

<epam>



**Termékkereső** Solr  **alapokon**

# Salló László

---



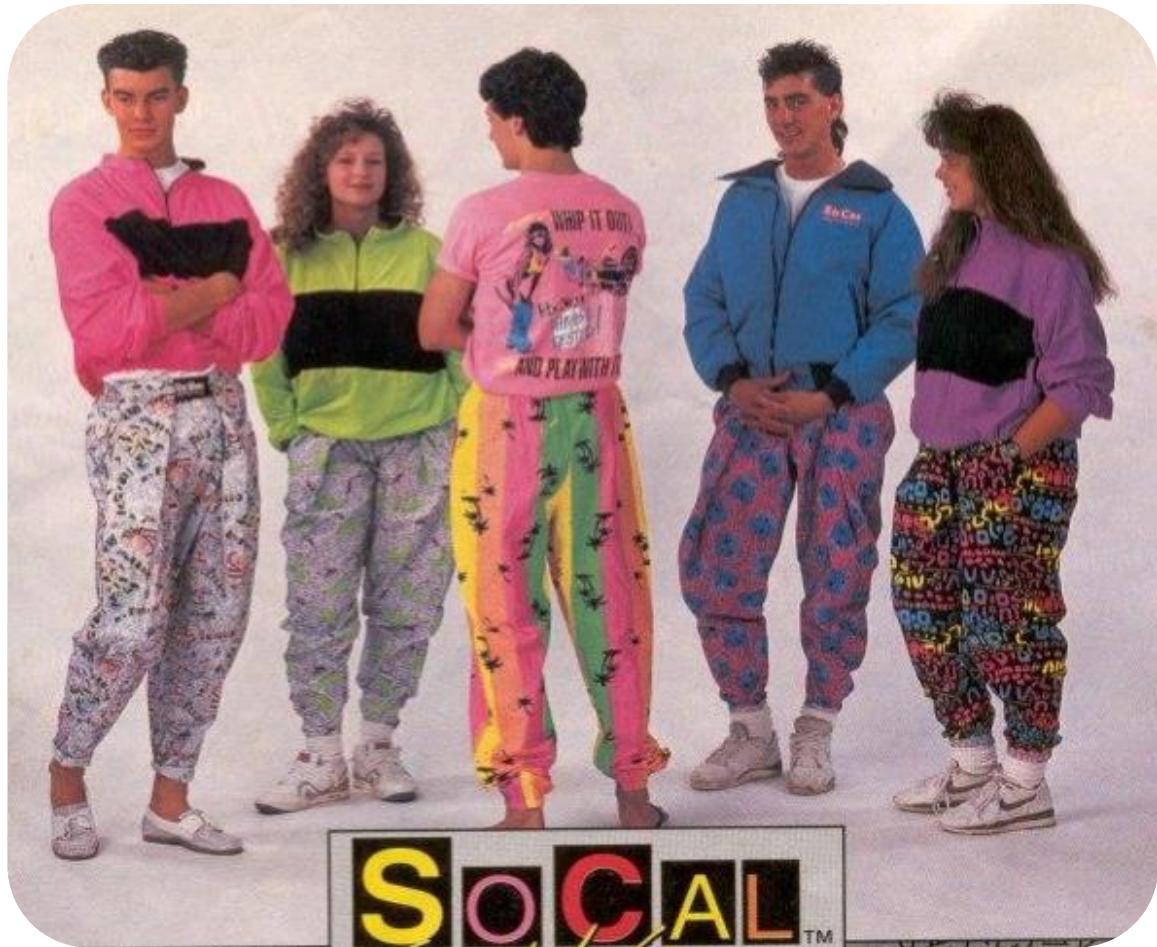
<https://hu.linkedin.com/in/laszlosallo>



<https://www.kaggle.com/salacika>



<http://business.ceu.edu/msc-in-business-analytics>



**S** **O** **C** **A** **L** <sup>TM</sup>

**S** **O** **C** **A** **L** <sup>TM</sup>





Megkeressük Önnek az álom vakációt ...

