

NoSQL - Analýza súčasného stavu a vymedzenie pojmu :)

Analýza súčasného stavu

Pojem databáza v dnešnej dobe pozná skoro každý. Už minulosti mali ľudia potrebu zhromažďovať dáta, zväčša išlo o evidenciu obyvateľov, zoznam kníh alebo klientov obchodov, bánk. V súčasnosti počty týchto záznamov presahujú stovky miliónov. S rapidným nárastom dát sa paralelne stretávame s javom, kedy začínajú byť dáta neštruktúrované. Toto má za následok, že tradičná správa dát pomocou relačných schém a referencií prestáva byť použiteľná. To viedlo k vzniku špecializovaných typov databázových systémov.

Teoretické vymedzenie pojmu NoSQL databáza

Samotný termín NoSQL sa vyznačuje svojou komplikovanou povahou, nakoľko neexistuje jeho presná definícia a taktiež nemá ani všeobecne ustálený význam (McCreary, 2013). Termín NoSQL sa prvýkrát použil v roku 1998 v prípade relačnej databázy, kde nebol využívaný jazyk SQL. V roku 2009 došlo k opätovnému oživeniu termínu, vzhľadom k tomu, že sa začal uvádzať na rôznych konferenciách obhajcov nerelačných databáz. Za znovuzrodením termínu stál Eric Evans, ktorý ho použil na konferencii v San Franciscu (Litho, Mattsson, 2010).

NoSQL je súbor konceptov, ktoré umožňujú rýchle a efektívne spracovanie dát so zameraním na výkon, spoľahlivosť a agilnosť (McCreary, Kelly, 2013). Z toho dôvodu skratku NoSQL vysvetľujeme skôr ako Not-onl-SQL, nie ako No-SQL (Strauch, 2011) aj keď spočiatku sa NoSQL chápalo skutočne ako No-SQL (teda striktné vyhradenie sa voči SQL).

NoSQL nie je názvom konkrétneho produktu, ale označuje nový databázový koncept, nový trend, ktorým sa mnoho databázových systémov vydalo. Samotné NoSQL databázové systémy nevyklúčujú využívanie jazyka SQL. Prínos NoSQL databáz spočíva v tom, že vnášajú nové myšlienkové pochody a riešenia do databázových systémov.

Známe sú aj také databázy, ktoré spĺňajú už uvedené a napriek tomu sa nezaradujú medzi NoSQL databázy. Najvýznamnejšou ideou pri takýchto databázach je predovšetkým vyriešenie problémov týkajúcich sa aplikácií, ktoré sú novšieho a modernejšieho charakteru, a ktoré disponujú väčšími schopnosťami ako tradičné a staršie databázové systémy (Klobása, 2012).

Kellyho definícia nevyklúča ani klasické relačné či SQL databázové systémy. Preto autori pokračujú výpočtom vlastností, ktoré NoSQL databázy spájajú. Tými hlavnými sú rozdielne formáty uloženia dát, nepoužívajú iba uloženie riadkov v tabuľkách, väčšina z nich nepodporuje JOIN, umožňujú ukladať dáta bez nutnosti definovať schému, fungujú na mnohých procesoroch, sú ľahko horizontálne škálovateľné, využívajú architektúru shared-nothing, môžu fungovať na mnohých lacných uzloch, bežné počítače prepojené do veľkého výkonného celku a v neposlednom rade sú inovatívne (McCreary; Kelly, 2013).

V súčasnom svete existujú desiatky rôznych NoSQL databázových systémov, medzi najpoužívanejšie patrí napríklad MongoDB, Redis a Memcached, ElasticSearch, alebo Apache Cassandra. Naším reprezentantom NoSQL databáz bude práve MongoDB, ktoré posluží ako stavebný pilier pre riešenie navrhnuté v tejto práci.

Základné vlastnosti, ktoré sú totožné pre väčšinu databáz NoSQL:

- nevyužívajú relačný model (záznamy neobmedzuje presná štruktúra),
- sú dostupné open-source,
- boli navrhnuté tak, aby vyhovovali podmienkam pre prácu v klastri,
- sú horizontálne škálovateľné.

