

Filip Vítek

-

Big data - dost' bolo bla, bla...

Ako to naozaj uchopit'?

eFOCUS



Dost' bolo bla, bla ...

Čo skutočne je BIG DATA...



V roku 2001 (!!!) označil Gartner group vo výskumnom článku BIG DATA ako “*hlavnú príležitosť rozvoja mnohých odvetví*”

Koncept 3V

*V*elocity (of data creation)

*V*olume of observations (# of lines)

*V*ariety of data (formats/sources)

Vyvinuli sa **obsesie** ohladne posledných 2 V:

VOLUME-isti = hlavne nech sú tie súbory dát veľké (desiatky milionov riadkov ...)

VARIET-isti = hlavne nech sú tie údaje, čo najmenej štruktúrované a komplikované na spracovanie [*prúd social media*]

VS.

talks about ...

BIG DATA
how to use already generated...

BIG DATA is like

sex among teenagers:

- everybody *talks* about it
- almost *no body* really did it
- and those who did, *didn't really have the best moments of life*

Dost' bolo bla, bla ...

V čom je teda problém ???

“Nomen Omen“

Viete ako to chodí ...

Technologické spoločnosti
vtlačili diskusiu skôr do “BIG“ ako do “DATA“

Možno by bolo lepšie... SMART DATA

Podstatou <doplň názov> DATA má byť získať viac užitočných postrehov , nie zbehnúť rovnako primitívne query nad oveľa väčším objemom riadkov.

opačné “Človeče,
nehnevaj sa“



*Chodíme po vonku a zúfalo hádzeme šestku,
aby sme vkročili na východziu pozíciu*

BIG DATA sa odpútava pozornosť od interných dát ... občas vedie až k utkvelej potrebe bezhlavo preferovať externé dáta pred internými dátovými zdrojmi .

2+5 úrovní napredovania v Big Data

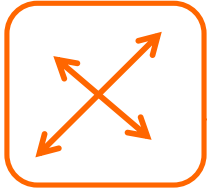
?

Príklady dát

Implem. obtiaže

<p>0. A</p>	<p>Nevyužité vlastné dáta Pre ktoré polia svojej databázy by ste ešte nenašli exter. kupca ?</p>	<ul style="list-style-type: none"> Formálne polia dokumentov Servisné údaje o využívaní produk. 	<p>Nekonvenčný pohľad na podceňované dáta</p>
<p>0. B</p>	<p>Nezbierané vlastné dáta Ktoré informácie o používaní Vašich produktov klientmi ešte nemáte podchytené v databáze ?</p>	<ul style="list-style-type: none"> Anomálie v používaní produktu Dáta neukladané elektron. Preferencie alebo opakujúce sa zvyky klienta 	<p>Manuálne procesy neprodukujú digitálne stopy. Profesionálna slepota.</p>
<p>1. vrstva</p>	<p>Údaje spoločných služieb Kde Vaši klienti používajú paralelne s vašimi produktmi aj iné (služby)?</p>	<ul style="list-style-type: none"> Autoservisy vs. Poistenie áut Telco signál + prezeraný obsah Banky + kde míňam Účet za vodu + účet za elektrinu 	<p>Rozhrania systémov, dátová kvalita. Ochota vymieňať si klientske dáta.</p>
<p>2. vrstva</p>	<p>Užívateľom vygenerované Aké údaje vytvárajú o sebe klienti pri používaní vášho produktu na sociálnych sieťach alebo inde?</p>	<ul style="list-style-type: none"> Telefonovanie + Geo lokácia Nike plus aplikácia Klienti v internetbankingu Sentiment o značke na Facebooku 	<p>Rozumná miera ochrany súkromia. Profesionálna slepota.</p>
<p>3. vrstva</p>	<p>Údaje iných firiem/odvetví Ktoré iné odvetvia by vám vedeli poskytnúť údaje o klientoch, ktoré vy sami nemáte ako zistiť?</p>	<ul style="list-style-type: none"> Externé obohacovanie dát výskumnou agentúrou Predaj „príznakov“ o klientoch Úverový register, blacklisty 	<p>Dátová kvalita u dodávateľa údajov. Pomer cena/výkon.</p>
<p>4. vrstva</p>	<p>Verejné registre v SR Ktoré verejné databázy by Vám dodali lepší obraz o klientoch?</p>	<ul style="list-style-type: none"> Databázy ORSR, ŽRSR, ... Databáza poistených vozidiel Kataster portál nehnuteľnosti 	<p>Spojovací kľúč databáz. Algoritmizácia dáta mining robota.</p>
<p>5. vrstva</p>	<p>Zahraničné dostup. registre Žiadnu otázku si nekladzte, proste googlite</p>	<ul style="list-style-type: none"> Eurostat databázy Zahraničné sesterské firmy Univerzitné výskumné databázy Asociácie a medzinár. organizácie 	<p>Časová aktuálnosť dát. Porovnateľnosť údajov naprieč krajinami.</p>

