

Challenge 07: Data Science & SQL

Data Science und Datenbanken
mit Python und SQLite

Prof. Dr. Johannes Schildgen

Data Science

Data Science mit Python

Prof. Dr. Johannes Schildgen

Strukturierte Daten

person_id	firstname	lastname	birthday
141	Jane	Miller	1995-01-23
178	Carl	Smith	1992-05-09
214	Sandra	Fox	1993-12-30
184	Bob	Fox	1990-09-08

Unstrukturierte Daten

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. A diam sollicitudin tempor id eu nisl nunc mi ipsum. Metus dictum at tempor commodo ullamcorper a lacus vestibulum sed. Velit euismod in pellentesque massa placerat dui ultricies lacus. Non blandit massa enim nec dui nunc. Accumsan sit amet nulla facilisi morbi tempus iaculis urna id. Integer eget aliquet nibh praesent tristique magna. Eu facilisis sed odio morbi quis commodo odio. Magna eget est lorem ipsum dolor. Ut ornare lectus sit amet. Lacinia at quis risus sed vulputate odio ut enim. Blandit cursus risus at ultrices mi. Sem integer vitae justo eget magna fermentum iaculis. Faucibus scelerisque eleifend donec pretium vulputate sapien. Quisque sagittis purus sit amet volutpat consequat mauris. Nisl rhoncus mattis rhoncus urna neque viverra justo nec.

Enim ut tellus elementum sagittis vitae. Condimentum id venenatis a condimentum. Nisi vitae suscipit tellus mauris a. Pharetra massa massa ultricies mi quis hendrerit dolor magna. Quisque egestas diam in arcu cursus euismod. Aliquet risus feugiat in ante metus dictum. Consectetur adipiscing elit pellentesque habitant morbi tristique senectus. Id cursus metus aliquam eleifend mi in nulla posuere. Volutpat commodo sed egestas egestas fringilla phasellus. Vitae congue mauris rhoncus aenean. Pellentesque habitant morbi tristique senectus et netus et malesuada. Venenatis cras sed felis eget velit aliquet. Urna nunc id cursus metus aliquam eleifend. Amet commodo nulla facilisi nullam vehicula ipsum a. Tortor pretium viverra suspendisse potenti nullam ac tortor. Egestas erat imperdiet sed euismod. Egestas diam in arcu cursus euismod quis. Facilisi nullam vehicula ipsum a. Odio tempor orci dapibus ultrices in iaculis nunc. Neque ornare aenean euismod elementum nisi quis eleifend.



customers.csv

```
customerno,firstname,lastname,email  
141,Jane,Miller,jane@example.com  
178,Carl,Smith,carl@example.com  
214,Sandra,Fox,sandra@example.com  
184,Bob,Fox,bob@example.com
```

customers.csv

```
customerno;firstname;lastname;email  
141;Jane;Miller;jane@example.com  
178;Carl;Smith;carl@example.com  
214;Sandra;Fox;sandra@example.com  
184;Bob;Fox;bob@example.com
```

customers.csv

```
customerno|firstname|lastname|email  
141|Jane|Miller|jane@example.com  
178|Carl|Smith|carl@example.com  
214|Sandra|Fox|sandra@example.com  
184|Bob|Fox|bob@example.com
```

customers.csv

141	Jane	Miller	jane@example.com
178	Carl	Smith	carl@example.com
214	Sandra	Fox	sandra@example.com
184	Bob	Fox	bob@example.com

Datei lesen

```
file = open("robot_log.csv", "r")  
data = file.read()  
print(data)  
file.close()
```

CSV-Datei lesen

```
import csv
file = open("robot_log.csv", "r")
data = csv.reader(file, delimiter=",")
for row in data:
    print(row[1])
file.close()
```

CSV-Datei lesen



Data Science

- Problembeschreibung
- Datenbeschaffung
- Data Cleaning
- Datenanalyse
- Optimierung

