

# DATA MODELING

---

AMIR DIRIN

## Kuvaus- ja mallintamismenetelmät TX00EC75-3006

### Perustiedot

**Tunnus**

TX00EC75-3006

**Ajoitus**

22.08.2022 - 16.10.2022

**Toimipiste**

Myllypurontie 1

**Opetuskieli**

Suomi

**Toteutustapa**

Lähiopetus

**Nimi**

Kuvaus- ja mallintamismenetelmät

**Ilmoittautumisaika**

02.05.2022 - 21.08.2022

**Yksikkö**

ICT ja tuotantotalous

**Opintopistemäärä**

5

# OPINTOJAKSON SISÄLTÖ

---

- rakenteiden mallintaminen ja kuvaaminen
- toiminnallisuuden mallintaminen ja kuvaaminen
- oliopohjainen mallintaminen ja kuvaaminen (UML)
- UML-työvälineohjelmiston käyttö.

## Opintojakson oppimistavoitteet

- Opintojakson suoritettuasi:
  1. ymmärrät datan ja ohjelmistojen mallinnuksen tavoitteet ja tarpeen
  2. tunnet ER- ja UML-kaaviotekniikat ja niiden käyttökohteet
  3. osaat laatia datan rakennetta sekä ohjelmistosi rakennetta ja toimintaa kuvaavia kuvauksia ja malleja
  4. osaat generoida mallinnetusta datasta relaatiotietokannan.
  5. osaat tulkita toisten kehittäjien tuottamia kuvauksia ja malleja
- Opintojaksolla opit siis
  - suunnittelemaan ja kuvamaan ohjelmistojen ratkaisuja
  - esittelemään ja kommunikoimaan ratkaisuja ammattimaisesti ja myös toisten kehittäjien ymmärtämällä tavalla
  - käyttämään ER- ja UML-kaaviotekniikoita omassa ohjelmistoprojektiasi sitä hyödyttävällä tavalla

- 
- Opintojakso kannattaa suorittaa samanaikaisesti Ohjelmistotuotantoprojekti 1 -opintojakson kanssa. Kurssin harjoitteet tukevat OTP-projektin etenemistä.

# ARVIOINTIPERUSTEET

---

- Opintojakson arvosana *muodostuu harjoituksista* (painoarvo 50%) ja *kokeesta* (painoarvo 50%).
- Opiskelijan on saatava *vähintään 40%* kurssin maksimipistemäärästä. Lisäksi koe on suoritettava hyväksytysti (vähintään 40% kokeen maksimipisteistä).
- Kukin harjoitus arvioidaan *asteikolla 0..3 pistettä*, ja arvioinnin perustana on ratkaisun vaatimusten- ja tavoitteidenmukaisuus. Harjoitus on palautettava Omassa olevaan määräaikaan mennessä, minkä jälkeen alkavalla tunnilla ratkaisu puretaan. Myöhässä palautetusta ratkaisusta voi saada korkeintaan puolet pisteistä. Tehtäväratkaisut on palautettava Omaan ladattavina tiedostoina (ei pilvipalvelulinkkeinä).
- Osa harjoituksista on tiimitehtäviä. Näiden osalta yksilökohtaisia pisteitä voidaan opintojakson lopussa painottaa jatkuvan näytön sekä itse- ja vertaisarvioinnin perusteella tilanteissa, joissa työpanokset ovat jakautuneet tiimin sisällä epätasaisesti.



- 
- opetusviikko (22.08.-28.08.):
    - Tietojärjestelmä, ohjelmisto ja data. Datan mallinnus ja käsitemalli. ERDPlus.
    - Käsitemallin muunnos tietokanta kaavioksi ja SQL-esitykseksi. Aikasidonnainen data.

