

DATA WAREHOUSES (DWH/DW) - INTRO

1. Introduction to the DWH discipline

- 2. Brief history** (Inmon & Linstedt, 2015) **and Data Architecture** (Kimball & Ross, 2016)
- 3. Dimensional Modeling Fundamentals** (Kimball & Ross, 2016)
- 4. Technical Architecture Considerations** (Kimball & Ross, 2016)
- 5. Extract Transform Load and Data Quality** (Kimball & Ross, 2016)
- 6. DWH Lifecycle** (Kimball & Ross, 2013)
- 7. Trends in the evolution of DWH:**
 - Extended RDBMS Architecture** (Kimball & Ross, 2013).
 - Pushing into the Future** (Reeves, 2009).
 - DWH 1.0 vs. 2.0** (Krishnan, 2013)

Assessment

Type of activity	10.1 Assessment criteria	10.2 Assessment methods	10.3 Share of final grade
Developing support components for a DWH application prototype	Real-world application, complexity, validity and originality	Six face-to-face lab presentations of the homework about: the design of DWHcubes & DM models implemented, the support MDX & DMX queries and the code execution behind .NET forms in weeks No.: 3, 5, 7, 9, 11 and 13.	40% (6 * 6.66%)
Theoretical presentations during lecture hours (see those 14 themes on domains at section 8.1, pp.2)	Format, consistent pro-or-cons arguments, originality of comments and conclusions	Theoretical presentation and responses to questions	20%
Theoretical exam	Knowledge about DWH theory, real-world scenarios and personal implementations	Final theoretical test with at least three open questions	40%
10.6 Minimum performance standard			
<ul style="list-style-type: none"> Design and implement a DWH cube, a DM model, two interfaces to programmatically connect to and query them by using MDX and DMX queries, alternative examples (user mode) with Microsoft Excel add-ins; Each master student must create, present and answer questions for a part of those at least 10 specific presentations (14 themes on domains - section 8.1, pp.2) during lecture hours; The average grade for all those six face-to-face lab presentations (section 8.2, pp.3) of the homework is greater than or equal to 5; The grade for the final theoretical test must be greater than or equal to 5. 			

Planning those at least 10 presentations (course hours, 20% in final grades) of master students (teams) about:

(1.) retail sales, (2.) inventory, (3.) procurement, (4.) order management, (5.) accounting, (6.) CRM, (7.) HRM, (8.) financial services, (9.) telecommunications, (10.) transportation, (11.) education, (12.) healthcare, (13.) e-commerce, (14.) insurance

details: **Kimball&Ross,2013.**






by e-mail: dan.homocianu@gmail.com

(subject: *DWH course pres. theme TITLE.. on DATE..*

content: *team: MEMBER1, MEMBER2, MEMBER3)*

Each master student (team member) must present a part of his team's support presentation for at least 5-10 minutes and answer questions.

Semestrul al II-lea

-  **15 februarie – 30 mai** – 14 săptămâni – Activitate didactică
-  **3 mai – 9 mai** – 1 săptămână liberă în perioada sărbătorilor de Paște
-  **31 mai – 13 iunie** – 2 săptămâni – Evaluare (sesiune)
-  **14 iunie – 4 iulie** – 3 săptămâni practică de specialitate și evaluare. În această perioadă, fiecare facultate organizează o săptămână de evaluare (reexaminări în vederea promovării sau măririi notei).
-  **5 iulie – 30 septembrie** – Vacanța de vară

În perioada 1-10 septembrie, fiecare facultate organizează o sesiune de evaluare (reexaminări în vederea promovării sau maririi notei).

Home

Daniel Homocianu - Outl

Inbox - dan.homocianu@

EnformationProfil acces

Web of Science [v.5.32] -

data warehouse - Căutare

Not secure | apps.webofknowledge.com.am.e-nformation.ro/RAMore.do?product=WOS&search_mode=GeneralSearch&SID=C4tou3di4U2ke1oinKO&qid=11&ra_mode=more&ra_name=PublicationYe... ☆

Web of Science | InCites | Journal Citation Reports | Essential Science Indicators | EndNote | Publons | Kopernio | Sign In | Help | English

Web of Science

Clarivate Analytics

Search | Tools | Searches and alerts | Search History | Marked List

Results: ...
(from Web of Science Core Collection)

You searched for: TOPIC: ("Business Intelligence") ...More

Create Alert

Refine Results

Search within results for...

Filter results by:

- Highly Cited in Field (11)
- Open Access (407)
- Associated Data (1)

Refine

Publication Years

Web of Science Categories

Publication Years | Refine | Exclude | Cancel | Sort these by: Alphabetical

The first 100 Publication Years (by record count) are shown. For advanced refine options, use Analyze results.

<input type="checkbox"/> 2019 (17)	<input type="checkbox"/> 2010 (211)	<input type="checkbox"/> 2001 (14)	<input type="checkbox"/> 1991 (2)
<input type="checkbox"/> 2018 (314)	<input type="checkbox"/> 2009 (220)	<input type="checkbox"/> 2000 (20)	<input type="checkbox"/> 1990 (8)
<input type="checkbox"/> 2017 (428)	<input type="checkbox"/> 2008 (192)	<input type="checkbox"/> 1999 (6)	<input type="checkbox"/> 1988 (1)
<input type="checkbox"/> 2016 (474)	<input type="checkbox"/> 2007 (121)	<input type="checkbox"/> 1998 (8)	<input type="checkbox"/> 1986 (4)
<input type="checkbox"/> 2015 (468)	<input type="checkbox"/> 2006 (80)	<input type="checkbox"/> 1997 (7)	<input type="checkbox"/> 1985 (2)
<input type="checkbox"/> 2014 (356)	<input type="checkbox"/> 2005 (78)	<input type="checkbox"/> 1996 (2)	<input type="checkbox"/> 1983 (2)
<input type="checkbox"/> 2013 (312)	<input type="checkbox"/> 2004 (46)	<input type="checkbox"/> 1995 (5)	<input type="checkbox"/> 1976 (1)
<input type="checkbox"/> 2012 (276)	<input type="checkbox"/> 2003 (32)	<input type="checkbox"/> 1994 (3)	<input type="checkbox"/> 1975 (1)
<input type="checkbox"/> 2011 (236)	<input type="checkbox"/> 2002 (34)	<input type="checkbox"/> 1993 (1)	

Refine | Exclude | Cancel | Sort these by: Alphabetical

3,982 results

Home

Daniel Homocianu - Outl

Inbox - dan.homocianu@

EnformationProfil acces

Web of Science [v.5.32]

data warehouse - Căutare

Not secure | apps.webofknowledge.com.am.e-nformation.ro/?product=WOS&search_mode=GeneralSearch&SID=C4tou3di4U2ke1oinKO&qid=10&ra_mode=more&ra_name=PublicationYe...

Web of Science | InCites | Journal Citation Reports | Essential Science Indicators | EndNote | Publons | Kopernio

Sign In | Help | English

Web of Science

Clarivate Analytics

Search

Tools | Searches and alerts | Search History | Marked List

Results: ...
(from Web of Science Core Collection)

You searched for: TOPIC: ("decision s
upport systems") ...More

Create Alert

Refine Results

Search within results for...

Filter results by:

☐ Highly Cited in Field (44)

☐ Hot Papers in Field (1)

☐ Open Access (1,931)

☐ Associated Data (20)

Refine

Publication Years

Publication Years

Refine Exclude Cancel Sort these by: Alphabetical

The first 100 Publication Years (by record count) are shown. For advanced refine options, use Analyze results.

