



SQL Server 2008 R2

Poslovna inteligencija

Josip Šaban, dipl. ing. rač.

Thanks to all our
sponsors,
they made this possible!

Sponsors:

Microsoft

Your potential. Our passion.™

Microsoft | Innovation Center
Varaždin

 PARK PRIRODE KOPAČKI RIT 

 insa

redgate®

SPAN*


ALGEBRA
UČILIŠTE

 **SOLID QUALITY**
MENTORS

 ineta

Media sponsor:

MREŽA



Ukratko o autoru

- **Formalno obrazovanje**
 - Diplomirao na FER-u 2004, smjer računarstvo
 - Pohađa doktorski studij na FOI-u te Cotrugli MBA
- **Stručno obrazovanje i dosadašnja karijera**
 - Više stručnih certifikata iz područja Microsoft i Oracle tehnologija te vođenja projekata
 - Višegodišnji predavač na Windaysima i ostalim lokalnim konferencijama
 - Trenutno radi na poziciji konzultanta za poslovnu inteligenciju u poduzeću Alfatec Group d.o.o



Ukratko o predavanju

- Što je poslovna inteligencija?
- Tko je tko u svijetu poslovne inteligencije?
- Od čega se sastoji poslovna inteligencija?
 - Alati za poslovnu inteligenciju SQL Server-u 2008 R2
- Kako izgleda “standardni” projekt poslovne inteligencije?
 - Zašto se projekti poslovne inteligencije, uz projekte implementacije ERP sustava, redovno proglašavaju kao “preskupi i predugi” te često “neuspješni”



Što nećete ovdje vidjeti

- Isključivu “reklamu” Microsoft-a već pokušaj objektivnog pregleda tržišta i tehnologija
- Detalje o pojedinim područjima radi vremenskog ograničenja predavanja
 - Pitanja su naravno dobrodošla u svakom trenutku
- “Tečaj” poslovne inteligencije na SQL Server 2008 R2 platformi



Definicija poslovne inteligencije

- “A broad category of applications and technologies for gathering, storing, analyzing, sharing and providing access to data to help enterprise users make better business decisions.” – Gartner
 - Tehnički gledano, poslovna inteligencija je skup metoda za skupljanje i analizu podataka
 - Ključ je u podatku i njegovom odnosu sa ostalim podacima a ne otkuda i kako smo došli do njega

Figure 1. Magic Quadrant for Business Intelligence Platforms



Gartnerov pogled na tržište

- Prevladavaju “veliki igrači”
- Brzo rastuće područje IT-a
- Potreba za analizom i konsolidacijom podataka se trajno povećava
- Dugoročno, nastati će stabilizacija tržišta
- Vjerodostojnost Gartnera?



Poslovna inteligencija

Što je poslovna inteligencija i čemu služi?



Što je poslovna inteligencija

- U prvom redu to je priča o znanju i ljudima a ne o tehnologiji
 - Tehnička izvedba je samo nadogradnja – izgradnja skladišta podataka, izvještaja, analiza, ...
 - Ukoliko nemamo razlog zašto krećemo u neki posao onda ga je bolje i ne počinjati



Kako stvoriti sustav poslovne inteligencije

- Koje podatke staviti u sustav i kako ih organizirati?
- Kako ih napuniti i dostaviti krajnjim korisnicima?
- Što je potrebno učiniti da se to sve ostvari?
- Kako održavati sustav u dugom vremenskom razdoblju?

<http://www.roseindia.net/technology/business-intelligence/history-of-business-intelligence.shtml>
History Of Business Intelligence - http://www.youtube.com/watch?v=_1y5jBESLPE



Poslovno – tehnički problemi

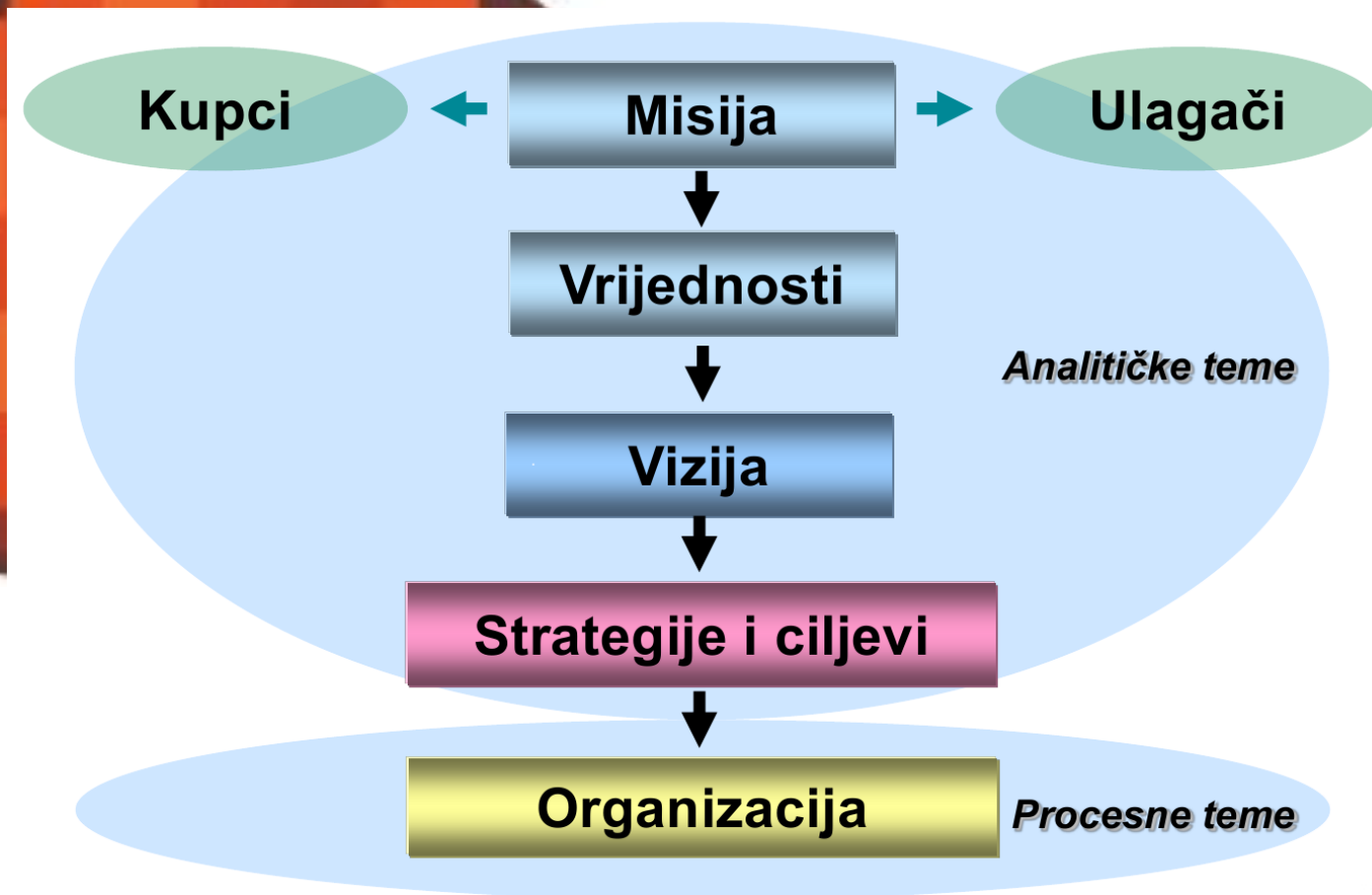
- Spori upiti nad transakcijskim podacima
- Općenito usporeenje sustava
- Ručno pisanje često vrlo složenih upita
- Različiti izvori podataka
- Podaci nisu dostupni svim korisnicima
- Previše podataka



Koje podatke staviti u BI sustav

- Sustavi poslovne inteligencije izrađuju se sa unaprijed definiranim ciljevima (uglavnom 😊) i služe za mjerenje organizacije
 - Što ne možete mjeriti ne možete kontrolirati
 - Što ne možete mjeriti ne možete poboljšati
- To je sustav kojim daje odgovore na dva ključna pitanja
 - Gdje smo?
 - Kamo idemo?

Koje podatke staviti u BI sustav





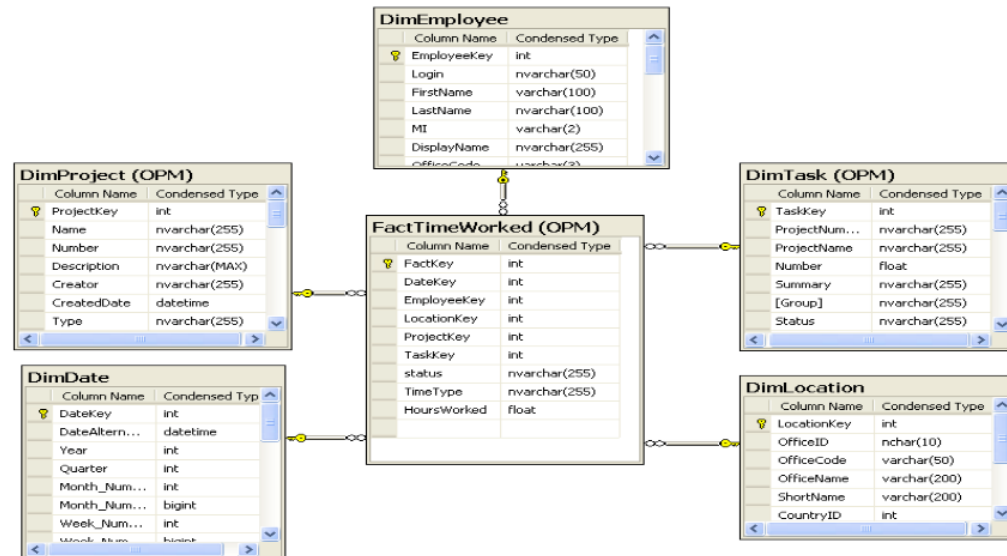
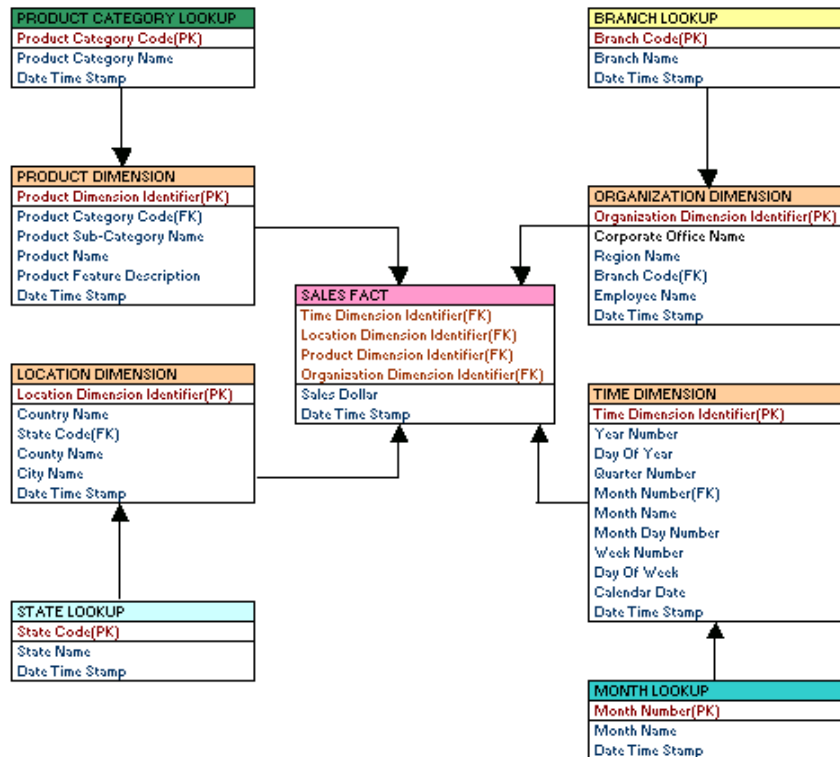
Kako organizirati podatke

