

NoSQL baze podataka

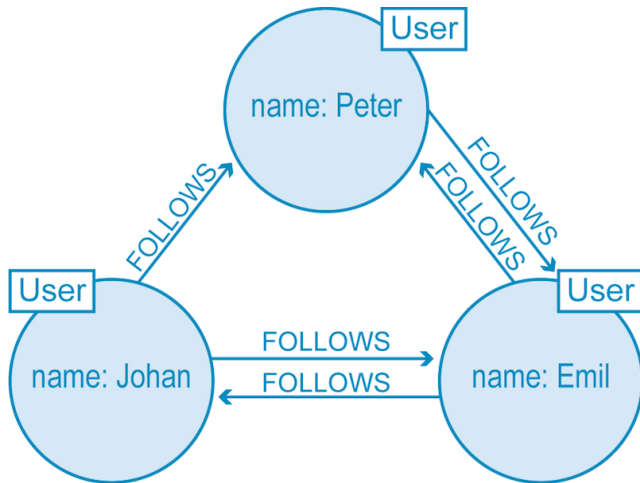
Predavanje 9: Graf orijentisane baze podataka, Jednostavan sistem preporuka



Univerzitet u Novom Sadu
Fakultet Tehničkih Nauka

Uvod

- ▶ Kao što im naziv sugeriše, struktura ovih baza podataka sledi strukturu grafa, sa više čvorova i veza medju njima.
- ▶ Graf baze podataka se kao rešenje primenjuju kada je potrebno rukovati povezanim podacima
- ▶ Podaci se smeštaju kao prarovi ključ-vrednost koji opisuju čvorove i veze.



(<https://neo4j.com/blog/why-graph-databases-are-the-future/>)

- ▶ Najvažniji argument za korišćenje graf baza podataka je njihova mogućnost prolaska kroz čvorove prateći veze
- ▶ Drugi bitan parametar je mogućnost da efikasno reše bilo koji graf algoritam, kao što su veze sa n-stepeni ili najkraći put.
- ▶ Za razliku od drugih vrsta NoSQL baza podataka graf baze su više prilagodjene da se izvršavaju na konfiguraciji s jednim serverom a ne na klasterima servera kao većina NoSQL baza podataka
- ▶ TitanDB je bio primer gde se model graf-a spušta na Apache Cassandra-u, ali to nije native graf baza
- ▶ Gotovo svaka operacija mora da prođe kroz nekoliko čvorova i veza.

- ▶ Baze podataka orijentisane ka grafovima su NoSQL tip baze podataka koje se baziraju na teoriji grafova.
- ▶ Karakteriše ih korišćenje graf strukture za skladištenje i prezentovanje podataka.
- ▶ Najčešće se koriste za skladištenje i prikaz kompleksnih i nestruktuiranih medjusobno povezanih podataka
- ▶ Kao i kod problema koji se teže mogu obradivati korišćenjem nekog drugog modela.

Model podataka

- ▶ Socijalne mreže su najbolji ali ne i jedini primer za tip podataka za koji je graf sistem baza podataka idealno optimizovan.
- ▶ Grafovi se sastoje od:
 - ▶ Čvorova – Koji predstavljaju entitete
 - ▶ Grana – Koje povezuju čvorove
 - ▶ Osobina – Predstavljaju atribut, reprezentuju i čuvaju informacije
- ▶ Osobine se uglavnom mogu dodavati i na same čvorove ali i na veze
- ▶ Osobine koji se dodaju, kao i kod ostalih NoSQL baza podataka su uglavnom ključ vrednost
- ▶ Čvorovi i veze istih grupa, ne moraju imati iste osobine

- ▶ Baze podataka orijentisane ka grafovima
 - ▶ Nastale sa popularizacijom društvenih mreža – Veliki broj korisnika koji poseduju veze sa drugim korisnicima, statusima, komentarima itd.
 - ▶ Pogodne za podatke kojima rukuje semantički Web
- ▶ Ovo npravno nisu jedine primene, ali su tu našle svoju prvu veliku ulogu
- ▶ Graf i reprezentacija znanja kod nekih internet pretraživača je često graf