# TODAY

---

- Why data modeling?
- Types of data modeling
- What is a database?
- What is data?
- Why not just Excel?
- What is RDBMS?
- Visa card in the US.
- Database
- Entity
- Attribute

# WHY DATA MODELING

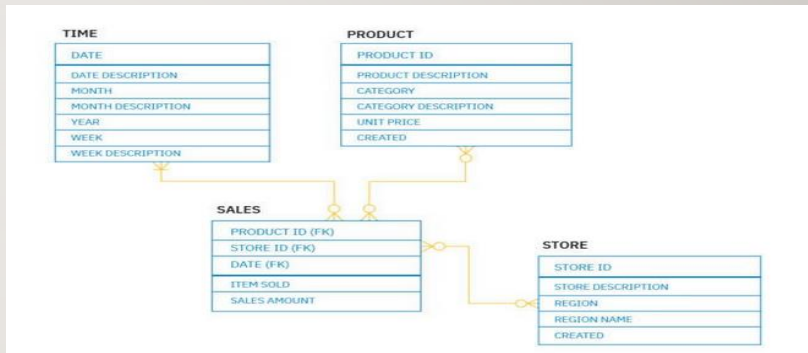
---

- **Data modeling** is an important stage of any software project because, without it, you cannot get a clear idea of what your database should look like and how your application will be built upon it.
  - High-Quality Applications
  - Reduce cost and time of application development
  - Early Detection of data issues and Errors
  - Faster application performance
  - Better Documentation

