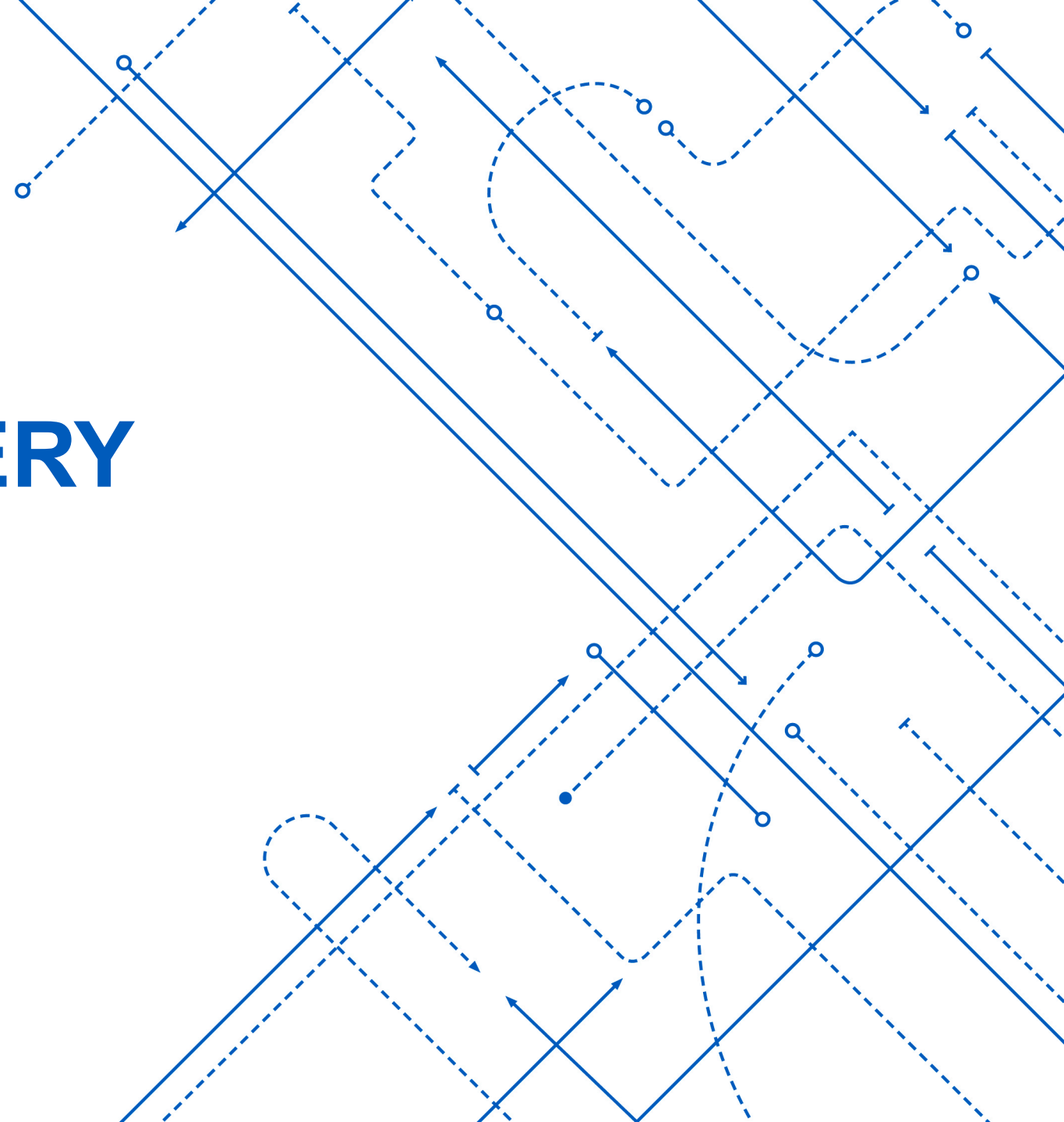

# CSE460/560 DATA MODELS AND QUERY LANGUAGES

## Relational Database Design

Cheng-En Chuang
(Slides Adopted from Jan Chomicki and Ning Deng)

University at Buffalo
Department of Computer Science and Engineering
School of Engineering and Applied Sciences

# Outline

1. Functional Dependencies
    1. Functional Dependencies
    2. Inference of FDs
2. Normal Forms
    1. BCNF and 3NF

# Outline

1. Functional Dependencies
    1. **Functional Dependencies**
    2. Inference of FDs
2. Normal Forms
    1. BCNF and 3NF

# Functional Dependencies (FDs)

- Keys of Relations
  - Given a set of attributes $\{A_1, A_2, \ldots, A_n\}$
    - Impossible for two distinct tuples of R
      - Agree on all $\{A_1, A_2, \ldots, A_n\}$
    - No proper subset of $\{A_1, A_2, \ldots, A_n\}$ determines all other attributes
      - Key must be ***minimal***
  - Key {title, year, starName}
    - Is {title, year} a key?

| *title* | *year* | *length* | *genre* | *studioName* | *starName* |
|---|---|---|---|---|---|
| Star Wars | 1977 | 124 | SciFi | Fox | Carrie Fisher |
| Star Wars | 1977 | 124 | SciFi | Fox | Mark Hamill |
| Star Wars | 1977 | 124 | SciFi | Fox | Harrison Ford |
| Gone With the Wind | 1939 | 231 | drama | MGM | Vivien Leigh |
| Wayne's World | 1992 | 95 | comedy | Paramount | Dana Carvey |
| Wayne's World | 1992 | 95 | comedy | Paramount | Mike Meyers |

# Functional Dependencies (FDs)

- Superkeys
  - Superset of a key
    - Every key is a superkey
    - Some superkeys are not keys
      - Why?
  - 🍭Example of superkey?

