

## Skladišta podataka

- Relacijski model ne može pružiti bogate analitičke sposobnosti koje moderna poslovanja zahtijevaju.
- Skladište podataka jest subjektivno orijentiran, integriran, postojan i vremenski različit skup podataka koji služi kao potpora odlučivanju.
- Razlike **transakcijski sustav** / **skladišta podataka**
  - o **Trenutni podaci** / **Povijesni podaci**
  - o **Podaci promjenjivi** / **Podaci postojani**
  - o **Velika učestalost transakcija** / **Srednja i mala uč. Transakcija**
  - o **Predvidljivo korištenje** / **Nepredvidljivo korištenje**
- Prikaz podataka u dimenzijskom modelu – predstavlja podatke u jednostavnom, intuitivnom obliku za dobar pregled
  - o Zvezdast – Činjenična tablica u sredini (Normalizirana, velik broj zapisa), okolo dimenzijske tablice (Ne normalizirane, mali broj zapisa), činj. tablica sadrži
    - Ključeve – FK na dimenzijske tablice
    - Mjere – numeričke vrijednosti
  - o Pahuljast – nije normaliziran
- Prednost dimenzijskog
  - o Predvidljiva, standardizirana struktura, omogućuje gomilu programa
  - o Lako proširiv model
  - o Otporan na neočekivane korisničke akcije
- Najzahtjevniji proces – ETL (Extracting, Transformation, Loading)
- OLAP – On Line Analytical Processing – analize, izvještaji, struktura kocke,
  - o F - fast
  - o A – analiza, kompleksna poslovna logika ali lako korištenje
  - o S – šeranje, tajnost podataka
  - o M – multidimenzijski, potpora za višestruke hijerarhije
  - o I – information, koliko ulaznih podataka može podnijeti, ne koliko troši na pohranu
- Modeli:
  - o ROLAP – relational OLAP – agregati i podaci u relacijske
  - o MOLAP – multidimensional OLAP – agregati i podaci na OLAP server
  - o HOLAP – hybrid OLAP – agregati OLAP, podaci relacijska

## Data mining

- Proces otkrivanja i izlučivanja znanja iz velikih baza podataka
- Financijske aplikacije, proizvodne aplikacije, medicinske aplikacije...
- Osnovni modeli
  - o Regresija – kvantificira odnos između dvije varijable
  - o Razvrstavanje – kao drvo, na temelju ulaza dobijemo izlaz (velika mudrost)
  - o Grupiranje – K najbližih susjeda, pridruživanje najbližim centrima
  - o Asocijativna pravila – A -> B, u market-basket analysis

## Business Intelligence

- Način kako iskoristiti 90% inače ne analiziranih podataka
- Podaci -> informacije
- Široka kategorija aplikacija i tehnologija koje pospješuju poslovne odluke, a svrha im je skupljanje, analiza i dijeljenje informacija
- Sustavi za podršku odlučivanju, OLAP, statistička analiza, predviđanje, dubinska analiza

## Peer-2-peer sustavi

- Svojstva
  - o Koriste raspodijeljena sredstva da obave zadaću na decentralizirani način
  - o Računalna snaga, podaci, mrežna propusnost, prisutnost...
  - o Stara ideja, ali zamah tek nedavno
  - o Skalabilnost – proširivati moguće do u nedogled
  - o Anonimnost – skriveni putovi, enkripcija,
  - o Samoorganizacija
  - o Cijena vlasništva – SETI@home 1% cijene super računala
  - o Namjenska spojnost
  - o Učinkovitost – temeljni problem je usmjeravanje poruka unutar mreže
  - o Sigurnost – osigurati korisnika, a opet zaobići firewall
  - o Transparentnost – imenovanja, administriranja, platforme
  - o Otpornost na pogreške – single point of failure [decentralizacijom], nedostupnost sadržaja, maliciozno ponašanje...
  - o Interoperabilnost – slabo zasad, JXTA dobri temelji
- Definicija
  - o Dijeljene računalnih resursa i usluga direktnom razmjenom između sustava
- Struktura
  - o Strukturirani – globalna raspodijeljena tablica raspršivanja, čvor se brine za podskup, logaritamski broj skokova potrebnih za nalazak podatka
  - o Nestrukturirani – random veze
  - o Polustrukturirani
- Generacije
  - o 1. – jedno ili više centralnih poslužitelja
  - o 2. – decentralizacija
  - o 3. – anonimnost korisnika
- Povijest
  - o UseNet 1979. – članci
  - o DNS 1983.
  - o FidoNet 1984.
  - o ICQ 1996. – chat (centralni poslužitelj uspostavlja vezu)
  - o SETI@home 1999. – analiziraju slike teleskopa, 5,000,000 korisnika
  - o FreeNet 1999. – 3. Generacija
  - o Napster 1999. – centralizirana arhitektura – danas legalan za razmjenu glazbe
  - o Gnutella 2000. – 2. Generacija
  - o eDonkey 2000.

