

# Tehnični izzivi sodobne poslovne analitike

Mateja Verlič Brunčič, Boris Ovcjak

Databox Inc., Ptuj, Slovenija  
mateja@databox.com, boris.ovcjak@databox.com

**Sinopsis** Poslovna analitika lahko pomaga podjetjem sprejeti boljše in informirane odločitve ne glede na velikost podjetja. Na podlagi podatkov podjetja boljše razumejo obnašanje svojih strank ali pridobijo vpogled v konkurenco. Pri tem ne gre zanemariti tehničnih izzivov, ki pridejo s področjem in na finančne, človeške, tehnične, pa tudi strateške omejitve pri reševanju. Podatke je potrebno prečrpati, obdelati, analizirati, vizualizirati in na koncu pripraviti v obliki, ki jo uporabniki razumejo in jo uporabijo kot osnovo za odločanje. Vsak korak predstavlja specifikum in ima svoj nabor omejitev, dodaten vir izzivov pa predstavljajo tudi zunanje storitve, od katerih pridobivamo podatke in nad katerimi načeloma nimamo kontrole. Databox ima več kot deset let izkušenj z reševanjem izzivov na tem področju. V prispevku bomo predstavili, kako smo se lotili njihovega reševanja, da bi uporabniku čim bolj približali podatke na intuitiven in razumljiv način, ki jim daje dodano vrednost, omenili bomo tudi, kakšni izzivi nas čakajo v bližnji in srednji prihodnosti.

## **Ključne besede:**

razvoj digitalnih storitvenih platform

poslovno obveščanje

poslovno usmerjen razvoj

## 1 Uvod

Poslovna analitika ni nova stvar, vendar vztrajno pridobiva na pomembnosti in vedno več je podjetij na trgu, ki ponujajo takšno ali drugačno storitev poslovne analitike. Čas Excela in preglednic še sicer ni minil, se pa vse več podjetij zaveda, kako pomembno je spremljati uspešnost, doseganje kralnikov in vključevati več ljudi iz tima z namenom povečanja transparentnosti, kar pa običajno presega zmožnosti in namembnost trenutnih aplikacij, še posebej namiznih. Srednja in manjša podjetja iščejo SaaS alternative za sodobno poslovno analitiko, saj večina nima ne virov in ne znanja, da bi si sami postavljali podatkovna skladišča in po meri narejeno poslovno analitiko. Na tem mestu vstopi Databox, ki že več kot deset let deluje na tem področju in omogoča manjšim in srednjim podjetjem relativno enostaven način vzpostavitve sodobne poslovne analitike, obdelave podatkov, vizualizacij in poročanja metrik poslovanja.