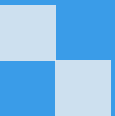




ER-Modelle

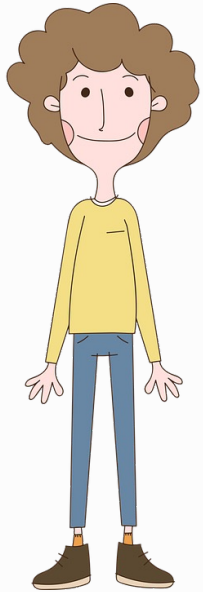
Entity-Relationship-Diagramme

Prof. Dr. Johannes Schildgen



ER-Diagramme

Entity-Relationship



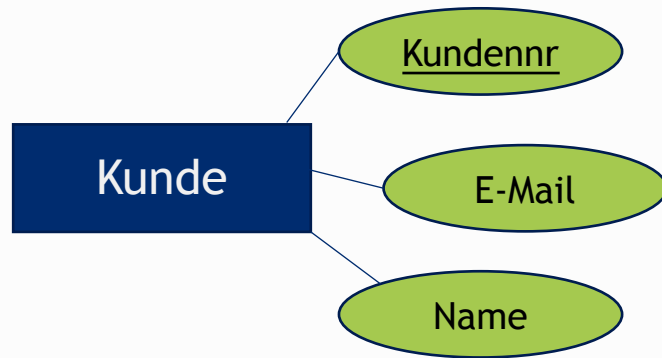
ER-Diagramme

Entity-Relationship



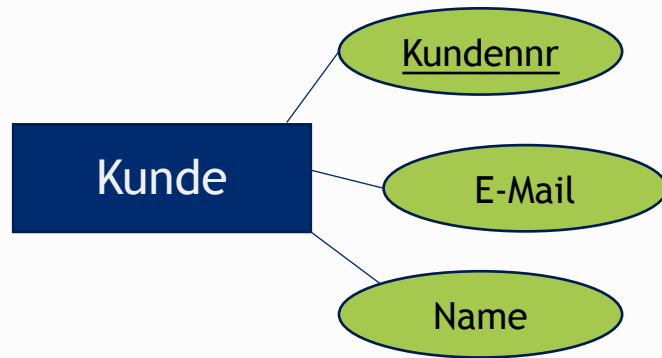
ER-Diagramme

Attribute



ER-Diagramme

Primärschlüssel