- ▶ Čvorovi predstavljaju entitete.
- ▶ Grane povezuju čvorove.
- ▶ Osobine su definisane atributima i služe za čuvanje informacija .
- ▶ Svaki element baze može da sadrži direktan pokazivač na neki drugi element.
- ▶ Karakteristike graf baze kao što je neograničen broj svojstava čvora
- ▶ Ali i neograničen broj veza između čvorova
- ▶ Obično više veza između dva ista entitea reprezentuju drugi tip veze

- ▶ Omogućuju kompleksnu strukturu i čine je pogodnom za čuvanje specifičnih podataka.
- ▶ Ova činjenica kao i specifična struktura omogućuju značajno bolje performanse u odnosu na druge tipove baza (relacione) kada su u pitanju kompleksni i dinamički podaci.
- ▶ U poredjenju sa relacionim bazama podataka, graf orijentisane baze su boljih performansi u pogledu brzine rada sa podacima.
- ▶ Moguće je grupisati velike skupove podataka jer tipično ne zahtevaju JOIN operacije.

- ▶
- ▶ Obzirom da ne zavise od šeme u potpunosti pogodne su za trenutnu manipulaciju sa podacima sa dinamičkom promenom šeme podataka.
- ▶ Najpopularnija baza podataka u ovom trenutku je open source Neo4j graf baza podataka
- ▶ Pprilično popularna medju aplikacijama društvenih mreža kod kojih se odnosi medju podacima predstavlja grafovima.
- ▶ Postoji dosta primera aplikacija gde je graf baza bila dovoljna za sve use-case-ve aplikacije
- ▶ Tipičan primer je dating aplikacija – Neo4J ima dosta whitepaper-a na tu temu

- ▶ Neo4j je jedna od (verovatno) najpopularnijih baza podataka iz grupe graf orijentisanih baza podataka.
- ▶ Postoje dva načina na koji se Neo4j može koristiti kao skladište podataka:
 - ▶ Prvi je korišćenje nezavisnog servera koji podržava REST servise za pristup podacima – aplikativni pristup
 - ▶ Drugi način korišćenje API — eksplorativni pristup
- ▶ Prednost korišćenja nezavisnog servera je podrška za pristup iz bilo kog programskog jezika.

- ▶ Neke od osnovnih karakteristika (ne samo) Neo4j skladišta podataka su:
 - ▶ visok nivo skalabilnosti – nekoliko milijardi čvorova/relacija
 - ▶ Dostupnost se povećava pri distribuciji na veći broj računarskih sistema
 - ▶ Transakcije podržavaju pun skup ACID osobina
 - ▶ Koristi se JSON za reprezentaciju podataka u oba smera (od klijenta ka serveru i obrnuto)
 - ▶ Ugradjena podrška za Cypher – deklarativni graf upitni jezik koji radi po principu pattern matching-a
 - ▶ Brza pretraga zahvaljujući algoritmima za graf upite
 - ▶ pristup preko RESTful API – ja.

