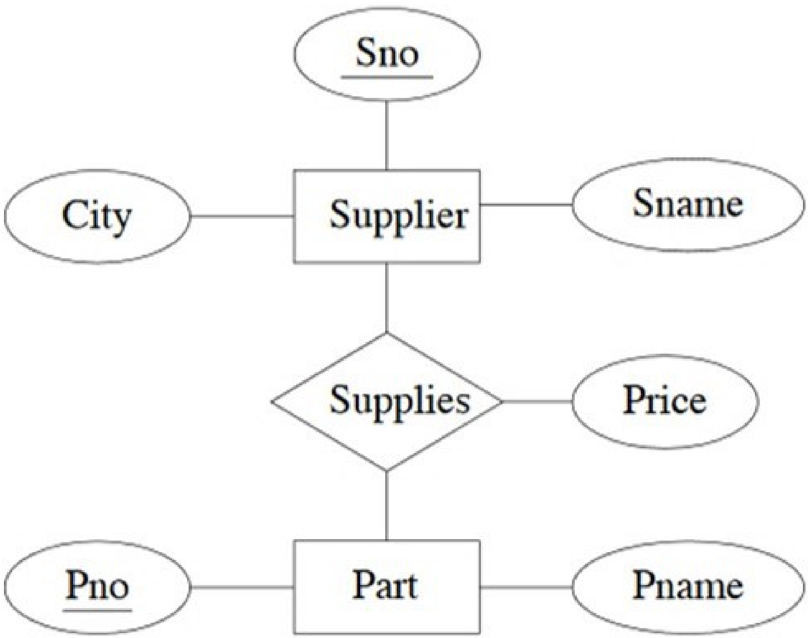
Description of a parts/suppliers database:

- Each type of part has a name and an identifying number, and may be supplied by zero or more suppliers.
- Each supplier may offer the part at a different price.
- Each supplier has an identifying number, a name, and a contact location for ordering parts.

Anomalies

- The single-table schema suffers from several kinds of problems:
  - Update problems (e.g. changing name of supplier)
  - Insert problems (e.g. add a new item)
  - Delete problems (e.g. Budd no longer supplies screws)
  - Likely increase in space requirements

Alternate Design



Suppliers

Sno	Sname	City
S1	Magna	Ajax
S2	Budd	Hull

Parts

Pno	Pname
P1	Bolt
P2	Nut
P3	Screw

Supplies

Sno	Pno	Price
S1	P1	0.50
S1	P2	0.25
S1	P3	0.30
S2	P3	0.40

- How do we know an anomaly exists?
  - Certain types of integrity constraints reveal regularities in database instances that lead to anomalies.
- What should we do if an anomaly exists?
  - Certain schema decompositions can avoid anomalies while retaining all information in the instances

Functional Dependencies

- In a relation schema(values of) a set of attributes uniquely determine(values of) another set of attributes.

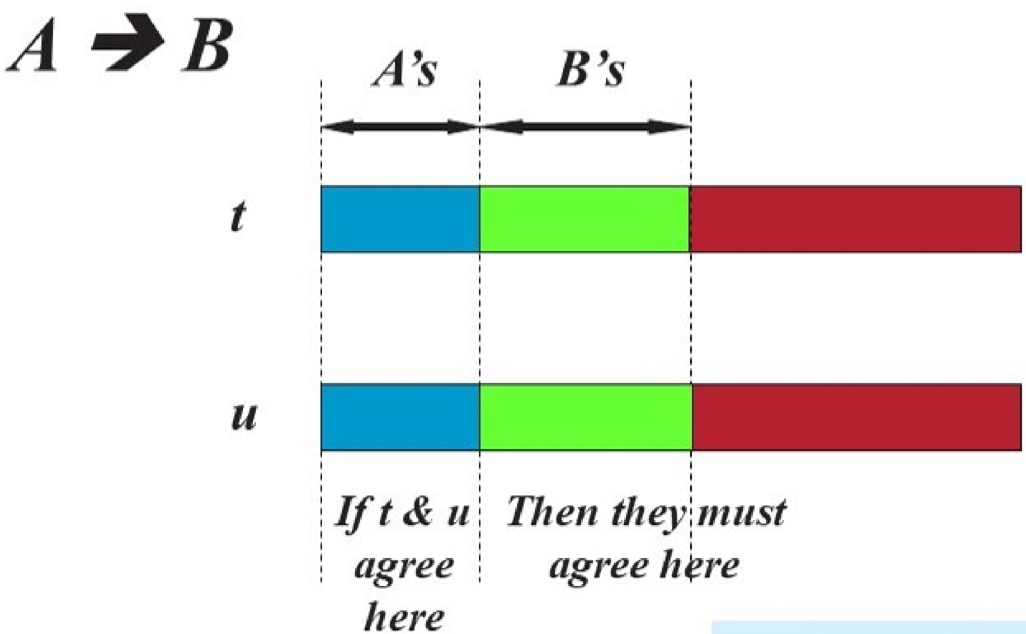
Let R be a relation schema, and X, $Y \subseteq R$ sets of attributes. The **functional dependency**

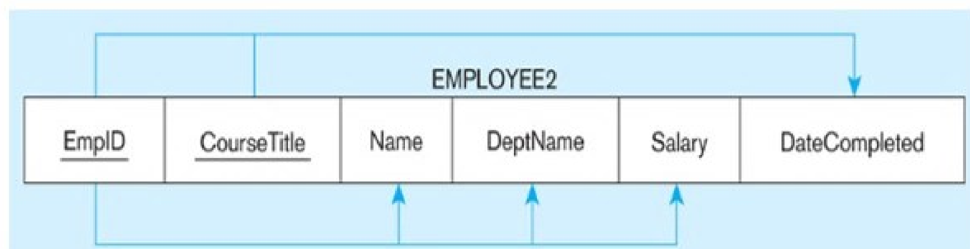
$$X \to Y$$

holds on R if whenever an instance of R contains two tuples t and u such that $t[X] = u[X]$ then it is also true that $t[Y] = u[Y]$.

