Zagrebačka banka
UniCredit Group

UEFA CHAMPIONS LEAGUE
UniCredit
official sponsor

EVOLUCIJA ARHITEKTURE BI SUSTAVA (Inženjerska perspektiva)

Osobno putovanje, tj. transformacija jednog DBA

Daniel Beden, Vodeći DBA specijalist

Rovinj, 18. listopada 2013.

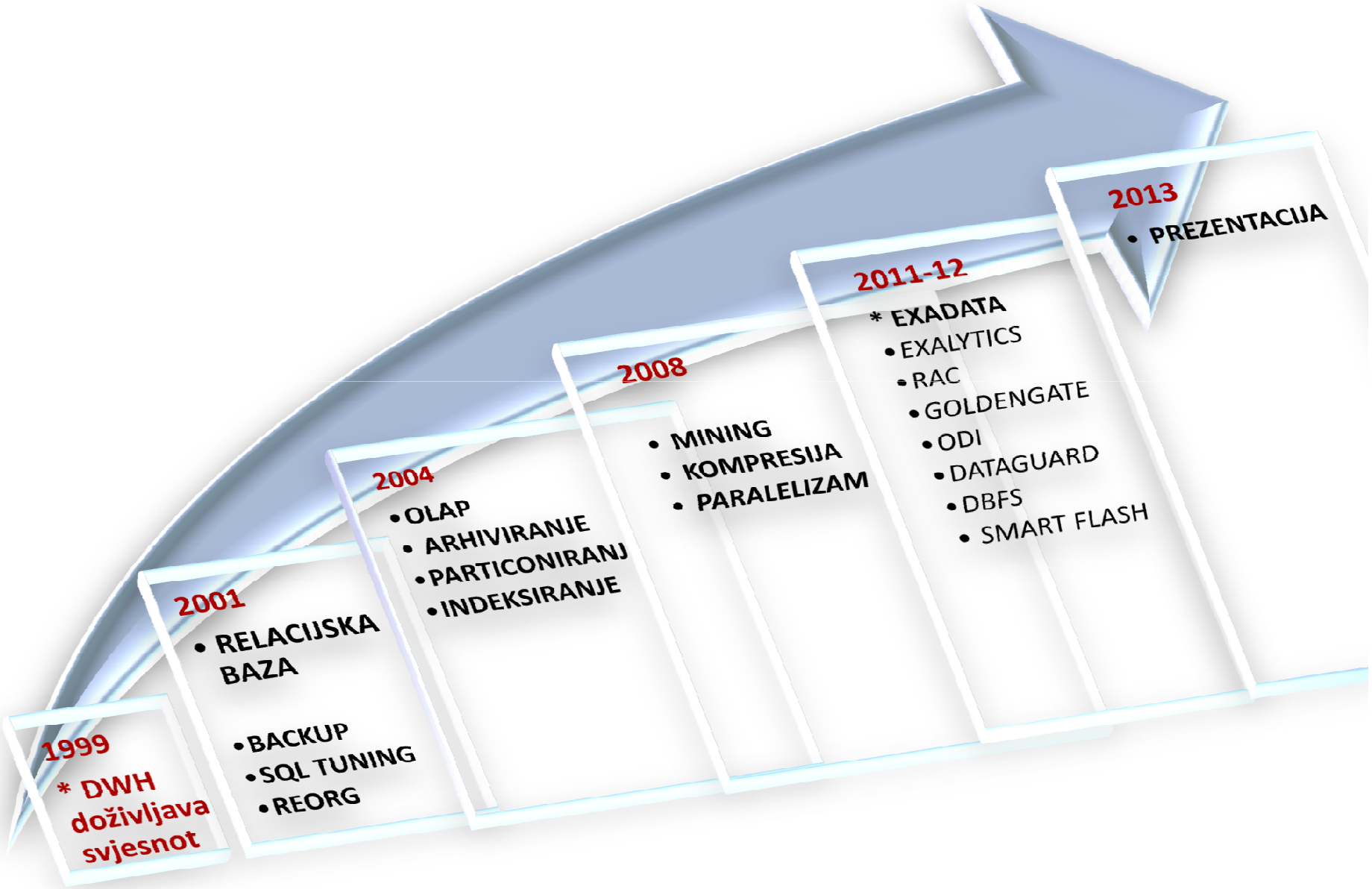
Sadržaj

- Tradicionalni DBA
- Promjena paradigme
- Šok i nevjerica
- Pionirski dani
- Planovi za budućnost ili što već voda nosi

