

UPRAVLJANJE POSLOVNIM PODACIMA



DIMENZIJSKO MODELIRANJE

PROF. DRAŽENA GAŠPAR

30.11.2015.

NAPREDNI KONCEPTI

1. Degenerativna dimenzija
2. Pahuljasta shema (Snowflaking)
3. Previše dimenzija
4. Surogatni ključevi
5. Periodični snapshot
6. Podudarne dimenzije
7. Podudarne činjenice/fakti
8. Sporo promjenjive dimenzije



DEGENERATIVNA DIMENZIJA

- Dimenzijska tablica bez atributa.
- Ima samo primarni ključ.

Degenerativna dimenzija je podatak koji je po svojoj prirodi dimenzijski ali je pohranjen u tablici činjenica.

Primjer: dimenzija koja samo ima Broj narudžbe

- 1:1 relacija s tablicom činjenica

POSLJEDICE ????



DEGENERATIVNA DIMENZIJA

Posljedice:

→ Dvije tablice s milijun redaka

Umjesto jedne tablice s milijun redaka.

→ To treba biti degenerativna dimenzija s
Brojem narudžbe pohranjenim u tablici
činjenica.



DEGENERATIVNA DIMENZIJA

Degenerativne dimenzije su uobičajene kada usitnjenost tablice činjenica odgovara jednoj transakciji ili jednoj stavci te transakcije.

Degenerativna dimenzija predstavlja jedinstveni identifikator roditelja u relaciji roditelj-dijete.

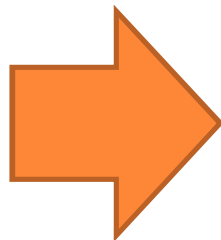
Na primjer, broj narudžbe, broj računa i broj teretnice se gotovo uvijek pojavljuju kao degenerativne dimenzije u dimenzijskom modelu.



DEGENERATIVNA DIMENZIJA

○ Primjer:

ORDERS TRANSACTIONS
order#
customer id
customer lname
customer fname
shipto street address
shipto city
shipto state
shipto zip
order total amount
discount amount
net order amount
payment amount
order date

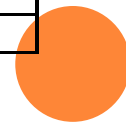


ORDERS FACTS
customer key
shipto address key
order date key
order total amount
discount amount
net order amount
payment amount
order#