## 2 Sodobna poslovna analitika

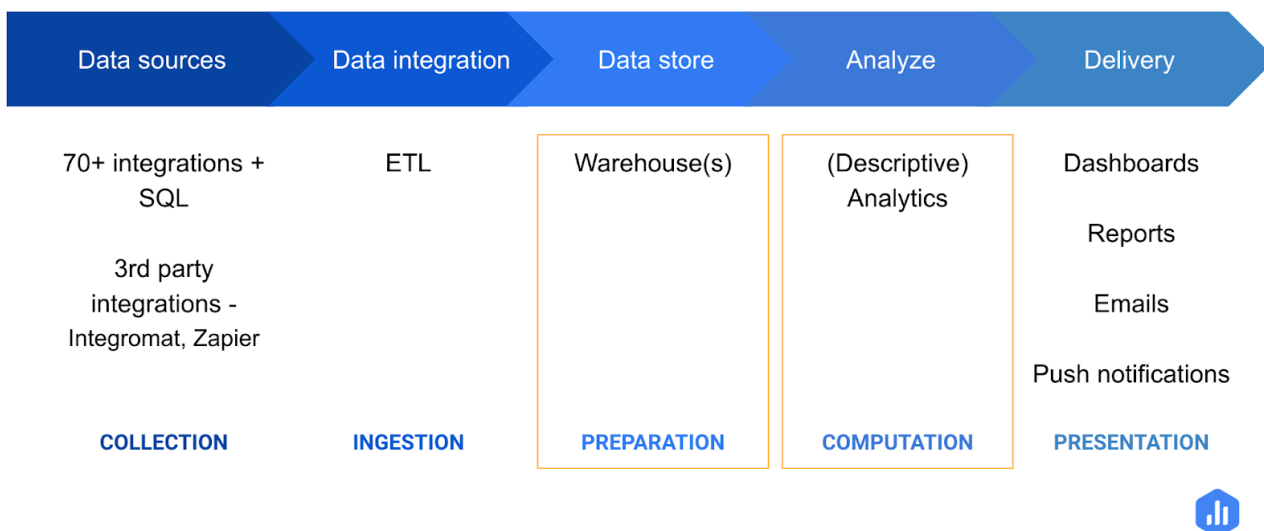
Storitve poslovne analitike se ne dotikajo samo marketinga, ki je sicer tipični uporabnik storitev poslovne analitike, ampak tudi financ, projektnega vodenja, prodaje in še mnogih drugih področij. Za vsako od teh področij najdemo ponudnike storitev s specifičnimi metrikami, lastnimi vmesniki ter drugačnim načinom delovanja. Njihova skupna značilnost pa je to, da jih večina nudi javne programske vmesnike (public API), preko katerih je možno dostopati do podatkov oziroma jih tudi pretakati. Vizualizacija in poročanje sta pomembni komponentni poslovne analitike in Databox omogoča uporabo storitev na poenoten način, agregacijo metrik iz različnih podatkovnih virov in prikaz metrik drugo ob drugi za lažje spremljanje in zaznavo korelacij.

Databox se ukvarja z izzivi in omejitvami v vsakem od korakov podatkovnega toka, pri čemer so ravno omejitve tiste, ki narekujejo način reševanja. Na poslovnem področju imajo podjetja finančne omejitve (stroški), omejitve virov (človeških ali strojnih) in omejitve s strani poslovnih odločitev, ki velikokrat postavijo smernice reševanja določenih težav. V teoriji je možno marsikaj, v praksi ponavadi malo manj. Naš sistem je porazdeljen, dinamičen in asinhron, kar lahko predstavlja dodaten vir izzivov, s katerimi se srečujemo, a hkrati nudi rešitve za to, kar potrebujemo.

## 3 Podatkovni cevovod in tehnični izzivi

Področje poslovne analitike je nasičeno, v večji meri storitve s tega področja pokrivajo posamezne segmente podatkovnega cevovoda, ena izmed naših prednosti pred konkurenco je pokrivanje celotnega cevovoda, kar pa seboj seveda prinese določene dodatne izzive.

S platformo pokrivamo vseh pet področij (Slika 1): 1) podatkovne vire, 2) integracijo podatkov, 3) shranjevanje podatkov, 4) analizo in 5) dostavo. Pri tem se srečujemo z različnimi tehničnimi izzivi, ki se dotikajo več tehnoloških področij.



Slika 1: Področja podatkovnega cevododa, ki jih pokriva platforma Databox.  
Vir: lasten.

### 3.1 Pridobivanje podatkov iz podatkovnih virov

Podatkovne vire v našem primeru predstavljajo ponudniki zunanjih storitev (angl. *3rd party service providers*), ki so integrirane v našo platformo in iz katerih črpamo podatke za metrike, ki se potem agregirajo pri nas za lažjo nadaljnjo obdelavo in hranjenje.

Na tem področju se srečujemo z izzivi milijonov dnevniških podatkovnih zahtev z odzivi od nekaj do tudi milijonov vrstic. Zaradi omenjenega števila zahtev se lahko hitro zgodi, da naletimo na limite zahtev, ki jih postavljajo ponudniki in so lahko tako na osnovi uporabnika kot tudi globalne. Ko pridobimo ustrezne odzive, pa je pomembno, da le-te tudi ustrezno shranimo in pripravimo na nadaljnjo obdelavo - za ETL (angl. *extract transform load*).

