DATABASES

Objectives

- After this unit you will be able to explain the concepts of:
  - Database
  - database system
  - Database schema
  - Table/relation
  - Query (language)
  - SQL

The Task

- Manifold information,
  accessed by users in manifold (often unanticipated) ways
  - Standard task
  - Many variations
- Solution: individually configurable standard tool
- *...is this marketing speak???*

[image: simplisafe.com]

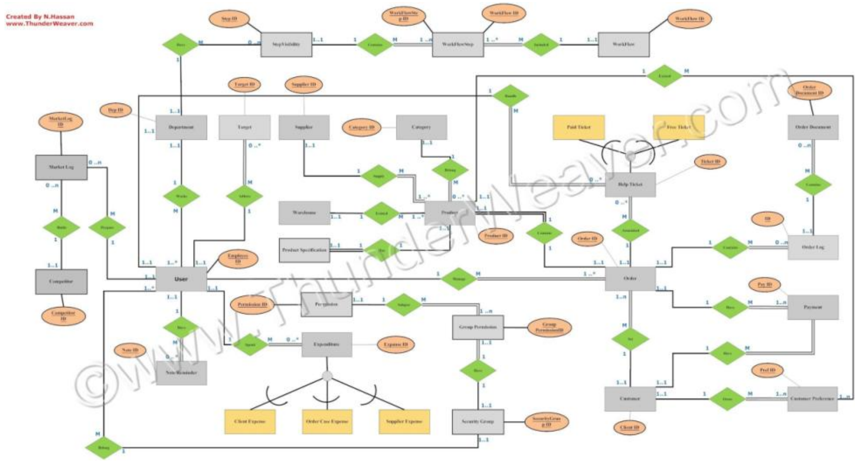
Describing Database Contents

- „A database is an organized collection of data“ [Wikipedia]

→ need to describe what database will contain, on „high level“

  - No implementation details like table schemas (later) or file structures (not here)

- canonical: Entity-Relationship Model (Peter Chen, 1976)

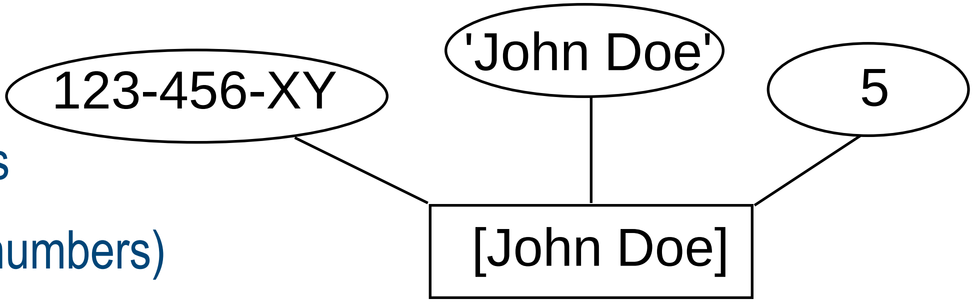


Created By N.Hassan
www.ThunderWeaver.com

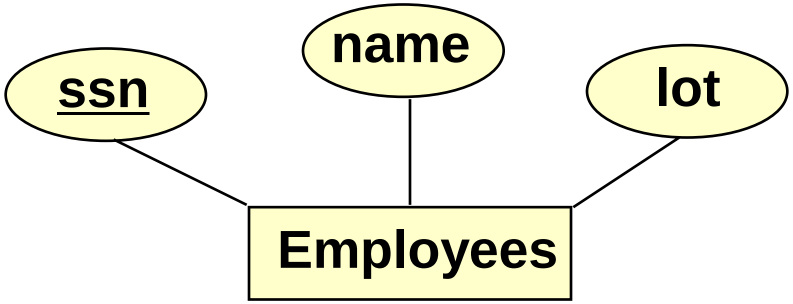
- Alternative: UML (Unified Modelling Language)

Entities

- Entity: Real-world object
  - distinguishable from other objects
  - Simple attribute values (strings, numbers)