We say that X *functionally determines* Y (in R).

Functional Dependencies





Functional Dependancy

SOMAIYA VIDYAVIHAR UNIVERSITY
K J Somaiya Institute of Management

- Consider the following relation schema

EmpProj

<u>SIN</u>	<u>PNum</u>	Hours	EName	PName	PLoc	Allowance

- SIN determines employee name

$$\text{SIN} \rightarrow \text{EName}$$

- project number determines project name and location

$$\text{PNum} \rightarrow \text{PName, PLoc}$$

- allowances are always the same for the same number of hours at the same location

$$\text{PLoc, Hours} \rightarrow \text{Allowance}$$

- **Consider the relation:**

PLOTS (prop#, state, plot#, area, price, Tax_rate)

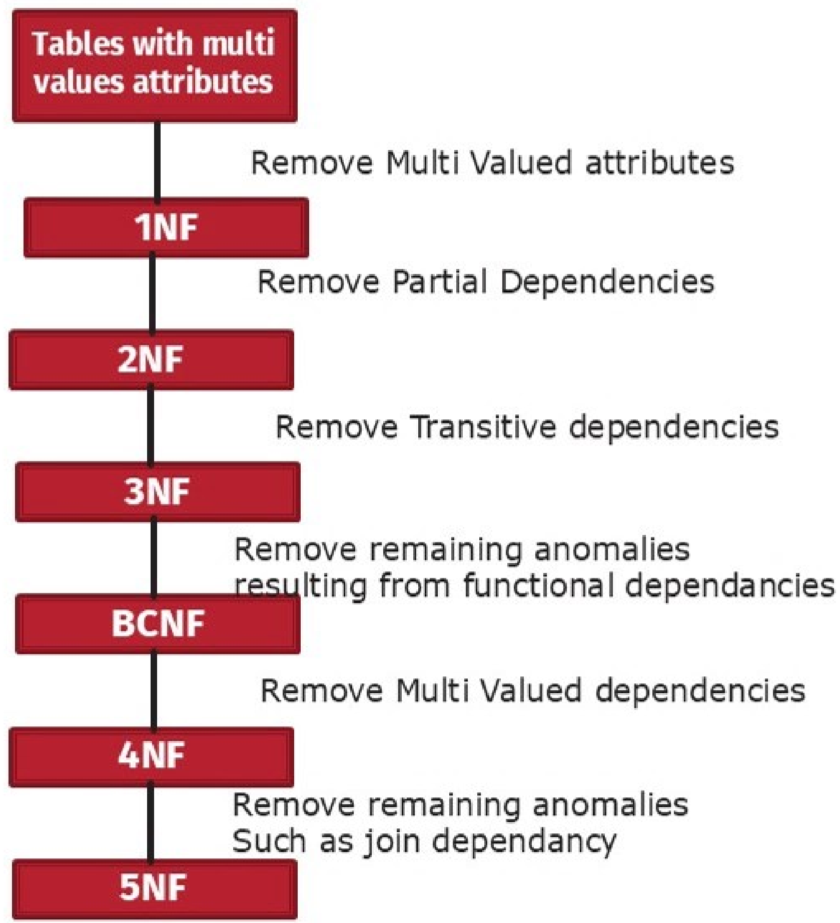
Information about plots available in India.

The constraints on the relation are:

  - Prop# is unique throughout India
  - Plot# are unique within a given state
  - For a given_state, tax_rate is fixed
  - Plots having the same area have the same price, irrespective of the state in which they are located

- **Write all the FDs on the relation PLOTS**

Normalization



1NF

A relation is in 1NF iff all underlying domains contain scalar values(atomic values) only.

S#	STATUS	CITY	P#	QTY
S1	20	Pune	P1,P4	300
S1	20	Pune	P2,P3	400
S1	20	Pune	P5,P6	100
S2	10	Mumbai	P1	300
S2	10	Mumbai	P2	400
S3	10	Mumbai	P2	200
S4	20	Pune	P2,P4	200
S4	20	Pune	P5	400

S#	STATUS	CITY	P#	QTY
S1	20	Pune	P1	300
S1	20	Pune	P2	400
S1	20	Pune	P3	400
S1	20	Pune	P4	300
S1	20	Pune	P5	100
S1	20	Pune	P6	100
S2	10	Mumbai	P1	300
S2	10	Mumbai	P2	400
S3	10	Mumbai	P2	200
S4	20	Pune	P2	200
S4	20	Pune	P4	200
S4	20	Pune	P5	400

Anomalies

- INSERT
  - Cannot insert the fact that a particular supplier is located in a particular city until that supplier supplies at least one part.
- DELETE
  - Deleting a tuple will remove the information about the city of a supplier.
- UPDATE
  - If a supplier moves from one city to another, update to be done in multiple tuples.

2NF

A relation is in 2NF iff it is in 1NF and every **non key attribute is irreducibly dependent on primary key**.

Given relation R as follows

R: (A,B,C,D)

PRIMARY KEY (A,B)

A→D

Then replace R by two projections R1 and R2

R1: (A,D)

PRIMARY KEY(A)

R2: (A,B,C)

PRIMARY KEY(A,B)

SUPPLIER (S#,STATUS,CITY,P#,QTY)

PRIMARY KEY: ???

