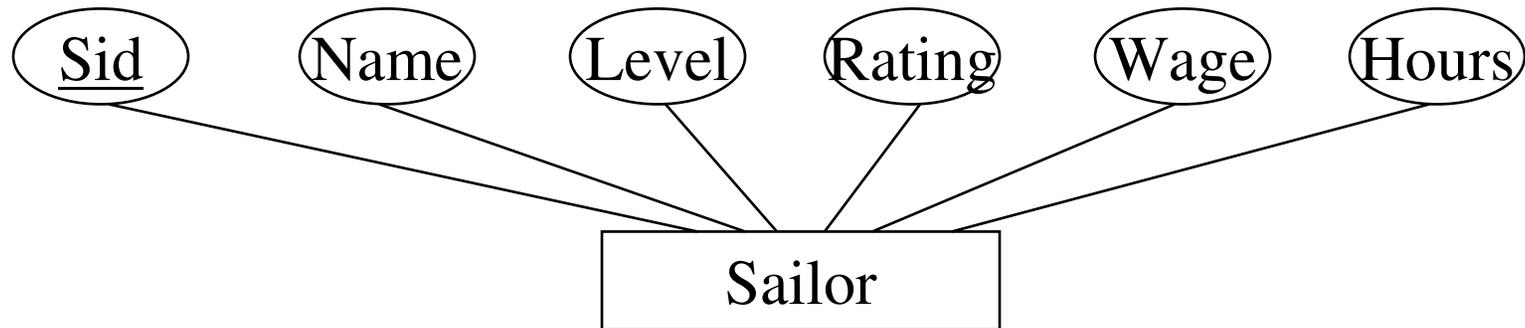


Schema Design and Normal Forms

From Chapter 19

Entity-Relationship Diagram



Data Redundancy

S	N	L	R	W	H
123-22-3666	Attishoo	48	8	10	40
231-31-5368	Smiley	22	8	10	30
131-24-3650	Smethurst	35	5	7	30
434-26-3751	Guldu	35	5	7	32
612-67-4134	Madayan	35	8	10	40

- Application constraint: all sailors with the same rating have the same wage ($R \rightarrow W$)
- Problems due to data redundancy?

Problems due to Data Redundancy

- Problems due to $R \rightarrow W$:
 - Update anomaly: Can we change W in just the first tuple of Sailor?
 - Insertion anomaly: What if we want to insert a new rating and wage, but there is no sailor for this rating?
 - Deletion anomaly: If we delete all sailors with rating 5, we lose the information about the wage for rating 5!
- Solution?

Relation Decomposition

S	N	L	R	W	H
123-22-3666	Attishoo	48	8	10	40
231-31-5368	Smiley	22	8	10	30
131-24-3650	Smethurst	35	5	7	30
434-26-3751	Guldu	35	5	7	32
612-67-4134	Madayan	35	8	10	40

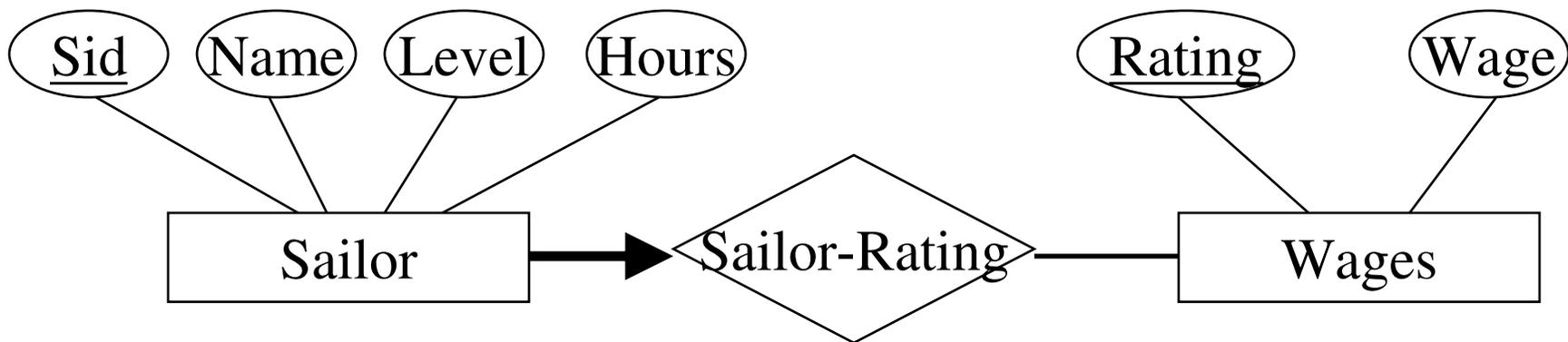
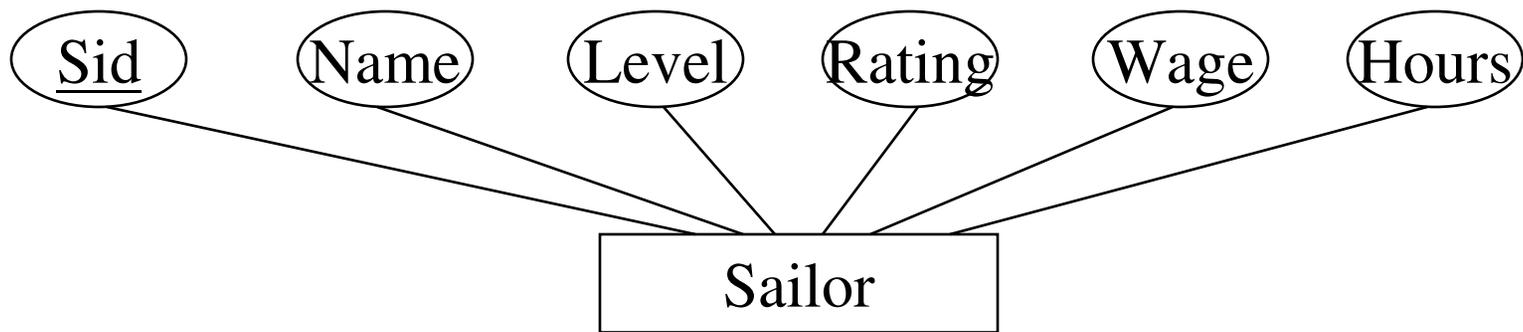
S	N	L	R	H
123-22-3666	Attishoo	48	8	40
231-31-5368	Smiley	22	8	30
131-24-3650	Smethurst	35	5	30
434-26-3751	Guldu	35	5	32
612-67-4134	Madayan	35	8	40