DIM CUSTOMER
Customer key
customer id
customer lname
customer fname

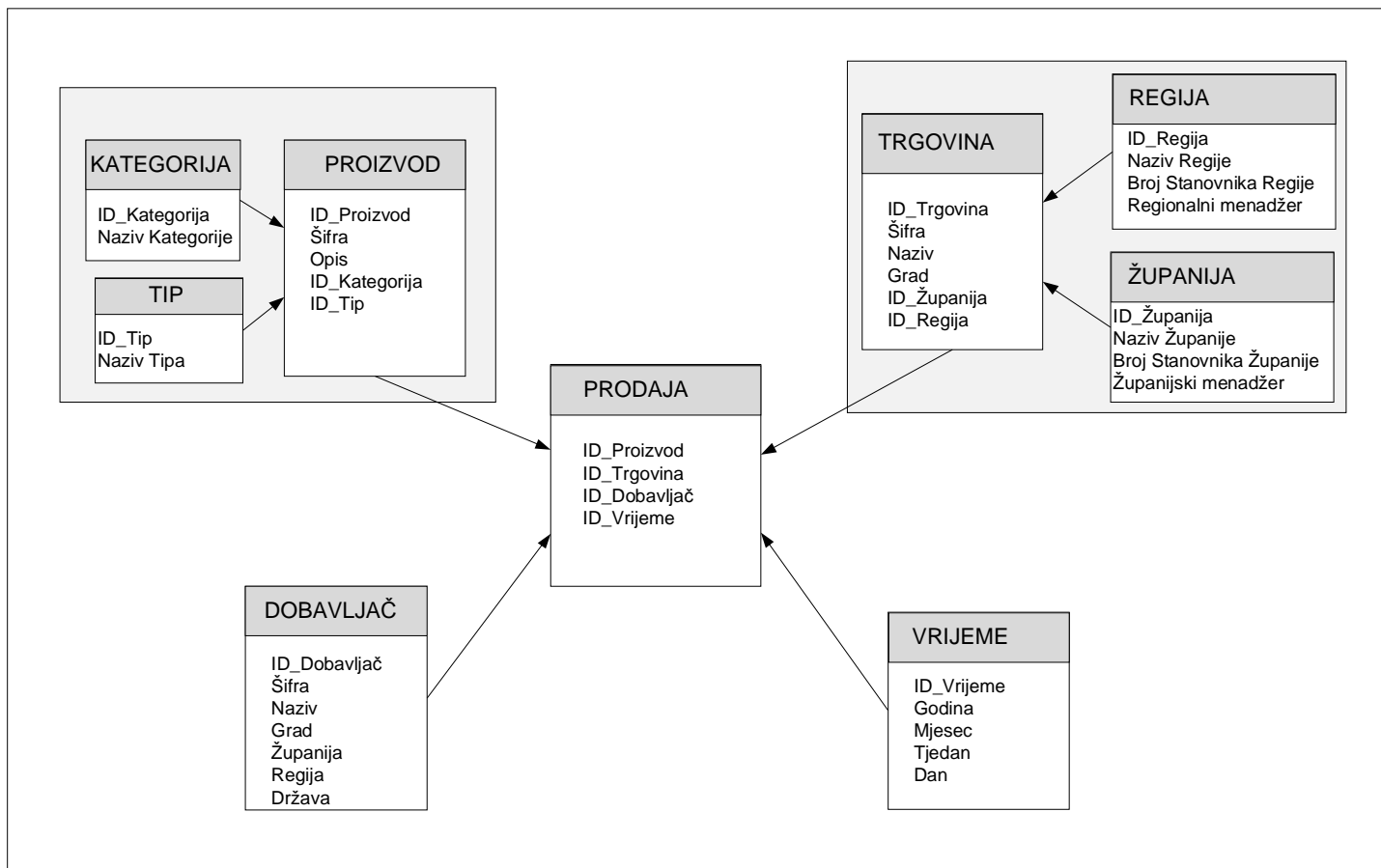
DIM SHIPTO ADDRESS
Shipto address key
shipto street address
shipto city
shipto state
shipto zip

DIM Order Date
Order date key
Calendar date
Calendar month
...



PAHULJASTA SHEMA (SNOWFLAKING)

Normalizirana star shema



PAHULJASTA SHEMA (SNOWFLAKING)



Problemi:

- Povećava kompleksnost za korisnike
- Umanjuje performanse (brojne tablice i spajanja)
- Umanjuje mogućnost pretraživanja unutar dimenzije
- Primjer problema: svi brendovi unutar kategorije



PAHULJASTA SHEMA (SNOWFLAKING)

