



RĪGAS TEHNISKĀ UNIVERSITĀTE

NoSQL

NoSQL bāzētu datubāzu vadības sistēmu apraksts

Autors: Reinis Veips

Studenta apl. nr.: 081RDB032

Grupa: RDGDB3

Rīga, 2013

NoSQL

- *Not only SQL / No SQL*
- Specializētas datubāzes, neizmanto SQL, specifiski interfeisi (*API*)
- Nav fiksētas datu shēmas, datu struktūra var būt mainīga
- Lielākajā daļā realizāciju neeksistē JOIN konstrukcija
- Horizontālās mērogošanas atbalsts (*replication, sharding*)

Kategorizācija/taksonomija

- Dokumentu krātuve (*document store*).
- Atslēga-vērtība krātuve (*key-value store*).
- Grafu datubāze.
- Tabulāra krātuve (*tabular store*).
- Objektu datubāze.
- XML datubāze.

Dokumentu krātuve

- Galvenais koncepts- dokuments
- Dokuments iekapsulē datus (formāti- XML, YAML, JSON, PDF..)
- Grupēšana- pēc birkām, metadatiem
- Relāciju datubāzēs, dokuments = rinda
- Meklēšana dokumentos- atkarīga no realizācijas.

Kolekcija vs tabula

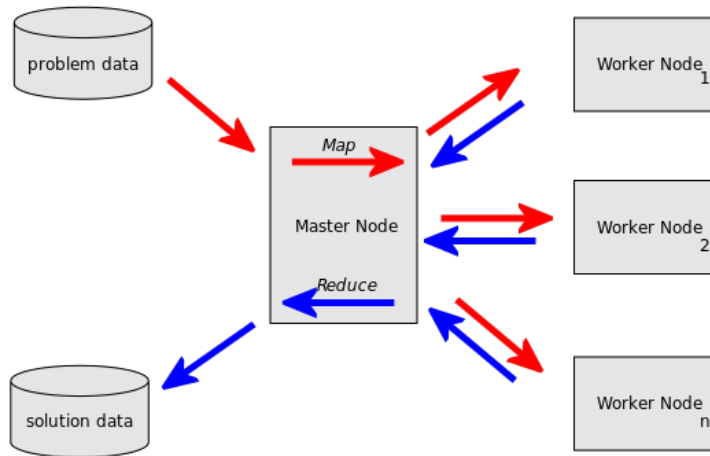
Last Name	First Name	Date of Birth
DUMONT	Jean	01-22-1963
PELLERIN	Franck	09-19-1983
GANNON	Dustin	11-12-1982

```
{
  "_id": ObjectId("4efa8d2b7d284dad101e4bc9"),
  "Last Name": "DUMONT",
  "First Name": "Jean",
  "Date of Birth": "01-22-1963"
},
{
  "_id": ObjectId("4efa8d2b7d284dad101e4bc7"),
  "Last Name": "PELLERIN",
  "First Name": "Franck",
  "Date of Birth": "09-19-1983",
  "Address": "1 chemin des Loges",
  "City": "VERSAILLES"
}
```

Map-Reduce

- Programmēšanas paņēmiens dokumentu apstrādei paralēli
- Datus apstrādā vairāki datori/serveri vienlaicīgi
- Reducē datus un atgriež rezultātus
- Reducēšanas operācijas var būt: filtrēšana, agregācija, skaitļošana

Map-Reduce 2



Pielietojumi

- Izmantojams, ja datu modelis "ietilpst" dokumentu formā (lietotājs, ieraksts)
- Ir jēga tikai gadījumā, ja datu ir ļoti daudz (vairāki simti GB)
- Realizācijas pārsvarā ir optimizētas biežai lasīšanai un pievienošanai, reti dzēšanai un modificēšanai
- Ja nepieciešama paralēla dokumentu apstrāde

Realizācijas

- **Apache CouchDB**

<http://couchdb.apache.org/>

- **MongoDB**

<http://mongodb.org/>

- **Oracle NoSQL Database**

<http://www.oracle.com/technetwork/products/nosqldb/overview>

<https://github.com/festlv/rtu-nosql>

Atslēga-vērtība krātuve (*key-value store*)

- Atslēga-vērtība koncepts (atslēga, vērtība) kortežs
- Relāciju datubāzēs- tabula ar divām kolonnām
- Parasti atslēga- primitīvs datu tips, savukārt vērtība var būt salikts datu tips
- Realizācijas parasti vērtību atgriež jau programmēšanas valodas objektos- kā masīvu, teksta virkni, int u.c.