- Koristimo dimenzionalni model koji nam služi da prevedemo poslovna pravila u strukturu skladišta podataka
 - Osnovni elementi modela su mjere i dimenzije
 - Mjere određuju što gledamo
 - Dimenzije ono po čemu gledamo (sadrže niz atributa)

- Prilikom izrade modela, najčešće korištenjem alata kao što su ERWin ili Microsoft Visio, slijedimo najčešće uobičajeni proces konstrukcije
 - Definiranje procesa
 - Definiranje dimenzija
 - Definiranje mjera
- Rezultat ovog procesa je dimenzionalni model u svojoj prvoj verziji

- Dimenzionalni model je specifikacija za implementaciju sustava
 - Definicija svih pojmova (podataka) u sustavu
 - Definicija „vlasnika” pojedinih područja, izvora podataka i transformacija
 - Pisan svima razumljivim jezikom ali istovremeno omogućava kasniju izradu skladišta podataka
 - Nikad nije konačan već ovisi o novim saznanjima

Izrada modela

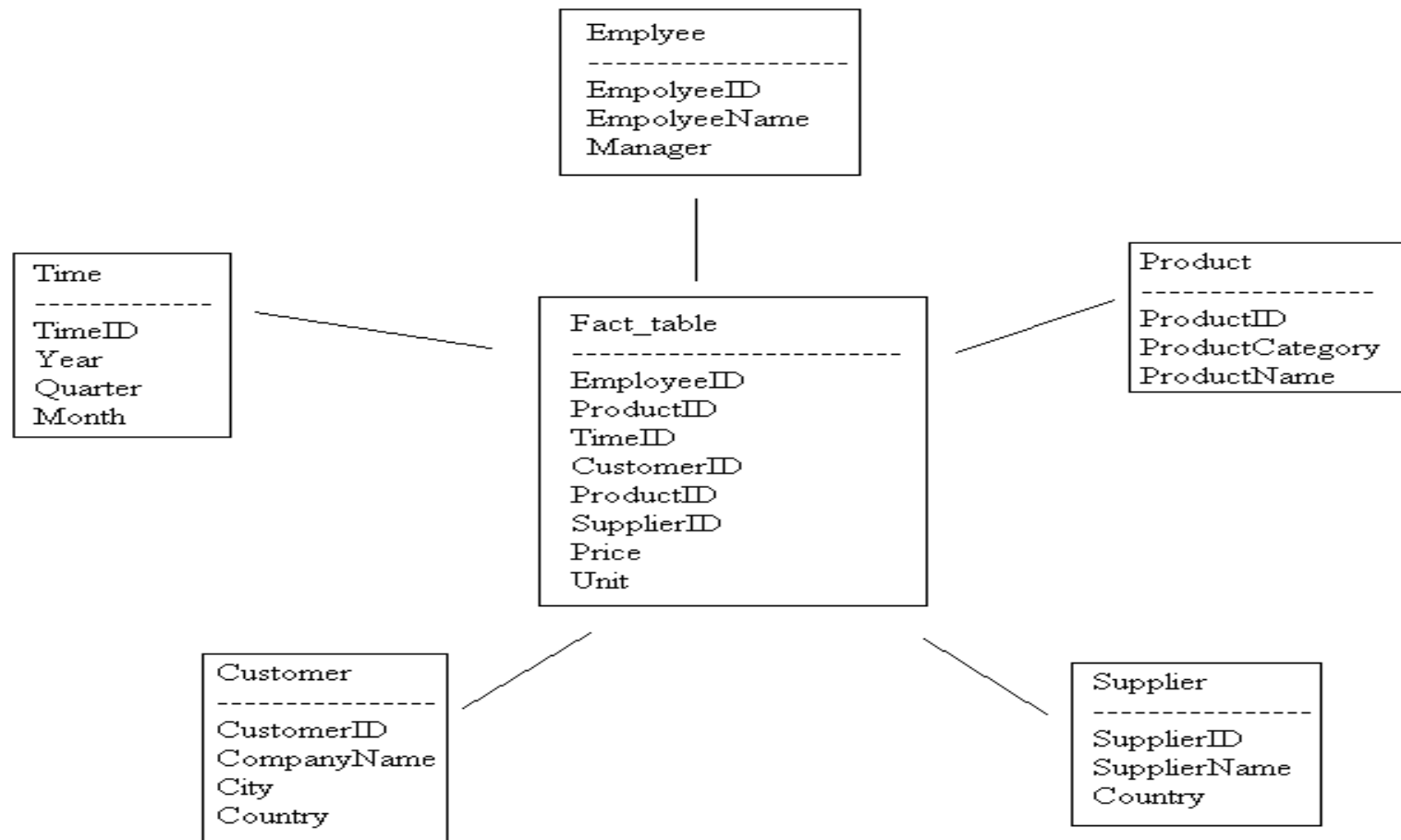




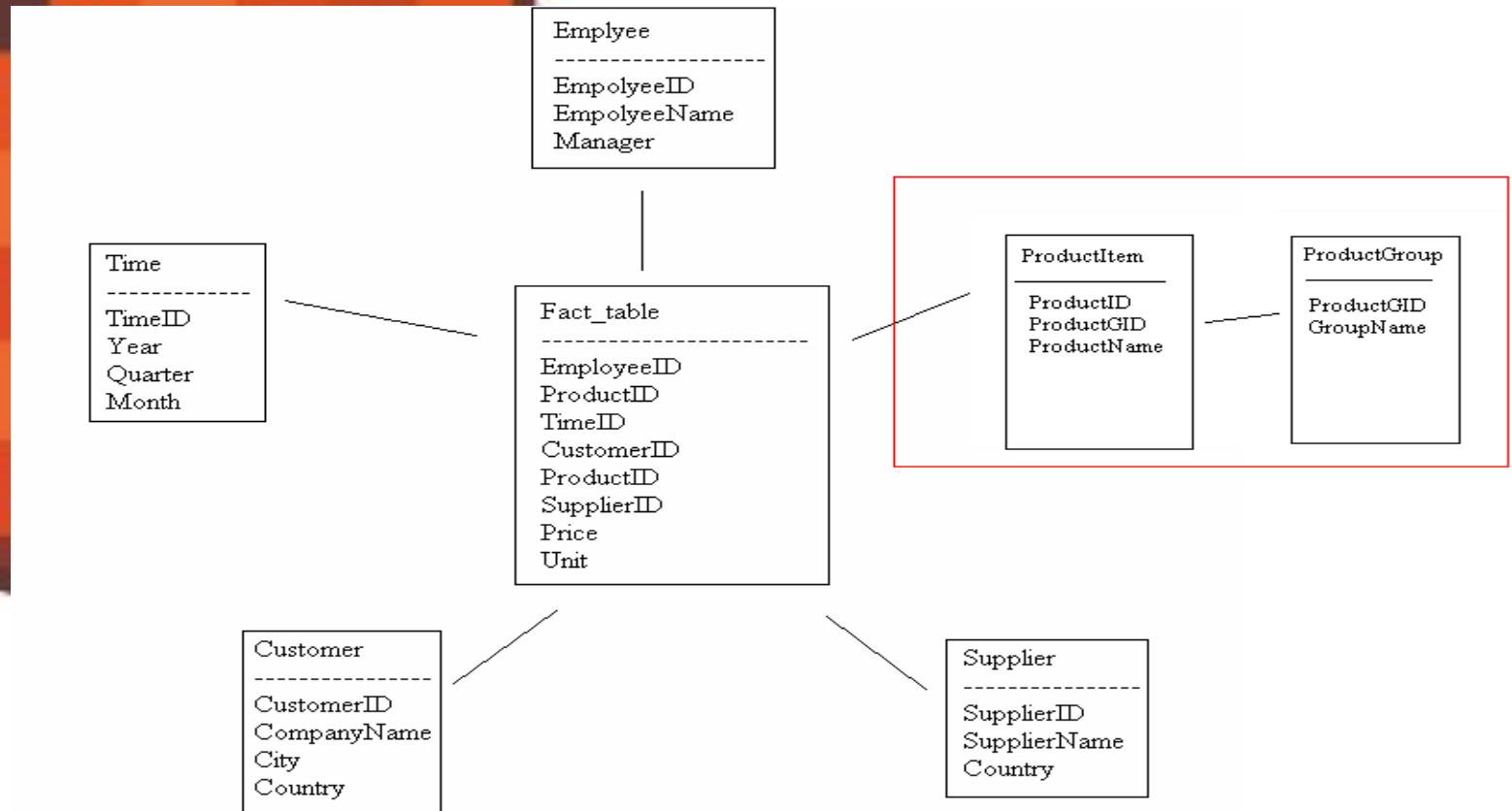
Zvezdasta (eng. star) shema

- Zvezdasta shema sadržava činjeničnu tablicu i jednu ili više dimenzijskih tablica
 - Činjenična tablica – centralna tablica koja sadržava numeričke vrijednosti kao što su prodaja u dolarima, cijene, ...
 - Dimenzijska tablica – okružuju činjeničnu tablicu i sadržavaju opisne podatke za činjeničnu tablicu
- Oblika je zvijezde

Zvijezdasta (eng. star) shema



Pahuljasta (eng. snowflake) shema



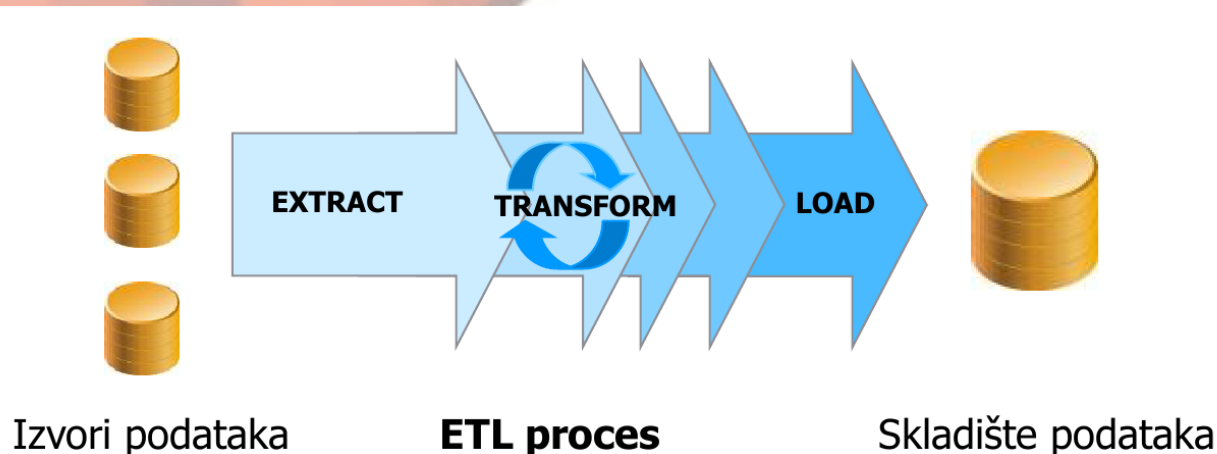


Kako podatke staviti u sustav

- Podaci postaju informacije ako i samo ako:
 - Imamo podatke
 - ZNAMO da ih imamo
 - Možemo im pristupiti
 - Možemo im vjerovati
- Podatke prebacujemo procesom koji se zove ETL (Extract – Transform – Load)
 - To je najveći dio posla na projektu

