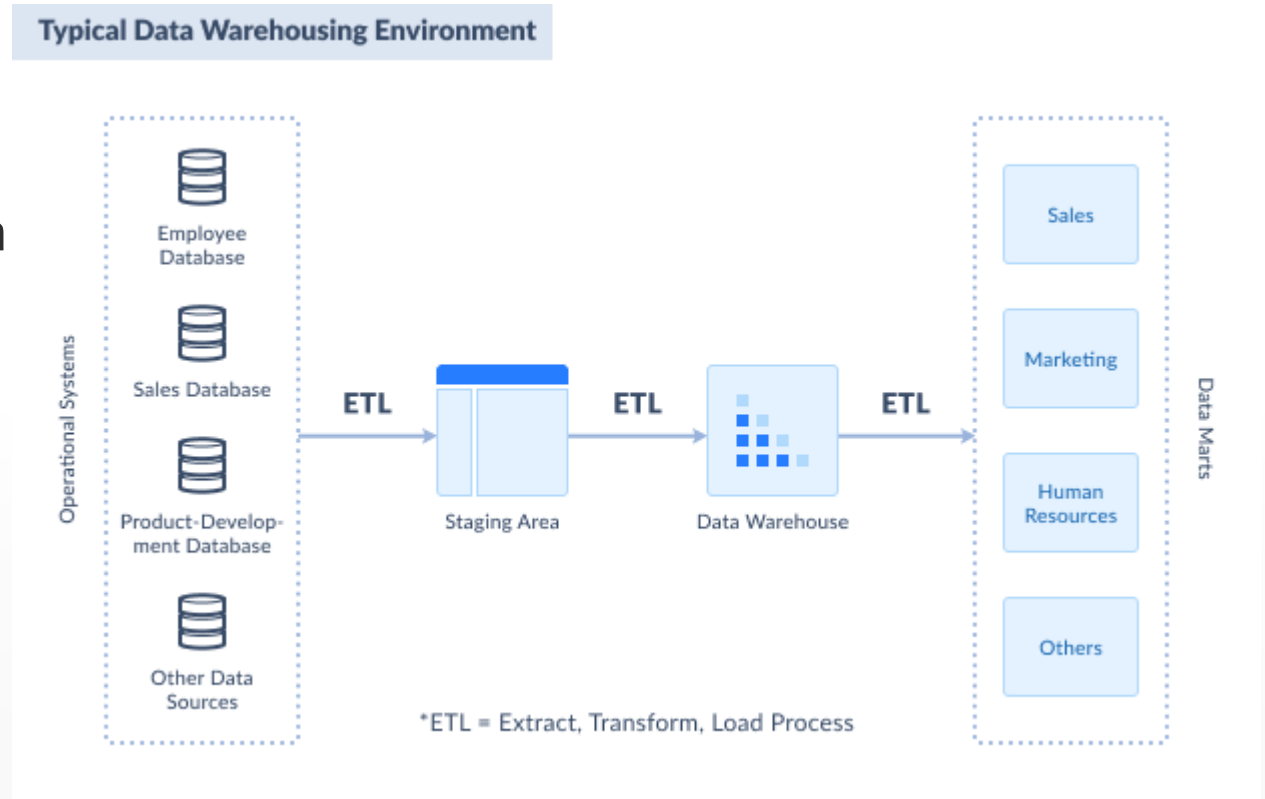


Baze podataka 2

2021/22

Data Warehouse

- Centralno skladište informacija pripremljenih za analizu
- Podaci se preuzimaju iz različitih izvora podataka.
- Podaci se redovno ažuriraju kako bi se odluke donele na vreme.
- Denormalizovana BP

