

Poslovanje protkano podacima

Svaka organizacija traži načine da otključa svoje podatke kako bi bila inovativna i ostala konkurentna. Većina se suočava sa značajnim izazovima prikupljanja, pohranjivanja i analize podataka, jer su podatkovna okruženja postala sve složenija

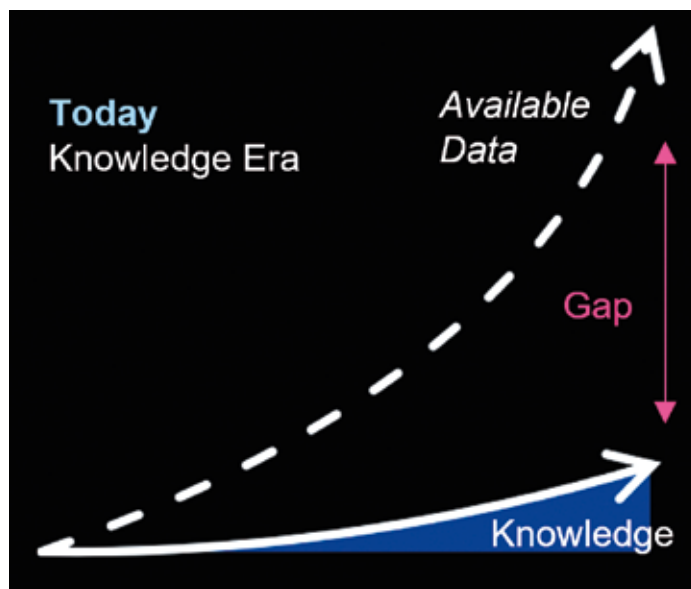
Mario Jurić

Složenost nastaje zbog nekoliko čimbenika. Podaci se nalaze na mnogo različitih mjestima, takozvanih podatkovnih silosa, zatim povećava se volumen podataka koji se generiraju i prikupljaju, potom – da bi podaci bili dostupni našim analitičkim procesima, često se dodatno kopiraju i transformiraju, i na kraju, što više podataka i izvora podataka imamo, to je složenije razumjeti, kako su podaci povezani, održavati visoku razinu kvalitete podataka.

JAZ IZMEĐU PODATAKA

Posljedica svega toga je jaz koji nastaje između podataka koji su nam dostupni i onih podataka koje razumijemo, odnosno koristimo s razumijevanjem. Bitno je raditi na smanjenju tog jaza, jer to može značiti veću prodaju, raniju detekciju prijevara, smanjenje troškova ili kreiranje novog inovativnog proizvoda.

Tvrtke koje to uspijevaju, s pravom se mogu nazivati *data driven* tvrtkama. Problem je što postizanje toga nije jednostavno, i na neki način postaje privilegij manjine koja ima snagu to ostvariti i koja može izgraditi talentirane i uspješne podatkovne timove. Jedan od načina kako tome pristupiti jest omogućiti da što više ljudi unutar organizacije ima pristup podacima i alatima potrebnima za rad s njima. Što više osoba koristi podatke i postavlja pitanja, generirat će se veći broj hipoteza. Dio tih hipoteza pokazat će se validnima, kao ideje



Jaz između podataka koji su nam dostupni i podataka koje koristimo s razumijevanjem

koje se mogu isprobati, a ako se pokažu korisnima, u konačnici mogu završiti i u upotrebi.

Drugi izazov koji se pojavljuje je problem integracije svih tih podataka. U prošlosti su organizacije pokušavale riješiti probleme pristupa podacima, bilo integracijom od točke do točke (point-to-point), bilo uvođenjem podatkovnih čvorišta (*data hub*). Niti jedno od njih nije prikladno kad su podaci visoko distribuirani i razdvojeni. Integracije od točke do točke dodaju veliki trošak za svaku dodatnu krajnju točku koja se treba povezati, što znači da taj pristup nije skalabilan. Podatkovna čvorišta vremenom postaju vrlo kompleksnima, i izazov je održavati svu logiku i sve podatke na jednom mjestu.

Za uspješnu borbu s tim iza-

zovima potrebno je izgraditi modernu podatkovnu infrastrukturu. "AI ljestvica" (AI ladder), koncept koji je razvio IBM, omogućuje modernizaciju informacijske arhitekture u četiri ključna koraka, pružajući pritom organizacijama razumijevanje gdje se nalaze na njihovom AI putu, i služi kao okvir za određivanje na koji korak se trebaju usredotočiti.

ČETIRI PODRUČJA

Kao vodeće načelo za organizacije u transformaciji poslovanja pruža četiri područja koja treba razmotriti:

Prikupljanje (collect): Učinite podatke jednostavnim i dostupnim – prikupljajte podatke svih vrsta, bez obzira na to gdje se oni nalaze, omogućujući flek-

sibilnost u odnosu na stalno mijenjajuće izvore podataka, istodobno smanjujući troškove, neučinkovitost i kvalitetu povezanu s tradicionalnim upravljanjem podacima.