Kako podatke staviti u sustav

- ETL je proces ekstrakcije podataka iz izvorišnih sustava, transformacije i integracije podataka te pohranjivanja pročišćenih podataka u skladište





Osnovni zahtjevi za ETL

- Prilikom izgradnje ETL procesa susrećemo se sa nekim uobičajenim zahtjevima
 - Minimalna interakcija sa izvorima i skladištem podataka
 - Maksimalno razdvajanje unutrašnjih procesa (nezavisna, asinkrona i parcijalna obrada)
 - Praćenje grešaka i performansi



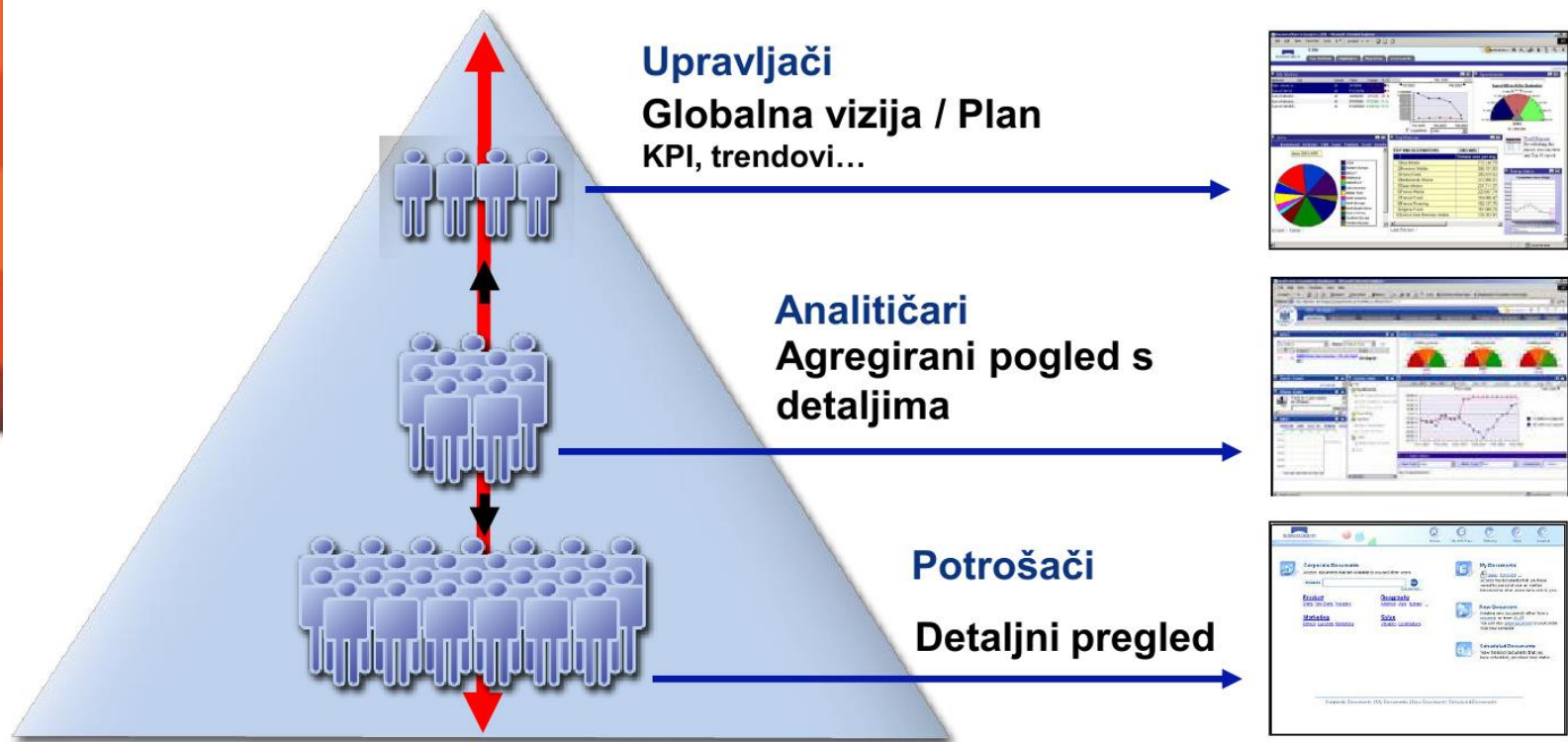
ETL alati

- Profesionalni alati specijaliziranih proizvođača
 - Informatica, SAP Data Integrator, ...
- Alati integrirani u baze podataka
 - Microsoft SSIS, IBM, Oracle
- Postoje i ETL alati otvorenog koda
 - Pentaho, JasperSoft
- Moguće je ETL ostvariti i pisanjem SQL-a
 - To je ponekad najbolje rješenje

Kako to vide krajnji korisnici

BI povezuje sve korisnike

Potrebe za informacijama su različite na svakom nivou unutar tvrtke





Što od svega mogu učiniti krajnji korisnici

- Krajnji korisnici mogu u potpunosti prilagoditi sustav
- No, realne mogućnosti su ograničene
 - Pravima pristupa u sustav
 - Poznavanjem podataka, alata i metoda razvoja
- Također je i stvar motivacije jer pokretači su uglavnom
 - Informatika
 - Određeni poslovni odjel
 - Organizacija / proces (najbolji slučaj)

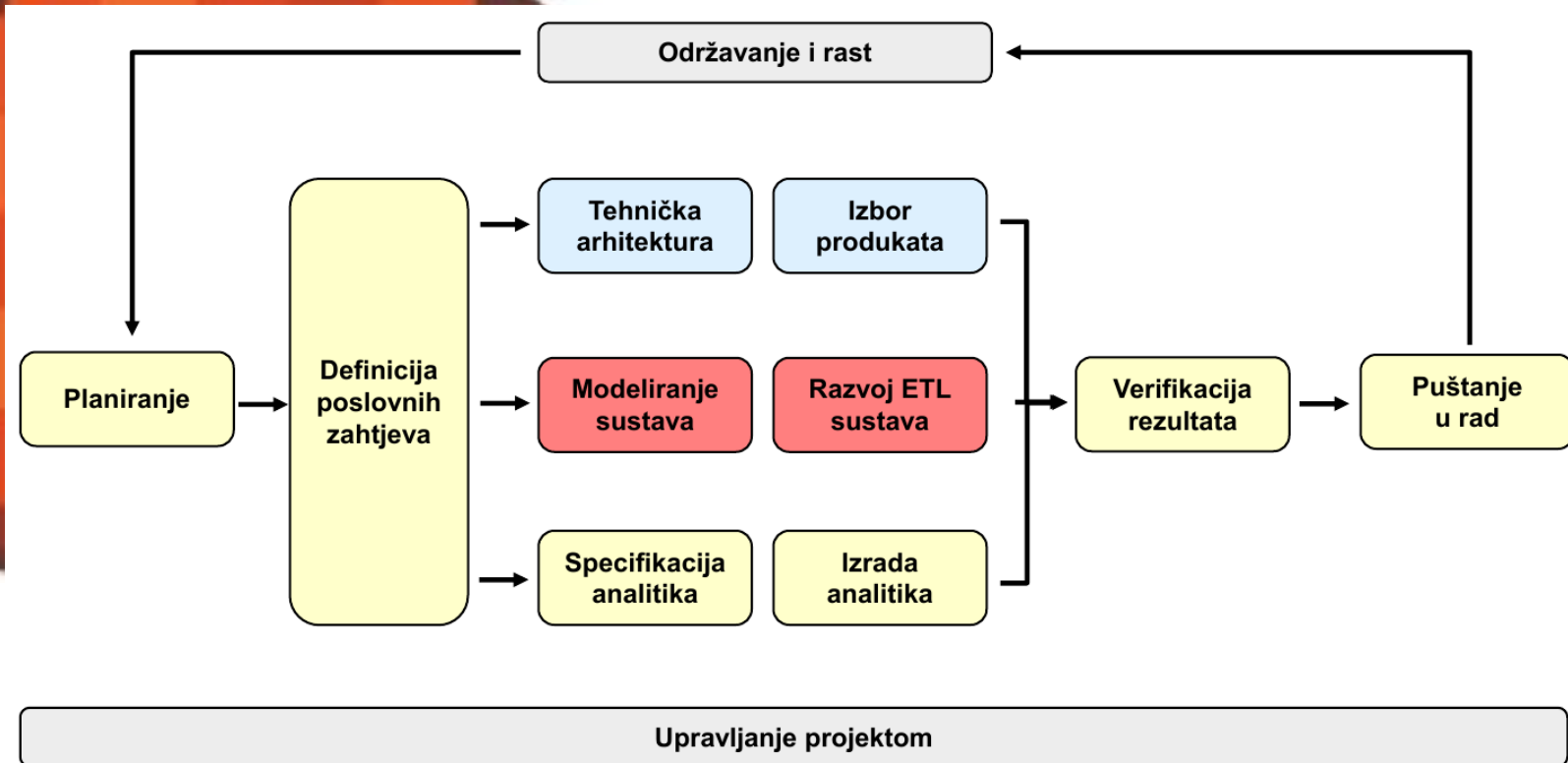


Preduvjeti uspjeha

- Sponzorska podrška sa što višeg nivoa u organizaciji
- Aktivno sudjelovanje poslovnih korisnika
 - Suradnja i raspoloživost tokom svih faza projekta
- Razumno određivanje ciljeva
 - Kratke iteracije sa jasno određenim ciljem
 - Prva iteracija je krajnje jednostavna
- Upravljanje očekivanjima

- **Spiralni pristup**
 - Pronaći mogućnost za ispunjavanje jaza između poslovnih potreba i postojećih informacija
 - Izabrati politički održivo rješenje
 - Razviti i implementirati rješenje
 - Objaviti uspjeh što glasnije (iznimno važno!)
 - Ponoviti sve od početka
- Osnovni pokretač ciklusa je POSLOVNA POTREBA

