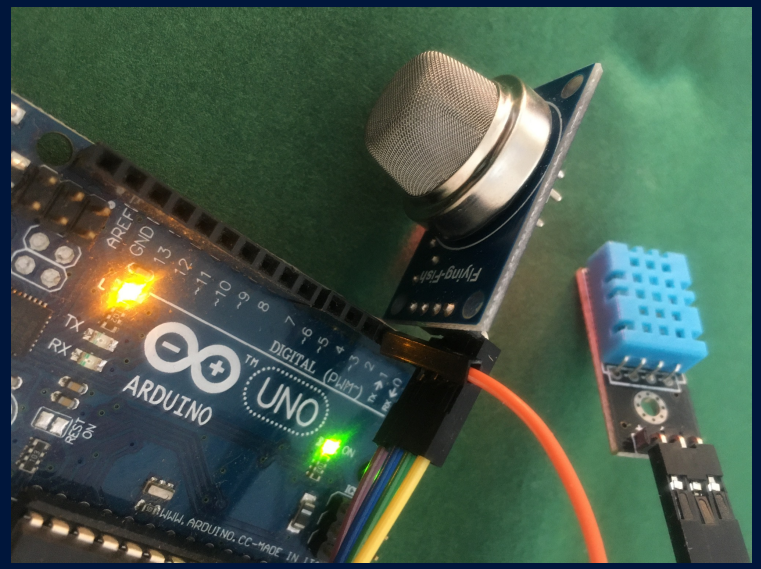
cphbusiness

COPENHAGEN BUSINESS ACADEMY





DATA ENGINEERING



FLOW 3 – Foreløbig plan

Uge 10	07.11.2022	intro til dataforespørgsler	Intro - API, Mongo, SQL og webscraping
	08.11.2022	Webscraping	Webscrapping: Case EDC, Bilbasen
	09.11.2022		
	10.11.2022	Webscraping / MongoDB	Mongodb
	11.11.2022	Webscraping	Præsentation af OLA
Uge 11	14.11.2022	SQL	MySQL: Case Northwind
	15.11.2022	ML with Baum	
	16.11.2022		
	17.11.2022	SQL	Advanced SQL, MySQL og R
	18.11.2022	SQL	MySQL, R og AWS (RDS)
Uge 12	21.11.2022	Cloud Computing	AWS - server og services
	22.11.2022	Cloud Computing	API og Mongo: Casse smart city Aarhus
	23.11.2022		
	24.11.2022	Cloud Computing	Case: PR Flights, R & Mongo på AWS
	25.11.2022	Cloud Computing	ML på AWS
Uge 13	28.11.2022	IOT	Internet of Things
	29.11.2022	IOT	Case: Afstands-sensor
	30.11.2022		
	01.12.2022	IOT	Case:Afstands-sensor
	02.12.2022	OLA	
Uge 14	05.12.2022	Webscraping & NLP	Intro til NLP
	06.12.2022	Webscraping & NLP	Sentiment på boligannoncer
	07.12.2022		
	08.12.2022		
	09.12.2022	Opsamling	Præsentation af OLA, eksamensforberedelse

Agenda - CRUD

- SQL
  - Northwind-øvelserne
  - EMP-database-øvelser
  - Views
- SQL i R
  - SQL-queries fra R
    - EMP-database-øvelser
  - Gennemgang af Jeff’s artikel
  - NEWS-api’et

MySQL from R – øvelse med emp-databasen

1. SQL
    1. Lav et ER-diagram
    2. Identificer fremmednøgler
    3. Er der afdelinger der ikke har medarbejdere?
        1. slide om joins
    4. Indsæt dig selv som medarbejder
    5. Opret en afdeling og indsæt dig selv
    6. Lav en liste over medarbejdere og tilhørende chefer
2. R

worker	managede by
SMITH	FORD
ALLEN	BLAKE
WARD	BLAKE
JONES	KING
MARTIN	BLAKE
BLAKE	KING
CLARK	KING
SCOTT	JONES
TURNER	BLAKE
ADAMS	SCOTT
JAMES	BLAKE
FORD	JONES
MILLER	CLARK

MySQL from R – øvelse med emp-databasen - joins

What is Inner Join?

An Inner Join returns only the rows that have matching values in both the tables (we are considering here the join is done between the two tables).