(S#,P#)

FDs

(S#,P#)→QTY

S#→CITY

S#→STATUS

STATUS→CITY

S#	STATUS	CITY	P#	QTY
S1	20	Pune	P1	300
S1	20	Pune	P2	400
S1	20	Pune	P3	400
S1	20	Pune	P4	300
S1	20	Pune	P5	100
S1	20	Pune	P6	100
S2	10	Mumbai	P1	300
S2	10	Mumbai	P2	400
S3	10	Mumbai	P2	200
S4	20	Pune	P2	200
S4	20	Pune	P4	200
S4	20	Pune	P5	400

Decompose into two:

SUPPLIER2(S#,STATUS,CITY)

SP(S#,P#,QTY)

S#	STATUS	CITY
S1	20	Pune
S2	10	Mumbai
S3	10	Mumbai
S4	20	Pune
S5	30	Thane

S#	P#	QTY
S1	P1	300
S1	P2	400
S1	P3	400
S1	P4	300
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	200
S4	P5	400

SP RELATION

Anomalies

- INSERT
  - Cannot insert the fact that a particular city has a particular status until some supplier actually located in the city.
- DELETE
  - If a tuple is deleted from SUPPLIER2 relation we destroy the information for the supplier concerned and the information that city has that particular status.
- UPDATE
  - STATUS of a given CITY appears may times.
  - To change the STATUS change must be done to multiple tuples.

3NF

A relation is in 3NF iff it is in 2NF and **every non key attribute is non transitively dependent on the primary key.**

Consider

R: (A,B,C)

PRIMARY KEY(A)

B→C

DECOMPOSE

R1: (B,C)

PRIMARY KEY (B)

R2: (A,B)

PRIMARY KEY(A)

S#	STATUS	CITY
S1	20	Pune
S2	10	Mumbai
S3	10	Mumbai
S4	20	Pune
S5	30	Thane

S#	CITY
S1	Pune
S2	Mumbai
S3	Mumbai
S4	Pune
S5	Thane

CITY	STATUS
Mumbai	10
Pune	20
Thane	30
Nashik	50

BCNF(Boyce/Codd normal form)

A relation is in BCNF iff **every non-trivial, left irreducible FD has a candidate key as its determinant.**

OR

A relation is in BCNF iff **the only determinants are candidate key.**

Consider the following example with relation containing two disjoint nonoverlapping candidate keys.

S {S#,SNAME,STATUS,CITY}

Candidate keys : S#,SNAME

Assume CITY and STATUS are mutually independent.

BCNF(Boyce/Codd normal form)

FD are

S# → SNAME

- STATUS and CITY dependent on S#

S# → STATUS,CITY

- STATUS and CITY dependent on SNAME

SNAME → STATUS, CITY

Determinant is candidate key. Then the relation is in BCNF.

Candidate key overlap

When they involve two or more attributes each and has at least one attribute in common.

Consider the relation

SSP {S#,SNAME,P#,QTY}

Candidate keys: (S#,P#)

(SNAME,P#)

This relation is not in BCNF because S# and SNAME are both determinants.

S# → SNAME

SNAME → S#

S#	SNAME	P#	QTY
S1	Sam	P1	300
S1	Sam	P2	200
S1	Sam	P3	400
S1	Sam	P4	200
S1	Sam	P5	300

- SSP involves same type of redundancies as that of SUPPLIER relation.
- It is in 3NF.
- But not in BCNF.
- Break the relation down into two projections.

SS (S#,SNAME)   SP(S#,P#,QTY)

OR

SS (S#,SNAME)   SP(SNAME,P#,QTY)

Avoiding Database Anomalies

Normalization

- A relation is in 1NF iff all underlying domains contain scalar values(atomic values) only.

Product ID	Product Name	Categories	Supplier ID	Supplier Name	Supplier Contact
101	Laptop	Electronics, Office	201	ABC Electronics	contact@abc.com
102	Printer	Electronics, Office	201	ABC Electronics	contact@abc.com
103	Desk	Furniture	202	XYZ Furnishings	info@xyzfurniture.com

- Atomicity

Normalization

Product ID	Product Name	Categories	Supplier ID	Supplier Name	Supplier Contact
101	Laptop	Electronics,	201	ABC Electronics	contact@abc.com
101	Laptop	Office	201	ABC Electronics	contact@abc.com
102	Printer	Electronics,	201	ABC Electronics	contact@abc.com
102	Printer	Office	201	ABC Electronics	contact@abc.com
103	Desk	Furniture	202	XYZ Furnishings	info@xyzfurniture.com

- Data Redundancy / Repeated Data

2NF

- A relation is in 2NF iff it is in 1NF and every **non key attribute is irreducibly dependent on primary key**.
- **Functional Dependency** : Value of one attribute uniquely determines the value of another attribute in the same table.

Functional Dependency

- How attributes (columns) in a table relate to each other
- Value of one attribute uniquely determines the value of another attribute in the same table.
- Employee ID functionally determines Employee Name and Department.

Employee ID	Employee Name	Department
101	John Doe	Finance
102	Jane Smith	Marketing
103	Bob Johnson	Finance
104	Mary Adams	HR

Functional Dependency

- helps to design and structure databases effectively.
- ensures data accuracy and reduces redundancy.

Customer ID	Customer Name	Total Purchase Amount	Loyalty Tier
101	Nikita	1500	Gold
102	Rahul	800	Silver
103	Anu	2200	Gold
104	Mary	500	Bronze

- **Customer ID → Customer Name**
- **Customer ID → Total Purchase Amount**
- **Total Purchase Amount → Loyalty Tier**

SOMAIYA VIDYAVIHAR UNIVERSITY
...e of Management

2NF

Primary Key – (CustID, ProductID)

Dependencies

(CustID, ProductID) → OrderID

ProductID → Category

Cust_ID	Product ID	Order ID
101	P1101	11
102	P1102	22
103	P1101	33
104	P1103	44
101	P1103	11

Cust_ID	Product ID	Order ID	Catgory
101	P1101	11	Electronics
102	P1102	22	Electronics
103	P1101	33	Electronics
104	P1103	44	Homedecor
101	P1103	11	Homedecor