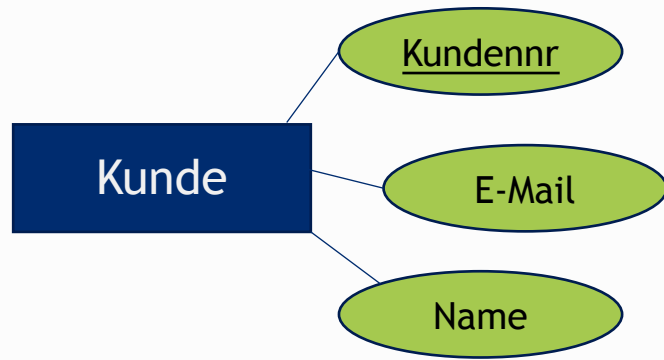
Kundenr. 7



Kundenr. 7

ER-Diagramme

Attribute



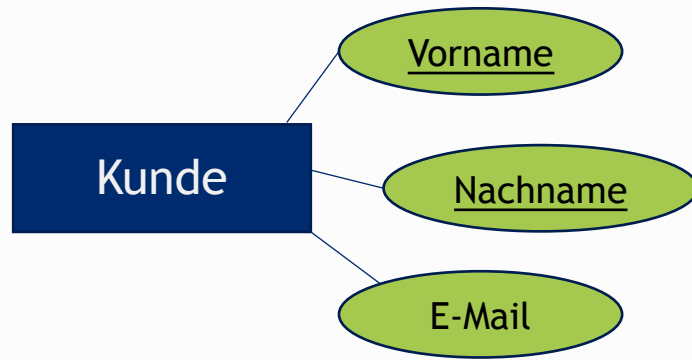
Kundenr. 7
Name: Ute



Kundenr. 8
Name: Ute

ER-Diagramme

Kombinierte Primärschlüssel



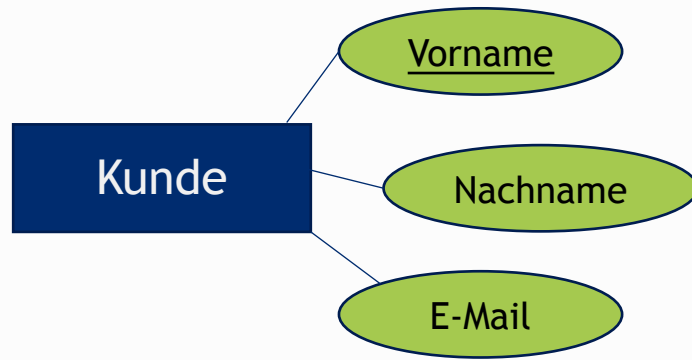
Ute
Müller



Ute
Müller

ER-Diagramme

Kombinierte Primärschlüssel

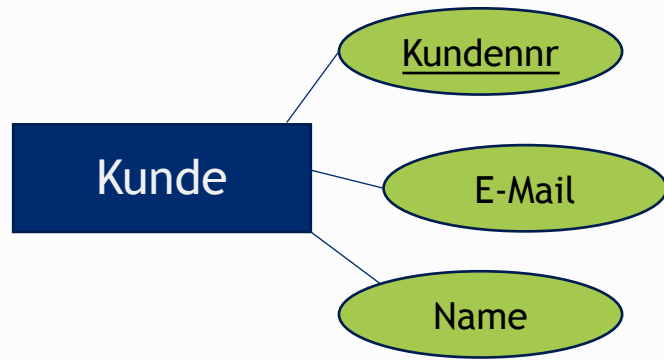