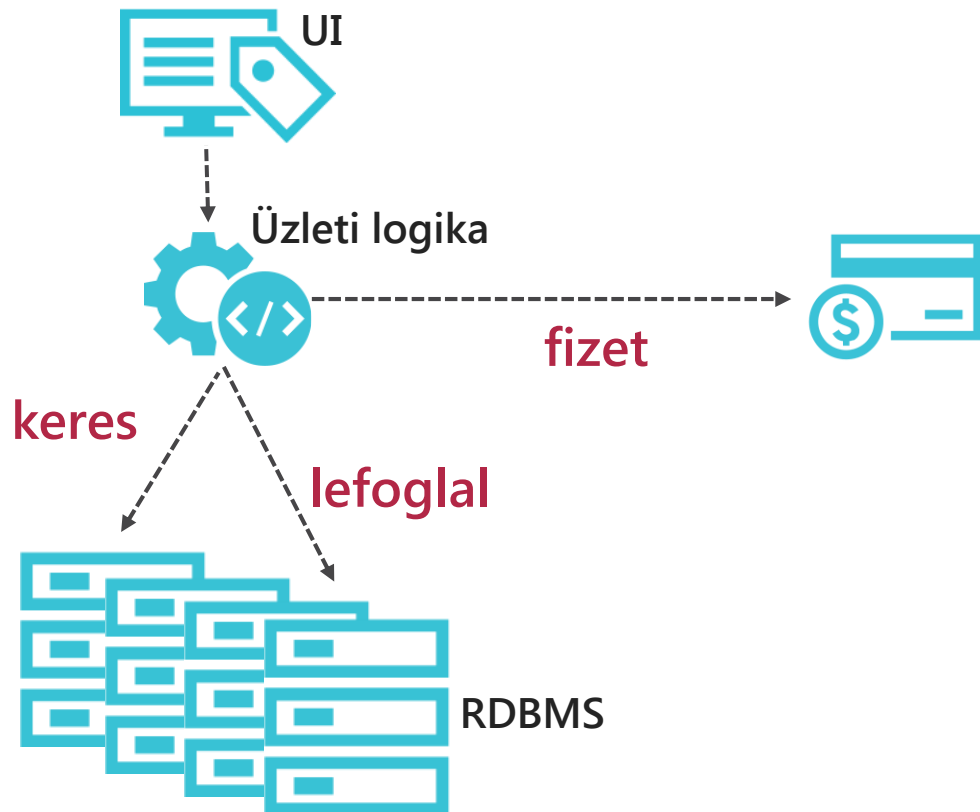
62%



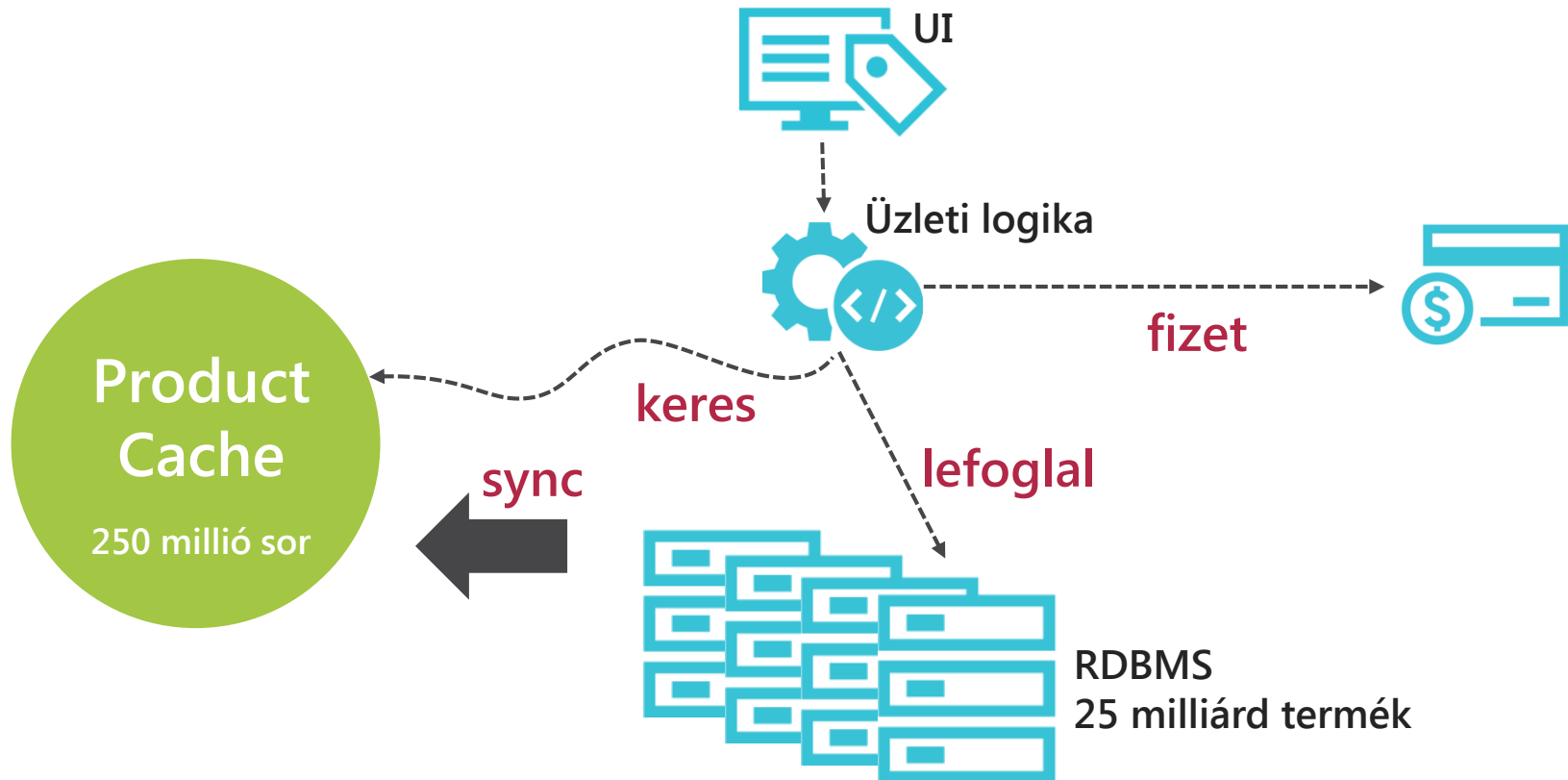
... ez eltarthat néhány percig.



# E-commerce



# E-commerce



Megoldás ?