Product ID	Catgory
P1101	Electronics
P1102	Electronics
P1103	Homedecor

Normalization

- Data Redundancy / Repeated Data

Product ID	Product Name	Categories	Supplier ID	Supplier Name	Supplier Contact
101	Laptop	Electronics,	201	ABC Electronics	contact@abc.com
101	Laptop	Office	201	ABC Electronics	contact@abc.com
102	Printer	Electronics,	201	ABC Electronics	contact@abc.com
102	Printer	Office	201	ABC Electronics	contact@abc.com
103	Desk	Furniture	202	XYZ Furnishings	info@xyzfurniture.com

PRODUCT

Product ID	Product Name	Category	Supplier Id
101	Laptop	Electronics ,	201
101	Laptop	Office	201
102	Printer	Electronics ,	201
102	Printer	Office	201
103	Desk	Furniture	202

SUPPLIER

Supplier ID	Supplier Name	Supplier Contact
201	ABC Electronics	contact@abc.com
202	XYZ Furnishings	info@xyzfurniture.com

3NF

A relation is in 3NF iff it is in 2NF and **every non key attribute is non transitively dependent on the primary key.**

Consider

R: (A,B,C)

PRIMARY KEY(A)

B→C

DECOMPOSE

R1: (B,C)

PRIMARY KEY (B)

R2: (A,B)

PRIMARY KEY(A)

Primary Key
(Product ID, Category)
OR
(ProductName,Category)

Product ID	Product Name	Category	Supplier Id
101	Laptop	Electronics	201
101	Laptop	Office	201
102	Printer	Electronics	201
102	Printer	Office	201
103	Desk	Furniture	202

Dependencies
ProductID → ProductName
ProdID→ SupplierID

Product ID	Product Name	Supplier ID
101	Laptop	201
102	Printer	201
103	Desk	202

Product Name	Category
Laptop	Electronics
Laptop	Office
Printer	Electronics
Printer	Office
Desk	Furniture

Perform Normalization

Student ID	Student Name	Course ID	Course Name	Instructor Name	Department
101	Roshan	C101	Math 101	Prof. Rekha	Mathematics
102	Shruti	C102	History 101	Prof. Megha	History
101	Roshan	C102	History 101	Prof. Megha	History
103	Shivam	C101	Math 101	Prof. Rekha	Mathematics
102	Shruti	C101	Math 101	Prof. Renu	Statistics

Table is in first normal form as all cells contain atomic values.
Identify the primary key

(StudentID,CourseID)

Student ID	Student Name	Course ID	Course Name	Instructor Name	Department

Identify Functional Dependencies

StudentID → Student Name

Course ID → Coursename

Instructorname → Department

CourseID → Department

2NF

STUDENT

Student ID	Student Name
101	Roshan
102	Shruti
101	Pallavi
103	Shivam
102	Shruti

COURSE_INSTR

Course ID	Course Name	Instructor Name	Department
C101	Math 101	Prof. Rekha	Mathematics
C102	History 101	Prof. Megha	History
C102	History 101	Prof. Megha	History
C101	Math 101	Prof. Rekha	Mathematics
C101	Math 101	Prof. Renu	Statistics

Student ID	Course ID
101	C101
102	C102
101	C102
103	C101
102	C101

STUD_COURSE

3NF

Course ID	Course Name	Instructor Name	Department
C101	Math 101	Prof. Rekha	Mathematics
C102	History 101	Prof. Megha	History
C102	History 101	Prof. Megha	History
C101	Math 101	Prof. Rekha	Mathematics
C101	Math 101	Prof. Renu	Statistics

COURSE

Course ID	Course Name
C101	Math 101
C102	History 101

COURSE_INSTR

Course ID	Instructor Name
C101	Prof. Rekha
C102	Prof. Megha
C101	Prof. Renu

INSTRUCTOR

Instructor Name	Department
Prof. Rekha	Mathematics
Prof. Megha	History
Prof. Renu	Statistics

STUDENT

Student ID	Student Name
101	Roshan
102	Shruti
101	Pallavi
103	Shivam
102	Shruti

STUD_COURSE

Student ID	Course ID
101	C101
102	C102
101	C102
103	C101
102	C101

COURSE

Course ID	Course Name
C101	Math 101
C102	History 101

COURSE_INSTR

Course ID	Instructor Name
C101	Prof. Rekha
C102	Prof. Megha
C101	Prof. Renu

INSTRUCTOR

Instructor Name	Department
Prof. Rekha	Mathematics
Prof. Megha	History

Unnormalized Relation

REVENUE TRANSACTIONS								
Invoice #	Transaction Date	Service Type #	Service Name	Hours	Hourly Rate	Amount	Customer #	Customer Name
101	1/5/2003	1	Bookkeeping	6	25	150	251	Stinson & Assoc.
		3	Partnership	25	50	1,250	251	Stinson & Assoc.
102	1/28/2003	4	Personal	6	40	240	136	Jack Randall
103	2/10/2003	4	Personal	4	40	160	200	Judy and June
104	2/18/2003	1	Bookkeeping	20	25	500	257	Warren Cleaning
		3	Partnership	35	50	1,750	257	Warren Cleaning
105	2/25/2003	1	Bookkeeping	8	25	200	102	Bernie Contrell
		4	Personal	16	40	640	102	Bernie Contrell
		5	Tax Planning	5	75	375	102	Bernie Contrell
106	3/5/2003	1	Bookkeeping	151	25	3,775	385	Kirk Company
107	3/9/2003	4	Personal	6	40	240	154	Amy Holt

First Normal Form (1NF)

A relation is in first normal form if it does not contain repeating groups.

