

Nagyvállalati adatintegráció és adatkezelés

az Informatica® eszközeivel

Biró Attila

fejlesztési igazgató, Areus Zrt.

Nagy Balázs

szoftverfejlesztő, Areus Zrt.



Miről lesz ma szó?

- **17:05 – 18:00:** Prezentáció /Biró Attila, Nagy Balázs/
 - A nagyvállalati adatkezelés problémakörei
 - Adatintegrációs szoftverek bemutatása
 - Areus Zrt. és Informatica Corporation
- **18:00 – 18:15:** Szünet
- **18:15 – ?:** Kérdések és válaszok

IT trendek – Mit hoz a jövő?

FUTURE



CLOUD



INTERACTIONS



MOBILE

PAST



ON-PREMISE

WHERE



TRANSACTIONS

WHAT



DESKTOP

HOW

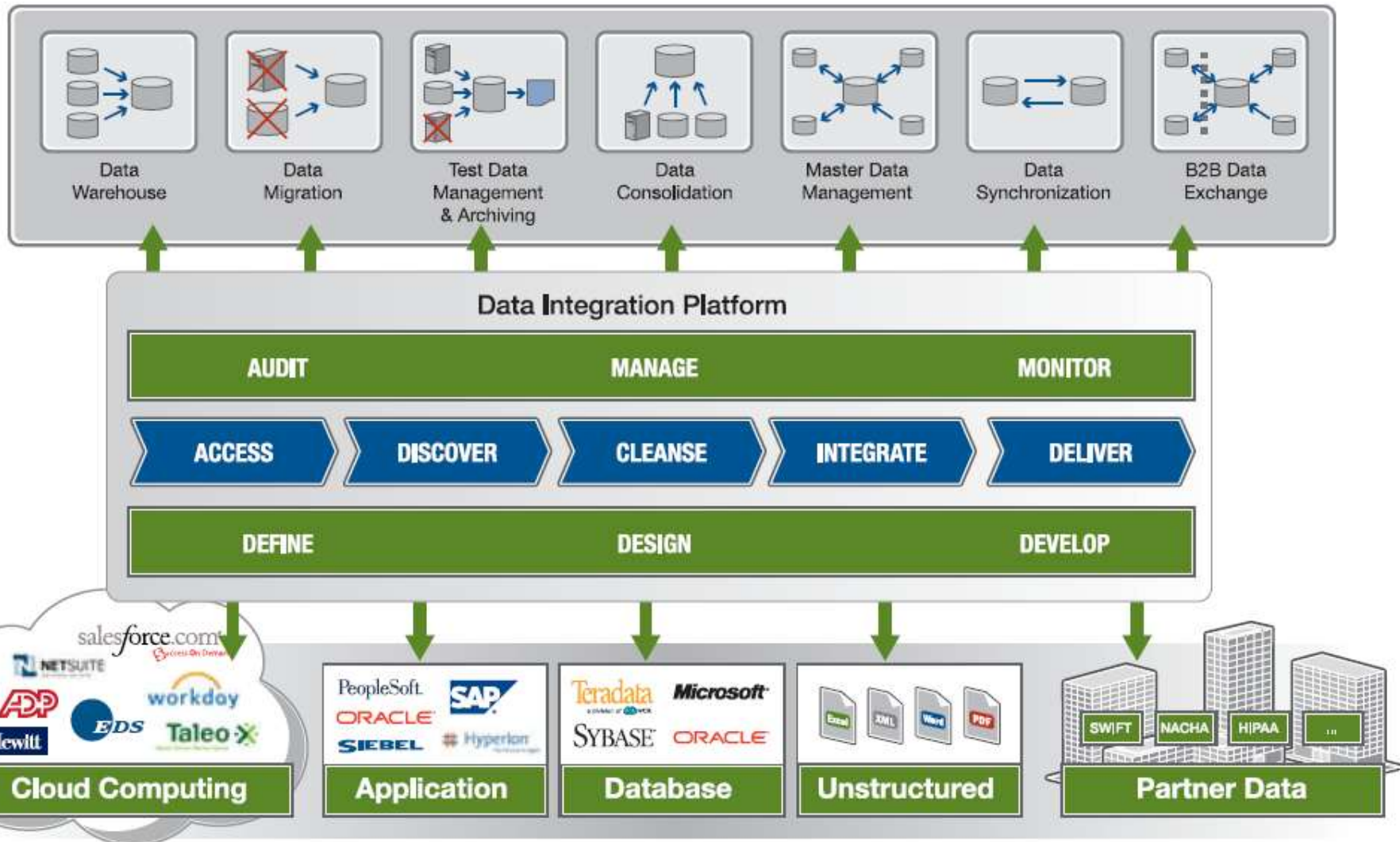


Gartner®

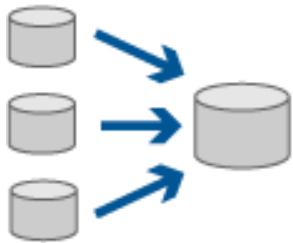



areus
more in one

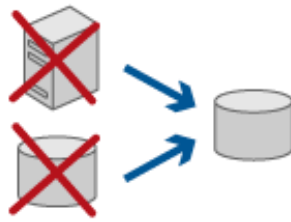
Adatintegráció



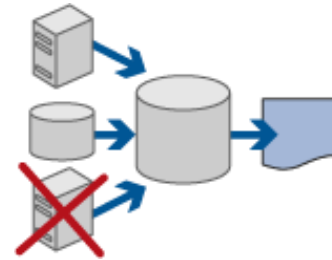
Adatintegráció – Használati esetek



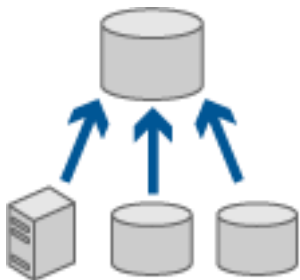
DW



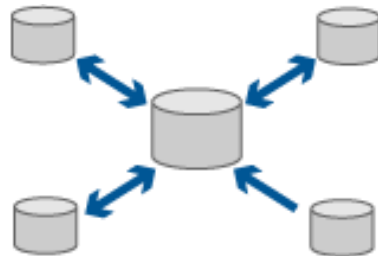
Adatmigráció



Tesztadat kezelés
és Archiválás



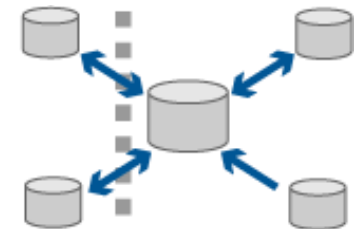
Adatkonzolidáció



MDM



Replikáció /
Szinkronizáció



B2B



INFORMATICA[®]
The Data Integration Company[™]



1990-es évek: az első kódolásmentes ETL eszköz

2002: az első 64 bites ETL eszköz

2006: az első grid támogatású ETL eszköz

2007: az első felhő alapú adatintegrációs megoldás

2011: Hadoop adatok natív feldolgozása (Hparser)

2012: Facebook, Twitter, LinkedIn, ...

