

FAKULTETA ZA INFORMACIJSKE ŠTUDIJE
V NOVEM MESTU

MAGISTRSKA NALOGA

ŠTUDIJSKEGA PROGRAMA DRUGE STOPNJE

NATAŠA KLENOVŠEK

FAKULTETA ZA INFORMACIJSKE ŠTUDIJE
V NOVEM MESTU

MAGISTRSKA NALOGA

VEČPARAMETRSKA PRIMERJALNA ANALIZA
IMPLEMENTACIJ PODATKOVNIH SKLADIŠČ ZA
POTREBE MIGRACIJE PODATKOVNE BAZE V MS
ACCESSU V PODATKOVNO SKLADIŠČE

Mentor: izr. prof. dr. Blaž Rodič

Novo mesto, september 2014

Nataša Klenovšek

IZJAVA O AVTORSTVU

Podpisana Nataša Klenovšek, študentka FIŠ Novo mesto, izjavljam:

- da sem magistrsko nalogo pripravljala samostojno na podlagi virov, ki so navedeni v magistrski nalogi,
- da dovoljujem objavo magistrske naloge v polnem tekstu, v prostem dostopu, na spletni strani FIŠ oz. v elektronski knjižnici FIŠ,
- da je magistrska naloga, ki sem jo oddala v elektronski obliki identična tiskani verziji,
- da je magistrska naloga lektorirana.

V Novem mestu, dne _____ Podpis avtorice _____

POVZETEK

V uvodnem delu raziskave smo predstavili in opredelili pojem podatkovnega skladišča ter preučili šest podatkovnih baz, ki se najpogosteje uporabljajo za namene podatkovnega skladiščenja: MySQL, PostgreSQL, MongoDB, MS SQL, Oracle in DB2. Izvorno podatkovno bazo v MS Accessu smo migrirali v podatkovna skladišča v prej naštetih bazah. Preučili smo kriterije, ki vplivajo na izbor podatkovnega skladišča za manjše podjetje ter na podlagi tega izoblikovali drevo kriterijev. S pomočjo sistema za podporo odločanju DEXi smo naredili večparametrsko primerjalno analizo obravnavnih podatkovnih skladišč, ter opravili dodatne analize v orodju VREDANA. Analizo najbolje ocenjenega podatkovnega skladišča smo razdelali še s pomočjo SWOT analize. Ugotovili smo, da MySQL najbolje zadovoljuje postavljene kriterije, saj je prejel najvišjo oceno. Sledijo mu v vrstnem redu po oceni PostgreSQL, DB2, MongoDB, MS SQL in Oracle.

KLJUČNE BESEDE: podatkovno skladišče, večparametrska odločitvena analiza, MySQL, PostgreSQL, MongoDB, MS SQL, Oracle, DB2

ABSTRACT

In the introductory part of the research we have presented and defined the concept of a data warehouse and examined the six databases which are most frequently used for the purpose of data warehousing: MySQL, PostgreSQL, MongoDB, MS SQL, Oracle and DB2. We have migrated a source database in MS Access into data warehouses using databases mentioned above. We have examined the criteria that influence the selection of a data warehouse for a small company, and based on that formed a criteria tree. With the decision support system DEXi, we have performed a multi-attribute comparative analysis of data warehouses and an additional analysis with the VREDANA tool. We have performed a SWOT analysis for the best scored data warehouse. We have concluded that MySQL best meets the criteria, as scored highest in final evaluation, followed by PostgreSQL, DB2, MongoDB, MS SQL and Oracle.

KEY WORDS: data warehouse, multi-attribute decision analysis, MySQL, PostgreSQL, MongoDB, MS SQL, Oracle, DB2

KAZALO

1.	UVOD.....	1
1.1	Opis problema.....	1
1.2	Cilji raziskave	2
1.3	Raziskovalna vprašanja.....	3
1.4	Metodologija	3
1.5	Pričakovani rezultati	3
1.6	Struktura magistrskega dela.....	4
2.	PODATKOVNA SKLADIŠČA.....	7
2.1	Zgodovina podatkovnih skladišč	7
2.2	Opredelitev pojmov	10
2.3	Arhitektura podatkovnih skladišč	14
2.3.1	<i>Centralizirana arhitektura podatkovnega skladišča</i>	15
2.3.2	<i>Distribuirana arhitektura podatkovnega skladišča</i>	16
2.3.3	<i>Federativna arhitektura podatkovnega skladišča</i>	17
2.4	OLAP sistemi.....	17
2.5	ETL proces.....	20
3.	OCENJEVANE IMPLEMENTACIJE PODATKOVNIH SKLADIŠČ	22
3.1	MySQL	26
3.2	PostgreSQL.....	29
3.3	MongoDB	31
3.4	Microsoft SQL Server.....	34
3.5	Oracle.....	36
3.6	DB2.....	38
4.	PROCES ODLOČANJA	39
4.1	Definicija odločanja.....	40
4.2	Faze odločitvenega procesa	40

4.3	Sistemi za podporo odločanju.....	41
4.3.1	<i>Modeli sistemov za podporo odločanju</i>	42
5.	PRIMERJALNA ANALIZA REFERENČNIH PODATKOVNIH SKLADIŠČ.....	44
5.1	Identifikacija problema.....	44
5.2	Izgradnja odločitvenega modela.....	45
5.3	Izbrani kriteriji.....	47
5.4	Identifikacija kriterijev.....	48
5.5	Opis zalog vrednosti kriterijev.....	52
5.6	Funkcija koristnosti.....	53
5.7	Odločitvena pravila.....	57
5.8	Opis variant.....	62
5.9	Vrednotenje in analiza variant.....	63
5.10	Analiza z uporabo orodja DEXi.....	70
5.11	Analiza z uporabo orodja VREDANA.....	77
5.12	SWOT analiza.....	79
6.	ZAKLJUČEK.....	81
6.1	Možnosti nadaljnjega razvoja.....	82
7.	LITERATURA IN VIRI.....	84
	PRILOGA	

KAZALO SLIK

Slika 1.1: Zelena končna arhitektura podatkovnega skladišča	5
Slika 2.1: Pomembna delitev različnih podatkovnih baz	10
Slika 2.2: Primer predmetne usmerjenosti podatkovnega skladišča.....	11
Slika 2.3: Različni nivoji podrobnosti podatkovnega skladišča	13
Slika 2.4: Arhitektura podatkovnega skladišča	14
Slika 2.5: Centralizirano podatkovno skladišče	15
Slika 2.6: Primer distribuiranega podatkovnega skladišča	16
Slika 2.7: Federativno podatkovno skladišče	17
Slika 2.8: Primer dimenzionalnega modela	18
Slika 2.9: Relacijska podatkovna baza	19
Slika 2.10: Večdimenzionalna podatkovna baza	19
Slika 2.11: Trije pristopi izvedbe ETL procesov.....	21
Slika 3.1: Priljubljene podatkovne baze	22
Slika 3.2: Struktura podatkovne baze v MS Access	23
Slika 3.3: Postopek migracije v MySQL	28
Slika 3.4: Migrirana podatkovna baza v MySQL.....	28
Slika 3.5: Postopek migracije v PostgreSQL	30
Slika 3.6: Migrirana podatkovna baza v PostgreSQL	30
Slika 3.7: Primer vstavljanja podatkov v MongoDB.....	31
Slika 3.8: Postopek migracije v MongoDB	32
Slika 3.9: Migrirana podatkovna baza v MongoDB.....	33
Slika 3.10: Primer prikaza podatkov v MongoDB	33
Slika 3.11: Postopek migracije v MS SQL.....	35
Slika 3.12: Migrirana podatkovna baza v MS SQL	35
Slika 3.13: Postopek migracije v Oracle 11g R2.....	37
Slika 3.14: Migrirana podatkovna baza v Oracle 11g	37
Slika 3.15: Postopek migracije v DB2	38
Slika 3.16: Migrirana podatkovna baza v DB2	39
Slika 4.1: Razvoj SPO koncepta.....	41
Slika 4.2: Večparametrski odločitveni model.....	43
Slika 5.1: Nadzor kakovosti podatkov.....	46

Slika 5.2: Drevo kriterijev	49
Slika 5.3: Izvajanje poizvedbe v MySQL.....	65
Slika 5.4: Izvajanje poizvedbe v PostgreSQL	65
Slika 5.5: Izvajanje poizvedbe v MongoDB.....	66
Slika 5.6: Izvajanje poizvedbe v MS SQL	66
Slika 5.7: Izvajanje poizvedbe v Oracle	67
Slika 5.8: Izvajanje poizvedbe v DB2	67
Slika 5.9: MySQL kot odprtokodna podatkovna baza	68
Slika 5.10: PostgreSQL kot odprtokodna podatkovna baza.....	68
Slika 5.11: MongoDB kot odprtokodna podatkovna baza	68
Slika 5.12: Strošek MS SQL	69
Slika 5.13: Strošek Oracle 11g	69
Slika 5.14: Strošek IBM DB2.....	70
Slika 5.15: Grafični prikaz končne ocene.....	71
Slika 5.16: Vpliv tehničnih in upravljaljskih kriterijev na podatkovno skladišče.....	72
Slika 5.17: Vpliv neposrednih in posrednih stroškov na celotne stroške	73
Slika 5.18: Vpliv administratorskega in uporabniškega vidika na tehnične kriterije.....	74
Slika 5.19: Vpliv kriterijev na migracijo	75
Slika 5.20: Vpliv orodij za dostop in funkcionalnosti poizvedb na podatkovno skladišče.....	76
Slika 5.21: Vpliv kriterijev združljivost oziroma kompatibilnost in kvalitete podatkov na končno oceno podatkovnega skladišča.....	77
Slika 5.22: Prikaz končne ocene variant v VREDANI.....	78
Slika 5.23: Sprememba zaloge vrednosti pri varianti MS SQL	79
Slika 5.24: »Kaj-če« analiza v VREDANI.....	79

KAZALO TABEL

Tabela 3.1: Virtualna okolja z virtualnimi sistemi	24
Tabela 5.1: Drevo kriterijev z opisi	51
Tabela 5.2: Zaloge vrednosti kriterijev	52
Tabela 5.3: Odločitvena pravila za kriterij podatkovno skladišče.....	54
Tabela 5.4: Odločitvena pravila za upravljavski kriterij	54
Tabela 5.5: Odločitvena pravila za tehnični kriterij	55
Tabela 5.6: Funkcije koristnosti	55
Tabela 5.7: Prikaz povprečnih uteži v odločitvenem drevesu	56
Tabela 5.8: Vpliv tehničnih in upravljavskih kriterijev na podatkovno skladišče	58
Tabela 5.9: Vpliv uporabniškega in administratorskega vidika na tehnični kriterij.....	58
Tabela 5.10: Vpliv uporabniškega vmesnika, orodij za dostop in funkcionalnost poizvedb na uporabniški vidik.....	59
Tabela 5.11: Vpliv poizvedb in orodij za analizo ter naprednih orodij za analizo na kriterij orodja za dostop	59
Tabela 5.12: Vpliv hitrosti poizvedb in grafičnega vmesnika za poizvedbe na funkcionalnost poizvedb	60
Tabela 5.13: Vpliv trajanja in težavnosti namestitve na kriterij namestitev	60
Tabela 5.14: Vpliv natančnosti, konsistence in celovitosti podatkov na kvaliteto prenesenih podatkov pri migraciji	61
Tabela 5.15: Vpliv obstoječe programske opreme in infrastrukture na kriterij kompatibilnost.....	61
Tabela 5.16: Vpliv izvornega sistema, drugih orodij za analizo, drugih orodij za podatkovno skladišče na kriterij združljivost.....	61
Tabela 5.17: Vpliv ekstrakcije, transformacije in nalaganja na kriterij ETL funkcionalnost ..	62
Tabela 5.18: Vpliv stroškov in ponudnika podatkovnega skladišča na upravljavski kriterij ...	62
Tabela 5.19: Vrednotenje kriterijev pri variantah	64
Tabela 5.20: Rezultati vrednotenja variant.....	70
Tabela 5.21: SWOT analiza variante MySQL	80

1. UVOD

1.1 Opis problema

V manjšem podjetju Car marketing iz Krškega se ukvarjajo s področjem telefonske prodaje. Trenutno implementiran informacijski sistem temelji na MS Access podatkovni bazi, ki pa se zaradi vse večje količine podatkov odziva in deluje vse počasneje. Prav tako informacijski sistem ne omogoča analiz in izvedb različnih poročil, ki bi podjetju koristila kot podpora pri nadaljnjemu odločanju in planiranju. Podjetje se je zato odločilo, da bodo obstoječo podatkovno bazo v MS Accessu migrirali v podatkovno skladišče, ki bo zadostovalo zahtevanim kriterijem.

Za namen magistrskega dela smo skupaj z vodstvom podjetja Car Marketing (z metodo delno strukturiranega intervjuja) preučili in definirali kriterije z uporabniškega vidika. S preučitvijo strokovne literature s področja migracij podatkovnih baz v podatkovna skladišča pa smo pridobili kriterije za tehnično izvedbo migracije, ki jih mora izpolnjevati podatkovno skladišče. S pomočjo definiranih kriterijev smo ovrednotili več različnih implementacij podatkovnih skladišč.

Podatkovno skladišče (ang. »*data warehouse*«) je sistem, ki pridobi in periodično združi podatke iz izvornih sistemov v dimenzionalne ali normalizirane podatkovne shrambe. Običajno vsebuje večletne podatke, v katerih se poizveduje za namen poslovnega obveščanja. Podatki se običajno posodablajo v serijah, vsaka transakcija pa se ne zgodi vedno na izvornem sistemu. Podatkovna skladišča se prav tako uporabljajo za različne analitične namene, kot so znanstvene raziskave, statistična poročila, ekonomske analize in napovedi, spremljanje uspešnosti in analize omrežnega prometa (Rainardi, 2008).

Tako lahko rečemo, da je podatkovno skladišče namenjeno podpori odločanju, saj je namen podatkovnih skladišč izboljšati odločanje, planiranje in nadzor nad različnimi poslovnimi področji (Jürgens, 2002).

Odločanje predstavlja sestavni del procesa reševanja problemov, pri čemer se srečujemo z različnimi tveganji in omejitvami. Agarwal in drugi (2011) poudarijo Zeleznikow vidik odločanja, ki pravi, da inteligentni sistemi za podporo odločanju vodijo k doslednemu načinu

odločanja in večjemu zaupanju, medtem ko le-to zagotavlja dodatno podporo in alternativo za reševanje sporov in težav.

Na vsakem koraku se soočamo z različnimi odločitvami, ki spreminjajo našo vsakdanjo poslovno ali osebno pot. V poslovnem svetu, kjer se operira z veliko količino podatkov, je še toliko bolj pomembno, da so podatki v primerni obliki in pripravljeni za različne analize in nadaljnje obdelave, saj lahko odločitve sprejemamo le na podlagi informacij. Za takšne namene uporabljamo podatkovna skladišča, ki omogočajo, da lahko z različnimi orodji za analizo podatkov (npr. OLAP, ang. »*on line analytical processing*«) oblikujemo informacije iz zbranih podatkov in izvajamo kompleksna ad-hoc poizvedovanja (Thomsen, 2002). Poznamo različne implementacije podatkovnih skladišč, med katerimi zasledimo odprtokodna (kot so MySQL, PostgreSQL, MongoDB) in komercialna (kot so Microsoft SQL Server, Oracle in IBM DB2). Glede na zahtevnost in podprtost različnih sistemov za poslovno inteligenco ter količino zbranih podatkov se pogosto pojavi dvom, katero implementacijo podatkovnega skladišča izbrati, da bomo kasneje z njo zadovoljni pri pridobivanju potrebnih informacij iz podatkov.

V ta namen smo postavili posamezna podatkovna skladišča v virtualnem okolju in preizkusili ter ovrednotili posamezne izbrane kriterije. Z orodjem DEXi in VREDANA ter SWOT analizo smo primerjalno analizirali ugotovljene vrednosti in predlagali optimalno podatkovno skladišče na osnovi izbranih kriterijev.

1.2 Cilji raziskave

Cilj raziskave je bil s kabinetno študijo iz področja podatkovnih skladišč:

- predstaviti pojem podatkovna skladišča,
- predstaviti vlogo in zgradbo podatkovnega skladišča,
- opisati pomen podatkovnih skladišč,
- določiti kriterije ocenjevanja in s pomočjo orodja DEXi izdelati večkriterijsko parametrsko analizo in z njo ugotoviti, kako ocenjevane implementacije podatkovnih skladišč dosegajo postavljene kriterije ter izpostaviti bistvene razlike, prednosti in slabosti.

1.3 Raziskovalna vprašanja

Za namen obravnavanega področja smo postavili dve raziskovalni vprašanji:

1. Kateri kriteriji so v podjetju Car Marketing najbolj pomembni za migracijo podatkovne baze iz MS Access v podatkovna skladišča?
2. Kako odprtokodna in komercialna podatkovna skladišča zadovoljujejo izbrane kriterije?

1.4 Metodologija

V magistrski nalogi smo sledili implementaciji raziskovalnega dela, kot ga opisuje Biggam (2008) :

1. podroben pregled literature in
2. zbiranje empiričnih podatkov.

Pri obravnavanem področju smo uporabili naslednje metode dela:

- analizo vsebine (deskriptivna metoda) – pregled domače in tuje literature, ki obravnava področje sistemov za podporo odločanju,
- komparativno metodo, kjer smo med seboj primerjali več različnih implementacij podatkovnih skladišč,
- kvalitativno in kvantitativno analizo zbranih podatkov, informacij in gradiva,
- metodo odločitvene analize,
- model kvalitativnega modeliranja odločitvenega znanja,
- sintezo analitičnih ugotovitev,
- delno strukturiran intervju.

1.5 Pričakovani rezultati

Pričakovan rezultat večparametrskane analize je bil ovrednotiti uporabnost obravnavanih podatkovnih skladišč na podlagi postavljenih odločitvenih kriterijev. Pri tem smo pridobili konkretne rezultate o uporabnosti več različnih implementacij podatkovnih skladišč, pri čemer smo le-te prikazali s pomočjo orodja DEXi in VREDANE.

1.6 Struktura magistrskega dela

V začetnem delu magistrskega dela smo se osredotočili predvsem na podrobno preučitev tuje in domače strokovne literature ter izsledkov raziskav s področja podatkovnih skladišč in sistemov za podpro odločanju. Obdelali in predstavili smo tudi OLAP in ETL proces, ki sta pomemben del analize in postavitve podatkovnih skladišč.

V nadaljevanju smo z uporabo primerjalne raziskovalne metode analizirali in primerjali komercialna podatkovna skladišča (Microsoft SQL Server, Oracle in DB2) ter odprtokodna podatkovna skladišča (MySQL, PostgreSQL in MongoDB).

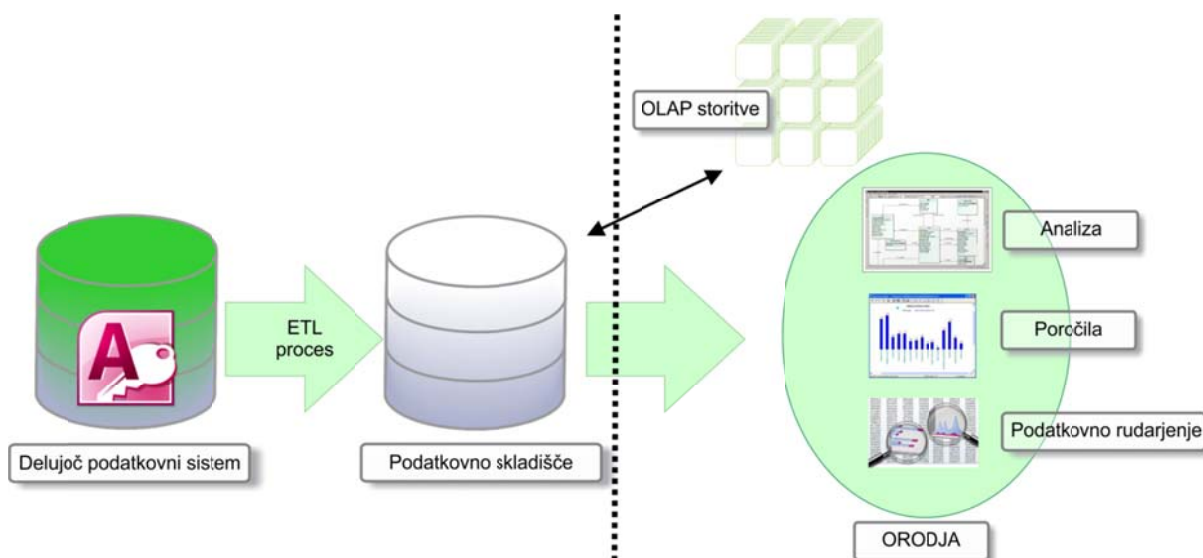
Predstavili smo proces odločanja, faze odločitvenega procesa ter sisteme za podporo odločanju, s pomočjo katerih smo v empiričnem delu magistrskega dela zgradili odločitveno drevo za podajanje konkretnih rezultatov.

V zadnjem (empiričnem) delu magistrskega dela smo na praktičnem primeru predstavili migracijo podatkovne baze v podatkovno skladišče. Postavili smo strategijo izgradnje podatkovnega skladišča, pri čemer smo se osredotočali na glavne tri elemente, ki jih mora strategija vsebovati (Nagabhushana, 2006):

1. predhodni načrt podatkovnega skladišča,
2. predhodno arhitekturo podatkovnega skladišča in
3. kratek seznam okolja in orodij podatkovnega skladišča.

Pri arhitekturi podatkovnega skladišča (Slika 1.1) smo izhajali iz arhitekture, kot jo navajata Prabhu in Venatesan (2007).

Slika 1.1: Zelena končna arhitektura podatkovnega skladišča



Vir: Povzeto po Prabhu in Venatesan (2007, str. 6)

Strategijo izgradnje podatkovnega skladišča smo izdelali na podlagi delno strukturiranega intervjuja, ki smo ga izvedli z vodstvom podjetja. Pri izvedbi intervjuja smo se osredotočili na naslednja ključna področja (Nagabhushana, 2006):

1. **Funkcionalnost.** Kaj je poslanstvo podjetja? Kako podjetje izpolnjuje in dosega zastavljeno poslanstvo? Kako podjetje ve, ali zastavljeno poslanstvo prinaša uspeh? Kaj so ključni kazalniki uspešnosti in kritični dejavniki uspeha?
2. **Stranke.** Na kakšen način podjetje grupira svoje stranke? Ali se te skupine spreminjajo v določenem časovnem obdobju? Ali grupiranje vpliva na obnašanje do strank? Katere informacije podjetje hrani o strankah? Katere demografske podatke uporabljate? Ali je potrebno slediti vsaki stranki posebej?
3. **Dobiček.** Na kateri ravni podjetje meri donosnost posamezne skupine? Na uslužbenca? Na stranko? Po regijah? Kako podrobno podjetje meri stroške in prihodke? Kako podjetje meri in spremlja stroške in prihodke? Kakšna poročila o dobičku podjetje trenutno uporablja?
4. **Sistem.** Katere sisteme podjetje uporablja pri svojem poslovanju? Katerih sistemov se zavedajo drugi udeleženci v procesu? Kakšen način ročnega prenosa podatkov ima podjetje v primeru, da podatki niso na voljo?
5. **Čas.** Koliko časa (mesecev, let) podjetje hrani podatke? Ali podjetje analizira podatke po letih? Kako pogosto podjetje želi poročilo (dnevno, tedensko, mesečno, na vsako

- četrletje, letno)? Kako hitro podjetje želi podatke (primer: želi včerajšnje podatke videti danes)? Kako hitro ob koncu določenega obdobja želi podjetje videti podatke?
6. **Poizvedbe in poročila.** Katera poročila podjetje trenutno uporablja? Katere informacije se uporabljajo v trenutnih poročilih? Kako pogosto se izdelujejo poročila? Ali se poročila izdelujejo dovolj pogosto? Kdo izdeluje poročila?
 7. **Izdelek.** Katere izdelke podjetje prodaja in kako jih klasificira? Ima podjetje hierarhijo izdelkov? Ali se analize izdelujejo za vse izdelke hkrati ali se analize izdelujejo za posamezni izdelek? Kako podjetje upravlja spremembo v hierarhiji izdelkov?
 8. **Geografija.** Ali podjetje deluje več kot samo na eni lokaciji? Ali podjetje deli tržišče na geografska območja? Ali podjetje sledi prodaji glede na geografsko območje?
 9. **Trenutna arhitektura.** Kakšna je trenutna tehnološka arhitektura podjetja? Kakšne vrste sistemov, strojne opreme, podatkovnih skladišč, omrežje, uporabniških orodij, razvojnih orodij in orodij za dostop podatkov so trenutno v uporabi?
 10. **Relacije izvornega sistema.** Ali so izvorni sistemi kakor koli medsebojno povezani? Ali en sistem zagotavlja informacije drugemu? Ali so sistemi na kakršen koli način integrirani? V primerih, kjer ima več sistemov podatke o strankah in produktih, kateri sistem »služi« kot kopija?
 11. **Omrežne zmogljivosti.** Ali je možno uporabiti en računalnik za dostop do različnih operacijskih sistemov iz katere koli lokacije?
 12. **Kvaliteta podatkov.** Koliko čiščenja, podvajanja in integracije mislite, da bo potrebno? Katera področja (tabele ali polja) v izvornih sistemih so trenutno znana, da vsebujejo slabo kvaliteto podatkov?
 13. **Dokumentacija.** Koliko dokumentacije je trenutno na voljo za izvorni sistem? Kako točni in posodobljeni so ti priročniki in referenčna gradiva? Poskusite dobiti sledeče podatke, kadar koli je to možno: kopijo priročnikov in referenčne dokumentacije, velikost podatkovne baze, načrtovane izboljšave, velikost varnostne kopije, obseg varnostne kopije in medij varnostne kopije, obseg podatkov v sistemu (primer: pomembne tabele in polja), kodeksi omrežja in njihov pomen ter ključne generacijske sheme.
 14. **Možni mehanizmi za ekstrakcijo.** Kateri mehanizmi ekstrakcije so možni s trenutnim sistemom? Katere mehanizme ekstrakcije ste že uporabljali s tem sistemom? Kateri mehanizmi ekstrakcije ne bodo delovali?

Na podlagi pridobljenih podatkov iz strukturiranega intervjuja ter preučitvijo drugih praktičnih primerov smo določili kriterije podatkovnega skladišča. Le-te smo nato s kvalitativno metodo analizirali v orodju DEXi, kjer smo izvedli komparativno analizo kvalitativnih podatkov, ki smo jih pridobili pri različnih migracijah v različna podatkovna skladišča. S funkcijo koristnosti, ki smo jo definirali z odločitvenimi pravili tipa če-potem, smo opredelili vpliv določenih kriterijev na končno odločitev. Dobljene rezultate smo grafično prikazali ter na koncu naredili še SWOT analizo za ocenjevano varianto z najboljšo končno oceno.

2. PODATKOVNA SKLADIŠČA

Sistemi podatkovnih skladišč so postali ena izmed ključnih komponent informacijske strukture podjetja in igrajo zelo pomembno vlogo pri vzpostavitvi poslovnega sistema za podporo odločanju. Z zbiranjem in konsolidacijo podatkov iz različnih notranjih in zunanjih virov poskušajo podatkovna skladišča zagotavljati homogeno informacijsko osnovo za uspešno poslovno planiranje in odločanje.

Kot je danes znano, ekonomija zelo hitro spreminja trg in poslovne priložnosti. Zato je za podjetje zelo pomembno, da lahko sprejema prave poslovne odločitve in tako ostaja v koraku ali pa celo pred konkurenco. Poslovne odločitve temeljijo na analizi preteklih in trenutnih pogojev poslovanja, analizi tržišča in predvidevanj za prihodnost. V ta namen se v času »življenjske dobe« podjetja zbirajo različni operativni podatki (Wrembel in Koncilia, 2007).

2.1 Zgodovina podatkovnih skladišč

Podatkovna skladišča poznamo približno dve desetletji in so postala pomemben del infrastrukture informacijskega sistema. Na začetku smo poznali enostavne mehanizme, ki so hranili podatke, kot denimo luknjane kartice (ang. »punched cards«), ki so bile kot papirni trakovi. Na začetku je bila takšna hramba zelo draga in tudi omejena.

Velik preskok pri shrambi podatkov se je pokazal z razvojem magnetnih trakov, na katere se je lahko shranjevalo velike količine podatkov in to zelo cenovno ugodno. Na magnetne trakove so se podatki lahko prepisovali in ponovno shranjevali. Vendar pa tudi magnetni trakovi niso bili idealni, saj je bil dostop do podatkov možen le sekvenčno. Prav tako pa magnetni trak z vidika varnosti ni bil ravno najbolj primeren. Še en korak naprej pa so predstavljali trdi diski, ki so omogočali direkten dostop do podatkov (Inmon in drugi, 2008).

Začetek razvoja podatkovnih skladišč beležimo po letu 1970, ko so na priljubljenosti pridobivali mini računalniki. Razvil se je nov koncept razmišljanja, s katerim bi se pohitrile poizvedbe v podatkovnih bazah in s tem razbremenil oddelek za obdelavo podatkov. Koncept je temeljil na kopiranju ključnih elementov podatkov, ki so bili pogosto uporabljeni, na podatkovne trakove ali mini računalnike, poslovni uporabniki pa bi lahko s pomočjo orodij za podporo odločanju pridobili podatke, ki jih potrebujejo.

Takšen pristop se je izkazal za učinkovitega z vidika razbremenitve oddelka za obdelavo podatkov. Težava se je pojavila pri nerednih osveževanjih podatkov, kar je otežilo proces pri podjetjih, ki so podatke pridobivala iz podatkovnih baz na različnih poslovnih lokacijah. Sčasoma so se začele pojavljati rešitve tudi za to oviro. Med letoma 1976 in 1979 so na inštitutu za raziskave in razvoj v Kaliforniji oblikovali sistem za upravljanje podatkovnih baz za paralelno procesiranje z več mikroprocesorji, še posebej namenjen sistemom za podporo odločanju (Hammergren in Simon, 2009).

Vse od prvega razvoja računalnikov sta shranjevanje in upravljanje podatkov predstavljala velik izziv. V 70. letih prejšnjega stoletja so se pojavili sistemi za upravljanje podatkovnih baz (DBMS, ang. »*data base management system*«) (Kedar, 2009).

DBMS programska oprema je bila razvita z namenom upravljanja podatkov na samem disku in je omogočala (Inmon in drugi, 2008):

- identifikacijo točne lokacije podatka na disku,
- reševanje konfliktov, kadar sta bili dve enoti podatkov preslikani na isto lokacijo,
- brisanje podatkov,
- razširitev fizične lokacije, ko je bil zapis večji od omejenega fizičnega prostora.

Ena izmed največjih prednosti shranjevanja na disk je bila vsekakor ta, da je možno hitro najti iskane podatke. In ravno pri uporabi DBMS je ta prednost diskov prišla najbolj do izraza.

Z uporabo trdih diskov in DBMS sistemov je bilo možno direktno dostopati do podatkov, tako se je kmalu začel tako imenovani razvoj spletnih aplikacij. Spletno računalništvo del obdelave seli (nazaj) na strežnike in ne zahteva zmogljivih osebnih računalnikov. Razvite so bile številne komercialne aplikacije za spletno obdelavo podatkov, ki so hitro postale zelo priljubljene (Inmon in drugi, 2008).

Edgar Codd, raziskovalec pri podjetju IBM, je v poznih 70. letih prejšnjega stoletja predlagal nov okvir predstavitve podatkov imenovan relacijski podatkovni model. To se je izkazalo kot prelomnica v razvoju sistema podatkovnih baz, saj je sprožilo hiter razvoj različnih DBMS sistemov, ki so temeljili na relacijskem modelu (Ramakrishnan in Gehrke, 2003).

Z razvojem se je povečevala tudi količina podatkov, ki so se hitro začeli kopičiti in so postali nepregledni ter razpršeni. Po letu 1980 se je pojavila nova ideja o izdelavi posebne programske opreme, ki se imenuje porazdeljen sistem za upravljanje podatkovnih baz (DDBMS, ang. »*distributed data base management system*«) (Hammergren in Simon, 2009).

Ralph Kimball je leta 1986 ustanovil Red Brick Systems z vizijo razvoja programske opreme, ki bi izboljšala dostop do podatkov. Promovirali so specializirano platformo za relacijske podatkovne baze, ki je omogočala kompleksne ad-hoc poizvedbe.

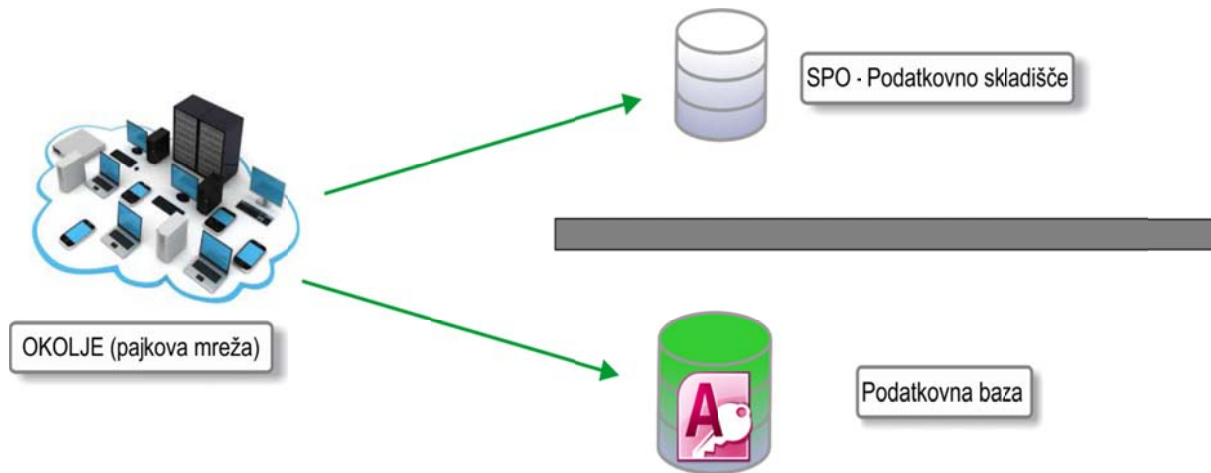
V letu 1988 sta Barry Devlin in Paul Murphy iz podjetja IBM predstavila izraz poslovna podatkovna skladišča kot sestavni del EBIS-a (Skupni informacijski sistem – Evropa, Bližnji Vzhod, Afrika, ang. »*Europe/Middle East/Africa Business Information System*«), ki je temeljil na fleksibilnosti in medsebojni podpori. Če bi povzeli, so bila osemdeseta prejšnjega stoletja rojstni čas podatkovnih skladišč.

