

Pitanja za odbranu prvog domaćeg zadatka:

1. Šta su transakcioni sistemi i čemu služe?
 - Svakodnevno beleženje podataka o izvršenim transakcijama
 - online transactional processing” (OLTP)
 - za podršku operativnom odlucivanju I beleženje transakcija
 2. Šta su SPO i čemu služe?
 - Sistemi koji imaju zadatak da DO obezbede informacije u svakom trenutku.
 - Treba da upotpune znanje i iskustvo menadžera, jer pomaže pri donošenju prave odluke vezanu za poslovanje preduzeća.
 - Racunarske aplikacije
 - podrzavaju odlucivanje, analitička obrada informacija
 - sastoje se iz baze podataka, baze modela I korisnickog interfejsa
 - Treba da obezbede info za DO u svakom trenutku , upotpune znanje menadzera kroz pomoc da donese pravu odluku za poslovanje
 3. Karakteristike transakcionih sistema.
 - Min zauzimanje skladišnog prostora
 - min vremensko kasnjenje (trenutno beleženje transakcija)
 - normalizovane strukture podataka
 - orijentisani na transakciju (proces)
 - Omogućili su trenutno evidentiranje poslovnih transakcija i beleženje ogromnih količina podataka
 4. Karakteristike SPO.
 - za podršku u donosenju polustrukturiranih (takticki nivo) I nestruktuiranih (strateski nivo) odluka
 - jednostavna interaktivna upotreba poslovnim korisnicima koji se ne razumeju u IT
 - Fleksibilni za prilagodjavanje promenama u okruzenju, kao I promenama u pristupu donosenja odluka o strane DO
 5. Razlika između transakcionih(OLTP) i SPO (OLAP)
 - Podaci – oltp obrada – olap citanje podataka
 - upiti – oltp standradni sql upiti – olap kompleksni analitički upiti
 - memorija – oltp mb/gb – olap gb/tb
 - agregacija – oltp sirovi podaci – olap agregirani podaci
 - pristup – oltp svi korisnici – olap DO, analitičari
 - vreme – oltp real time podaci – olap istorijski podaci
 6. Šta je to normalizacija podataka? Šta je prednost a šta nedostatak normalizovanih podataka? (Ima odgovor ali sam ga našla na netu jer nema po slajdovima)
 - Normalizacija sprečava unos redudantnih podataka.
 - Omogućava brz unos podataka.
 - Zauzima malo skladišnog prostora.
 - Kompleksni relacioni modeli, gde su relacije (tabele) povezane na više nivoa.
- Prednosti normalizacije:
- Smanjenje fizičkog prostora za smeštanje podataka
 - Bolja organizacija podataka

- Promene podataka se rade na nivou samo jednog rekorda

Mane normalizacije:

- Fizički prostor diska je danas jeftin
 - Minimizacija dovodi do visoke granularnosti
7. Definicija DW.

DW je baza podataka sa karakteristikama:

- predmetno orijentisana
- nepromenljiva
- integrisana
- ima vremensku dimenziju
- Skladište podataka je baza podataka u kojoj se prikupljaju, integrišu i strukturiraju podaci potrebni za podršku odlučivanju. DW je jedinstveno mesto u preduzeću u kome se dešava ovaj složen proces. Zasebna kolekcija podataka nezavisna od operativnog rada BP. Osnovna fja analitička obrada podataka a ne transakciona. 36

8. Objasniti pojam integrisanosti kod DW.

- Integrisanost - skladišta podataka pune iz više operacionih baza jednog sistema
- Različiti formati datuma, različiti kodovi zapisa
- Svi podaci sa istim značenjem moraju da imaju iste oznake

9. Objasniti pojam nepromenljivosti kod **DW**.

- Podaci koji se jednom ubace u DW više nisu podložni promenama

10. Objasniti pojam granularnosti kod DW.

Granularnost – izbor osnovne jedinice mere posmatrane dimenzije (dimenzije su: Vreme, Lokacija, Proizvod (kategorija, podkategorija), Kupac (industrija, podgrana, niša) Zaposleni?)

fina granulacija:

- viši stepen granularnosti
- najviši stepen granularnosti (ekstremni slučaj) – operativni podaci iz OLTP sistema –

gruba granulacija:

- niži stepen granularnosti
- podaci iz DW

nivo granularnosti je definisan nivoom agregacije podataka u DW – obrnuto proporcionalan nivou agregacije podataka

11. Šta je to OLAP kocka i čemu služi?