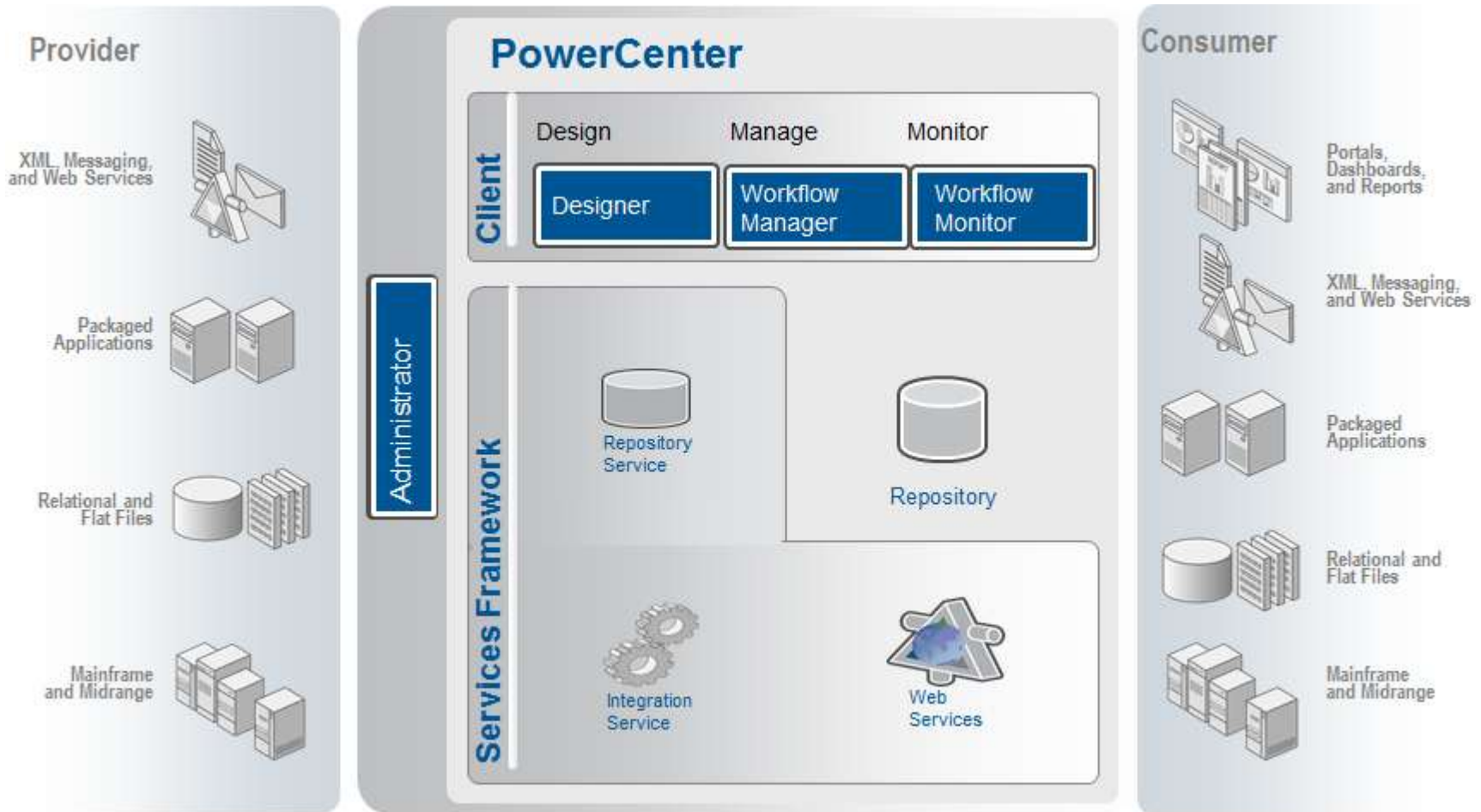


Informatica PowerCenter

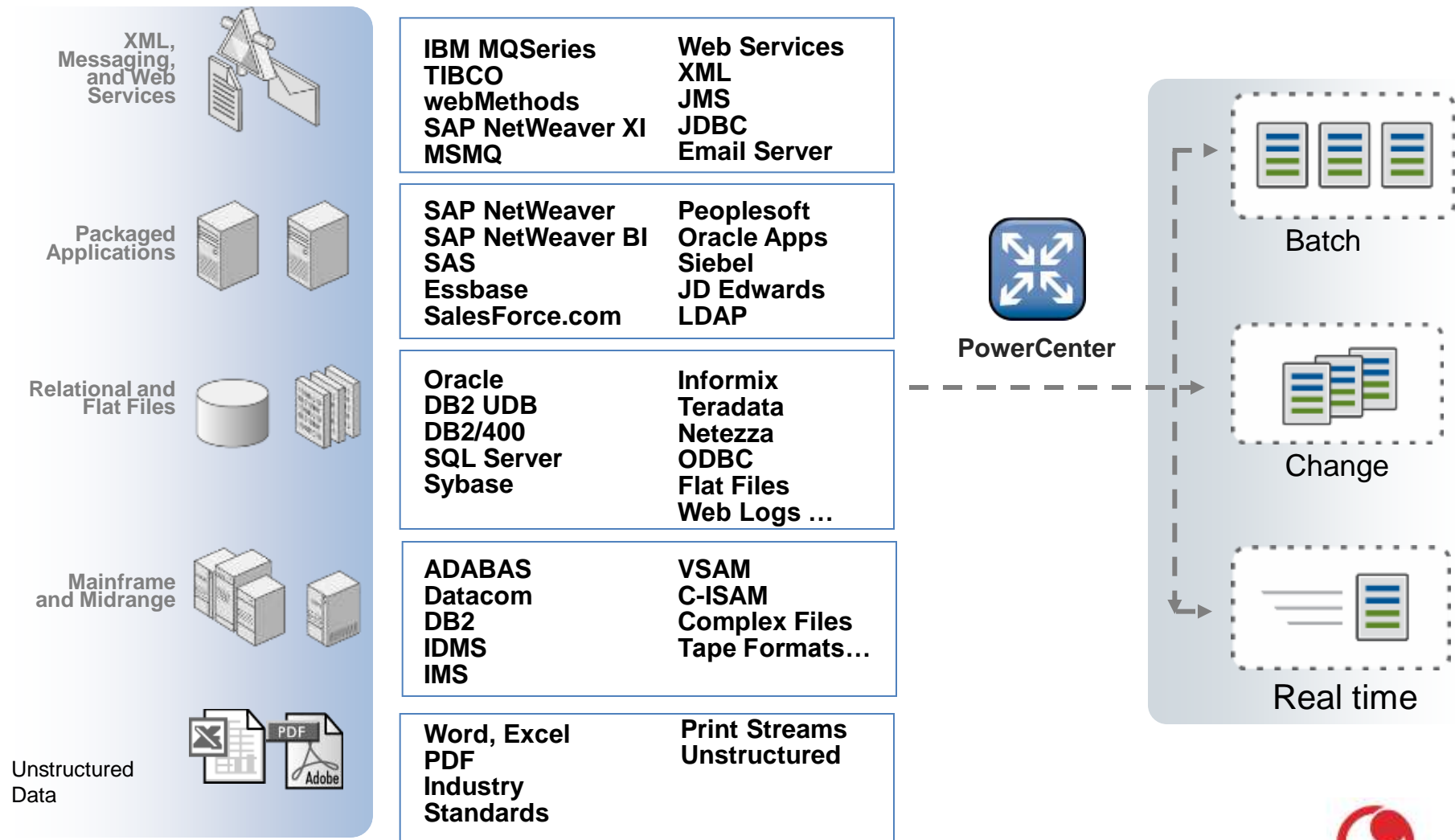
Informatica PowerCenter

- ◉ egységes, alkalmazás és platform független SOA architektúra
- ◉ univerzális csatlakoztathatóság
- ◉ késleltetésmentes, valós idejű adatintegráció, tranzakció figyelés és esemény feldolgozás
- ◉ nagy teljesítmény, jó skálázhatóság
- ◉ H/A, terheléselosztás és hibatűrés
- ◉ webszolgáltatásként és API-n keresztül is hívható




PowerCenter architektúra

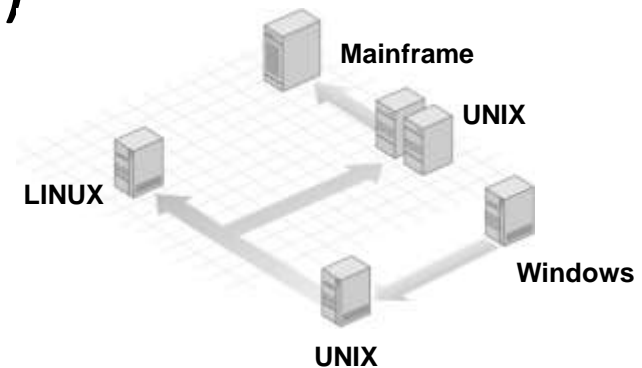


Univerzális csatlakoztathatóság



Optimalizációs lehetőségek

- ◉ Lineáris skálázhatóság, 64 bites architektúra
- ◉ Heterogén Grid + Terheléselosztás
- ◉ Dinamikus memória allokáció
- ◉ Pushdown Optimization (EL-T)
- ◉ Dinamikus particionálás
- ◉ Többféle adatbetöltési mód:
 - Batch 
 - Change Data 
 - Real-Time 



PowerCenter GUI

The screenshot shows the Informatica PowerCenter Designer interface. The title bar reads "Informatica PowerCenter Designer - [Mapping Designer - MappingLab - [Workshop]]". The menu bar includes Repository, Edit, View, Tools, Layout, Versioning, Mappings, Transformation, Window, and Help. The toolbar contains various icons for file operations and data processing. The left pane shows a tree view of the "MappingLab" repository, including sources like "FlatFile" and "TRANSACTIONS", and targets like "CUSTOMERS" and "GOOD_CUST_STG". The main workspace, titled "Mapping Designer", displays a data flow diagram. It starts with two source components: "TRANSACTIONS (Flat File)" and "CUSTOMERS (Oracle)". Both feed into "SQ_TRANSACTIONS" and "SQ_CUSTOMERS" respectively. These then merge into a join component "int_many_to_one". The output of the join goes to a filter component "ift_check_customer_name". Finally, the data is split into two target components: "CUSTOMER_NAME (Oracle)" and "GOOD_CUSTOMERS (Oracle)".

This section displays four panels related to data mapping configuration:

- Source Definition (Flat File):** A table with columns Name and Datatype.

K	Name	Datatype
	NEWFIELD	number
	NEWFIELD1	number
	NEWFIELD2	string
- Source Qualifier (SQL):** A table with columns Name and Datatype.

Name	Datatype
NEWFIELD	decimal
NEWFIELD1	decimal
NEWFIELD2	string
- Java Transformation (Java Transformation):** A table with columns Name and Datatype.

Name	Datatype
INPUT	
input1	integer
input2	integer
output1	integer
- Target Definition (Flat File):** A table with columns Name and Datatype.

K	Name	Datatype
	NEWFIELD	number
	NEWFIELD1	number
	NEWFIELD2	string

Blue arrows indicate the data flow from the Source Definition to the Source Qualifier, then to the Java Transformation, and finally to the Target Definition.

Adatprofilozás

INFORMATICA Analyst Administrator Sign Off | Manage | Help Search All Go

Browse: Projects CustomersProfile_Dev... Profile_Customers_St...