Leta 1993 je Bill Inmon, ki ga imenujejo tudi oče podatkovnih skladišč, napisal svojo knjigo Gradnja podatkovnih skladišč. Programska in strojna oprema sta postajali vse bolj kompatibilni pri izgradnji podatkovnih skladišč, pojavile so se različne tehnike dostopov do podatkov. Z uporabo tehnologije odjemalec/strežnik pa je bilo možno izboljšati orodja za analizo podatkov (Hammergren in Simon, 2009).

Razvoj podatkovnih skladišč je predstavljal nov izziv v razmišljanju za IT strokovnjake. Pred prihodom podatkovnih skladišč se je zdelo, da bi podatkovna baza morala služiti za vse

namene, vendar pa je kmalu postalo jasno, da obstaja več različnih vrst podatkovnih baz (Slika 2.1) (Inmon in drugi, 2008).

Slika 2.1: Pomembna delitev različnih podatkovnih baz



Vir: Povzeto po Inmon in drugi (2008, str. 6)

Svoj razcvet so podatkovna skladišča doživela v začetku 21. stoletja, saj se je količina podatkov nenehno povečevala. Leta 2006 je Microsoft prevzel ProClarity (podjetje specializirano za poslovno inteligenco in programsko opremo za analizo podatkov) in tako vstopil v svet podatkovnih skladišč. Leta 2007 je podjetje Oracle kupilo Hyperion (vodilni ponudnik programske opreme za upravljanje), SAP je prevzel podjetje Business Objects, podjetje IBM pa se je združilo s podjetjem Cognos. Tako se je začel razvoj podatkovnih skladišč največjih ponudnikov (Hammergren in Simon, 2009).

2.2 Opredelitev pojmov

V literaturi zasledimo več različnih definicij pojma podatkovno skladišče. Prabhu (2008) definira podatkovno skladišče kot odlagališče subjektivno izbranih in prilagojenih operativnih podatkov, ki lahko uspešno opravljajo vse ad-hoc kompleksne, statistične in analitične poizvedbe. Podatkovno skladišče se nahaja v središču sistemov za podporo odločanju (SPO, ang. »*decision support systems*«) in vsebuje vse celovite zgodovinske podatke, tako povzete kot tudi podrobnejše informacije, ki so skupne v celotni organizaciji.

Podatkovno skladišče je podatkovna baza posebne vrste, ki periodično zbira informacije o dejavnostih, ki se izvajajo v organizaciji. Te informacije so nato uporabljene za namene

analize in odkrivanja strateških informacij, kot so trendi, korelacije in podobno. Podatkovna skladišča se vse bolj uporabljajo v različnih organizacijah, da lahko le-te izboljšajo svoje poslovanje in tako dosežejo zastavljene cilje (Malinowski in Zimányi, 2009).

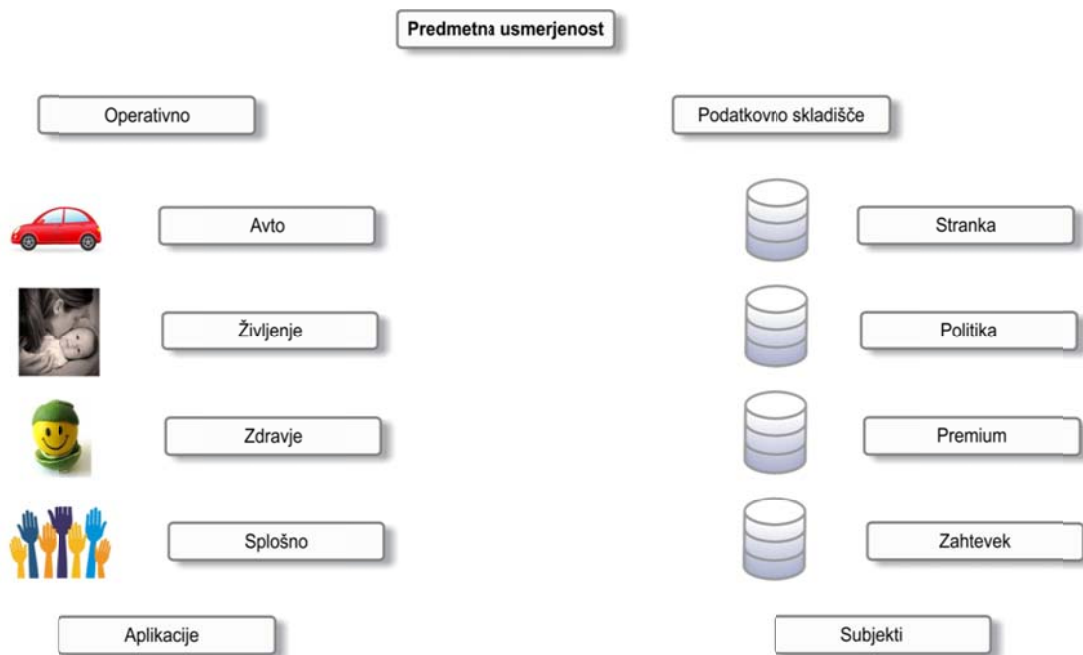
Podatkovno skladišče je »dom« vseh pomembnih podatkov, ki izvirajo iz drugih poslovnih aplikacij, kot so tiste, ki jih uporabljamo v podjetjih, na primer za izpolnitev naročil kupcev ali podatkov iz zunanjih virov, kot so javno dostopne baze podatkov, ki vsebujejo podatke o konkurenci (Hammergren in Simon, 2009).

Prav tako zaznamo odmevni definiciji pojma podatkovno skladišče med Inmonom in Kimballom.

Inmon (2005) pravi, da je podatkovno skladišče srce zgrajenega okolja in predstavlja temelj vseh obdelav SPO. Podatkovno skladišče je predmetno usmerjena, integrirana, stalna in časovna različica zbirke podatkov pri podpori odločanja. Vsebuje spisek vseh poslovnih podatkov.

Predmetna usmerjenost (Slika 2.2) – kjer so podatki grupirani po subjektu, kar omogoča shranjevanje podatkov v eno skupino ne glede na sam izvor.

Slika 2.2: Primer predmetne usmerjenosti podatkovnega skladišča



Vir: Povzeto po Inmon (2005, str. 32)

Integracija – je najbolj pomemben vidik podatkovnega skladišča. Podatke, ki prihajajo iz različnih virov, je potrebno poenotiti oziroma povezati.

Stalnost – podatki se v podatkovnem skladišču hranijo zelo dolgo. Operativni podatki so vedno dostopni in se iz podatkovnega skladišča izbrišejo, ko niso več potrebni.

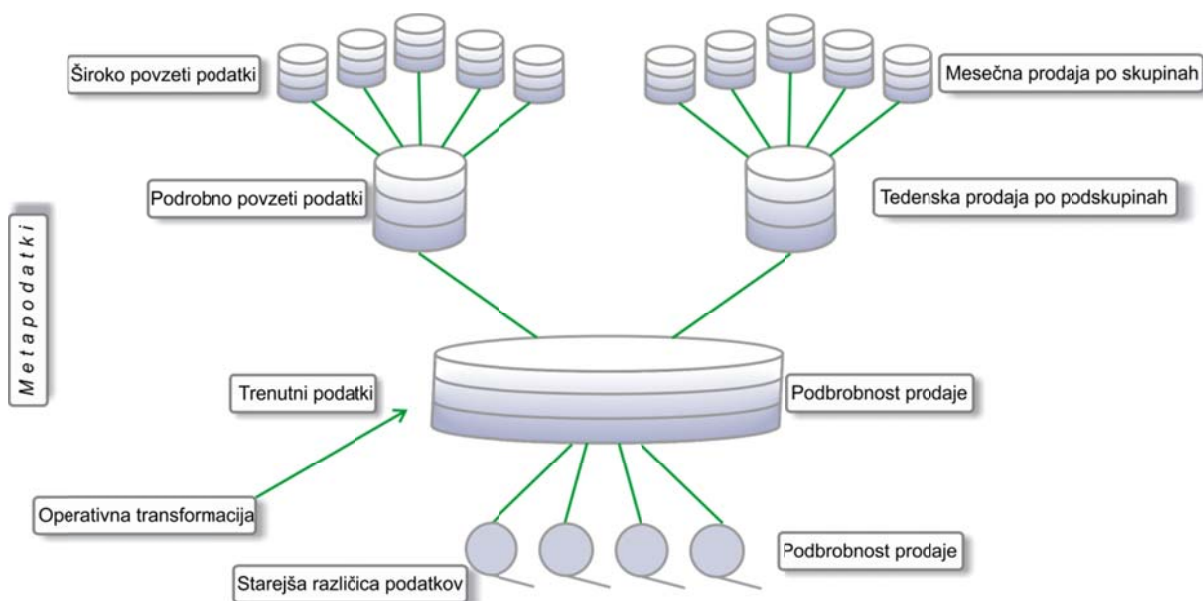
Časovna različica – različna okolja imajo različna časovna obdobja, zato podatki v podatkovnem skladišču vsebujejo podatek o času nastanka.

Podatkovno skladišče je usmerjeno v glavna tematska področja, ki se določijo v višjih nivojih podatkovnega modela. Tipična tematska področja vključujejo (Inmon, 2005):

- stranke,
- izdelek,
- transakcije,
- politiko,
- terjatve,
- račune.

Slika 2.3 prikazuje, da ima podatkovno skladišče več različnih nivojev podrobnosti. Vsebuje starejšo različico podatkov (običajno masovno shranjevanje), stopnjo trenutnih podatkov, stopnjo široko povzetih podatkov in stopnjo podrobno povzetih podatkov. Podatki se pretakajo v podatkovno skladišče iz operativnega okolja. Običajno prihaja do pomembnega preoblikovanja podatkov pri prehodu iz operativne ravni na raven podatkovnega skladišča (Inmon, 2005).

Slika 2.3: Različni nivoji podrobnosti podatkovnega skladišča



Vir: Povzeto po Inmon (2005, str. 36)

Kimball in drugi (2008) pa pravijo, da je podatkovno skladišče temelj za poslovno inteligenco ter zagotavlja podporo programski opremi. Eden izmed ciljev podatkovnega skladišča je zgraditi ugled za zagotavljanje pravočasnih, doslednih in zanesljivih podatkov, s čimer lahko okrepimo podjetje. Podatkovno skladišče je kopija podatkov, ki so posebej oblikovani za namene poizvedbe in analiz ter služijo kot podpora sistemu za odločanje. Asemi in drugi (2011) ga definirajo po Raymond (1990), ki razlaga SPO kot računalniški informacijski sistem, ki je namenjen določenemu upravljavcu ali običajno skupini menedžerjev na organizacijski ravni v procesu odločanja pri reševanju strukturiranih odločitev. SPO poda izhod v obliki periodičnih ali posebnih poročil, lahko pa je rezultat tudi matematična simulacija.

Rainardi (2008) definira podatkovno skladišče kot sistem, ki pridobi in redno združuje podatke iz izvornih sistemov v dimenzionalne ali normalizirane podatkovne shrambe. Običajno se podatki hranijo več let, na podlagi le-teh pa se izvajajo poizvedbe za poslovno inteligenco ali druge analitične dejavnosti. Podatkovno skladišče običajno deluje po sistemu samo za branje (ang. »*read-only*«). To pomeni, da uporabniki ne morejo posodabljan ali brisati podatkov v podatkovnem skladišču, s čimer se zagotavlja konsistenca podatkov. Podatki v podatkovnem skladišču se posodabljanjo z uporabo standardnega mehanizma, ki se imenuje ETL (predstavili ga bomo v nadaljevanju), v določenih obdobjih, pri čemer se

podatki iz operativnega sistema prenesejo v podatkovno skladišče. Prenos podatkov v določenih obdobjih in ne v realnem času zagotavlja stabilnost podatkov v podatkovnem skladišču.

2.3 Arhitektura podatkovnih skladišč

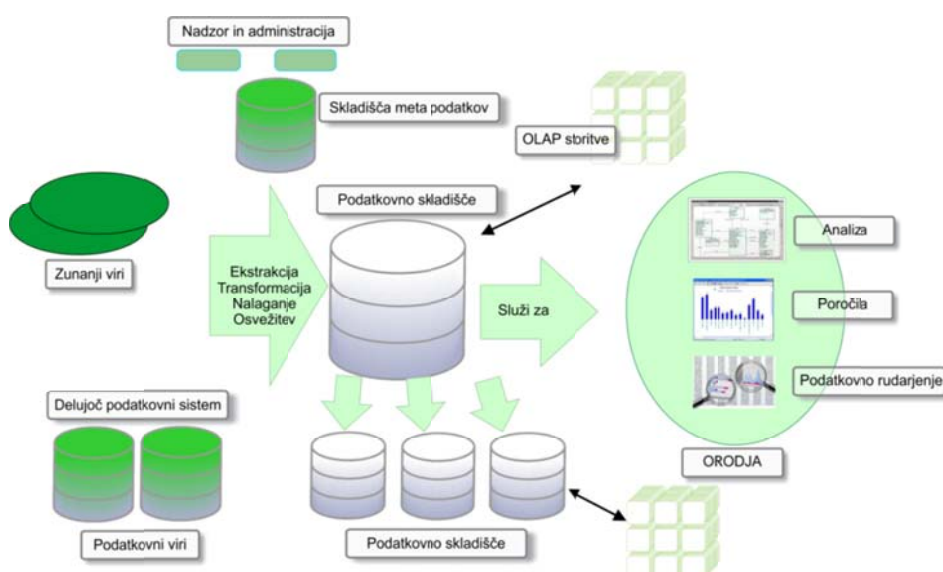
Podatkovno skladišče ima dve glavni arhitekturi (Rainardi, 2008) :

- arhitektura pretoka podatkov, katere namen je razporeditev podatkovne shrambe v podatkovnem skladišču ter kako se podatki pretakajo iz izvornih sistemov do uporabnikov s pomočjo urejenih podatkovnih shramb in
- arhitektura sistema, ki prikazuje fizično konfiguracijo strežnikov, omrežja, programske opreme in shrambe.

Arhitektura podatkovnega skladišča (Slika 2.4) vsebuje orodja za pridobivanje podatkov iz različnih operativnih podatkovnih zbirk in zunanjih virov (Prabhu in Venatesan, 2007):

- za čiščenje, preoblikovanje in integracijo teh podatkov,
- za nalaganje podatkov v podatkovno skladišče in
- za periodično osveževanje podatkovnega skladišča.

Slika 2.4: Arhitektura podatkovnega skladišča



Vir: Povzeto po Prabhu in Venatesan (2007, str. 6)

Pri načrtovanju podatkovnega skladišča je izbor pravilne arhitekture ključnih korakov pri izgradnji.

V nadaljevanju predstavljamo tri najpogosteje uporabljene arhitekture pri izgradnji podatkovnih skladišč. To so: centralizirana, distribuirana in federativna arhitektura podatkovnega skladišča.

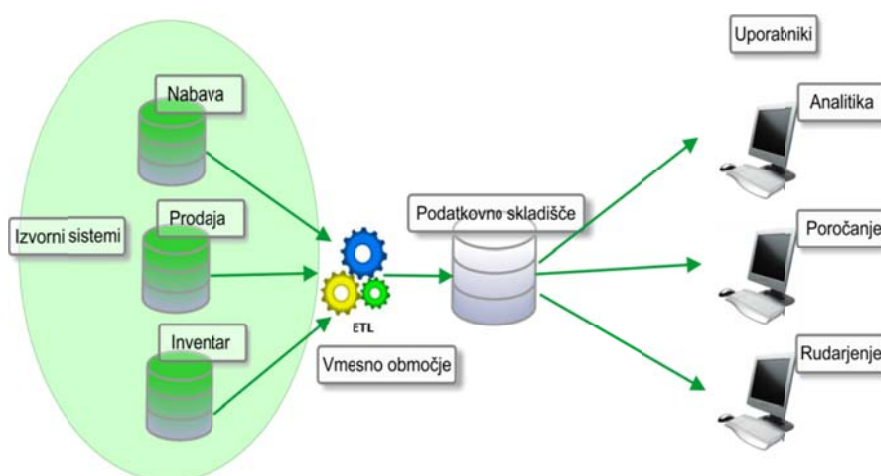
2.3.1 Centralizirana arhitektura podatkovnega skladišča

Večina organizacij se danes odloča za gradnjo centraliziranih podatkovnih skladišč. Takšna izgradnja podatkovnega skladišča je smiselna zaradi več razlogov (Inmon, 2005):

- podatki in podatkovno skladišče so integrirani skozi celotno organizacijo, integrirani vpogled pa se uporablja samo na sedežu,
- organizacija operira na centraliziranem poslovnem modelu,
- četudi bi podatki lahko bili integrirani in bi bili le-ti razpršeni po več lokacijah, bi bil dostop do njih otežen.

Centralizirano podatkovno skladišče (Slika 2.5) vsebuje podatke o atomski ravni, nekaj povzetih podatkov in logične dimenzijske poglede podatkov. Poizvedbe in aplikacije dostopajo do podatkov tako iz relacijskih podatkov kot tudi iz dimenzionalnih pogledov. Ta arhitektura je pogosto logična, ne pa tudi fizična implementacija informacijske strukture.

Slika 2.5: Centralizirano podatkovno skladišče



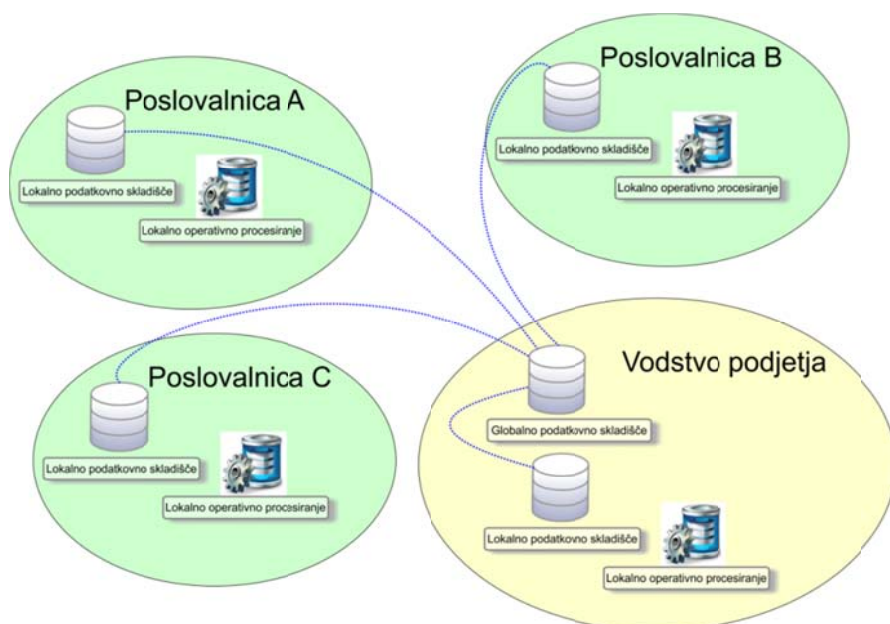
Vir: Povzeto po Ariyachandra in Watson (2005)

2.3.2 Distribuirana arhitektura podatkovnega skladišča

Ločimo tri vrste distribuiranih podatkovnih skladišč (Inmon, 2005):

1. Lokalno in globalno podatkovno skladišče. Lokalno podatkovno skladišče predstavlja podatke in njihovo obdelavo na oddaljeni lokaciji, globalno podatkovno skladišče pa predstavlja tisti del poslovanja, ki je vgrajen znotraj podjetja.
2. Tehnološko distribuirano podatkovno skladišče. Okolje podatkovnega skladišča bo vsebovalo, količina podatkov bo razdeljena na več procesorjev. Logična rešitev je eno podatkovno skladišče, vendar fizično obstaja veliko podatkovnih skladišč, ki so vsa tesno povezana, vendar se nahajajo na ločenih strežnikih.
3. Neodvisno distribuirano podatkovno skladišče. Okolje podatkovnega skladišča nastaja na neuskladen način, najprej eno podatkovno skladišče, nato drugo. Pomanjkanje usklajevanja rasti različnih podatkovnih skladišč, je običajno posledica političnih in organizacijskih razlik.

Slika 2.6: Primer distribuiranega podatkovnega skladišča

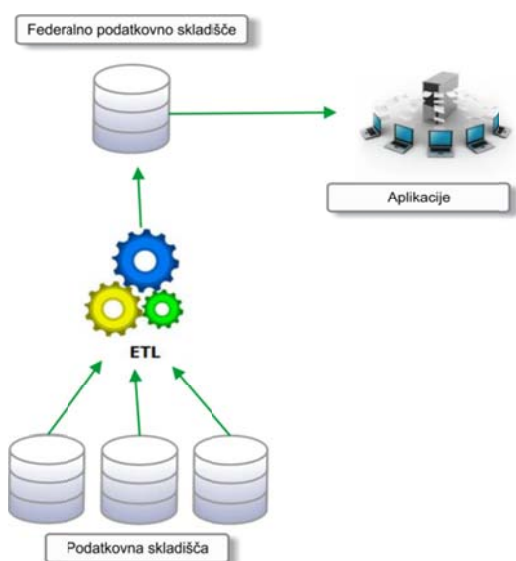


Vir: Povzeto po Inmon (2005, str. 207)

2.3.3 Federativna arhitektura podatkovnega skladišča

Federativno podatkovno skladišče (FPS) (Slika 2.7) je sestavljeno iz več podatkovnih skladišč. FPS pridobi podatke iz obstoječih podatkovnih skladišč z uporabo ETL (ang. »extract, transform, load«) in naloži podatke v novo dimenzionalno podatkovno shrambo (Rainardi, 2008).

Slika 2.7: Federativno podatkovno skladišče



Vir: Povzeto po Rainardi (2008, str. 40)

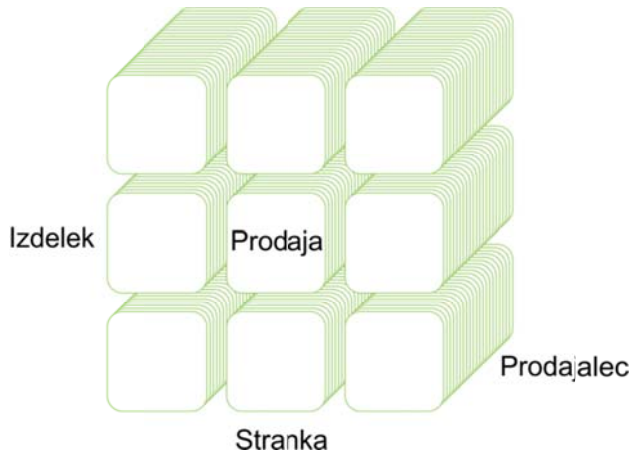
Granulacija podatkov v FPS je enaka kot najvišja granularnost izvornih podatkovnih skladišč. Če so granularnosti izvornih podatkovnih skladišč G1, G2 in G3, potem je najvišja vrednost granulacije FPS ena izmed G1, G2 ali G3. Če pogledamo na primeru, če je G1 dnevni, G2 tedenski in G3 mesečni, potem je granulacija FPS mesečna. To je zaradi tega, ker se ne da pretvoriti G1 in G2 v mesečno, lahko pa pretvorimo G3 v dnevno ali tedensko.

2.4 OLAP sistemi

OLAP (ang. »OnLine Analytical Processing«) – sprotna analitična obdelava – je proces, ki omogoča analizo velikih količin podatkov v zelo kratkem času. Podatke prikazuje večdimenzionalno, čemur rečemo OLAP kocka (Hammergren in Simon, 2009).

Za analizo prodaje po kupcih in izdelkih (glede na prodajalca) lahko za primer procesa prodaje v podjetju Car marketing prikažemo z dimenzionalno kocko kot na Sliki 2.8.

Slika 2.8: Primer dimenzionalnega modela



Vir: Klenovšek, lastni prikaz (2014)

Nagabhushana (2006) definira OLAP s petimi ključnimi elementi, ki so hitra analiza deljenih večdimenzionalnih informacij.

Hitrost pomeni, da sistem odgovarja na zahteve uporabnike v okviru petih sekund, kjer najbolj enostavna analiza ne sme trajati več kot eno sekundo. Takšne hitrosti ni enostavno doseči pri veliki količini podatkov, še posebno, če se zahtevajo ad-hoc kalkulacije.

Analiza pomeni, da se sistem lahko spopada s kakršno koli logiko ali statistično analizo, ki je pomembna za aplikacijo in končnega uporabnika, pri tem pa še vedno stremi k enostavnosti. Uporabniku mora omogočati definicijo kakršnih koli ad-hoc kalkulacij, ki so del analize in poročila.

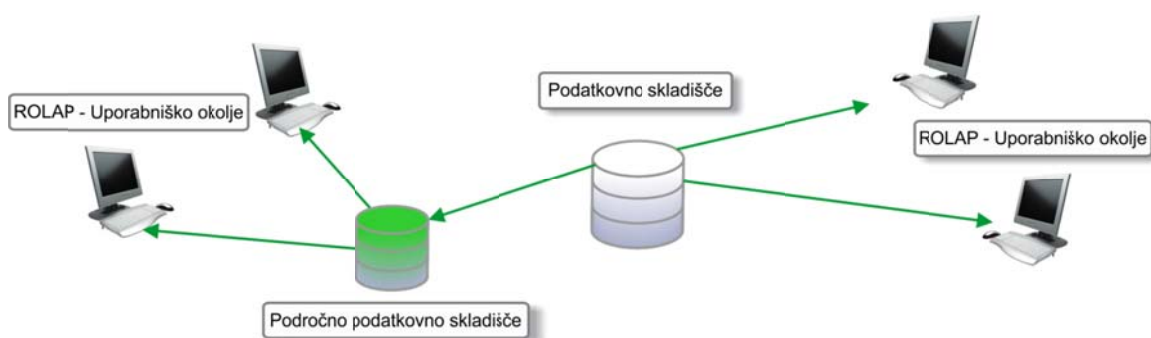
Deljen pomeni, da sistem implementira vse varnostne zahteve za zaupnost. Vse aplikacije namreč ne potrebujejo uporabnikov, ki bi zapisovali podatke, za tiste, ki pa jih potrebujejo, pa mora sistem omogočati izvedbo večkratnih posodobitev pravočasno in na varen način. To je velika slabost OLAP izdelkov, ki predvidevajo, da so OLAP aplikacije samo za branje. Tudi izdelki, ki omogočajo več uporabnikov, zapisovanje in branje, imajo slab varnostni model.

Večdimenzionalnost je ključna zahteva. Sistem mora zagotavljati večdimenzionalni pogled podatkov, kar vključuje tudi celotno podporo hierarhiji, saj je le-to najbolj zanesljiv način za analizo poslovanja in organizacije.

Informacije so vsi podatki, ki jih potrebujemo in so pomembni za aplikacijo. Pomembna pri OLAP produktih je količina podatkov, ki jih lahko vnesemo. Pri tem pa so pomembni tudi drugi dejavniki, kot so podvajanje podatkov, zahteva bralno-pisalnega pomnilnika, velikost diska, zmogljivost ter integracija s podatkovnim skladiščem.

Podatkovna skladišča so zgrajena na tehnologiji relacijskih podatkovnih baz. Zato se uporabijo OLAP orodja za neposredno interakcijo z relacijskim podatkovnim skladiščem ali podatkovnim središčem (Nagabhushana, 2006). Relacijski OLAP oziroma ROLAP orodja prepoznajo relacijsko podatkovno bazo (Slika 2.9), vendar pa uporabniku predstavijo večdimenzionalen pogled na podatke.

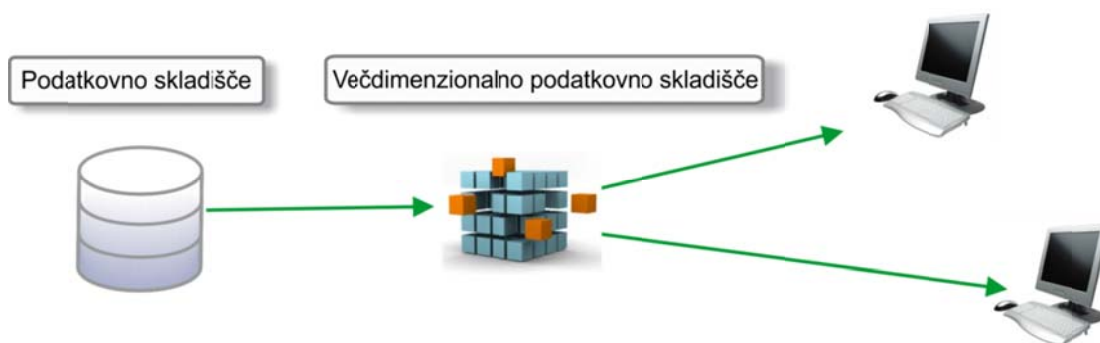
Slika 2.9: Relacijska podatkovna baza



Vir: Povzeto po Nagabhushana (2006, str. 81)

V svetu podatkovnih skladišč dosega poslovna analiza veliko vrednost, če je možno delati z večdimenzionalnimi podatkovnimi bazami (Slika 2.10). Takšnim okoljem pravimo večdimenzionalni OLAP oziroma s kratico MOLAP (Hammergren in Simon, 2009).

Slika 2.10: Večdimenzionalna podatkovna baza



Vir: Povzeto po Nagabhushana (2006, str. 81)

2.5 ETL proces

Poleg relacijske zbirke podatkov podatkovno skladišče vključuje ekstrakcijo, transformacijo in nalaganje (ETL), OLAP, orodja za analizo strank in druge aplikacije, ki upravljajo proces zbiranja in dostavljanja podatkov (Venatesen, 2007).

ETL je proces, ki pridobiva in pretvarja podatke iz izvornega sistema in pošilja podatke v podatkovno skladišče (Rainardi, 2008).

Poznamo več pristopov k implementaciji ETL. Tradicionalen pristop temelji na pridobivanju podatkov iz izvornih sistemov, nato ga prestavi v vmesno območje, za tem sledi transformacija in nalaganje v podatkovno skladišče. Alternativno lahko ETL namesto prenosa podatkov v vmesno območje le-te shrani v spomin in ga zatem prenese direktno v podatkovno skladišče.

Pri izbiri prenosa podatkov iz izvornega sistema se pri transformaciji pojavita dve možnosti (Rainardi, 2008):

- prenesemo podatke iz izvornega sistema, jih naložimo v podatkovno skladišče in šele nato izvedemo transformacijo (s posodobitvijo podatkov v podatkovnem skladišču),
- prenesemo podatke iz izvornega sistema v podatkovno skladišče in tukaj izvedemo transformacijo.

Pri prenosu podatkov iz izvornega sistema, lahko kategoriziramo štiri pristope (Rainardi, 2008):

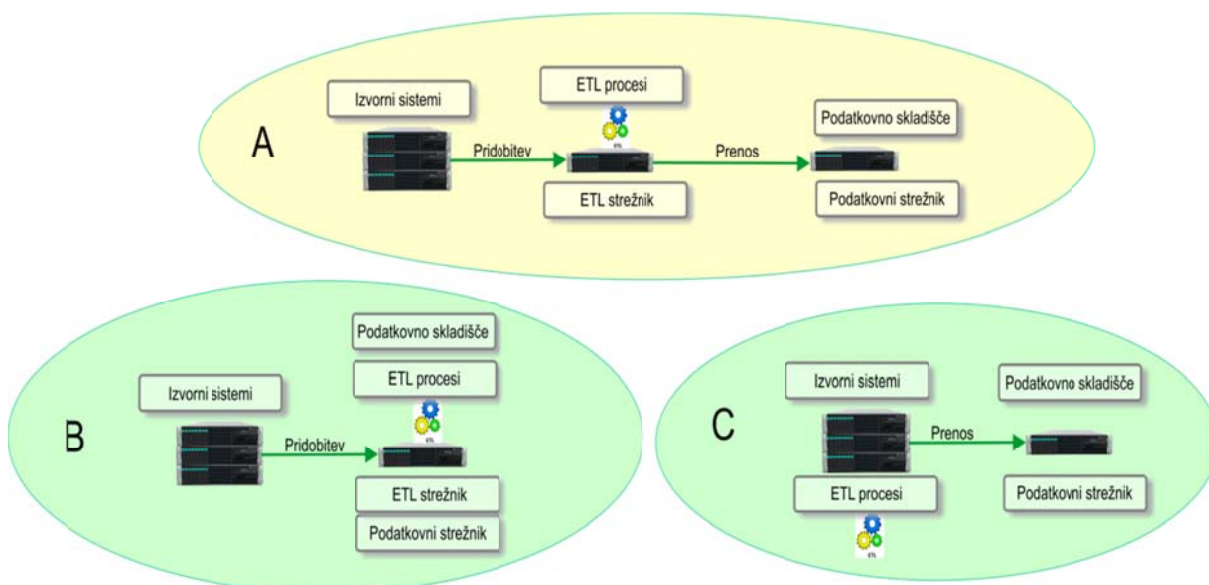
1. ETL proces prenese podatke s pogostimi poizvedbami. Ta pristop je najbolj pogost. ETL se poveže v podatkovno bazo izvornega sistema, izvede poizvedbe in prebere podatke.
2. Prožilci na podatkovni bazi izvornega sistema prenesejo spremembe v podatkih. Prožilec v podatkovni bazi je skupek SQL stavkov, ki se izvajajo vsakič ob izvedbi vstavljanja, posodobitve ali brisanja podatkov v tabeli.
3. Planiran proces na izvornem sistemu redno izvaža podatke. Ta pristop je zelo podoben prvemu, vendar pa program, ki izvaja poizvedbe v podatkovni bazi, ni zunanji ETL program.

4. Bralniki transakcijskega dnevnika (ang. »log«) berejo transakcijski dnevnik podatkovne baze in beležijo spremembe. V transakcijskem dnevniku se beležijo vse spremembe, ki so bile narejene v podatkovni bazi.

