# Normalization Examples

# Example I

Take the following table.

StudentID is the primary key.

| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 1959433ZX | Mary Watson | 10 Charles Street | Bob | Red | English<br>Maths<br>Info Tech | $50<br>$50<br>$100 | B<br>A<br>B+ |

# Is it 1NF?

No. There are repeating groups (subject, subjectcost, grade)

| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | English<br>Maths<br>Info Tech | $50<br>$50<br>$100 | B<br>A<br>B+ |

How can you make it 1NF?

# Create new rows so each cell contains only one value

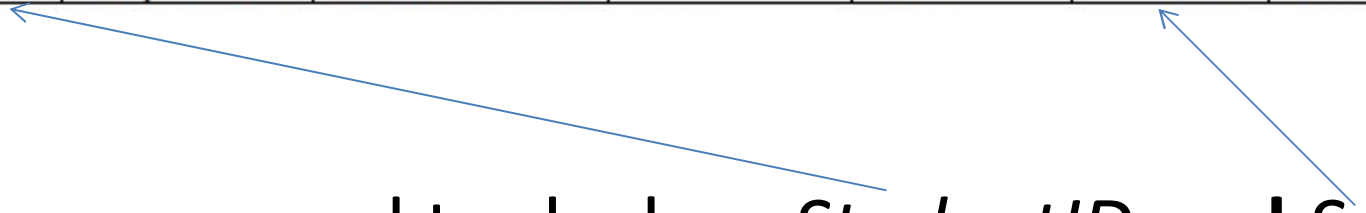
| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | English<br>Maths<br>Info Tech | $50<br>$50<br>$100 | B<br>A<br>B+ |

| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | English | $50 | B |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Maths | $50 | A |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Info Tech | $100 | B+ |

But now look – is the *studentID* primary key still valid?

# No – the studentID no longer uniquely identifies each row

| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | English | $50 | B |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Maths | $50 | A |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Info Tech | $100 | B+ |



You now need to declare *StudentID* **and** *Subject* **together** to uniquely identify each row.

So the new **key** is StudentID *and* Subject.

# So. We now have 1NF.

| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | English | $50 | B |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Maths | $50 | A |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Info Tech | $100 | B+ |

# Is it 2NF?

(StudentID, Subject)

StudentID -> StudentName

**StudentName** and **Address** are dependent on studentID (which is part of the key)
This is good.

| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | English | $50 | B |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Maths | $50 | A |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Info Tech | $100 | B+ |

But they are **not** dependent on *Subject* (the *other* part of the key)

# And 2NF requires…

| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | English | $50 | B |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Maths | $50 | A |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Info Tech | $100 | B+ |

All non-key fields are dependent on the ENTIRE key (StudentID + Subject)

# So it's not 2NF

| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | English | $50 | B |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Maths | $50 | A |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Info Tech | $100 | B+ |

How can we fix it?

# Make new tables

- Make a new table for each primary key field
- Give each new table its own primary key
- Move columns from the original table to the new table that matches their primary key

# Step 1

| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | English | $50 | B |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Maths | $50 | A |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Info Tech | $100 | B+ |

STUDENT TABLE (key = StudentID)

# Step 2

| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | English | $50 | B |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Maths | $50 | A |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Info Tech | $100 | B+ |

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

SUBJECTS TABLE (key = Subject)

# Step 3

| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | English | $50 | B |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Maths | $50 | A |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Info Tech | $100 | B+ |

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

SUBJECTS TABLE (key = Subject)

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

RESULTS TABLE (key = StudentID + Subject)

# Step 3

| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | English | $50 | B |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Maths | $50 | A |
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | Info Tech | $100 | B+ |

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

SUBJECTS TABLE (key = Subject)

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

RESULTS TABLE (key = StudentID + Subject)

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

# Step 4 - relationships

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

SUBJECTS TABLE (key = Subject)

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

# Step 4 - cardinality

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

Each student can only appear ONCE in the student table

SUBJECTS TABLE (key = Subject)

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

# Step 4 - cardinality

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

SUBJECTS TABLE (key = Subject)

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

1

Each subject can only appear ONCE in the subjects table

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

# Step 4 - cardinality

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

SUBJECTS TABLE (key = Subject)

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

1

A subject can be listed MANY times in the results table (for different students)

∞

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

# Step 4 - cardinality

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

SUBJECTS TABLE (key = Subject)

