

# MonetDB

Natália Potočková

# Obsah

- ◇ 1. Predstavenie systému
- ◇ 2. Podporované modely
- ◇ 3. Podporované jazyky pre dotazovanie
- ◇ 4. Príklady
- ◇ 5. Využitie
- ◇ 6. Zhodnotenie

# Základné vlastnosti

- ◇ Ukladá dáta v stĺpcoch
- ◇ Open source
- ◇ Implementovaná v jazyku C
- ◇ OS
  - ◇ Linux, OS X, Solaris, Windows
- ◇ Podporované jazyky:
  - ◇ C, C++, Java, JavaScript (Node.js), Perl, PHP, Python, R, Ruby
- ◇ Transakcie: ACID

# Podporované modely

- ◇ Primárny model
  - ◇ Relačný DBMS
- ◇ Sekundárne modely
  - ◇ Document store → JSON document
  - ◇ Spatial DBMS → GeoSpatial dáta

# Podporované jazyky pre dotazovanie

- ◇ SQL
- ◇ Extenzie:
  - ◇ C/C++-UDF Funkcie
  - ◇ Data Vaults (manažment externých dát)
  - ◇ GeoSpatial
  - ◇ Embedded R
  - ◇ Embedded Python

# Príklady

- ◇ Vytvorenie tabuľky, vloženie hodnôt, vypísanie obsahu tabuľky:

```
sql>CREATE TABLE players(id int, name varchar(30), surname varchar(50), sponsor varchar(40));
operation successful
sql>INSERT INTO players VALUES (1, 'Novak', 'Djokovic', 'Head');
1 affected row
sql>INSERT INTO players VALUES (2, 'Daniil', 'Medvedev', 'Dunlop');
1 affected row
sql>INSERT INTO players VALUES (3, 'Alexander', 'Zverev', 'Head');
1 affected row
sql>INSERT INTO players VALUES (4, 'Rafael', 'Nadal', 'Babolat');
1 affected row
sql>SELECT * FROM players;
+-----+-----+-----+-----+
| id   | name      | surname | sponsor |
+-----+-----+-----+-----+
| 1   | Novak     | Djokovic | Head    |
| 2   | Daniil   | Medvedev | Dunlop  |
| 3   | Alexander | Zverev   | Head    |
| 4   | Rafael    | Nadal    | Babolat |
+-----+-----+-----+-----+
4 tuples
sql>_
```

# Príklady

- ◆ Agregácia a triedenie:

```
sql>SELECT sponsor, COUNT(id) AS sponsored_players FROM players GROUP BY sponsor ORDER BY sponsor;
+-----+-----+
| sponsor | sponsored_players |
+-----+-----+
| Babolat |          1 |
| Dunlop  |          1 |
| Head    |          2 |
+-----+-----+
3 tuples
sql>_
```

- ◆ Join a ostatné štandardné funkcie by sa napísali ako v SQL



# Príklady

## ◆ Update hodnoty

```
sql>UPDATE players SET sponsor = 'Tecnifibre' WHERE id = 2;
1 affected row
sql>SELECT * FROM players
more>;
+-----+-----+-----+-----+
| id    | name      | surname | sponsor |
+=====+=====+=====+=====+
| 1     | Novak     | Djokovic | Head    |
| 2     | Daniil   | Medvedev | Tecnifibre |
| 3     | Alexander | Zverev   | Head    |
| 4     | Rafael    | Nadal    | Babolat |
+-----+-----+-----+-----+
4 tuples
sql>
```



# Príklady

## ◆ C/C++ UDF funkcie

```
sql>CREATE FUNCTION multiply(input INTEGER)
more>RETURNS INTEGER
more>LANGUAGE C {
more>    result->initialize(result, input.count);
more>    // loop over the input values
more>    for(size_t i = 0; i < input.count; i++) {
more>        if (input.is_null(input.data[i])) {
more>            // handle NULL values
more>            result->data[i] = result->null_value;
more>        } else {
more>            // handle regular values
more>            result->data[i] = input.data[i] * 2;
more>        }
more>    }
more>};
sql>INSERT INTO integers VALUES (1), (2), (NULL), (3), (4);
5 affected rows
sql>SELECT i, multiply(i) FROM integers;
```

# Príklady

- ◆ GeoSpatial data
  - ◆ Package geos

```
sql>CREATE TABLE forests(id INT,name TEXT,shape MULTIPOLYGON);
sql>CREATE TABLE buildings(id INT,name TEXT,location POINT,outline POLYGON);
sql>INSERT INTO forests VALUES(109, 'Green Forest',
more>'MULTIPOLYGON( ((28 26,28 0,84 0,84 42,28 26), (52 18,66 23,73 9,48 6,52 18)), ((59 18,67 18,67 13,59 13,59 18)))');
sql>INSERT INTO buildings VALUES(113, '123 Main Street',
more>'POINT( 52 30 )',
more>'POLYGON( ( 50 31, 54 31, 54 29, 50 29, 50 31) )');
sql>INSERT INTO buildings VALUES(114, '215 Main Street',
more>'POINT( 64 33 )',
more>'POLYGON( ( 66 34, 62 34, 62 32, 66 32, 66 34) )');
sql>SELECT forests.name,buildings.name
more>FROM forests,buildings
more>WHERE forests.name = 'Green Forest' and
more>    Overlaps(forests.shape, buildings.outline) = true;
```

# Využitie

- ◇ Analytické databázy
- ◇ Data warehouse
- ◇ OLAP
- ◇ Data mining
- ◇ Vedecké databázy

# Zhodnotenie

## ◇ Plusy

- ◇ Data warehousing
- ◇ Dobrá a efektívna práca s obrovským množstvom dát
- ◇ Podpora jazyka R → ľahšie použitie pre štatistikov
- ◇ Používaná v Big Data science

## ◇ Mínusy

- ◇ Predpokladá analytické dáta (veľa agregácií)
- ◇ Predpokladá, že sa hot dataset (dáta, s ktorými pracujeme) vojdú do hlavnej pamäte

# Zdroje

- ◇ <https://www.monetdb.org/documentation-Jan2022/user-guide/>
- ◇ <https://www.monetdb.org/documentation-Jan2022/admin-guide/>
- ◇ <https://www.monetdb.org/documentation-Jan2022/dev-guide/>
- ◇ <https://db-engines.com/en/system/MonetDB>

Ďakujem za pozornosť