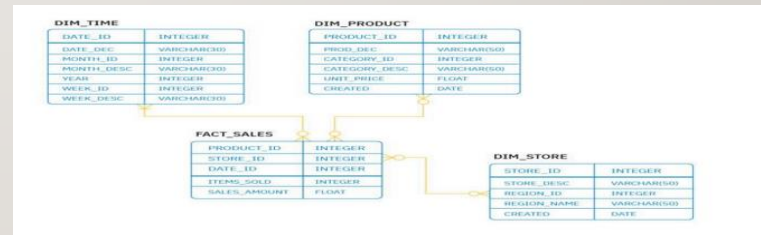


# TYPES OF DATA MODELING

- Conceptual (logical) data modeling



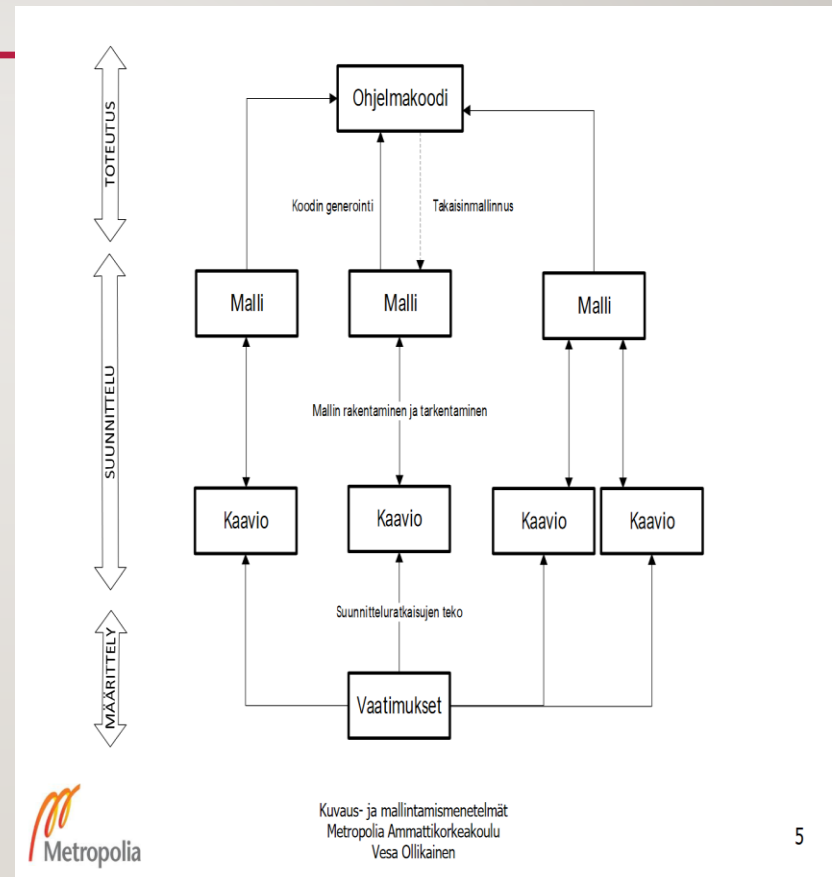
- Physical Data modeling



- Theoretical Data modeling

# STEPS IN DATA MODELING

- REVIEW THE BUSINESS CHALLENGE
- PULL THE RIGHT DATA FROM THE BUSINESS
- COLLECT AND ORGANIZE DATA
- CREATE A CONCEPTUAL MODEL
- BUILD THE LOGICAL DATABASE DESIGN
- BUILD THE PHYSICAL DATABASE DESIGN
- MAP STAKEHOLDERS AND THEIR REQUIREMENTS FOR THE DATA MODEL
- PERFORM A GAP ANALYSIS OF REQUIREMENTS VS. DATASETS
- DEPLOYMENT & DOCUMENTATION OF RESULTS
- MEASURE & MODIFY DATA MODEL TO MEET CHANGING REQUIREMENTS



# INFORMATION SYSTEM (TIETOJÄRJESTELMÄ)

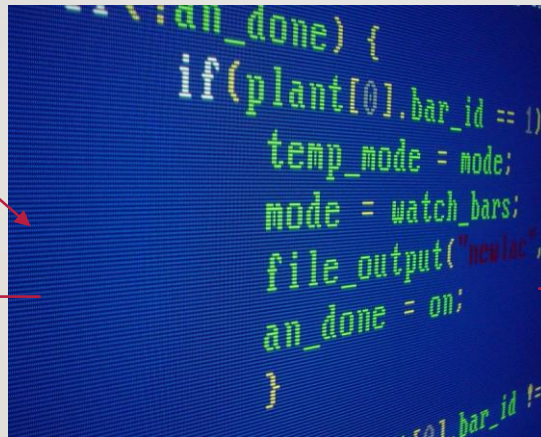
---

- 90% of the data on the internet has been created since 2016, according to an **IBM Marketing Cloud study**.
- See yourself
  - <https://www.internetworldstats.com/stats.htm>
- Data are generated all the time
  - Social Media : 550 new social media each minute meaning 1,209600 new data producing social media each day.
  - Tweets: 474000 tweets per minute in 2019
  - Youtube: 400 hours of new video each minute
  - Instagram: 100 milloin photos and videos every day
  - Facebook: 510000 comments, 293,000 statuses are updated every minute
  - 3.5 billion google searches each minute
  - SMS: 100 million messages each minute
- Do not agree with me see it yourself
  - <https://www.internetlivestats.com/>

