

# Analiza podataka o gledanosti IPTV sadržaja

---

**Bodulušić, Matea**

**Master's thesis / Diplomski rad**

**2019**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Organization and Informatics / Sveučilište u Zagrebu, Fakultet organizacije i informatike**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:211:049725>

*Rights / Prava:* [Attribution-NonCommercial-NoDerivs 3.0 Unported](#) / [Imenovanje-Nekomercijalno-Bez prerada 3.0](#)

*Download date / Datum preuzimanja:* **2024-07-06**



*Repository / Repozitorij:*

[Faculty of Organization and Informatics - Digital Repository](#)



**SVEUČILIŠTE U ZAGREBU  
FAKULTET ORGANIZACIJE I INFORMATIKE  
VARAŽDIN**

**Matea Bodulušić**

**ANALIZA PODATAKA O GLEDANOSTI  
IPTV SADRŽAJA  
DIPLOMSKI RAD**

**Varaždin, 2019.**

**SVEUČILIŠTE U ZAGREBU**  
**FAKULTET ORGANIZACIJE I INFORMATIKE**  
**V A R A Ž D I N**

**Matea Bodulušić**

**Matični broj: 46335/17–R**

**Studij: *Baze podataka i baze znanja***

**ANALIZA PODATAKA O GLEDANOSTI IPTV SADRŽAJA**  
**DIPLOMSKI RAD**

**Mentor:**

Prof. dr. sc. Rabuzin Kornelije

**Varaždin, rujan 2019.**

*Matea Bodulušić*

### **Izjava o izvornosti**

Izjavljujem da je moj završni/diplomski rad izvorni rezultat mojeg rada te da se u izradi istoga nisam koristila drugim izvorima osim onima koji su u njemu navedeni. Za izradu rada su korištene etički prikladne i prihvatljive metode i tehnike rada.

*Autor/Autorica potvrdio/potvrdila prihvaćanjem odredbi u sustavu FOI-radovi*

---

## Sažetak

Prava informacija u pravom trenutku donosi uspjeh. U današnje vrijeme podaci su dostupni, ali treba znati doći do prave informacije. Proces ekstrakcije (eng. Extract) podataka s izvora (eng. Source) , transformiranja (eng. Transform) te učitavanja (eng. Load) podataka u željene tablice, uz uvjet da su podaci istiniti, omogućuje pohranjivanje podataka u skladište podataka (eng. data warehouse) ili u njegov podskup orijentiran na određenu poslovnu liniju ili tim, engleskim nazivom data mart. Skladište podataka čuva podatke i omogućuje dolazak do informacija na brži i učinkovitiji način.

Internetski protokol televizije (IPTV, eng. Internet Protocol Television) omogućuje isporuku televizijskog sadržaja izvornog medija putem internetske mreže u vremenu kada se sadržaj emitira.

U svrhu ovog rada razvijen je data mart za tvrtku A1 Hrvatska koji omogućuje pregled gledanosti IPTV sadržaja.

**Ključne riječi:** Skladište podataka; Analiza podataka; ETL; IPTV; Logički model podataka; Model pahuljice; Podatak; Informacija

# Sadržaj

1. Uvod.....	1
2. Metode i tehnike rada .....	2
3. Skladište podataka i data mart.....	3
3.1. Skladišta podataka.....	3
3.1.1. Višedimenzionalni modeli skladišta podataka.....	7
3.1.1.1. Model zvijezde.....	7
3.1.1.2. Model pahuljice .....	8
3.2. Data mart.....	8
4. Proces razvoja skladišta podataka.....	12
4.1. Priprema za ETL.....	12
4.2. ETL Process .....	18
4.2.1. Ekstraktiranje .....	19
4.2.2. Transformiranje .....	19
4.2.3. Učitavanje .....	20
4.2.3.1. Tri tipa učitavanja .....	20
4.2.3.2. Vrste učitavanja.....	21
4.3. Testiranje ETL-a .....	24
4.4. Poslovna inteligencija.....	25
5. Korišteni alati .....	27
5.1. Toad.....	27
5.2. Informatica .....	27
5.2.1. Informatica PowerCenter.....	27
6. Implementacija skladišta podataka .....	28
6.1. Osnovni podaci o djelatnosti tvrtke i poslovanju .....	28
6.2. IPTV.....	29

6.3. Faze implementacije .....	30
6.3.1. Dizajn.....	30
6.3.2. Razvoj.....	34
7. Zaključak .....	45
8. Literatura .....	46
Popis slika.....	48
Popis tablica.....	49

# 1. Uvod

„Gradovi i općine moraju postati više svjesni kako su informacije nafta 21. stoljeća i da se time mogu ostvariti značajni prihodi" izjavio je Gerd Landsberg 2018. godine, ali informacije su značajne i za puno manje kao i za puno veće organizacije. Prije informacija važno je znati podatke.

Varga, Čerić i Birolla (1998) podatak definiraju kao pojam koji se opisuje i kvantificira funkciju objektivne stvarnosti, odnosno elementarno stanje objektivne stvarnosti pod čime mislimo na nedjeljivu cjelinu unutar sistema sa bar jednom vezom sa nekim drugim elementom, u određenom trenutku. Leksikografski zavod „Miroslav Krleža“ navodi da podatak sam za sebe nema nikakvo značenje, ali skup podataka čini osnovu za definiranje informacije.

Dok je informacija „skup podataka s pripisanim značenjem, osnovni element komunikacije koji, primljen u određenoj situaciji, povećava čovjekovo znanje.“(Leksikografski zavod Miroslav Krleža, bez dat.).

„Baza podataka nije ništa drugo nego skup povezanih, organiziranih podataka te da korisnici (u pravilu) bazu podataka doživljavaju kao skup (povezanih) tablica.“(Rabuzin, 2011, str 1).

Rabuzin (2014) navodi da se svaka baza podataka temelji na odgovarajućem modelu podataka po kojemu nose ime tako da postoje:

- hijerarhijske baze podataka
- mrežne baze podataka
- relacijske baze podataka
- temporalne baze podataka
- objektno/relacijske baze podataka
- deduktivne baze podataka
- objektno orijentirane baze podataka

Tijekom vremena su ljudi prepoznali moć informacija i tu dolazi potreba ne samo za spremanjem podataka već za bržim dolaskom do informacija, a to omogućuju skladišta podataka.



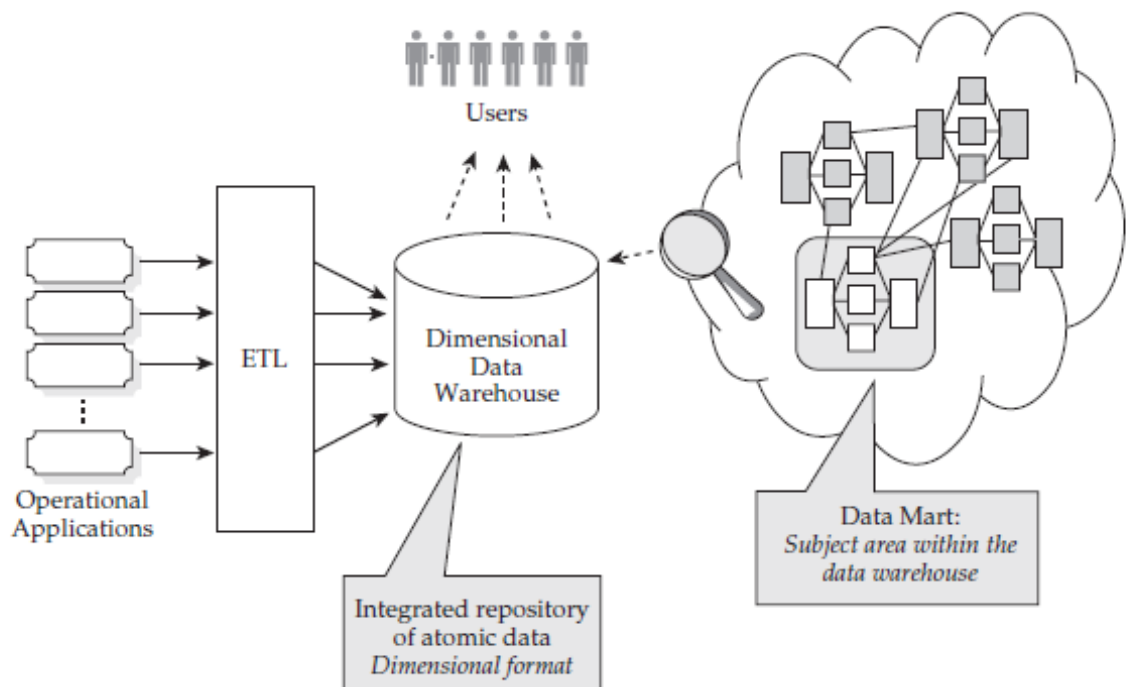
## 2. Metode i tehnike rada

Prije nego što sam počela s pisanjem rada, tijekom čitanja knjiga, članaka i drugih izvora zapisivala sam ono što sam smatrala potrebnim objasniti u ovom radu. Popis literature čije sam informacije koristila u radu nalazi se na kraju ovog rada. U radu su korišteni različiti tipovi izvora literature, primjerice; primarni, sekundarni i tercijarni izvori iako sam se trudila doći do primarnih izvora. Većina knjige bile su preporuka direktora odjela koji mi je omogućio razvitak praktičnog dijela rada, Tomislava Lisaka, kojemu ovim putem zahvaljujem, te članci s Interneta i dijelovi prezentacija s kojim mi je mentor približio ovo područje.

Za potrebe praktičnog dijela koristila sam Informatica Power Centar s kojim sam se upoznavala prilikom izrade rada. Informatica je korištena u svrhu razvitka skladišta podataka. Za pregled podataka s izvora, kreiranje tablica skladišta i pregled učitanih podataka u skladište korišten je alat Toad for Oracle. Dijagrami izvora i skladišta podataka izrađeni su u VisualParadigm alatu.

### 3. Skladište podataka i data mart

Prije nego što objasnim što je data mart bitno je znati što je skladište podataka, od kuda potreba za skladištem podataka i kako se ono razvijalo.



Slika 1: Ralph Kimballova arhitektura skladišta podataka (Izvor: Adamson, 2010)

#### 3.1. Skladišta podataka

Prema R. Kimball-u i M. Ross (2004), organizacije sadrže informacije koje su sredstvo za operativno vođenje evidencije i analitičko odlučivanje. Transakcijski sustavi omogućuju spremanje podataka, a skladišta podataka su mjesta gdje se čuvaju velike količine potrebnih podataka i iz njih se iščitavaju informacije.

Postoji više definicija skladišta podataka, Billa Inmon, kojeg nazivaju jednu od njih je ona „kumom“ skladištenja podataka, definirao je skladište podataka kao „skup subjektno orijentiranih, integriranih, vremenski ovisnih i nepromjenjivih podataka za podršku poslovnom odlučivanju“ dok Oreščanin (2013) pojašnjava te značajke kao:

- Subjektna orijentiranost znači da su podaci u skladištu podataka organizirani tako da daju informacije o pojedinom poslovnom subjektu (prodaji, naplati). Subjekt je jednoznačno definiran bez obzira da li podaci o subjektu dolaze iz jednog ili više sustava. Subjektna orijentiranost često se naziva podatkovna orijentiranost.
- Integriranost podrazumijeva dosljednost u sadržaju, nazivima i formatima podataka koji mogu doći iz različitih izvora, a najveći problem joj predstavljaju podaci na izvorištu koji su često nekvalitetni. Najpoznatiji takav primjer je format zapisa datuma. Datumu je određen jedinstven format, on je zapisan u dimenzijsku tablicu datuma i njegov vanjski ključ nam služi za dodavanje podatka o datumu.
- Vremenska ovisnost nam omogućuje analitičku funkcionalnost i analizu trendova poslovanja kroz vrijeme, a to se ostvaruje uz pomoć snimaka (engl. snapshot) transakcijskih podataka uzetih u redovnim vremenskim periodima i uz podatke je zapisano kada su uzeti ti snimci.
- Nepromjenjivost određuje da se podaci jednom učitani u skladište podataka više ne mogu promijeniti. Često su razvijeni načini koji poništavaju prethodne upise, ali taj dio mora biti dio sporazuma s naručiteljima.

Prema Kimball i Caserta (2004) skladištenje podataka je proces uzimanja podataka iz naslijeđenih i transakcijskih baza podataka, te zatim pretvaranja u organizirane informacije u pristupačnom formatu kako bi se potaknula analiza podataka i podržalo donošenje poslovnih odluka utemeljenih na činjenicama.

Prema Oreščaninu (2005), najveće razlike između skladišta podataka i transakcijskog sustava su te da je skladište podataka zasnovano na bazama podataka koje su orijentirane na velik broj istovremenih korisnika i brzo zapisivanje, čitanje i izmjenjivanje malih količina podataka.

Cilj skladišta podataka (Kimball i Caserta , 2013) je:

- učiniti informacije lako dostupnima
  - pod navedenim se misli da informacije moraju biti lako razumljive, intuitivne i jasne poslovnim korisnicima, što podrazumijeva također da aplikacije i alati koji prikazuju te podatke moraju biti jednostavni za korištenje poslovnim korisnicima i brzo isporučivati podatke
- dosljedno prikazivati informacije
  - podaci moraju biti vjerodostojni, a vjerodostojnost se ostvaruje pročišćavanjem podataka, osiguravanjem kvalitete podataka i odabirom prikaza za iste mjere koje se u različitim izvorima prikazuju različito

- prilagođavanje promjenama
  - nakon kreiranja skladišta podataka moguće su promjene kod uvjeta poslovanja, potreba korisnika, podataka i tehnologija kojima rukuju poslovni korisnici, skladište podataka se mora moći prilagoditi tim novim promjenama jer čuva povijest
- pravovremeno prikazivanje podataka
  - sirovi podaci moraju se moći pretvoriti u informacije u kratkom vremenskom roku od nekoliko sati, minuta ili sekundi (zavisi o potrebama poslovnih korisnika)
- učinkovito upravljati pristupom povjerljivih informacija organizacije
- služiti kao autoritativna i pouzdana osnova za bolje donošenje odluka
- biti prihvaćen od strane poslovne zajednice naručitelja
  - rješenje skladišta podataka je opcionalno, za razliku od rješenja transakcijskih sustava, te se smatra uspješnim samo ako ga poslovni korisnici koriste bez obzira na eleganciju rješenja

Prema Kimball i Caserta (2004), top pet „pogrešnih“ shvaćanja pojma skladišta podataka su to da je skladište podataka:

- Proizvod
  - Skladište podataka obuhvaća analizu sustava, manipulaciju i čišćenje podataka, praćenje kretanja podataka, modeliranje spremanja podataka i pristup podacima. Nijedan proizvod ne može biti definiran istim koracima kako bi ostvario isti cilj i postići sve zadatke uključene u izgradnju skladišta podataka te se skladište podataka, u skladu s time, ne može prodati kao proizvod.
- Jezik
  - Skladište podataka se ne kodira na način na koji se uči primjena XML, SQL, C ili bilo kojeg drugog programskog jezika te se iz tog razloga ne može nazvati jezikom. Također, skladište podataka se sastoji od nekoliko komponenti od kojih svaka može zahtijevati jedan ili više jezika programiranja ili specifikacije podataka.
- Projekt
  - Skladište podataka sastoji se od mnogih projekata (i faza projekata) te je bolje shvaćanje da je to proces koji se sastoji od projekata planiranih na razini poduzeća, ali uspostavljenih vlastitim vremenskim okvirom i proračunom. Najbolji primjer su data mart-ovi koji sadrže usklađene dimenzije i

standardizirane činjenice te se svaki integrira u jednu jedinstvenu jedinicu skladišta podataka poduzeća koje se razvija i raste kako se završava svaki projekt s data mart-om.

- Model podataka
  - Sam model podataka ne čini skladište podataka jer ono mora uključivati ETL proces koji podacima puni model podataka bez kojih bi taj model bio beskoristan.
- Kopija transakcijskog sustava
  - Skladište podataka nije kopija transakcijskog sustava jer kopija, kao ni model podataka, ne transformira podatke što je jedan od najbitnijih čimbenika skladištenja

Uvođenje skladišta podataka sve je popularnije zbog svojih prednosti, a to su prema Kimballu i Casertu (2004):

- Smanjenje troškova
  - Troškovi skladištenja podataka su početni troškovi postavljanja i tekući troškovi održavanja te su ti troškovi znatno manji od drugih opcija s kojima se postiže isti rezultat
- Pобољшanje performansi
  - uspostavljanjem dobre preraspodjele radnih opterećenja dobiva se istovremeno i lako pokretanje izvješća, upita i složene analize te su time poboljšane performanse
- Široka dostupnost
  - Ugradnjom novih rješenja konstrukcije omogućuje se dostupnost; kao rješenja navedeni su ugradnja selektivne redundancije, dvostruki izvori napajanja, sigurnosne mreže, zrcaljenjem diska i automatskim rješenjima kvarova poslužitelja
- Smanjenje administracijskih poslova
  - Postiže se automatskim dodjeljivanjem prostora, održavanjem indeksa, smanjenjem podešavanja performansi te integriranjem komponenti
- Skalabilnost
  - ostvarena modularnim dizajnom za svaku komponentu
- Pouzdanost
  - Otklanjanjem zahtjeva za tešku i neizvjesnu integraciju diskovne memorije, transakcijskih sustava, DBMS-a i procesora te njihovu međusobnu skladnost, odnosno zajednički rad ostvaruje se pouzdanost

- Brža provedba
  - DW uređaji se mogu implementirati bez regresijskog ili integracijskog testiranja što uvelike smanjuje sveukupno vrijeme implementacije. U današnje vrijeme postoji puno alata, poput PowerBI-a i Tableau-a, koji omogućuju da se do lako čitljivih izvještaja dolazi jednim klikom nakon što je njegova struktura spremljena

Prema Oreščaninu (2004), skladište podataka je namijenjeno za korištenje od strane pojedine poslovne funkcije ili djelatnika na jednoj lokaciji dok Rabuzin (2014) ističe da su menadžeri (ili druge osobe zadužene za donošenje odluka) najčešći korisnici skladišta podataka i to u određenim vremenskim intervalima.