Zagrebačka banka
UniCredit Group

Partneri projekta

DBA U BI SVIJETU



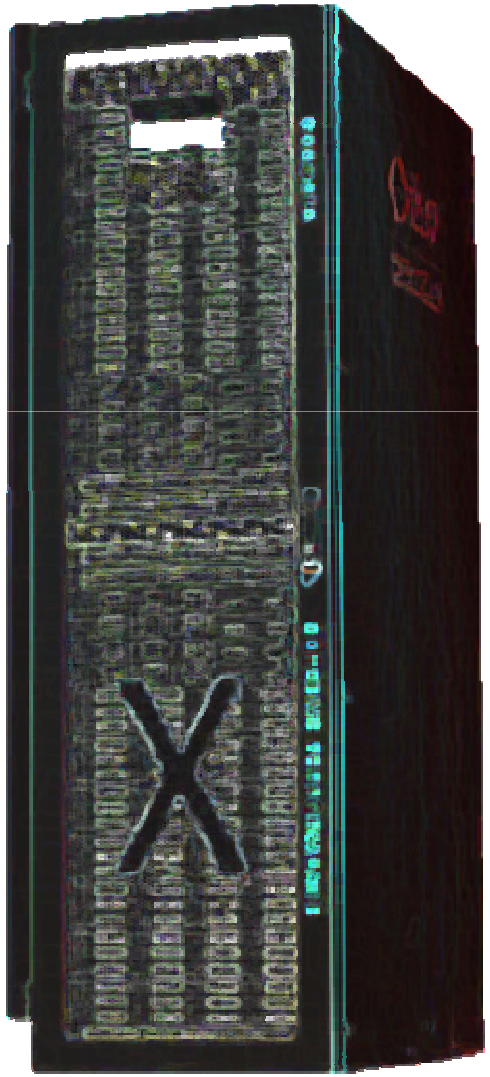
DBA (kako smo živjeli nekad)

- Režim rada
  - 23,6/7 fokus na operativni dio i fizičke operacije s podacima
  - 0,4/7 optimizacija DW jobova
  - 0/365 razvoj
- Izazovi
  - Sustav neoptimiziran za veliki I/O
  - Malo vremena za razvoj i poboljšanje modela
  - Nervozni programeri
  - Performanse
  - Diverzificirana DW okolina
  - Neispavanost, loša probava.

Promjena paradigme : pokretači

- Količina podataka je eksponencijalno rasla (x3)
- Novi zatjevi (regulatora, managamenta, analitičara)
- Diverzifikacija DW na manja skladišta, svaki sa svojom specifičnom svrhom
- Izvještavanje u realnom vremenu, nad živim podacima

I tada jednog lijepog dana, stigla je ...

