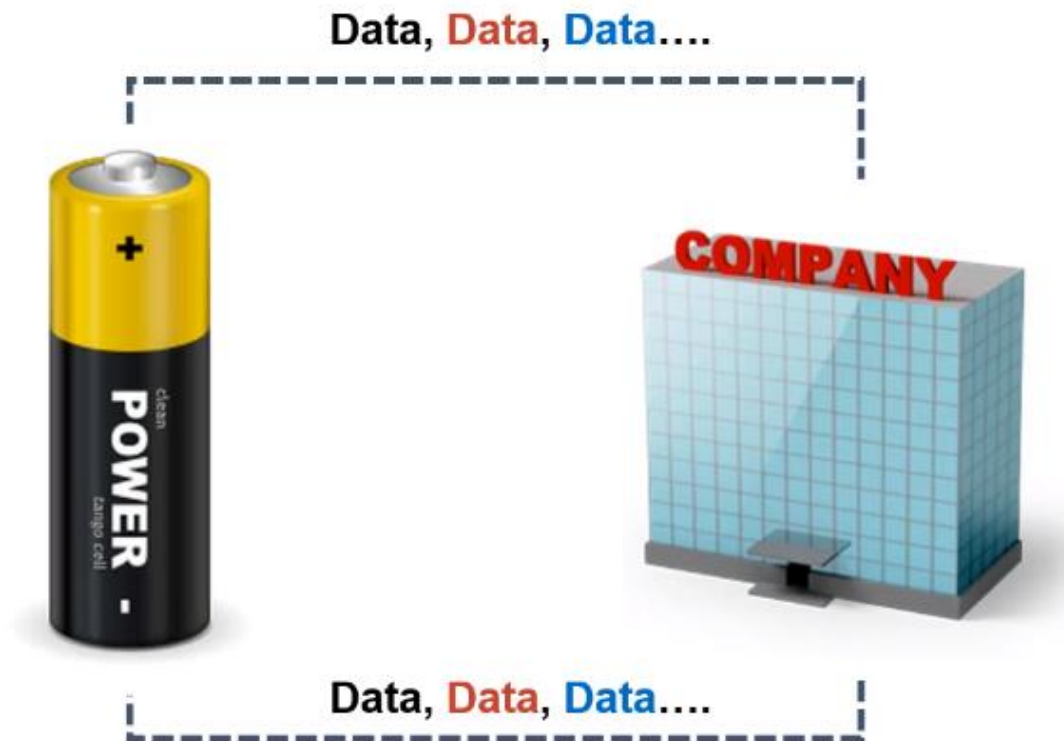


Terminologija u bazama podataka

Predmet: Administriranje Baze Podataka
Predavač: dr Dušan Stefanović

PODACI SU SVUDA OKO NAS

- Podaci su energija koja pokreće kompanije da posluju
- Ako su podaci energija, potrebno je za njih obezbediti mesto gde će se čuvati



PODACI SU SVUDA OKO NAS

- Podaci su energija koja pokreće kompanije da posluju
- Ako su podaci energija, potrebno je za njih obezbediti mesto gde će se čuvati

Kontejner za podatke & Engine

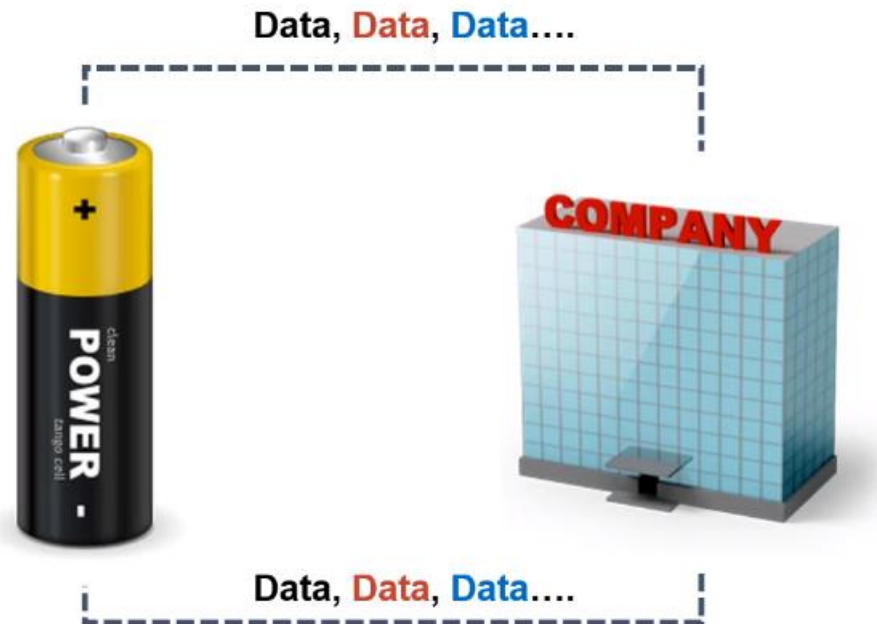
Fleksibilan

Brz

Pouzdan

Lako upravljiv

Isplativ



ČUVANJE I OBRADA PODATAKA

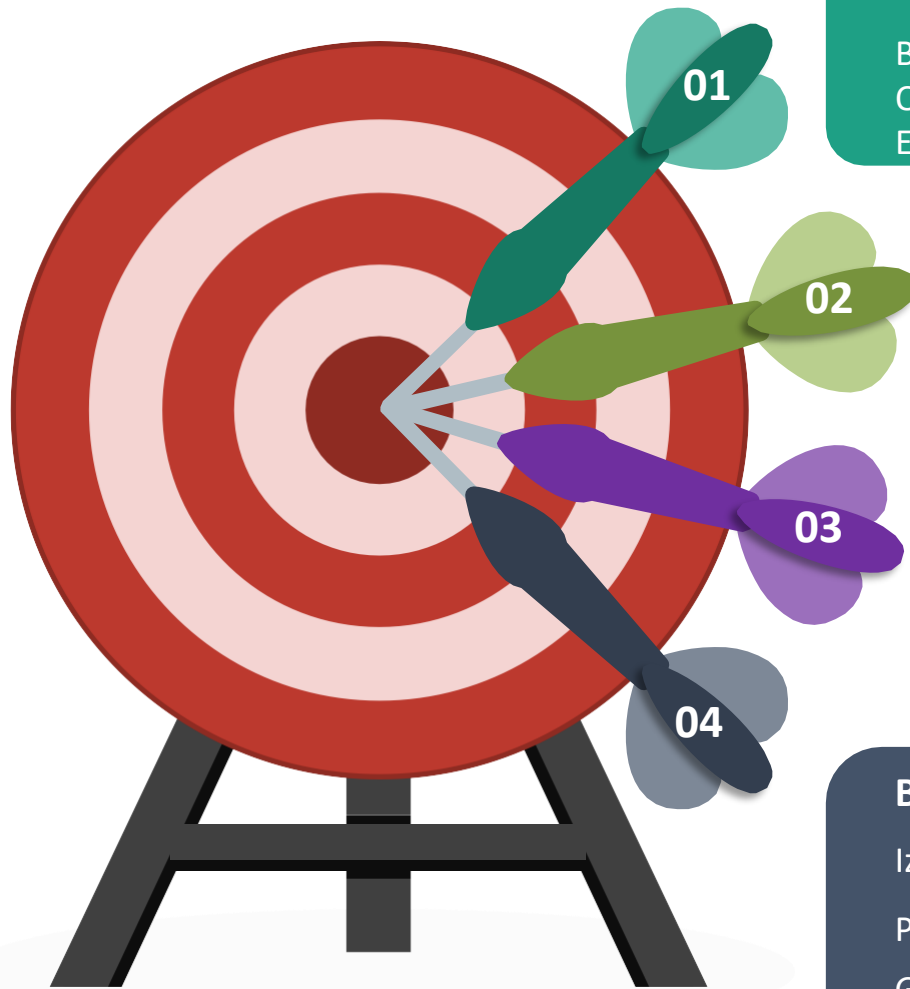
- **Zahtevi modernih aplikacija**

- Složeni tipovi podataka
 - Transakcije, tekst, log, IoT senzori, slike, video, audio...
- Veći kapacitet podataka
 - Milioni IoT senzora, Milioni korisnika koji pristupaju web platformama...
- Duže vreme zadržavanja
 - Analiza podataka (istorijski paterni, predviđanja, prognoze)
 - Mašinsko učenje

