

SKLADIŠTE PODATAKA - UVOD

Definicije

Skladište podataka (data warehouse): Domenski orijentisan, integrisan, vremensko promenljiv i neuništiv skup podataka namenjen podršci odlučivanju kod upravljanja nekim sistemom.

Podskladište podataka (data mart): Poseban izdvojeni deo skladišta podataka namenjen potrebama nekog dela sistema.

Web skladište podataka (data webhouse): Distribuirano skladište podataka implementirano preko web-a (za koje ne postoji centralizovano čuvanje podataka).

U sva tri slučaja, reč je o strateškom IS, za razliku od operacionog IS.

Pogodnosti

Visok stepen povraćaja ulaganja
Povećanje konkurentnosti
Povećanje produktivnosti odlučivanja
Povećanje kvaliteta odlučivanja

Poređenje

Operacioni (OLTP) IS

Trenutni podaci stanja i prometa
Detaljne podatke
Dinamički podaci
Ponavljajuća predefinisana obrada
Visok nivo transakcione aktivnosti
Predvidiv način korišćenja
Transakciono orijentisan
Aplikativno orijentisan
Podrška svakodnevnom odlučivanju
Opislučuje veliki broj korisnika

Strateški (OLAP) IS

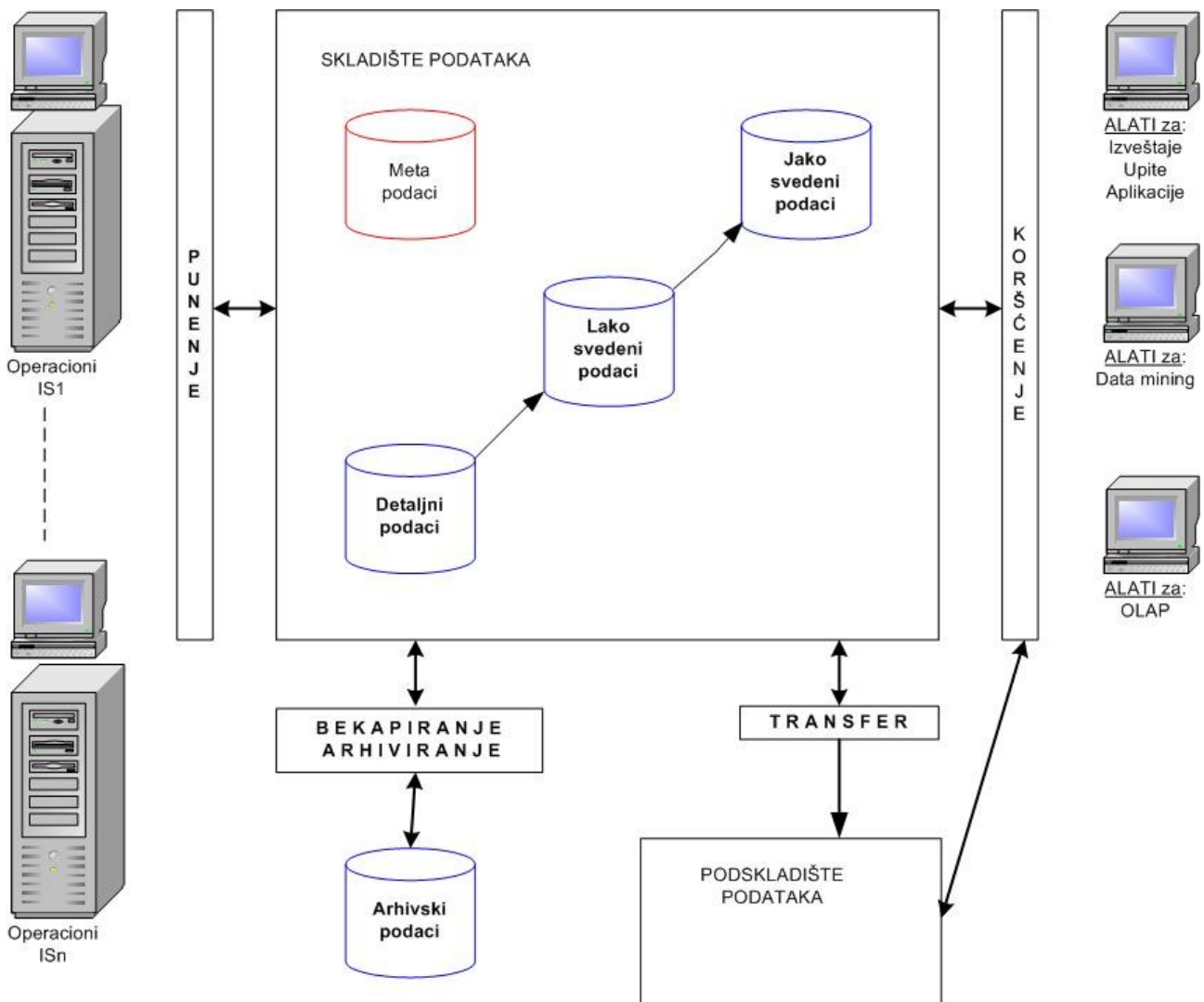
Istorijski podaci stanja i prometa
Detaljni i srednje i visoko svodni podaci
Ugalvnom statički podaci
Od hoc obrada po zahtevu
Nizak / srednji nivo transakcione aktivnosti
Nepredvidiv način korišćenja
Analitički orijentisan
Domenski orijentisan
Podrška strateškom odlučivanju
Opislučuje manji broj korisnika