Column Profiling Properties Data Preview Actions

Profile_Customers... Sampling Policy: All 131 Rows | Last Run On: Sep 28, 2009 5:39:56 AM CDT

Name	Unique Value	Unique %	Null Values	Null %	Inferred Data Type	Preview
City	60	45.00	0	0.00	Varchar(19) [100.00%]	<input type="checkbox"/>
State	30	22.90	2	1.53	Fixed Length String(2) [100.00%]	<input type="checkbox"/>
PostalCode	74	56.49	12	9.16	Integer(5) [100.00%]	<input type="checkbox"/>
Country	9	6.87	19	14.50	Varchar(13) [100.00%]	<input type="checkbox"/>
Phone	87	66.41	28	21.37	Varchar(18) [100.00%]	<input type="checkbox"/>
Last_Order_Dt	69	52.67	4	3.05	Varchar(10) [100.00%]	<input checked="" type="checkbox"/>
Status	6	4.58	4	3.05	Varchar(9) [100.00%]	<input type="checkbox"/>

RULE: Format_Date_Rule2

Name	Unique Value	Unique %	Null Values	Null %	Inferred Data Type	Preview
Format_Date_Rule2	62	47.33	4	3.05	Varchar(10) [100.00%]	<input checked="" type="checkbox"/>

Frequency **Patterns** Statistics

Pattern	Frequency	Percent	Chart
Others	4	3.05	
9/99/9999	14	10.69	
9-9-9999	13	9.92	
9/9/9999	92	70.23	
99/99/9999	8	6.11	

Last_Order_Dt = '9-9-9999' (All 13 rows)

Customer_ID	Name	Last_Order_Dt	Format_Date_Rule2
1	1029 Andre Hayes	3-9-2006	3/9/2006
2	1030 Annika Larsson	8-2-2006	8/2/2006
3	1031 Anthony Dillon	2-7-2006	2/7/2006
4	1035 Brad Poorman	7-4-2006	7/4/2006
5	1036 Brent Salzer	8-1-2006	8/1/2006
6	1037 Carol Brookins	2-5-2006	2/5/2006
7	1041 Tim Diefendorf	6-3-2006	6/3/2006
8	1042 Craig Perkins	3-3-2006	3/3/2006

Done

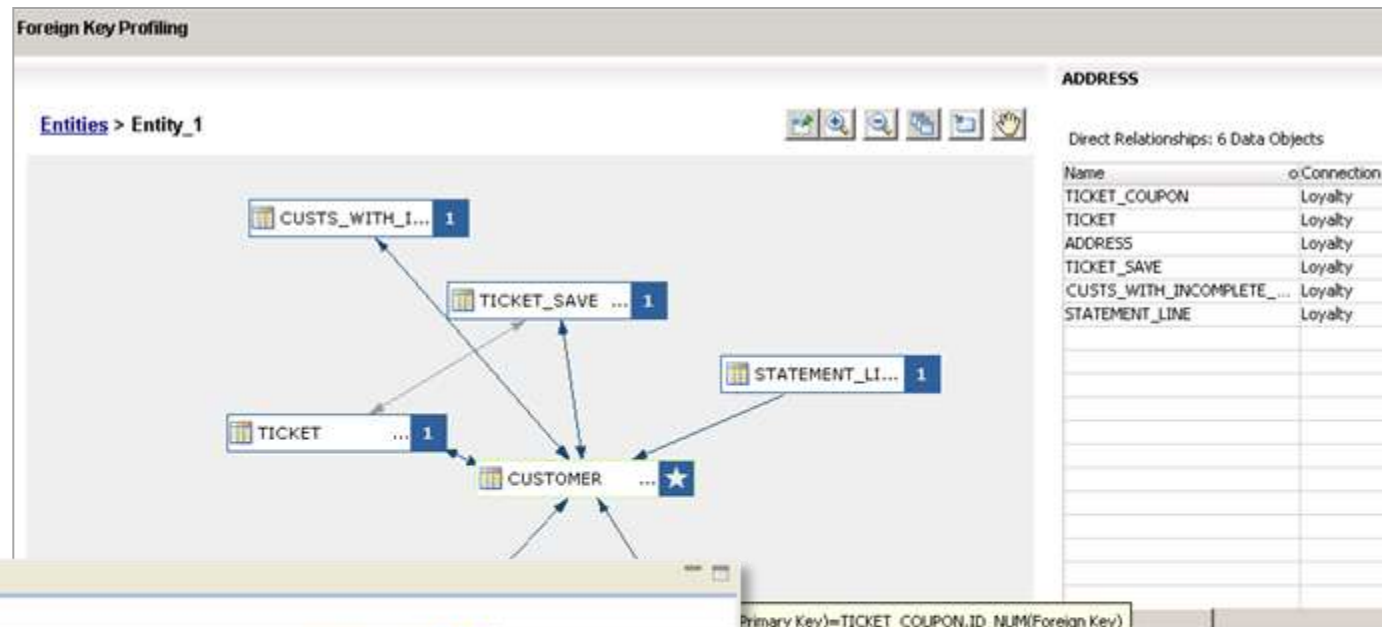
Values Patterns Statistics

Value	Frequency	Percent	Chart
2/4/2006	8	6.15	
3/1/2006	7	5.38	
7/1/2006	5	3.85	
3/3/2006	5	3.85	
NULL	4	3.08	
3/2/2006	4	3.08	
2/1/2006	4	3.08	
9/4/2006	3	2.31	
8/4/2006	3	2.31	
7/4/2006	3	2.31	
7/3/2006	3	2.31	
6/6/2006	3	2.31	
6/3/2006	3	2.31	

Values Patterns **Statistics**

Statistic	Value
Maximum Length	10
Minimum Length	8
Bottom (5)	1/1/2006
	1/19/2007
	1/2/2006
	1/4/2006
	1/7/2006
Top (5)	9/4/2006
	9/3/2006
	9/22/2001
	9/2/2006
	8/4/2006

Tartomány / kapcsolat feltérképezés



Primary Key)=TICKET_COUPON.ID_NUM(Foreign Key)

Level 2: 1
Level 3: 0
Selected Data Object: ★
1st Level Data Objects: 1
2nd Level Data Objects: 2
3rd Level Data Objects: 3
Other Level Data Objects: ...

Adatminőség monitorozása

Add to Scorecard: Step 3 of 3

Select each score and adjust its settings. Scores in the scorecard are the percentage of valid values.