---

ORACLE®

atg<sup>®</sup>

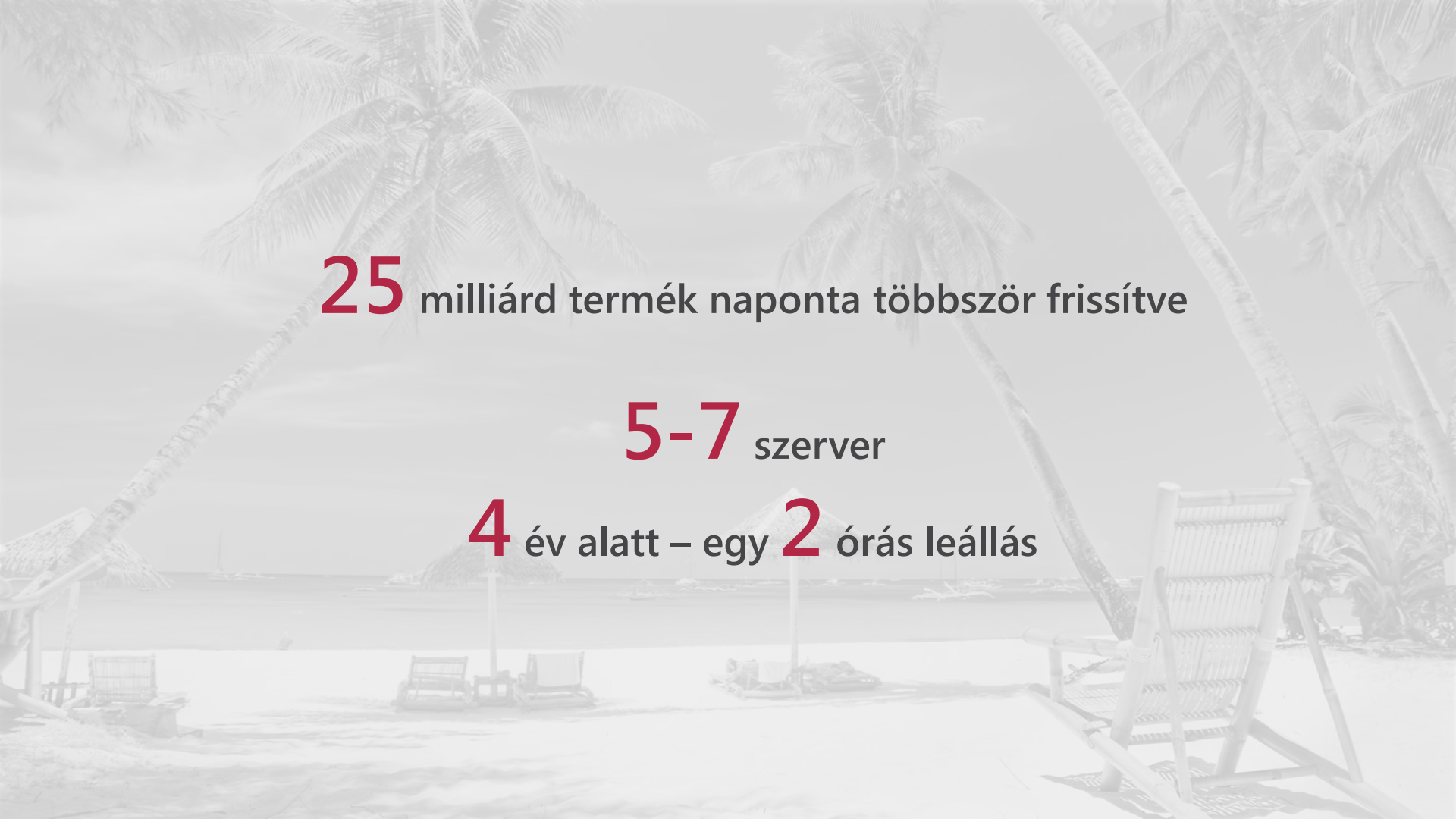
SDL | Fredhopper



A tropical beach scene with palm trees, lounge chairs, and a thatched hut, serving as a background for the text.

**Open Source +**

**NoSQL**

A tropical beach scene with palm trees, a thatched hut, and a lounge chair. The background is a soft-focus image of a tropical beach with several palm trees, a thatched hut on the left, and a lounge chair on the right. The sky is overcast.

**25** milliárd termék naponta többször frissítve

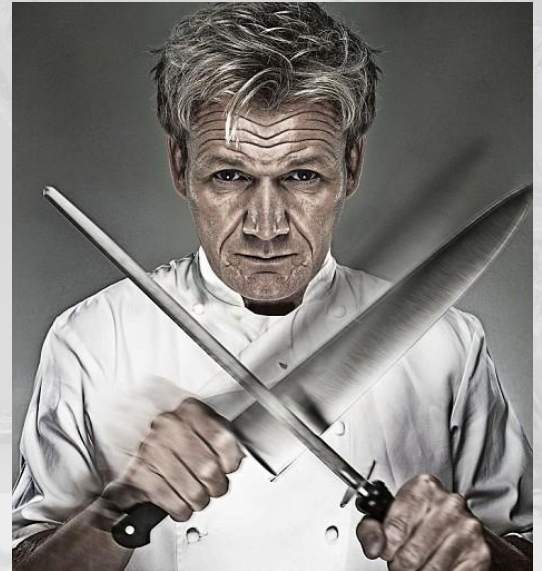
**5-7** szerver

**4** év alatt – egy **2** órás leállítás



**RDBMS**

**NOSQL**



Mi kell a  
termék  
kereséshez ?