<input type="checkbox"/> 2019 (67)	<input type="checkbox"/> 2008 (479)	<input type="checkbox"/> 1997 (297)	<input type="checkbox"/> 1986 (28)
<input type="checkbox"/> 2018 (706)	<input type="checkbox"/> 2007 (500)	<input type="checkbox"/> 1996 (202)	<input type="checkbox"/> 1985 (27)
<input type="checkbox"/> 2017 (936)	<input type="checkbox"/> 2006 (418)	<input type="checkbox"/> 1995 (216)	<input type="checkbox"/> 1984 (28)
<input type="checkbox"/> 2016 (880)	<input type="checkbox"/> 2005 (362)	<input type="checkbox"/> 1994 (209)	<input type="checkbox"/> 1983 (22)
<input type="checkbox"/> 2015 (871)	<input type="checkbox"/> 2004 (376)	<input type="checkbox"/> 1993 (180)	<input type="checkbox"/> 1982 (12)
<input type="checkbox"/> 2014 (657)	<input type="checkbox"/> 2003 (326)	<input type="checkbox"/> 1992 (206)	<input type="checkbox"/> 1981 (15)
<input type="checkbox"/> 2013 (702)	<input type="checkbox"/> 2002 (273)	<input type="checkbox"/> 1991 (152)	<input type="checkbox"/> 1980 (11)
<input type="checkbox"/> 2012 (615)	<input type="checkbox"/> 2001 (278)	<input type="checkbox"/> 1990 (58)	<input type="checkbox"/> 1979 (7)
<input type="checkbox"/> 2011 (629)	<input type="checkbox"/> 2000 (305)	<input type="checkbox"/> 1989 (42)	<input type="checkbox"/> 1978 (4)
<input type="checkbox"/> 2010 (563)	<input type="checkbox"/> 1999 (262)	<input type="checkbox"/> 1988 (35)	<input type="checkbox"/> 1977 (3)
<input type="checkbox"/> 2009 (600)	<input type="checkbox"/> 1998 (327)	<input type="checkbox"/> 1987 (29)	<input type="checkbox"/> 1976 (1)

Refine Exclude Cancel Sort these by: Alphabetical

12,916 results

Home

Daniel Homocianu - Outl

Inbox - dan.homocianu@

EnformationProfil acces

Web of Science [v.5.32] -

data warehouse - Căutare

Not secure | apps.webofknowledge.com.am.e-nformation.ro/RAMore.do?product=WOS&search_mode=GeneralSearch&SID=C4tou3di4U2ke1oinKO&qid=9&ra_mode=more&ra_name=PublicationYea...

Web of Science | InCites | Journal Citation Reports | Essential Science Indicators | EndNote | Publons | Kopernio | Sign In | Help | English

Web of Science

Clarivate Analytics

Search | Tools | Searches and alerts | Search History | Marked List

Results: ...
(from Web of Science Core Collection)

You searched for: TOPIC: ("data ware house") ...More

Create Alert

Refine Results

Search within results for...

Filter results by:

- Highly Cited in Field (13)
- Hot Papers in Field (2)
- Open Access (705)
- Associated Data (18)

Refine

Publication Years

Publication Years | Refine | Exclude | Cancel | Sort these by: Alphabetical

The first 100 Publication Years (by record count) are shown. For advanced refine options, use Analyze results.

<input type="checkbox"/> 2019 (19)	<input type="checkbox"/> 2011 (250)	<input type="checkbox"/> 2004 (190)	<input type="checkbox"/> 1997 (76)
<input type="checkbox"/> 2018 (254)	<input type="checkbox"/> 2010 (272)	<input type="checkbox"/> 2003 (170)	<input type="checkbox"/> 1996 (30)
<input type="checkbox"/> 2017 (417)	<input type="checkbox"/> 2009 (332)	<input type="checkbox"/> 2002 (133)	<input type="checkbox"/> 1995 (6)
<input type="checkbox"/> 2016 (414)	<input type="checkbox"/> 2008 (300)	<input type="checkbox"/> 2001 (130)	<input type="checkbox"/> 1993 (2)
<input type="checkbox"/> 2015 (417)	<input type="checkbox"/> 2007 (241)	<input type="checkbox"/> 2000 (117)	<input type="checkbox"/> 1992 (1)
<input type="checkbox"/> 2014 (364)	<input type="checkbox"/> 2006 (234)	<input type="checkbox"/> 1999 (76)	<input type="checkbox"/> 1991 (1)
<input type="checkbox"/> 2013 (285)	<input type="checkbox"/> 2005 (214)	<input type="checkbox"/> 1998 (110)	<input type="checkbox"/> 1989 (1)
<input type="checkbox"/> 2012 (296)			

Refine | Exclude | Cancel | Sort these by: Alphabetical

5,352 results

375

376

377

Reporting as Democratization

In turn, democratizing your data does not render a reporting program obsolete—it enhances it. Consider a system architected as in Figure 16-2.

