

UPRAVLJANJE PODACIMA

1. ORGANIZACIJA I SKLADIŠTENJE PODATAKA U POSLOVNIM SISTEMIMA

2. ANALITIKA I PRETRAŽIVANJE PODATAKA

Prof. dr Jasna Soldić-Aleksić

1. ORGANIZACIJA I SKLADIŠTENJE PODATAKA

Imajući u vidu da **podaci predstavljaju resurs preduzeća**, neminovno se postavlja pitanja kako organizovati ovaj resurs i kako upravljati njime? Ali, pre nego što odgovorimo na ovo pitanje, važno je da sagledamo neke karakteristike ovog resursa sa sledećih aspekata: koji su izvori podataka, metode prikupljanja podataka, odrednice kvaliteta podataka, priroda (tip) podataka u pogledu struktuiranosti, organizacije i čuvanja podataka.

U pogledu izvora podataka u organizaciji, možemo razlikovati sledeće:

1. Unutrašnji podaci

- odnose se na ljude, usluge i procese;
- mogu se čuvati na jednom ili više mesta;
- mogu da budu dostupni i preko intraneta.

2. Lični podaci se odnose na podatke o pojedincima (ime, pol, školska sprema, zanimanje, i sl.), kao i različite percepcije, mišljenja, kritike i sl. (na primer, mišljenje o tome šta će učiniti konkurencija, procene prodaje i sl.).

3. Spoljašnji podaci se odnose na sledeće:

- Opšti izveštaji vlada i njenih tela (ministarstva, zavoda, biroa, agencija i sl.), izveštaji lokalne uprave, kao i mnogih nevladinog organizacija, udruženja građana i sl.
- Internet i komercijalni izvori podataka
 - Spoljašnji podaci pritiču u organizacije putem elektronske razmene podataka (EDI), ili kroz druge kanale između kompanija. Mnogo spoljašnjih podataka dostupno je preko ektraneta ili interneta.
 - Hiljade baza podataka širom sveta dostupno je preko interneta, pristup ovim podacima može biti besplatan, ali i komercijalan.
 - komercijalno onlajn izdavaštvo - onlajn izdavač baza podataka prodaje pristup specijalnim bazama podataka, novinama, časopisima i izveštajima. On može po razumnoj ceni da obezbedi podatke korisnicima.

Ovi podaci mogu da se čuvaju kao tekst, slike, dijagrami, atlas, filmovi, zvučne animacije i sl. na različitim nosiocima podataka – papiru, trakama, elektronskim nosiocima podataka - CD-ROM, DVD, USB, hard drive, na internet serverima.

Kako se podaci prikupljaju i kreiraju u poslovnom sistemu? Metode prikupljanja podataka mogu biti:

- Ručno prikupljanje podataka i ručno unošenje podataka na različite nosioce podataka;
- Elektronski unos podataka u elektronske obrasce,
- Elektronski prenos podataka iz registar kasa, pasivno prikupljanje podataka (engl. *machine to machine transactions*), Elektronska razmena podataka (EDI) i sl.;
- Pomoću instrumenata (kamera, različitih čitača, na primer, RFID tehnologija), senzora;
- Podaci “koji se dobijaju na pritisak tastera miša” “*Clickstream data*”.

Na primer, u marketingu posebna oblast – Analiza *clickstream* podataka.

Bez obzira kako se podaci prikupljaju, bitna odrednica prikupljenih – raspoloživih podataka jeste **kvalitet podataka**. Kvalitet podataka određuje njihovu korisnost u smislu poslovne vrednosti koja se kreira upotrebom tih podataka, kao i kvaliteta poslovnih odluka koje se zasnivaju na podacima. Problemi sa kojima se organizacije mogu suočiti su: greške u podacima, dupli podaci (koji se najčešće razlikuju), zastareli (ne ažurni) podaci, kompromitovani podaci (problem zaštite), nedostajući podaci i dr. U literaturi nailazimo na termin „prljavi“ podaci (engl. **dirty data**). Pod ovim terminom podrazumevaju se podaci koji sadrže neku pogrešnu informaciju. Na primer, sledeći podaci se mogu smatrati da pripadaju ovoj grupi¹:

- Pogrešni podaci,
- Dupli podaci,
- Netačni podaci,
- Ne-integrirani podaci,
- Podaci koji narušavaju poslovna pravila,
- Podaci neodgovarajućeg formata,
- Pogrešno zapisani (spelovani) podaci.

Kada govorimo o kvalitetu podataka možemo izdvojiti nekoliko bitnih dimenzija kvaliteta podataka:

1. Suštinski kvalitet podataka povezan sa **sadržajem** podataka
 - Tačnost, objektivnost, verodostojnost, integritet, ažurnost i reputacija.
2. **Dostupnost** podataka
 - Dostupnost i sigurnost pristupa.
3. **Kontekstualni okvir** podataka
 - Relevantnost, konzistentnost, potpunost, kredibilitet, dodata vrednost, vremenska neograničenost.
4. **Prikazivanje** podataka
 - Mogućnost tumačenja, lakoća razumevanja, sažetost predstavljanja, dosledno prikazivanje.

Važan princip koji se odnosi na kvalitet podataka – **GIGO princip** (engl. “*Garbage In, Garbage Out*” – *GIGO princip*). Smatra se da je termin prvi put upotrebio IBM programmer i instruktor George Fuechsel². Suština ovog principa je da loš kvalitet input vrednosti dovodi do lošeg kvaliteta izlaznih rezultata. Odnosno, drugim rečima **kvalitet izlaznih rezultata nekog sistema ne može biti veći od kvaliteta ulaznih podataka**. To nam govori da je kvalitet podataka od suštinske važnosti za bilo kakvu upotrebu podataka. Sam princip je baziran na osnovama filozofije upravljanje kvalitetom (engl. *quality management*), a nalazi primenu u mnogim oblastima, kao što su: planiranje, proces donošenja odluka, finansijski menadžment, finansijsko modeliranje, upravljanje proizvodnjom, prediktivno modeliranje, simulacije i u svim drugim oblastima analitike podataka i podrške poslovnom odlučivanju.

U nastavku su prikazani osnovni elementi organizovanog pristupa pitanjima upravljanja podacima.

¹ <https://www.techopedia.com/definition/1194/dirty-data>

² <https://www.techopedia.com/definition/3801/garbage-in-garbage-out-gigo>

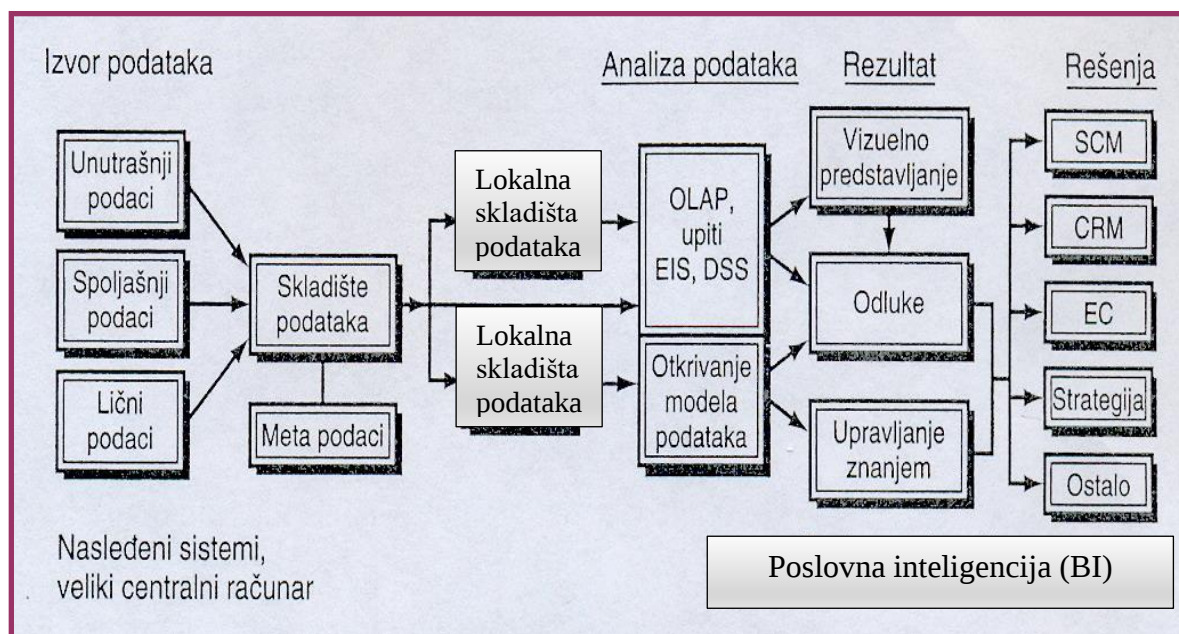
1.1. Upravljanje podacima (eng. *Data management*)

Upravljanje podacima danas predstavlja jedan od najvećih poslovnih izazova za organizacije. Moderne kompanije kreiraju i prikupljaju velike količine podataka, koje je potrebno adekvatno skladištiti, organizovati i koristiti. Stoga je pitanje upravljanja podacima jedno od ključnih pitanja kako za poslovne sisteme u privredi, tako i za organizacije i institucije u sektoru vanprivrede.

Upravljanje podacima se odnosi na upravljanje toka podataka od kreiranja podataka i inicijalnog skladištenja sve do faze kada podaci postaju zastareli i brišu se.

Upravljanje podacima je *struktuirani pristup* za efektivno prikupljanje, skladištenje, obradu, integrisanje, distribuciju, zaštitu i arhiviranje podataka u toku životnog ciklusa podataka.

Na sledećoj slici je prikazan životni ciklus podataka. Životni ciklus podataka pokazuje način na koji podaci „putuju“ kroz organizaciju, od njihovog inicijalnog prikupljanja i kreiranja sve do njihovog korišćenja kao podrške za donošenje poslovnih odluka u okviru: upravljanja odnosom sa potrošačima - SCM, upravljanja lancem snabdevanja - CRM, elektronske trgovine i dr.



Slika 1. Životni ciklus podataka³

Neposredno povezano sa životnim ciklusom podataka izdvajaju se tri principa upravljanja podacima. Reč je o sledećim principima:

³ Na osnovu Turban E., Leidner D., McLean E., Wetherbe J., (2006), Information Technology and Management, 5th ed. John Wiley & Sons, Inc, USA, str. 410

1. **Princip opadajuće vrednosti podataka** – ovo je jedan veoma jednostavan, ali moćan princip (intuitivno lako prihvatljiv): što su „svežiji“ podaci utoliko imaju veću vrednost i obrnuto. Dakle, sa protokom vremena opada vrednost podataka.

2. **Princip “90/90” u korišćenju podataka** – prema ovom principu ogromnoj većini podataka (oko 90% podataka) pristup se dešava u okviru 90 dana (izuzetak su podaci koji su važni za potrebe revizije). Ovaj princip sugerise da većina podataka gubi značajan deo svoje vrednosti nakon 90 dana.

