

Databázové systémy a SQL

Daniel Klimeš

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

Lectures

Každou středu/wednesday od 10:00 – do 11:40

Teoretická přednáška → navazující praktické cvičení

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

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

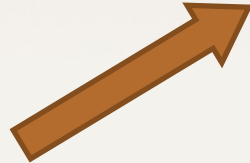
- Základy SQL
 - SELECT, UPDATE, DELTE, INSERT
- Funkce a operátory
- Základy agregace dat
 - GROUP BY, HAVING, AVG, SUM, ...
- Práce s více tabulkami
 - JOIN, UNION
- Zanořené dotazy
- Window funkce
 - Rank, lag, lead, OVER
- Vyhledávání v textu
 - Like, regulární výrazy

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

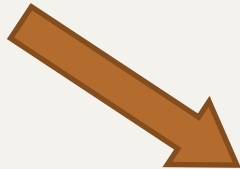
Databáze SQL

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



Statistický SW

Statistika for Windows, R,
SPSS, SAS, MS Excel



Matematicko - statistický skriptový SW

R, Matlab, Maple, programovací jazyk

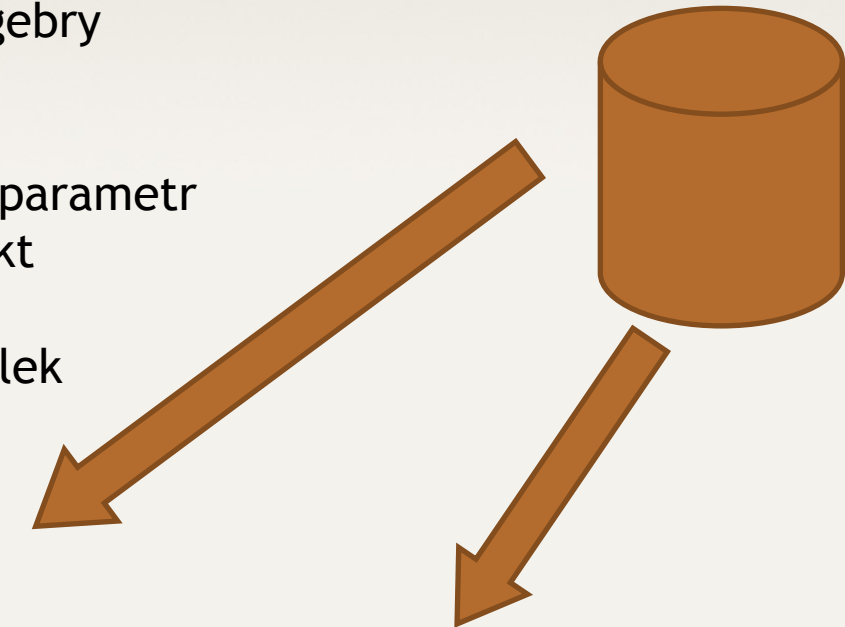
Relační databáze (Relational database management system - RDBMS)
 Relace/relation - termín z relační algebry

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

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

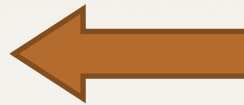
ORACLE
 MS SQL
 DB2
 MySQL
 PostgreSQL
 Firebird