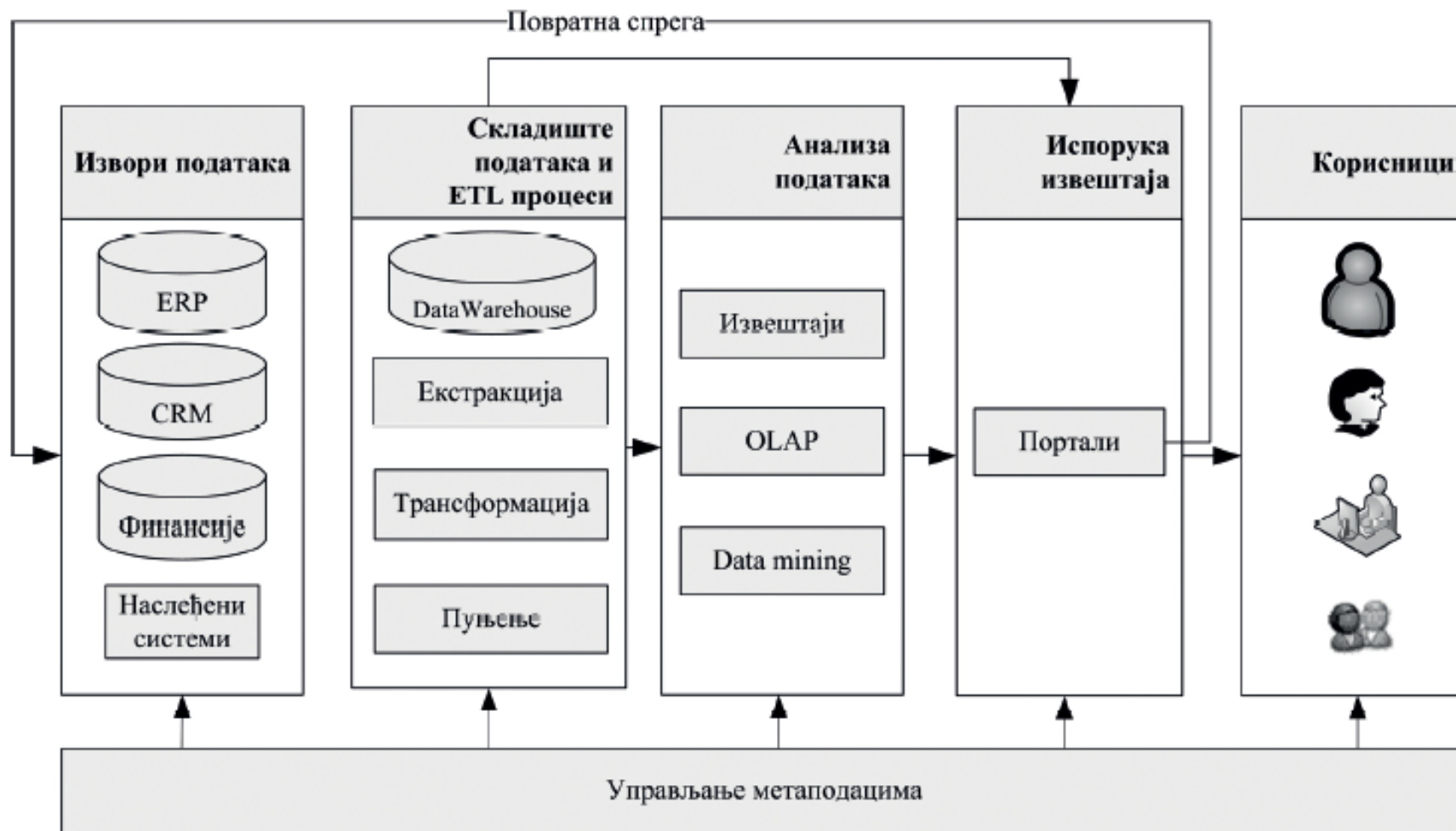


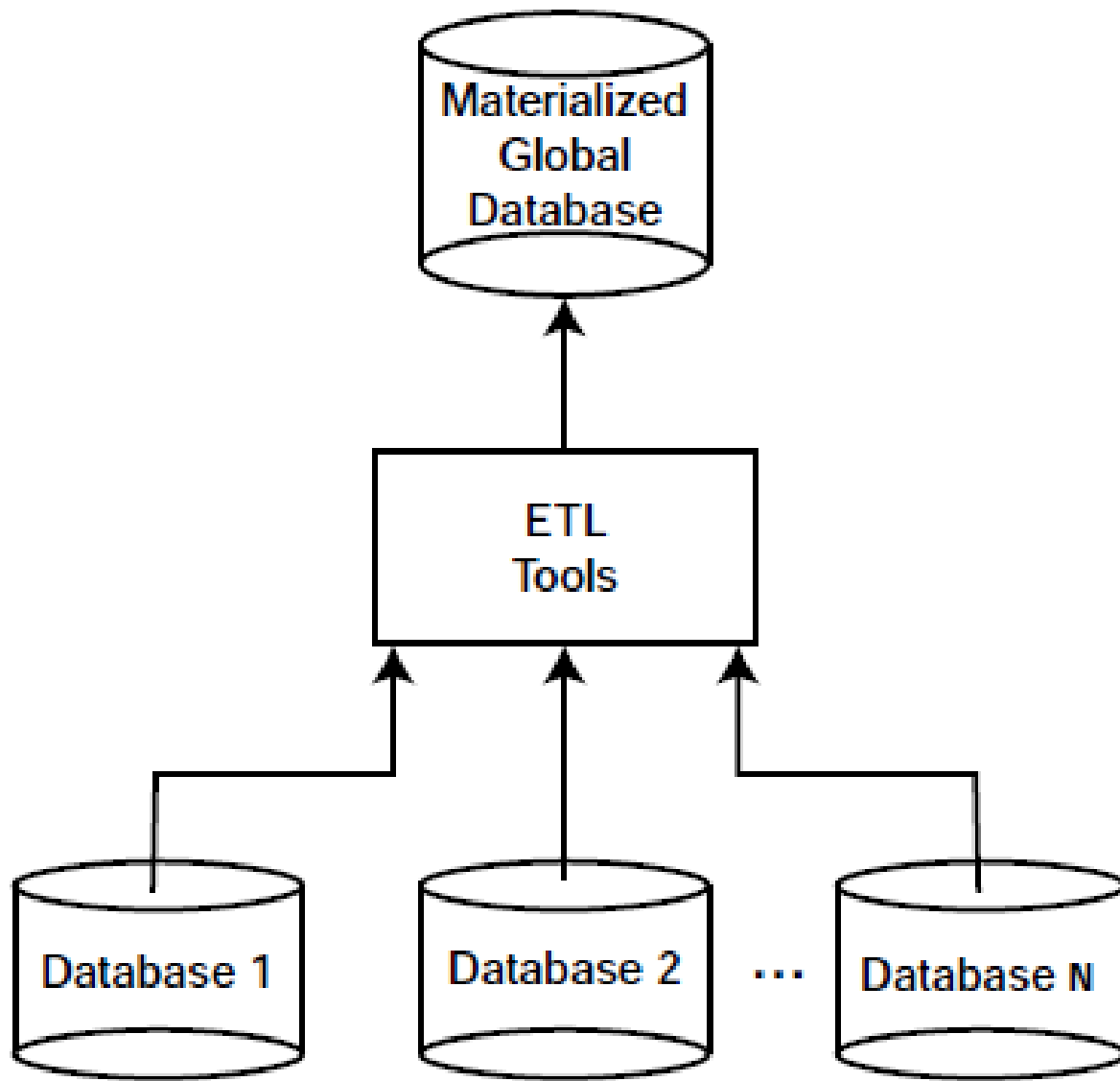
Data Warehouse