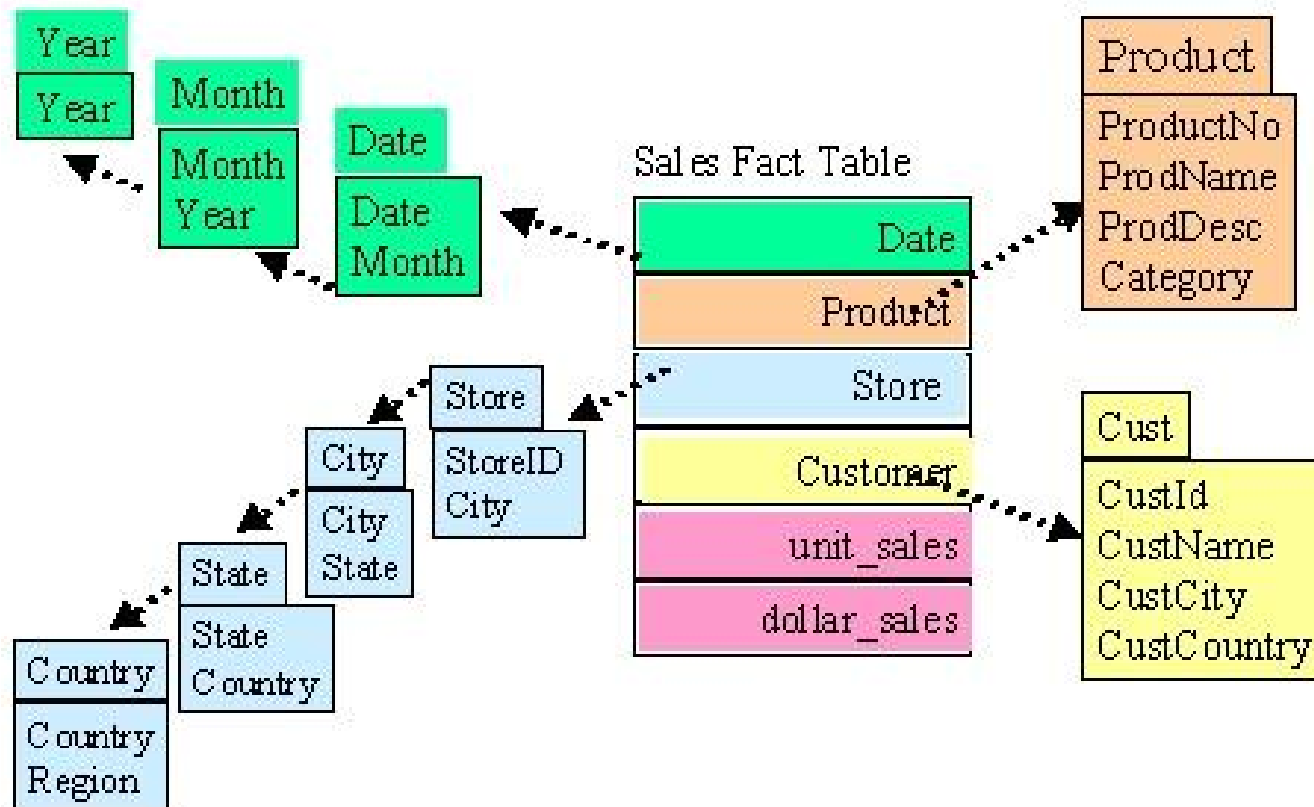


- Kimball navodi sljedeća tri slučaja u kojima je njena uporaba neophodna kako bi se osigurao uspješan dizajn:
 - Velika dimenzija Kupaca (različite kategorije kupaca)
 - Dimenzije financijskih proizvoda (usluga)
 - Dimenzija vremena (kalendar) za multinacionalne kompanije.



PREVIŠE DIMENZIJA

Example of Snowflake Schema



(Source: JRD)



PREVIŠE DIMENZIJA

- Veliki broj dimenzija je obično znak da neke od dimenzija nisu potpuno neovisne i da bi se trebale objediniti u jednu dimenziju.
- Pogrešno je u dimenzijskom modelu predstaviti elemente hijerarhije kao odvojene dimenzije.




SUROGATNI KLJUČEVI

- Surogatni (umjetni, neprirodni, sintetički...) ključevi su cjelobrojne vrijednosti kojima se vrijednost dodjeljuje sekvencijalno.
- Surogatni ključ je zamjena za prirodni primarni ključ.
- Nema značenje.
- To je samo jedinstveni identifikator ili broj za svaki redak koji se koristi kao primarni ključ tablice.
- Služi za povezivanje dimenzijskih tablica s tablicom činjenica.
- Koristan je jer se prirodni primarni ključ (npr. Šifra kupca u tablici Kupac) može promijeniti i otežati ažuriranja.



SUROGATNI KLJUČEVI

- Prednosti uprabe surogatnih ključeva
 - Performanse
 - Efikasna spajanja
 - Manji indeksi
 - Više redaka po bloku
 - Integritet podataka
 - Kada se ključevi u transakcijskim sustavima mogu ponovno koristiti
 - Diskontinuitet proizvoda, Kupci koji više ne postoje i sl.
 - Mapiranje kada se integriraju podaci iz različitih izvora
 - Ključevi iz različitih izvora mogu biti različiti
 - Tablice mapiranja surogatnih ključeva i ključeva iz različitih izvora
- 

SUROGATNI KLJUČEVI

- Prednosti uporabe surogatnih ključeva (nastavak)
 - Rad s nepoznatim ili N/A vrijednostima
 - Jednostavna dodjela surogatnih ključeva redovima s tim vrijednostima
 - Praćenje promjena vrijednosti atributa u dimenzijama
 - Kreiranje novih atributa i dodjela sljedećeg raspoloživog surogatnog ključa



