

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Marko Horvat

POSLOVNA INTELIGENCA

DIPLOMSKA NALOGA
VISOKOŠOLSKEGA STROKOVNEGA ŠTUDIJA

MENTOR: dr. Igor Rožanc

Ljubljana, 2009



Št. naloge: 00444/2009

Datum: 05.04.2009

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Kandidat: **MARKO HORVAT**

Naslov: **POSLOVNA INTELIGENCA**
BUSINESS INTELLIGENCE

Vrsta naloge: Diplomsko delo visokošolskega strokovnega študija

Tematika naloge:

Pojem poslovna inteligenca obsega različna znanja, tehnologije in orodja, ki omogočajo vodstvu boljši vpogled in poslovno odločanje na podlagi zbranih podatkov organizacije.

V diplomski nalogi sistematično in poglobljeno predstavite tri glavna področja poslovne inteligenca: podatkovna skladišča, orodja za pripravo podatkov ter orodja poslovne inteligenca. Jedro naloge naj bo celovita predstavitev orodij, ki sta jih v ta namen razvila dva najpomembnejša ponudnika na tem področju v Sloveniji: podjetji Oracle in Microsoft. V zadnjem delu naloge opravite v skladu z izbranimi kriteriji še objektivno primerjavo orodij obeh ponudnikov.

Mentor:

viš. pred. dr. Igor Rožanc



Dekan:

prof. dr. Franc Solina

IZJAVA O AVTORSTVU
diplomskega dela

Spodaj podpisani/-a **Marko Horvat**,

z vpisno številko **63000041**,

sem avtor/-ica diplomskega dela z naslovom:

Poslovna inteligenca

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal/-a samostojno pod mentorstvom (naziv, ime in priimek)
dr. Igor Rožanc
in somentorstvom (naziv, ime in priimek)

/
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela
- soglašam z javno objavo elektronske oblike diplomskega dela v zbirki »Dela FRI«.

V Ljubljani, dne **15.09.2009**

Podpis avtorja/-ice: _____

Zahvala

Zahvaljujem se mentorju dr. Igorju Rožancu, ki mi je bil v veliko pomoč in oporo. Poleg tega bi se zahvalil staršem in vsem prijateljem, ki so me podpirali in vzpodbujali pri izdelavi diplomskega dela. Izpostavil bi tudi prijateljico Marušo, ki mi je pomagala pri lektoriranju besedila in se ji ob tej priložnosti še posebej zahvalil.

Kazalo

Povzetek	1
Abstract.....	2
1. Uvod	3
1.1. Povod.....	3
1.2. Namen.....	3
1.3. Cilji	3
2. Poslovna inteligenca.....	4
2.1. Kaj je poslovna inteligenca?.....	4
2.2. Zgodovina poslovne inteligence.....	4
2.3. Delovanje sistemov za poslovno inteligenco.....	5
2.4. Rešitve	7
3. Orodja	8
3.1. Ključne kategorije orodij poslovne inteligence.....	9
3.1.1. Preglednice	9
3.1.2. Programi za poročanje in poizvedbe	10
3.1.3. Sprotne transakcijske in analitične obdelave.....	10
3.1.4. Digitalna pregledna plošča	16
3.1.5. Podatkovno rudarjenje.....	16
3.1.6. Procesno rudarjenje	17
3.1.7. Upravljanje poslovne uspešnosti	18
3.2. Orodje za pripravo podatkov	19
3.2.1. Črpanje podatkov iz zunanjih virov.....	19
3.1.1. Preverjanje pravilnosti in kakovosti podatkov	19
3.2.2. Nalaganje v podatkovno skladišče	20
3.2.3. Življenjski cikel procesa za pripravo podatkov.....	20
3.1.2. Elementi za kvaliteten proces priprave podatkov.....	21
3.3. Podatkovno skladišče	22
4. Ponudniki rešitev	23
4.1. Oracle	23
4.1.1. Orodja poslovne inteligence	23
4.1.2. Orodja za pripravo podatkov	27
4.1.3. Podatkovno skladišče	28
4.2. Microsoft	28
4.2.1. Orodja poslovne inteligence	28
4.2.2. Orodja za pripravo podatkov	31
4.2.3. Podatkovno skladišče	32
5. Primerjava ponudnikov	34
5.1. Varnost	34
5.2. Razvojna orodja.....	34
5.3. Poslovna inteligenca.....	35
5.4. Integracija z standardnimi produkti.....	36
5.5. Cena.....	36
5.6. Neodvisnost od platforme.....	37
5.7. Gartner Group analiza	38
5.8. Analize TPC	39
6. Sklepne ugotovitve	44
7. Seznam uporabljenih virov	46

Seznam slik

Slika 1: Shema sistema poslovne inteligence	9
Slika 2: Primer OLAP kocke	12
Slika 3: Primer zvezdne sheme	13
Slika 4: Primer snežinkaste sheme	14
Slika 5: Primer digitalne pregledne plošče	16
Slika 6: Shema sistema poslovne inteligence	19
Slika 7: Shema sistema poslovne inteligence	22
Slika 8: Uporabniški vmesnik paketa Oracle Business Intelligence Answers	23
Slika 9: Uporabniški vmesnik BI Interactive Dashboards	24
Slika 10: Zaslonska slika Business Intelligence Delivers	25
Slika 11: Primer spletnega uporabniškega vmesnika	26
Slika 12: Primer obdelave podatkov v Excelu	26
Slika 13: Zaslonska slika izdelave procesa ETL z ODI	27
Slika 14: Zaslonska slika izdelave ETL z OWB	28
Slika 15: Primer zaslonske slike SQL Server Reporting Services	29
Slika 16: Uporaba BI funkcij v Excelu	30
Slika 17: Primer uporabe SharePointServer	31
Slika 18: Grafični prikaz dveh pristopov za zajem podatkov	32
Slika 19: Priprava ETL procesa z MS SQL Server	32
Slika 20: Varnost Oracle	34
Slika 21: Varnost MS SQL Server	34
Slika 22: Gartner: Razporejenost ponudnikov v magičnem kvadrantu	38

Seznam preglednic

Preglednica 1: Formalna vprašanja in odgovori , ki si jih postavljamo pri odločanju	5
Preglednica 2: Definicija spremenljivk	6
Preglednica 3: Prihodki in tržni deleži BI proizvajalcev	8
Preglednica 4: Primerjava sprotnih transakcijskih obdelav in sprotnih analitičnih obdelav	11
Preglednica 5: Primerjava ponudnikov Oracle in Microsoft	35
Preglednica 6: Sistemi ocenjeni glede na hitrost izvajanja transakcij po kriteriju TPC-C	39
Preglednica 7: Sistemi ocenjeni glede na hitrost izvajanja transakcij in ceno po kriteriju TPC-C	40
Preglednica 8: Sistemi ocenjeni glede na hitrost izvajanja po kriteriju TPC-E	41
Preglednica 9: Sistemi ocenjeni glede na hitrost in ceno po kriteriju TPC-E	41
Preglednica 10: Sistemi ocenjeni glede na število sestavljenih poizvedb za 100 GB podatkov po kriteriju TPC-E	42
Preglednica 11: Sistemi ocenjeni glede na število sestavljenih poizvedb za 10.000 GB podatkov po kriteriju TPC-E	43

Povzetek

Namen pričujoče diplomske naloge je predstaviti sisteme za poslovno inteligenco, ki so izredno kompleksni, vendar pomembni za sodobno tržno konkurenčnost. Učinkovitost se seveda odraža tudi v ceni, zato je pomembna panoga, v katero se spleča finančno vlagati.

V zadnjih dvajsetih letih je bilo razvitih veliko tovrstnih orodij, vendar so se redka uspela obdržati.

Sistem poslovne inteligence je razdeljen na tri glavna področja: podatkovna skladišča, orodja za pripravo podatkov in orodja poslovne inteligence.

Podatkovno skladišče služi hranjenju podatkov v elektronski obliki. Pomembno je, da se ti podatki hranijo na način, ki omogoča hitre poizvedbe.

Glavni nameni orodja za pripravo podatkov so prepisovanje podatkov iz različnih virov, preureditev teh podatkov in njihov vpis v podatkovno skladišče.

Orodja poslovne inteligence lahko pomagajo uporabniku, da na čim lažji način dobiti in preurediti podatke v željeno obliko informacij. Orodja lahko uporabljajo tudi tehnično nepodkovani uporabniki, zato so ti sistemi ogledalo celotnega sistema. Med drugimi so to digitalne pregledne plošče, orodja za podatkovno rudarjenje, procesno rudarjenje, preglednice, OLAP in drugi.

Na tržišču orodij za poslovno inteligenco prevladuje nekaj velikih ponudnikov. V nalogi sta predstavljena dva ponudnika na tem področju in sicer Microsoft in Oracle. Podrobneje so opisana orodja in njihova razvrstitev v opisane tri skupine. Za vsako skupino orodij na kratko opišemo prednosti in slabosti.

V zadnjem delu naloge opisane ponudnike primerjamo po kriterijih varnosti, ceni, neodvisnosti od platforme in drugih. Pokazalo se je, da so Oracle orodja dražja, vendar lahko delujejo na različnih platformah.

Po oceni organizacije Gartner Group je Oracle boljši pri upoštevanju zastavljene vizije, Microsoft pa po izvedljivost zastavljenih ciljev. Po hitrosti sta orodji primerljivi.

Ključne besede: poslovna inteligenca, podatkovno skladišče, podatkovno rudarjenje, OLAP, procesno rudarjenje, Microsoft SQL Server, Oracle

Abstract

The purpose of this work is to present business intelligence systems. These systems can be extremely complex and important in modern market competition. Its effectiveness also reflects in price, so we have to explore their financial potential before investment. The systems have 20 years long history and during that time many of such tools have been developed, but they are rarely still in use.

Business intelligence system consists of three main areas: Data Warehouse, ETL tools and tools for Business intelligence.

Data Warehouse is system for storing data in electronic form. It is important that this information is stored in a manner which allows quick searches. Main purpose of an ETL tool is to extract, translate and load data into data warehouse. Business intelligence tools help end-user to transform data into information. They are usually used by non-professional users so they must be as simple as possible. Some of those tools are digital dashboard, tools for data mining, tools for process mining, spreadsheets, OLAP, etc.

Many companies produce tools in this area. In this thesis we present Oracle and Microsoft BI products. After detailed presentation we compare tools by different criteria: safety, price, platform independence and others.

We discovered Oracle has more expensive products, but they are not limited to Microsoft platform only. Gartner Group estimates that Oracle has better vision, Microsoft has better ability to execute, while they prove similar concerning speed.

Keywords: business intelligence, spreadsheet, data mining, OLAP, mining process, Microsoft SQL Server, Oracle

1. Uvod

1.1. Povod

V zadnjih nekaj desetletjih je informacijska tehnologija zelo napredovala. S tem tudi sama informatizacija procesov v podjetjih. Med izvajanjem procesov se zberejo ogromne količine podatkov. V najboljšem primeru se iz njih naredijo osnovna poročila. Možnost za izrabo teh podatkov pa so velike. Verjetno bo prišel čas, ko se bo interes za uporabo zaradi konkurenčnosti na trgu prisiljen povečati, do česar v tujini že prihaja. V določenih panogah prej kot v drugih. Trgovina je panoga, pri kateri je tekmovalnost na trgu zelo velika in vsaka prednost pred konkurenco je dobrodošla. Posledično je zbiranje in obdelava teh informacij izredno koristna pri pomembnejših odločitvah. V svetu ta način ni novost, saj ga veliko podjetij s pridom izrablja, saj je to edini način za ohranjanje konkurenčnosti. Menim, da se bodo slovenska podjetja morala z novostmi kmalu spopasti, če želijo še naprej konkurirati ostalim ponudnikom na globalnem trgu. [1]

1.2. Namen

Namen diplomskega dela je narediti splošen pregled znanj, tehnologij in orodij na področju poslovne inteligence (angl. Business Intelligence – BI). Poizkusili bomo sistematično opisati vsako od tehnologij, povezav med njimi v smislu celotnega poslovnega odločanja, kot to vidi uporabnik. Predstavil bom tudi dva ponudnika, ki sodita v vrh ponudbe na trgu: Microsoft in Oracle. Razlog za izbiro teh dveh ponudnikov temelji na mojem poznavanju njihovih produktov, še pomembnejša pa je prisotnost na slovenskem trgu programov za poslovno inteligenco.

1.3. Cilji

Cilj diplomskega dela je, da bralec po prebranem delu dojame zgradbo in delovanje sistemov za poslovno inteligenco. Zanima nas še zlasti:

- kateri glavni deli sistema morajo obstajati za delovanje,
- kakšni tipi uporabniških aplikacij obstajajo in
- kdo so ponudniki teh aplikacij.

Na koncu še ocena lastnosti ponudnikov teh storitev.

2. Poslovna inteligenca

2.1. Kaj je poslovna inteligenca?

Kadar govorimo o poslovni inteligenci mislimo na znanje, tehnike, programe in ustaljene postopke, ki pomagajo podjetjem pridobiti lažje razumevanje poslovanja podjetja ali organizacije. BI se lahko nanaša neposredno na zbrane podatke ali pa posredno iz teh črpa dodane informacije o poslovanju podjetja.

BI aplikacije poizkušajo zagotoviti pretekle, sedanje in bodoče vidike poslovanja. Med aplikacije poslovne inteligence tako sodijo poročanje, OLAP, analitika, podatkovno rudarjenje, poslovno upravljanje uspešnosti, merila uspešnosti in napovedna analitika.

Poslovna inteligenca si prizadeva podpirati boljše poslovno odločanje. BI sistem lahko imenujemo tudi sistem za podporo pri odločanju (angl. Decision Support System - DSS). [2]

2.2. Zgodovina poslovne inteligence

Pojem poslovne inteligence je prvič uporabil IBM-jev raziskovalec Hans Peter Luhn v članku leta 1958. Pojem BI je opredelil kot zmožnost doumeti razmerja iz dejstev in si ustvariti predstavo na način, ki vodi proces odločanja v smeri zelenega cilja.

Leta 1989 je Howard Dresner (pozneje je postal analitik v organizaciji Gartner Group) predlagal, da bi BI postal krovni izraz za opisovanje konceptov in metod za izboljšanje poslovnega odločanja z uporabo sistemov, katerih odločanje temelji na dejstvih (angl. Fact-Based Support Systems). Ta koncept je bil splošno razširjen v zadnjih letih 20. stoletja, predvsem po letu 1990.

BI pogosto uporablja zbrane podatke iz podatkovnih skladišč, vendar pa za delovanje BI sistemov to ni vedno potrebno. Vsa skladišča podatkov niso primerna za poslovno inteligenco, enako kot ne zahtevajo vse aplikacije poslovne inteligence podatkovnih skladišč.

Z razcvetom informacijske tehnologije (IT) v 20. in 21. stoletju, je vse odvisno od razpoložljivosti informacij in vlaganja v inovacije ter nove tehnologije, ki ne koristijo samo tehnološkemu, temveč tudi poslovnemu razvoju. Najznačilnejši primer tehnološke in poslovne inovacije je svetovni splet, ki zagotavlja odprto okolje (angl. Platform) za poslovanje po celem svetu. Za pripravo uspešne poslovne strategije je potrebno veliko število uporabnih informacij o stanju trga, inovacij in uporabnih konceptov BI, ki omogočajo enostaven dostop in upravljanje informacij vseh vrst.

Pred prihodom informacijske dobe je bil dostop do velikega števila informacij zelo težak, zato so poslovne odločitve večinoma temeljile na intuiciji. V današnjem svetu globalizacije si pomembna podjetja ne predstavljajo poslovanja brez uporabe BI sistemov. Včasih je prevladovala miselnost, da samo velike organizacije potrebujejo BI sisteme, vendar tudi manjše niso izključene. Informacije pomagajo napovedovati prihodnost in tako podjetjem omogočiti lažje razumevanje prodajnih trendov in njihov položaj na trgu v primerjavi s konkurenco. V ta namen je na voljo široka paleta različnih sodobnih tehnologij, ki se ukvarjajo s spremljanjem tržnih trendov, obnašanjem kupcev, povpraševanjem, analizo

dobavne verige in s številnimi drugimi vidiki poslovanja, ki izboljšujejo uspešnost podjetja. [3]

Tehnološki napredek nam omogoča enostavnejše in hitreje pridobivanje ter lažje analiziranje podatkov za dolgoročno strateško odločanje. Vendar je bil šele pred kratkim razvit sistem za poenostavljanje poslovnega odločanja. Leta 1989 je Howard Dresner, raziskovalec iz podjetja Garter Group, ki se ukvarja z raziskavami in svetovanjem, izraz BI definiriral kot niz metod in konceptov za izboljšanje poslovnega odločanja z uporabo obogatenih virov podatkov.

Ob uporabi konceptov BI s pomočjo podatkovnega skladišča in dobro razvitih orodji lahko sproti izvajamo analitične obdelave podatkov ter tako hitreje zberemo in analiziramo podatke ali raziskave. BI tehnologija pomaga pri pridobivanju informacij na podlagi razpoložljivih podatkov ter njihovo uporabo pri razvoju inovativnih poslovnih strategij. [12]

2.3. Delovanje sistemov za poslovno inteligenco

V večini definicij se BI nanaša na spretnosti, tehnologije, aplikacije in prakse, ki se uporabljajo kot pomoč pri razumevanju poslovnega konteksta.