# DATA IS MONEY



Physical Store



Programing to manipulate  
date



Retrieve the data / user interface

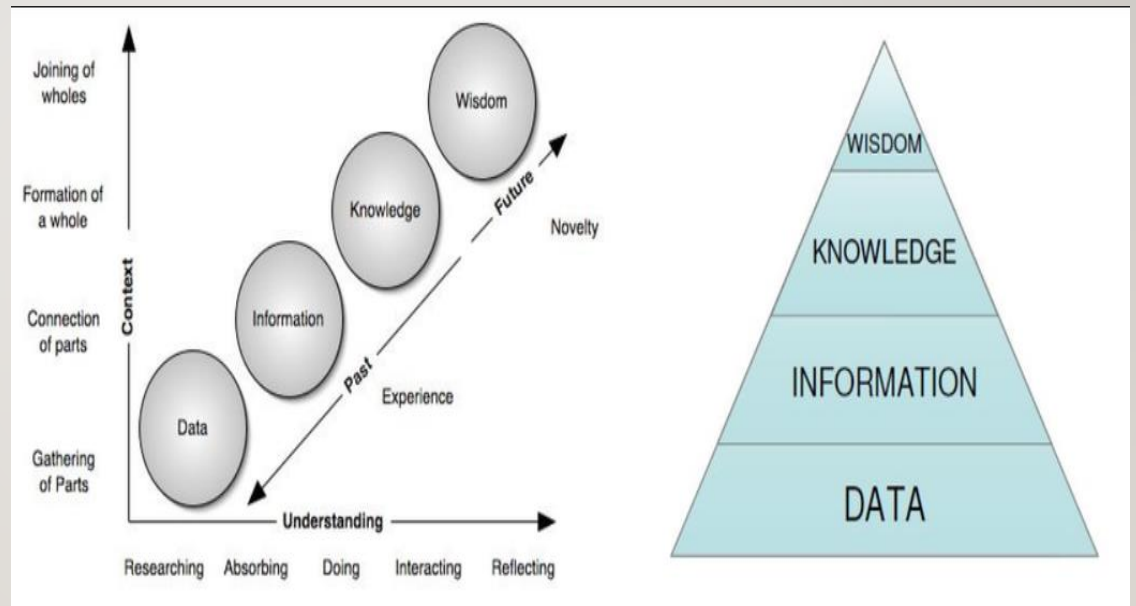


# EXAMPLE





- Data
- Information
- Knowledge



# DATABASES

---

- Spreadsheet
- ER database
- Big data

# SOFTWARE

---

- Computer programs and possible documents
- Data: data are facts that are used by programs to produce useful information
- Procedure: procedure are the policy that governs the operation of a computer system
- People mean the end user of every computer-based information system needs people if it is to be useful.

# SYSTEM

---

- System: Set of interrelated components that work together towards a particular goal.
- Human body is a system, or what?
- Esim: hengitus
- Computer!?