Osnovne faze





Prije početka

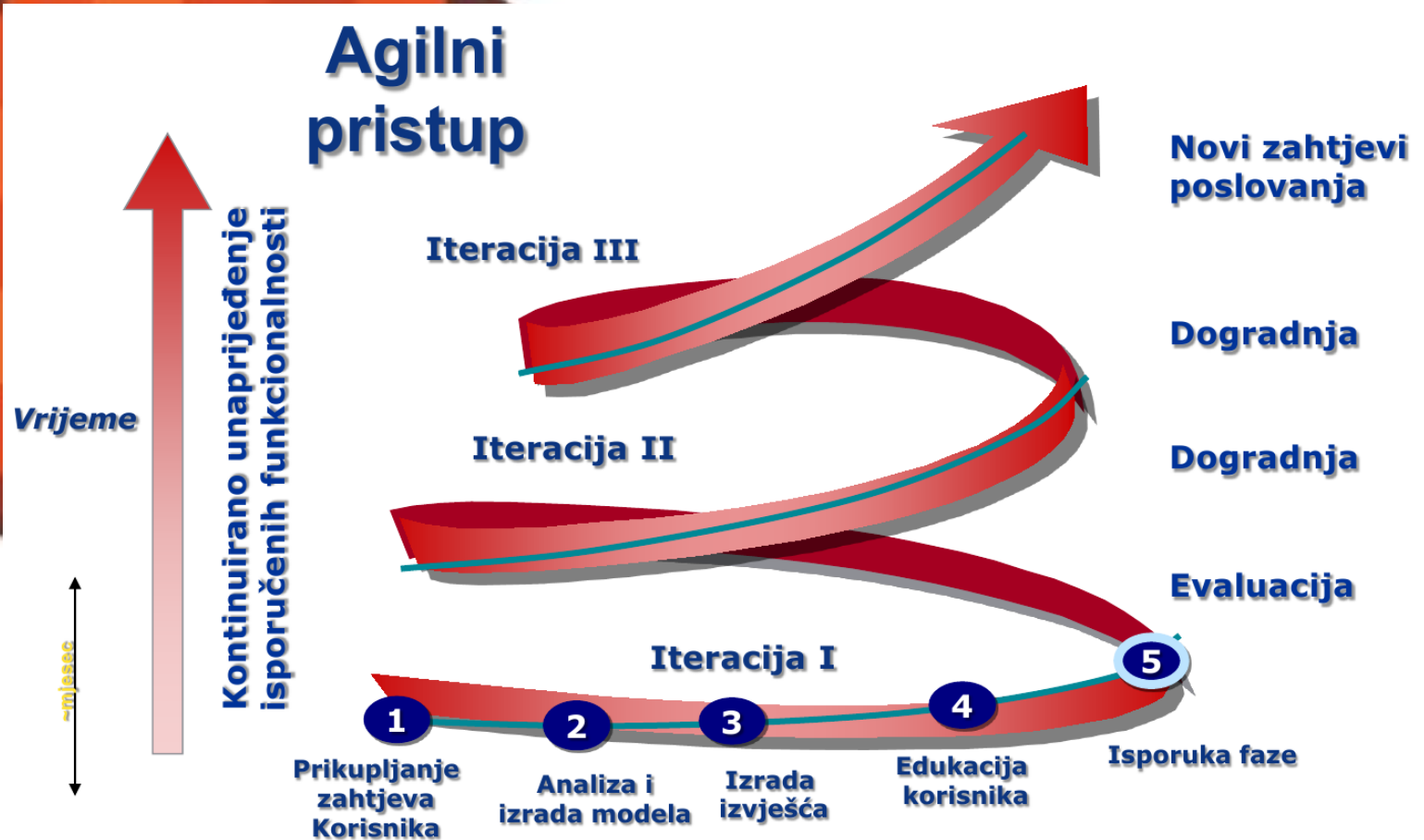
- Prije početka projekta potrebno je odrediti
 - Što prvo napraviti (izgraditi)
 - Tko će uspostaviti standarde
 - Tko će uspostaviti semantičku definiciju pojmova
 - Tko će postaviti poslovna pravila
 - Kako će sustav na kraju izgledati



Plan implementacije

- Gruba snimka stanja i izrada početnog modela
- Određivanje prioriteta i reda implementacije
- Implementacija izabrane teme
 - Detaljna specifikacija modela
 - Izrada ETL modela
 - Provjera i potvrda rezultata
 - Izrada izvještaja i/ili modela za rudarenje podataka
 - Školovanje korisnika
 - Prihvaćanje sustava od strane korisnika (!)
- Implementacija slijedeće teme

Višegodišnja podrška sustavu





Podaci, informacije i znanje

- Podaci – skup neobrađenih vrijednosti ili činjenica koje se koristi u daljnjim izračunima, odlučivanjima ili mjerenjima
- Informacija – rezultat skupljanja i organiziranja podataka u obliku ovisnosti između podatkovnih jedinica pružajući im na taj način kontekst i značenje
- Znanje – koncept razumijevanja informacije temeljen na prepoznatim uzorcima u cilju uvida u informaciju



Proces poslovne inteligencije

- Podaci → Informacija → Znanje → Plan djelovanja
 - Podaci → Informacija – proces određivanja koji podaci će se prikupljati i u kojem kontekstu
 - Informacija → Znanje – Proces koji uključuje analitičke komponente, kao što su skladište podataka, OLAP, kvaliteta podataka, analiza podataka, primjena poslovnih pravila i rudarenje podataka
 - Znanje → Plan djelovanja – najvažniji dio procesa poslovne inteligencije



Plan djelovanja

- Informacija zadržava svoju vrijednost ako i samo ako je znanje koje je iz nje izvučeno upotrebljivo, to jest na temelju njega možemo poduzeti neku akciju
 - Potrebna nam je metoda izdvajanja vrijednosti iz znanja
 - Ovo nije tehničko već organizacijsko pitanje – potrebni su ljudi sa znanjem i mogućnostima u organizaciji koji mogu nešto poduzeti
 - Ovo je ujedno i pitanje povrata uloženog (ROI – Return Of Investment)



Razvoj poslovne inteligencije

- Prva generacija – tradicionalna analitika (upiti i izvještavanje)
- Druga generacija – tradicionalni BI (OLAP, skladišta podataka)
- Treća generacija – Nova generacija sustava za rad sa podacima
 - Dinamički izvještaji, KPI indikatori, informacijske ploče, ...
 - Korisnici sami dohvaćaju potrebne informacije
- Četvrta generacija – Napredna analitika
 - Implementacija dinamičkih pravila
 - Predviđajuća analitika
 - Rudarenje podataka u stvarnom vremenu



SQL Server – trenutno stanje

Pregled verzija od 2000 do 2008



SQL Server 2000...doviđenja

- Potpuna podrška za SQL Server 2000 prestaje 8.4.2008
 - Posljednji servisni paket je SP4
- Proširena podrška traje do travnja 2013
 - Plaćanje po incidentu
 - Rješavanje sigurnosnih problema je besplatno
- SQL Server 2000 ne radi na Windows Visti/7/2008




SQL Server 2005...još će dugo biti s nama

- Potpuna podrška za SQL Server 2005 do nedefiniranog (pretpostavljamo dalekog) trenutka
- Trenutno vrlo aktivna migracija sa SQL-a 2000 na 2005
- Većina današnjih produkcijskih sustava u Microsoft okolini na kombinaciji Windows Server 2003 + SQL Server 2005



SQL Server 2008 R2

- Neke stvari su već rečene...ja ću samo ukratko napomenuti stvari vezane uz poslovnu inteligenciju
- Novosti:
 - Integraciju sa Office-om 2007/2010
 - Brži prikaz podataka kod velikih skupova podataka
 - Kompresija na nivou retka i stranice
 - Integration Services (SSIS) može skalirati na više od dva procesora i poboljšava podršku za rad sa velikim skupovima podataka
 - Analysis Services (SSAS) radi bolje na velikim skupovima podataka
 - Report Services (SSRS) se spaja sa Office 2007/2010 – izvoz direktno u Word, korištenje Worda i Excela kao predložaka za izvještaje u Report Designer-u



SQL Server 2008 R2 BI

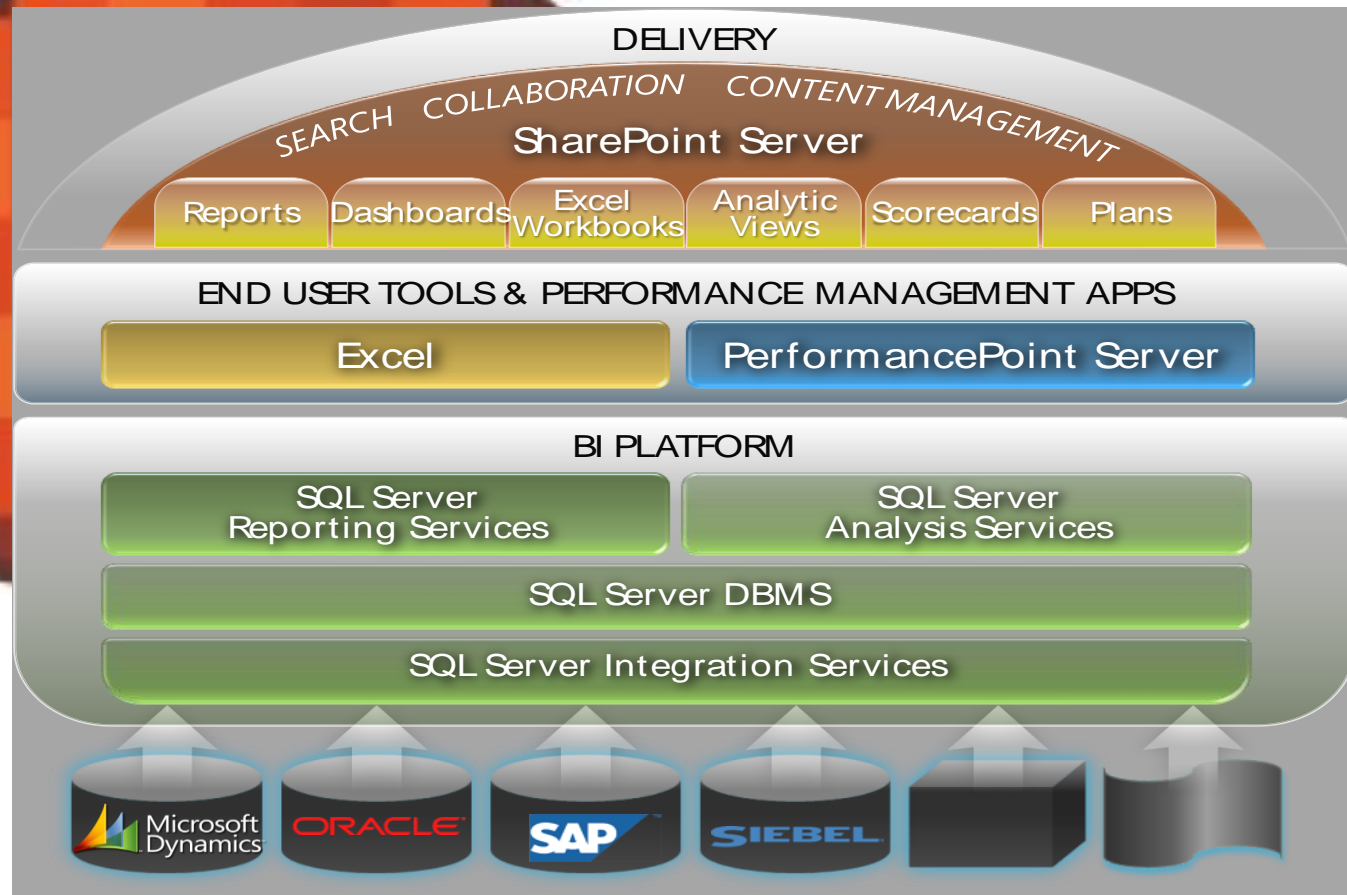
Poslovna inteligencija na SQL Serveru 2008 R2



SQL Server kao baza podataka

- SQL Server je klijent – server relacijska baza podataka
 - To je čini ravnopravnom u borbi za tržište sa Oracle-om ili IBM-om
- Ali problem je tradicije – dok su Oracle i DB2 desecima godina neupitno enterprise produkti, SQL Server u očima mnogih ljudi to još mora zaslužiti
 - Nažalost, često su to ljudi koji odlučuju
 - Tehnički gledano, u zadnjih 5 godina SQL Server je potpuno konkurentna platforma ... i više od toga

