

Databázové systémy a SQL

Daniel Klimeš

Daniel Klimeš

- Vzdělání: Obecná biologie
- PGS: onkologie
- Specializace: klinické databáze
- Databáze ORACLE

- klimes@iba.muni.cz
- Kotlářská 2, budova 11

Každé pondělí od 16:00 – do 17:40

Teoretická přednáška → navazující praktické cvičení

Praktická část : databáze ORACLE 11g

Domácí cvičení: PostgreSQL

Zakončení: zápočet – domácí úkol

zkouška – praktický test, pomůcky bez omezení, časový limit

Grafický statistický SW
Statistika for Windows, MS Excel
SPSS, SAS

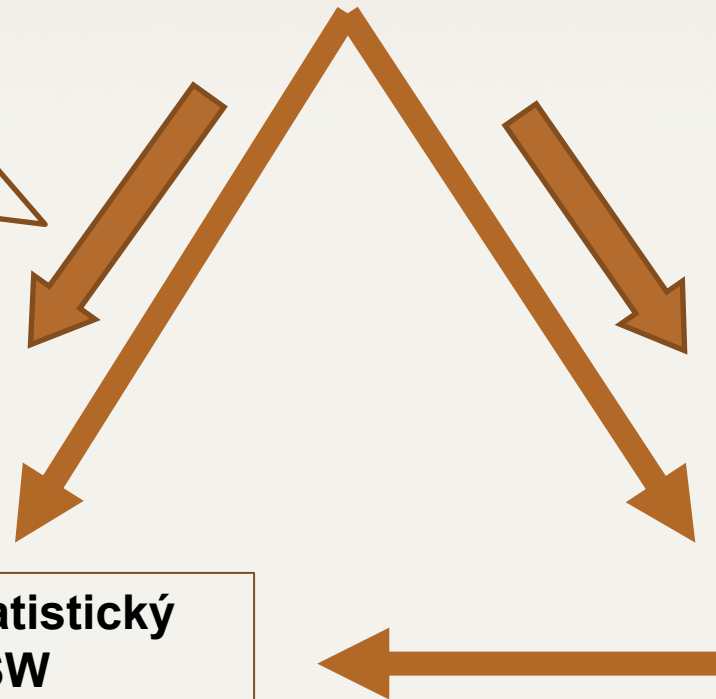
- Potřeba matematického aparátu
- Automatizace zpracování

Rostoucí objem dat
(miliony záznamů)

Matematicko - statistický skriptový SW
R, Matlab, Maple, SPSS, SAS,
programovací jazyk

Databáze SQL

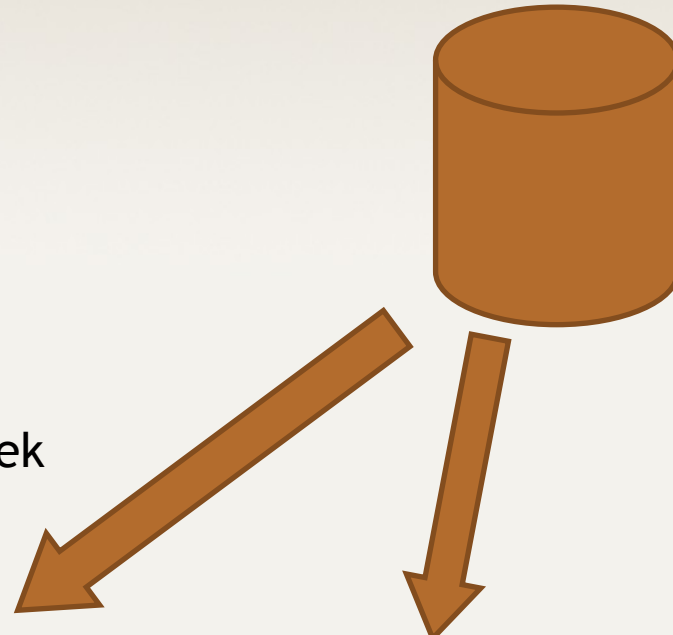
Předzpracování dat
čištění dat
popisná analýza



Relační databáze (RDBMS)
 Relace - termín z relační algebry

Základ: tabulka
 sloupec = atribut/parametr
 řádek = popsáný objekt

Databáze = systém provázaných tabulek



ID pacienta	Jmeno	Prijmeni
1	Jan	Novák
2	Jana	Nová
3	Karel	Starý

ID pacienta	Datum vysetreni	Vysledek vysetreni
1	12.1.2011	39,5
1	15.3.2011	36,8
2	2.2.2011	37,5