- **Baze podataka**

- Kontejneri za skladištenje podataka
- Backend tehnologija tzv. *power-horse* svih aplikacija
- Izbor prave baze podataka

PODACI SU SVUGDE OKO NAS



Terminologija

Baze podataka, DBMS, Šema, Operational/Analytics Data Warehouse/Lake, ETL/ELT, Skaliranje

Tehnologije

SQL DB, NoSQL DB, Distribuirane SQL DB
In-memory DB, Time-series DB
[Koncept + Prednosti + Primena]

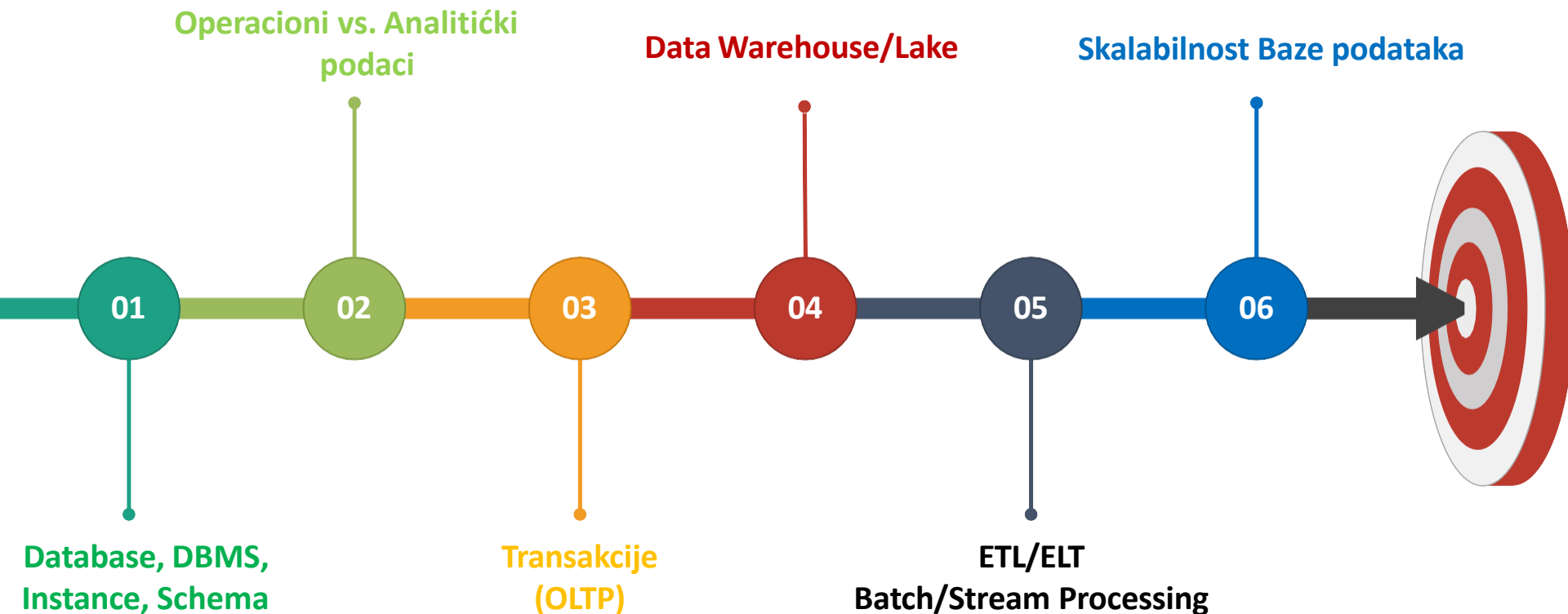
Vrste

#1 - Key-Value, #2 - Document
#3 - Wide Column, #4 - Graph

Baze podataka u oblaku (DBaaS)

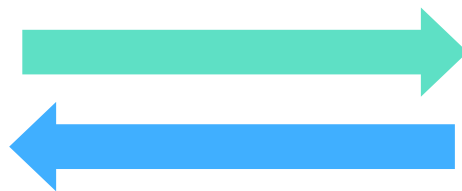
Izazovi kod tradicionalnih baza podataka
Prelazak na DBaaS model
Glavne prednosti
DBaaS – Azure, AWS, GCP

Terminologija Baze podataka



BAZA PODATAKA

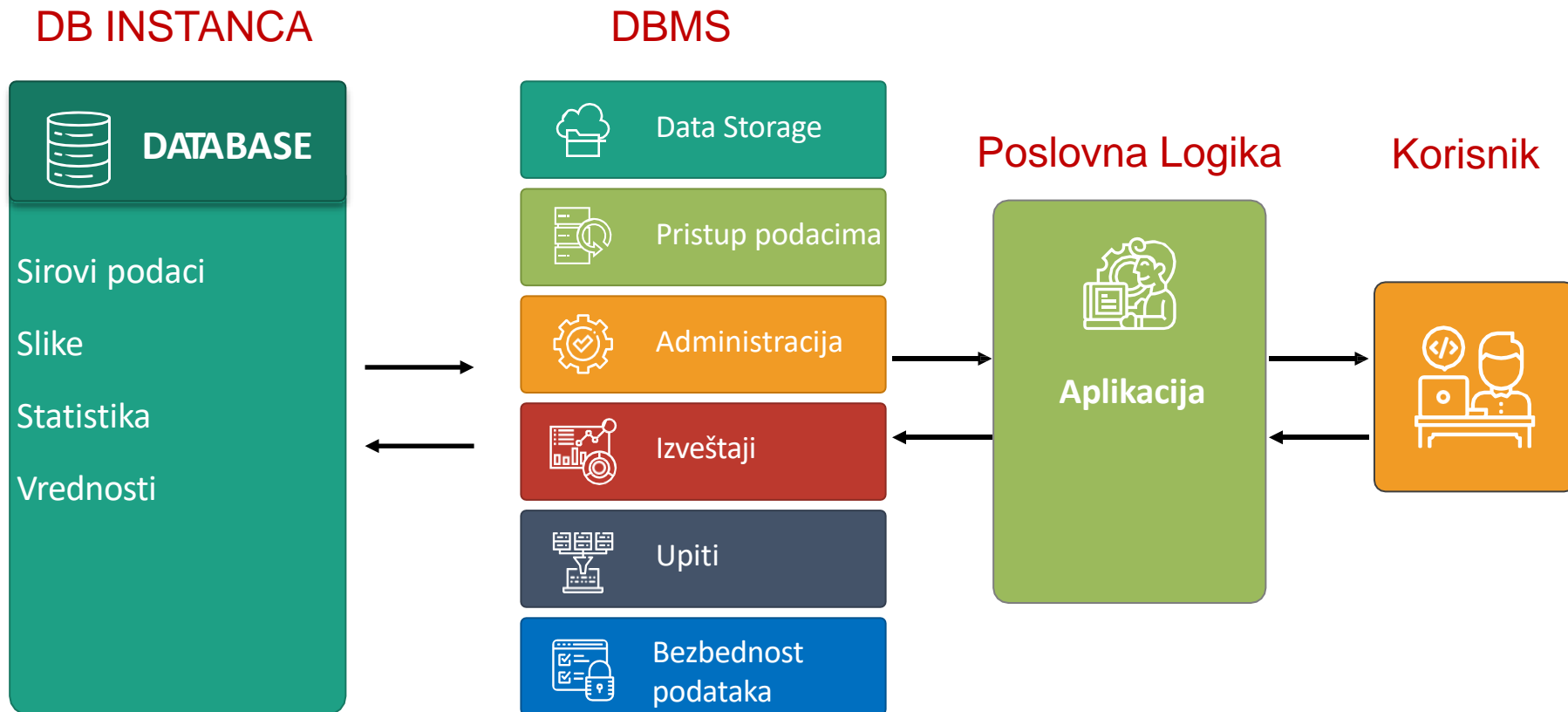
Baza podataka je organizovana kolekcija podataka koja čuva podatke i kojoj se pristupa od strane aplikacije



