

Tehnologije VR

Delo s podatki
(Baza časovnih vrst)

Podatki

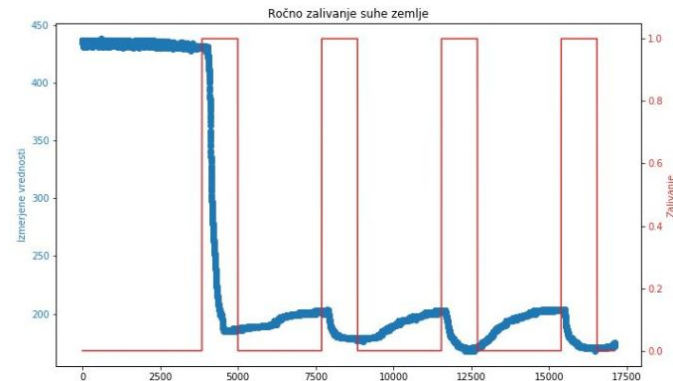
- Kako bi odgovorili na spodnja vprašanja?
 - Kakšni podatki?
 - Strukture podatkov?
 - Količina podatkov?
 - Poizvedovajna?
 - Analiza podatkov?
- Kako bi implementirali podporo tem podatkom?
 - Katere tehnologije?
 - Infrastrukturo?

Baze časovnih vrst (Time series database - TSdb)

- Kaj je časovni podatek?
 - Podatek, ki je opremljen s časovno značko. Npr. milisekunde, nanosekunde, ...
- Kaj je časovna vrsta?
 - Je vrsta v kateri so časovni podatki, ki opisujejo isto lastnost ob različnem času.
- Kaj je baza časovnih vrst?
 - Je podatkovna baza, ki je optimizirana za delo s časovnimi vrstami.

Primer podatkov

- Področja
 - Vremenski podatki
 - Borza (real time analitika)
 - Spremljanje delovanja avtomobila
 - Spremljanje strežnikov, prog. opreme, ...
 - Obremenitve, ...
- Senzorji
 - Pospeškometer
 - GPS, ...
 - Gibanje človeka
- Frekvenca
 - Od 1 krat na dan, mesec, ...
 - Do kilo, mega Hz (1000000 Hz)



Lastnosti časovnih podatkov

- Sam podatek ne pomeni veliko
 - Lahko hitro zastari
 - Natančnost je lahko vprašljiva (npr. na borzi se zgodi trenutno velik nakup/prodaja, motnja,...)
- Vrednost/koristnost podatkov v časovni vrsti raste
 - Trendi
 - Ciklično ponavljanje
 - Obdelava signalov
 - S časom lahko vrednost/pomembnost/natančnost podatkov pade
 - Temperatura na minuto natančno pred 100 leti ;) ,
- Število podatkov narašča veliko hitreje
 - Npr. 1 x na sekundo 8 bajtov (*64 bitov*) = ? v 1 letu
 - Časovna značka med 4 do 32 bajtov (odvisno od predstavitve, natančnosti, obdobja).
 - 85% podatkov je nastalo v zadnjih 2 letih (primer izračuna [WWW](#))

Vrsta časovnih podatkov

- Časovno ekvidistančno pravilni podatki (regular data series oz. evenly spaced time series)
 - Čas med dvema podatkom v vrsti je konstanten.
 - Senzor deluje sinhrono in bere v konstantnih zamikih.
- Časovno ekvidistančno nepravilni podatki (irregular data series oz. unevenly spaced time series)
 - Čas med dvema podatkom je različen.
 - Primer odpiranje in zapiranje vrat.
 - Senzor deluje po principu dogodkov.
 - Asinhrono delovanje (mreža, čakalna vrsta procesa).
- Mešano
 - Manjkajoči podatki ali generiranje ekvidistančnih podatkov
 - Tehnike preprocesiranja
 - Proces časovnega gladenja da dobimo časovno ekvidistančne podatke

Katera predstavitev je boljša (prednosti, slabosti)?

Brez časovne značke?

Ekvidistančno pravilna?

Ekvidistančno nepravilna?

Primerjava različnih baz podatkov

- Dokumentno usmirjene baze
 - Popravljaš, združuješ, dodajaš dokumente.
 - Dodajanje posamezne vrednosti ni učinkovito (nov dokument).
 - MongoDB, CouchDB, NoSQL kategorija,...[WWW](#)
- Iskalne baze (search database)
 - Hierarchical databases, Network databases, Graph databases
 - Primerne za iskanje po velikih bazah, po vseh podatkih.
 - Bolj povezujejo relacije med podatki kot pa časovno komponento.
- Relacijske baze (MySQL...)
 - Delujejo po principu tabel, ki so sestavljene iz stolpcev in vrstic.
 - Relacije med stolpci in tabelami.
 - Povpraševanje po stolpcih..
 - Neprimerno za velike količine podatkov, ki jih hitro zavržemo.

