

UPRAVLJANJE POSLOVNIM PODACIMA

SKLADIŠTE PODATAKA 2.0

PROF. DR.SC. DRAŽENA GAŠPAR

05.12.2016.

PREZENTACIJE

IZVJEŠĆA ZA DIMENZIJSKI MODEL



Max. 10 minuta



PITANJA ZA DISKUSIJU

- Prva generacija skladišta podataka – otvorena pitanja
- DW 2.0
- Osnovne značajke DW 2.0
- DW 2.0 komponente



PRVA DW GENERACIJA – OTVORENA PITANJA

- Naglasak na izgradnji DW i pohranjivanju podataka, danas na održavanju
- Integriranje podataka se dominatno odnosilo na numeričke, transakcijske podatke
- Nije bilo razmišljanja o životnom ciklusu podataka
- Upravljanje meta i matičnim podacima
- U početku je DW bilo novitet, danas je osnovica za potporu odlučivanju
- U početku je DW viđeno kao podatkovna osnovica za statističke analize, danas je to istraživačko (exploration) DW



DW 2.0. GENERACIJA

- Nova paradigma se fokusira na:
 - Osnovne tipove podataka
 - Podatkovne strukture
 - Kako povezati podatke i formirati moćno skladište podataka koje ispunjava potrebe organizacije za informacijama



DW 2.0. GENERACIJA

- Troškovi DW infrastrukture se ne povećavaju kontinuirano
- Infrastruktura se drži na okupu pomoću meta podataka, meta podaci su kičmeni stup DW 2.0
- Podaci su brzo dostupni. Podaci se smještaju prema vjerojatnosti pristupa
- Prepoznata je potreba za arhiviranjem
- DW privlači ogromne količine podataka. Ali, u DW 2.0 podaci su smješteni po sektorima, tako da krajnji korisnik ima posla sa značajno manje podataka