Dle dostupnosti

Komerční

ORACLE - databáze*
MS SQL server*
DB2
MS ACCESS
FOX PRO

Freeware

MySQL
PostgreSQL
Firebird

* Okleštěné verze jsou k dispozici zdarma

Dle počtu uživatelů

Jednouživatelské

MS ACCESS
FOX PRO

Víceuživatelské

ORACLE
MS SQL
DB2
MySQL
PostgreSQL
Firebird

Definovaná struktura, do které se vkládají záznamy

Definují se sloupce

- jméno
- datový typ
 - text
 - číslo
 - datum
 - BLOB
- doplňující vlastnosti

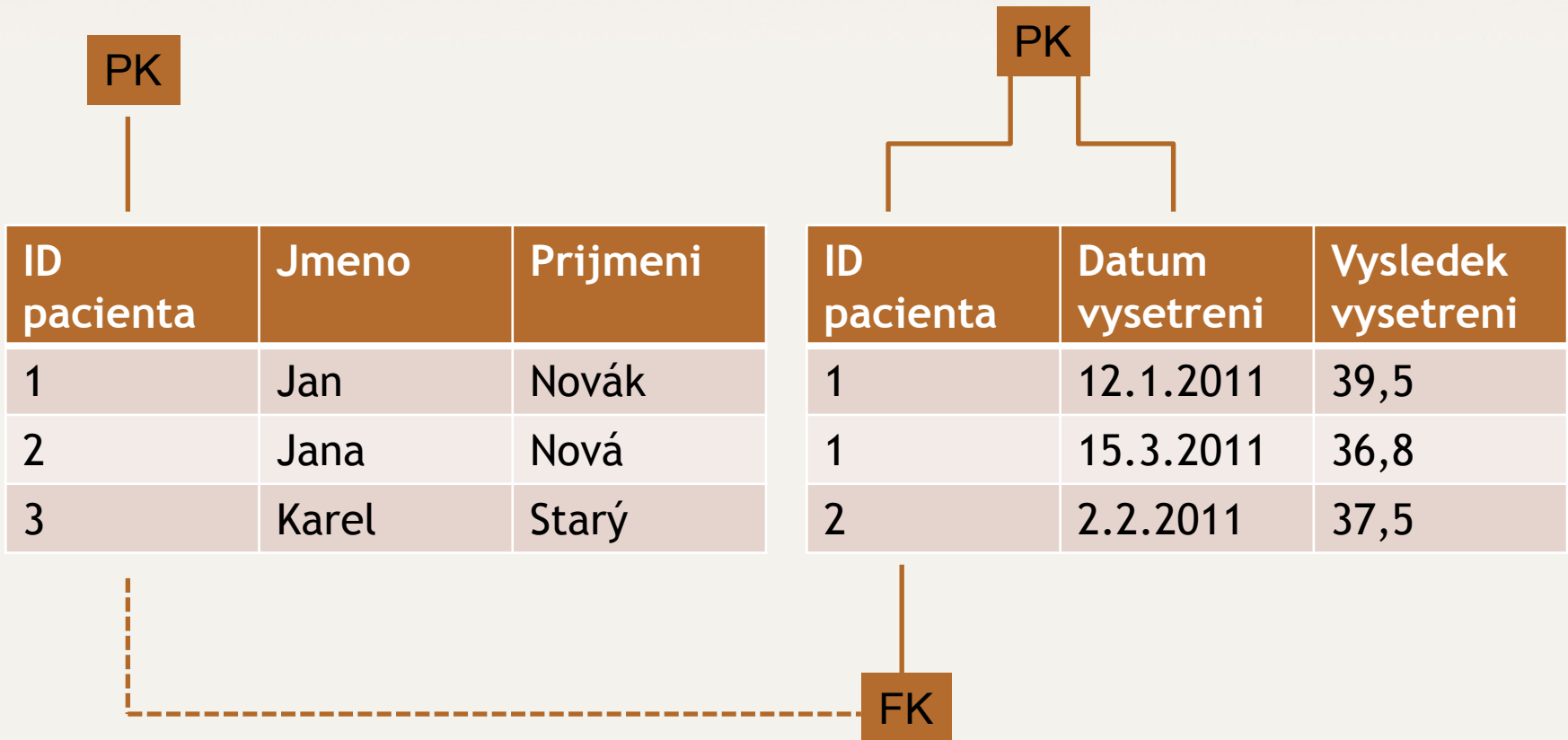


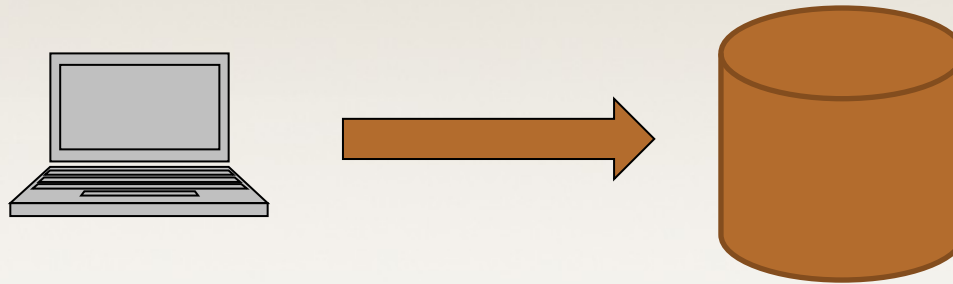
ID pacienta	Datum vyšetření	Vysledek vyšetření
1	12.1.2011	39,5
1	15.3.2011	36,8
2	2.2.2011	37,5

Vybrané sloupce se označují jako klíče (keys)

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

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





Klient = SW umožňující ověření uživatele a spuštění řídicích příkazů

Řídicí příkazy = Structured Query Language - SQL

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

Klient

sqlplus - textový
SQLDeveloper - grafický

Identifikace ORACLE databáze

IP adresa + SID nebo síťový alias (network alias)