# ERDPLUS

---





# ERD

---



—+ = One

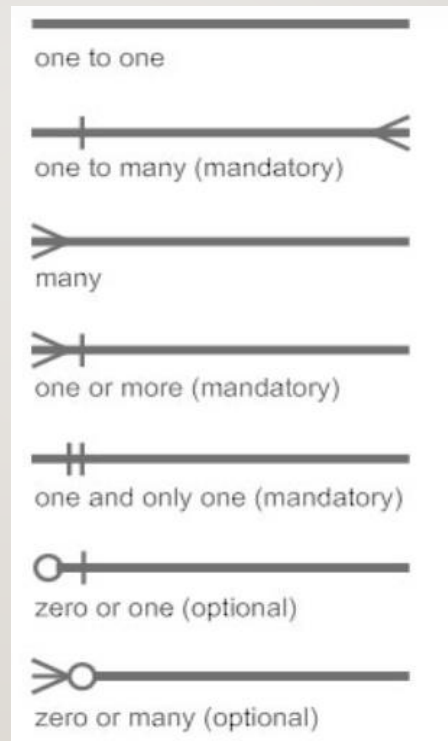
—○ = Zero or more, Optional

—> = Many

# RELATIONS

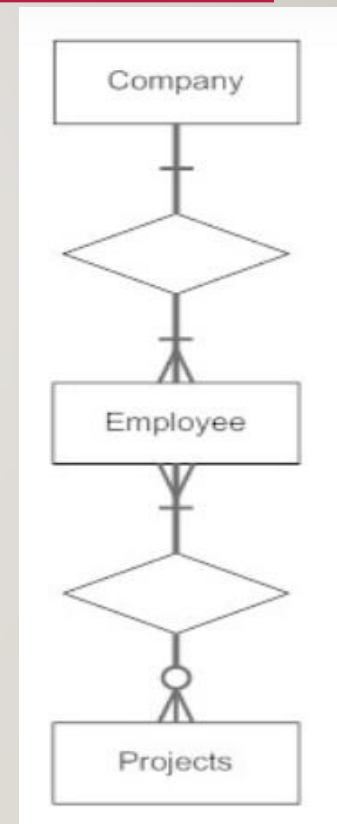
---

- one-to-one
- one-to-many
- many-to-many



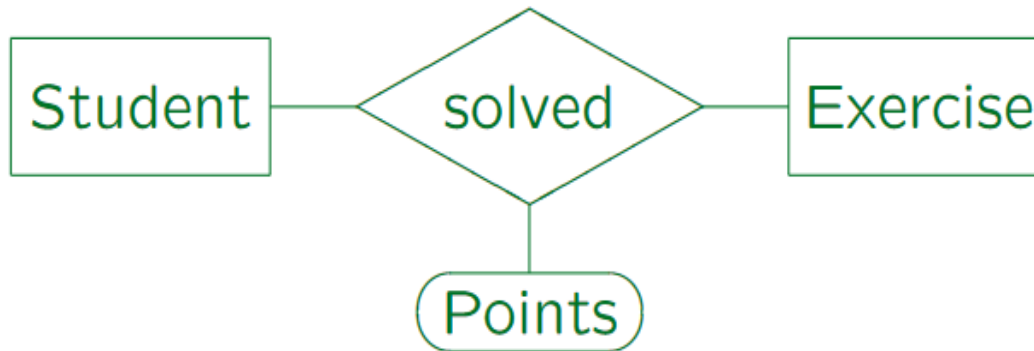
# EXERCISE

---



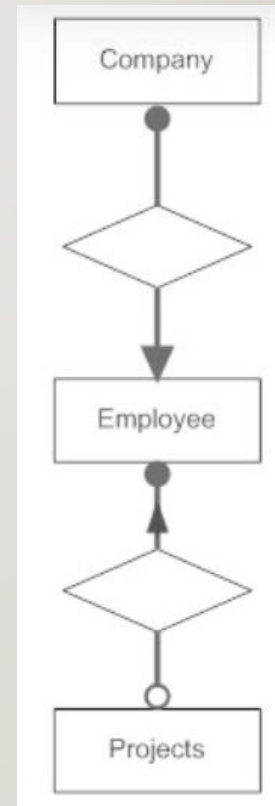
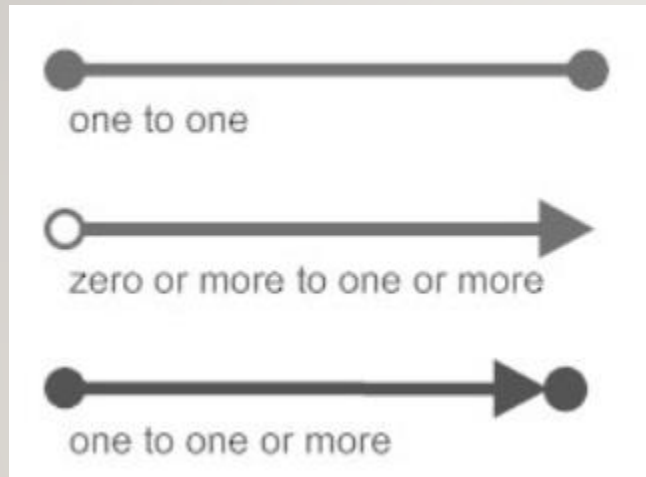
# EXAMPLE