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

1

A student can be listed MANY times in the results table (for different subjects)

∞

∞

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

# A 2NF check

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

∞

SUBJECTS TABLE (key = Subject)

1

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

∞

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

**SubjectCost** is only dependent on the primary key, *Subject*

# A 2NF check

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

∞

SUBJECTS TABLE (key = Subject)

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

1

∞

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

**Grade** is only dependent on the primary key (*StudentID* + *Subject*)

# A 2NF check

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

∞

**Name, Address** are only dependent on the primary key (*StudentID*)

SUBJECTS TABLE (key = Subject)

1

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

∞

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

∞

**So it is 2NF!**

SUBJECTS TABLE (key = Subject)

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

1

∞

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

# But is it 3NF?

# A 3NF check

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

**Oh oh...**

**What?**

SUBJECTS TABLE (key = Subject)

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

1

∞

∞

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

# A 3NF check

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

**HouseName** is dependent on both ***StudentID + HouseColour***

SUBJECTS TABLE (key = Subject)

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

1

∞

∞

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

# A 3NF check

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

***Or* HouseColour** is dependent on both **StudentID + HouseName**

SUBJECTS TABLE (key = Subject)

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

1

∞

∞

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

# A 3NF check

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

*But either way, non-key fields are dependent on MORE THAN THE PRIMARY KEY (StudentID)*

SUBJECTS TABLE (key = Subject)

1

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

∞

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

# A 3NF check

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

*And 3NF says that non-key fields must depend on **nothing but the key***

SUBJECTS TABLE (key = Subject)

| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

1

∞

∞

| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

# A 3NF check

STUDENT TABLE (key = StudentID)

| StudentID | StudentName | Address | HouseName | HouseColor |
|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red |

1

∞

*WHAT DO WE DO?*

SUBJECTS TABLE (key = Subject)

1

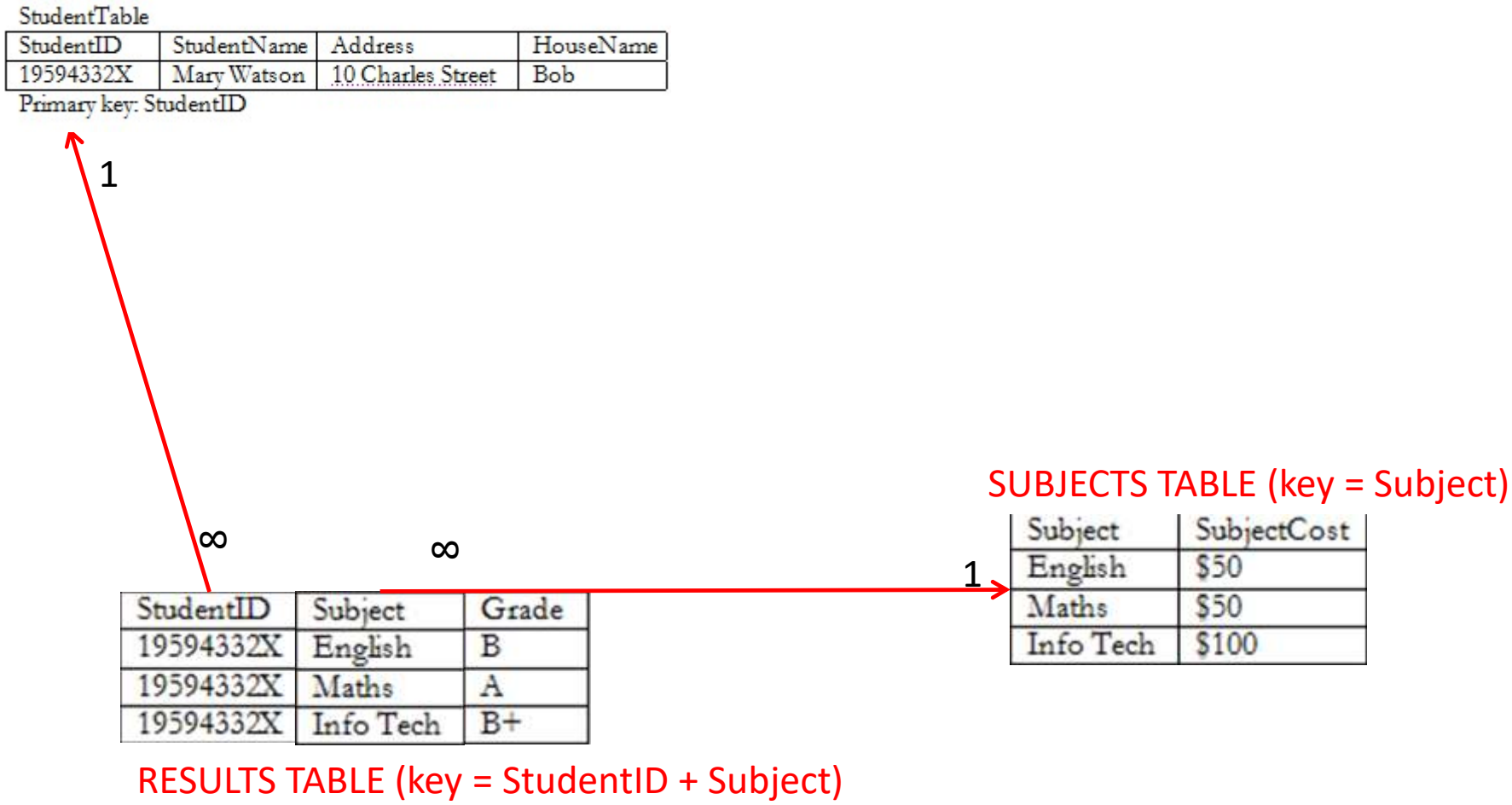
| Subject | SubjectCost |
|---|---|
| English | $50 |
| Maths | $50 |
| Info Tech | $100 |