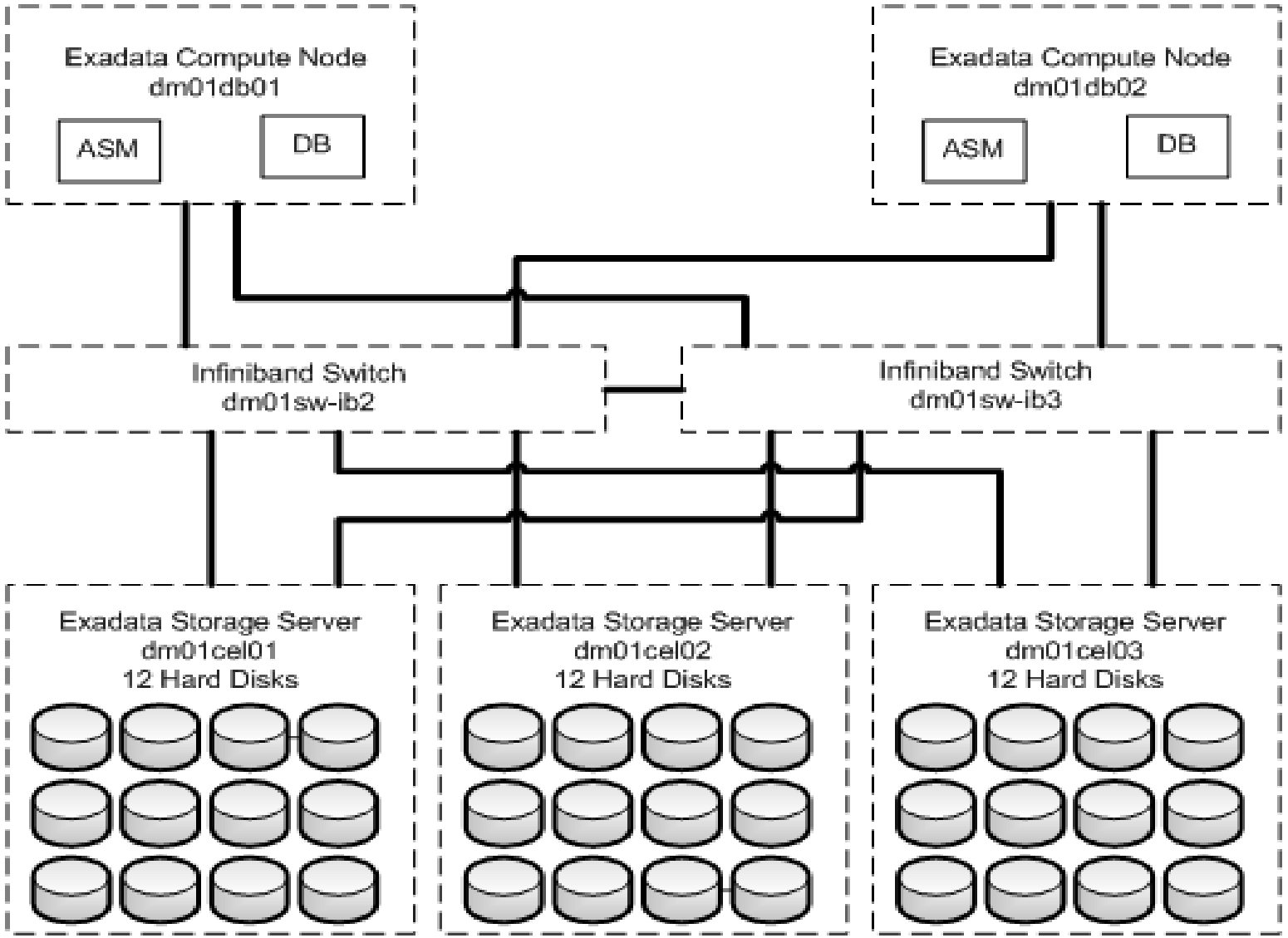


IT nakon otvaranja kutije

Čitanje deklaracije na kutiji

Exadata Database Machine X2-2 Hardware		
Exadata Database Machine X2-2 Full Rack	**Exadata Database Machine X2-2 Half Rack**	**Exadata Database Machine X2-2 Quarter Rack**
8 x Database Servers, each with: • 2 x Six-Core Intel® Xeon® X5675 Processors (3.06 GHz) • 96 GB Memory (expandable to 144GB) • Disk Controller HBA with 512MB Battery Backed Write Cache • 4 x 300 GB 10,000 RPM SAS Disks • 2 x QDR (40Gb/s) Ports • 2 x 10 Gb Ethernet Ports based on the Intel 82599 10GbE Controller • 4 x 1 Gb Ethernet Ports • 1 x ILOM Ethernet Port • 2 x Redundant Hot-Swappable Power Supplies	4 x Database Servers, each with: • 2 x Six-Core Intel® Xeon® X5675 Processors (3.06 GHz) • 96 GB Memory (expandable to 144GB) • Disk Controller HBA with 512MB Battery Backed Write Cache • 4 x 300 GB 10,000 RPM SAS Disks • 2 x QDR (40Gb/s) Ports • 2 x 10 Gb Ethernet Ports based on the Intel 82599 10GbE Controller • 4 x 1 Gb Ethernet Ports • 1 x ILOM Ethernet Port • 2 x Redundant Hot-Swappable Power Supplies	2 x Database Servers, each with: • 2 x Six-Core Intel® Xeon® X5675 Processors (3.06 GHz) • 96 GB Memory (expandable to 144GB) • Disk Controller HBA with 512MB Battery Backed Write Cache • 4 x 300 GB 10,000 RPM SAS Disks • 2 x QDR (40Gb/s) Ports • 2 x 10 Gb Ethernet Ports based on the Intel 82599 10GbE Controller • 4 x 1 Gb Ethernet Ports • 1 x ILOM Ethernet Port • 2 x Redundant Hot-Swappable Power Supplies
3 x 36 port QDR (40 Gb/sec) InfiniBand Switches	3 x 36 port QDR (40 Gb/sec) InfiniBand Switches	2 x 36 port QDR (40 Gb/sec) InfiniBand Switches
14 x Exadata Storage Servers X2-2 with 12 x 600 GB 15,000 RPM High Performance SAS disks or 12 x 2 TB 7,200 RPM High Capacity SAS disks Includes: • 168 CPU cores for SQL processing • 5.3 TB Exadata Smart Flash Cache	7 x Exadata Storage Servers X2-2 with 12 x 600 GB 15,000 RPM High Performance SAS disks or 12 x 2 TB 7,200 RPM High Capacity SAS disks Includes: • 84 CPU cores for SQL processing • 2.6 TB Exadata Smart Flash Cache	3 x Exadata Storage Servers X2-2 with 12 x 600 GB 15,000 RPM High Performance SAS disks or 12 x 2 TB 7,200 RPM High Capacity SAS disks Includes: • 36 CPU cores for SQL processing • 1.1 TB Exadata Smart Flash Cache