https://en.wikipedia.org/wiki/List_of_relational_database_management_systems

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



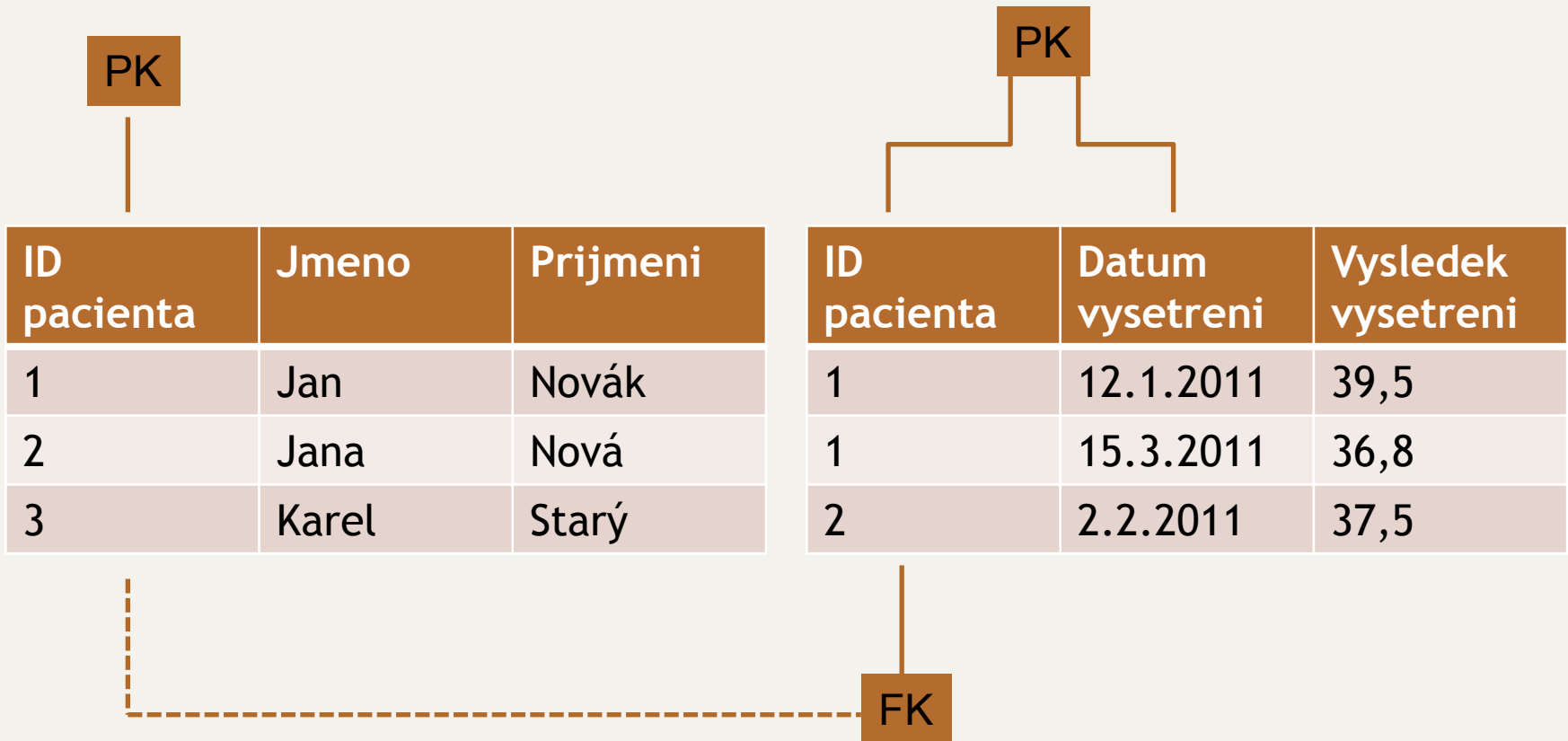
PatientID	ExaminationDate	Result
1	12.1.2011	39,5
1	15.3.2011	36,8
2	2.2.2011	37,5

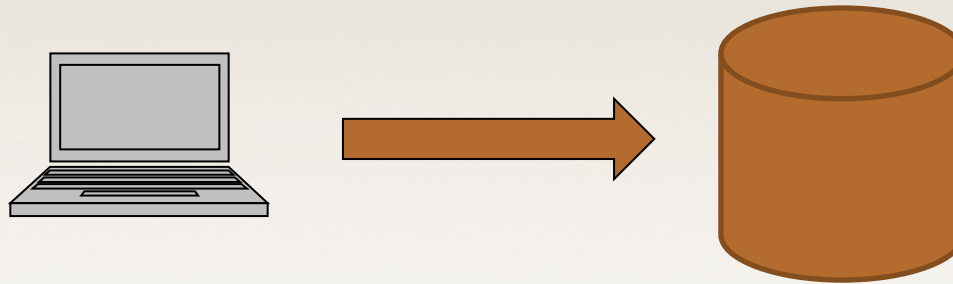
Definice tabulky předchází načtení dat
A table must be created before data import