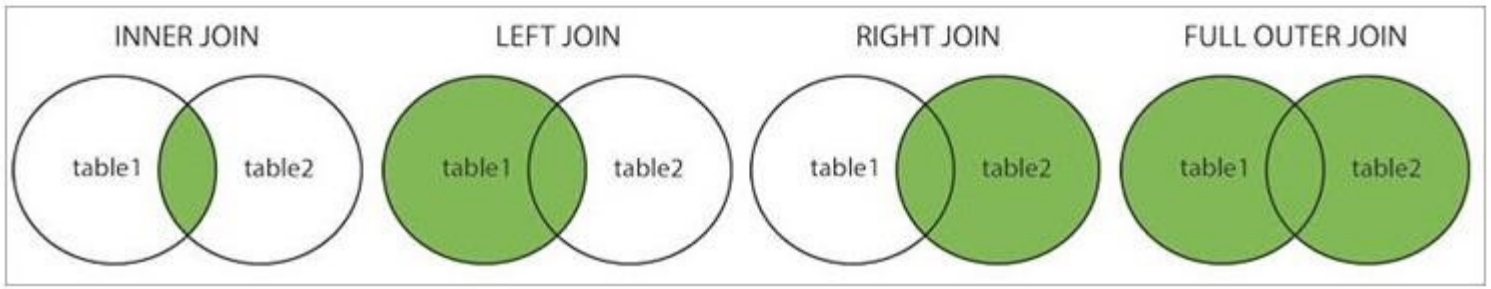
What is Outer Join?

The Outer Join includes the matching rows as well as some of the non-matching rows between the two tables. An Outer join basically differs from the Inner join in how it handles the false match condition.

There are 3 types of Outer Join:

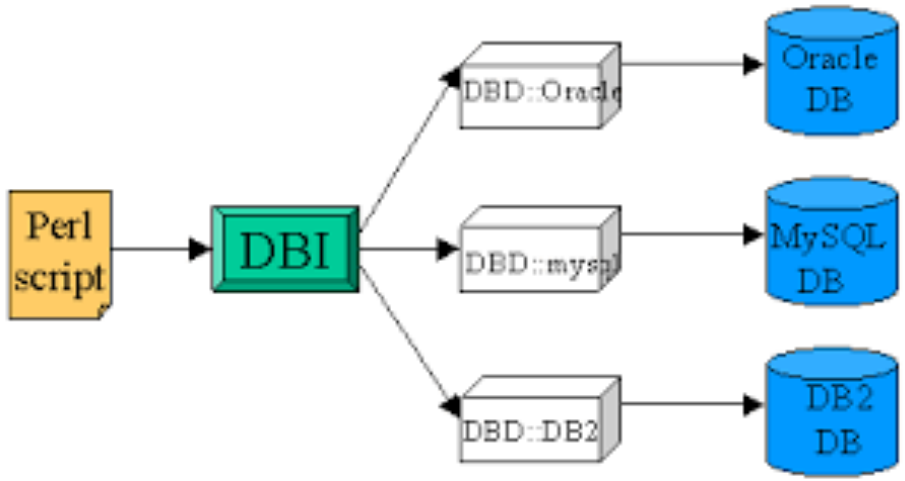
- **Left Outer Join**: Returns all the rows from the LEFT table and matching records between both the tables.
- **Right Outer Join**: Returns all the rows from the RIGHT table and matching records between both the tables.
- **Full Outer Join**: It combines the result of the Left Outer Join and Right Outer Join.