Organizacija (organize):

Stvorite analitički temelj spreman za poslovanje – organizirajte sve podatke u pouzdanu poslovno spremnu podlogu s ugrađenim upravljanjem, zaštitom i usklađenošću (data governance), bez obzira na to gdje se podaci nalaze.

Analiziranje (analyze):

Gradite i skalirajte AI s povjerenjem transparentnošću – Analizirajte podatke na brže i pametnije načine, i iskoristite modele AI koji organizacijama omogućuju da steknu nove uvide i donesu bolje, pametnije odluke, bez tradicionalnih izazova i silosa.

Unošenje (infuse):

operacionalizirajte AI tijekom cijelog poslovanja – primijenite AI u cijeloj tvrtki, u više odjela i unutar različitih procesa, oslanjajući se na predviđanja, automatizaciju i optimizaciju koju podržava unificirana platforma.

Osim prolaska kroz četiri navedena koraka AI ljestvice, velika korist može se dobiti uspostavom *data fabric* arhitekture. To je napredni način upravljanja podacima, koji je usmjeren na podatke kako bi se omogućila dinamična i inteligentna orkestracija podataka u visoko distribuiranom podatkovnom krajoliku.

TKIVO PODATAKA

Primjenom *data fabric* arhitekture, koja je opisana u okviru uz tekst, izbjegava se prekomjerno kopiranje podataka i optimizira se način kako se podaci premještaju ili im se pristupa. Nekada ne možemo smanjiti količinu podataka ili složenost

DATA FABRIC

Data fabric zapravo je apstrakcijski sloj koji omogućuje korištenje i dijeljenje podataka duž cijelog podatkovnog krajolika, i može se logično podijeliti na četiri mogućnosti:

Znanje, uvidi i semantika

- Automatski obogaćuje otkrivena sredstva podataka znanjem i semantikom, omogućujući korisnicima da lakše pronađu i razumiju podatke

Jedinstveno upravljanje i usklađenost

- Omogućuje lokalno upravljanje metapodacima, ali podržava globalni jedinstveni pogled i provedbu politike
- Automatski primjenjuje pravila nad podacima
- Koristi napredne mogućnosti za automatsku klasifikaciju podataka

Inteligentna integracija

- Ubrzava rad podatkovnih inženjera kroz automatizirano kreiranje podatkovnih cjevovoda
- Omogućuje samoposlušno učitavanje i pristup svim podacima
- Automatski određuje najbolju način izvođenja upita, uz optimiziranu raspodjelu radnog opterećenja

Orkestracija i životni ciklus







- Omogućuje kreiranje, testiranje, izvršavanje i nadzor podatkovnih cjevovoda
- Unosi AI sposobnosti u životni ciklus podataka

podatkovne mreže, ali možemo koristiti tehnike tako da se podaci dohvate najboljom mogućom metodom. Također, tako se može uspostaviti visoko automatizirano i vođeno korisničko iskustvo.

Kako bismo sve to ostvarili, trebamo platformu, a unutar IBM-a platforma koja omogućuje uspostavu *data fabric* arhitekture zove se IBM Cloud Pak for Data. IBM Cloud Pak for Data sastoji se od bogatog skupa usluga za prikupljanja, organiziranja, upravljanja i analizu podataka. Najveća vrijednost tog sustava leži u činjenici da je kompletan set alata i servisa potrebnih za kompletan rad s podacima implementiran u jednu integriranu platformu, čineći troškove za hardversku infrastrukturu manjima, te smanjuje potrebno vrijeme i resurse za implementaciju i ukupno održavanje. U konačnici, primjenom *data fabric* arhitekture i IBM Cloud Pak for Data, može se omogućiti pristup bilo kome, bilo kojim podacima u bilo koje vrijeme, i na bilo kojem mjestu, naravno, ako je taj pristup autoriziran.



Data fabric arhitektura kao vezivno tkivo između različitih podatkovnih izvora

Data consumers / Business processes					
Unified lifecycle – Design – Compose – Build – Test – Orchestrate – Deploy – Operate – Monitor	Data self-service Exploration, collaboration, data marketplace, provisioning	Unified data governance, security, compliance, financial governance – Data policies – Data lineage – Data quality – Data privacy – Cost optimization			
	Smart integration Data ingestion, transformation, preparation, virtualization, real-time data				
	Augmented knowledge Discovery, catalog, semantics, ontologies, graph, metadata, master data, reference data, event-driven data				
IBM public cloud 	AWS 	Microsoft 	Google 	Edge 	Private 

Mogućnosti platforme IBM Cloud Pak for Data