Na Sliki 2.11 so prikazani trije pristopi, kje se izvedejo ETL procesi, ki izvlečejo podatke iz izvornega sistema (Rainardi, 2008):

- A. Izvedba ETL procesov na ločenem ETL strežniku, ki se nahaja med izvornim sistemom in podatkovnim skladiščem. Takšen pristop zagotavlja najvišjo učinkovitost. ETL teče na svojem strežniku, tako da ne izkorišča virov podatkovnega skladišča.
- B. Izvedba ETL procesov na strežniku podatkovnega skladišča. Takšen pristop lahko uporabimo, če imamo dodatno kapaciteto strežnika podatkovnega skladišča ali če imamo časovni prožilec, kadar podatkovno skladišče ni v uporabi (na primer zvečer). Takšen pristop je cenejši, saj ne potrebujemo dodatnih strežnikov.
- C. Izvedba ETL procesov na strežniku, ki gosti izvorni sistem. Takšen pristop je implementiran kadar potrebujemo podatke v realnem času. To pomeni, da se v trenutku, ko se podatki na izvornem sistemu spremenijo, ta sprememba izvede tudi v podatkovnem skladišču.

Slika 2.11: Trije pristopi izvedbe ETL procesov

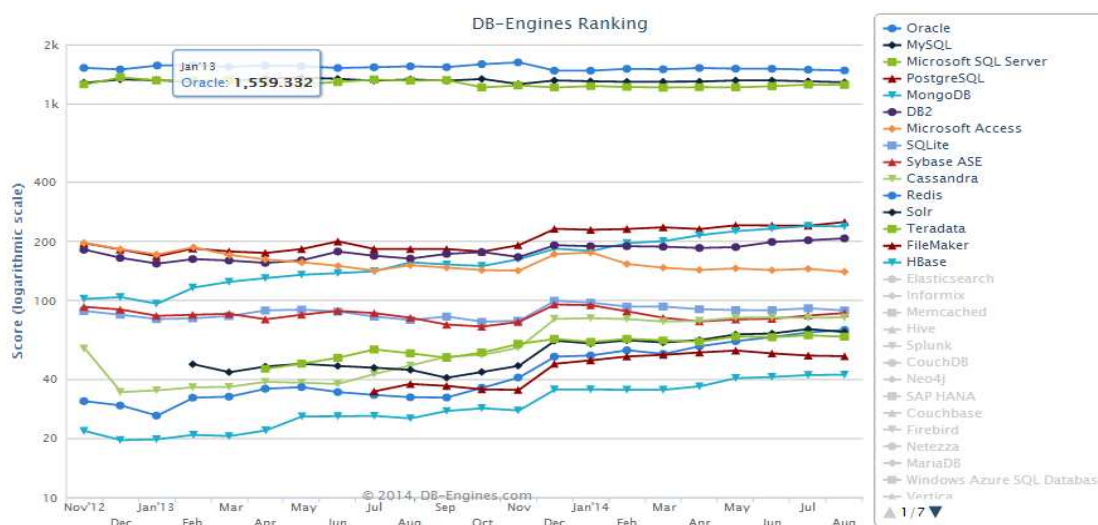


Vir: Povzeto po Rainardi (2008, str. 177)

3. OCENJEVANE IMPLEMENTACIJE PODATKOVNIH SKLADIŠČ

Pri migraciji podatkovne baze v MS Access-u v podatkovno skladišče smo na podlagi DB-Engines (DB-Engines, 2. julij 2014) izbrali šest najbolj priljubljenih podatkovnih baz (Slika 3.1), ki jih lahko uporabimo za namene podatkovnega skladiščenja.

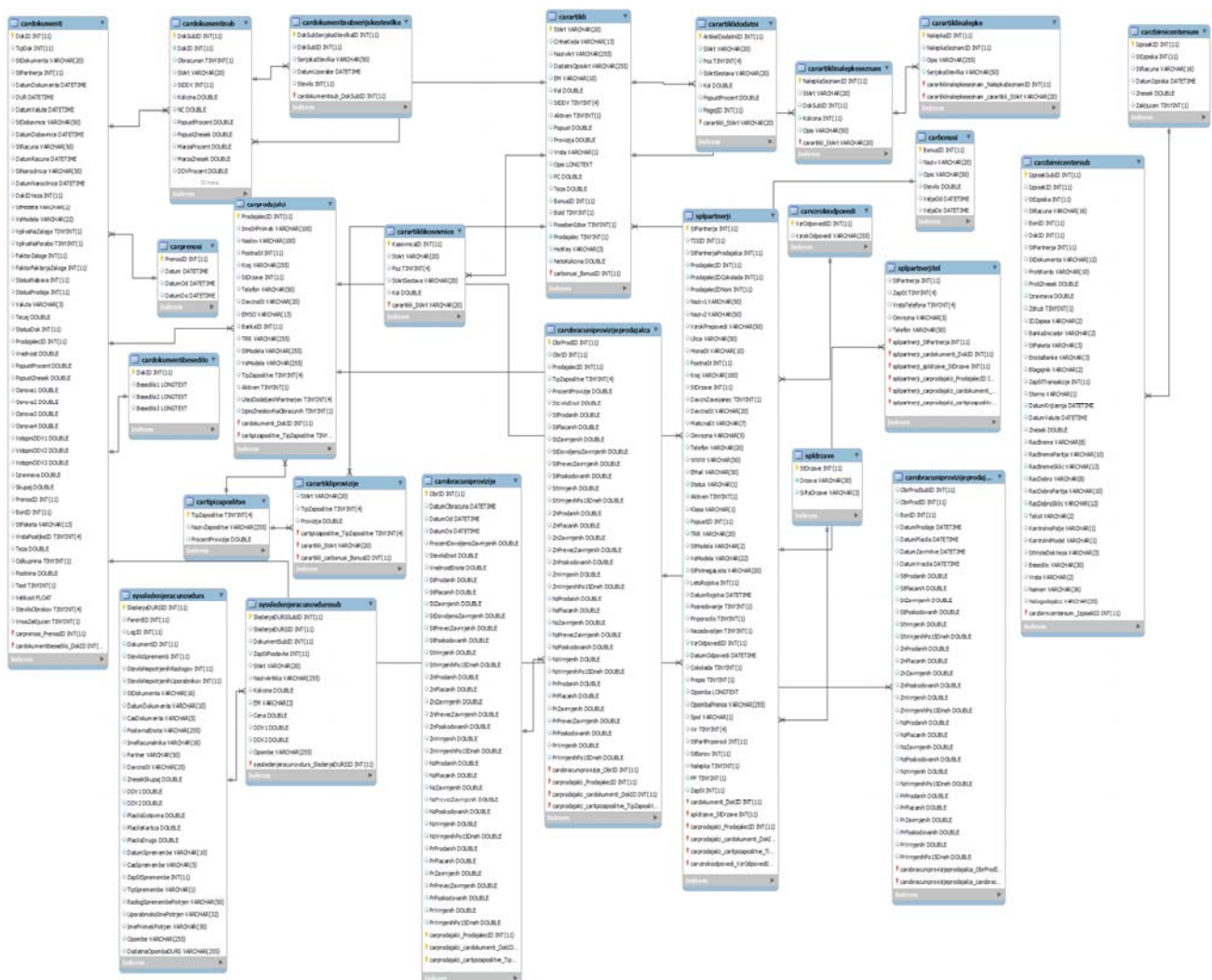
Slika 3.1: Priljubljene podatkovne baze



Vir: DB-Engines (2. julij 2014)

Struktura tabel, polj in relacij v MS Access podatkovni bazi, ki smo jih migrirali v omenjena podatkovna skladišča, je prikazana na Sliki 3.2.

Slika 3.2: Struktura podatkovne baze v MS Access



Vir: Klenovšek, lastni prikaz (2014)

Za namen migracije izvorne podatkovne baze v podatkovno skladišče smo postavili testno okolje na naslednji strojni in programski opremi:

- Procesor: AMD A-10 6790K z integrirano grafično kartico Radeon HD.
- RAM: 8 GB.
- Primarni disk: Kingston SSD 120 GB.
- Sekundarni diski: 2 x WD 1 TB.
- Operacijski sistem: Windows 8.1 Pro 64 bit.

Za potrebe vzpostavitve in testiranja posameznega podatkovnega skladišča so bila postavljena virtualna okolja v Oracle VM VirtualBox v. 4.3.12, z virtualnimi sistemi, kot je prikazano v Tabeli 3.1.

Tabela 3.1: Virtualna okolja z virtualnimi sistemi

Podatkovna baza	Virtualno testno okolje
MySQL v. 5.6.19.0	Virtualna strojna oprema: 2x CPU RAM 4 GB Virtualni HD: 25 GB SATA port Video memory 256 MB LAN: Intel PRO/1000 MT Desktop (NAT) OS: Win 7 pro MySQL Workbench
PostgreSQL v. 9.3.5.1	Virtualna strojna oprema: 2x CPU RAM 4 GB Virtualni HD: 25 GB SATA port Video memory 256 MB LAN: Intel PRO/1000 MT Desktop (NAT) OS: Win 7 pro PGAdmin3
MongoDB	Virtualna strojna oprema:

<p>v. 2.6.3</p>	<p>2x CPU</p> <p>RAM 4 GB</p> <p>Virtualni HD: 25 GB SATA port</p> <p>Video memory 256 MB</p> <p>LAN: Intel PRO/1000 MT Desktop (NAT)</p> <p>OS: Win 7 pro</p> <p>MongoDB Shell</p>
<p>MS SQL Server 2014</p> <p>v.12.0.2000.8</p>	<p>Virtualna strojna oprema:</p> <p>2x CPU</p> <p>RAM 4 GB</p> <p>Virtualni HD: 25 GB SATA port</p> <p>Video memory 256 MB</p> <p>LAN: Intel PRO/1000 MT Desktop (NAT)</p> <p>OS: Win Server 2012r2</p> <p>MS SQL Management Studio 12.0.2000.8</p>
<p>Oracle 11g R2</p> <p>v.11.2.0.3.0</p>	<p>Virtualna strojna oprema:</p> <p>2x CPU</p> <p>RAM 4 GB</p> <p>Virtualni HD: 25 GB SATA port</p> <p>Video memory 256 MB</p> <p>LAN: Intel PRO/1000 MT Desktop (NAT)</p>

	OS: Win Server 2008r2 Oracle SQL Developer 4.0.2
IBM DB2 v.10.5	Virtualna strojna oprema: 2x CPU RAM 4 GB Virtualni HD: 25 GB SATA port Video memory 256 MB LAN: Intel PRO/1000 MT Desktop (NAT) OS: Win Server 2008r2 IBM Data Studio 4.1.0.1

Vir: Klenovšek, lastni prikaz (2014)

3.1 MySQL

MySQL je najbolj priljubljena odprtokodna podatkovna baza, ki so jo uporabniki z uradne spletne strani prenesli že več kot 100 milijon-krat. S svojo vrhunsko hitrostjo, zanesljivostjo in enostavno uporabo je MySQL postala najboljša izbira za različna podjetja, saj odpravlja glavne težave, ki so povezani z različnimi izpadi, vzdrževanjem in sodobnim upravljanjem (MySQL, 23. junij 2014).

MySQL je edinstveno zasnovana podatkovna baza za uporabo kot podatkovno skladišče ter tudi kot tradicionalno podatkovno skladišče, veliko zgodovinsko/arhivsko podatkovno skladišče ali podatkovno skladišče v realnem času (MySQL, 23. junij 2014).

MySQL ima naslednje lastnosti (MySQL AB, 2007):

- implementira particioniranje podatkov/indeksov (obseg, ključ, seznam, kompozitno delitev) v verziji 5.1 in novejših,

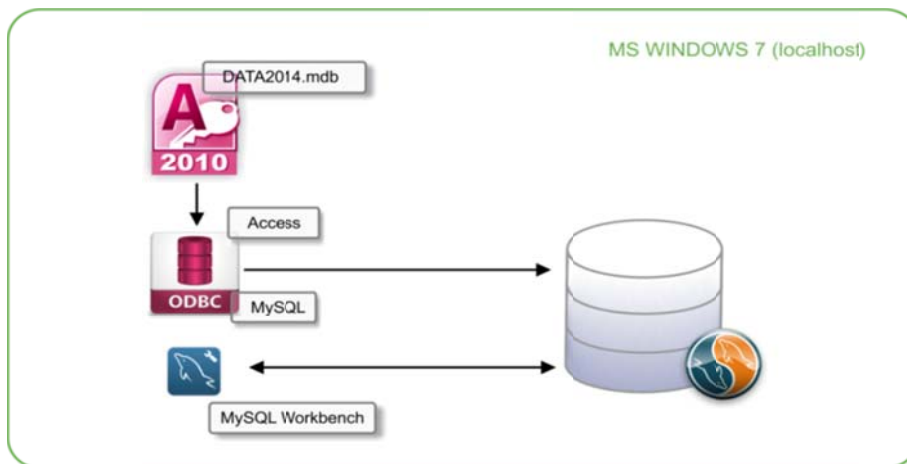
- nima omejitev za shranjevanje (primer: velikost 1 tabele je lahko velika 110 terabajtov (TB)),
- omogoča samodejno upravljanje shranjevanja,
- podpira vse podatkovne tipe za ANSI-SQL (tudi BLOB in XML),
- ima vgrajeno replikacijo, ki je preprosta in enostavna za namestitev,
- omogoča odlično podporo za indeksiranje (B-drevesa, gruča (ang. »cluster«), GIS),
- podpira predhodno nalaganje podatkov v predpomnilnik,
- ima unikaten predpomnilnik za poizvedbe, ki zajame rezultat poizvedbe kot tudi samo poizvedbo in ne samo podatke ter tako zagotavlja skoraj istočasni odziv vseh ponavljajočih se poizvedb, kot so poizvedbe, ki se uporabljajo v podatkovnem skladišču,
- omogoča vzporedno nalaganje podatkov, saj lahko naloži več datotek hkrati,
- podpira stiskanje podatkov, kar zagotavlja velike prihranke pri shranjevanju,
- omogoča tabele samo za banje (ang. »read-only tables«), s čimer lahko zaščitimo občutljive podatke,
- omogoča šifriranje za dodatno varstvo občutljivih podatkov.

Mnoge vodilne in hitrorastoče organizacije, kot so Yahoo!, Google, Nokia, YouTube, Wikipedia in Booking.com, uporabljajo MySQL, da prihranijo čas in denar (MySQL, 23. junij 2014).

V podjetju Car marketing se odločajo za celovito prenovitev informacijskega sistema in posodobitev oziroma vpeljavo sistema, ki bo omogočal večjo preglednost ter pregled različnih analiz in poročil, ki bodo podjetju v pomoč za lažje nadaljnje odločanje pri poslovanju.

MySQL strežnik smo namestili na operacijski sistem Windows 7 (localhost) . Izvorno podatkovno bazo v MS Accessu smo prenesli v MySQL (Slika 3.3) in ga preizkusili kot podatkovno skladišče.

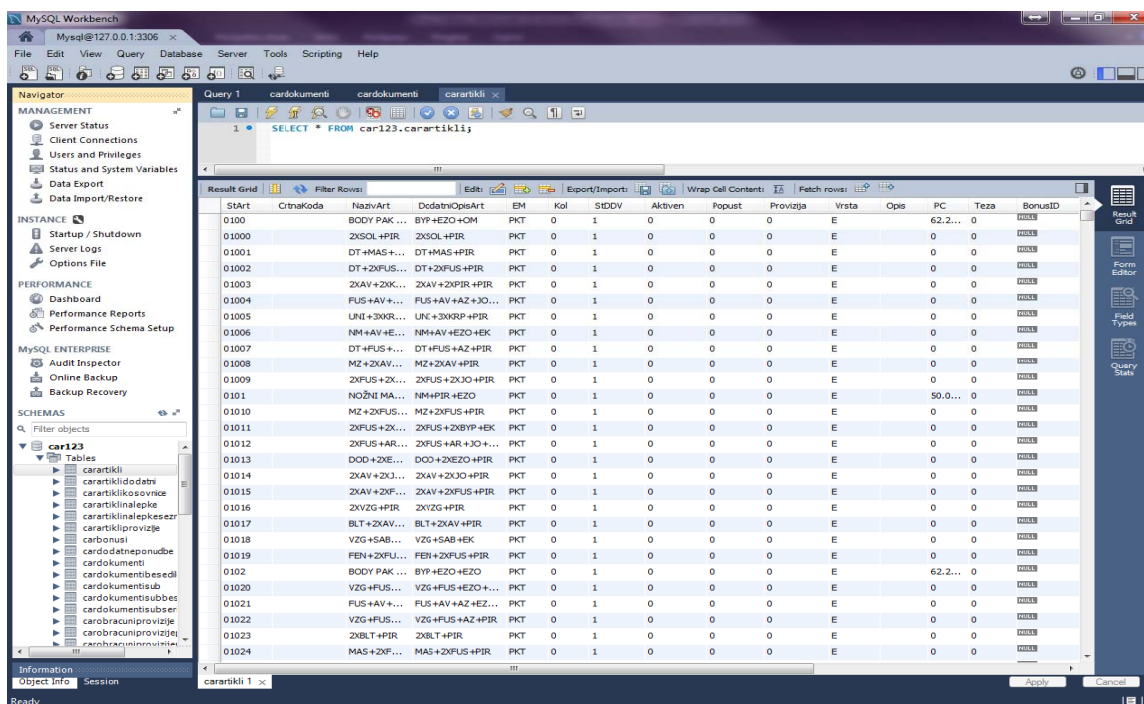
Slika 3.3: Postopek migracije v MySQL



Vir: Klenovšek, lastni prikaz (2014)

Podatke smo iz MS Accessa 2010 prenesli s pomočjo ODBC (ang. »*Open DataBase Connectivity*«) podatkovnega vira, ki smo ga nastavili za izvirno podatkovno bazo, kot tudi podatkovno bazo za prenos v MySQL (Slika 3.4). Pri migraciji nismo imeli večjih težav, pojavila se je težava le pri prenosu nekaterih relacij, ki smo jih nato ročno dodali v podatkovni bazi MySQL.

Slika 3.4: Migrirana podatkovna baza v MySQL



Vir: Klenovšek, lastni prikaz (2014)

3.2 PostgreSQL

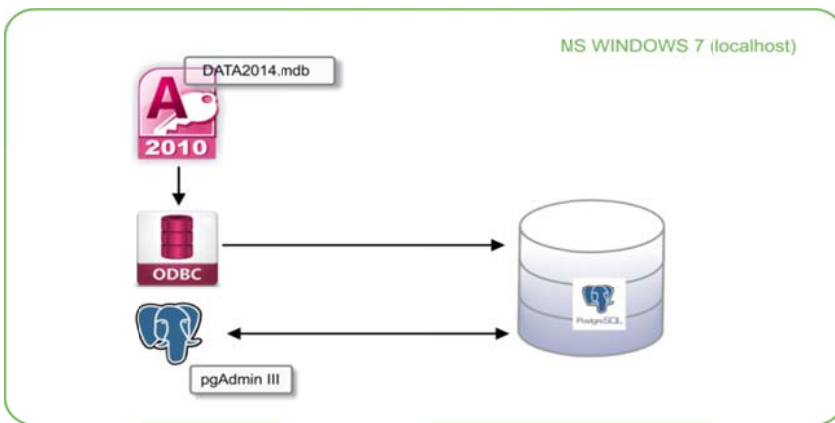
PostgreSQL je zmogljiv odprtokodni objektno-relacijsko orientiran podatkovni sistem. Slovi po zanesljivosti, integriteti podatkov in korektnosti. Deluje na vseh večjih operacijskih sistemih, vključno z Linux, Unix in Windows. Ima polno podporo za tuje ključne, poglede, prožilce in shranjene procedure (v več različnih jezikih, kot so Java, Perl, Python, C/C++ in PL/pgSQL). V standardni knjižnici vključuje ogromno vgrajenih funkcij, vse od osnovnih matematičnih operacij do kriptografske in Oracle kompatibilnosti (PostgreSQL, 2. julij 2014).

PostgreSQL ima naslednje lastnosti (PostgreSQL, 2. julij 2014):

- omogoča integracijo z različnimi drugimi orodji in aplikacijami za OLAP in izvajanje poročil,
- podpira tabele in poglede samo za banje (ang. »*read-only*«), s čimer lahko zaščitimo občutljive podatke,
- omogoča organizacijo objektov podatkovne baze v logične skupine,
- nima omejitev za shranjevanje (primer: velikost 1 tabele je lahko velika 32 terabajtov (TB)),
- vsebuje možnost indeksiranja kot je generalizirano iskalno drevo (ang. »*GiST – Generalized Search Tree*«), ki združuje različne postopke sortiranih algoritmov kot so B-drevo, B+ drevo, R-drevo, delna vsota dreves in mnogo drugih.

Za migracijo smo PostgreSQL strežnik postavili na operacijski sistem Windows 7 (localhost). Pri migraciji izvorne podatkovne baze v PostgreSQL (Slika 3.5) smo uporabili ODBC podatkovni vir, ki smo ga nastavili za PostgreSQL.

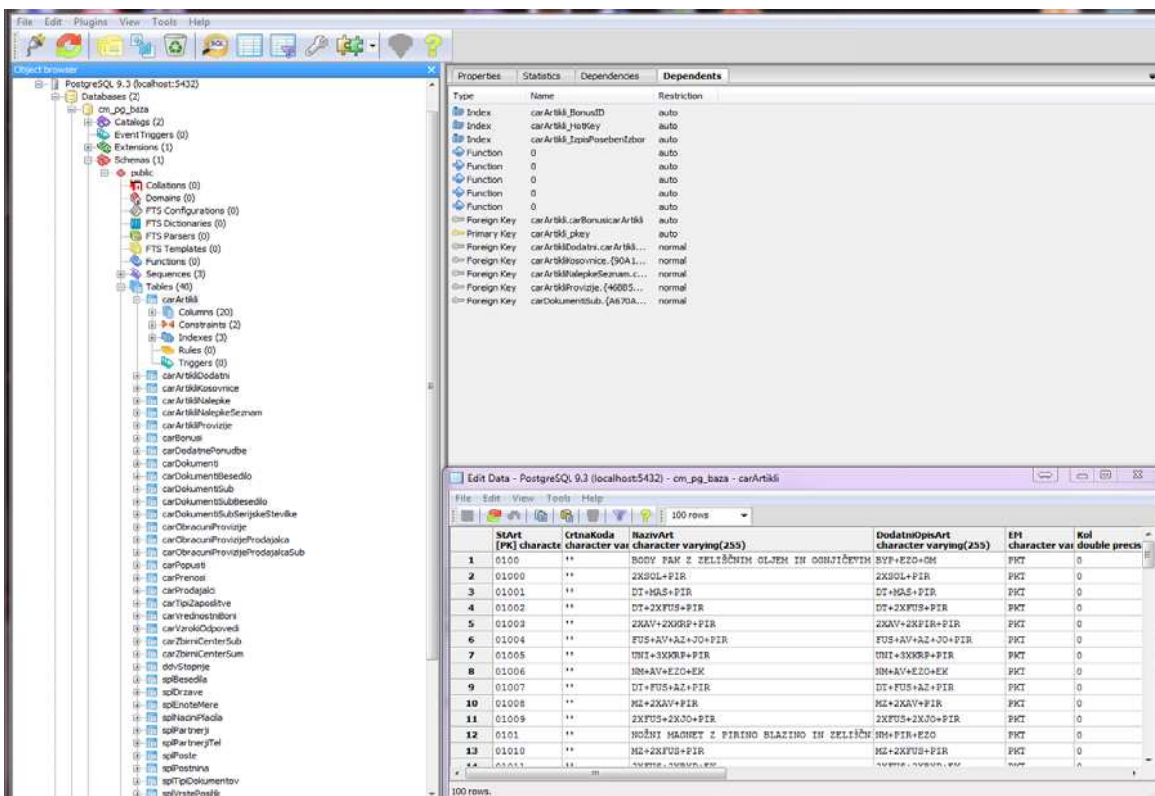
Slika 3.5: Postopek migracije v PostgreSQL



Vir: Klenovšek, lastni prikaz (2014)

Pri vzpostavitvi samega delovnega okolja nismo imeli težav, tudi pri migraciji so se v PostgreSQL (Slika 3.6) prenesli vsi elementi izvirne podatkovne baze (tabele, polja, tipi polj, privzete vrednosti, relacije). PgAdmin III nam omogoča celovit pregled nad celotno strukturo podatkovne baze, relacijami, kot tudi podatki, ki so se pri prenosu v celoti uspešno prenesli.

Slika 3.6: Migrirana podatkovna baza v PostgreSQL



Vir: Klenovšek, lastni prikaz (2014)

3.3 MongoDB

MongoDB je dokumentna (ne SQL) podatkovna baza, ki zagotavlja visoko zmogljivost, razpoložljivost in enostavno skalabilnost (MongoDB, 5. julij 2014).

MongoDB poleg kreiranja, branja, posodobitve in brisanja podatkov zagotavlja sledeče lastnosti (Chodorow, 2013):

- indeksiranje – podpira generične sekundarne indekse, kar omogoča hitrejše poizvedbe,
- združevanje – MongoDB omogoča gradnjo kompleksnega združevanja,
- posebno vrsto zbirk – podpora zbirk, ki imajo določen čas trajanja (primer seja), podpira fiksno velikost zbirk, ki so koristne pri datotekah s trenutnimi podatki (primer dnevnik (ang. »logs«)),
- shranjevanje datotek – podpora enostavnega protokola za shranjevanje velikih datotek,
- na enem MongoDB strežniku lahko gostimo več neodvisnih podatkovnih baz, kjer ima lahko vsaka svoje zbirke in dovoljenja,
- vsak dokument ima svoj poseben ključ, »_id«, ki je edinstven skozi celotno dokumentno zbirko.

Kot že omenjeno, je MongoDB dokumentna podatkovna baza, to pomeni, da se podatki shranjujejo kot dokumenti, kar lahko primerjamo z na primer JavaScript objekti. Dokumenti so osnovna enota podatkov, kar približno ustreza vrstici v relacijski podatkovni bazi (MongoDB, 5. julij 2014). Način vstavljanja podatkov v podatkovno bazo MongoDB je prikazan na Sliki 3.7.

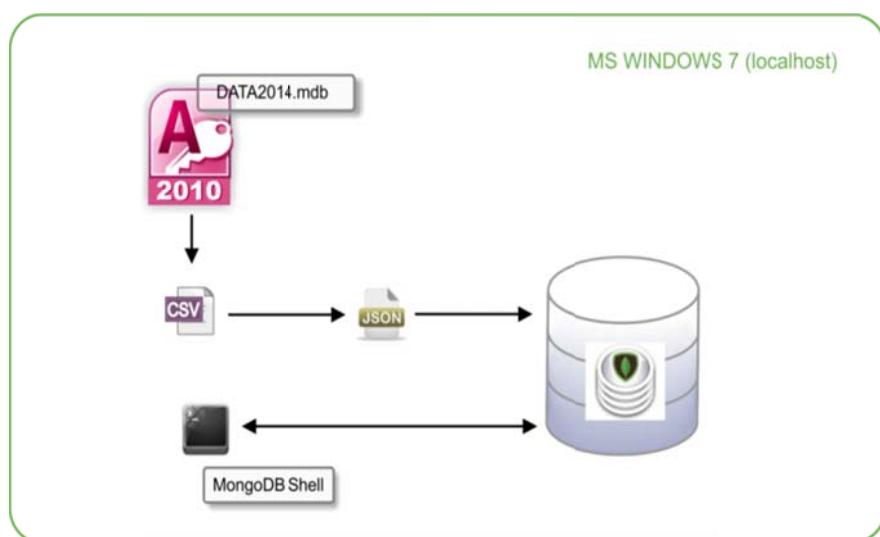
Slika 3.7: Primer vstavljanja podatkov v MongoDB

```
{
  "StArt":100,
  "CrtnaKoda":"","
  "NazivArt": "BODY PAK Z ZELISCNIM OLJEM IN OGNJICEVIM MAZILOM",
  "DodatniOpisArt": "BYP+EZO+OM",
  "EM": "PKT",
  "Kol": "0,00",
  "StDDV": "1",
  "Aktiven": "0",
  "Popust": "0,00",
  "Provizija": "0,00",
  "Vrsta": "E",
  "Opis": null,
  "PC": "62,24",
  "Teza": "0,00",
  "BonusID": "0",
  "Bold": "0",
  "PosebenIzbor": "0",
  "Prodajalec": "0,00",
  "HotKey": "",
  "Netokolicina": ""
}
```

Vir: Klenovšek, lastni prikaz (2014)

Za migracijo smo MongoDB strežnik postavili na operacijski sistem Windows 7 (localhost). Pri migraciji izvorne podatkovne baze v MongoDB je bilo potrebno celotno vsebino prenesti v tekstovno datoteko .csv, le-to pa prilagoditi in pretvoriti v ustrežno strukturo (Slika 3.8) in shraniti v .json datoteko. Pri tem smo morali prilagoditi vse relacije, vrednosti in tipe polj, da so ustrezali BSON zapisu. BSON je binarni format, ki se uporablja za shranjevanje dokumentov in omogoča klice oddaljenih procedur v MongoDB (MongoDB, 5. julij 2014). Prav tako smo imeli težave pri sami migraciji podatkov, saj se podatki numerične vrednosti, kot je na primer šifra, ki je v izvorni datoteki kot tekstovno polje, niso mogli uspešno prenesti, če je bil prvi znak 0 (nič), tako da je bil potreben popravek podatkov. Celoten postopek ustvarjanje podatkovne baze, prenosa podatkov iz izvorne podatkovne baze v MongoDB in poizvedb poteka prek MongoDB Shell ukazne vrstice.

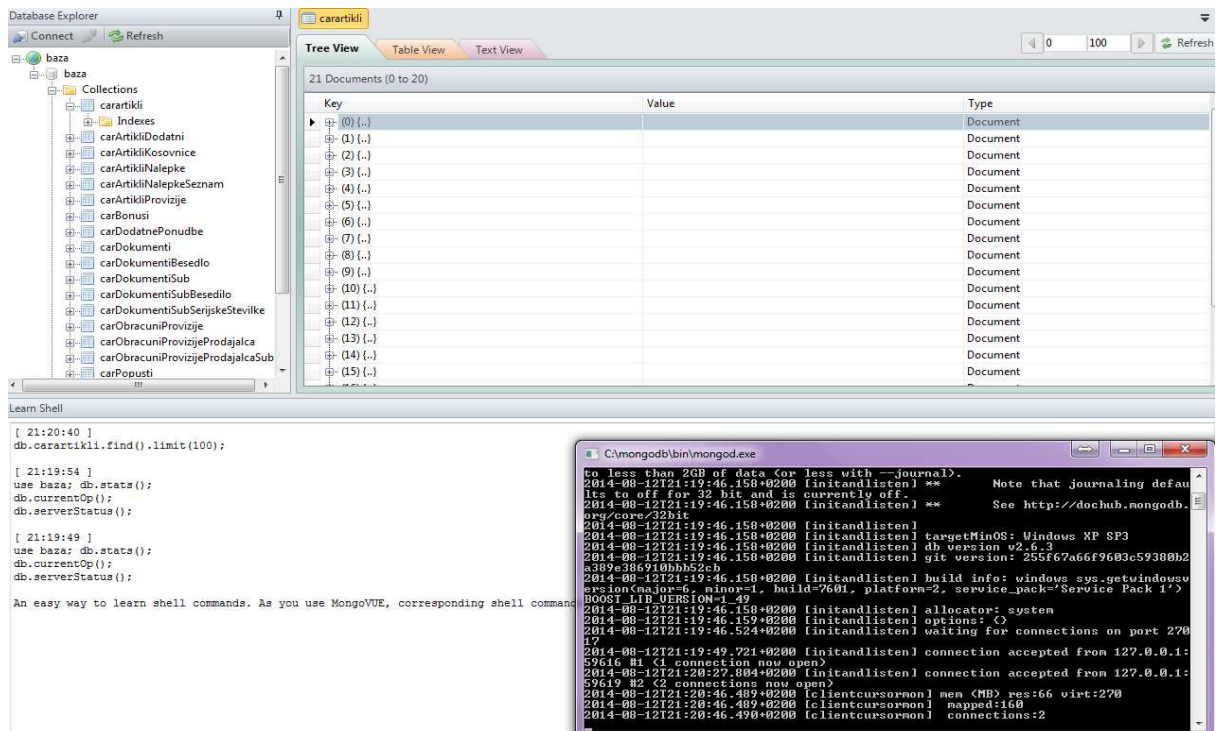
Slika 3.8: Postopek migracije v MongoDB



Vir: Klenovšek, lastni prikaz (2014)

Za pregled podatkov in izvajanje poizvedb, smo uporabili zunanje orodje MongoVUE, ki je aplikacija z enostavnim in uporabnim grafičnim vmesnikom in omogoča delo z MongoDB (Slika 3.9).

Slika 3.9: Migrirana podatkovna baza v MongoDB



Vir: Klenovšek, lastni prikaz (2014)

Na Sliki 3.10 je prikazan primer pregleda vpisanega podatka v podatkovno bazo MongoDB.

Slika 3.10: Primer prikaza podatkov v MongoDB

Key	Value	Type
(0) {...}		Document
_id	53e913bf868e5e105887d655	ObjectId
StArt	100	Int32
CrtnaKoda	0	Int32
NazivArt	BODY PAK Z ZELISNIM OLJEM IN OGNJICEVIM MAZILOM	String
DodatniOpisArt	BYP+EZO+OM	String
EM	PKT	String
Kol	0,00	String
StDDV	1	Int32
Aktiven	0	String
Popust	0,00	String
Provizija	0,00	String
Vrsta	E	String
Opis	brez opisa	String
PC	62.24	String
Teza	0,00	String
BonusID		Null
Bold	0	String
PosebenZhor	0	String

Vir: Klenovšek, lastni prikaz (2014)

3.4 Microsoft SQL Server

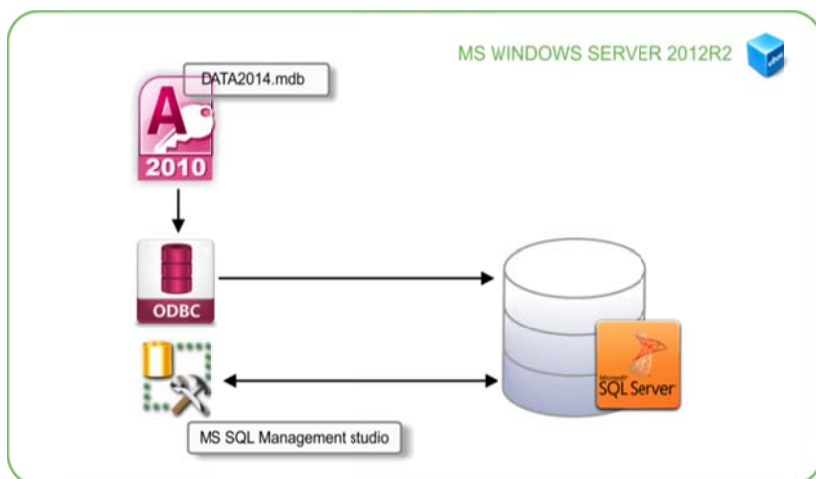
Microsoft SQL Server je uporabniku prijazno in učinkovito orodje. Za obvladovanje Microsoft SQL Server-ja je potreben čas in trud, vendar pa je že velika večina vsebovanih orodij zasnovana tako, da so preprosta za uporabo (Root in Caryn, 2012).