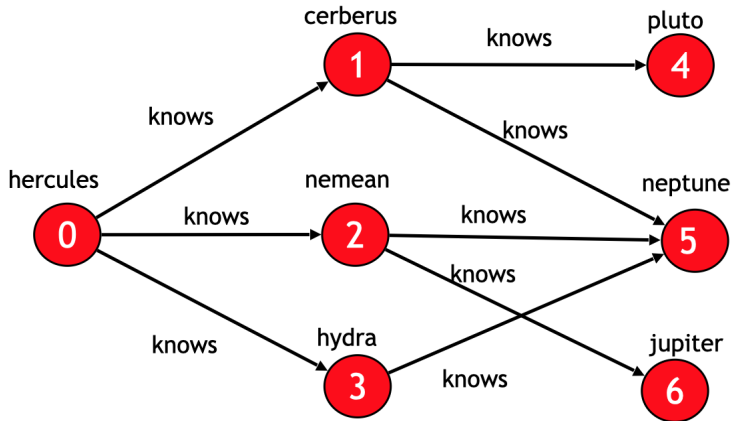
- ▶ Veliki broj slabo struktuiranih podataka povezanih medjusobnim konekcijama doveli su do primene graf baze podataka u nekoliko oblasti kao što su:
 - ▶ socijalne mreže
 - ▶ telekomunikacione mreže,
 - ▶ zatim finansijske transakcije
 - ▶ sigurnosna analiza itd.
- ▶ Nedostaci se (uglavnom) ogledaju u komplikovanim procesima prilikom migracije podataka
- ▶ Složeni upiti predstavljaju problem prilikom pisanja

Uvod

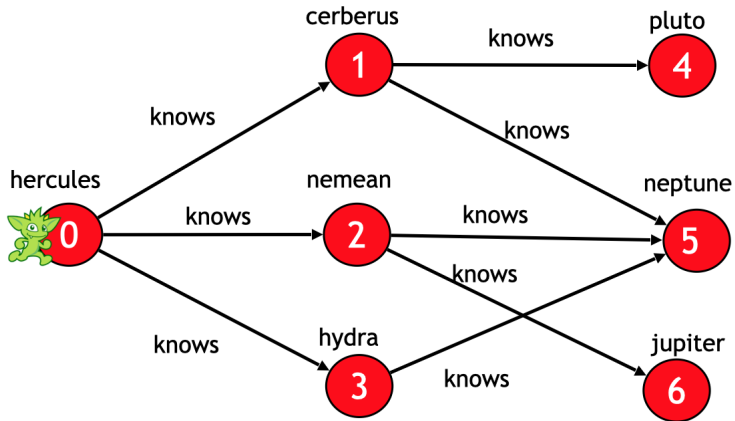
- ▶ Sistem preporuka je mehanizam kojim se korisnicima preporučuje novi sadržaj
- ▶ Novi sadržaj se preporučuje na osnovu nekakvih osobina korisnika
- ▶ Ove osobine mogu biti unapred pozante, ili se mogu zaključiti iz njegovog ponašanja i njegovih veza itd.
- ▶ Kada se zaključuje na osnovu veza, bitno je da dodajemo konkretne veze u podatke na osnovu kojih bi mogli da redimo preporuke
- ▶ Znači moramo imati nekakvu istoriju i povezanost u podacima

- ▶ Pošto radili graf orijentisane baze gde se entiteti povezuju nekih vezama
- ▶ Ovaj mehanizam možemo iskoristiti za sistem preporuka
- ▶ Sve što treba da uradimo jeste da imamo entitete u sistemu, i da dodamo odgovarajuće veze
- ▶ Nije loše dodati i neke osobine na veze i entitete zbog bolje preciznosti
- ▶ I vrlo jednostavno smo dobili brz i prilično precizan sistem preporuka koji može da radi prilično generično
- ▶ Ovo nam može biti od koristi kada aplikacija započinje svoj život, i ako brzo želimo da dodamo tu funkcionalnost