3. **Princip korišćenja podataka u određenom kontekstu** – da bi podaci imali svrsishodnu vrednost, oni moraju biti integrisani, obrađeni, analizirani i formatirani u smislenom formatu, tj. kontekstu - “akcione informacije”. To znači da krajnji korisnici moraju da vide podatke u smislenom formatu i kontekstu, ako se očekuje da ti podaci utiču na njihove odluke i planove.

Generalno se može reći da je **glavni cilj upravljanja podacima** da se pomogne poslovnim sistemima da unaprede svoju produktivnost, time što će da omoguće poslovnim akterima kojima su potrebni određeni podaci da ih lako nađu, bez dugih odlaganja i napornog pretraživanja. Ili, drugim rečima, reč je o **disciplini koja treba da obezbedi infrastrukturu i alate koji će omogućiti poslovnom sistemu da transformiše sirove podatke u informacije i znanje koji uvećavaju i/ili stvaraju novu poslovnu vrednost – *business value*** (podsetimo se uloge različitih nivoa u hijerarhiji znanja u generisanju poslovne vrednosti).



Slika 2. Hijerarhija znanja

Specifični ciljevi upravljanja podacima su sledeći:

- Zadovoljenje pravnih zahteva (zakona i drugih pravnih normativa), koje poslovni sistem treba da ispuni;
- Umanjenje rizika i troškova vezanih za ispunjenje pravne regulative;
- Obezbeđenje **sigurnosti** podataka;
- Održavanje **tačnosti i raspoloživosti** podataka;

- Obezbeđenje **konzistentnosti** podataka koji dolaze iz različitih izvora;
- Obezbeđenje usaglašenosti podataka sa najboljom praksom pristupa, skladištenja, čuvanja, raspolaganja i održavanja podataka.

Nosioci funkcija u procesu upravljanja podacima suočavaju se sa mnogim **izazovima**, među kojima se mogu izdvojiti sledeći:

1. Količina podataka se povećava **eksponencijalno** (fenomen *Big data*);
2. Podaci su **rasuti širom organizacija**, a prikupljaju ih pojedinci različitim metodama;
3. Podaci se čuvaju na **nekoliko servera i lokacija** u organizaciji – u različitim računarskim sistemima, bazama podataka, formatima, aplikacija i sl.;
4. Različite aplikacije koje pristupaju podacima na različite načine;
5. Razmatranje i uključanje **spoljašnjih podataka** (podataka iz okruženja) koji su neophodni za donošenje poslovnih odluka;
6. Najčešće je za donošenje konkretnih poslovnih odluka bitan **mali deo podataka – kako ih izdvojiti?**
7. **Bezbednost, kvalitet podataka i integritet podataka su kritični**, i lako mogu da se ugroze;
8. **Čuvanje podataka** na različite načine, pripremanje odgovarajućih **back-up i data-recovery** alata.
9. **Privatnost i etička pitanja** prikupljanja i obrade podataka;
10. **Zakonska rešenja** koja se odnose na podatke najčešće se **razlikuju** od jedne do druge zemlje i podložna su čestim promenama (za zemlje Evropske unije važi Opšta direktiva za zaštitu podataka -*General Data Protection Regulation* GDPR, koja je stupila na snagu 25. 05. 2018);
11. **Biranje alata** za upravljanje podacima je takođe **značajan izazov**.

Upravljanje ključnim podacima - *master data mangement* (MDM)

U ICT oblasti jedna od solucija za konstruktivno rešenje prethodnih izazova jeste primena **koncepta ključnih podataka - master data** i upravljanja ovim podacima - **master data management** (MDM).

Termin ključni podaci (engl. *master data*) se koristi da opiše bitne informacije o različitim entitetima u poslovnoj organizaciji, koje su bitne za obavljanje poslovnih operacija i transakcija. Pod **entiteom** (engl. *data entity*) se podrazumeva **svaki realan ili apstraktan pojam** o kome organizacija želi da prikuplja i čuva podatke. Naime, svaka organizacija ima podatke o svojim proizvodima, uslugama, kupcima, dobavljačima, lokacijama, fizičkoj imovini, zaposlenima i sl., ali ovi podaci se retko kada čuvaju na jednoj lokaciji. Najčešće su rasuti u organizaciji u raznim aplikacijama, radnim tabelama, ili nekim papirnim dokumentima. Nije redak slučaj da

u jednoj organizaciji u različitim poslovnim sektorima postoje različite definicije za jedan isti poslovni entitet ili poslovni odnos. Na primer, podaci o jednom istom zaposlenom mogu biti sačuvani na različite načine u sektoru ljudskih resursa, u finansijama – obračun zarada, u operativnom menadžmentu i sl. Ovim pitanjima se bavi upravljanje glavnim podacima - MDM. **Pod MDM se podrazumeva skup disciplina i metoda koji ima za cilj da osigura aktuelnost, kvalitet, značenje i upotrebu bitnih podataka kompanije u različitim njenim oblastima.** Naime, MDM povezuje i sinhronizuje sve krirične podatke iz različitih oblasti i sistema u jedan dokument (engl. *master file*), koji predstavlja zajedničku referentnu tačku (engl. *common point of reference*). Kako većina MDM opcija može biti veoma kompleksna i skupa, čitav ovaj koncept u mnogim slučajevima nije dostupan malim i srednjim preduzećima. U tom smislu važno je pojavljivanje provajdera koji nude MDM usluge u oblaku (engl. *cloud-based MDM service*). Na primer, kompanija Dell Software je ponudila Dell Boomi MDM, koji obezbeđuje usluge upravljanja ključnim podacima, upravljanje podacima i usluge upravljanja kvalitetom podataka zasnovane na cloud platformi.

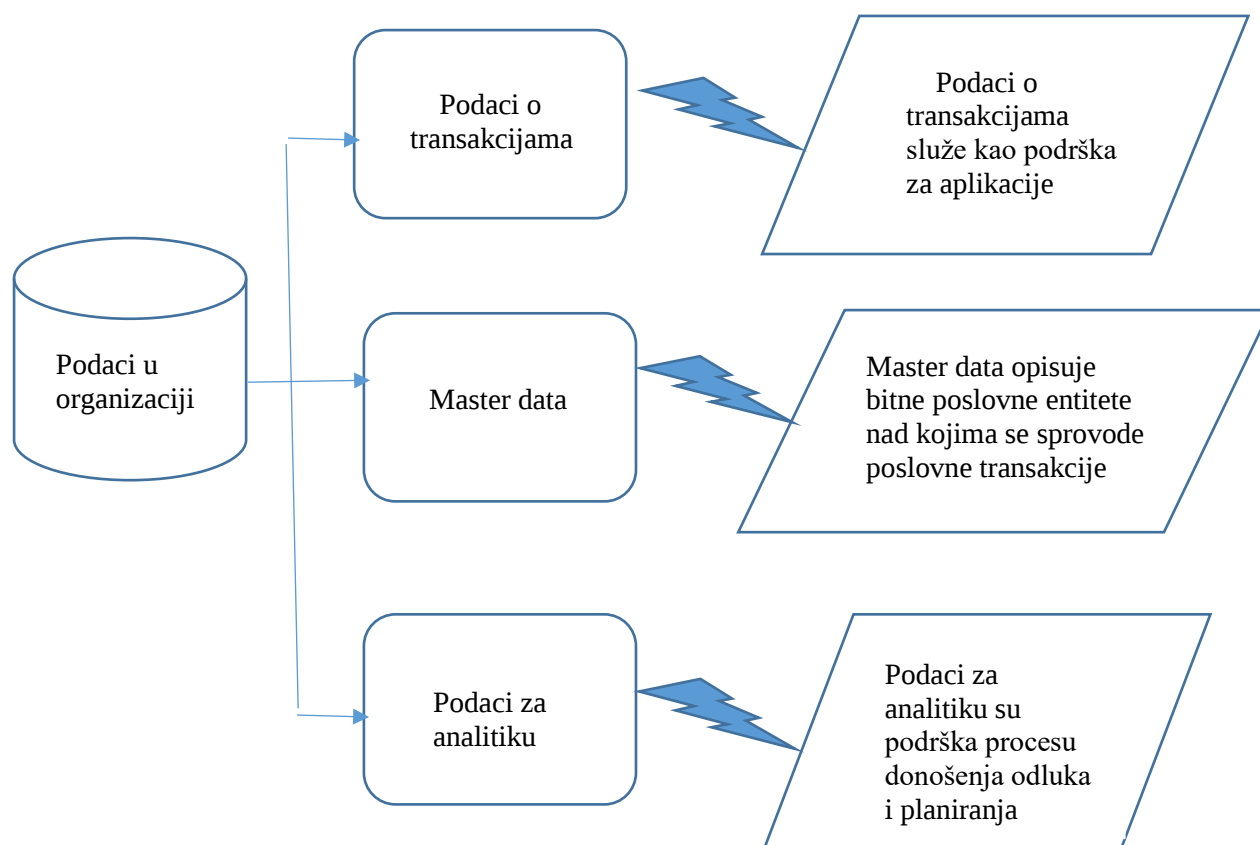
Posledice slabog upravljanja ključnim podacima mogu biti:

- Dupli podaci koji dovode do postojanja tzv. **silosa podataka** (engl. *Data silos*);
- Nekonzistentnost u značenju i nivoima detalja u prikazu pojedinih elemenata podataka;
- Korisnici nemaju poverenja u podatke i gube mnogo vremena u postupku validacije podataka, umesto da se koncentrišu na njihovu analizu radi donošenja odgovarajućih poslovnih odluka;
- Netačna analitika podataka;
- Donošenje odluka na osnovu percepcija, a ne na osnovu realnih činjenica;
- Povećan obim posla u vezi podataka – njihova validacija, kao i vreme obrade.

Može se reći da se podaci u poslovnom sistemu mogu generalno koristiti na dva načina:

za odvijanje i vođenje poslovanja – transakciona ili operativna upotreba podataka
i
za unapređenje poslovanja – analitička upotreba.

Imajući to u vidu, na sledećoj slici se može videti kako ključni podaci (engl. *master data*) mogu da posluže kao srednji sloj između transakcionih podataka (koji se čuvaju u, na primer, bazama podataka) i podataka za analitiku (u skladištima podataka).



Slika 3. Organizacija podataka u poslovnoj organizaciji⁴

Upravljanje podacima se sprovodi u skladu sa pripremljenom Strategijom podataka (*Data Strategy*) na nivou poslovnog sistema. A strategija podataka mora biti u skladu sa poslovnom strategijom organizacije (*Business strategy*).

Promenjen ekosistem podataka

U poslednjih nekoliko godina značajno se promenila priroda podataka u poslovnim sistemima, odnosno može se reći da je *ekosistem podataka promenjen i da je postao veoma složen*. Šta se pod time podrazumeva?

