IN4331: WEB DATA MANAGEMENT

Development assignment group 25 by:

Maria Gatou
Ioannis Papakonstantinopoulos
Konstantinos Touloumis
Ingmar Wever

THEORY

RELATIONAL

PostgreSQL

DOCUMENT-BASED

mongoDB

GRAPH-BASED

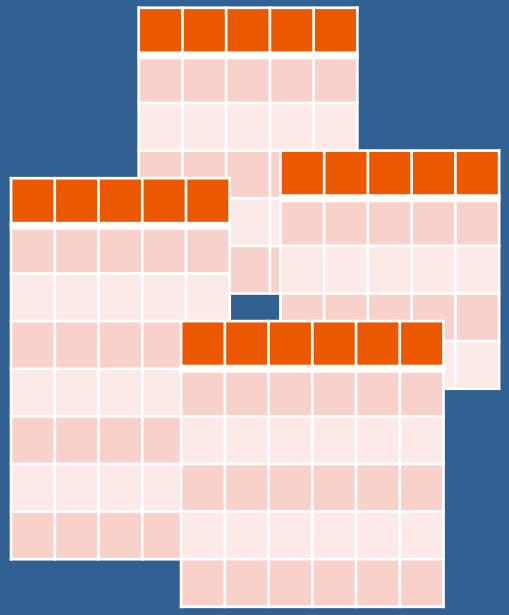
neo4j

COMPARISON

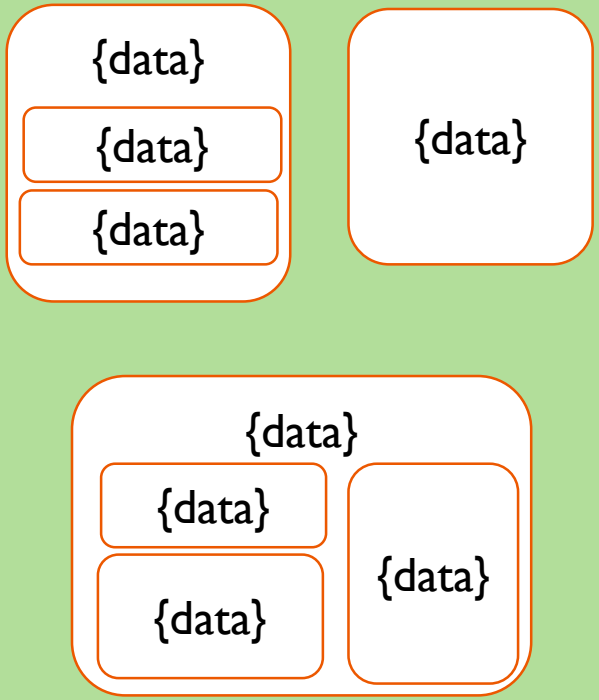
THEORY

SQL NoSQL

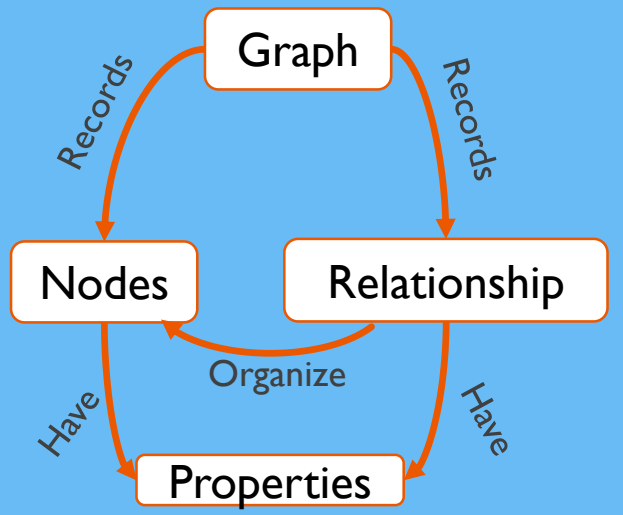
RELATIONAL



DOCUMENT-BASED



GRAPH-BASED



SHOWCASE

RELATIONAL

DOCUMENT-BASED

GRAPH-BASED

POSTMAN

Pre-processing

- Easy import of data
- Vacuum. Reclaim storage occupied by dead tuples
- Reindex. Rebuilds indexes

Implementation

- Python with Psycopg library
- Built-in functions (e.g. count) to create statistics for genres and actors
- String partial match (like expressions)

mongoDB

Pre-processing

- Export csv files from PostgreSQL with joined tables
- Import csv files to MongoDB using Pyhton scripts
- Create indexes to speed up access to the database

Implementation

2 types of queries were used

- Search query
- Bulk query to retrieve the wanted documents IDs

Pre-processing

- Extraction of PostgreSQL data to CSV-files
- Import of the CSV-files into Neo4j
- Creation of Indexes
- Creation of relationships

Implementation

- Similar functionality to SQL with some differences in the syntax
- JOINS replaced by relationships eg:
  - `MATCH m:movies{idmovies:3}-[r:MOVIES_GENRE]->(g:genres) RETURN a.title`
- Five queries for Full Info, one query for specific Info

COMPARISON

RELATIONAL	DOCUMENT-BASED	GRAPH-BASED
PostgreSQL	mongoDB	neo4j
Complex queries Combing of related tables Good support	Schema-less Easy scale-out Cost	Absence of Foreign Keys Quick traverse of relationships
Easy tasks need many queries	Less efficient disk usage No support for transactions Functionality	Lack of advanced data types New product

EXPERIENCE

RELATIONAL	DOCUMENT-BASED	GRAPH-BASED
		
Very complex queries can be carried out. But this is also means that for example when full info is needed. Five queries have to be made.	The immediate and fundamental difference between MongoDB and an RDBMS is the underlying data model. A relational database structures data into tables and rows, while MongoDB structures data into collections of JSON documents. JSON is a self-describing, human readable data format.	It was hard to create relationships using CSV files. Lack of advanced data types and parametrized queries. Low quantity and simplicity of Relationships. This means there is no reason for using it instead of PostgreSQL in this case.

QUESTIONS?

Development assignment group 25 by:

Maria Gatou Ioannis Papakonstantinopoulos
Konstantinos Touloumis Ingmar Wever