

Az NI adattárházának evolúciója

Adattárház Fórum 2013

2013. június 5.

Németh Márk

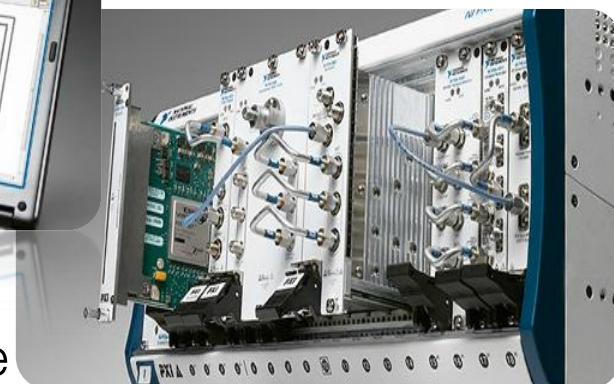
Senior Programmer Analyst

National Instruments

National Instruments

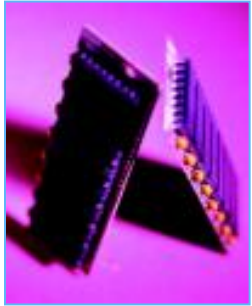
National Instruments

- Piacvezető a számítógép-alapú mérés és irányítás automatizálásban
- 1,052 milliárd \$ árbevétel (2011)
- > 6 500 munkavállaló
> 45 országban
- Fortune felmérés: legjobb 100 munkaadó (Amerika)



Mérnökök és kutatók használják
tesztelésre, mérésre és tervezésre

A változatos felhasználás



Félvezetőgyártás

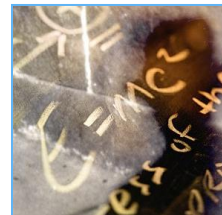
Autóipar

Elektronika

Tele-
kommunikáció

Környezet-
védelem

Egészségügy



Légierő/Űrkutatás

Oktatás

Big physics

Élelmiszeripar

Textilgyártás

Több mint 35.000 ügyfél

köztük a Fortune 500 cégek 90%-a



Az NI adattárháza

Az adattárház története

- 1998
 - indulás, WWW értékesítési adatok
 - frissítéskor felülírás
 - saját ETL keretrendszer
 - napi frissítés, TTS, 1+ óra állásidő
 - más üzleti területek fokozatos megjelenése

Az adattárház története

- 2005
 - magyar DW csapat megalakulása 2 fővel
- 2007
 - Történetiség tárolása a nyersadat szintjén (snapshot → Change Data Capture)
 - Egyes ETL-ek és adatpiacok már felhasználják record history-t
 - ETL Tool bevezetés: Oracle Warehouse Builder
 - Frissítés adatbázis linken keresztül → nincs downtime