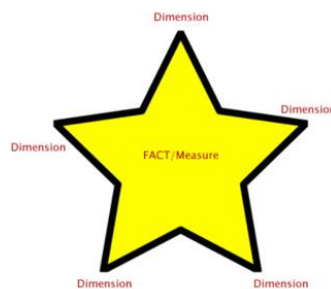
### 3.1.1. Višedimenzionalni modeli skladišta podataka

Najpopularniji višedimenzionalni modeli skladišta podataka su sheme zvijezde i snježne pahulje, a razlikuju se u tome što shema zvijezde ne koristi normalizaciju dok shema snježne pahulje koristi normalizaciju za uklanjanje suvišnih podataka. Međutim, za oba modela su potrebne tablice dimenzija i činjenica („Tech Differences“, 2017). Oba modela bit će detaljnije objašnjena u nastavku .

#### 3.1.1.1. Model zvijezde

Adamson (2010) je za dimenzionalni dizajn zvijezde napomenuo da su povezane dimenzije grupirane kao stupci u tablicama dimenzija dok su činjenice pohranjene kao stupci u tablici činjenica te kad se nacrtaju tablice dimenzija i činjenična tablica u sredini dobiva se izgled poput zvijezde ili zvjezdice.

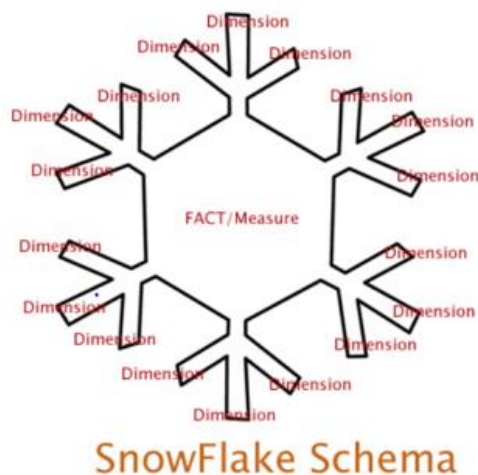
Rabuzin (2018) ističe da je model, odnosno shema, zvijezde idealan za procesiranje zbog redundantnosti.



Slika 2: Model zvijezde (Izvor: QuestPond, 2017)

### 3.1.1.2. Model pahuljice

Adamson (2010) navodi da je za dimenzije snježne pahulje potreban proces sličan procesu normalizacije koji se provodi kako bi se osigurao referentni integritet podataka u transakcijskim sustavima koji podržavaju široku paletu istodobnih transakcija koje su vrlo detaljne, ali sama normalizacija nije potrebna kod izgradnje skladišta jer bi zbog nje došlo do umanjenja upotrebljivosti, kompliciranja ETL-a i mogućeg narušavanja performansi. Ono zbog čega se ljudi odlučuju za model pahuljice, prema Adamsonu (2010) je to što se temelji na tehnologiji, odnosno u arhitekturi pahuljice možda najbolje proizvodi funkcioniraju s dizajnom snježne pahulje te je još jedan razlog što se neki specifični izazovi modeliranja ne mogu ispuniti bez dekompozicije dimenzija na više tablica.



Slika 3: Model pahuljice (Izvor: QuestPond, 2017)

Rabuzin (2018) navodi da se model snježne pahuljice načelno izbjegava zbog toga što model korisniku izgleda znatno složeniji te su upiti koji se postavljaju sporiji od onih koji se postavljaju kada je model zvijezde u pitanju.

## 3.2. Data mart

Skladište podataka omogućuje pregled poduzeća dok je data mart podskup skladišta podataka koji pruža pogled na odjel, odnosno decentraliziranu pohranu. U vezi s tim, skladište podataka je vrlo veliko i integrirano, ima visoki rizik od neuspjeha i poteškoće u njegovoj izgradnji dok je data mart lako sastaviti i rizik je manji jer se razvijaju u manjim dijelovima. („TechDifferences“, 2017)

Adamson (2010) smatra da su rezultati uvođenja brži i jeftiniji te se zbog toga mnoge organizacije odlučuju upravo na uvođenje data mart-a te da se one nerijetko grade od nule. Zapakirani data mart-ovi dostupni su kao dodaci paketa (eng. add-onsi) kod paketa za operacijske aplikacije. Temeljene su na različitim tehnologijama, a korisnička se publika može osloniti na zasebne infrastrukture za upite i izvješćivanje.

Sljedeća tablica prikazuje razlike između skladišta podataka i data mart-a nakon čega će biti predstavljeni tipovi arhitekture data mart-a.

Tablica 1: Razlike skladišta podataka i data mart-a

BASIS FOR COMPARISON	SKLADIŠTE PODATAKA	DATA MART
<b>Tip sistema</b>	Centralizirani tip	Decentralizirani tip
<b>Detaljizacija</b>	Detaljno	Sumirani
<b>Korist denormalizacije</b>	Blaga denormalizacija	Visok stupanj denormalizacije
<b>Model podataka</b>	„Prema gore“	„Prema dolje“
<b>Priroda</b>	Fleksibilan, orijentiran na podatke i dugoročnost.	Ograničen, orijentiran na projekt, a ne na dugoročnost.
<b>Tipovi sheme</b>	Zvijezda i pahuljica	Zvijezda i pahuljica
<b>Jednostavnost izgradnje</b>	Teža izgradnja	Jednostavnija izgradnja

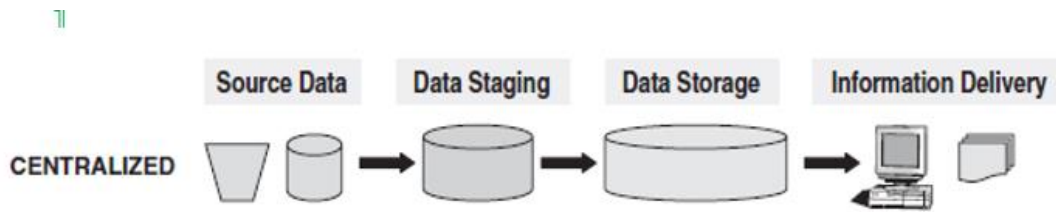
(Izvor: TeshDifferences, 2017)

Podskum skladišta podataka o kakvom se govori u ovom poglavlju može biti u tipu:

- Centraliziranog data mart-a
- Neovisnog data mart-a
- Saveznog data mart-a
- „Hub-and-Spoke“
- Bus

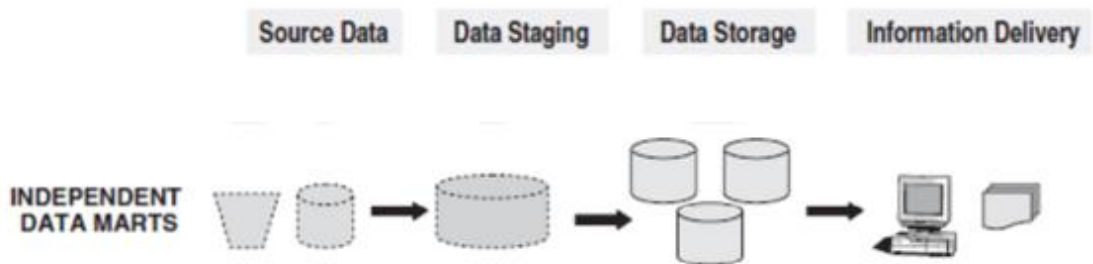
Prema Ponniah-u (2010) centraliziran data mart uzima u obzir zahtjeve za informacijama na razini poduzeća. Podaci se nalaze na najnižoj razini zrnatosti, pohranjeni u trećem normalnom obliku dok su povremeno uključeni i neki sažeti podaci. Pristupiti se može normaliziranim podacima u središnjem skladištu podataka i važna značajka je da ne postoje zasebni data mart-ovi.





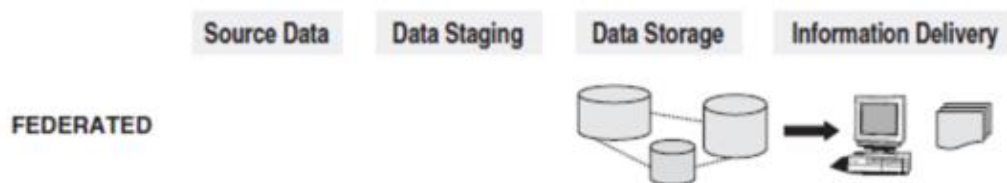
Slika 4: Centraliziran data mart(Izvor: Ponniah, 2010)

Ponniah (2010) za neovisan data mart navodi da se razvija u tvrtkama u kojima se razvijaju data mart-ovi za zasebne organizacijske jedinice i njihove specifične svrhe. Iako svaki data mart služi određenoj organizacijskoj jedinici, te odvojene podatkovne baze ne pružaju “jednu verziju istine“. Data mart-ovi neovisni su jedan o drugome. Kao rezultat, različiti data mart-ovi vjerojatno imaju nedosljedne definicije i standarde podataka. Takva odstupanja ometaju analiza podataka preko podatkovnih centara.



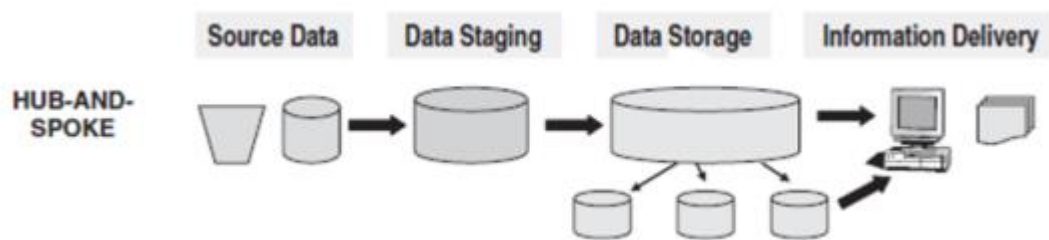
Slika 5: Neovisan data mart (Izvor: Ponniah, 2010)

Ponniah (2010) ističe da su neke tvrtke prije investiranja u skladište podataka investirale u alate za podršku o odlučivanju u obliku transakcijskih sustava, ekstrahiranih skupova podataka, primitivnih data mart-ova i slično, te se takve tvrtke odlučuju ne odbaciti sva ta velika ulaganja i krenuti od nule, već se odlučuju za federalni arhitektonski tip gdje se podaci mogu fizički ili logički integrirati kroz područja zajedničkih ključeva globalnih meta podataka, distribuiranih upita i drugih metoda. U ovom arhitektonskom tipu ne postoji jedno jedinstveno skladište podataka.



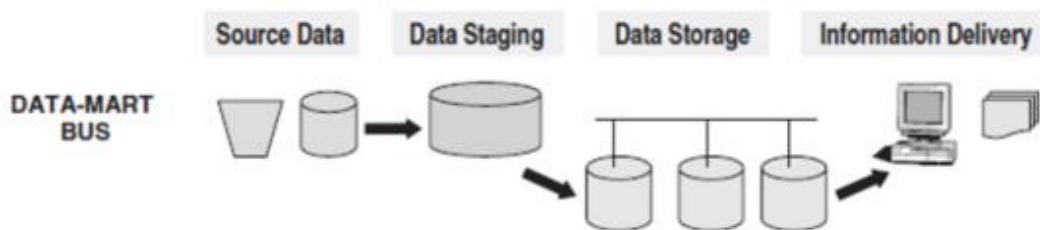
Slika 6: Savezni data mart-ovi (Izvor: Ponniah, 2010)

Ponniah (2010) uspoređuje hub-and-spoke arhitekturu s centraliziranom arhitekturom skladišta podataka i ističe njihovu sličnost: ovisni data mart-ovi dobivaju podatke iz centraliziranog skladišta podataka te se razvijaju za različite svrhe poput analitičkih potreba odjela, specijaliziranih upiti, rudarenje podacima i tako dalje. Svaki ovisan data mart može biti normaliziran, denormaliziran, sažet ili dimenzionalne strukture podataka na temelju individualnih zahtjeva. Većina upita upućuje se na ovisne data mart-ove iako se centralizirana skladišta mogu koristiti za upite. Ovaj arhitektonski tip proizlazi iz usvajanja „odozgo prema dolje“ (eng. „top-down“) pristupa razvoju skladišta podataka.



Slika 7: "Hub-and-Spoke" data mart (Izvor: Ponniah, 2010)

Prethodna arhitektura predstavlja Inmonov pristup, a Bus arhitektura je razvijena prema Kimbalu. Ponniah (2010) ističe da je najznačajnije što pruža ova arhitektura to što nakon izgradnje prvog data mart-a (tako zvanog „supermart-a“) uz pomoć poslovnih dimenzija i podataka o mjerama, nastavlja se izgradnja u koju se može uključiti postojeći data mart, tj. rezultat se dobiva logičkim integriranjem data mart-ova koje će pružiti poduzeću pogled na podatke. Ovaj arhitektonski tip je rezultat usvajanja poboljšanog pristupa odozdo prema gore za razvoj skladišta podataka. Naziva se bus zato što se ostvaruje spajanjem dijelova, koje čine zasebne data mart-ove.



Slika 8: Bus arhitektura(Izvor: Ponniah, 2010)

## 4. Proces razvoja skladišta podataka

Prije procesa ekstrahiranja, transformiranja i učitavanja podataka (ETL) u skladište podataka, kao i razvitka data marta, potrebno je izraditi logički model podataka, provesti dizajnersko modeliranje podataka te krenuti u proces ETL-a. Nakon što je proces izvršen potrebno ga je testirati kako bi se uvjerali da će poslovni analitičari moći dobiti sve potrebne informacije.

### 4.1. Priprema za ETL

Prema Kimball i Caserta (2004), prije razvoja skladišta podataka bitno je definirati nazive tablica, nazive stupaca, vrstu tablice, vrste polako mijenjajućih dimenzija SCD (eng. Slowly Changing Dimension), izvornu bazu podataka, ime izvornog stupca te transformacije. Pojmovi koji su specifični za skladišta podataka pojašnjeni su u nastavku.

Vrsta tablice pokazuje je li tablica činjenica (eng. Fact), dimenzijska (eng. Dimension) ili subdimenzija (eng. outrigger).

Polako mijenjajuće dimenzije mogu biti tipa 1, 2 i 3. U dimenzijama postoje komponente koje se polako mijenjaju i važno je naznačiti njihove promjene. Ovaj pokazatelj može varirati za svaki stupac u dimenziji. Na primjer, unutar dimenzije korisnika, prezime može zahtijevati ponašanje tipa 2 (zadržati povijest), dok će za prvo ime trebati vrsta 1 (prebrisati). Često se za dimenziju 3 kaže da je izvedenica iz prve dvije te se rjeđe koristi. Prema Adamsonu (2010):

- promjene tipa 1 mogu se identificirati usporedbom izvornog zapisa s aktualnim zapisom dimenzija za prirodnu ključnu vrijednost. Ako se bilo koji atribut tipa 1 ne podudara, došlo je do promjene tipa 1 te je potrebno primijeniti promjenu tipa 1 na svaki redak dimenzijske tablice koji dijeli isti prirodni ključ, to jest prepisati novu vrijednost preko trenutne vrijednosti u skladištu podataka s čime se gubi vrijednost koja je bila prethodno zapisana. Odnosno, ukoliko ja promijenim prezime, a atribut prezime sam zadala da je tipa 1, tada se gubi podatak da je moje prezime bilo Bodulušić i zna se samo moje trenutno prezime.
- promjene tipa 2 mogu se identificirati usporedbom izvornog zapisa s zapisima dimenzija koji imaju isti prirodni ključ kao i kod promjene tipa 1, ali ukoliko ne postoji odgovarajući zapis s istim vrijednostima za sve attribute tipa 2, a došlo je do promjene tipa 2 što zahtijeva zatvaranje trenutnog zapisa, odnosno u

odgovarajući atribut se zapisuje da je podatak vrijedio do ovog trenutka u vremenu, te se novi zapis kreira s promijenjenim vrijednostima podataka i ovaj novi zapis postaje trenutni zapis s nepoznatim vremenom do kojeg traje zapis ili dogovorenim datumom koji će to označavati. Svaki zapis sadrži efektivno vrijeme i vrijeme isteka za određivanje vremenskog razdoblja između kojeg je zapis bio aktivan.

- Promjena tipa 3, pohranjuje dvije vrijednosti za jedan podatak, a to su: prethodnu vrijednost i trenutnu vrijednost odabranog atributa tako da u svakom trenutku imamo poznatu trenutnu vrijednost koja se zapisuje kao stara vrijednost i novu vrijednost koja postaje trenutna. („Oracle“, bez dat.)

Prema Kimball i Caserta (2004, str. 61), ključni kriterij uspjeha skladišta podataka su čistoća i kohezivnost podataka u njemu. Za ostvarivanje čistoće i kohezivnosti u skladištu podataka posebnu pažnju potrebno je posvetiti:

- Jedinstvenim identifikatorima i prirodnim ključevima (eng Unique identifiers and natural keys)
  - „Jedinstveni identifikatori prikazuju stupce koji jedinstveno predstavljaju red u tablici.“ S referentnog stajališta integriteta, jedinstveni identifikator je primarni ključ tablice; u skladištima primarni ključ je najčešće umjetan, a iako je jedinstven sa stajališta ETL-a, nema dovoljno podataka da bi se utvrdilo je li red jedinstven. U tu svrhu uveden je jedan prirodni ključ koji se koristi za jedinstveno opisivanje retka.
- Tipovima podataka (eng. data types)
  - U skladištima podataka, čest je slučaj da je atribut datum tipa podataka text jer on predstavlja ključ na dimenzijsku tablicu datuma. To je samo jedan od primjera kada naziv atributa nije u skladu s tipom podatka.
- Odnosima među tablica
  - Razumijevanje povezanosti tablica od vitalnog je značaja za osiguravanje točnosti spajanja tijekom preuzimanja podataka
- Diskretnom odnosu
  - Uobičajeno je da postoji jedna pregledna tablica koja pohranjuje sve statičke referentne podatke za sve tablice u bazi podataka. Ona sadrži stupac koji identificira koja tablica i stupac podržana skupina redaka podržava.
- Kardinalnosti odnosa i stupaca.
  - Kardinalnost prikazuje maksimalan broj instanci jednog entiteta koji može biti povezan s instancom drugog entiteta

- Odnosi jedan na jedan se prepoznaje promatranjem da je odnos između primarnih ključeva svake tablice
- Jedan-na-više je odnos koji spaja reference vanjskih ključeva. Lako je prepoznati primijetivši da se atribut koji nije ključ u tablici odnosi na primarni ključ druge tablice. Taj atribut naziva se vanjskim ključem i potrebno je da su svi vanjski ključevi dobri, odnosno da postoje primarni ključevi na koje pokazuju
- Više-prema-više je odnos koji uključuje tri tablice od kojih su dvije tablice s asocijativnom tablicom te središnja tablica koja sadrži složeni primarni ključ i dva vanjska ključa, jedan za primarni ključ jedne tablice i drugi za primarni ključ druge tablice.