Te definicije se osredotočajo na lastnosti in funkcionalnosti, ki jih je potrebno združiti v sistem. Za gradnjo BI rešitev je namreč potreben celovit sistem, ki je tesno povezan. Te definicije se osredotočajo na lastnosti in funkcionalnosti, vendar je pred uporabo le-teh potrebno zagotoviti sistem. Obstaja razlika med BI sistemom in BI dejavnostjo. BI sistem predstavlja pripravo kakovostnih informacij, BI dejavnost pa odločanje na podlagi le-teh podatkov. Za gradnjo BI rešitev je pomemben celovit sistem, ki temelji na tesni povezanosti.

BI omogoča osebi, ki v podjetju odloča, razumeti dosedanje poslovanje podjetja in poslovno okolje. Odločanje zahteva ocenjevanje dosedanje uspešnosti (kaj se je zgodilo), testiranje hipotez (zakaj in kako so se stvari zgodile) in napovedovanje prihodnjih dogodkov (kaj se lahko zgodi). Vodstvo mora vedeti, da so njihove strategije dobre in da se pravilno izvajajo. Sistem BI omogoča uporabniku, da pridobi odgovore na večino formalnih vprašanj, ki jih najdemo v preglednici 1 (opis spremenljivk se nahaja v preglednici 2):

Vprašanje	Formula	Odgovori
Zakaj se je to zgodilo?	$X \rightarrow Y$	Ali je X vzrok, da se je zgodil Y? Če naredimo X ali se bo zgodil Y?
Kako se je zgodilo?	$X_1 \rightarrow Y_1$	Kako je X povzročil Y? Ali smo lahko prepričani, da bo X povzročil Y ali da bo Z povzročil Y
Kaj se je zgodilo?	Y_1 proti Y_Y	Kaj se je zgodilo v primerjavi s tem, kar smo pričakovali? Če predpostavimo, da X povzroči Y, ali smo prepričani, da se X izvaja?
Kaj se lahko zgodi?	$X_f \rightarrow Y_f$	Če X_f zgodi v prihodnosti ali se bo zgodil tudi Y_f ? Če predpostavimo, da X povzroči Y ali lahko še v prihodnje pričakujemo enak rezultat?

Preglednica 1: Formalna vprašanja in odgovori, ki si jih postavljamo pri odločanju

Spremenljivka	Opis
X	Neodvisna spremenljivka
Y	Odvisna spremenljivka
X1	Posebni primer X-a
Y1	Posebni primer Y-a
Xf	X v prihodnosti
Yf	Y v prihodnosti

Preglednica 2: Definicija spremenljivk

Proces razvoja BI sistema naj bi omogočil, da analitik dobi odgovore na ta vprašanja. Za lažje nadaljevanje je potrebno opredeliti še nekatere pojme. Predvsem je pomembno razumeti, kaj zajema pojem »sistem«. Ena od teorij pravi, da je ta sestavljen iz:

- arhitekture (načrta) oziroma pravil in strukture, ki že obstajajo ter zagotavljajo okvir za celotno zasnovo sistema ali izdelkov,
- infrastrukture, ki je opredeljena kot tehnologija, okolje (angl. Platform), podatkovna baza, usmerjevalniki in druge komponente, ki so potrebne za delovanje arhitekture v organizaciji. Ostale komponente so ljudje in postopki, ki so potrebni za namestitev in vzdrževanje infrastrukture, procesov ter tehnologije za delovanje arhitekture BI sistema.

Metodologija izgradnje BI sistemov je lahko:

- od zgoraj navzdol (angl. top-down) in se nanaša na razumevanje poslovne strategije in izgradnjo sistema na podlagi le-tega. Odgovorna oseba preceni učinkovitost in uspešnost izvajanja strategije,
- okvirni pristop (angl. Framework) neposredno vključuje strategijo v arhitekturo sistema BI. Najprej formaliziramo strategijo s pomočjo diagramov, potem pa te uporabimo za oblikovanje sistema. Tako diagrami postanejo del arhitekture BI.

Bistvena je točna analiza strategije, saj osnova BI temelji tudi potrjevanju zastavljene strategije na podlagi predhodno analizirane. Na podlagi tega lahko pridemo do ugotovitev o izboljšanju učinkovitosti. S preverjanjem strategije se doseže zaupanje zaposlenih v pravilnost le-te. Potrjena strategija omogoča vodstvu spremljanje učinkovitosti poslovanja. Ob finančni neuspešnosti poslovanja ter hkratni pravilnosti izvajanja strategije pa mora vodstvo premisliti o preoblikovanju le-te.

Namen BI je zagotavljanje informacij nosilcem odločanja, saj so potrebne za sprejemanje preišljenih in kakovostnih odločitev. BI sistem običajno pripravi informacije v obliki poročil. Arhitektura sistema zagotavlja okvir za celotno zasnovo sistema ali izdelkov. Sistem je sestavljen iz:

- zainteresiranja strank in njihovih potreb po informacijah,
- vnosa podatkov,
- shranjevanja podatkov,
- analize podatkov in
- samodejnega in selektivnega razširjanja informacij.

Za izvajanje BI arhitekture mora arhitekt sistema najprej zagotoviti infrastrukturo, ki je temelji za nadaljevanje izgradnje arhitekture BI.

Nosilci odločitev potrebujejo poročila, ki jim bodo omogočila boljše razumevanje lastne organizacije in nanjo delujočih dejavnikov, kar posledično pomeni boljše odločanje.

2.4. Rešitve

Infrastruktura BI sistemov se zanaša na tehnologijo za pridobivanje informacij iz podatkov. Proizvajalci BI sistemov so jedro inovacij. Stopnja inovacij od leta 1990 ves čas narašča. Ne samo, da proizvajalci ponujajo nove produkte in izboljšujejo obstoječe, temveč se tudi razne inovacije hitro širijo med ponudniki.

Poleg tega proizvodi postajajo vse bolj napredni in enostavni za uporabo. Ti dve lastnosti pa sta si nasprotni, saj z večanjem funkcionalnosti raste kompleksnost rešitev, vendar se v tem primeru ni izkazalo tako. Sistemi so kljub vse večji kompleksnosti uporabniku vse bolj prijazni, kar je posledica avtomatizacije postopkov. Nekateri deli BI sistemov, kot so lahko podatkovna skladišča, se spreminjajo počasneje, saj so že dosegli zrelo fazo.

Enako velja za tudi systemske gradnike BI sistema. Potrebe uporabnikov in arhitekture se spreminjajo veliko počasneje kot sama tehnologija, vendar mora biti kljub temu tehnologija prirejena potrebam uporabnika.

3. Orodja

BI sisteme so začeli razvijati leta 1990, zato je to področje danes zrelo. To pomeni, da jasen obstaja osnovni nabor izdelkov in storitev, ki jih ponujajo prodajalci.

Sistem BI predstavlja skupek orodij, ki so namenjena za poročanje, analiziranje in predstavitev podatkov. Orodja v splošnem berejo podatke, ki so shranjeni v nekem mediju, ta je pogosto podatkovno skladišče.

Najpomembnejši način razvoja med velikimi proizvajalci je pogosto kar prevzem manjših podjetij z izdelanimi rešitvami. Tako večja podjetja povečujejo svoj tržni delež. Nekaj večjih poslovnih transakcij na tem področju v zadnjih letih:

- 2008 - IBM kupi Cognos za 5 milijard dolarjev
- 2008 - SAP kupi Business Objects za 4,8 milijarde evrov
- 2007 - Oracle kupi Hyperion za 3.3 milijard dolarjev

Podjetje \ leto	Dobiček (M\$)			Tržni delež (%)		
	2006	2007	2008	2006	2007	2008
SAP	1,214	1,350	1,589	24.1	23.8	25.4
IBM	685	755	794	13.6	13.3	12.7
Oracle	502	588	692	9.9	10.3	11.0
Microsoft	462	533	623	9.2	9.4	10.0
SAS	297	352	373	5.9	6.2	6.0
MicroStrategy	249	266	282	4.9	4.7	4.5
InformationBuildersInc.	182	180	178	3.6	3.2	2.8
ActuateCorp.	99	111	106	2.0	2.0	1.7
QlikTech	39	70	104	0.8	1.2	1.7
PanoramaSoftware	38	57	62	0.7	1.0	1.0
TIBCO	35	38	51	0.7	0.7	0.8
Fujitsu	27	28	33	0.5	0.5	0.5
Arcplan	26	29	31	0.5	0.5	0.5
OpenText	28	26	25	0.5	0.5	0.4
LawsonSoftware	28	19	23	0.6	0.3	0.4
CA	29	34	20	0.6	0.6	0.3
SPSS	24	21	19	0.5	0.4	0.3
Kalido	15	16	17	0.3	0.3	0.3
Targit	9	11	14	0.2	0.2	0.2
Infor	12	13	14	0.2	0.2	0.2
DatawatchCorp.	13	14	13	0.3	0.3	0.2
AdvizorSolutions	9	10	11	0.2	0.2	0.2
Ostali	1,025	1,155	1,186	20.3	20.3	18.9

Preglednica 3: Prihodki in tržni deleži BI proizvajalcev

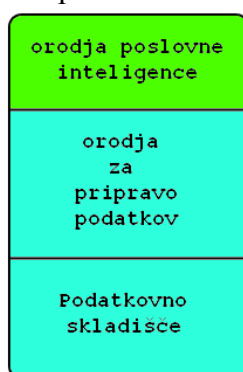
Treba je upoštevati, da imajo poleg določenih rešitev podjetja tudi velik tržni delež. Tako je imeli recimo podjetje Cognos skoraj 14% tržni delež, podjetje Business Objects pa 19% delež.

Preglednica 3 Preglednica 3 prikazuje deleže in dobičke največjih podjetij na trgu od leta 2006 do 2008, pa čeprav še ne vključuje treh zgoraj naštetih poslovnih transakcij [5].

3.1. Ključne kategorije orodij poslovne inteligence

BI orodja predstavljajo vrhni del sistema, ki ga prikazuje slika 1. Delimo jih lahko v naslednje skupine:

- preglednice,
- orodja za poročanje in poizvedbe,
- sprotne transakcijske in analitične obdelave,
- digitalne pregledne plošče,
- sistemi za podatkovno rudarjenje,
- sistemi za procesno rudarjenje in
- sistemi za upravljanje poslovne uspešnosti.



Slika 1: Shema sistema poslovne inteligence

Običajno orodja na tržišču ne nastopajo samostojno, temveč v zbirkah, s čimer se bolj približajo željam uporabnikov. Vsako od orodij se lahko prodaja na različne načine in sicer kot:

- samostojno orodje,
- zbirka orodij,
- sestavni del programske opreme, ki cilja na specifične industrije ali
- drugi.

3.1.1. Preglednice

Preglednica (angl. Spreadsheet) je računalniška aplikacija, ki simulira delovni list papirja. Prikaže več celic, ki skupaj tvorijo mrežo, sestavljeno iz vrstic in stolpcev. Vsaka celica pa vsebuje bodisi besedilo ali številčne vrednosti. Celica preglednice lahko vsebuje tudi formulo, torej se vsebina celice izračuna iz vsebine katere koli druge celice (ali kombinacije celic) in se posodobi vsakič, ko se posodobi katerakoli izmed celic. Vsako spremembo podatka zazna in posledično izračuna novo stanje celotne preglednice. Zaradi te zmogljivosti aplikacije se preglednice pogosto uporabljajo za finančne informacije. Najbolj poznana tovrstna programa sta Microsoft Excel in OpenOffice Calc.

3.1.2. Programi za poročanje in poizvedbe

To so orodja, ki izvlečejo, uredijo in povzamejo izbrane podatke. Kot take omogočajo 4 glavne faze priprave poročil:

- branje podatkov iz raznih virov,
- urejanje podatkov,
- povzemanje bistva in
- predstavitev podatkov.

Za to obstajajo več plačljivih in odprto-kodnih orodj:

- Plačljivi so Actuate, Alchemex, Crystal Reports, DevExpress Reporting, I-net Crystal-Clear, Information Builders' FOCUS and WebFOCUS, Izenda Self-Service Reports, Jreport by Jinfonet Software, LogiXML, MicroStrategy, MicroStrategy Reporting Suite, Oracle XML Publisher, Proclarity, SQL Server Reporting Services, Style Report, Telerik Reporting in drugi.
- Odprto-kodni pa so Agata Report, Eclipse BIRT Project, JasperReports, Pentaho in drugi.

3.1.3. Sprotne transakcijske in analitične obdelave

Cilj informacijskega procesiranja je pretvoriti podatke v informacije. Pri upravljanju z informacijami se v podjetju srečujemo z dvema vrstama podatkovnih tehnologij:

- sprotne transakcijske obdelave (angl. Online Transaction Processing - OLTP) in
- sprotne analitične obdelave (angl. Online Analytical Processing - OLAP).

Operativne podatkovne zbirke vsebujejo informacije o dogodkih, ki se dogajajo dnevno in so pogojene z odvijanjem poslovnih procesov v podjetju. Podjetje ima posledično lahko več OLTP informacijskih sistemov, ki se bodo odvijali na različnih operacijskih sistemih ter hranili podatke v različnih podatkovnih zbirkah. Kljub temu je njihov cilj podoben, to je zadovoljiti uporabnikovo potrebo po zbiranju in obdelovanju tekočih podatkov. Operativne podatkovne zbirke OLTP so običajno visoko normalizirane in temeljijo na relacijskih podatkovnih zbirkah, s čimer se izboljšajo hitrost vnosa, urejanje in brisanje podatkov in drugo. Razvijalcem in ponudnikom relacijskih podatkovnih zbirk je sčasoma uspelo močno nadgraditi komponente podatkovnih zbirk, zato so te sposobne učinkovito izvajati marsikatero zahtevo OLTP sistemov.

Po drugi strani pa se OLAP tehnologija uporablja za analizo poslovanja ogromnih količin podatkov. Predstavitve podatkov v OLAP tehnologiji deluje podobno kot proces pomnjenja pri človeških možganih. Podatke si kar naprej urejamo in združujemo v skupine, kar nam pomaga pri preglednejši predstavitvi določenega pojava. Za lažje razumevanje je primerjava prikazana v preglednici 4.

Značilnost	OLAP	OLTP
Namen	Podpora poslovnim procesom.	Podpora odločitvenim procesom.
Vir podatkov	Tekoči vnosi podatkov, POS terminali, različni uporabniki.	OLTP baze podatkov, podatkovna skladišča in svetovni splet.
Kritičnost	Življenjsko pomembni podatki. Potrebno dnevno arhiviranje.	Pomembni za podporo odločanju. Podjetje lahko deluje tudi brez njih. Manj pogosto arhiviranje.
Količina podatkov	Relativno malo podatkov.	Ogromna količina podatkov.
Dostop	Veliko število uporabnikov, zelo pogost dostop.	Manjše število uporabnikov.
Odzivni čas	Zelo hiter odzivni čas, navadno le delček sekunde.	Daljši odzivni čas. Včasih več sekund, minut ali celo dlje.
Spremembe	Ogromno vnosov, popravkov in brisanja s strani uporabnikov.	Ni neposrednega spreminjanja. Za to funkcijo skrbi sveženj transformacijskih programov.
Namensko poizvedovanje (lati. AD-hoc)	Zelo slabe zmogljivosti.	Dobre zmogljivosti.
Zahtevnost poizvedb	Zahtevne poizvedbe. Poizvedba zahteva veliko povezav med tabelami.	Enostavne za poizvedovanje. Zaradi denormalizacije le nekaj povezav.

Preglednica 4: Primerjava sprotnih transakcijskih obdelav in sprotnih analitičnih obdelav

- **OLAP arhitektura**

Vhodni podatki v OLAP okolje se navadno hranijo v podatkovnem skladišču, ki je vzpostavljen kot relacijska baza podatkov. [4] Vsebuje podatke o vseh merah in dimenzijah, ki jih potrebuje OLAP sistem za obdelavo podatkov. OLAP sistem zajema podatke iz podatkovnegaskladišča, jih obdela in shrani v OLAP podatkovno zbirko, ki jo imenujemo tudi OLAP kocka. Računanje seštevkomov poteka s seštevanjem podatkov iz vhodnih vrstic, ki se shranjujejo v OLAP kocko. Ob tem imamo opravka z dvema tipoma podatkov: z vhodnim podatkom in seštevkom. Glede na to, kako se ta dva tipa podatkov shranita, ločimo tri načine shranjevanja v OLAP podatkovno bazo, in sicer MOLAP, ROLAP in HOLAP.

MOLAP je večdimenzijski OLAP (angl. Multidimensional OLAP - MOLAP), ki predstavlja arhitekturo, kjer so podatki shranjeni v večdimenzionalni kocki. Vsi izračuni so izvedeni pred tvorbo kocke, s čemer nudi OLAP večje hitrosti poizvedovanja. Ker se pri MOLAP arhitekturi vsi podatki prestavijo in pretvorijo v primeren format z namenom zadostovati večdimenzionalni podatkovni zbirki, je MOLAP primeren za majhne ali srednje količine podatkov.