Az adattárház története

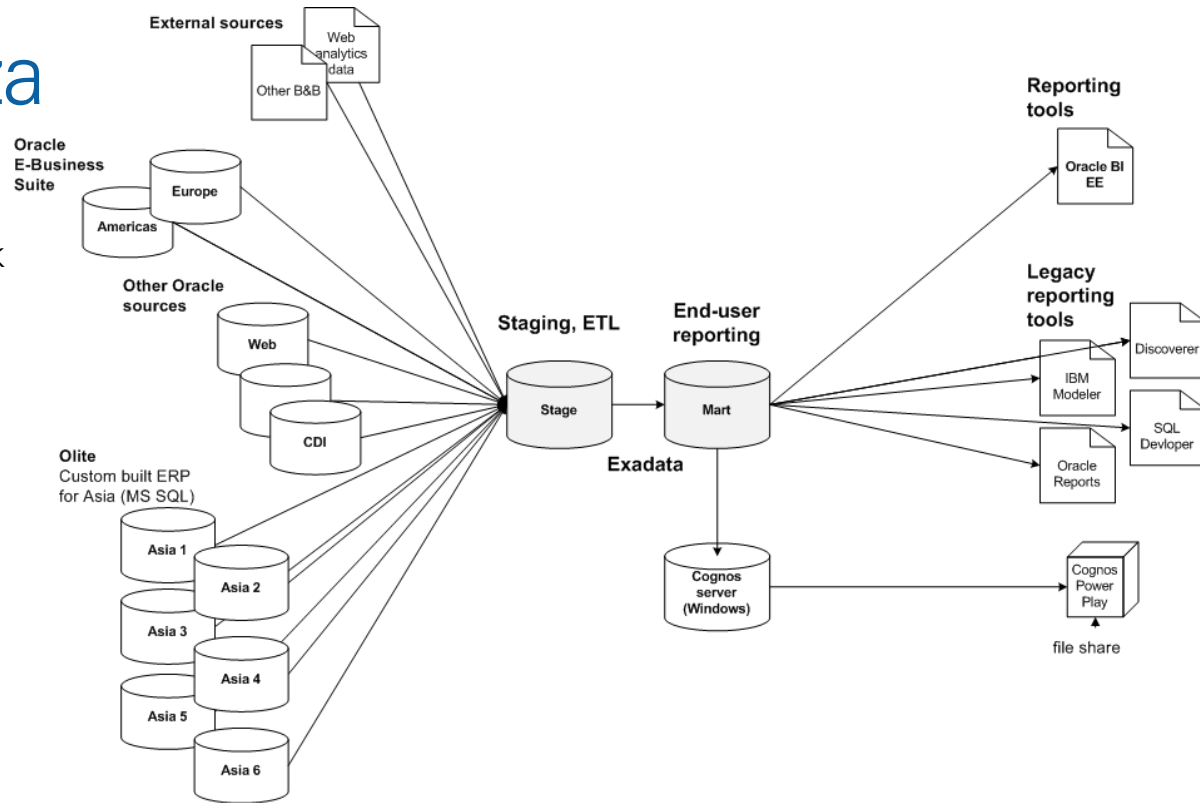
- 2008
 - Oracle BI EE
- 2011
 - magyar csapat 11 fősre bővül

Az adattárház története

- 2012
 - adattárház migrációja Exadata-ra (ami egyben RAC migráció is)
 - új, jobban strukturált ETL architektúra, saját fejlesztésű engine-ek
- 2013
 - ETL Tool cseréje
 - Két adatbázis egyesítése

