

**M U N I**  
**M E D**

Institut  
biostatistiky  
a analýz

# Databázové systémy v biomedicíně

Lekce I - Úvod

## About me

Daniel Klimeš

- Vzdělání: Obecná biologie
- PGS: onkologie
- Specializace: Screeningové projekty v onkologii, Národní zdravotní registry
- Databáze MS SQL, PostgreSQL, ORACLE
- Zaměstnání: Datové centrum ÚZIS

□ [klimes@iba.muni.cz](mailto:klimes@iba.muni.cz)

# E5447 Databázové systémy v biomedicině

Lectures

Každou druhou středu/wednesday od 14:00 – do 17:50

Teoretická přednáška → praktické cvičení

Praktická část/practical tasks : PostgreSQL,  
Domácí cvičení/homework: PostgreSQL

Zakončení: zápočet/credit – domácí úkol/homework  
zkouška/exam – praktický **test**, časový limit

# Rozsah předmětu

- Základy SQL
  - SELECT, UPDATE, DELETE, INSERT
- Funkce a operátory
- Základy agregace dat
  - GROUP BY, HAVING, AVG, SUM, ...
- Práce s více tabulkami
  - JOIN, UNION
- Zanořené dotazy
- Window funkce
  - Rank, lag, lead, OVER
- Vyhledávání v textu
  - Like, regulární výrazy
- Umělá inteligence a SQL
  - ChatGPT

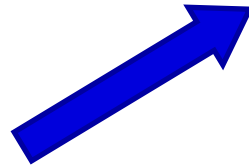
# Kdy zpracovávat data v databázi

1. Data jsou primárně v databázi uložena
2. Zpracováváme objemná data v řádu sto tisíc záznamů a více
3. Zpracování dat plánujeme provádět opakovaně
4. S daty bude pracovat více uživatelů
5. Způsob zpracování potřebujete dokladovat
6. Nechcete se zbláznit z excelu

**Relační databáze** x NoSQL databáze

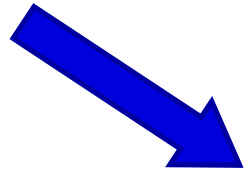
# Význam databáze pro analytika

**Databáze**  
SQL



**Statistický SW**  
Statistika for Windows, R,  
SPSS, SAS, MS Excel

- **Předzpracování dat**
- **Čištění dat**
- Popisná analýza
- Filtrování
- Propojení dat
- Agregace dat



**Matematicko - statistický  
skriptový SW**  
R, Matlab, Maple, programovací jazyk (např.  
Python)

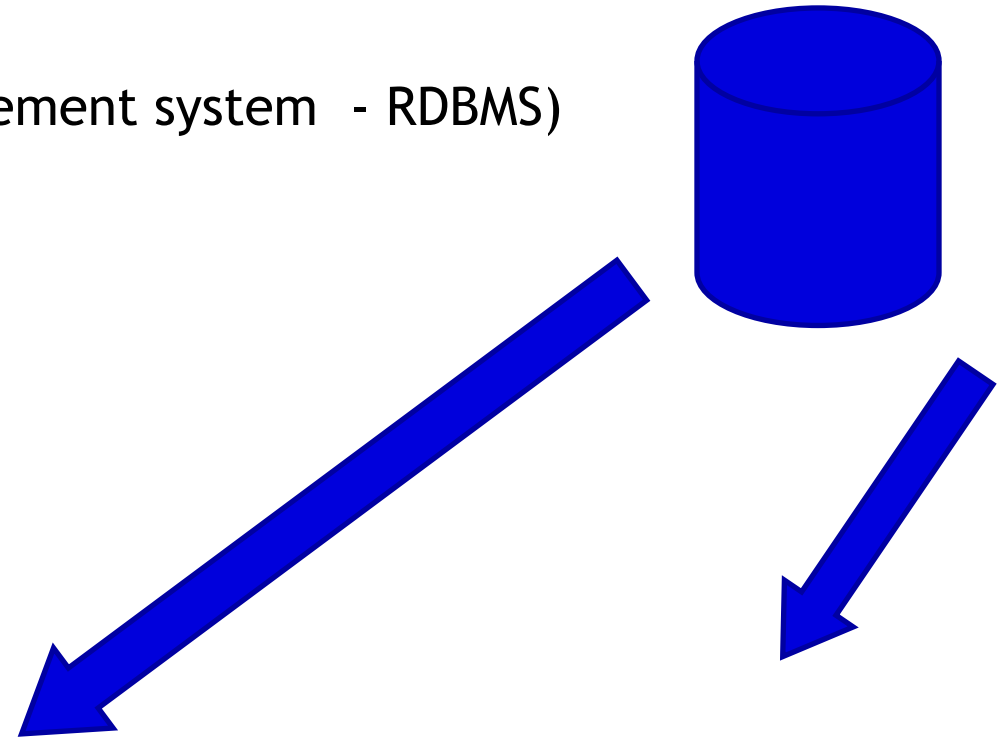
# Databázové systémy

Relační databáze (Relational database management system - RDBMS)  
Relace/relation - termín z relační algebry