| *title* | *year* | *length* | *genre* | *studioName* | *starName* |
|---|---|---|---|---|---|
| Star Wars | 1977 | 124 | SciFi | Fox | Carrie Fisher |
| Star Wars | 1977 | 124 | SciFi | Fox | Mark Hamill |
| Star Wars | 1977 | 124 | SciFi | Fox | Harrison Ford |
| Gone With the Wind | 1939 | 231 | drama | MGM | Vivien Leigh |
| Wayne's World | 1992 | 95 | comedy | Paramount | Dana Carvey |
| Wayne's World | 1992 | 95 | comedy | Paramount | Mike Meyers |

# Functional Dependencies (FDs)

- Dependency Implication
  - A set of FDs *F* implies an FD $X \rightarrow Y$
    - If every relations instance that satisfies all dependencies in F
      - Also satisfies $X \rightarrow Y$
  - Notation
    - $F \vDash X \rightarrow Y$
  - Closure of a dependency set *F*
    - $F^+ = \{X \rightarrow Y : F \vDash X \rightarrow Y\}$
    - Suppose $\{A_1, \ldots, A_n\}$ is a set of attributes and F is a set of FDs
      - The closure of $\{A_1, \ldots, A_n\}$ under *F* is set of attribute *B*
        - such that every relation that satisfies all the *FDs* in *F* also satisfies $A_1, \ldots, A_n \rightarrow B$
    - Notation: $\{A_1, \ldots, A_n\}^+$

# Functional Dependencies (FDs)

- Formal Specification of Key
  - $X \subseteq \{A_1, \ldots, A_n\}$ is a key of R if
    - The dependency $X \rightarrow A_1, \ldots, A_n$ is in $F^+$
    - For all proper subsets *Y* of *X*
      - The dependency $Y \rightarrow A_1, \ldots, A_n$ is not in $F^+$
- Notations
  - Superkey: supetset of a key
  - Primary key: one designated key
  - Candidate key: one of the keys

# Outline

1. Functional Dependencies
    1. Functional Dependencies
    2. **Inference of FDs**
2. Normal Forms
    1. **BCNF and 3NF**

# Inference of FDs

| *title* | *year* | *length* | *genre* | *studioName* | *starName* |
|---|---|---|---|---|---|
| Star Wars | 1977 | 124 | SciFi | Fox | Carrie Fisher |
| Star Wars | 1977 | 124 | SciFi | Fox | Mark Hamill |
| Star Wars | 1977 | 124 | SciFi | Fox | Harrison Ford |
| Gone With the Wind | 1939 | 231 | drama | MGM | Vivien Leigh |
| Wayne's World | 1992 | 95 | comedy | Paramount | Dana Carvey |
| Wayne's World | 1992 | 95 | comedy | Paramount | Mike Meyers |

- How to tell **whether** $\boldsymbol{X \rightarrow Y \in F^{+}}$
- Inference Rules (Armstrong Axioms)
  - Reflexivity: infer $X \rightarrow Y \; if \; Y \subseteq X \subseteq attr(R)$ (trivial dependency)
  - Augmentations: from $X \rightarrow Y$ infer $XZ \rightarrow YZ$ if $Z \subseteq attr(R)$
  - Transitivity: from $X \rightarrow Y$ and $Y \rightarrow Z$, infer $X \rightarrow Z$
- Armstrong axioms are
  - Sound: if $X \rightarrow Y$ is derived from *F*, then $X \rightarrow Y \in F^{+}$
  - Complete: if $X \rightarrow Y \in F^{+}$, then $X \rightarrow Y$ is derived from *F*
- Additional (implied) inference rules
  - Union: from $X \rightarrow Y$ and $X \rightarrow Z$, infer $X \rightarrow YZ$
  - Decomposition: from $X \rightarrow Y$ infer $X \rightarrow Z$, if $Z \subseteq Y$

# Outline

1. Functional Dependencies
    1. Functional Dependencies
    2. Inference of FDs
2. Normal Forms
    1. **BCNF and 3NF**

# Normal Forms

- To eliminate or reduce redundancy in database relations
  - 1NF: a relation is in 1NF
    - Every attribute in that relation is atomic
      - Reduce the need for restructuring the collection of relations, as new types of data are introduced
      - Make the relational model more informative to users
  - 2NF: a relation is in 2NF
    - iff it's in 1NF
    - And has no partial dependency
      - No non-prime attribute (attributes which are not part of any candidate key) is dependent on any proper subset of any candidate key of the relation

# Normal Forms (3NF, BCNF)

- A relation *R* is in 3NF if for every non-trivial FD $X \rightarrow A \in F$
  - *either*
    - *X* contains a key of *R*
    - *A* is part of some key of *R*
- A relation *R* is in BCNF if
  - For every non-trivial FD $X \rightarrow A \in F$
    - *X* contains a key of *R*
  - BCNF does not contain a redundancy
- 3NF vs. BCNF
  - If R is in BCNF, it is also in 3NF
  - A relations the are in 3NF but not in BCNF

# Normal Forms (3NF, BCNF)

- 3NF vs. BCNF
  - A relations the are in 3NF but not in BCNF
    - R(A,B,C)
    - FD: {A,B} -> {C}, {C} -> {B}
    - Candidate Keys: {A,B}, {A,C}
    - In 3NF
      - {A,B} -> {C}  (*A* contains a key of *R)*
      - {C} -> {B} (*B* is part of some key of *R*)
    - Not BCNF
      - {C} -> {B} (C does NOT contain a key of R)

# Recommended Reading

Database Systems: The Complete Book
Chapter 3.1 – 3.6