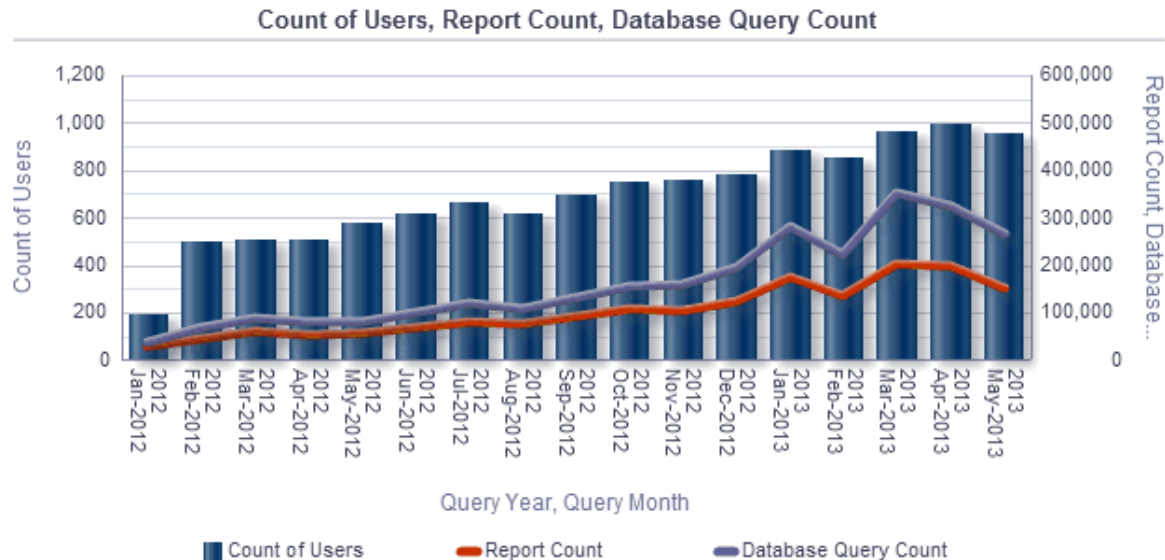
Az NI adattárháza

- Hardware
 - Exadata X2-2 Quarter Rack
- Adatbázis
 - Oracle Database
 - stage: 5,2 TB
 - mart: 1,7 TB
- ETL
 - Oracle Warehouse Builder
 - saját SQL PL/SQL kód
- Riportolás
 - Oracle BI EE
 - (Discoverer, Powerplay, SQL Developer, Crystal Reports, SPSS, IBM Modeler)
- Forrásrendszerek:
 - 2x Oracle E-Business Suite
 - 6x MS SQL Server (házon belül fejlesztett ERP)
 - Egyéb Oracle adatbázisok (CDI, ni.com)
 - Eloqua, Lotus Notes, Cloud (XML), szöveges fájlok



Az NI adattárháza

- Felhasználók száma:
 - Oracle BI: ~ 1000



- Fejlesztők száma:
 - DW: 14 állandó és 9 alkalmi (más csapatokból)
 - ETL Core Team: 4 fő (architektúra, engine fejlesztés, sztenderdek, code és design review)

Kihívások

Kihívások - Fejlesztési idő lerövidítése

- Exadata
 - gyorsabb fejlesztői környezet
- Architektúra
 - metaadat-vezérelt engine-ek
 - ETL sablonok és sztenderdek
- ETL tool

Kihívások – Point-in-time riportok

- Architektúra
 - alapértelmezés lett a történetiség tárolása
 - saját fejlesztésű engine a duplikációk kiszűréséhez (valódi változások azonosítása)
- Exadata
 - megnövekedett adatvolumen