Korisnički profili

Cene na tržištu (berzi)

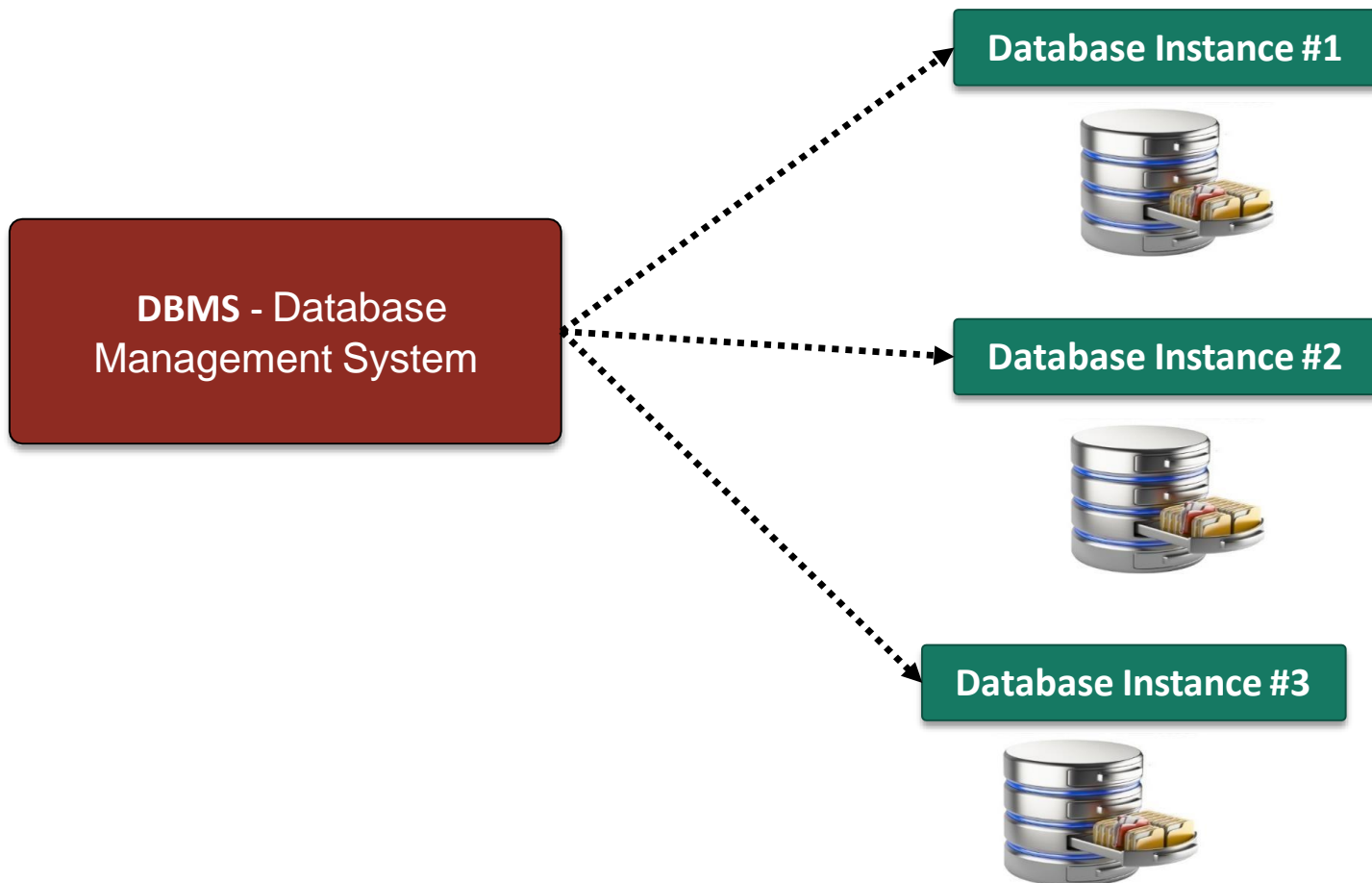
SISTEM ZA UPRAVLJANJE BAZOM PODATAKA



Instanca Baze podataka

Instanca baze podataka je logički entitet (kontejner) koji je kreiran od strane korisnika i spreman je da se popuni podacima

Kontejner sa podacima

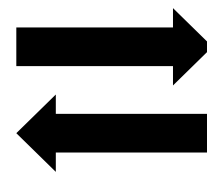
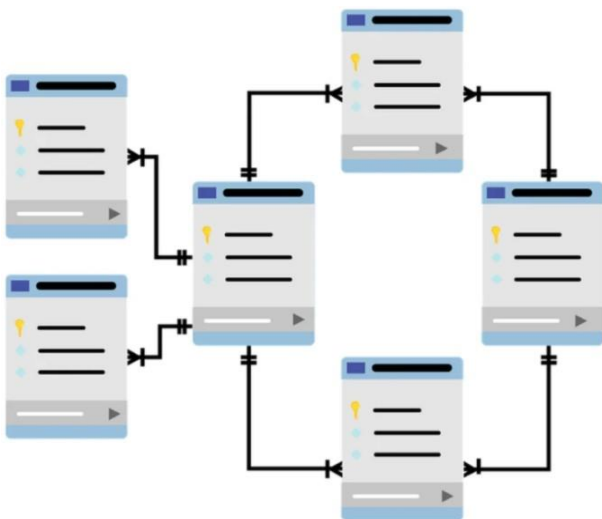
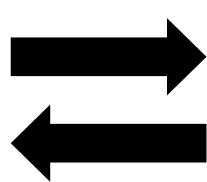


Šema Baze podataka

Šema baze podataka obezbeđuje da podaci budu organizovani u strukturu sa jasno definisanim setom pravila i ograničenjima koja su nametnuta od strane DBMS-a.

Šema baze podataka je šablon koji opisuje strukturu i koja se definiše nakon kreiranja instance baze podataka

Postoje DBMS koji se striktno drže definisane šeme ali postoje i baze koje su fleksibilnije



Read / Write
Data

Tip Baze podataka

Tip baze podataka se bira na osnovu tipa podataka koji se čuva i na osnovu primene samih podataka (use case).

Dve osnovne vrste primene podataka u kompaniji su **operacioni podaci** i **podaci za analizu**.

Operacioni podaci su svakodnevni podaci koji se kreiraju na dnevnom nivou u realnom vremenu primenom transakcija

Analitički podaci i **operacioni podaci** se drugačije skladište i obrađuju u bazama podataka

Operacioni Podaci

- Porudžbine
- Proizvodi
- Klijenti/Snabdevači
- Servisi
- Tiketi
- Prodaja



Analitički Podaci

OLTP (Online Transactional Processing)

Transakcije

Transakcija u bazama podataka je jedinica posla koja se izvrši unutar baze

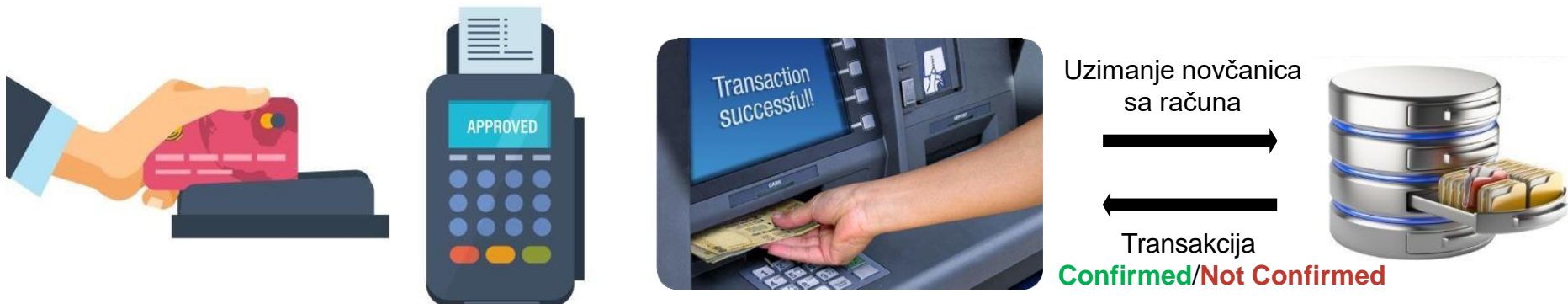
Primer je plaćanje robe gde je transakcijom obuhvaćeno skidanje novca sa računa upatioca, prebacivanje novca na račun primaoca i ažuriranje brojčanog stanja kupljenog artikla.