Unnormalized Relation

SOMAIYA VIDYAVIHAR UNIVERSITY
K J Somaiya Institute of Management

REVENUE TRANSACTIONS								
Invoice #	Transaction Date	Service Type #	Service Name	Hours	Hourly Rate	Amount	Customer #	Customer Name
101	05-01-2003	1	Bookkeeping	6	25	150	251	Stinson & Assoc.
		3	Partnership	25	50	1,250	251	Stinson & Assoc.
102	28-01-2003	4	Personal	6	40	240	136	Jack Randall
103	2/10/2003	4	Personal	4	40	160	200	Judy and June
104	2/18/2003	1	Bookkeeping	20	25	500	257	Warren Cleaning
		3	Partnership	35	50	1,750	257	Warren Cleaning
105	25-02-2003	1	Bookkeeping	8	25	200	102	Bernie Contrell
		4	Personal	16	40	640	102	Bernie Contrell
		5	Tax Planning	5	75	375	102	Bernie Contrell
106	05-03-2003	1	Bookkeeping	151	25	3,775	385	Kirk Company
107	09-03-2003	4	Personal	6	40	240	154	Amy Holt

First Normal Form (1NF)

Changes to service names have to be made on each line on which the the name appears.

REVENUE TRANSACTIONS								
Invoice #	Transaction Date	Service Type #	Service Name	Hours	Hourly Rate	Amount	Customer #	Customer Name
101	05-01-2003	1	Bookkeeping	6	25	150	251	Stinson & Assoc.
101		3	Partnership	25	50	1,250	251	Stinson & Assoc.
102	28-01-2003	4	Personal	6	40	240	136	Jack Randall
103	2/10/2003	4	Personal	4	40	160	200	Judy and June
104	2/18/2003	1	Bookkeeping	20	25	500	257	Warren Cleaning
104		3	Partnership	35	50	1,750	257	Warren Cleaning
105	25-02-2003	1	Bookkeeping	8	25	200	102	Bernie Contrell
105		4	Personal	16	40	640	102	Bernie Contrell
105		5	Tax Planning	5	75	375	102	Bernie Contrell
106	05-03-2003	1	Bookkeeping	151	25	3,775	385	Kirk Company
107	09-03-2003	4	Personal	6	40	240	154	Amy Holt

First Normal Form (1NF)

Inconsistent data could result when the same service name has several name variations.

REVENUE TRANSACTIONS								
Invoice #	Transaction Date	Service Type #	Service Name	Hours	Hourly Rate	Amount	Customer #	Customer Name
101	1/5/2003	1	Bookkeeping	6	25	150	251	Stinson & Assoc.
101		3	Partnership	25	50	1,250	251	Stinson & Assoc.
102	1/28/2003	4	Personal	6	40	240	136	Jack Randall
103	2/10/2003	4	Personal	4	40	160	200	Judy and June
104	2/18/2003	1	Bookkeeping	20	25	500	257	Warren Cleaning
104		3	Partnership	35	50	1,750	257	Warren Cleaning
105	2/25/2003	1	Bookkeeping	8	25	200	102	Bernie Contrell
105		4	Personal	16	40	640	102	Bernie Contrell
105		5	Tax Planning	5	75	375	102	Bernie Contrell
106	3/5/2003	1	Bookkeeping	151	25	3,775	385	Kirk Company
107	3/9/2003	4	Personal	6	40	240	154	Amy Holt

First Normal Form (1NF)

New services can not be added unless there is an existing transaction.

REVENUE TRANSACTIONS								
Invoice #	Transaction Date	Service Type #	Service Name	Hours	Hourly Rate	Amount	Customer #	Customer Name
101	1/5/2003	1	Bookkeeping	6	25	150	251	Stinson & Assoc.
101		3	Partnership	25	50	1,250	251	Stinson & Assoc.
102	1/28/2003	4	Personal	6	40	240	136	Jack Randall
103	2/10/2003	4	Personal	4	40	160	200	Judy and June
104	2/18/2003	1	Bookkeeping	20	25	500	257	Warren Cleaning
104		3	Partnership	35	50	1,750	257	Warren Cleaning
105	2/25/2003	1	Bookkeeping	8	25	200	102	Bernie Contrell
105		4	Personal	16	40	640	102	Bernie Contrell
105		5	Tax Planning	5	75	375	102	Bernie Contrell
106	3/5/2003	1	Bookkeeping	151	25	3,775	385	Kirk Company
107	3/9/2003	4	Personal	6	40	240	154	Amy Holt

First Normal Form (1NF)

Deletion of an invoice could result in a loss of service information.

REVENUE TRANSACTIONS								
Invoice #	Transaction Date	Service Type #	Service Name	Hours	Hourly Rate	Amount	Customer #	Customer Name
101	1/5/2003	1	Bookkeeping	6	25	150	251	Stinson & Assoc.
101		3	Partnership	25	50	1,250	251	Stinson & Assoc.
102	1/28/2003	4	Personal	6	40	240	136	Jack Randall
103	2/10/2003	4	Personal	4	40	160	200	Judy and June
104	2/18/2003	1	Bookkeeping	20	25	500	257	Warren Cleaning
104		3	Partnership	35	50	1,750	257	Warren Cleaning
105	2/25/2003	1	Bookkeeping	8	25	200	102	Bernie Contrell
105		4	Personal	16	40	640	102	Bernie Contrell
105		5	Tax Planning	5	75	375	102	Bernie Contrell
106	3/5/2003	1	Bookkeeping	151	25	3,775	385	Kirk Company
107	3/9/2003	4	Personal	6	40	240	154	Amy Holt

Second Normal Form (2NF)

A relation is in second normal form if it is in first normal form and there are **no partial dependencies**. A partial dependency occurs when a table has a composite primary key and a non-key attribute is dependent on only part of the primary key.

REVENUE TRANSACTIONS								
Invoice #	Transaction Date	Service Type #	Service Name	Hours	Hourly Rate	Amount	Customer #	Customer Name

Relations in Second Normal Form (2NF)

SOMAIYA VIDYAVIHAR UNIVERSITY
K J Somaiya Institute of Management

SERVICES		
Service Type #	Service Name	Rate
1	Bookkeeping	25
2	Bus. Consult.	60
3	Partnership	50
4	Personal	40
5	Tax Planning	75

Changes to customer names have to be made on each line on which the customer appears.