Adamson (2010) dimenzionalno modeliranje predstavlja kao dizajnerski pristup optimiziran za analitičke sustave koji bilježi kako se mjere procesi, te podatke o mjerenjima naziva činjenicama dok podatke koji pružaju kontekst za mjerenja naziva dimenzijama.

Kimball, prema Kimball-u i Ross-u (2013) definira 4 osnovna koraka dimenzijskog modeliranja u kojima odabiremo poslovni proces koji želimo modelirati, zrnatost tog procesa te dimenzije koje se odnose na redak u činjeničnoj tablici i identificiramo činjenice kojima će se popuniti činjenična tablica.

Dok su Kimball i Caserta (2004) te korake razdvojili i definirali na način da nam je potrebno:

**1. imati plan.**

- a. ETL proces mora biti logičan i dokumentiran. Kartu logičkih podataka osigurava arhivara skladišta podataka i predstavlja specifikaciju za ETL tim za kreiranje fizičkog ETL posla, takozvanog mappiranja. Karta logičkih podataka je dokument s metapodacima koji se na kraju prezentira testerima osiguranja kvalitete i krajnjim korisnicima da opišu točno što se radi između izvorišnih sustava i skladišta podataka.

**2. identificirati kandidate izvora podataka**

- a. Nakon što se utvrde poslovni ciljevi najviše razine, potrebno je identificirati potencijalne izvore podataka koji će omogućiti ispostavu podataka koji vode do informacija za postizanje tih ciljeva. Nakon identificiranja tih podataka potrebno je profilirati te podatke.

**3. analizirati izvorne sustave pomoću alata za profiliranje podataka**

- a. Podaci u izvornim sustavima moraju se pomno ispitati radi kvalitete podataka, cjelovitosti i prikladnosti jer punjenjem skladišta podacima

koji nisu kvalitetni ne ostvarujemo svrhu samog skladišta. Kvaliteta podataka u nadležnosti je osoba koje su najkvalificiranije za obavljanje tog posla: to može biti ETL tim ili naručitelj s kojim ETL tim surađuje, ali bitno je doći do podataka koji su kvalitetni i moguće anomalije ukloniti u ovom koraku.

**4. definiranje poslovnih pravila (eng Receive walk-through of data lineage and business rules.)**

- a. Nakon što su izvori podataka profiltrirani i poznat je ciljani model podataka, arhitekt skladišta podataka i poslovni analitičari moraju upoznati ETL arhitekta i programere s područjem nad kojim oni trebaju provesti ETL proces kako bi podaci u skladištu bili pravilno zapisani.

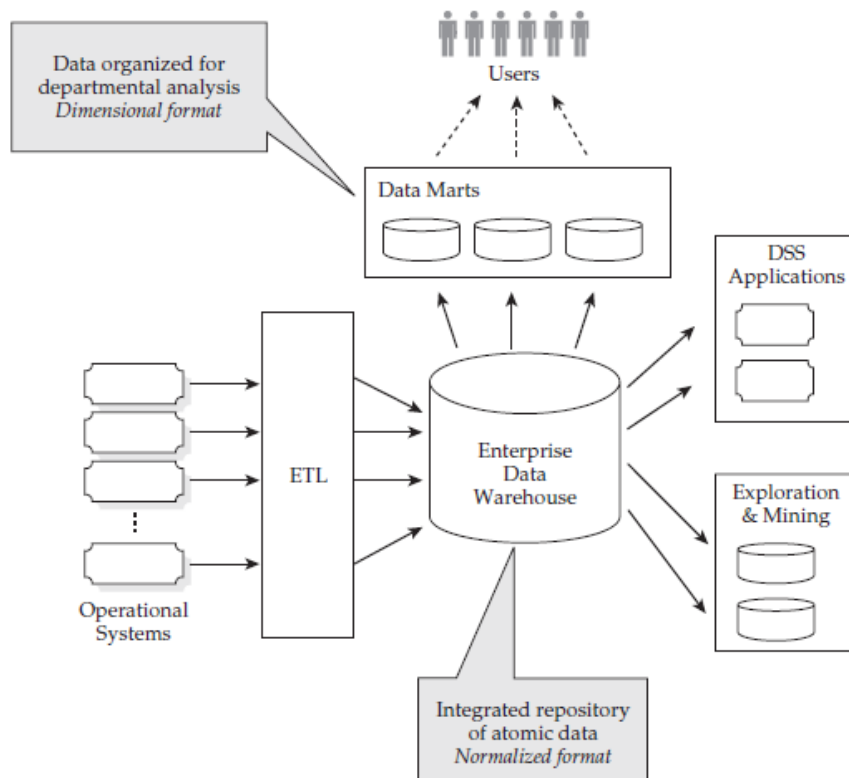
**5. provjera modela podataka o skladištu podataka**

- a. ETL tim mora u potpunosti razumjeti fizički model skladišta podataka što uključuje i temeljito razumijevanje kako dimenzije, činjenice i druge posebne tablice u dimenzionalnom modelu rade zajedno kako bi implementirali uspješna ETL rješenja.

**6. potvrditi izračune i formule.**

- a. S krajnjim korisnicima provjerite sve izračune navedene u podatkovnoj liniji kako ne bi potrošili vrijeme na kodiranje pogrešnih algoritama u ETL procesu.

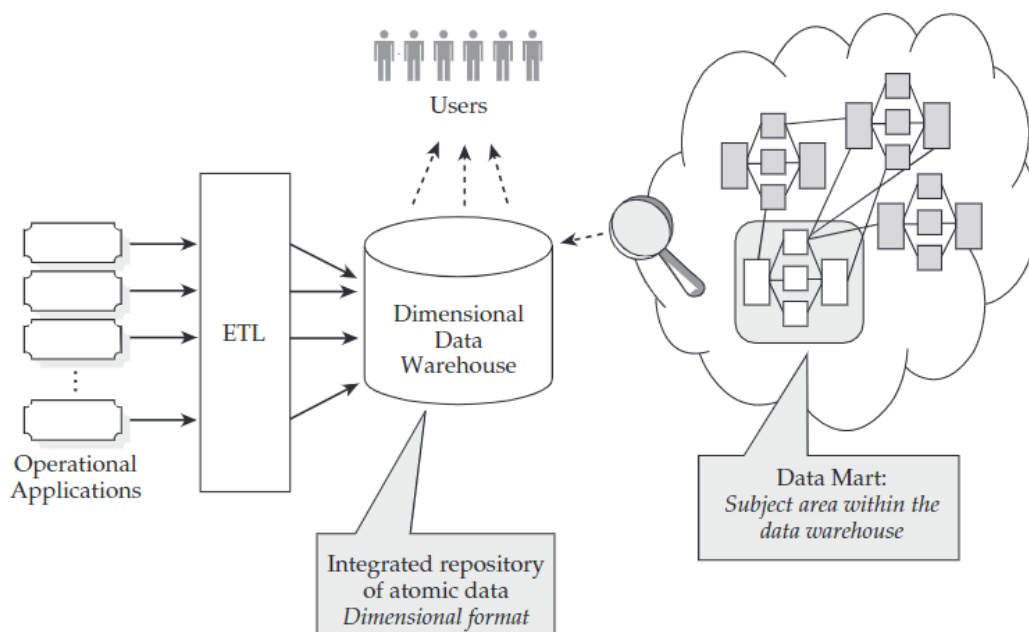
*W. H. Inmonu svoj pristup izgradnje skladišta podataka naziva „Corporate Information Factory“ te ističe „...if I had to design a data mart tomorrow, I would not consider using any other approach.“ (Adamson, 2010, str. 19) U arhitekturi „Corporate Information Factory“ nalazi se podatkovno skladište poduzeća koje ne služi za izravno ispitivanje pomoću analitičkih aplikacija niti za alate poslovne inteligencije već je njegova svrha uvesti dodatne pohrane podataka posvećene različitim analitičkim sustavima. Skladište podataka poduzeća obično se izrađuje iz podataka koji su zapisani u relacijskim bazama podataka, a Inmon zagovara korištenje treće normalne forme u dizajnu skladišta podataka te da data mart-ovi mogu agregirati podatke iz atomarne reprezentacije u skladište podataka tvrtke. (Adamson, 2010).*



Slika 9 Inmonova arhitektura(Izvor: Adamson, 2010)

Za razliku od Inmona, prema Adamsonu (2010), Kimball je razvio arhitekturu poduzeća za skladište podataka koja omogućava integrirano spremište atomarnih podataka i oslanja se na dimenzionalni dizajn koji podržava analitiku. Kao i kod Inmonove arhitekture, potrebno je postupak početi odvajanjem skladišta od transakcijskog i analitičkog sustava te nakon toga pomoću ETL procesa objediniti informacije iz različitih transakcijskih sustava, integrirati ih i učitati u jedno skladište. Dimenzijsko podatkovno skladište razlikuje se od Inmonovog skladišta podataka poduzeća jer je

- projektiran u skladu s principima dimenzionalnog modeliranja, tj. sastoji se od zvijezda sheme ili kocke. To je u suprotnosti s Inmonovim pristupom, gdje se skladište podataka poduzeća dizajnira na principima ER modeliranja.
- analitički sustavi mogu izravno pristupiti dimenzionalnom skladištu podataka.



Slika 10: Arhitektura Ralph Kimballa (Izvor: Adamson, 2010)

Sve te arhitekture uspješno koriste tvrtke i organizacije diljem svijeta. Arhitektura skladišta podataka može se u potpunosti podudarati s jednom od tih paradigmi, ili biti ukomponirane od elemenata svake od njih. Prema Adamsonu (2010), Inmonova Tvornica korporativnih informacija i Kimballovo skladište podataka o dimenzijama imaju u fokusu poduzeće, cilj im je podržati analitičke potrebe u poslovanju ili organizaciji. Dok obje arhitekture poduzeća uključuju integrirano spremište atomarnih podataka razlikuju se u načinu na koji je dizajniran i korišten pristup izgradnje. Kimball zagovara uporabu dimenzionalnog dizajna, Inmon ne. Kimball dopušta izravno spremanje repozitorija, Inmon ne. Svaka od tih arhitekture ima svoje mjesto u dizajnu skladišta podataka. (Adamson, 2010)

Karakteristike dimenzijskog modeliranja su zrnatost, mjere i umjetni ključevi.

Zrnatost podataka u skladištu podataka odnosi se na razinu detaljnosti s time da se smanjenjem razine detalja povećava zrnatost podataka. Najniža razina detalja zahtjeva mnogo prostora za pohranu podataka. (Ponniah, 2010). Dok Rabuzin (2014) napominje da će se zrnatost odrediti ovisno o poslovnom slučaju, odnosno o tome na koje upite želimo moći dobiti odgovore. Bill Inmon je izjavio: „Najvažnije pitanje s kojim se IT menadžer suočava ove godine jest hoće li najprije izgraditi skladište podataka ili data mart.“ (Ponniah, 2010, str. 29).



Mjere su numerički i aditivni podaci u činjeničnoj tablici s kojima ima smisla pratiti kretanje kroz vrijeme i zbrajati ih ili računati prosječne vrijednosti. Sve mjere u tablici moraju biti iste zrnatosti.

Prema Kimball i Caserta (2004), jedinstveni identifikatori prikazuju stupce koji jedinstveno predstavljaju red u tablici, što se kosi s referentnim integritetom koji kaže da je jedinstveni identifikator primarni ključ za tablicu. Razlog tomu je to što je primarni ključ u skladištima podataka umjetan, i iako je jedinstven s ETL stajališta, nema dovoljno informacija s kojima možemo odrediti njegovu jedinstvenost.

U svakoj pravilno dizajniranoj transakcijskoj tablici, osim primarnog ključa, postoji barem jedan prirodni ključ kojeg tvrtka definira i koristi za jedinstveno opisivanje retka.

Poslovni ključ (eng. Business key), koji se naziva još i prirodni ključ, je indeks koji identificira jedinstvenost retka na temelju stupaca koji prirodno postoje u tablici prema poslovnim pravilima. (Wood, 2011) Na primjer, poslovni ključevi su korisnički kôd u tablici korisnika, kompozitni broj zaglavlja prodajnog naloga i broj retka stavke prodajnog naloga unutar tablice s pojedinostima prodajnog naloga.

Alternativa poslovnom ključu je zamjenski ključ, nazivan još i tehnički ključ (eng. technical key), koji je obično veliki cijeli broj s automatskim generiranim sjemenom identiteta koji učinkovito povezuje tablicu činjenica s dimenzijom, a on nastaje u trenutku odvajanja činjenica i dimenzija izvora podataka unutar područja za opuštanje. On osigurava održavanje veze između činjenice i dimenzije. Autor ističe da su zamjenski ključevi učinkovitiji od poslovnih ključeva jer imaju dodatnu mjeru s kojom lakše dođemo do podatka o subjektu.

## 4.2. ETL Process

Adamson (2010) napominje da je bilo koji proces koji premješta podatke iz jednog sustava ili strukture u drugi ETL proces. Za taj proces učitavanja je uvijek pod vremenskim pritiskom i zbog toga je potrebno da bude brz i učinkovit. Brzina i učinkovitost se postižu tehnikama izvođenja identifikacije novih i promijenjenih podataka na izvoru prije nego što se izvuku informacije obrada. To može dramatično smanjiti količinu informacija koje se moraju obraditi, ali može imati utjecaja na transakcijski sustav.

ETL sustav mora, prema Kimballu i Casertu (2004, str. 28):

- Najučinkovitije dostavljati podatke krajnjim korisnicima
- Dodavati vrijednost podacima u koracima čišćenja i usklađivanja
- Zaštititi i dokumentirati liniju podataka

Kimball i Casert (2004) također napominju da se cjelokupni proces sastoji od izdvajanja podataka iz originalnog izvora, osiguravanja kvalitete i čišćenja podataka, te usaglašavanje oznaka i mjera u podacima kako bi se postigla dosljednost i to nad svim uključenim sustavima iz kojih se uzimaju podaci, a ti sustavi mogu biti sustavi za upravljanje bazama podataka, transakcijski sustavi, neki oblik hardvera i/ ili komunikacijski protokoli.

### 4.2.1. Ekstraktiranje

Prema Kimball i Caserta (2004), prvi korak kod izgradnje skladišta ili data mart-a je izraditi logičnu mapu koja dokumentira odnos između izvornog sustava i konačnog odredišta koje čine tablice skladišta podataka, odnosno ciljane tablice.

Svako skladište podataka, bez obzira na arhitekturu ili okruženje, uključuje područje za čišćenje i usklađivanje podataka (eng. Stage). To područje može biti realizirano na različite načine. O tome hoćemo li imati mjesto gdje ćemo pohraniti sirove podatke i koliko dugo ćemo ih čuvati na tom mjestu odlučuje se na temelju nekoliko pretpostavki.

Kimball i Caserta (2004) ističu da uspostavljanje stage-a ovisi o brzini prebacivanja podataka s izvora do krajnjeg cilja . Moguće je stagirati podatke u bazu podataka i/ ili u datotečni sustav. Stage se najčešće koristi kao točka oporavka kako bismo mogli doći do podataka bez ponovnog ulaza u izvorni sustav, kako bismo došli do podataka bez ponovne transformacije podataka, kod neuspjelog procesa (statusa procesa „aborted“ ili „failed“), te čuvanja podataka iz transakcijskih sustava koji se prepisuju preko svojih podataka.

### 4.2.2. Transformiranje

Prema Kimballu i Rossu (2013), nakon što su podaci ekstrahirani u ETL sustav, postoje brojne transformacije koje dodaju vrijednost podacima.

Najčešći tipovi transformacija su („Stitchdata““, bez dat.):

- čišćenje koje osigurava konzistentnost formata podataka, ujednačavanje oznaka (npr. ako se negdje koristi 'Male', a negdje 'M', onda ih treba uskladiti tako da je oznaka jedinstvena)
- micanje duplih, odnosno potpuno istih zapisa (uzimajući u obzir SCD)
- utvrđivanje odnosa ključeva među tablicama
- derivacije, odnosno izvučene mjere koje se koriste u drugim mjerama
- filtriranje u obliku odabira samo određenih redaka ili stupaca
- povezivanje podataka iz različitih izvora
- dijeljenje jednog stupca u više njih

- validacija podataka
- sumiranje podataka
- agregiranje različitih izvora podataka
- integriranje, pod čime se misli na uspostavljanje jedinstvenog imena i definicije za isti podatak.

### 4.2.3. Učitavanje

Adamson (2010) ističe da je učitavanje podataka u tablice dimenzija postupan proces zbog identificiranja novih i promijenjenih podataka, upravljanja zamjenskim ključevima i dodavanja prema potrebi. Glavni izazov za ovaj proces uključuje zamjena prirodnih ključeva koji opisuju poslovne transakcije sa zamjenskim ključevima koji se odnose na tablice dimenzija.

Prije bilo kakvog učitavanja, potrebno je odabrati tip učitavanja.

#### 4.2.3.1. Tri tipa učitavanja

Postoji nekoliko tipova učitavanja podataka u tablice skladišta podataka. Razlikujemo učitavanje (eng. load) i dodavanje (eng. append) sadržaja. Prema Ponniah (2010), ukoliko već postoji ciljano tablica, proces učitavanja će ju prebrisati i dodati željene podatke nakon brisanja te će sadržavati samo upravo dodani sadržaj dok bi postupkom dodavanja pridodali željeni sadržaj ciljanoj tablici te bi nakon toga imali sve podatke, odnosno imali bi podatke koji su bili prije ovog dodavanja u tablici i nove upravo dodane podatke.