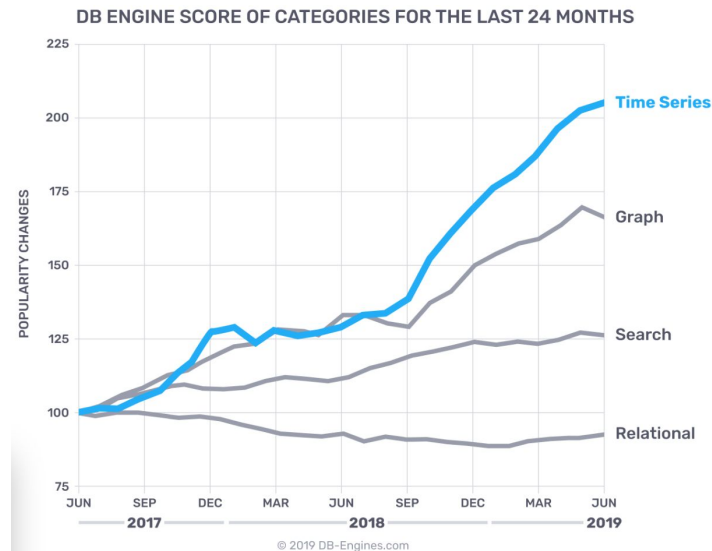
Prednosti TSdb

- Primernost za veliko zapisov (več sto milijonov zapisov)
 - Učinkovit format
 - Narejeno v mislih za skalabilnost
- Učinkovito povzemanje podatkov (summarization of data)
 - Pridobi podatke v določenih časovnih obdobje z natančnostjo na 15minut (time tick)
- Politika vzdrževanja podatkov (RETENTION POLICY)
 - Odbrži vse
 - Po X urah hrani natančnost na Y minut
 - Po X dnevih hrani natančnost na Y ur
 - ...



Primeri baz

- Trend popularnosti [WWW](#)
- Izbrane [WWW](#)
 - InfluxDB
 - Najpopularnejša baza
 - Zastoj za single computer
 - TimescaleDB
 - Dodatna plast nad bazo PostgreSQL
 - OpenTSDB
 - Graphite
 - Narejena nad bazo Apache HBase
 - Graphite
 - Orodje za nadzor podatkov
 - Vsebuje TSdb



Rank			DBMS	Database Model	Score		
Jun 2019	May 2019	Jun 2018			Jun 2019	May 2019	Jun 2018
1.	1.	1.	InfluxDB	Time Series	17.98	-0.09	+6.65
2.	2.	2.	Kdb+	Time Series, Multi-model	5.80	+0.21	+2.79
3.	3.	4.	Graphite	Time Series	3.33	+0.10	+0.95
4.	4.	6.	Prometheus	Time Series	3.32	+0.21	+2.06
5.	5.	3.	RRDtool	Time Series	2.67	-0.23	0.00
6.	6.	5.	OpenTSDB	Time Series	2.24	-0.23	+0.68
7.	7.	7.	Druid	Multi-model	1.78	+0.09	+0.65
8.	8.	17.	TimescaleDB	Time Series, Multi-model	1.11	-0.05	+1.06
9.	9.	8.	KairosDB	Time Series	0.50	-0.04	+0.09
10.	10.	9.	eXtremeDB	Multi-model	0.41	+0.03	+0.13
11.	13.	20.	Heroic	Time Series	0.40	+0.06	+0.40
12.	12.	15.	GridDB	Multi-model	0.36	+0.02	+0.28
13.	11.	12.	FaunaDB	Multi-model	0.36	-0.03	+0.24

Influx DB (primer)

- Legende, metadata tags (oznake) so indeksirane.
- Seznam oznak za posamezno meritev imenujemo tag set.
- Vrednosti polja (fields) je Y os
 - int, floats, strings,
 - Na sliki trenutno samo eno polje stock price! lahko bi bil še volume, ...HiPrice, lowPrice
 - Vsa polja field set.
- X Os
 - Timestamp (lahko nanoseconds)
- Predstavitve podatkov lineProtokol [www](http://www.influxdb.org/docs/quickstart/#line-protocol)
 - meritev, množica značk " " množica polj " " značka časovna
 - Vrsta podatkov so vse točke v meritvi ki imajo skupne značkr (tag set)
- Single point (časovna značka in tagi)



Primer

```
airSensors,sensor_id=TLM0100 temperature=71.21393552024017,humidity=35.11986335555151,co=0.48876906097308576 17168763970000000000
airSensors,sensor_id=TLM0101 temperature=71.81023072166334,humidity=34.917392327544555,co=0.5188692080407862 17168763970000000000
airSensors,sensor_id=TLM0102 temperature=72.02379208500484,humidity=34.923310510163766,co=0.5129489294426316 17168763970000000000
airSensors,sensor_id=TLM0103 temperature=71.27729749861177,humidity=35.15148465300235,co=0.4080512061483494 17168763970000000000
airSensors,sensor_id=TLM0200 temperature=73.55049538070678,humidity=35.841357175269216,co=0.525280327831053 17168763970000000000
```

Influx DB

Cmd example [WWW](#)

Primer [WWW](#)

Primer 2 [WWW](#)