Ute
Müller

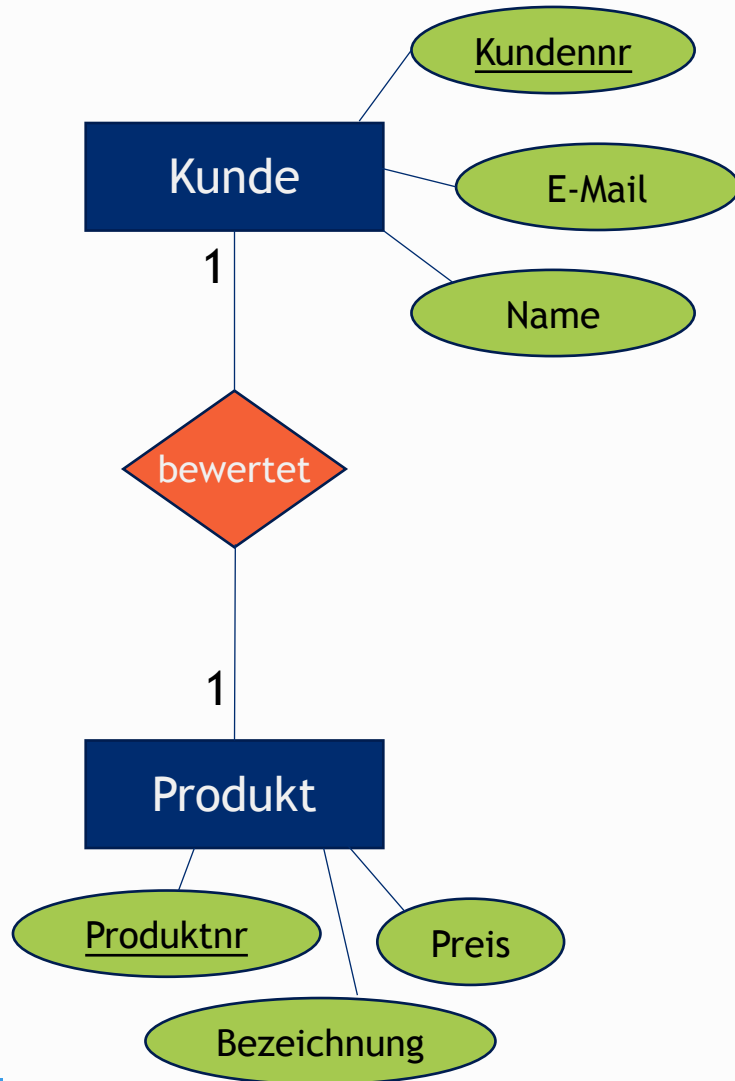


Ute
Müller

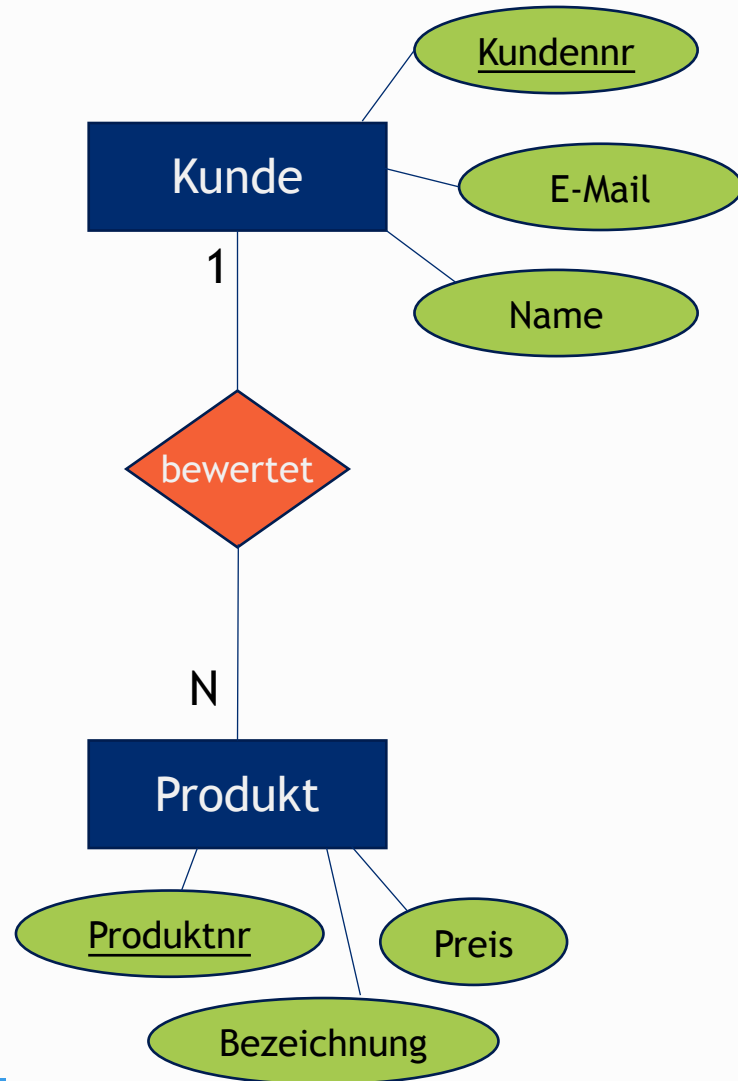
ER-Diagramme



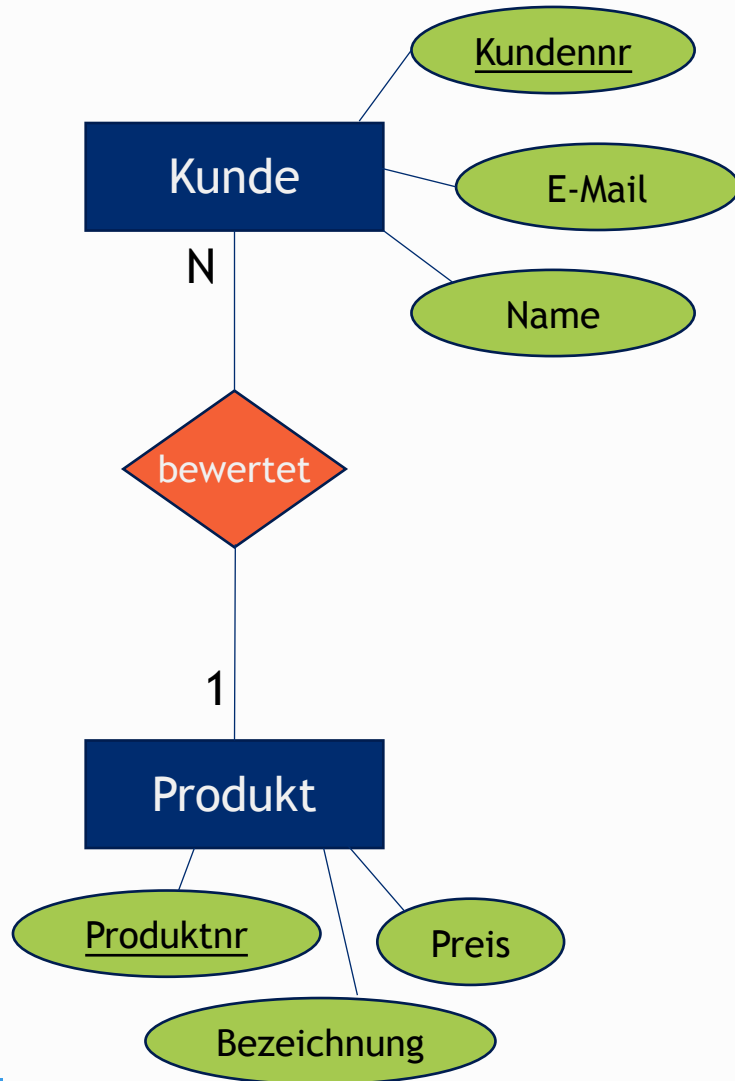
ER-Diagramme



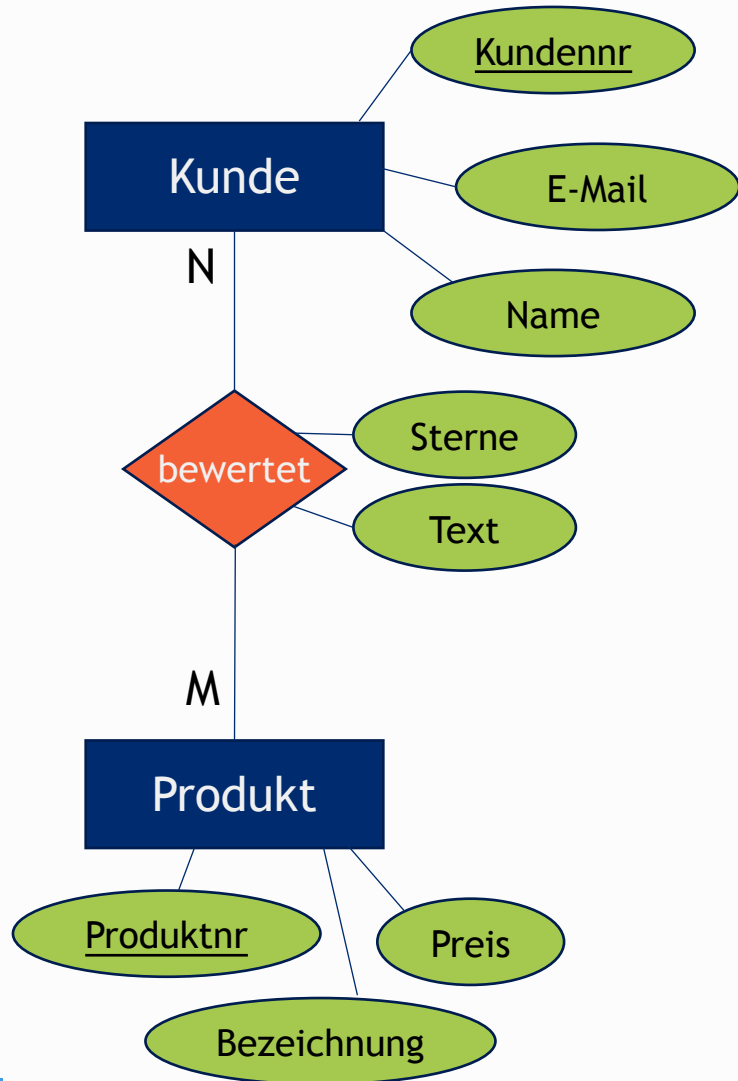
ER-Diagramme



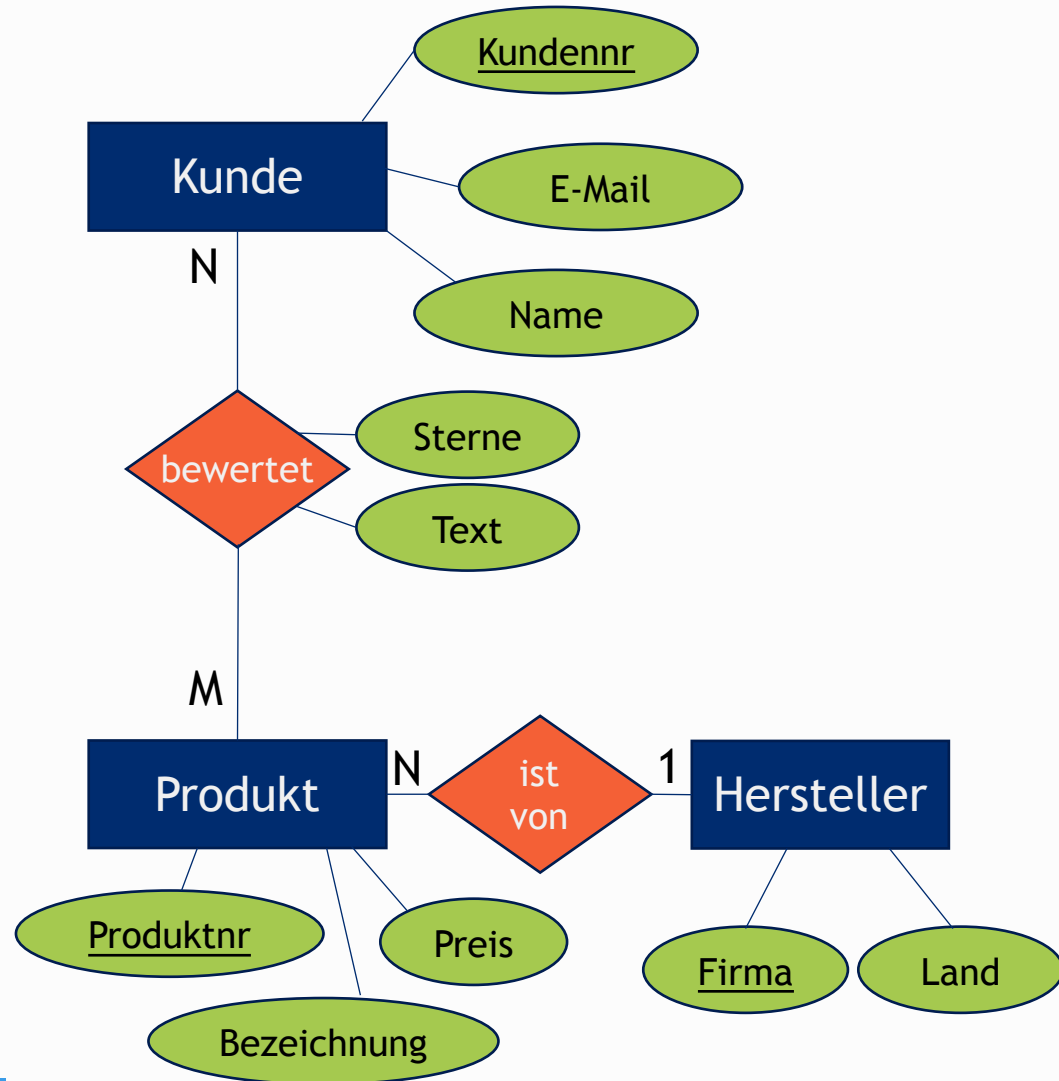
ER-Diagramme



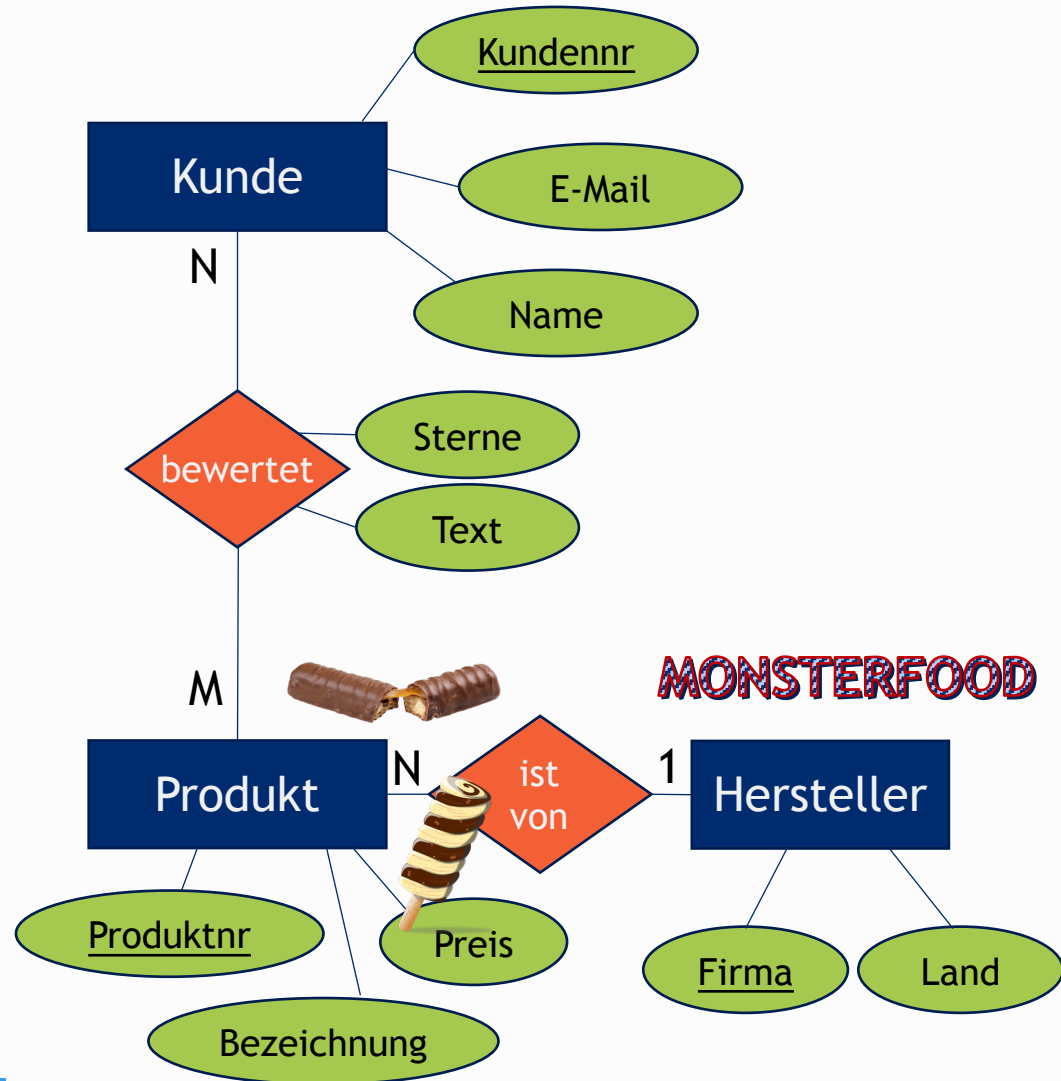
ER-Diagramme