- FastTrack 2001. – superčvorovi, ugašena
- BitTorrent 2001. – potpuno decentraliziran, uz n centraliziranih računala
- Tablice raspršenog adresiranja – neće bit

## Semantički web

Ideja o mogućnosti da podaci na webu budu definirani i povezani na način na koji bi se mogli koristiti za automatizaciju, integraciju i ponovnu iskoristivost u različitim aplikacijama.

- RDF – resource description framework
  - Subjekt – predikat – objekt
  - Trojna notacija, XML notacija, graf
- Ontologija – model podataka koji predstavlja skup pojmova unutar neke domene i veze među tim pojmovima
  - RDF – iskazivanje tvrdnji
  - RDFS – vokabular
  - OWL – opisivanje veza između vokabulara, ograničenja...
    - Proširuje mogućnosti RDF-a i RDFS-a
    - Ograničenja
    - Preslikavanja
    - Složeni razredi
  - GRDDL – XML->RDF
  - D2RQ – baza -> RDF
  - Upit - SPARQL endpoint -> RDF
- RDF spremišta
  - In memory – Jena, Sesame
  - Prirodna – najbolje performanse - Virtuoso, Sesame Native, Oracle 11g, Allegro graph
  - Baze – Jena + (mySql, PostgreSQL, Oracle, DB2)

## XQuery

```
for $varijabla in doc(„/put/“)/root/some-child//some-descendant
for $varijabla in doc()//tag[@atribut="asda"]/tag[vrijednost>5]
for $varijabla in doc()//tag[podelement]

let $sub := $varijabla/some-sub-tag

where $sub/vrijednost > 8

return <element>{concat($sub/@attr, $sub/podElementData)}</element>
return <element>{ for statement } </element>
return <element>{ dana($/sub/../../nekitag) }</element>
```

## OO baze

```
var n = new Natjecatelj();
db.Store(n);
var grupa = db.Query<Grupa>( p => p.Name == „a1”).Single();
grupa.Add(n);
db.Store(grupa);

var query = db.Query();
query.Constrain(typeof(Grupa));
query.Discard(„propertyName”).Discard(„anotherProperty”).Constrain(„
maraton”).Equal();
var result = query.Execute();

query.Discard(„naziv”).Constrain(„B”).StartsWith().Or(otherQuery...)

Db4o.Configure().ObjectClass
```

## GIS

- ST\_Area(geom)
- ST\_Touches(geom,geom)
- ST\_Within(geom, geomContainer)
- ST\_Equals(geom, geom)
- ST\_GeometryType(geom)
- ST\_Distance(geom)

## OR – mapiranje

Pljuga

## Objektno relacijske

```
CREATE ROW TYPE tKnjiga (  
    ime    LIST(tAuthor NOT NULL),  
    drugo  SET(CHAR(20) NOT NULL),  
    vezni  ROW(naz TIP, prp TIP ...)  
  
CREATE TABLE Knjiga OF TYPE tKnjiga;  
  
INSERT INTO Knjiga VALUES( LIST{ROW(...),ROW(...)}, SET{"", ""}...  
  
CREATE ROW TYPE subType UNDER tKnjiga;  
  
CREATE TABLE sss OF TYPE subType UNDER Knjiga;  
  
CREATE FUNCTION povrsina(krug tKrug) RETURNING FLOAT;  
    RETURN krug.fdklklsdlg;  
END FUNCTION;  
  
SELECT * from Knjiga  
WHERE 'df' NOT IN drugo  
  
SELECT CARDINALITY(ime) from Knjiga
```