SQL Server 2008 R2 BI arhitektura



Microsoftova BI strategija



- Poznato okruženje
- Integracija sa Office-om
- Bazirano na SQL Serveru

Poslovna inteligencija po Microsoftu

Business User Experience



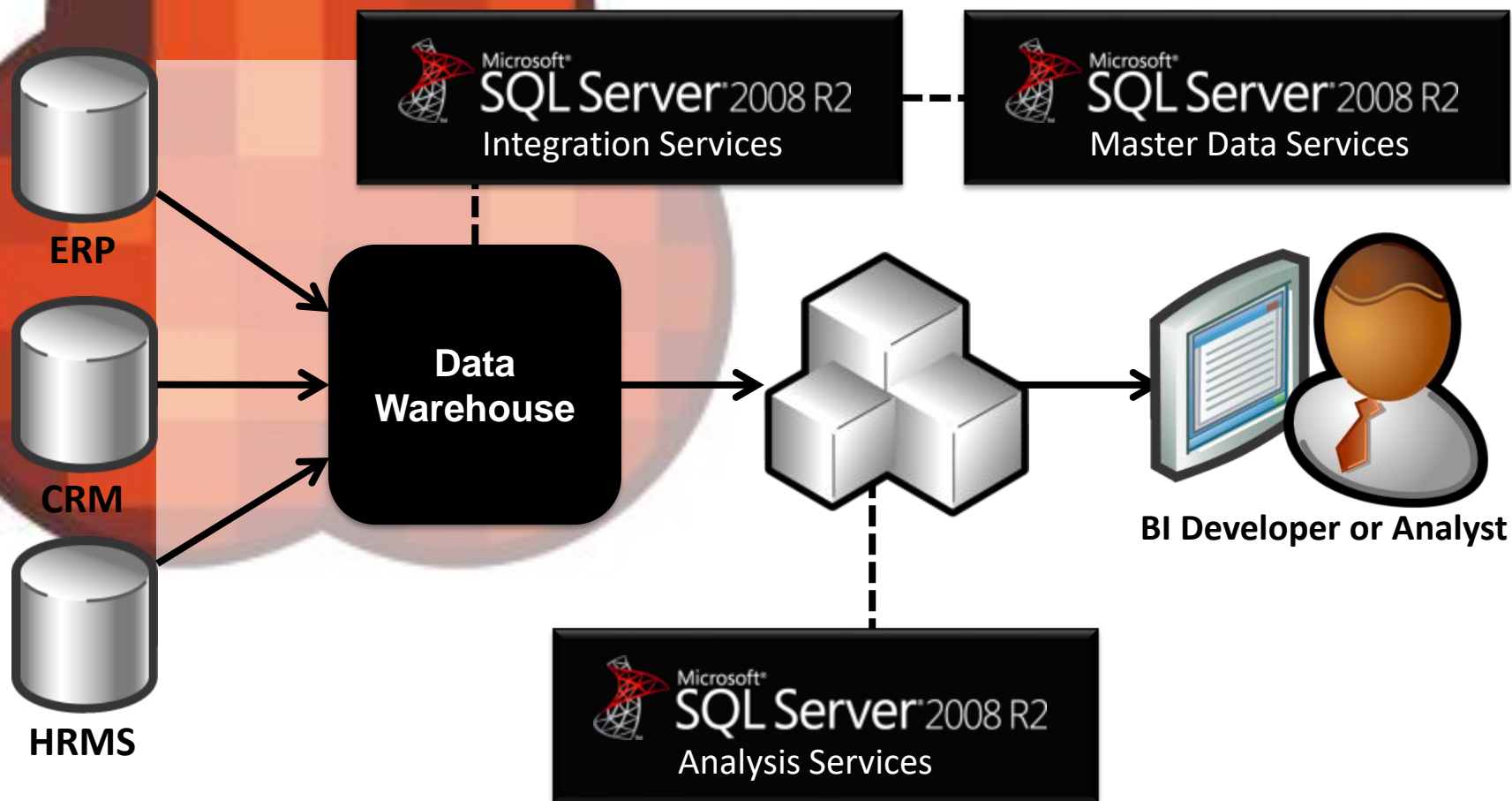
Business Collaboration Platform



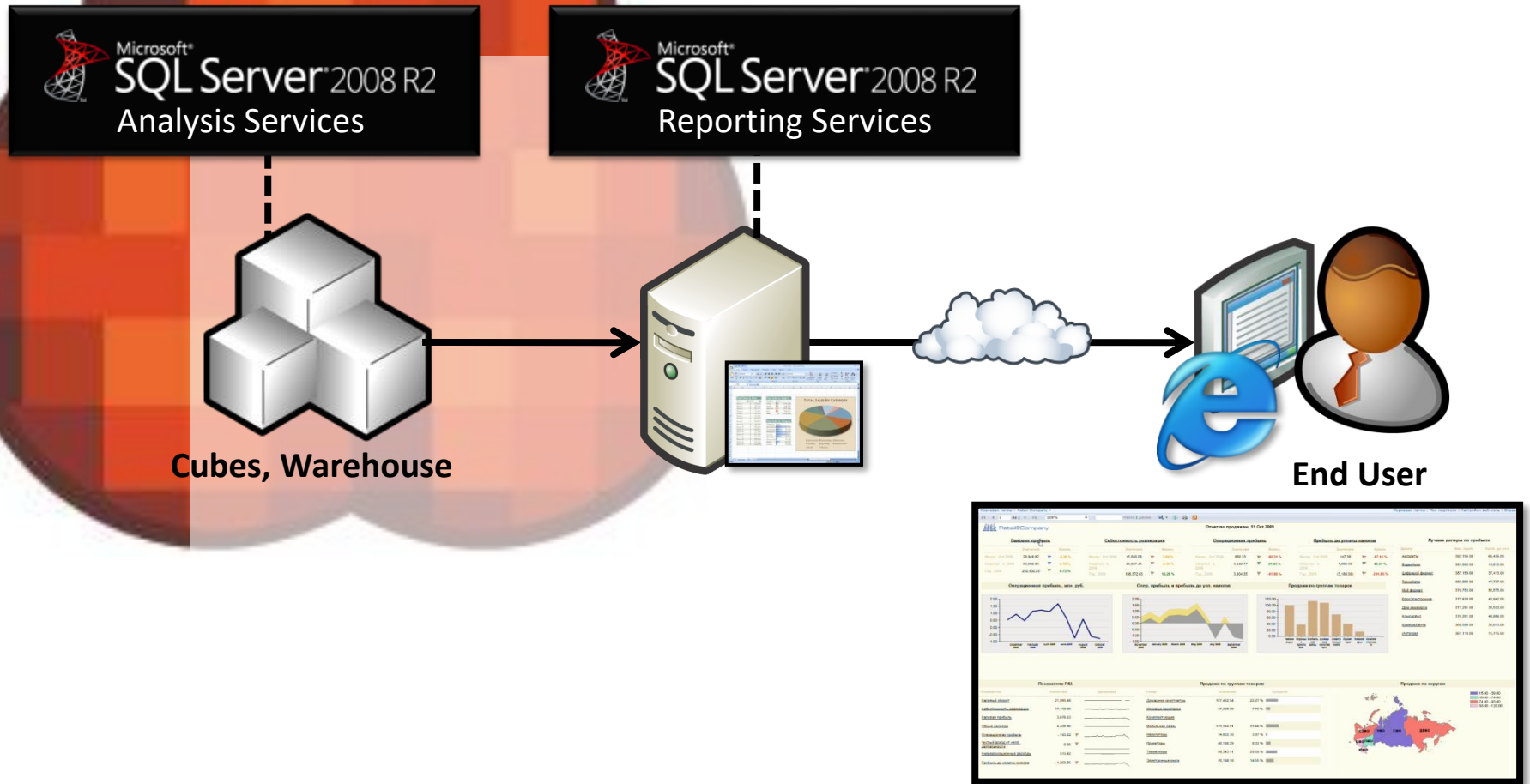
Data Infrastructure & BI Platform



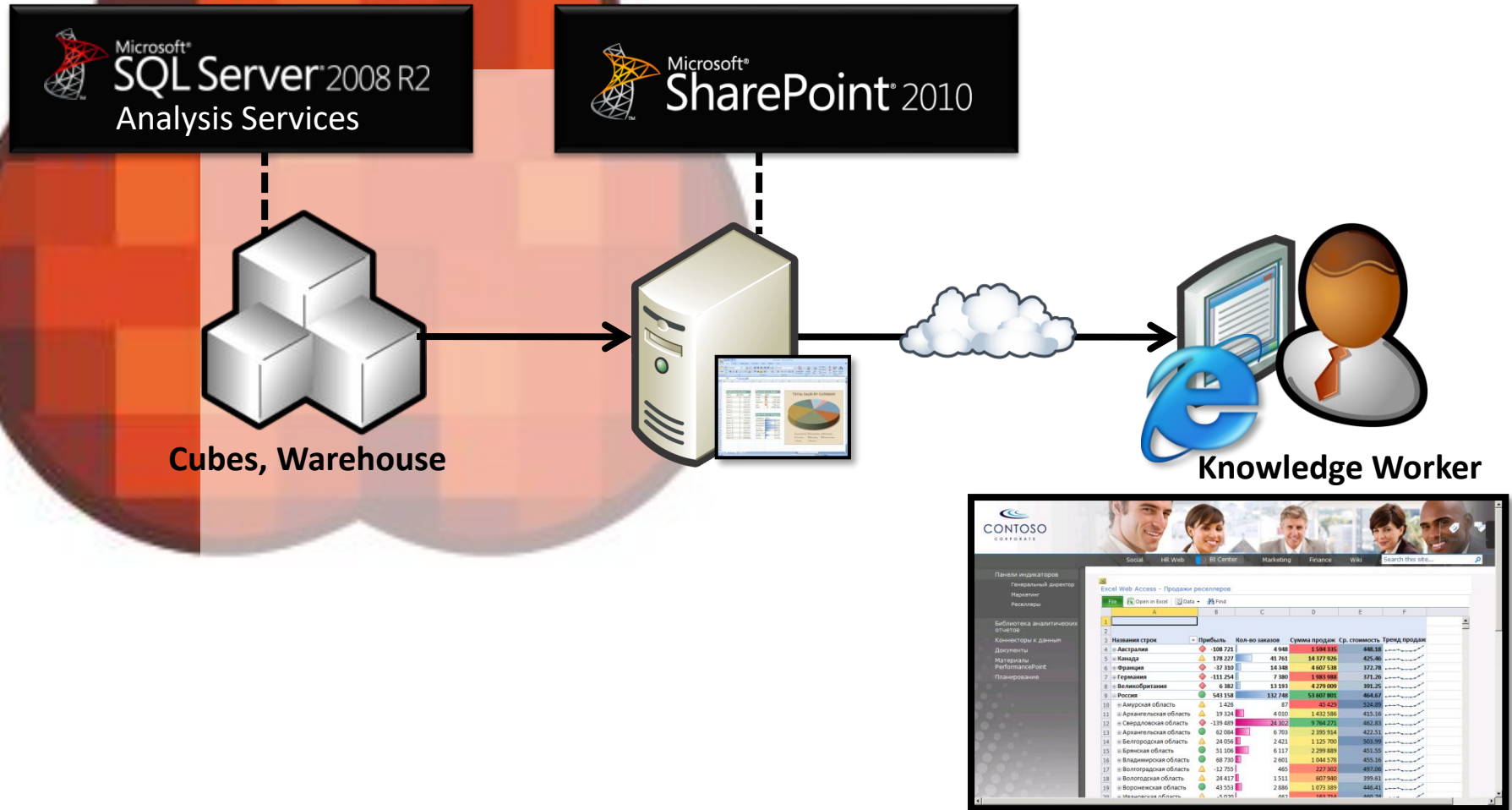
Šira slika – upravljanje informacijama



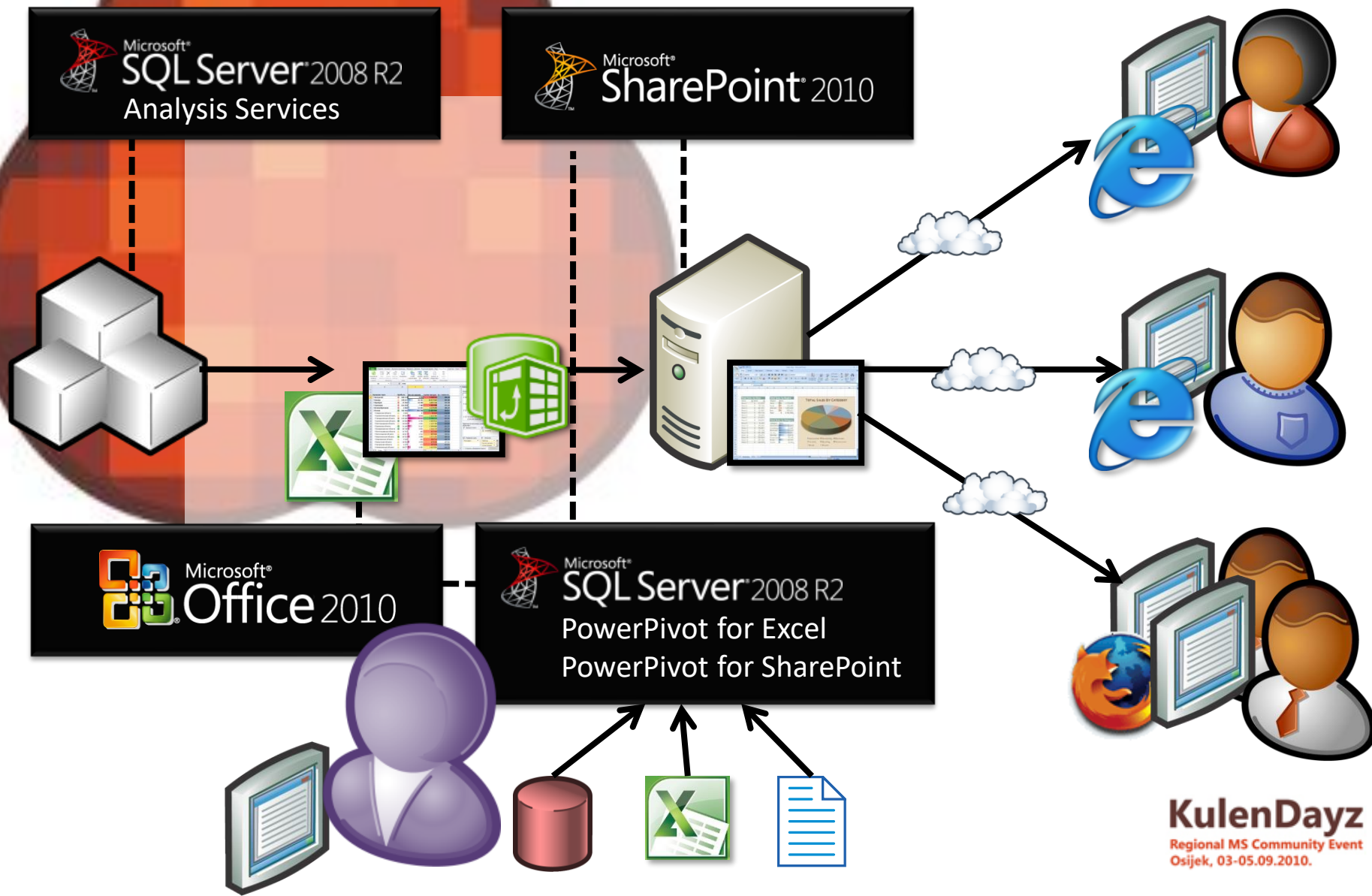
Šira slika - izvještavanje



Šira slika – dashboards & scorecards



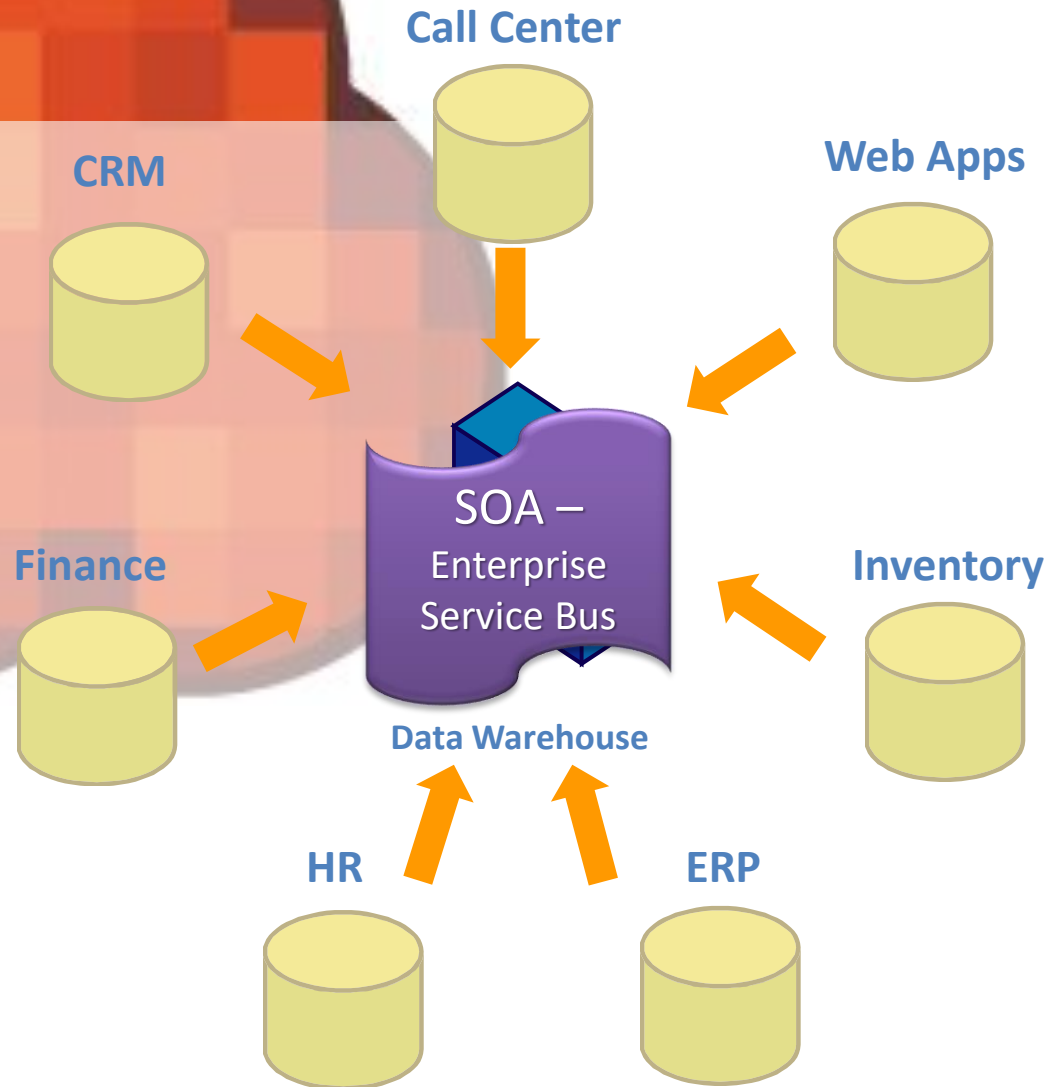
Šira slika – poslovni analitičari i napredni korisnici



Poslovni podaci



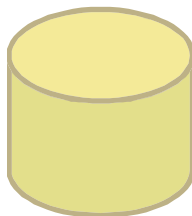
Problem integracija podataka



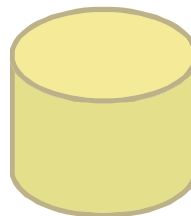
Izvorišni sustavi

- Obraduje transakcije u stvarnom vremenu
- Optimizirana za promjene podataka
 - Normalizirana
- Ograničena sposobnost odlučivanja
- Najčešći nazivi:
 - Online transaction processing (OLTP) sustavi
 - Operacijski sustavi

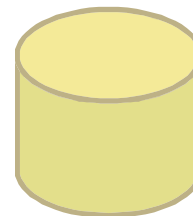
HR



Finance



Inventory



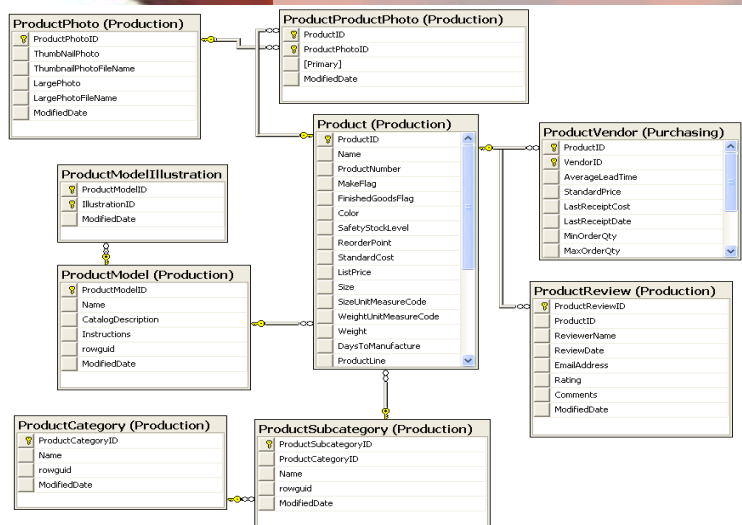


Skladište podataka

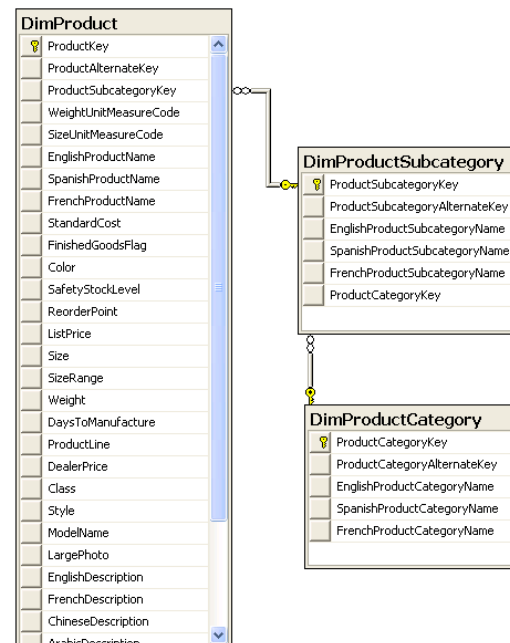
- Pruža podatke za poslovnu analizu