Piemērs

```
1 >>> import redis
2 >>> r = redis.Redis()
3 >>> r.sadd('atslega', (1,2,3))
4 1
5 >>> r.smembers('atslega')
6 set(['(1, 2, 3)'])
7
8 >>> r.set('vienkarsaatslega', 'abc')
9 True
10 >>> r.get('vienkarsaatslega')
11 'abc'
```

Pielietojumi

- Relatīvi īsu, bet daudzskaitlīgu, kopumā, apjomīgu datu glabāšanai
- It īpaši- izplatītās sistēmās
- Iegultajās sistēmās (HTML5 LocalStorage)
- Dati bieži tiek glabāti operatīvajā atmiņā, un tikai periodiski saglabāti cietajā diskā (ātrdarbība > konsistence)

Realizācijas

- **Redis**

<http://redis.io/>

- **Memcached**

<http://memcached.org/>

- **Apache Cassandra**

<http://cassandra.apache.org/>

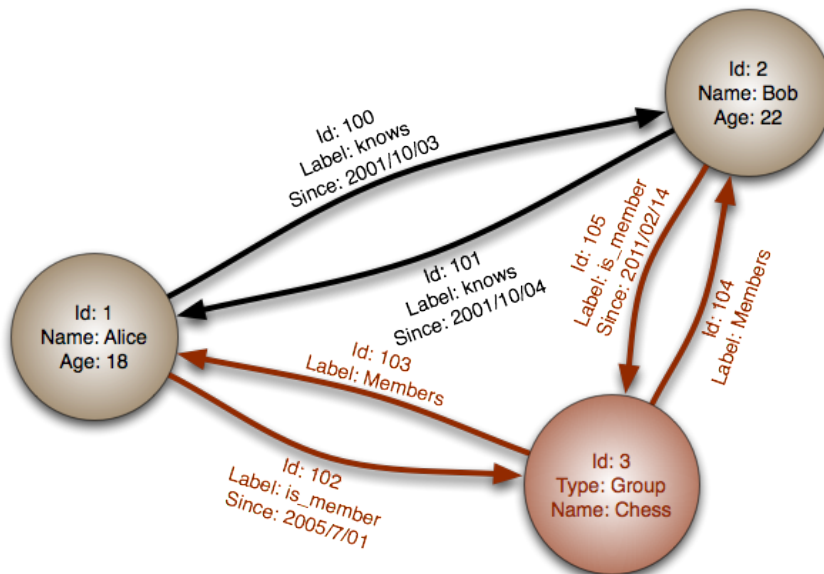
Grafu datubāzes

- Ideāli piemērotas, ja datu kopa ir aprakstāma kā grafs
- Iebūvēti grafu algoritmi- īsākā ceļa meklēšana u.c.
- Gan virsotnēm, gan šķautnēm var būt piesaistīti dati (jēdziena "svars" paplašinājums)

Pielietojumi:

- Draudzības saites starp cilvēkiem sociālajos tīklos
- Sabiedriskā transporta sistēmas aprakstīšana (virsotnes=pieturas, šķautnes apzīmē pārvietošanos starp pieturām, īpašības uz šķautnēm- pārvietošanās laiks)

Grafu datubāzes- vizualizācija



Realizācijas

- **Neo4j**
<http://neo4j.org/>
- **Titan**
<http://thinkaurelius.github.com/titan/>
- **VertexDB**
<http://www.dekorte.com/projects/opensource/vertexdb/>

Secinājumi

- Nav labu vai sliktu darbarīku. Ir piemēroti un nepiemēroti darbarīki.
- Relāciju datubāzes ir neaizvietošanas datiem, kuriem obligāti jānodrošina ACID īpašības.
- NoSQL datubāzes dažus ACID burtiņus aizvieto ar veikspēju.
- Potenciāla uzturēšanas izmaksu samazināšana (viena jaudīga un dārga servera vietā izmanto vairākus lētus)