DW 2.0

Architecture for the next generation of data warehousing

Interactive

Very current

Integrated

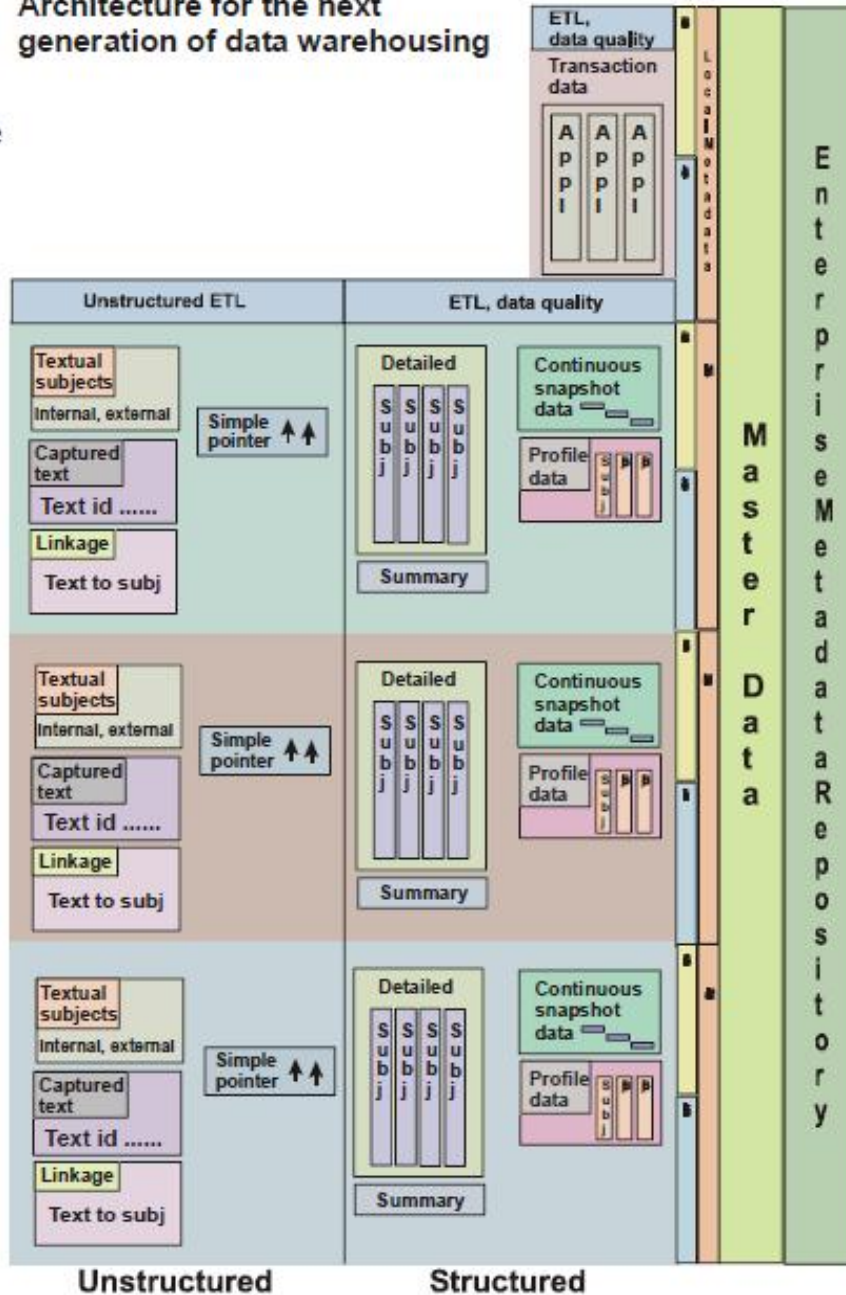
Current++

Near line

Less than current

Archival

Older



DW 2.0. OSNOVNE ZNAČAJKE

- Životni ciklus podataka
- Meta podaci
- Pristup podacima
- Strukturirani / nestrukturirani podaci
- Tekstualna analitika
- Tehnološka osnovica
- Promjene poslovnih zahtjeva
- Tok podataka unutar DW 2.0
- Količine podataka
- Korisne aplikacije



DW 2.0. OSNOVNE ZNAČAJKE

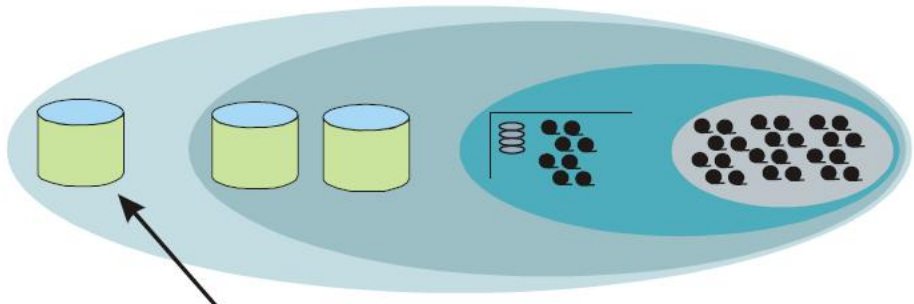
ŽIVOTNI CIKLUS PODATAKA

- DW 1.0 nije prepoznavalo potrebu za praćenjem životnog ciklusa podataka
- DW 2.0 prepoznaje životni ciklus podataka – uključuje različite sektore:
 1. podaci brzo ulaze u Interaktivni (Interactive) sektor
 2. podaci se integriraju i prosljeđuju u Integrirani (Integrated) sektor i ostaju tu dok se ne smanji vjerojatnost pristupa (3-4 godine)
 3. Iz Integriranog sektora podaci se mogu prebaciti u jedan od dva sektora. Jedan je Blizu linije (Near line) sektor (opcionalan), kao proširenje Integriranog sektora u situaciji kada postoji iznimno velika količina podataka i gdje se vjerojatnost pristupa značajno razlikuje
 4. Posljednji sektor – arhivski (archival) sektor