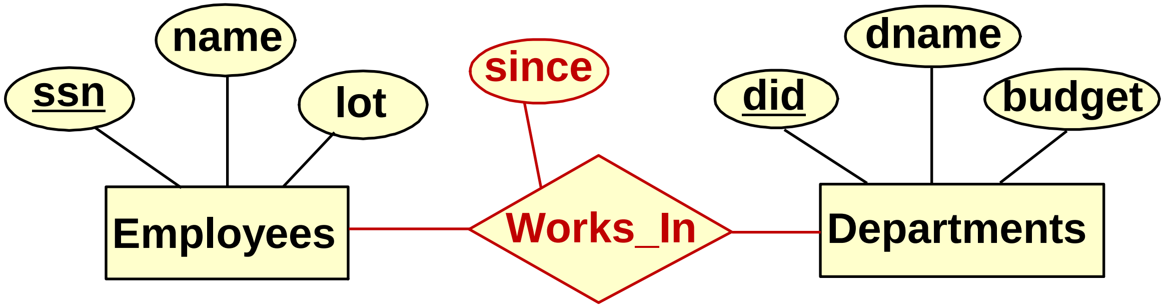


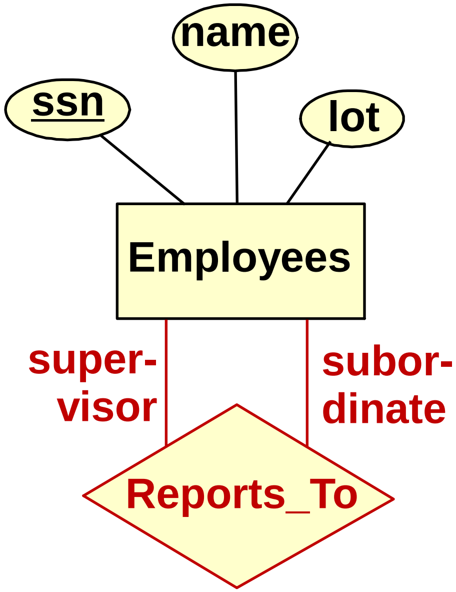
- Entity set: collection of similar entities
  - same set of attributes
    - *Until we consider ISA hierarchies*
  - key = unique identifier
  - attribute has domain = data type



Relationships

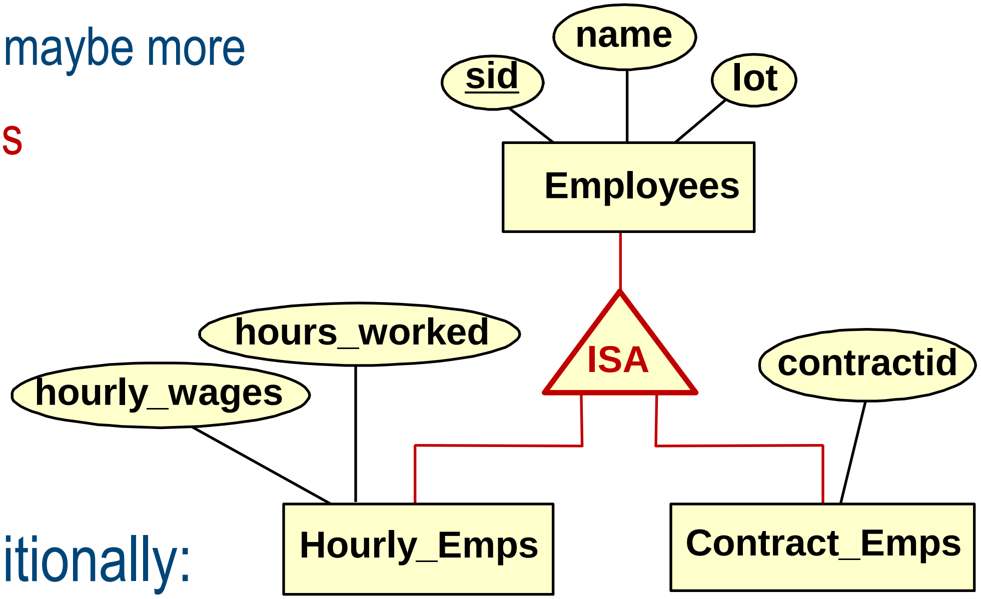
- Relationship connects two or more entities
  - Ex: Attishoo works_in Pharmacy_Department
- Relationship Set: Collection of similar relationships
  - each relationship in R∈E1×...×En involves entities e1∈E1, ..., en∈En
  - Roles to differentiate different “legs”





ISA (`is a') Hierarchies

- A ISA B :<=> every A entity is also a B entity ("A inherits from B")
  - A entities have B attributes, plus maybe more
  - A is called subclass, B superclass



- Constraints to be captured additionally:
  - Overlap constraints: Can Joe be an Hourly_Emps as well as a Contract_Emps entity? (Allowed/disallowed)
  - Covering constraints: Does every Employees entity also have to be an Hourly_Emps or a Contract_Emps entity? (Yes/no)

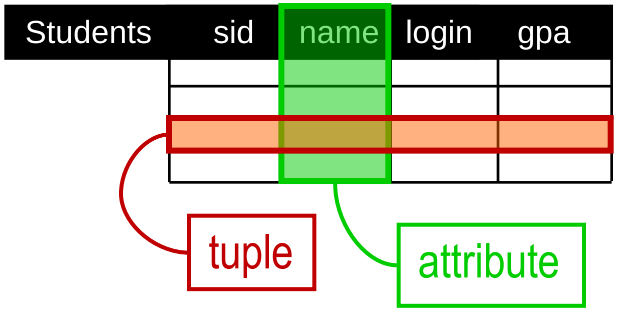
Relational Database: Definitions

- Technically: Relation made up of 2 parts:
  - Schema: name of relation, plus name & type of each column
    - *Ex: Students(sid: integer, name: string, login: string, gpa: real)*
  - Table: data arranged in rows & columns