Score Name	Column Name	Remove
Standardize_Status_Rule	Standardize_Status_Rule	

Score using: Values

Value	Frequency	Is Valid
A	110	<input checked="" type="checkbox"/>
I	16	<input checked="" type="checkbox"/>
?	4	<input type="checkbox"/>

Score Settings: Standardize_Status_Rule

Set Custom Thresholds for this Score

67 to 100 % ■ Good

33 to 66 % ■ Acceptable

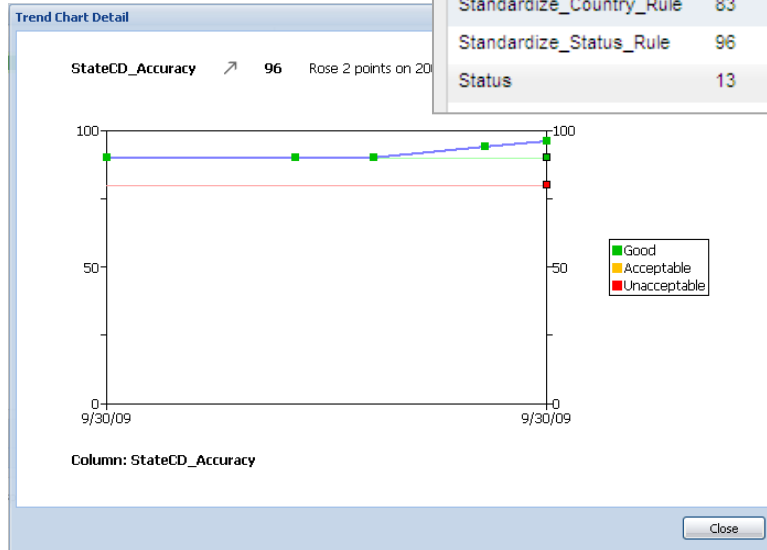
0 to 32 % ■ Unacceptable

Scorecard

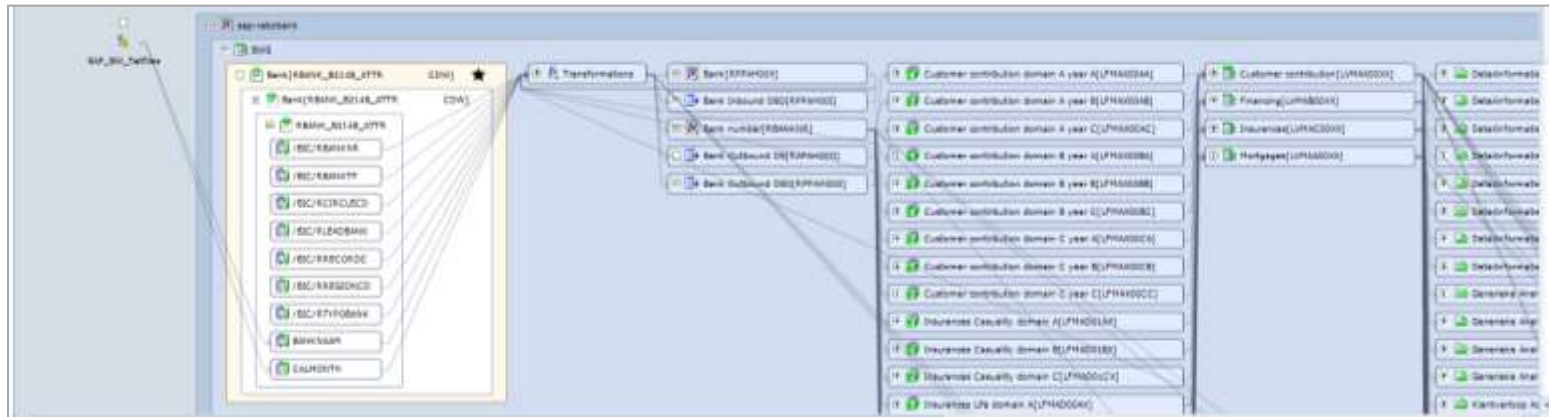
Customer_Scorecard_f... x

Customer_Scorecard_for_Country_and_Status - Scores Last Run On: Jun 12, 2010 8:01:28 AM EDT

Name	Score(%)	Score Bar	Data Object	Source	Source Type
Country	63		Customers	Country	Column
Standardize_Country_Rule	83		Customers	Standardize_Country_Rule	Rule
Standardize_Status_Rule	96		Customers	Standardize_Status_Rule	Rule
Status	13		Customers	Status	Column



Adatvizualizáció



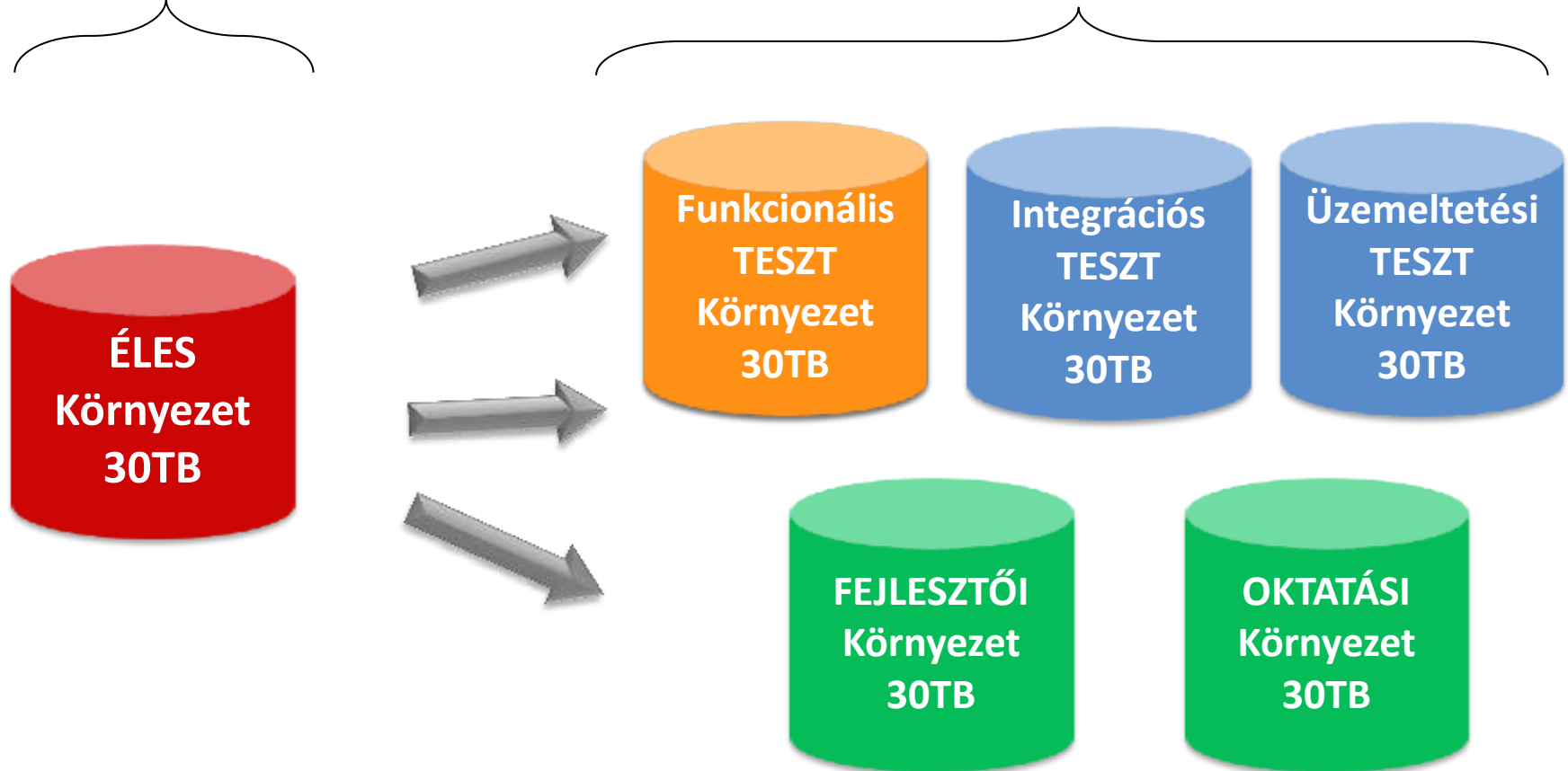
Miért jó ez a nagyvállalatoknak?