Additional Hardware Components Included: • Ethernet switch for administration of the Database Machine • Keyboard, Video or Visual Display Unit, Mouse (KVM) hardware for local administration	Additional Hardware Components Included: • Ethernet switch for administration of the Database Machine • Keyboard, Video or Visual Display Unit, Mouse (KVM) hardware for local administration	Additional Hardware Components Included: • Ethernet switch for administration of the Database Machine • Keyboard, Video or Visual Display Unit, Mouse (KVM) hardware for local administration

Pa onda sheme za spajanje



Odrazi na površini - poznata tehnologija

- OS : Oracle Linux 5 (RH)
- Baza : Oracle 11g2 (sa dodanim vrijednostima)
  - DBRM
  - DBFS
  - IDB
  - Particioniranje
  - Kompresija
- Cluster : RAC sa 2 noda (u osnovnom modelu)
- Storage : ASM
- GoldenGate

Ispod površine – tajni sastojci

- Smart Flash Cache (kralježnica) ♡
- Smart Scan (algoritam za optimizaciju) ♡
- Smart Flash Logging
- IO Resource Manager
- Hybrid columnar compression ♡
- Storage Indexes ♡
- Neka optimizacija za Data Mining

Izazovi dizajna

- Migracija podataka (online, offline)
- Logička organizacija istih
  - Particioniranje
  - Arhiviranje
  - Indeksiranje
- Prilagodba programa (ODI konfiguracija)
- Automatizacija obrada
- Zaštita (iliti backup)
- D/R
- Sigurnost (integracija s centralnom administracijom)

Priprema podataka : migracija

- Data Pump
- DBFS
  - ASM datafile-ovi izloženi kao direktoriji na OS-u, kako bi ih mogao koristiti program za izdvajanje ulaznih podataka
- External tables
- Fazni pristup
- Paralelan rad

Rezultat : migracija 6TB sa stare okoline u 3TB na Exadati, u jednom komadu

Dizajn nove okoline

- Konkuretan pristup podacima
  - paralelan dohvat i punjenje istih
  - Problem zaključavanja
- Podrška za automatizirano arhiviranje
- Indeksiranje (što s njim, a što s integritetom podatka)
- Testna/razvojna okolina

Dizajn podataka : Particioniranje

- Koncept korišten u starom sustavu
- U novom okruženje dobiva novu dimenziju
- Smanjuje zaključavanje podataka
  - u kombinaciji s particijiskim indeksima i izmjenama particija
- Povećava konkurentnost pristupa
  - u kombinaciji s paralelizmom
- Pojednostavljuje arhiviranje
- Povećava perfomanse
  - kombinaciji sa Smart Scanom i Storage Indeksima
- Zahtijeva određenu administraciju ako se želi fleksibilnost