∞

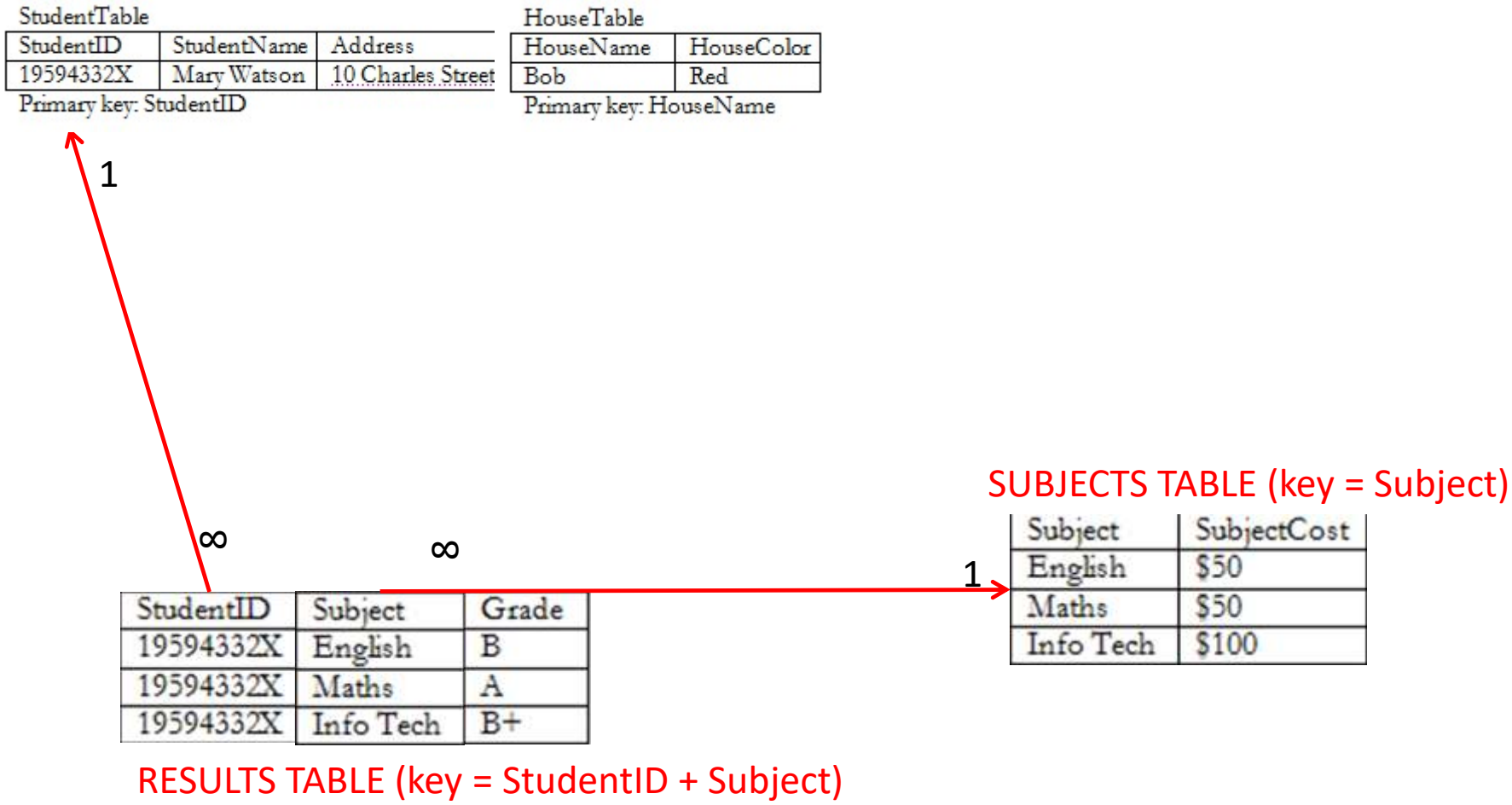
| StudentID | Subject | Grade |
|---|---|---|
| 19594332X | English | B |
| 19594332X | Maths | A |
| 19594332X | Info Tech | B+ |