Typical Data Warehousing Environment



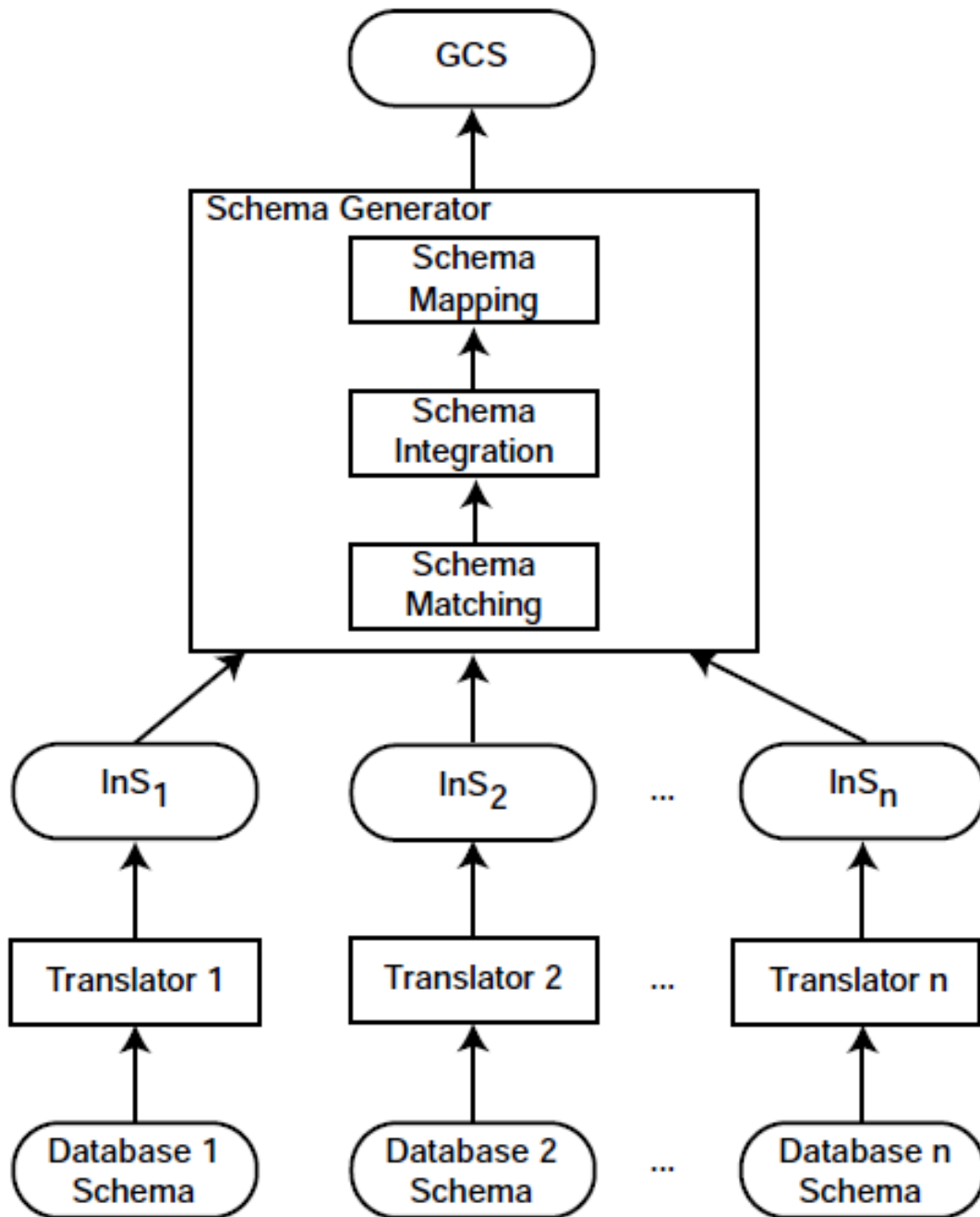