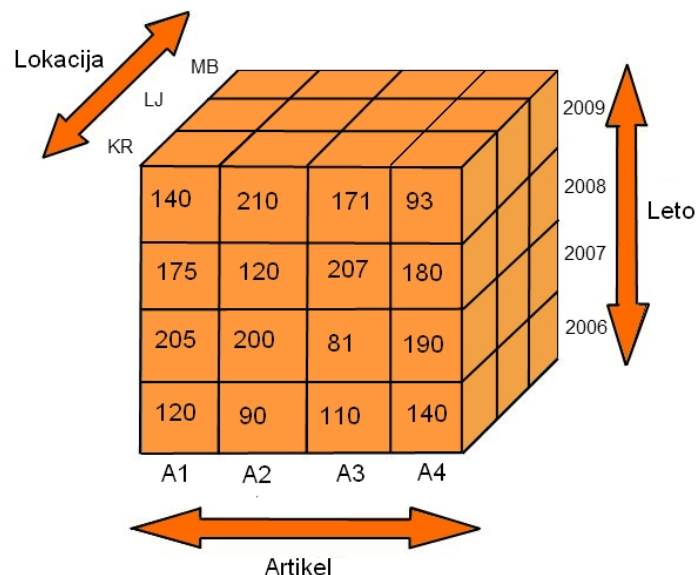
ROLAP je relacijski OLAP (angl. Relational OLAP - ROLAP). Podatke pusti v izvornih relacijskih tabelah. Izračunani podatki in povezave so shranjene v OLTP sistemu v posebnih tabelah. Tabele, ki vsebujejo seštete podatke imenujemo fizični pregledi (angl. Materialized View - MV). Izračunani podatki so shranjeni v skladu z definicijo dimenzij v kocki. Tabele vsebujejo polja za vsako od dimenzij in mero, vsaka dimenzijska kolona pa je označena (angl. Indexed). Prav tako je izdelana dimenzijska oznaka (angl. Index) za vsa dimenzijska polja, ki je primerna za iskanje po ogromnih količinah podatkov. Hitrost poizvedb je odvisna od postavljenih označb (angl. Index) nivoja predhodnega združevanja v povprečju je slabša kot

pri orodjih MOLAP, saj poizvedeni jezik SQL (angl. Structured Query Language - SQL) ni razširjen za izvajanje tipičnih OLAP operacij (kot sta zvijanje in vrtanje v globino).

HOLAP je hibridni OLAP (angl. hibride – OLAP - HOLAP) in shranjuje izvirne podatke v relacijski bazi, izračunane podatke pa v več dimenzijskem modelu oziroma podatkovni kocki. Prednost se kaže v tem, da HOLAP po eni strani omogoča hiter dostop do velikih količin podatkov v relacijskih tabelah, po drugi strani pa gre za kombinacijo med MOLAP in ROLAP arhitekturo. Slabost te izbire je v tem, da lahko število operacij med ROLAP in OLAP resno ogrozi samo delovanje sistema.

- **OLAP kocka**

OLAP kocka oziroma podatkovna kocka (angl. Data Cube) je naziv za dimenzijsko strukturo v večdimenzionalni ali OLAP podatkovni platformi. Prvotno se nanaša na tri preproste dimenzije in sicer *izdelek*, *tržišče* in *čas*. Podatkovna kocka dopolnjuje relacijsko tabelo z izračuni glede na kombinacije atributov v relacijskih tabelah. Podatki so v OLAP kocki grupirani v posamezne skupine, ki so razdeljene v hierarhične nivoje. Skupine in nivoji se tvorijo na podlagi dimenzijskih funkcij ter algebre za združevanje podatkov, kar prikazuje slika 2.



Slika 2: Primer OLAP kocke

- **Shema kocke**

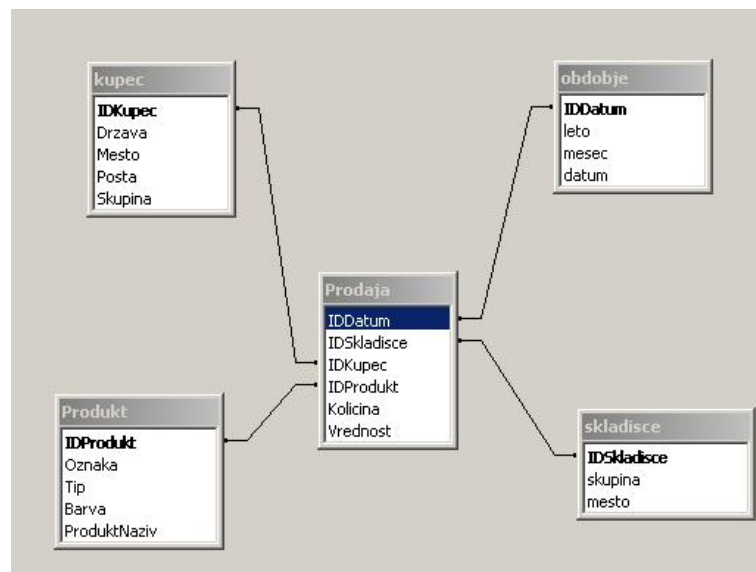
Shema kocke je množica povezanih tabel v podatkovnem skladišču, od koder kocka črpa podatke. Tabela, ki se nahaja v središču sheme, imenujemo tabela dejstev (angl. Fact Table), ostale tabele pa so tabele dimenzij (angl. Dimension Table). Strukturo kocke definirajo njene meritve in tabele dimenzij. Množica tabel, iz katere izhajajo meritve in dimenzije se imenuje shema kocke. Vsaka shema kocke je sestavljena iz tabel dejstev in ene ali več tabel dimenzij. Obstajata dve tipični shemi:

- a. **Zvezdna shema**

Ime izhaja iz oblike podatkovnega modela, oziroma razporeditve tabel, ki je prikazan na sliki 3. Model vsebuje osrednjo tabelo, ki ji pravimo tabela dejstev, ta pa je obkrožena z večjim številom dimenzijskih tabel.

Tabela dejstev zajema podatke o prodaji in o ključih vseh dimenzij, preko katerih je tabela dejstev povezana z ostalimi dimenzijskimi tabelami. V našem primeru zvezdne sheme so to obdobje, kupec, produkt in skladišče.

Tabela dejstev vsebuje poleg ključev za povezavo z ostalimi dimenzijskimi tabelami tudi dva atributa, ki vsebujeta vrednost. V našem primeru sta to količina prodanih izdelkov in vrednost v denarju.



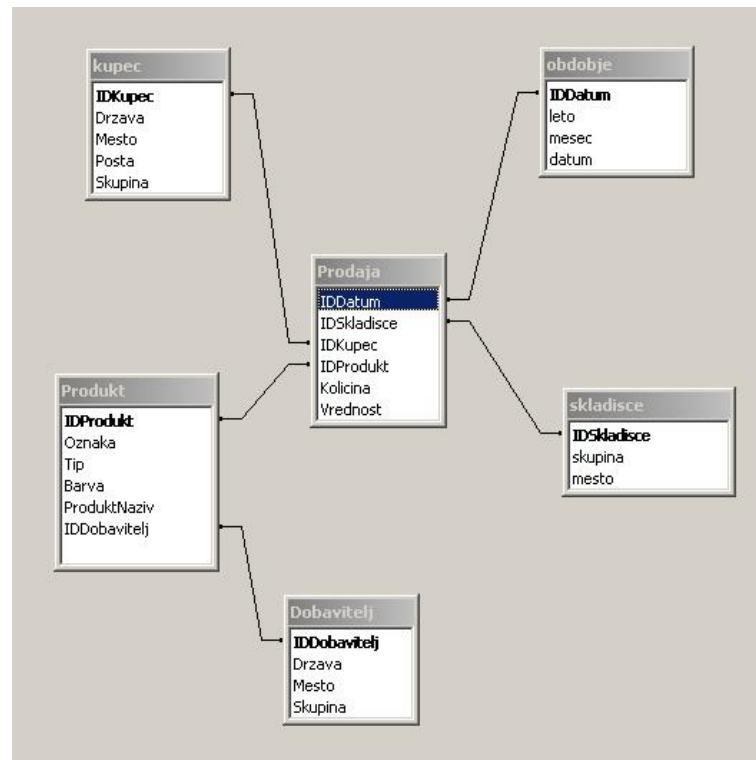
Slika 3: Primer zvezdne sheme

- b. **Snežinkasta shema**

Model snežinkaste sheme na sliki 4 je izpeljan iz zvezdne sheme. Dimenzijske tabele v snežinkasti shemi vsebujejo normalizirane podatke, kar dosežemo z dodatno tabelo, ki preprečuje odvečne podatke (podvajanje atributov). Tako prihranimo prostor za shranjevanje podatkov, vendar izgubimo učinkovitost zaradi večje zahtevnosti poizvedb. Zato snežinkasta shema ni tako priljubljena kakor zvezdna.

Glavna razlika med zvezdno in snežinkasto shemo je vidna pri določanju dimenzijskih tabel. Iz dimenzijske tabele, v kateri se vrednosti posameznih atributov ponavljajo, se te attribute

prenese v pomožno tabelo, ki je z dimenzijsko tabelo povezana s ključem. Nadaljnja normalizacija lahko sledi tudi v pomožni tabeli, na katero je v tem primeru povezana še ena pomožna tabela.



Slika 4: Primer snežinkaste sheme

• Dimenzije

Dimenzije so sestavni del lastnosti (angl. Attribute) kocke. Vsaka dimenzija je v bazi podatkov predstavljena kot dimenzijska tabela, ki vsebuje lastnosti, ki opisujejo to dimenzijo. Dimenzija lahko vsebuje enega ali več nivojev, kar nam omogoča, da se lahko spuščamo in dvigamo znotraj ene dimenzije. Tako si jih lahko ogledujemo na različnih hierarhičnih ravneh. Dimenzije opredelijo organizacijo podatkov kock na hierarhični način. Nivo pa združuje vse elemente, ki so enako oddaljeni od korena, v primeru v odvisnosti od časovne dimenzije bi lahko vsebovala tri nivoje (leto, mesec dan), lahko pa jih tudi več. To je odvisno predvsem od naše želje po natančnosti in predmeta obdelave. Včasih nam tudi ure, minute in sekunde veliko pomenijo.

V našem primeru bodo elementi na posameznih nivojih:

- nivo Leto lahko vsebuje vrednosti: 2003, 2004, 2005, 2006, 2007,
- nivo Mesec lahko vsebuje vrednosti: januar 2003, februar 2003, marec 2003, ..., december 2003,
- nivo Dan vsebuje datum od prvega dne meritve (recimo 1.1.2004) do zadnjega dne meritve.

Elementi so znotraj dimenzije urejeni stopensko, kar pomeni, da je nadrejeni član vedno seštevek vseh njemu podrejenih članov. Iz stolpcev tabele dimenzije so izpeljani nivoji, ki razvrščajo člane po stopnjah. Glede na dimenzijo so podatki predstavljeni od najvišjega (najbolj splošnega) do najnižjega (najbolj podrobnega). Dimenzije so vidiki, ki jih podjetje

želi obravnavati (npr. obdobje, kupec, produkt,...), zato lahko na podatke gledamo iz več vidikov. Lahko recimo gledamo prodajo izdelka glede na kupca, prodajo izdelka glede na lokacijo, prodajo izdelka glede na čas in kupca, prodajo izdelka glede na kupca, skladišče in čas in tako dalje.

a. Privatna dimenzija

Taka dimenzija se uporablja samo v eni kocki, zato si kocka take dimenzije ne more deliti z drugo kocko. Če izpostavimo primer kocke *prodaja* in *finance*, ki obe analizirata isto dimenzijo, zato ne bi imelo nobenega pomena, da si kocki delita dimenzijo *produkt*.

b. Deljene dimenzije

Kocke si lahko deljene dimenzije delijo med sabo. Pogosto si kocke med seboj delijo časovno dimenzijo, recimo kot kocki *prodaja* in *kupec*, ki si lahko delita dimenzijo *izdelek*.

• Meritve

S pojmom meritev v kocki zajamemo množico vrednosti, ki temelji na stolpcu v tabeli dejstev kocke in je zato pogosto numerična. Vrednosti za posamezno meritev vsebuje celica, ki je najosnovnejši element kocke. Meritev ima podano funkcijo izračuna, ki določa, na kakšen način so vrednosti iz nižjih nivojev povzete in združene na višjem nivoju. Izračuni so vnaprej določeni in shranjeni povzetki podatkov, ki izboljšajo odzivne čase poizvedb. V primeru kocke *prodaj* imamo meri za količino in vrednost prodanih izdelkov.

• Najpogostejše operacije

Operacije, ki se najpogosteje uporabljajo nad podatki so:

- a. **Zvijanje** (angl. Roll-Up): podatke prikažemo manj podrobno. Torej gre za združevanje podatkov kocke, ki se lahko izvrši bodisi z dvigovanjem po hierarhiji bodisi s krčenjem dimenzije na kocki.
- b. **Vrtanje v globino** (angl. Drill-Down): podatke prikažemo podrobneje. Ta operacija je nasprotna operaciji zvijanja, kar pomeni, da s pomočjo vrtanja v globino dobimo podrobnejši vpogled v podatke, ki se skrivajo za določeno vrednostjo. Letno vrednost prodaje lahko tudi s pomočjo vrtanja v globino razčlenimo na vrednost po posameznih mesecih, tednih oziroma dnevih.
- c. **Rezanje** (angl. Slice And Dice): naredimo prilagojen izbor podatkov kocke na podlagi enega ali več kriterijev.
- d. **Vrtenje** (angl. Pivot): obračamo pogled na podatke. Ta operacija vizualno rotira osi pri pregledu, ter s tem omogoči drugačen pogled na predstavljene podatke.
- e. **Prikaz izjem** (angl. Exception Report): prikazujemo podatke, ki odstopajo od ostalih vrednosti glede na določen kriterij.

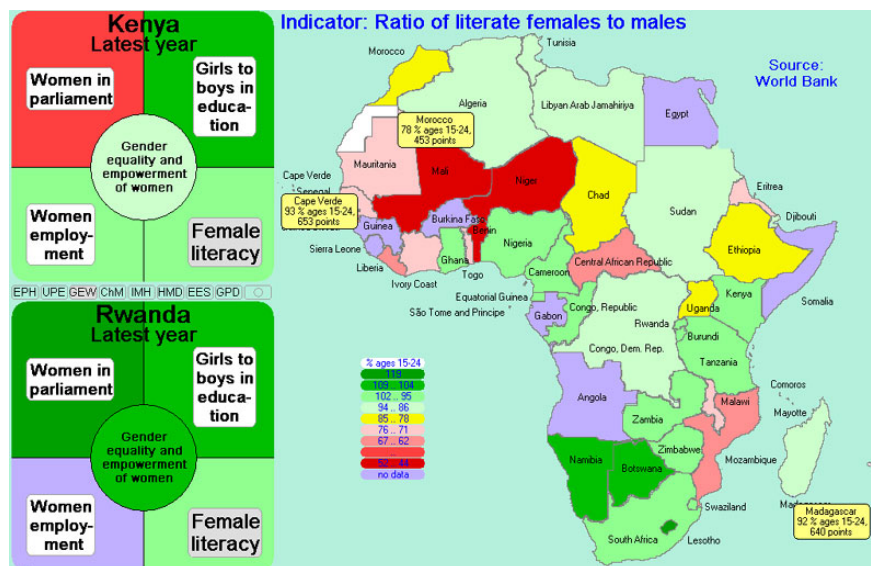
3.1.4. Digitalna pregledna plošča

Digitalna pregledna plošča (angl. Digital Dashboard) predstavlja uporabniški vmesnik izvršilnega informacijskega sistema. Podobna je armaturni plošči avtomobila. Namenjena je enostavnemu branju in tolmačenju podatkov na njej. Sistem lahko črpa podatke iz več različnih virov in jih prikazuje na enem prilagojenem pregledu. Primer pregledne plošče prikazuje slika 5.

Namen digitalnih preglednih plošč je sledenje toku poslovnih procesov, ki jih sistem nadzoruje. S pomočjo grafičnega prikaza lahko uporabnik opazuje delovanje na zelo visoki ravni procesov, po želji pa lahko vrta navzdol do nivoja podatkov. Najnižja stopnja podatkov je ponavadi nedostopna vodstvu, ker je zakopana v različnih transakcijskih sistemih [8].

Digitalne pregledne plošče delimo v glavne tri skupine [18]:

- samostojni programi,
- programi na osnovi spletnih brskalnikov in
- namizni programi ali namizni gradniki (angl. Widget).



Slika 5: Primer digitalne pregledne plošče

3.1.5. Podatkovno rudarjenje

Podatkovno rudarjenje (angl. Data Mining) je proces pridobivanja vzorcev iz podatkov. Količina zbranih podatkov se podvoji vsake tri leta, zato postaja rudarjenje podatkov vse bolj pomembno orodje za preoblikovanje teh podatkov v informacije. To se običajno uporablja na področjih kot so trženje, nadzor, odkrivanje goljufij in znanstvenih odkritij.

Podatke rudarjenj lahko uporabijo pri odkrivanju vzorcev v podatkovnih zbirkah. Pri tem je pomembno, da lahko uporaba neobičajnih vzorcev pripelje do nepravilnih rezultatov. Drugi problem je odsotnost posameznih podtipov v izbranem vzorcu podatkov, ki v celotni populaciji podatkov še obstajajo.