Tri tipa učitavanja su:

- Početno učitavanje (eng. Initial Load)
- Inkrementalno učitavanje (eng. Incremental Load)
- Potpuno osvježavanje (eng. „Full Refresh“) učitavanje

Prema Ponniah (2010), inicijalno učitavanje se bazira na učitavanju cijelog skladišta podataka u jednom ciklusu i to svaki put od nule, te se ovaj način preporuča kada je dovoljno jedno učitavanje u skladište sa svim potrebnim podacima. Moguće je odvojiti učitavanje u više faza, ali samo u slučaju da su sadržani svi potrebni i prethodno učitani podaci u svakom od pod učitavanja. Prilikom početnog učitavanja tablice koristi se ovaj način učitavanja.

Kod inkrementalnog učitavanja, Ponniah (2010) navodi da se prikupljaju promjene za one zapise kod kojih je došlo do promjene nakon zadnjeg učitavanja, zatim se za te zapise pokreće proces integracije i transformacije i kreiraju se izlazni zapisi koji će se primijeniti na dimenziji. Ono što naznačuje da je došlo do promjene je zapisano specifično vrijeme u

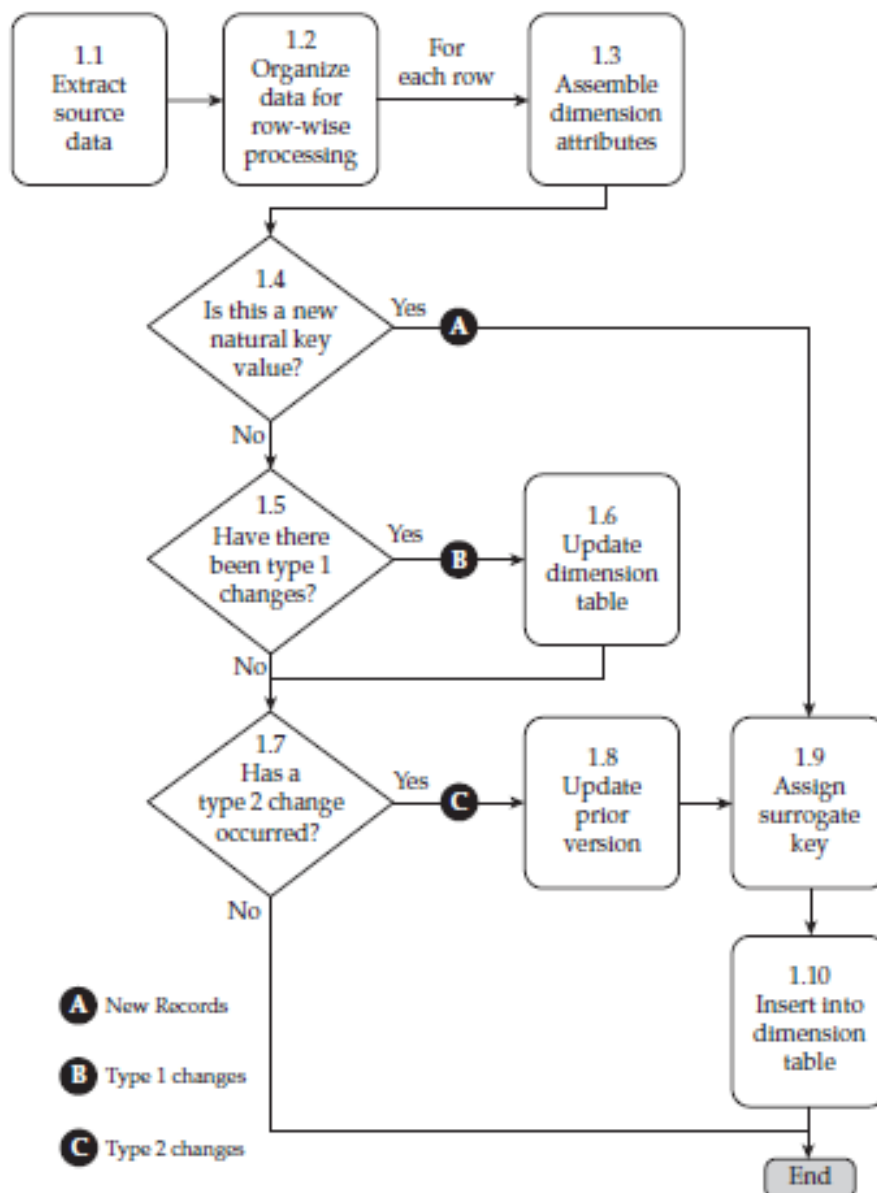
kojemu je došlo do učitavanja podataka u skladište. Na osnovu zapisanog vremena učitavanja i definiranih ključeva možemo provjeriti imamo li od prije te podatke zapisane u skladište, da li je došlo do promjene zapisanih podataka u skladište i imamo li nove podatke, odnosno postoje li podaci koji još nisu pohranjeni u skladište. Nakon definiranja statusa o pohrani podatka postupamo prema dogovorenom/ željenom SCD postupku.

Ponniah (2010) navodi, kao još jedan primjer učitavanja, potpuno osvježavanje (eng. Full Refresh) koje se koristi kada je najbolje rješenje prebrisati sav sadržaj skladišta podataka novim sadržajem. Moguća varijanta je i prebrisati sadržaj novim sadržajem samo nad nekim tablicama. Ova varijanta slična je varijanti početnog učitavanja, a razlika je u tome što u slučaju potpunog osvježavanja podaci postoje u ciljanim tablicama prije primjene dolaznih podataka dok kod varijante potpunog osvježanja se postojeći podaci moraju izbrisati prije dodavanja dolaznih podataka.

#### **4.2.3.2. Vrste učitavanja**

Adamson(2010) razdvaja učitavanje podataka u skladište podataka na dvije glavne kategorije aktivnosti, a to su obrada podataka za tablice dimenzija i obrada podataka za tablice činjenica.

Učitavanje dimenzijske tablice inkrementalni je proces, a ne samo jednokratna stvar. Nužno je redovito pregledati podatke u dimenzijskim tablicama kako bi identificirali nove i promijenjene informacije, te po potrebi unositi ili ažurirati zapise. (Adamson, 2010) Novi zapisi sadrži prirodni ključ koji prethodno nije učitao u dimenzijsku tablicu. Izvorni zapis može se identificirati kao novi tako što će se tražiti njegov prirodni zapis u dimenzijskoj tablici te, ukoliko se ne pronađe, to je novi zapis i potrebno ga je dodati u tablicu. (Adamson, 2010)



Slika 11: Učitavanje dimenzijske tablice (Izvor: Adamson, 2010)

Postupak učitavanja tablice dimenzija mora sadržavati sljedeće glavne zahtjeve:

- Izdvajanje izvornih podataka iz izvora
- Sastavite dimenzijske atribute
- Identificirati nove i izmijenjene zapise o dimenzijama
- Upravljajte zamjenskim ključevima
- Obradite nove zapise
- Promjena tipa procesa 1
- Promjena tipa procesa 2

Prema Adamsonu (2010), novi zapisi su oni zapisi koji imaju prirodni ključ koji još nije pohranjen u dimenzijsku tablicu. Kada se identificira novi zapis, mora mu se dodijeliti zamjenski ključ i umetnuti u dimenzijsku tablicu. Da biste utvrdili da li je dolazni zapis nov, potrebno je provjeriti postoji li prirodni ključ u dimenzijskoj tablici.

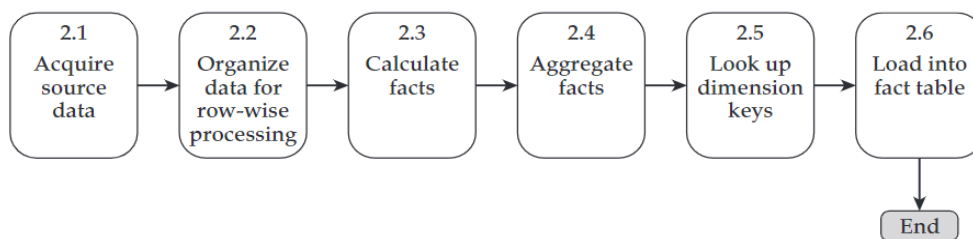
Razumijevanje načina na koji se učitava zvjezdana shema počinje s prepoznavanjem temeljne ovisnosti: svaki redak činjenične tablice sadrži reference vanjskog ključa na retke tablice dimenzija. To podrazumijeva da prije nego se redak može umetnuti u tablicu činjenica, odgovarajuća dimenzija mora sadržavati taj redak. (Adamson, 2010)

Tablica činjenica u dimenzijskom modelu pohranjuje mjerenja učinka koja proizlaze iz događaja poslovnih procesa organizacije. Budući da su mjerni podaci u većini slučajeva najveći skup podataka, ne bi se trebalo ponoviti na više mjesta za višestruke organizacijske funkcije oko poduzeća. Omogućuje pristup poslovnim korisnicima iz više organizacija jedan centralizirani repozitorij za svaki skup podataka mjerenja osigurava korištenje dosljedne podatke u cijelom poduzeću. (Kimball i Ross, 2013)

Izraz činjenica predstavlja poslovnu mjeru. Zamislite da stojite na tržištu gledajući proizvode koji se prodaju i zapisujući jediničnu količinu i prodaju u dolarima iznos za svaki proizvod u svakoj transakciji prodaje. Ova mjerenja su snimljena kao što se proizvodi skeniraju u registar, kao što je prikazano na slici 1-2. Svaki redak tablice činjenica odgovara mjernom događaju. Podaci o svakom od njih redak je na specifičnoj razini detalja, koja se naziva žito, kao što je jedan redak po proizvodu prodan u transakciji prodaje. Jedno od temeljnih načela dimenzijskog modeliranja je to da svi redovi mjerenja u tablici činjenica moraju biti u istom zrnu. Postojanje discipline za izradu tablica činjenica s jednom razinom detalja osigurava da mjerenja nisu neprikladno dvostruko prebrojavana.

Proces učitavanja tablice činjenica uzima transakcije iz izvornog sustava, izračunava činjenice, identificira odgovarajuće vanjske ključeve i dodaje zapise u činjeničnu tablicu. (Adamson, 2010)

Učitavanje redaka u činjeničnu tablicu zahtijeva identifikaciju vrijednosti vanjskih ključeva pretraživanjem u dimenzijskim tablicama. Izvorni sustavi osiguravaju transakcije s prirodnim ključevima; one se moraju prevesti u zamjenske ključeve kako bi se mogla pratiti njihova promjena kroz vrijeme (Adamson, 2010)



Slika 12: Učitavanje činjenične tablice (Izvor: Adamson, 2010)

Činjenična tablica regulirano je sljedećim zahtjevima (Adamson, 2010):

- Izdvajanje podataka iz izvornih sustava
- Izračunajte činjenice
- Zbrojite činjenice koje odgovaraju u tablice činjenica
- Nabavite zamjenske ključeve za svaku dimenziju
- Učitajte zapise tablica činjenica u skladište

Obrada podataka počinje prikupljanjem izvornih podataka, prenošenjem po potrebi, izračunavanjem činjenica, i agregiranjem ako je potrebno.

Tamo gdje postoji više izvora, taj se postupak može ponoviti za svaki izvor. Alternativno, početne ekstrakcije mogu se izvoditi odvojeno, s ekstrahiranim podacima objedinjenima u jedno područje. Zatim se izvršavaju sljedeći zadaci pomoću ovog skupa podataka.

Kada postoji jedan izvor koji je relacijski, ova prva četiri koraka se mogu nositi u jednom upitu. Upit odabire odgovarajuće prirodne ključeve, a zatim agregate i izračunava činjenice. Sljedeći korak će biti dodjeljivanje zamjenskih ključeva.

Identifikacija vrijednosti surogatnih ključeva provodi se kao proces traženja. Dimenzijska tablica se pretražuje kako bi pronašla red koji sadrži vrijednost prirodnog ključa iz izvornih sustava. Kada se pronađe odgovarajući zapis u dimenziji, dodaje mu se zamjenski ključ.

### 4.3. Testiranje ETL-a

Testiranje je bitan proces kod uspostavljanja skladišta podataka. Ponniah (2010) ističe da se testiranje skladišta podataka odnosi na nekoliko općih ciljeva za testiranje:

- Izvlačenje podataka

- Potrebno je osigurati da su svi podaci označeni za vađenje iz različitih izvora potpuno i ispravno izvađeni te da su cjeloviti.
- Transformacija i čišćenje podataka
  - Transformacije podataka moraju biti ispravno izvedene prema poslovnim pravilima za konverzije jer se time uspostavlja kvaliteta podataka.
- Učitavanje podataka
  - Nakon provjere jesu li svi moduli opterećenja ispravni na temelju transformacija i čišćenja podataka, potrebno je osigurati da su svi podaci za tablice dimenzija, činjenica i sažetak ispravno smještene u odgovarajuće datoteke.
- Ispitne staze
  - Testiranje kretanja podataka, i ukupne kontrole tijekom ETL-a, osigurava se da ništa nije izgubljeno ili korumpirano na putu od vađenja podataka, transformacija i konačnog rješenja
- Integracija
- Osigurajte da cijeli ETL proces dobro funkcionira sa svim ostalim uzvodnim i nizvodnim procesima.

## 4.4. Poslovna inteligencija

Poslovna inteligencija bavi se jednostavnijim i bržim donošenjem kvalitetnijih odluka; informacije se pretvaraju u znanje. Prikupljanje, pohranjivanje i analiziranje prikupljenih podataka neophodne su radnje za poslovnu inteligenciju koje se izvršavaju tijekom uspostave skladišta podataka kojim osnovna svrha jeste pružanje potpore alatima poslovne inteligencije. Poslovna inteligencija predstavlja skup procesa, tehnologija i alata čijom se primjenom iz podataka dobivaju informacije ključne za poslovanje nekog poslovnog subjekta, a pojam obuhvaća skladištenje podataka, analitičke i poslovne alate te upravljanje sadržajem. (ZAP FER, 2018).

Alati poslovne inteligencije omogućavaju provođenje (ZAP FER, 2018):

- više dimenzijskih analiza
- upita i izvještavanja
- naprednih vrsta izračunavanja
- dubinskih analiza
- predviđanja



- rudarenja podacima
- i dr.

Ćurko (2001) analitičku obradu predstavlja kao način pronalaženja različitih oblika informacija potrebnih u procesu odlučivanja, a. kao postupke analitičke obrade navodi:

- selekciju,
- istijecanje,
- izdvajanje i kombiniranje svih dimenzija (eng. slice and dice requirement),
- rotaciju - isticanje jedne dimenzije dok su druge u pozadini (eng. pivoting),
- grafičko prikazivanje informacija (eng. charting),
- prognoziranje,
- modeliranje,
- statističke analize poput trenda, klasteriranja i slično,
- "što-ako" funkcije,
- agregaciju i detaljizaciju koja može biti prema gore ili prema dolje (eng. drill up, eng. drill down).

## 5. Korišteni alati

Podatke sam čitala iz Oracle baze pišući upite u Toad alatu. Model podataka sam napravila u alatu VisualPaardigm nakon čega sam proces ETL razvijala u Informatica Power Center alatu.

### 5.1. Toad za Oracle

Toad je skup alata za upravljanje bazama podataka razvijan zadnjih 20 godina radi korištenja od strane programera baza podataka, administratora baze podataka i analitičara podataka koji ga koriste za upravljanje bazama podataka. (Quest, bez dat.)

Omogućuje rad u najkorištenijim bazama podataka (uključujući RDBMS i NoSQL) uz smanjenje vremena za učenje, implementaciju i upravljanje novim i postojećim platformama baza podataka te automatizaciju rada. (Quest, bez dat.)

### 5.2. Informatica

Informatica, kao vodeći svjetski proizvođač u upravljanju podacima u oblaku, nudi bolju uslugu, bržu isporuku, s manje troškova kako bi korisnici bili prilagodljivi promjenama, ostvarili nove mogućnosti rasta i stvorili nove izume. (Informatica, bez dat.)

#### 5.2.1. Informatica PowerCenter

PowerCenter omogućuje integraciju podataka, analitiku i skladištenje podataka, migraciju aplikacija ili konsolidaciju i upravljanje podacima. Ono što ga razlikuje od drugih, to jest zbog čega je najbolji alat u svojoj skupini alata, jest brzo prototipiranje, profiliranje i provjera valjanosti, ponovna uporabljivost, automatizacija i jednostavnost upotrebe, univerzalno povezivanje (to znači besprijekoran pristup i integracija podataka iz svih vrsta izvora, korištenjem visoko učinkovitih, vanjskih konektora), skalabilnost, automatizirano testiranje provjere podataka, nadzor rada i upravljanja (odnosno nadgledanje proizvodnje i pojačanje najboljih praksi kodiranja upozorenjima koja kasnije sprječavaju skupu kontrolu štete), napredna transformacija podataka, podaci u stvarnom vremenu za aplikacije i analitičke podatke i povezivanje s Cloud aplikacijama. (Informatica, bez dat.)

## 6. Implementacija skladišta podataka

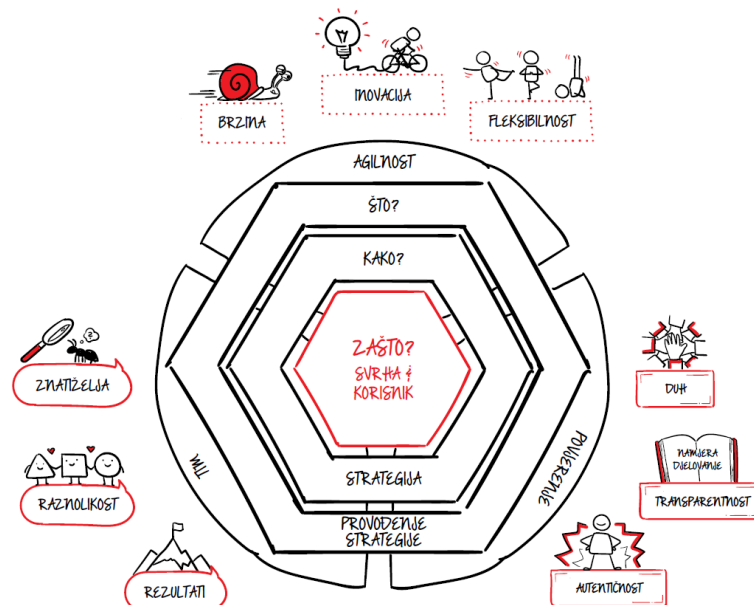
Potrebu za implementacijom skladišta podataka, odnosno njegovog podskupa nazvanog data mart, IPTV sadržaja prepoznala je tvrtka A1 Hrvatska.