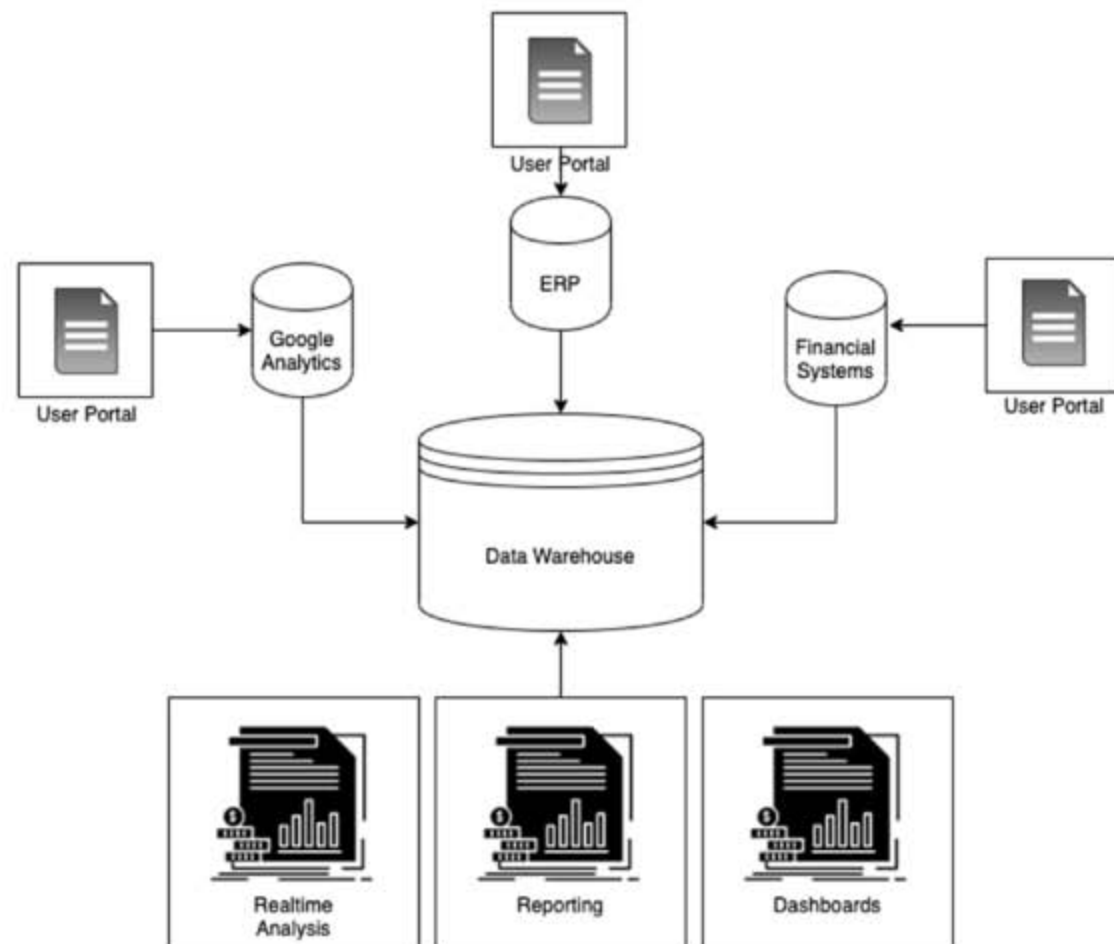
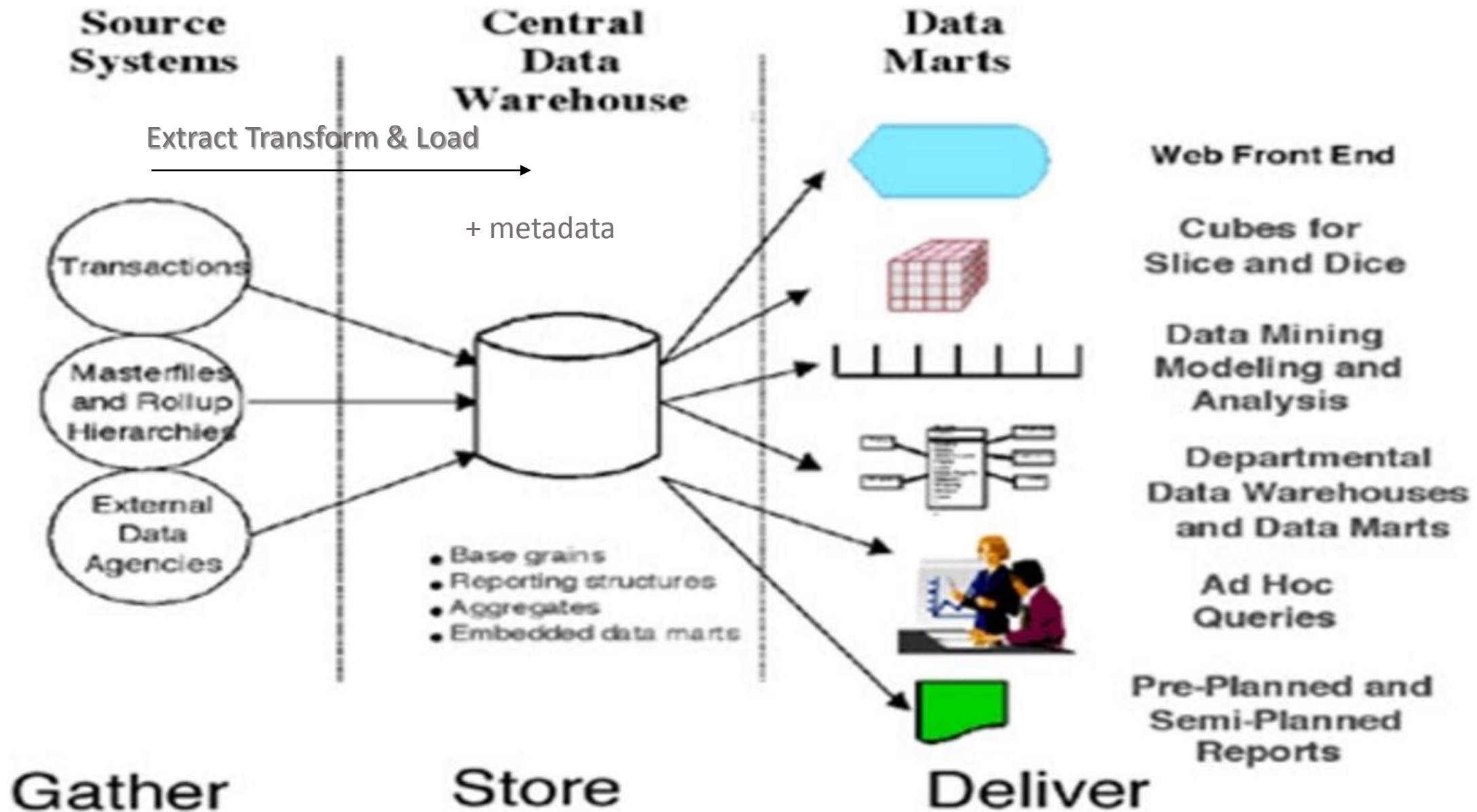
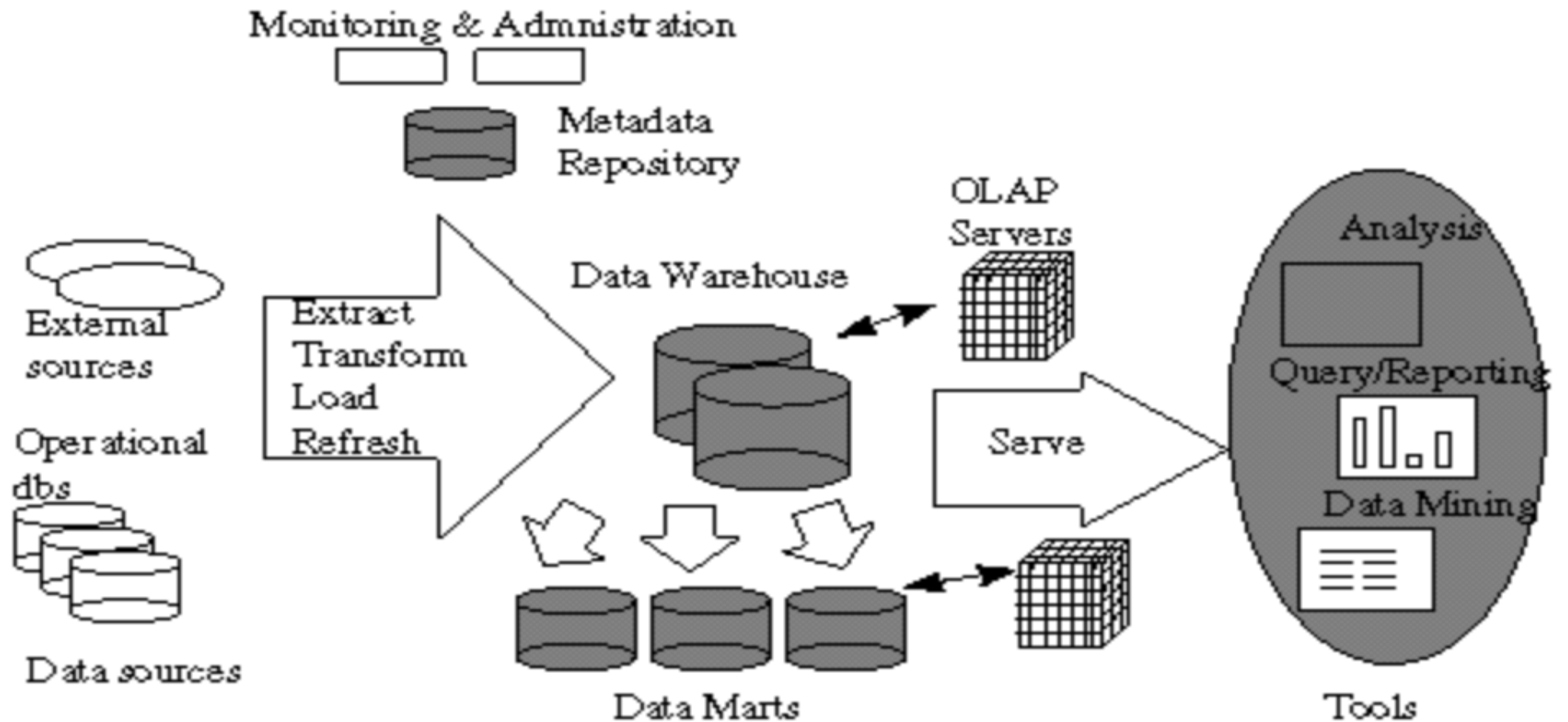