Tako kot katerokoli drugo orodje tudi podatkovno rudarjenje za svoje delovanje potrebuje podatke, ki jih mora predčasno zagotoviti uporabnik. Izredno pomemben del procesa je zato preverjanje in potrjevanje vzorcev. Ko vzorce preverimo, ugotovitve o njih lahko uporabimo pri potrjevanju novih vzorcev podatkov. Termin podatkovno rudarjenje je lahko uporabljen tudi v negativnem smislu, kadar zahtevamo iskanje očitnih informacij v slabem vzorcu.

Podatkovno rudarjenje je sestavljeno iz 4 glavnih opravil:

- a. razvrščanje (angl. Classification) je urejanje podatkov po vnaprej določenih skupinah (kot pri razvrščanju elektronske pošte v skupini zelena in neželena elektronska pošta). Pogosto se za to uporabljajo naslednji algoritmi:
 - Najbližji sosed (angl. Nearest Neighbor) [33],
 - Naivni Bayesov klasifikator (angl. Naive Bayes Classifier) [34],
 - Nevronske mreže (angl. Neural Network) [35] in
 - Drugi,
- b. združevanje (angl. Clustering) je podobno kot klasifikacija, le da skupine pred postopkom niso definirane in jih algoritem skuša optimalno določiti,
- c. zmanševanje (angl. Regression) poizkuša najti način, ki naredi optimalno podobo podatkov (z najmanj napakami). Najpogosteje uporabljena metoda je genetsko programiranje (angl. Genetic Programming) in
- d. učenje povezovalnih pravil je iskanje povezav med spremenljivkami. Če imamo v trgovini zbrane podatke o kupljenih artiklih, lahko ugotavljamo, kateri artikli se najpogosteje kupujejo skupaj.

Zadnji korak za odkrivanje znanja iz podatkov je preverjanje vzorcev. Pri tem skušamo ugotoviti, če porazdelitev podatkov v izbranem vzorcu ustreza tudi porazdelitvi podatkov na vzorcu, ki temelji na večji količini podatkov. V nasprotnem primeru porazdelitev podatkov v našem vzorcu ne predstavlja celotne populacije podatkov.

Kadar naučeni vzorci ne izpolnjujejo želenih meril, jih moramo spremeniti ali preurediti pred obdelavo podatkov. Ob pogoju, da naučeni vzorci sovpadajo z merili, lahko te vzorce pretvorimo v znanje.

3.1.6. Procesno rudarjenje

Procesno rudarjenje (angl. Process Mining) je tehnika upravljanja procesov. Omogoča nam preučevanje poslovnih procesov na podlagi dnevnikov dogodkov. Glavna ideja je potegniti informacije iz dnevniških zapisov, ki jih je zabeležil informacijski sistem.

Obstajajo tri vrste procesnega rudarjenja. Osnovni pogoj za izvajanje procesnega rudarjenja so modeli procesov. Ob obstoju teh lahko delimo vrste procesnega rudarjenja na:

- odkritje – ni nujno, da obstaja model, lahko pa se te zgradi iz dogodkov.
- skladnosti – model mora obstajati. Primerja se z dnevnikom dogodkov, neskladja med dnevnikom dogodkov in modelom se preučijo.
- razširitev – model obstaja in služi kot osnova za razširitev ali obogatitev pogleda.

Primer programskega okvirja (angl. Framework) za ocenjevanje algoritmov postopka izkopavanja je bil razvit na eindhovenški tehniški univerzi pod okriljem inženirja Wila Van der Aalsta. Zastavljen je bil kot odprtokodni projekt.

Med komercialnimi ponudniki najdemo naslednje: Futura Reflect (Futura Technology), Interstage Automated Process Discovery (Fujitsu, Ltd) in druge.

3.1.7. Upravljanje poslovne uspešnosti

Upravljanje poslovne uspešnosti (angl. Business Performance Management - BPM) se sestoji iz procesov, ki pomagajo organizacijam izboljšati njihovo poslovanje. Zagotavlja okvir za organiziranje, avtomatizacijo in preučevanje poslovne metodologije, matrik, procesov ter sistemov.

Colin White meni, da obstaja povezava med BI in BPM. Največja rast BI je namreč ravno zaradi preučevanja sistemov BPM. Vloga operativnega BPM-ja ni samo v preučevanju uspešnosti, temveč tudi v primerjavi izmerjene uspešnosti glede na poslovne cilje in v opozarjanju uporabnikov, ko uspešnost ni dovolj visoka.

BPM omogoča podjetjem učinkovitejše izrabljanje finančnih, človeških, materialnih in drugih virov. Ključna razlika med BI in BPM je, da drugi vsebuje koncept nadzora ali povratne zanke, ki pomaga vodstvu podjetij doseči svoje cilje. BI sistem lahko s pomočjo analitike pridobi nabor poslovnih ciljev in spremlja napredek pri njihovem uresničevanju.

Programska oprema in metode za upravljanje poslovne uspešnosti (angl. Corporate Performance Management - CPM) omogočajo sistematičen in celosten pristop, ki povezuje strategije podjetja v osnovne procese in dejavnosti podjetja.

- **Pristopi**

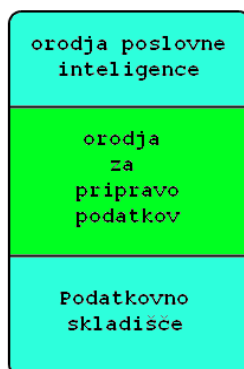
Obstajajo različni pristopi za izvajanje BPM. Disciplina omogoča podjetjem izvajanje od zgoraj navzdol, s katerim se uskladi načrtovanje in izvedbo, strategijo in taktiko ter poslovanje enote in cilje podjetja.

Obstaja več znanih pristopov:

- šest sigma (angl. Six Sigma) [27],
- urejen seznam točk (angl. Balanced Scorecard) [28],
- preučevanje stroškov po aktivnostih (angl. Activity-Based - ABC) [29],
- celovito upravljanje kvalitete (angl. Total Quality Management -TQM) [30],
- dodane ekonomske vrednosti (angl. Economic Value-Add) [31] in
- celovito stroškovno merjenje (angl. Integrated Strategic Measurement) [32].

Najbolj uporabljen je urejen seznam točk (angl. Balanced Scorecard). Metodologije same ne morejo zagotoviti popolnega CPM sistema, večina metodologij zaradi celo preslabe povezave s osnovnimi CPM procesi ne prinaša želenega učinka. Glavni razlog naj bi tičal v tem, da so ti pristopi narejeni neodvisno in jih da je potrebno za vgradnjo v CPM prirediti ali dopolniti.[7]

3.2. Orodje za pripravo podatkov



Slika 6: Shema sistema poslovne inteligence

Orodje za pripravo podatkov [6] (angl. Extract, transform, load - ETL) je osrednji del BI sistema, kar je razvidno iz slike 6. Uporablja se predvsem pri polnjenju podatkovnih skladišč. Ta postopek zajema tri faze:

- črpanje podatkov iz zunanjih virov,
- preverjanje pravilnosti in kakovosti podatkov in
- nalaganje v podatkovno skladišče.

3.2.1. Črpanje podatkov iz zunanjih virov

Prvi del procesa ETL vključuje pridobivanje podatkov iz izvornih sistemov. Večina projektov skladiščenja pridobiva podatke iz različnih sistemov/virov. To pomeni prepisovanje iz različnih podatkovnih baz (različnih ali enakih ponudnikov), prepisovanje datotek ali celo nerelacijskih podatkov (strukture IMS, VSAM, ISAM,..). Pri zajemu se preveri ali so podatki prišli v zelenem formatu, nato se po uspešni pretvorbi predajo naprej procesu za preoblikovanje. Kadar podatki niso pričakovane oblike, se lahko v tej fazi delno ali v celoti zavrnejo, kar je odvisno od izvedbe.

3.1.1. Preverjanje pravilnosti in kakovosti podatkov

Vsak sistem lahko uporablja drugačen format ali organizacijo podatkov, zato je potrebno pred prepisovanjem v enotno strukturo preoblikovati podatke. V tem procesu se tudi preverja semantična pravilnost podatkov. Od same izvedbe je odvisno, kaj se izvaja v ETL procesu. Pred zapisom v podatkovno skladišče je potrebno odpraviti vse napake. Kadar napak ni mogoče odpraviti, se podatki zavrnejo.

Preoblikovanje je stopnja, ki se sestoji iz vrste pravil, funkcij in posredovanih podatkov. Pri nekaterih virih je potrebno zelo malo ali skoraj nič preurejanja, v drugih primerih pa je potrebno izvesti eno ali več izmed naslednjih vrst preoblikovanja za doseganje poslovnih in tehničnih potreb ciljne zbirke podatkov:

- izbiro samo določenih stolpcov (ne uvažamo praznih vrednosti NULL),

- Z pretvarjanjem kodiranih vrednosti (shranjeno v nekem sistemu za spol 1 je moški in 2 je ženska, v skladišču pa to poenotimo kot M in F) je potrebno avtomatsko čiščenje podatkov. V samem ETL postopku ni ročnega čiščenja,
- dešifriranje vrednosti iz obrazcev (npr., Moški je lahko g . ne more pa biti ga .),
- izračunavanje novih vrednosti (npr., prodaja=količina * cena enote),
- filtriranje,
- razvrščanje / urejanje,
- zlivanje podatkov iz različnih virov (npr., iskanje, zlivanje),
- združevanje (npr., izračuni prodaje za posamezne regije),
- ustvarjanje / nadomeščanje ključev (v vsakem viru so ključi drugačni kot v podatkovnem skladišču),
- premeščanje kolon (angl. Transposing) in vrtilne tabele (angl. Pivot) ,
- delitev stolpca v več samostojnih stolpcev (npr., vrednosti v enem stolpcu, z vejico razdelimo v več stolpcev),
- razčlenitev ponavljajočih se podatkov in njihov zapis v tabelo podrobnosti (npr., podatki o naslovih zapisani v več zapisih, združitve v svoj zapis v tabeli naslovov in priprava povezav na to tabelo),
- uporabo kakršnikoli oblik (enostavnih ali kompleksnih) za preverjanje veljavnosti podatkov. Če preverjanje uspe, se podatki posredujejo naprej. Če pa katera od zgornjih transformacij povzroči izjemo, se podatke zavrne.

3.2.2. Nalaganje v podatkovno skladišče

Nalaganje v podatkovno skladišče je faza prepisa podatkov na končno lokacijo, večinoma v podatkovno skladišče. Ta proces je odvisen od zahtev organizacije. Nekateri podatki iz podatkovnega skladišča lahko povezujejo obstoječe podatke s skupnimi. Gledano s časovnega vidika je način dodajanja podatkov lahko intervalni (enkrat tedensko) ali pa obstaja kot beleženje sledi izvajanja (na eno uro). Čas in obseg prepisa je odvisen predvsem od potreb posamezne organizacije. Najzahtevnejši sistemi lahko vzdržujejo celotno zgodovino in pregled sledi izvajanja za vse spremembe v samem podatkovnem skladišču.

Med nalaganjem podatkov v podatkovno skladišče pride do še ene vrste preverjanja pravilnosti uvoženih podatkov. To so omejitve sheme, kot je recimo omejitev po ključu, obvezna polja in celostna omejitev, ki vse tudi prispevajo k celotni kvaliteti prepisa podatkov v ETL procesu.

3.2.3. Življenjski cikel procesa za pripravo podatkov

V realnem življenju ETL cikel obsega naslednje korake izvedbe:

- inicializacijo cikla,
- pripravo referenčnih podatkov,
- črpanje podatkov iz virov,
- preverjanje pravilnosti,
- pretvorbo (čiščenje, preverjanje poslovnih pravil, preverjanje celovitosti podatkov, pripravo agregatov),
- odlaganje v začasno odložišče,

- priprava revizijskih poročil (poročilo o skladnosti poslovnih pravil. V primeru neuspeha pomaga pri popravilu.),
- objavo (v ciljne tabele),
- arhiviranje in
- čiščenje.

3.1.2. Elementi za kvaliteten proces priprave podatkov

Elementi za kvaliteten proces priprave podatkov so:

- a. večstopenjski pristop za model arhitekture ETL (vsaj 4 stopnje):
 - funkcijska plast: glavne funkcionalnosti ETL procesa,
 - plast za operativno upravljanje: opredelitev toka nalog, parametrov, časovnikov, nadzorov, komunikacij in opozarjanj,
 - plast preverjanja in nadzora: statistika izvajanja procesov, uravnavanje in nadzor, zavračanje in obravnavanje napak, upravljanje kode ter
 - plast orodij: skupne komponente za podporo ostalim plastem,
- b. uporaba datotečnega procesiranja ETL, kadar je le-to mogoče. Vzroki za to so:
 - relativno majhni stroški skladiščenja,
 - vmesne datoteke služijo več namenom,
 - koristnost za testiranje in razhroščevanje,
 - enostavni ponovni zagon procesa ali obnavljanje podatkov,
 - možnost izračuna kontrolnih statistik,
 - zmanjšajo se odvisnosti – omogoča modularno programiranje,
 - omogoča prilagodljivost izvajanja in načrtovanje nalog in
 - obstaja večja učinkovitost izvajanja, če je pravilno kodirano. Pri tem izkorišča vzporedno procesiranje,
- c. postavljanje podatkov v središče za zmanjšanje prilagoditev ETL procesa:
 - usmerjenost k razčlenitvi (angl. Parameterization) nalog, funkcij in krmiljenja procesov,
 - definicija kod za preslikave v bazi in
 - težnja k uporabi tabel za definicijo bolj zahtevnih preslikav in poslovnih pravil in
- d. Karakteristike zasnove dobre arhitekture ETL procesa so: zmogljivost, nadgradljivost / razširljivost, prenosljivost, popravljivost, izvedljivost (možnost delitev opravil na faze, kontrolne točke,...) in preverljivost (preverjanje v dveh dimenzijah: poslovnih zahtev in tehničnih težav).

3.3. Podatkovno skladišče



Slika 7: Shema sistema poslovne inteligence

Podatkovno skladišče se nahaja na dnu sheme na sliki 7. Skladišče skrbi za elektronsko hranjenje podatkov posamezne organizacije. Podatki so shranjeni v obliki, ki omogoča lažje poročanje in analiziranje. Podatkovna skladišča vsebujejo standardizirane, skladne in prečiščene podatke, ki so združeni iz več različnih virov. Naloga podatkovnega skladišča je osredotočena na skladiščenje podatkov, potrebno pa je tudi precej podpornih sistemov za polnjenje, preoblikovanje, urejanje in zajemanje podatkov, kar je nujno potrebno za delovanje takega skladišča. Skladišče je podatkovna baza, ki je prirejena za velike količine podatkov in izvajanje poizvedb na teh podatkih.

4. Ponudniki rešitev

Ponudnikov rešitev je na trgu veliko. Nekateri ponujajo celostne rešitve, medtem ko drugi pokrivajo samo določene dele funkcionalnosti. V nadaljevanju bom predstavil dva za Slovenske razmere pomembna ponudnika, ki ponujata celovito rešitev, saj pokrivata področja podatkovnega skladišča, ETL ter sam zajem in pripravo podatkov (sestavljanje poročil). To sta ponudnika Oracle in Microsoft.

4.1. Oracle

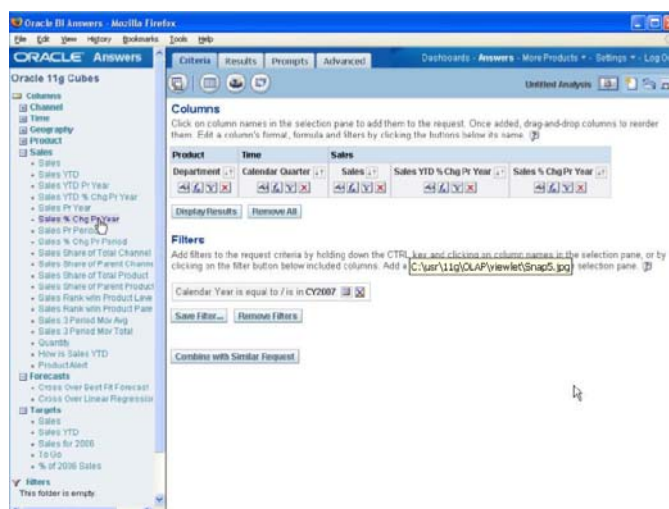
4.1.1. Orodja poslovne inteligence

Oracle svoje BI storitve ponuja pod imenom Oracle Business Intelligence Enterprise Edition OBIEE. Vsebina tega programskega paketa je v veliki meri proizvod podjetja Siebel, ki ga je Oracle prevzel v letu 2005 in njihove funkcionalnosti priključil v svoj programski paket. S tem pa je Oracle pridobil velik tržni delež, ki je v tistem času pokrival kar 45% celotnega trga orodij poslovne inteligence.

