

Graf baze

Baze podataka služe za pohranjivanje, održavanje i pristupanje podacima. Najpoznatiji tip baza podataka su relacijske baze podataka te se za izvođenje upita nad tim podacima koristi SQL (engl. *Structured Query Language*). Kod relacijskih podataka podaci su prikazani u tablicama koje su međusobno povezane. S druge strane postoje NoSQL baze podataka koje koriste druge načine upravljanja podacima. Primjer takvih su graf baze korištene u ovom eksperimentalnom radu. Graf baze podataka temeljene su na grafu, a graf sadrži čvorove (engl. *Nodes*) i veze među čvorovima (engl. *Relationships*), a čvorovi i veze nadalje mogu imati dodatne attribute (engl. *Properties*). Ovakav tip baza podataka poseban je zbog njihove efikasnosti prilikom obrade vrlo povezanih podataka te olakšano izvođenje samih upita.

Korištene tehnologije

Prilikom izrade eksperimentalnog rada korišteni su Docker i Neo4j za postavljanje baze podataka i Apache JMeter u svrhu benchmarkinga baze podataka.

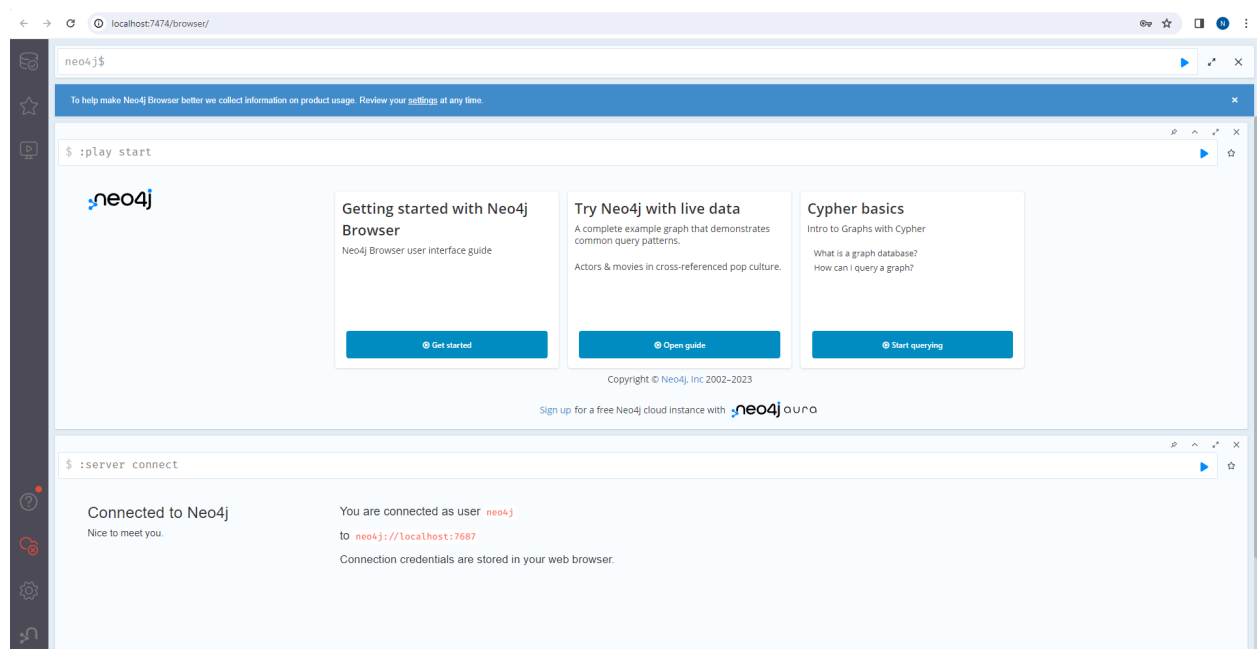
Neo4j i Docker-compose

Neo4j je sustav za upravljanje graf bazama podataka te je otvorenog koda. Koristi Cypher upitni jezik, omogućuje jednostavno dodavanje čvorova i veza te drive za različite programske jezike kao što su Java, JavaScript, Python itd. Koristi se u različitim domenama, od svijeta financija, preko prirodnih znanosti i farmaceutskih kompanija pa sve do telekomunikacija i upravljanja mrežom. Docker-compose je alat koji u YAML formatu omogućuje ručno definiranje i pokretanje kontejnera. To također omogućuje da se aplikacije željene aplikacije pokreću bez potrebe za instalacijom istih na računalo.