U poslovnim sistemima postoje različiti tipovi podataka u pogledu njihove struktuiranosti, načina skladištenja i brzine kojom se skladište i obrađuju. U tom smislu razlikujemo sledeće vrste podataka:

⁴ Na osnovu: Turban, E., Pollard, C., Wood, G. (2018), *Information Technology for Management, On-Demand Strategies for Performance, Growth and Sustainability*, 11th ed. John Wiley & Sons, Inc. , str. 78

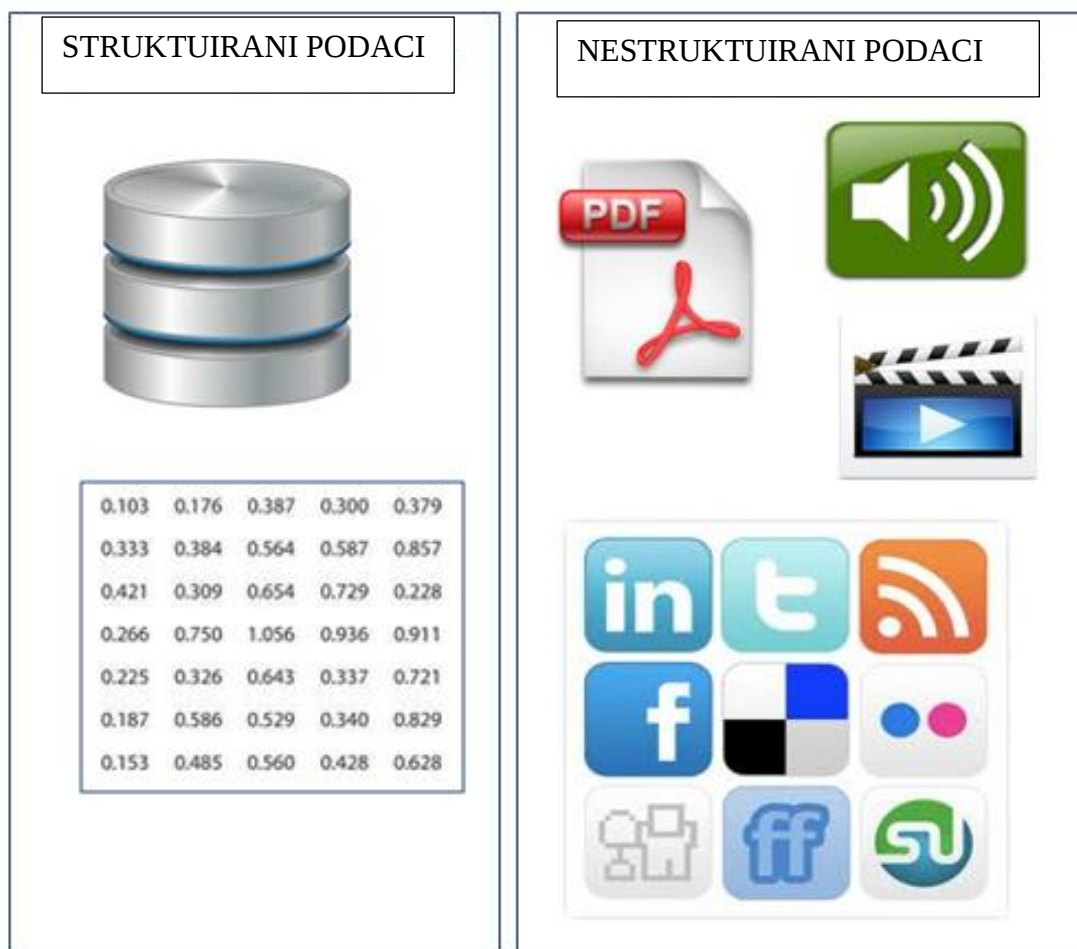
- **Strukturirani** versus **nestrukturirani** podaci;
- **On-premises** versus **u cloud-u** — predviđanje da će do 2022. godine više od 1/3 podataka prolaziti kroz *cloud*;
- **Podaci u mestu (*at rest*)** versus **podaci u pokretu (*in motion*)** — sve više se generišu podaci u kretanju (podaci koji potiču od senzora, mobilnih uređaja, transakcija, raznih mašina itd.).

U pogledu strukturiranosti podaci u poslovnim sistemima mogu biti:

- **Strukturirani podaci:** imaju strogo određenu formu, obično se radi o podacima koji su uneti u datoteke i baze podataka – tabelarni podaci (Dbase, Excel) čija je bitna karakteristika da se mogu lako pretraživati. Na primer reč je o podacima o prodaji, poslovnim transakcijama, poslovnim izveštajima, rezultatima kontrola i dr.
- **Polustrukturirani podaci:** e-mails, Log files, Markup language XML, CSV – *comma separated values fajl*, EDI, objektno orijentisane baze podataka, Open standard JSON (JavaScript Object Notation), NoSQL. Ovi podaci čine 5% -10% ukupnog obima podataka u poslovnim sistemima. Postoje izvesne informacije o podacima, kao što su metapodaci ili tagovi.
- **Nestrukturirani podaci:** tekst, slika, karte, dijagrami, planovi, audio zapisi, video zapisi i sl.

Opšta je procena da polustrukturirani i nestrukturirani podaci čine preko **70%** svih podataka kojima raspolaže jedan poslovni sistem. IBM analitičari su dali procenu da čak 85% podataka u poslovnom sistemu čine nestrukturirani podaci, a da su u okviru njih 50% duplikati. Takođe, njihova je procena da su stope rasta količine nestrukturiranih podataka izuzetno visoke: 55%-65% godišnje⁵.

⁵ <https://www.datamation.com/big-data/structured-vs-unstructured-data.html>



Slika 4. Struktuirani i nestruktuirani podaci u poslovnom sistemu⁶

U sledećoj tabeli prikazane su razlike između struktuiranih i nestruktuiranih podataka.

Tabela 1. Razlika između struktuiranih i nestruktuiranih podataka

	Struktuirani podaci	Nestruktuirani podaci*
Karakteristike	Prethodno definisani formati i modeli podataka, Obično su numerici i tekst, Laki za pretraživanje.	Ne postoje prethodno definisani formati i modeli, Tekst, slika, zvuk, video i sl. Teški za pretraživanje
Skladište se u:	Relacione baze podataka Skladišta podataka	NoSQL baze Aplikacije skladišta podataka Jezera podataka

6

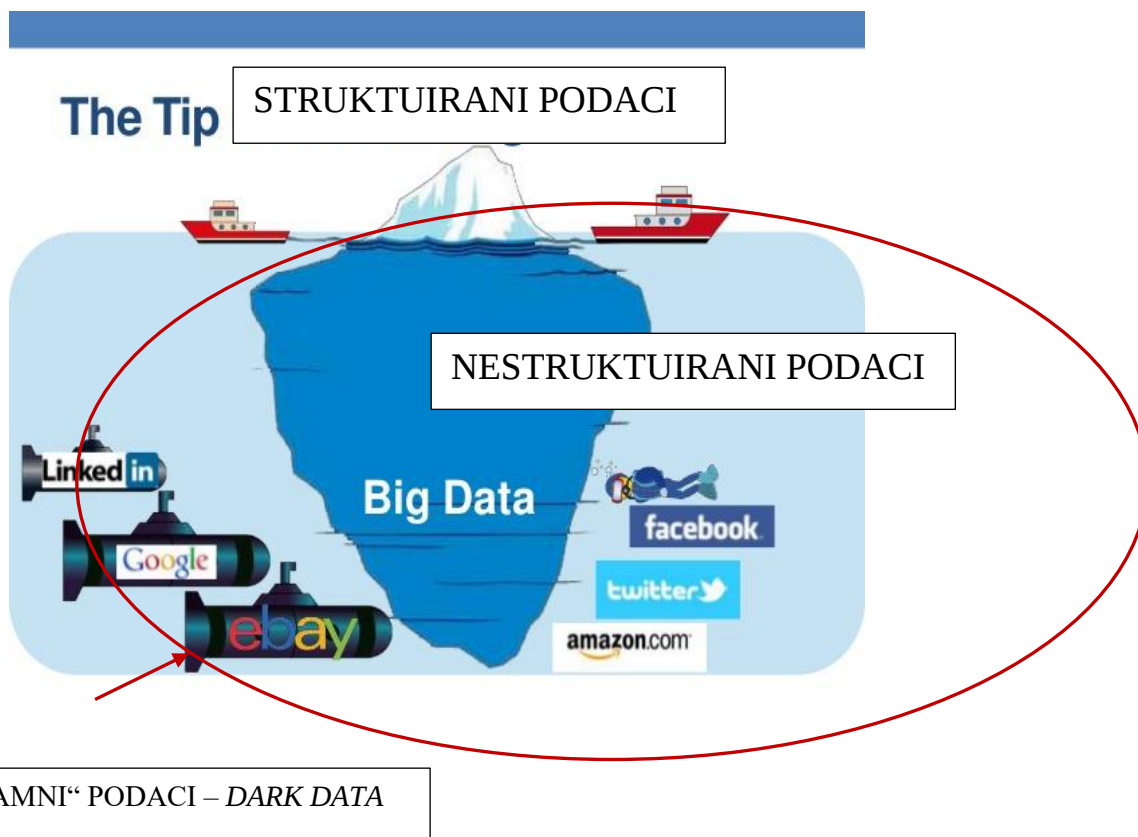
<https://www.bing.com/search?q=structured+and+unstructured+data&form=PREXEN&mkt=en-us&httpsmsn=1&msnews=1&refig=ea3515fbfb34e64a2e05c5a6c7e3da4&sp=1&q=SC&pq=structured+and+nonstructured+data&sk=PREs1&sc=833&cvid=ea3515fbfb34e64a2e05c5a6c7e3da4>

Generiše ih	Čovek ili mašina	Čovek ili mašina
Tipične aplikacije	Kontrola zaliha Obračun plata Rezervacija avionskih karata ERP sistem SCM sistem	Tekst procesori Softver za prezentacije e-mail klijenti alati za pregled i/ili editovanje medija
Primeri	Ime potrošača Adresa Datumske veličine Broj telefona JMBG Broj kreditne kartice Ime proizvoda/usluge Broj transakcije	Tekstualna dokumenta Planovi Izveštaji e-mail poruke Slike Audio dokumenta Video dokumenta

*Napomena: obuhvaćeni su polustrukturirani i nestrukturirani podaci

Izvor: <https://www.datamation.com/big-data/structured-vs-unstructured-data.html>

U organizacijama za poslovne svrhe uglavnom se koriste strukturirani podaci, a znatno manjim delom nestrukturirani podaci. Obično se dešava da su nestrukturirani podaci neiskorišćeni, pa se još i nazivaju „tamni“ podaci (engl. *dark data*).