Difference between Inner and Outer Join



avg sal	dname
NULL	OPERATIONS
1567	SALES
2175	RESEARCH
2917	ACCOUNTING

MySQL from R


Perl
script
DBI
DBD::Oracle
DBD::mysql
DBD::DB2
Oracle
DB
MySQL
DB
DB2
DB

MySQL from R

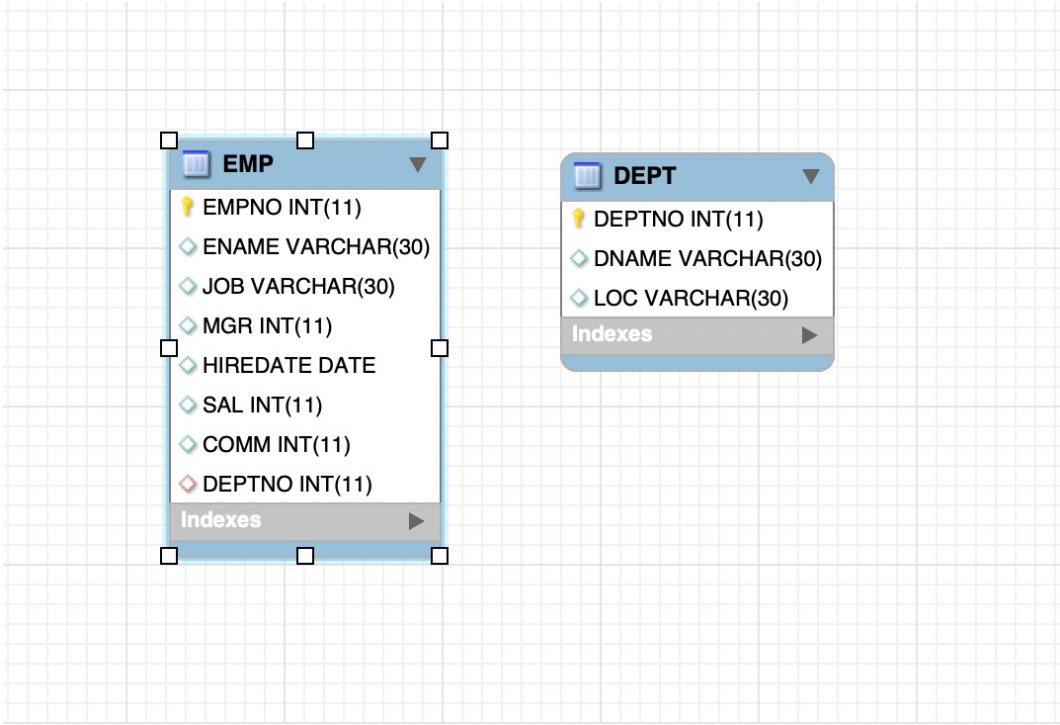
- Connecting and disconnecting
  - Connecting to and disconnecting from databases
    - dbConnect(MariaDB(), ..)
    - dbDisconnect(con)

- Tables
  - Reading and writing entire tables
    - dbWriteTable(con, "mycars-table",mycarsdf)
    - dbReadTable(con, "mycarstable")

- Queries
  - More control for sending queries and executing statements
    - dbGetQuery(con, "SELECT * FROM tab") (all in one)
    - dbSendQuery(con, "SELECT * FROM city limit 3") (get in batches)
    - dbExecute (con,"INSERT INTO city (Name,Population) VALUES ('Lviv',123123)")