### 6.1. Osnovni podaci o djelatnosti tvrtke i poslovanju

A1 Hrvatska d.o.o. dio je A1 Telekom Austria Grupe, u nastavku teksta Grupa. Grupa je vodeći pružatelj digitalnih usluga i komunikacijskih rješenja u centralnoj i istočnoj Europi. A1 Hrvatska ima oko 2000 zaposlenika i oko 2 milijuna korisnika.

A1 (2019) u rubrici za medije navodi da prema P3 mjerenju imaju najbolju mobilnu mrežu, koja pokriva 99% populacije, od čega je 98% pokrivenost 4G mrežom, te najbrži download u fiksnoj mreži, a to je omogućeno uz 13 tisuća kilometara gradske i međugradske optičke mreže te 530 tisuća kućanstva u više od 50 gradova uz super-brzi internet do 500 Mbps u komercijalnoj upotrebi. Posebno su ponosni jer se ističu kao prvi i jedini pružatelj nacionalne pokrivenosti NB IoT mrežom koja omogućava velikom broju uređaja slanje potrebnih informacija s lokacija gdje standardni mobilni signal nema prodornost.

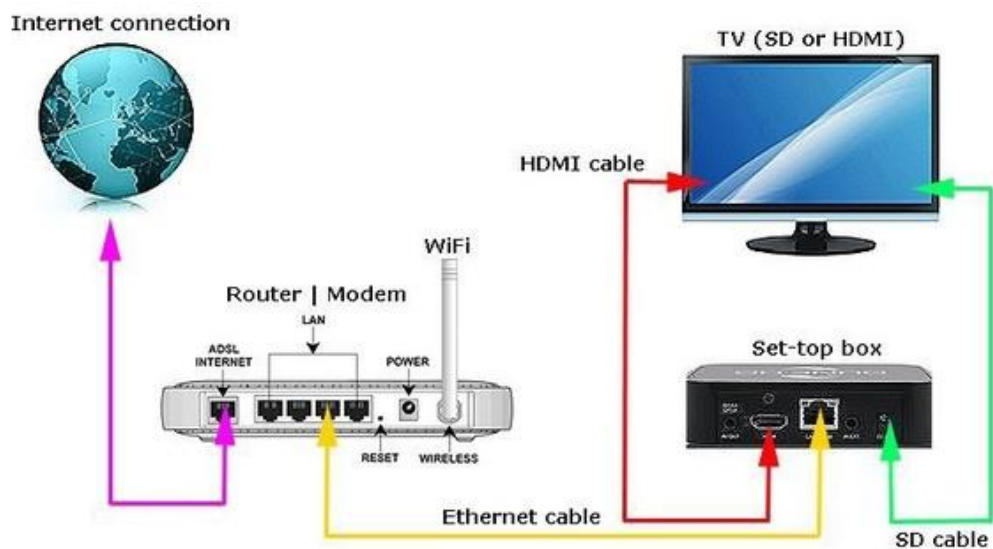
Vizija A1 Telekom Austria Grupe je „Osnaživanje digitalnog života“ dok kao vodeća načela navode „Povjerenje, Tim, Agilnost i Korisničko iskustvo“. (A1, 2019)



Slika 13: Vodeća načela A1 elekoma (Izvor: A1 Telekom Austria Group, 2019)

## 6.2. IPTV

IPTV (eng. Internet Protocol Television) koristi internetski protokol za isporuku medija, a u osnovi je to protokol na kojem je zasnovan i djeluje cijeli internet. Pod isporukom medija misli se na televizijski ili video sadržaj kojeg potrošači gledaju, a do njih sadržaj dolazi jer set-top box, povezan s normalnom širokopojasnom vezom, prima informacije koje omogućuju prikaz tog željenog sadržaja. („Department of Computer Science and Engineering", 2006).



Slika 14: IPTV (Izvor: Quora, bez dat.)

Najveća razlika s nekadašnjom distribucijom televizije je ta što korisnik sam izabire što će gledati iz izbora koji mu je ponuđen, omogućena mu je visokokvalitetna dvosmjerna komunikacija i mogućnost interakcije s davateljem usluga. (Department of Computer Science and Engineering, 2006)

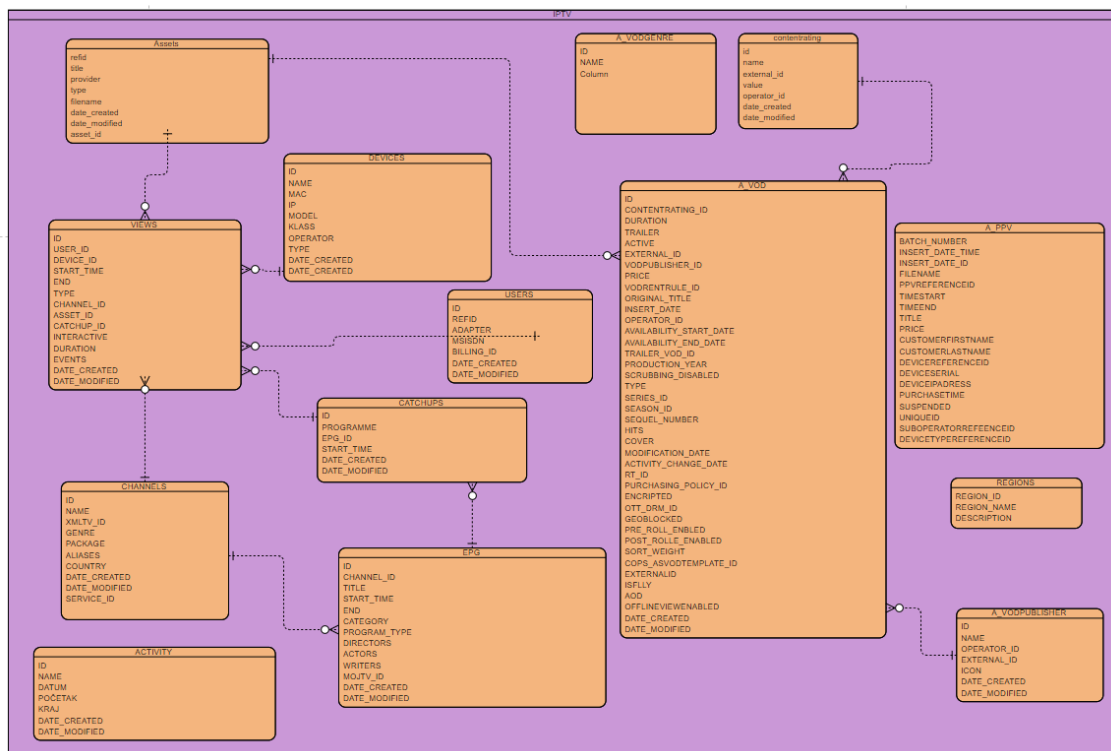
Pojmovi s kojima sam se susrela vezano za ovo područje, a važni su prilikom implementacije su EPG (eng. Electronic Program Guide), odnosno elektronički programski vodič, te PVR (eng. Personal video recorder), odnosno osobni video snimač. Dodatne mogućnosti koje su omogućeno su stanke, premotavanje naprijed i premotavanje natrag tijekom gledanja izabranog sadržaja.

## 6.3. Faze implementacije

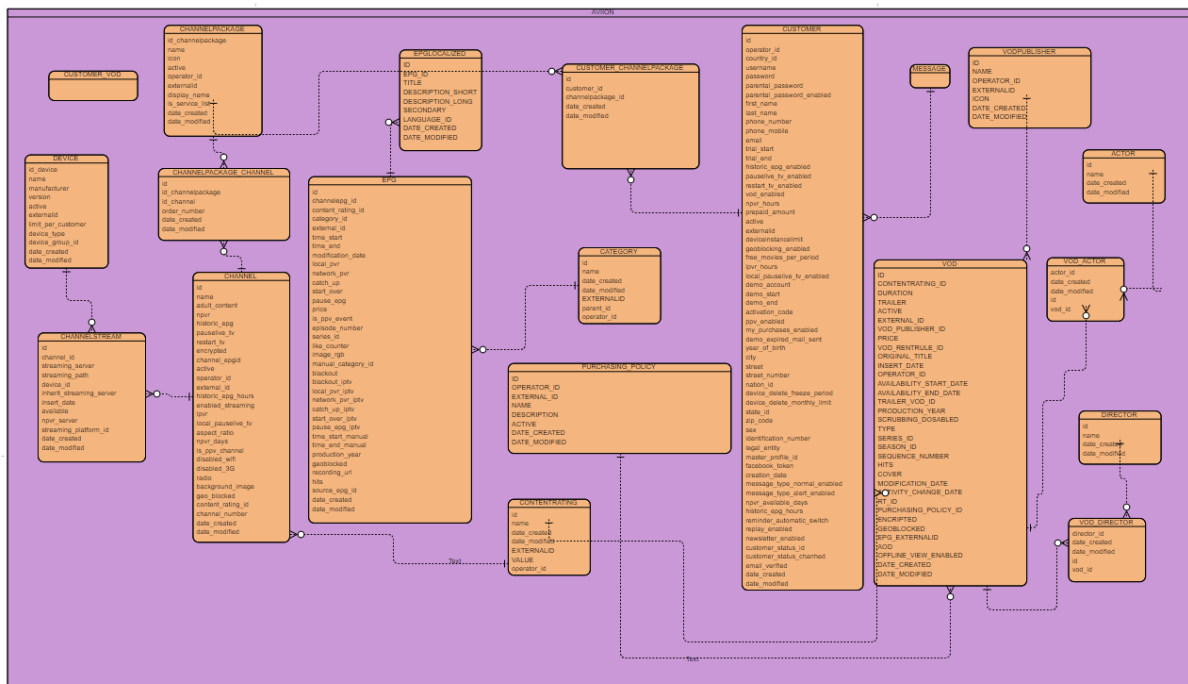
Proces implementacije krenuo je od definiranja informacija koje se žele dobiti kao krajnji rezultat. Nakon definiranja željenih informacija potrebno je pregledati podatke s izvora s kojima se raspolaže na početku, definirati logički model podataka, razmisliti o svemu navedenom u poglavlju o dimenzijskom modeliranju, kreirati potrebne dimenzijske i činjenične tablice u bazi podataka u koju će se spremati podaci te nakon toga prijeći na proces mapiranja zadatka (eng. jobova) koji će dovesti podatke s izvora do odredišta, odnosno do kreiranih dimenzijskih i činjeničnih tablica.

### 6.3.1. Dizajn

Prije izgradnje skladišta podataka potrebno je odrediti izvore podataka, proučiti podatke s izvora i razmisliti o dizajnu skladišta podataka i transformacijama koje će podatke dovesti na prava mjesta u skladište podataka. U implementaciji skladišta podataka za navedeno poduzeće kao osnovni izvori podataka poslužiti će podaci dobiveni s platformi IPTV i AVIION, odnosno kopije podataka spremljene na stage. Podaci se na stage-u čuvaju ovisno o potrebama. Prema dogovoru, podaci se čuvaju mjesec dana. Važno je napomenuti da se razvoj odnosi samo na Hrvatsku. Na slici se vide entiteti, atributi i veze.



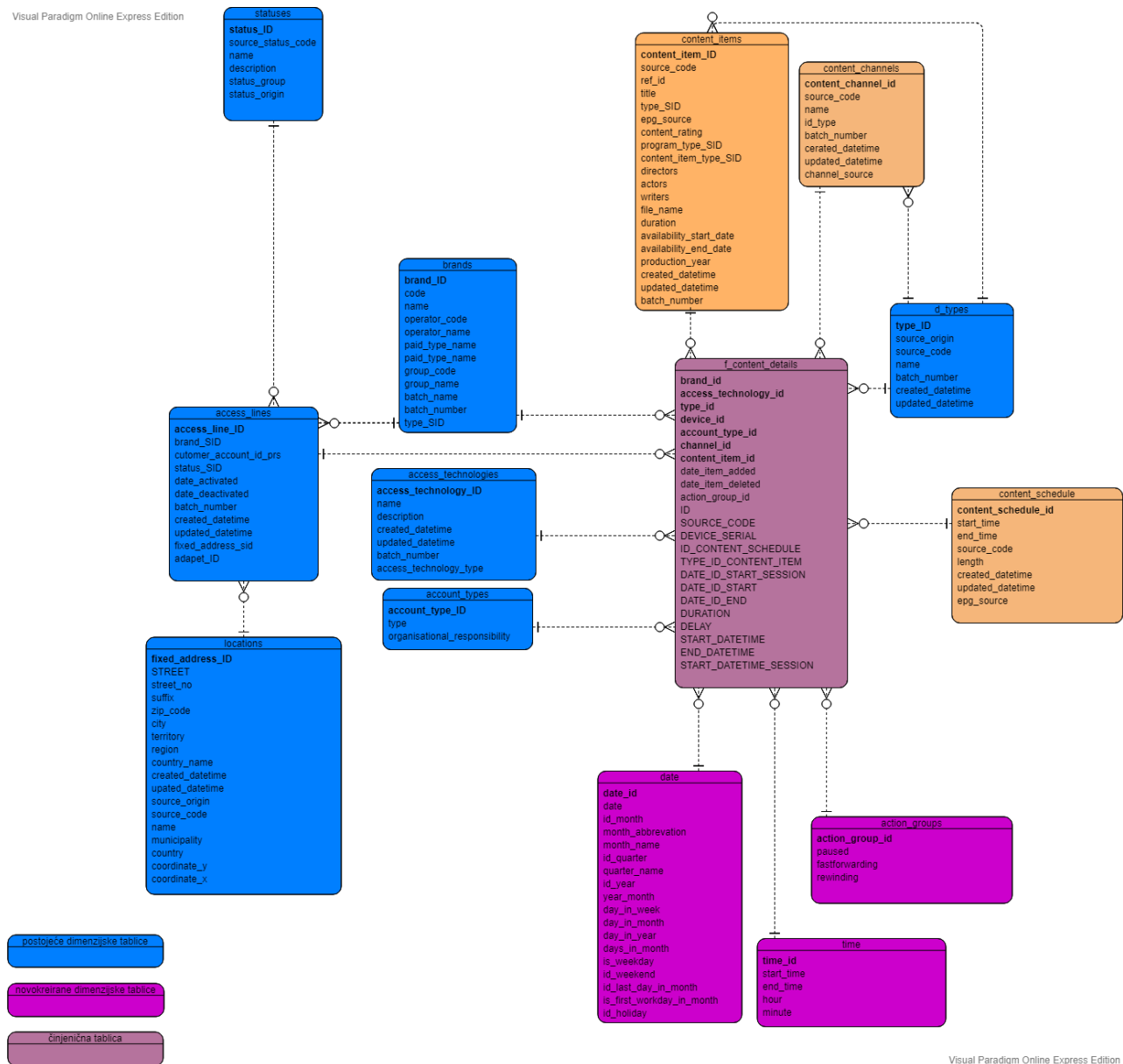
Slika 15: ERA model izvora podataka: IPTV



Slika 16: ERA model izvora podataka: AVIION

Skladište koje će biti implementirano sastojati će se od 13 dimenzijskih tablica i jedne činjenične tablice.

Neke dimenzije postoje od prije u tvrtki A1 Hrvatska te nije potrebno razvijati ih ni dodavati im attribute. Kao što je navedeno u radu, ponekad nema potrebe razvijati od početka nešto što je već razvijeno. Kroz bus arhitekturu moguće je razvijene dimenzije koristiti i kod činjeničnih tablica za koje one nisu bile prvobitno razvijene kao što sam ja napravila za potrebe ovog razvoja. To su entiteti o korisnicima, tipovima korisnika, brandovima i statusima u telekomu te vremenska i datumska dimenzija. U ovom radu će biti prikazan razvoj dimenzije *d\_content\_channel*, koja čuva stavke kanala, dimenzije *d\_content\_item*, u kojoj su stavke sadržaja, i činjenične tablice *f\_content\_details*. Model skladišta podataka, kao i izvora podataka, izrađen je u alatu Visual Paradigm.



Slika 17: Model skladišta podataka

Plavom bojom nanačene su dimenzije koje su implementirane prije ovog projekta. Žarko ružičastom bojom obojene su dimenzije koje se ne učitavaju svaki dan, a to su vremenska dimenzija koja se učitava jednom, datumska dimenzija koja se učitava prije početka kodine i puni se podacima za tu godinu i action\_group dimenzija koja sadrži podatke o akcijskim grupama, a one su definirane u malom broju i učitavaju se iz CSV datoteke. Žutom bojom su označene dimenzije koje sam razvila tijekom ovog projekta i činjenična tablica koja je svijetlo ljubičasta.