REVENUE TRANSACTIONS			
Invoice #	Transaction Date	Customer #	Customer Name
101	1/5/2003	251	Stinson & Assoc.
102	1/28/2003	136	Jack Randall
103	2/10/2003	200	Judy and June
104	2/18/2003	257	Warren Cleaning
105	2/25/2003	102	Bernie Contrell
106	3/5/2003	385	Kirk Company
107	3/9/2003	154	Amy Holt

REVENUE TRANSACTION LINES		
Invoice #	Service Type #	Hours
101	1	6
101	3	25
102	4	6
103	4	4
104	1	20
104	3	35
105	1	8
105	4	16
105	5	5
106	1	151
107	4	6

Relations in Second Normal Form (2NF)

SOMAIYA VIDYAVIHAR UNIVERSITY
K J Somaiya Institute of Management

SERVICES		
Service Type #	Service Name	Rate
1	Bookkeeping	25
2	Bus. Consult.	60
3	Partnership	50
4	Personal	40
5	Tax Planning	75

Inconsistent data could result when the same customer has several name variations.

REVENUE TRANSACTIONS			
Invoice #	Transaction Date	Customer #	Customer Name
101	1/5/2003	251	Stinson & Assoc.
102	1/28/2003	136	Jack Randall
103	2/10/2003	200	Judy and June
104	2/18/2003	257	Warren Cleaning
105	2/25/2003	102	Bernie Contrell
106	3/5/2003	385	Kirk Company
107	3/9/2003	154	Amy Holt

REVENUE TRANSACTION LINES		
Invoice #	Service Type #	Hours
101	1	6
101	3	25
102	4	6
103	4	4
104	1	20
104	3	35
105	1	8
105	4	16
105	5	5
106	1	151
107	4	6

Relations in Second Normal Form (2NF)

SOMAIYA VIDYAVIHAR UNIVERSITY
K J Somaiya Institute of Management

SERVICES

Service Type #	Service Name	Rate
1	Bookkeeping	25
2	Bus. Consult.	60
3	Partnership	50
4	Personal	40
5	Tax Planning	75

New customers can not be added unless there is an existing transaction for the customer.

REVENUE TRANSACTIONS

Invoice #	Transaction Date	Customer #	Customer Name
101	1/5/2003	251	Stinson & Assoc.
102	1/28/2003	136	Jack Randall
103	2/10/2003	200	Judy and June
104	2/18/2003	257	Warren Cleaning
105	2/25/2003	102	Bernie Contrell
106	3/5/2003	385	Kirk Company
107	3/9/2003	154	Amy Holt

REVENUE TRANSACTION LINES

Invoice #	Service Type #	Hours
101	1	6
101	3	25
102	4	6
103	4	4
104	1	20
104	3	35
105	1	8
105	4	16
105	5	5
106	1	151
107	4	6

Relations in Second Normal Form (2NF)

SOMAIYA VIDYAVIHAR UNIVERSITY
K J Somaiya Institute of Management

SERVICES		
Service Type #	Service Name	Rate
1	Bookkeeping	25
2	Bus. Consult.	60
3	Partnership	50
4	Personal	40
5	Tax Planning	75

Deletion of an invoice could result in the loss of customer information.

REVENUE TRANSACTION LINES		
Invoice #	Service Type #	Hours
101	1	6
101	3	25
102	4	6
103	4	4
104	1	20
104	3	35
105	1	8
105	4	16
105	5	5
106	1	151
107	4	6

REVENUE TRANSACTIONS			
Invoice #	Transaction Date	Customer #	Customer Name
101	1/5/2003	251	Stinson & Assoc.
102	1/28/2003	136	Jack Randall
103	2/10/2003	200	Judy and June
104	2/18/2003	257	Warren Cleaning
105	2/25/2003	102	Bernie Contrell
106	3/5/2003	385	Kirk Company
107	3/9/2003	154	Amy Holt

Third Normal Form (3NF)

A relation is in third normal form if it is in second normal form and there are **no transitive dependencies**. A transitive dependency occurs when a table has a non-key attribute that is dependent on another non-key attribute.

REVENUE TRANSACTIONS			
Invoice #	Transaction Date	Customer #	Customer Name

Relations in Third Normal Form (3NF)

SOMAIYA VIDYAVIHAR UNIVERSITY
K J Somaiya Institute of Management

SERVICES		
Service Type #	Service Name	Rate
1	Bookkeeping	25
2	Bus. Consult.	60
3	Partnership	50
4	Personal	40
5	Tax Planning	75

Customers Accounts	
Customer #	Customer Name
102	Bernie Contrell
136	Jack Randall
154	Amy Holt
200	Judy and June
251	Stinson & Assoc.
257	Warren Cleaning
385	Kirk Company

REVENUE TRANSACTIONS		
Invoice #	Transaction Date	Customer #
101	1/5/2003	251
102	1/28/2003	136
103	2/10/2003	200
104	2/18/2003	257
105	2/25/2003	102
106	3/5/2003	385
107	3/9/2003	154

REVENUE TRANSACTION LINES		
Invoice #	Service Type #	Hours
101	1	6
101	3	25
102	4	6
103	4	4
104	1	20
104	3	35
105	1	8
105	4	16
105	5	5
106	1	151
107	4	6


Success!