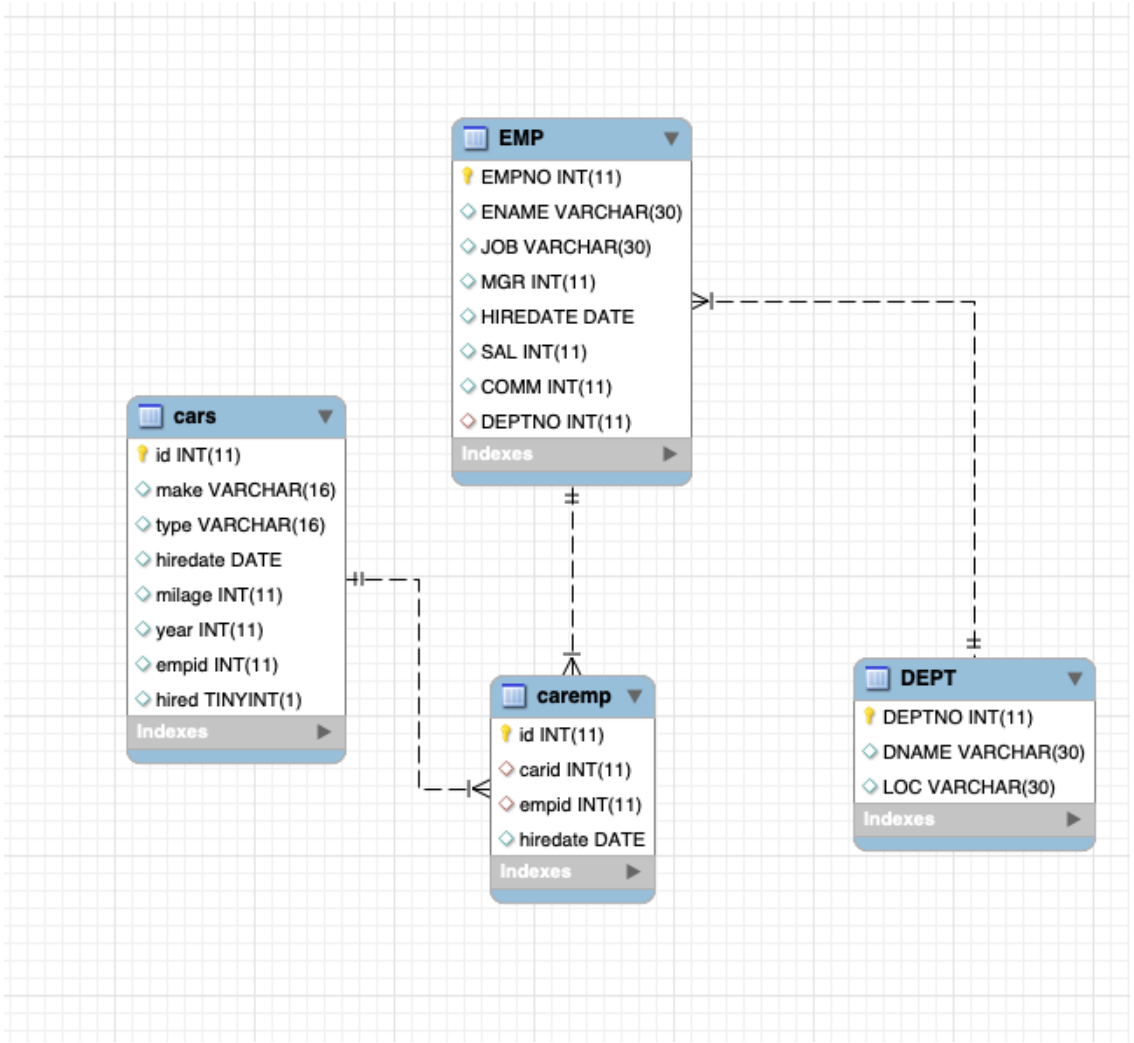
MySQL from R – øvelse med emp-databasen

1. SQL
2. R
    1. Hent de to tabeller ud af databasen.
    2. Merge de to dataframes
        1. join i dplyr
    3. Lav listen, hvor afdelinger uden medarbejdere vises
    4.



MySQL from R – øvelse med emp-databasen

1. SQL
2. R
    1. Hent de to tabeller ud af databasen.
    2. Merge de to dataframes
        1. join i dplyr
    3. Lav listen, hvor afdelinger uden medarbejd
    4. Udvid skemaet som på tegningen



MySQL from R – Jeff

1. Gemmer pw i
    1. sys.env
    2. my.cnf
2. INSERT i WB
3. INSERT i R vha **dbSendQuery**
4. TRUNCATE i WB
5. INSERT i R, paste vars og dbSendQuery
    1. dates in R **as.Date**(entryPublished, "%d %B %Y")
    2. date-slide
6. Indlæs csv i R
    1. sampleGardenData <- read.csv(”allot.csv”)
7. tilpas data
    1. title til 99 vha substr
    2. date
8. Send **HELE** df til DB vha **dbWriteTable**(conn, df, ”name-of-table”)
9. SELECT .. <antal-stories> GROUP BY <> vha **dbSendQuery** og **dbFetch**(results)

MySQL from R – Jeff & Dates

1. case tt="2022/16/11 22:10:00” til
    1. 2022-11-16
    2. 2022-11-16 22:10:00 CET

1. case tt=”24-1-1988 22:10

Symbol	Meaning	Example
%a	Abbreviated weekday name	Tue
%A	Full weekday name	Tuesday
%b	Abbreviated month name	Apr
%B	Full month name	April
%C	Century: the integer part of the year divided by 100	20
%d	Day of the month	09
%H	Hours as decimal number (00–23)	13
%I	Hours as decimal number (01–12)	1
%m	Month as number (01–12)	04
%M	Minute as number (00–59)	12
%p	AM/PM indicator for 12-hour time (%I)	PM
%S	Second as integer (00–61)	12
%u	Weekday as a decimal number (1–7, Monday is 1)	2
%w	Weekday as decimal number (0–6, Sunday is 0)	2
%y	2-Digit Year (00–99)	19
%Y	4-Digit Year	2019

realdate <- as.Date("09/May/2022:00:09:24",format='%d/%m%Y:%H%M%S')

MySQL from R – SoccerDB

```
SELECT * FROM player AS t1 JOIN (SELECT id FROM player
ORDER BY RAND() LIMIT 10) as t2 ON t1.id=t2.id
```