Okrem toho mnohé ponúkajú nekomplikovanú replikáciu, nenáročnú programovaciu rozhranie a sú schopné spracovať i väčšie objemy dát (Fowler, 2012). História pojmu NoSQL databáza nesiahá ďaleko. Ako názov hovorí, pojem NoSQL vznikol až po vytvorení štandardizovaného jazyka SQL. Prvá spoločnosť, ktorá musela riešiť problém s veľkým objemom štruktúrovaných dát, bola Google. Medzi dátami, ktoré spoločnosť potrebovala ukladať, prevažovali obsahy stránok, prepojenie medzi stránkami a textové odkazy (Kopal, 2015). Spoločnosť Google sa rozhodla vyriešiť tento vzniknutý problém za pomoci projektu pod názvom BigTable. Ten odštartoval väčší rozmach databáz s neskorším pomenovaním NoSQL databázy. Vo väčšine prípadov je základom týchto databáz princíp kľúč-hodnota. BigTable spôsobil výrazný zvrät v rámci databázového sveta, keďže sa nejedná o databázu, ktorá by bola riadkovo zameraná, ako je tomu v prípade tradičných databáz. Využíva sa databáza stĺpcovo orientovaná (Kopal, 2015). K podobnej situácii došlo aj v prípade spoločnosti Amazon. Najväčší internetový predajca sa inšpiroval týmto projektom a rovnako našiel riešenie v ukladaní dát za použitia princípu

klúč-hodnota. Riešenie už konkrétneho charakteru sa využilo v distribuovanej databáze, ktorá vystupuje pod názvom Dynamo. V nej sú ukladané dáta o predaných produktoch. Obidva spomínané projekty, BigTable a Dynamo, sú inšpirujúce pre vytvorenie a vývoj mnohých ďalších NoSQL databáz.

Časovo najstaršia je databáza Oracle Berkeley DB, ktorá sa na rozdiel od väčšiny NoSQL databáz neinšpirovala projektom BigTable. Jej vznik sa odhaduje na dobu pred 16 rokmi, pričom sa jej škála vlastností postupne rozširovala. Dnes sa delí do troch jedinečných verzií: Berkeley DB, Berkeley DB JavaEdition, Berkeley DB XML. Bola vytvorená na univerzite v Berkeley. Uvedenú črtu NoSQL databázy týkajúcu sa používania rozličných dotazovacích jazykov alebo rozhraní (nielen SQL) možno nájsť aj u Oracle Berkeley DB, ktorá poskytuje možnosť ukladať a čítať dáta za pomoci:

- SQL syntaxe,
- XQuery,
- Java objektov,
- párov klúč/hodnota.

Na základe svojich vlastností možno NoSQL využiť predovšetkým pre zálohovanie a ukladanie logovacích záznamov. Jeho využitie je adekvátne aj v tých prípadoch, kde sa požaduje realizácia vkladania a dopytovania v skutočnom čase, ak je konzistencia dát menšou prioritou. Mnoho NoSQL databáz sa využíva v takých podmienkach, kde sa vyžaduje škálovateľnosť. Ako príklad môžeme uviesť sociálnu sieť Foursquare, pracujúcu s NoSQL databázou MongoDB, slúžiacou na ukladanie dát o zaujímavých lokalitách (dáta relatívne statického charakteru), ale aj o aktivitách používateľov (Prمود, 2013). NoSQL disponuje výraznými medzerami napríklad v takých situáciách, kde je nevyhnutná transakčnosť, napríklad banky a v situáciách, kde je oveľa jednoduchšie využiť statické analytické prostriedky s tým, že sa použije SQL.

MongoDB sa považuje za open-source databázové riešenie, ktoré sa zaraďuje do skupiny NoSQL. Za výhodu MongoDB sa považuje to, že návrh je jednoduchý a škálovanie horizontálne (MongoDB, 2015). Za hlavnú entitu v MongoDB sa považuje dokument, ktorý je možné prezentovať ako slovníkovú štruktúru dát na princípe klúč-hodnota. Hodnotou môže byť iný slovník, pole, či niekoľko hlavných typov ako je číselná hodnota, eventuálne reťazec znakov. Dokumenty sú zhrnuté do jednotlivých kolekcí. Je nevyhnutné podotknúť, že kolekcie nedisponujú predpísanou schémou. Inak povedané, dokumenty v jednej kolekcii môžu disponovať rôznymi štruktúrami. Viacero kolekcí spolu vytvára databázu. Dokument sa považuje za analógiu k riadku v SQL databáze, kolekcia je tabuľkou v SQL databáze. Dotazovacieho jazyka implementovaný v MongoDB napomáha k vyhľadávaniu podľa hodnoty klúča, rozsahu hodnôt, či za pomoci regulárnych výrazov a vlastných JavaScript funkcií.

Zdroje

Použitá, citovaná a odporúčaná interná „literatúra“:

- Michal Székely, *Využitie NoSQL databáz v súčasných webových aplikáciách*, Nitra 2016.