SUROGATNI KLJUČEVI

- Nedostaci uporabe surogatnih ključeva
 - Dodjela i upravljanje surogatnim ključevima i odgovarajuća zamjena prirodnih ključeva surogatnim – dodatno učitavanje za ETL sustav
 - Mnogi ETL alati imaju ugrađene mogućnosti za potporu u obradi surogatnih ključeva
 - Jedanput kada se proces razvije, može se jednostavno koristiti i za druge dimenzije



TIPOVI TABLICA ČINJENICA

- Kimball razlikuje tri osnovna tipa tablica činjenica (Kimball, 2002):
 - Transakcijska
 - Periodični snapshot
 - Akumulirani snapshot



Osobine	Transakcijska Usitnjenost	Periodični snapshot usitnjenost	Akumulirani snapshot usitnjenost
Predstavljeno vremensko razdoblje	Točka u vremenu	Regularni, predvidivi intervali	Neodređeni vremenski razmak, obično kraći
Usitnjenost	Jedan redak za svaki transakcijski događaj	Jedan redak za period	Jedan redak za razdoblje
Punjenje tablice (engl. <i>load</i>)	Dodavanje	Dodavanje	Dodavanje i ažuriranje
Ažuriranje redaka	Nema revizije	Nema revizije	Revizija kad god postoji aktivnost
Vremenska dimenzija	Datum transakcije	Datum kraja razdoblja	Više datuma za standardne kontrolne točke
Činjenice (vrijednosti)	Transakcijska aktivnost	Izvršenje za predefiniirano vremensko razdoblje	Izvršenje za konačni životni vijek



TRANSAKCIJSKA TABLICA ČINJENICA

- Temeljni prikaz poslovnih operacija je na razini svake pojedinačne transakcije.
- Transakcijske tablice činjenica predstavljaju događaje koji se odigravaju u vremenskim trenucima, odnosno jedan redak u transakcijskoj tablici činjenica odgovara mjerenju događaja u jednoj točki i jednom vremenskom trenutku.
- Detaljna (atomizirana) razina usitnjenosti je veoma pogodna za pretraživanja po različitim dimenzijama.



TABLICA ČINJENICA TIPA PERIODIČNI SNAPSHOT

- Koriste se kako bi se vidjele kumulativne performanse poslovanja u redovitim, predvidljivim vremenskim razdobljima.
- Za razliku od transakcijske tablice činjenica gdje se dodaje novi redak u tablicu za svako pojavljivanje događaja, kod periodičnog snapshota se radi „slika“ aktivnosti na kraju dana, tjedna ili mjeseca, zatim se radi sljedeća „slika“ na kraju sljedećeg razdoblja itd.
- Periodični snapshoti se sukcesivno pohranjuju u tablicu činjenica



TABLICA ČINJENICA TIPA PERIODIČNI SNAPSHOT

- Redak u tablici činjenica koja sadrži periodične snapshotte zbraja više mjerenja događaja koji se pojavljuju tijekom standardnog razdoblja kao što je dan, tjedan ili mjesec.
- Usitnjenost čini razdoblje, a ne pojedinačna transakcija.
- Ove tablice činjenica su uniformne gustoće u odnosu na strane ključeve jer čak i ako nije bilo nikakve aktivnosti tijekom razdoblja, redak se obično dodaje u tablicu činjenica i sadrži nula ili nul vrijednosti (null)



TABLICA ČINJENICA TIPA AKUMULIRANI SNAPSHOT

- Akumulirani snapshoti predstavljaju neograničeni vremenski razmak i tako pokrivaju čitav životni vijek transakcije, određenog proizvoda ili kupca.
- Redak u tablici činjenica koja je tipa akumulirani snapshot zbraja mjerenja događaja koja su se pojavljivala u predviđenim koracima od početka do kraja procesa.
- Tokovi procesa, kao što su naručivanje ili obrada zahtjeva, koji imaju definiranu polaznu točku, standardne među korake i definiranu krajnju točku, mogu biti modelirani korištenjem ovog tipa tablice činjenica.