Figure 16-2. Architectural diagram of systems

DWH – overview

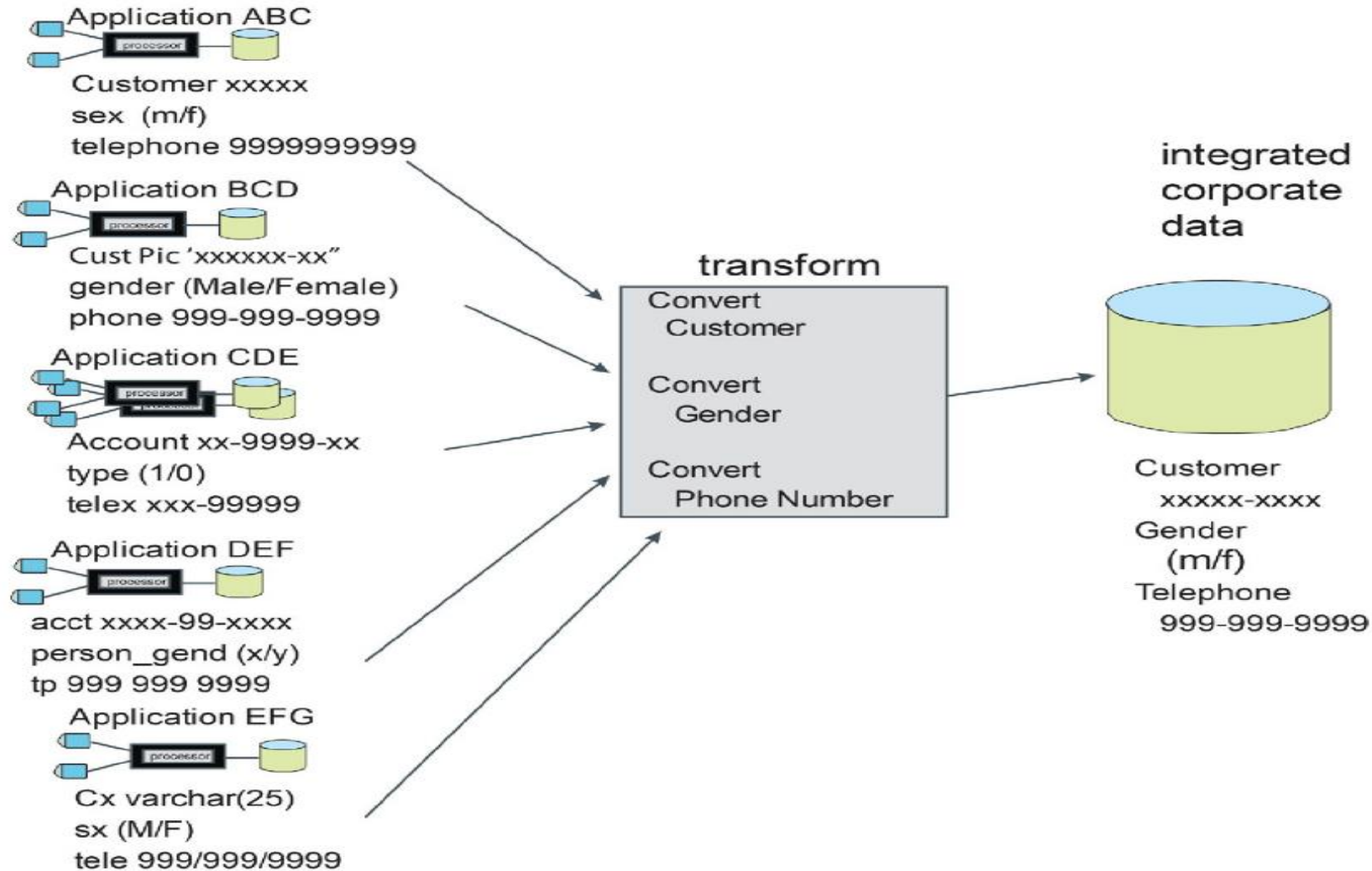


DWH – overview



Extract Transform & Load (ETL)

Inmon & Linstedt, 2015



Extract Transform & Load (ETL)

Nesstar WebView | Inbox - dan.homocianu@gmail.com | InceptionStudy source xls - C x

127.0.0.1:3333/project?project=1975359136494

Google refine InceptionStudy source xls Permalink

Open... Export Help

Facet / Filter

Filter:

0. Create project

1. Split 12 cells into several columns

2. Rename column

3. Rename column

4. Transpose cells in columns starting with BACAU: Nu are clienti vechi into rows in one new column named mix

5. Split 432 cells into columns by

6. Rename column

7. Rename column

8. Rename column

Extract Operation History

Extract and save parts of your operation history as JSON that you can apply to this or other projects in the future.

- ☒ Split column Column by separator
- ☒ Rename column Column 1 to CA
- ☒ Rename column Column 2 to Domeniu
- ☒ Transpose cells in columns starting with BACAU: Nu are clienti vechi into rows in one new column named mix
- ☒ Split column mix by separator
- ☒ Rename column mix 1 to Judet
- ☒ Rename column mix 2 to Tip_CL
- ☒ Rename column mix 3 to Nr.rasp.

Select All Unselect All

Close

```
[
  {
    "op": "core/column-split",
    "description": "Split column Column by separator",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "Column",
    "guessCellType": true,
    "removeOriginalColumn": true,
    "mode": "separator",
    "separator": ":",
    "regex": false,
    "maxColumns": 2
  },
  {
    "op": "core/column-rename",
    "description": "Rename column Column 1 to CA",
    "oldColumnName": "Column 1",
    "newColumnName": "CA"
  },
  {
    "op": "core/column-rename",
    "description": "Rename column Column 2 to Domeniu",
    "oldColumnName": "Column 2",
    "newColumnName": "Domeniu"
  }
]
```

Uses With ETL

tfmorris edited this page on Jan 9, 2013 · 2 revisions

Is OpenRefine an ETL tool ? or can it be used in batch operations or job control ?

OpenRefine is a discovery and cleanup tool mostly, as it stands now at release 2.5. It also has the benefit of aligning, reconciling, and exporting data, to Freebase or somewhere else. Some folks in the community have expressed the idea or need for batching transform operations or evolving OpenRefine into an automated pipeline process.

One feature of OpenRefine that folks find intriguing is it's ability to display inconsistencies in your data and clean it up easily.

Usage with Enterprise data or large databases,

- For instance, you might leverage OpenRefine to analyze and find faults in the output of some of your stored procedures within an existing database and then fix them outright.
- Or, you might make slight job changes inside your program or ETL tool after you discover (using OpenRefine) that your output isn't always correct, and the job needs to be modified.