Zaštita podataka

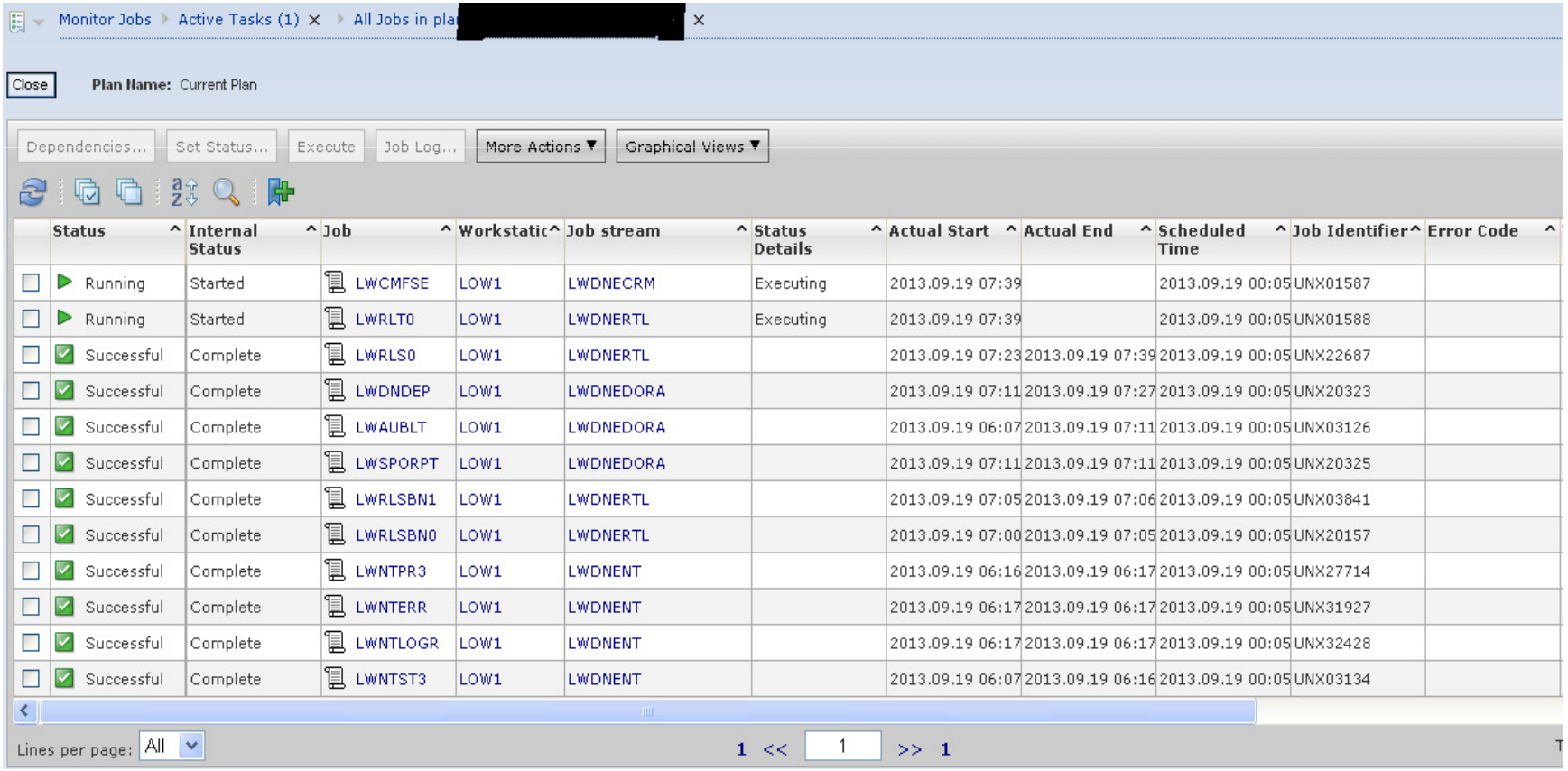
- Integracija s postojećim sustavom za zaštitu podataka na „virtualne” trake
- Agent za Linux/Oracle
- Full zaštita koristi 100% resursa postojećeg sustava
- Razmisliti o dediciranom sustavu samo za Exadata-u, odnosno kombinirati zaštitu sa export schema-om, odnosno zaštitama konkretnih podataka