docker-compose.yml:

```
docker-compose.yml
1  version: ""
2  services:
3    neo4j:
4      image: neo4j:5.13.0
5      container_name: neo4j
6      ports:
7        - "7474:7474" # HTTP Bolt port
8        - "7687:7687" # Bolt port
9      environment:
10       NEO4J_AUTH: admin/lozinka123 # Set your preferred username and password
11      volumes:
12        - ./data:/data
13        - ./logs:/logs
14        - ./conf:/conf
15        - ./plugins:/plugins
16
```

Naredbom docker-compose up podiže se kontejner aplikacije Neo4j. Preko porta 7474 pristupa se Neo4j aplikaciji te se korisnik sa svojim korisničkim imenom i lozinkom (autentifikacija) spaja na bazu podataka.



Rad sa grafovskim bazama podataka

Važni pojmovi u svim bazama podataka su učitavanje, ograničenja, indeksi, pretraživanje i upiti. Grafovske baze posebno nude rješenje pronalaženja najkraćeg puta između dva čvora. Za ovaj dio korištena je isključivo službena dokumentacija koja se pojavljuje prilikom pokretanja kontejnera, odnosno preglednika na odgovarajućem portu i povezivanjem na bazu.

Učitavanje (engl. *Load*)

Navedene operacije odvijaju se u pregledniku u editoru. Učitavanje, odnosno stvaranje čvorova vrši se naredbom CREATE, primjerice:

```
CREATE (Keanu:Person {name:'Keanu Reeves', born:1964})
```

Gdje () označava čvor, ee:Person - ee je čvor varijable i Person je oznaka čvora, a { } sadrži svojstva koja opisuju čvor. Na kraju se klikom na dugme za pokretanje izvrši operacija stvaranja čvora.

Ograničenja (engl. *Constraints*)

Kreiranjem jedinstvenih ograničenja svojstava čvorova osigurava se da vrijednosti svojstava budu jedinstvene za sve čvorove s određenom oznakom. Dodavanjem jedinstvenog ograničenja implicitno se dodaju indeks na to svojstvo. Primjerice:

```
CREATE CONSTRAINT FOR (n:Person) REQUIRE (n.name) IS UNIQUE
```

Indeksiranje (engl. *Index*)

Općenito rečeno, indeksi u bazama podataka služe za poboljšanje performansi pretrage i dohvaćanja podataka. Moguće je kreirati indekse na jednom ili više svojstava za sve čvorove koji imaju određenu oznaku. Primjerice:

```
CREATE INDEX FOR (m:Movie) ON (m.released)
```

Pretraživanje (engl. Find)

Naredbom MATCH vrši se pretraživanje na sljedeći način. Primjerice, želimo pronaći glumca koji se zove Tom Hanks:

```
MATCH (tom:Person {name: "Tom Hanks"}) RETURN tom
```

Ili pronaći filmove objavljene u 1990-ima i vratiti njihova imena:

```
MATCH (nineties:Movie) WHERE nineties.released >= 1990 AND  
nineties.released < 2000 RETURN nineties.title
```

Upiti (engl. Query)

Moguće je koristiti veze kako bi se našli odgovarajući uzorci u grafu. Primjer veza u korištenoj bazi podataka može biti ACTED_IN ili DIRECTED. Primjerice, želimo saznati u kojim filmovima je Tom Hanks glumio:

```
MATCH (tom:Person {name: "Tom  
Hanks"})-[:ACTED_IN]->(tomHanksMovies) RETURN tom,tomHanksMovies
```

Ili tko je glumio sa Tom Hanksom:

```
MATCH (tom:Person {name: "Tom  
Hanks"})-[:ACTED_IN]->(m)<-[:ACTED_IN]-(coActors) RETURN  
DISTINCT coActors.name
```

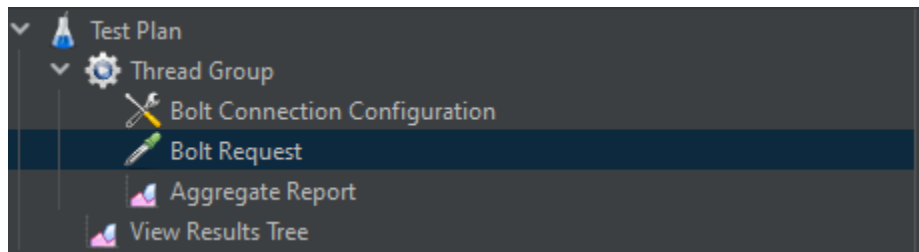
Najkraći put (engl. *Shortest Path*)