Podaci se čuvaju u dvodimenzionalnim strukturama

Transakcije obrađuju sistemi poznati kao OLTP (ERP, CRM, Plaćanja, ...) aplikacije

OLTP aplikacije koriste OLTP bazu podataka tj bazu podataka koja podržava transakcije

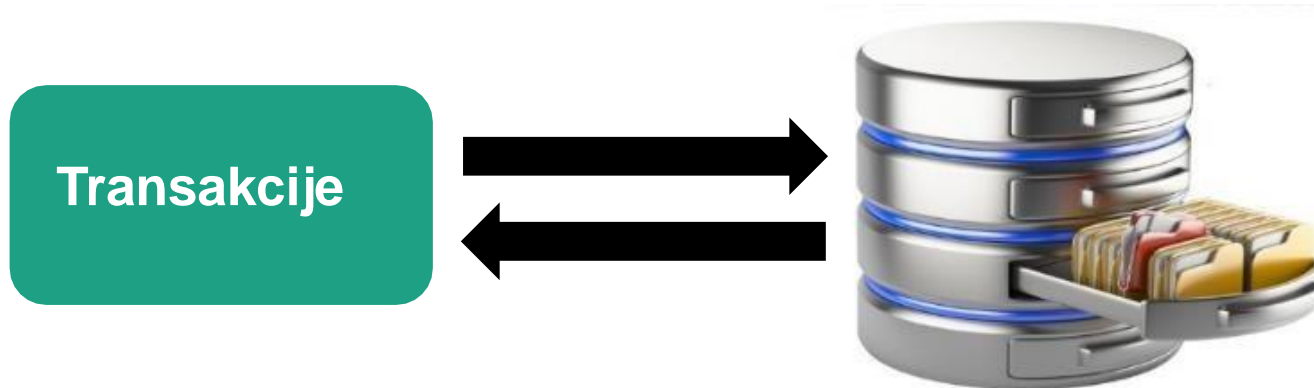
ACID Model (**A**tomic, **C**onsistent, **I**solation, and **D**urable)



OLTP (Online Transactional Processing)

Transakcije u OLTP sistemima

- Sve aktivnosti se čuvaju u bazu podataka preko transakcija
- Obrada velikog broja transakcija u realnom vremenu (**High-volume**)
- Brz pristup podacima u bazi (**Low-latency**)
- Česte promene podataka (updated transakcije)



OLAP (Online Analytical Processing)

■ Analitički podaci

- Poslovne odluke na osnovu analize istorijskih podataka
- Ulaz za *Business Intelligence* (BI) sisteme
- Analitički podaci su operacioni podaci ali u mnogo većem obimu
- Kompleksna analiza
 - Utiče na bolje poslovanje
 - Fokus na novim prilikama
 - Prepozna trend na tržištu
 - Prepoznaju anomalije

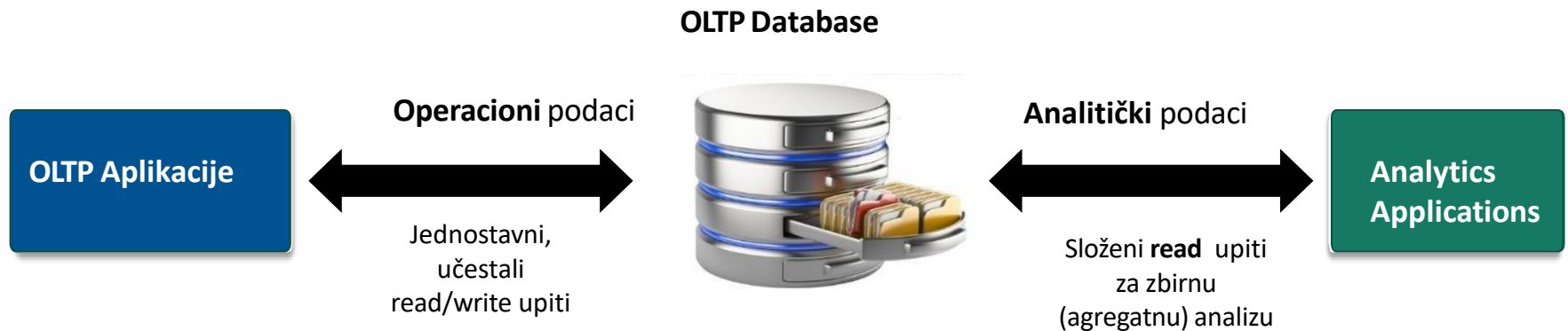


OLAP (Online Analytical Processing)

Opcija #1 – Ista OLTP Baza podataka za transakcije i analizu podataka

OLTP baza podataka nije dizajnirana da obrađuje veliku količinu agregiranih podataka kroz kompleksne upite koje bi uticale na performanse sistema

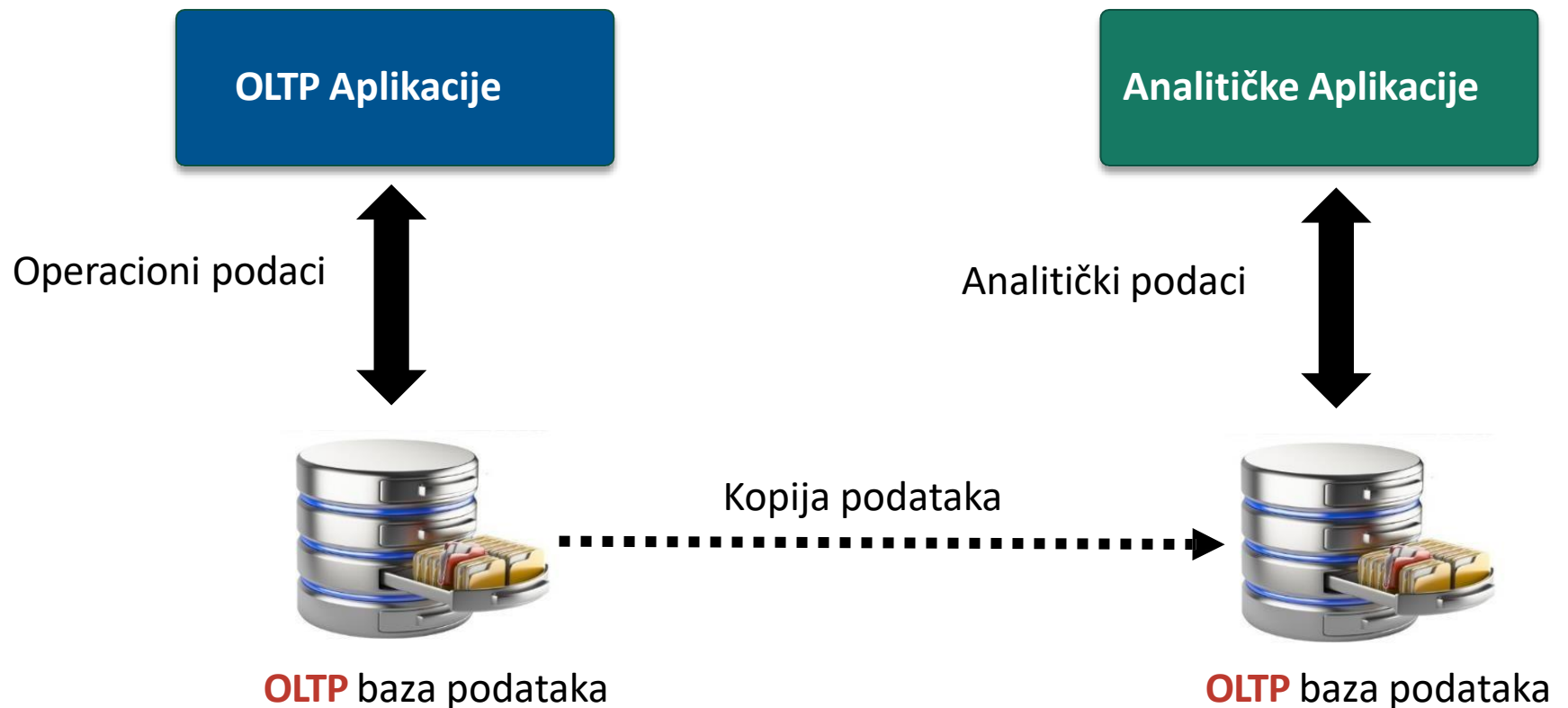
OLAP je tehnologija baze podataka je optimizovana za upite i izveštavanje, umesto obrade transakcija