Wages

R	W
8	10
5	7

Problem?

Modifying ER Diagram



Decisions

- Do we have to decompose / merge?
- How do we identify problems caused by redundancy?
 - Functional dependencies

Functional Dependency (FD)

- $X \rightarrow Y$ (X determines Y)
 - If same value for X then same value for Y
- A functional dependency is a statement about *all allowable* instances of a relation!
- Examples:
 - Any primary or candidate key
 - Alpha \rightarrow (Name, Class, DateOfBirth)
 - Rating \rightarrow Wage
 - (NbHours, HourlyPrice) \rightarrow Charge

Functional Dependency (FD) Rules

- If $A \rightarrow (B, C)$, then $A \rightarrow B$ and $A \rightarrow C$
- If $(A, B) \rightarrow C$, then
 - neither A nor B determines C by itself
 - A and B determine C

Normal Forms

- Relations are categorized as a **normal form** based on which modification anomalies or other problems they are subject to:

Source of Anomaly	Normal Forms	Design Principles
Functional dependencies	1NF, 2NF, 3NF, BCNF	BCNF: Design tables so that every determinant is a candidate key
Multivalued dependencies	4NF	4NF: Move each multivalued dependency to a table of its own
Data constraints and oddities	5NF, DK/NF	DK/NF: Make every constraint a logical consequence of candidate keys and domains

Normal Forms

Number	Last Name	First Name	Email	Company	Wing
190	Smith	John	jsmith@usna.edu	12	2
673	Doe	Jane	jdoe@usna.edu	7	4
312	Doe	Bob	bred@usna.edu	6	6
152	Johnson	Matt	mat@usna.edu	7	4

- **1NF** – A table that qualifies as a relation is in 1NF (Ex: no column contains a list!)
- **Boyce-Codd Normal Form (BCNF)** – A relation is in BCNF if every determinant is a (candidate) key

“I swear to construct my tables so that all nonkey columns are dependent on the key, the whole key and nothing but the key, so help me Codd.”

Eliminating Modification Anomalies from Functional Dependencies in Relations

- Put all relations into Boyce-Codd Normal Form (BCNF):

1. Identify every functional dependency
2. Identify every candidate key
3. If there is a functional dependency that has a determinant that is not a candidate key:
 - A. Move the columns of that functional dependency to a new relation
 - B. Make the determinant of that functional dependency the primary key of the new relation
 - C. Leave a copy of the determinant as a foreign key in the original relation
 - D. Create a referential integrity constraint between the original relation and the new relation
4. Repeat step 3 until every determinant of every relation is a candidate key

(Note: In step 3, if there is more than one such functional dependency, start with the one with the most columns.)

Putting a Relation into BCNF: ASSIGNMENT_GRADES

Alpha	Assignment	Points	PointsTotal
129722	QUIZ1	10	10
129722	QUIZ2	2.5	10
129722	QUIZ3	2	20
122422	QUIZ1	6	10
122422	QUIZ2	7	10
122422	QUIZ3	18	20
129936	QUIZ1	6	10
129936	QUIZ2	8	10
129936	QUIZ3	20	20

Putting a Relation into BCNF: ASSIGNMENT_GRADES

ASSIGNMENT_GRADES (Alpha, Assignment,
Points, PointsTotal)

(Alpha, Assignment) → (Points, PointsTotal)
Assignment → (PointsTotal)

ASSIGNMENT (Assignment, PointsTotal)
GRADES (Alpha, Assignment (FK), Points)

Class Exercise

- $R(A, B, C, D, E, F)$
 $A \rightarrow (B, C, D, E, F)$
 $B \rightarrow C$
 $(D, E) \rightarrow F$
- Is A a key? Why?
- Is R in BCNF? Why?
- If R not in BCNF, decompose to BCNF

Class Exercise

ID	Name	University	MainCampus
1	John Smith	Cornell	Ithaca
2	John Smith	MIT	Boston
3	Matt Johnson	Ithaca College	Ithaca
4	Chris Brown	USNA	Annapolis
5	Jane Doe	Cornell	Ithaca
6	Ric Crabbe	USNA	Annapolis

- Example of deletion anomaly?
- Do these FDs hold? Why?
 - ID → University
 - Name → ID
 - University → MainCampus
 - MainCampus → Name

Summary of Schema Refinement

- BCNF implies free of redundancies due to FDs
- If a relation is not in BCNF, we can try to decompose it into a collection of BCNF relations.
- Decompositions should be carried out and/or re-examined keeping *performance issues* in mind