Project Management Report

Project Code: PC010
Project Title: Pensions System

Project Manager: M Phillips **Project Budget:**
£24,500

Employee No.	Employee Name	Department No.	Department Name	Hourly Rate
S10001	A Smith	L004	IT	£22.00
S10030	L Jones	L023	Pensions	£18.50
S21010	P Lewis	L004	IT	£21.00
S00232	R Smith	L003	Programming	£26.00

Total Staff on Project: 4 **Average Hourly Rate:** £21.88

Calculated Fields

Project Code	Project Title	Project Manager	Project Budget	Employee No.	Employee Name	Department No.	Department Name	Hourly Rate
PC010	Pensions System	M Phillips	24500	S10001	A Smith	L004	IT	22.00
PC010	Pensions System	M Phillips	24500	S10030	L Jones	L023	Pensions	18.50
PC010	Pensions System	M Phillips	24500	S21010	P Lewis	L004	IT	21.00
PC045	Salaries System	H Martin	17400	S10010	B Jones	L004	IT	21.75
PC045	Salaries System	H Martin	17400	S10001	A Smith	L004	IT	18.00
PC045	Salaries System	H Martin	17400	S31002	T Gilbert	L028	Database	25.50
PC045	Salaries System	H Martin	17400	S13210	W Richards	L008	Salary	17.00
PC064	HR System	K Lewis	12250	S31002	T Gilbert	L028	Database	23.25
PC064	HR System	K Lewis	12250	S21010	P Lewis	L004	IT	17.50
PC064	HR System	K Lewis	12250	S10034	B James	L009	HR	16.50
PC010	Pensions System	M Phillips	24500	S10001	A Smith	L004	IT	22.00
				S10030	L Jones	L023	Pensions	18.50
				S21010	P Lewis	L004	IT	21.00
PC045	Salaries System	H Martin	17400	S10010	B Jones	L004	IT	21.75
				S10001	A Smith	L004	IT	18.00
				S31002	T Gilbert	L028	Database	25.50
				S13210	W Richards	L008	Salary	17.00
PC064	HR System	K Lewis	12250	S31002	T Gilbert	L028	Database	23.25
				S21010	P Lewis	L004	IT	17.50
				S10034	B James	L009	HR	16.50

UNF: Unnormalised Table

3NF Tables: Non-Key Dependencies Removed

Project table

Project Code	**Project Title**	**Project Manager**	**Project Budget**
PC010	Pensions System	M Phillips	24500
PC045	Salaries System	H Martin	17400
PC064	HR System	K Lewis	12250

Employee table

Employee No.	**Employee Name**	**Department No. ***
S10001	A Smith	L004
S10030	L Jones	L023
S21010	P Lewis	L004
S10010	B Jones	L004
S31002	T Gilbert	L023
S13210	W Richards	L008
S10034	B James	L0009

Employee-project table

Project Code	**Employee No.**	**Hourly Rate**
PC010	S10001	22.00
PC010	S10030	18.50
PC010	S21010	21.00
PC045	S10010	21.75
PC045	S10001	18.00
PC045	S31002	25.50
PC045	S13210	17.00
PC064	S31002	23.25
PC064	S21010	17.50
PC064	S10034	16.50

Department table

Department No.	**Department Name**
L004	IT
L023	Pensions
L028	Database
L008	Salary
L009	HR

4NF

SOMAIYA VIDYAVIHAR UNIVERSITY
K J Somaiya Institute of Management

A relation R is in 4NF iff, whenever there exist subsets A and B of the attributes R such that the nontrivial multivalued dependency $A \twoheadrightarrow B$ is satisfied, then all attributes of R are also functional dependent on A.

OR

A relation R is in 4NF if it is in BCNF and all MVDs in R are in fact FDs "out of candidate keys".

multivalued dependency

SOMAIYA
VIDYAVIHAR UNIVERSITY
K J Somaiya Institute of Management

- There are at least 3 attributes A, B, C in a relation and
- For each value of A there is a well defined set of values for B, and a well defined set of values for C,
- But the set of values for B is independent on the set of values for C
- Example: **JointAppoint(facId, dept, committee)** assuming a faculty member can belong to more than one department and belong to more than one committee
- Given a relation R(A,B,C) the MVD $A \twoheadrightarrow B$ holds iff the MVD $A \twoheadrightarrow C$ also holds.

Consider the following relation CTX:

In relational databases, repeating groups are not allowed

Course	Teacher	Texts
DBS	N Goyal J P Misra	Garcia Raghu
ADBS	J P Misra	Connolly Garcia

4 NF

1 NF Version

CTX

COURSE	TEACHER	TEXTS
DBS	N GOYAL	GARCIA
DBS	N GOYAL	RAGHU R
DBS	J P MISRA	GARCIA
DBS	J P MISRA	RAGHU R
ADBS	J P MISRA	GARCIA
ADBS	J P MISRA	CONNOLLY

NO FDs in this relation

4 NF

- **Highest Normal Form?** BCNF?

CTX

COURSE	TEACHER	TEXTS
DBS	N GOYAL	GARCIA
DBS	N GOYAL	RAGHU R
DBS	J P MISRA	GARCIA
DBS	J P MISRA	RAGHU R
ADBS	J P MISRA	GARCIA
ADBS	J P MISRA	CONNOLLY

SOMAIYA
VIDYAVIHAR UNIVERSITY
K J Somaiya Institute of Management

4 NF

- **Anomalies?**

MANY!!

CTX

COURSE	TEACHER	TEXTS
DBS	N GOYAL	GARCIA
DBS	N GOYAL	RAGHU R
DBS	J P MISRA	GARCIA
DBS	J P MISRA	RAGHU R
ADBS	J P MISRA	GARCIA
ADBS	J P MISRA	CONNOLLY

4 NF

Anomalies

- New Teacher for DBS
- New Text for ADBS
- Teacher teaching DBS leaves

4 NF

- **Decompose CTX into CT & TX**

CT

COURSE	TEACHER
DBS	N GOYAL
DBS	J P MISRA
ADBS	J P MISRA

TX

COURSE	TEXT
DBS	GARCIA
DBS	RAGHU R
ADBS	GARCIA
ADBS	CONNOLLY

4 NF

- **Decompose CTX into CT & TX is not done on the basis of FDs (as there are no FDs)**
- **Decompose CTX into CT & TX is done on the basis of MVDs**

  $course \rightarrow\!\!\!\rightarrow teacher$ (course multi-determines teacher)

