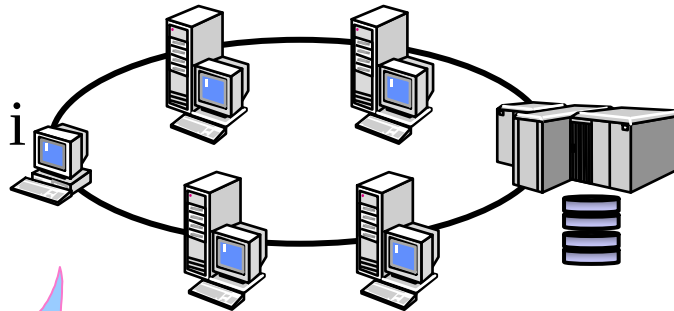


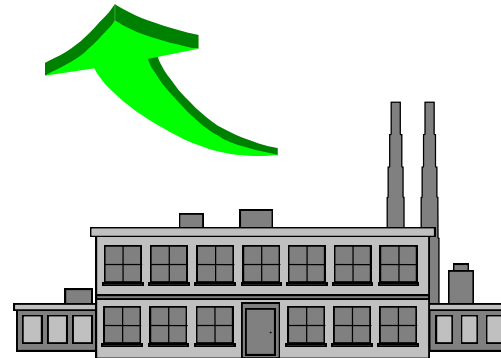
Organizacija podataka

Eksterno okruženje

Danas e-biznis preduzeće ne može opstati niti uspješno poslovati bez kvalitetnih podataka o sopstvenom poslovanju i externom okruženju.



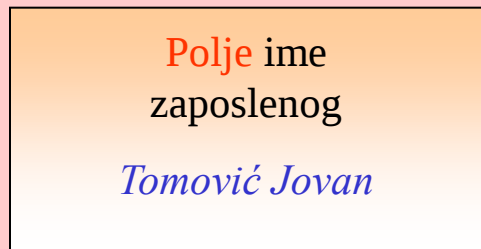
Podaci predstavljaju vitalni organizacioni resurs kojim se mora upravljati kao bilo kojim drugim poslovnim sredstvom !



Podatak - osnovni koncepti

- **Karakter** je elementarna jedinica podatka koja se sastoji od jednog alfabetskog, numeričkog ili nekog drugog znaka (**L**)
- **Polje (field)** je niz karaktera koji predstavlja atribut tj. osobinu nekog entiteta, na primjer *ime zaposlenog* (**Ivan**), *iznos plate* (**3.456,78**)
- **Slog ili zapis (record)** čini više povezanih polja koja predstavljaju skup atributa koji opisuju neki entitet, na primjer skup atributa platnog spiska *ime, broj osiguranja, iznos plate* (**Janković Ivan; 23/ 345; 230,70**)

- **Datoteka (file)** skup slogova istog tipa, koji predstavlja skup entiteta istog tipa, na primjer **Zaposleni**
- **Baza podataka (database)** je integrisani skup logički povezanih slogova koji su smješteni u posebnim datotekama, skup logički povezanih datoteka tj. skupova entiteta, na primjer baza podataka **Personal** koja može sadržati datoteke tj. skupove entiteta *Zaposleni, Izdržavana lica, Radno iskustvo, Platni spisak* itd..



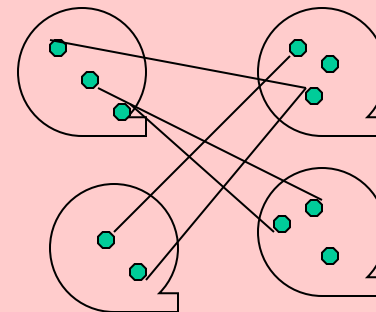
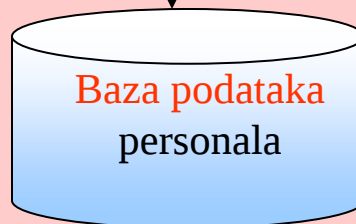
ATRIBUT



ENTITET



**SKUP
ENTITETA**



Sistemi za upravljanje bazama podataka DBMS

- DBMS (Data Base Management System) - softverski paket koji predstavlja interfejs između korisnika i baze podataka.
- Aktivnosti DBMS-a : razvoj, direktno pretraživanje, održavanje baze podataka, razvoj aplikacija
- Oracle, IBM DB2, Microsoft Access, Lotus Approach ...

Razvoj

- Korisnik (manje lične baze)
- Projektanti i administratori (DBA -Database Administrator) (velike organizacije sa klijent/server ili mainframe baziranim sistemom)
- koriste DDL (Data Definition Language) koji se nalazi u okviru DBMS-a za specifikaciju i generisanje podataka, odnosa i strukture baze, kao i za njihovu modifikaciju ukoliko je to potrebno.
- Te informacije tj. specifikacije i definicije podataka čuvaju se u okviru *rečnika podataka* kojim upravlja DBA.