Slika 5. Ilustracija strukturiranih i nestrukturiranih podataka

Pod terminom „**tamni**“ **podaci** (engl. *dark data*) podrazumevaju se podaci koji su skladišteni u poslovnom informacionom sistemu, ali **koji se ne koriste u poslovne svrhe** (digitalna informacija koja se ne koristi). Prema Gartner definiciji (jedna od prvih definicija) pod ovim terminom se podrazumeva sledeće:

Tamni podaci (Dark data) su informacioni resurs koji organizacija prikuplja, obrađuje i skladišti tokom redovnih poslovnih aktivnosti, ali ih **ne koristi** za druge poslovne svrhe, kao što su analitika, uspostavljanje poslovnih odnosa i direktnu monetizaciju (Gartner def.).

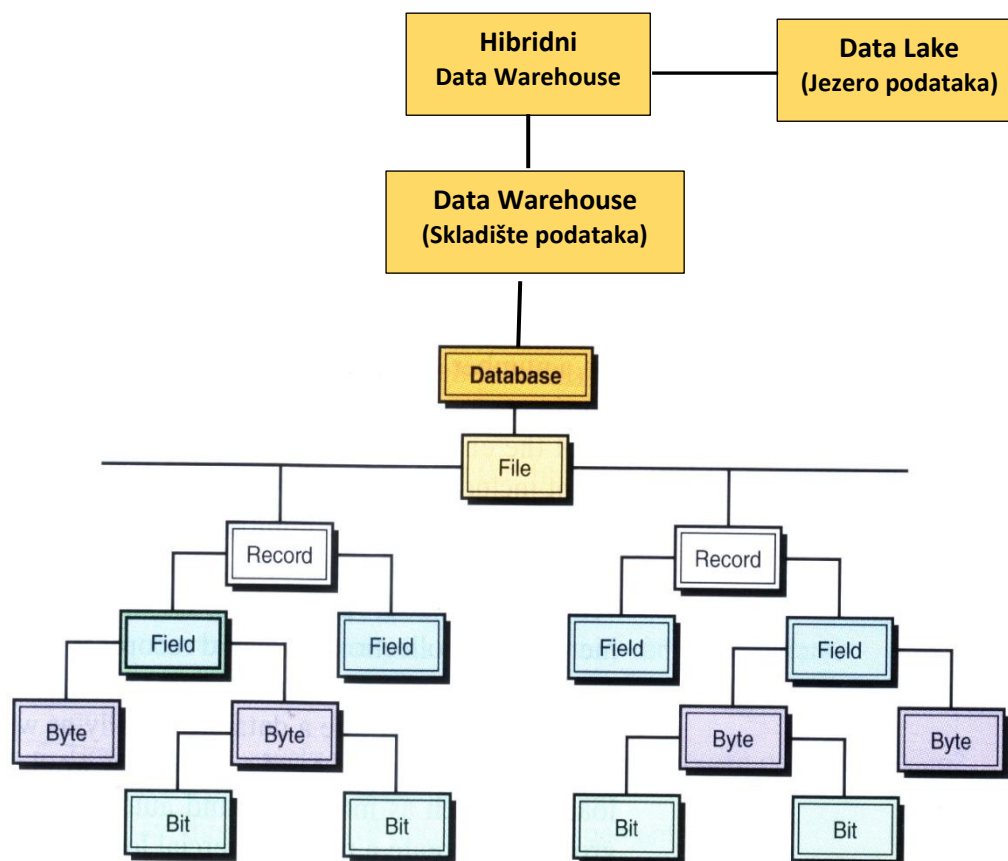
U nekim situacijama donosioci poslovnih odluka možda i nisu svesni koji sve podaci postoje, a koji bi mogli da se koriste. Procene stručnjaka IBM-a su da se 90% podataka koji su generisani pomoću senzora i koji se prevode iz analognog u digitalni oblik (engl. *analog-to-digital conversions*) nikada se ne koriste⁷. Takođe, procena je da većina organizacija za potrebe analitike koristi samo oko 1% svojih podataka⁸. Problemi vezani za prisustvo *dark data* posebno dolaze do izražaja danas, kada živimo u vremenu fenomena *Big data* i rastuće potrebe za analitikom podataka, o čemu će kasnije biti reči.

1.2. Organizacija podataka u poslovnom sistemu

Organizacija podataka u poslovnom sistemu može se uglavnom predstaviti u hijerarhijskom obliku. Jedna takva hijerarhija je prikazana na sledećoj slici.

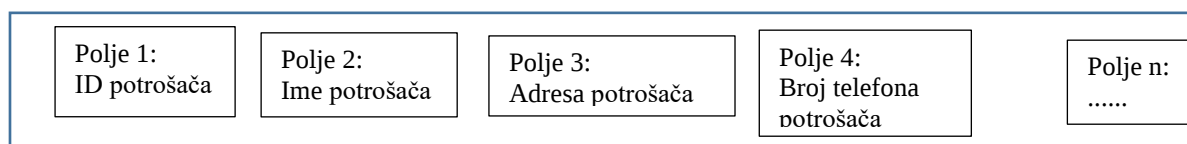
⁷ <https://siliconangle.com/2015/10/30/ibm-is-at-the-forefront-of-insight-economy-ibminsight/>

⁸ <https://www.ibmbigdatahub.com/blog/big-data-challenge-transformation-manufacturing-industry>



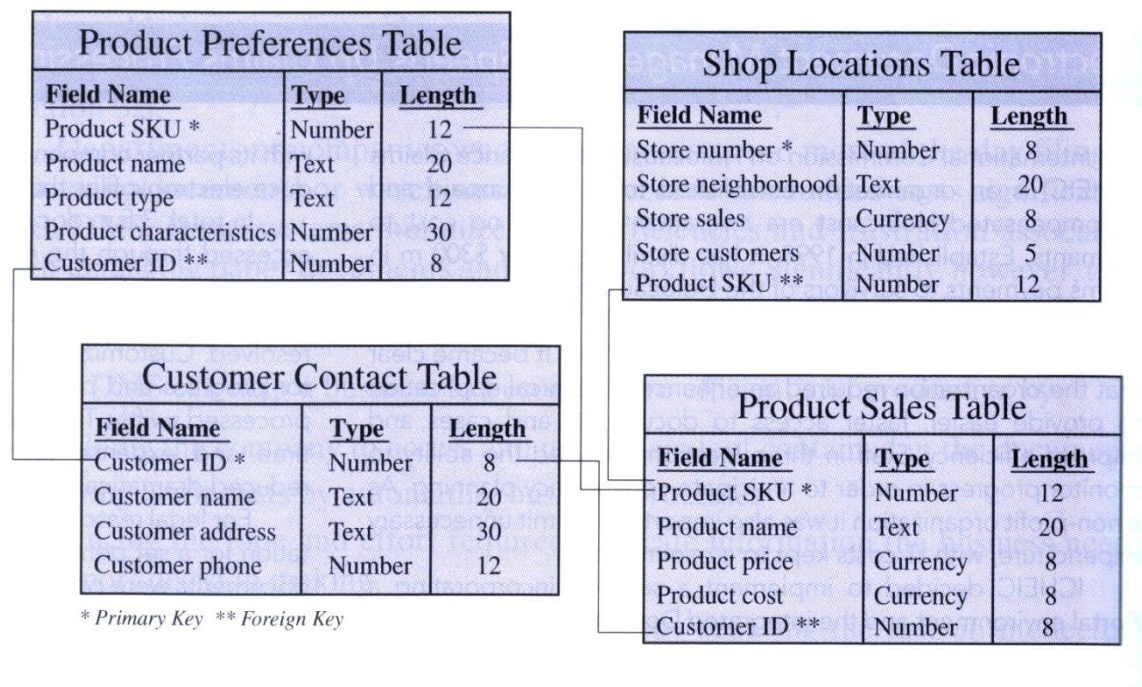
Slika 6. Hijerarhija organizacije skladištenja podataka

Bit predstavlja najmanju jedinicu podataka koju računar može da obradi, a koji uzima vrednost 0 ili 1. Niz od 8 bitova čini bajt (*byte*) koji predstavlja jedan karakter (znak) koji može biti slovo, broj ili neki simbol. Više karaktera formira polje (*field*): na primer polje *Ime_proizvoda*, *Tip_proizvoda*, *ID_potrosaća*, itd., a više polja čini slog (*record*). Jedan slog može da sadrži, na primer, sledeća polja: *ID_potrosaća*, *Ime_potrosaća*, *Adresa_potrosaća*, *Broj_telefona_potrosaća*, itd. Svaki slog prikazuje jedan *entitet*, a polja sloga prikazuju karakteristike, odnosno, *attribute* entiteta.



Slika 7. Struktura sloga

Ako se nastavi dalje (prema vrhu hijerarhije), više slogova formira datoteku (*data file*), a više logički povezanih datoteka formira bazu podataka. Jedna organizacija relacione baze podataka može se videti na sledećoj slici:



Slika 8. Organizacija relacione baze podataka

O strukturi i organizaciji skladišta podataka biće reči nešto kasnije u ovom poglavlju.

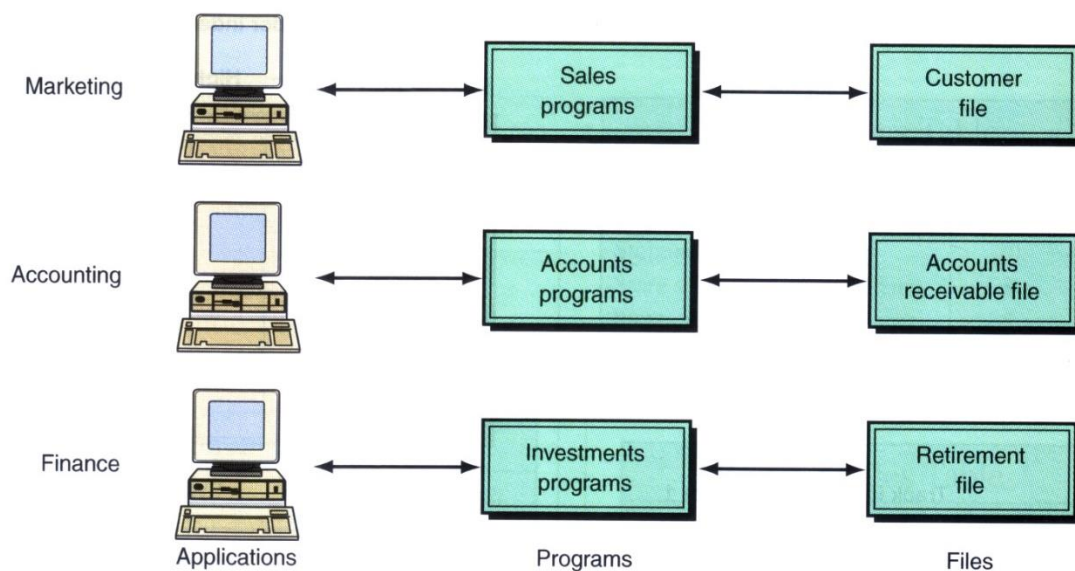
U nastavku ćemo ilustrovati kako se upravlja sistemom datoteka, odnosno bazom podataka i koje su prednosti sistema upravljanja bazom podataka?