ER-Diagramme



ER-Diagramme

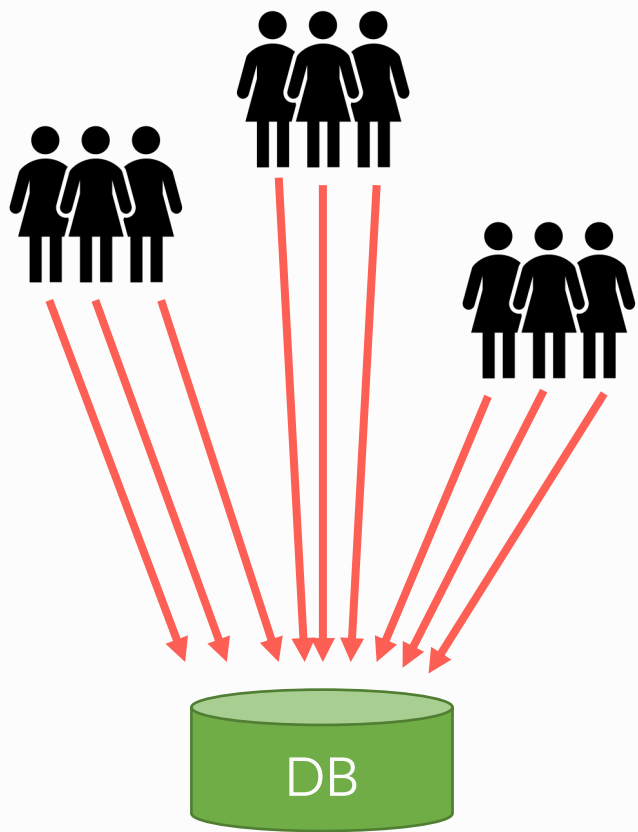


Datenbanken

Relationale Datenbanken

Prof. Dr. Johannes Schildgen





SQL



INSERT



UPDATE



DELETE



```
SELECT bezeichnung, preis  
FROM produkte  
WHERE produktnr = 29;
```





```
INSERT INTO bewertungen  
(kundenr, produktnr, sterne,  
text)  
VALUES (214, 29, 3,  
      'Ganz okay.');
```