Jedan od načina korištenja je pomoću ugrađene funkcije shortestPath(). Primjerice, želimo naći najkraći put između čvorova Kevin Bacon i Meg Ryan:

```
MATCH p=shortestPath(  
(bacon:Person {name: "Kevin Bacon"})-[*]-(meg:Person {name: "Meg  
Ryan"})  
)  
RETURN p
```

Benchmarking

Testiranje baze podataka izvršavati će se pomoću alata Apache JMeter. Temeljen je na Javi, stoga je potrebno imati instaliranu Javu na računalima prije njegovog korištenja. Potrebno je izraditi plan testiranja. Plan testiranja opterećenja i performansi baze će izgledati ovako:

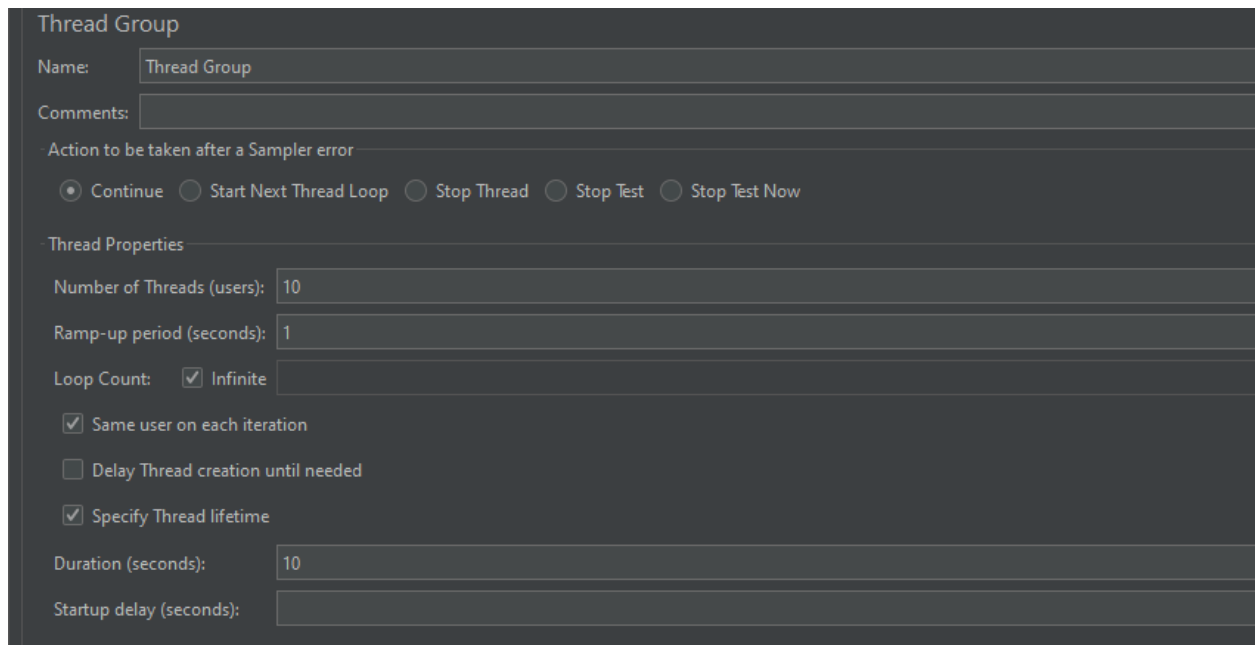


Gdje Thread Group određuje broj korisničkih niti (virtualnih korisnika) koji će istovremeno izvršavati zahtjeve prema bazi podataka. Bolt Connection Configuration omogućuje pristupanje Neo4j graf bazi putem JMetera. Bolt Request predstavlja zahtjev koji se šalje Neo4j bazi podataka putem Bolt protokola te može uključivati razne upite ili operacije koje se izvršavaju nad bazom podataka prilikom testiranja. Aggregate Report daje statistiku testiranja te sadrži informacije kao što su prosječno vrijeme odgovora baze podataka na zahtjeve, minimalnog i maksimalnog vremena odgovora, broja uspješnih zahtjeva itd. Zadnje, View Results Tree daje detaljan pregled rezultata testiranja u obliku stabla, odnosno pregled pojedinačnih zahtjeva, odgovora i drugo.

Za početak je najvažnije konfigurirati Bolt protokol. Znamo da je default port 7687, kao i username i lozinku naše baze podataka.