Document Store ?

Search engine !

Big Data ?

Wide column store ?

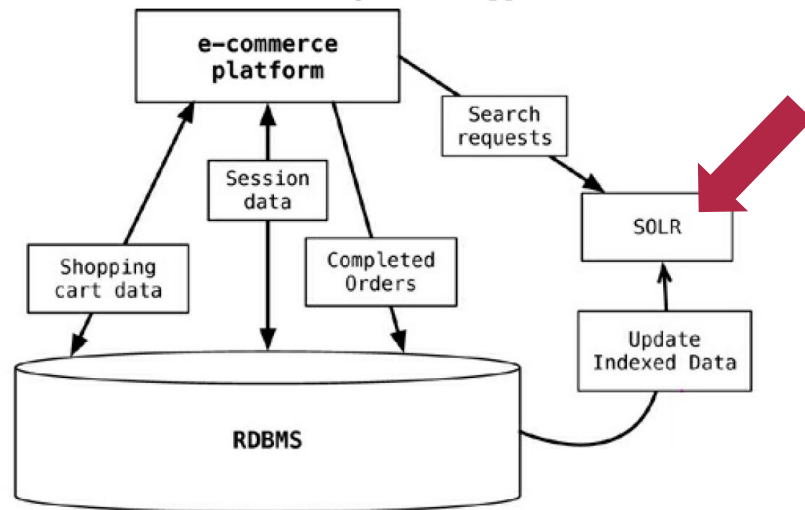
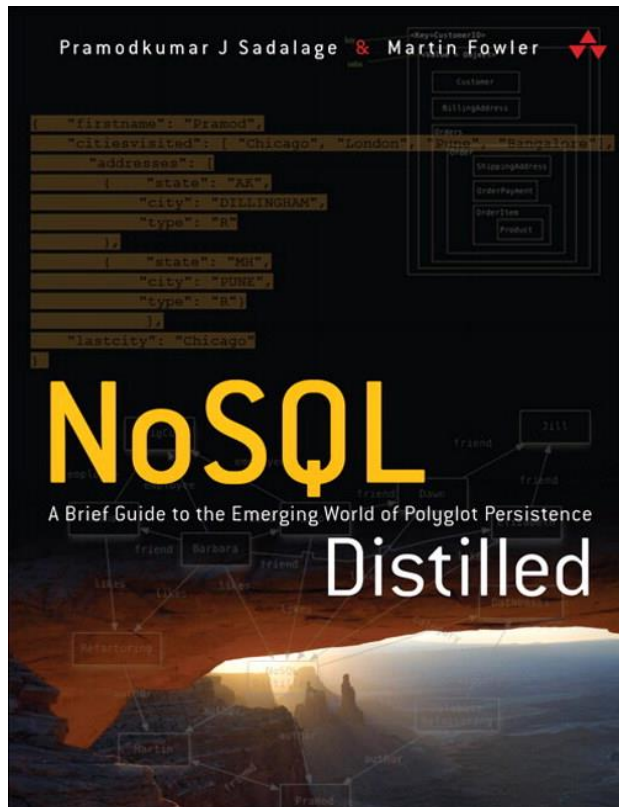
Key value store ?

# Megfelelő technológia ?



	Search Engine	Score
1.	Elasticsearch_	80.17
2.	Solr	69.37
3.	Splunk	43.73
4.	MarkLogic	9.36
5.	Sphinx	8.72
6.	Google Search Appliance	3.78
7.	Amazon CloudSearch	2.39
8.	Microsoft Azure Search	1.28
9.	Xapian	0.50
10.	Indica	0.33

# Martin Fowler



# Keresés



Java API / Rest API (JSON/XML)