Na sledećim slikama dati su uporedni prikazi

- **sistema za upravljanje dokumentima (*File Management Systems*) i**
- **sistemi za upravljanje bazom podataka (*Database Management Systems - DBMS*).**

U vreme kada su organizacije počele da koriste računare za automatizaciju svojih poslovnih procesa dominantno je bilo okruženje sa posebnim datotekama i aplikacijama. To je praktično značilo da je svaka aplikacija dizajnirana da bude samostalna aplikacija koja radi nezavisno od drugih aplikacija i radi nad „svojim“ podacima odnosno datotekama. Na primer, služba marketinga koristi svoju aplikaciju, koja koristi podatke o potrošačima smeštenim u marketinškoj datoteci o potrošačima. Druga služba, na primer operativni menadžment, koristi takođe „svoju“ datoteku podataka o potrošačima. Ovakav pristup ima niz nedostataka, kao što su:

- **redundatnost** podataka – nepotrebno dupliranje istih podataka u različitim datotekama;
- **nekonzistentnost** podataka – mogućnost da postoje različiti podaci o istom entitetu u različitim datotekama. Na primer, pretpostavimo da jedna finansijska institucija ima klijente sa nekoliko različitih kreditnih zaduženja i da za svako zaduženje ima podatke o klijetima u posebnoj datoteci. Ako dođe do, na primer, promene adrese klijenta u jednoj datoteci, to znači da će ovaj podatak biti nekonzistentan sa podatkom o adresi u drugim datotekama;
- **izolacija** podataka – organizacija podataka po datotekama dovodi do kreiranja tzv. **silosa podataka**, što dovodi do toga da je izuzetno teško da se pristupi podacima iz različitih aplikacija;
- **nedostatak integriteta** podataka – u okruženju sa datotekama mnogo je teže da se ugrade i primene pravila o integritetu podataka, koja, na primer, mogu da spreče da se unose podaci sa greškom;
- **problem sa sigurnošću** – u sistemu rada sa datotekama, javlja se veliki broj aplikacija na ad hoc bazi, što znači da veliki broj ljudi može da pristupi podacima u datoteci, čime se povećava rizik zloupotrebe podataka.

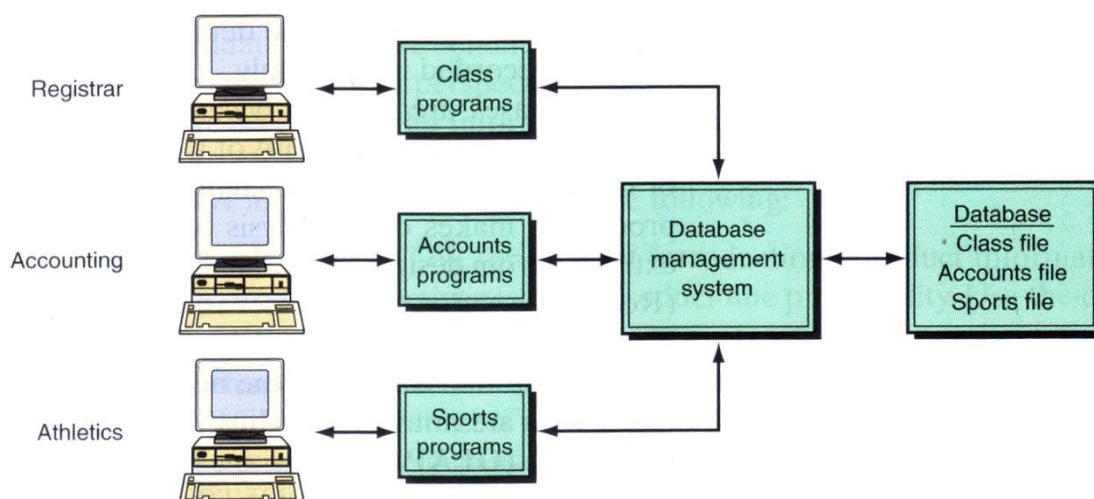


Slika 9. Sistem za upravljanje datotekama

Sistemi za upravljanje bazama podataka otklanjaju sve prethodno navedene probleme u radu sa datotekama i zasebnim aplikacijama. Primenom ovih sistema minimizirani su problemi redundatnosti podataka, izolacije podataka i nekonzistentnosti podataka. Podacima mogu da

pristupe različite aplikacije, pri čemu su aplikacije nezavisne od podataka koje obrađuju. Takođe, u okruženju baza podataka lakše je kontrolisati integritet i sigurnost podataka u odnosu na okruženje rada sa datotekama.

Program koji obezbeđuje pristup bazi podataka jeste sistem za upravljanje bazom podataka - **Database Management System (DBMS)**. DBMS omogućava organizaciji da centralizuje podatke, da efikasno upravlja podacima i omogući različitim aplikacionim programima da pristupe podacima. Ovaj sistem predstavlja vezu (*interface*) između aplikacionih programa i datoteka sa podacima, što se može videti na sledećoj slici. Može se postaviti pitanje kako može jedna baza podataka da odgovori zahtevima većeg broja korisnika? Na primer, kako može biti struktuirana jedna baza podataka a da omogući osoblju prodaje da vidi potrošače, zalihe i sl., dok osoblje ljudskih resursa pristupa ličnim podacima zaposlenih? Odgovor na ovo pitanje je sledeći: DBMS obezbeđuje dva „pogleda“ na podatke: **fizički pogled i logički pogled**. Fizički pogled, odnosno pristup odnosi se na fizičku organizaciju i lokaciju podataka na uređajima na kojima se čuva baza podataka. Sa druge strane, za korisnike nije bitan fizički aspekt čuvanja i organizacije baze podataka, već je za njih bitno da mogu da koriste potrebne podatke. Ovde je bitna uloga DBMS, jer on obezbeđuje mnogobrojne logičke poglede na podatke. Naime, **korisnicima je bitno da vide podatke sa stanovišta poslovne perspektive, a ne sa stanovišta tehničke perspektive**. Jasno je da logički, odnosno korisnički pogled na podatke može da varira, dok fizički pogled, tj fizičko skladištenje i čuvanje podataka je fiksno.



Slika 10. Sistem za upravljanje bazom podataka (*Database Management Systems - DBMS*)

DBMS ima sledeće funkcionalnosti:

- **Filtriranje podataka** (*data filtering and profiling*) – provera podataka u smislu grešaka, nekonzistentnosti, redundantnosti i nekompletnosti;
- **Obezbeđenje kvaliteta podataka** (*data quality*) – korekcija, standardizacija i verifikacija integriteta podataka;
- **Sinhronizacija podataka** (*data synchronization*) – integrisanje, uparivanje ili povezivanje podataka iz različitih izvora;
- **Obogaćivanje podataka** (*data enrichment*) – proširenje podataka upotrebom informacija iz unutrašnjih i spoljašnjih izvora;
- **Održavanje podataka** (*data maintenance*) – provera i kontrolisanje integriteta podataka tokom vremena.

Na osnovu prethodnog mogu se ukratko prikazati mogućnosti koje pruža DBMS-a.

- **Stalnost** (*Permanence*) – atributi su permanentno sačuvani na hard disku ili nekom drugom mediju, sve dok se eksplicitno ne promene ili izbrišu;
- **Pretraživanje** (*Querying*) – proces traženja određenih informacija atributa iz različitih perspektiva;
- **Istovremeni pristup** (*Concurrency*) – DBMS obezbeđuje niz alata i tehnika da se reguliše istovremeni pristup pojedinim poljima.
- **Pravljenje kopija** (*Backup and replications*) – DBMS obično obezbeđuje alate za pravljenje kopija i odgovarajućih backup-ova.
- **Uključenje pravila** (*Rule enforcement*) – primena pravila vezana za attribute kako bi se obezbedilo da oni ostanu verodostojni i čisti.
- **Sigurnost** (*Security*) – postavljanje ograničenja ko može da vidi ili promeni vrednost atributa;
- **Računske operacije** (*Computation*) – mnoge računске operacije može da obavi DBMS;
- **Praćenje promena** (*Change and Access logging*) – DBMS ima funkcionalnosti da prati i beleži promene nad atributima (ko, gde i kada je izvršio promenu);
- **Automatska optimizacija** (*Automated optimization*) – mnogi DBMS imaju mogućnost da prilagode svoje funkcionisanje u skladu sa učestalim zahtevima.

1.3. Skladište podataka – Data Warehouse i drugi načini skladištenja i čuvanja podataka

Prethodno smo se upoznali sa bazama podataka, a sledeći viši nivo u prikazanoj hijerarhiji organizacije skladištenja i čuvanja podataka su skladišta podataka (engl. *Data Warehouse - DW*). Naime, kako je obim podataka postajao sve veći, kompanije nisu mogle da efikasno funkcionišu sa više baza podataka koje nisu bile međusobom povezane, a kreirane su za posebne korisnike. Time su se stvarali **silosi podataka (*data silos*)**, što znači da su se podaci decentralizovano i fragmentirano čuvali u okviru jednog poslovnog sistema. Skladišta podataka su upravo nastala sa potrebom da se povežu i ujedine struktuirani podaci baza podataka na nivou poslovnih organizacija. Od skladišta podataka se očekivalo da podrže ideju sagledavanja kupaca iz različitih uglova tj. *“360 degree view of the customer”*⁹.

Istorijski posmatrano, inicijalne ideje na kojima se zasniva skladište podataka pojavile su se 60-tih godina. Međutim, može se reći da se skladišta podataka pojavljuju 1980-tih godina kada su *Paul Murphy* i *Barry Devlin* (zaposleni u IBM) razvili *Business Data Warehouse*. Ipak, sam koncept skladišta podataka pripisuje se *Inmon Bill-u*, za koga se smatra da je otac skladišta podataka¹⁰.

Šta je skladište podatka?

Skladište podataka predstavlja **centralni repozitorijum** podataka koji su integrisani iz različitih izvora. Ovaj repozitorijum je dizajniran **za analitičke svrhe** (kreiranje izveštaja, postavljanja upita i napredno pretraživanje podataka), a **ne za obradu transakcija**.