does not change often

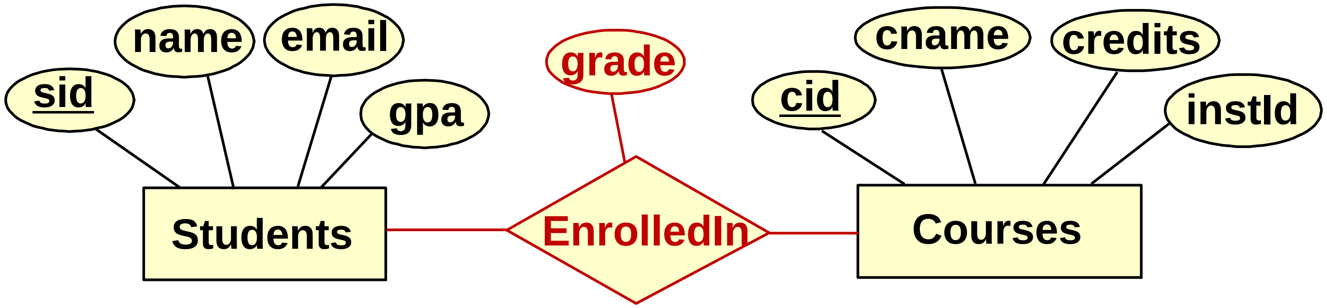
changes all the time

- Database = a set of tables



- Mathematically, table = relation:
  - Let A1, …, An (n>0) be value sets
  - relation $R \subseteq A1 \times \ldots \times An = \{ (a1,\ldots,an) \mid a1 \in A1, \ldots, an \in An \}$
- Schema maintained by database administrator
  - Plus all other housekeeping work

Ex: University Miniworld



```
Students:
sid name     email     gpa
-------------------------
42  Neo      n.eo      1.7
43  Trinity  t.rinity  2.3
44  Cypher   c.ypher   3.0
```

```
EnrolledIn:
sid cid grade
-------------
42  142  1.3
43  142  2.0
44  142  2.0
42  143  1.0
44  143  2.0
43  144  1.7
```

```
Courses:
cid cname  credits  instId
----------------------------
142 AlgDS  5        242
143 DBWA   5        243
144 CGVis  5        244
```

SQL, Structured English Query Language

- "students with GPA better than 3.0":

```
SELECT  *
FROM    Students
WHERE   gpa < 3.0
```

```
sid name     email     gpa
--------------------------
42  Neo      n.eo      1.7
43  Trinity  t.rinity  2.3
~~44  Cypher   c.ypher   3.0~~
```

- "names & emails of…" :

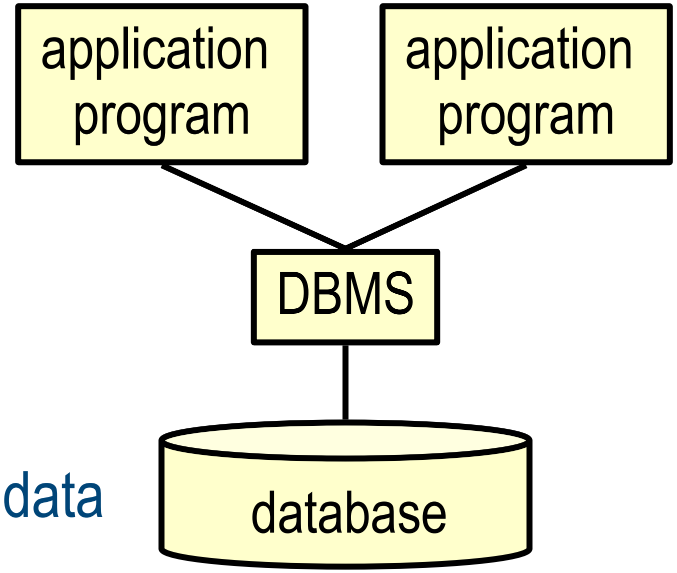
```
SELECT  name, email
FROM    Students
WHERE   gpa < 3.0
```

```
sid name     email     gpa
--------------------------
42  Neo      n.eo      1.7
43  Trinity  t.rinity  2.3
~~44  Cypher   c.ypher   3.0~~
```

SQL query accesses tables, query returns a table

Some Vocabulary

- Applications request data, update data, etc.
- Database [Management] System = DBMS
  = software to store and manage databases
  - …and no one else!
- Database = DB = an integrated collection of data
  - Stored on disk
  - With a well-described structure = schema



Importance of Schema

- data are structured (not simply text) + schema describes it
- → more complex evaluations are possible
  - Ex: How many homeworks has each student done?
- Access restrictions can be imposed on rows & columns ↳
- also used internally to speed up query processing

DBMS Technology

- Vendors:
  - Leading: Oracle, IBM, Microsoft
  - Further: Teradata, Sybase, SAP HANA, ...
- Open source:
  - PostgreSQL, MySQL/MariaDB, ...and many more
- Recently: NoSQL databases
  - NotOnlySQL

Summary

- Database (system) for managing data in a structured way
  - DB knows about its structure and how to maintain integrity
- High-level visual information modelling: Entity-Relationship Model
- Prevailing technology: Relational Databases
  - Relations = tables
  - Flexible queries on tables → SQL
- Main advantages of DBMSs over flat files & self-built systems:
  - Flexibility
  - Scalability
  - Information integration