OBIEE je sestavljen iz treh delov:

- a. **Oracle BI Presentation Services** delimo na 3 področja:
 - o **Oracle Business Intelligence Answers**

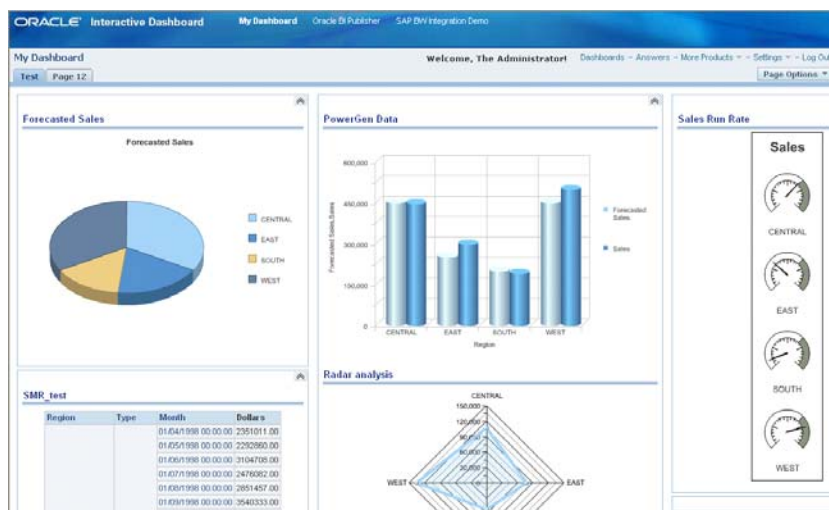
To orodje pokriva področje urejanja obstoječih in priprave novih poročil. Namenjen je zahtevnejšim uporabnikom, ki poznajo podatke in njihov pomen. Njihova naloga je sestaviti poročila, da jih lahko manj večji uporabniki uporabljajo in pregledujejo tako, da opravljajo razne poizvedbe, razvijajo grafe, vrtilne tabele, preglednice in podobno. Obstajajo panožno usmerjene predloge poizvedb, ki olajšajo pripravo poročil. Vsa pripravljena poročila je možno tudi deliti z ostalimi uporabniki. Poizvedbe je možno nastaviti tako, da osvežujejo podatke v realnem času ali ne. Vse poizvedbe se lahko shrani v poseben katalog z imenom BI Presentation Catalog, katerega vsebino lahko vključimo v katerokoli stran ali vmesnik. Prikaz rezultatov se lahko izboljšuje s prikazom grafov, prilagajanjem izgledom, izračuni, vrtnjem v globino in podobnim. Primer uporabniškega vmesnika je prikazan na sliki 8.



Slika 8: Uporabniški vmesnik paketa Oracle Business Intelligence Answers

○ Oracle Business Intelligence Interactive Dashboards

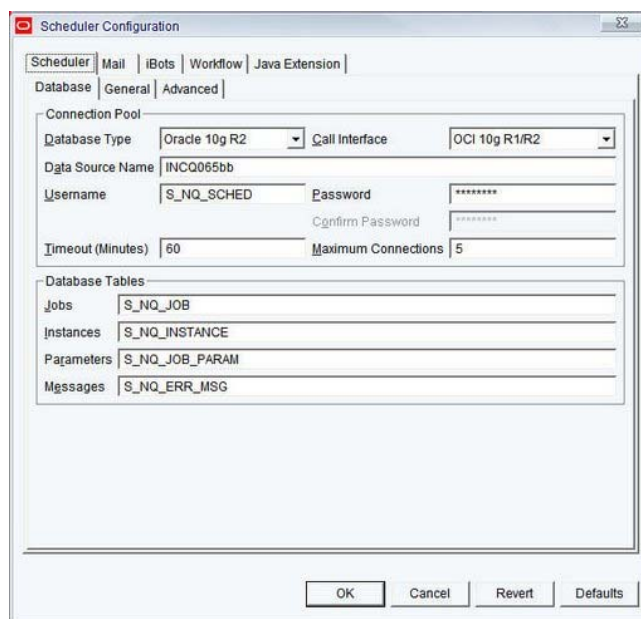
Ta del omogoča interaktiven dostop do analitičnih informacij. Ko se končni uporabnik prijavi, lahko s privzetim vmesnikom sam izvaja analize. Orodje omogoča urejanje uporabnikov v skupine in temu primeren izris vmesnika (pri tem lahko nastavimo opcije ali poročila, ki jih posameznik potrebuje). Primer uporabniškega vmesnika je prikazan na sliki 9.



Slika 9: Uporabniški vmesnik BI Interactive Dashboards

○ Oracle Business Intelligence Delivers

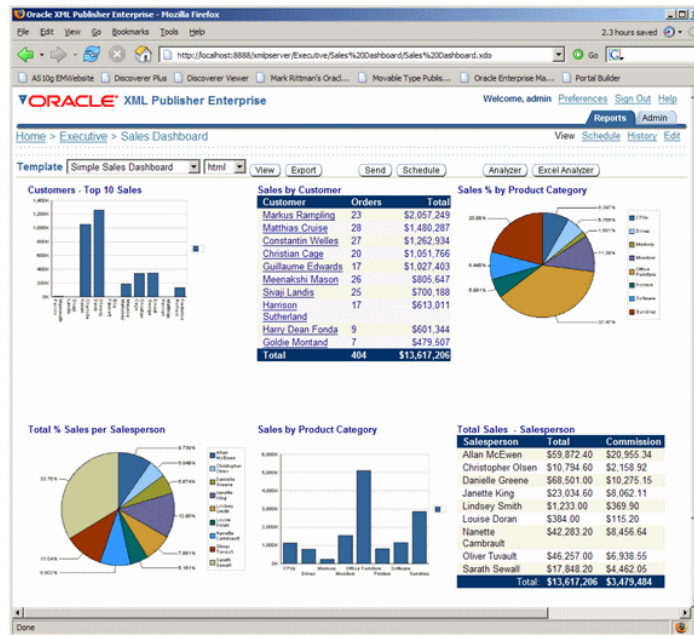
Orodje omogoča nastavljanje alarmov na rezultatih. Če v rezultatih pride do prekršitve pravila, se na predhodno izbran način sproži alarm. Obstaja več načinov alarmov: obvestilo preko spleta, brezžičnih komunikacij, mobilnih komunikacij in podobno. Sistem deluje s programskimi agenti, ki jih Oracle imenuje iBots. IBots-i se lahko sprožijo s pomočjo časovnika ali dogodka. Imajo dostop do podatkov, ki jih potrebujejo pri opravljanju analitike in filtriranja. iBots-a lahko nastavimo, da preverja vsakih 10 minut ali pa se odzove na določen dogodek. Pri tem ugotavlja če je nastavljeno pravilo prekršeno ter o tem obvesti uporabnika. Možnost obveščanja je tudi SMS sporočilo. Primer nastavljanja alarmov prikazuje slika 10.



Slika 10: Zaslonska slika Business Intelligence Delivers

b. Oracle Business Intelligence Applications

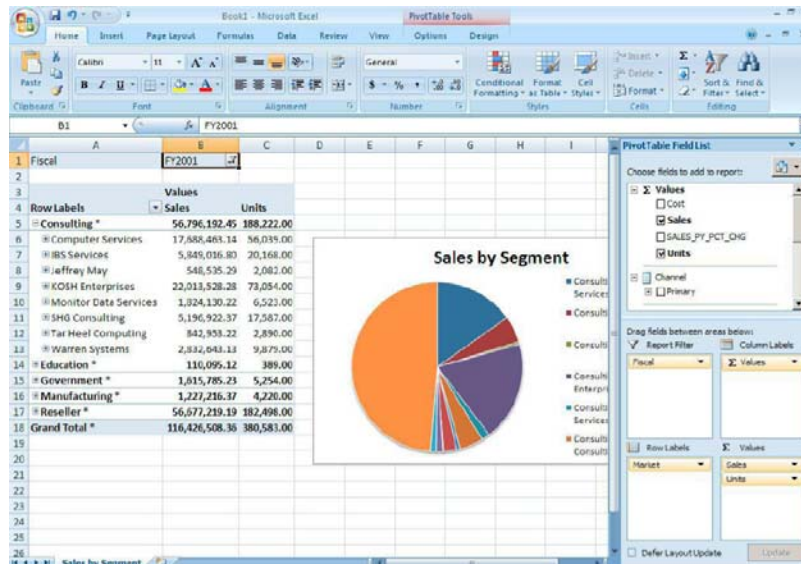
Vsebuje predhodno opisane panožno specifične digitalne pregledne plošče pod nazivom Oracle BI Interactive Dashboards in poizvedbe. Te izdelke je možno urediti glede na uporabnike in vlogo teh v podjetju. Slika 11 prikazuje primer digitalne pregledne plošče za področje prodaje.



Slika 11: Primer spletnega uporabniškega vmesnika

c. integracija z Excelom

Oracle ponuja vtičnik za Excel, s katerim lahko dostopate do podatkov v skladišču. To omogoča dodatne funkcije, saj je Excel tudi močno orodje za pripravo poročil in grafov, kar prikazuje slika 12.



Slika 12: Primer obdelave podatkov v Excelu

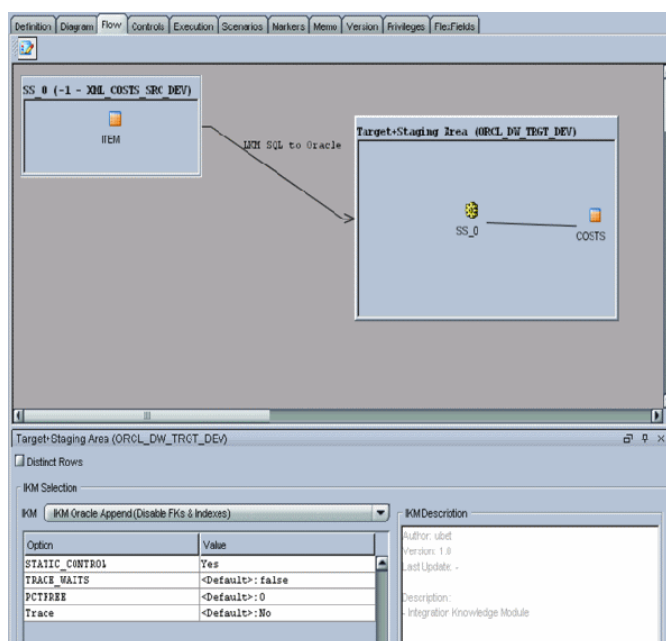
4.1.2. Orodja za pripravo podatkov

Oracle trenutno ponuja dve zelo močni orodji za izvajanje ETL procesa. Oracle Data Integrator – ODI in Oracle Warehouse Builder – OWB [11].

a. Oracle Data Integrator [17]

Gre za novo orodje, ki ga je Oracle pridobil z nakupom podjetja Sunopsis leta 2006. Poleg funkcionalnosti OWB poleg tega pa ODI ponuja še dodatne prednosti:

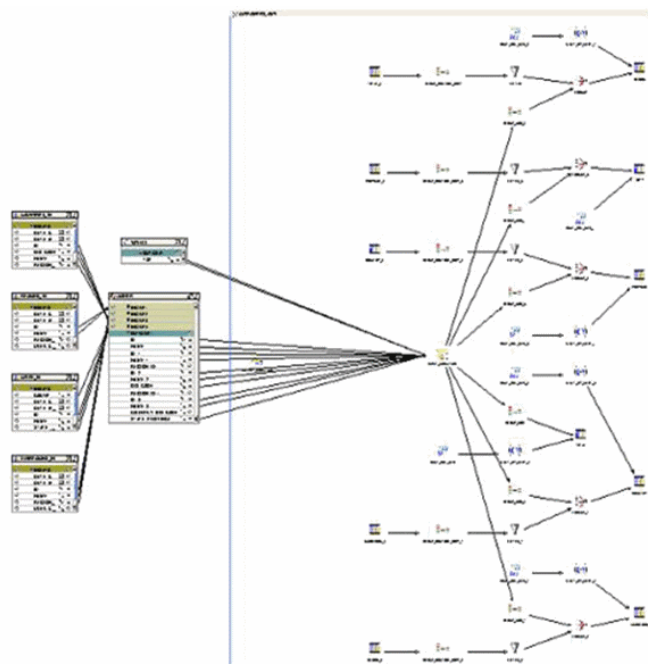
- podpora velikem številu različnih podatkovnih baz,
- bolj pregleden in enostaven pristop ob načrtovanju ETL procesa,
- glede na to, da okoli 80-90% tabel dejstev uporablja enako strategijo polnjenja, lahko tega izvedemo s pomočjo predlog (angl. Knowledge Module), ki so enake za večino tabel,
- deklarativen pristop povečuje razumevanje in tako zmanjšuje stroške razvoja. Primer modeliranja razvoja je prikazan na sliki 13,
- CDC predloge (angl. Change Data Capture) pomagajo pri vzpostavitvi zajema podatkov v realnem času in
- ima podporo beleženju različic, ki deluje na osnovi primerjanja objektov v podatkovni bazi. Poleg primerjave omogoča tudi vračanje v predhodna stanja.



Slika 13: Zaslonska slika izdelave procesa ETL z ODI

b. Oracle Warehouse Builder - OWB je primarno orodje za izvajanje ETL procesa. Druge glavne lastnosti:

- podatkovna baza je lahko samo Oracle,
- uporablja tradicionalni ETL pristop z gradnjo preslikav za vsako tabelo. Pri modeliranju takih procesov je kompleksnost velika. Primer takega modela prikazuje slika 14.



Slika 14: Zaslonska slika izdelave ETL z OWB

- c. Skupne lastnosti obeh orodij
- Obe orodji izkoriščata tako imenovan ELT pristop, kar pomeni, da se podatki ne pretvarjajo pred prepisom v bazo, temveč med samim zapisom. Podatkovna baza poskrbi za potrebne transformacije, tako da ni potrebno odvečno prepisovanje podatkov v začasno odložišče za transformacijo,
 - Orodji se počasi združujeta v eno. Predvidoma se bo to zgodilo leta 2011.

4.1.3. Podatkovno skladišče

Oracle 11g ponuja veliko izboljšav na področju podatkovne baze, še zlasti z vidika podatkovnih skladišč. [16] Nekaj novih funkcionalnosti:

- Gradnja programskih skupin (angl. Real Application Cluster - RAC) je dodatne funkcionalnost avtomatskega usklajevanja skupin (angl. Automatic Database Diagnostic Monitor - ADDM)
- upravljalci podatkovnih baz lahko za pohitritev sistema uporabljajo namige baze.

4.2. Microsoft

4.2.1. Orodja poslovne inteligence

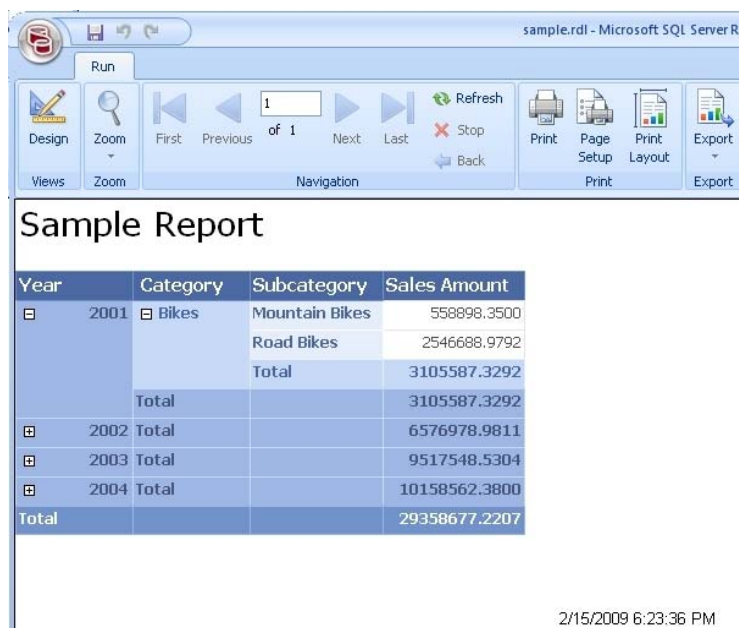
Microsoft ponuja celovite rešitve za BI [10], ki so zbrane predvsem okrog orodja SQL Server. Storitve so integrirane z večino znanih orodji, kot so recimo MS Excel 2007, kar naredi MS SQL Server še toliko bolj zanimiv in enostaven za uporabo. Glavne tri lastnosti, ki jih MS SQL Server 2008 ponuja so:

- poenoteno shranjevanje in dostop do podatkov v podjetju,
- enostavno ustvarjanje in upravljanje BI rešitev in
- enostavnejše rešitve, ki jih lahko uporablja večje število uporabnikov.

Cel sistem je razdeljen na štiri segmente, od katerih BI poročanje zajema dva:

- a. SQL Server Analysis Services – ima osnove za:
 - OLAP, ki vključuje možnost združevanja preko več dimenzij,
 - ključne kazalce uspešnosti (angl. Key Performance Indicators - KPIs) in
 - rešitve za podatkovno rudarjenje, ki uporabljajo specializirane algoritme za identifikacijo vzorcev, trendov in povezave s poslovnimi podatki. [13]

b. SQL Server Reporting Services je orodje za ustvarjanje poročil. Primer prikazuje Slika 15. Omogoča enostavno ustvarjanje, objavljanje in razdeljevanje poročil v podjetju in zunaj njega.