RESULTS TABLE (key = StudentID + Subject)

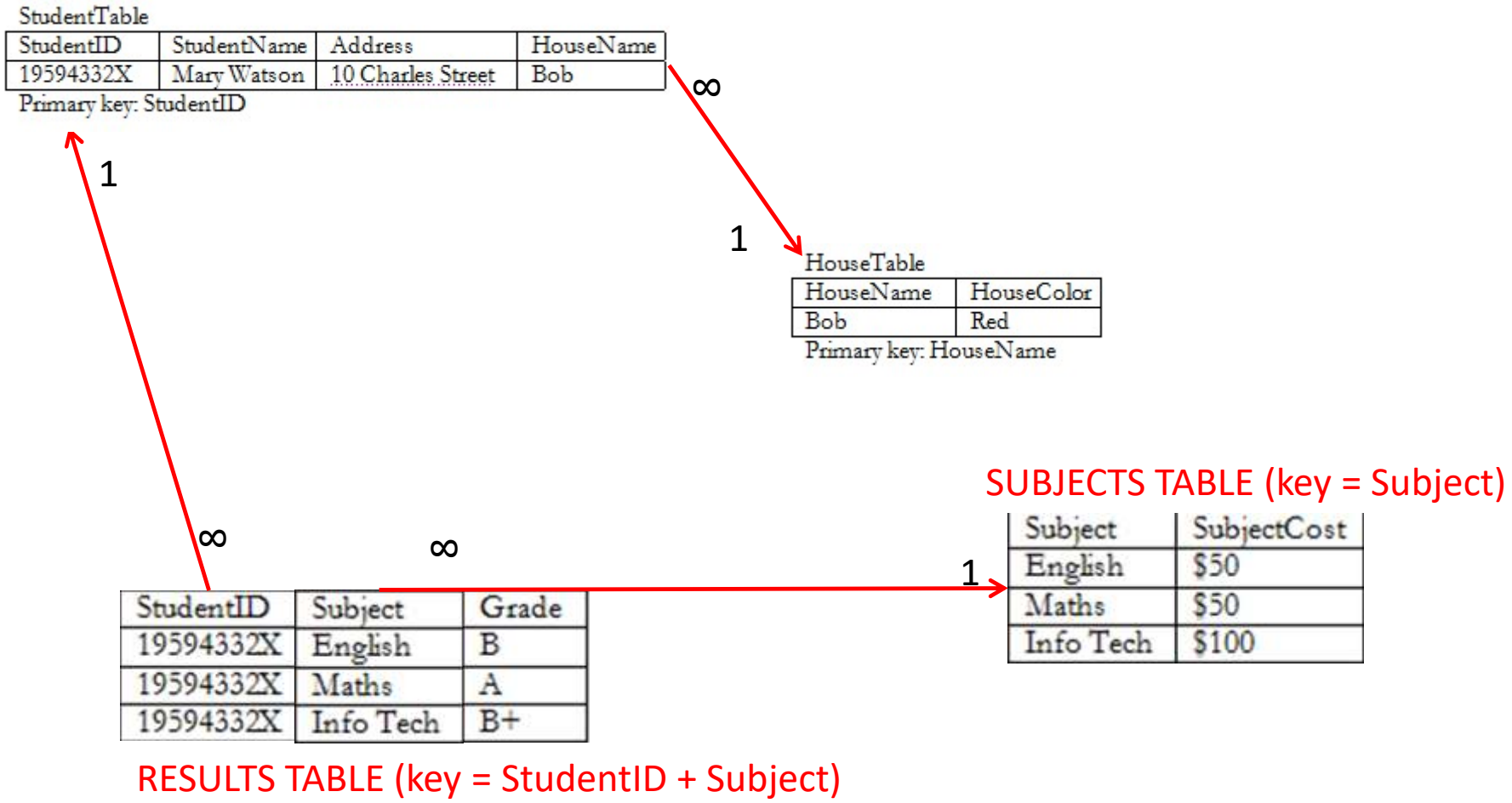
# Again, carve off the offending fields



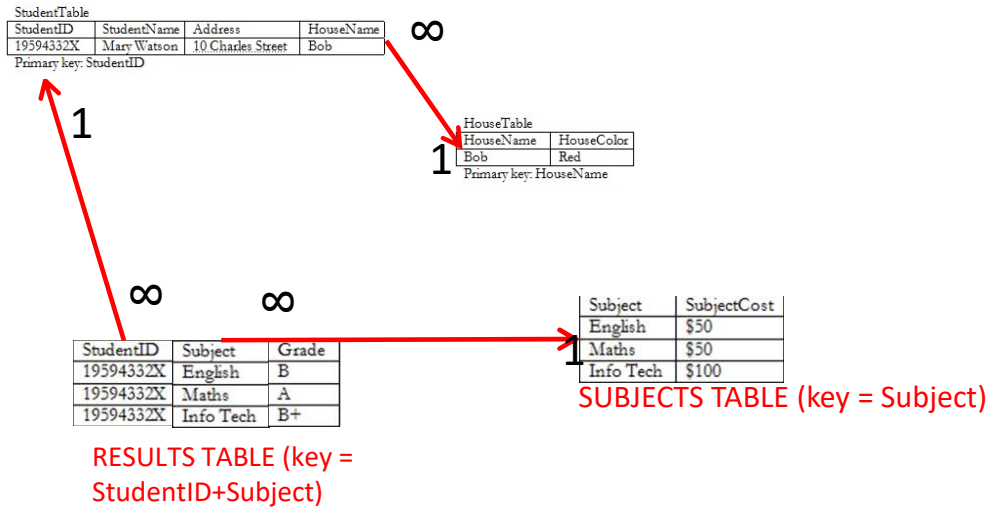
# A 3NF fix


StudentTable
StudentID | StudentName | Address
19594332X | Mary Watson | 10 Charles Street
Primary key: StudentID
HouseTable
HouseName | HouseColor
Bob | Red
Primary key: HouseName
1
∞
∞
1
SUBJECTS TABLE (key = Subject)
Subject | SubjectCost
English | $50
Maths | $50
Info Tech | $100
StudentID | Subject | Grade
19594332X | English | B
19594332X | Maths | A
19594332X | Info Tech | B+
RESULTS TABLE (key = StudentID + Subject)