Olap kocka je softversko rešenje koje služi poboljšanju klasičnog načina postavljanja upita i izveštavanja. Pruža konzistentnost podataka, velike količine podataka, brzo izvršenje upita, kvalitetna analiza, unapred izračunate kalkulacije, prijatan za krajnjeg korisnika.

12. Šta je to pivot tabela?

Pivot Tabela je Alat (Pivot Table servis) koji služi za pravljenje različitih izveštaja iz OLAP kocke u Ekselu.

13. **Koji su osnovni elementi pivot tabele. na internetu nadjen odgovor**

A pivot table is constructed from four elements.

- The PIVOTTABLE element represents the table itself and all other elements are contained in it.
- The columns of the table are described by HIERARCHY elements (in case of a dimension column) or MEASURE elements (in case of a value column).
- MEASURE elements are grouped in FACT elements.

14. Šta predstavljaju dimenzije kod OLAP kocke?

Dimenzije su podaci koji cine strukturu kocke, tj tabele iz DW su dimenzije kod OLAP kocke . Sluze da bi se mere mogle prikazati iz razlicitih perspektiva i dovesti u određeni kontekst. Posebno je bitna dimenzija vremena. Npr kupci, proizvodi itd 41,59

15. Šta predstavljaju mere kod OLAP kocke?

Mere su agregirani podaci po dimenzijama i clanovima dimenzija, na osnovu kojih se donose odluke. Svaki DO ima potrebu za određenim brojem pokazatelja (mera) da bi mogao efikasno da donosi odluke. To su numerički (kvatifikovani) podaci. Npr prosečni profit po proizvodu 41, 59

16. Razlika između dimenzija i mera (kako ih prepoznati?).

Dimenzije su podaci koji cine strukturu kocke, a mere su agregirani podaci po dimenzijama i clanovima dimenzija. Zajedno cine strukturu kocke. Mere i dimenzije su bitne : Ako određena tabela u sebi nosi podatke koje je potrebno agregirati (postace mera), ta tabela postaje tabela mera ili tabela cinjenica. Ako tabela ne sadrzi podatke iz kojih treba da se generisu mere, onda se takva tabela naziva dimenziona tabela. 41, 69

17. Razlika između Star sheme i Snowflake sheme.

- Star shema / mere u centralnoj tabeli, dimenzije oko. Zauzima više mesta od pahuljicaste sheme, bolje performanse kod upita. Jednostavniji problemi, manje spajanja tabela
- Pahuljicasta / mere u centralnoj tabeli, dimenzije rasclanjenje (normalizovane) u više tabela. Lošije performanse od Star sheme zbog složenije strukture, i spajanja više tabela. Složeniji problemi, više spajanja tabela 55

18. Opisati MOLAP, HOLAP, ROLAP

- predstavljaju načine skladištenja (balans između smanjivanja skladišnog prostora i brzog izvršavanja upita)
- ROLAP – relaciona (klasična) struktura. Podaci i agregacije su skladišteni u relacionom sistemu (OLTP tj DW) i zauzima najmanje prostora na disku, ali ima najlošije performanse
- HOLAP – hibridna struktura, gde su samo izabrani delovi u multidimen. Strukturi (vrlo često agregacije). Podaci su skladišteni u relacionom sistemu a agregacije (ono što se najčešće koristi) u multidimen kocki. Više prostora od ROLAPa ali ima bolje performanse. Zadovoljavajuće performanse
- MOLAP – multidimenzionalna struktura. Podaci i agregacije skladišteni su u multidimen kocki, zauzima najviše prostora, najbolje performanse. 61

19. Šta je to drill down?