Also, if your interest is transforming large amounts of data, you might also want to look at existing database ETL tools (Extract, Transform, Load) which some we have listed here: [Related Software](#)

(Currently, using OpenRefine with extremely large amounts of data and millions of rows does require plenty of RAM and Java 64 bit, so you might find a hard time absorbing and utilizing, for example, a 1 billion row Enterprise database table.)

We're interested in your ideas and comments ! and patches are always welcome !

▼ Pages 101

[Home](#)

[Architecture](#)

[Back Up OpenRefine Data](#)

[Broker Protocol](#)

[Building OpenRefine From Source](#)

[Cell Editing](#)

[Changes for 2.0](#)

[Changes for 2.1](#)

[Changes for 2.5](#)

[Changes for 2.6](#)

[Changes for 3.1](#)

[Client Side Architecture](#)

[Clustering](#)

[Clustering In Depth](#)

[Column Editing](#)

Show 86 more pages...



OpenRefine

A free, open source,
powerful tool for working
with messy data



Home
Community
Documentation
Download
Contact Us
Blog

Enhanced with Java profiler



Documentation for Developers

- [Wiki for Developers](#)
- [Developer Discussion List](#)

Documentation for User

Official Documentation

- [FAQ](#)
- [Official Documentation Wiki](#)

Support and Questions

- [OpenRefine Support Mailing List](#)
- [StackOverflow Tag](#)
- [File a Bug or Feature Request](#)

Online Courses

- [Cognitive Class: OpenRefine 101 \(with video\)](#)
- [DataCarpentry: OpenRefine for Social Science Data](#)
- [DataCarpentry: OpenRefine for Ecology](#)
- [The Programming Historian: Cleaning Data with OpenRefine \(en\) and \(es\)](#)
- [The Programming Historian: Fetching and Parsing Data from the Web with OpenRefine](#)
- [Nettoyer et préparer des données avec OpenRefine : atelier pour les journées du consortium MASA - 14/11/2018 \(fr\)](#)
- [Curated Tutorial List](#)

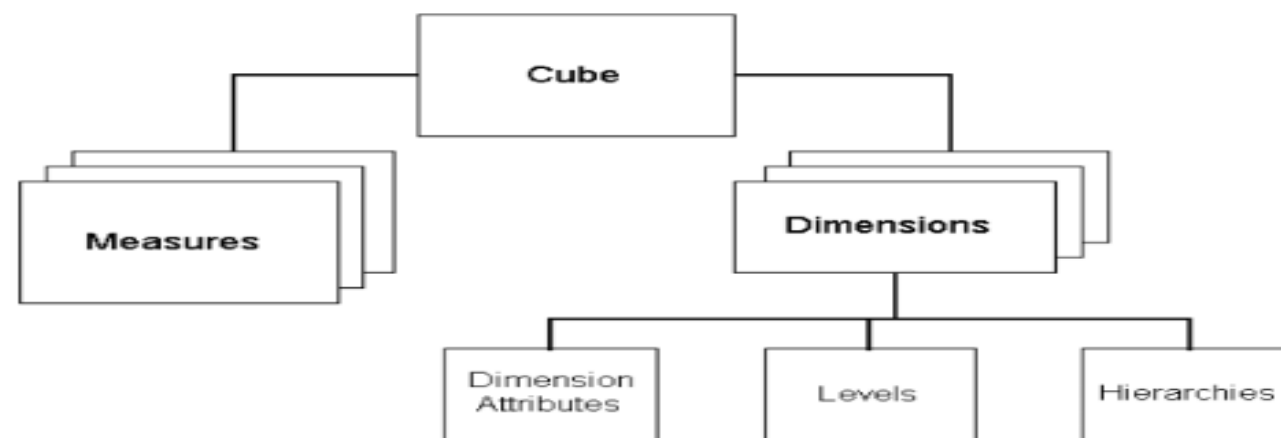
Using OpenRefine - The Book



Using OpenRefine by Ruben Verborgh, Max De Wilde offers anyone a great introduction to OpenRefine. Organized by recipes with hands on examples, the book covers the following topics:

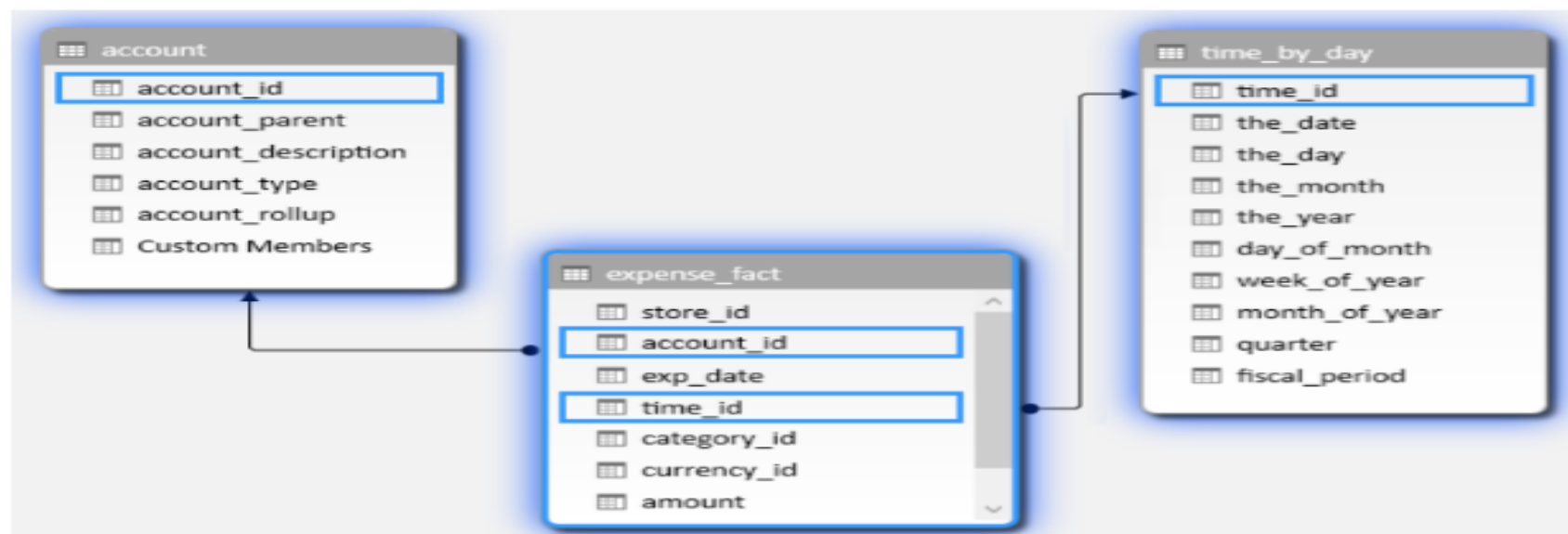
1. Import data in various formats

Figure 1. The diagram of the multidimensional model



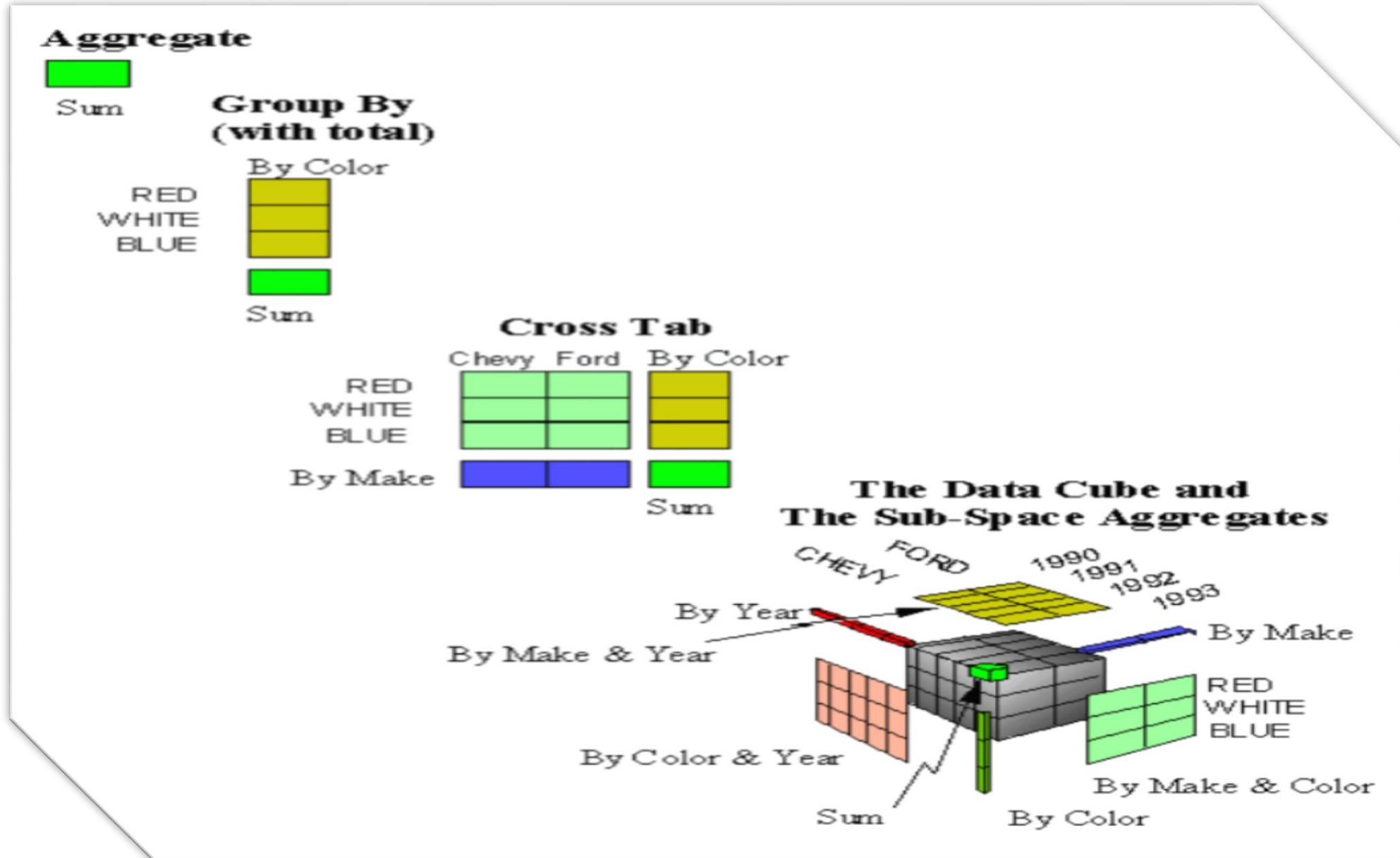
Source: <http://web.stanford.edu/dept/itss/docs/oracle/10g/olap.101/b10333/multimodel.htm>

Figure 2. Example of the simplest scheme (star) of relational implementation of the multi-dimensional model with three tables containing attributes about account, expenses (expense_fact) and time (time_by_day)



Source: The video tutorial developed by the author: <https://www.youtube.com/watch?v=zR370Z18mPY>

THE DATA CUBE



Data Marts

■ **Data mart** has been replaced with *business process dimensional model*, *business process subject area*, or just *subject area*, depending on the context. (Kimball&Ross,2016)

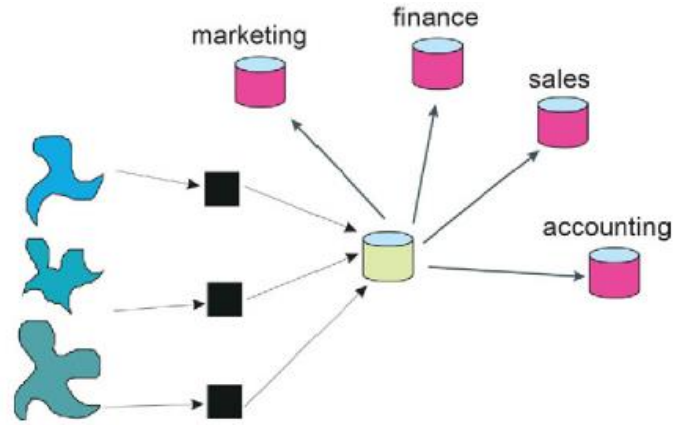


Figure 3.4.3

Figure 3.4.3 shows that each different organization has its own perspective of data. All data begins with the granular data found in the data warehouse. But the different interpretation of data is created for each different department from that granular data. Note that while each department has its own interpretation of data, *all* data is still reconcilable to the common data warehouse. Also note that if it is desired to build a new data mart, the data in the data warehouse is available for the immediate construction of a new perspective of data.

The database design for the data mart that is optimal is the dimensional model. Figure 3.4.4 shows that there is a different and unique dimensional model for each data mart.