- Produktivitás javul
- Gyorsabb time-to-market, time-to-delivery
- Alacsonyabb IT üzemeltetési költségek
- Megbízható, átlátható és időszerű adatok
- Jobb üzleti döntések
- **Nagyobb bevétel**

Az adatkezelés problémakörei

Produktív (éles) környezet

Nem produktív üzemű környezetek



Az adatkezelés kritikus problémái

Heartland, Visa Announce \$60 Million Settlement

★ Credit Funds Would Reimburse Card Issuers for Breach-Related Losses
Eligible January 8, 2010 - Linda McGlasson, Managing Editor



Heartland Payment Systems announced to branded credit and debit card issuers up to losses incurred from the Heartland data breach, eclipsing the TJX settlement of \$4.5 billion in 2007.

In a statement, Heartland and Visa say they will be subject to certain conditions, including participation by Visa issuers. Visa says it will

Credit Card Companies Say TJ Maxx Breach Affected 94 Million Accounts

By Meg Marco on October 26, 2007 9:18 PM



According to new court papers, Visa and Mastercard are saying that the TJ Maxx security breach actually affected 94 million accounts—more than double the amount that TJ Maxx reported.

The new estimate is part of a lawsuit by the credit card companies against Fifth Third Bancorp and TJX (parent company of

and TJ Maxx, etc.).

that their fraud etc. are in the

Sony data breach could be most expensive ever

Sony Corp.'s PlayStation Network and Sony Online Entertainment suffered data breaches that could cost up to \$2 billion.



Hackers zap Zappos: Info from 24 million users stolen

Published January 16, 2012 | FoxNews.com

Print Email Share Comments Recommend 258 Tweet 300 LinkedIn Share 37



Popular online shoe retailer Zappos.com said late Sunday that hackers had accessed its network, stealing customer account information from as many as 24 million customers.

Credit card information was not stolen, company CEO Tony Hsieh said in a statement sent to users, but email addresses, billing and shipping addresses, phone numbers, the last four digits from credit cards -- and more -- may have been compromised.





Informatica

Persistent Data Masking & Data Subset

Mi a megoldás?

Forrás rendszerek

Informatica
Persistent Data Masking
and Data Subset

Oktatási, fejlesztői és teszt
környezetek

ORACLE PeopleSoft
SIEBEL

SAP

Custom
Apps



IBM DB2 Microsoft SQL Server

ORACLE SYBASE

TERADATA

CSONKOLÁS

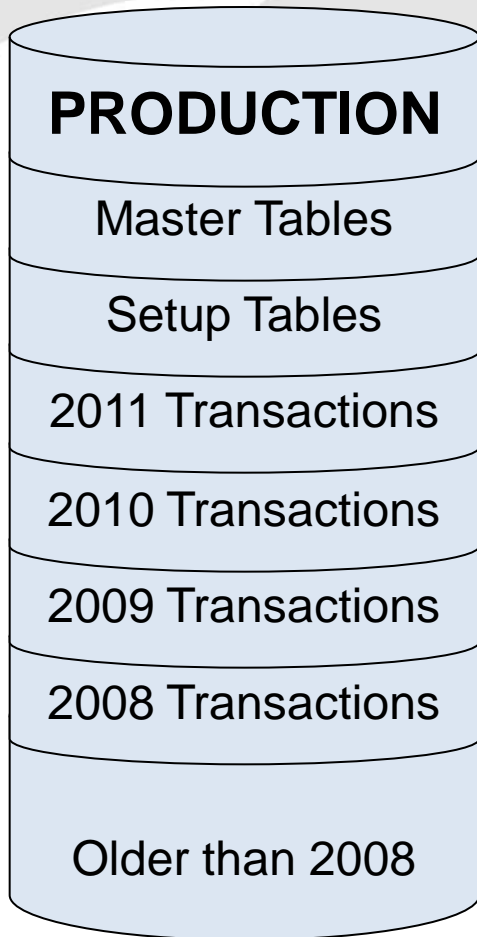
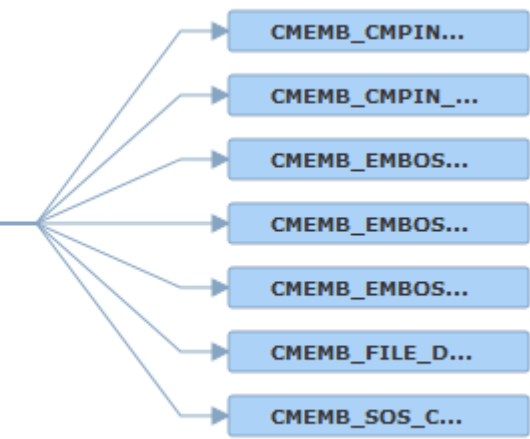
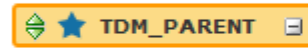
PERZISZTENS ADATMASZKOLÁS

