

# Podaci – podaci, svuda oko nas...

**Trenutačno aktualna problematika GDPR-a** natjerala je tvrtke da se upitaju gdje im se sve nalaze osobni podaci, jesu li neki od njih još uvijek potrebni ili se mogu obrisati, kako tvrtka mora upravljati podacima...



**Boris Stipetić, Technology Consulting Manager, Megatrend poslovna rješenja**

**S**vakodnevno prikupljamo i generiramo velike količine podataka. Kako se tehnologija razvija, sve to postaje lakše i brže, a rastu i volumeni koji se smatraju "normalnim" – USB *stickovi*, ograničenje na veličinu *e-maila*, razlučivost "obične" fotografije napravljene mobitelom... Radimo izvještaje i razne analize, spremamo razne verzije, i količina samo raste i raste. Prije pet godina samo spominjanje sustava koji ima desetak ili više terabajta zvučalo je impresivno, davalo naznaku da se radi o velikoj organizaciji, a danas i male tvrtke koriste nekoliko terabaj-

ta, i nitko se tome više ne čudi. Istraživanja govore da se u većini organizacija volumen digitalnih podataka udvostručuje svakih 18-24 mjeseca. Jasno, ako želimo te podatke sačuvati, treba raditi *backup*, a na kraju sve to košta. Nevolja s podacima jest to što vremenom njihova vrijednost opada, ali troškovi i rizici povezani uz njih ne (pohrana jednog petabajta košta približno četiri milijuna dolara).

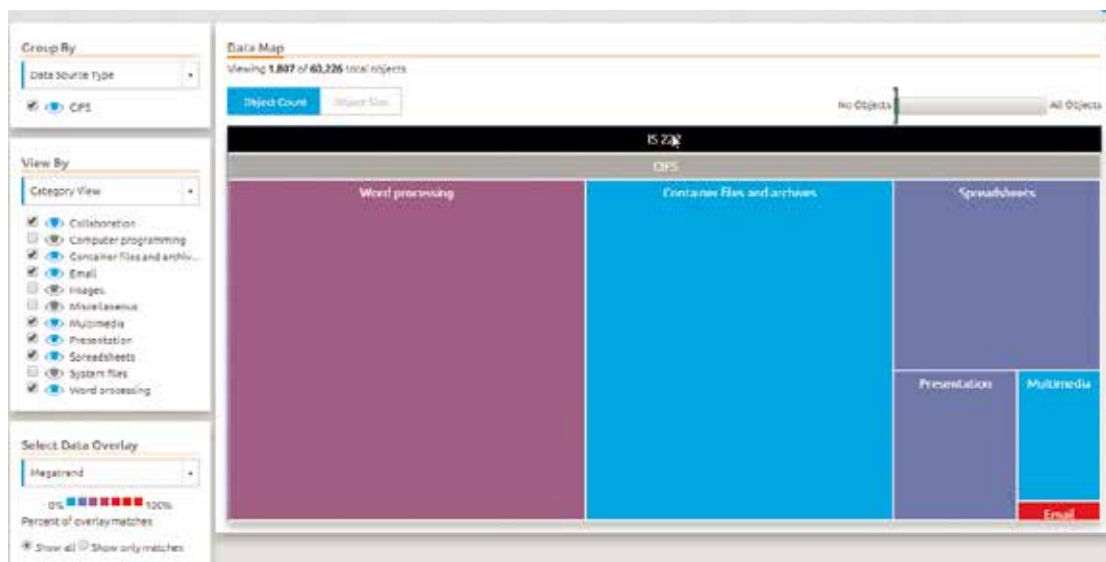
## UPRAVLJANJE PODACIMA

Dakle, jasno je da je iznimno važno dobro upravljati svim tim podacima, njihovim du-

plikatima, nepotrebnim datotekama i sličnim. To znači da treba ukloniti nepotrebne te se fokusirati na uistinu bitne podatke, posebne kategorije podataka premjestiti na zasebne repozitorije i slično, sve u skladu s politikom organizacije i procedurama koje idu u smjeru boljeg poslovanja i smanjenja rizika. Trenutačno aktualna problematika GDPR-a natjerala je tvrtke da se upitaju gdje im se sve nalaze osobni podaci, jesu li neki od njih još uvijek potrebni ili se mogu obrisati, kako tvrtka mora upravljati podacima da bi bila u skladu sa zahtjevima koje pred njju predstavlja regulativa...

Kako bi se na tom planu učinili pomaci, potrebno je shvatiti što sve od podataka posjedujemo, u kakvom formatu, kao i što je s njima potrebno učiniti. Već sama činjenica da tvrtka želi zadovoljiti regulatorne zahtjeve tjera tu istu tvrtku da dobro istraži podatke koje posjeduje. Kad je već tako, zašto ne iskoristiti priliku i promijeniti paradigmu tako da podaci više ne predstavljaju obavezu i rizik, nego vrijednost kojom tvrtka raspolaže.

Kako bi se zaista najbolje upravljalo podacima, potrebno je spoznati gdje su,



**IBM StoredIQ razvijen je s idejom da se aktivno upravlja podacima i smanji rizik koji oni nose**

## ALAT IBM STOREDIQ

**IBM StoredIQ** razvijen je s idejom aktivnog upravljanja podacima i smanjenja rizika koji oni nose, omogućavanja pretraživanja i pronalaženja prema pojedinim kriterijima i omogućavanja naknadne akcije nad objektima. Također, spomenute postupke moguće je automatizirati na temelju definiranih kriterija, s ciljem poštivanja propisanih procedura vezanih uz podatke i njihovu obradu.