Microsoft SQL Server ima naslednje lastnosti (SQL Server 2012 – 2014, 26. junij 2014):

- dobro izrablja pomnilniško zmogljivost za pospešitev hitrosti delovanja ter s tem hitrejšo obdelavo transakcij in delo s podatkovnim skladiščem,
- omogoča visoko razpoložljivost in obnovitev po nesreči (ang. »*disaster recovery*«),
- omogoča varnost in skladnost, saj pomaga varovati podatke s šifriranjem, razširjenim upravljanjem s ključi in šifriranimi varnostnimi kopijami,
- omogoča merjenje modelov poslovne inteligence, pri tem pa zagotavlja kakovost in natančnost analitike,
- podpira skalabilnost podatkovnega skladišča in omogoča integracijo z ne-relacijskimi viri, kot na primer Hadoop,
- omogoča integracijo storitev, ki vključujejo podporo za storitve ekstrakcije, transformacije in nalaganja (ETL),
- ponuja orodja za enostavno upravljanje z infrastrukturo podatkovne baze,
- omogoča robustna razvojna orodja.

Microsoft Access omogoča enostavno združljivost z MS SQL, zato pri migraciji izvorne podatkovne baze nismo imeli večjih težav (Slika 3.11). Za migracijo smo postavili virtualno okolje MS Windows Server 2012R2, na katerega smo namestili MS SQL Server 2014. Pred migracijo so bili potrebni nekateri popravki elementov izvorne podatkovne baze, kot so imena tabel, ki ne smejo biti daljša od 30 znakov ter popravek polj v izvorni datoteki iz tipa Date/Time na tip Text za ustrezno pretvorbo v tip Varchar v MS SQL podatkovni bazi. Pri izvozu smo morali paziti na odvisnosti med tabelami, kjer smo najprej prenesli tabele brez tujih ključev, nato tabele odvisnosti.

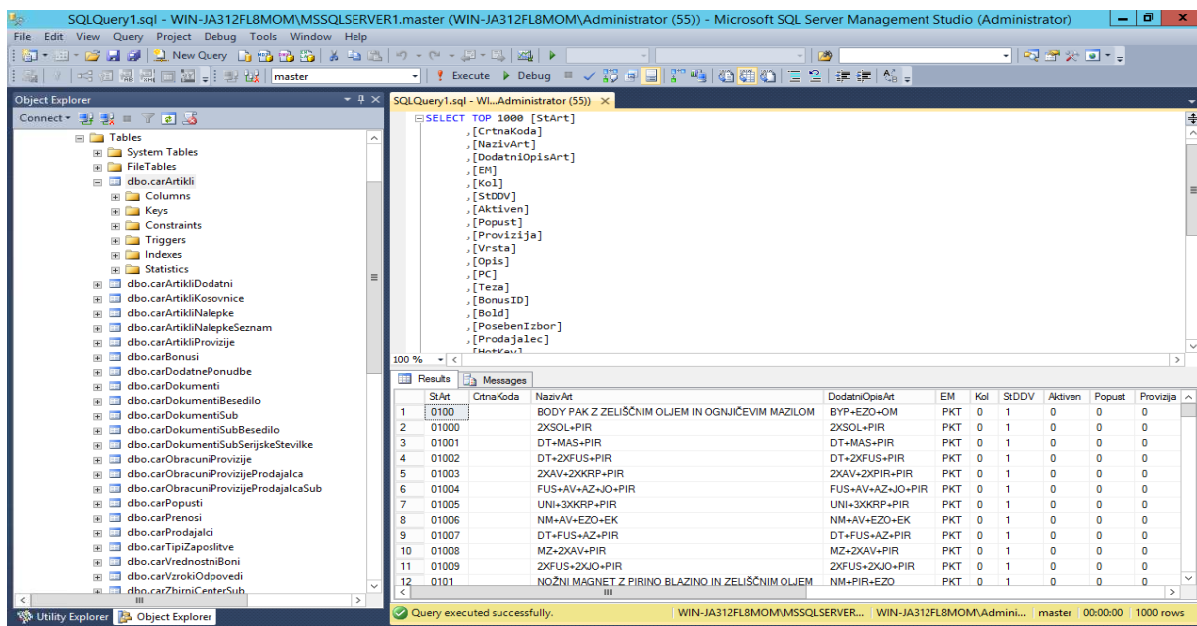
Slika 3.11: Postopek migracije v MS SQL



Vir: Klenovšek, lastni prikaz (2014)

Celovit pregled nad migriranimi podatki v MS SQL nam nato omogoča MS SQL Server Management Studio (Slika 3.12).

Slika 3.12: Migrirana podatkovna baza v MS SQL



Vir: Klenovšek, lastni prikaz (2014)

3.5 Oracle

Oracle Warehouse Builder je orodje, ki ga lahko uporabimo v kateri koli fazi implementacije podatkovnega skladišča od začetka načrtovanja in izgradnje strukture tabel do procesa ETL in upravljanja kakovosti podatkov (Griesemer, 2011).

Oracle Warehouse Builder je celovito orodje, primerno za vse vidike povezovanja podatkov. Oracle Warehouse Builder uporablja Oracle podatkovno bazo za pretvarjanje podatkov v visoko kakovostne informacije. Zagotavlja kvaliteto podatkov, preverjanje podatkov, popolnoma integrirane relacijske in dimenzionalne modele in popolno upravljanje podatkov in meta podatkov (Oracle Warehouse Builder User's Guide, 5. julij 2014).

Oracle 11g ima naslednje lastnosti (Oracle Database, 5. julij 2014):

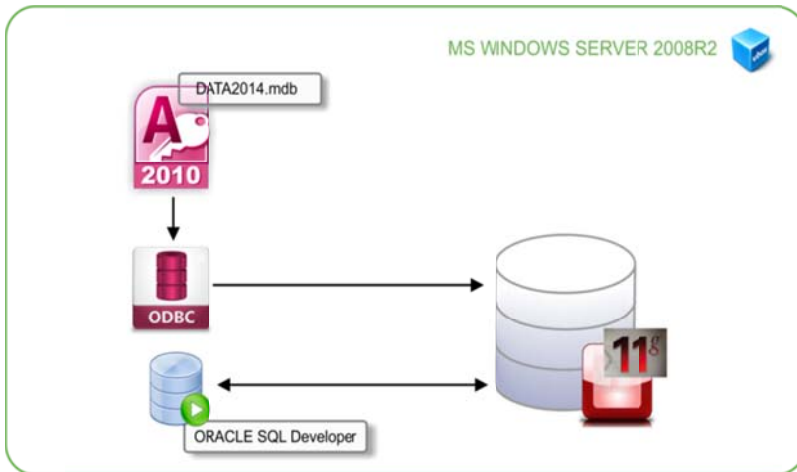
- omogoča različne analitične funkcije, ki omogočajo združevanje zmerjenih vrednosti znotraj skupine,
- omogoča pridobivanje datotek iz zunanjih tabel, pri čemer lahko določimo kateri koli program, ki se bo izvedel in obdelal datoteke ter omogočil Oracle podatkovni bazi uporabo izhodne datoteke tega programa,
- vsebuje OLAP orodja,
- omogoča izdelavo dimenzijskih podatkovnih modelov,
- podpira napredne poizvedbe.

Podatkovno skladišče Oracle prav tako dobro služi za namene poslovne inteligence. Relacijske podatke je mogoče prikazati kot večdimenzionalno matriko, pri kateri lahko uporabljamo različne kalkulacije v obliki preglednice.

Eden izmed težjih elementov pri reševanju težav je le-te reproducirati in zajeti dogajanje v sistemu, ko se je težava pojavila. Oracle podatkovna baza omogoča statistiko zmogljivosti, ki predstavlja stanje sistema, le-ta pa se lahko samodejno zajema in shranjuje v podatkovno bazo (Hobbs, 2005).

Oracle 11g smo namestili na virtualno okolje Windows Server 2008R2 (Slika 3.13), kjer smo podatke iz izvorne podatkovne baze prenesli preko ODBC podatkovnega vira. Pri prenosu podatkov so bili vsi podatki uspešno preneseni, pri tem pa ni bilo težav s strukturo izvorne podatkovne baze.

Slika 3.13: Postopek migracije v Oracle 11g R2



Vir: Klenovšek, lastni prikaz (2014)

Pregled prenesenih podatkov nam omogoča Oracle SQL Developer (Slika 3.14), s pomočjo katerega smo lahko izvajali poizvedbe in pridobili OLAP poročila.

Slika 3.14: Migrirana podatkovna baza v Oracle 11g

The screenshot shows the Oracle SQL Developer interface. The main window displays a table with the following columns: StArt, OtnaKoda, NazivArt, DodadniOpisArt, EM, Kol, STDOV, Aktiven, Popust, Provizija, Vrsta, Opis, and PC. The table contains 27 rows of data, including items like 'BODY PAK Z ZELIŠČNIM OLJEM IIN OGNJIČEVIM MAZILOM' and 'NOŽNI MAGNET Z PIRINO BLAZINO IN ZELIŠČNIM OLJEM'.

StArt	OtnaKoda	NazivArt	DodadniOpisArt	EM	Kol	STDOV	Aktiven	Popust	Provizija	Vrsta	Opis	PC
1	0100	(null)	BODY PAK Z ZELIŠČNIM OLJEM IIN OGNJIČEVIM MAZILOM	BYP+EZO+GM	PKT	0.01	false	0.0	0.0	E	(null)	62.247471345
2	0100	(null)	2XSOL+PIR	2XSOL+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
3	0101	(null)	DI+MAS+PIR	DI+MAS+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
4	0102	(null)	DI+2XFUS+PIR	DI+2XFUS+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
5	0103	(null)	2XAV+2XKRP+PIR	2XAV+2XKRP+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
6	0104	(null)	FUS+AV+AZ+JO+PIR	FUS+AV+AZ+JO+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
7	0105	(null)	UNI+3XKRP+PIR	UNI+3XKRP+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
8	0106	(null)	NM+AV+EZO+EK	NM+AV+EZO+EK	PKT	0.01	false	0.0	0.0	E	(null)	
9	0107	(null)	DI+FUS+AZ+PIR	DI+FUS+AZ+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
10	0108	(null)	MZ+2XAV+PIR	MZ+2XAV+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
11	0109	(null)	2XFUS+2XJO+PIR	2XFUS+2XJO+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
12	0101	(null)	NOŽNI MAGNET Z PIRINO BLAZINO IN ZELIŠČNIM OLJEM	NM+PIR+EZO	PKT	0.01	false	0.0	0.0	E	(null)	50.07509263
13	0101	(null)	MZ+2XFUS+PIR	MZ+2XFUS+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
14	0101	(null)	2XFUS+2XBYP+EK	2XFUS+2XBYP+EK	PKT	0.01	false	0.0	0.0	E	(null)	
15	0102	(null)	2XFUS+AR+JO+PIR	2XFUS+AR+JO+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
16	0103	(null)	DOD+2XEZO+PIR	DOD+2XEZO+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
17	0104	(null)	2XAV+2XJO+PIR	2XAV+2XJO+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
18	0105	(null)	2XAV+2XFUS+PIR	2XAV+2XFUS+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
19	0106	(null)	2XVZG+PIR	2XVZG+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
20	0107	(null)	BLI+2XAV+PIR	BLI+2XAV+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
21	0108	(null)	VZG+SAB+EK	VZG+SAB+EK	PKT	0.01	false	0.0	0.0	E	(null)	
22	0109	(null)	FEN+2XFUS+PIR	FEN+2XFUS+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
23	0102	(null)	BODY PAK IN 2 x ZELIŠČNO OLJE	BYP+EZO+EZO	PKT	0.01	false	0.0	0.0	E	(null)	62.247471345
24	0102	(null)	VZG+FUS+EZO+PIR	VZG+FUS+EZO+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
25	0102	(null)	FUS+AV+AZ+EZO+PIR	FUS+AV+AZ+EZO+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
26	0102	(null)	VZG+FUS+AZ+PIR	VZG+FUS+AZ+PIR	PKT	0.01	false	0.0	0.0	E	(null)	
27	0103	(null)	2XBLI+PIR	2XBLI+PIR	PKT	0.01	false	0.0	0.0	E	(null)	

Vir: Klenovšek, lastni prikaz (2014)

3.6 DB2

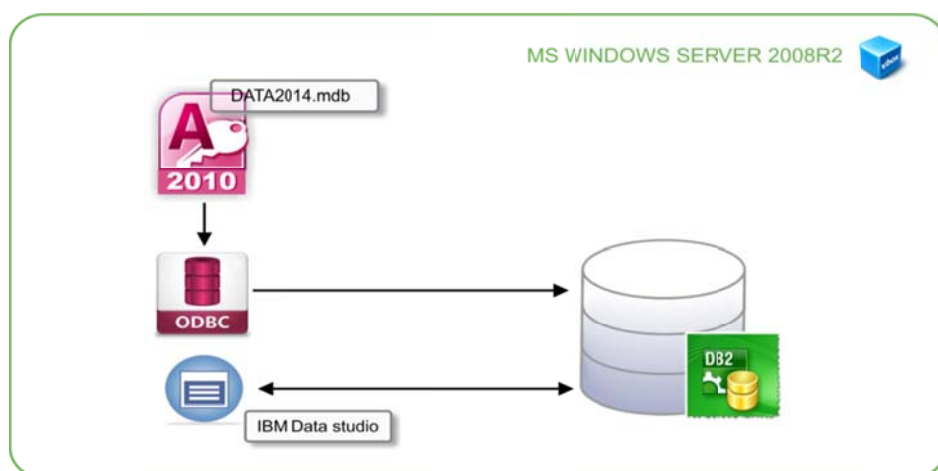
IBM DB2 prihaja iz družine sistemov za upravljanje relacijske podatkovne baze. DB2 vsebuje lastnosti, ki omogočajo analitiko za kakovostno sprejemanje odločitev, izboljšanje kakovosti, ter enostavno upravljanje podatkovne baze (DB2, 10. julij 2014).

IBM DB2 ima naslednje lastnosti (Gonzales, 2003 in DB2, 10. julij 2014):

- omogoča večdimenzionalno združevanje v gruče,
- omogoča posodobitve v realnem času,
- podpira dinamično upravljanje vmesnega pomnilnika,
- omogoča dinamično konfiguracijo baze podatkov,
- podpira izredno hitre analitične poizvedbe,
- vsebuje OLAP orodja,
- omogoča pregled podatkov za časovna obdobja v preteklosti.

IBM DB2 smo namestili na virtualno okolje Windows Server 2008R2 (Slika 3.15), kjer smo podatke iz izvorne podatkovne baze prenesli preko ODBC podatkovnega vira. Pri prenosu smo naleteli na težavo pri migraciji tabele splPatnerji, ki vsebuje 60.031 zapisov. Tako smo pri tej tabeli prenesli le polja, tipe in relacije, podatke pa smo iz izvorne tabele izvozili v datoteko .csv, le-to pa potem ročno uvozili v DB2. Drugih težav pa pri samem prenosu nismo imeli. Podatki so se pri drugih tabelah prenesli v celoti.

Slika 3.15: Postopek migracije v DB2



Vir: Klenovšek, lastni prikaz (2014)

Pregled prenesenih podatkov nam omogoča IBM Data Studio (Slika 3.16), s pomočjo katerega smo lahko izvajali poizvedbe.

Slika 3.16: Migrirana podatkovna baza v DB2

Start [VARCHAR(20)]	Grtnakoda [VARCHAR(13)]	NazivArt [VARCHAR(255)]	DodatniOpisArt [VARCHAR(255)]	EM [VARCHAR(10)]	Kol [DOUBLE]	SIDDV [SMALL]
0100		BODY PAK Z ZELIŠČNIM...	BYP+HEZO+OM	PKT	0.0	1
01000		ZXSOL+PIR	ZXSOL+PIR	PKT	0.0	1
01001		DT+MAS+PIR	DT+MAS+PIR	PKT	0.0	1
01002		DT+2XFUS+PIR	DT+2XFUS+PIR	PKT	0.0	1
01003		ZKAV+2XKRP+PIR	ZKAV+2XKRP+PIR	PKT	0.0	1
01004		FUS+AV+AZ+JO+PIR	FUS+AV+AZ+JO+PIR	PKT	0.0	1
01005		UNI+2XKRP+PIR	UNI+2XKRP+PIR	PKT	0.0	1
01006		NM+AV+HEZO+HEK	NM+AV+HEZO+HEK	PKT	0.0	1
01007		DT+FUS+AZ+PIR	DT+FUS+AZ+PIR	PKT	0.0	1
01008		MZ+2KAV+PIR	MZ+2KAV+PIR	PKT	0.0	1
01009		ZKFUS+2XJO+PIR	ZKFUS+2XJO+PIR	PKT	0.0	1
0101		NOŽNE MAGNET Z PIRI...	NM+PIR+HEZO	PKT	0.0	1
01010		MZ+2XFUS+PIR	MZ+2XFUS+PIR	PKT	0.0	1
01011		ZKFUS+2XBYP+HEK	ZKFUS+2XBYP+HEK	PKT	0.0	1
01012		ZKFUS+AR+JO+PIR	ZKFUS+AR+JO+PIR	PKT	0.0	1
01013		DOD+2KEZO+PIR	DOD+2KEZO+PIR	PKT	0.0	1
01014		ZKAV+2XJO+PIR	ZKAV+2XJO+PIR	PKT	0.0	1
01015		ZKAV+2XFUS+PIR	ZKAV+2XFUS+PIR	PKT	0.0	1
01016		ZKVGZ+PIR	ZKVGZ+PIR	PKT	0.0	1
01017		BLT+2KAV+PIR	BLT+2KAV+PIR	PKT	0.0	1
01018		VZG+SAB+HEK	VZG+SAB+HEK	PKT	0.0	1
01019		FEN+2XFUS+PIR	FEN+2XFUS+PIR	PKT	0.0	1
0102		BODY PAK IN 2 x ZELIŠ...	BYP+HEZO+HEZO	PKT	0.0	1
01020		VZG+FUS+HEZO+PIR	VZG+FUS+HEZO+PIR	PKT	0.0	1
01021		FUS+AV+AZ+HEZO+PIR	FUS+AV+AZ+HEZO+PIR	PKT	0.0	1
01022		VZG+FUS+AZ+PIR	VZG+FUS+AZ+PIR	PKT	0.0	1
01023		ZXBLT+PIR	ZXBLT+PIR	PKT	0.0	1
01024		MAS+2XFUS+PIR	MAS+2XFUS+PIR	PKT	0.0	1
01025		FEN+2XKRP+PIR	FEN+2XKRP+PIR	PKT	0.0	1
01026		DT+UNI+KRP+PIR	DT+UNI+KRP+PIR	PKT	0.0	1
01027		DOD+2XKRP+PIR	DOD+2XKRP+PIR	PKT	0.0	1
01028		DT+AV+AZ+PIR	DT+AV+AZ+PIR	PKT	0.0	1
01029		ZKFUS+AV+JO+PIR	ZKFUS+AV+JO+PIR	PKT	0.0	1
0103		BODY PAK Z OGNJIČEV...	BYP+OM+PIR	PKT	0.0	1
01030		SOL+2KAZ+HEK	SOL+2KAZ+HEK	PKT	0.0	1
01031		VZG+DT+PIR	VZG+DT+PIR	PKT	0.0	1
01032		DOD+AV+HEZO+HEK	DOD+AV+HEZO+HEK	PKT	0.0	1
01033		3XBYP+HEZO+HEK	3XBYP+HEZO+HEK	PKT	0.0	1
01034		FEN+ABC+AZ+PIR	FEN+ABC+AZ+PIR	PKT	0.0	1
01035		ZKAV+FUS+UNI+HEK	ZKAV+FUS+UNI+HEK	PKT	0.0	1
01036		SOL+2XBYP+PIR	SOL+2XBYP+PIR	PKT	0.0	1
01037		NM+2XFUS+PIR	NM+2XFUS+PIR	PKT	0.0	1
01038		SOL+2XJO+HEK	SOL+2XJO+HEK	PKT	0.0	1
01039		BLT+2XBYP+PIR	BLT+2XBYP+PIR	PKT	0.0	1

Vir: Klenovšek, lastni prikaz (2014)

4. PROCES ODLOČANJA

Vsakodnevno se srečujemo z različnimi odločitvami, ki tako ali drugače vplivajo na naše življenje. Ne glede na to, v kakšnem okolju ali situaciji se nahajamo, nenehno se odločamo za smeri, ki nas vodijo do zastavljenega cilja. Prave odločitve so pomemben dejavnik uspešnosti tako za posameznika kot tudi podjetje. Po besedah Ravindrahatha (2003, str. 3) so določitve velikokrat odvisne od različnih dejavnikov, kot so izkušnje, intuicija, ustvarjalne sposobnosti, presoja in notranji občutek.

Pri odločanju za izbiro najbolj ustreznega podatkovnega skladišča smo sledili fazam odločitvenega procesa, ki jih definiramo v nadaljevanju. S pomočjo določenih kriterijev smo podali končno odločitveno analizo izbire najbolj ustreznega podatkovnega skladišča.

4.1 Definicija odločanja

Odločanje je proces razreševanja problemov, kateremu sledi proces sprejemanja odločitve o rešitvi (Kralj in drugi, 2002). Pri procesu odločanja lahko sledimo osmim korakom, ki jih definirata Robbins in Decenzo (2001), pri čemer so glavni trije prepoznavanje problema, izbira rešitve in ocenjevanje uspešnosti izbrane rešitve.

Odločitve so na različnih nivojih postale veliko bolj kompleksne. Hitre spremembe in večja tveganja, s katerimi se srečujejo menedžerji v organizacijah, so dejavniki, ki stremijo k potrebi po programskih rešitvah sistemov za podporo odločanju (Power, 2009).

Kot v pomoč pri lažjemu odločanju poznamo različne programske rešitve sistemov za podporo odločanju, ki so poseben razred informacijskega sistema, ki podpira poslovne in organizacijske dejavnosti odločanja (Sauter, 2010).

4.2 Faze odločitvenega procesa

Pri odločanju je za učinkovito in najbolj optimalno izbiro potrebno upoštevati faze odločitvenega procesa. Praviloma proces poteka po sledečih fazah (Bohanec in drugi, 2003):

1. Identifikacija problema
Identificiramo naš problem, poskušamo določiti cilje in zahteve. V primeru, da je naš problem kompleksnejši, pa v proces vključimo tudi strokovnjake s področja oblikovanja odločitvenega modela.
2. Identifikacija kriterijev
Strukturiramo kriterije, na podlagi katerih smo ocenjevali variante. Izbrane kriterije smiselno hierarhično uredimo, tako dobimo drevo kriterijev, katerim določimo zalogo vrednosti.
3. Definicija funkcije koristnosti

Funkcija koristnosti določa vpliv kriterijev v drevesu. Funkcija je lahko preprostejša, kot je utežna vsota, ki jo bom uporabili tudi mi, ali pa bolj zahtevna, kot so: funkcija zvezne logike, funkcija na osnovi Bayesovega pravila, odločitvena pravila.

4. Opis variant

Vsako varianto, ki smo jo ocenjevali, v našem primeru so to ocenjevana podatkovna skladišča, opišemo z vrednostnimi osnovami kriterijev.

5. Vrednotenje in analiza variant

Pridobimo končno oceno variant na podlagi opisa določenih kriterijev. Postopek vrednotenja poteka od listov do korena drevesa kriterijev. Na podlagi pridobljene ocene naredimo analizo.

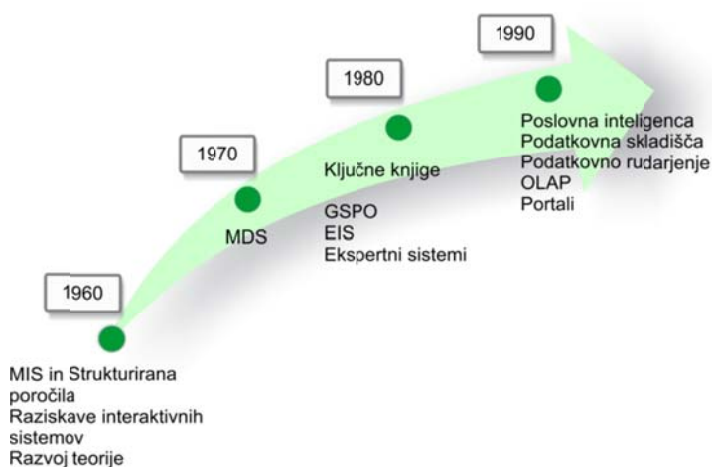
6. Odločitev

S podrobno analizo vrednotenja lahko podamo končno odločitev, katero podatkovno skladišče bo v največji meri zadostilo kriterijem.

4.3 Sistemi za podporo odločanju

Začetke sistemov za podporo odločanju zasledimo že v 60. letih prejšnjega stoletja, kjer so mnogi prodajalci, praktiki in akademiki spodbujali razvoj računalniško podprtih SPO. Z nadaljnjim razvojem (Slika 4.1) so se v kontekstu SPO razvila številka orodja, ki jih uporabljamo za podporo pri odločanju (Power, 2002).

Slika 4.1: Razvoj SPO koncepta



Vir: Povzeto po Power (2002, str. 5)

Številne akademske discipline zagotavljajo vsebinske temelje za razvoj in raziskave SPO. Raziskave s področja podatkovnih baz so prispevale k razvoju orodij in nadaljnjim raziskavam o upravljanju podatkov in dokumentov. Raziskave s področja upravljanja in operacijske raziskave so pripomogle k razvoju matematičnih modelov za uporabo modelov SPO in s tem potrdile prednosti takšnih modelov pri reševanju problemov odločanja. Kognitivna znanost, še posebej raziskave s področja vedenjskega odločanja, so zagotovile opisne in empirične podatke, ki so pripomogli pri oblikovanju SPO. Nekatera druga pomembna področja, povezana s SPO, so umetna inteligenca (AI), interakcija človek – računalnik (HCI), programski inženiring in telekomunikacij. Zgodovina SPO je relativno kratka, vendar se koncepti in tehnologije še vedno razvijajo (Power, 2002).

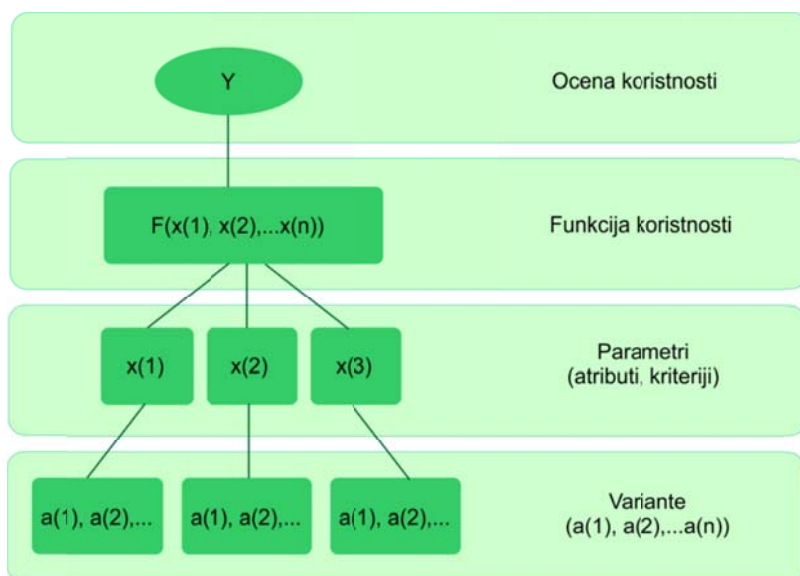
4.3.1 Modeli sistemov za podporo odločanju

V teoriji zasledimo tri modele SPO, to so:

- odločitvena drevesa,
- večparametrski modeli,
- diagrami vpliva.

Osredotočili smo se na večparametrške sisteme odločanja (Slika 4.2), ki temeljijo na razdrobitvi določenega problema na manjše podprobleme. Variante razdelimo na posamezne parametre, kot so kriteriji in atributi, ter jih ločeno ocenimo glede na vsak podan parameter. Končno oceno pa pridobimo z združevanjem. Na takšen način potem pridemo do izbora najustreznejše variante (Bohanec in drugi, 2003).

Slika 4.2: Večparametrski odločitveni model



Vir: Bohanec in drugi (2003)

Pri večparametrskem odločanju se srečamo z različnimi metodami, kot sta metodi analitičnega hierarhičnega procesiranja (ang. »*Analytic Hierarchy Process (AHP)*«) in ekspertni sistemi (ang. »*Decision EXpert (DEX)*«).

AHP metoda, ki smo jo uporabili pri izgradnji drevesa kriterijev, je splošna teorija merjenja. Uporablja se za izpeljavo razmerja lestvic, tako iz diskretnih kot tudi zveznih parnih primerjav na več ravneh hierarhične strukture. Te primerjave je možno vzeti iz dejanskih meritev ali iz temeljne lestvice, ki odraža relativno jakost preferenc ali občutkov. Pri uporabi AHP za modeliranje problemov potrebuje hierarhično ali omrežno strukturo, s katero predstavlja problem (Saaty in Vargas, 2012).

DEX je kvalitativna metodologija za podporo odločanju pri vrednotenju in analiziranju alternativ odločanja. Zasnovana je bila pred več kot 30 leti, metodologija pa ima za sabo dolgo zgodovino znanstvenih, tehničnih in praktičnih prispevkov. Predstavlja pristop združevanja »klasičnega« numeričnega večparametrskega odločitvenega modela, ki temelji na pravih ekspertnih sistemih. Takšen pristop je vodil k razvoju novih algoritmov in tehnik za pridobivanje in zastopanje odločitvenega znanja in ocenitev ter odločitvene analize. DEX je bil implementiran v treh generacijah programske opreme – to so DECMAK, DEX in DEXi – ter vgrajen v veliko računalniških programov in sistemov. Uporabljen je v več sto praktičnih aplikacijah na nacionalni kot tudi mednarodni ravni.

Za vrednotenje in podajanje analize smo uporabili odprtokodno programsko rešitev sistemov za podporo odločanju, to je DEXi.

DEXi je preprost računalniški program za večparametrsko odločanje. Njegov namen je interaktivni razvoj kvalitativnih modelov večparametrskega odločanja in ocenitev možnosti. Uporaben je za podporo kompleksnih nalog odločanja, kjer je potrebno izbrati eno izmed množice možnosti glede na izpolnjevanje ciljev. Večparametrski model je hierarhično strukturiran in predstavlja razgradnjo odločitvenega problema v podprobleme, ki so manjši, manj zapleteni in morda tudi lažje rešljivi kot celoten problem (Bohanec, 2013).

5. PRIMERJALNA ANALIZA REFERENČNIH PODATKOVNIH SKLADIŠČ

Podatkovno skladišče predstavlja osnovo za proces odločanja. Zato je pomembno zgraditi učinkovito metodo implementacije podatkovnega skladišča, v naslednjem koraku pa pomembno vlogo igra izgradnja učinkovitega programa za doseg inkrementalnega podatkovnega rudarjenja, s pomočjo katerega izboljšamo pridobljene rezultate (Zhang, 2012).

Živimo v svetu velikih količin podatkov. Konkurenčna prednost podjetja je zagotovo tudi ta, da zna te podatke pravilno izkoristiti in se pravilno odločiti za nadaljnji korak v procesu poslovanja. Izgradnja podatkovnega skladišča je kompleksen postopek, ki mora biti dobro premišljen, zato se mora podjetje odgovorno odločiti, katero podatkovno skladišče bo izbralo, da bo zadostovalo kriterijem (Krishnan, 2013).

Za namen izbora podatkovnega skladišča, ki ga bo podjetje uporabilo za nadaljnje operacije, smo uporabili metodo večparametrške odločitvene analize in jo izvedli s sistemom za podporo odločanju DEXi.

5.1 Identifikacija problema

Podjetje Car marketing ima lasten informacijski sistem, ki temelji na podatkovni bazi MS (Microsoft) Access in programskem jeziku Visual Basic. Podatkovna baza v MS Access

vsebuje veliko količino podatkov, nekatere tabele vsebujejo tudi več kot 50.000 zapisov in zaradi preobsežnosti podatkovne baze ni možno več procesiranje poizvedb s sprejemljivo zakasnitvijo. Ker podjetje za svoje dobro poslovanje potrebuje prave odločitve, želi implementirati sistem, ki bo omogočal različne analize in poročila, s čimer bi podprli proces odločanja. Za to pa potrebuje podatkovno skladišče, ki bo zmogljivo in bo ustrezalo zahtevanim kriterijem.

V ta namen smo postavili posamezna podatkovna skladišča v virtualnem okolju in preizkusili ter ovrednotili uporabnosti obravnavanih podatkovnih skladišč. Pri migraciji izvirne podatkovne baze smo na podlagi postavljenih odločitvenih kriterijev podali konkretno primerjalno analizo o uporabnosti več različnih implementacij podatkovnih skladišč. Podali smo tudi konkretne rezultate s pomočjo orodja DEXi in VREDANE

5.2 Izgradnja odločitvenega modela

Cilj fizične izgradnje podatkovne baze je določiti, kako so zapisi v podatkovni bazi shranjeni, dostopni in povezani, da lahko zagotovimo ustrezno delovanje aplikacije, s katero dostopamo do podatkovne baze. Upoštevati moramo različne dejavnike, ki se uporabljajo za merjenje učinkovitosti tovrstnih aplikacij. Odzivni čas je eden izmed ključnih dejavnikov, ki ga je potrebno upoštevati pri izgradnji, saj je pomembno, kako hitro lahko pridobimo informacije. Pri fizični izgradnji je zelo pomembno spremljanje sistema (nadzor in opozorila) in možnost prilagoditve ob spremembah (Malinowski in Zimányi, 2009).