Soočamo se tudi s specifikami in nekonsistentnostjo različnih ponudnikov podatkov, kjer je pomanjkanje formalnih modelov za predstavitev aktivnosti ETL, ki bi mapirali vhodne podatke virov, da bi bili v primernem formatu za nalaganje v ciljna podatkovna skladišča [1]. Ker je teh veliko se moramo prilagajati različnim aplikacijskim vmesnikom, spreminjajočimi se specifikacijami, omejitvam glede pridobivanja metrik ter limitami ter tudi različnimi časovnimi razponi [2]. Med glavnimi omejitvami so predvsem omejitve glede pridobivanja zgodovine metrik, saj vsi ne omogočajo pridobivanja zgodovine ali pa je le-ta omejena, kar nas naredi dovzetnejše za posledice izgube podatkov pri nas. Pri delu z zunanjimi ponudniki storitev za beleženje analitičnih podatkov se srečujemo z različnimi vrstami, strukturami in formati podatkov, predvsem pa tudi z različnimi interpretacijami pomena posameznih metrik.

### 3.2 Integracija podatkov

Ko pridobimo podatke s strani ponudnikov, sledi pomemben in hkrati najbolj zahteven del procesa priprave podatkov - ETL. V tej fazi se moramo zavedati različnosti naših podatkovnih virov, nabora ponujenih metrik ter njihovih specifik in na koncu tudi samih časovnih intervalov podatkov. Pri tem pa zelo pomembno vlogo igrajo tudi sami časovni pasovi, ki se lahko razlikujejo tako med samimi podatkovnimi viri kot pa samim izhodišnim časovnim pasom naših uporabnikov.

V tej fazi je naša glavna naloga, da znamo pridobljene podatke ekstrahirati iz samih odzivov na kar se da enoten način, poskrbeti da metrike ustrezno obravnavamo (različni tipi metrik), ter jih transformirati v obliko, ki bo olajšala obdelavo, shranjevanje ter omogočila ponovno uporabo tako za platformo kot podatkovno rudarjenje.

### 3.3 Shranjevanje podatkov

V tej fazi smo že poskrbeli za ustrezno obliko podatkov, primerno za shranjevanje. Potrebno se je zavedati da gre v tem primeru za velike količine podatkov, za katere moramo zagotoviti tako hitro shranjevanje kot tudi kasnejše branje. Pri tem pa so pomembne tudi programske optimizacijske tehnike, kjer se je potrebno zavedati narave naših metrik in posledično podatkov. Glede na tip metrike se lahko namreč zgodi, da se vrednosti določenih podatkov lahko s časom spreminjajo, medtem ko so v drugih primerih večni.

Poleh samih podatkov pa se moramo zavedati tudi količine le-teh. S številom uporabnikov in njihovih podatkovnih virov se število podatkov nenehno večja in naša odgovornost je, da skrbimo za različne metode skaliranja podatkovnih shramb, s tem pa ohraniti dobro uporabniško izkušnjo.

### 3.4 Analiza podatkov

Preden lahko uporabniku dostavimo podatke je potrebno opraviti še enega izmed pomembnejših korakov, in sicer sama analiza podatkov, kalkulacija in priprava le-teh za vizualizacije. Izziv pri tem predstavljajo predvsem raznovrstni tipi metrik, njihove enote ter specifike njihovih podatkovnih virov, kot so različni časovni pasovi ter sam prekop na poletni/zimski čas. Pri tem pa omogočamo pa tudi različne agregacije ter izračunane metrike (iz različnih metrik).

### 3.5 Dostava (vizualizacija) podatkov

Pri vizualizaciji in poročanju se srečujemo spet z drugim naborom izzivov, ki jih s seboj prinesejo pravočasna priprava podatkov za vizualizacijo, asinhronost sistema, kjer so podatki eventualno konsistentni, zagotavljanje dobre uporabniške izkušnje in pravočasno dostavljanje poročil po različnih distribucijskih kanalih.

## 4 Rešitve s sklopu platforme Databox

V sklopu platforme smo reševali izzive z vseh prej omenjenih področij, da bi pokrili potrebe naših uporabnikov in zagotovili najboljšo uporabniško izkušnjo. V nadaljevanju na kratko opisujemo pristope k iskanju in uporabi rešitev v najbolj kritičnih zgoraj omenjenih korakih.