Kihívások - Gyakoribb frissítés

- Exadata
 - parallel query és parallel DML
 - auto DOP
- Architektúra

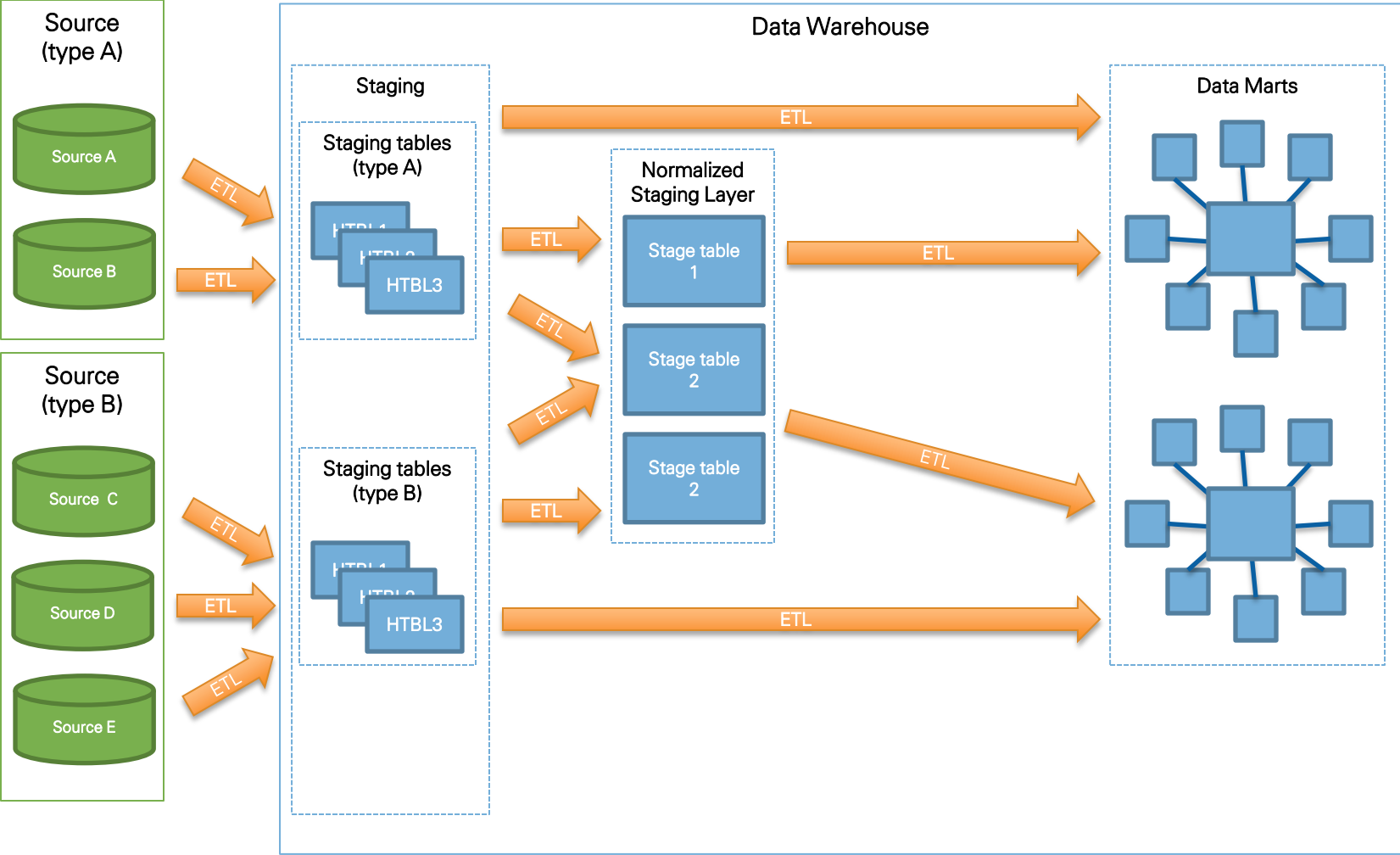
Kihívások - Adatminőség

- Architektúra
 - Dedikált réteg (cleanse)