Základ: **tabulka/table**

- sloupec/column = atribut/parametr
- řádek/row = popsáný objekt

Databáze = systém provázaných tabulek



PatientID	FirstName	LastName
1	Jan	Novák
2	Jana	Nová
3	Karel	Starý

PatientID	ExaminationDate	Result
1	12.1.2011	39,5
1	15.3.2011	36,8
2	2.2.2011	37,5

# Databázové systémy - Produkty

## Dle dostupnosti

### Komerční

ORACLE\*

MS SQL\*

DB2

MS ACCESS

FOX PRO

### Freeware

MySQL

PostgreSQL

Firebird

\* Okleštěné verze jsou k dispozici zdarma

## Dle počtu uživatelů

### Jednouživatelské/single user

MS ACCESS

FOX PRO

### Víceuživatelské/multiuser

ORACLE

MS SQL

DB2

MySQL

PostgreSQL

Firebird

[https://en.wikipedia.org/wiki/List\\_of\\_relational\\_database\\_management\\_systems](https://en.wikipedia.org/wiki/List_of_relational_database_management_systems)

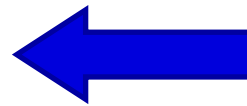


# Tabulka/Table

Definovaná struktura, do které se vkládají záznamy

Definují se sloupce/columns

- jméno/name
- datový typ /datatype
  - text
  - číslo/number
  - datum/date
  - BLOB
- doplňující vlastnosti/other properties



PatientID	ExaminationDate	Result
1	12.1.2011	39,5
1	15.3.2011	36,8
2	2.2.2011	37,5

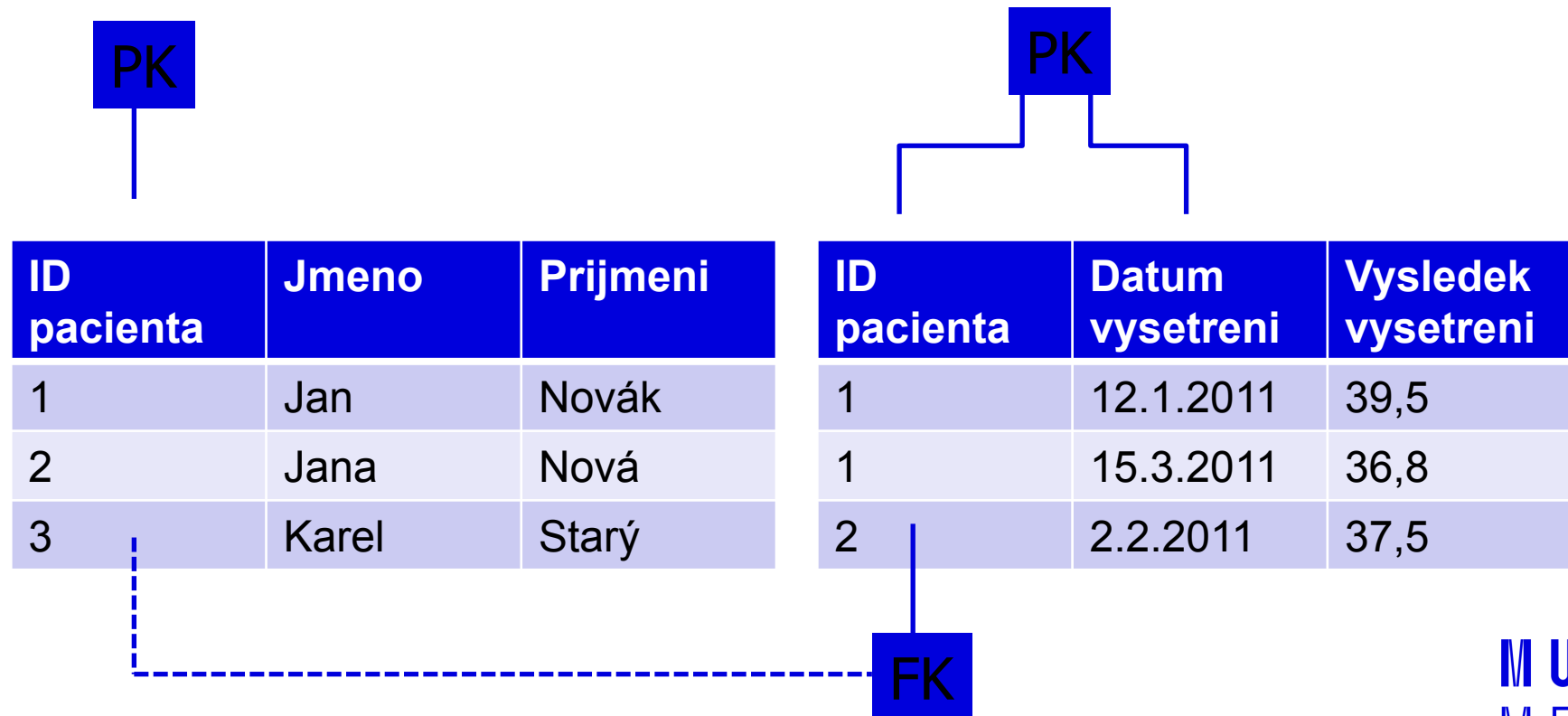
**Definice tabulky předchází načtení dat**  
**A table must be created before data import**