OLAP (Online Analytical Processing)

Opcija #2 – Dedicated OLTP Baza podataka samo za analizu podataka

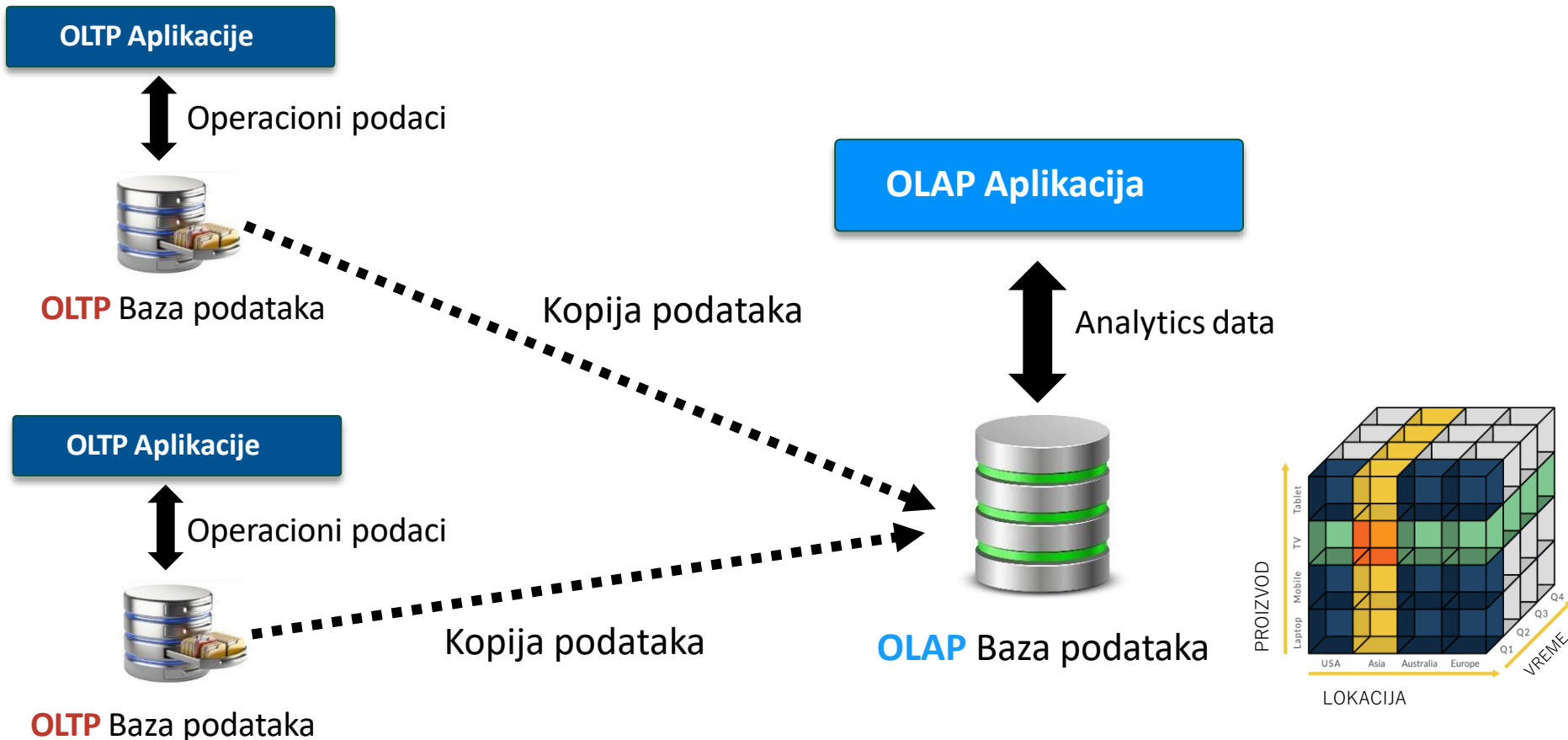
Ovakva arhitektura za analizu podataka radi racionalno samo za jednostavne analitičke zahteve



OLAP (Online Analytical Processing)

Opcija #3 – OLAP

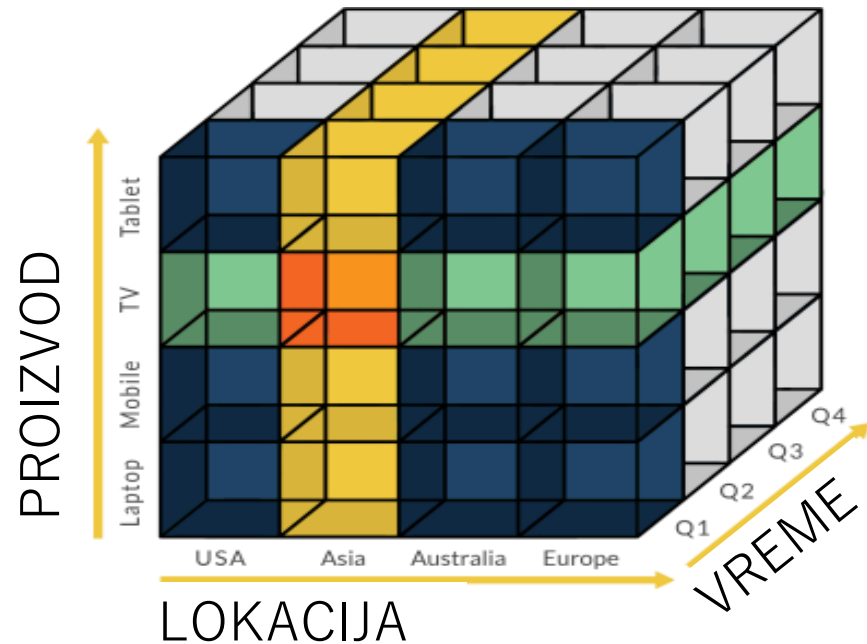
Podaci se čuvaju u drugačijoj strukturi u odnosu na OLTP bazu podataka (dvodimenzionalna) koja je optimizovana za analitičke upite – **višedimenzionalna struktura podataka Kocka**