- **Rečnik podataka** je alat za administraciju baze podataka.
- kompjuterski katalog (direktorijum) koji sadrži podatke o podacima tj. **metapodatke**.
- Na primjer, rečnik podataka može sadržati nazive i opise svih tipova slogova kao i njihove međusobne veze, ograničenja pristupa krajnjih korisnika i aplikativnih programa i drugo.

Direktno pretraživanje

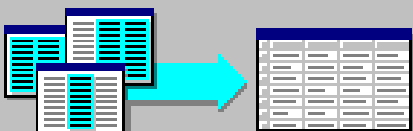
- Direktno pretraživanje baze podataka korisnici obavljaju pomoću upitnih jezika ili generatora izvještaja koje DBMS sadrži u sebi.
- Neposredne odgovore dobijaju ili u vidu ekranskih prikaza ili štampanih izvještaja.
- Upitni jezici ne zahtijevaju kompleksno programiranje već se definiše kratki i jednostavni ad hoc zahtjev.
- **SQL (Structured Query Language)** je upitni jezik kojeg posjeduje veliki broj DBMS-ova. Osnovni oblik SQL upita je

SELECT ... FROM ... WHERE

GUI - *point and click* metod za definisanje upita

Simple Query Wizard

Which fields do you want in your query?
You can choose from more than one table or query.



Tables/Queries:
Table: kupci

Available Fields:
ADRESA
TELEFON

Selected Fields:
SIF_KUPCA
NAZIV

Buttons: >>, <, <<

Cancel < Back Next > Finish

Komparacija upita prirodnim jezikom i SQL upita za Microsoft Access.

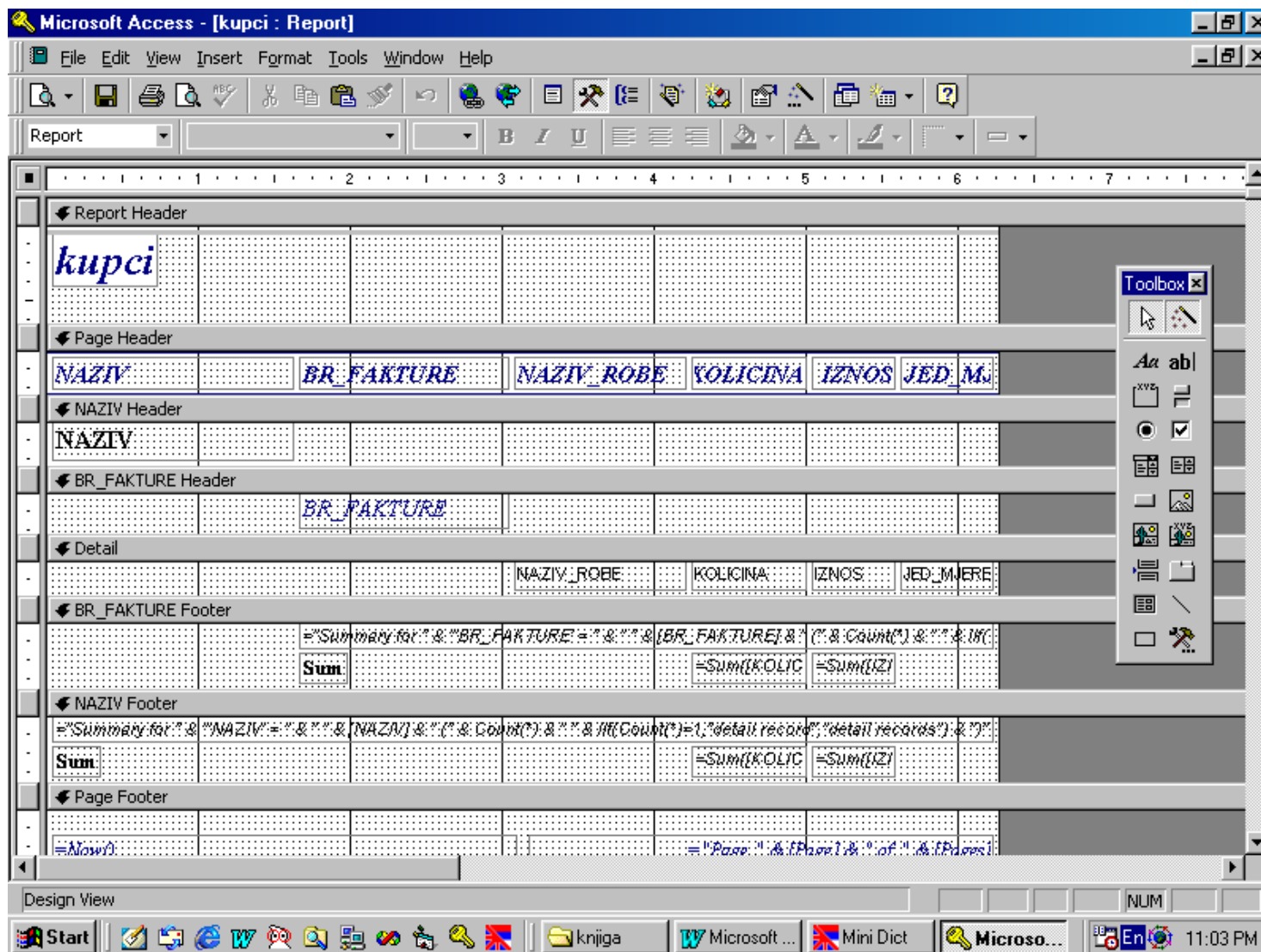
Prirodni jezik

WHAT CUSTOMERS DIDN'T HAVE ANY ORDERS LAST MONTH ?