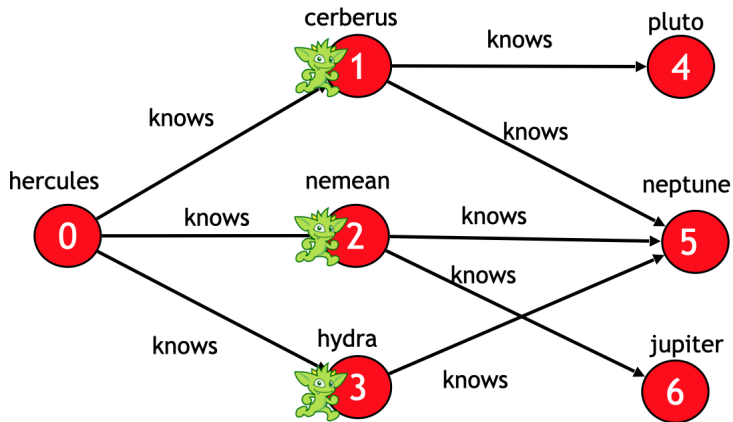
Jednostavan primer



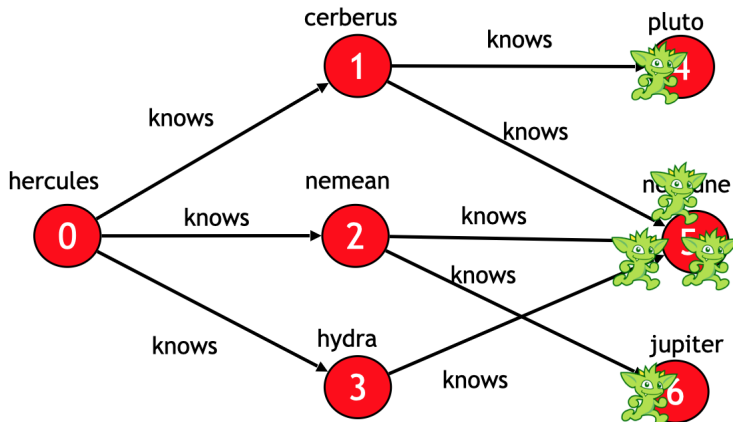
Jednostavan primer



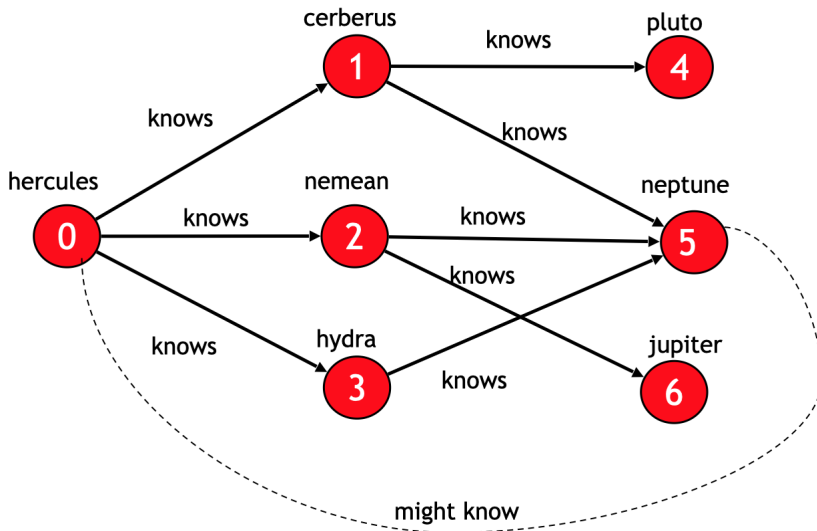
Jednostavan primer



Jednostavan primer



Jednostavan primer



- ▶ Ovo je vro jednosavan primer
- ▶ Ali demonstrira kako se može smao uz upotrebu pravog modela rešiti relativno težak problem na lak način
- ▶ Ovaj sistme se može vremenom proširiti
- ▶ Moguće je dodavati težine, ili neke druge parametre na veze ili čvorove
- ▶ I uz upotrebu malo matematike napraviti sistme još preciznijim

Dodatni materijali

- ▶ Making Sense of NoSQL A guide for managers and the rest of us
- ▶ Database Internals
- ▶ NoSQL Distilled A Brief Guide to the Emerging World of Polyglot Persistence
- ▶ Seven Databases in Seven Weeks
- ▶ Use Cases: Recommendation Engine & Product Recommendation System

Pitanja

Pitanja :) ?