Network alias musí být definován na klientském počítači:

TNS (Transparent Network Substrate)

přístupné databáze jsou definované v lokálním souboru tnsnames.ora
../network/Admin

SID = identifikace instance databáze na serveru

SQL developer - připojení

New / Select Database Connection

Connection Name	Connection Details
ord	klimes@//147.251....
ordscr	klimes@//147.251....
SYS	SYS@//147.251.26...

Connection Name: ord
Username: student
Password:
 Save Password

Oracle Access

Role: default
Connection Type: Basic
OS Authentication
Kerberos Authentication
Proxy Connection

Hostname: 147.251.26.130
Port: 1521
 SID: ord
 Service name

Status :

Nápověda Save Clear Test Connect Zrušit

Role: default
Connection Type: TNS
OS Authentication
Kerberos Authentication
Proxy Connection

Network Alias: ORCL
 Connect Identifier

Status :

Nápověda Save Clear Test Connect Zrušit

Připojení k databázi

příkazová řádka - cmd

```
sqlplus login@network_alias
```

```
sqlplus student@TESTORCL
```

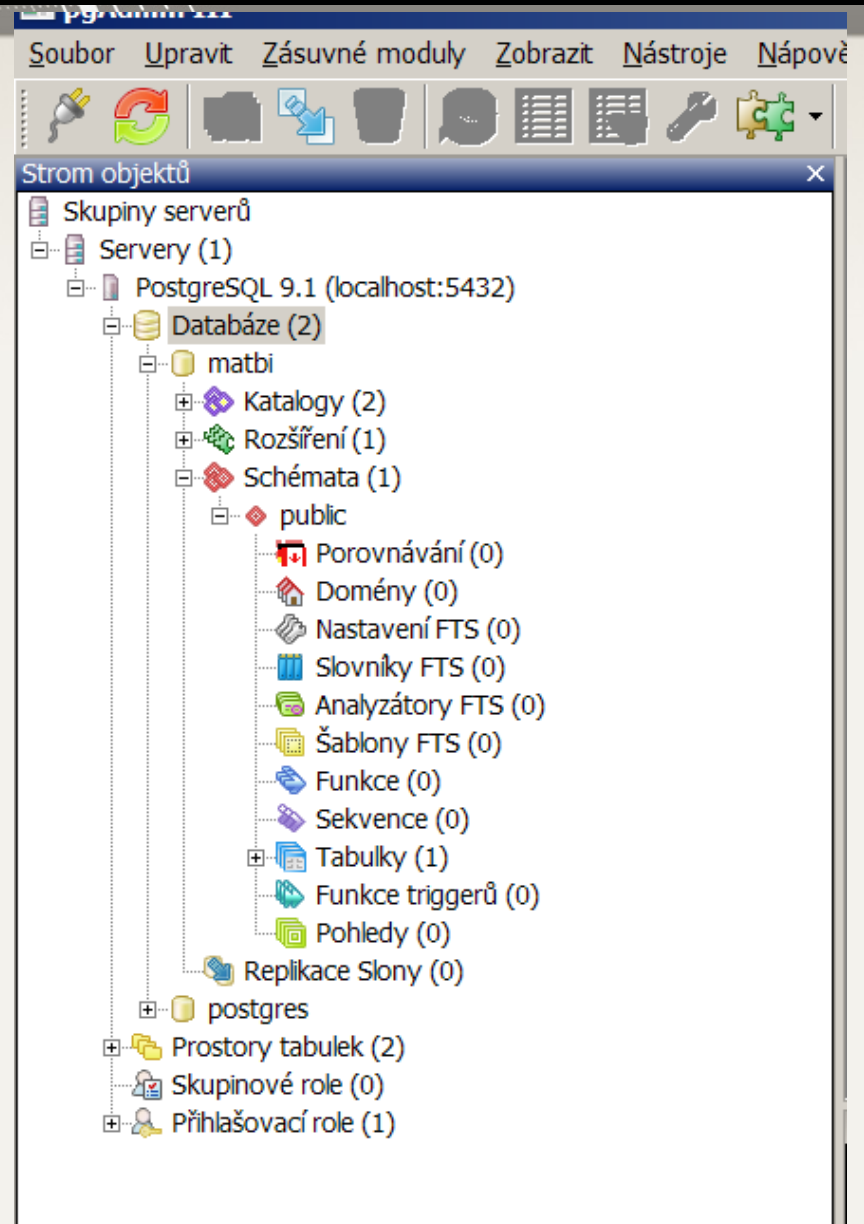
Varianta s přímým zadáním hesla:

```
sqlplus login/heslo@network_alias  
student/DBM753
```