SQL upit (MS ACCESS)

```
SELECT [Customers].[Company Name], [Customers].[Contact Name]
FROM [Customers]
WHERE not Exists {SELECT [Ship Name] FROM [Orders] where Month {[Order
Date]} = 12 and Year {[Order Date]} =2004 and [Customer].[Customer ID]= [Orders].
[Customer ID]}
```

Generator izvještaja omogućava jednostavno formatiranje informacija tj. izvještaja i njihovo generisanje- štampu.



Održavanje

- kontinuirane promjene nad podacima koje odražavaju nove poslovne transakcije ili neke druge poslovne događaje
- druge vrste promjena koje osiguravaju tačnost (konzistentnost) podataka u bazi
- proces održavanja baze podataka obavljaju programi za procesiranje transakcija i korisnički aplikativni programi

Razvoj aplikacija

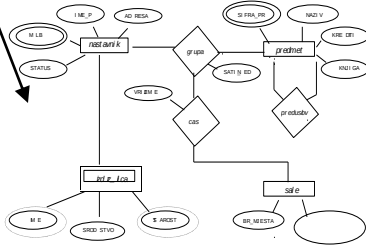
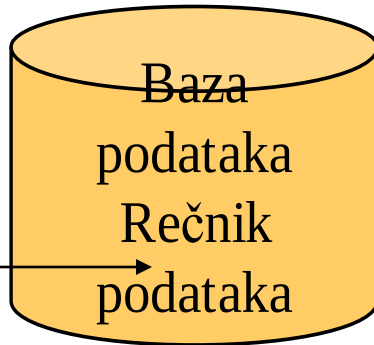
- krajnjim korisnicima, sistem analitičarima i programerima obezbjeđuje interne 4GL jezike i ugrađene alate za razvoj softvera.
- Posao aplikacionih programera je znatno olakšan jer ne moraju razvijati detaljne procedure za upravljanje podacima.
- Oni jednostavno koriste naredbe DML (Data Manipulation Language) jezika koje pozivaju DBMS da izvrši potrebne aktivnosti u vezi upravljanja podacima (unos novih podataka, brisanje, izmjene, pretraživanje).

Funkcija	DBMS komponenta
Definisanje baze podataka	Jezici i grafički alati za definisanje entiteta, veza, integritetskih ograničenja i prava pristupa
Neproceduralni pristup	Jezici i grafički alati za pristup podacima bez kompletnog kodiranja
Razvoj aplikacija	Grafički alati za razvoj menija, formi za unos podataka i izvještaja, 4GL jezici
Proceduralni jezički interfejs	Jezici koji kombinuju neproceduralni pristup podacima sa svim mogućnostima višeg programskog jezika
Procesiranje transakcija	Kontrolni mehanizmi koji sprečavaju konflikte prilikom konkurentnog dostupa i oporavljaju izgubljene podatke u slučaju kraha sistema
Podešavanje baze podataka	Alati za praćenje i poboljšanje performansi baze podataka.