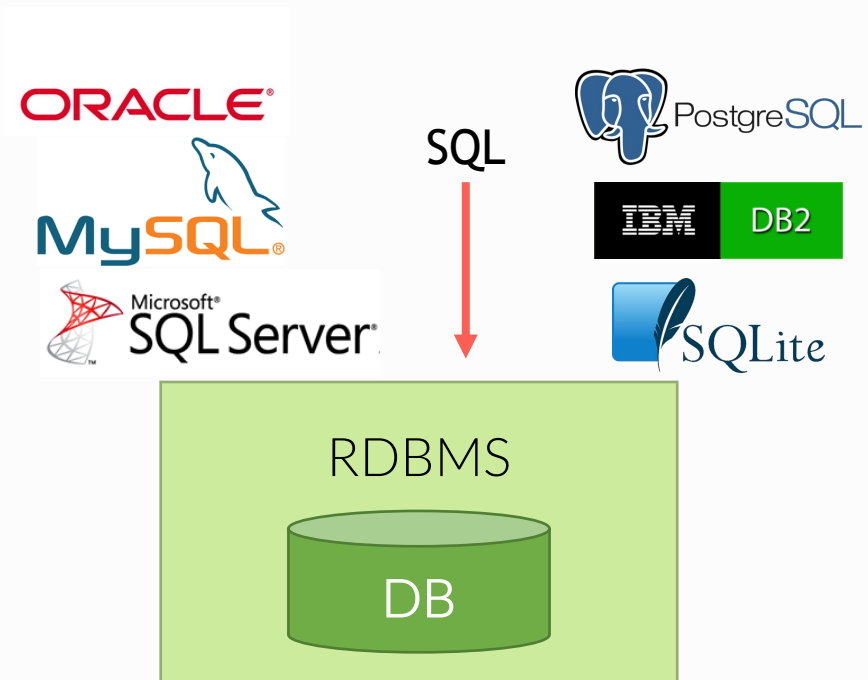
Tabellen

(Relationen)

```
CREATE TABLE people  
(person_id INT PRIMARY KEY,  
  firstname VARCHAR(100),  
  lastname VARCHAR(100),  
  birthday DATE);
```

person_id	firstname	lastname	birthday
141	Jane	Miller	1995-01-23
178	Carl	Smith	1992-05-09
214	Sandra	Fox	1993-12-30
184	Bob	Fox	1990-09-08

Relationale Datenbank- management- systeme



SQL

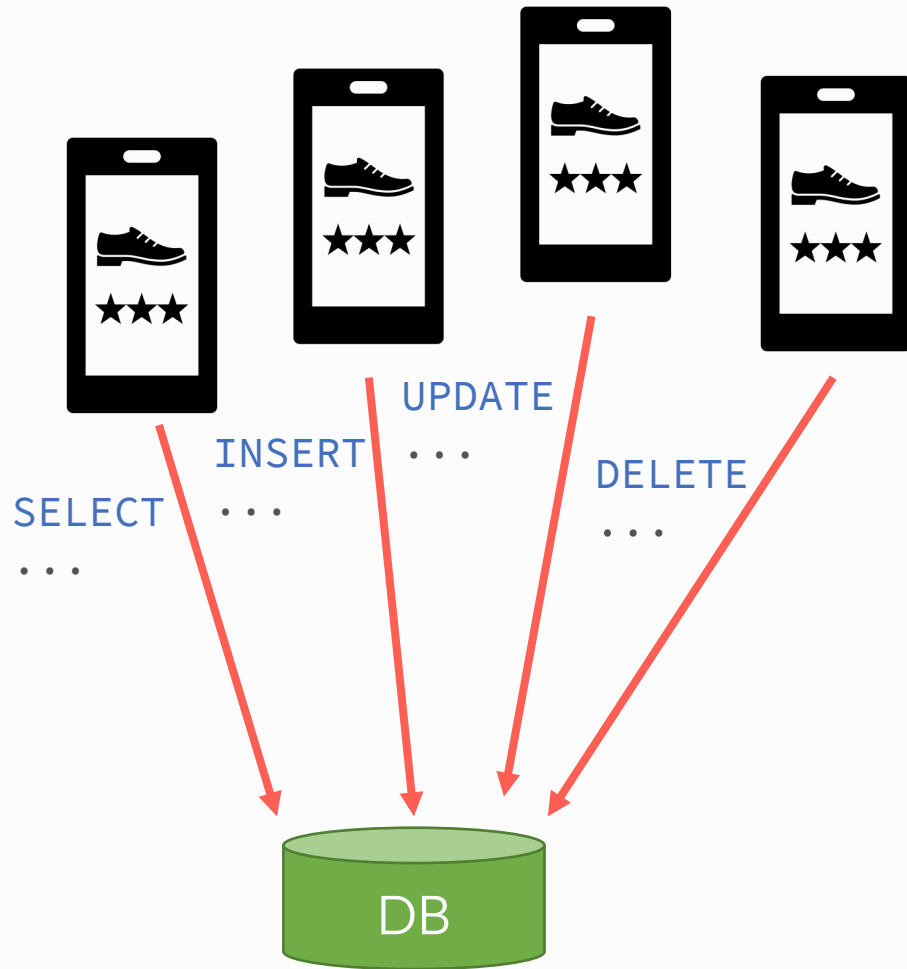


```
SELECT bezeichnung, preis  
FROM produkte  
WHERE produktnr = 29;
```

Was?