BI lanac





Data Warehouse Approach

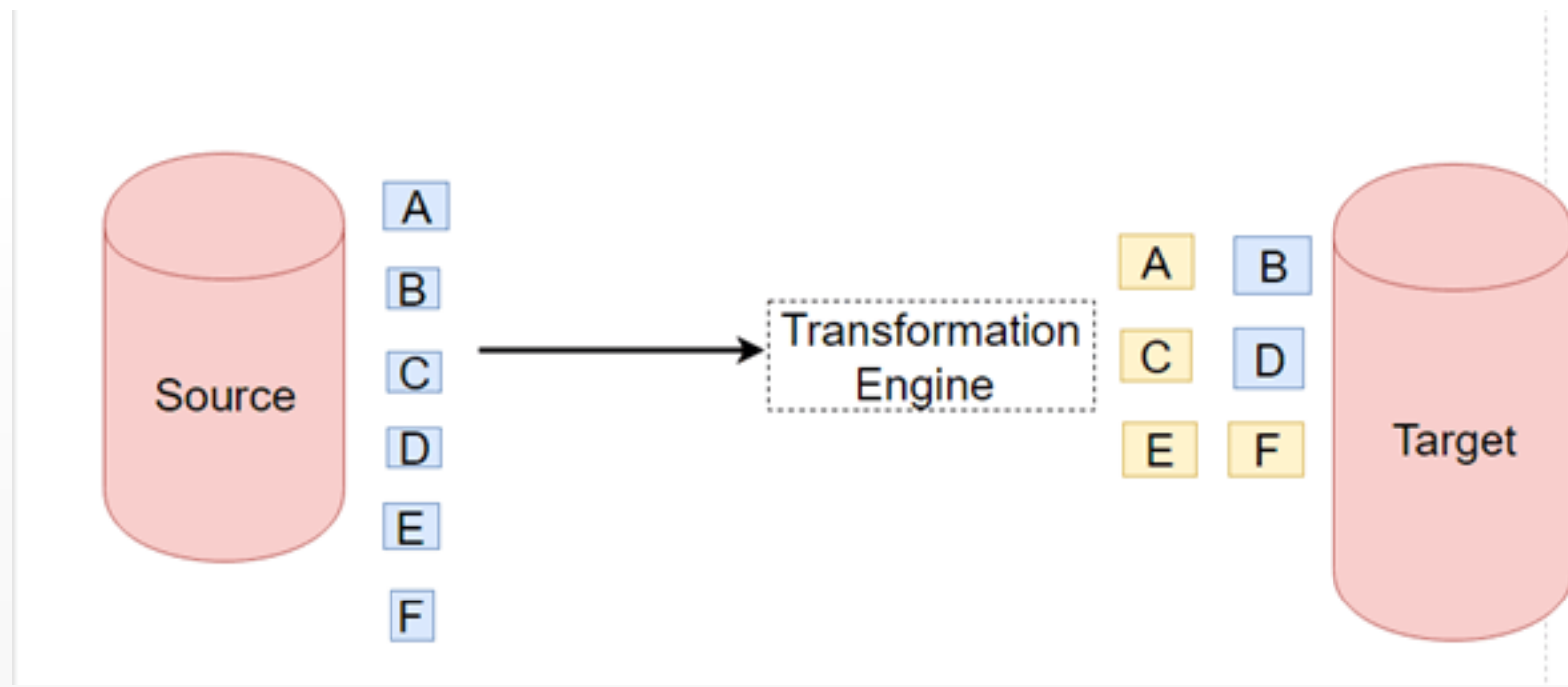
Data Warehouse