Projektanti i
administratori

1. Razvoj

Konceptualni
model
podataka

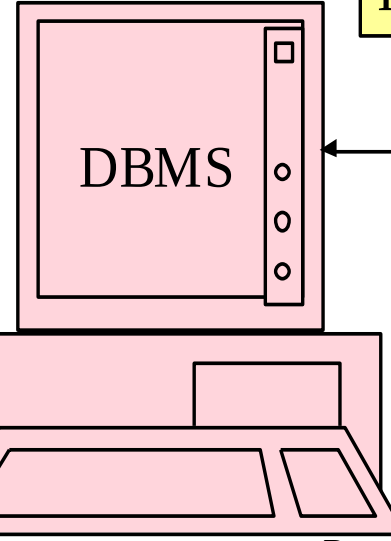


Programeri

4. Razvoj
aplikacija

DDL,
DML

DDL



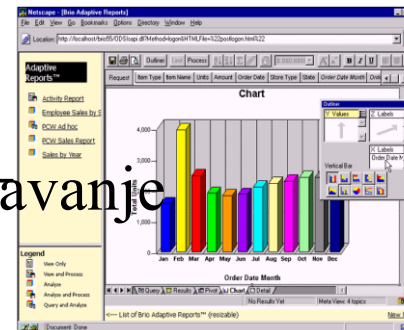
Korisnici

2. Direktno
pretraživanje

SQL

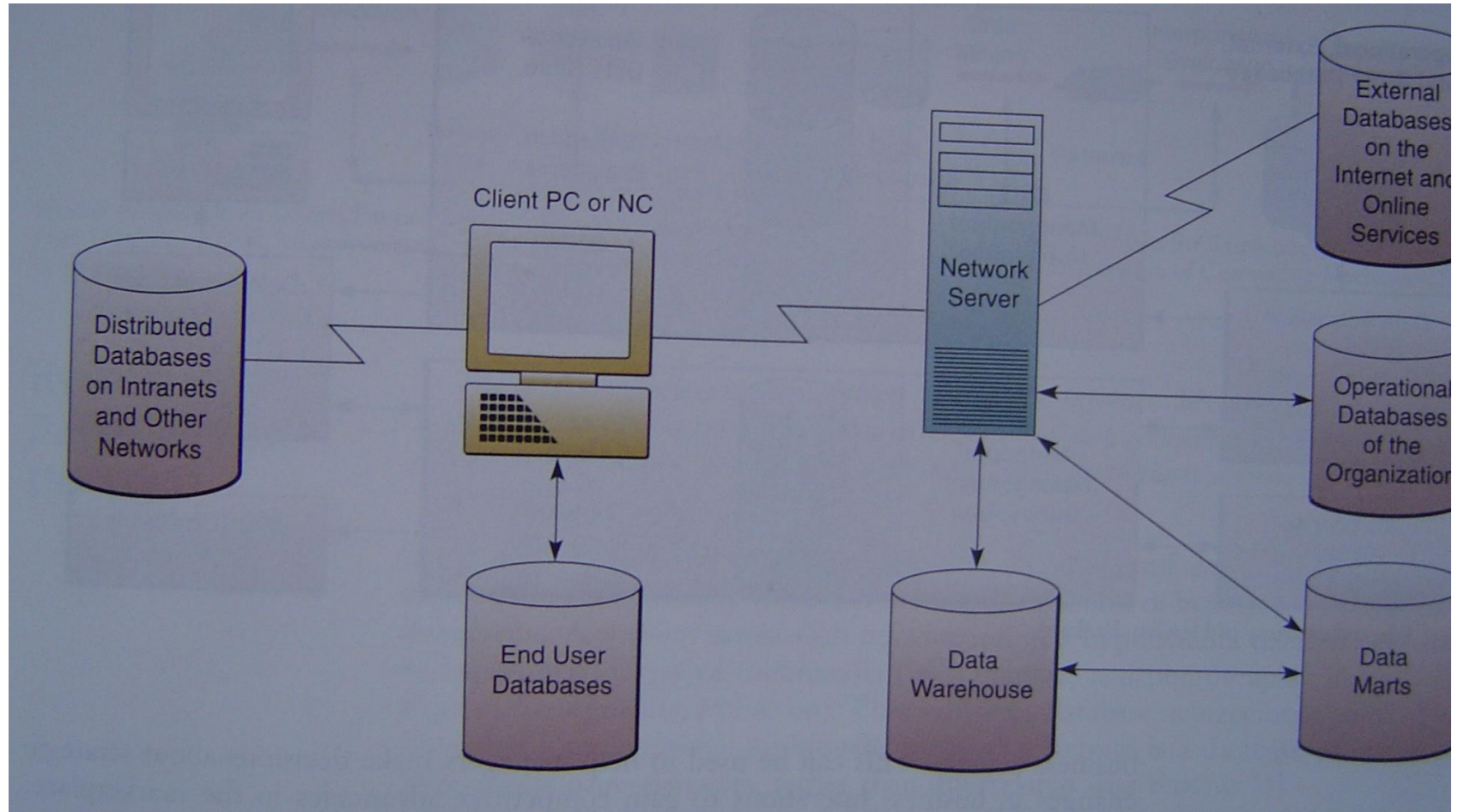
Programski
jezici i
DML, 4GL

3. Održavanje



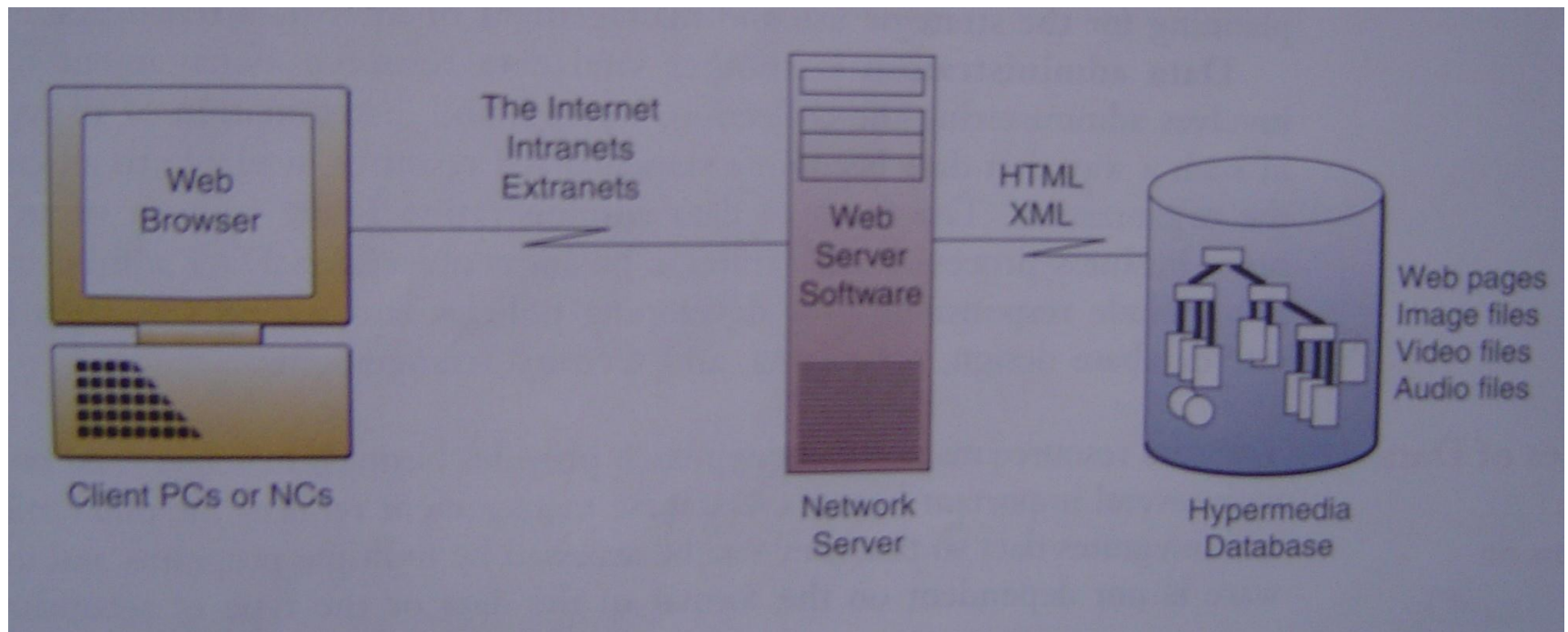
Aplikativni
program

Tipovi baza podataka



Tipovi baza podataka

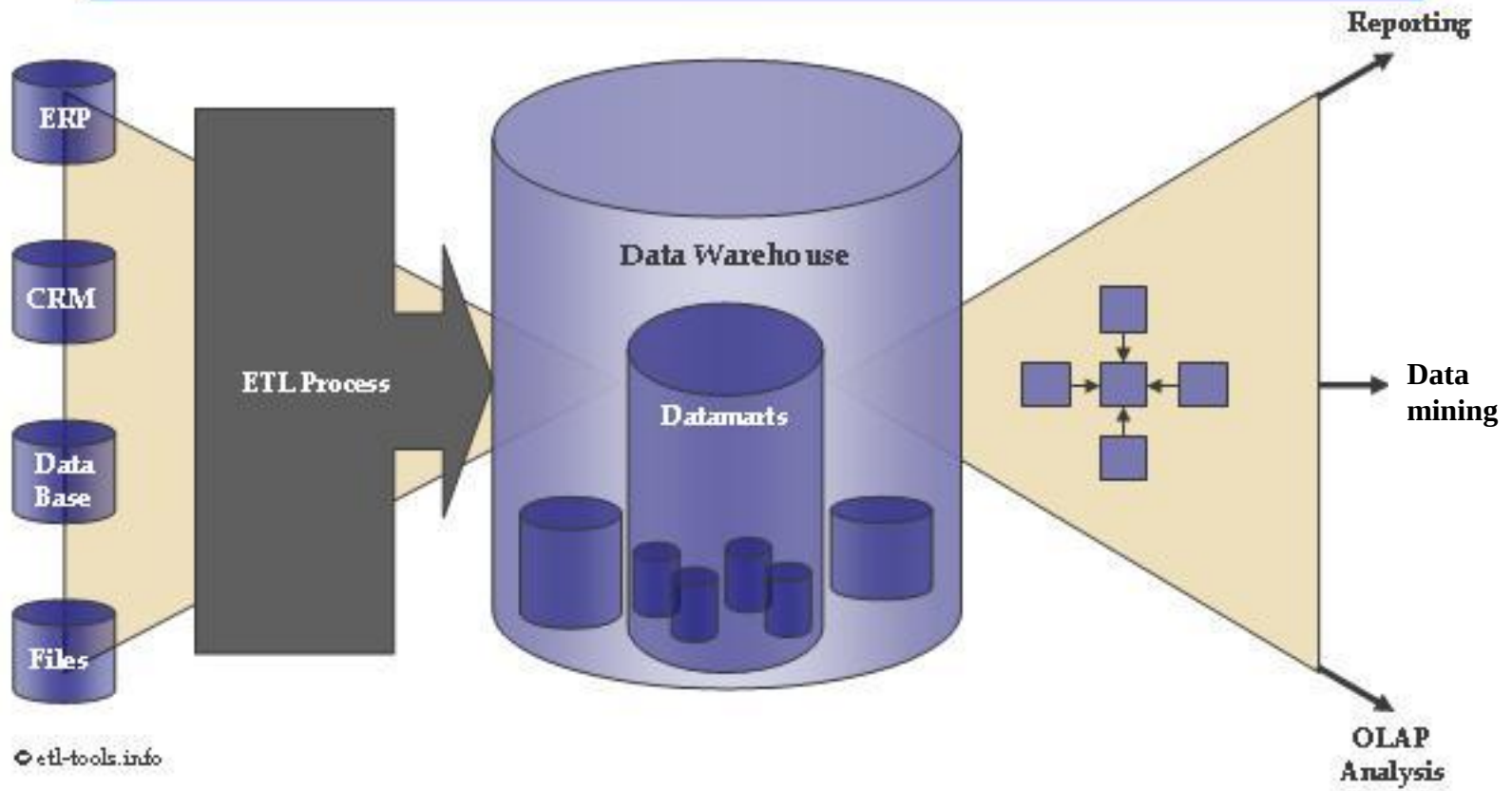
- **Operacione (transakcije)** baze podataka, sadrže aktuelne podatke nastale u poslovnim transakcijama (npr. podaci o zalihama (ulaz-izlaz), računovodstveni podaci...)
- **Distribuirane baze** podataka su kopije djelova transakcione baze koje su dislocirane na umreženim serverima (replikacija).
- **Eksterne** baze podataka, na primjer na Internetu statističke baze, baze publikacija,...
- **Hypermedijalne Web** baze podataka su baze koje sadrže hiperlinkovane strane multimedije (tekst, grafiku, video i audio sadržaje i drugo). Ova baza je skup povezanih multimedijalnih web strana.