# Klíče/keys

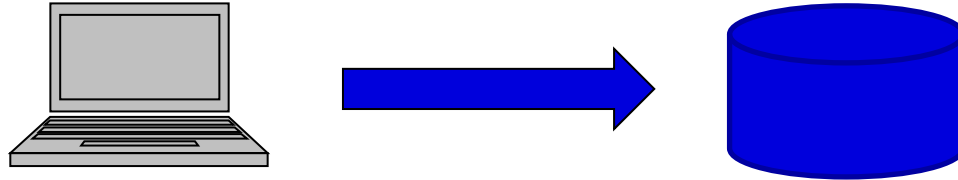
Vybrané sloupce se označují jako klíče (keys)

**Primární klíč** (primary key - PK)- 1 až n sloupců jednoznačně identifikující řádek (unique values)

**Cizí klíč** (foreign key - FK) - identifikuje nadřazený řádek v rodičovské tabulce



# Přístup do databáze/Access to a database



**Klient/client** = SW umožňující ověření uživatele a spuštění řídicích příkazů  
**Ovladač/Driver** = komponenta klienta

**Řídící příkazy/commands** = Structured Query Language - SQL

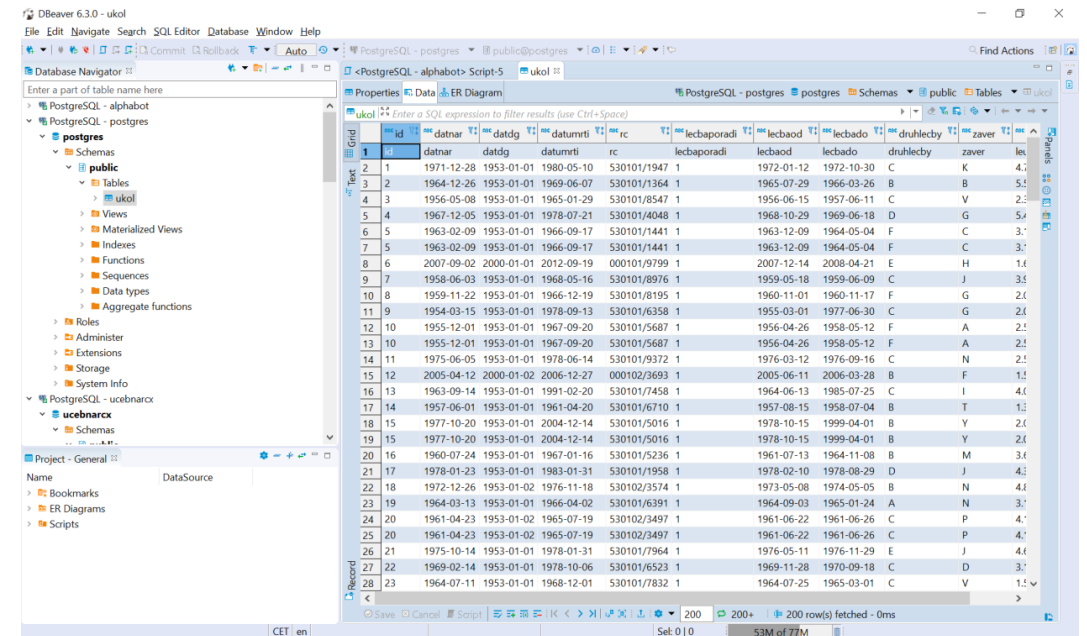
- DDL = data definition language
  - vytváření, změna, rušení objektů (tabulka, index, pohled, ...)
  - CREATE / ALTER / DROP
- DML = data manipulation language
  - SELECT - získávání dat z databáze
  - INSERT - vkládání dat do databáze
  - DELETE - mazání dat v databázi
  - UPDATE - změna/aktualizace dat
- transakční příkazy / commands for transactions
  - COMMIT - potvrzení transakce
  - ROLLBACK - odvolání transakce

# Instalace klienta

- **pgAdmin – specifický pro postgres**
- Přihlášení k serveru/connection
- Nová databáze/new database - matbi
- Schéma/schema – public

- **Dbeaver**
- Univerzální

<https://dbeaver.io/download/>  
Community edition



# SQL

## SQL jazyk (Structured Query Language)

- case insensitive
- klíčová slova /keywords – pro názornost VELKÝM písmem/uppercase
- názvy objektů (tabulek, sloupců) / name of objects
  - pouze alfanumerické znaky / only alphanumeric
  - první znak písmeno / first letter
  - omezená délka (ORACLE 32 znaků) / length limit
- operátory / operators
- funkce / function
- různá rozšíření v jednotlivých DB produktech
- SQL příkazy – ve skriptu ukončeny defaultně středníkem (;)
- komentáře / comments odděleny - - nebo v bloku /\* komentar \*/