Sigurnost

- Grupe na OS-u i Role u Oracle-u
- Integracija s centralnim sustavom pomoću agenta za Linux i Oracle
- Automatizacija dodjeljivanja profila
- ODI je riješen pomoću posebnih procedura, no za automatizaciju je potrebno napisati skripte (još čekamo da se netko javi)

Automatizacija obrada

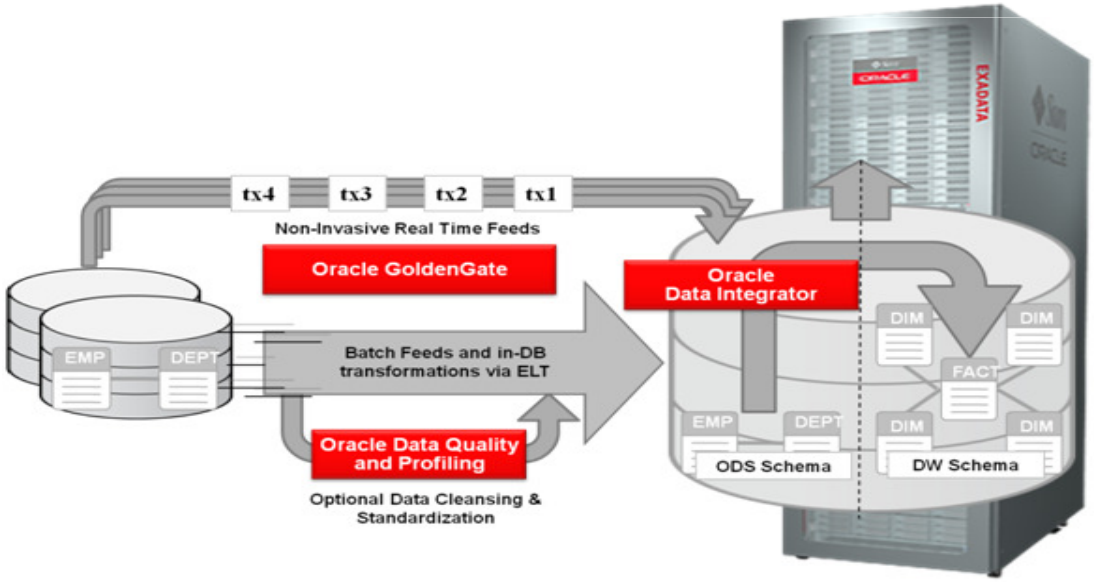
- ODI jobovi uključeni u centralni sustav pomoću agenta i shell skripti
- Integrirani lanci obrada po svim sustavima



Status	Internal Status	Job	Workstatic	Job stream	Status Details	Actual Start	Actual End	Scheduled Time	Job Identifier	Error Code
Running	Started	LWCMFSE	LOW1	LWDNECRM	Executing	2013.09.19 07:39		2013.09.19 00:05	UNX01587	
Running	Started	LWRLT0	LOW1	LWDNERTL	Executing	2013.09.19 07:39		2013.09.19 00:05	UNX01588	
Successful	Complete	LWRLS0	LOW1	LWDNERTL		2013.09.19 07:23	2013.09.19 07:39	2013.09.19 00:05	UNX22687	
Successful	Complete	LWDNDEP	LOW1	LWDNEDORA		2013.09.19 07:11	2013.09.19 07:27	2013.09.19 00:05	UNX20323	
Successful	Complete	LWAUBLT	LOW1	LWDNEDORA		2013.09.19 06:07	2013.09.19 07:11	2013.09.19 00:05	UNX03126	
Successful	Complete	LWSPORPT	LOW1	LWDNEDORA		2013.09.19 07:11	2013.09.19 07:11	2013.09.19 00:05	UNX20325	
Successful	Complete	LWRLSBN1	LOW1	LWDNERTL		2013.09.19 07:05	2013.09.19 07:06	2013.09.19 00:05	UNX03841	
Successful	Complete	LWRLSBN0	LOW1	LWDNERTL		2013.09.19 07:00	2013.09.19 07:05	2013.09.19 00:05	UNX20157	
Successful	Complete	LWNTPR3	LOW1	LWDNENT		2013.09.19 06:16	2013.09.19 06:17	2013.09.19 00:05	UNX27714	
Successful	Complete	LWNTERR	LOW1	LWDNENT		2013.09.19 06:17	2013.09.19 06:17	2013.09.19 00:05	UNX31927	
Successful	Complete	LWNTLOGR	LOW1	LWDNENT		2013.09.19 06:17	2013.09.19 06:17	2013.09.19 00:05	UNX32428	
Successful	Complete	LWNTST3	LOW1	LWDNENT		2013.09.19 06:07	2013.09.19 06:16	2013.09.19 00:05	UNX03134	