---



# BACHMAN STYLE

---





# EXAMPLE

---

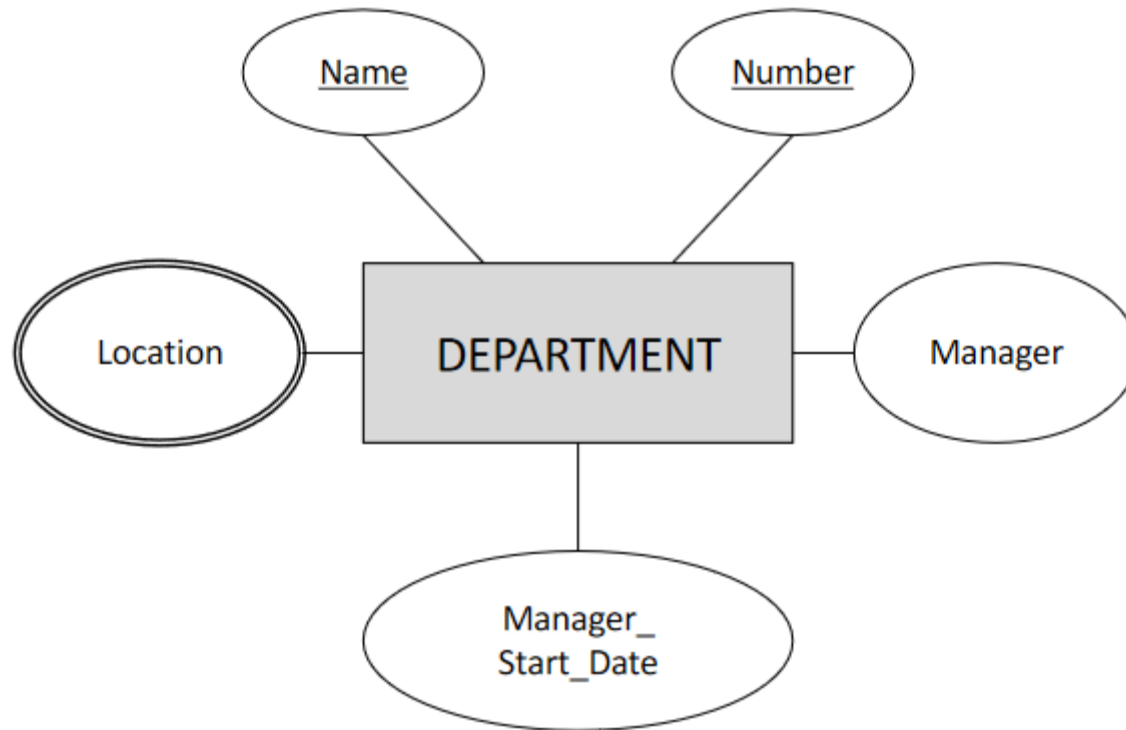


# EXERCISE

---

1. Each department has a unique name, a unique number, and a particular employee who manages the department. We keep track of the start date when that employee began managing the department. A department may have several locations.

# ANSWER



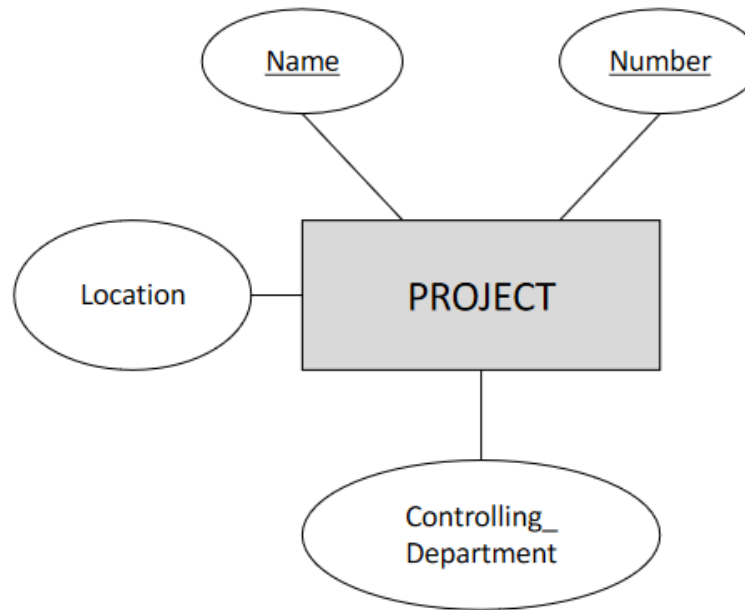
# EXERCISE 2

---

- A department controls a number of projects, each of which has a unique name, a unique number, and a single location.