Činjenična tablica povezana je sa svim dimenzijskim tablicama preko vanjskih ključeva, a ona sama nema primarni ključ jer nije utvrđena potreba za njezinim primarnim ključem. Dimenzija stavki sadržaja i kanala sastoji se od atributa:

Tablica 2: Tablica Content\_item

Naziv atributa	Tip podatka	Mjesto izvorišta podatka
<b>ID</b>	NUMBER	
<b>SOURCE_CODE</b>	NUMBER	ODS_IPTV
<b>REFID</b>	NUMBER	ODS_IPTV
<b>CHANNEL_ID</b>	NUMBER	
<b>TITLE</b>	VARCHAR2(100)	ODS_IPTV
<b>TYPE_ID</b>	NUMBER	DWH
<b>CONTENT_RATING</b>	VARCHAR2(50)	ODS_IPTV
<b>PROGRAM_TYPE</b>	VARCHAR2(50)	
<b>CONTENT_ITEM_TYPE</b>	VARCHAR2(10)	
<b>DIRECTORS</b>	VARCHAR2(300)	ODS_IPTV
<b>ACTORS</b>	VARCHAR2(200)	ODS_IPTV
<b>WRITERS</b>	VARCHAR2(200)	
<b>FILENAME</b>	VARCHAR2(100)	ODS_IPTV
<b>LENGTH</b>	NUMBER	ODS_IPTV
<b>AVAILABILITY_START_DATE</b>	DATE	ODS_IPTV
<b>AVAILABILITY_END_DATE</b>	DATE	ODS_IPTV
<b>PRODUCTION_YEAR</b>	INTEGER	ODS_IPTV
<b>BATCH_NUMBER</b>	NUMBER	
<b>CREATED_DATETIME</b>	DATETIME	
<b>UPDATED_DATETIME</b>	DATETIME	

Tablica 3: Tablica Content\_channel

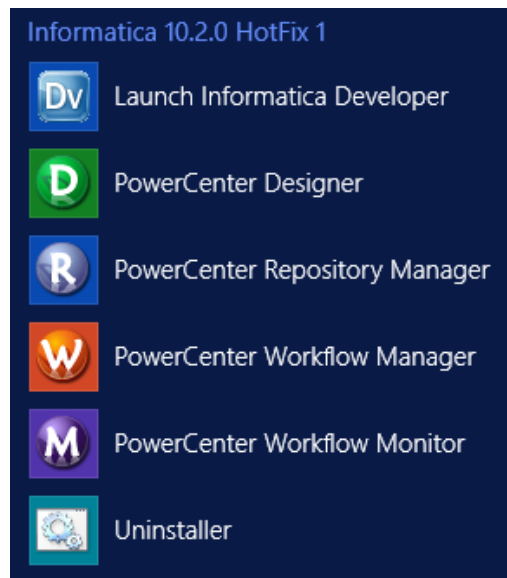
Naziv atributa	Tip podatka	Mjesto izvorišta podatka
<b>ID</b>	NUMBER	
<b>SOURCE_CODE</b>	NUMBER	ODS_IPTV
<b>NAME</b>	VARCHAR2(64)	ODS_IPTV
<b>TYPE_ID</b>	NUMBER	DWH
<b>XMLTV</b>	NUMBER	ODS_IPTV
<b>CONTENT_CHANNEL_PACKAGE_ID</b>	NUMBER	ODS_IPTV
<b>BATCH_NUMBER</b>	NUMBER	
<b>CREATED_DATETIME</b>	DATE	
<b>UPDATED_DATETIME</b>	DATE	



Nakon što je utvrđeno sve potrebno, kreirane su skripte koje kreiraju postojeće tablice i prenose njihove podatke u shemu područja za razvoj (eng development, skraćeno dev) te su u istoj shemi kreirane tablice koje ću puniti preko mapiranja u Informatici. Kada mapiranja budu potpuno točna, biti će prebačena u razvojni dio produkcije gdje će biti, pomoću vanjske datoteke koja sadrži sve potrebne parametre, pokretana svaku noć kako bi se tablice punile novim podacima.

### 6.3.2.Razvoj

Cjelokupni proces integracije podataka biti će proveden u prethodno predstavljenom alatu Informatica PowerCenter. Informatica se sastoji od dijelova navedenih na slici, a za potrebe ovog razvoja bili su korišteni: Power Center Designer, Power Center Repository Workflow Manager i PowerCenter Workflow Monitor.



Slika 18: Dijelovi Informatice

PowerCenter Designer sadrži foldere u kojima se spremaju Izvori, Targeti, Dimenzije, Transformacije, Mapleti i Mapiranja.

Glavni dio se obavlja u mapiranju, a prije njega je potrebno dodati sve potrebne tablice izvora u source dio kako bi mogli krenuti s mapiranjem, odnosno kako bi Informatica mogla generirati upit koji će se izvršiti nakon svih transformacija i spremiti u target tablicu koju smo dodali u mapiranje, nakon što smo je dodali u target dio. Jedna tablica može biti ciljana u jednom mapiranju i izvorišna u drugom, ali mora postojati u dijelu iz kojega ga tehnikom drag and drop-a unosimo u mapiranje.



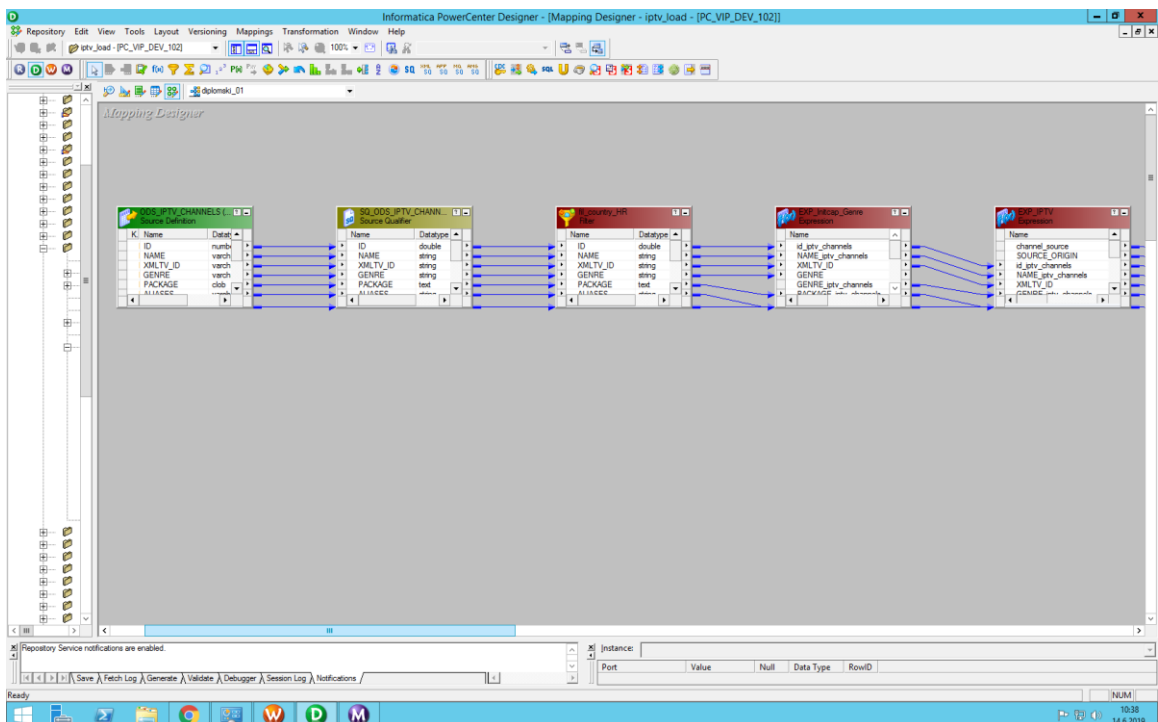
Slika 19: Pogledi u alatu Informatica

20. Mapiranje omogućava transformacije. Korištene transformacije prikazane su na slici

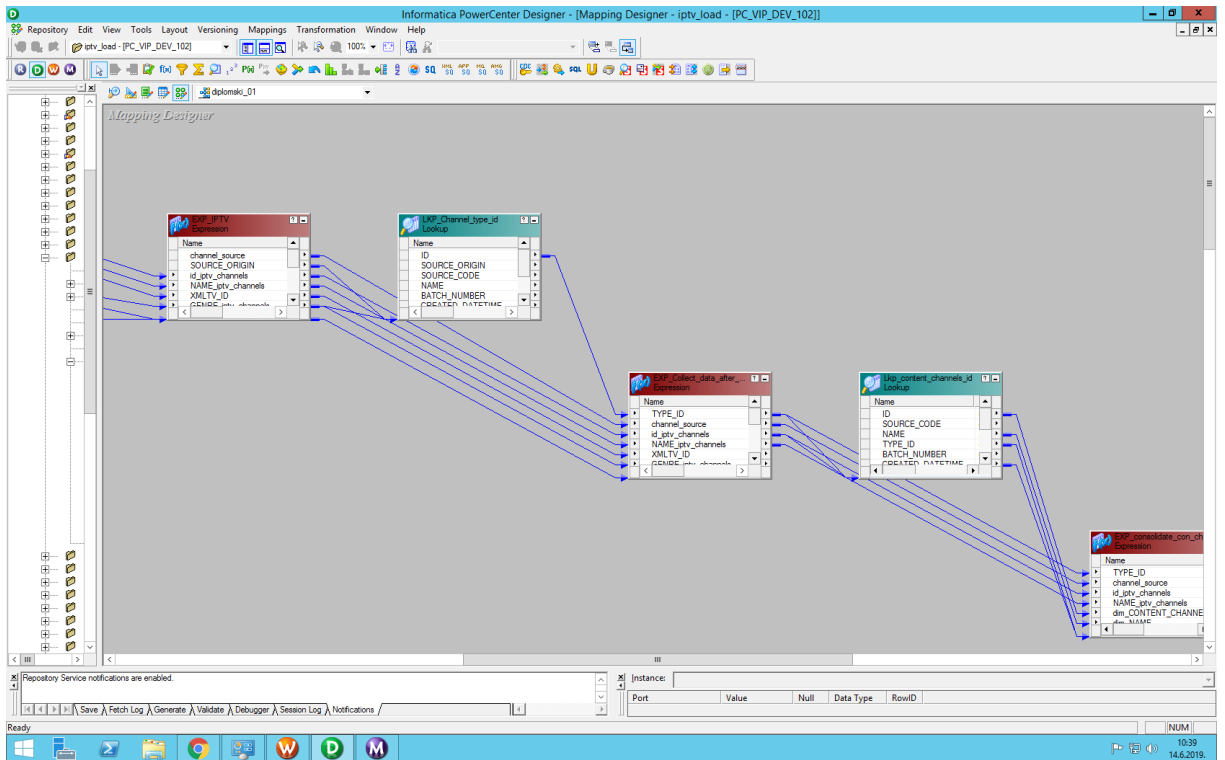


Slika 20: Mogući elementi transformacija u Informatici

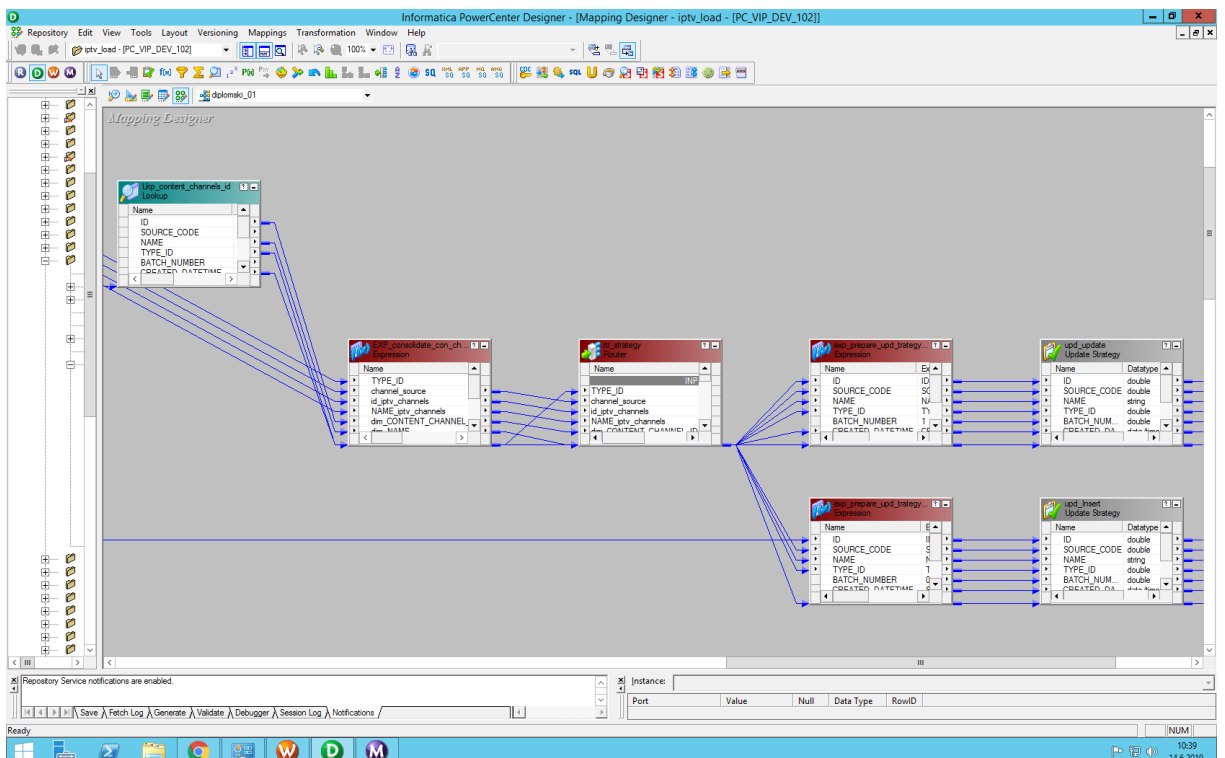
Mapiranje dimenzijske tablice sadržaja kanala prikazano je u nastavku. Nakon dodavanja izvora iz kojeg uzimamo podatke, filtriramo samo one koji su iz Hrvatske te se dodaju podaci o vrsti i tipu. Nakon što imamo te podatke komponentom „LOOKUP“ dobivamo jedinstveni identifikator tipa kanala na tablici kanala i jedinstveni identifikator sadržaja kanala. Router komponentom ispitujemo da li je podatak nov i trebamo mu dodati pomoću komponente „sequence“ jedinstveni identifikator ili ga je potrebno ažurirati (eng. Update). Na kraju je potrebno podatak pohraniti u dimenzijsku tablicu sadržaja kanala.



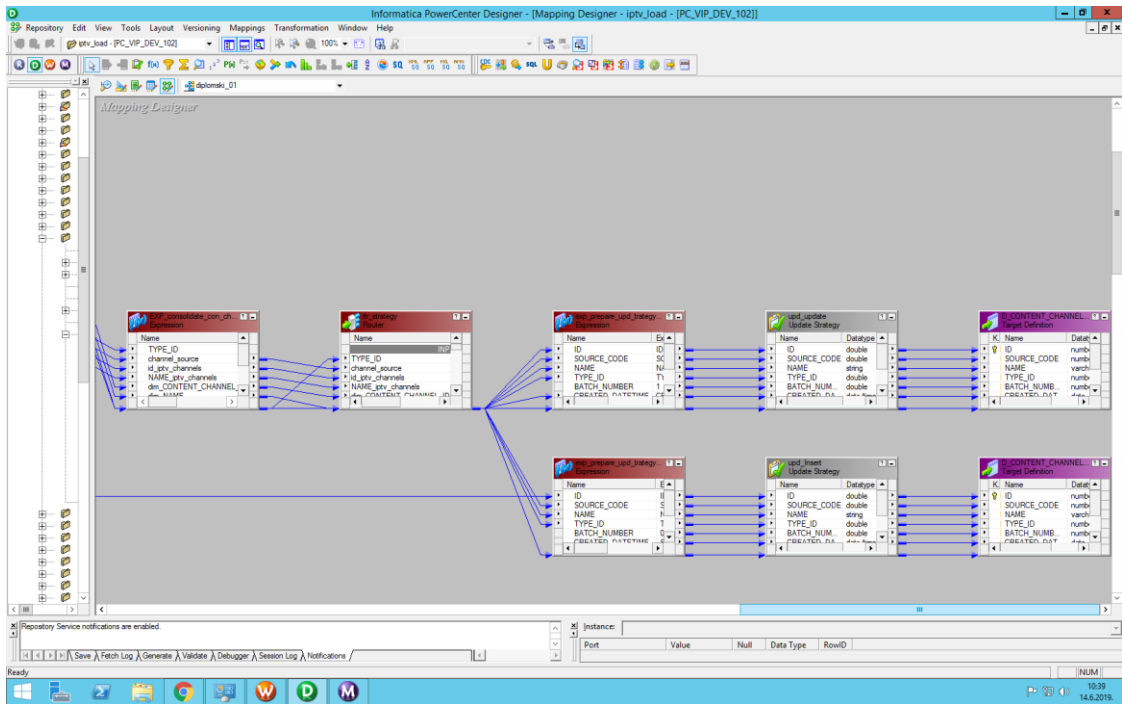
Slika 21: Mapiranje kanala od izvora do filtriranja