# Určete primární a cizí klíč v tabulkách

Lékař

ID pracoviště	ID lékaře	Jméno lékaře	Příjmení lékaře
1	1	Petr	Moudrý
1	2	Josef	Mladý
2	3	Jana	Malá
2	4	Petr	Mladý

Pracoviště

ID pracoviště	Název pracoviště
1	Interna
2	Chirurgie
3	Neurologie
4	Urologie
5	Gynekologie

# Praktické cvičení

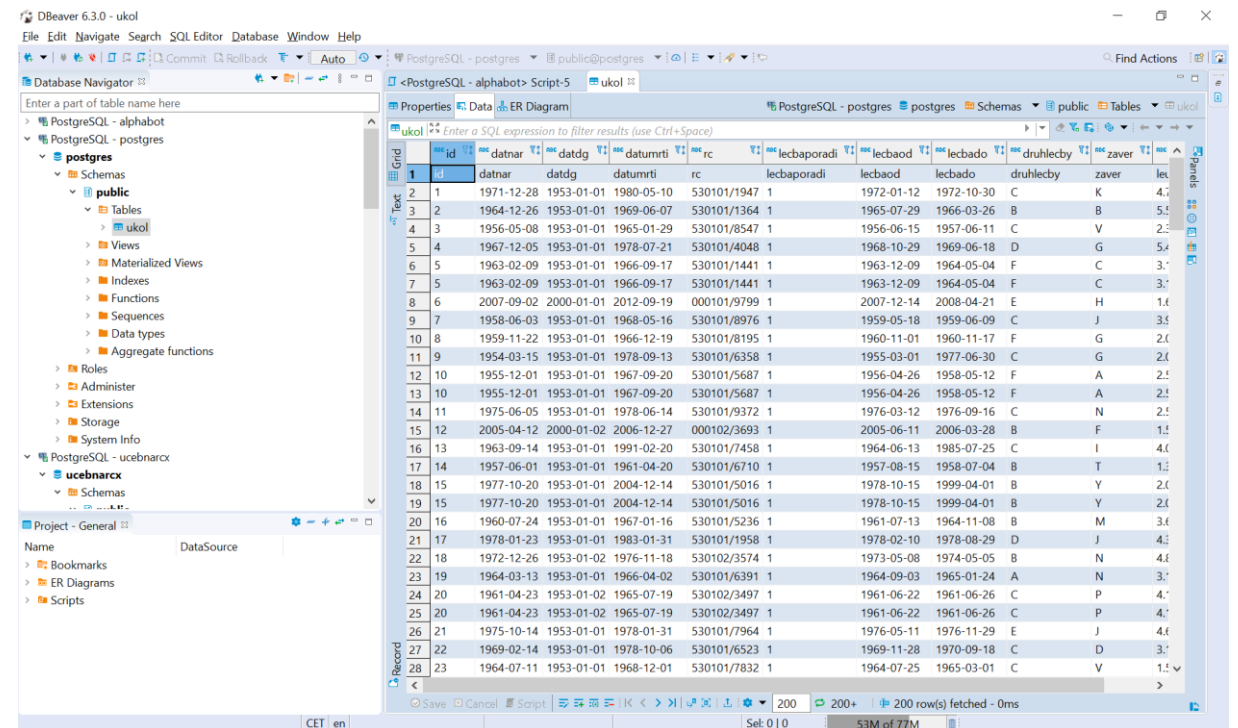


# Instalace a použití klienta

Klient/client

- **Dbeaver**
- Univerzální

<https://dbeaver.io/download/>  
Community edition





# Dbeaver

The screenshot displays the DBeaver 24.2.1 interface. On the left, the Database Navigator shows a tree view of the 'ucebnarncx' database, with the 'public' schema expanded to show various tables. The main window contains a SQL editor with the following query:

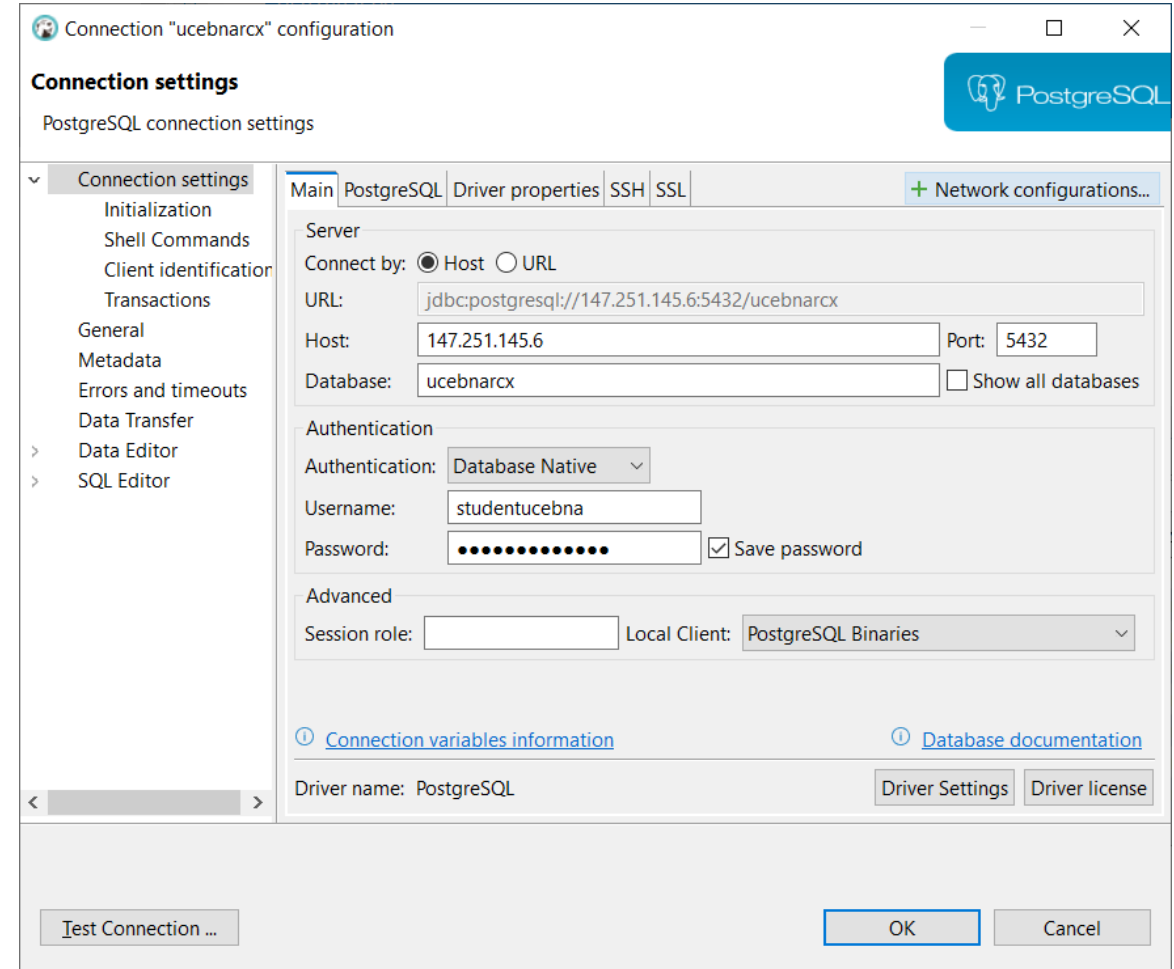
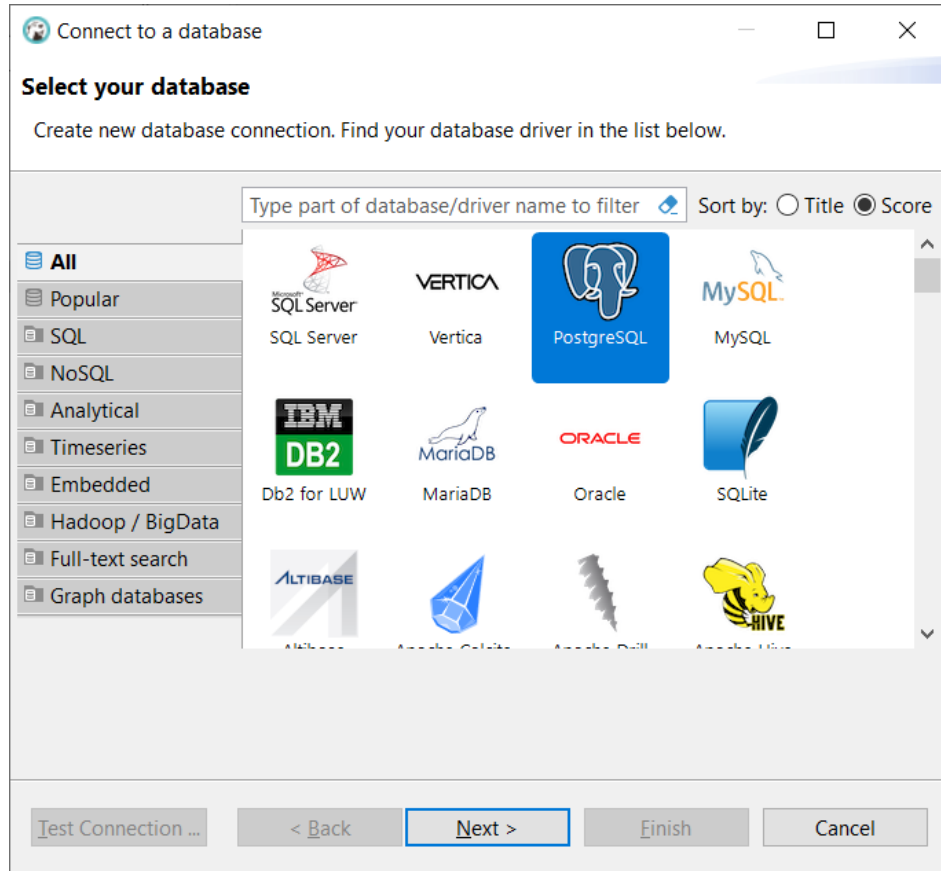
```
select kodzarizeni
,den
,ordinujeod
,ordinujedo
,poznamka
,zdroj
,datumovereni
from ordinacnidoba od
```