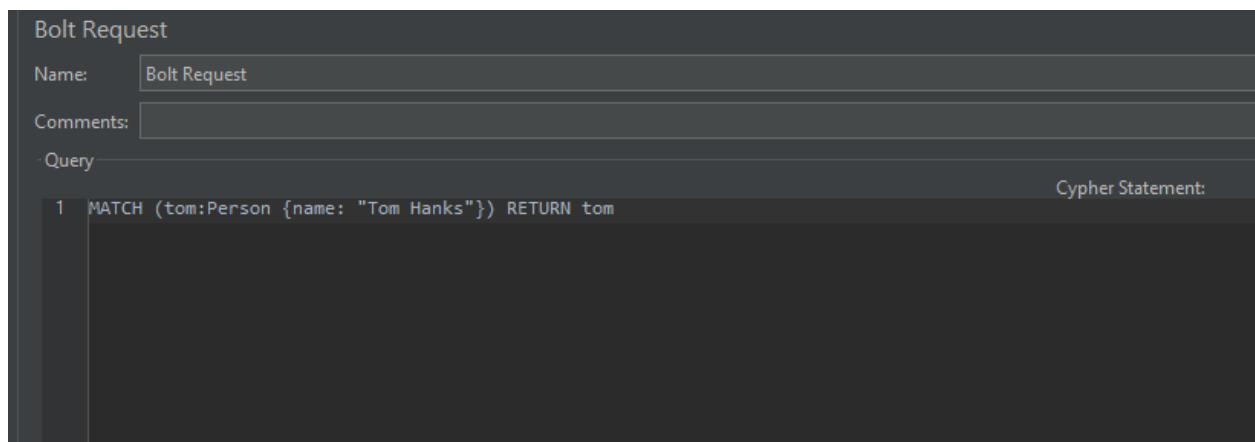
Bolt Connection Configuration	
Name:	Bolt Connection Configuration
Comments:	
Bolt Configuration	
Bolt URI:	bolt://localhost:7687
Username:	neo4j
Password:	*****
Connection Pool Max Size:	100

Postavljaju se opcije o broju virtualnih korisnika te o životnom vijeku niti unutar Thread Group opcija:



The screenshot shows the 'Thread Group' configuration window. It includes fields for 'Name' (Thread Group), 'Comments', and 'Action to be taken after a Sampler error' with radio buttons for 'Continue', 'Start Next Thread Loop', 'Stop Thread', 'Stop Test', and 'Stop Test Now'. The 'Thread Properties' section contains: 'Number of Threads (users): 10', 'Ramp-up period (seconds): 1', 'Loop Count: [checked] Infinite', 'Same user on each iteration: [checked]', 'Delay Thread creation until needed: [unchecked]', 'Specify Thread lifetime: [checked]', 'Duration (seconds): 10', and 'Startup delay (seconds):'.

Te na kraju se zadaje određena operacija ili upit unutar Bolt Requesta:



The screenshot shows the 'Bolt Request' configuration window. It includes fields for 'Name' (Bolt Request), 'Comments', and 'Query'. The 'Query' field contains a Cypher statement: `1 MATCH (tom:Person {name: "Tom Hanks"}) RETURN tom`. A label 'Cypher Statement:' is positioned to the right of the query text.

Pokretanjem ove skripte pod Aggregate Report biti će vidljiva statistika mjerenja opterećenja i performansi baze podataka. Slike prikazuju mjerenje performansi gore navedene operacije pretraživanja. Prva slika prikazuje kada imamo jednog virtualnog korisnika, a druga kada imamo 10, virtualnih korisnika (obraćamo pažnju na Bolt Request).

Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Maximum	Error %	Throughput	Received KB/sec	Sent KB/sec
Debug Sampler	4460	0	0	0	0	1	0	1	0,00%	446,3/sec	138,93	0,00
Bolt Request	4460	2	2	3	3	3	1	12	0,00%	446,2/sec	115,91	0,00
TOTAL	8920	1	1	2	3	3	0	12	0,00%	892,4/sec	254,81	0,00

Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Maximum	Error %	Throughput	Received KB/sec	Sent KB/sec
Debug Sampler	25275	0	0	0	0	1	0	1	0,00%	2528,0/sec	783,73	0,00
Bolt Request	25275	3	3	5	5	7	2	22	0,00%	2526,7/sec	656,36	0,00
TOTAL	50550	1	1	4	5	6	0	22	0,00%	5053,5/sec	1439,70	0,00

Sljedeće objašnjenja testiranih značajki:

- Samples: ukupan broj zahtjeva ili transakcija koji su izvršeni tijekom testiranja pod određenom oznakom
- Average: prosječno vrijeme odgovora baze podataka na zahtjeve ili transakcije izraženo u milisekundama
- Median: središnja vrijednost u skupu podataka, odnosno vrijednost koja dijeli podatke na dva jednaka dijela
- 90% Line, 95% Line, 99% Line: vrijeme odgovora na zahtjeve koji su se dogodili unutar određenog postotka svih zahtjeva
- Min: minimalno vrijeme odgovora baze podataka na zahtjeve ili transakcije izraženo u milisekundama
- Max: maksimalno vrijeme odgovora baze podataka na zahtjeve ili transakcije izraženo u milisekundama
- Error %: postotak grešaka, omjer neuspješnih zahtjeva u odnosu na ukupan broj zahtjeva izražen u postocima
- Throughput: broj izvršenih zahtjeva u sekundi tijekom testiranja
- Received KB/sec: brzina primanja podataka iz baze podataka izražena u kilobajtima po sekundi
- Sent KB/sec: Brzina slanja podataka prema bazi podataka izražena u kilobajtima po sekundi

Sljedeće će se obavljati nekoliko operacija pretraživanja odjednom (4). Broj virtualnih korisnika iznosi 10. Mjerenja su sljedeća:

Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Maximum	Error %	Throughput	Received KB/sec	Sent KB/sec
Debug Sampler	7939	0	0	0	0	1	0	1	0,00%	795,0/sec	247,32	0,00
Bolt Request	7939	11	10	19	22	30	4	48	100,00%	794,0/sec	134,91	0,00
TOTAL	15878	5	1	15	19	27	0	48	50,00%	1588,0/sec	381,92	0,00

U ovom slučaju opterećenje iznosi 100%, odnosno došlo je do pogreške. Radi se o tome da izvršavanje više neovisnih MATCH naredbi nije direktno dozvoljeno

```
Query
1 MATCH (cloudAtlas:Movie {title: "Cloud Atlas"})
2 RETURN cloudAtlas;
3
4 MATCH (tom:Person {name: "Tom Hanks"})
5 RETURN tom;
```

Zbog toga se kombiniraju odvojene MATCH naredbe u jedan upit te omogućuje i jedan i drugi rezultat odjednom pomoću naredbe WITH.

```
Query
1 MATCH (cloudAtlas:Movie {title: "Cloud Atlas"}), (tom:Person {name: "Tom Hanks"})
2 RETURN cloudAtlas, tom;
3
```

Mjerenja su ponovno izvršena te nema pogrešaka.

Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Maximum	Error %	Throughput	Received KB/sec	Sent KB/sec
Debug Sampler	23025	0	0	0	0	1	0	2	0,00%	2303,4/sec	711,09	0,00
Bolt Request	23025	3	4	5	6	8	2	27	0,00%	2302,5/sec	649,83	0,00
TOTAL	46050	2	2	5	5	7	0	27	0,00%	4605,0/sec	1360,63	0,00

Do sada su izvršavale operacije i upiti vezani uz metodu READ. Sada će se testirati izvod operacije CREATE koji je vezan uz WRITE metodu. Kreirati će se dva čvora i pratiti performanse istog. Broj virtualnih korisnika sada iznosi jedan, a petlja traje sekundu. Mjerenja:

Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Maximum	Error %	Throughput	Received KB/sec	Sent KB/sec
Debug Sampler	1	0	0	0	0	0	0	0	0,00%	∞/sec	0,00	0,00
Bolt Request	1	30	30	30	30	30	30	30	0,00%	33,3/sec	7,85	0,00
TOTAL	2	15	0	30	30	30	0	30	0,00%	66,7/sec	18,13	0,00

Nema grešaka. Mjerenjima za veći broj korisnika je prošao bez grešaka te su stvoreni svi novi čvorovi istih parametara.

Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Maximum	Error %	Throughput	Received KB/sec	Sent KB/sec
Debug Sampler	12791	0	0	0	0	1	0	2	0,00%	1280,1/sec	397,46	0,00
Bolt Request	12791	7	6	10	12	17	3	33	0,00%	1279,4/sec	301,10	0,00
TOTAL	25582	3	2	8	10	15	0	33	0,00%	2558,5/sec	698,25	0,00

Zaključak

Neo4j je izdržljiva baza podataka koja omogućuje brzu obradu povezanih podataka, optimizaciju upita i skalabilnost. Funkcionalnosti kao što su modeliranje grafova, Cypher jezik, fleksibilnost, mogućnost integracije te vizualizacija podataka čine ovu bazu podataka jednom od najpopularnijih kada su u pitanju graf baze.