VIŠEDIMENZIONALNOST

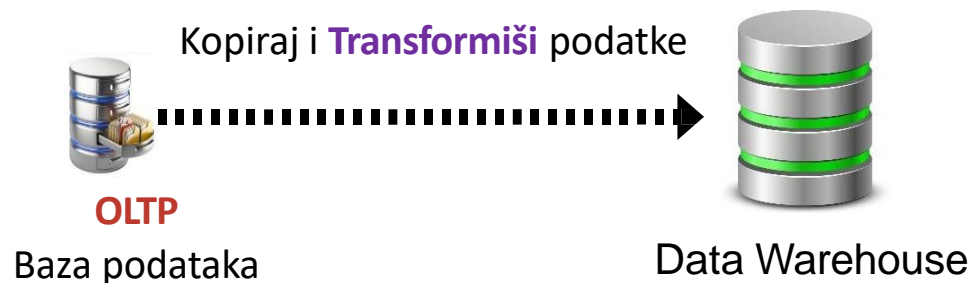
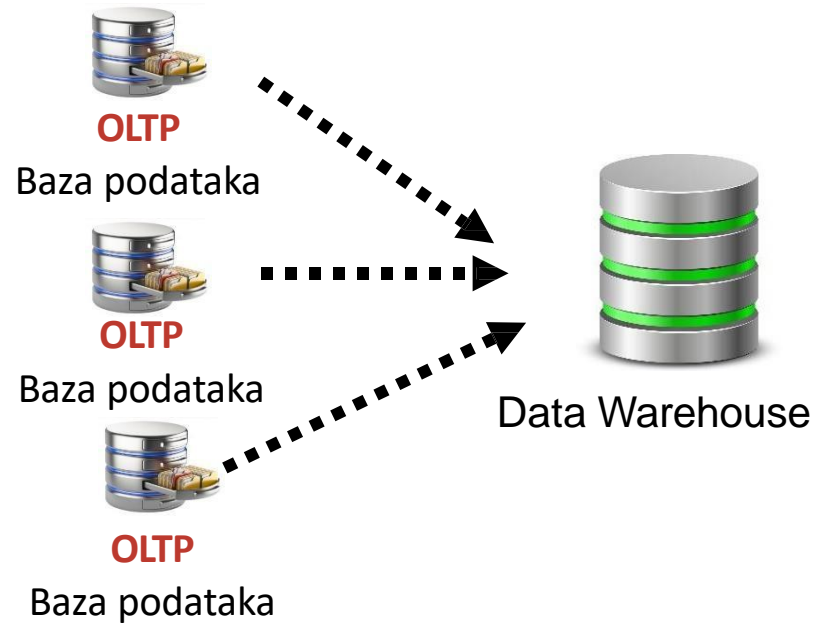
Opcija #3 – OLAP

- Višedimenzionalnost je osnovna karakteristika OLAP baze podataka
- Korisnik ima mogućnost da parametre poslovanja vidi u preseku dimenzija koje opisuju te parametre
- Ljudima je prirodno da podatke analiziraju kroz dimenzije
- Ako se pojava prati u **tri dimenzije**, reč je o **kocki**, a u **više dimenzija** o **hiperkocki**.
 - Svakoj **dimenziji kocke** pripada jedan parametar **pojave**.
 - Svaka **tačka u kocki** ima tačno **određene vrednosti** svake posmatrane dimenzije.



Data Warehouse

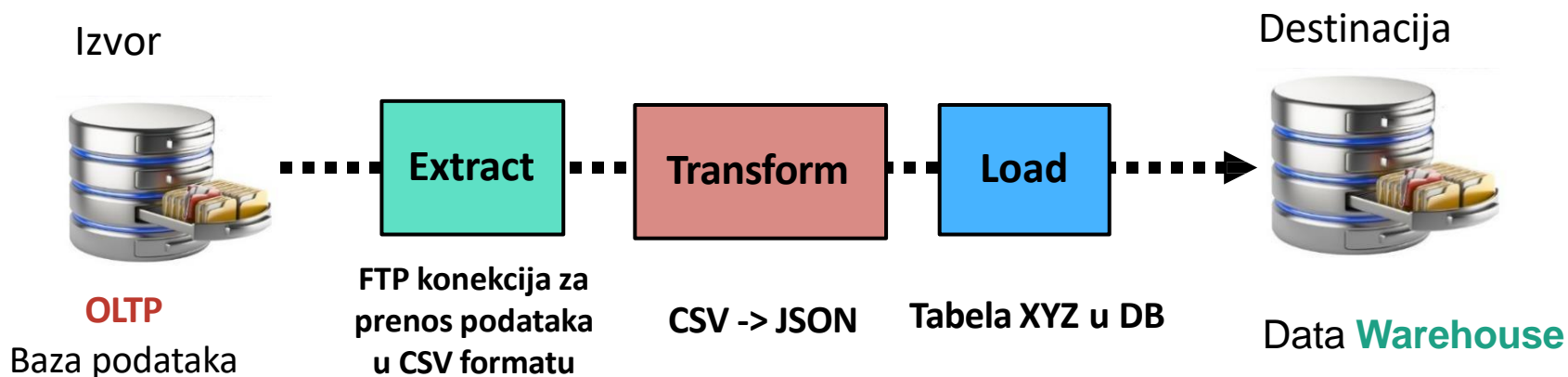
- Prikuplja podatke na jednom mestu sa različitih izvora podataka
- Radi sa velikom količinom podataka
- Optimizovan za OLAP aplikacije (kompleksne read upite)
- Čuva istorijske podatke na duže vreme (količina je u terabajtima i petabajtima)
- Sirovi podaci se prebacuju (transformišu) u novu strukturu optimizovanu za analizu podataka



Premeštanje podataka između sistema OLTP - OLAP

- **ETL - Extract, Transform, Load**

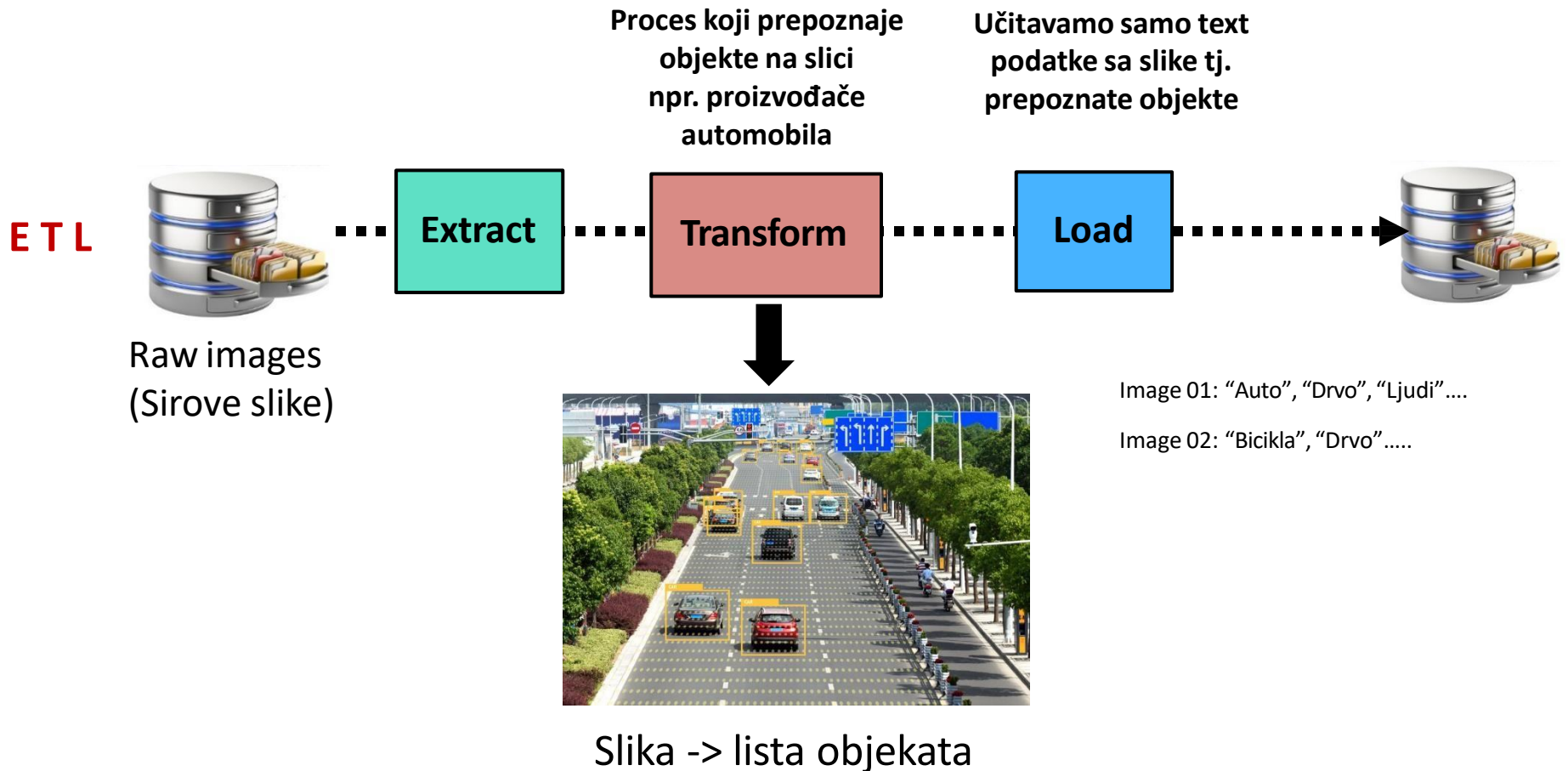
- Proces uzimanja (**extracting**) podataka iz jednog sistema (izvor podataka), **transformacija** podataka u novu strukturu, i unos (**load**) podataka u odredišni sistem
- Podaci se transformišu pre nego da se unesu u odredišni sistem
- Tipičan primer je Data Warehouse
- **Izvorišna aplikacija** radi transformaciju podataka