Dobar sistem za obezbeđivanje podataka treba da omogućí podršku:

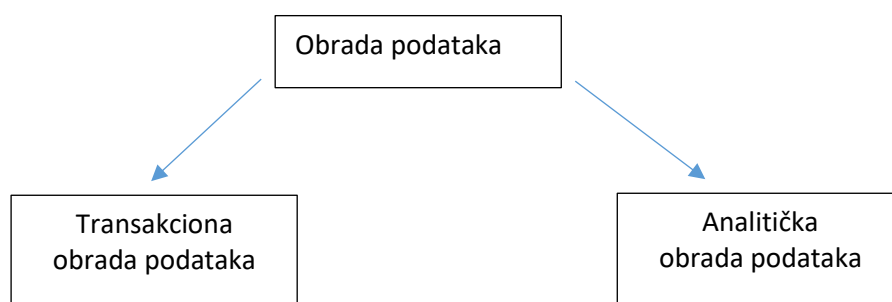
1. da krajnji korisnici lakše pristupaju podacima iz različitih oblasti;
2. da se mogu obaviti kompleksne pretrage poslovnih podataka;
3. da se mogu obaviti različiti analitički postupci na podacima;

⁹ <https://databricks.com/discover/data-lakes/history>

¹⁰ <https://www.guru99.com/data-warehousing-pdf.html>

4. brzo donošenje poslovnih odluka na osnovu relevantnih podataka;
5. precizno i efikasno donošenje odluka;
6. fleksibilno donošenje odluka.

Poznato je da se najčešće za obradu poslovnih transakcija koriste relacione baze podataka, a rezultati su uglavnom rezimei - sumarni prikazi i razni izveštaji. Takođe koriste se SQL (engl. *Structured Query Language*) upiti nad relacionim bazama podataka. Za ovaj način obrade poslovnih podataka kaže se da predstavlja **transakcionu obradu podataka**. Za razliku od ovog načina obrade podataka javlja se potreba za sprovođenjem različitih analiza podataka – **analitička obrada podataka**, za koju je potrebna drugačija organizacija podataka u odnosu na klasične baze podataka. Za te svrhe koristi se skladište podatka (engl. **Data Warehouse - DW**), ili kako se često označava i kao poslovno skladište podatka (engl. **Enterprise Data Warehouse - EDW**).



Slika 11. Vrste obrada podataka u poslovnom sistemu

Videli smo da je kod baza podataka, koje predstavljaju *online transactional processing system* (OLTP), neophodno da se svaka transakcija brzo obradi i sve izmene da se sačuvaju u odgovarajućem slogu. Kod baza podataka podaci se konstantno dodaju, edituju i/ili ažuriraju, što ukazuje na visoku **volatilnost (nepostojanost - promenljivost)** ovih podataka. Na primer, u bankarskom poslovanju svaki depozit, svako podizanje ili ulaganje novca na neki od bankarskih računa registruje se u realnom vremenu, što znači da se podaci u bankarskim bazama podataka neprekidno menjaju i ažuriraju. To je dobra osobina za potrebe obrade transakcija, ali nije poželjno za analitičku obradu podataka. U tom smislu skladišta podataka predstavljaju rešenje, jer su dizajnirana kao *online analytical processing system* (OLAP), dakle sistem u kome se mogu obavljati složene pretrage i analiza na mnogo efikasniji način nego sa podacima iz baza podataka.

Osnovne **karakteristike** skladišta podataka su:

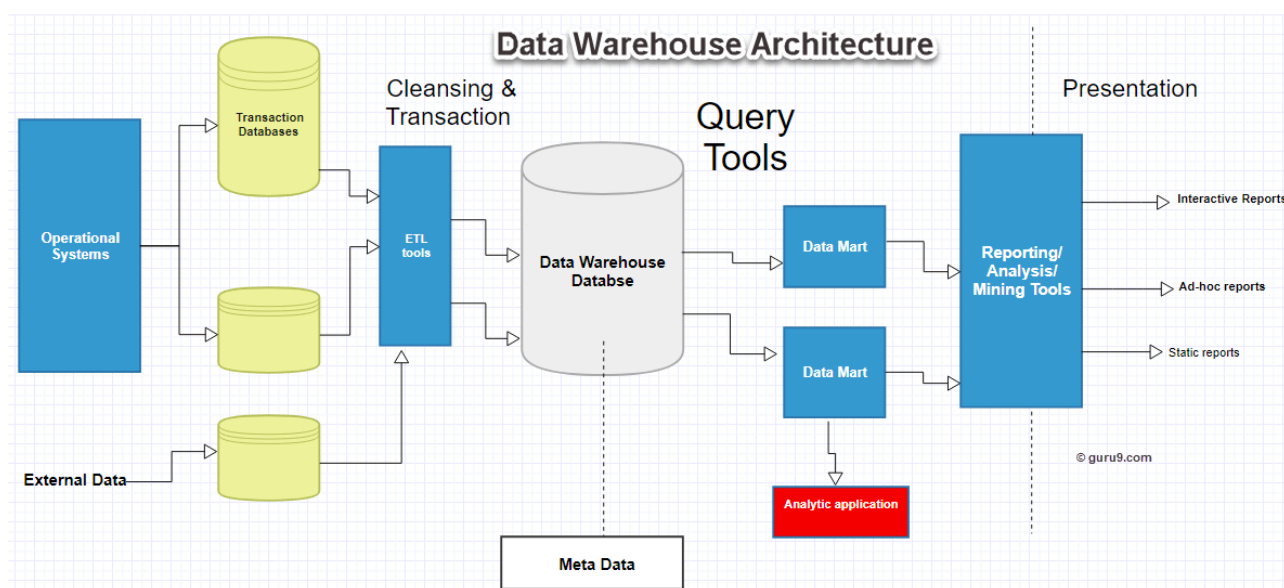
- **Organizacija podataka prema subjektima** – podaci su organizovani *po određenim predmetima* (temama ili subjektima) kao što su kupac, prodavac, proizvod, itd., a *ne prema poslovnim transakcijama*, kao što su isporuka, nabavka, kontrola zaliha;
- **Integrisanost podataka** – u skladištu podataka su integrisani podaci iz različitih poslovnih izvora;
- **Konzistentnost** podataka – skladište podataka je izgrađeno tako što su integrisani podaci iz različitih izvora; relacionih baza podataka, mainframe sistema, *flat* datoteka i sl. Pri tome podaci preuzeti iz različitih sistema moraju biti konzistentni – na isti način šifrirani (na primer, nekonzistentnost - pol: M i Z, pol: 0 i 1);
- **Trajnost** podataka – podaci u skladištu podataka se **ne ažuriraju** kao što je to slučaj sa podacima u bazama podataka. Ovi podaci su *read-only* i **periodično se mogu obnavljati** (kada se obavljaju ETL procesi). Opcije kao što su brisanje, ažuriranje i umetanje (*delete, update and insert*) kod skladišta podataka ne postoje, već su ključne dve operacije: učitavanje podataka (*data loading*) i pristup (*data access*).
- **Vremenska dimenzija** – skladište podataka ima izraženu vremensku dimenziju, jer sadrži podatke iz više meseci, kvartala, godina tako da se mogu sprovoditi poređenja kroz vreme, kao i trend analiza;
- **Relaciona struktura** – tipično je da skladišta podatka koriste relacionu strukturu;
- **Klijent/server** ili web-based arhitektura;
- **Sadrži metapodatke** – podaci o samim podacima.

Radi boljeg sagledavanja karakteristika skladišta podataka, u narednoj tabeli su prikazane razlike između baza podataka i skladišta podataka.

Tabela 2. Karakteristike baza podataka i skladišta podataka

Karakteristike	Baza podataka	Skladište podataka
Cilj	Da se čuvaju i procesiraju slogovi	Da se analiziraju podaci
Korišćenje	Da se obave fundamentalne poslovne transakcije	Da se izvrši analitika poslovnih podataka
Orijentacija	Aplikaciono-orijentisana kolekcija podataka	Kolekcija podataka orijentisana prema subjektima
Metod obrade	Online Transactional Processing (OLTP)	Online Analytical Processing (OLAP)
Tabele i veze	Izvršena normalizacija	Nije izvršena normalizacija
Raspoloživost podataka	Podaci su raspoloživi real-time	Podaci se osvežavaju iz izvornih datoteka po potrebi
Volatilnost – (ne/postojanost) podataka	Nepostojanost podataka	Postojanost podataka
Tip podatka	Samo tekući podaci. Podaci u bazama podataka su <i>up-to –date</i>	Tekući i istorijski podaci se unose, nisu obavezno <i>up-to-date</i> .
Tip pretrage koji se koristi	Koriste se jednostavni upiti na transakcione podatke	Koriste se kompleksni upiti za analitičke svrhe
Sumarni podaci	U bazama podataka se čuvaju samo detaljni podaci , a ne sumarni podaci	Čuvaju se sumarni podaci

Struktura skladišta podataka je prikazana na sledećoj slici:



Slika 12. Skladište podataka¹¹

Generalno se mogu izdvojiti četiri komponente skladišta podataka:

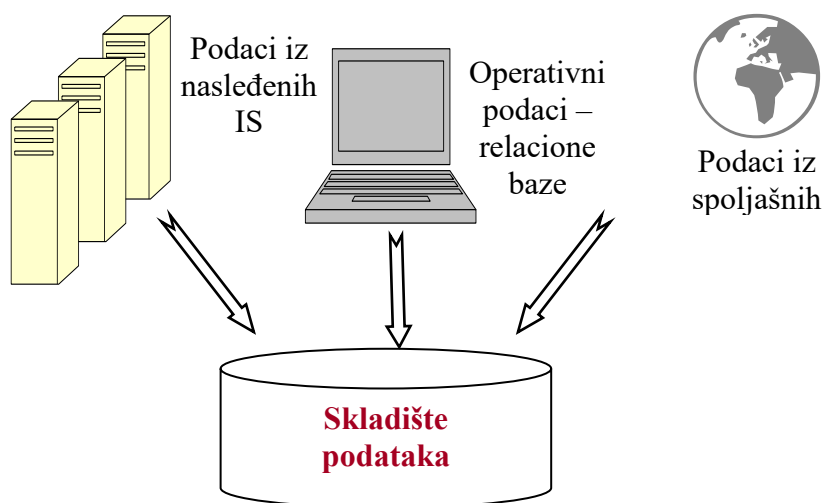
1. Komponenta za učitavanje podataka (**Load manager**) u skladište podataka. Ova komponente, poznata i kao *front-end* komponenta, obavlja sve operacije ekstrakcije, transformacije i učitavanja podataka u skladište podataka (*Extraction, Transformation and Loading - ETL procesi*).
2. Komponenta za upravljanje podacima u skladištu podataka – **Warehouse manager**. Ova komponenta obavlja sve operacije koje su povezane sa upravljanjem podataka u samom skladištu podataka, kao što su: analiza podataka radi održavanja konzistentnosti podataka, kreiranje indeksa, sprovođenje denormalizacije i agregacije, transformacija i udruživanje podataka, arhiviranje i stvaranje *back-up-a*.
3. Sistem za upite – **Query manager**. Ova komponenta, poznata i kao *back-end* komponenta, obavlja sve operacije koje su povezane sa upravljanjem upitima korisnika.
4. Korisnički alati - **End-user access tools**. Ova komponenta obuhvata nekoliko grupa alata koje su korisnicima na raspolaganju. Reč je o sledećem:

¹¹ <https://www.guru99.com/data-warehouse-architecture.html>

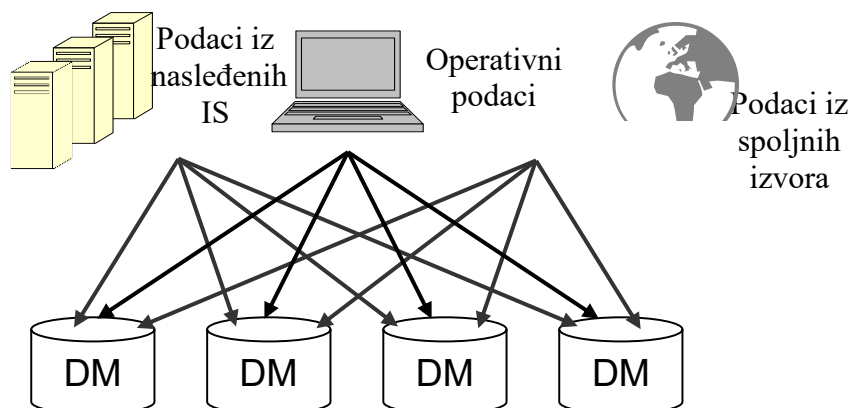
- Alati za generisanje izveštaja,
- Alati za postavljanje upita,
- OLAP alati i data mining alati,
- EIS alati, DSS alati,
- Alati za razvoj aplikacija.