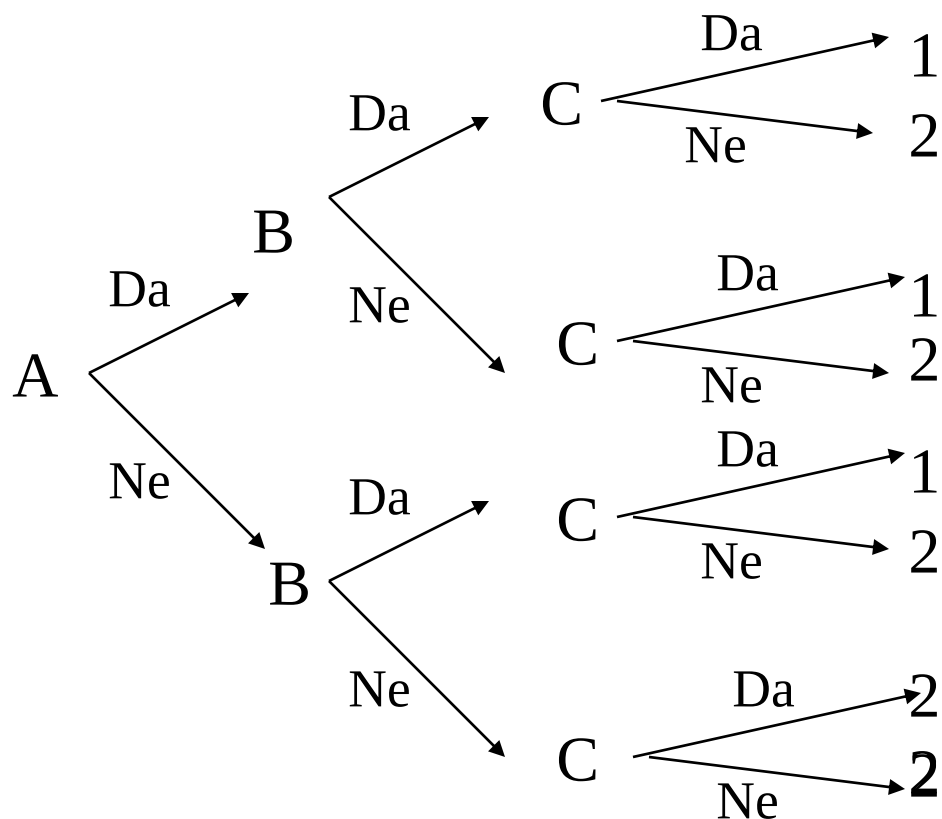
Tipovi baza podataka

- **Data Warehouse** sadrži tekuće i istorijske podatke ekstrahovane iz više različitih operacionih i eksternih baza. Ima višedimenzionalnu strukturu. Koristi se za poslovne analize OLAP, Data Mining i drugim tehnikama, kao podrška odlučivanju.
- **Data mining** procesom podaci se u data warehouse-u analiziraju radi automatskog otkrivanja skrivenih obrazaca ponašanja i trendova u velikoj količini podataka nastalim u poslovnim aktivnostima organizacije u dužem periodu.
- Koristi napredne algoritme za učenje iz podataka – mašinsko učenje (drveta odlučivanja, neuronske mreže, genetski algoritam) kao i različite statističke₁₉ tehnike.