Drill Down podrazumeva navigaciju kroz hijerarhijske dimenzije podataka sačuvanih u DW. Drill Down se koristi u okviru OLAP-a da biste zumirali na detaljnije podatke promenom dimenzije

20. Šta je to ETL?

- Složen i oduzima 90% vremena potrebnog za razvoj skladišta podataka.
- Iščitavanje podataka – podaci iz različitih izvora se pronalaze i učitavaju u alat za ETL
 - Spajanje više izvora
- Transformisanje podataka – podaci se transformisu da bi bili pogodni za učitavanje u DW. Bitno je dostići stepen kvaliteta podataka da bi bili upotrebljivi za analizu.
 - Prenosenje podataka iz različitih izvora u skladište
 - Konsolidacija različitih formata zapisa (4/29/2012, 2 sept 2013)
 - Ispravljanje gresaka
 - Denormalizacija podataka
- Učitavanje podataka – podaci se učitavaju u DW 48, 56

21. Šta je to Data Mart (Centri podataka)?

- DM su funkcionalno orijentisana skladišta namenjena za podršku odlučivanja određene funkcije preduzeća. Nastaju iz potrebe za podacima za analizu za konkretan posao ili oblast. Takođe iz razloga preobimne baze podataka ili skladišta podataka za korišćenje, i iz sigurnosnih razloga, pravi se DM koji korisniku omogućava rad nad određenim delom te baze/skladišta. 44

22. Da generišu OLAP izveštaj na naš zahtev (jednostavan i složen)

23. Kako određuje da li podatak može biti dimenzija ili mera? (numerički, ne-numerički; “dani u mesecu”??)

Mera su kvantitativni, numericki podaci. Dimenzije sadrže kategoricke informacije.

24. Navedite nekoliko načina agregiranja mera. (count, sum, avg, stdDev, ...)

Odgovor je zapravo ovo sto je u zagradi. Count-ako zelimo prebrojati podatke. Sum-suma podataka. Avg-sr vrednost. St devijacija itd.

25. U čemu je prednost drill-down tehnike?

Drill-down je tehnika koja korisnicima omogućuje navigaciju kroz detalje podataka (npr. kada se traži prodaja po proizvodima u različitim regionima).

26. U čemu je razlika između ColumnFields i PageFields delova Pivot Tabele.

- Koriste se u Ekselu za OLAP kocke, kod Pivot tabela

- ColumnFields: služi za smestanje dimenzija, koje treba da budu redovi u zeljenom izveštaju. Paziti kojim redom se postavljaju dimenzije
- PageFields: predstavlja svojevrsni filter koji ceo prostor kocke suzava izborom jedne ili vise konkretnih vrednosti, bez uticaja na izgled Pivot Tabele. Ovo može biti npr dan u toku nedelje, kategorija proizvoda itd. U PageFields se ubacuju dimenzije, i može ih biti vise. 67

27. Šta je SingleTable schema?

- Najjednostavnija vrste seme za koriscenje(ako se uzmu o obzir Star i Pahuljicasta). Sve dimenzije i mere se cuvaju u jednoj tabeli. Jednostavna za koriscenje, neefikasna za cuvanje velike kolicine podataka. Kako se zahtevi korisnika uslozjavaju ova sema se transformise u pahulju ili zvezdu. 55

28. U čemu je razlika između DW i DataMart-a. **Njihov primer je šta od ta dva?**

- DW umeju biti prilično velika, pa se klasifikuju u DM koja su znatno manja
- DW integrisu podatke iz svih izvora podataka, a DM integrisu informacije vezane za određenu oblast/funkciju/projekat
- DM sadrži podatke bitne za jednu funkciju/projekat, a DW podatke bitne za više oblasti
- DW za stratesko odlucivanje, a DM za taktičko

29. Da li je moguće sve mere primeniti na sve kombinacije dimenzija? Obrazložite.

30. Objasniti Bar Chart (kada se koristi, šta je potrebno da bi se generisao)

Bar chart koristimo da uporedimo podatke kroz kategorije(prodaju kroz godine npr).