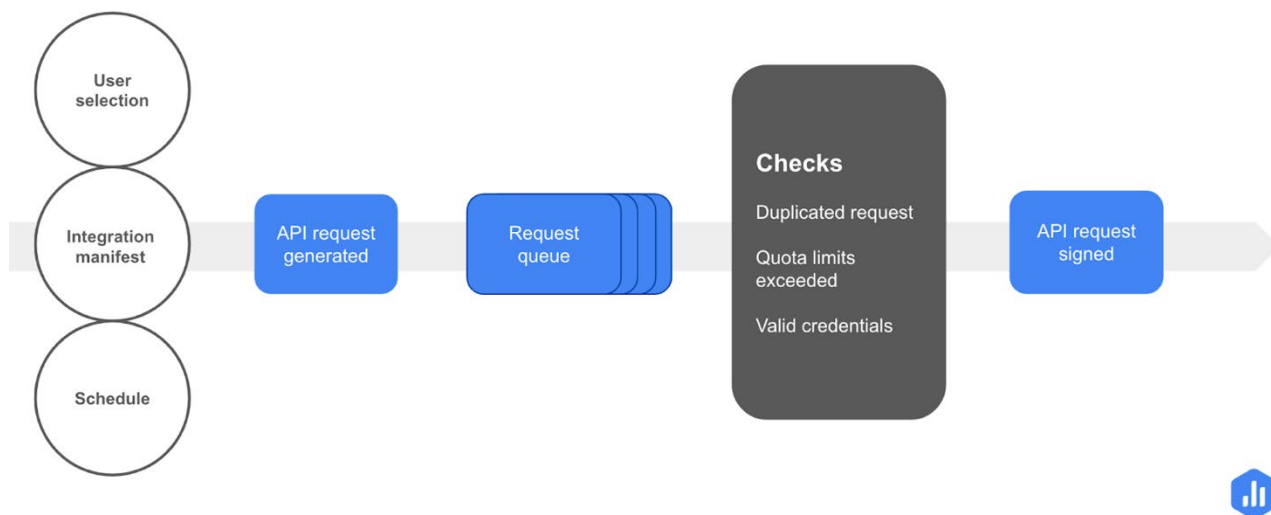
### 4.1 Pridobivanje podatkov iz podatkovnih virov

Problematike pridobivanja ogromne količine podatkov se lotevamo tako, da pridobivamo oziroma sinhroniziramo le metrike, ki jih je uporabnik izbral in jih dejansko uporablja (pregleduje) v aplikaciji. Na tak način smo zagotovili, da imamo podatke, ki jih uporabnik dejansko potrebuje. S tem sistema ne obremenjujemo po nepotrebnem, hkrati pa skrbimo za optimizacijo stroškov za infrastrukturo. Tudi podatke, ki smo jih že sinhronizirali, ne sinhroniziramo ponovno oz. sinhroniziramo le tiste, katerih narava nam ne omogoča trajne shrambe (metrikam se lahko spreminja zgodovina ali vrednost skozi čas). V ozadju je kompleksen sistem shranjevanja nastavitve posameznih metrik, s katerim smo dosegli relativno optimiziran proces sinhronizacije podatkov.

Poslužujemo se tudi razporejanja in prioritizacije zahtev glede po urniku in glede na izbrane plane naših uporabnikov, običajno so prioritizirani uporabniki, ki plačujejo. Da bi preprečili doseganje limit števila zahtevkov

pri ponudnikih zunanjih storitev, moramo znati zahteve tudi zakasnit in razporediti čez čas, da imajo uporabniki karseda aktualne podatke. S problemom večjih količin odzivov se soočamo tako, da začasno shranjujemo zahteve, dokler ne pridejo na vrsto za obdelavo in shranjevanje v podatkovno skladišče.