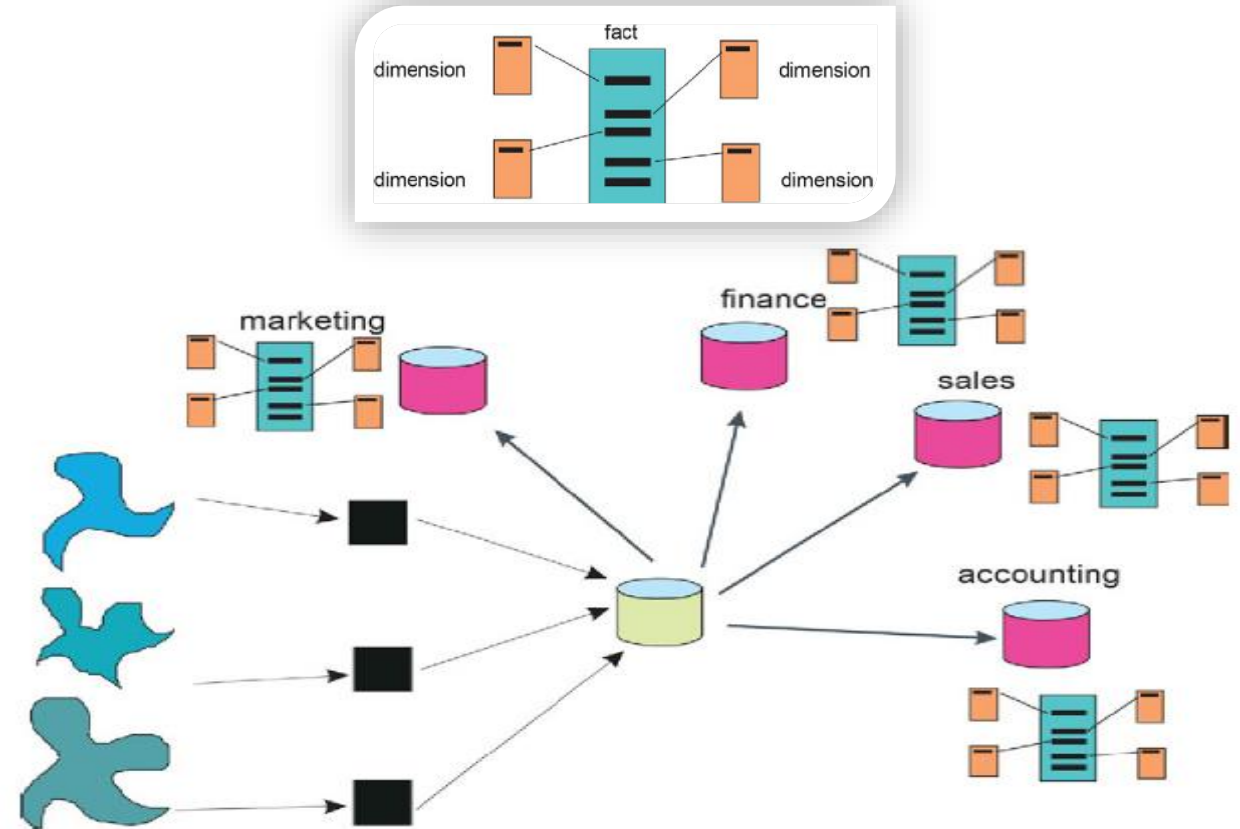


Figure 3.4.4

(Inmon & Linstedt, 2015)

Business Processes & Common Dimensions

(Kimball Group)

BUSINESS PROCESSES	COMMON DIMENSIONS						
	Date	Product	Warehouse	Store	Promotion	Customer	Employee
Issue Purchase Orders	X	X	X				
Receive Warehouse Deliveries	X	X	X				X
Warehouse Inventory	X	X	X				
Receive Store Deliveries	X	X	X	X			X
Store Inventory	X	X		X			
Retail Sales	X	X		X	X	X	X
Retail Sales Forecast	X	X		X			
Retail Promotion Tracking	X	X		X	X		
Customer Returns	X	X		X	X	X	X
Returns to Vendor	X	X		X			X
Frequent Shopper Sign-Ups	X			X		X	X

Business Process / Event	COMMON DIMENSIONS														
	Time	Customer	Service	Rate Category	Local Svc Provider	Calling Party	Called Party	Long Dist Provider	Internal Organization	Employee	Location	Equipment Type	Supplier	Item Shipped	Account Status
Customer Billing	X	X	X	X	X		X			X					X
Service Orders	X	X	X		X		X	X	X	X	X				X
Trouble Reports	X	X	X		X	X	X	X	X	X	X	X	X	X	X
Yellow Page Ads	X	X		X	X			X	X	X					X
Customer Inquiries	X	X	X	X	X	X	X	X	X	X					X
Promotions & Communication	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Billing Call Detail	X	X	X	X	X	X	X	X		X	X	X	X	X	X
Network Call Detail	X	X	X	X	X	X	X	X		X	X	X	X	X	X
Customer Inventory	X	X	X	X	X		X	X		X	X	X	X	X	X
Network Inventory	X		X					X	X	X	X	X	X		
Real Estate	X							X	X	X	X				
Labor & Payroll	X							X	X	X					
Computer Charges	X	X	X		X		X	X	X	X	X	X	X	X	
Purchase Orders	X							X	X	X	X	X	X	X	
Supplier Deliveries	X							X	X	X	X	X	X	X	

Web-based OLAP & Questionnaire variables

Nesstar WebView Inbox - dan.homocianu@gmail.com

www.cityaudit.ro/webview/index.jsp?mode=documentation&submode=catalog&catalog=http%3A//89.38.215.251% ☆ ☰

cityaudit

Acest site este cofinanțat din Fondul Social European prin Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007-2013

DESCRIPTION TABULATION ANALYSIS Home Contact

ue analiza, cu date reale din
piața
domeniu activitate -
extins
Care este cifra de afaceri
a companiei?
Si care este numarul de
angajati ai companiei?
JUDET
- VARIABLE CROSS -
REGIUNE
CIFRA AFACERI
NUMAR ANGAJATI
TIP CLIENT
PIATA LOCALA
PIATA REGIONALA
PIATA NATIONALA
PIATA EXTERNA
DOMENIU ACTIVITATE
DACA AU PERSOANA DE
MARKETING DEDICATA
DACA VINDE LA
PERSOANE FIZICE SAU
JURIDICE

Turism
Auto
Lifestyle
Haine copii
Caracteristici piata de consultanta si
asistenta juridica in problema identitatii
marcii_City55
Caracteristici piata financiar-
contabila_City54
Alegerea furnizorului de instalatii
electrice si telecomunicatii_City46
Caracteristici piata produse
alimentare fara gluten_City48
Consumul de peste in Romania si
particularitatile crapului_City50
Cursuri la bi se ne City15

Dataset: InceptionStudy
Variable JUDET: JUDET

Values Categories N

1	BACAU	108	10.4%
2	BIHOR	106	10.2%
3	BISTRITA NASAUD	42	4.0%
4	BOTOSANI	32	3.1%
5	CLUJ	234	22.5%
6	IASI	126	12.1%
7	MARAMURES	83	8.0%
8	NEAMT	83	8.0%
9	SALAJ	35	3.4%
10	SATU MARE	41	3.9%
11	SUCEAVA	106	10.2%
12	VASLUI	45	4.3%

SUMMARY STATISTICS
Valid cases 1041
Missing cases 0
This variable is numeric

0:21 / 4:00

bit.ly/2T2INKY

Web-based OLAP & Questionnaire variables

Nesstar WebView x Inbox - dan.homocianu@gm x

www.cityaudit.ro/webview/index.jsp?mode=documentation&submode=catalog&catalog=http%3A//89.38.215.251%

cityaudit

Acest site este cofințat din Fondul Social European prin Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007-2013

DESCRIPTION TABULATION ANALYSIS Home Contact

de analiza, cu date reale din piața

- ☒ domeniu activitate - extins
- ☒ Care este cifra de afaceri a companiei ?
- ☒ Si care este numarul de angajati ai companiei ?
- ☒ JUDET
- ☒ - VARIABILE CROSS -
- ☒ REGIUNE
- ☒ CIFRA AFACERI
- ☒ NUMAR ANGAJATI
- ☒ TIP CLIENT
- ☒ PIATA LOCALA
- ☒ PIATA REGIONALA
- ☒ PIATA NATIONALA
- ☒ PIATA EXTERNA
- ☒ DOMENIU ACTIVITATE
- ☒ DACA AU PERSOANA DE MARKETING DEDICATA
- ☒ DACA VINDE LA PERSOANE FIZICE SAU JURIDICE

Turism

Auto

Lifestyle

Haine copii

Caracteristici piata de consultanta si asistenta juridica in problema identitatii marcii_City55

Caracteristici piata financiar-contabila_City54

Alegerea furnizorului de instalatii electrice si telecomunicatii_City46

Caracteristici piata produse alimentare fara gluten_City48

Consumul de peste in Romania si particularitatile crapului_City50

Dataset: InceptionStudy

Variable TIPCLIENT: TIP CLIENT

Values	Categories	N	
0	Nu are clienti vechi	34	3.4%
1	Majoritatea clienti noi	137	13.8%
2	Majoritatea clienti vechi	819	82.7%
99		51	

SUMMARY STATISTICS

Valid cases 990

Missing cases 51

This variable is numeric

0:27 / 4:00

bit.ly/2T2INKY

Web-based OLAP & Questionnaire variables

Nesstar WebView x Inbox - dan.homocianu@gm x

www.cityaudit.ro/webview/index.jsp?mode=documentation&submode=catalog&catalog=http%3A//89.38.215.251%

cityaudit

Accest site este cofinatat din Fondul Social European prin Programul Operational Sectorial Dezvoltarea Resurselor Umane 2007-2013

