

Denormalizacija ir kitos SQL optimizacijos

Aurelijus Banelis

Apie mane

Aurelijus Banelis

aurelijus@banelis.lt
aurelijus.banelis.lt

PHP programuotojas
Dirbu NFQ

Jau 2 pranešimas VilniusPHP



Pristatymo eiga

Alternatyvos

(DBVS, Sphinx, NoSql)

Kas

yra denormalizacija

Praktikoje

(el. parduotuvės pavyzdys)

Kodėl veikia

(žemo lygmens realizacija)

Denormalizacija ~ spartinimas

Normalizavimas

~

D_{ont} **R**_{epeat} **Y**_{ourself}

duomenų bazėms

Formalios taisyklės:

1NF laukelyje neskaidomi duomenys

2NF laukelis nėra išskaičiuojamas

3NF nėra tranzityvių priklausomybių

...

DeNormalizavimas

~

rankinis spartinimas
(angl. catching)

duomenų bazėms

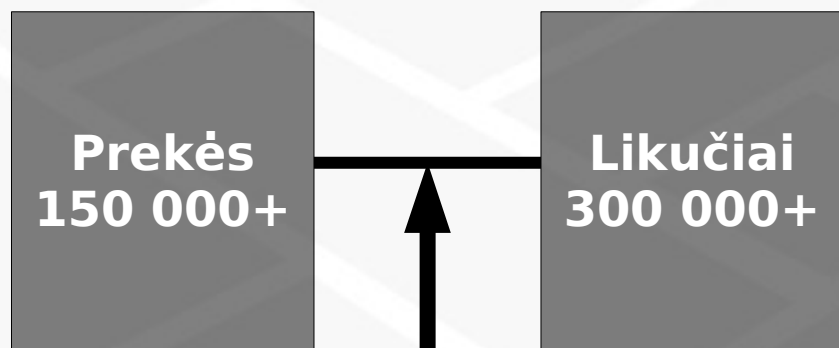
**Tikslingai pažeidžiamos
norminės formos, kad**

**būtų greitesnis
duomenų skaitymas**

„Uždėk indeksus“ - ar pakanka?

EXPLAIN SELECT ...

| ... | |



Indeksai yra greiti

**EXPLAIN SELECT ...
ORDER BY ...**

.. |... Using filesort |
ID+SortField netilpo į sort_buffer_size



Jei trumpi, jų mažai ir iš pirmos lentelės

Praktikoje: sprendimas (1/3)

CREATE TABLE cache123 ...

**INSERT INTO cache123 SELECT ... JOIN ...
WHERE ... GROUP BY ... ORDER BY ...**

**Papildomos mažesnės lentelės,
Kai lieka tik laukai, su kuriais dirbama.**

Pagal ilgiausias užklausas

Pagal verslo logiką

Ir jokių indeksų

Gamintojas

Produktai

Kategorija

Likutis

Praktikoje: sprendimas (2/3)

CREATE TABLE bigcache123 ...

**INSERT INTO bigcache123 SELECT ... JOIN ...
ORDER BY ...**

**Papildomos jungtinės lentelės,
Filtruojama pagal 1 lentelę,
Filtruojama pagal indeksą!**

Prekės
150 000+

Analogai
3 000 000+

Modeliai
500 000+



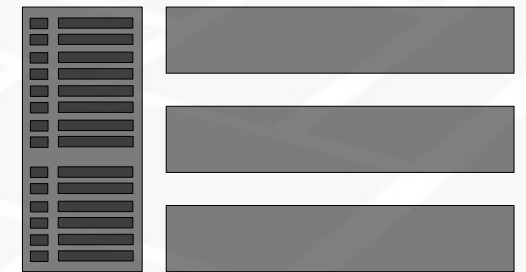
Modeliai pagal prekes
70 000 000+

Praktikoje: sprendimas (3/3)

ALTER TABLE normal ADD COLUMN comp ...

UPDATE TABLE normal ...

SET comp = ...



Papildomas išskaičiuojamas laukas.

Kažkoks ilgas atributas=ilga atributo reikšmė

5F98BDE6F446E135211A799521AEE4B0

0 250 202 022 0.250.202.022 0-250-202-022 0/250/202/022

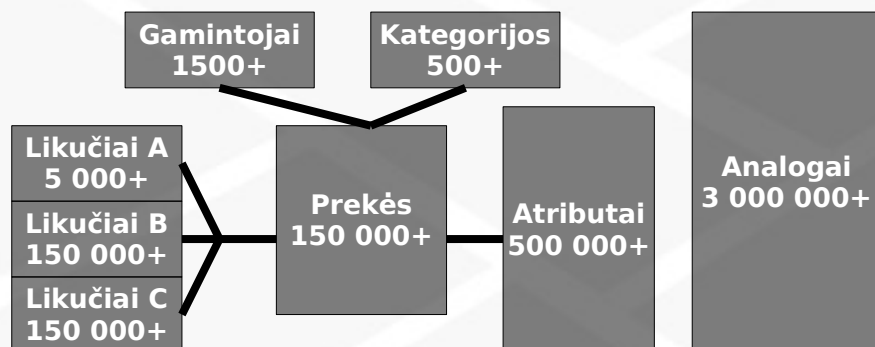
0250202022

BET: Prekės kaina asortimente ir prekės kaina užsakyme? (Ne)Kopija?

