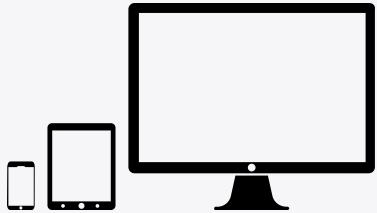


# System Design Fundamentals

## Data Normalization

Client Application



Network Layer



Application Layer



Cache Layer



Database Layer



# Decomposition example

R (Book name, Author, Publisher, Reader)

FDs: Book name  $\Rightarrow$  Author & Publisher , Author  $\Rightarrow$  Publisher

R1(Book name, Author, Publisher)

R2(Book name, Reader)

# Properties of Decomposition

Decompose a relation R into relations R1 and R2,

## **Lossless**

$$R1 \bowtie R2 = R$$

## **Dependency preserving**

All dependencies of R either must be a part of R1 or R2 or must be derivable from combination of FD of R1 and R2.

# Keys

**Super Key** : Set of one or more columns, which can uniquely identify a row.

**Candidate key** : minimal super key i.e. no redundant attribute present.

# Normalization Forms

- 1NF
- 2NF
- 3NF
- Boyce–Codd normal form (BCNF)

# 1NF

All the entries are atomic

Book name	Author	Publisher	Reader
Harry Potter	JK Rowling	Bloomsbury	Arjun, Sunil
Sherlock holmes	Conan Doyle	George Newnes	Ashwin
Nancy drew	Edward Stratemeyer	Stratemeyer Syndicate	Krishna

Book name	Author	Publisher	Reader
Harry Potter	JK Rowling	Bloomsbury	Arjun
Harry Potter	JK Rowling	Bloomsbury	Sunil
Sherlock holmes	Conan Doyle	George Newnes	Ashwin
Nancy drew	Edward Stratemeyer	Stratemeyer Syndicate	Krishna

# 2NF

Every non-key column must depend on the whole key, not just part of it.

FDs: Book name => Author & Publisher , Author => Publisher

Book name	Author	Publisher	Reader
Harry Potter	JK Rowling	Bloomsbury	Arjun
Sherlock holmes	Conan Doyle	George Newnes	Ashwin
Nancy drew	Edward Stratemeyer	Stratemeyer Syndicate	Krishna

Book name	Author	Publisher
Harry Potter	JK Rowling	Bloomsbury
Sherlock holmes	Conan Doyle	George Newnes
Nancy drew	Edward Stratemeyer	Stratemeyer Syndicate

Book name	Reader
Harry Potter	Arjun
Sherlock holmes	Ashwin
Nancy drew	Krishna

# 3NF

## 2NF + no transitive dependencies

FDs: Book name => Author & Publisher , Author => Publisher

Book name	Author	Publisher	Reader
Harry Potter	JK Rowling	Bloomsbury	Arjun
Sherlock holmes	Conan Doyle	George Newnes	Ashwin
Nancy drew	Edward Stratemeyer	Stratemeyer Syndicate	Krishna

Book name	Author
Harry Potter	JK Rowling
Sherlock holmes	Conan Doyle
Nancy drew	Edward Stratemeyer

Author	Publisher
JK Rowling	Bloomsbury
Conan Doyle	George Newnes
Edward Stratemeyer	Stratemeyer Syndicate

Book name	Reader
Harry Potter	Arjun
Sherlock holmes	Ashwin
Nancy drew	Krishna



# Boyce–Codd normal form (BCNF)

For every one of its dependencies  $X \rightarrow Y$ ,  
 $X \rightarrow Y \Rightarrow Y$  is a subset of  $X$  or  $X$  is a superkey

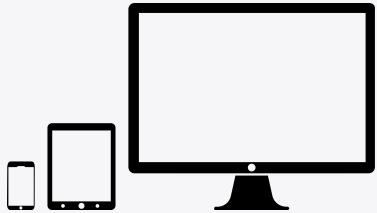
BCNF need not be a dependency preserving and lossless!

Eg:  $R(X,Y,Z)$ , FD  $\{XY \rightarrow Z, Z \rightarrow Y\}$ . only possible lossless decomposition is  $XZ$  and  $ZY$  which destroys the relation  $XY \rightarrow Z$

# System Design Fundamentals

## Data Normalization

Client Application



Network Layer



Application Layer



Cache Layer



Database Layer