CSONKOLÁS + MASZKOLÁS

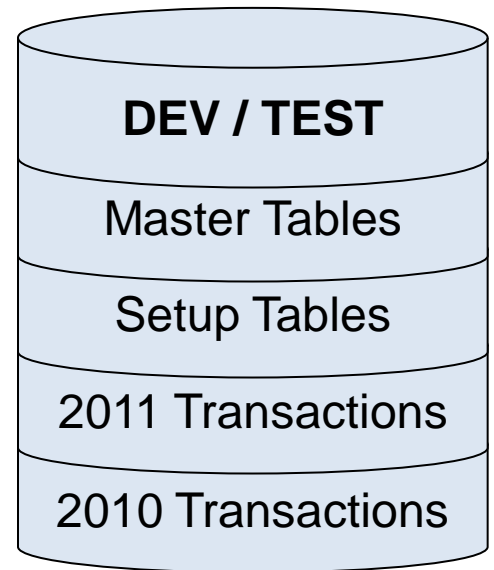
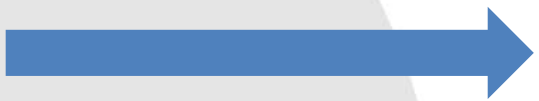


Csonkolás

Entitások és csoportok segítségével, a megfelelő szűrőfeltételeket megfogalmazva.



2010.01.01. utáni tranzakciók megőrzése



25 largest tables (GB)

Table	Schema	Module	Size (Table+Index)
ACT	ODSSUPER	Activity	275.8 (180.1+95.7)
FACT_ACTIVITY_ATTR	ODSSUPER	Activity	241.9 (183.1+58.8)
FACT_ACTIVITY_PRESE...	ODSSUPER	Activity	101.3 (91.6+9.7)
FACT_ACT_ATTR_BKUP_...	ODSSUPER	Other	78.3 (74.8+3.4)
BKP_0809_FACT_ACTIV...	ODSSUPER	Backup	76.6 (69.7+7.0)
FAA_BKUP_FACT_ACT	ODSSUPER	Other	76.1 (69.7+6.4)
FACT_ACTIVITY_05_06	ODSSUPER	Activity	44.4 (44.4+0.0)
FACT_ACTVT_ATTR_05_...	ODSSUPER	Activity	42.9 (42.9+0.0)
FACT_ACTIVITY_ATTR_...	ODSSUPER	Activity	39.3 (36.0+3.3)
FACT_ACTIVITY_STP	ODSSUPER	Activity	32.4 (29.7+2.8)
CMG_CONTRACTS	ODSSUPER	Contract Mgmt	10.0 (10.0+0.0)
FACT_ACTIVITY_FACT_...	ODSSUPER	Activity	9.9 (4.0+6.0)
CMG_CUST_MON_FEES	ODSSUPER	Contract Mgmt	9.5 (9.5+0.0)
FACT_ACTIVITY_ACT_A...	ODSSUPER	Activity	9.2 (3.7+5.5)
MV_FACT_ACT_EXP_SMP	ODSSUPER	Samples	8.6 (8.3+0.3)
TMP_SMP_FACT	ODSSUPER	Samples	8.3 (8.3+0.0)
ACT_ACTIVITY_SFA	ODSSUPER	Activity	6.9 (6.3+0.7)
SAMPLE_QCI_STG	ODSSUPER	Samples	5.4 (5.4+0.0)
TMP_ACT	ODSSUPER	Other	5.0 (4.7+0.3)
CONTRACT_MANAGEMENT	ODSSUPER	Contract Mgmt	4.3 (2.9+1.4)
SEGMENTATION	ODSSUPER	Segmentation	4.2 (2.5+1.7)
TMP_JAN_ACT	ODSSUPER	Other	3.0 (2.8+0.2)
BKP_06_09_SEGMENTAT...	ODSSUPER	Backup	3.0 (3.0+0.0)
CONT_MGMT_CUST_MON_...	ODSSUPER	Contract Mgmt	2.7 (2.0+0.7)
ACT_SAMPLE_ITEM_DRO...	ODSSUPER	Samples	2.5 (2.5+0.0)



Maszkolási szabályok

Véglegesen anonimizálja a szenzitív adatokat, előre definiált maszkolási technikákkal, megfelelve az adatbiztonsági szabványoknak (HIPAA, PCI DSS).

New Rule: Step 1 of 2 Steps

Specify rule properties.

*Name:

Description:

*Datatype:

*Masking Type:

Standard

Credit Card

Credit Card

Email Address

Encryption

Expression

IP Address

Key

Nullification

Phone

Random

Shuffle

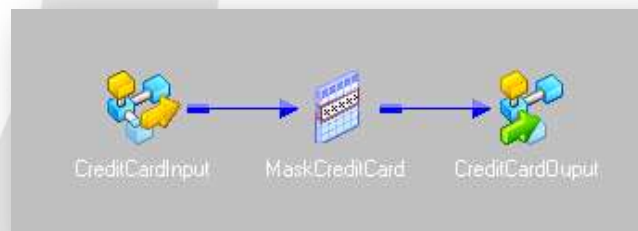
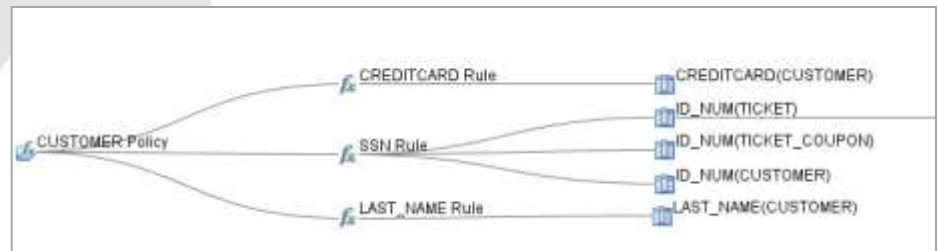
SIN