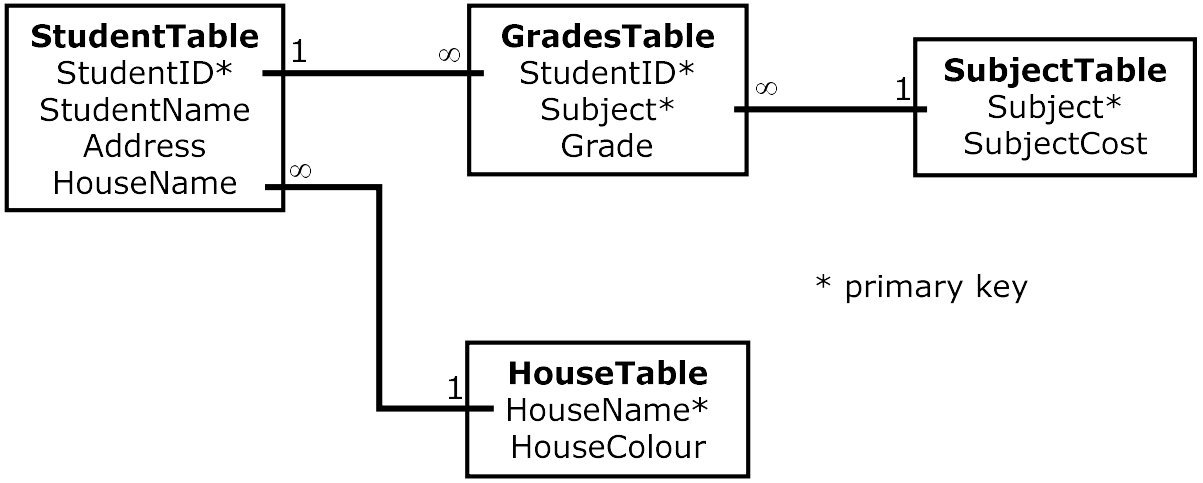
# A 3NF fix


StudentTable
StudentID | StudentName | Address | HouseName
19594332X | Mary Watson | 10 Charles Street | Bob
Primary key: StudentID
∞
1
1
HouseTable
HouseName | HouseColor
Bob | Red
Primary key: HouseName
∞
∞
1
SUBJECTS TABLE (key = Subject)
Subject | SubjectCost
English | $50
Maths | $50
Info Tech | $100
StudentID | Subject | Grade
19594332X | English | B
19594332X | Maths | A
19594332X | Info Tech | B+
RESULTS TABLE (key = StudentID + Subject)

# A 3NF win!


StudentTable
StudentID | StudentName | Address | HouseName
19594332X | Mary Watson | 10 Charles Street | Bob
Primary key: StudentID
∞
1
HouseTable
HouseName | HouseColor
Bob | Red
Primary key: HouseName
1
∞
∞
StudentID | Subject | Grade
19594332X | English | B
19594332X | Maths | A
19594332X | Info Tech | B+
RESULTS TABLE (key = StudentID+Subject)
1
Subject | SubjectCost
English | $50
Maths | $50
Info Tech | $100
SUBJECTS TABLE (key = Subject)