Problemi koji se javljaju u vezi skladišta podataka

Podcenjivanje resursa potrebnih za punjenje podacima
Skriveni problemi unutar izvornih IS
Neobuhvatanje neophodnih podataka unutar izvornih IS
Semantika i homogenizacija podataka
Visoki zahtevi za resursima
Vlasništvo/pristup podacima
Obimno naknadno održavanje
Dugoročnost projekta (≥ 3 godine)
Kompleksnost integracije sistema

SKLADIŠTE PODATAKA – USTROJSTVO I RAD

Struktura skladišta podataka



DATA WAREHOUSE
DATA MART

- skladište podataka (globalna namena)
- podskladište podataka (specifična namena)

Implementacija skladišta podataka

Uobičajeno je da se vremenom iz jednog skladišta podataka formiraju podskladišta podataka, pri čemu isti podaci i dalje ostaju u skladištu podataka.

U praksi se dešava obrnuto - prvo se realizuju pojedina kritična podskladišta podataka koja se direktno pune iz operacionih IS, a zatim se naknadno formira skladište podataka. Od tog trenutka, tok podataka iz operacionih IS je ka skladištu podataka, a podskladišta podataka se pune transferom iz skladišta podataka.

Implementacija skladišta podataka je preko baze podataka (uglavnom relacione) sa visokim stepenom redundanse, a mehanizam svođenja je zasnovan na okidačima.

Po drugoj implementacionoj klasifikaciji, imamo:

- ✓ **Realno skladište podataka:** Podaci skladišta realno postoje, odvojeno od informacionog sistema operativne namene. Operacije nad skladištem podataka ne opterećuju informacioni sistem operativne namene.
- ✓ **Virtuelno skladište podataka:** Podaci skladišta realno ne postoje, nego sve svaki put izvode kao pogledi nad podacima informacionog sistema operativne namene. Operacije nad skladištem podataka opterećuju informacioni sistem operativne namene, pa je ova varijanta prihvatljiva samo kod manjih sistema.

Punjenje skladišta podataka

U principu postoje dva načina punjenja podacima iz operacionih IS:

- ✓ **Totalno punjenje:** U određenim vremenskim trenucima, skladište se isprazni a zatim ponovo napuni podacima iz operacionih IS.
- ✓ **Inkrementalno punjenje:** Prilikom punjenja, u skladište se prenose samo izmene nastale u operacionim IS nakon prethodnog punjenja.

Postoje dve varijante inkrementalnog punjenja:

- ✓ **Paketno inkrementalno punjenje:** Vršiti se u određenim vremenskim trenucima. Zahteva izmene u operacionom IS (bazi podataka) koje će implementirati mehanizam prepoznavanja nastalih izmena.
- ✓ **Neprekidno inkrementalno punjenje:** Vršiti se neprekidno. Nakon svake promene u operacionim IS mehanizmom okidača vrši se prenos podataka ka skladištu podataka.

Konkretna tehnika inkrementalnog punjenja skladišta podataka su:

- ✓ Eksport promena u log-fajlu baze podataka (paketno).
- ✓ Eksport efekata transakcija (paketno).
- ✓ Eksport promena u bazi podataka preko medijatora (međusloj, paketno).
- ✓ Eksport promena u bazi podataka preko servisa replikacije (direktno).

Mogući problemi kod punjenja skladišta podataka, naročito izraženi kod heterogenih operacionih IS, su:

- ✓ **Nedostatak podataka:** Kod nekih operativnih informacionih sistema iz kojih se skladište puni nedostaju neke dimenzije.
- ✓ **Netačnost podataka:** Pri punjenju je neophodno filtriranje podataka, odnosno odbacivanje netačnih podataka.
- ✓ **Neusaglašenost podataka po tipu/preciznosti:** Pri punjenju je neophodno usaglašavanje po oba osnova.