Facet keresés



Text keresés



Geo spacial keresés

**Categories**

Cell Phones & Accessories

Cell Phones & Smartphones

Smart Watches

Smart Watch Accessories

Cell Phone Accessories

Display Phones

Phone Cards & SIM Cards

More ▼

**Features**

see all

- 3G Data Capable (31,735)
- 4G Data Capable (20,628)
- Bluetooth Enabled (34,541)
- GPS (30,787)
- Internet Browser (27,341)
- Music Player (30,791)
- Touchscreen (28,508)
- Wi-Fi Capable (28,179)

**Brand**

see all

- Apple (30,278)
- BlackBerry (4,872)
- HTC (8,151)

Sort: **Best Match** ▼

View:  ▼

Cell Phones & Accessories > Cell Phones & Smartphones

**Cell Phones and Smartphones** 296,477 listings [+ Follow this search](#)

Shop by Color



Black



White



Gold



Gray



Factory Unlocked Apple iPhone 5S A1533 16GB 4G LTE Touch ID Mobile Smartphone

**68,264.75 HUF**

or Best Offer

From Hong Kong

**Bundled Items** [see all](#)

- Case or Bag (841)
- Extra Battery (630)
- Extra Battery Charger (331)
- Flash (389)
- Lens (3,319)
- Memory Card (926)
- Strap (Neck or Wrist) (1,266)
- Tripod (337)

**Series** [see all](#)

- Canon EOS (2,643)
- Canon EOS Rebel (426)
- Canon PowerShot (44)
- Fujifilm X Series (77)
- Nikon COOLPIX (56)
- Nikon D (2,951)
- Olympus OM-D (89)
- Sony Alpha (463)

**Color** [see all](#)

- Blue (3,014)
- Blue (40)
- Gray (60)
- Pink (14)
- Red (162)
- Silver (265)
- White (172)

**Bundled Items** [see all](#)

- Case or Bag (415)
- Extra Battery (378)
- Extra Battery Charger (205)
- Flash (150)
- Lens (1,160)
- Memory Card (390)
- Strap (Neck or Wrist) (704)
- Tripod (117)

**Series** [see all](#)

- Canon EOS (879)
- Canon EOS Rebel (221)
- Canon PowerShot (17)
- Fujifilm X Series (35)
- Nikon COOLPIX (18)
- Nikon D (1,100)
- Olympus OM-D (38)
- Sony Alpha (148)

**Color** [see all](#)

- Black

**Megapixels** [see all](#)

**Optical Zoom** [see all](#)

# A facet paradigm

## Zoom in search

Black (10,028)

Blue (9)

Brown (1)

Gold (30)

Gray (3)

Green (2)

Orange (2)

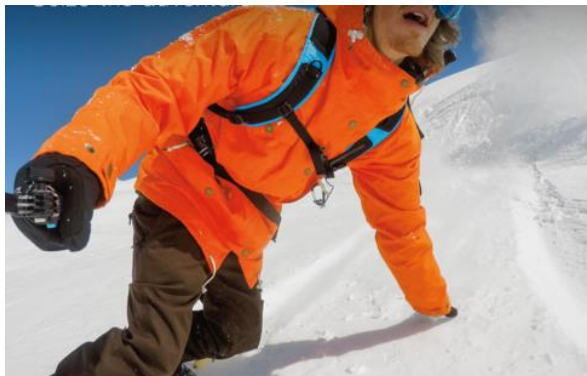
Pink (1)

Red (5)

Silver (14)

White (2)





### Shop by Category

Digital SLRs



Mirrorless Cameras



Point-and-Shoots



Lenses



Camcorders



Sports & Action Cameras



Photography Drones



Security Cameras



Optics



Accessories



## Language

- English (903,528)
- German (58,485)
- French (36,211)
- Spanish (25,259)
- Italian (18,987)
- Russian (3,854)
- Japanese (27,069)

+ See more

# Partial facets



« Previous | Page: 1 2 | Next »

Find:

English (902,475)	Comish (1)	Indonesian (156)
German (58,059)	Croatian (33)	Irish (233)
French (34,194)	Czech (155)	Japanese (27,443)
Spanish (25,212)	Danish (500)	Kannada (1)
Italian (17,908)	Dutch (1,366)	Korean (872)
Russian (3,849)	Esperanto (8)	Latin (878)
Chinese (2,699)	Estonian (7)	Latvian (16)
Afrikaans (251)	Farsi (31)	Lithuanian (13)
Albanian (15)	Finnish (218)	Macedonian (1)
Ancient Greek (28)	Galician (114)	Malay (3)
Arabic (107)	Georgian (10)	Malayalam (1)
Armenian (43)	Greek (3,843)	Maori (2)
Basque (75)	Gujarati (1)	Middle English (20)
Bengali (15)	Hebrew (96)	Norwegian (180)
Bulgarian (37)	Hindi (26)	Old English (17)
Cantonese Chinese (3)	Hungarian (170)	Persian (85)
Catalan (642)	Icelandic (55)	Polish (241)

# Testre szabott keresés

---

Solr  solrconfig.xml




# Testre szabott keresés



```
<fieldType name="date" class="solr.TrieDateField"  
            sortMissingLast="true" omitNorms="true"/>
```


```
<fieldType name="hu_date" class="com.epam.MagyarNemzetiIdo"  
            sortMissingLast="true" omitNorms="true"/>
```




# Full text search / Phrase search / Fuzzy search

 kolumbusz kristóf ki volt ?  

kolumbusz kristóf **hajói**  
kolumbusz kristóf **élete**  
kolumbusz kristóf **utazásai**  
kolumbusz kristóf **felfedezései**

About 26,600 results (0.64 seconds)

[Kolumbusz Kristóf – Wikipédia](#)  
[https://hu.wikipedia.org/wiki/Kolumbusz\\_Kristóf](https://hu.wikipedia.org/wiki/Kolumbusz_Kristóf)  Translate this page  
Cristoforo Colombo vagy hagyományos magyaros nevén **Kolumbusz Kristóf** ..... de

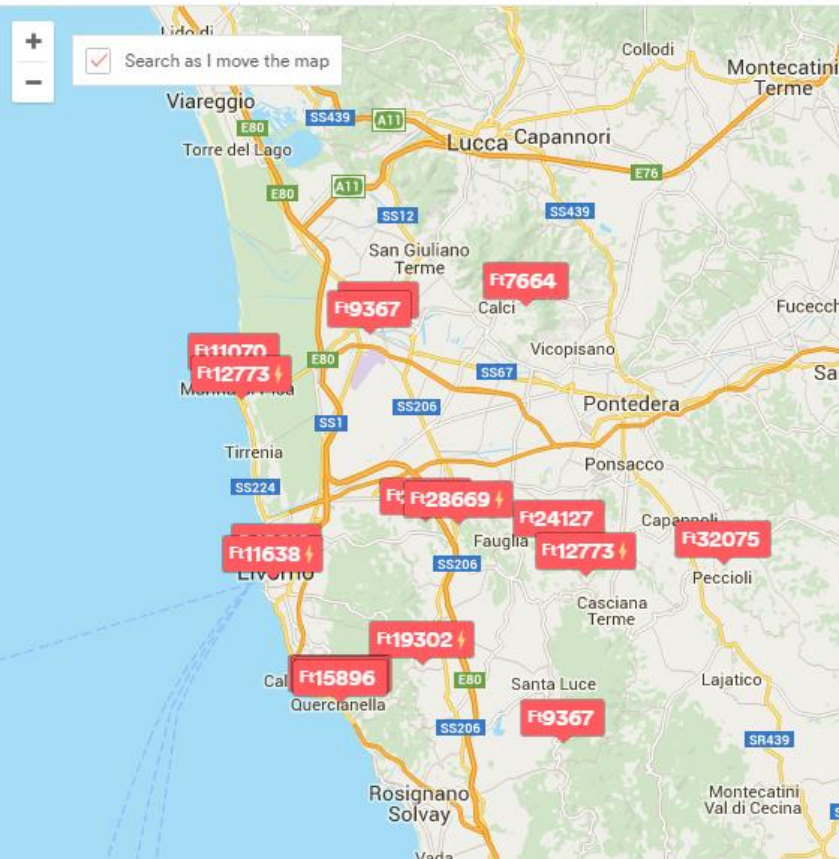
 budapest  

budapest  
**budapest**  
**budapest lyrics**  
**budapest weather**  
**budapest song**

Press Enter to search.