Slika 2 prikazuje korake v procesu pridobivanja podatkov in implementirane točke za preverjanje in optimizacijo. Glede na izbiro metrik s strani uporabnika, meta podatki o metriki in podatkovnem viru v integracijskem manifestu in urnikom predvidenega pridobivanja podatkov zgeneriramo zahteve in jih damo v vrsto za izvajanje. Pred dejanskim pošiljanjem zahtev na ponudnike storitev izvedemo preverjanja, da izločimo zahteve, ki jih ni potrebno pošiljati ali jih ne moremo. Če zahteva preстане validacijo, jo ustrezno opremimo in pošljemo naprej do ponudnikov storitev.

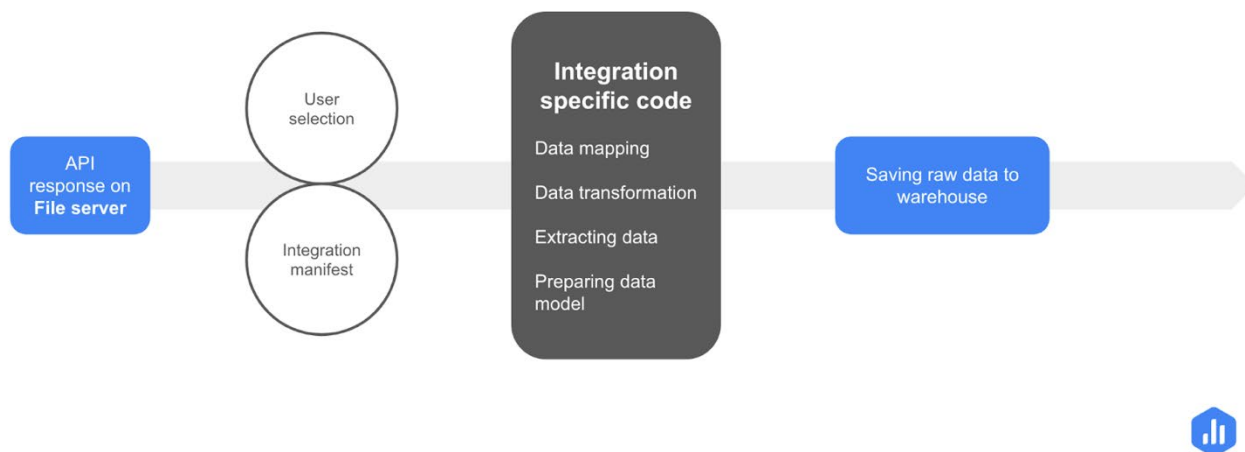


**Slika 2:** Proces pridobivanja podatkov iz različnih podatkovnih virov in implementiranimi točkami preverjanja.  
Vir: lasten.

## 4.2 Integracija podatkov

Raznolikost ponudnikov zunanjih storitev pridobivanja analitičnih podatkov ter njihovih metrik smo rešili z uvedbo integracijskega manifesta, preko katerega poskrbimo za harmonizacijo podatkov. Vsebuje tudi navodila za procesiranje vhodnih odzivov in preslikavo vhodnih podatkov v format podatkov, ki ga uporablja naš sistem. Na tak način smo podatke poenotili in našim uporabnikom omogočili, da metrike iz različnih virov uporabljajo na enak način, jih kombinirajo in relativno enostavno uporabijo na svojih preglednih ploščah z metrikami (angl. *dashboard, databoard*).

Slika 3 ilustrira procesiranje pridobljenega odziva s strani podatkovnega vira, ki ga glede na uporabnikovo izbiro metrik in integracijskega manifesta spustimo skozi kodo, ki je specifična za izbrano integracijo, podatke ustrezno preslikamo, transformiramo, in izluščimo podatke, ki jih nato shranimo v podatkovno skladišče.