 - Grupirano u sadržajno specifična područja koja se zovu data martovi
- Optimizirano za brzi ad-hoc dohvat informacija
- Integrira heterogene izvorišne sustave
- Konzistentan izvor povijesnih podataka

ETL: Extract, Transform and Load

- Izdvajanje podataka iz izvorišnog sustava
- Transformiranje podataka u izlazni oblik
- Učitavanje podataka u skladište podataka



ETL





Dimenzije i činjenice – osnova BI-a

- Činjenica – nešto što se dogodilo
 - Prodaja, kupovina, slanje, ...
 - Transakcija ili događaj
 - „Akcija”
 - U osnovi to je mjera
- Dimenzija – opis činjenice
 - Kupac, proizvod, račun, ...
 - Objekt
 - „Subjekt”
- Činjenica (mjera) izražava se preko dimenzija
 - 42 „proizvoda” prodana su Marku 01.10.2009

Dimenzije

- Opis poslovnih entiteta
- Sadrži attribute koji daju kontekst numeričkim podacima
- Prikaz podataka organiziran u hijerarhije

DimDate	
🔑	DateKey
	FullDateAlternateKey
	DayNumberOfWeek
	EnglishDayNameOfWeek
	SpanishDayNameOfWeek
	FrenchDayNameOfWeek
	DayNumberOfMonth
	DayNumberOfYear
	WeekNumberOfYear
	EnglishMonthName
	SpanishMonthName
	FrenchMonthName

DimProduct	
🔑	ProductKey
	ProductAlternateKey
	ProductSubcategoryKey
	EnglishProductName
	StandardCost
	Color
	Class
	Style
	ModelName
	SafetyStockLevel
	...