Arhitektura skladišta podataka

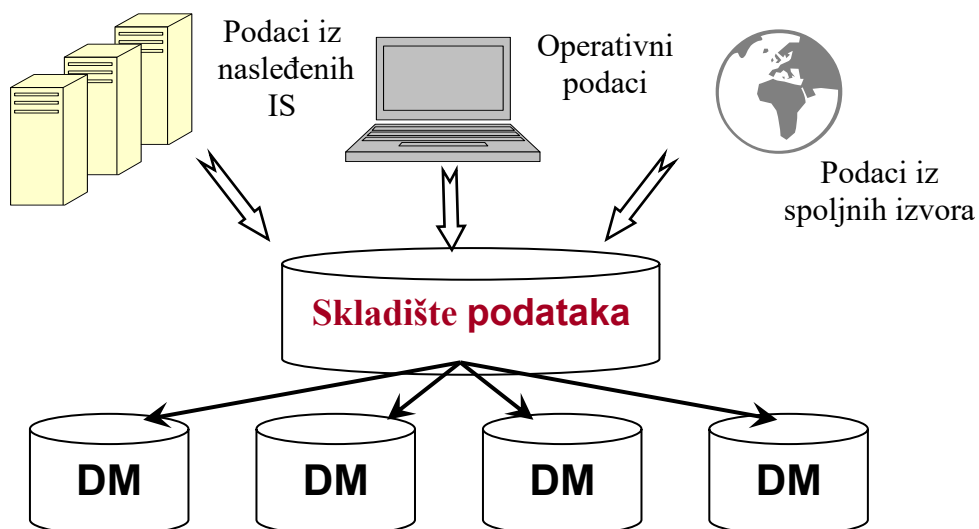
- Dvoslojna arhitektura sa zajedničkim skladištem podataka
- Dvoslojna arhitektura sa većim brojem nezavisnih lokalnih skladišta podataka (*Data Mart – DM*)
- Troslojna arhitektura skladišta podataka



Slika 13. Dvoslojna arhitektura sa zajedničkim skladištem podataka



Slika 14. Dvoslojna arhitektura sa većim brojem nezavisnih lokalnih skladišta podataka (*Data Mart – DM*)



Slika 15. Troslojna arhitektura skladišta podataka

Koji su razlozi da se poslovni sistem opredeli za izgradnju skladišta podataka? U nastavku su prikazane **osnovne prednosti** koje pružaju ovi sistemi:

1. Skladište podataka omogućava kompanijama da **restruktuiraju** svoju strategiju informacionih sistema;

2. obezbeđuje **objedinjen pregled na podatke** na nivou preduzeća. Poslovni korisnici mogu na jednom mestu brzo da pristupe kritičnim podacima iz različitih izvora;
3. skladišta podataka doprinose da se **odvoji analitička obrada podataka od obrade transakcija**, čime se doprinosi poboljšanju performansi i jednog i drugog sistema;
4. omogućuje da se **skupi operativni sistemi oslobode** obrade podataka, koja se prebacuje na jeftinije servere. Naime, skladišta podataka smanjuju pritisak na operativne sisteme;
5. skladišta podataka sadrže podatke iz različitih izvora o mnogim međusektorskim aktivnostima i **podržavaju ad-hoc izveštavanje i postavljanje kompleksnih upita**;
6. Skladišta podataka obično sadrže i velike količine **istorijskih podataka**. To pruža mogućnost poslovnim analitičarima da analiziraju pojave kroz vreme (trendove), kao i da predviđaju kretanja poslovnih pojava.

Navedene prednosti mogu da:

- poboljšaju poslovno znanje,
- obezbede konkurentsku prednost,
- poboljšaju usluge koje se pružaju klijentima,
- olakšavaju donošenje odluka i
- pomognu u racionalizaciji poslovnih odluka.

Pored navedenih prednosti koje pružaju skladišta podataka, postoje i **izvesni problemi**, odnosno **nedostaci**, koji se vezuju za ove sisteme, a koje je potrebno imati u vidu pri razmatranju projekta izgradnje ovih sistema. Reč je o sledećem:

- Skladišta podataka **nisu idealni** repozitorijumi **za nestruktuirane podatke**;
- Skladišta podataka su **skupi sistemi** koji zahtevaju održavanje: ETL procesi su obično veoma zahtevni što se tiče vremena i kompleksnosti operacija;
- Skladišta podataka **lako mogu postati zastarela**;
- Nije jednostavno da se prave promene u tipovima podataka, šemama izvora podataka, indeksima i upitima;

- Mada na prvi pogled može delovati da je jednostavno korišćenje skladišta podataka, obično to nije slučaj, pa se **prosečni korisnik često suočava sa njegovom kompleknošću**;
- Projekti izgradnje skladišta podataka su izuzetno **zahtevni projekti** u pogledu svih resursa: vremena, novca, organizacije;
- Poslovni sistemi moraju **uložiti značajne resurse** ne samo za implementaciju DW, već i za obuku korisnika;
- Ponekad problem sa skladištima podataka mogu biti neotkriveni godinama.

Imajući u vidu sve prethodno izložene prednosti, ali i nedostatke skladišta podataka, može se sugerisati sledeće - primena skladišta podatka se preporučuje u sledećim situacijama:

- Donosioci odluka u svom radu se oslanjaju na velike količine podataka;
- Posebno je interesantno za donosiocce odluka koji biraju sistematičan pristup odlučivanju zasnovanom na podacima (primena naprednih tehnika pretraživanja podataka – *data mining*);
- Korisnicima su potrebni među-sektorski podaci iz različitih izvora;
- Korisnicima su potrebne brze performanse na velikoj količini podataka: brzo izveštavanje, brzo postavljanje kompleksnih upita i sl.

Skladišta podataka se primenjuju u mnogim poslovnim oblastima. Ali, posebno se izdvajaju sledeće oblasti:

- vazdušni saobraćaj,
- finansijska industrija: bankarstvo, osiguranje, investiranje,
- javni sektor,
- zdravstveni sistemi,
- telekomunikacije.

Skladišta podataka predstavljaju ključan sistem za razvijanje koncepta Poslovne inteligencije (engl. *Business Intelligence* - *BI*) u jednom poslovnom sistemu, o kome će kasnije biti više reči.

Lokalna skladišta podataka (*Data mart* - *DM*)

Veoma visoka cena izgradnje skladišta podataka predstavlja jedan od glavnih razloga zašto mnoge organizacije ne ulaze u projekat implementacije ovog centralnog repozitorijuma

podataka. Umesto toga često se odlučuju za implementaciju lokalnog skladišta podataka – **data mart**. Šta je data mart?

Lokalno skladište podataka – **data mart** je **kondenzovana verzija skladišta podataka** koja je obično dizajnirana za jednu stratezijsku poslovnu jedinicu ili odeljenje: marketing, prodaja, finansije ili ljudski resursi.

U lokalnom skladištu podataka obično su uneti podaci iz manjeg broja izvora u poređenju sa skladištem podataka. Lokalna skladišta podataka su ne samo manja po veličini od skladišta podataka, već su mnogo fleksibilnija u poređenju sa skladištem podataka. Cena lokalnih skladišta podataka je značajno niža od cene skladišta podataka (cena lokalnog skladišta podataka obično je niža od 100 000 \$ u poređenju sa cenom skladišta podataka od 1 000 000 \$). Takođe vreme potrebno za dizajniranje lokalnog skladišta podataka je nekoliko puta kraće od vremena potrebnog za izgradnju skladišta podataka. Ukratko, **prednosti** gradnje lokalnih skladišta podataka su sledeća:

- relativno **niža cena**,
- vreme implementacije je **znatno kraće**,
- **brži pristup** podacima,
- **upravljanje je lokalno**, a ne centralno,
- sadrže **manje podataka** (iz neke posebne poslovne oblasti), koji su preuzeti iz manjeg broja izvora, što ih čini fleksibilnijim od skladišta podataka,
- omogućavaju poslovnoj jedinici da gradi sopstvene sisteme podrške pri odlučivanju – **sopstvene DSS sisteme**.

Sledeća tabela daje sumaran prikaz razlika između Skladišta podataka i lokalnih skladišta podataka.