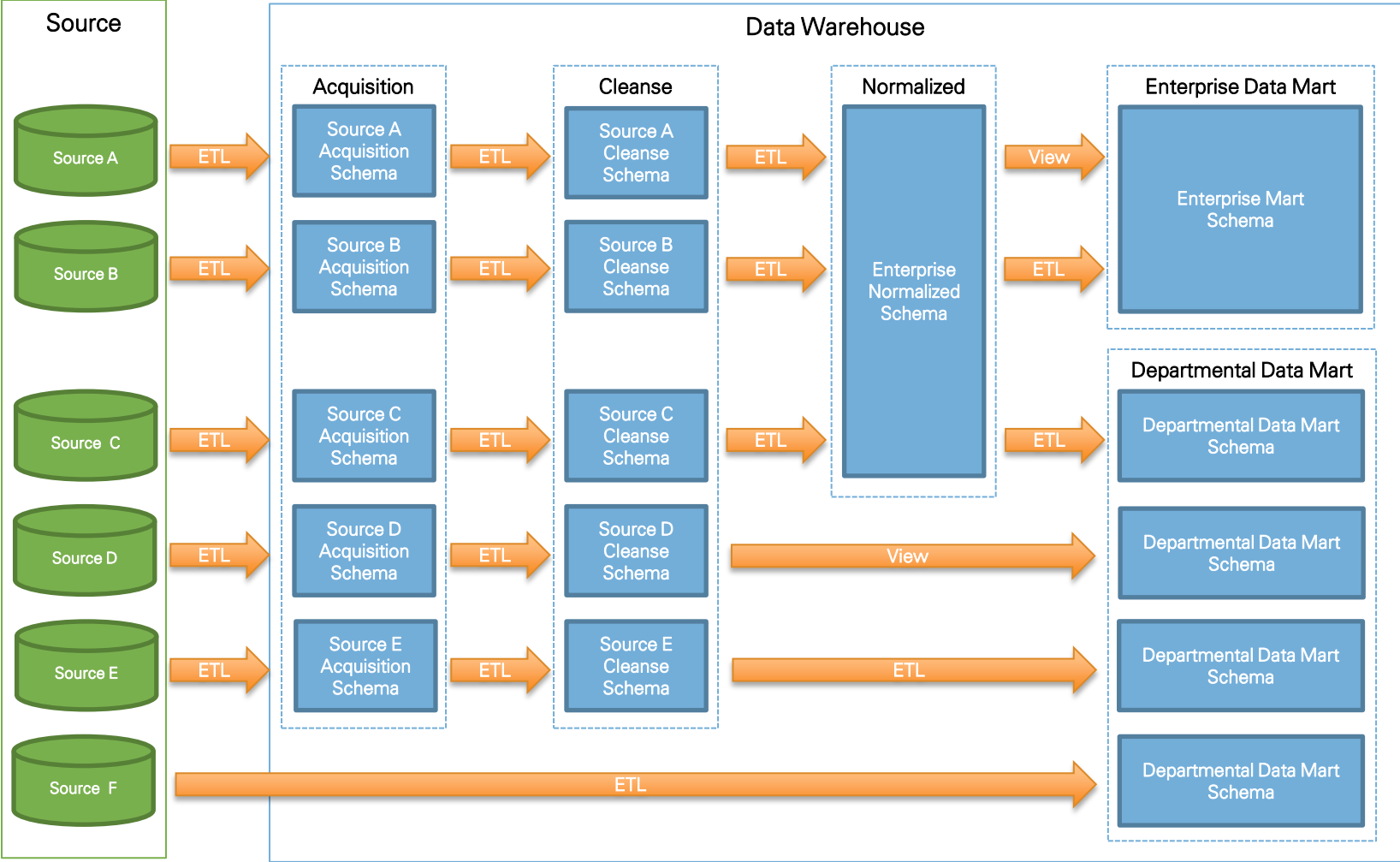
Kihívások - DW mint platform

- Architektúra
- ETL Core Team
- Dokumentáció, tréningek és konzultáció

Korábbi ETL architektúra

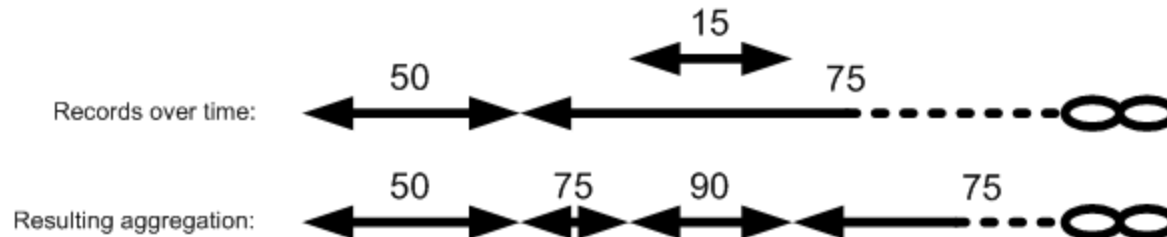


Új ETL architektúra



Történetiség kezelés az ETL-ben

seat_type	seat_count	row_start_date	row_end_date
computer	50	5/5/2010 10:00	5/1/2011 15:25
computer	75	5/1/2011 15:25	1/1/4000
user	15	6/12/2011 23:35	6/12/2012 13:00



row_start_date	row_end_date	SUM(seat_count)
5/5/2010 10:00	5/1/2011 15:25	50
5/1/2011 15:25	6/12/2011 23:34	75
6/12/2011 23:35	6/12/2012 13:00	90
6/12/2012 13:00	1/1/4000	75

Az adattárház migrációja Exadatára

Előzmények

- Korábbi konfiguráció: Sun M5000 + Hitachi USP-V
- Rendszer gyengeségeinek felismerése egy ERP verziófrissítés projekt kapcsán
- Szóba jöhető megoldások:
 - Oracle Exadata Database Machine X2-2 Quarter Rack
 - Előre konfigurált
 - Oracle Server (M/T-Series) és HDS VSP Storage
 - Házon belül konfigurált megoldás
- Eredmények:
 - ETL és Report: 4,5x és 10,5x gyorsulás
 - Hasonló, Smart Scan esetén az Exadata gyorsabb

Elvárások az Exadatával szemben

Üzleti előnyök	Oracle Exadata
Projektek időtartamának csökkenése	25%
Projektek terjedelmének csökkenése	20%
BI hatékonyságának javulása	25%
Riportfejlesztés hatékonyságának javulása	25%