Klient

- pgAdmin
- Přihlášení k serveru
- Nová databáze - matbi
- Schémata – public

- psql
- ve Windows nutné nastavení jazykové sady a fontu



SQL jazyk

- case insensitive
- různá rozšíření v jednotlivých DB produktech
- klíčová slova – pro názornost VELKÝM písmem
- názvy objektů (tabulek, sloupců)
 - pouze alfanumerické znaky
 - první znak písmeno
 - omezená délka (ORACLE 32 znaků)
- operátory
- funkce
- SQL příkazy – ve skriptu ukončeny defaultně středníkem (;)
- komentáře odděleny - - nebo v bloku /* komentar */

SELECT * FROM tabulka; -- všechny řádky i sloupce tabulky

/* vybrané sloupce, všechny řádky */

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

/* všechny sloupce, vybrané řádky */

SELECT * FROM tabulka

WHERE sloupec1 = 1 AND sloupec2 > 10 AND sloupec3 < sloupec 4

--Sumární výstupy

SELECT COUNT(*) FROM tabulka -- počet řádků v tabulce

-- Agregáční funkce

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

SELECT COUNT(*), sloupec1 FROM tabulka - nelze

Tabulka **STUDIES**

- Kolik má sloupců?
- Kolik má řádků?
- Minimum, maximum a průměrná hodnota sloupce *study_id*?
- Minimum, maximum hodnota sloupce *study_name*
pouze řádky *is_active*= 1 ?
- Výpis sloupců *study_name*, *description*, všechny řádky, dle abecedy
podle *study_name*

Seskupení položek

```
SELECT sloupec, count(*), MAX(sloupec2), MIN(sloupec2) FROM tabulka
GROUP BY sloupec;
```

```
SELECT sloupec, count(*), MAX(sloupec2), MIN(sloupec2) FROM tabulka
WHERE sloupec2 > 1 and ...
GROUP BY sloupec;
```

```
SELECT sloupec, count(*), MAX(sloupec2), MIN(sloupec2) FROM tabulka
GROUP BY sloupec
HAVING count(*) > 1
```

Výpis počtu studií pro jednotlivé verze trialdb → 2 sloupce

trialdbversion, počet řádků

To samé pouze pro ***is_active*** = 1

Výpis ***principle_investigator***, kteří mají na starosti více jak 5 aktivních studií
principle_ivestigator, počet studií

Vytvoření tabulky (ORACLE)

- DDL příkazem
- v grafickém prostředí

CREATE TABLE *jmeno*

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

PostgreSQL

CREATE TABLE jmeno

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

jmeno = do 30 znaků (písmena, čísla, podtržítka) bez mezer, začíná písmenem
Řádkování příkazu – nepovinné, pouze pro lepší čitelnost

```
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 ('05.03.2011', 'dd.mm.yyyy'));
```

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

```
INSERT INTO jmeno (cislo, text)
SELECT study_id, text FROM studies
WHERE is_active = 2;
```

```
INSERT INTO jmeno (cislo, text)
SELECT MAX(study_id), principal_investigator FROM studies
GROUP BY principal_investigator
```

```
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;
```

TRANSAKCE = sada DML příkazů – všechny nebo žádný

Transakci zahajuje první příkaz

Ukončení transakce

COMMIT; = potvrzení změn (DDL příkazy => automatický commit)

ROLLBACK; = zrušení změn

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

Co nejkratší transakce!

V PGADMIN automatický commit !!

- Nainstalovat PostgreSQL
- Vytvořit databázi matbi
- Vytvořit tabulku student se sloupci
 - jmeno, prijmeni, datum_narozeni, rok_prijeti
- Vložit řádek se svým jménem
- Pomocí update prohodíte jméno a příjmení