Business Intelligence



Klasifikaciona stabla



Legenda:

A - stalan posao

B- odgovarajuća
aktiva

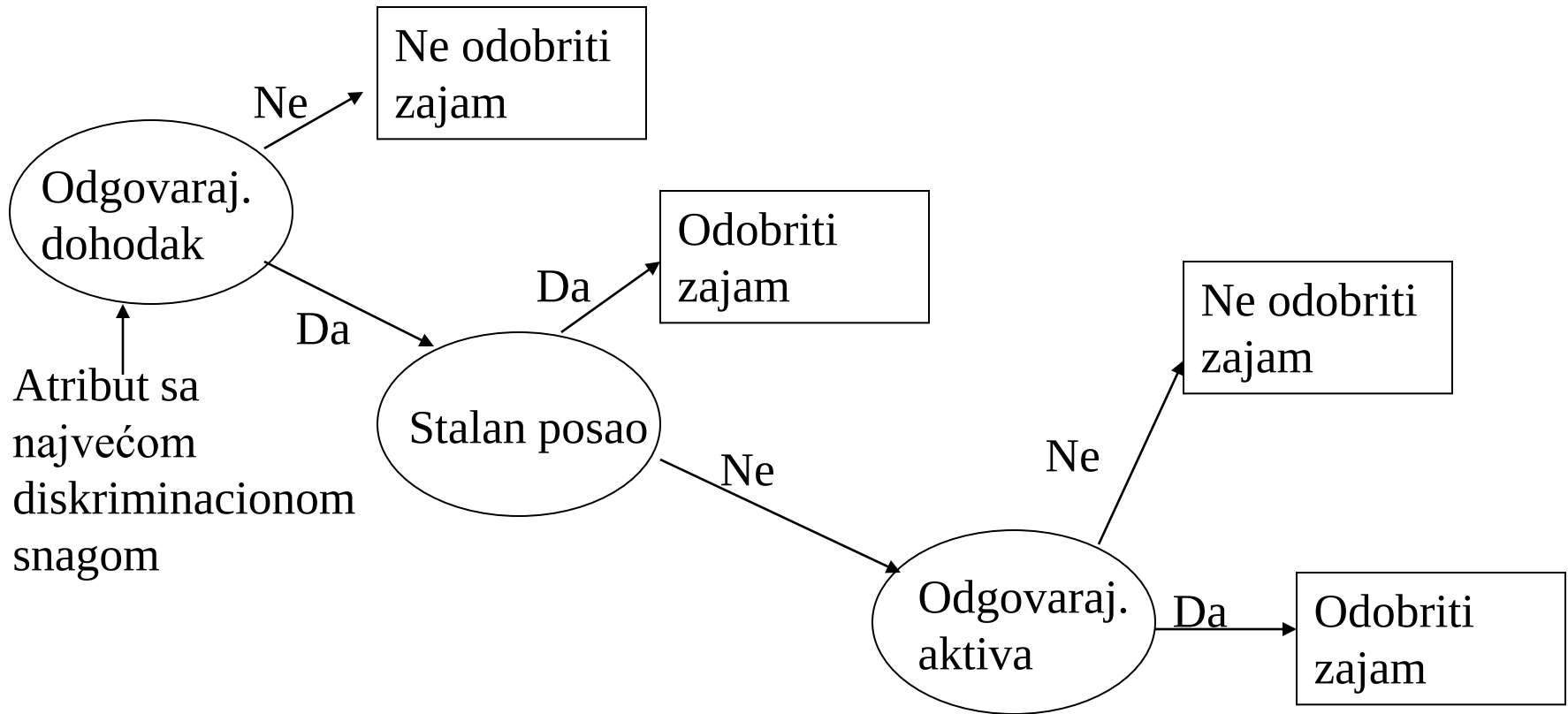
C- odgovarajući
dohodak

1- odobriti kredit

2- ne odobriti
kredit

Stablo se formira staističkim metodama
(HI-KVADRAT test) !!! CHAID

Klasifikaciona stabla



ID3 algoritam daje optimalno stablo (bez nepotrebnih čvorova i grana) !!!

Data Mining-primjer

- Standardni upitni alati
mod verifikovanja

Koliki je iznos prodaje na sjeveru a koliki na jugu zemlje ?

- Data mining alati
mod otkrivanja

Koji su to faktori koji utiču na iznos prodaje ?

Problemi i izazovi upravljanja podacima

- definisanje i primjena polisa za vlasništvo i pristup podacima,
- definisanje strateških i tehničkih planova baze podataka,
- setovanje i primjena operativnih procedura,
- upravljanje dizajniranjem i administracijom baze podataka,
- ocjena i izbor softvera za bazu podataka

Fajl sistemi - problemi

Svako funkcionalno područje ima svoj skup nezavisnih fajlova a svaki fajl ima svoj skup programa koji njime upravljaju.