TABLICA ČINJENICA TIPA AKUMULIRANI SNAPSHOT

- Akumulirani snapshoti gotovo uvijek moraju imati višestruke vremenske oznake koje predstavljaju predvidljive glavne događaje ili faze koje se odvijaju tijekom životnog vijeka samog događaja.
- Postoji datumski strani ključ u tablici činjenica za svaku kritičnu kontrolnu točku u procesu.
- Pojedinačni redak u tablici činjenica koja je tipa akumuliranog snapshota odgovara, na primjer, retku narudžbe kakav je inicijalno unesen u trenutku kreiranja retka narudžbe.
- Kako se proces odvija, tako se redak u tablici činjenica tipa akumuliranog snapshota pregledava i ažurira. Ovo konzistentno ažuriranje retka kod akumuliranog snapshota je specifično samo za ovaj tip tablice činjenica.



PODUDARNE (CONFORMED) DIMENZIJE

- Matične (master) ili zajedničke dimenzije
- Djeljive su u DW okruženju i spajaju višestruke činjenične tablice koje predstavljaju različite poslovne procese
- 2 tipa
 - Identične dimenzije
 - Smanjene dimenzije (engl. *shrunk dimensions*) - jedna dimenzija je podskup druge detaljnije dimenzije



PODUDARNE (CONFORMED) DIMENZIJE

Identične dimenzije

- Za dimenzijske tablice se kaže da su podudarne ako atributi u različitim dimenzijskim tablicama imaju iste ključeve, iste nazive stupaca (atributa), iste definicije atributa i iste domenske vrijednosti (tip i veličina podataka, korisnička domena).
- Podudarne dimenzije imaju isti sadržaj, interpretaciju i prezentaciju neovisno o poslovnom procesu.
- Na primjer: dimenzija Proizvod je ista bez obzira poziva li se na nju narudžba ili skladište.