Tabela 3. Razlike između skladišta podataka (DW) i lokalnih skladišta podataka (DM)

Karakteristika	Skladišta podataka – DW	Lokalnih skladišta podataka - DM
Cilj	Da obezbedi integrisano okruženje podataka i koherentnu sliku svih poslovnih funkcija u poslovnom sistemu	Da obezbedi podatke iz određene poslovne jedinice ili odeljenja u organizaciji
Upotreba	Pomaže u donošenju strateških odluka	Pomaže u donošenju taktičkih odluka
Dizajn	Proces dizajna i izgradnje može biti veoma kompleksan	Proces je mnogo jednostavniji

Rad sa podacima	Može biti veoma složen zbog obimnosti podata u skladištu podataka	Mnogo jednostavniji za korisnika
Fokus	Na celu kompaniju	Na pojedina odeljenja poslovnog sistema
Izvor podataka	Mnogobrojni i raznovrsni izvori podataka	Mali broj izvora podataka
Veličina	100GB do nekoliko TB +	Obično manje od 100GB
Vreme implementacije	Do nekoliko godina	Nekoliko meseci (često manje od 90 dana)

U pogledu strukture postoje tri vrste lokalnih skladišta podataka:

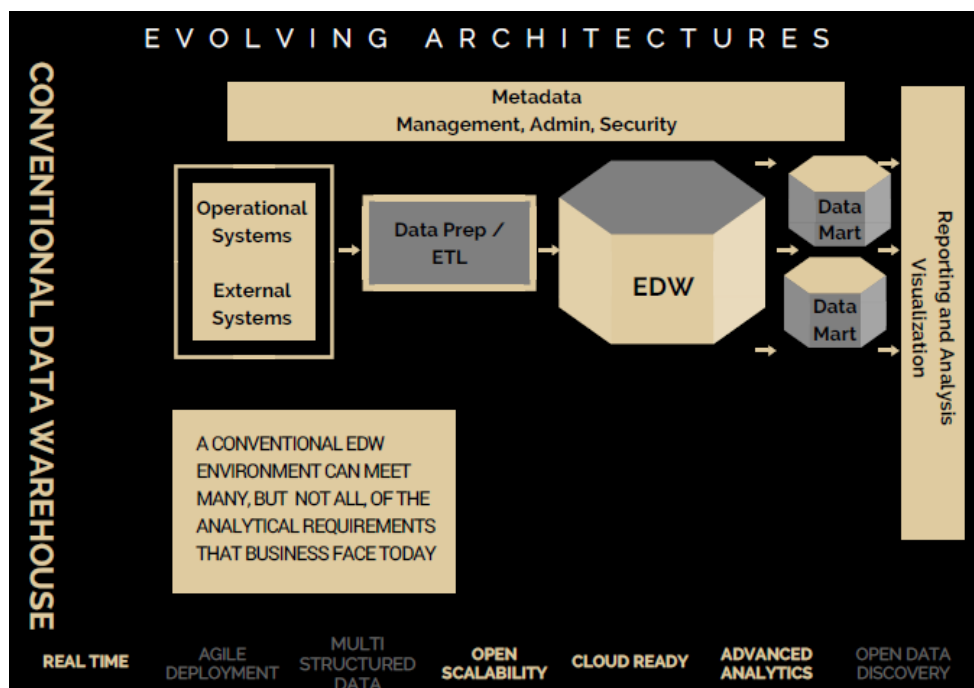
- **Samostalna (nezavisna)** lokalna skladišta podataka;
- **Zavisna (kopirana)** lokalna skladišta podataka;
- **Hibridna** lokalna skladišta podataka.

Samostalna lokalna skladišta podataka se kreiraju bez korišćenja centralnog skladišta podataka. Dakle u ovom slučaju ne postoji skladište podataka. Ovaj tip DM nije povezan niti sa nekim skladištem podataka niti sa drugim DM. Ovaj tip DM je pogodan za manje grupe ili odeljenja u okviru poslovnog sistema (pogledati sliku 14).

Zavisna (kopirana) lokalna skladišta podataka su ona skladišta koja su dodata postojećem skladištu podataka (videti prethodnu sliku 15). Ovaj pristup pruža sve prednosti centralne organizacije podataka, a zatim kopiranje na posebne poslovne celine. Ovaj tip DM može se izgraditi na dva načina: prvi, kada je korisniku omogućeno da pristupi i skladištu podataka i lokalnom skladištu podataka, a drugi pristup je kada je korisniku omogućen samo pristup lokalnom skladištu podataka.

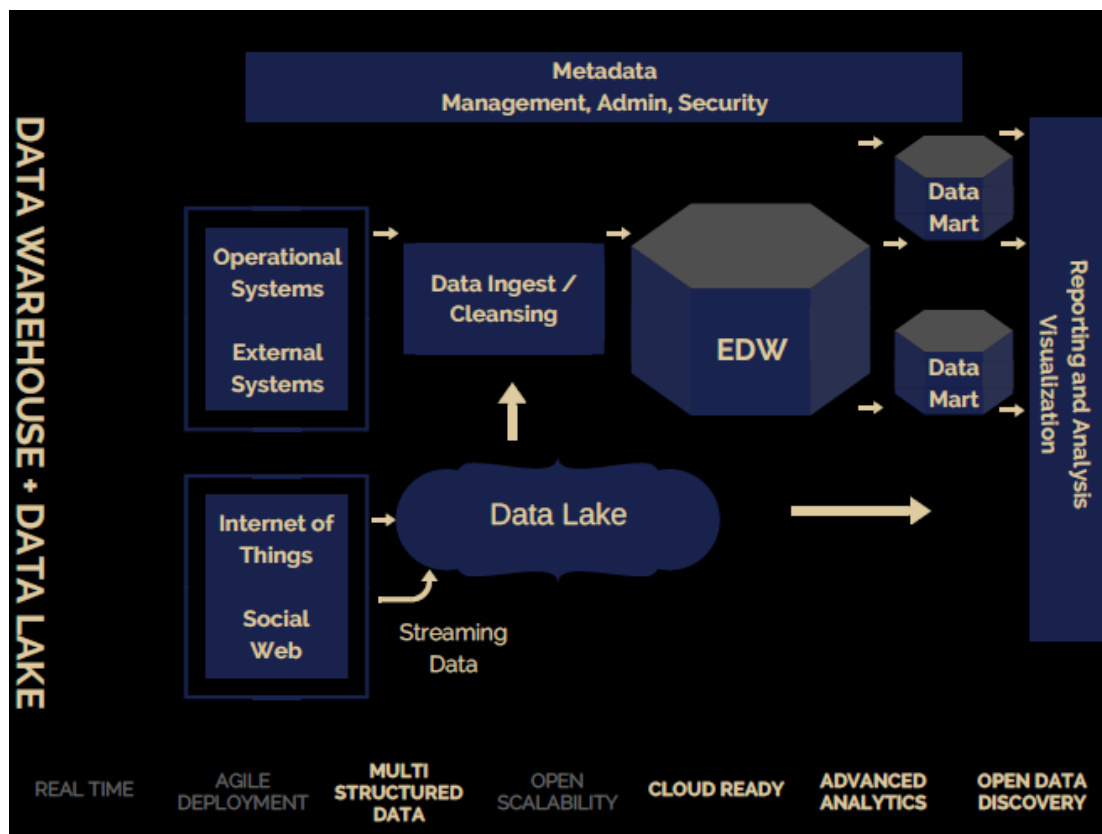
Hibridna lokalna skladišta podataka – kombinuju inpute iz različitih izvora, tj. i izvora van DW, dakle podaci se mogu uneti ili iz skladišta podataka ili iz drugih izvora. Ovaj tip DM je pogodan kada se želi ad-hoc integracija, ili može da podrži obimne strukture skladištenja podataka, a takođe i za fleksibilne male *data-centric* aplikacije.

BUDUĆNOST SKLADIŠTA PODATAKA¹²



Slika 17. Konvencionalna arhitektura skladišta podataka

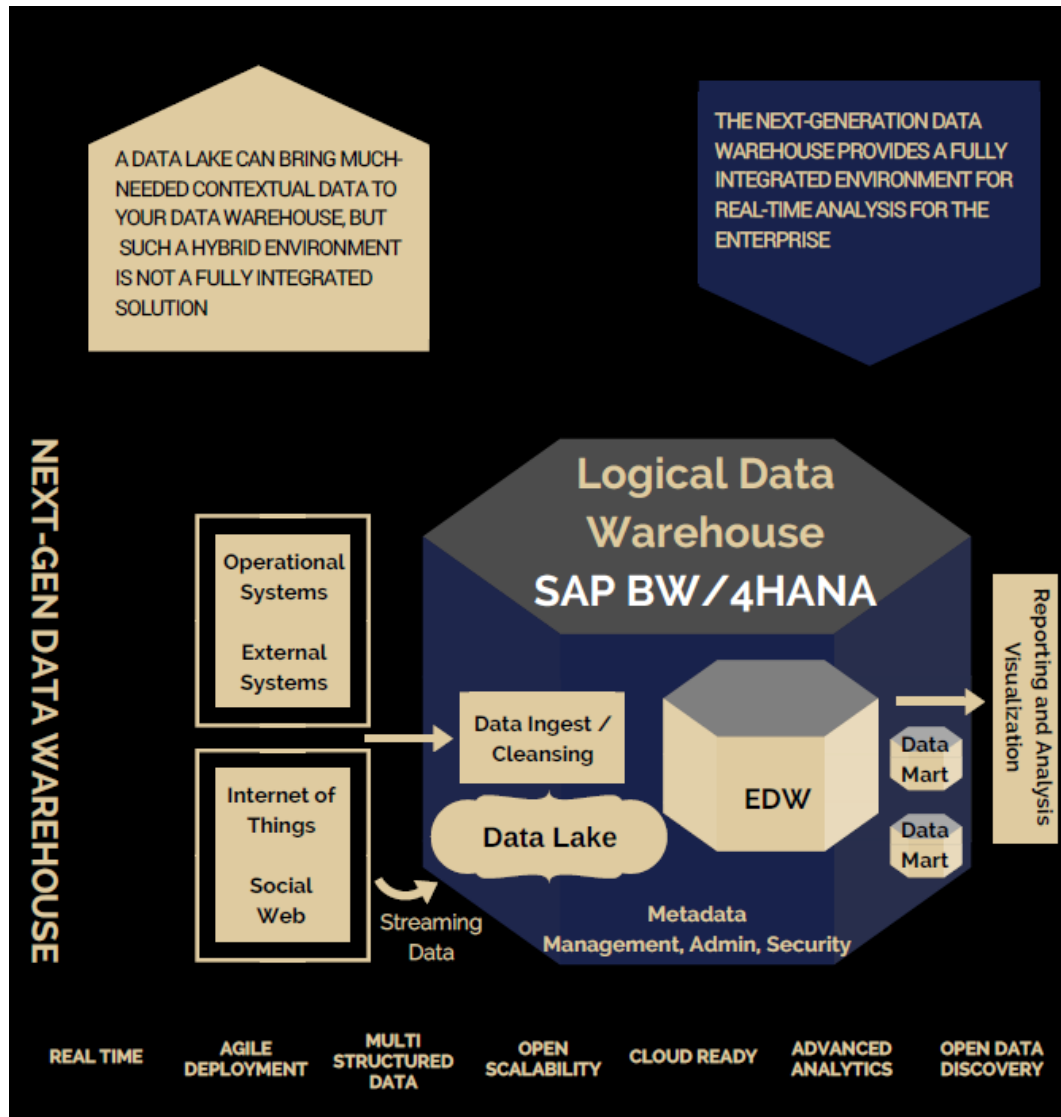
¹² THE EVOLUTION OF THE DATA WAREHOUSE, www.sap.com/bw4hana



Slika 18. Skladište podataka i *Data lake* (jezero podataka) - hibridno okruženje¹³

¹³ THE EVOLUTION OF THE DATA WAREHOUSE, www.sap.com/bw4hana

Sledeća generacija skladišta podataka



Slika 19. Integrirano okruženje¹⁴

¹⁴ THE EVOLUTION OF THE DATA WAREHOUSE, www.sap.com/bw4hana