Below the editor, the results are displayed in a grid format. The grid has columns for 'Grid', 'A:z kodzarizeni', 'A:z den', 'O:ordinujeod', 'O:ordinujedo', 'A:z poznamka', 'A:z zdroj', and 'D:datumovereni'. The data includes rows with patient IDs, dates, times, and notes.

Grid	A:z kodzarizeni	A:z den	O:ordinujeod	O:ordinujedo	A:z poznamka	A:z zdroj	D:datumovereni
152	751072100000041A00	Pá	10:00:00	12:00:00	pouze nemocni	internet	2024-03-03
153	283005640000043A00	Po	07:00:00	18:00:00		internet	2024-03-03
154	283005640000043A00	Út	07:00:00	18:00:00		internet	2024-03-03
155	283005640000043A00	St	07:00:00	18:00:00		internet	2024-03-03
156	283005640000043A00	Čt	08:00:00	18:00:00		internet	2024-03-03
157	283005640000043A00	Pá	07:00:00	12:00:00		internet	2024-03-03
158	040659130000040A00	Po	07:00:00	13:00:00		internet	2024-03-03
159	040659130000040A00	Út	12:00:00	18:00:00		internet	2024-03-03

The status bar at the bottom indicates that 170 rows were fetched in 0.105s (0.002s per row) on 2025-02-16 at 10:49:27. The connection is identified as 'ucebnarncx'.

# Nová konexe



server (host) 147.251.145.6

database:ucebnarcx

studentucebna / RCX2019ucebna

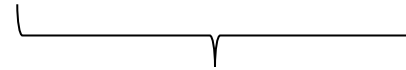
# Data definition language

DDL

# Datové typy/datatypes

Obsah sloupců tabulky určuje přiřazený tzv. datový typ

Verze PostgreSQL



<https://www.postgresql.org/docs/15/static/datatype.html>

Skupina / Group	Name of datatype
Číslo	numeric(x,y)
Text (omezený)	varchar(x)
Text neomezený	text
Datum /date	date
Datum + čas / date with time	timestamp
Časový interval/time interval	interval

# CREATE TABLE/DROP TABLE

PostgreSQL

```
CREATE TABLE jmeno  
(  
    text Varchar(200),  
    cislo Numeric(5,2),  
    datum Timestamp  
);
```

ORACLE

```
CREATE TABLE jmeno  
(  
    text          VARCHAR2(200),  
    cislo         NUMBER(9,1),  
    datum        DATE  
);
```

**DROP TABLE tabulka;**

# ALTER TABLE

- ALTER TABLE tabulka ADD sloupec typ;
- ALTER TABLE tabulka DROP sloupec;
- ALTER TABLE tabulka ADD PRIMARY KEY (sloupec);

<https://www.postgresql.org/docs/15/static/ddl.html>

**M U N I**  
**M E D**

Institut  
biostatistiky  
a analýz

# Data **MANIPULATION** language

**DML**



# SELECT – zobrazení dat v tabulce

```
SELECT sloupec1,sloupec2, sloupec3  
FROM tabulka
```

```
WHERE sloupec1 = 50
```

```
ORDER BY sloupec3
```

# SQL - SELECT

SELECT \* FROM tabulka; -- všechny řádky i sloupce tabulky / all rows , all columns

/\* vybrané sloupce, všechny řádky /selected columns \*/

SELECT sloupec1, sloupec2, sloupec1 + sloupec2 AS soucet FROM tabulka;

/\* všechny sloupce, vybrané řádky / selected rows, all columns \*/

SELECT \* FROM tabulka WHERE sloupec1 = 1;

SELECT \* FROM tabulka WHERE sloupec2 = 'Jan' ; -- text do apostrofů

SELECT \* FROM tabulka  
WHERE sloupec1 = 1 AND sloupec2 > 10 AND sloupec3 < sloupec 4

# SQL - SELECT

--Setřídění výstupu (ORDER BY)