Povezivanje s ostalim izvorima

- Jdbc database driveri
- Mapiranje tipova podataka u ODI KM
- GOLDENGATE za replikaciju iz transakcijskog sustava
- Korištenje DBFS-ova za SQLLDR

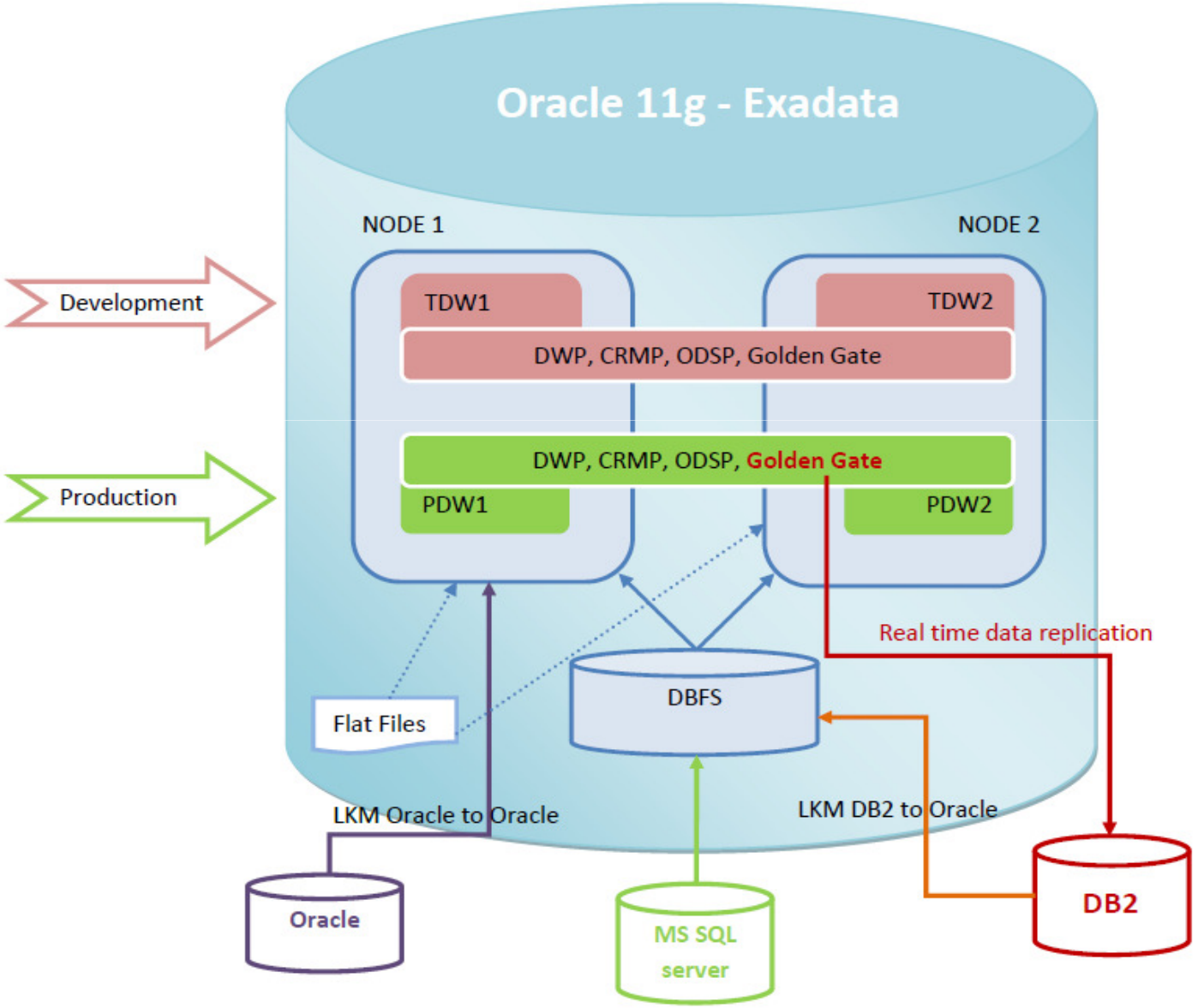


Arhiviranje

- Najjednostavnije u kombinaciji s particioniranjem
- Omogućuje automatsku kompresiranje na ARCHIVE mod
- Bitno zbog veličine backupa i D/R strategije
- Potrebno napisati skripte za parametrizaciju kreiranja i arhiviranja particija
- Može i kao najobičniji append u insert mode-u