Or…

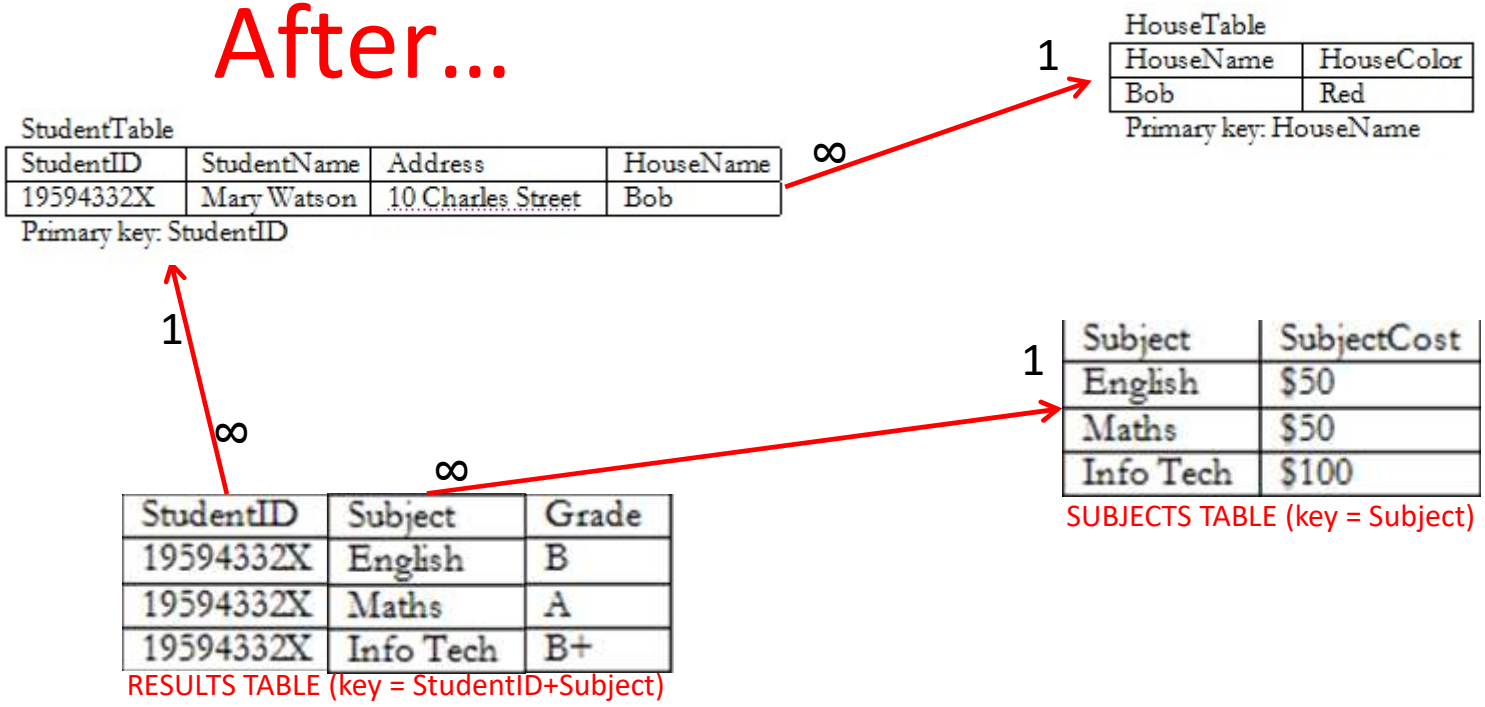

StudentTable
StudentID*
StudentName
Address
HouseName
1
∞
GradesTable
StudentID*
Subject*
Grade
∞
1
SubjectTable
Subject*
SubjectCost
∞
1
HouseTable
HouseName*
HouseColour
* primary key

# The Reveal

Before…

| StudentID | StudentName | Address | HouseName | HouseColor | Subject | SubjectCost | Grade |
|---|---|---|---|---|---|---|---|
| 19594332X | Mary Watson | 10 Charles Street | Bob | Red | English<br>Maths<br>Info Tech | $50<br>$50<br>$100 | B<br>A<br>B+ |

After…



# Example II

# Normalization Example

Orders(Order, Product, Quantity, UnitPrice, Customer, Address)

- We have a table representing orders in an online store
- Each row represents an item on a particular order
- Primary key is {Order, Product}

# Functional Dependencies

Orders(Order, Product, Quantity, UnitPrice, Customer, Address)

- Each order is for a single customer:
  - Order $\rightarrow$ Customer
- Each customer has a single address
  - Customer $\rightarrow$ Address
- Each product has a single price
  - Product $\rightarrow$ UnitPrice
- As Order $\rightarrow$ Customer and Customer $\rightarrow$ Address
  - Order $\rightarrow$ Address

Order -> Customer, Address

# 2NF Solution (I)

- ***First decomposition***
  - First table

| Order | Product | Quantity | UnitPrice |
|---|---|---|---|

  - Second table

| Order | Customer | Address |
|---|---|---|

# 2NF Solution (II)

- ***Second decomposition***
  - First table