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





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

Řídicí 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

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 */

Instalace PostgreSQL

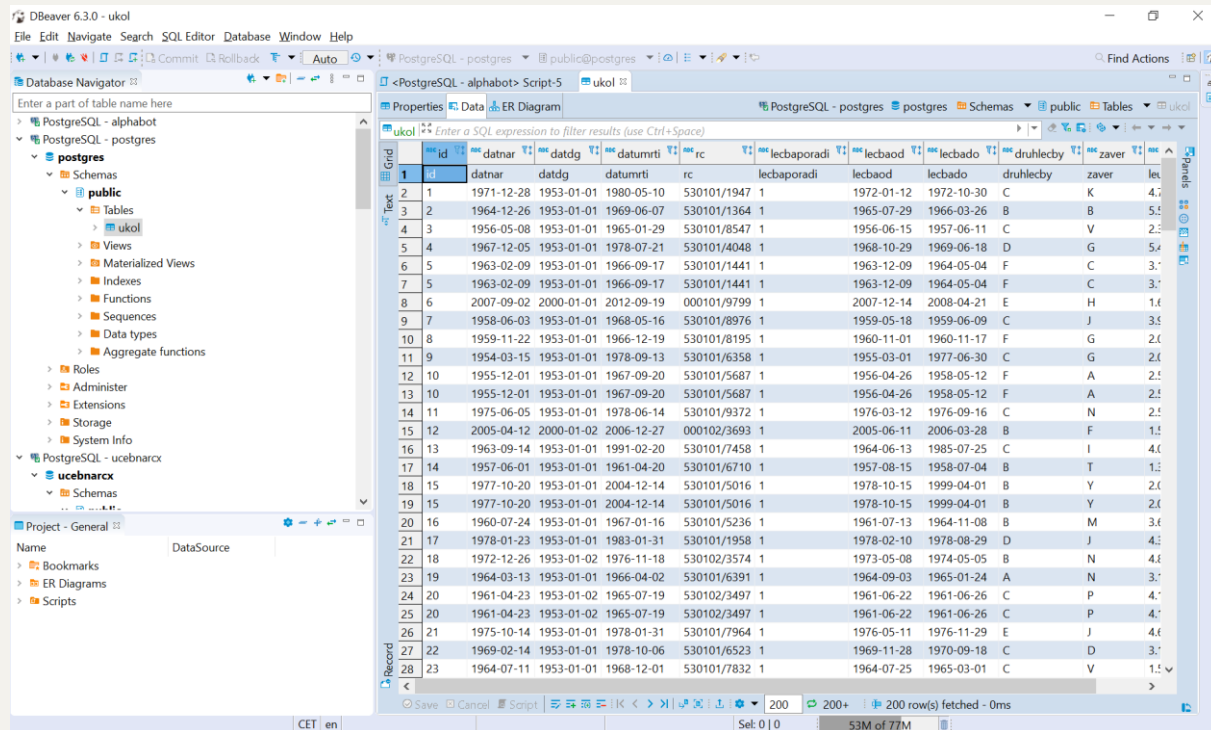
<https://www.enterprisedb.com/downloads/postgres-postgresql-downloads>

Klient/client

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

- **Dbeaver**
- Univerzální

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



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

--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

--Sumární výstupy = Agregční 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

Spuštění prvního skriptu (skript1.sql) / Run script

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 studia, jména / order by study
- 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?