DimEmployee	
🔑	EmployeeKey
	ParentEmployeeKey
	EmployeeNationalIDAlternateKey
	FirstName
	LastName
	Title
	EmailAddress
	Gender
	...



Slowly changing dimensions (SCD)

- Vrijednosti nekih atributa mogu se mijenjati sa vremenom
- Postoje dva osnovna načina praćenja promjena:
 - Promjene tipa 1
 - Promjene tipa 2

SCD – Tip 1

- Jednostavno prepisujemo staru vrijednost novom

Employees				
Key	EmployeeID	EmployeeName	Position	Manager
109	566778999	Steve Smith	Programmer	Mary Jordan
110	344981000	David Wu	Programmer	Mary Jordan
Employess				
Key	EmployeeID	EmployeeName	Position	Manager
109	566778999	Steve Smith	Programmer	Mary Jordan
110	344981000	David Wu	Test Engineer	Peter Mundy

SCD – Tip 2

- Stvaramo novi zapis u dimenziji sa novom vrijednošću i surogatnim ključem te označavamo stari zapis kao neaktivan
 - Činjenična tablica koristi novi surogatni ključ za povezivanje na pojedine mjere

Employees						
Key	EmployeeID	EmployeeName	Position	Manager	StartDate	EndDate
109	566778999	Steve Smith	Programmer	Mary Jordan	8/19/1999	
110	344981000	David Wu	Programmer	Mary Jordan	7/3/1998	
Employess						
Key	EmployeeID	EmployeeName	Position	Manager	StartDate	EndDate
109	566778999	Steve Smith	Programmer	Mary Jordan	8/19/1999	
110	344981000	David Wu	Programmer	Mary Jordan	7/3/1998	4/8/2000
111	344981000	David Wu	Test Engineer	Peter Mundy	4/9/2000	

Prediktivna analiza

Role of software

Data mining

Predictive Analysis

Self-service Analysis

OLAP

Ad-hoc reporting

Canned reporting

Presentation

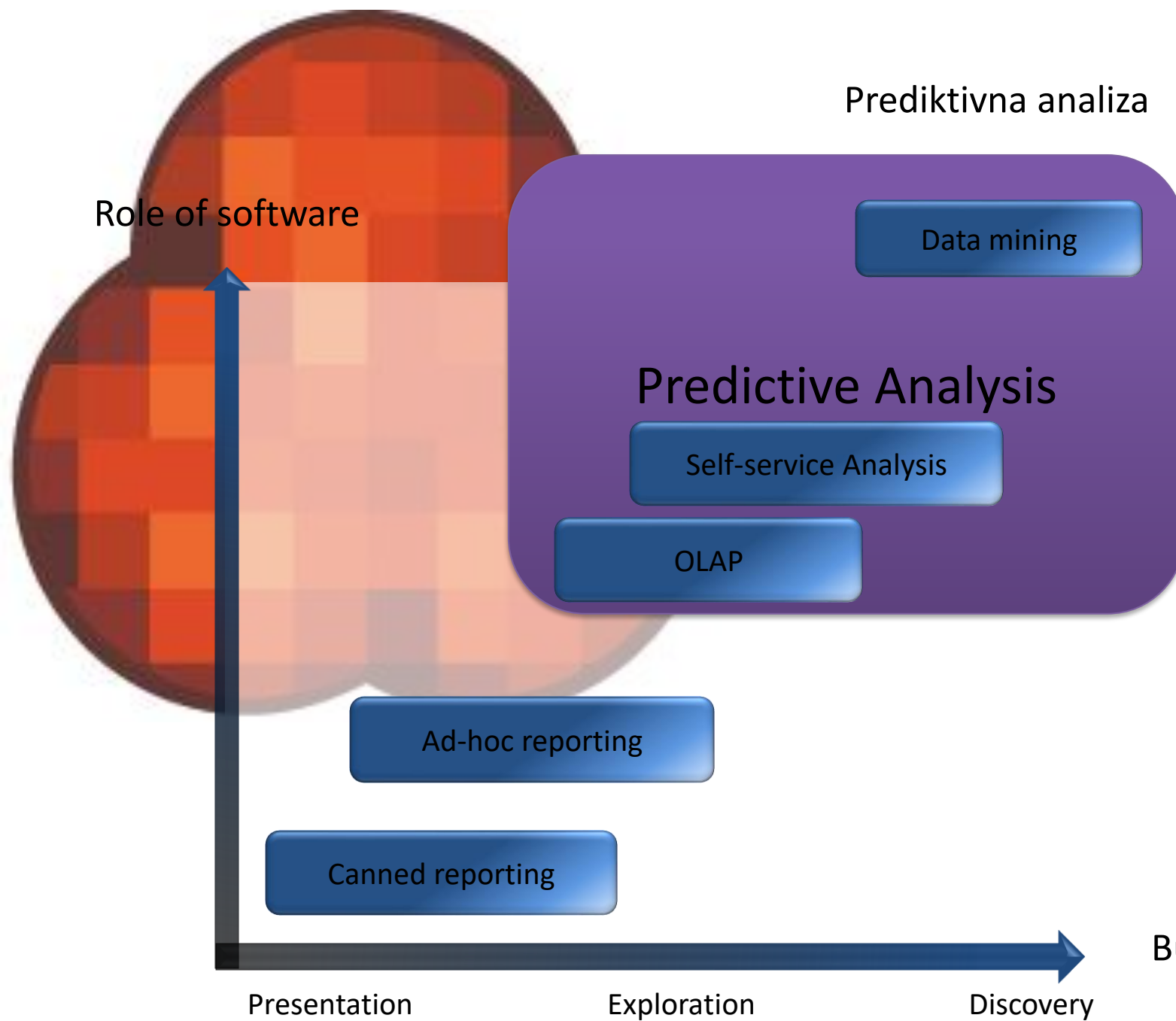
Exploration

Discovery

Business Insight

KulenDayz

Regional MS Community Event
Osijek, 03-05.09.2010.





OLAP ili višedimenzijski podaci

- Online Analytical Processing = Višedimenzijski podaci
- Mjere i dimenzije
- Koristi unaprijed izračunate agregirane vrijednosti za brz dohvat i transformaciju ulaznih podataka
- Podržava otkrivanje poslovnih trendova i statističkih podataka koji su teško vidljivi iz direktnih upita nad bazu

Kocka

Products

Dates

Measures

Cars

Parts

Accessories

2009

Q1

Jan

Feb

Mar

Items Sold

Cost

Sales \$

KulenDayz

Regional MS Community Event
Osijek, 03-05.09.2010.

Analiza kocke

Products

Dates

Measures

Cars
Parts
Accessories

2009

Q1

Jan

Feb

Mar

25

6

2

3

1

Items Sold

Cost

Sales \$

Analiza kocke

Product

Donut

Sandwich

Milk

Soda

Beer

Region

North

South

East

West

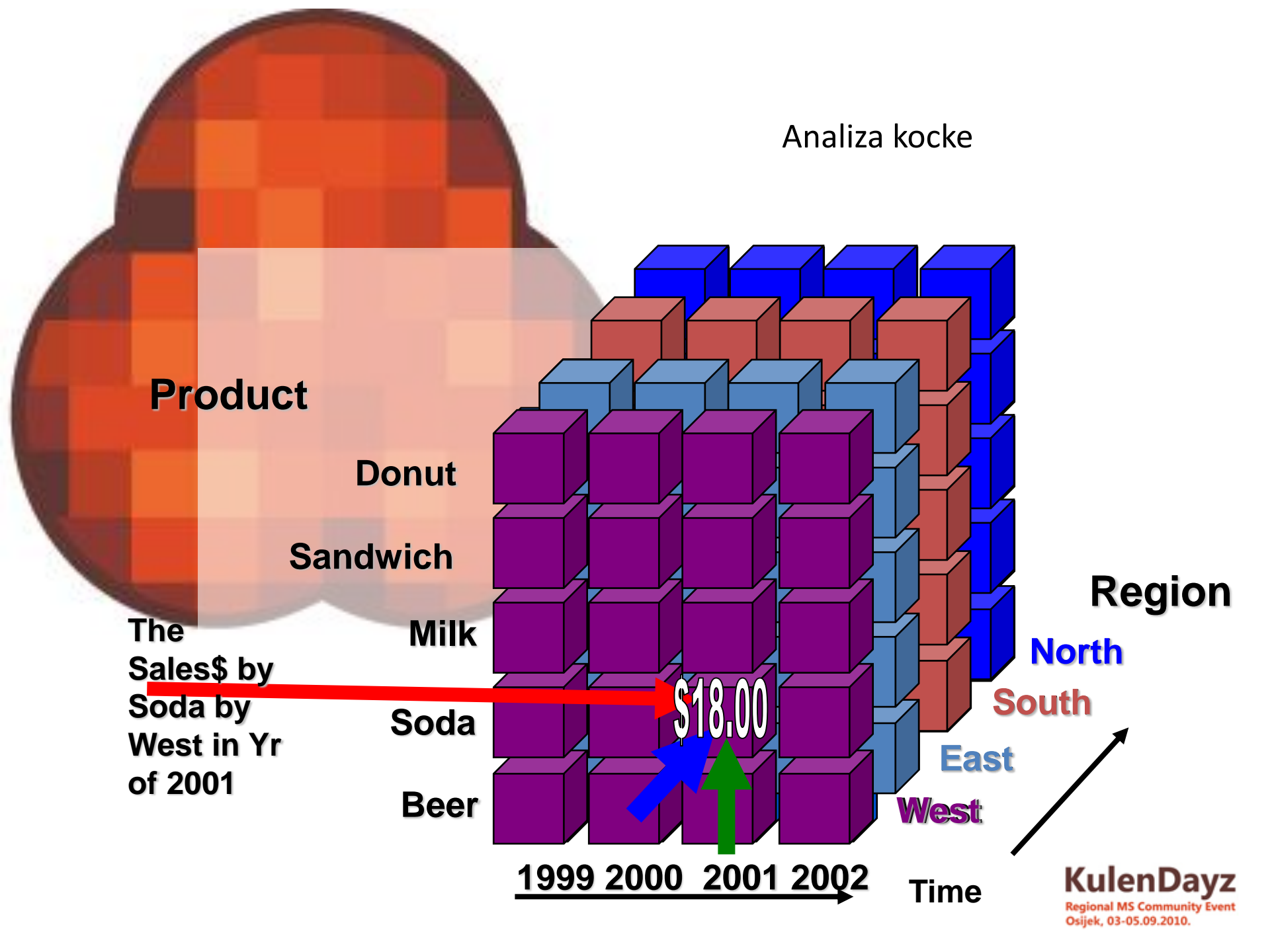
The
Sales\$ by
Soda by
West in Yr
of 2001

\$18.00

1999 2000 2001 2002

Time

KulenDayz
Regional MS Community Event
Osijek, 03-05.09.2010.