Pravimo ga tako sto stavimo dimension u red, a measure u colonu, ili obrnuto.

31. Objasniti Histogram (kada se koristi, šta je potrebno da bi se generisao)

Histogram prikazuje vrednosti kao oblike. Izgleda kao bar chart, ali vrednosti prikazuje u pravougaonicima. Koristi se za raspodelu vrednosti

Measures su obavezne, a dimenzije mogu a ne moraju da se unose(mogu samo brojevi)

32. Objasniti Pie Chart (kada se koristi, šta je potrebno da bi se generisao)

Pie chart nam pomaze da sagledamo proporcije.

U kolone unosimo measure a u redove dimension.

Sve ovo nam show me pokazuje da li je moguće i koji tip charta može za date podatke.

33. Objasniti Line Chart (kada se koristi, šta je potrebno da bi se generisao)

Line charts su jako dobre za vizualizaciju trendova kroz vreme, ili da predvidimo buduće trendove.

Primer pravljenja: Godine(one su dimension jer su broj) u kolone, prodaja u redove, zatim profit u redove.

34. Objasniti Box Plot (kada se koristi, šta je potrebno da bi se generisao)

Box plox pokazuje raspodelu vrednosti duz ose.

U kolone stavljamo dimensions(npr pol ljudi), a u kolone dimension-neki brojni podatak koji nas zanima.

35. Objasniti Scatter Plot (kada se koristi, šta je potrebno da bi se generisao)

- Scatter plot se koristi da uvidimo vezu izmedju brojnih vrednosti.
- Moze se napraviti tako sto stavimo measures i u red i u kolonu(npr zavisnost prodaje od broja radnika).
- Potrebne su kontinualne vrednosti

36. Razlika između Live i Extract konekcije ka podacima.

- Live konekcija ce kod upita koristiti podatke u realnom vremenu, te ce izvestaji biti formirani na osnovu poslednjih unetih podataka.
- Extract konekcija je konekcija sa staticnom bazom podataka, koja se refresuje u određenim periodima na zahtev korisnika. Svi podaci se kopiraju iz izvora podataka u tablo server. Da bismo dobili nove podatke moramo da ponovimo extract.

37. Navesti barem 3 osnovna tipa grafova u Tableau-u.

Bilo koja 3 od ovih iznad. Bar pie line....

38. Opisati Trend Line funkcionalnost u Tableau-u?

Trend line koristimo za konstruisanje modela ponasanja. Sa trend lines mozemo da damo odgovor na pitanja tipa: da li je profit predvidjen prodajom? da li su kasnjenja u korelaciji sa mesecima u godini?

39. Opisati Forecast funkcionalnost u Tableau-u?

Forecast- tehinka u tableau koja koristi algoritme da bi predvidela kako ce se u buducnosti neki trend odvijatu.(measures tj brojevi)

40. Razlika između Trendline i Forecast u Tableau-u.

Forecast se bavi predvidjanjem tj buducnoscu, dok trend line modelima ponasanja(stvari koje su se desile)

41. Razlika između Filter i Page Fields u Tableau-u.

Filter koristimo da bi "filtrirali"(izdvojili) podatke koje zelimo da prikazemo na sheet-u. Npr zelimo prvih 15 kupaca.

Pages koristimo za uvid vise stranica da bismo analizirali kako neko polje utice na ostatak podataka.

Npr: Dobijamo jednu umesto cetiri razlicite skale koja su jedna ispod druge, koju mozemo menjati sa promenom dana u nedelji.