Slika 22: Mapiranje kanala uz prikupljanje ključeva pomoću "LOOKUP"-ova

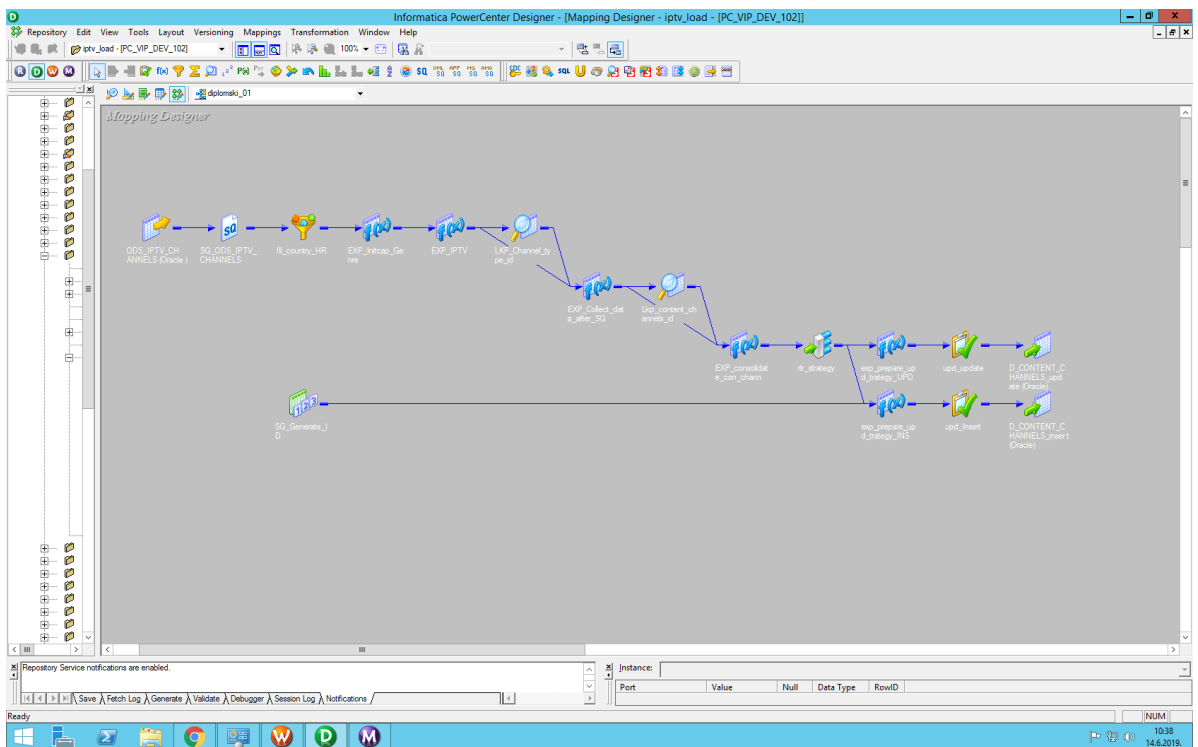


Slika 23: Mapiranje kanala do elementa odluke

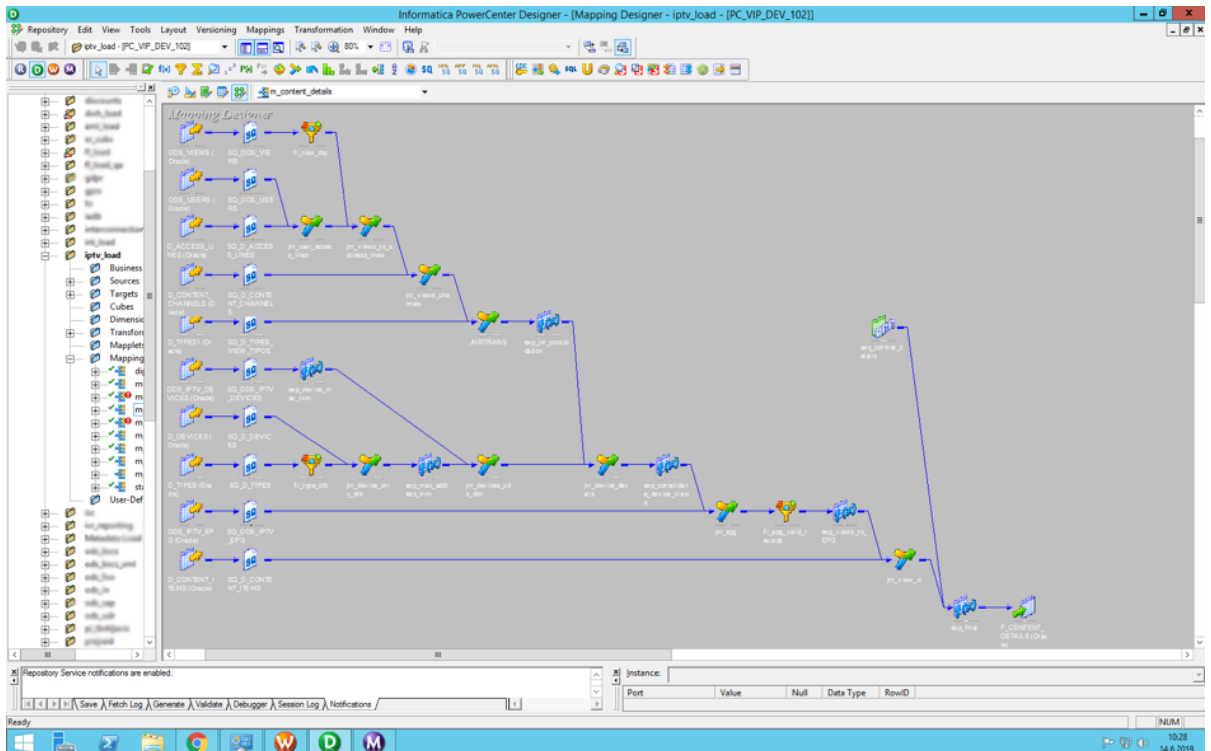


Slika 24: Mapiranje podatka do odredišta

Informatica nudi i dizajnerski prikaz koji je pregledniji. Mapiranje dimenzije je prikazano na slici ispod. Mapiranje činjenične tablice prikazano je na slici 26.



Slika 25: Mapiranje kanala

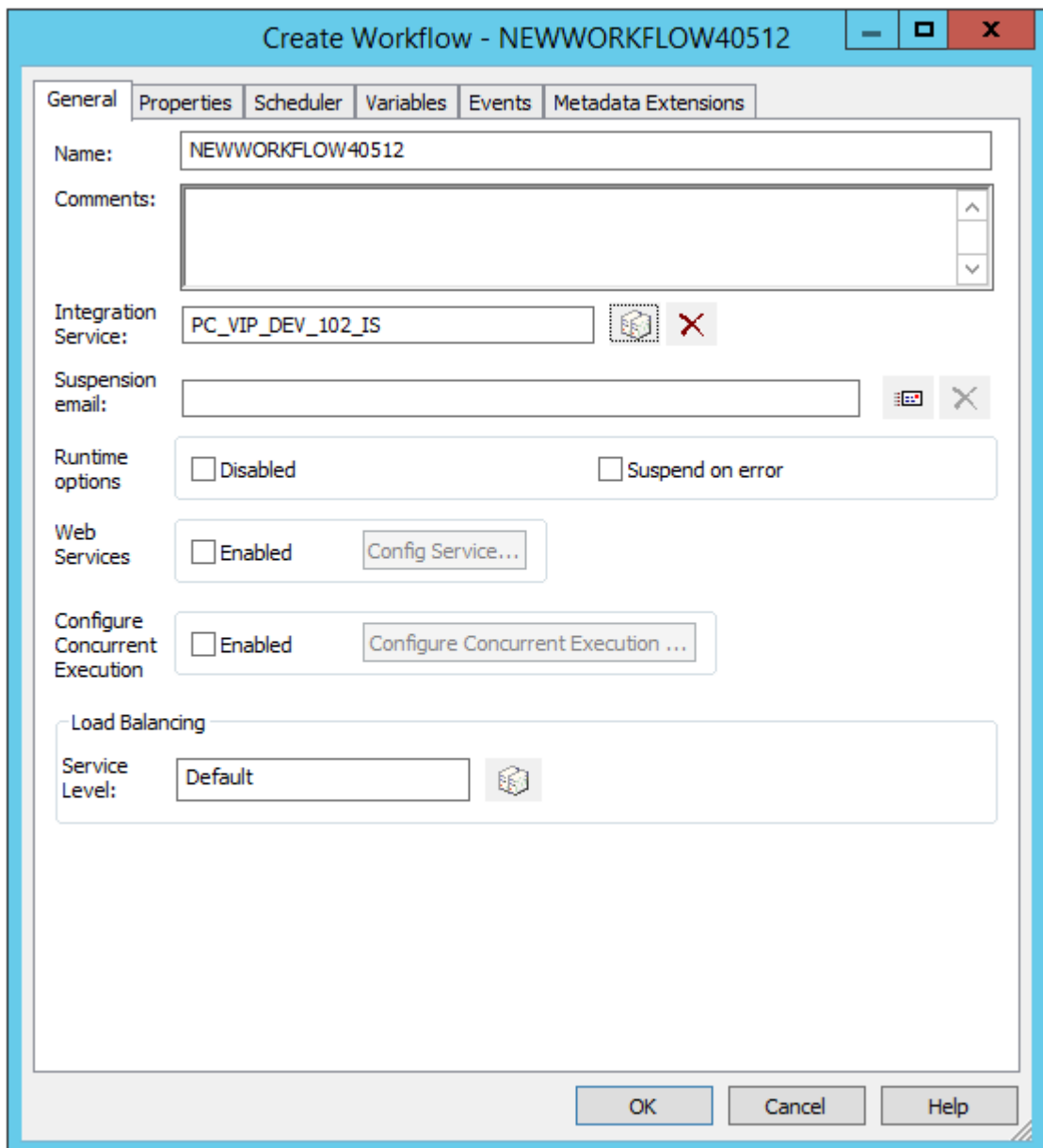


Slika 26: Mapiranje činjenične tablice

Na slici se vidi da postoji 10 izvora koji su potrebni za činjeničnu tablicu, a neki od njih su i same dimenzije. Za spajanje su korišteni „JOIN“ elementi. Filteri su ovdje u službi filtriranja podataka samo za Hrvatsku.

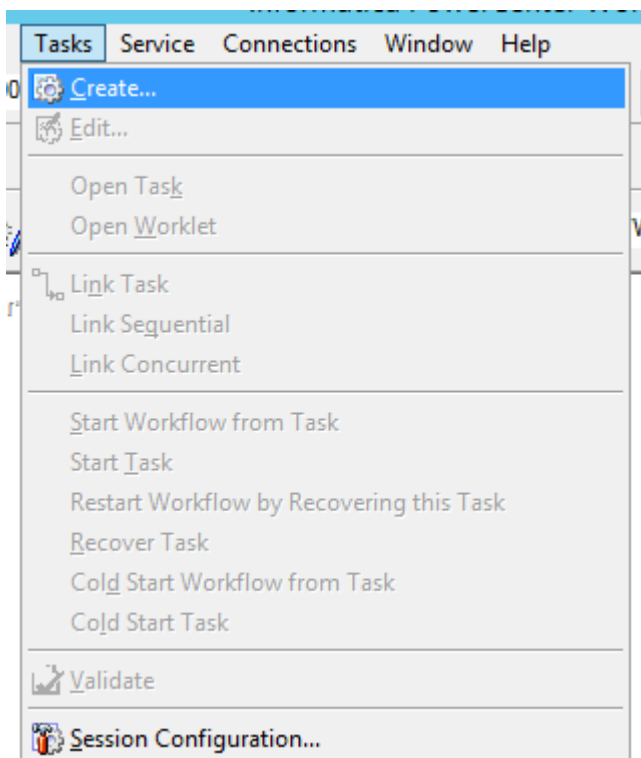
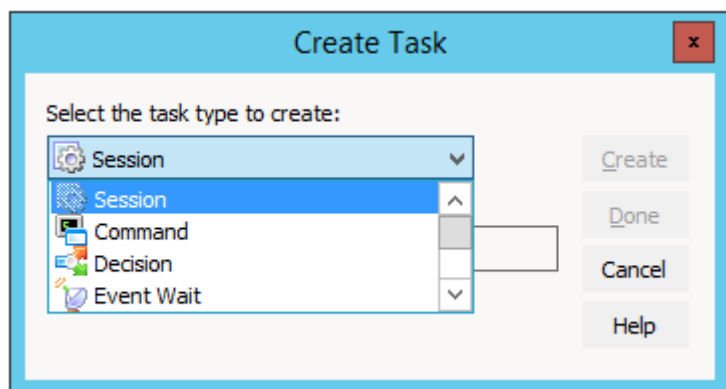
Kada su sva mapiranja tablica gotova moguće je odrediti kada će se pokretati mapiranja i u kojem redoslijedu.

Workflow Manager omogućuje kreiranje Workflowa. Ono što se definira tijekom kreiranja prikazano je na slici 27.



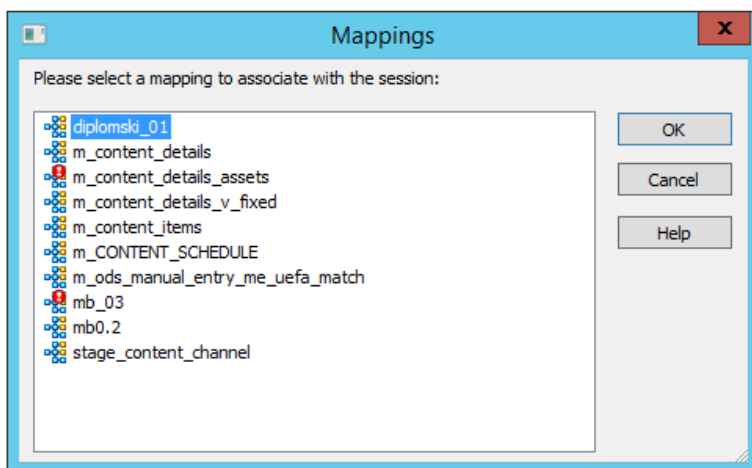
Slika 27: Kreiranje Workflow-a

Nakon što imamo workflow, možemo kreirati zadatak (eng Task) za kojeg smo odabrali da će biti Session kako bi ispunili sve planirano.



Slika 28: Kreiranje zadatka

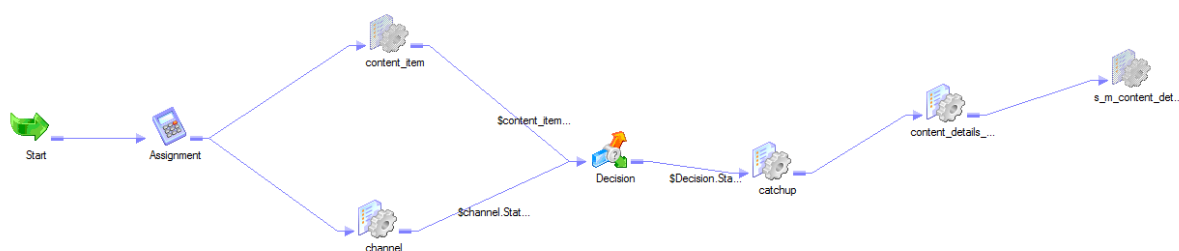
Nakon što smo kreirali potrebna mapiranja, u dijelu WORKFLOW MANAGER-a nazvanog Workflow Designer možemo dodati odabrana mapiranja i, uz još neke elemente koje on nudi, definirati cijeli proces s kojim će se popuniti sve dimenzije i to u dopuštenom slijedu.



Slika 29: Dodavanje

komponente u Workflow Manager

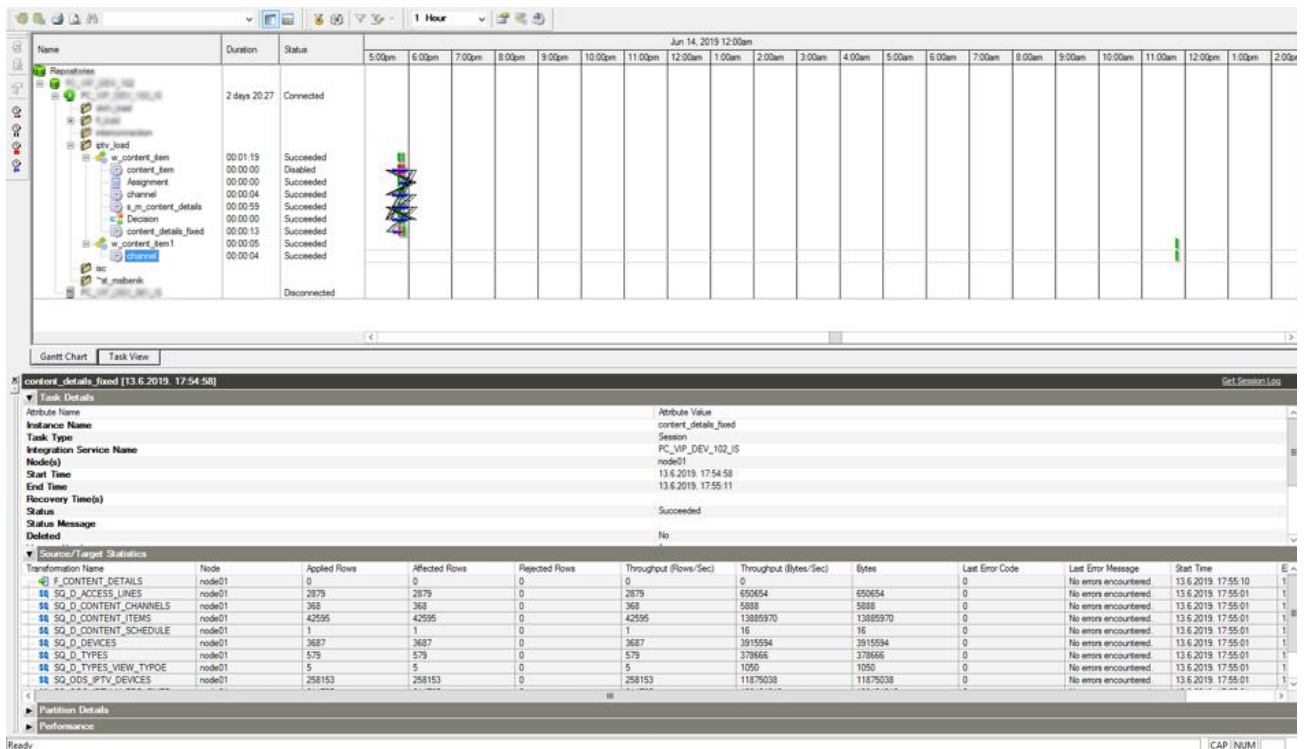
Pod dopuštenim slijedom se misli da je prvo potrebno napuniti dimenzije te zatim činjeničnu tablicu koja se sastoji od podataka iz tih dimenzija. Ukoliko bi činjeničnu tablicu punili prije dimenzija, mogli bi naići na probleme ne pronalaska ključa što vodi do spremanja nepoznatog ključa (ključa definiranog za spremanje kada vidimo da se radi o podatku, ali ga nemamo spremljenog u odabranoj dimenzijskoj tablici iz koje dolazi) ili do neke druge dodatne komplikacije. Do ne pronalaska ključa iz dimenzijske tablice može doći jer ključ ne mora biti prethodno pohranjen u dimenziji ukoliko nije postojao taj podatak prije punjenja ili je nadodan iz razloga što je došlo do promjene, a definiran je SCD veći od 1.