- **redudansa podataka** (dupliciranje podataka u više različitih fajlova),
- **konfuzija** (isti podaci mogu imati sasvim različita značenja u različitim aplikacijama),
- **zavisnost programa od podataka** (promjena formata podatka u nekom fajlu zahtijeva izmjenu svih programa koji pristupaju tom fajlu),

Fajl sistemi - problemi

- **nefleksibilnost** (izvještavanje je samo na osnovu zahtjevnog programiranja, nema mogućnosti direktnih upita i ad hoc izvještaja),
- **nesigurnost podataka** (nema mogućnosti definisanja prava pristupa) i
- **nemogućnost dijeljenja podataka između više aplikacija.**

DBMS pristup

- reducira se dupliciranje podataka
- podaci se integrišu i omogućava se pristup od strane više različitih korisničkih aplikacija
- aplikacije su nezavisne od formata podataka
- ad hoc upiti i automatsko generisanje izvještaja bez kompleksnog programiranja
- podržava integritet i sigurnost podataka

DBMS pristup - problemi

- tehnološka kompleksnost i velike količine poslovnih podataka kojim se mora upravljati
- Razvoj velikih baza podataka sa kompleksnim tipovima podataka i instalacija data warehouse-a može biti jako teška i skupa.
- Komleksnost podataka i softvera može dovesti do produženja vremena procesiranja i sporog odziva.
- Kod distribuiranih baza podataka može doći do nekonzistentnosti podataka a podležnost greškama, upadima i krahovima se povećava

1. Zaokružite u tabeli jedan karakter, polje, slog, fajl, atribut, entitet, skup entiteta.

Broj indeksa	Ime studenta	Adresa	Telefon
22/04	Marković Jovan	Podgorica	2222
97/04	Adžić Ana	Bijelo Polje	3333
111/04	Burić Igor	Bar	5555

2. Fakultetska studentska služba čuva svoje podatke o studentima i ispitima u odvojenim fajlovima na svom PC računaru. Odjel za plate zaposlenih na fakultetu takođe ima svoj PC i svoju nezavisnu kolekciju fajlova. Uprava je odlučila da profesorima u sklopu plate isplati povišicu zavisno od broja studenata koje su ispitali u toku mjeseca. Može li se u odjelu za plate izračunati iznos povišice za profesore ? Zašto? Da li ovdje ima redudanse podataka ? Može li aplikacija iz odjela za plate pristupiti fajlu o ispitima ? Može li studentska služba odmah obezbijediti izvještaj o profesorima i broju obavljenih ispitivanja? Zašto? U slučaju da se format broja indeksa promijenio pa sada više nije tekstualni podatak maksimalne dužine 6 već dužine 8, može li studentska služba proći bez izmjene programa ? Šta bi trebalo uraditi pa da sve ovo bude moguće.

3.3 firme koje čine konzorcijum, imaju svoje operativne baze podataka. Jedna ima Oracle DBMS a ostale dvije SQL server DBMS. Firme su povezane extranet mrežom, baziranom na Internet osiguranim linkovima. Strateški menadžment konzorcijuma želi da analizira sumarne poslovne podatke za sve 3 firme kroz više različitih dimenzija. Za analizu želi da ima aktuelne ali i istorijske podatke. Konzorcijum je odlučio da u tu svrhu koristi server firme koja ima Oracle DBMS kao i ovaj sistem za upravljanje podacima. Za OLAP analizu koristiće Oracle Discoverer softver iz Oracle-ovog *business intelligence* paketa. Šta konzorcijum mora uraditi sa podacima ? Strateški menadžment bi takođe želio da automatski otkriva faktore koji najviše utiču na količinu prodate robe, kao i karakteristike kupaca čiju su lojalnost izgubili, tj. zašto im kupci otkazuju lojalnost ? Šta se mora nabaviti da bi ovo bilo moguće ?

4. Jedna avio kompanija sa više širom svijeta dislociranih dispečera letova treba da dizajnira svoju bazu podataka. Svaki udaljeni dispečer ima sopstvenu LAN mrežu sa serverom. Intranet ove kompanije baziran je na osiguranim Internet linkovima. Da li je za ovu kompaniju dobra centralizovana baza podataka ? Zašto? Kod distribuiranog koncepta kompanija ima na raspolaganju dvije opcije.

- 1) Centralna baza se dijeli na djelove i svaki dislocirani server ima svoj dio podataka sa kojim radi. Promjene na lokalnim fajlovima će biti usklađene sa centralnom bazom na batch osnovi, noću.
- 2) Kompletna centralna baza će biti replicirana (duplicirana) na svim izmještenim lokacijama. Svaka promjena napravljena u Frankfurtu, automatski će biti replicirana u Njujorku ili Hong Kongu. Ova strategija zahtijeva ažuriranje centralne baze na manje od jednog sata.

Koja od ove dvije opcije je bolja za ovu avio kompaniju. Od čega to zavisi ?