Platforma IBM StoredIQ koristi tri sloja:

- **Data Server** – sloj koji se spaja na pojedine izvore podataka (podržano preko 75 platformi),
- **Gateway** – element koji povezuje jedan ili više Data Servera s aplikacijskim slojem sustava,
- **Application server** – na njemu se nalaze aplikacije koje omogućuju konfiguriranje i rad u aplikacijama nad podacima koji preko Gatewaya dolaze s Data Servera.

kakva im je vrijednost, zatim se složiti oko procedura i postupaka za njihovo tretiranje, i napokon provesti akcije nad njima. Za otkrivanje podataka prijeko je potreban kvalitetan alat, primjerice, IBM-ov StoredIQ. To je platforma koja omogućuje povezivanje na podatke (koji mogu biti pohranjeni na preko 75 različitih platformi) te kroz sustav filtera i ekstraktora dolazi do relevantnih informacija o samim podacima. Pomoću tog alata moguće je pretraživati sve raspoložive podatke prema metapodacima (datum nastanka, ime autora, format i slično), ali i prema sadržaju koji se nalazi unutar pojedine datoteke (pretraga prema definiranom *stringu* i slično).


### PRIMJER PRONALASKA PODATAKA

Na primjeru je prikazano kako IBM-ov StoredIQ vizualizira pretraživanje nad skupom od preko 60.000 datoteka. Na zadanim skupom podataka traže se datoteke u kojima se nalazi definirani *string* "Megatrend". Iz skupa podataka izbacuju se datoteke koje nisu zadovoljile kriterij, a brojčano su velike i maskiraju pravu sliku (Images, Miscellaneous, System files). Tako je ostalo samo 1.800 objekata koji se mapiraju u bojama, ovisno o njihovoj relevantnosti. Plavi su tipovi datoteka koji ne zadovoljavaju kriterij, a što je boja bliža crvenoj, to je viši postotak "pogodaka" u tom tipu datoteka (očito je da u ovom primjeru *e-mail* ima najveći postotak). Tako se vizualno i jednostavnim operacijama lako dolazi do vrlo malog skupa unutar 60.000 dokumenata koji zadovoljavaju kriterij.

Nakon toga je jednostavno utvrditi točno mjesto gdje se dokumenti nalaze, što je nužno za obavljanje bilo kakve akcije nad objektima. Tako je moguće pronaći ROT (redundant, obsolete, trival) podatke, koje je u pravilu najbolje obrisati. Primjeri trivijalnih podataka su razni *temporary*, privatni i logovi, koji su u većini slučajeva nepotrebni. Jasno, prilikom brisanja privatnih podataka (unatoč politikama tvrtke) dobro je biti oprezan jer će direktor teško prihvatiti racionalnost upravljanja podacima ako su mu obrisane fotografije s vjenčanja kćeri. Tu dobro dođe zrnice kreativnosti pa se sve privatne datoteke mogu premjestiti na posebni repozitorij, tako da korisnici misle da su obrisani, pa ako nitko u idućih šest mjeseci ne zatraži svoje privatne slike, onda ih se doista može izbrisati na tom drugom (tajnom) mjestu.

Kroz proces otkrivanja digitalnih podataka može se još mnogo toga učiniti za boljitak budućeg poslovanja: pronaći podatke koji

su vrijedni (klasificirati ih i dalje njima upravljati sukladno vrijednosti koju imaju za tvrtku), uočiti kršenje regulatornih obveza (tko ih krši, gdje se podaci nalaze, kako spriječiti kršenje u budućnosti), ukloniti ROT podatke, uskladiti upravljanje podacima s postojećim politikama tvrtke u spremanju podataka, smanjiti volumen podataka koji se koriste (i

njihov rast u budućnosti) te time uštedjeti na infrastrukturi i administraciji, pronaći relevantne podatke za pravne potrebe (prilikom tužbi i slično). Kad se jednom utvrdi postojeće stanje pojedine klase podataka, podatke treba migrirati na adekvatna mjesta, gdje će biti ispravno zaštićeni u smislu pristupa i ovlasti pojedinim rolama unutar tvrtke. 

## KOMPONENTE STOREDIQ-a

- **StoredIQ Platform Data Server** – Data Server služi za spajanje na različite izvore podataka, a kroz sučelje omogućuje se administriranje servera i konfiguriranje sustava,
- **StoredIQ Administrator** – upravlja instancama platforme kao što su volumeni, indeksi, konfiguracije i povezuje platformu s aplikacijama, ali i elementima aplikacije (infoset, akcije, destinacije i slično),
- **StoredIQ Data Workbench** – vizualizacija indeksiranih podataka i identificiranje kritičnih stanja u skladu sa zadanim kriterijima. Kroz ovo sučelje kreiraju se kriteriji i povećava preciznost u otkrivanju pojedinih tipova podataka,
- **StoredIQ eDiscovery** – aplikacija koja pomaže u otkrivanju, identifikaciji, očuvanju i prikupljanju podataka koji su relevantni za potrebe pravne službe. Zahtijeva manju razinu stručnosti i koristi modele kreirane kroz rad u Workbenchu,
- **StoredIQ Policy Manager** – provodi automatizirane akcije nad podacima, provodeći zadane politike organizacije,
- **StoredIQ Desktop Data Collector** – prikuplja podatke s računala

(indeksira ih kao zasebne volumene) kako bi bilo moguće analizirati i provoditi akcije nad podacima distribuiranima po računalima.

Više o IBM-ovoj platformi StoredIQ saznajte na [gdpr.megatrend.com](http://gdpr.megatrend.com), gdje možete zatražiti i demo probu usluge. ◀