The screenshot displays the SSRS report viewer interface. The title bar shows 'sample.rdl - Microsoft SQL Server R'. The toolbar includes buttons for Design, Zoom, First, Previous, Next, Last, Refresh, Stop, Back, Print, Page Setup, Print Layout, and Export. The report content is titled 'Sample Report' and contains a table with the following data:

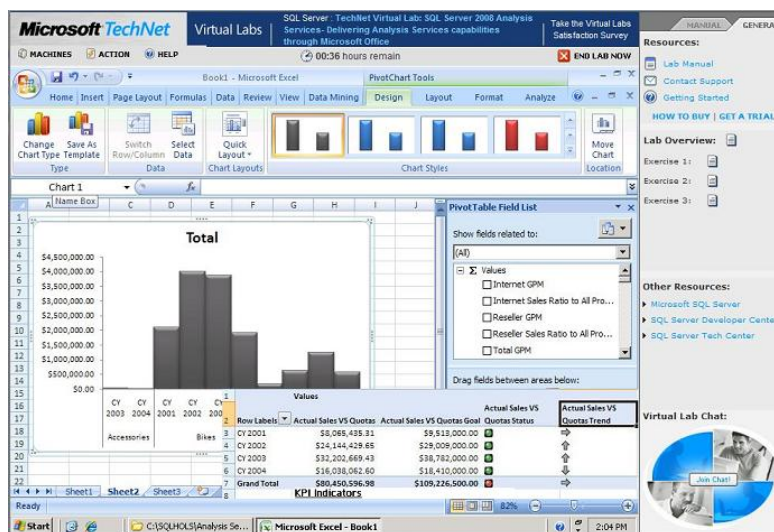
Year	Category	Subcategory	Sales Amount
2001	Bikes	Mountain Bikes	558898.3500
		Road Bikes	2546688.9792
		Total	3105587.3292
	Total		3105587.3292
2002	Total		6576978.9811
2003	Total		9517548.5304
2004	Total		10158562.3800
Total			29358677.2207

The date and time at the bottom right of the report is 2/15/2009 6:23:36 PM.

Slika 15: Primer zaslonske slike SQL Server Reporting Services

SQL Server 2008 zagotavlja celovito BI platformo s pomočjo integracije z ostalimi Microsoftovimi orodji (predvsem iz palete Microsoft Office 2007), ki lahko za uporabo pooblasti zaposlene v celotnem podjetju, kar pripomore k večji učinkovitosti. Urejanje poročila je prikazano na sliki 15. Združevanje z raznimi produkti Microsoft Office 2007:

a. Združevanje z Excelom je prikazano na sliki 16:



Slika 16: Uporaba BI funkcij v Excelu

S pomočjo Excela lahko uporabnik izvaja OLAP analize, podatkovno rudarjenje in oblikuje poročila. Glavne prednosti so:

- Izboljššan je dostop do orodja Server 2008 Analysis Services. Z Excelom uporabnik lahko pregleduje podatke, ki so shranjeni v večdimenzionalnih OLAP kockah. Te se nahajajo v Analysis Services. Excel 2007 uporabnikom omogoča enostavno izgradnjo vrtilnih tabel (angl. Pivot Tables) in dinamičnih pregledov podatkov na poljuben način s pomočjo orodij, ki so že nameščena na večini namiznih računalnikov.
 - Tesno povezovanje med orodji Excel in Analysis Services omogoča končnim uporabnikom, da preprosto uporabljajo lastnosti orodja Analysis Services, kot so: prevodi, ključni kazalci uspešnosti, izračuni, strežniške akcije in druge.
 - Napovedne analize podatkov postanejo dostopne tudi tehnično nepodkovanim uporabnikom, saj omogoča enostavno uporabo zahtevnih algoritmov za podatkovno rudarjenje v znanem okolju Excela. Obliko poročila lahko končni uporabnik po želji preuredi. Dodatki za podatkovno rudarjenje omogočajo uporabnikom, da izvajajo zahtevne analize neposredno v Excelu.
 - Obstaja možnost dodajanja alarmov pri podatkovnem rudarjenju, kadar v drugih predelih tabele podatki niso podobno porazdeljeni.
 - Poročila do končnih uporabnikov pridejo v obliki in formatu, ki ga poznajo. Urejanje samega dokumenta se ne razlikuje od urejanja ostalih dokumentov teh formatov.
- b. Povezave z ostalimi produkti Microsoft Office:
- **Microsoft Word** je program za urejanje besedil, v katerem lahko uredimo poročilo, ki ga tvori Reporting Services. SQL Server 2008 omogoča tudi težko pričakovani izvoz poročil v Word.

- **Microsoft Visio** je namenjen risanju diagramov, s katerimi izboljšamo prikaz in razumevanje rezultatov podatkovnega rudarjenja. Novi izrisi, ki jih s pomočjo Microsoft SQL Server 2008 lahko uporabljamo so: odločitvena drevesa, drevesa z poglobljanjem, diagrami gruč, odvisnostne mreže in drugo.

Poleg tega Visio omogoča tudi shranjevanje teh modelov in njihovo uporabo kot predloge za nadaljnjo uporabo ali kot spletno stran.

- SharePoint Server, (slika 17) omogoča celovito sodelovanje in objavljanje poročil na skupni mestu (angl. Table). To mesto lahko uporabimo, kot odložišče za vse BI vsebine in orodja. Tako lahko vsakdo z njegovo pomočjo pregleduje in izvaja pomembna analitična poročila.



Slika 17: Primer uporabe SharePointServer

4.2.2. Orodja za pripravo podatkov

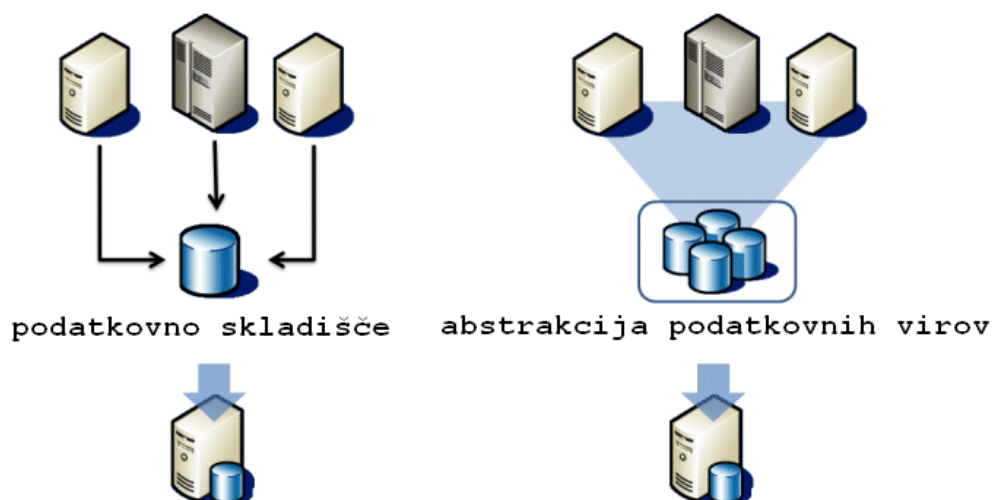
Večina organizacij za svoje delovanje uporablja več poslovnih sistemov, vsak od njih vsebuje svoje podatkovno odložišče. Kljub temu, da običajno lahko pripravimo poročila v vsakem sistemu, potrebujemo za popoln in konsistenten vpogled organizacije združene različne podatke iz različnih sistemov. Tako dobimo centralni vir poslovnih podatkov za poročila in analize.

Za prepis podatkov v podatkovno skladišče Microsoft ponuja dva pristopa:

- prepis podatkov s pomočjo ETL procesa v podatkovno skladišče ali
- abstrakcijo podatkovnih virov (angl. Data Source Abstraction), ki ne prepisuje podatkov iz virov. Za delovanje ustvari preglede po različnih virih. Za vsak tak vir ustvari preslikavo v virtualno strukturo, ki jo vidi uporabnik BI sistema. Ta abstraktna plast omogoča prijazno poimenovanje tabel, kar olajša delo in razumevanje uporabnikom. Podatki so na voljo v realnem času, saj se pobirajo direktno iz virov. Slaba stran tega pristopa je upočasnitev transakcijskih sistemov, iz katerih se podatki črpajo.

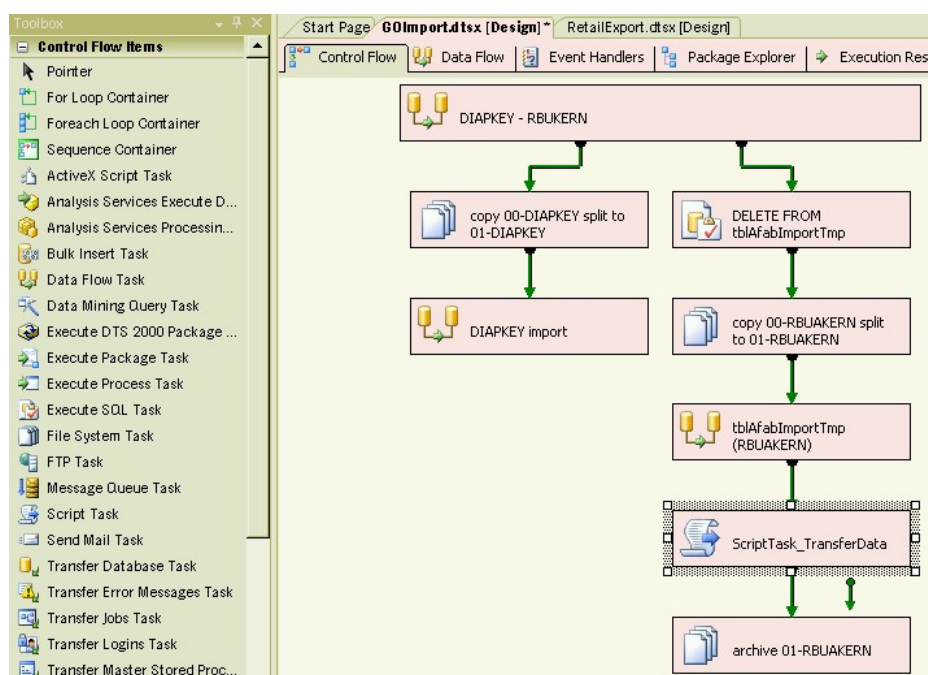
Pristopa sta grafično prikazana na slika 18. Z orodjem SQL Server 2008 lahko poljubno uporabljamo različne kombinacije. Za pregled podatkovnih virov omogoča integracijo in

dostop do podatkov, ki izvirajo iz heterogenih podatkovnih virov organizacije. Ti podatkovni viri so lahko od različnih ponudnikov, kot so recimo IBM, Oracle, Teradata in drugi.



Slika 18: Grafični prikaz dveh pristopov za zajem podatkov

SQL Server Integration Services služi kot izhodišče za gradnjo operacij ETL, ki omogočajo populacijo in sinhronizacijo podatkov iz različnih virov po organizaciji. Primer gradnje takega ETL procesa prikazuje slika 19.



Slika 19: Priprava ETL procesa z MS SQL Server

4.2.3. Podatkovno skladišče

SQL Server Database Engine [19] je podatkovno skladišče, ki je zelo razširljivo in zmogljivo. Primeren je za hranjenje velikih količin podatkov, ki se kasneje uporabljajo za analizo in poročanje. Izboljšano je bilo naključno poizvedovanje in poizvedovanje na velikih količinah

podatkov. Novost je tudi beleženje sprememb, kar zagotavlja lažje preverjanje in zagotavljanje konsistentnosti. Ob vpisu in spreminjanju podatkov lahko navedemo razloge (za vpis ali spreminjanje), kar lahko kasneje ob preverjanju pravilnosti postopkov pride zelo prav.

Nekaj novosti v različici 2008, ki so pospešila delovanje:

- **stiskanje podatkov** omogoča bolj učinkovito shranjevanje podatkov, ter posledično tudi zmanjšanje stroškov za trde diske. Prav tako stiskanje pospeši prenos velikih količin podatkov,
- **stiskanje varnostnih kopij** omogoča stiskanje že pri ustvarjanju. Tako ne potrebujemo posebnih orodij za to,
- bolj učinkovito shranjevanje za podobne podatke in kolone, kjer so podatki redki (vsebujejo veliko NULL vrednosti ali enakih vrednosti),
- **dinamično dodeljevanje večje količine pomnilnika** (Address Windowing Extension - AWE) je možnost, ki omogoča Windows Server 2003 izkoristiti do 64GB pomnilnika oziroma celo 2 TB spomina pri 64-bitnem sistemu. Tako dobimo podporo za zelo velika podatkovna skladišča in
- **upravljalca virov** (angl. Resource Governor) skrbnikom sistemov omogoča, da določijo meje za izrabo virov in prioritete za različna opravila. Tako dosežemo konstantno delovanje saj formalno določimo, kdo, kdaj in koliko bo potrošil.

Podatkovna baza vsebuje relacijski pogon, ki omogoča visoko zmogljivost in s tem gradnjo učinkovitih rešitev za podatkovno skladišče. Optimizacija poizvedb (angl. Query Optimizations) za zvezdasto shemo in orodja, ki pripomorejo k lažji uporabi indeksov in podatkovnih struktur so pomembne inovacije. Nova funkcionalnost združevanja opravljenih poizvedb omogoča razvijalcem, da bolj učinkovito obravnavajo skupne scenarije kot so preverjanje obstoja vrstic in tudi posodobitev vrstice. Novost je tudi, da je možno definirati več grupiranj za enako poizvedbo, kar je enako UNION ALL z različnimi grupiranjmi. S tem postane združevanje poizvedb in poročil še bolj enostavno in hitrejše.

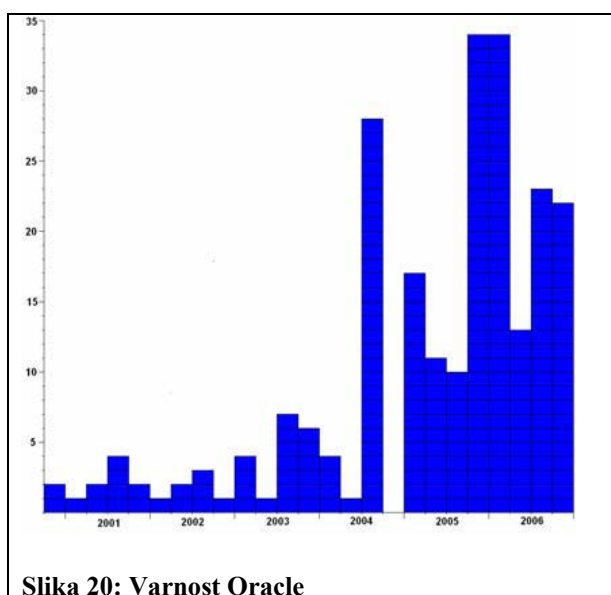
Podpora razdeljenim tabelam pomaga pri optimizaciji in vzdrževanju delovanja velikih tabel. Sedaj je tu podpora tudi za vzporedno delovanje, ki lahko v kombinaciji s porazdeljenimi tabelami lahko močno izboljša delovanje podatkovnega skladišča. Dodan je bil tudi mehanizem za beleženje meta podatkov o delovanju podatkovnega skladišča, kar omogoča nadzor in pregled porabe virov raznih rešitev. [9]

5. Primerjava ponudnikov

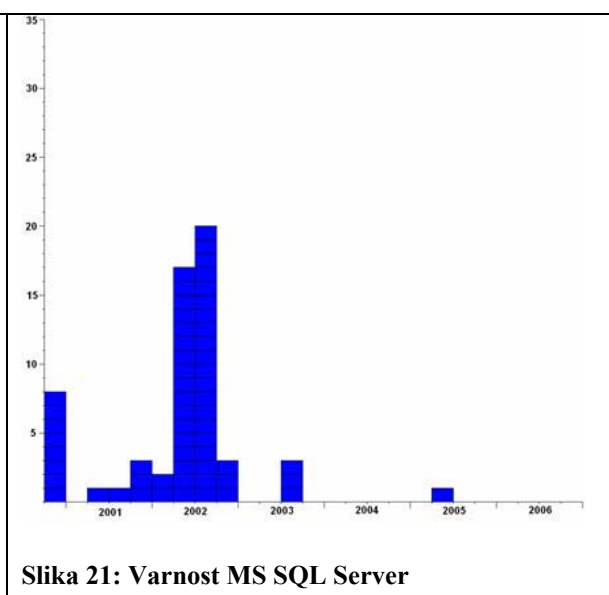
Primerjava ponudnikov Oracle in Microsoft je izvedena glede na kriterije varnosti, razvojnih orodij, poslovne inteligence, integracije, cene, neodvisnosti od platforme, Gartner Group analize in hitrosti transakcij po TPC.

5.1. Varnost

Varnost sistema se pokaže skozi čas. Zato se analiza nanaša na podatke, ki so splošni in ne vezani na verzijo podatkovne baze. V podjetju NGSSoftware so izdali poročilo o primerjavi varnosti med MS SQL Serverjem in Oracleovo podatkovno bazo. [20]



Slika 20: Varnost Oracle



Slika 21: Varnost MS SQL Server

Grafa na sliki 20 in sliki 21 prikazujeta število prijavljenih varnostnih pomanjkljivosti po letih. Primerjava je obsegala produkte Oracle 8, 9, 10 ter MS SQL Server 7, 2000 in 2005. Raziskava je bila opravljena med leti 2000 in 2006. Iz grafa je razvidno, da je Microsoft SQL po letu 2002 zelo učinkovito odpravil varnostne pomanjkljivosti, kar pa ne moremo reči za Oracle. Microsoft uporablja metodologijo SDL (angl. Security Development Lifecycle). Glavna prednost tega je, da se informacije o napakah ne izgubijo, ampak postanejo del cikla. Tako ne prihaja do ponavljanja napak. To ne pomeni, da bo nova koda brez napak, bo pa gotovo varnejša od stare. Glede na graf Oracle očitno nima podobnega pristopa k varnostnim problemom, saj se njihovo število varnostnih groženj povečuje. Res pa je, da Oracle domuje na mnogih različnih platformah, s čimer tvega tovrstne težave. Ne glede na to se veliko napak ponavlja na vseh platformah.