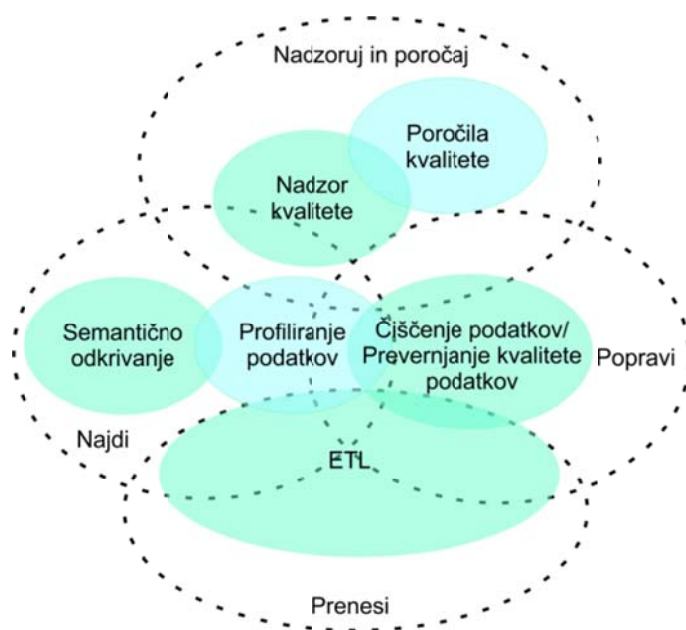
Infrastruktura podatkovnega skladišča mora zadoščati tudi tehničnemu vidiku, kar vključuje povezljivost med podedovanim okoljem in okoljem novega podatkovnega skladišča na omrežju kot tudi na nivoju podatkovne baze ter nivoju implementacije polnilnega, administrativnega in upravljaljskega podsistema (Prabhu, 2008).

Pri izgradnji podatkovnega skladišča moramo upoštevati tudi upravljaljski vidik (Humphries in drugi, 1999). Strošek podatkovnega skladišča raste eksponentno s količino podatkov, kar je potrebno upoštevati pri izgradnji podatkovnega skladišča (Inmon in drugi, 2008). Kot strošek pri izgradnji podatkovnega skladišča moramo upoštevati strojno in programsko opremo, storitve (to so sistemsko osebje, svetovalci in izobraževalci), poleg tega pa moramo upoštevati tudi stroške internega osebja, ki ga je potrebno izobraziti za uporabo podatkovnega skladišča (Humphries in drugi, 1999).

Orodja za nadzor in spremljanje podatkovnega skladišča igrajo pomembno vlogo pri kakovostnem delovanju le-tega. Na voljo so orodja, ki nadzirajo kakovost podatkov, in takšna, ki poročajo o kakovosti podatkov. ETL orodja, ki so odgovorna za ekstrakcijo, prenos in nalaganje podatkov, imajo pogosto tudi možnost nadzora kakovosti podatkov. Štiri glavne lastnosti (Slika 5.1) orodja za kakovost podatkov so (Inmon in drugi, 2008):

- *Najdi* – odkritje nepravilnosti,
- *Popravi* – očiščenje podatkov glede na določena pravila,
- *Prenesi* – ETL orodje prenese podatke v podatkovno skladišče,
- *Nadzoruj in poročaj* – nadzor in poročila o kakovosti podatkov.

Slika 5.1: Nadzor kakovosti podatkov



Vir: Povzeto po Inmon in drugi (2008, str. 287)

Kvaliteta informacij je eden izmed ključnih dejavnikov uspeha podatkovnega skladišča, saj vpliva na dobičkonosnost podjetja. Če podatki niso točni ali so nezanesljivi, podatkovno skladišče izgubi svojo vrednost. Če so netočni podatki uporabljeni pri odločitvenem procesu, je kvaliteta odločitev resno ogrožena, pričakujemo pa lahko tudi nezaželene posledice. Zato morajo biti podatki v podatkovnem skladišču natančni in dosegati visoke standarde kakovosti z naslednjimi značilnostmi (Khan, 2003):

- natančnost,

- popolnost,
- zanesljivost,
- pomembno in uporabno,
- razumljivo,
- trenutno.

Z vodstvom podjetja Car marketing smo opravili delno strukturiran intervju, kjer smo se osredotočili na ključna področja, kot jih omenja Nagabhushana (2006). Iz intervjuja smo pridobili informacijo, da podjetje pričakuje enostavnost in natančno prenesene podatke iz izvorne podatkovne baze v podatkovno skladišče. Želijo, da bi podatkovno skladišče omogočalo analize in poročila, ki jih sedaj ne morejo pridobiti. Glede na finančno zmogljivost bi podjetje želelo ostati znotraj določenega proračuna (med 1.000 in 2.000 EUR), pri tem pa jih zanima tudi sama podpora, ki bi je bili deležni pri uporabi podatkovnega skladišča. Tako smo se tukaj osredotočili na naslednje kriterije:

- namestitev (trajanje in težavnost) v primeru, da bi se podjetje samostojno odločilo za celotno izvedbo migracije in vzpostavitev sistema podatkovnega skladišča,
- migracija (prenos vseh elementov podatkovne baze),
- kvaliteta podatkov (natančnost, konsistenca in celovitost),
- kompatibilnost (v primeru, da bi podatkovno skladišče bilo kompatibilno z obstoječo programsko opremo in infrastrukturo),
- združljivost podatkovnega skladišča z izvornim sistemom in podprtost podatkovnega skladišča za različna orodja za analizo.

Prabhu (2008) omenja pomembnost tehničnega vidika, zato smo tudi naše drevo razdelili na dve glavni veji, to sta tehnični kriterij in upravljavski kriterij, ki ju razlaga Humphries (in drugi, 1999).

5.3 Izbrani kriteriji

Na podlagi analize smo prišli do sledečih končnih kriterijev, ki smo jih upoštevali pri migraciji podatkovne baze iz MS Access-a v podatkovno skladišče:

- enostavnost dostopa do podatkovnega skladišča (uporabniški vmesnik),
- podprtost naprednih poizvedb in orodij za analizo,

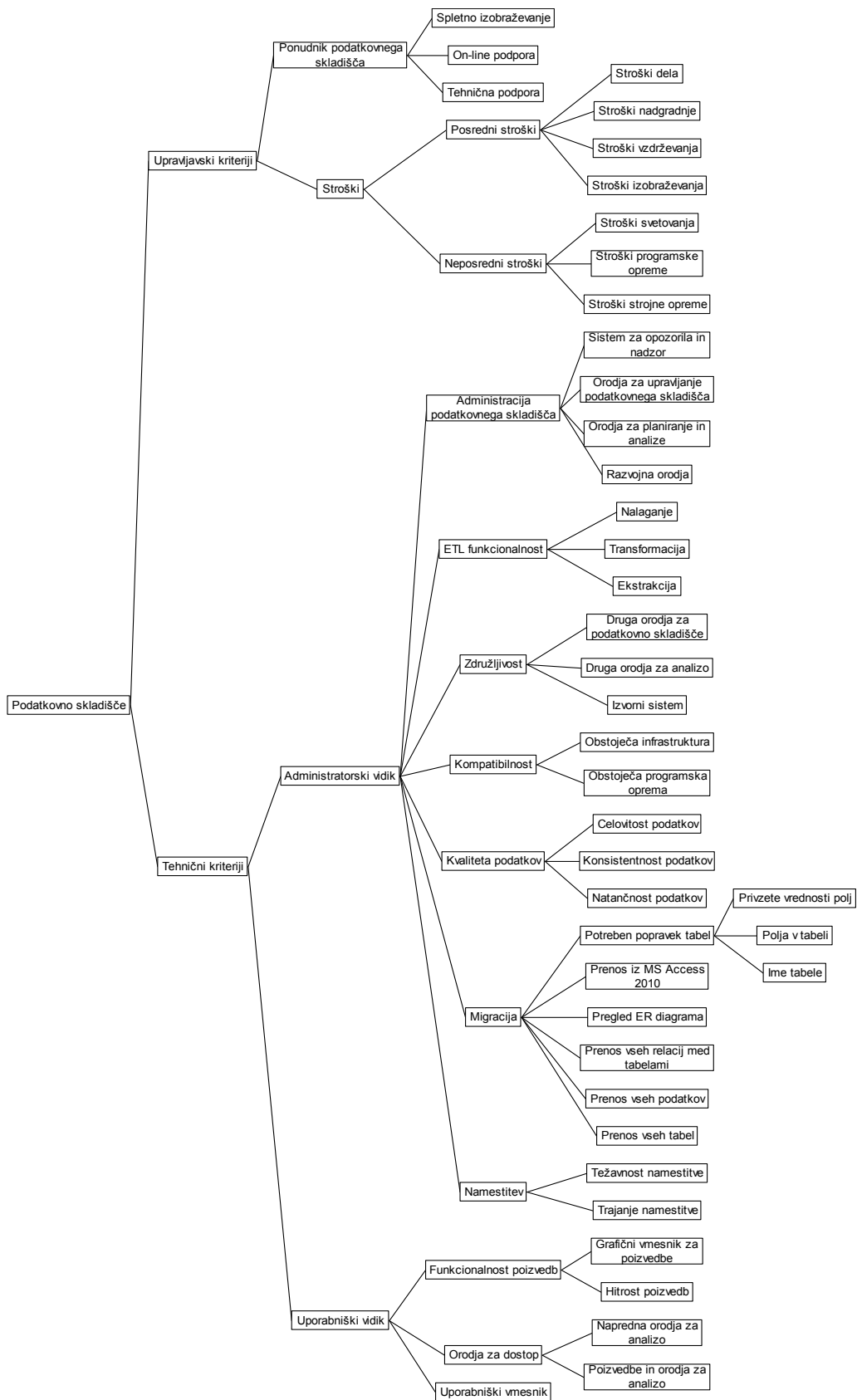
- hitrost pridobitve zelenih informacij,
- trajanje izvedbe in težavnost postavitve delovnega okolja,
- uspešen prenos vseh elementov izvorne podatkovne baze (tabele, polja, relacije) brez dodatnih prilagoditev in sprememb v izvorni podatkovni bazi,
- pregled diagrama relacij,
- neposreden prenos iz izvorne podatkovne baze (brez posrednikov),
- kvaliteta prenesenih podatkov (natančnost, konsistenca, celovitost),
- kompatibilnost podatkovnega skladišča z obstoječo programsko opremo in infrastrukturo,
- združljivost izvornega sistema s podatkovnim skladiščem in drugimi orodji za analize,
- podprtost podatkovnega skladišča z relacijskimi in večdimenzionalnimi podatkovnimi bazami,
- zmožnosti ETL procesa,
- podprtost z razvojnimi orodji,
- orodja za planiranje in analize,
- orodja za upravljanje podatkovnega skladišča,
- sistem za nadzor in opozorila,
- stroški stojne in programske opreme,
- stroški osebja,
- podpora uporabnikom.

5.4 Identifikacija kriterijev

S kabinetno študijo literature na področju podatkovnih skladišč ter na podlagi opravljenega intervjuja smo v orodju DEXi zgradili drevo kriterijev (Slika 5.2), s pomočjo katerega smo podali končno oceno obravnavanih podatkovnih skladišč.

Kriterije smo razdelili na dve glavni veji, in sicer na tehnične in upravljalvske kriterije, ker želimo ugotoviti, ali ocenjevana podatkovna skladišča dosegajo uporabniški in administratorski vidik (Chaundhuri in Dayal, 1997; Kimball in Ross, 2013), ki sta naslednji dve glavni veji tehničnega vidika, ter stroškovni in podporni kriterij, ki sta veji upravljalvskega kriterija.

Slika 5.2: Drevo kriterijev



Vir: Klenovšek, lastni prikaz (2014)

Z uporabniškega vidika želimo ugotoviti, kako prijazen je uporabniški vmesnik, uporabnost podatkovnega skladišča za uporabnika in njegove potrebe po iskanju, poizvedovanju in poročilih.

Administratorski vidik smo razvejali na sedem podkriterijev in predstavlja odgovornost zbiranja in priprave podatkov ter upravljanja podatkovnega skladišča. Administratorski vidik je sestavljen iz kriterijev: namestitev, migracija, kvaliteta podatkov, kompatibilnost, združljivost, podatkovna baza, ETL funkcionalnost in administracija podatkovnega skladišča.

Namestitev smo ocenili z vidika trajanja in težavnosti v primeru, da bi v podjetju želeli celotno infrastrukturo postaviti sami.

Ker so pravilno preneseni podatki ključnega pomena pri migraciji, smo to vejo razvejali na kriterije, kjer smo ocenjevali pravilno prenesene elemente podatkovne baze (tabele, podatki, relacije). Kot pravi Khan (2003), je kvaliteta tista, ki bo vplivala na kredibilnost podatkovnega skladišča, zato smo preverjali natančnost, konsistentnost in celovitost prenesenih podatkov.

Preverili smo tudi, ali je podatkovna baza kompatibilna s programsko opremo in infrastrukturo, ki jo podjetje že ima.

S kriterijem združljivost smo ocenili, ali lahko podatkovno skladišče integriramo z izvornim sistemom in pri tem uporabimo druga orodja za analizo (kot so podatkovno rudarjenje, sistemi za podporo odločanju) ali podatkovno skladišče (v primeru, da podatkovno skladišče ne zagotavlja takšnih orodij, lahko podjetje uporabi orodje drugega ponudnika).

ETL je pomembno orodje podatkovnega skladišča, saj integrira podatke v podatkovno skladišče (Kimball in Ross, 2013). Zato smo pri ocenjevanju upoštevali tudi ta kriterij.

Kot pomemben dejavnik za končno oceno smo podatkovna skladišča ocenili tudi z administratorskega vidika, saj lahko administratorska orodja pomagajo sistemskim administratorjem pri upravljanju, načrtovanju in planiranju podatkovnega skladišča (Chaundhuri in Dayal, 1997). Tako je pomembno orodje za poizvedbe in analize ter sistem za opozorila in nadzor.

Že smo omenili, da so tudi stroški ključni dejavnik pri izgradnji podatkovnega skladišča, zato smo v odločitveno drevo (skupaj z opisi predstavljeno v Tabeli 5.1) le-te vključili pri upravljaljskih kriterijih.

Tabela 5.1: Drevo kriterijev z opisi

Kriterij	Opis
Podatkovno skladišče	Kriteriji izbora optimalnega podatkovnega skladišča pri migraciji podatkovne baze v MS Accessu
Tehnični kriteriji	Tehnični vidik pri vzpostavitvi podatkovnega skladišča
Uporabniški vidik	Uporabniški vidik pri vzpostavitvi podatkovnega skladišča
Uporabniški vmesnik	Kako prijazen (enostaven) je uporabniški vmesnik
Orodja za dostop	Orodja za dostop do analize in poizvedbe
Poizvedbe in orodja za analizo	Orodja za poizvedbe in analize
Napredna orodja za analizo	Napredna orodja za analizo (statistika, vizualizacija, podatkovno rudarjenje)
Funkcionalnost poizvedb	Hitrost in enostavnost pridobljenih informacij
Hitrost poizvedb	Hitrost pridobljenih informacij
Grafični vmesnik za poizvedbe	Enostavna uporaba in z GUI podprto orodje za poizvedbe
Administratorski vidik	Administratorski vidik pri vzpostavitvi podatkovnega skladišča
Namestitve	Nemestitev okolja za podatkovno skladišče, kot tudi samo podatkovno skladišče
Trajanje namestitve	Koliko časa je trajala celotna namestitve
Težavnost namestitve	Kako težavna je bila namestitve
Migracija	Kriteriji pri migraciji podatkovne baze v podatkovno skladišče
Prenos vseh tabel	Vse tabele so se prenesle brez težav
Prenos vseh podatkov	Vsi podatki so se prenesli brez težav
Prenos vseh relacij med tabelami	Vse relacije so se prenesle brez težav
Pregled ER diagrama	Pregled diagrama relacij
Prenos iz MS Access 2010	Prenos podatkovne baze smo lahko opravili neposredno iz MS Accessa
Potreben popravek tabel	Pred prenos ni bil potreben popravek tabel
Ime tabele	Popravek imena tabel
Polja v tabeli	Popravek polj v tabeli (tipi)
Privzete vrednosti polj	Popravek privzetih vrednosti v poljih
Kvaliteta podatkov	Kvaliteta prenesenih podatkov
Natančnost podatkov	Vsi podatki so bili natančno preneseni
Konsistentnost podatkov	Skladnost prenešenih podatkov
Celovitost podatkov	Celovitost prenešenih podatkov
Kompatibilnost	Kompatibilnost podatkovnega skladišča s trenutnim sistemom
Obstoječa programska oprema	Kompatibilnost z obstoječo programsko opremo (os, aplikacije)
Obstoječa infrastruktura	Kompatibilnost z obstoječo infrastrukturo (strojna oprema)
Združljivost	Združljivost podatkovnega skladišča z obstoječim sistemom
Izvorni sistem	Združljivost z izvornim sistemom
Druga orodja za analizo	Združljivost podatkovnega skladišča z drugimi orodji za analize
Druga orodja za podatkovno skladišče	Združljivost podatkovnega skladišča z drugimi orodji za podatkovno skladiščenje
ETL funkcionalnost	ETL funkcionalnost podatkovnega skladišča
Ekstrakcija	Ekstrakcija podatkov
Transformacija	Transformacija podatkov
Nalaganje	Nalaganje podatkov
Administracija podatkovnega skladišča	Kriterij za administracijo podatkovnega skladišča
Razvojna orodja	Razvojna orodja za poizvedbe, pravila, skripte
Orodja za planiranje in analize	Orodja za planiranje in izvedbo analiz
Orodja za upravljanje podatkovnega skladišča	Orodja za nadzorni sistem, statistična poročila, varnost, časovno načrtovanje, dobro delovanje
Sistem za opozorila in nadzor	Podatkovno skladišče vsebuje sistem za opozorila in nadzor
Upravljaljski kriteriji	Upravljaljski vidik pri vzpostavitvi podatkovnega skladišča
Stroški	Višina stroškov
Neposredni stroški	Neposredni stroški pri vzpostavitvi podatkovnega skladišča
Stroški strojne opreme	Višina stroškov strojne opreme
Stroški programske opreme	Višina stroškov programske opreme
Stroški svetovanja	Višina stroškov svetovanja
Posredni stroški	Posredni stroški pri vzpostavitvi podatkovnega skladišča
Stroški izobraževanja	Višina stroškov izobraževanja
Stroški vzdrževanja	Višina stroškov vzdrževanja
Stroški nadgradnje	Višina stroškov nadgradnje
Stroški dela	Višina stroškov dela
Ponudnik podatkovnega skladišča	Podpora ponudnika podatkovnega skladišča
Tehnična podpora	Ponudnik podatkovnega skladišča nudi tehnično podporo
On-line podpora	Ponudnik podatkovnega skladišča nudi on-line podporo
Spletno izobraževanje	Ponudnik podatkovnega skladišča omogoča spletno izobraževanje

Vir: Klenovšek, lastni prikaz (2014)

5.5 Opis zalog vrednosti kriterijev

V Tabeli 5.2 smo prikazali določene zaloge vrednosti izbranim kriterijem. Zaradi kompleksnosti odločitvenega drevesa smo se pri nekaterih kriterijih omejili na zalogo vrednosti samo »da« ali »ne«.

Tabela 5.2: Zaloga vrednosti kriterijev

Kriterij	Zaloga vrednosti
Podatkovno skladišče	1; 2; 3; 4; 5
Tehnični kriteriji	1; 2; 3; 4; 5
Uporabniški vidik	slabo; dobro; odlično
Uporabniški vmesnik	Brez; Kompleksen; Enostaven
Orodja za dostop	ne; delno; v celoti
Poizvedbe in orodja za analizo	ne; da
Napredna orodja za analizo	ne; delno; da
Funkcionalnost poizvedb	slaba; dobra; odlična
Hitrost poizvedb	več kot 5 sekund; med 2 in 5 sekund; do 2 sekundi
Grafični vmesnik za poizvedbe	ne; da
Administratorski vidik	slabo; dobro
Namestitve	zahtevno; enostavno
Trajanje namestitve	več kot 2 ure; do 2 ur
Težavnost namestitve	zahtevno; enostavno
Migracija	slaba; dobra
Prenos vseh tabel	ne; da
Prenos vseh podatkov	ne; da
Prenos vseh relacij med tabelami	ne; da
Pregled ER diagrama	ne; da
Prenos iz MS Access 2010	ne; da
Potreben popravek tabel	da; ne
Ime tabele	da; ne
Polja v tabeli	da; ne
Privzete vrednosti polj	da; ne
Kvaliteta podatkov	nepopolna; popolna
Natančnost podatkov	nepopolna; popolna
Konsistentnost podatkov	nepopolna; v celoti
Celovitost podatkov	nepopolna; v celoti
Kompatibilnost	ne; da
Obstoječa programska oprema	ne; delno; da
Obstoječa infrastruktura	ne; delno; da
Združljivost	ne; delno; da
Izvorni sistem	ne; da
Druga orodja za analizo	ne; da
Druga orodja za podatkovno skladišče	ne; da
ETL funkcionalnost	slaba; dobra
Ekstrakcija	ne; da
Transformacija	ne; da
Nalaganje	ne; da
Administracija podatkovnega skladišča	slaba; dobra
Razvojna orodja	ne; da
Orodja za planiranje in analize	ne; da
Orodja za upravljanje podatkovnega skladišča	ne; da
Sistem za opozorila in nadzor	ne; da
Upravljaljski kriteriji	1; 2; 3; 4; 5
Stroški	visoki; srednji; nizki
Neposredni stroški	visoki; srednji; nizki
Stroški strojne opreme	nad 10.000 EUR; med 5.000 EUR in 10.000 EUR; do 5.000 EUR; do 1.000 EUR
Stroški programske opreme	nad 5.000 EUR; med 1.000 EUR in 5.000 EUR; brezplačno
Stroški svetovanja	nad 2.000 EUR; med 1.000 EUR in 2.000 EUR; do 1.000 EUR
Posredni stroški	visoki; srednji; nizki
Stroški izobraževanja	nad 2.000 EUR; med 1.000 EUR in 2.000 EUR; do 1.000 EUR
Stroški vzdrževanja	nad 1.000 EUR; med 500 EUR in 1.000 EUR; do 500 EUR
Stroški nadgradnje	nad 1.000 EUR; med 500 in 1.000 EUR; do 500 EUR
Stroški dela	nad 1.000 EUR; med 500 EUR in 1.000 EUR; do 500 EUR
Ponudnik podatkovnega skladišča	slab; dober; odličen
Tehnična podpora	slaba; zadovoljiva; odlična
On-line podpora	ne; da
Spletno izobraževanje	ne; da

Vir: Klenovšek, lastni prikaz (2014)

Pri določanju zaloge vrednosti smo kriteriju podatkovno skladišče določili kvalitativno zalogo vrednosti (vrednosti od 1 do 5, kjer 5 predstavlja najboljšo vrednost). Prav tako smo kvalitativno zalogo vrednosti določili tehničnemu in upravljavskemu kriteriju, da smo tako lahko dobili širši spekter natančnosti ocene. Pri drugih kriterijih smo zalogam vrednosti določili kvalitativno vrednost, pri čemer smo se pri nekaterih kriterijih odločili le za varianti »ne« in »da«, predvsem zaradi večje razvejanosti drevesa in morebitne kompleksnosti pri določanju funkcije koristnosti.

5.6 Funkcija koristnosti

Funkcija koristnosti določa vpliv kriterija v drevesu (Bohanec in drugi, 2003). Funkcija koristnosti je predstavljena v obliki tabel, kot je to na primeru funkcije koristnosti za podatkovno skladišče v Tabeli 5.3. Pri določanju funkcije koristnosti smo uporabili utežno vsoto.

Pri odločitveni analizi je funkcija koristnosti velikokrat definirana kot vsota uteži, na primer (Bohanec, 2013):

$$f(X_1, X_2, \dots, X_n) = w_1 \times X_1 + w_2 \times X_2 + \dots + w_n \times X_n$$

V predstavljeni formuli so w_i uteži, X_i pa atributi.

Program DEXi lahko samodejno izračuna funkcijo koristnosti, pri tem pa določimo upoštevanje uteži. Za izračun funkcije koristnosti je vsem listom v drevesu predhodno potrebno določiti zalogo vrednosti.

Tabela 5.3: Odločitvena pravila za kriterij podatkovno skladišče

	Tehnični kriteriji	Upravljavski kriteriji	Podatkovno skladišče
1	1	1	1
2	1	2	1
3	1	3	2
4	1	4	3
5	1	5	3
6	2	1	1
7	2	2	2
8	2	3	3
9	2	4	3
10	2	5	3
11	3	1	2
12	3	2	3
13	3	3	3
14	3	4	3
15	3	5	4
16	4	1	3
17	4	2	3
18	4	3	3
19	4	4	4
20	4	5	5
21	5	1	3
22	5	2	3
23	5	3	4
24	5	4	5
25	5	5	5

Vir: Klenovšek, lastni prikaz (2014)

V Tabeli 5.4 smo predstavili funkcijo koristnosti za upravljavski kriterij odločitvenega drevesa.

Tabela 5.4: Odločitvena pravila za upravljavski kriterij

	Stroški	Ponudnik podatkovnega skladišča	Upravljavski kriteriji
1	visoki	slab	1
2	visoki	dober	2
3	visoki	odličen	3
4	srednji	slab	2
5	srednji	dober	3
6	srednji	odličen	4
7	nizki	slab	3
8	nizki	dober	4
9	nizki	odličen	5

Vir: Klenovšek, lastni prikaz (2014)

V Tabeli 5.5 smo predstavili funkcijo koristnosti za tehnični kriterij odločitvenega drevesa.

Tabela 5.5: Odločitvena pravila za tehnični kriterij

	Uporabniški vidik	Administratorski vidik	Tehnični kriteriji
1	slabo	slabo	1
2	slabo	dobro	2
3	dobro	slabo	2
4	dobro	dobro	4
5	odlično	slabo	4
6	odlično	dobro	5

Vir: Klenovšek, lastni prikaz (2014)

Vse funkcije koristnosti so prikazane v Tabeli 5.6. Funkcija koristnosti združi vse kombinacije na kriterijih nižje vrednosti v vrednost za koren drevesa. Združevanje je predstavljeno v tabelah, kjer vsaka vrstica predstavlja eno kombinacijo kriterija nižje vrednosti. Vrstice pa prav tako imenujemo odločitvena pravila, saj lahko vsako vrstico predstavimo kot pravilo »kaj-če« (Bohanec, 2013).

Tabela 5.6: Funkcije koristnosti

Kriterij	Pravil	Definiranost	Določenost	Vrednosti
Podatkovno skladišče	2/25	8,00%	100,00%	1:3,2:3,3:13,4:3,5:3
Tehnični kriteriji	2/6	33,33%	100,00%	1:1,2:2,3:0,4:2,5:1
Uporabniški vidik	2/27	7,41%	100,00%	slabo:4,dobro:19,odlično:4
Uporabniški vmesnik				
Orodja za dostop	2/6	33,33%	100,00%	ne:1,delno:4,v celoti:1
Poizvedbe in orodja za analizo				
Napredna orodja za analizo				
Funktionalnost poizvedb	2/6	33,33%	100,00%	slaba:1,dobra:4,odlična:1
Hitrost poizvedb				
Grafični vmesnik za poizvedbe				
Administratorski vidik	2/192	1,04%	100,00%	slabo:121,dobro:71
Namestitve	2/4	50,00%	100,00%	zahtevno:3,enostavno:1
Trajanje namestitve				
Težavnost namestitve				
Migracija	2/64	3,13%	100,00%	slaba:42,dobra:22
Prenos vseh tabel				
Prenos vseh podatkov				
Prenos vseh relacij med tabelami				
Pregled ER diagrama				
Prenos iz MS Access 2010				
Potreben popravek tabel	3/8	37,50%	100,00%	da:7,ne:1
Ime tabele				
Polja v tabeli				
Privzete vrednosti polj				
Kvaliteta podatkov	2/8	25,00%	100,00%	nepopolna:4,popolna:4
Natančnost podatkov				
Konsistentnost podatkov				
Celovitost podatkov				
Kompatibilnost	2/9	22,22%	100,00%	ne:6,da:3
Obstoječa programska oprema				
Obstoječa infrastruktura				
Zdržljivost	2/8	25,00%	100,00%	ne:1,delno:6,da:1
Izvorni sistem				
Druga orodja za analizo				
Druga orodja za podatkovno skladišče				
ETL funkcionalnost	2/8	25,00%	100,00%	slaba:4,dobra:4
Ekstrakcija				
Transformacija				
Nalaganje				
Administracija podatkovnega skladišča	2/16	12,50%	100,00%	slaba:11,dobra:5
Razvojna orodja				
Orodja za planiranje in analize				
Orodja za upravljanje podatkovnega skladišča				
Sistem za opozorila in nadzor				
Upravljalški kriteriji	2/9	22,22%	100,00%	1:1,2:2,3:3,4:2,5:1
Stroški	2/9	22,22%	100,00%	visoki:3,srednji:3,nizki:3

Neposredni stroški	2/36	5,56%	100,00%	visoki:4,srednji:28,nizki:4
Stroški strojne opreme				
Stroški programske opreme				
Stroški svetovanja				
Posredni stroški	2/81	2,47%	100,00%	visoki:15,srednji:51,nizki:15
Stroški izobraževanja				
Stroški vzdrževanja				
Stroški nadgradnje				
Stroški dela				
Ponudnik podatkovnega skladišča	2/12	16,67%	100,00%	slab:4,dober:4,odličen:4
Tehnična podpora				
On-line podpora				
Spletno izobraževanje				

Vir: Klenovšek, lastni prikaz (2014)

DEXi uporablja štiri različne tipe uteži, kot lahko vidimo v Tabeli 5.7. Razlika med lokalno in globalno utežjo je v drevesnih atributih. Lokalne uteži se vedno nanašajo na en sam skupni atribut in ustrezni funkciji koristnosti, tako da je vsota uteži neposrednih potomcev 100 %. Globalne uteži pa upoštevajo drevo in relativno pomembnost poddrevesa. Globalna utež se izračuna kot produkt lokalne uteži in globalne uteži atributa, ki se nahaja eno stopnjo nad njim. Globalna utež korenskega atributa je 100 %.

Uteži lahko normaliziramo, saj lahko nekatere lestvice vsebujejo več vrednosti kot druge. Večje lestvice so običajno v nižjih stopnjah drevesa in imajo posledično tudi manjše uteži. Normalizacija se nanaša na postopek, v katerem so vse lestvice prilagojene na enako dolžino, predno se jim določijo uteži (Bohanec, 2013).

Tabela 5.7: Prikaz povprečnih uteži v odločitvenem drevesu

<u>Kriterij</u>	<u>Lokalne</u>	<u>Globalne</u>	<u>Lok.norm.</u>	<u>Glob.norm.</u>
Podatkovno skladišče				
Tehnični kriteriji	50	50	50	50
Uporabniški vidik	53	26	63	31
Uporabniški vmesnik	33	9	33	10
Orodja za dostop	33	9	33	10
Poizvedbe in orodja za analizo	57	5	47	5
Napredna orodja za analizo	43	4	53	6
Funkcionalnost poizvedb	33	9	33	10
Hitrost poizvedb	43	4	53	6
Grafični vmesnik za poizvedbe	57	5	47	5
Administatorski vidik	47	24	37	19
Namestitev	14	3	13	2
Trajanje namestitve	50	2	50	1
Težavnost namestitve	50	2	50	1
Migracija	14	3	13	2
Prenos vseh tabel	17	1	17	0
Prenos vseh podatkov	17	1	17	0
Prenos vseh relacij med tabelami	17	1	17	0
Pregled ER diagrama	17	1	17	0
Prenos iz MS Access 2010	17	1	17	0
Potreben popravek tabel	17	1	17	0
Ime tabele	33	0	33	0
Polja v tabeli	33	0	33	0
Privzete vrednosti polj	33	0	33	0

—Kvaliteta podatkov	14	3	13	2
—Natančnost podatkov	33	1	33	1
—Konsistentnost podatkov	33	1	33	1
—Celovitost podatkov	33	1	33	1
—Kompatibilnost	14	3	13	2
—Obstoječa programska oprema	50	2	50	1
—Obstoječa infrastruktura	50	2	50	1
—Združljivost	15	4	21	4
—Izvorni sistem	33	1	33	1
—Druga orodja za analizo	33	1	33	1
—Druga orodja za podatkovno skladišče	33	1	33	1
—ETL funkcionalnost	14	3	13	2
—Ekstrakcija	33	1	33	1
—Transformacija	33	1	33	1
—Nalaganje	33	1	33	1
—Administracija podatkovnega skladišča	14	3	13	2
—Razvojna orodja	25	1	25	1
—Orodja za planiranje in analize	25	1	25	1
—Orodja za upravljanje podatkovnega skladišča	25	1	25	1
—Sistem za opozorila in nadzor	25	1	25	1
—Upravljavski kriteriji	50	50	50	50
—Stroški	50	25	50	25
—Neposredni stroški	50	12	50	12
—Stroški strojne opreme	31	4	37	5
—Stroški programske opreme	35	4	31	4
—Stroški svetovanja	35	4	31	4
—Posredni stroški	50	12	50	12
—Stroški izobraževanja	25	3	25	3
—Stroški vzdrževanja	25	3	25	3
—Stroški nadgradnje	25	3	25	3
—Stroški dela	25	3	25	3
—Ponudnik podatkovnega skladišča	50	25	50	25
—Tehnična podpora	36	9	46	11
—On-line podpora	32	8	27	7
—Spletno izobraževanje	32	8	27	7

Vir: Klenovšek, lastni prikaz (2014)

5.7 Odločitvena pravila

DEXi uporablja kompleksna pravila za predstavitev funkcije koristnosti v bolj kompaktnem in enostavnem načinu. Odločitvena pravila so pridobljena z združitvijo nekaterih osnovnih pravil, ki imajo enako funkcijsko vrednost (Bohanec, 2013).

V drevesu kriterijev vsakemu poddrevesu nastavimo odločitveno pravilo. Pravila nam povedo, kako skupina kriterijev vpliva na končno odločitev izbire podatkovnega skladišča.

V tabelah predstavljamo odločitvena pravila za posamezna poddrevesa.

Tabela 5.8 prikazuje vpliv tehničnih in upravljavskih kriterijev na izbor podatkovnega skladišča. Vrednost 1 predstavlja najmanjšo oceno, vrednost 5 pa največjo oceno. Iz tabele lahko razberemo, da bo najboljše tisto podatkovno skladišče, ki bo imelo tehnični kriterij ≤ 4 , upravljavski kriterij pa 5, ali pa podatkovno skladišče s tehničnim kriterijem 5 in upravljavskim kriterijem ≥ 4 . Tehnični in upravljavski kriterij imata oba enak vpliv, saj oba na odločitev vplivata 50 %.