DW 2.0. OSNOVNE ZNAČAJKE

ŽIVOTNI CIKLUS PODATAKA



- a customer's account balance
- the status of a flight for an airline
- the status of an insurance policy
- the amount of money owed on a loan

Fig inter.1
The interactive sector of the data warehouse 2.0

The interactive sector is the place in DW2.0 where online processing occurs. You can get true OLTP response time in the interactive sector. Fig inter.2 shows that online response time is a feature of the DW2.0 environment.

OnLine obrada

Dozvoljeno ažuriranje

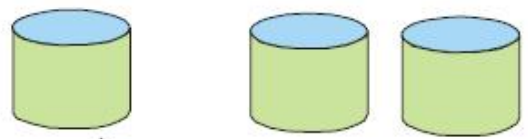


Fig inter.2
Consistent 2/3 second response time

Another possibility in the interactive environment is the possibility of update – update in the sense of finding data in place and making changes to it. Fig inter.3 shows the possibility of update. And of course there is the activity of loading data into the interactive sector.

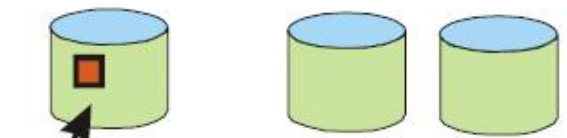


Fig inter.3
Update is a possibility

DW 2.0. MAIN CHARACTERISTICS

THE LIFE CYCLE OF DATA

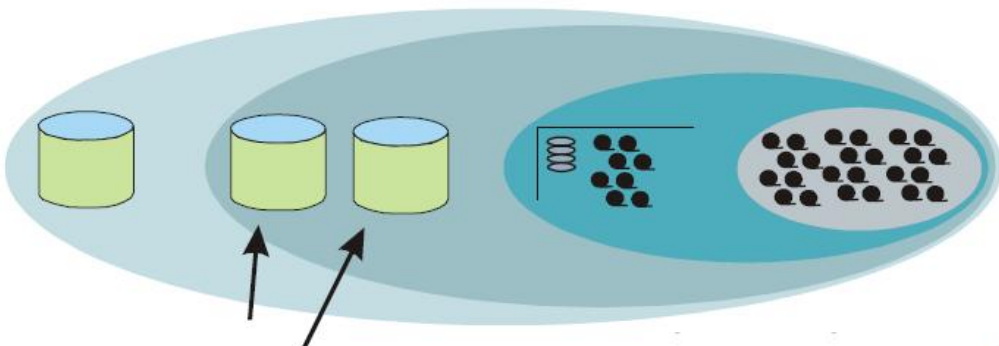


Fig integ.1
The integrated sector of the data warehouse 2.0

Podaci se agregiraju, zbrajaju, rekonstruiraju

Nema ažuriranja
Pohrana kao serija snapshota

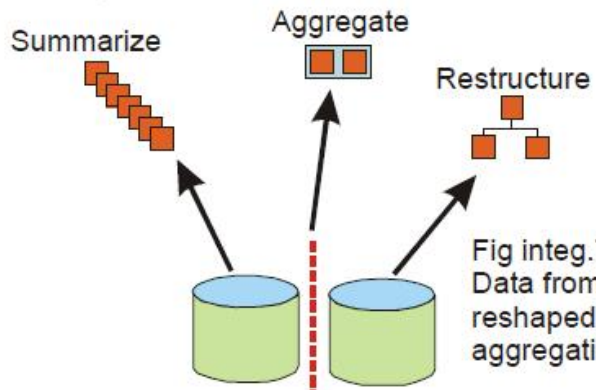


Fig integ.7
Data from the integrated sector can be reshaped in many ways - summarization, aggregation, restructuring, etc.