Úroveň 0a Nevyužité interné dáta

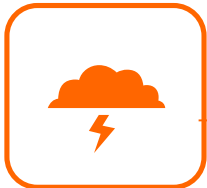


KDE HĽADAŤ takéto dáta:

- Dáta, ktoré majú **druhý** (a ďalšie) **významy**
- Dáta, ktoré vyzerajú ako **nudné povinné polia**



*Pre ktoré polia
svojej databázy
by ste zatiaľ
nenašli kupca?*



NA ČO POZOR pri implementácii tejto úrovne

- Podceňovanie dát, úzkoprsý pohľad na obsah
- Lipnutie na **formáte** a nie informačnej hodnote (... alfanumerické ID)



REÁLNE PRÍKLADY (zo Slovenska a blízkeho okolia):

Príklady “prekvapivých” interných dát

1 Preferencia banky

(mBank vs. Poštová banka)



2 Dĺžka vzťahu s bankou

(133333333 /xxxx
133353333 /xxxx)



3 Cielenie X-sell ponuky

(priemerná cena v danom PSČ/ulici)



5 Distance to yes



4 Druh / štandard bývania

(rodinné domy, byty)

8 Stupeň oficiálnosti

(Jozef.Smely @hotmail.com
Johnyfrajer @hotmail.com)



9 Vplyv na iných klientov z portfólia

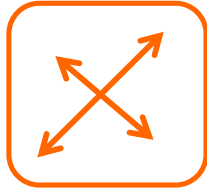
Jana.Nastvana@sberbank.cz
Michal.Spokojny@sberbank.cz



7 Telef. operátori

(a ich podiel na vašich klientoch)

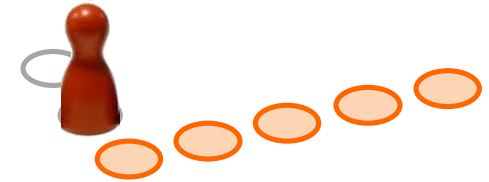
Úroveň 0b Nezberané klientske dáta



KDE HĽADAŤ takéto dáta:

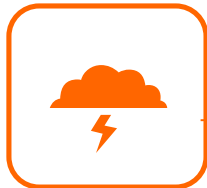


- **Netypicky vysoké** alebo nízke hodnoty (outliers)
- Nízko početné **zvláštnosti** (Rybana Lakatošová)
- Dáta, ktoré sa doposiaľ **neukladajú do elektronickej podoby** (kniha sťažností a prianí)
- **Opakujúce sa správanie** klienta (zvyk alebo preferencia)



?

Keď si sadnem na deň do Vašej pobočky/ predajne, čo z toho, čo odpozorujem, nenájdem v žiadnej databáze?



NA ČO POZOR pri implementácii tejto úrovne

- **Manuálne procesy** negenerujú žiadne informácie o priebehu danej operácie (z dátového pohľadu ako keby sa neudiali)
- **Profesionálna slepota**, veľký nepriateľ tejto vrstvy (pýtajte sa nezainteresovaných)



REÁLNE PRÍKLADY (zo Slovenska a blízkeho okolia):

Nemáme dost' času ísť úplne do hĺbky,
ale tu je niekoľko inšpirácií ...



Akademický titul
klienta



Počet miest, v ktorých
klient zaplatil kartou za
posledných 6 mesiacov



Z koľkých IP adres
sa nalogoval človek?



Znamenie
zverokruhu



Zamestnávateľ
klienta

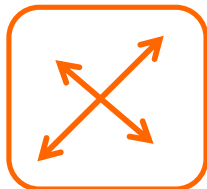


Najväčší výber z
bankomatu



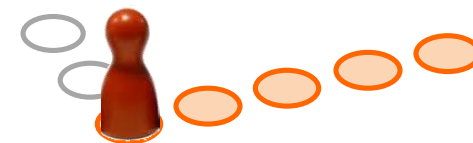
Zvyčajný čas
prihlásenia do EB
alebo e-shopu

Úroveň 1 Údaje spoločných služieb

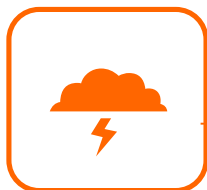


KDE HĽADAŤ takéto dáta:

- Čo robil klient pred tým než prišiel na moju stránku? Kam z nej odišiel ?
- Kde sa nachádzal, **v akej bol situácii**, keď využil moje služby? (volal na call centrum zo zahraničia)
- Každý musí **jesť, presúvať sa, telefonovať**, čítať niečo, oblieť sa, nakupovať, ...