Transaktionen



Produkte

produktnr	beschreibung	Preis
17	Schokoriegel	0.89
29	Spülmaschinentabs	3.99

Kunden

knr	vorname	nachname	email
141	Jane	Miller	j@example.com
178	Carl	Smith	c@example.com
214	Sandra	Fox	s@example.com

Bewertungen

knr	produktnr	sterne	text
214	29	3	Ganz okay.
141	29	5	Super gut!

Users

user_id	name	...
141	Jane Miller	...

Pages

page_id	title	...
1508	Lustige Bilder	...

Groups

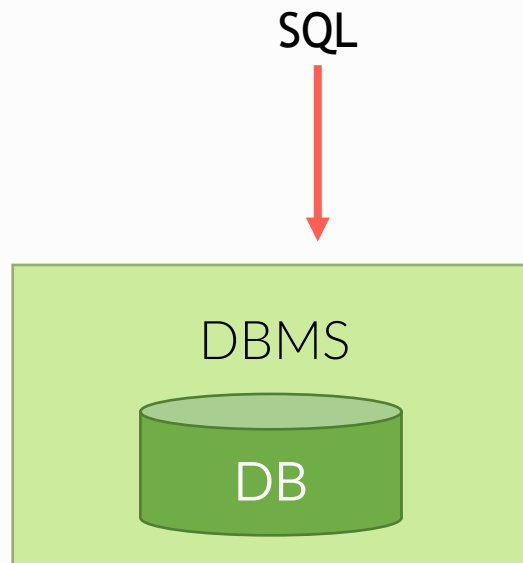
group_id	title	...
4698	OTH Regensburg	...