Local -> Global Schema

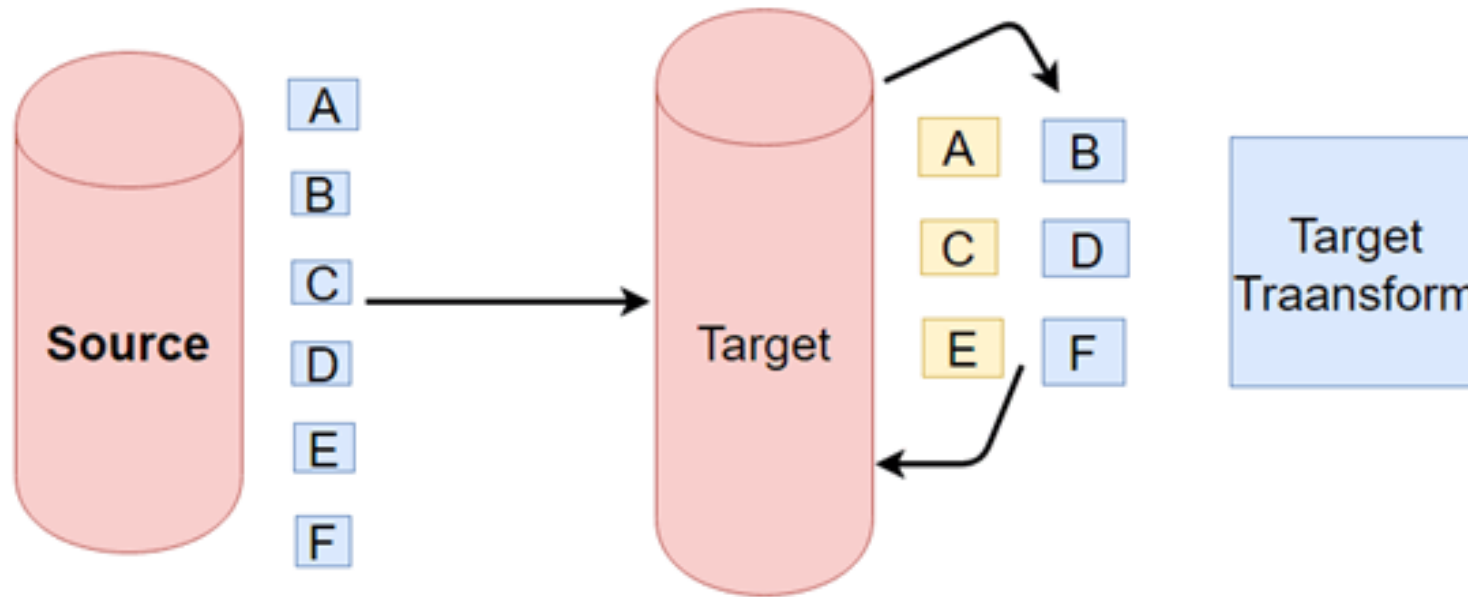
ETL (Extract, Transform, Load)