SELECT \* FROM tabulka ORDER BY sloupec1; -- vzestupné třídění / ascending

SELECT \* FROM tabulka ORDER BY sloupec2 DESC; -- sestupné třídění / descending

SELECT \* FROM tabulka ORDER BY sloupec1, sloupec2 DESC --kombinace

# SQL - SELECT

--Sumární výstupy = Agregační funkce

SELECT COUNT(\*) FROM tabulka -- počet řádků v tabulce / number of rows

/\* suma, aritmetický průměr, minimum, maximum \*/

SELECT SUM(sloupec1), AVG(sloupec2), MIN(sloupec3), MAX(sloupec4)  
FROM tabulka

SELECT COUNT(\*), sloupec1 FROM tabulka - nelze / error

# INSERT DATA

```
INSERT INTO tabulka (sloupec1, sloupec2, sloupec3)
VALUES (cislo, 'text', TO_DATE ('datum', 'dd.mm.yyyy'));
```

```
INSERT INTO jmeno (CISLO, TEXT, DATUM)
VALUES (2.3,'testovací řetězec', TO_DATE ('05032011','ddmmyyyy'));
```

```
INSERT INTO tabulka (sloupec1, sloupec2, sloupec3)
SELECT sloupec1,sloupec2, sloupec3 FROM tabulka2;
```

```
INSERT INTO jmeno (cislo, text)
SELECT uco, lastname FROM student
WHERE sex= 'muž' ;
```

# UPDATE, DELETE

```
UPDATE tabulka SET sloupec = hodnota;  
UPDATE jmeno SET cislo = cislo+1;
```

```
UPDATE tabulka SET sloupec = hodnota WHERE sloupec2 = hodnota;  
UPDATE jmeno SET datum = SYSDATE WHERE text = 'Klimeš';
```

```
DELETE FROM tabulka;  
DELETE FROM jmeno;
```

```
DELETE FROM tabulka WHERE ...;  
DELETE FROM jmeno WHERE cislo > 5;
```

<https://www.postgresql.org/docs/15/static/dml.html>

# TRANSAKCE

TRANSAKCE/transaction = set DML příkazů/commands – všechny nebo žádný  
(all or none)

Ukončení transakce/ end of transaction

COMMIT; = potvrzení změn

or

ROLLBACK; = zrušení/cancel změn

V PGSQL automatický commit !!

Zahájení transakce / start transaction

BEGIN TRANSACTION

Nepotvrzené transakce nevidí ostatní, brání provedení změn jiných uživatelů  
(zamykání sloupců, řádků, tabulek)

**=> Co nejkratší transakce!**

<https://www.postgresql.org/docs/15/static/sql-begin.html>

# NULL, prázdná hodnota

NULL nerovná se 0 / different meaning 0 and NULL

NULL nelze testovat standardními operátory / don't use standard operators with NULL  
~~WHERE X = NULL OR X <> NULL~~

Správně / correct: WHERE **sloupec IS NULL**  
**sloupec IS NOT NULL**

**ALL / BUT:**

**UPDATE tabulka SET sloupec = NULL WHERE sloupec IS NOT NULL**  
**UPDATE tabulka SET sloupec = 5 WHERE sloupec = 1**

**NULL – téměř veškeré operace (funkce, operátory) nad hodnotou NULL  
opět vrací NULL**  
**5 + NULL = NULL**

<https://www.postgresql.org/docs/15/static/functions-comparison.html>



# Cvičení 1 / task 1

- Vytvořte tabulku (vaše příjmení jako název tabulky)

create table (your lastname as table name)

- textový sloupec *firstname*
- textový sloupec *lastname*
- datumový sloupec *date\_of\_enrollment*

- Vložte záznam, naplňte první 2 sloupce tabulky svým jménem (firstname, lastname)

insert your name as a record to the table (without date)

- Doplňte datum zápisu na aktuální datum

fill in date of enrollment to an existing row

- Přidejte libovolný další řádek

insert another record

- Ověřte počet řádků v tabulce

check the number of rows

- Smažte řádek se svým jménem

remove the record with your name

# Cvičení 2 / task 2

1. Tabulka student – vložení vlastního záznamu

```
INSERT INTO student (uco, firstname, lastname, enrollmentdate)
VALUES (514000, 'Hana', 'Malá', '2024-02-02');
```

2. Tabulka student – doplnění pohlaví a data narození u vlastního záznamu

```
UPDATE student
SET sex = , birthdate =
WHERE uco =
```

# Cvičení 3 / Task 3

## Tabulka/Table STUDENT

- Zobrazte celý obsah tabulky / select all data from table
- Zobrazte jen jméno a příjmení / select only firstname and lastname columns
- Setříd'te výstup podle příjmení, jména / order by lastname, firstname
- Kolik má tabulka řádků? / how many rows are in the table?
- Vyberte pouze svůj záznam (své UČO) / select only your UCO row
- Minimum, maximum a průměrná hodnota sloupce *UCO*?

Min, max , average of UCO

- Minimum, maximum z jména / min, max of firstname
- Kolik máme v seznamu mužů? / How many men are in the table?