  $course \rightarrow\!\!\!\rightarrow text$ (text multi-dependent on course)

5NF(Projection-join normal form)

- R is in 5NF if and only if every join dependency in R is implied by the candidate keys of R
- **Join dependency**

  Let R be a relation. Let A, B, ..., Z be arbitrary subsets of R's attributes. R satisfies the JD

  R ∗ (A, B, ..., Z)

  if and only if R is equal to the join of its projections on A, B, ..., Z

Travelling Salesman	Brand	Product Type
Jack Schneider	Acme	Vacuum Cleaner
Jack Schneider	Acme	Breadbox
Willy Loman	Robusto	Pruning Shears
Willy Loman	Robusto	Vacuum Cleaner
Willy Loman	Robusto	Breadbox
Willy Loman	Robusto	Umbrella Stand
Louis Ferguson	Robusto	Vacuum Cleaner
Louis Ferguson	Robusto	Telescope
Louis Ferguson	Acme	Vacuum Cleaner
Louis Ferguson	Acme	Lava Lamp
Louis Ferguson	Nimbus	Tie Rack

The table's predicate is: Products of the type designated by *Product Type*, made by the brand designated by *Brand*, are available from the travelling salesman designated by *Travelling Salesman*.

A Travelling Salesman has certain Brands and certain Product Types in his repertoire. If Brand B is in his repertoire, and Product Type P is in his stock, then (assuming Brand B makes Product Type P), the Travelling Salesman must offer products of Product Type P made by Brand B.

Travelling Salesman	Product Type
Jack Schneider	Vacuum Cleaner
Jack Schneider	Breadbox
Willy Loman	Pruning Shears
Willy Loman	Vacuum Cleaner
Willy Loman	Umbrella Stand
Louis Ferguson	Telescope
Louis Ferguson	Vacuum Cleaner
Louis Ferguson	Lava Lamp
Louis Ferguson	Tie Rack

Travelling Salesman	Brand
Jack Schneider	Acme
Willy Loman	Robusto
Louis Ferguson	Robusto
Louis Ferguson	Acme
Louis Ferguson	Nimbus

Brand	Product Type
Acne	Vacuum Cleaner
Acne	Breadbox
Acne	Lava Lamp
Robusto	Pruning Shears
Robusto	Vacuum Cleaner
Robusto	Breadbox
Robusto	Umbrella Stand
Robusto	Telescope
Nimbus	Tie Rack

Definition of Decomposition

Let R be a relation schema

A set of relation schemas { R1, R2,…, Rn } is a decomposition of R if

- R = R1 U R2 U …..U Rn
- each Ri is a subset of R (for i = 1,2…,n)

Example of Decomposition

For relation R(x,y,z) there can be 2 subsets:

R1(x,z) and R2(y,z)

If we union R1 and R2, we get R

R = R1 U R2

- Given instances of the decomposed relations, we may not be able to reconstruct the corresponding instance of the original relation – **information loss**

Example : Problem with Decomposition

SOMAIYA VIDYAVIHAR UNIVERSITY
K J Somaiya Institute of Management

R

Model Name	Price	Category
a11	100	Canon
s20	200	Nikon
a70	150	Canon

R1

Model Name	Category
a11	Canon
s20	Nikon
a70	Canon

R2

Price	Category
100	Canon
200	Nikon
150	Canon

Example : Problem with Decomposition

R1 U R2

Model Name	Price	Category
a11	100	Canon
a11	150	Canon
s20	200	Nikon
a70	100	Canon
a70	150	Canon

R

Model Name	Price	Category
a11	100	Canon
s20	200	Nikon
a70	150	Canon

Lossy decomposition

- In previous example, additional tuples are obtained along with original tuples
- Although there are more tuples, this leads to less information
- Due to the loss of information, decomposition for previous example is called lossy decomposition or lossy-join decomposition

Lossy decomposition (more example)

SOMAIYA VIDYAVIHAR UNIVERSITY
K J Somaiya Institute of Management

T

Employee	Project	Branch
Brown	Mars	L.A.
Green	Jupiter	San Jose
Green	Venus	San Jose
Hoskins	Saturn	San Jose
Hoskins	Venus	San Jose

Functional dependencies:

Employee $\longrightarrow$ Branch, Project $\longrightarrow$ Branch

Decomposition of the previous relation

T1

Employee	Branch
Brown	L.A
Green	San Jose
Hoskins	San Jose

T2

Project	Branch
Mars	L.A.
Jupiter	San Jose
Saturn	San Jose
Venus	San Jose

SOMAIYA
VIDYAVIHAR UNIVERSITY
K J Somaiya Institute of Management

After Natural Join

Employee	Project	Branch
Brown	Mars	L.A.
Green	Jupiter	San Jose
Green	Venus	San Jose
Hoskins	Saturn	San Jose
Hoskins	Venus	San Jose
Green	Saturn	San Jose
Hoskins	Jupiter	San Jose

Original Relation

Employee	Project	Branch
Brown	Mars	L.A.
Green	Jupiter	San Jose
Green	Venus	San Jose
Hoskins	Saturn	San Jose
Hoskins	Venus	San Jose

After Natural Join, we get two extra tuples.
Thus, there is loss of information.

A decomposition {R1, R2,…, Rn} of a relation R is called a lossless decomposition for R if the natural join of R1, R2,…, Rn produces exactly the relation R.

Lossless Decomposition

A decomposition is lossless if we can recover:

R(A, B, C)

Decompose

R1(A, B)   R2(A, C)

Recover

R'(A, B, C)

Thus, R' = R

Dependency Preservation

- Let F_i be the set of dependencies in F^+ that include only attributes in R_i.
  - A decomposition is **dependency preserving**, if $(F_1 \cup F_2 \cup \ldots \cup F_n)^+ = F^+$
  - If it is not, then checking updates for violation of functional dependencies may require computing joins, which is expensive.

THANK YOU