Slika 30: Kompletan proces učitavanja dimenzija i činjenične tablice

Statusi o uspješnosti izvršavanja prikazani su na slici 31. Prikazano nam je koliko je zapisa bilo na izvoru, koliko zapisa je dodano u dimenziju, jesu li obavljene sve potrebne transformacije, odnosno da li je cjelokupni proces završio uspješno.

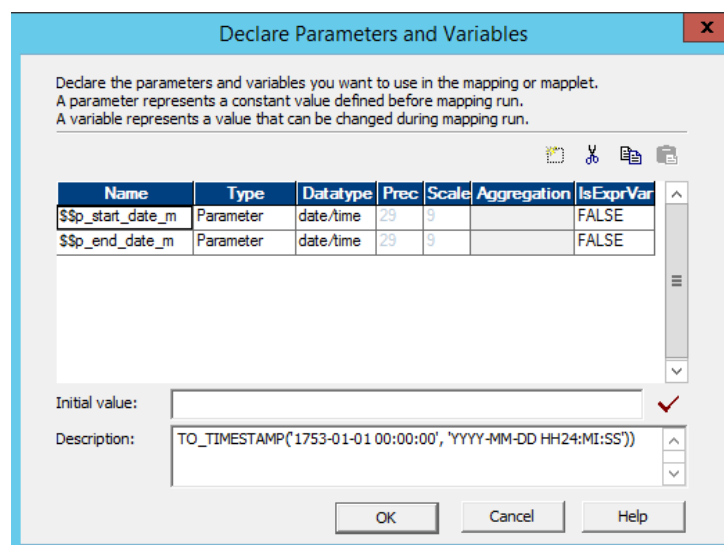




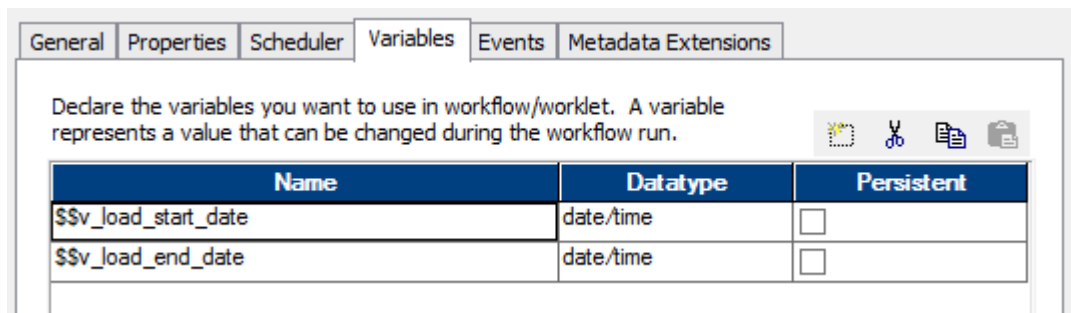
Slika 31: Pregled učitavanja

Rezultat cijelog procesa je prikazan na slici 31 gdje se vidi da nakon početka obrađuje komponenta Assigment u kojoj su pomoću varijabli definirani datumi koji će biti obrađivani.

Varijable i parametri su prethodno definirani nad workflow-om, a to prikazuju slike 32 i 33.



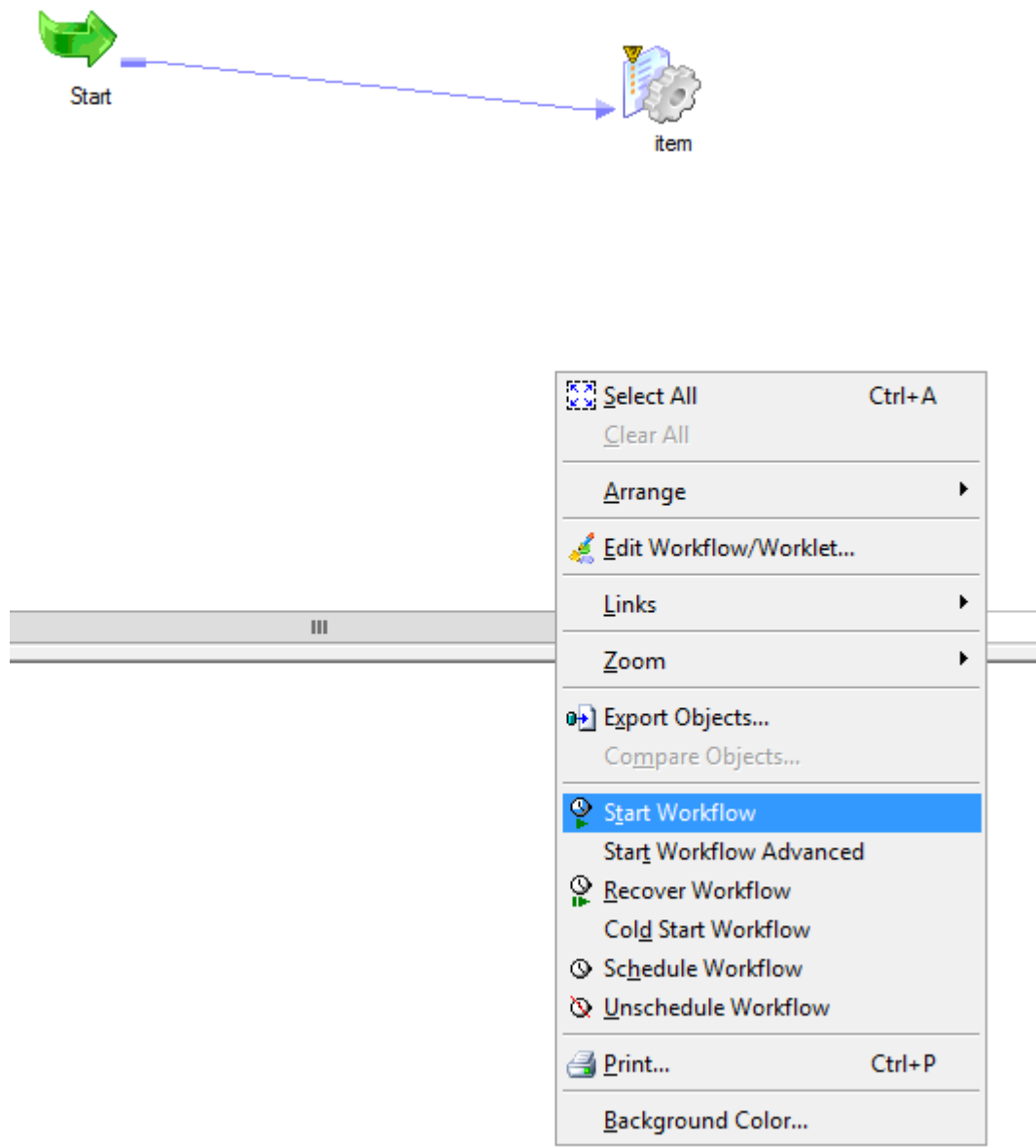
Slika 32: Parametri



Slika 33: Varijable

Kada se zna vrijeme koje će se promatrati, kreće kreiranje stavki sadržaja i stavki kanala u isto vrijeme te ukoliko ne dođe do nikakvog problema, odnosno ukoliko oba mapiranja završe sa statusom Ok, dolazi do mapiranja činjenične tablice koju sam razdvojila na tri mapiranja zbog toga što podaci dolaze s dvije platforme i sama mapiranja su izuzetno kompleksna.

WORKFLOW MANAGER omogućuje samostalno pokretanje programa u određeno vrijeme (eng. scheduling), ali omogućeno je i ručno pokretanje koje se koristi kod testiranja odnosno razvijanja i ukoliko dođe do pogreške koja obustavlja punjenje tablica. Ukoliko se tablice ne napune, ljudi koji su razvili ETL proces moraju osigurati izvještavanje o neuspjehu kako bi greška mogla biti ispravljena i podaci nadodani.



Slika 34: Pokretanje Workflow-a

## 7. Zaključak

„Podaci su nafta 21. stoljeća“ (eng. “Data is the new oil”) poznata je izreka koja opisuje stanje u kojemu onaj tko ima pravu informaciju u pravo vrijeme može ostvariti uspjeh. Podatak sam po sebi nema „težinu“ dok ga se ne stavi u kontekst s kojim postaje informacija. Kako bi došli do informacije potrebno je prikupljati podatke i provoditi transformacije koje će nas dovesti do zaključivanja. Podaci mogu biti zapisani na papiru, u bazi podataka, u mislima osobe (iako na taj način često bivaju izgubljene), a mogu biti pohranjena i u skladišta podataka kako bi se iz njih dobila najveća moguća vrijednost.

Skladište podataka se razvija procesom ekstrahiranja, transformiranja i učitavanja. Često se puni na dnevnoj razini, a to je omogućeno kroz samostalno pokretanje sustava preko datoteka ili na neki drugi način. Izvori podataka mogu biti u različitim aplikacijama, sustavima i slično, ali odredište podataka, odnosno dimenzijske i činjenične tablice, mora imati definiranu strukturu i pravila koja se poštuju kod spremanja podataka. Ukoliko se proces pravilno provede do informacija je puno jednostavnije i brže doći. Najčešći korisnici su menadžeri iako sve više ljudi otkriva pomoć koji im pruža razvitak skladišta podataka.

Podskup skladišta podataka čini data mart koji je implementiran u tvrtki A1 Hrvatska za potrebe ovog rada. Prilikom implementacije korištene su neke prethodno definirane dimenzije koje su popunjene potrebnim podacima i za koje nije bilo potrebno razvijati mapiranja, odnosno proces u kojemu će biti provedeno spremanje podataka s izvora u odredište, dok sam za ostale dimenzije i činjeničnu tablicu napravila mapiranja.

Leksikografski zavod „Miroslav Krleža“ u svojoj definiciji informacije navodi da „ključnu ulogu u pretvorbi podataka u informaciju ima znanje kojim čovjek raspolaže.“ Skladište podataka, kao i data mart, omogućuje ispostavu informacija osobi koja zna čitati te informacije bez da treba provesti neki proces transformacije podataka jer su te transformacije učinjene kroz proces razvitka skladišta.

Za skladište podataka kažu da vrijedi onoliko koliko se koristi.

## 8. Literatura

1. A1 (2019) Preuzeto 08.06.2019 s [https://www.a1.hr/tko-smo-mi/objave-za-medije/-/objave/clanak/20-god/851740433?\\_pressreleases\\_WAR\\_press\\_returnUrl=%2Ftko-smo-mi%2Fobjave-za-medije](https://www.a1.hr/tko-smo-mi/objave-za-medije/-/objave/clanak/20-god/851740433?_pressreleases_WAR_press_returnUrl=%2Ftko-smo-mi%2Fobjave-za-medije)
2. A1: Kodeks ponašanja (2019) Preuzeto 08.06.2019. s [https://ssc.a1.hr/documents/10307706/287534663/A1+Group+Code+of+Conduct+2017+Croatia+v2\\_IB\\_28.11.2017+Arnoldner+%28002%29.pdf/dc586ae7-8875-91da-f8ff-c7d2b8262070](https://ssc.a1.hr/documents/10307706/287534663/A1+Group+Code+of+Conduct+2017+Croatia+v2_IB_28.11.2017+Arnoldner+%28002%29.pdf/dc586ae7-8875-91da-f8ff-c7d2b8262070)
3. A1 Telekom Austria Group: Guiding Principles (2019) Preuzeto 09.09.2019. s <https://www.a1.group/de/group/guiding-principles>
4. Adamson, C. (2010). Star Schema. Unated States: The McGraw-Hill Companies
5. Behin, S., i Parlov M. (2005). Active Data Warehouse. Dostupno 3.6.2019. na <http://www.skladistenje.com/active-data-warehouse/>
6. Ćurko, K. (2001).Skladište podataka - sustav za potporu odlučivanju (Izvorni znanstveni rad. Dostupno 08.06.2019. na <https://hrcak.srce.hr/file/45126>
7. Department of Computer Science and Engineering (2006) IPTV the future of television? Dostupno 15.06.2019. na <http://www.cse.chalmers.se/~tsigas/Courses/DCDSeminar/Files/IPTVrapport.pdf>
8. Informatica (bez dat.) Dostupno 13.6.2019. na <https://www.informatica.com/about-us.html#fbid=tF6USbp3RJs>.
9. Informaica (bez dat.) Preuzeto 8.6.2019. s <https://www.informatica.com/products/data-integration/powercenter.html>
10. Kimball, R., Caserta, J. (2004). The Data Warehouse Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Indianapolis: Wiley.
11. Kimball, R., Ross, M. (2002). The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling
12. Kimball, R., Ross, M. (2013). The Data Warehouse Toolkit, Third Edition. Canada: Wiley.
13. Leksikografski zavod „Miroslav Krleža“ (bez dat.) informacija Dostupno 29.08.2019. na <http://www.enciklopedija.hr/natuknica.aspx?id=27405>
14. Maleković, M., i Rabuzin, K. (2016). Uvod u baze podataka. Varaždin: Fakultet organizacije i informatike.
15. Oracle (bez dat.) Preuzeto 10.08.2019. s [https://www.oracle.com/webfolder/technetwork/tutorials/obe/db/10g/r2/owb/owb10gr2\\_gs/owb/lesson3/slowlychangingdimensions.htm](https://www.oracle.com/webfolder/technetwork/tutorials/obe/db/10g/r2/owb/owb10gr2_gs/owb/lesson3/slowlychangingdimensions.htm)

16. Oreščanin, D. (2004). Skladno skladištenje. Dostupno 3.6.2019. na <http://www.skladistenje.com/skladno-skladistenje/>.
17. Oreščanin, D. (2004). Skladno skladištenje. Dostupno 19.7.2019. na <http://www.skladistenje.com/integracija-podataka-i-aplikacija/>
18. Oreščanin, D. (2011). Osiguranje kvalitete podataka u skladištima podataka (Magistarski rad). Fakultet elektrotehnike i računarstva, Sveučilište u Zagrebu.
19. Ponniah, P. (2010). Data warehousing: Fundamentals for IT professionals (Second Edition). Hoboken: John Wiley & Sons.
20. Rabuzin, K.(2011).Predavanje 7. Skladišta podataka i poslovna inteligencija [Moodle]. Sveučilište u Zagrebu, Fakultet organizacije i informatike, Varaždin.
21. Rabuzin, K. (2011). Uvod u SQL.Varaždin: Fakultet organizacija i informatike.
22. Rabuzin, K. (2014). SQL –Napredne teme.Varaždin: Fakultet organizacije i informatike.
23. Stitchdata (bez dat.) Preuzeto 10.08.2019. s <https://www.stitchdata.com/etl/database/etl-transform/>
24. Tech Differences (2017). Preuzeto 08.06.2019. s <https://techdifferences.com/difference-between-data-warehouse-and-data-mart.html#Definition>
25. Varga M., Čerić V., Birolla H. (1998) Poslovno računarstvo, Znak
26. Wood, K. P. (2011) Preuzeto 10.08.2019. s <https://social.technet.microsoft.com/Forums/en-US/0c946b78-fd1f-4207-9f7b-c5eb0081f509/what-is-business-key?forum=sqldatawarehousing>
27. QuestPond (2017) Preuzeto 08.06.2019. s <http://www.learnmsbitutorials.net/ssis-dimension-fact-table-snowflake-star-schema-part1.php>
28. Quest (bez dat.) Dostupno 13.6.2019. na <https://www.quest.com/toad/>
29. ZPR FER. (2018). 1. Uvod. Poslovna inteligencija [Nastavni materijali]. Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva, Zagreb.

# Popis slika

Slika 1: Ralph Kimballova arhitektura skladišta podataka (Izvor: Adamson, 2010).....	3
Slika 2: Model zvijezde (Izvor: QuestPond, 2017).....	7
Slika 3: Model pahuljice (Izvor: QuestPond, 2017) .....	8
Slika 4: Centraliziran data mart(Izvor: Ponniah, 2010) .....	10
Slika 5: Neovisan data mart (Izvor: Ponniah, 2010) .....	10
Slika 6: Savezni data mart-ovi (Izvor: Ponniah, 2010).....	10
Slika 7: "Hub-and-Spoke" data mart (Izvor: Ponniah, 2010).....	11
Slika 8: Bus arhitektura(Izvor: Ponniah, 2010) .....	11
Slika 9 Inmonova arhitektura(Izvor: Adamson, 2010).....	16
Slika 10: Arhitektura Ralph Kimballa (Izvor: Adamson, 2010) .....	17
Slika 11: Učitavanje dimenzijske tablice (Izvor: Adamson, 2010).....	22
Slika 12: Učitavanje činjenične tablice (Izvor: Adamson, 2010) .....	24
Slika 13: Vodeća načela A1 elekoma (Izvor: A1 Telekom Austria Group, 2019) .....	28
Slika 14: IPTV (Izvor: Quora,bez dat.) .....	29
Slika 15: ERA model izvora podataka: IPTV.....	30
Slika 16: ERA model izvora podataka: AVIION.....	31
Slika 17: Model skladišta podataka.....	32
Slika 18: Dijelovi Informatice.....	34
Slika 19: Pogledi u alatu Informatica.....	35
Slika 20: Mogući elementi transformacija u Informatici .....	35
Slika 21: Mapiranje kanala od izvora do filtriranja .....	35
Slika 22: Mapiranje kanala uz prikupljanje ključeva pomoću "LOOKUP"-ova.....	36
Slika 23: Mapiranje kanala do elementa odluke.....	36
Slika 24: Mapiranje podatka do odredišta .....	37
Slika 25: Mapiranje kanala .....	37
Slika 26: Mapiranje činjenične tablice .....	38
Slika 27: Kreiranje Workflow-a.....	39
Slika 28: Kreiranje zadatka.....	40
Slika 29: Dodavanje komponente u Workflow Manager .....	41
Slika 30: Kompletan proces učitavanja dimenzija i činjenične tablice.....	41
Slika 31: Pregled učitavanja .....	42
Slika 32: Parametri .....	42
Slika 33: Varijable.....	43
Slika 34: Pokretanje Workflow-a.....	44

## Popis tablica

Tablica 1: Razlike skladišta podataka i data mart-a .....	9
Tablica 3: Tablica Content_item .....	33
Tablica 4: Tablica Content_channel.....	33