SSN

Substitution

URL

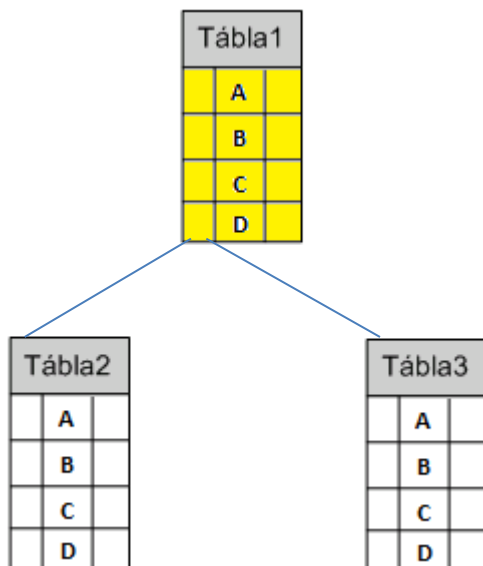
ID	Name	City	Credit Card
0964	Mike Wilson	Pleasanton	4417 9741 1949 9471
9388	Mark Moore	Modesto	4981 1341 0854 0508
2586	Rod Davis	Hartford	4298 9341 9544 9114
7310	Jeff Richards	Tampa	4198 9481 9147 0521



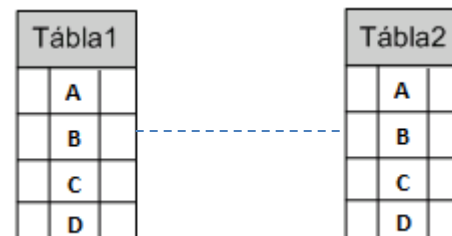
Konzisztens maszkolás

(munkafolyamaton belül)

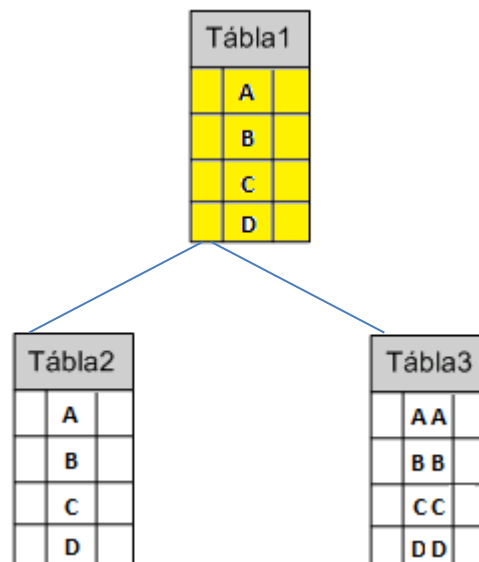
Auto cascade



Value cascade



Expression cascade



Konzisztens maszkolás

(alkalmazásokon, munkafolyamatokon keresztül is)

New Rule: Step 2 of 2 Steps

Specify masking properties.

SUBSTITUTION - Masking Parameters

Replaces a column of data with similar but unrelated data from a dictionary

Repeatable Output i

Seed:

Unique Substitution Data i

Dictionary Information:

Flat File i

Relational i

Dictionary:

File Encoding:

Serial Number Column:

Output Column:

Lookup Condition i:

Input Column:

Dictionary Column:

Exception Handling

Specify the default behavior for exception handling

Null and Empty Spaces:

Constant:

Default Server Value

Log error and continue

Treat as a value

Ignore

Error Handling:

Constant:

Default Server Value

Log Exception and Continue

Ignore and Continue

Error Out

Leading or Trailing Spaces:

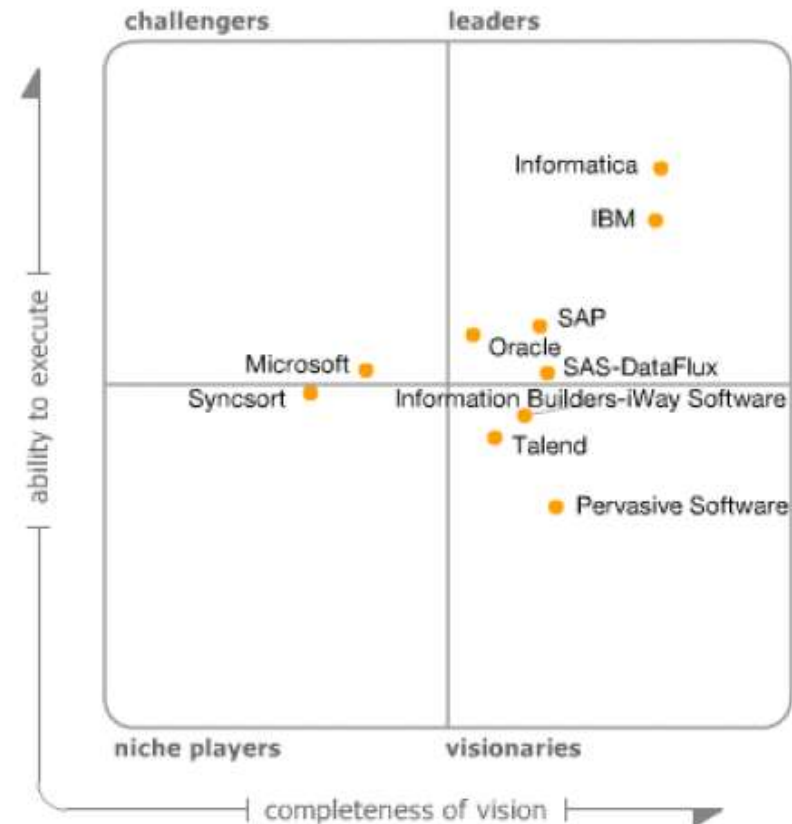
Trim

Miért jó ez a nagyvállalatoknak?

- ◉ Adatvesztés, adatlopás megelőzése, törvényi megfelelés
- ◉ Alacsonyabb tárhely és karbantartási költség
- ◉ Növekvő produktivitás:
 - a tesztkörnyezet előállítása 2-3 hónap helyett néhány nap
 - realisztikus tesztadatok -> hatékony fejlesztés

Megéri kompetensnek lenni!

- Új trend a nagyvállalatok körében, növekvő kereslet, piacképes tudás
- Innovatív technológiák, „best of breed” eszközök
- Adatbázis kezelő rendszerek ismerete előny



As of October 2012



Ügyfelek világszerte

<p>Financial Services and Insurance</p>	
<p>Government and Public Sector</p>	
<p>Healthcare and Life Sciences</p>	
<p>Manufacturing</p>	
<p>Retail and Services</p>	
<p>Telecommunications</p>	
<p>Transportation and Distribution</p>	
<p>Energy and Utilities</p>	

**Hivatalos magyarországi disztribútor és
kompetenciaközpont vagyunk.**

INFORMATICA | AUTHORIZED
DISTRIBUTOR

areus.hu/adatintegracio

informatica.com

Köszönjük a figyelmet!



Kérdések és Válaszok