The pattern of access of data across the integrated sector can be described as "sequentially random". Fig integ.20 shows a sequentially random pattern of access.

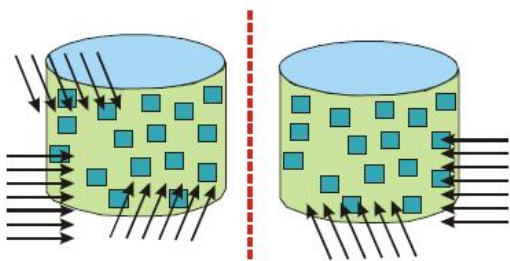


Fig integ.20
The pattern of access of data in the integrated sector can be described as "sequentially random"

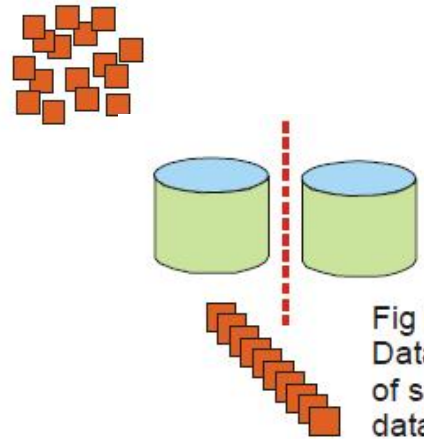


Fig integ.13
Data is stored as a series of snapshots in the integrated data sector

No update

DW 2.0. MAIN CHARACTERISTICS

THE LIFE CYCLE OF DATA

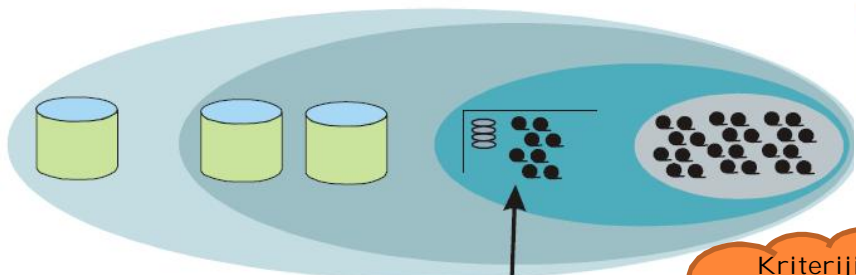


Fig n1.1
The near line sector

Kriteriji:
-Starost
-Vjerojatnost pristupa

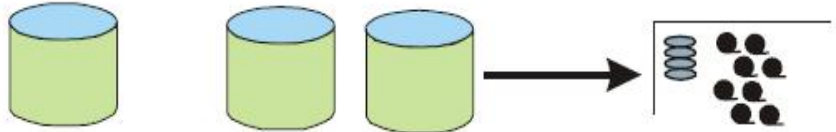


Fig n1.2
The near line sector is fed exclusively from the integrated sector

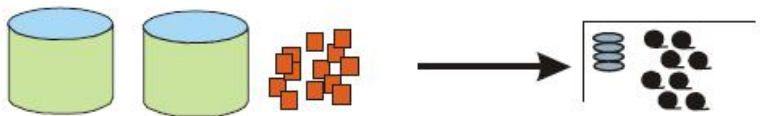


Fig n1.3
Data is fed to the near line sector based on the 1) aging of the data or 2) the lowered probability of access of the day or 3) both factors.

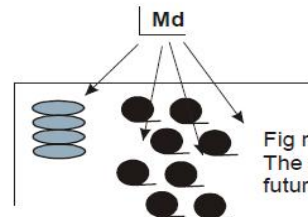


Fig n1.5
The purpose of the metadata is for future location of the data

If by some chance metadata and indexes are not created as data enters the near line sector, then the near line sector turns into a garbage dump, as seen in Fig n1.6.