Apibendrinus sprendimus

- **Tik WHERE** → **Didelė kopija**
- **GROUP, ORDER** → **Maža kopija**
- **Skaičiavimai()** → **Papildomas laukas**

Sudėtinga/daug → Duomenų kopijos

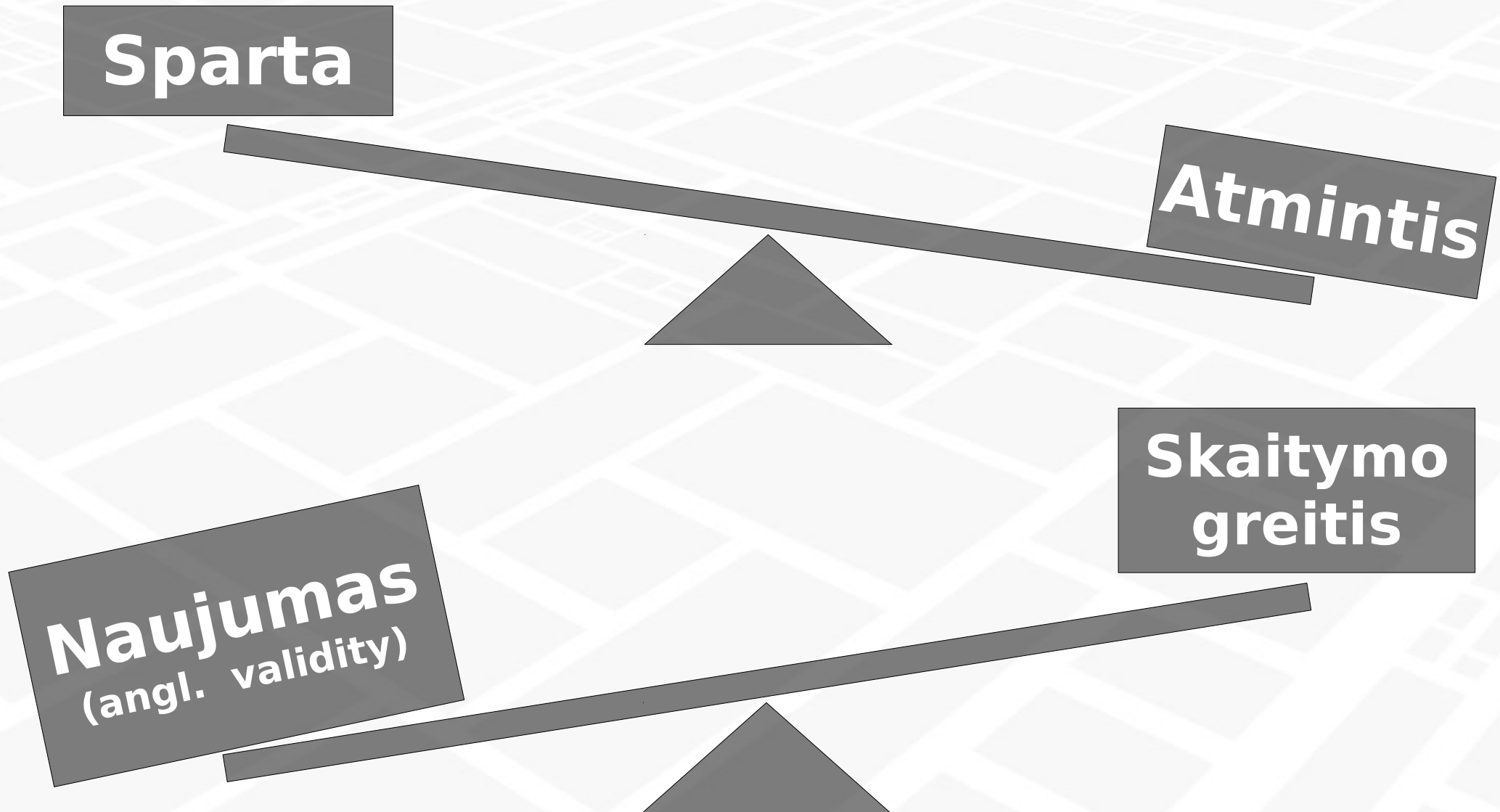


Modeliai pagal prekes
70 000 000+

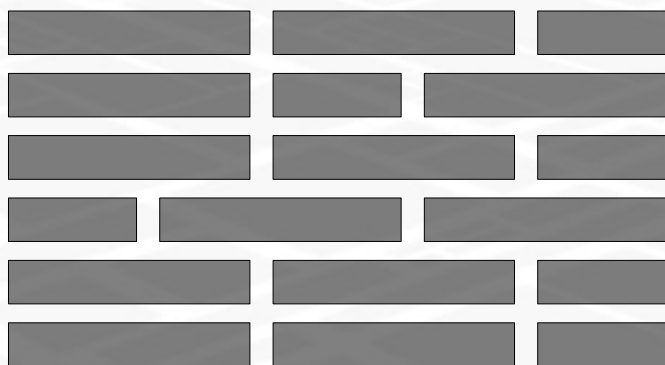
Gamintojas
Produktai
Kategorija
Likutis



Praktikoje: minusai

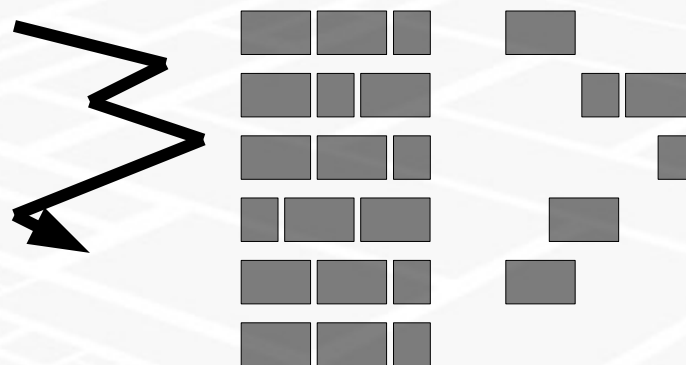


Realizacija žemame lygmenyje



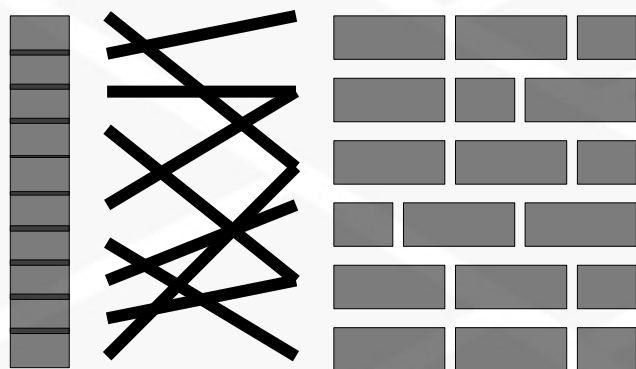
Duomenys

„123“ vs 123.0



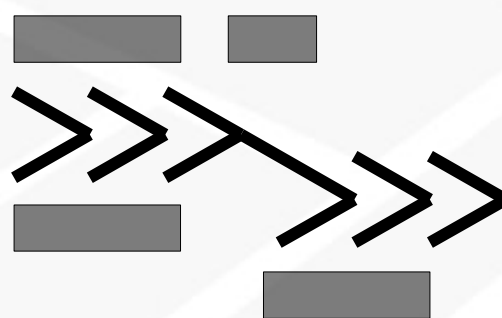
Filtravimas

WHERE vs JOIN



Indeksas

KEY(A,B) vs KEY(A),KEY(B)



Rūšiavimas

BTREE vs HASH, WHERE

**Vienodi
dydžiai,
tipai**

**Šokinėti/
skaityti**

**Sudėtinis/
skirtingi
indeksai**

**Surikiuoti,
prafiltruoti
duomenys**

Alternatyvos

Automatizavimas iš DBVS pusės:

- **MsSql: *indexed views***
- **Oracle: *materialised views***

Rezultatai užima tik mažą dalį, gauti ID:

- **Sphinx - *inverse index***

NoSql/distribution:

- ***Eventual Consistency***
- ***Consistent Hash Rings***
- ***Map reduce***

Tai va. Klausimai?

Alternatyvos

(DBVS, Sphinx, NoSql)

Kas

yra denormalizacija

Praktikoje

(el. parduotuvės pavyzdys)

Kodėl veikia

(žemo lygmens realizacija)

Verta pasiskaityti

- <http://dev.mysql.com/doc/refman/5.0/en/select-optimization.html>
- <http://dev.mysql.com/doc/refman/5.0/en/execution-plan-information.html>
- <http://www.aosabook.org/en/nosql.html>