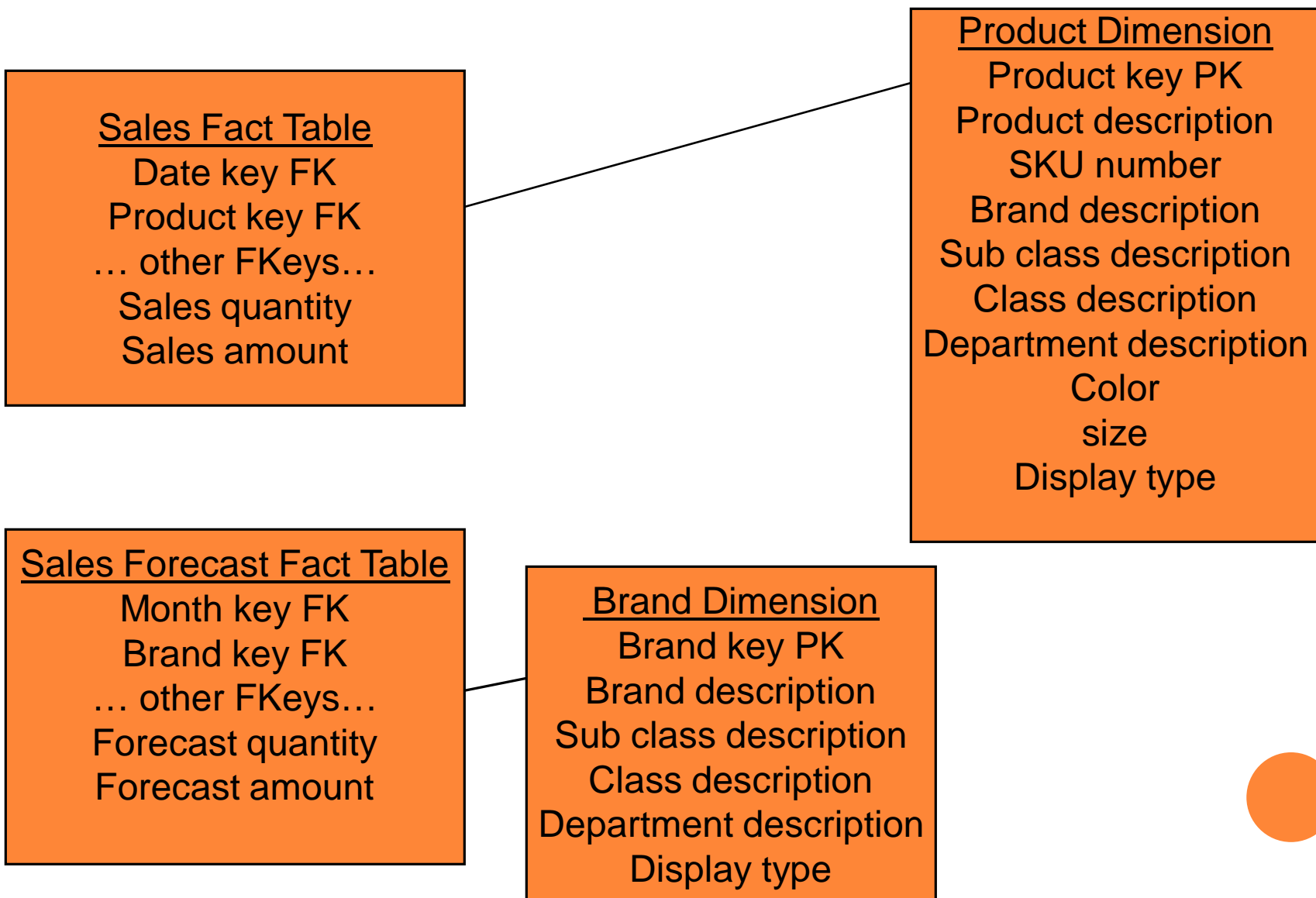
PODUDARNE (CONFORMED) DIMENZIJE

Smanjene dimenzije

- Podudarne dimenzije koje su podskup redaka i/ili stupaca osnovne dimenzije.
- Jedna dimenzija može biti podskup detaljnije, usitnjene dimenzijske tablice.
- Na primjer: prodaja je povezana na dimenzijsku tablicu na pojedinačnoj razini proizvoda, dok je predviđanje prodaje povezano s razinom grupe proizvoda.



PODUDARNE (CONFORMED) DIMENZIJE



PODUDARNE (CONFORMED) DIMENZIJE

Prednosti podudarnih dimenzija su (Kimball & Ross, 2013) :

- Konzistentnost - svaka tablica činjenica se filtrira konzistentno, a rezultati se označavaju na isti način.
- Integracija - korisnici mogu kreirati upite koji svrdlaju kroz tablice činjenica i predstavljaju pojedinačno različite procese, a zatim spojiti rezultate po zajedničkim atributima dimenzije.
- Smanjuje vrijeme razvoja - jedanput kreirane, podudarne dimenzije se mogu višestruko koristiti.



PODUDARNE ČINJENICE

- Ako se dvije iste činjenice (mjere) pojavljuju u odvojenim tablicama činjenica – mora se osigurati da su tehničke definicije činjenica identične, kako bi se mogle uspoređivati ili zajedno koristiti u izračunima.
- Ako vrijednosti postoje u više tablica činjenica, njihove definicije i izračuni moraju biti isti ako se želi reći da se odnose na istu stvar.
- Ako su činjenice označene identično tj. imaju isti naziv, onda moraju biti definirane u istom kontekstu dimenzija i sa istim jedinicama mjere u svim zvijezda shemama.
- Primjeri: prihod, profit, cijene, troškovi, mjerenje kvaliteta, mjerenje zadovoljstva kupaca i drugi KPI.