Tabela 5.8: Vpliv tehničnih in upravljaljskih kriterijev na podatkovno skladišče

Tehnični kriteriji	Upravljaljski kriteriji	Podatkovno skladišče
50%	50%	
1 1	<=2	1
2 <=2	1	1
3 1	3	2
4 2	2	2
5 3	1	2
6 <=2	>=4	3
7 <=3	4	3
8 2	>=3	3
9 2:3	3:4	3
10 2:4	3	3
11 3	2:4	3
12 3:4	2:3	3
13 >=3	2	3
14 4	<=3	3
15 >=4	<=2	3
16 3	5	4
17 4	4	4
18 5	3	4
19 >=4	5	5
20 5	>=4	5

Vir: Klenovšek, lastni prikaz (2014)

V Tabeli 5.9 smo prikazali vpliv uporabniškega in administratorskega vidika na tehnične kriterije. Uporabniški in administratorski vidik smo ocenjevali z vrednostmi »slabo«, »dobro« in »odlično«, kjer tehnični vidik pridobi vrednost 1 v primeru, da sta oba vidika ocenjena »slabo« in vrednost 5 v primeru, da sta oba vidika ocenjena »odlično«.

Tabela 5.9: Vpliv uporabniškega in administratorskega vidika na tehnični kriterij

Uporabniški vidik	Administratorski vidik	Tehnični kriteriji
53%	47%	
1 slabo	slabo	1
2 slabo	dobro	2
3 dobro	slabo	2
4 dobro	dobro	4
5 odlično	slabo	4
6 odlično	dobro	5

Vir: Klenovšek, lastni prikaz (2014)

Uporabniški vmesnik, orodja za dostop in funkcionalnost poizvedb v enaki meri vplivajo na uporabniški vidik (Tabela 5.10). Uporabniški vmesnik smo ocenjevali z kvantitativnimi vrednostmi »slabo«, »dobro« in »odlično«.

Tabela 5.10: Vpliv uporabniškega vmesnika, orodij za dostop in funkcionalnost poizvedb na uporabniški vidik

Uporabniški vmesnik	Orodja za dostop	Funkcionalnost poizvedb	Uporabniški vidik
33%	33%	33%	
1 Brez	ne	<=dobra	slabo
2 Brez	<=delno	slaba	slabo
3 <=Kompleksen	ne	slaba	slabo
4 Brez	*	odlična	dobro
5 <=Kompleksen	<=delno	odlična	dobro
6 *	ne	odlična	dobro
7 Brez	>=delno	>=dobra	dobro
8 <=Kompleksen	delno	>=dobra	dobro
9 <=Kompleksen	>=delno	dobra	dobro
10 *	delno	dobra	dobro
11 Brez	v celoti	*	dobro
12 <=Kompleksen	v celoti	<=dobra	dobro
13 *	v celoti	slaba	dobro
14 Kompleksen	<=delno	>=dobra	dobro
15 Kompleksen	*	dobra	dobro
16 >=Kompleksen	ne	>=dobra	dobro
17 >=Kompleksen	<=delno	dobra	dobro
18 Kompleksen	delno	*	dobro
19 Kompleksen	>=delno	<=dobra	dobro
20 >=Kompleksen	delno	<=dobra	dobro
21 >=Kompleksen	>=delno	slaba	dobro
22 Enostaven	ne	*	dobro
23 Enostaven	<=delno	<=dobra	dobro
24 Enostaven	*	slaba	dobro
25 >=Kompleksen	v celoti	odlična	odlično
26 Enostaven	>=delno	odlična	odlično
27 Enostaven	v celoti	>=dobra	odlično

Vir: Klenovšek, lastni prikaz (2014)

Poizvedbe ter osnovna in napredna orodja za analize v približno enaki meri vplivajo na kriterij orodja za dostop (Tabela 5.11). V primeru, da podatkovno skladišče ne vsebuje orodij za poizvedbe in analize, je kvalitativna vrednost kriterija orodja za dostop »ne«, v nasprotnem primeru je kvalitativna vrednost »v celoti«, vmesne kombinacije pa so določene z vrednostjo »delno«.

Tabela 5.11: Vpliv poizvedb in orodij za analizo ter naprednih orodij za analizo na kriterij orodja za dostop

Poizvedbe in orodja za analizo	Napredna orodja za analizo	Orodja za dostop
57%	43%	
1 ne	ne	ne
2 ne	>=delno	delno
3 *	delno	delno
4 da	<=delno	delno
5 da	da	v celoti

Vir: Klenovšek, lastni prikaz (2014)

V primeru, da poizvedba v tabeli z veliko količino podatkov (več kot 50.000) traja več kot 5 sekund in za ustvarjanje poizvedb ne obstaja grafični vmesnik, je funkcionalnost poizvedb »slaba« (Tabela 5.12).

Tabela 5.12: Vpliv hitrosti poizvedb in grafičnega vmesnika za poizvedbe na funkcionalnost poizvedb

Hitrost poizvedb	Grafični vmesnik za poizvedbe	Funkcionalnost poizvedb
43%	57%	
1 več kot 5 sekund	ne	slaba
2 <=med 2 in 5 sekund	da	dobra
3 med 2 in 5 sekund	*	dobra
4 >=med 2 in 5 sekund	ne	dobra
5 do 2 sekundi	da	odlična

Vir: Klenovšek, lastni prikaz (2014)

Tabela 5.13 prikazuje odvisnost trajanja in težavnosti na kriterij namestitev. Oba kriterija, trajanje in namestitev, v enaki meri vplivata na končno odločitev kriterija namestitev

Tabela 5.13: Vpliv trajanja in težavnosti namestitve na kriterij namestitev

Trajanje namestitve	Težavnost namestitve	Namestitev
50%	50%	
1 več kot 2 ure	*	zahtevno
2 *	zahtevno	zahtevno
3 do 2 ur	enos tavno	enos tavno

Vir: Klenovšek, lastni prikaz (2014)

Tabela 5.14 prikazuje, kako kriteriji natančnost, konsistentnost in celovitost vplivajo na kriterij kvaliteta podatkov. Vsi trije kriteriji predstavljajo enako težo vrednosti na končno oceno kriterija kvaliteta podatkov.

Tabela 5.14: Vpliv natančnosti, konsistence in celovitosti podatkov na kvaliteto prenesenih podatkov pri migraciji

	Natančnost podatkov	Konsistentnost podatkov	Celovitost podatkov	Kvaliteta podatkov
	33%	33%	33%	
1	nepopoln	nepopoln	*	nepopolna
2	nepopoln	*	nepopolna	nepopolna
3	*	nepopoln	nepopolna	nepopolna
4	*	v celoti	v celoti	popolna
5	popoln	*	v celoti	popolna
6	popoln	v celoti	*	popolna

Vir: Klenovšek, lastni prikaz (2014)

Kompatibilnost obstoječe programske opreme in infrastrukture sta dejavnika, ki vplivata na kriterij kompatibilnost. Kriterij je določen s kvalitativno vrednostjo »da« in »ne« (Tabela 5.15).

Tabela 5.15: Vpliv obstoječe programske opreme in infrastrukture na kriterij kompatibilnost

	Obstoječa programska oprema	Obstoječa infrastruktura	Kompatibilnost
	50%	50%	
1	ne	*	ne
2	<=delno	<=delno	ne
3	*	ne	ne
4	>=delno	da	da
5	da	>=delno	da

Vir: Klenovšek, lastni prikaz (2014)

Združljivost je kriterij, ki je odvisen od izvornega sistema, drugih orodij za analizo in drugih orodij za podatkovno skladiščenje (Tabela 5.16). Določen je s kvantitativno vrednostjo »ne«, »delno« ali »da«.

Tabela 5.16: Vpliv izvornega sistema, drugih orodij za analizo, drugih orodij za podatkovno skladišče na kriterij združljivost

	Izvorni sistem	Druqa orodja za analizo	Druqa orodja za podatkovno skladišče	Združljivost
	33%	33%	33%	
1	ne	ne	ne	ne
2	ne	*	da	delno
3	*	ne	da	delno
4	ne	da	*	delno
5	*	da	ne	delno
6	da	ne	*	delno
7	da	*	ne	delno
8	da	da	da	da

Vir: Klenovšek, lastni prikaz (2014)

Tabela 5.17 prikazuje vpliv ekstrakcije, transformacije in nalaganja na ETL funkcionalnost, ki predstavlja velik pomen pri podatkovnem skladišču. Kriterij ETL funkcionalnost smo ovrednotili z vrednostjo »slaba« in »dobra«.

Tabela 5.17: Vpliv ekstrakcije, transformacije in nalaganja na kriterij ETL funkcionalnost

	Ekstrakcija	Transformacija	Nalaganje	ETL funkcionalnost
	33%	33%	33%	
1	ne	ne	*	slaba
2	ne	*	ne	slaba
3	*	ne	ne	slaba
4	*	da	da	dobra
5	da	*	da	dobra
6	da	da	*	dobra

Vir: Klenovšek, lastni prikaz (2014)

Upravljavski kriterij je določen s kvalitativno vrednostjo od 1 do 5, kjer 5 predstavlja odlično oceno, 1 pa slabo. Na kriterij vplivata stroški in ponudnik podatkovnega skladišča, oba v enaki meri 50 % (Tabela 5.18).

Tabela 5.18: Vpliv stroškov in ponudnika podatkovnega skladišča na upravljavski kriterij

	Stroški	Ponudnik podatkovnega skladišča	Upravljavski kriterij
	50%	50%	
1	visoki	slab	1
2	visoki	dober	2
3	srednji	slab	2
4	visoki	odličen	3
5	srednji	dober	3
6	nizki	slab	3
7	srednji	odličen	4
8	nizki	dober	4
9	nizki	odličen	5

Vir: Klenovšek, lastni prikaz (2014)

5.8 Opis variant

Na podlagi postavljenih kriterijev smo ocenjevali podatkovna skladišča ter podali oceno, katero je tisto podatkovno skladišče, ki bi bilo za podjetje Car marketing najbolj primerno.

Ocenjevana podatkovna skladišča so MySQL, PostgreSQL, MongoDB, MS SQL, Oracle 11g in DB2, ki smo jih natančneje že opredelili v Poglavju 3.

5.9 Vrednotenje in analiza variant

Analiza je eden izmed ključnih konceptov v procesu odločanja. Z razliko od ocenjevanja, ki je zgolj izračun usmerjenih vhodov do rezultatov, se izraz analiza razume kot aktivno sodelovanje udeležencev, ki poskušajo najti odgovore na vprašanja, kot so (Bohanec, 2013):

- Ali so kriteriji vrednotenja v skladu s pričakovanji? Ali so »pravilni«? Če niso, zakaj ne?
- Kako kriterije medsebojno primerjati? Kateri so boljši in zakaj?
- Ali lahko razložimo in utemeljimo ocene? Katere so najpomembnejše močne in šibke točke posameznih kriterijev?
- Kaj se zgodi, če se nekaj spremeni: Kaj če poskusimo nov kriterij? Kaj če kriterij postane nedosegljiv? Kaj če se spremenijo karakteristike kriterija?
- Kako občutljiva je ocena na male spremembe v odločitvenem modelu (kot je na primer dodajanje ali brisanje atributov, sprememba nekaterih pravil odločanja)?

Z drugimi besedami je analiza ustvarjalna in ponavljajoča se aplikacija procesa odločanja, ki stremi k boljšemu razumevanju odločitvenega problema, boljšemu razumevanju kriterijev, njihovih karakteristik in posledic ter omogoča boljše utemeljene odločitve. Vsekakor na končno oceno vpliva veliko dejavnikov, ki se lahko kadar koli spremenijo, zato se lahko pri ocenjevanju pojavijo tudi napake (Bohanec, 2013).

Pri vrednotenju variant smo temeljito preučili vse kriterije v odločitvenem drevesu in prišli do vrednosti, kot so prikazane v Tabeli 5.19.

Tabela 5.19: Vrednotenje kriterijev pri variantah

Varianta	MySQL	PostgreSQL	MongoDB	MS SQL	Oracle 11g	DB2
Uporabniški vmesnik	Enostaven	Enostaven	Brez	Kompleksen	Kompleksen	Enostaven
Poizvedbe in orodja za analizo	da	ne	ne	da	da	da
Napredna orodja za analizo	da	delno	delno	da	da	da
Hitrost poizvedb	med 2 in 5 sekund	več kot 5 sekund	več kot 5 sekund	do 2 sekundi	več kot 5 sekund	med 2 in 5 sekund
Grafični vmesnik za poizvedbe	ne	da	ne	ne	ne	ne
Trajanje namestitve	do 2 ur	do 2 ur	do 2 ur	več kot 2 ure	več kot 2 ure	več kot 2 ure
Težavnost namestitve	enostavno	enostavno	zahtevno	zahtevno	zahtevno	zahtevno
Prenos vseh tabel	da	da	ne	da	da	da
Prenos vseh podatkov	da	da	ne	da	da	da
Prenos vseh relacij med tabelami	ne	da	ne	da	da	da
Pregled ER diagrama	da	ne	ne	ne	ne	ne
Prenos iz MS Access 2010	da	da	ne	da	da	da
Ime tabele	ne	ne	ne	da	ne	ne
Polja v tabeli	ne	ne	da	da	ne	ne
Privzete vrednosti polj	ne	ne	da	ne	ne	ne
Natančnost podatkov	popolna	popolna	nepopolna	popolna	popolna	popolna
Konsistentnost podatkov	v celoti	v celoti	nepopolna	v celoti	v celoti	v celoti
Celovitost podatkov	v celoti	v celoti	nepopolna	v celoti	v celoti	v celoti
Obstoječa programska oprema	delno	ne	ne	delno	ne	ne
Obstoječa infrastruktura	ne	ne	ne	ne	ne	ne
Izvorni sistem	ne	ne	ne	ne	ne	ne
Druga orodja za analizo	da	da	da	da	da	ne
Druga orodja za podatkovno skladišče	da	ne	da	da	ne	ne
Ekstrakcija	da	da	ne	da	da	da
Transformacija	da	da	ne	da	da	da
Nalaganje	da	da	ne	da	da	da
Razvojna orodja	da	da	ne	da	da	da
Orodja za planiranje in analize	da	da	da	da	da	da
Orodja za upravljanje podatkovnega skladišča	da	da	ne	da	da	da
Sistem za opozorila in nadzor	da	ne	da	da	da	da
Stroški strojne opreme	do 1.000 EUR	do 1.000 EUR	do 1.000 EUR	do 5.000 EUR	do 5.000 EUR	do 5.000 EUR
Stroški programske opreme	brezplačno	brezplačno	brezplačno	med 1.000 EUR in 5.000 EUR	nad 5.000 EUR	med 1.000 EUR in 5.000 EUR
Stroški svetovanja	do 1.000 EUR	do 1.000 EUR	do 1.000 EUR	med 1.000 EUR in 2.000 EUR	med 1.000 EUR in 2.000 EUR	med 1.000 EUR in 2.000 EUR
Stroški izobraževanja	do 1.000 EUR	do 1.000 EUR	do 1.000 EUR	med 1.000 EUR in 2.000 EUR	med 1.000 EUR in 2.000 EUR	med 1.000 EUR in 2.000 EUR
Stroški vzdrževanja	do 500 EUR	do 500 EUR	do 500 EUR	med 500 EUR in 1.000 EUR	nad 1.000 EUR	med 500 EUR in 1.000 EUR
Stroški nadgradnje	do 500 EUR	do 500 EUR	do 500 EUR	nad 1.000 EUR	nad 1.000 EUR	med 500 in 1.000 EUR
Stroški dela	do 500 EUR	do 500 EUR	med 500 EUR in 1.000 EUR	nad 1.000 EUR	nad 1.000 EUR	nad 1.000 EUR
Tehnična podpora	odlična	zadovoljiva	zadovoljiva	odlična	odlična	zadovoljiva
On-line podpora	da	ne	ne	da	da	da
Spletno izobraževanje	da	da	da	da	da	da

Vir: Klenovšek, lastni prikaz (2014)

Vrednotenje kriterijev smo prikazali na dveh primerih in sicer za kriterij »hitrost poizvedbe«, ki smo jo izvedli na izbranih variantah, ter za kriterij »strošek programske opreme« za posamezno varianto.

Poizvedba, ki smo jih izvedli na vseh variantah, je bila enostavna, in sicer je vrnila prikaz vseh podatkov v tabeli, ki vsebuje največ podatkov, to je 60.031 vrstic. Tabela se imenuje splPartnerji.

SQL poizvedba, ki smo jo izvedli, je sledeča:

```
SELECT * FROM splPartnerji
```

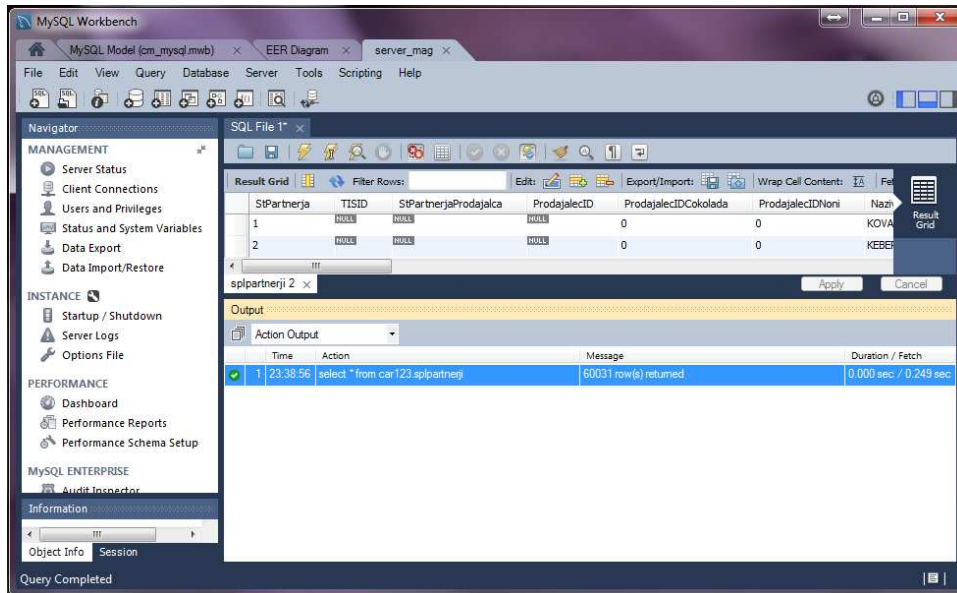
Pri MongoDB, ki ne podpira SQL, smo poizvedbo izvedli s sledečo poizvedbo:

```
{"StPartnerja" : {"$gt" : 0}}
```

Navedena poizvedba prikaže vse zapise, kjer je StPartnerja večja od 0.

Pri varianti MySQL je poizvedba trajala 0,249 sekunde (Slika 5.3), zato smo varianto ocenili z vrednostjo »med 2 in 5 sekund«.

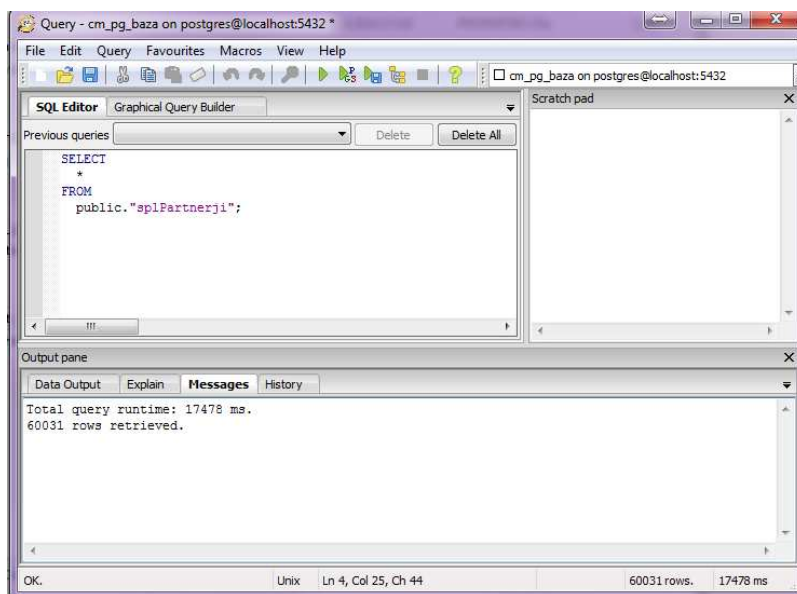
Slika 5.3: Izvajanje poizvedbe v MySQL



Vir: Klenovšek, lastni prikaz (2014)

Poizvedba v PostgreSQL je trajala 17,478 sekund (Slika 5.4), zato smo varianti na naši lestvici določili vrednost »več kot 5 sekund«.

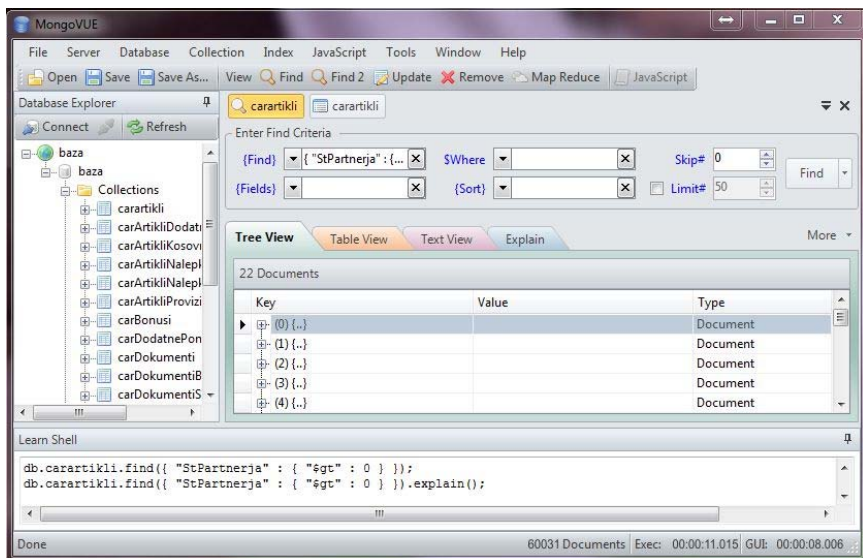
Slika 5.4: Izvajanje poizvedbe v PostgreSQL



Vir: Klenovšek, lastni prikaz (2014)

Poizvedbo v MongoDB smo izvedli s pomočjo grafičnega vmesnika MongoVUE (Slika 5.5), ki je samostojno orodje in omogoča delo s podatkovno bazo MongoDB. Poizvedba je trajala 11,015 sekund, zato smo varianti določili vrednost »več kot 5 sekund«.

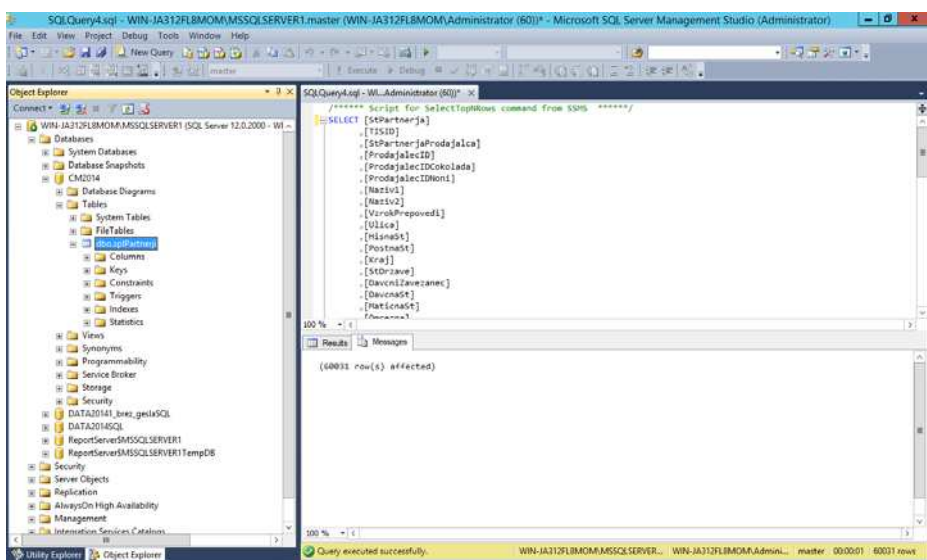
Slika 5.5: Izvajanje poizvedbe v MongoDB



Vir: Klenovšek, lastni prikaz (2014)

Poizvedba v MS SQL je trajala 1 sekundo (Slika 5.6), zato smo varianti določili vrednost »manj kot 2 sekundi«.

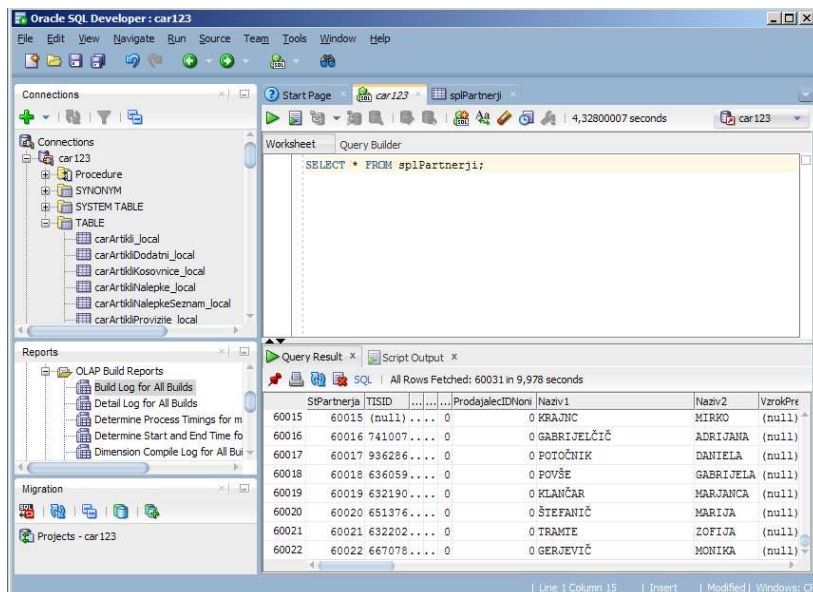
Slika 5.6: Izvajanje poizvedbe v MS SQL



Vir: Klenovšek, lastni prikaz (2014)

Poizvedba v Oracle 11g je potekala 9,978 sekund (Slika 5.7), zato smo varianti Oracle 11g določili vrednost »več kot 5 sekund«.

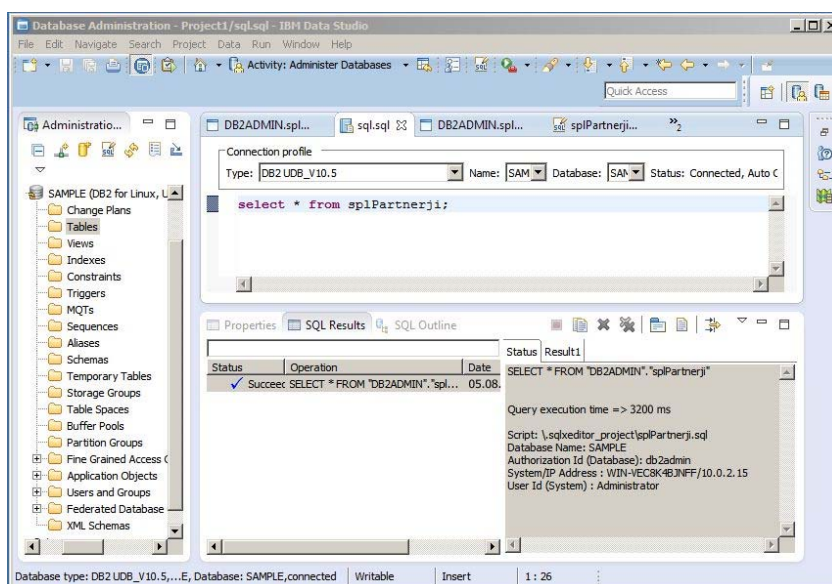
Slika 5.7: Izvajanje poizvedbe v Oracle



Vir: Klenovšek, lastni prikaz (2014)

Poizvedba v DB2 je trajala 3,2 sekunde (Slika 5.8), zato smo varianti DB2 določili vrednost »med 2 in 5 sekundami«.

Slika 5.8: Izvajanje poizvedbe v DB2



Vir: Klenovšek, lastni prikaz (2014)

Na uradnih spletnih straneh ponudnikov programske opreme smo pregledali cenike in ugotovili, da je programska oprema MySQL (Slika 5.9) brezplačna.

Slika 5.9: MySQL kot odprtokodna podatkovna baza



Vir: Klenovšek, lastni prikaz (2014)

Prav tako ugotovimo, da je programska oprema PostgreSQL (Slika 5.10) brezplačna.

Slika 5.10: PostgreSQL kot odprtokodna podatkovna baza



Q: How is PostgreSQL licensed? How much does it cost?

A: PostgreSQL is released under the OSI-approved PostgreSQL License. There is no fee, even for use in commercial software products. Please see <http://www.postgresql.org/about/licence>

Vir: Klenovšek, lastni prikaz (2014)

Programska oprema MongoDB (Slika 5.11) je prav tako brezplačna.

Slika 5.11: MongoDB kot odprtokodna podatkovna baza

MongoDB Licensing

Software and Documentation

MongoDB Database Server and Tools

- Free Software Foundation's [GNU AGPL v3.0](#).
- Commercial licenses are also available from [MongoDB, Inc.](#), including free evaluation licenses.

Vir: Klenovšek, lastni prikaz (2014)

Strošek programske opreme za MS SQL Server 2014 Standard Edition smo pridobili pri podjetju Repro – MS 03 d.o.o., Ljubljana (Slika 5.12). Strošek strežniške licence znaša 1.348,39 EUR (z DDV), strošek licence za enega uporabnika (za dostop do strežnika) pa znaša 313,81 EUR (z DDV). V podjetju bi potrebovali pet uporabniških licenc, tako bi skupni strošek nakupa znašal 2.917,43 EUR (z DDV), zato smo to varianto uvrstili v vrednost »med 1.000 EUR in 5.000 EUR«.

Slika 5.12: Strošek MS SQL

Pozdravljeni,
 Repro – MS 03 d.o.o. se ukvarja samo z distribucijo, zato lahko licence prodamo samo partnerjem za nadaljnjo prodajo.

Vseeno pa vam spodaj pošiljam **informativne PPC** za končnega kupca:

228-10344	SQLSvrStd 2014 SNGL OLP NL	1.105,24 €/kos brez DDV (strežniška licenca)
359-06098	SQLCAL 2014 SNGL OLP NL UstrCAL dostop do strežnika)	257,22 €/kos brez DDV (licenca za uporabnika – za

Vir: Klenovšek, lastni prikaz (2014)

Strošek programske opreme za varianto Oracle 11g smo preverili pri Oracle Slovenija (Slika 5.13). Strošek programske opreme za letni najem (vključene posodobitve) znaša 1.007,15 EUR, strošek programske opreme za strežnik znaša 4.578 EUR, strošek licence za enega uporabnika (za dostop do strežnika in enoletna podpora) znaša skupaj 173,26. Celoti strošek nakupa, skupaj s petimi uporabniškimi licencami, bi tako znašal 6.451,45 EUR, zato smo to varianto uvrstili v vrednost »več kot 5.000 EUR«.

Slika 5.13: Strošek Oracle 11g

Glede na število uporabnikov, pride v poštev le verzija Oracle Database Standard Edition one, ki ima minimalno število uporabnikov 5.
 Cena so v EUR in so sledeče:

Database Products	Named User Plus	Software Update License & Support	Processor License	Software Update License & Support	Notes
Oracle Database Standard Edition One	142	31,26	4.578	1.007,15	8, 32, 48

Kar pomeni, da je cena po uporabniku 142 EUR + 31,26 EUR enoletna podpora (nadgradnje, patchi, bug fixi). Na te cene dodate še davek.

Vir: Klenovšek, lastni prikaz (2014)

Strošek programske opreme za varianto DB2 (Slika 5.14) smo preverili pri partnerskem podjetju IBM-a, to je podjetje Marmis d.o.o. iz Ljubljane. Letni strošek nakupa petih uporabniških licenc bi znašal 1.256,60 EUR, zato smo varianto uvrstili v vrednost »med 1.000 EUR in 5.000 EUR«.

Slika 5.14: Strošek IBM DB2

Pozdravljeni,

podatkovna baza DB2 je na voljo v večih izdajah, izbira je odvisna od potreb in zahtev stranke.

Vse izdaje imajo enotno programsko jedro, razlikujejo se pa po naboru dodatnih funkcionalnosti in glede omejitev sistemskih resursov (število CPU jeder in količina RAM pomnilnika). Več o tem na:

<http://www.ibm.com/developerworks/data/library/techarticle/dm-1311db2compare/>

Posamezne izdaje in cene:

- DB2 Express (max 8 CPU jeder in 64GB RAM, minimalno naročilo je 5 uporabniških licenc) - licenca za 1. leto 206 € + DDV, naslednja leta 41,20 € + DDV

Vir: Klenovšek, lastni prikaz (2014)

5.10 Analiza z uporabo orodja DEXi

Variante smo ovrednotili z orodjem za podporo odločanju DEXi. Rezultati vrednotenja po celotni drevesni strukturi so prikazani v Tabeli 5.20. Glede na rezultate vrednotenja ugotavljamo, da je najbolj primerno podatkovno skladišče MySQL, ki bi v največji meri ustrezalo postavljenim kriterijem.

V nadaljevanju predstavljamo natančno grafično analizo, ki smo jo pridobili v orodju DEXi.