Tako smo mi složili našle puzzle

Nova arhitektura



I što smo na koncu naučili

Sve je u životu relativno, osim .....

Kompresije

- Pokazala se odlična u svim situacijama
- Preko 90% podataka je trenutno kompresirano (u nekom stupnju), dok CPU opterećenje sistema ne prelazi 30% (ni u našim morama)
- U većini situacija i performanse su bolje
- Naravno, u nekom OLTP okruženju , priča bi mogla biti kompletno drugačija, ali ovo nije ta priča
- HCC kompresija se aktivira samo kod bulk operacija (direct load, insert append, paralelni load), ne i kod operacija sa pojedinačnim slogovima

Što smo na koncu naučili (2)

Zaboravite na .....

Indekse

- Ne trebaju vam, Exadata ima svoje (ne radi se o klasičnom stablu, ali su 100 brži)
- Ako vam trebaju za nametanje integriteta. Jednostavno ih proglasite nevidljivim
- Bilo kakva operacija s logičkim indeksima, isključuje Smart Scan

Što smo na koncu naučili (da, ima još)

Kako isključiti .....

Smart Scan

(no ne mislimo to ponavljati)

- Isključite automatske statistike. Nemojte puštatii ručne. Kompletno zbunite Oracle
- Pišite složene union operacije
- Kreirajte gomilu indeksa
- Kad želite usporiti upit do 100X

Što smo na koncu naučili (ovo je sad zadnje)

Particioniranje je dobro ali u tek kombinaciji sa Smart Scanom postaje ...

ZAKON

- Kad su podaci grupirani kao kod particija, Smart Scan postaje izrazito efikasan, pogovo ako su kolone međusobno korelirane
- Omogućuje paralelni dohvat podataka iz Smart Cache i diska
- Omogućuje eliminaciju particija
- U kombinaciji s lokalnom indeksima, smanjuje zaključavanje i time povećava konkurentnog RW operacija
- U kombinacij s paralelizmom i Smart Scanom, ubrzava upit do 1000x (dobro, to sam sad izmislio, ali siguran sam kako sam blizu)

Što još učimo

- Paralelizam (eskplozivna mješavina)
  - Izrazito moćan, ali i zahtjevan za upravljanje
  - U kombinaciji s TEMP SPACE-om, veliki potrošač resursa
  - Kako upravljati raspodjelu paralelizma između instanci
  - Ciljano namještati za najteže operacije, neka ostalo odradi Smart Scan s minimalnim paralelizmom (2)
  - PARALLEL_DEGREE_POLICY=AUTO („In memory” paralelizam)
- DBRM
  - Upravljanje dodjelom resura, prema tipu korisnika, aplikacija
  - Nužnost kod više instanci
- Data Mining
  - Upravljanje CPU potrošnjom, općenito resursima
  - Integracija s DBRM i postavkama za paralelizam

Osjetljiva područja

- Temp tablespace :
  - raspodjela diskovnog TEMP prostora između node-ova u klusteru izgleda nasumično
- Korisnici :
  - Exadata nije optimizirana sa sve moguće kombinacije predikata i funkcija koje korisnici mogu smisliti (razmisliti o uvođenju politike resursa, posebno u slučaju paralelizma)
- Virtualizacija :
  - korištenje više okolina na jednoj fizičkoj Exadata je nespretno sa stajališta sigurnosti i upravljanja zakrpama

Jedan dan s Exadatom

- Enterprise Manager 11gR2
- Upravljanje zakrpama
- Change Management
- Platinum support

Planovi za budućnost

- Enteprise Manager 12c
- Data Mining (upravljanje resursima)
- D/R unapređenje
- Exalytics integracija (ODI agent + nadzor resura)