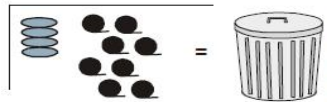


Fig n1.6
Without metadata, the near line sector turns into a garbage dump

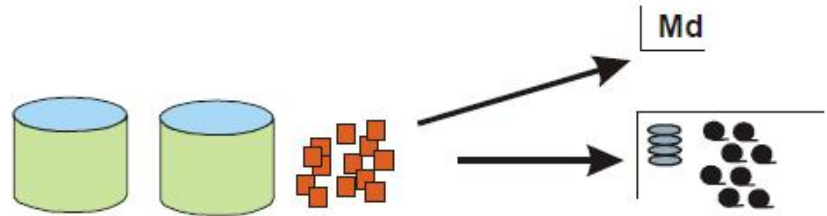


Fig n1.4
Immediately upon entering the near line sector, extensive metadata is created

Meta podaci

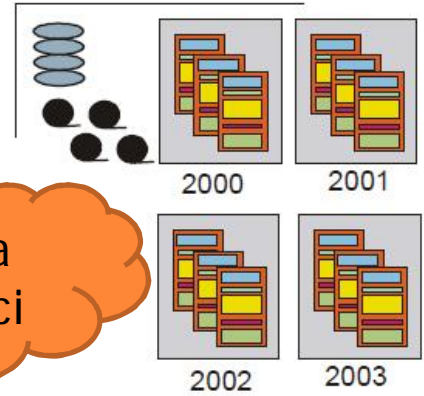


Fig n1.10
As a rule data in the near line sector is organized at the highest level by date, usually year

DW 2.0. MAIN CHARACTERISTICS

THE LIFE CYCLE OF DATA

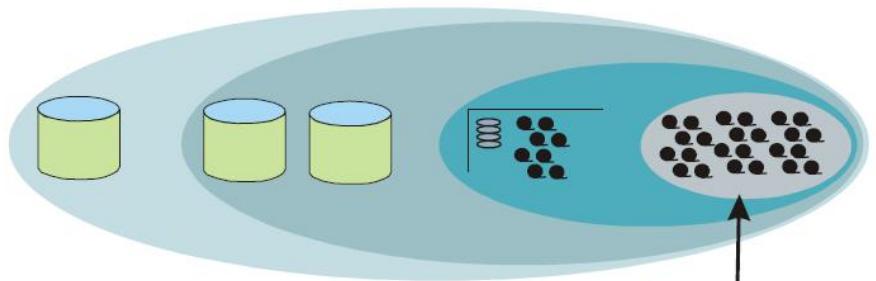


Fig arch.1
The archival sector

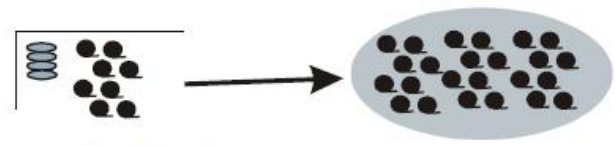


Fig arch.2
The source of data for the archival sector is the near line sector

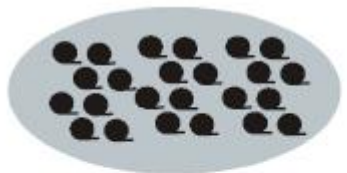


Fig arch.5
Archived data is almost never stored on disk storage

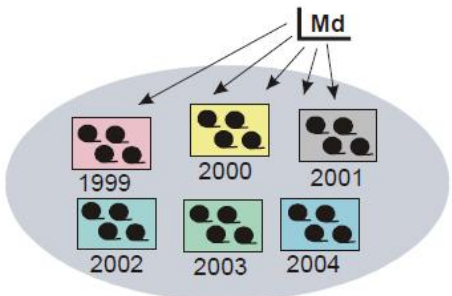


Fig arch.8
Metadata is a very important component of the archival sector

The metadata that is created needs to be stored as an actual part of the archival sector itself. It needs to be stored in the actual data set itself. The reason for storing metadata as part of the actual data is so that over time the data and the metadata won't become separated. Fig arch.12 shows that metadata is part of the archival sector and is stored with the data itself.