5.2. Razvojna orodja

Razvojna orodja so zelo pomemben člen pri odločanju za sistem, saj z dobrim orodjem lahko izdelujemo boljše produkte z manj vložena dela. MS SQL Server 2008 prisega na .NET ogrodje, ki lahko neposredno dostopa do podatkovne baze (angl. Transact-SQL), brez uporabe API-jev za dostop kot to počne Oracle. Poleg tega je delo s podatkovno bazo SQL Server 2008 popolnoma integrirano v Visual Studio .NET, kar poenostavi razvoj, saj

imamo opravka s samo enim orodjem od načrtovanja podatkovne baze do implementacije produkta.

Na drugi strani Oracle nima direktnega dostopa iz .NET ogrodja, vendar kljub temu omogoča veliko posebnih funkcionalnosti kot so:

- Nastavitve za pohitritev ali uravnavanje delovanj:
 - shranjevanje rezultatov poizvedb na klientu, kar zmanjšuje potrebe po količini pomnilnika na strežniku (angl. Client Result Cache),
 - nastavljenost načina sprejemanja rezultatov poizvedb (angl. Configurable Result Set Data Retrieval),
 - posredovanje tabelarnih parametrov baznim funkcijam in proceduram (angl. Array Parameters),
 - referenčni kazalci omogočajo, da je rezultat bazne funkcije ali procedure kazalec na rezultat (angl. REF Cursors) in
 - drugi.
- Zagotavljanje večje stopnje dosegljivosti z orodjem Oracle Data Guard, ki se zaveda .NET povezav in avtomatsko čisti sproščene povezave
- Hranjenje in zaganjanje .NET procedur v podatkovni bazi

5.3. Poslovna inteligenca

Preglednica 5 Preglednica 5 prikazuje pokritost različnih področjih poslovne inteligenca. Oba ponudnika ponujata orodja za poročanje. Microsoft je v prednosti pri oblikovanju samih poročil, saj je to podobno urejanju v Wordu ali Excelu. Oracle pa ponuja široko paleto panožnih predlog poročil, ki poenostavijo in pospešijo vzpostavitev novih poročil.

Ponudnik	Microsoft	Oracle
Orodja poslovne inteligenca		
OLAP	Zelo dobro podprt	Dobro podprt
Preglednice	Zelo dobro podprte	Dobro podprte
Digitalne pregledne plošče	podprte	Dobro podprte
Ostala orodja	Slabo podprte	Dobro podprte
Izvozi podatkov v standardne programe	Zelo dobro podprto	Dobro podprto
Alarmi in obveščanje	Malo	Zelo dobro podprto
Orodja za pripravo podatkov		
Orodje	Dobro	Zelo dobro
Razvojna orodja	Zelo dobro	dobro
Podatkovno skladišče		
Hitrost	Dobra	Dobra
Omogoča velika skladišča	da	da
Veliko referenc	Ne	da

Preglednica 5: Primerjava ponudnikov Oracle in Microsoft

5.4. Integracija z standardnimi produkti

Oba ponudnika ponujata visoko stopnjo združevanja s standardnimi produkti. Microsoft je seveda osredotočen na svoje produkte. Zelo dobra je integracija BI funkcionalnosti v razne pisarniške pakete (MS Office). Seveda tudi Oracle ponuja vtičnike za MS Excel. Oba ponudnika pa sta zelo veliko naredila na področju vključevanja vnaprej pripravljenih poročil v razne uporabniške aplikacije.

5.5. Cena

Cena je zelo kompleksen element, pri katerem moramo upoštevati več dejavnikov. Za celoto se uporablja kratica (Total Cost Of Ownership - TCO). Znano/tehtno analizo TCO [21] je pripravilo podjetje Gartner Group leta 1987. Korenine te analize segajo v prvo četrletje 20. stoletja. TCO ponuja veliko število različnih postopkov in orodij, ki pomagajo celostno določiti finančne stroške informacijske tehnologije v celem življenjskem ciklu delovanja. V odvisnosti od izvedbe lahko upoštevamo sledeče dejavnike:

- strošek strojne opreme,
- strošek programske opreme,
- strošek vzpostavitve/razvoja strojne/programske opreme,
- strošek garancije strojne opreme in vzdrževalni stroški,
- strošek licenciranja,
- strošek delovanja infrastrukture,
- strošek prostorov za infrastrukturo,
- strošek elektrike in hlajenja,
- strošek mrežne opreme,
- strošek strežniške strojne in programske opreme,
- stroški testiranja,
- stroški posodobitev in nadgraditev,
- stroški IT strokovnjakov,
- stroški izvršnega vodstva,
- stroški arhiviranja in restavriranja,
- stroški izpadov,
- stroški počasnega delovanja (uporabniki morajo čakati),
- stroški varnostnih udorov (izguba ugleda in stroški vzpostavitve stanja),
- stroški tečajev za uporabnike in IT strokovnjake,
- stroški revizije,
- stroški zavarovanj,
- stroški migracije in
- stroški ustavitve sistema

Na te dejavnike lahko gledamo z vidika točno določenega projekta ali nasploh, kadar imamo določenih več dejavnikov in ne le izbor podatkovne baze. Podrobneje bo predstavljen strošek licenciranja pri obeh ponudnikih:

- MS SQL Server 2008 se lahko plača na tri načine:
 - po številu uporabnikov se plača licenca za vsakega uporabnika in za vsako namestitev SQL Server programa,

- po številu naprav se plača licenca za vsako napravo, ki dostopa do SQL Server-ja in za vsako namestitev SQL Server programa in
- glede na procesor (jedro) se plača licenca glede na število procesorjev ali jeder.

Oracle 11g se lahko plača na dva načina v skladu z Oraclovim modelom licenciranja (Manufacturer's Suggested Retail Price - MSRP):

- po številu uporabnikov in
- glede na število procesorjev ali jeder.

Več o načinu licenciranja si lahko preberete na spletu. [25]

Pomemben podatek je, da Microsoftu pri različici Enterprise v ceni podatkovne baze vključuje tudi orodja za BI, pri Oraclu pa je storitev potrebno plačati dodatno. Glede na to, da je cena podatkovne baze primerljiva, pomeni da je Oracle dražji toliko kot nas stanejo licence za BI.

Podjetje Alinean je izvedlo raziskavo stroška vzdrževanja posameznega sistema [26]. Raziskava je glede na velikost zajemala vse vrste podjetij od malih do velikih. Merili so sledeče podatke:

- število podatkovnih baz na podjetje,
- število uporabnikov na podatkovno bazo,
- število podatkovnih baz z kritičnimi podatki,
- število transakcijskih podatkovnih baz,
- število podatkovnih baz podpora odločanju,
- število baz, ki jih vzdržuje ena oseba,
- število uporabnikov, ki jih vzdržuje ena oseba,
- letna cena vzdrževanja na podatkovno bazo in
- letna cena vzdrževanja na uporabnika.

Prišli so do zanimivih podatkov. V celoti gledano je Oracle na podatkovno bazo dražji kar za 350% od MS SQL Server-ja, vendar je treba upoštevati, da je Oraclova podpora namenjena predvsem velikim strežnikom. To se sicer počasi spreminja, saj je prihod Itanium 2 procesorja, izenačil oziroma presegel procesorsko moč RISC procesorjev za UNIX operacijski sistem. Druga pomanjkljivost MS SQL Serverja pa je bila neobstoječa podpora za programski paket SAP. S podporo temu produktu se bo odločitev za nabavo podatkovne baze gotovo povečala. [26]

5.6. Neodvisnost od platforme

Platforma označuje strojno arhitekturo in programsko arhitekturo, na kateri aplikacija lahko deluje. V našem primeru je za delovanje podatkovne baze potrebna ustrezna strojna in programska podprtost v okviru operacijskega sistema. Neodvisnost od platforme se razdeli na dva sklopa:

a. Strojna arhitektura

- MS SQL Server 2008 lahko deluje na procesorjih tipa Itanium, x86 in x86-64
- Oracle 11g lahko deluje na veliko večjem naboru procesorjev. Poleg teh, ki jih podpira Microsoft deluje še na PowerPC, Alpha, Aplha Itanium, IBM PowerPC, zSeries in SPARC procesorjih.

b. Programska arhitektura (operacijski sistem)

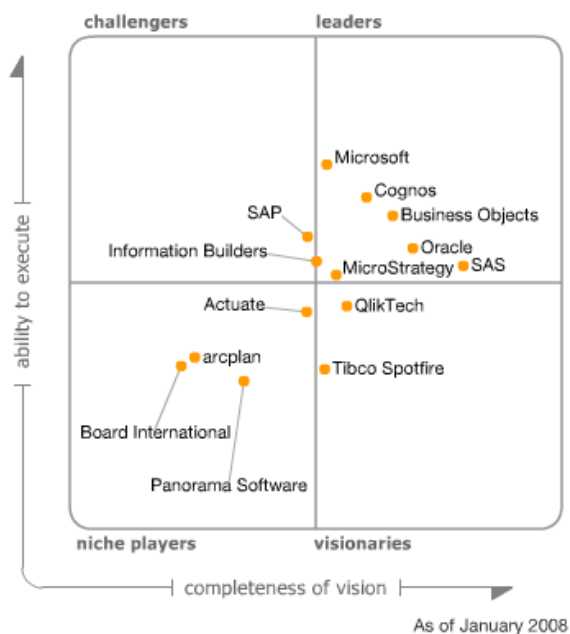
- MS SQL Server 2008 lahko namestimo le na sledeče Microsoftove operacijske sisteme: Windows XP, Windows Vista, Windows Server 2003 in Windows Server 2008.
- Oracle 11g nudi pod operacijskimi sistemi poleg palete, ki jo ponuja kot Microsoft, še več dodatnih operacijskih sistemov, ki so Mac OS X, Linux, Unix, AIX, Solaris, Tru64 UNIX, z/OS, Open VMS in Fujitsu Siemens BS2000/OSD.

Microsoft je na tem področju vidno slabši, saj Oracle podpira skoraj vse večje sisteme. Vendar ob uvedbi sistema, nam platforma ne pomeni kaj dosti. Imamo namreč prost izbor, kaj želimo. Če pa imajo podjetja že vzpostavljene sisteme, ki jih ne želijo zavreči, je Oracle v prednosti. Saj lahko ponudi svoje produkte vsem, kar je lahko zelo močan argument za izbiro.

5.7. Gartner Group analiza

Podjetje Gartner Group [15] je med 1400 direktorji informatike izvedlo analizo o pomembnosti projektov. Rezultati so pokazali, da so projekti poslovnega obveščanja na prvem mestu glede pomembnost v letu 2007. Januarja 2008 so objavili na grafu vodilne ponudnike za poslovno obveščanje, ki je prikazan na sliki 22. Graf ponazarja uspešnost ponudnikov poslovne inteligence z vidika izvedljivosti in vizije za določeno časovno obdobje. Raziskava ni bila narejena za komercialne namene, kar mu daje večjo kakovost. Gartner Group ocenjuje na trgu prisotne ponudnike na podlagi jasnosti vizije poslovnih rešitev, to pa je le eden izmed mnogih kriterijev, ki jih ocenjujejo. Zgornji desni kvadrant na grafu prikazuje podjetja Microsoft, Cognos, Business Objects, SAP, Oracle, SAS.

Ti ponudniki so vodilni v panogi. Sposobnost izvedbe je na grafu prikazana na navpični osi. Višje, ko se nahajamo, boljše je. Celostna vizija je na vodoravni osi. Bolj, ko smo na desni, boljše je naša vizija.[15] Če pogledamo Oracle ima zelo dobro vizijo, Microsoft pa izstopa glede sposobnosti izvedbe.



Slika 22: Gartner: Razporejenost ponudnikov v magičnem kvadrantu

5.8. Analize TPC

TPC (Transaction Processing Performance Council) je nedobičkonosna organizacija, ki se ukvarja s preizkušanjem in zagotavljanjem objektivnih testov hitrosti delovanja podatkovni baz.[14] Organizacija skrbi za pripravo pristopov testiranja hitrosti raznih vrst procesiranja. Trenutno so v uporabi 3 vrste testiranja: TPC-C, TPC-H, TPC-E.

TPC-C

Glavno merilo preizkusnih rezultatov pristopa TPC-C je transakcija. Pristop simulira celotno računalniško okolje, kar zajema množico uporabnikov, ki izvajajo različne transakcije na podatkovni bazi. Proces skuša posnemati večino standardnih opravil določenega sistema. Za izvajanje testov uporablja široko paleto sistemskih komponent, ki so povezane s tem okoljem. Te pa delimo v slednje skupine testov:

- za sočasno izvajanje množice različnih vrst transakcij različnih stopenj kompleksnosti,
- za izvajanje odloženih in transakcij v realnem času,
- za izvajanje več sočasnih transakcij,
- za merjenje sistemskega in aplikativnega časa izvajanja,
- za preverjanje celovitosti transakcij,
- za preverjanje enakomernosti dostopa do podatkov preko primarnih in sekundarnih ključev,
- za ugotavljanje kvalitete algoritmov za reševanje sporov pri dostopu do podatkov in
- druge.

#	Sistem	število transakcij na minuto	Cena na trans. v USD	Podatkovna baza	Operacijski sistem
1	IBM Power 595	6,085,166	2,81	IBM DB2	IBM AIX
	Bull Escala	6,085,166	2,81	IBM DB2	IBM AIX
2	HP - Itanium2	4,092,799	2,93	Oracle 10g	HP-UX
3	IBM p5 595	4,033,378	2,97	IBM DB2	IBM AIX
4	IBM p5 595	3,210,540	5,07	IBM DB2	IBM AIX
5	PRIMEQUEST	2,382,032	3,76	Oracle 10g	Red Hat
6	PRIMEQUEST	2,196,268	4,70	Oracle10g	Red Hat
7	IBM p 570	1,616,162	3,54	IBM DB2	IBM AIX
	Bull Escala	1,616,162	3,54	IBM DB2	IBM AIX
8	IBM p5 595	1,601,784	5,05	Oracle 10g	IBM AIX
9	PRIMEQUEST	1,354,086	3,25	Oracle 10g	Red Hat
10	NEC	1,245,516	4,57	Oracle 10g	Red Hat

Preglednica 6: Sistemi ocenjeni glede na hitrost izvajanja transakcij po kriteriju TPC-C

Uspešnost preizkusnih rezultatov pristopa TPC-C je število izvedenih transakcij v minuti v zgoraj naštetih načinih testiranja. Poleg tega se pripravi tudi razvrstitev rezultatov glede na ceno samega sistema. Te meritve so namenjene ocenjevanju hitrosti transakcijskih sistemov.

#	Sistem	število transakcij na minuto	Cena na transakcijo v USD	Podatkovna baza	Operacijski sistem
1	HP ProLiant	232,002	0,54	Oracle 11g	Oracle Linux
2	Dell PowerEdge	104,492	0,6	Oracle 11g	MSW Server 2003
3	Dell PowerEdge	97,083	0,68	Oracle 11g	MSW Server 2003
4	HP ProLiant	102,454	0,73	Oracle 11g	MSW Server 2003
5	HP ProLiant	100,926	0,74	Oracle 10g	Oracle Linux
6	HP ProLiant	82,774	0,84	SQL Server 2005	MSW Server 2003
7	Dell PowerEdge	20,705	0,85	Sybase SQL	MSW Server 2003
8	PowerEdge	69,564	0,91	SQL Server 2005	MSW Server 2003
9	HP ProLiant	579,814	0,96	SQL Server 2005	MSW Server 2003
10	HP ProLiant	639,253	0,97	Oracle 11g	Oracle Linux

Preglednica 7: Sistemi ocenjeni glede na hitrost izvajanja transakcij in ceno po kriteriju TPC-C

V preglednici 6 je prikazana ocena prvih desetih sistemov glede na hitrost procesiranja. Kot je razvidno sta trenutno prevladujoča dva ponudnika in sicer IBM in Oracle. Microsoft se na to lestvico ni uspel uvrstiti. Ko pa pogledamo preglednico glede na ceno transakcije pa so rezultati malo drugačni. Preglednica 7 prikazuje, da ob upoštevanju cene ponudnika IBM ni več na lestvici. Microsoft, Oracle, Sybase pa še ostajajo. Oracle je še vedno prevladujoč, vendar se moramo zavedati, da so to v povprečju 25-krat manjši sistemi. Razvidno je, da je Oracle prisoten tako na velikih kot majhnih transakcijskih sistemih, Microsoft pa zaenkrat velikih transakcijskih sistemov še ni uspel osvojiti.

TPC-E

TPC-E je nov obremenitveni test za OLAP. Kot merilo se uporablja model podatkovne baze za potrebe borzno posredniške hiše. Količina podatkov je nastavljen parameter, kar omogoča testiranje sistemov različnih velikosti. Obremenitveni test določa skupek različnih transakcij, ki jih test mora ohranjati skozi čas. Rezultat testa je število transakcij na sekundo.