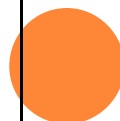
SPORO PROMJENJIVE DIMENZIJE

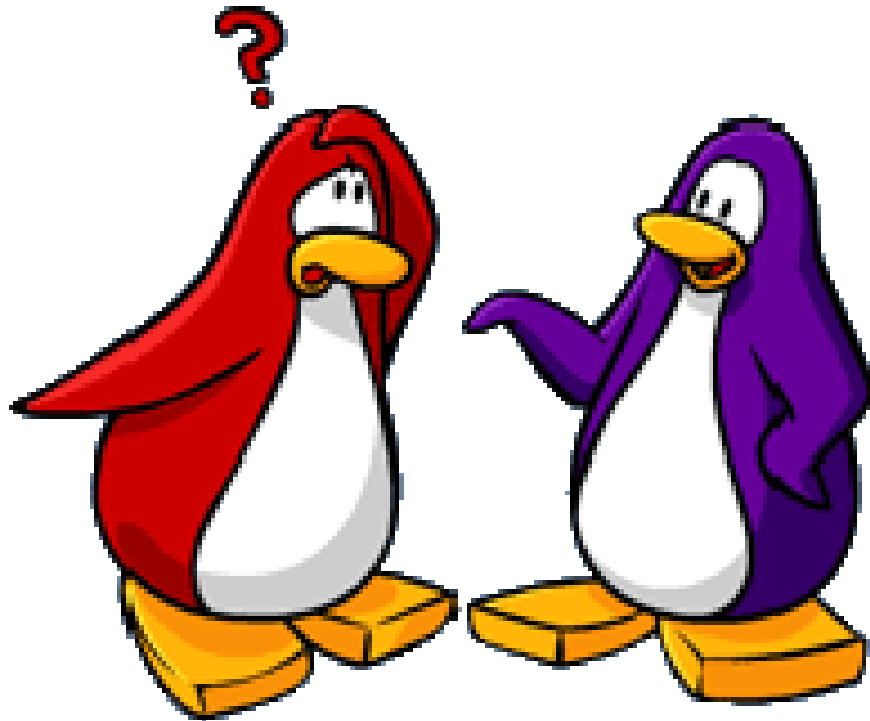
- Ako se atributi dimenzijske tablice ne mijenjaju često - riječ je o sporo promjenjivim dimenzijama.
- Kimball razlikuje osam osnovnih tehnika za rad sa sporo promjenjivim dimenzijama (Kimball & Ross, 2013):
 - *Tip 0: Zadržavanje originala*
 - *Tip 1: Prepisivanje preko originala*
 - *Tip 2: Dodavanje novog retka*
 - *Tip 3: Dodavanje novog atributa*
 - *Tip 4: Dodavanje mini dimenzije*
 - *Tip 5: dodavanje mini dimenzije i tip 1*
 - *Tip 6: Dodavanje atributa Tipa 1 dimenziji Tipa 2*
 - *Tip 7: Dualne (dvojne) Tip1 i Tip2 dimenzije*



SPORO PROMJENJIVE DIMENZIJE

Tip tehnike	Dimenzijska tablica - poduzeta akcija -	Utjecaj na analizu tablice činjenica
Tip 0	Nema promjene vrijednosti atributa.	Činjenice su povezane s originalnom vrijednošću atributa.
Tip 1	Prepisivanje preko vrijednosti atributa.	Činjenice su povezane s tekućom (aktualnom) vrijednošću atributa.
Tip 2	Dodavanje novog retka s novom vrijednošću atributa.	Činjenice su povezane s vrijednošću atributa ovisno o tome kada su se činjenice pojavile.
Tip 3	Dodavanje novog stupca kako bi se sačuvala i tekuća i ranija vrijednost atributa.	Činjenice su povezane i s tekućom i s prethodnom, alternativnom vrijednošću atributa.
Tip 4	Dodavanje mini dimenzijske tablice koja sadrži brzo promjenjive atribute.	Činjenice se povezuju s brzo promjenjivim atributima ovisno o tome kada su se činjenice pojavile.
Tip 5	Dodavanje tip 4 mini dimenzije zajedno s tip 1 ažuriranim ključem mini dimenzije u osnovnoj dimenziji.	Činjenice se povezuju s brzo promjenjivim atributima ovisno o tome kada su se činjenice pojavile, i dodatno s trenutnim vrijednostima brzo promjenjivih atributa.
Tip 6	Dodavanje tip 1 ažuriranih atributa retku dimenzije tipa 2 i ažuriranje svih prethodnih redaka dimenzije.	Činjenice se povezuju s vrijednošću atributa ovisno o tome kada su se činjenice pojavile, i dodatno s tekućim vrijednostima.
Tip 7	Dodavanje tip 2 retka dimenzije s novom vrijednošću atributa, i dodatno, ograničavanje pogleda na tekuće retke i/ili vrijednosti atributa.	Činjenice se povezuju s vrijednošću atributa ovisno o tome kada su se činjenice pojavile, i dodatno s tekućim vrijednostima.





Questions..