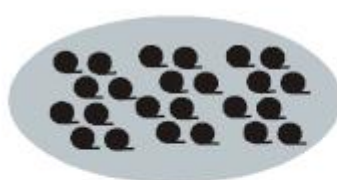


Fig arch.6
Archived data is stored for a long time

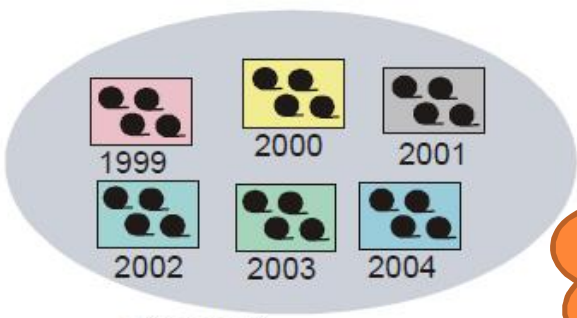


Fig arch.7
All data inside the archival environment is related to time

Povezanost s vremenom, Rijetko kada pohrana na diskove, Matapodaci

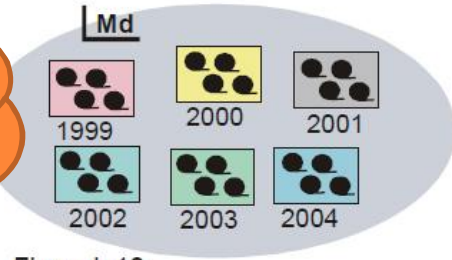
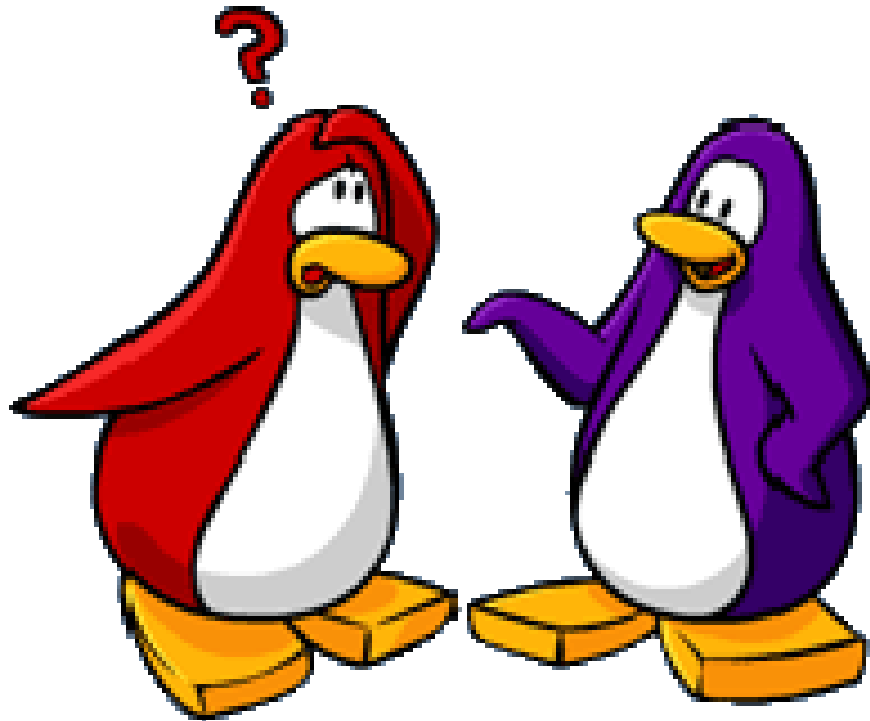


Fig arch.12
The metadata needs to be stored as a close and integral part of the archival sector



Questions..