| Order | Product | Quantity |
|---|---|---|

  - Second table

| Order | Customer | Address |
|---|---|---|

  - Third table

| Product | UnitPrice |
|---|---|

# 3NF

- In second table

| Order | Customer | Address |
|---|---|---|

  - Customer $\rightarrow$ Address
- Split second table into

| Order | Customer |
|---|---|

| Customer | Address |
|---|---|

# Normalization to 2NF

- Second normal form means no partial dependencies on candidate keys
  - {Order} $\rightarrow$ {Customer, Address}
  - {Product} $\rightarrow$ {UnitPrice}

- To remove the first FD we project over

  {Order, Customer, Address} (R1)

and

  {Order, Product, Quantity, UnitPrice} (R2)

# Normalization to 2NF

- R1 is now in 2NF, but there is still a partial FD in R2

  {Product} $\rightarrow$ {UnitPrice}

- To remove this we project over

  {Product, UnitPrice} (R3)

  and

  {Order, Product,  Quantity} (R4)

# Normalization to 3NF

- R has now been split into 3 relations - R1, R3, and R4
  - R3 and R4 are in 3NF
  - R1 has a transitive FD on its key
- To remove

  $\{Order\} \rightarrow \{Customer\} \rightarrow \{Address\}$
- we project R1 over
  - {Order, Customer}
  - {Customer, Address}

# Normalization

- 1NF:
  - {Order, Product, Customer, Address, Quantity, UnitPrice}
- 2NF:
  - {Order, Customer, Address}, {Product, UnitPrice}, and {Order, Product, Quantity}
- 3NF:
  - {Product, UnitPrice}, {Order, Product, Quantity}, {Order, Customer}, and {Customer, Address}