Na začetku je bilo omenjeno, da je sistem zasnovan na podatkih borzno posredniške hiše, vendar s tem ne izgubi reprezentativnosti. Glede na preglednico 8 in preglednico 9 lahko sklepamo, da na tem področju prevladuje Microsoft, saj je na obeh lestvicah (po hitrosti in ceni), edini prisoten med prvimi desetimi sistemi. Zanimivo je tudi, da v vseh primerih nastopa v enaki kombinaciji in sicer z operacijskim sistemom Windows Server 2008 in podatkovno bazo SQL Server 2008. Očitno je Microsoftu z najnovejšim operacijskim sistemom in podatkovno bazo uspelo zasenčiti konkurenco na OLAP področju.

#	Sistem	Število transakcij na sekund	Cena za sistem v USD	Podatkovna baza	Operacijski sistem	Procesorji / jedra/ niti
1	NEC	1,568	1,180	SQL Server 08	MSW Server 08	16 / 64 /64
2	Unisys	1,493	917	SQL Server 08	MSW Server 08	16 / 64 / 64
3	NEC	1,400	1,19	SQL Server 08	MSW Server 08	12 / 64 / 64
4	IBM x3950	1,250	1,311	SQL Server 08	MSW Server 08	16 / 64 / 64
5	Unisys	1,165	783	SQL Server 08	MSW Server 08	8 / 48 / 48
6	NEC	1,126	2,771	SQL Server 08	MSW Server 08	32 / 64 / 64
7	IBM x3650	817	319	SQL Server 08	MSW Server 08	2 / 8 / 16
8	IBM x3950	804	1,45	SQL Server 08	MSW Server 08	8 / 32 / 32
9	PRIMERGY	800	343	SQL Server 08	MSW Server 08	2 / 8 / 16
10	IBM x3650	798	378	SQL Server 08	MSW Server 08	2 / 8 / 16

Preglednica 8: Sistemi ocenjeni glede na hitrost izvajanja po kriteriju TPC-E

#	Sistem	Število transakcij na sekundo	Cena za sistem v USD	Podatkovna baza	Operacijski sistem	Procesorji / jedra/ niti
1	IBM x3650	817	319	SQL Server 08	MSW Server 08	2 / 8 / 16
2	Dell	766	334	SQL Server 08	MSW Server 08	2 / 8 / 16
3	PRIMERGY	800	343	SQL Server 08	MSW Server 08	2 / 8 / 16
4	IBM x3650	798	378	SQL Server 08	MSW Server 08	2 / 8 / 16
5	Dell	635	403	SQL Server 08	MSW Server 08	4 / 16 / 16
6	IBM x3850	729	457	SQL Server 08	MSW Server 08	4 / 24 / 24
7	PRIMERGY	721	459	SQL Server 08	MSW Server 08	4 / 24 / 24
8	Dell	671	500	SQL Server 08	MSW Server 08	4 / 24 / 24
9	PRIMERGY	317	523	SQL Server 08	MSW Server 08	2 / 8 / 8
10	PRIMERGY	492	559	SQL Server 08	MSW Server 08	4 / 16 / 16

Preglednica 9: Sistemi ocenjeni glede na hitrost in ceno po kriteriju TPC-E

TPC-H

TPC-H služi merjenju hitrosti sistemov za odločanje. Testiranje temelji na hitrih poizvedbah in sočasnemu spreminjanju podatkov. Poizvedbe in podatki so izbrani z namenom zadovoljiti potrebe gospodarstva. Naloga teh testov je ponazarjanje učinkovitosti sistemov za poslovno inteligenco. Testi ponazarjajo preučevanje velikih količin podatkov, izvajanje poizvedb visoke zahtevnosti in iskanje odgovorov na kritična vprašanja.

Rezultat tega testa je število sestavljenih poizvedb na uro za določeno količino podatkov v podatkovnem skladišču, kar odraža različne zmožnosti procesiranja podatkov.

Za razliko od prejšnjih dveh kriterijev ocenjevanja je pri tem število ponudnikov večje. Zaradi obsežnosti analize bom prikazal samo najmanjšo kategorijo (100GB) in eno večjih (10.000 GB). V preglednici 10 so prikazani rezultati za velikost skladišča 100 GB.

Rezultati za velikost skladišča: 100 GB					
#	Sistem	Število sestavljenih poizvedb na uro	Cena na sestavljeno poizvedbo v USD	Podatkovna baza	Operacijski sistem
1	Kickfire	49,228	0,7	MySQL	Kickfire
2	HP	51,422	1,07	SQL Server 08	MSW Server 08
3	HP	51,085	1,09	SQL Server 08	MSW Server 08
4	CPI Phoenix	209,298	1,25	EXASOL	EXASOL
5	PowerEdge	28,772	1,46	SQL Server 08	MSW Server 08
6	SunFire	98,857	2,65	ParAccel	Red Hat
7	SunFire	8,337	3,3	Sybase	Solaris
8	HP	34,989	3,97	SQL Server 05	MSW Server 03
9	HP	4,521	4,3	SQL Server 05	MSW Server 03
10	SunFire	4,132	4,61	Sybase	Solaris

Preglednica 10: Sistemi ocenjeni glede na število sestavljenih poizvedb za 100 GB podatkov po kriteriju TPC-E

Zopet prevladuje ponudnik Microsoft, saj je kar polovica od prvih 10 sistemov last Microsoft-a. Oracle tukaj ni prisoten, se pa začne pojavljati v večjih skladiščih. Tako da je pri skladiščih velikosti 10.000 GB, na lestvici več kot polovica njihovih strežnikov. Ta primerjava je prikazana v preglednici 11. Sledi mu Microsoft in IBM. Vidimo lahko, da je pri majhnih količinah podatkov ponudnikov zelo veliko, ko pa ta količina naraste se številka skrči na tri ponudnike. To so Microsoft, Oracle in IBM.

Rezultati za velikost skladišča: 10.000 GB					
#	Sistem	Število sestavljenih poizvedb na uro	Cena na sestavljeno poizvedbo v USD	Podatkovna baza	Operacijski sistem
1	Unisys	80,172	18.95	SQL Server 08	MSW Server 08
2	HP	208,457	27.97	Oracle 11g	HP-UX
3	IBM	343,551	32.89	IBM DB2	IBM AIX
4	HP	171,38	32.91	Oracle 10g	HP-UX
5	HP	63,65	38.54	SQL Server 08	MSW Server 08
6	IBM	180,108	47.00	IBM DB2	IBM AIX
7	Sun	108,099	53.80	Oracle 10g	Solaris 10
8	HP	49,104	118.13	Oracle 10g	HP UX
9	HP	86,282	161.24	Oracle 10g	HP UX

Preglednica 11: Sistemi ocenjeni glede na število sestavljenih poizvedb za 10.000 GB podatkov po kriteriju TPC-E

6. Sklepne ugotovitve

Poslovno inteligenca je veliko, predvsem pa ekonomsko pomembno področje, zato ga je potrebno obravnavati čim bolj celovito in poglobljeno. Beseda poslovna inteligenca se velikokrat uporablja splošno, kar otežuje pravo razumevanje pomena.

Z BI mislimo na znanje, tehnike, programe in ustaljene postopke, ki pomagajo podjetjem pridobiti lažje razumevanje samega poslovanja ter posledično omogočajo lažje odločanje. V pričujočem diplomskem delu smo opredelili BI na tri glavne sklope: podatkovno skladišče, orodja za pripravo podatkov in orodja poslovne inteligence. Vsa tri področja so enako pomembna za delovanje sistema.

Glavne možnosti izboljšav:

- hitrost delovanja je zaradi vse večjih količin podatkov problem. Zaradi velike konkurence se ponudniki na vse pretege trudijo najti boljše in hitrejše načine vpisa in branja podatkov.
- poenostavitev uporabniških vmesnikov na vseh delih BI sistema. Od prijazne rešitve za končnega uporabnika do enostavnih orodij za skrbnike sistemov in razvijalce.

V opisu ponudnikov smo izpostavili dva: Oracle in Microsoft. Razlog je, da sta oba primerna tako za veliko kot tudi majhne sisteme. Majhnih je v Sloveniji veliko. Sami smo se srečali z obema ponudnikoma in številne storitve preizkusil v delovanju. Zaradi aktualnosti smo se osredotočili na zadnje različice in sicer Oracle 11g in Microsoft SQL Server 2008. Na spletu zanj še ni prav veliko primerjav, ker sta izdelka sorazmerno nova.

Za primerjavo smo izbrali nekaj kriterijev, ki najbolj vplivajo na izbiro. Razvrstiti kriterije po pomembnosti je težko, ker je potrebo pred tem vedeti kakšen sistem potrebujemo in kakšne so naše omejitve glede tehnologije in cene.

Za primerjavo smo izbrali nekaj najpomembnejših kriterijev:

- cena je v današnjem svetu eden od najpomembnejših, vendar nanjo ne moremo gledati enoznačno. Cena sicer je pokazatelj, vendar samo v primeru, ko izbiramo med več ponudniki, ki zadoščajo vsem ostalim kriterijem,
- varnost je tudi vse bolj pomemben kriterij, saj si podjetja v tekmovalnem okolju težko privoščijo neskrbno ravnanje s podatki. Tukaj je predvsem pomembno kako ponudnik rešuje odkrite varnostne luknje. Z uporabniškega stališča mora biti sistem posodobitev čim bolj avtomatiziran. Microsoft je znan po tem, da ima zelo dobro rešeno posodabljanje programov, pri Oraclu pa se ti popravki nalagajo ročno. Poleg tega je sam proces posodobitve precej kompleksen,
- nova razvojna orodja in tehnologije predstavlja dodatne možnosti razvoja lastnih rešitev. Oba ponudnika sta zelo dobra v .NET tehnologiji. Ker je to Microsoftova tehnologija, jo je Microsoft zelo dobro integriral v svoje izdelke. Z drugega zornega kota pa Oracle ne zaostaja, saj omogoča veliko naprednih funkcij, ki pa jih Microsoft nima,
- poslovna inteligenca predstavlja, kako se celoten sistem obnaša v praksi. Prednost Oracla je v tem, da ima dobre reference oziroma zelo veliko namestitvev svojih sistemov, kar Microsoftu še manjka. V nekaj letih se to zna zelo spremeniti in takrat bo lahko slika popolnoma drugačna. Kar se tiče funkcionalnosti BI sistema oba ponudnika ponujata dobre rešitve, ki jih tudi neprestano izboljšujeta,

- Gartner Group ocenjuje, da ima podjetje Microsoft večjo sposobnost izvedbe, podjetje Oracle pa izstopa po dobri viziji. Microsoft je v zadnjih letih izredno izboljšal svoje tehnologije, predvsem na področju velikih sistemov. Raziskave pa kažejo, da je kljub temu prisotnost Microsofta na področju večjih sistemov bolj izjema kot pravilo in
- hitrost delovanja je pomembna. Predstavljena je analiza TPC, ki primerja hitrostne zmogljivosti podatkovnih baz. Sodeč po lestvicah lahko sklepamo, da je pri transakcijskih sistemih Oracle v veliki prednosti, na področju sistemov OLAP pa je slika popolnoma obrnjena. Microsoft prevladuje tako v kategoriji manjših kot večjih sistemov. Analiza, ki je bila izvedena za različne velikosti podatkovnih skladišč, je ugotovila, da obstaja za manjše sisteme zelo veliko različnih ponudnikov, pri velikih sistemih pa med najboljšimi ostaneta samo IBM in Oracle.

V zadnjem času je bilo kar nekaj inovacij pri obeh ponudnikih. Ena večjih inovacij ETL procesa podjetja Oracle je pohitritev izvedbe procesa. V konkretnem primeru je svoji stranki (italijanski pošti) pospešila prepis podatkov iz trinšestdeset ur na deset minut. Nov način ELT procesa namreč izvaja preureditev in preverjanje podatkov v sami podatkovni bazi. Tako je lahko podjetje pričelo s polnjenjem podatkovnega skladišča dnevno, kar pred to inovacijo ni bilo možno. Na drugi strani je Microsoft ponudil orodja za preverjanje kvalitete rešitev, ki že v razvoju opozorijo na slabo načrtovane rešitve.

Oracle je dlje časa na tem področju in ima veliko izkušenj, zato je zelo dober za zahtevne sisteme, kjer je potrebno veliko usposobljenih in izkušenih strokovnjakov. Pri Microsoftu po drugi strani prisegajo predvsem na enostavnost brez potrebe po strokovnjakih, zato se poraja vprašanje, kako bi Microsoftovi produkti delovali na zelo zahtevnih sistemih.

Vseeno je razglasitev Oracla za zmagovalca napačna, saj je ključno, da je izbrani produkt prilagojen tipu organizacije in njenim zahtevam. Najboljše pravilo je potemtakem: preučimo prednosti in slabosti v konkretnem primeru in se na podlagi tega odločimo za ponudnika.

7. Seznam uporabljenih virov

- [1] Turban, McLean, Wetherbe – Information technology for management – transforming business in the digital economy 3rd edition
- [2] Splošno o poslovni inteligenci. Vir na spletu: http://en.wikipedia.org/wiki/Business_intelligence
- [3] Splošno o poslovni inteligenci. Vir na spletu: http://en.wikibooks.org/wiki/Business_Intelligence/Introduction
- [4] OLAP kocka. Vir na spletu: http://wiki.fmf.uni-lj.si/wiki/OLAP#Kaj_je_OLAP_Kocka.3F
- [5] Poročila o ponudnikih storitev poslovnega obveščanja. Vir na spletu: <http://www.sas.com/news/analysts/idc-ww-bi-tools-2008.pdf>
- [6] Proces ETL. Vir na spletu: http://en.wikipedia.org/wiki/Extract,_transform,_load
- [7] Orodja poslovne inteligence. Vir na spletu: http://en.wikipedia.org/wiki/Business_intelligence_tools
- [8] Digitalna pregledna plošča. Vir na spletu: <http://en.wikipedia.org/wiki/File:DoSscreenshot.jpg>
- [9] Novosti v MS SQL Server 2008. Vir na spletu: <http://www.microsoft.com/sqlserver/2008/en/us/whats-new.aspx>
- [10] Poslovna inteligenca v MS SQL Server 2008. Vir na spletu: <http://www.microsoft.com/sqlserver/2008/en/us/business-intelligence.aspx>
- [11] Primerjava Oracle ODI in OWB. Vir na spletu: <http://www.business-intelligence-quotient.com/?tag=odi-owb-comparison>
- [12] Poslovna inteligenca. Vir na spletu: http://en.wikibooks.org/wiki/Category:Business_Intelligence
- [13] MS SQL Server Analysis services. Vir na spletu: <http://www.microsoft.com/sqlserver/2008/en/us/analysis-services.aspx>
- [14] Organizacija za izvajanje meritev hitrosti. Vir na spletu: <http://www.tpc.org/>
- [15] Organizacija Gartner. Vir na spletu: <http://www.gartner.com/technology/home.jsp>
- [16] Oracle na splošno. Vir na spletu: <http://www.oracle.com/technology/index.html>
- [17] Oracle ODI primer uporabe. Vir na spletu: <http://www.rittmanmead.com/files/An%20Introduction%20to%20Oracle%20Data%20Integrator.pdf>
- [18] Primer digitalne pregledne plošče. Vir na spletu: <http://www.axelerate.com>
- [19] Microsoft SQL Server 2008 Business Intelligence - White Paper
- [20] Primerjava varnosti produktov. Vir na spletu: <http://www.ngssoftware.com>
- [21] Kaj pomeni TCO. Vir na spletu: http://en.wikipedia.org/wiki/Total_cost_of_ownership
- [22] Cenovna politika podjetja Oracle. Vir na spletu: http://oraclestore.oracle.com/OA_HTML/ibeCCTpSctDspRte.jsp?section=11365&media=os_user_minimums&minisite=10061&respid=1298373&grp=STORE&language=US
- [23] Minimalne zahteve za delovanje MS SQL Server 2008. Vir na spletu: <http://msdn.microsoft.com/en-us/library/ms143506.aspx>
- [24] Zahteve za delovanje Oracle. Vir na spletu: http://en.wikipedia.org/wiki/Oracle_database
- [25] Licenciranje Oracle 11g. Vir na spletu: http://oraclestore.oracle.com/OA_HTML/ibeCCTpSctDspRte.jsp?section=11365&media=os_user_minimums&minisite=10061&respid=1298373&grp=STORE&language=US
- [26] Analiza stroškov vzdrževanja sistema. Vir na spletu: <http://www.alinean.com/>
- [27] Šest sigma. Vir na spletu: http://en.wikipedia.org/wiki/Six_Sigma

- [28] preučevanje stroškov po aktivnostih. Vir na spletu:
http://en.wikipedia.org/wiki/Balanced_scorecard
- [29] celovito upravljanje kvalitete. Vir na spletu: http://en.wikipedia.org/wiki/Activity-based_costing
- [30] celovito upravljanje kvalitete. Vir na spletu:
http://en.wikipedia.org/wiki/Total_quality_management
- [31] dodane ekonomske vrednosti. Vir na spletu:
http://en.wikipedia.org/wiki/Total_quality_management
- [32] celovito stroškovno merjenje. Vir na spletu: <http://www.mqi.com/msmt.htm>
- [33] Najbližji sosed. Vir na spletu: http://en.wikipedia.org/wiki/Nearest-neighbor_interpolation
- [34] Naivni Bayes klasifikator. Vir na spletu:
http://en.wikipedia.org/wiki/Naive_Bayes_classifier
- [35] Nevronske mreže. Vir na spletu: http://en.wikipedia.org/wiki/Neural_network