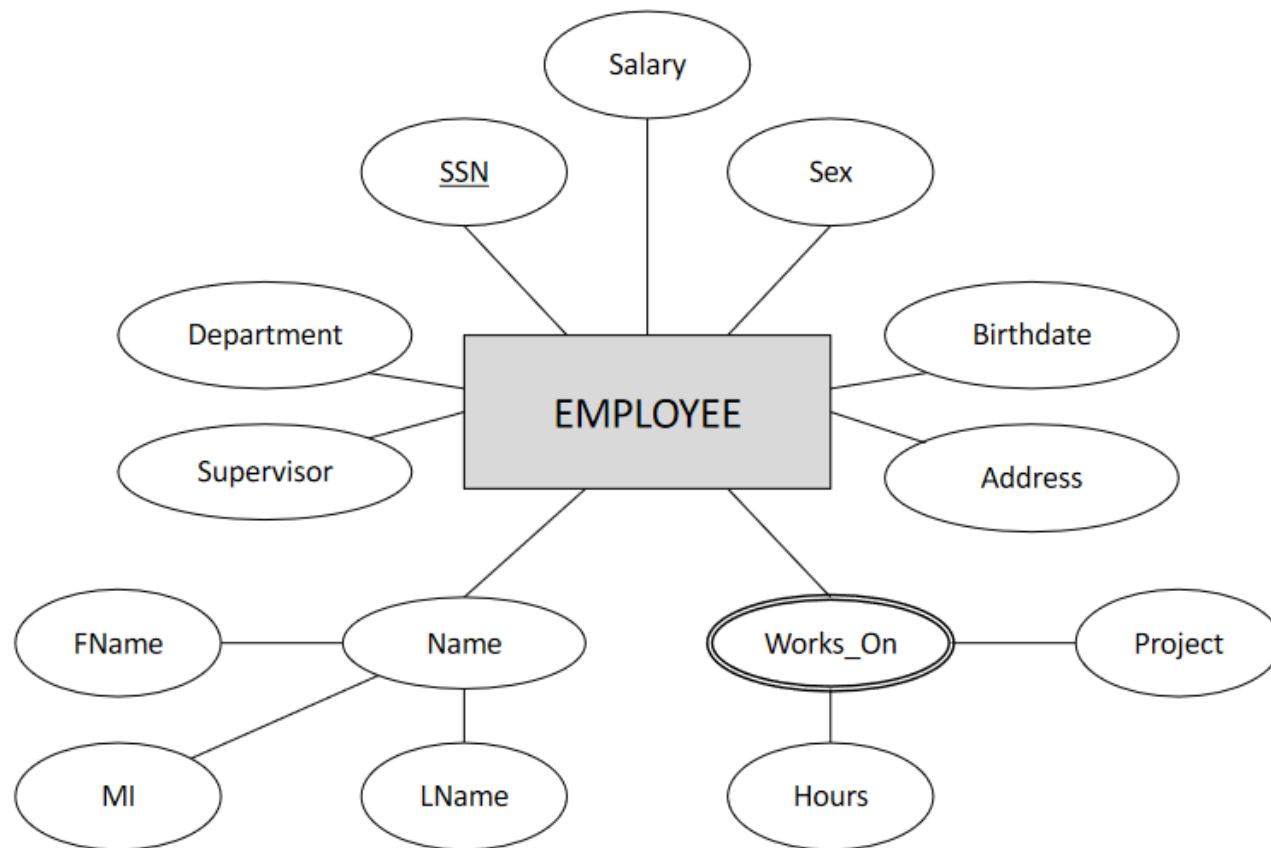
# ANSWER

---





- 
- We store each employee's name (first, last, MI), Social Security number (SSN), street address, salary, sex (gender), and birth date. An employee is assigned to one department, but may work on several projects, which are not necessarily controlled by the same department. We keep track of the current number of hours per week that an employee works on each project. We also keep track of the direct supervisor of each employee (who is another employee).