# Spacial search / Geo search



Szállások egy régióban.

Milyen mesze van a tenger ?

# Indexelés



Java API / Rest API (CSV)



Near Real Time vs  
Full Update

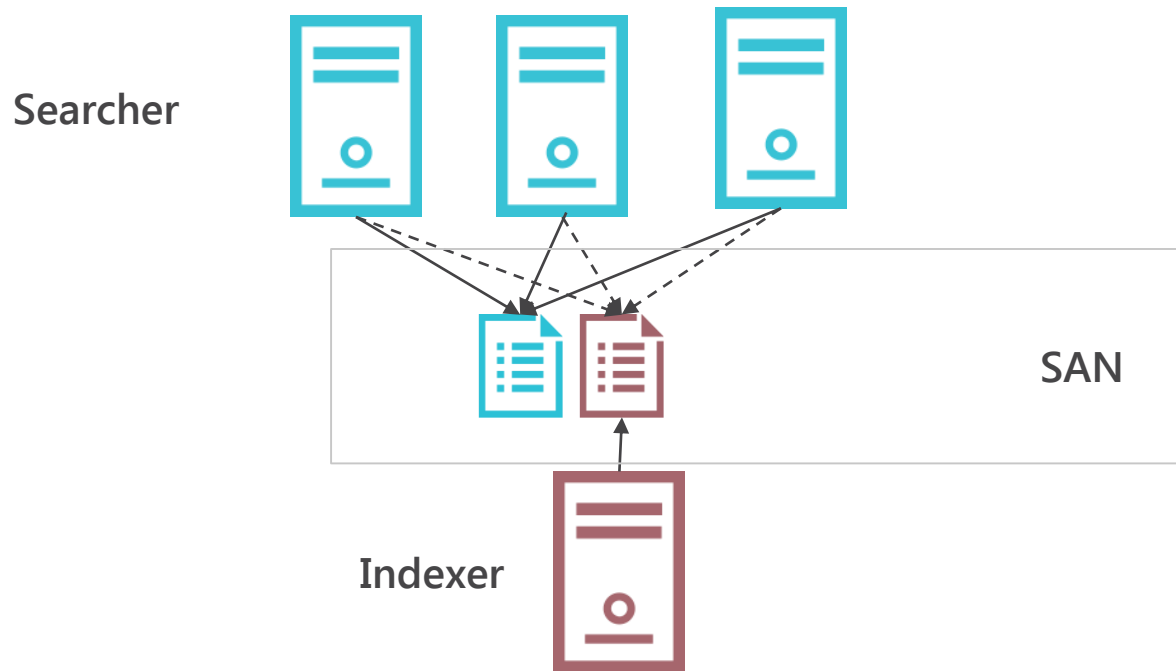


A tökéletesség = \$\$\$



1 instance, 250 millió sor  
~ 6 óra

# Indexelés



## *mikor mit* keresnek a felhasználók ?

Karácsony – egyedül állók foglalnak

November – síelők az Alpokba foglalnak

A keresések 95% az adatok 20-25% célozza



## többszintű cache:

1 szint

gyakran keresett termékek (millisec)

2 szint

ritkábban keresett termékek (< 2sec)

## ARANYSZABÁLYOK

Az üzletnek semmi nem elég

Az adatskálázás ritkán lineáris



# Particionálás

---

Digital SLRs



Mirrorless Cameras



Point-and-Shoots



Lenses



Camcorders



Sports & Action Cameras



Photography Drones



Security Cameras



Optics



Accessories





# Mennyit bír meg egy node ?

---

100 millió / box

- Gyorsabb algoritmusok
- Response cache
- Csak azt tárold ami a kereséshez kell

# Kérdések