Tabela 5.20: Rezultati vrednotenja variant

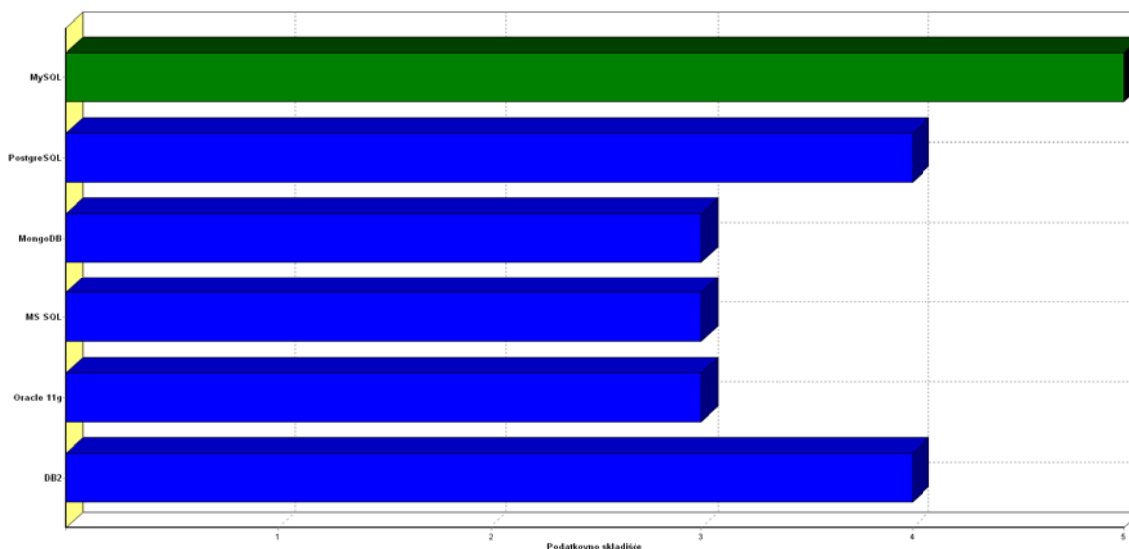
Kriterij	MySQL	PostgreSQL	MongoDB	MS SQL	Oracle 11g	DB2
Podatkovno skladišče	5	4	3	3	3	4
Tehnični kriteriji	5	4	1	4	4	4
Uporabniški vidik	odlično	dobro	slabo	dobro	dobro	odlično
Uporabniški vmesnik	Enostaven	Enostaven	Brez	Kompleksen	Kompleksen	Enostaven
Orodja za dostop	v celoti	delno	delno	v celoti	v celoti	v celoti
Poizvedbe in orodja za analizo	da	ne	ne	da	da	da
Napredna orodja za analizo	da	delno	delno	da	da	da
Funkcionalnost poizvedb	dobra	dobro	slaba	dobra	slaba	dobra
Hitrost poizvedb	med 2 in 5 sekund	več kot 5 sekund	več kot 5 sekund	do 2 sekund	več kot 5 sekund	med 2 in 5 sekund
Grafični vmesnik za poizvedbe	ne	da	ne	ne	ne	ne
Administratorski vidik	dobro	dobro	slabo	dobro	dobro	slabo
Namestitve	enostavno	enostavno	zahtevno	zahtevno	zahtevno	zahtevno
Trajanje namestitve	do 2 ur	do 2 ur	do 2 ur	več kot 2 ure	več kot 2 ure	več kot 2 ure
Težavnost namestitve	enostavno	enostavno	zahtevno	zahtevno	zahtevno	zahtevno
Migracija	dobra	dobro	slaba	dobra	dobro	dobra
Prenos vseh tabel	da	da	ne	da	da	da
Prenos vseh podatkov	da	da	ne	da	da	da
Prenos vseh relacij med tabelami	ne	da	ne	da	da	da
Pregled ER diagrama	da	ne	ne	ne	ne	ne
Prenos iz MS Access 2010	da	da	ne	da	da	da
Potreben popravek tabel	ne	ne	da	da	ne	ne
Ime tabele	ne	ne	ne	da	ne	ne
Polja v tabeli	ne	ne	da	da	ne	ne
Privzete vrednosti polj	ne	ne	da	ne	ne	ne
Kvaliteta podatkov	popolna	popolna	nepopolna	popolna	popolna	popolna
Natančnost podatkov	popoln	popoln	nepopoln	popoln	popoln	popoln
Konsistentnost podatkov	v celoti	v celoti	nepopoln	v celoti	v celoti	v celoti
Celovitost podatkov	v celoti	v celoti	nepopolna	v celoti	v celoti	v celoti
Kompatibilnost	ne	ne	ne	ne	ne	ne
Obstoječa programska oprema	delno	ne	ne	delno	ne	ne
Obstoječa infrastruktura	ne	ne	ne	ne	ne	ne
Zdržljivost	delno	delno	delno	delno	delno	ne
Izvorni sistem	ne	ne	ne	ne	ne	ne
Druga orodja za analizo	da	da	ne	da	ne	ne
Druga orodja za podatkovno skladišče	da	ne	da	da	da	ne
ETL funkcionalnost	dobra	dobro	slaba	dobra	dobro	dobra
Ekstrakcija	da	da	ne	da	da	da
Transformacija	da	da	ne	da	da	da
Nalaganje	da	da	ne	da	da	da

Administracija podatkovnega skladišča	dobri	dobra	slaba	dobra	dobra	dobra
Razvojna orodja	da	da	ne	da	da	da
Orodja za planiranje in analize	da	da	da	da	da	da
Orodja za upravljanje podatkovnega skladišča	da	da	ne	da	da	da
Sistem za opozorila in nadzor	da	ne	da	da	ne	da
Upravljaljski kriteriji	5	4	4	3	3	4
Stroški	nizki	nizki	nizki	visoki	visoki	srednji
Neposredni stroški	nizki	nizki	nizki	srednji	srednji	srednji
Stroški strojne opreme	do 1.000 EUR	do 1.000 EUR	do 1.000 EUR	do 5.000 EUR	do 5.000 EUR	do 5.000 EUR
Stroški programske opreme	brezplačno	brezplačno	brezplačno	med 1.000 EUR in 5.000 EUR	nad 5.000 EUR	med 1.000 EUR in 5.000 EUR
Stroški svetovanja	do 1.000 EUR	do 1.000 EUR	do 1.000 EUR	med 1.000 EUR in 2.000 EUR	med 1.000 EUR in 2.000 EUR	med 1.000 EUR in 2.000 EUR
Posredni stroški	nizki	nizki	nizki	visoki	visoki	srednji
Stroški izobraževanja	do 1.000 EUR	do 1.000 EUR	do 1.000 EUR	med 1.000 EUR in 2.000 EUR	med 1.000 EUR in 2.000 EUR	med 1.000 EUR in 2.000 EUR
Stroški vzdrževanja	do 500 EUR	do 500 EUR	do 500 EUR	med 500 EUR in 1.000 EUR	nad 1.000 EUR	med 500 EUR in 1.000 EUR
Stroški nadgradnje	do 500 EUR	do 500 EUR	do 500 EUR	nad 1.000 EUR	nad 1.000 EUR	med 500 in 1.000 EUR
Stroški dela	do 500 EUR	do 500 EUR	med 500 EUR in 1.000 EUR	nad 1.000 EUR	nad 1.000 EUR	nad 1.000 EUR
Ponudnik podatkovnega skladišča	odlični	dobri	dobri	odlični	odlični	odlični
Tehnična podpora	odlična	zadovoljiva	zadovoljiva	odlična	odlična	zadovoljiva
On-line podpora	da	ne	ne	da	da	da
Spletno izobraževanje	da	da	da	da	da	da

Vir: Klenovšek, lastni prikaz (2014)

Na podlagi vrednotenja (Slika 5.15) je varianta MySQL dobila oceno 5 (od 5) in je tako najbolj primerna izbira za podjetje Car marketing. Zelo se mu približata tudi odprtokodni PostgreSQL in komercialni IBM DB2, ki sta dosegla oceno 4, kar pomeni, da dokaj dobro izpolnjujeta postavljene kriterije, »slabo« oceno pa dobita pri kriterijih kompatibilnost in namestitvev programske opreme (za varianto DB2). Z oceno 3 pa sledijo ostale tri ocenjevane variante: MongoDB, MS SQL in Oracle 11g.

Slika 5.15: Grafični prikaz končne ocene

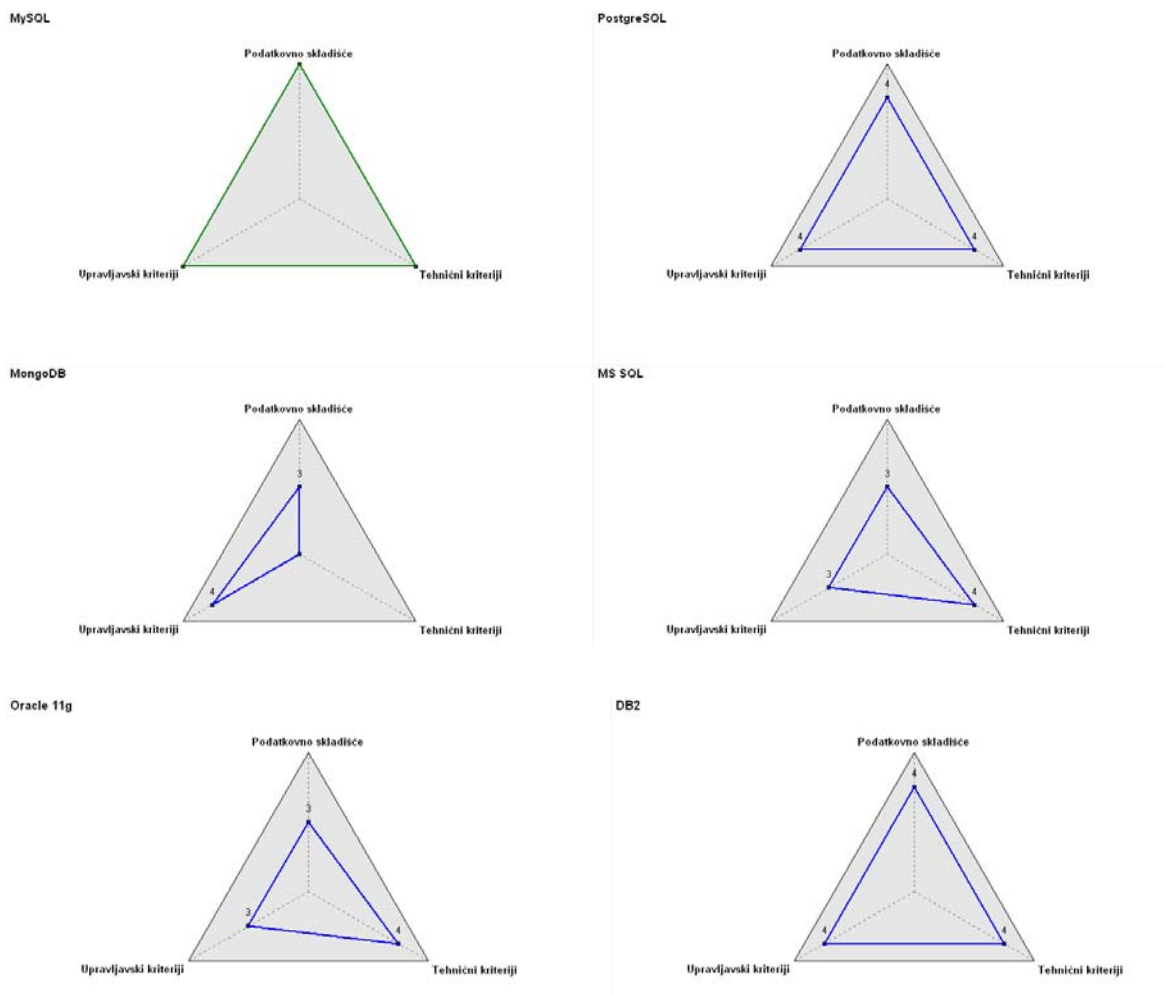


Vir: Klenovšek, lastni prikaz (2014)

V nadaljevanju bomo podrobneje grafično predstavili vpliv posameznih kriterijev na končno oceno ocenjevane variante.

Slika 5.16 ponazarja vpliv tehničnih in upravljaljskih kriterijev na končno oceno ocenjevanih podatkovnih skladišč, iz česar lahko razberemo, da MySQL v največji meri izpolnjuje oba kriterija z obeh področij, medtem ko MongoDB ne doseže zadovoljive ocene tehničnih kriterijev. Najslabšo oceno upravljaljskih kriterijev sta dosegla MS SQL in Oracle 11g, to je ocena 3, kar pomeni, da sta z vidika podpore in stroškov najmanj ugodni.

Slika 5.16: Vpliv tehničnih in upravljaljskih kriterijev na podatkovno skladišče

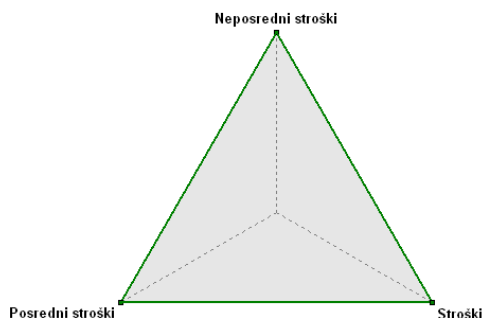


Vir: Klenovšek, lastni prikaz (2014)

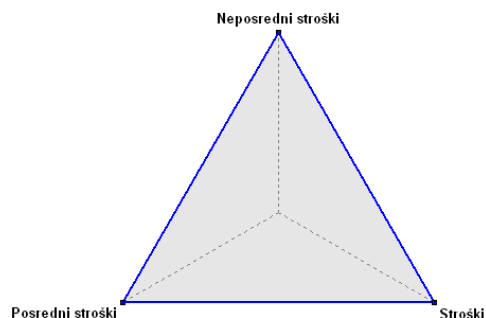
Če podrobneje pogledamo vpliv neposrednih in posrednih stroškov (Slika 5.17), so ocenjevane variante MySQL, PostgreSQL in MongoDB dosegle najvišjo možno oceno, medtem ko sta stroškovno najmanj ugodni MS SQL in Oracle 11g. Varianta DB2 je dosegla »srednjo« oceno za ta kriterij.

Slika 5.17: Vpliv neposrednih in posrednih stroškov na celotne stroške

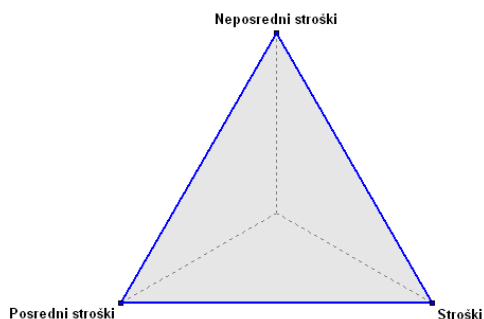
MySQL



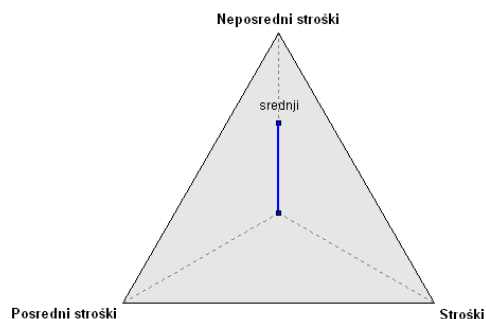
PostgreSQL



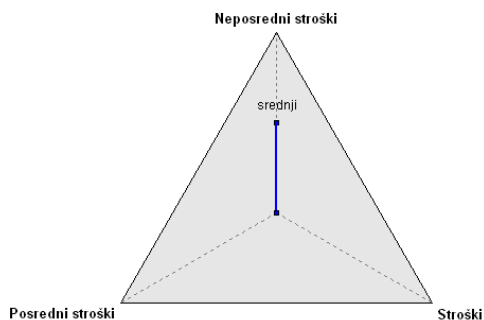
MongoDB



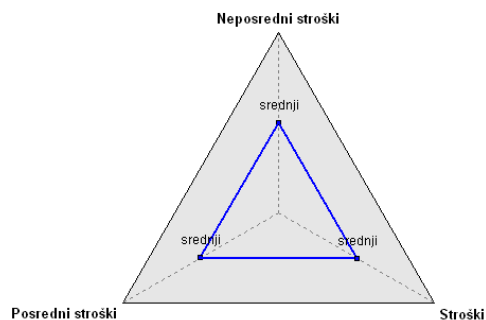
MS SQL



Oracle 11g



DB2

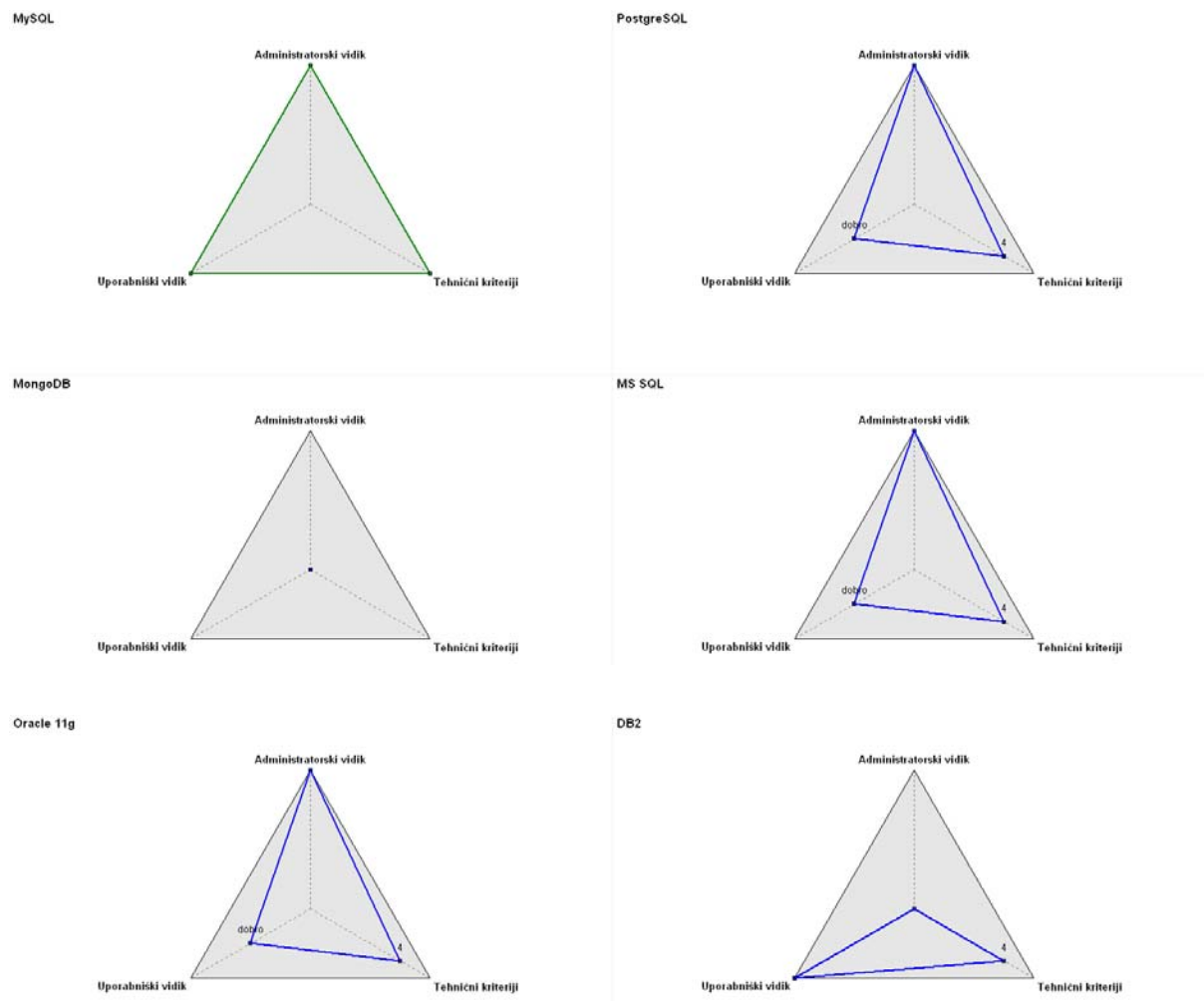


Vir: Klenovšek, lastni prikaz (2014)

Tehnični kriterij sestavljata administratorski in uporabniški vidik. Na Sliki 5.18 smo s pomočjo DEXi grafično ponazorili, v kolikšni meri posamezna varianta vpliva na oceno tehničnega kriterija. MySQL je tako pri administratorskem kot tudi uporabniškem vidiku dosegel »dobro« oziroma »odlično« oceno. Skupna ocena tehničnega kriterija pri varianti MongoDB je bila najslabša, saj je varianta dosegla oceno 1. Ocena administratorskega in uporabniškega vidika je bila »slaba«. Ostale variante so dosegle oceno 4, kjer so PostgreSQL,

MS SQL in Oracle 11g nekoliko odstopale pri uporabniškem vidiku, DB2 pa pri administratorskem vidiku, saj je dosegla oceno »slabo«.

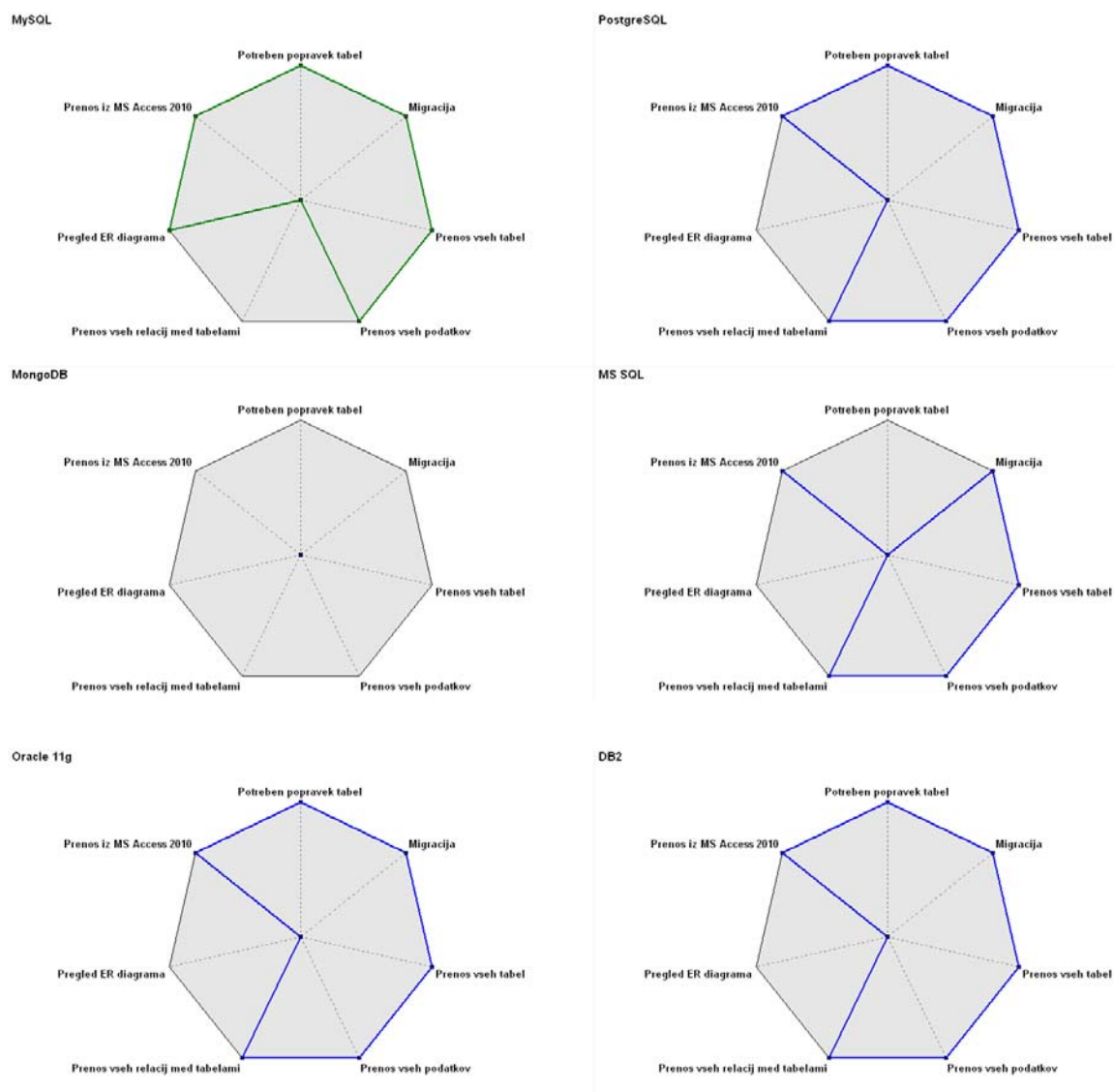
Slika 5.18: Vpliv administratorskega in uporabniškega vidika na tehnične kriterije



Vir: Klenovšek, lastni prikaz (2014)

Glavni kriteriji, ki so vplivali na oceno poddrevesa »Migracija«, so bili sledeči: potreben popravek tabel, prenos iz MS Accessa 2010, pregled ER diagrama, prenos vseh relacij med tabelami, prenos vseh podatkov in prenos vseh tabel. Na Sliki 5.19 je predstavljena grafična analiza, ki prikazuje, da MySQL najbolje dosega določene kriterije, težava se je pojavila le pri prenosu vseh relacij med tabelami. Najslabšo oceno je dosegla varianta Mongo DB, kjer smo imeli težave pri vseh navedenih kriterijih, saj je bil potreben korenit poseg v izvorno podatkovno bazo, da smo lahko le-to migrirali v MongoDB. Variante PostgreSQL, MS SQL, Oracle 11g ter DB2 dosegajo srednjo oceno kriterijev.

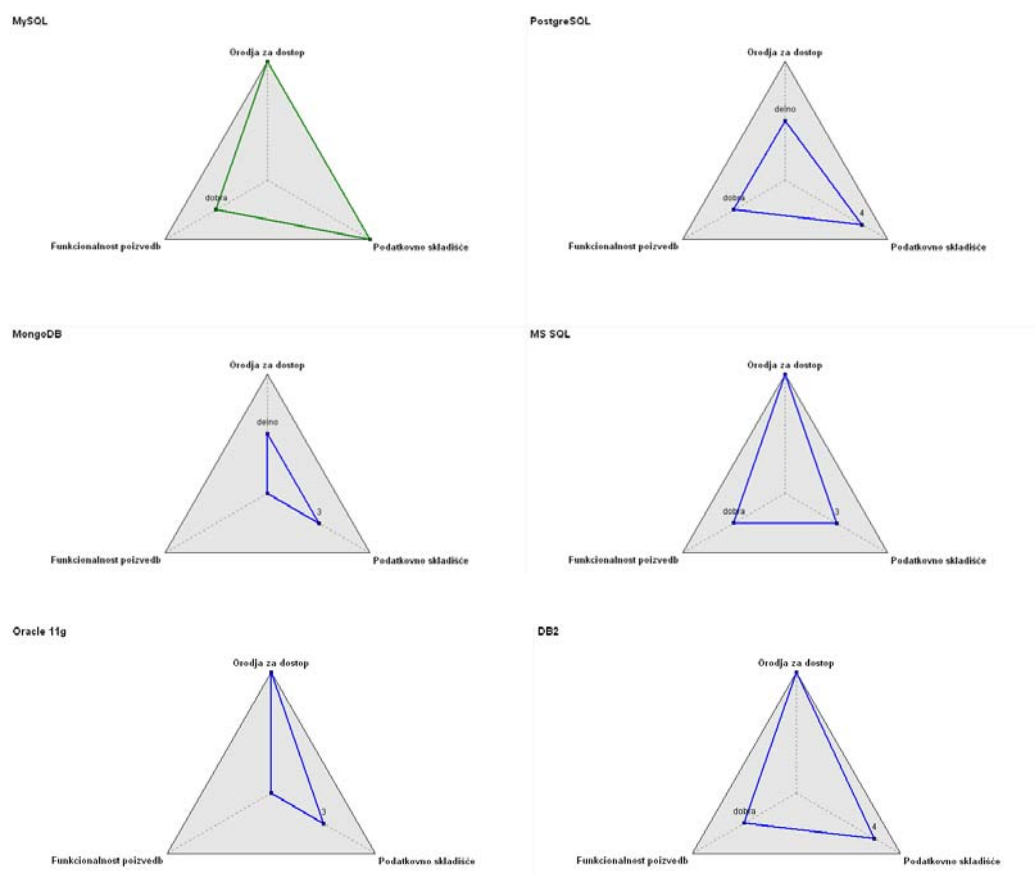
Slika 5.19: Vpliv kriterijev na migracijo



Vir: Klenovšek, lastni prikaz (2014)

Na Sliki 5.20 smo grafično ponazorili, kako dosežena ocena kriterija orodja za dostop in funkcionalnost poizvedb vplivata na končno oceno ocenjevanega podatkovnega skladišča. Variante MySQL, MS SQL, Oracle 11g in DB2 so dosegle odlično oceno pri kriteriju orodja za dostop, saj vse 4 variante podpirajo osnovna in napredna orodja za poizvedbe in analize. MongoDB je pri obeh kriterijih prejel slabo oceno.

Slika 5.20: Vpliv orodij za dostop in funkcionalnosti poizvedb na podatkovno skladišče



Vir: Klenovšek, lastni prikaz (2014)

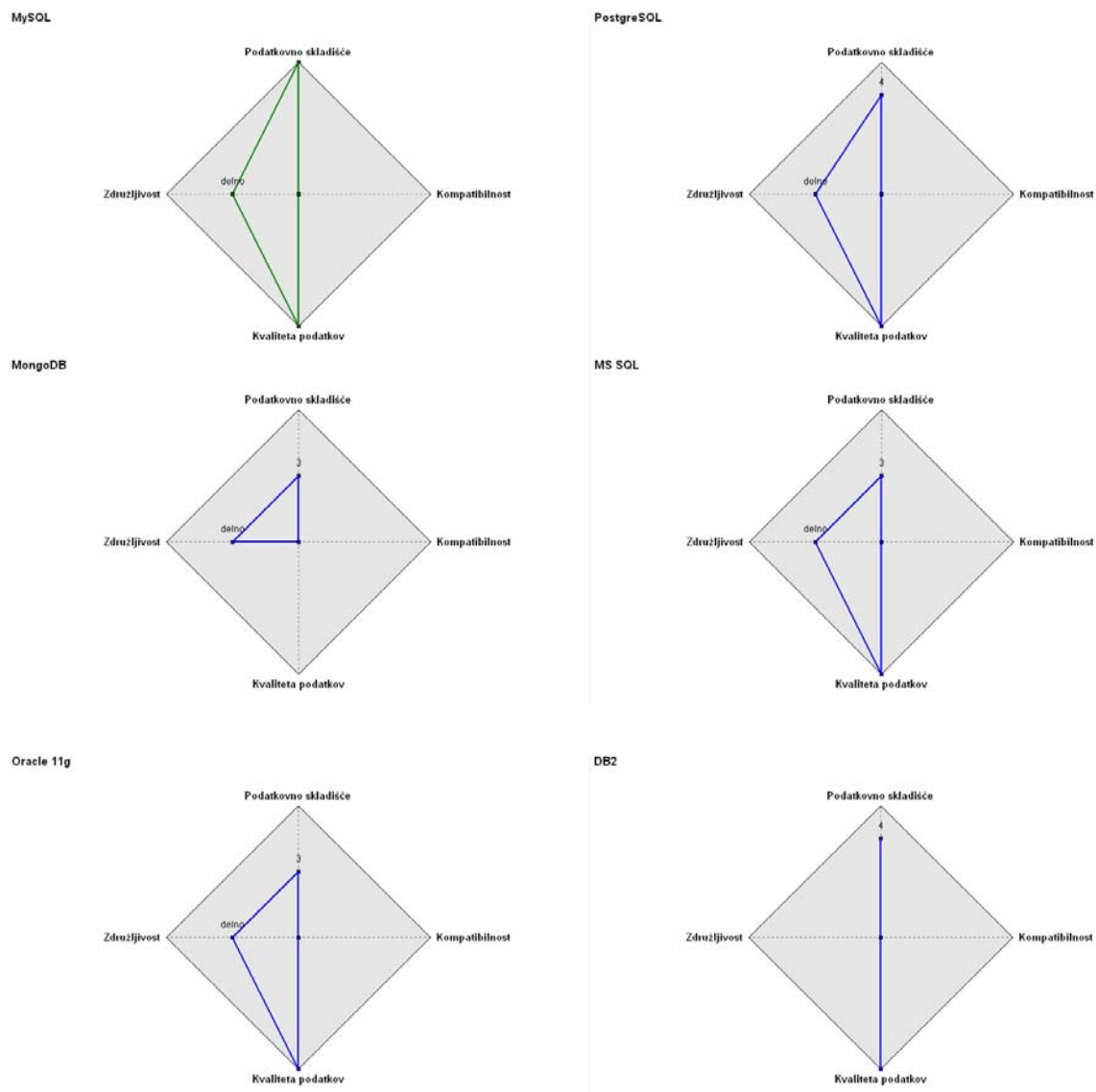
Združljivost oziroma kompatibilnost in kvaliteta podatkov so kriteriji administratorskega vidika odločitvenega drevesa. Kot lahko vidimo iz grafične analize na Sliki 5.21, nobena izmed ocenjevanih variant ne izpolnjuje kriterija kompatibilnosti z obstoječim sistemom, saj trenutna programska oprema in infrastruktura celotnega informacijskega sistema nista pripravljena za to, da bi lahko zagotovili kompatibilnost s podatkovnim skladiščem.

Pri kriteriju združljivost so vse ocenjevane variante razen DB2 ocenjene z oceno »delno«. Pri vseh variantah smo pri kriteriju združljivost z izvornim sistemom podali vrednost ne, saj trenutni sistem ni pripravljen za delovanje s katerim koli podatkovnim skladiščem. Pri kriteriju združljivost podatkovnega skladišča z drugimi orodji pri varianti DB2 ne zasledimo drugih orodij, ki bi podpirala tovrstno podatkovno skladišče.

Kvaliteta podatkov je pri vseh ocenjevanih variantah, razen MongoDB, ocenjena z vrednostjo popolna, saj so se vsi podatki prenesli v celoti. Kot smo že omenili, smo pri migraciji izvorne podatkovne baze v MongoDB morali celovito poseči v strukturo podatkovne baze in tudi v

same podatke, ki jih je bilo potrebno prilagoditi, zato je kriterij kvaliteta podatkov ocenjen kot »nepopolna«.