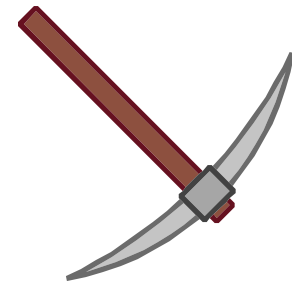


Alati za spajanje na kocke

- SQL Server Management Studio (SSMS)
- Business Intelligence Development Studio (BIDS)
- Query Analyzer (SSMS) – Za pisanje MDX-a
- Excel 2007/2010 – Koristi MDX

Rudarenje podataka

- Otkrivanje skrivenih uzoraka u velikim količinama podataka
- Kombinacija statistike, analize vjerojatnosti, tehnologija baza podataka, strojnog učenja i umjetne inteligencije





Key performance indicator (KPI)

- Mjere koje služe za usporedbu trenutnih performansi sa zadanim ciljevima
- Grupirane u poslovne scorecard-ove koje pokazuju „zdravlje” poduzeća
 - Idealno, sa izbalansiranom perspektivnom grupa KPI-eva
- Izgrađuju se
 - Korištenjem OLAP-a (KPI-evi na nivou poduzeća)
 - U Sharepoint PerformancePoint Services-ima
 - Korištenjem rudarenja podataka (prediktivni KPI-evi)

KPI karakteristike

	Actual	Goal and Status	Trend
<input type="checkbox"/> Sales Targets			
<input type="checkbox"/> Sales Target			
Pak, Jae	4172459	4000000 	↓
Reiter, Tsvi	2222128	2225000 	↓
Saraiva, José	1837614	1800000 	↓
Tsoflias, Lynn	701487	545000 	↑
Valdez-Smythe, Rachel	961128	1100000 	↓
Vargas, Garrett	1378290	1450000 	↓

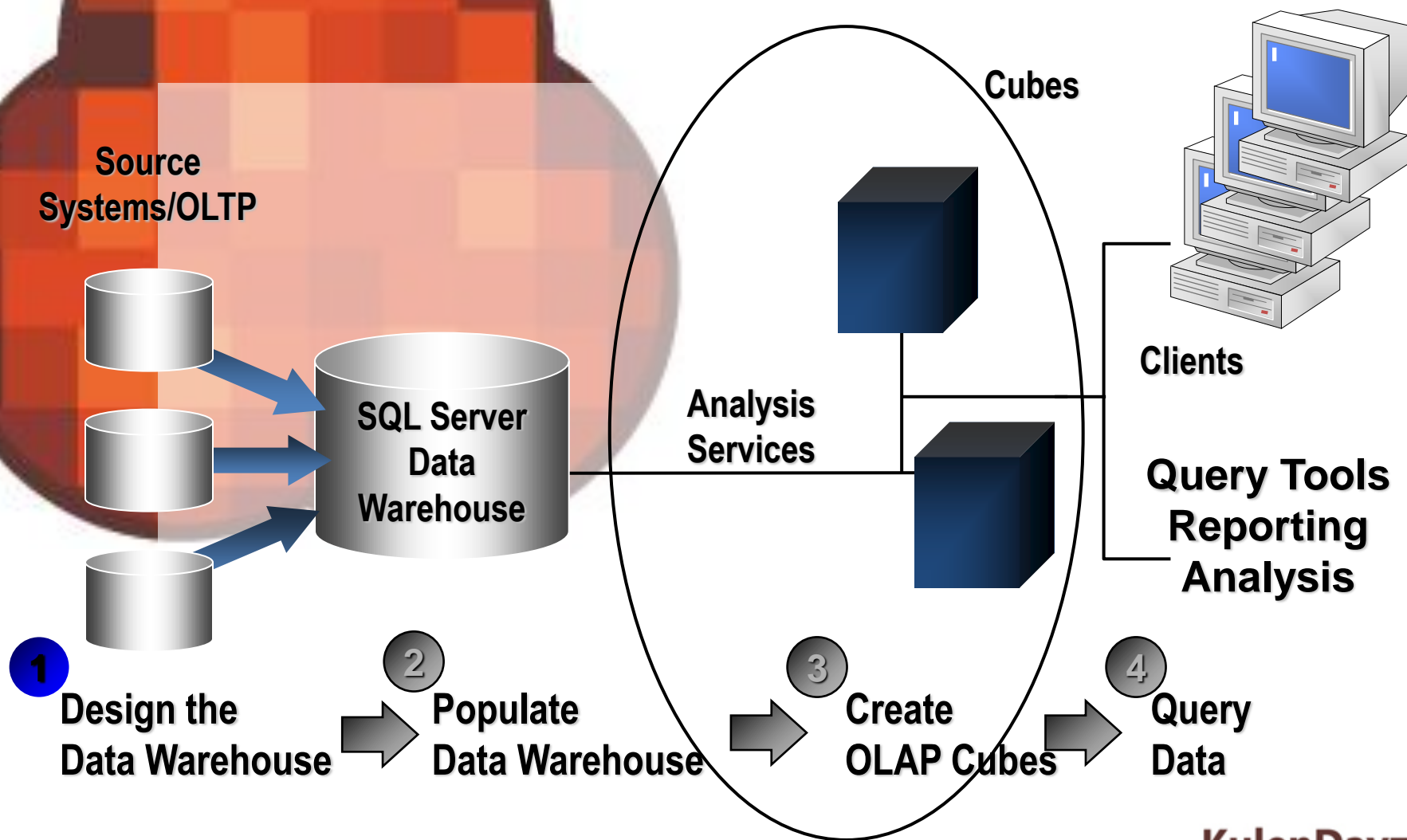
- Vrijednosti
- Ciljevi
- Statusi
- Trendovi

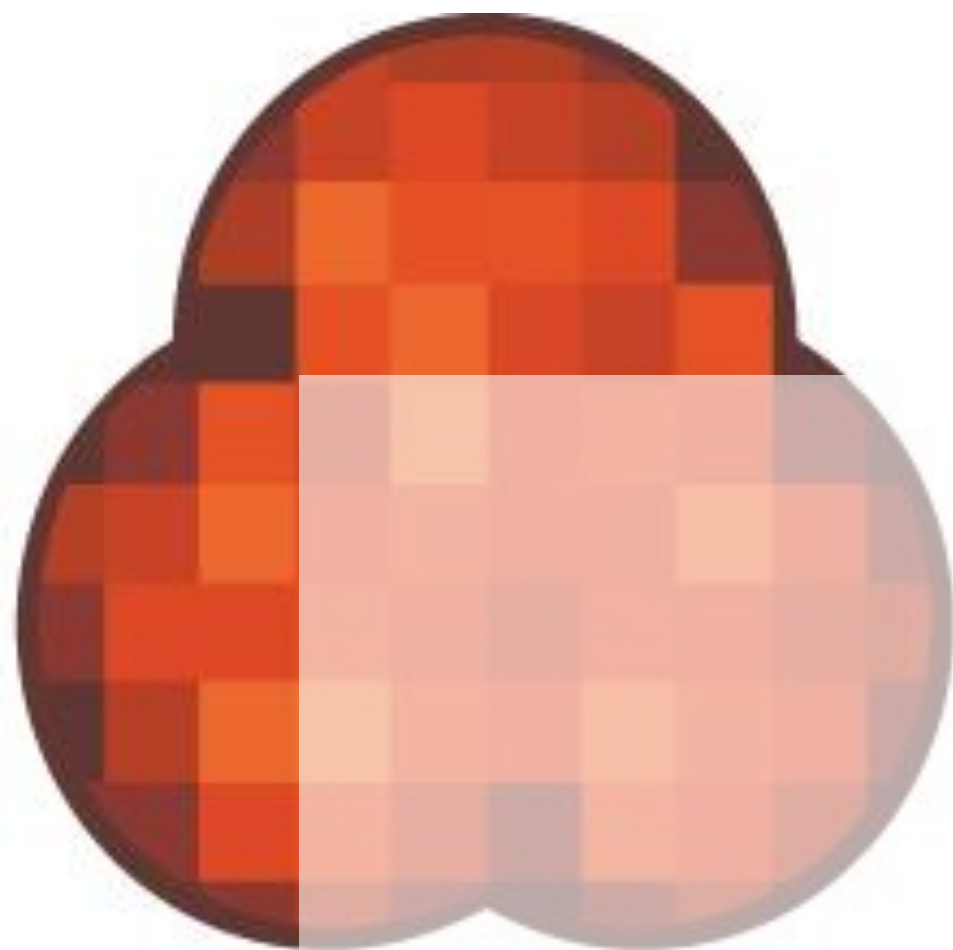


Dashboards & scorecards

- Scorecard – tablica (pivotirana) KPI-eva
- Dashboard
 - Sadrži scorecard-ove, analitičke izvještaje i ostale analitičke vizualizacije
- Stvaraju se
 - Korištenjem PowerPivot-a, Sharepoint 2010 PerformancePoint servisima
 - Razvoj u Sharepointu, Silverlightu i .NET-u

Kako to sve funkcionira zajedno





DEMO



Poslovna inteligencija “u praksi”

Problemi u projektima poslovne inteligencije

Kako bi projekt trebao izgledati



<http://blog.maia-intelligence.com/2008/06/09/bi-successful-project-management/>

<http://blog.maia-intelligence.com/2008/09/22/business-intelligence-implementation>



Problemi poslovne inteligencije

- Niska prihvaćenost kod krajnjih korisnika i velika ovisnost o IT-u
 - Problemi poslovnih analitičara
 - Teško dostupni organizacijski podaci
 - Izještavanje uglavnom ovisi o IT-u (što nije nužno loše ☺)
 - Složeno dijeljenje otkrivenih informacija (novih znanja)
 - Problemi IT osoblja
 - Nedostatak vremena (i motiva) za “ad-hoc” zahtjeve
 - Nedostak kontrole
 - Organizacijska poslovna inteligencija je često “preskupa”
 - <http://campustechnology.com/articles/2009/02/01/project-management.aspx>



Problemi poslovne inteligencije

- Tehnički i socijalni problemi
 - Korisnički alati ili presloženi ili prejednostavni
 - Više ponuđača programske podrške na istom projektu – problem međusobne kompatibilnosti i “politike”
 - Otkrivaju se “nepoželjni” podaci i uzorci
 - Projekti su često “politički” motivirani
 - Različite vizije uspjeha projekta klijenta i konzultantskog poduzeća
 - Nepostojeća poslovna podrška kod klijenta
 - Često neskloni krajnji korisnici jer se “boje” rezultata projekta
 - Nepostojanje jasne vizije što se projektom želi postići
 - ...



Uz ovoliko problema...zašto sve ovo

- Dobici poslovne inteligencije na rad nekog poduzeća su nemjerljivi - od izvještavanja do otkrivanja novog znanja
- U nekim slučajevima su nužni za obavljanje ključnih poslovnih djelatnosti – marketing, analiza rizika, ...
- Zanimljivi linkovi
 - <http://www.information-management.com/news/2707-1.html>
 - <http://searchbusinessanalytics.techtarget.com/news/1507200/Business-intelligence-project-management-Five-must-have-pieces-of-advice>



Kontakti i literatura

- josipsaban@gmail.com
- <http://www.josipsaban.net>
 - Svi materijali sa ovog i ostalih predavanja su slobodno dostupni za preuzimanje
- Preporučena literatura
 - McGraw-Hill: “Delivering Business Intelligence with Microsoft SQL Server 2008”
 - Wiley&Sons: “Data mining with SQL Server 2008”