**Slika 3:** Proces integracije pridobljenih podatkov v platformo.

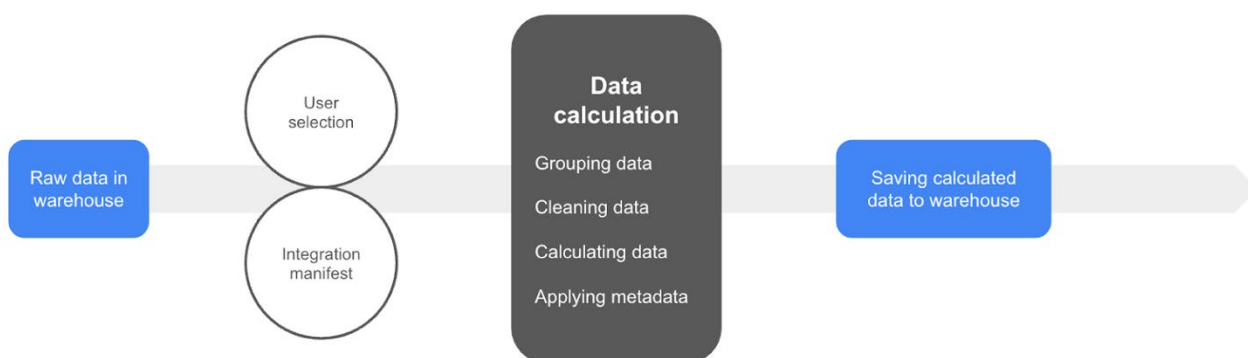
Vir: lasten.

### 4.3 Shranjevanje podatkov

Problema reševanja shranjevanja podatkov iz različnih virov smo se lotili z uporabo poenotenega podatkovnega modela, s katerim nam je uspelo vsesplošno raznolikost metrik postaviti na isti skupni imenovalec. To smo uspeli tako, da smo zajeli skupen minimalni nabor podatkov, ki jih potrebujemo, da lahko omogočimo nadaljnjo obdelavo in analizo podatkov na preprost in enoten način.

### 4.4 Analiza podatkov in vizualizacija

Za pripravo, analizo podatkov in hitrejši prikaz uporabljamo posebno storitev, ki uporabi navodila na podlagi uporabnikove izbire na grafičnem vmesniku, da zbere in pridobi shranjene podatke o metriki, jih primerno združi, izvede potencialne izračune (npr. agregacijo) in vrne izračunane podatke odjemalcu za prikaz (Slika 4). Pri tem upoštevamo metapodatke z informacijo o poteku veljavnosti podatkov, da optimiziramo osveževanje podatkov na podlagi veljavnosti in dejanskega obstoja izračunanih podatkov.



**Slika 4:** Proces analize podatkov in priprava na vizualizacijo.

Vir: lasten.

Tako pripravljene podatke so potem uporabljeni s strani odjemalca, ki jih lažje in hitreje vizualizira in se ne rabi ukvarjati še s transformacijo.

## 5 Odprti izzivi

V sodobni analitiki nikoli ne zmanjka novih. Z zgoraj opisanimi implementiranimi rešitvami smo sicer določene probleme odpravili, druge blažimo oziroma nadzorujemo, imamo pa tudi še nekaj odprtih izzivov, ki so običajno pridruženi rasti podjetja. Mednje sodijo upravljanje naraščajočih obremenitev sistema zaradi vedno več uporabnikov in novih integracij, skaliranje arhitekture, uvedba in izvajanje procesov odzivanja na nepričakovano delovanje sistema, priprava okvirjev referenčnih performančnih vrednosti, in preprečevanje izgube podatkov. S temi izzivi se ukvarjajo različni inženirski timi na Databoxu, ki skrbijo za posamezne domene.

## 6 Zaključek

Trenutna platforma za poslovno analitiko v oblaku pokriva večino potreb naših uporabnikov, vendar se nenehno trudimo prehiteti konkurenco z predvidevanjem potreb uporabnikov in uvajanjem sodobnih pristopov, med drugim tudi z apliciranjem podatkovnih ved. Skoraj vsakih nekaj mesecev se pojavi nova storitev, ki uporabnikom omogoča poenostavljeno spremljanje poslovanja, zato je še toliko pomembnejše, da jim zagotovimo dodatno dodano vrednost na podlagi zbranih podatkov, podatkovnega modeliranja, prediktivne analitike in drugih sodobnih pristopov.

## Literatura

- [1] KIMBALL R., CASERTA J. "The Data Warehouse ETL Toolkit. Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data", Wiley, 2004.
- [2] LOSHIN David "Data Integration", MK Series on Business Intelligence: Business Intelligence (Second Edition), Morgan Kaufmann, 2013, 189-210.