Slika 5.21: Vpliv kriterijev združljivost oziroma kompatibilnost in kvalitete podatkov na končno oceno podatkovnega skladišča



Vir: Klenovšek, lastni prikaz (2014)

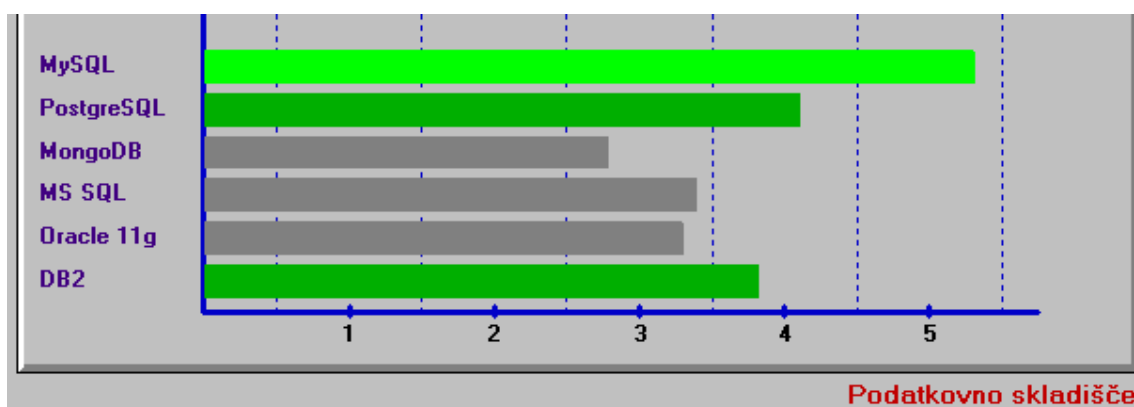
5.11 Analiza z uporabo orodja VREDANA

Programska oprema VREDANA nadgrajuje funkcionalnosti programa DEXi., kjer je vhodni podatek DEXi-jeva baza znanja s končnico .dax. VREDANA ima naslednjih pet glavnih funkcionalnosti, ki jih podpira. To so (VREDANA, 10. julij 2014):

- branje in pisanje datotek s končnico .dax,
- kombinirano vrednotenje kvalitativnih in kvantitativnih variant,
- grafični prikaz rezultatov (s stolpčnimi, korelacijskimi in zvezdnimi grafikoni),
- branje podatkov o variantah iz datotek s končnico .dif,
- »kaj-če« analiza, ki smo jo izvedli tudi mi.

Z orodjem VREDANA smo dobili grafični prikaz končne ocene izbora podatkovnega skladišča, ki je prikazan na Sliki 5.22. Kot lahko vidimo, rezultati ne odstopajo od grafičnega prikaza v orodju DEXi.

Slika 5.22: Prikaz končne ocene variant v VREDANI



Vir: Klenovšek, lastni prikaz (2014)

VREDANA omogoča »kaj-če« analizo, kar pomeni, da lahko spreminjamo zaloge vrednosti kriterijev pri posamezni varianti in spremljamo posledične spremembe rezultatov vrednotenja. Na takšen način lahko pridemo do boljših rezultatov vrednotenja, kar pomeni tudi boljše odločitve. »Kaj-če« analiza je dobrodošla v primerih, kadar okolja ne poznamo dovolj dobro ali pa za ugotovitev, kako naj bi se (se ne smejo) spreminjati zaloge vrednosti posameznih variant, da bi dosegli pričakovane rezultate (VREDANA, 10. julij 2014).

V okviru »kaj-če« analize smo se odločili, da bomo pri variantah MS SQL (Slika 5.23), Oracle 11g in DB2 spremenili kriterij neposrednih stroškov v primeru, da bi imeli možnost pridobiti brezplačne licence.

Slika 5.23: Sprememba zaloge vrednosti pri varianti MS SQL

Sprememba vrednosti variante

Varianta

Naziv: MS SQL

Opis:

Vrednosti

id.	Parameter	Vrednost
43.	Orodja za upravljanj	2
44.	Sistem za opozorila	2
48.	Stroški strojne opre	4
49.	Stroški programske o	3
50.	Stroški svetovanja	3
52.	Stroški izobraževanj	2
53.	Stroški vzdrževanja	2
54.	Stroški nadgradnje	1

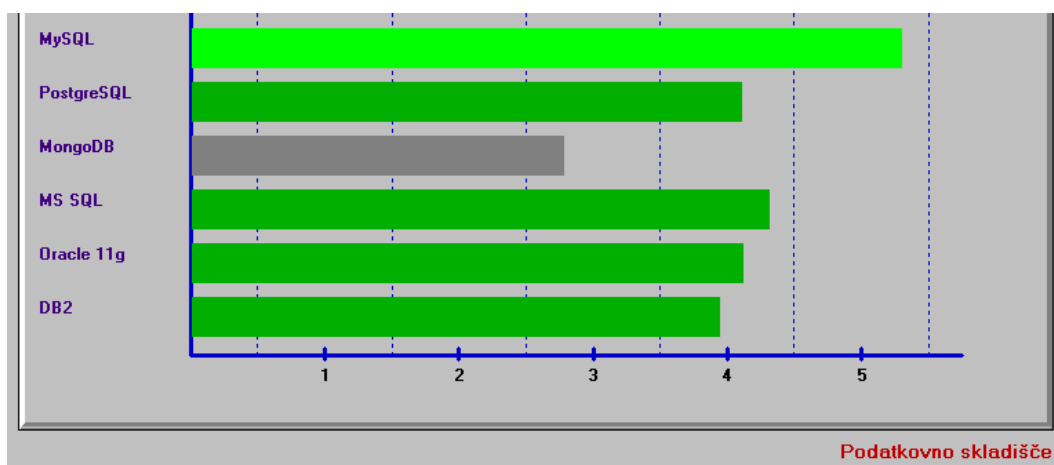
OK

Nezabi

Vir: Klenovšek, lastni prikaz (2014)

Po spremembi zaloge vrednosti ugotovimo (Slika 5.24), da sta se varianti MS SQL in Oracle 11g veliko bolj približali optimalni končni oceni, saj sta dosegli oceno več kot 4.5. Pri DB2 pa se navkljub izločitvi vpliv stroškov ocena ne izboljša.

Slika 5.24: »Kaj-če« analiza v VREDANI



Vir: Klenovšek, lastni prikaz (2014)

5.12 SWOT analiza

SWOT analiza je skupek ključnih elementov, ki izvirajo iz temeljite analize. Povzema, kaj želimo doseči in poudari bodoča dejanja. SWOT analiza naj bi odgovarjala na sledeča vprašanja (McDonald, 2007):

- Kaj stranke želijo?
- Kako kupujejo?
- Kaj počne konkurenca?
- Kako dobro izpolnjujemo potrebe strank?

SWOT je analiza, kjer smo za ocenjevano varianto MySQL prikazali prednosti (S, ang. »strengths«), slabosti (W, ang. »weaknesses«), priložnosti (O, ang. »opportunities«) in pasti (T, ang. »threats«).

Ker je ocenjevana varianta MySQL dobila največjo oceno, smo v Tabeli 5.21 predstavili še SWOT analizo.

Tabela 5.21: SWOT analiza variante MySQL

PREDNOSTI	SLABOSTI
<ul style="list-style-type: none"> • Brezplačno • Dobra povezljivost z MS Accessom (migracija je bila dokaj enostavna) • Dobra podprtost z orodji za poslovno inteligenco • Nizki stroški upravljanja • Enostavnost namestitve • ETL funkcionalnost 	<ul style="list-style-type: none"> • Nekompatibilnost z obstoječo infrastrukturo • Neuspešno migrirane relacije (kar pomeni dodatno delo)
PRILOŽNOSTI	PASTI
<ul style="list-style-type: none"> • Integracija z obstoječim informacijskim sistemom • Uporaba MySQL za izgradnjo novega informacijskega sistema, ki bo deloval v oblaku 	<ul style="list-style-type: none"> • Človeški faktor • Nepoznavanje arhitekture podatkovnega skladišča

Vir: Klenovšek, lastni prikaz (2014)

Ugotovili smo, da ima MySQL veliko prednosti, ena izmed teh je tudi ta, da je brezplačen. Pri migraciji iz izvirne podatkovne baze nismo imeli večjih težav, saj smo celoten postopek

migriranja lahko opravili preko ODBC podatkovnega vira. Preveriti in dopolniti smo morali le relacije med tabelami, kar je prineslo dodatno delo. Slabost vidimo v nekompatibilnosti z obstoječo infrastrukturo, vendar lahko to pripisujemo tudi zastareli strojni opremi in neustrezni programski opremi v podjetju.

Ker je informacijski sistem, ki uporablja izvorno podatkovno bazo, napisan v programskem jeziku Visual Basic, bi z nekaj dela in prilagoditvami lahko informacijski sistem prilagodili tudi na podatkovno bazo MySQL ter tako, po izvedbi ETL procesa, ustvarili infrastrukturo podatkovnega skladišča.

6. ZAKLJUČEK

Podatkovna skladišča in poslovno obveščanje sta tehniki, ki zagotavljata informacije in orodja podjetjem, da lahko na podlagi le-teh sprejemajo operativne in strateške poslovne odločitve. Vzpostavitev in upravljanje podatkovnega skladišča za namene poslovne inteligence je proces, ki je poln izzivov (Mundy in drugi, 2011).

Pri naši raziskavi smo sledili ciljem in predstavili pojem podatkovnega skladišča, vlogo in zgradbo podatkovnega skladišča, opisali pomen podatkovnih skladišč ter na podlagi določenih kriterijev s pomočjo orodja DEXi izdelali večparametrsko primerjalno analizo, kjer smo ugotovili, da je ocenjevana varianta MySQL dosegla največjo oceno in najbolje izpolnjuje zastavljene kriterije.

Pri naši raziskavi smo si postavili sledeča raziskovalna vprašanja:

1. Kateri kriteriji so v podjetju Car Marketing najbolj pomembni za migracijo podatkovne baze iz MS Access v podatkovna skladišča?
2. Kako odprtokodna in komercialna podatkovna skladišča zadovoljujejo izbrane kriterije?

Na podlagi strukturiranega intervjuja, ki smo ga izvedli z dvema osebama iz vodstva podjetja Car marketing (z direktorjem in skrbnikom sistema informatike) ter s preučitvijo strokovne literature s področja migracij podatkovnih baz v podatkovna skladišča smo pridobili kriterije

za oceno variant. S pomočjo definiranih kriterijev smo ovrednotili različne implementacije izbranih podatkovnih skladišč.

Ugotovili smo, da so pomembni kriteriji migracije tehnični in upravljavski. Pri tehničnih kriterijih smo upoštevali tehnični vidik izvedbe (migracija, poizvedbe, orodja za analize, ETL...), pri upravljavskem vidiku pa smo se osredotočili na stroške, ki nastanejo pri tovrstni migraciji in nadaljnjem vzdrževanju sistema. Glede na to, da je naročnik manjše podjetje, so to dokaj pomembni vidiki, vendar pa vseeno ne smemo zanemariti dejstva, da so celovite odločitve vseeno zelo pomembne in vplivajo na nadaljnji razvoj in obstoj samega podjetja.

Po vrednotenju ocenjevanih variant v orodju DEXi ugotavljamo, da je med odprtokodnimi in komercialnimi podatkovnimi skladišči največ odstopanj ravno pri kriteriju »stroški«, odstopanja pa se pojavijo tudi pri kriteriju »namestitvev«, saj je za namestitvev komercialnih podatkovnih skladišč potrebno dlje časa in so nekoliko bolj zahtevni. Ugotovili smo, da z vidika kvalitete prenesenih podatkov, ETL funkcionalnosti, podpore in uporabniškega vmesnika tako odprtokodna kot tudi komercialna v enaki meri zadostujejo kriterijem. Večja odstopanja se pojavijo le pri odprtokodni varianti MongoDB, ki omenjene kriterije zadovoljuje v najmanjši meri.

6.1 Možnosti nadaljnjega razvoja

Da bi v podjetju Car marketing lažje sledili poslovanju in sprejemali pomembne odločitve o prodaji, nabavi, strankah ... vsekakor potrebujejo sistem, ki bi to omogočal. Postavitev arhitekture podatkovnega skladišča bi zagotovo lahko izboljšala pregled nad vsemi poslovnimi procesi.

V nadaljnjem razvoju bi lahko s pomočjo že postavljenega virtualnega okolja in migriranih podatkov v ocenjevane variante izoblikovali ETL proces, ki bi samodejno potekal med izvorno podatkovno bazo in podatkovnim skladiščem. S pomočjo podatkov v podatkovnem skladišču bi preizkusili dejanska orodja za poslovno inteligenco ter tako preverili uspešnost odločitvenega procesa na različnih področjih poslovanja.

V podjetju bi želeli prenoviti informacijski sistem na način, da bi bil le-ta dostopen iz »oblaka«. V ta namen bi lahko s pomočjo nove podatkovne baze (predvidevamo MySQL),

vzpostavili povezavo med spletno aplikacijo in podatkovnim skladiščem ter tako zgradili celotno arhitekturo podatkovnega skladišča, ki bi služila za namene analiz in poročil.

7. LITERATURA IN VIRI

1. AGARWAL, ANURAG, RAMAMOORTI, SRIDHAR in JAYARAMAN, VAIDYANATHAN (2011) Decision support systems for strategic dispute resolution. *International Journal of Management & Information Systems*, 15 (4), str. 13–29.
2. ARIYACHANDRA, THILINI in WATSON, HUGH (2005) Key factors in selecting a data warehouse architecture. *Business Intelligence Journal*, 10 (2), str. 19–26.
3. ASEMI, ASEFEH, SAFARI, ALI in ZAVEREH, ADELEH (2011) The Role of Management Information System (MIS) and Decision Support System (DSS) for Manager's Decision Making Process. *International Journal of Business and Management*, 6 (7), str. 164–173.
4. BIGGAM, JOHN (2008) *Succeeding with your Master's Dissertation, A step-by-step handbook*. Velika Britanija: Bell and Bain Ltd.
5. BOHANEC, MARKO, JEREB, EVA in RAJKOVIČ, VLADISLAV (2003) *DEXi – Računalniški program za večparametrsko odločanje*. Kranj: Moderna organizacija.
6. BOHANEC, MARKO (2013) *DEXi: Program for Multi-Attribute Decision Making User's Manual*. Dostopno prek: <http://kt.ijs.si/MarkoBohanec/pub/DEXiManual400.pdf> (26.6.2014).
7. BOHANEC, MARKO (2013) *DEXi: Program for Multi-Attribute Decision Making User's Manual*. Dostopno prek: <http://kt.ijs.si/MarkoBohanec/pub/DEXiManual400.pdf> (26.6.2014).
8. BOUYSSOU, DENIS, DUBOIS, DIDIER, PRADE, HENRI in PIRLOT, MARC (2009) *Decision-Making Process: Concepts and Methods*. ZDA: John Wiley & Sons.
9. CHAUNDHURI, SURAJIT in DAYAL, UMESHWAR (1997) An Overview of Data Warehousing and OLAP Technology. *SIGMUND Record* 26 (1), str. 15–20.
10. CHODOROW, KRISTINA (2013) *MongoDB, The Definitive Guide, Second Edition*. Kalifornija: O'Reilly Media Inc.
11. DB2. Dostopno prek: <http://www-01.ibm.com/software/data/db2/linux-unix-windows/> (10.7.2014).
12. DB-Engines Ranking – Trend Popularity. Dostopno prek: http://db-engines.com/en/ranking_trend (2.7.2014).
13. GONZALES, MICHAEL (2003) *IBM Data Warehousing*. Indiana: Wiley Publishing, Inc.

14. GRIESEMER, BOB (2011) *Oracle Warehouse Builder 11gR2*. Velika Britanija: Packt Publishing Ltd.
15. HAMMERGREN, THOMAS in SIMON, ALAN (2009) *Data Warehousing For Dummies, Second Edition*. Hoboken: Wiley Publishing, Inc.
16. HOBBS, LILIAN (2005) *Oracle database 10g data warehousing*. Oxford: Elsevier Inc.
17. HUMPHRIES, MARK, HAWKINS, MICHAEL in DY MICHELLE (1999) *Data Warehousing: Architecture and Implementation*. New Jersey: Prentice Hall PTR.
18. INMON, WILLIAM H. (2005) *Building the Data Warehouse, Fourth Edition*. Kanada: John Wiley & Sons, Inc.
19. INMON, WILLIAM, STRAUSS, DEREK in NEUSHLOSS, GENIA (2008) *DW2.0: The Architecture for the Next Generation of Data Warehousing*. Združene države Amerike: Elsevier Inc.
20. JÜRGENS, MARCUS (2002) *Index Structures for Data Warehouses*. Tokyo: Springer.
21. KEDAR, SEEMA (2009) *Database Management Systems*. India: Technical Publications Pune.
22. KHAN, ARSHAD (2003) *Data Warehousing 101: Concepts and Implementation*. Združene države Amerike. The Canton Group, LLC.
23. KIMBALL, RALPH, ROSS, MARGY, THORNTHWAITTE, WARREN, MUNDY, JOY in BECKER, BOB (2008) *The Data Warehouse Lifecycle Toolkit, Second Edition*. Indiana: Wiley Publishing Inc.
24. KIMBALL, RALPH in ROSS, MARGY (2013) *The Data Warehouse Toolkit, Third Edition*. Indiana: John Wiley & Sons, Inc.
25. MALINOWSKI, ELZBIETA in ZIMÁNYI, ESTEBAN (2009) *Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications (Data-Centric Systems and Applications)*. Berlin: Springer.
26. MCDONALD, MALCOM (2007) *Marketing Plans: How to prepare them, how to use them, Sixth edition*. Velika Britanija: Elsevier Ltd.
27. MongoDB. Dostopno prek: <http://www.mongodb.org/> (5.7.2014).
28. MySQL AB (2007) *Enterprise Data Warehousing with MySQL*. Dostopno prek: <http://www.scribd.com/doc/3003152/Enterprise-Data-Warehousing-with-MySQL> (2.8.2014).
29. MySQL. Dostopno prek: <http://www.mysql.com/> (23.6.2014).
30. NAGABHUSHANA, S. (2006) *Data Warehousing: OLAP and Data Mining*. New Delhi: New Age International (P) Ltd.

31. Oracle Database. Dostopno prek: http://docs.oracle.com/cd/E11882_01/server.112/e25554/whatsnew.htm#DWHSG8061 (5.7.2014).
32. Oracle Warehouse Builder User's Guide. Dostopno prek: http://docs.oracle.com/cd/B28359_01/owb.111/b31278/concept_overview.htm#WBDOD10100 (5.7.2014).
33. PostgreSQL. Dostopno prek: <http://www.postgresql.org> (2.7.2014).
34. POWER, DANIEL (2002) *Decision support systems, Concepts and resources for managers*. Združene države Amerike: Greenwood Publishing Group, Inc.
35. POWER, DANIEL (2009) *Decision Support Basics*. New York: Business Expert Press.
36. PRABHU, S. in VENATESAN, N. (2007) *Data Mining and Warehousing*. New Delhi: New Age International Ltd.
37. PRABHU, C. S. R. (2008) *Data Warehousing: Concepts, Techniques, Products and Applications, Third Edition*. New Delhi: PHI Learning Private Limited.
38. RAINARDI, VINCENT (2008) *Building a Data Warehouse: With Examples in SQL Server*. New York: Springer.
39. RAMAKRISHNAN, RAGHU in GEHRKE, JOHANNES (2003) *Database Management Systems, Third Edition*. New York: McGraw-Hill.
40. RAVINDRANATH, B. (2003) *Decision Support Systems and Data Warehouses*. New Delhi: New Age International Ltd.
41. ROBBINS, STEPHEN in DECENZO, DAVID (2001) *Fundamentals of Management*. New Jersey: Prentice Hall.
42. ROOT, RANDAL in MASON, CARYN (2012) *Pro SQL Server 2012 BI Solutions*. New York: Apress Media LLC.
43. SAATY, THOMAS in VARGAS, LUIS (2012) *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process, Second Edition*. New York: Springer.
44. SAUTER, VICKI (2010) *Decision Support Systems for Business Intelligence*. New Jersey: John Wiley & Sons, Inc.
45. SQL Server 2012 - 2014. Dostopno prek: <http://www.microsoft.com/en-us/server-cloud/products/sql-server/Explore.aspx#fbid=fna3fTH1Lux> (26.6.2014).
46. THOMSEN, ERIK (2002) *OLAP Solutions: Building Multidimensional Information Systems, Second Edition*. Kanada: John Wiley & Sons, Inc.
47. VREDANA. Dostopno prek: <http://lopes1.fov.uni-mb.si/dex/vredana/> (10.7.2014).

48. WREMBEL, ROBERT in KONCILIA, CHRISTIAN (2007) *Data Warehouses and OLAP: Concepts, Architectures, and Solutions*. Združene države Amerike: IRM Press.
49. ZHANG, CHUNSHENG (2012) *The Intelligent Decision Supporting Technology of Cloud-Based Public Data Warehouse*. *Computer and Information Science*, 5 (2), str. 75–78.

PRILOGA

Delno strukturiran intervju

Priloga: Delno strukturiran intervju

Intervju z osebama: Marjan Kunaj (direktor) in skrbnik informacijskega sistema

Datum intervjuja: 4.8.2014

Intervju vodila: Nataša Klenovšek

VPRAŠANJA:

1. Funkcionalnost.

a. *Kaj je poslanstvo podjetja?*

Poslanstvo podjetja je, da bi stranke podjetja, z uporabo naših izdelkov, imele več življenjske energije, ker jim bo to pomagalo, da bodo še bolj uspešne na vseh področjih svojega življenja.

b. *Kako podjetje izpolnjuje in dosega zastavljeno poslanstvo?*

Z nenehno raziskavo tržišča in zahtev potrošnikov ter z inovativnimi idejami prodajne ekipe, oziroma celotnega tima. Nabor izdelkov nenehno dopolnjujemo in poskušamo biti prvi na tržišču, ki izdelek nudimo strankam. Pri izdelku poskušamo doseči najvišjo kvaliteto, ki je podprta s certifikati kakovosti.

c. *Kako podjetje ve, ali zastavljeno poslanstvo prinaša uspeh?*

Z zadovoljnimi strankami, ki samodejno naročajo izdelke podjetja in pohvalijo njihovo kakovostno delovanje. In seveda s pozitivnim poslovnim izidom.

d. *Kaj so ključni kazalniki uspešnosti in kritični dejavniki uspeha?*

Ključni kazalniki uspešnosti: zadovoljne svetovalke in stranke, samodejna mesečna naročila, donosnost ter likvidnost podjetja.

Kritični dejavniki uspeha: vodstvo podjetja, skupna vizija vseh udeležениh v poslovnem procesu, zaupanje in ekipno sodelovanje.

2. Stranke.

a. *Na kakšen način podjetje grupira svoje stranke?*

Po starostnih skupinah in geografskih območjih.

b. *Se te skupine spreminjajo v določenem časovnem obdobju?*

Ne.

c. *Ali grupiranje vpliva na obnašanje do strank?*

Za obdobje 5 let zelo vpliva.

d. *Katere informacije podjetje hrani o strankah?*

Ime, priimek, naslov, telefonska številka, letnica rojstva, naročeni izdelki

e. *Katere demografske podatke uporablja podjetje?*

Območna skupina (01, 02, 03, 04, 05, 07)

f. *Ali je potrebno slediti vsaki stranki posebej?*

Da, še posebno glede naročil izdelkov in spremljanja njihovih potreb glede na starost stranke.

3. Dobiček.

a. *Na kateri ravni podjetje meri donosnost posamezne skupine? Na uslužbenca? Na stranko? Po regijah?*

Donosnost se meri glede na svetovalko in stranko. Se pravi, kolikor svetovalke prodajo in stranke dejansko plačajo, iz tega nastane donosnost podjetja.

b. *Kako podrobno podjetje meri stroške in prihodke?*

Stroški in prihodki so razdeljeni na tri segmente, ki jih mesečno spremljamo.

c. *Kako podjetje meri in spremlja stroške in prihodke?*

Na 3 segmentih, kjer imamo stroške in prihodke z analizo posameznih podsegmentov. Podsegmenti so točkovani glede na procese, ki jih ima podjetje, da se lahko potem odločamo mesečno glede stroškov in prihodkov.

d. *Kakšno poročilo o dobičku podjetje trenutno uporablja?*

EBTN.

4. Sistem.

a. *Katere sisteme podjetje uporablja pri svojem poslovanju?*

Sistem za poslovanje (izdaja računov, pregled nad strankami, pregled in obračun plač), posebej sistem za računovodstvo in pa CRM za stranke.

b. *Katerih sistemov se zavedajo drugi udeleženci v procesu?*

Svetovalke, ki skrbijo za komunikacijo s strankami, niso seznanjene z vsemi procesi, ki so potrebni za uspešno delovanje sistema, poznajo pa del, kjer se izdajajo računi.

c. *Kakšen način ročnega prenosa podatkov ima podjetje, v primeru, da podatki niso na voljo?*

Ročni prenos podatkov je možen samo pri prenosu podatkov o naročilu strank, saj se ti podatki vsakodnevno ročno vpisujejo v za to namenjene obrazce, ki se potem vnašajo v sistem za izdajo računov.

5. Čas.

a. *Koliko časa (mesecev, let) podjetje hrani podatke?*

1 leto.

b. *Ali podjetje analizira podatke po letih?*

Da, ampak samo za segmente, ki so nam zelo pomembni.

c. *Kako pogosto podjetje želi poročilo (dnevno, tedensko, mesečno, na vsako četrtletje, letno)?*

Mesečno.

d. *Kako hitro podjetje želi podatke (primer: želi včerajšnje podatke videti danes)?*

V roku 48ur.

e. *Kako hitro, ob koncu določenega obdobja, želi podjetje videti podatke?*

V roku 48ur.

6. Poizvedbe in poročila.

a. *Katera poročila podjetje trenutno uporablja?*

Finančno poročilo, poročilo o izdelkih, poročilo o strankah, poročilo o prodajnih rezultatih svetovalke.

b. *Katere informacije se uporabljajo v trenutnih poročilih?*

Finančno : odhodki in prihodki, stroški za nabavo artiklov, drugi stroški

O izdelkih: strošek posameznega izdelka, preden ga stranka prevzame

O strankah: interval nakupov strank, starost stranke, kaj kupuje,

O prodajnih rezultatih svetovalk : % plačanih in % zavrženih,

c. *Kako pogosto se izdelujejo poročila?*

Enkrat mesečno.

d. *Ali se poročila izdelujejo dovolj pogosto?*

Nekatera ja, nekatera pa tudi, ker je obsežnost podatkov prevelika, da bi jih analizirali.

e. *Kdo izdeluje poročila?*

Skrbnik sistema informatike.

7. Izdelek.

a. *Katere izdelek podjetje prodaja in kako jih klasificira?*

Podjetje prodaja izdelke s področja dodatkov k prehrani.

b. *Ima podjetje hierarhijo izdelkov?*

Ima glede na segment strank.

c. *Ali se analize izdelujejo za vse izdelke hkrati ali se analize izdelujejo za posamezni izdelek?*

Analize se izdelujejo za vsak izdelek posamezno in sicer se izdeluje mesečna prodaja vsakega izdelka (po kosih).

d. *Kako podjetje upravlja spremembo v hierarhiji izdelkov?*

Spremlja trend na trgu, časovno obdobje v letu glede na to za kaj stranke potrebujejo izdelke, dobičkonosnost in obrat izdelka.

8. Geografija.

a. *Ali podjetje deluje več kot samo na eni lokaciji?*

Podjetje deluje na lokaciji v Krškem.

b. *Ali podjetje deli tržišče na geografska območja?*

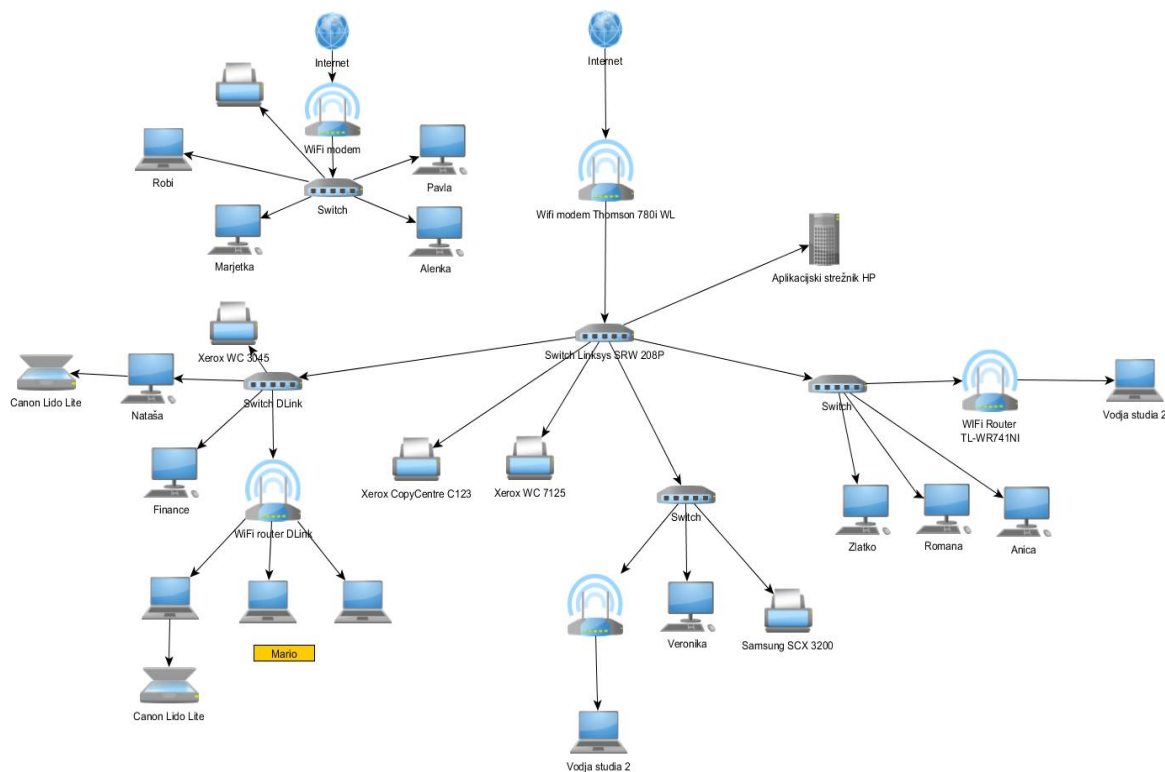
Da.

c. *Ali podjetje sledi prodaji glede na geografsko območje?*

Ne, ker nima zato analitike.

9. Trenutna arhitektura.

a. *Kakšna je trenutna tehnološka arhitektura podjetja?*



Strojna oprema

V sistemu je 11 osebnih računalnikov in prenosnih računalnikov različnih proizvajalcev. Za vzdrževanje in servis teh računalnikov skrbi zaposlen informatik, občasno tudi zunanji izvajalec.

Programska oprema

Na delovnih postajah so različni operacijski sistemi od Windows XP do Windows 7. Na vseh delovnih postajah je tudi programski paket Microsoft Office. Nekateri naprednejši uporabniki pa imajo instalirane še dodatne programe, ki niso potrebni za delovanje poslovnih procesov.

Podatki

Vsi uporabniki imajo svoje osebne in nekatere delovne podatke na osebnih računalnikih, predvsem zaradi pomanjkanja prostora na datotečnem strežniku. Nekateri uporabniki med seboj izmenjujejo podatke preko map v skupni rabi.

Vodstvo podjetja ima večino podatkov zaupne narave na svojih računalnikih, da ne bi prihajalo do zlorab.

Protivirusna in druge zaščite

Na vseh delovnih postajah je nameščen protivirus Nod32, ki je nastavljen na avtomatsko posodabljanje. Vse delovne postaje so zaščitene z gesli.

b. Kakšne vrste sistemov, strojne opreme, podatkovnih skladišč, omrežje, uporabniških orodij, razvojnih orodij in orodij za dostop podatkov so trenutno v uporabi?

Kot že omenjeno, podjetje uporablja dva glavna sistem za poslovanje. Eden je »centralni« v katerega se vnašajo vsi podatki, drugi pa je računovodski, kamor se (preko prenosljive datoteke) prenašajo podatki iz centralnega sistema.

10. Relacije izvornega sistema.

a. *Ali so izvorni sistemi kakor koli medsebojno povezani?*

Preko prenosljive datoteke, ste med seboj povezana poslovni in računovodski sistem.

b. *Ali en sistem zagotavlja informacije drugemu?*

Da.

c. *Ali so sistemi na kakršen koli način integrirani?*

Ne.

d. *V primerih, kjer ima več sistemov podatke o strankah in produktih, kateri sistem »služi« kot kopija?*

Podjetje nima kopije sistema.

11. Omrežne zmogljivosti.

a. *Ali je možno uporabiti en računalnik za dostop do različnih operacijskih sistemov, iz katere koli lokacije?*

Ne.

12. Kvaliteta podatkov.

a. *Koliko čiščenja, podvajanja in integracije mislite, da bo potrebno?*

Po hitrem pregledu podatkov v bazi predvidevamo, da bo kar nekaj dela s samimi podatki, da se bo dosegel želeni rezultat.

b. *Katera področja (tabele ali polja) v izvornih sistemih so trenutno poznani, da vsebujejo slabo kvaliteto podatkov?*

Segment finance, prodajne svetovalke in pa segment izdelki ima slabo podatkovno bazo podatkov.

13. Dokumentacija.

a. *Koliko dokumentacije je trenutno na voljo za izvorni sistem?*

Za fakturiranje strank je narejeno.

b. *Kako točni in posodobljeni so ti priročniki in referenčna gradiva?*

Za uporabniško izkušnjo so narejeni.

c. *Poskusite dobiti sledeče podatke kadar koli je to možno: kopijo priročnikov in referenčne dokumentacije, velikost podatkovne baze, načrtovane izboljšave, velikost varnostne kopije, obseg varnostne kopije in medij varnostne kopije, obseg podatkov v sistemu (primer: pomembne tabele in polja), kodeksi omrežja in njihov pomen ter ključne generacijske sheme.*

Nimamo podatkov.

14. Možni mehanizmi za ekstrakcijo.

a. *Kateri mehanizmi ekstrakcije so možni s trenutnim sistemom?*

Sistem je postavljen v bazi MS Access, tako da so podatki dostopni neposredno iz podatkovne baze. Ni pa možno ugotoviti kateri podatki so bili spremenjeni ali posodobljeni.

b. *Katere mehanizme ekstrakcije ste že uporabljali s tem sistemom?*

Še nismo delali na tem.

15. Kakšna so pričakovanja po migraciji podatkovne baze v podatkovno skladišče?

Želeli bi preizkusiti podatkovno skladišče, ter preveriti kakovost prenesenih podatkov, saj je to za nas zelo pomembno. Ne bi želeli, da je s tem preveč dela, želimo enostavnost in funkcionalnost. Da bo za uporabnike enostavno.

16. Koliko ste finančno pripravljeni vložiti v izvedbo?

Nekje med 1.000 in 2.000 EUR.

17. Na kakšen način sedaj izvajate analize?

Ročno, v Excelovi tabeli.