Aké iné služby Vaši klienti používajú paralelne (resp. tesne pred alebo po) s Vašou službou ?



NA ČO POZOR pri implementácii tejto úrovne

- Rozhrania systémov a hlavne **Dátová kvalita** premostených údajov
- Ochota druhých strán vymieňať si tieto dáta s Vami
- Postoj klientov k tomu, že disponujete takýmito dátami



REÁLNE PRÍKLADY (zo Slovenska a blízkeho okolia):

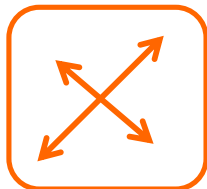
Na čo míňame?



Servisy a poistenie áut



Úroveň 2 Užívateľom generované dáta



KDE HĽADAŤ takéto dáta:

- Zmieňuje sa užívateľ o vašej značke alebo produkte na **sociálnych sieťach alebo na fórach/blogoch**?
- Aké **interakcie** s vašimi obslužnými kanálmi má ? [call centrum, osobná náštava, web portál, ...]
- Aký **sentiment prejavil** klient k Vašej službe/produktu?



Aké údajové stopy po sebe nechávajú klienti pri používaní vašich služieb na sociálnych sieťach alebo inde?



NA ČO POZOR pri implementácii tejto úrovne

- Rozumná miera súkromia, vyhnite sa stalkingu
- Stávka na jednu technológiu (FB, Twitter, ,...)
- **Out of the box** – kategória údajov, ktoré sú v podstate irelevantné pre beh samotného biznisu



REÁLNE PRÍKLADY (zo Slovenska a blízkeho okolia):



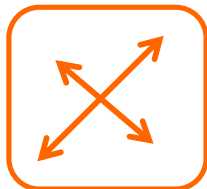
Phone usage



CallinspectorMobile

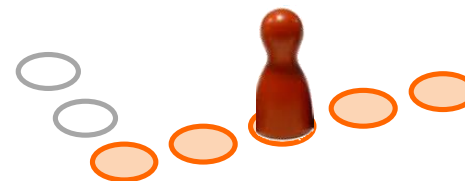


Úroveň 3 Údaje iných spoločností [data monetizácia]

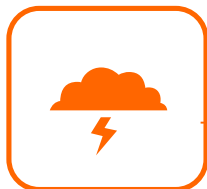


KDE HĽADAŤ takéto dáta:

- Externé obohacovanie dát vzorkou z prieskumu (external profiling napr. pri segmentácii)
- Nákup / výmena príznakov o spoločnosti
- Spoločné registre spoločností (bank. register, Telco, ...)



Ktoré iné odvetvia by vám vedeli poskytnúť informácie o Vašich klientoch?



NA ČO POZOR pri implementácii tejto úrovne

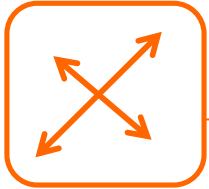
- Dátová kvalita dodávateľov ako aj jednoznačnosť namapovania údajov (3x Peter Kováč na tej istej adrese)
- Update nakúpených údajov vo Vašich databázach (ťažko sa obhaja opakovaný nákup)
- Pomer cena/výkon – ide totiž iba o doplnkové údaje



REÁLNE PRÍKLADY (zo Slovenska a blízkeho okolia):

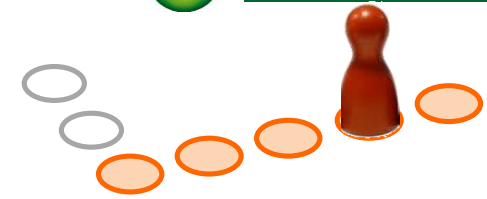


Úroveň 4 Verejné registre SR

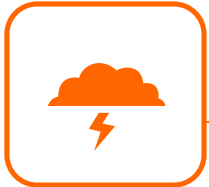


KDE HĽADAŤ takéto dáta:

- Web stránky vlády, štátnych organizácií a samospráv
- Asociácie, profesné združenia a komory
- Organizácie so špecifickou úlohou (SKP, soc.poiťovňa, ...)