page_followers group_members

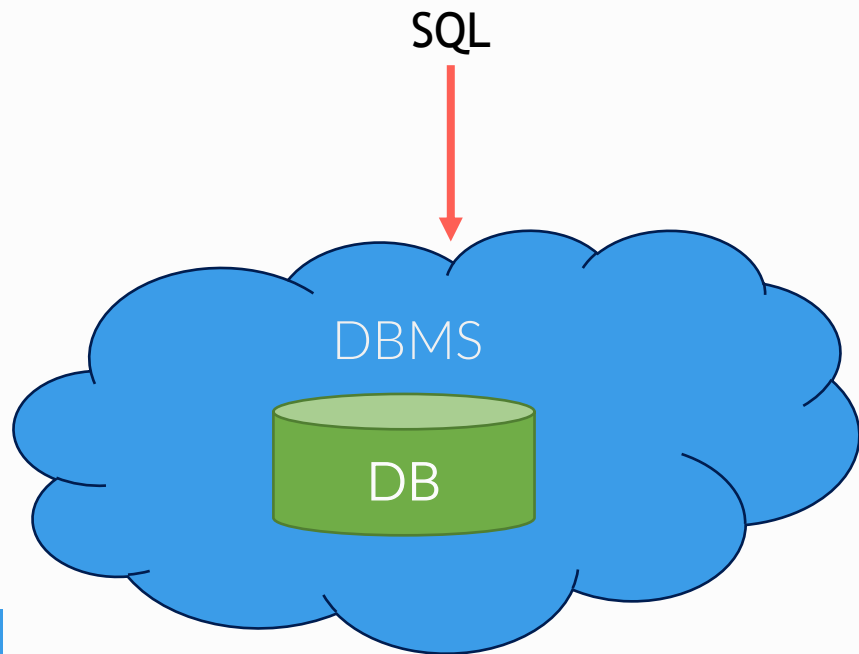
user_id	page_id
141	1508

user_id	page_id
141	4698

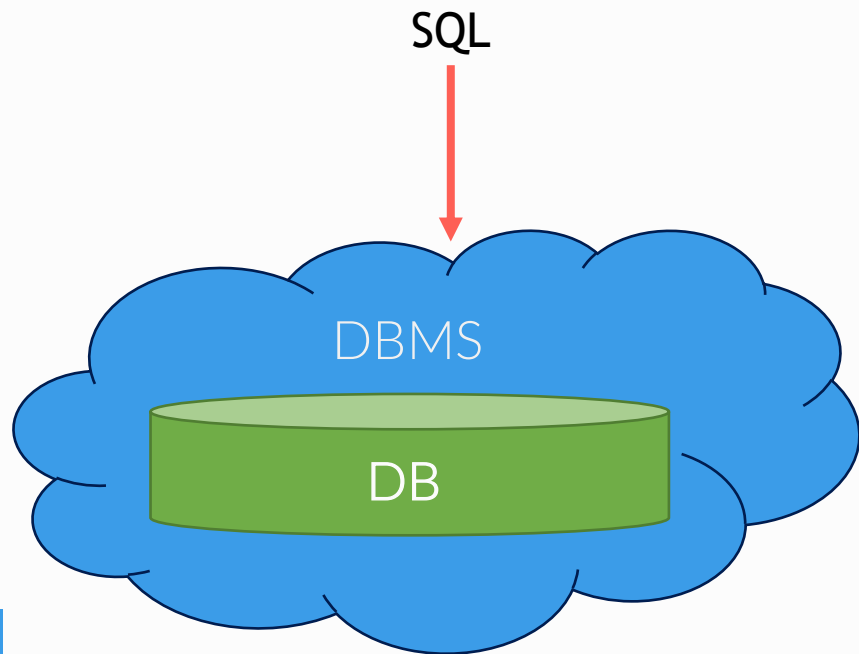
Datenbank-Server



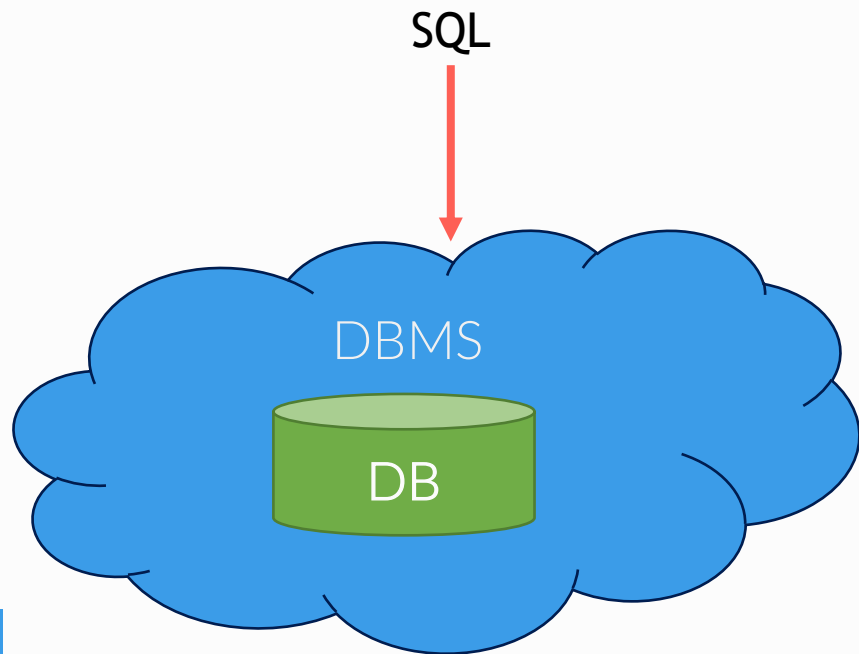
Cloud



Cloud



Cloud





`https://example.com`

HTTPS

Webserver

`SELECT bezeichnung
FROM produkte;`

SQL

DBMS

DB

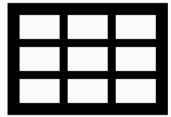


HTTPS

```
<html><head>  
<title>Shop</title>  
...
```

Webserver

SQL



DBMS

DB

ORACLE®

MySQL®

Microsoft®
SQL Server®

PostgreSQL

IBM

DB2

SQL

DBMS

DB



SQL



SQLite

DBMS



webshop.db

kundennummer	name	email
5	Peter	peter@example.com
8	Anna	anna@example.com

SQL-Datentypen

INT	Ganze Zahlen	8, 0, -3, 2319
DECIMAL(p,q)	Kommazahlen (p Stellen, davon q vor dem Dezimalpunkt)	DECIMAL(9,2): 29.95 -5.00 1234567.89
DOUBLE	Kommazahlen (ohne feste Anzahl Nachkommastellen)	36.57211619201
DATE	Datum (Jahr-Monat-Tag)	'2022-11-30'
TIMESTAMP	Zeitstempel (Datum + Uhrzeit)	'2022-11-30 18:30:05.123'
CHAR(l)	Zeichenkette fester Länge l	CHAR(5): 'Hallo', 'Hi '
VARCHAR(l)	Zeichenkette mit maximal l Zeichen	VARCHAR(5): 'Hallo', 'Hi'

<u>firma</u>	land
Calgonte	Italien

<u>firma</u>	land
--------------	------

<u>firma</u>	land
Calgonte	Italien
Monsterfood	<i>NULL</i>
Holzkopf	Österreich

<u>produktnummer</u>	bezeichnung	preis	hersteller
17	Schokoriegel	1.89	Monsterfood
29	Spülmaschinentabs	4.99	Calgonte
88	Katzenfutter	5.99	<i>NULL</i>

<u>produktnummer</u>	bezeichnung	preis	hersteller
----------------------	-------------	-------	------------

Fremdschlüssel

Hersteller

<u>firma</u>	land
Monsterfood	USA
Holzkopf	Österreich

Produkte

<u>produktnummer</u>	beschreibung	preis	hersteller
17	Schokoriegel	0.89	Monsterfood
18	Müsliriegel	0.99	Monsterfood
88	Katzenfutter	4.99	NULL

Join (Verbund) in SQL

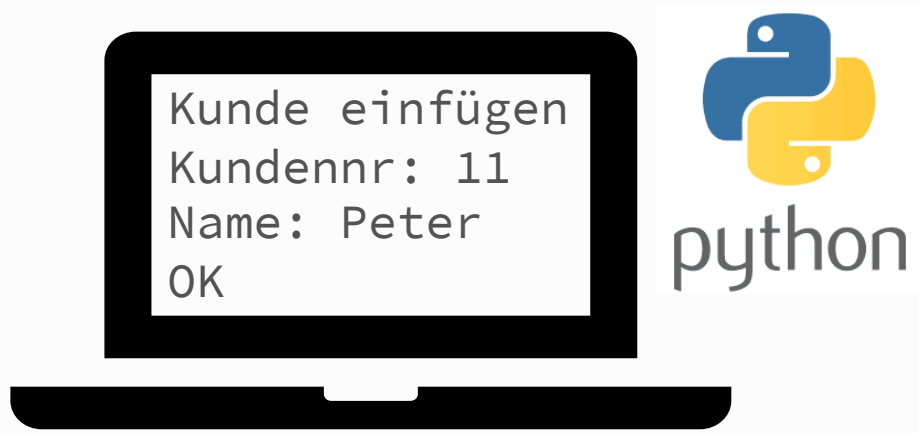
Produkte

<u>produktnummer</u>	beschreibung	preis	hersteller
17	Schokoriegel	0.89	Monsterfood
18	Müsliriegel	0.99	Monsterfood
88	Katzenfutter	4.99	NULL

Hersteller

<u>firma</u>	land
Monsterfood	USA
Holzkopf	Österreich

```
SELECT ...  
FROM produkte JOIN hersteller  
ON produkte.hersteller = hersteller.firma  
...;
```



INSERT INTO kunden ...





1. Treiber laden
2. Verbindung zur DB aufbauen
3. SQL-Anfragen ausführen

`SELECT kundennummer, name`
`FROM kunden ...`