# ASSIGNMENT (GROUP)

---



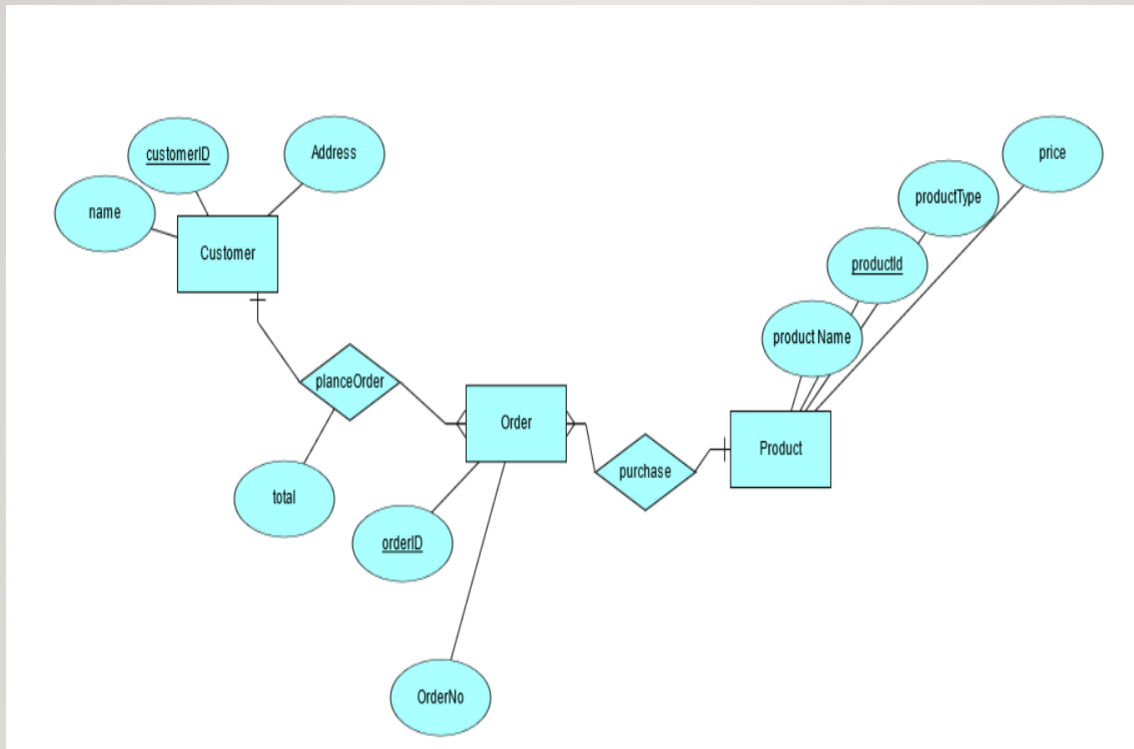
Define an ER schema (diagram) for the following application:

- Information about researchers in the database field must be stored.
- For each researcher, his/her last name, first name, email address, and homepage (URL) are needed.
- Also his/her current affiliation (employer) is needed (assume that all researchers work at universities).
- For each university, its name, URL, and country should be stored

# EXERCISE

---

- Design an ERD with the following scenario
- Customers may place an order for one or more products
- The following information are associated with **customer**:
  - address, name, age, city
- **Order**: order number
- **Product**: Product name, product type, and price





# ASSIGNMENT 2

---

- Courses, Students, Professors
  - ☐ Courses have ids, titles, and credits. The id is unique.
  - ☐ Courses have multiple sections that have time, a room and exactly one teacher
  - ☐ Professors have a unique name
  - ☐ Students take courses and receive a grade
  - ☐ Students may repeat a course
  - ☐ Must track students' course schedules and transcripts including grades, semester taken, etc.
  - ☐ Must track which classes a professor has taught
  - ☐ The database should work over multiple semesters