Ktoré verejne databázy obsahujú info na dokreslenie obrazu Vášho klienta?



NA ČO POZOR pri implementácii tejto úrovne

- Spojovací kľúč pre jednoznačne (automatizované) prepojenie databáz
- Algoritmizácie dopytovacích robotov (pre skutočne veľké databázy)
- Data hunt prevencia – (Captcha, IP slowdown, IP denial, ...)



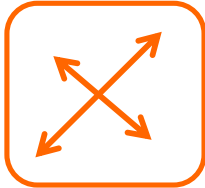
REÁLNE PRÍKLADY (zo Slovenska a blízkeho okolia):

SKP

ORSR

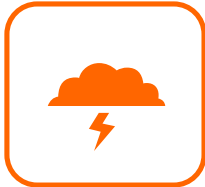
FOAF

Úroveň 5 Zahraničné dostupné registre



KDE HĽADAŤ takéto dáta:

- Databázy naprieč biznis skupinami (zahraničné sestry)
- Univerzitné prieskumy (opensource zaujímavej kvality)
- Asociácie a medzinárodné organizácie (MMF, OECD, EU, OSN,...)



NA ČO POZOR pri implementácii tejto úrovne

- Aplikovanie vzorky na domestikované dáta
- Aktuálnosť a aktualizácia (pre účely biznis rozhodnutí)
- Data hunt prevencia – (Captcha, IP slowdown, IP denial, ...)
- Overiť si kultúrne rozdiely (napr. priemerný počet detí v rodine)



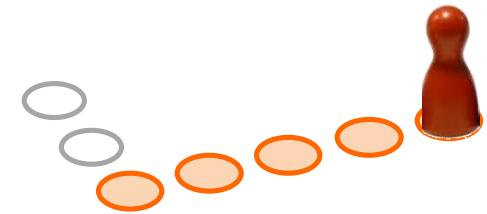
REÁLNE PRÍKLADY (zo Slovenska a blízkeho okolia):

Správanie klientov v mobilnej aplikácii

Allianz 

Predikcia príjmu klienta

[register nehnuteľností a register áut]



Nič sa nepýtajte samých seba.

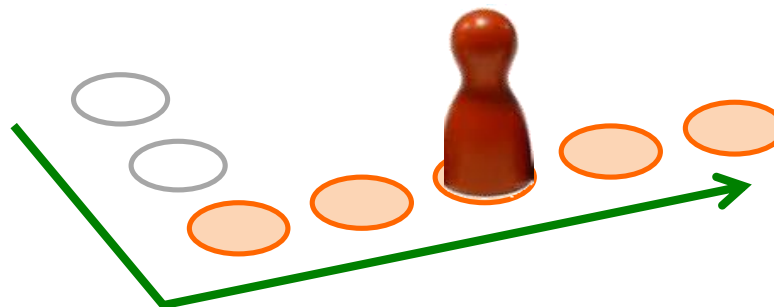
Proste Googlite :)

Dost' bolo bla, bla ...
Krátke zhrnutie



Think not just **BIG DATA**, ...
think rather
SMART data!

Nehrajte človeče
opačne ...



Vaše otázky, prosím

Spojte se
se mnou

Mgr. Filip Vitek

Head of department
Customer value mng. dept.
Sberbank Slovensko, a.s.



+ 421 903 457379

filip.vitek@sberbank.sk



<http://sk.linkedin.com/pub/filip-vitek-700-connections/4/3a/16a>

@FilipVitek



Kde nájdem
viac?

Zaujalo Vás to?

Máte na témy
iný názor ako ja ?

Píšem CRM blog, zastavte sa:

<http://blog.etrend.sk/filip-vitek/>

Blogy sú v Slovenčine, alebo bilinguálne v Angličtine. Ak sa Vám páčia, pošlite ďalej vo svojom okolí.

Vaše otázky, prosím

Spojte sa
so mnou

Mgr. Filip Vítek

Head of department
Customer value mng. dept.
Sberbank Slovensko, a.s.



+ 421 903 457379

filip.vitek@sberbank.sk



<http://sk.linkedin.com/pub/filip-vitek-700-connections/4/3a/16a>

@FilipVitek



Kde
nájdem
viac?

Zaujalo Vás to?

Máte na témy
iný názor ako ja ?

Píšem CRM blog, zastavte sa:

<http://blog.etrend.sk/filip-vitek/>

Blogy sú v Slovenčine, alebo bilingválne v Angličtine. Ak sa Vám páčia, pošlite ďalej vo svojom okolí.

eFOCUS