Premeštanje podataka između sistema OLTP - OLAP

- App X – Prepoznaje listu objekata na slici

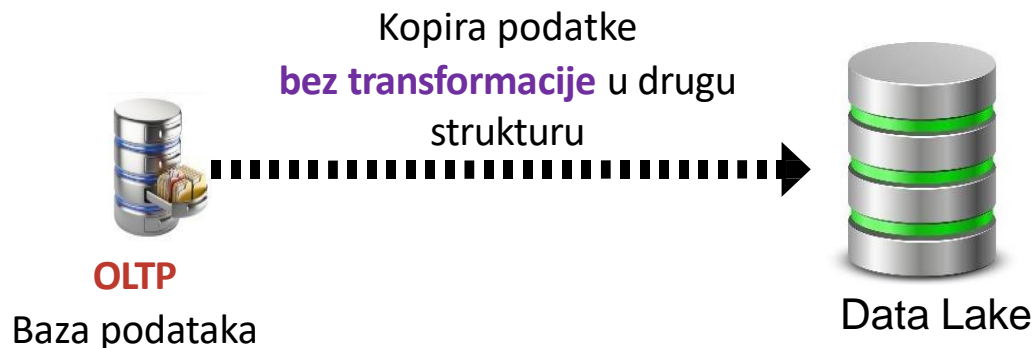
Efikanan metod za **umanjenje količine podataka** koji se čuva



Data Lake

■ Data Lake

- Centralizovani repozitorijum za čuvanje strukturiranih i ne strukturiranih podataka
- Prikupljeni podaci se čuvaju bez promene jer želimo da sačuvamo originalne podatke
 - Tekst, Slike, log podaci, IoT, Video fajlovi, ...
- Primena je u korišćenju podataka za kreiranje AI i mašinsko učenje
 - E.g. data science – ML\AI
- Kompanija može da ima warehouse i data lake



Premeštanje podataka između sistema OLTP - OLAP

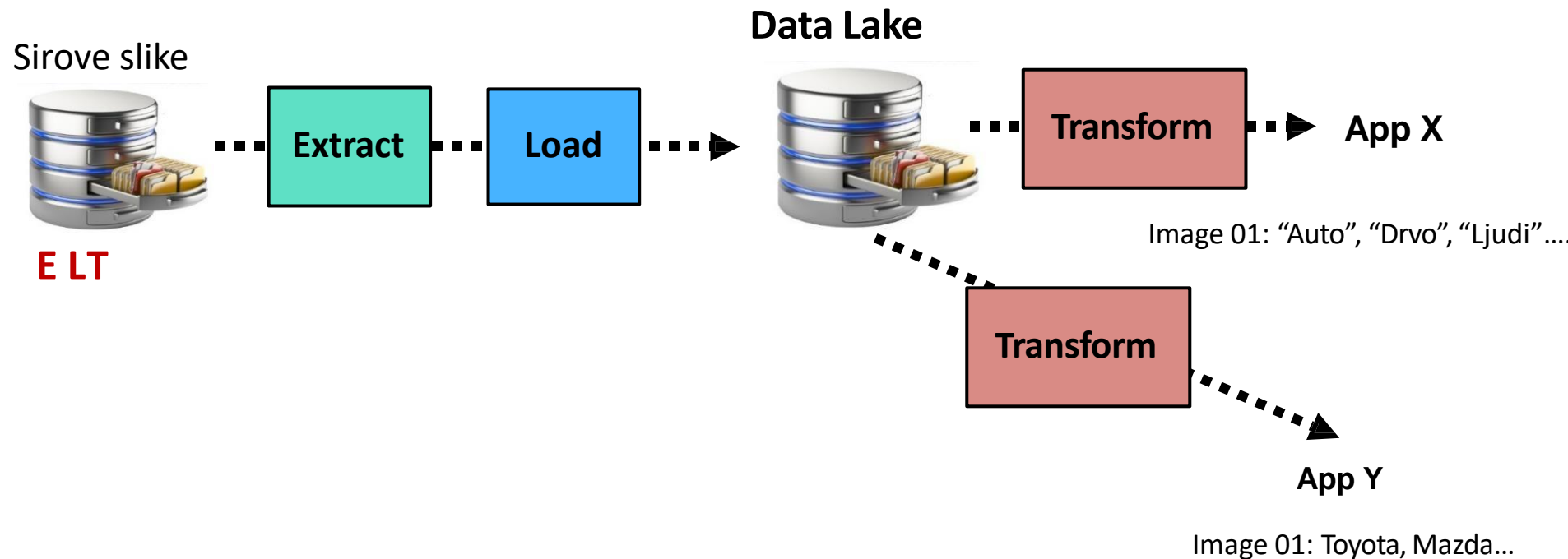
- **ELT - Extract, Load, Transform**

- Proces uzimanja (**extracting**) sirovog podatka (izvor podataka) i unos (**load**) podataka u bazu podataka bez transformacije originalne strukture
- **Ciljna aplikacija** radi transformaciju strukture podataka



Premeštanje podataka između sistema OLTP - OLAP

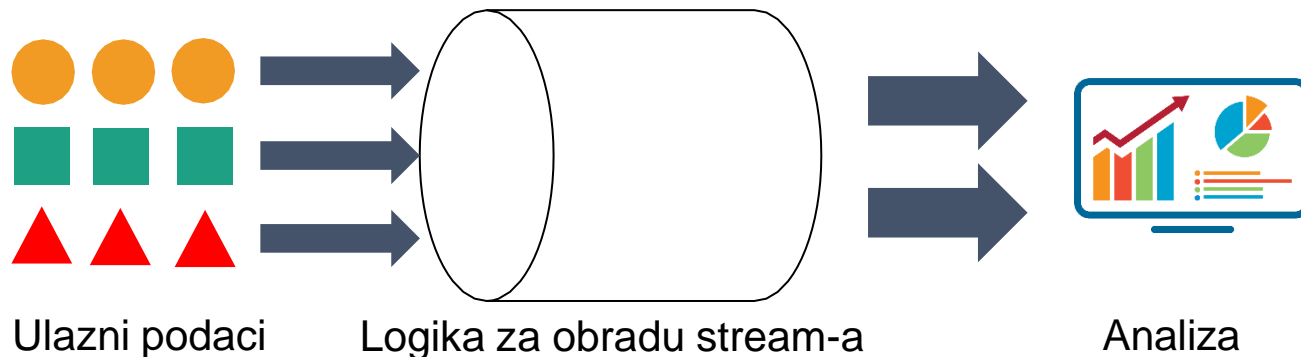
- App X i App Y
 - Postiže se veća fleksibilnost i brža obrada podataka jer je proces transformacije vremenski zahtevan
 - Nedostatak je što se zahteva veći prostor i što podaci nisu optimizovani za datu aplikaciju (use case)