A csapat és az időzítés

- 1x DBA Architect
- + időszakosan további DBA-k (OEM, HP PPM)
- 1x DW Architect
- 2x Unix Admin
- 1x PM
- 1x DW programozó + még kettő, rövidebb időre

6 fő

- 2012 május – Döntés
- 2012 június – Az Exadata megérkezik
- 2012 július 27. – Első migrációs kísérlet
- 2012 aug – okt. – DBA és fejlesztői munka
- 2012 okt. 20. – Átállás
csak szükséges módosítások
optimalizálás nélkül

3 hónap

5 hónap

- Átfedés a Q4 negyedéves release-zel (aug. – nov. 10)

A migráció kihívásai

- Az elkerülhetetlen
 - Eltérő SW és HW platform (SPARC Solaris → x86 Linux, little endian vs. big endian)
 - RAC
 - ASM
 - OWB
- Ha már egyszer költöztetjük a DW-t, akkor...
 - Karakterkészlet-váltás
 - Szervernév- és adatbázisnév-változtatás
 - Adatfájl- és táblatér-beállítások változtatása
- Felhasználók és alkalmazásaik beazonosítása

Az átállítás

- A migráció maga problémamentesen lezajlott
- Az első néhány nap
 - Exadata backup / refresh
 - saját fejlesztésű alkalmazás hibája
 - Oracle Gateway
- 4x ETL futási idő javulás
- 6x riport futási idő javulás
- Optimalizálás nélkül

ETL futási idők

Módosítások nélkül

ETL	Régi	Exadata	Régi / Exadata
B&B Incremental	3:05:00	0:57:00	3,2 X
Golden Sites Incremental	2:06:00	0:22:00	5,7 X
Contact Marketable Flag	2:33:00	0:52:00	2,9 X
Deferred Revenue Incremental	0:25:05	0:06:40	3,8 X
End User Contact Incremental	1:21:41	0:17:14	4,7 X

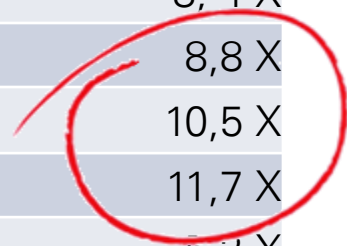
Parallel query és DML engedélyezésével, ill. minimális optimalizálással

ETL	Régi	Exadata	Régi / Exadata
B&B Full	8:54:00	1:34:00	5,7 X

OBIEE Dashboard futási idők

Dashboard Page	Régi rendszer (mp)	Exadata (mp)	Gyorsulás
Accounts	50	12	4,2 X
Run Rate	62	8	7,8 X
Territories	5	4	6,3 X
Summary	3	17	8,4 X
Opportunities	5	4	8,8 X
YoY, Trends Products	3	15	10,5 X
RT & Energy Summary	20	23	11,7 X
Billings Analysis	31	6	5,2 X
Bookings Analysis	14	3	4,7 X
Bookings Forecast	4	1	4,0 X
Hedge Billings Forecast	71	29	2,4 X
Order Size Analysis	23	6	3,8 X
Quater-to-Date Run Rate Analysis	18	3	6,0 X
Year-to-Date Run Rate Analysis	87	17	5,1 X

8x-12x



NetInsight lekérdezések futási ideje

Lekérdezés	Futási idő		Javulás
	Régi rendszer	Exadata	
Complex Analytics	0:57:06	0:01:30	38 X
Direct Marketing Query	0:19:44	0:00:24	49 X
Contact View History	0:15:28	0:00:16	58 X
w/ UPID	0:13:28	0:00:14	58 X
Search Query	0:11:55	0:00:11	65 X
w/o UPID	0:02:55	0:00:01	175 X

38x-175x

További tervek

- Exadata Optimalizálás
 - ETL és riportok
 - Parallel Query és DML, Auto DOP
 - Indexek felülvizsgálata
 - EHCC
- A két DW adatbázis egyesítése
- Advanced Analytics on Exadata

Kérdések?