# Funkce a operátory

Functions and operators

# Vyzkoušejte si funkce

- SELECT function()
  - SELECT function(parameter)
  - SELECT function(parameter1, parameter2)
  - SELECT ABS(-5)
- 
- SELECT 1/2
  - SELECT 1/2.0

# Operators and functions for number

Function	Description
+, -, *, /	Aritmetické operace
ABS(cislo)	Absolutní hodnota
SIN(cislo), COS(cislo), TAN(cislo)	Číslo v radiánech
POWER(cislo,exp)	Mocnina
SQRT(cislo)	Druhá odmocnina
MOD(cislo, cislo)	Zbytek po dělení
LN(cislo)	Přirozený logaritmus
LOG(cislo)	Dekadický logaritmus
EXP(x)	$e^x$
ROUND(x,[n])	zaokrouhlení
CEIL()	zaokrouhlení nahoru

<http://www.postgresql.org/docs/15/static/functions.html>

# Operators and function for text

Funkce	Popis	
	Spojení textových řetězců	
SUBSTR(text, od, počet)	Vrací podřetězec textu dle pozice	
INSTR(text, subtext)	Hledání podřetězce v textu, vrací pozici	ORACLE
STRPOS(text, subtext)		PG
REPLACE(text, puvodni, nove)	Nahrazení podřetězce	
LOWER(text)	Převod na malá písmena	
UPPER(text)	Převod na velká písmena	
LTRIM(text), RTRIM(text)	Odstranění mezer zleva zprava	
LENGTH(text)	Délka řetězce	
TRANSLATE(text, znaky,znaky)	Nahrazení po znacích	
SPLIT_PART(text,oddelovac, poradi)	Rozdělení řetězce	

# Operators and functions for datetime

Funkce	Popis	
CURRENT_DATE	Aktuální datum	Operátor
CURRENT_TIMESTAMP NOW()	Aktuální datum a čas	Operátor Funkce()
Datum +/- počet dnů	Přičítání, odečítání dnů	
Datum - datum	Počet dnů mezi daty (desetinná část udává časový rozdíl)	
MONTHS_BETWEEN(datum, datum)	Rozdíl datumů	ORACLE
AGE(datum, datum)	Rozdíl datumů	PG
interval '1 year 2 months 3 days 4 hours 5 minutes 6 seconds'	Datový typ, možnost přičítat, odčítat	PG
ADD_MONTHS(datum, počet)	Přičtení měsíců	ORACLE



# Operators and functions for datetime

DATE_PART(text, timestamp)	Extrakce component	PG
Century, day, dow, doy, hour, isoyear, minute, month, second, week, year		
TO_CHAR, TO_DATE	Konverze Datum ↔ Text	PG, ORACLE
dd,mm,yyyy, HH, HH24, mi, ss, Month, Day, D, DDD, W, WW, IYYY, IDDD, IW		
EXTRACT(co FROM interval)	Extrakce z intervalu	PG

# Other operators and functions

[NOT] IN (hodnota, hodnota, ...)	Rovnost [NEROVNOST] se skupinou hodnot
COALESCE	Vrací první NOT NULL argument

# Cvičení

- Textové funkce
  - Spojte jméno a příjmení do jednoho sloupce
  - Znovu rozdělte jméno a příjmení do vlastních sloupců
- Datumové funkce
  - Spočítejte aktuální věk studentů v tabulce student
  - `SELECT DATE_PART('year', AGE(CURRENT_DATE, birth))  
FROM student`
  - Zjistěte , který den v týdnu odpovídá vašemu datu narození

# Cvičení

<https://www.postgresql.org/docs/15/static/functions-formatting.html>

```
SELECT TO_CHAR(TO_DATE('2018-01-01','yyyy-mm-dd'), 'DAY')
```

## **ISO rok, týden – „zlomový“ čtvrtek**

```
SELECT TO_CHAR(TO_DATE('2017-01-01'),'yyyy-mm-dd'), 'iyyy')
```

```
SELECT TO_CHAR(TO_DATE('2017-01-01'),'yyyy-mm-dd'), 'iw')
```

# Cvičení

Přidejte do tabulky student sloupce / add columns to the table student

- Studium varchar(3)
- Rocnik int
- facebook varchar(1)
- instagram varchar(1)
- X varchar(1)
- other varchar(500)

Doplňte hodnoty sloupců u svého řádku / fill in columns for your row (uco)

- Studium – B=bakalar, M = magistr, D= doktorand
- Používání sociálních sítí / using social network (Y/N)
- Jiné aplikace jako text / other apps as text

# Cvičení

- **Vytvořte si kopii tabulky** Student s názvem student\_prijmeni
  - Sloupec příjmení převed'te na velká písmena / convert lastname to uppercase
  - Řádky se studiem =M - přičt'ete k ročníku +3, ke studiu D přičt'ete 5
  - Odháčkujte sloupec jmeno
  - Zkraťte jméno na 1. znak a tečku
  - Ověřte, zda ženy mají v příjmení „ová“ / compare the end of lastname with sex
- 
- Kolikátého bude za 7 let 7 měsíců a 7 dní /  
what date will be after 7 years 7 months and 7 days
- 
- Vymažte řádky ze **své** tabulky / delete all rows from your table
  - Zrušte **svoji** tabulku / remove your table

# Homework

Instalace PostgreSQL

<https://www.enterprisedb.com/downloads/postgres-postgresql-downloads>

[Instalace Dbeaver](https://dbeaver.io/download/)

<https://dbeaver.io/download/>

Community edition

Plnění tabulky Ordinacnidoba