Batch i Stream Obrada

Učestalost premeštanja podataka između sistema

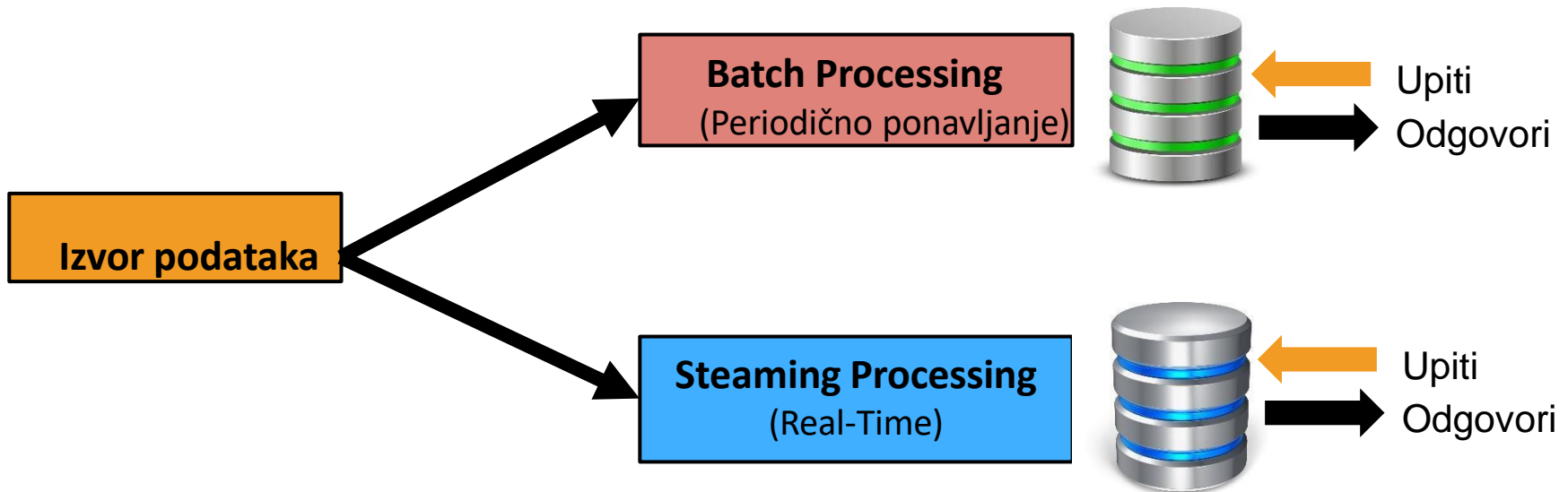
- **Batch obrada**
 - Premeštanje velike količine podataka odjednom na odredišni sistem
 - Obično u noćnim satima (**off-peak times**), **ponavlja se u zakazanim intervalima**
 - Npr. na svakih 24h , u 02:00
 - Korisno je ukoliko odredište **ne zahteva podatke u realnom vremenu**
 - Npr. obrada finansijskih podataka u serijama (in batches)
- **Stream obrada**
 - Akcija nad podatkom se sprovodi odmah čim je kreiran
 - Aplikacije u realnom vremenu (**Analytic real time app**)
 - Procesiranje je dizajnirano da se stream konstatno obrađuje



Batch i Stream Obrada

▪ Lambda Arhitektura

- Sistem kombinuje obe opcije procesiranja paralelno tj obe opcije su ugrađene unutar iste arhitekture
- Arhitektura je dizajnirana da obradi masivnu količinu podataka tako što se koriste prednosti obe metode prenosa (procesiranja) podataka
- Aplikacije koje rade obradu u realnom vremenu i aplikacije koje rade obradu konsolidovanih podataka



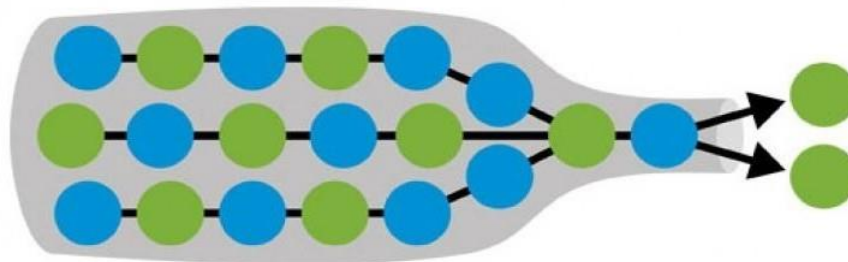
Skaliranje

▪ Skalabilnost sistema

- Skaliranje je proces upravljanja resursima u sistemu da bi se **zadovoljile tražene performanse**
 - Resursi - CPU, Memorija, Prostor za skladištenje....
 - Prekomerno iskorišćeno (Over-utilized), Nedovoljno iskorišćeno (Under-utilized)
 - Tekući procesi (Ongoing process)

▪ Skaliranje Baze podataka

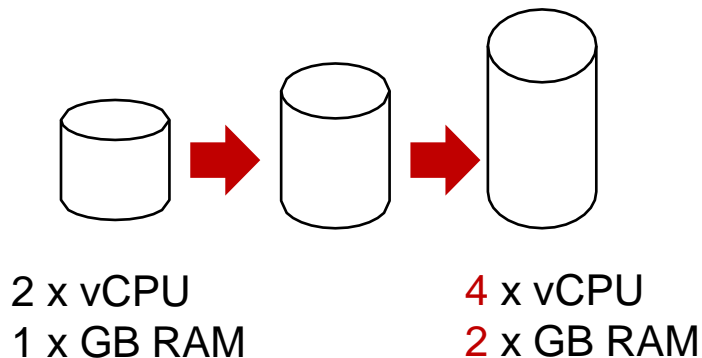
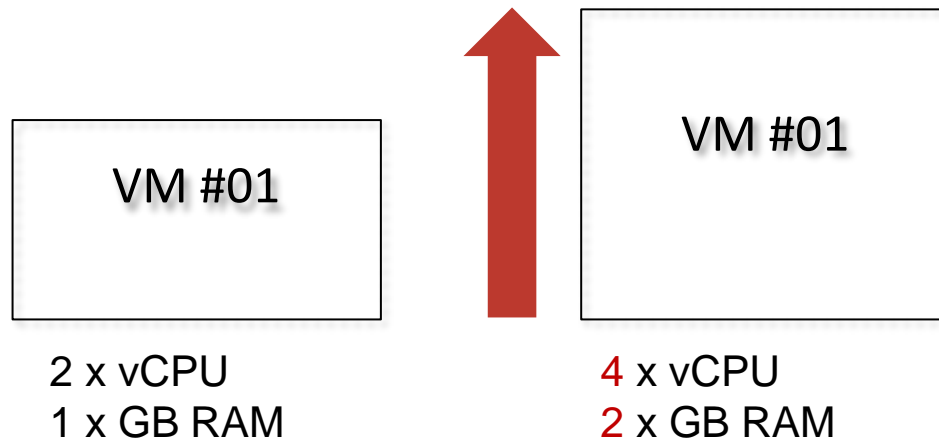
- Baze podataka su najčešće **usko grlo** (bottleneck) u aplikaciji
- Skaliranje baze podataka je vrlo bitno
- **Dve opcije skaliranja baze podataka**
 - Vertikalno skaliranje
 - Horizontalno skaliranje



Vertikalno Vs Horizontalno Skaliranje

- Skaliranje aplikacije i baze podataka je danas najveći izazov

Vertikalno Skaliranje (up/down)



Horizontalno Skaliranje (out/in)