- ETL alat izvaja podatke iz različnih RDBMS izvora, zatim transformiše podatke i učitava ih u DW sistem.



ELT (Extract, Load, Transform)

- ELT dozvoljava ciljnom sistemu da izvrši transformaciju. Podaci su prvo kopirani na cilj, a zatim transformisani na mestu.



| | DATA LAKES | DATA WAREHOUSE |
|---------------------|---|--|
| Svrha | Isplativo skladištenje velikih podataka | Analitika za poslovne odluke |
| Podaci | Nestrukturirani i strukturirani podaci iz različitih izvora podataka | Istorijski podaci koji su strukturirani tako da odgovaraju šemi relacione baze podataka |
| Skladištenje | Sadrži sve podatke u sirovom, nestrukturiranom obliku i može ih čuvati neograničeno – za trenutnu ili buduću upotrebu. | Sadrži strukturirane podatke koji su očišćeni i obrađeni. |
| Korisnici | Data scientis i inženjeri koji više vole da proučavaju podatke u njihovom sirovom obliku da bi stekli nove, jedinstvene poslovne uvide. | Menadžeri i poslovni korisnici koji žele da steknu uvid iz poslovnih KPI-ja. Podaci već strukturirani da daju odgovore na unapred definisana pitanja za analizu. |

| | DATA LAKES | DATA WAREHOUSE |
|----------------|---|---|
| Analiza | Predictive analytics, machine learning, data visualization, BI, big data analytics. | Data visualization, BI, data analytics. |
| Šema | Šema se definiše nakon što se podaci uskladište. | Šema se definiše pre skladištenja. Ovo produžava vreme potrebno za obradu podataka, ali kada se završe, podaci su spremni za doslednu i pouzdanu upotrebu u celoj organizaciji. |
| Procesiranje | ELT (Extract, Load, Transform). podaci se izdvajaju iz izvora i čuvaju DL skladišti, a strukturiraju se samo kada je to potrebno | ETL (Extract, Transform, Load). podaci se izdvajaju iz izvora(a), pročišćavaju, a zatim strukturiraju tako da budu spremni za analizu |
| Cena | Troškovi skladištenja - manji Održavanje i upravljanje jednostavnije - troškovi manji | |
| Zadaci | Storing data and big data analytics, like deep learning and real-time analytics | Typically read-only queries for aggregating and summarizing data |
| Opseg podataka | Čuva sve podatke koji se mogu koristiti—mogu zauzeti petabajte! | Čuva samo podatke relevantne za analizu |

| | DATA LAKE | DATA LAKEHOUSE | DATA WAREHOUSE |
|---------------|---|---|--|
| Type | Structured, semi-structured, unstructured | Structured, semi-structured, unstructured | Structured |
| Not available | Relational, non-relational | Relational, non-relational | Relational |
| Schema | Schema on read | Schema on read, schema on write | Schema on write |
| Format | Raw, unfiltered, processed, curated | Raw, unfiltered, processed, curated, delta format files | Processed, vetted |
| Sources | Big data, IoT, social media, streaming data | Big data, IoT, social media, streaming data, application, business, transactional data, batch reporting | Application, business, transactional data, batch reporting |
| Scalability | Easy to scale at a low cost | Easy to scale at a low cost | Difficult and expensive to scale |
| Users | Data scientists | Business analysts, data engineers, data scientists | Data warehouse professionals, business analysts |
| Use cases | Machine learning, predictive analytics | Core reporting, BI, machine learning, predictive analytics | Core reporting, BI |