DESCRIPTION TABULATION ANALYSIS

Dataset: InceptionStudy

Variable CIFAF: CIFRA AFACERI

Values Categories N

1	Sub 100 mii euro	602	57.8%
2	Intre 101 - 200 mii euro	185	17.8%
3	Peste 200 mii euro	119	11.4%
99	Nu stiu / nu raspund	135	13.0%

SUMMARY STATISTICS

Valid cases 1041

Missing cases 0

This variable is numeric

bit.ly/2T2INKY

de analiza, cu date reale din piata

- domeniu activitate - extins
- Care este cifra de afaceri a companiei?
- Si care este numarul de angajati ai companiei?
- JUDET
- VARIABLE CROSS - REGIUNE
- CIFRA AFACERI
- NUMAR ANGAJATI
- TIP CLIENT
- PIATA LOCALA
- PIATA REGIONALA
- PIATA NATIONALA
- PIATA EXTERNA
- DOMENIU ACTIVITATE
- DACA AU PERSOANA DE MARKETING DEDICATA
- DACA VINDE LA PERSOANE FIZICE SAU JURIDICE

Turism

Auto

Lifestyle

Haine copii

Caracteristici piata de consultanta si asistenta juridica in problema identitatii marci City55

Caracteristici piata financiar-contabila City54

Alegerea furnizorului de instalatii electrice si telecomunicatii City46

Caracteristici piata produse alimentare fara gluten City48

Consumul de peste in Romania si particularitatile crapului City50

Cursuri la bi si ne bi City51

0:84 / 4:00

[Nesstar WebView](#)
[Inbox - dan.hornocianu@gmail.com](#)

[www.cityaudit.ro/webview/index.jsp?mode=documentation&submode=catalog&catalog=http%3A//89.38.215.251%](#)

[cityaudit](#)

[DESCRIPTION](#)
[TABULATION](#)
[ANALYSIS](#)

[Home](#)
[Contact](#)

[Export table to spreadsheet](#)

Dataset: InceptionStudy

CIFRA AFACERI: Categories
 DOMENIU ACTIVITATE: Categories
 JUDET: Categories

TIP CLIENT: Categories
 Type: Raw numbers

JUDET		BACAU			BIHOR			BISTRITA NASAU	
TIP CLIENT		Nu are clienti vechi	Majoritatea clienti noi	Majoritatea clienti vechi	Nu are clienti vechi	Majoritatea clienti noi	Majoritatea clienti vechi	Nu are clienti vechi	Majoritatea clienti noi
CIFRA AFACERI	DOMENIU ACTIVITATE								
Sub 100 mii euro	Comert	2	2	19	1	2	11	0	1
	Industrie	2	2	8	0	0	10	0	1
	Servicii	0	2	27	2	2	22	0	1
Intre 101 - 200 mii euro	Comert	0	1	8	0	0	9	0	0
	Industrie	0	0	3	0	0	4	0	0
	Servicii	0	2	5	1	1	6	0	0
Peste 200 mii euro	Comert	0	0	3	0	2	7	0	1
	Industrie	0	1	3	0	0	3	0	1
	Servicii	0	0	2	0	0	2	0	0
Nu stiu / nu raspund	Comert	0	0	5	0	1	8	0	1
	Industrie	1	2	1	1	0	2	0	0
	Servicii	0	0	2	1	2	3	0	1
N=		5	12	86	6	10	87	0	7

[bit.ly/2T2INKY](#)

JUDET		BACAU			BIHOR			BISTRITA NAS	
TIP CLIENT		Nu are clienti vechi	Majoritatea clienti noi	Majoritatea clienti vechi	Nu are clienti vechi	Majoritatea clienti noi	Majoritatea clienti vechi	Nu are clienti vechi	Majoritatea clienti noi
CIFRA AFACERI	DOMENIU ACTIVITATE								
Sub 100 mii euro	Comert	2	2	19	1	2	11	0	1
	Industrie	2	2	8	0	0	10	0	1
	Servicii	0	2	27	2	2	22	0	1
Intre 101 - 200 mii euro	Comert	0	1	8	0	0	9	0	0
	Industrie	0	0	3	0	0	4	0	0
	Servicii	0	2	5	1	1	6	0	0
Peste 200 mii euro	Comert	0	0	3	0	2	7	0	1
	Industrie	0	1	3	0	0	3	0	1
	Servicii	0	0	2	0	0	2	0	0
Nu stiu / nu raspund	Comert	0	0	5	0	1	8	0	1
	Industrie	1	2	1	1	0	2	0	0
	Servicii	0	0	2	1	2	3	0	1
N=		5	12	86	6	10	87	0	7

Web-based OLAP & Questionnaire variables

Nesstar WebView

← → ↺ Ww

cityau

bit.ly/2T2INKY

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	InceptionStudy												
2	Raw numbers												
3													
4	JUDET		BACAU			BIHOR				BISTRITA NASAUD			BOTOSANI
5	TIP CLIENT		Nu are cli	Majoritate	Majoritate	Nu are cli	Majoritate	Majoritate	Nu are cli	Majoritate	Majoritate	Nu are cli	Majoritate
6	CIFRA AF. DOMENIU ACTIVITATE												
7	Sub 100 m	Comert	2	2	19	1	2	11	0	1	7	1	0
8		Industrie	2	2	8	0	0	10	0	1	2	0	0
9		Servicii	0	2	27	2	2	22	0	1	7	2	0
10	Intre 101 -	Comert	0	1	8	0	0	9	0	0	6	0	0
11		Industrie	0	0	3	0	0	4	0	0	0	0	0
12		Servicii	0	2	5	1	1	6	0	0	1	0	0
13	Peste 200	Comert	0	0	3	0	2	7	0	1	2	0	0
14		Industrie	0	1	3	0	0	3	0	1	4	0	0
15		Servicii	0	0	2	0	0	2	0	0	0	0	0
16	Nu stiu / n	Comert	0	0	5	0	1	8	0	1	1	0	1
17		Industrie	1	2	1	1	0	2	0	0	2	0	0
18		Servicii	0	0	2	1	2	3	0	1	0	0	0
19	N=		5	12	86	6	10	87	0	7	32	3	1
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													
31													
32													
33													
34													
35													
36													
37													
38													
39													
40													
41													

Turism
Auto
Lifestyle
Haine copii
Caracteristici p
asistenta juridic
marcii_City55
Caracteristici p
contabila_City5
Alegerea furniz
electrice si tele
Caracteristici p
alimentare fara

InceptionStudy.xls Sheet 0
1:17 / 4:00

	<u>Kimball</u>	<u>Inmon</u>
Need	Immediate	Longer time scale
Drive	Business areas	Enterprise
Budget	Smaller budget	Larger budget
Requirements	Volatile	More stable and growing
Customer	User base	Corporate
Sources	Stable	Changeable
Startup cost	Lower	Higher
Projects	Same cost as start up	Cheaper than start up

Common Goals

- Inmon – stresses on iterative approach
- Kimball – stresses on integration for consistency

Approaches

- Bill Inmon
 - Building large centralized enterprise-wide DW (top-down)
- Ralph Kimball
 - Integrating data marts for consistency through an information bus (bottom-up)

Structure

- Bill Inmon
 - Relational model(3NF)
- Ralph Kimball
 - Multidimensional(star schema & snowflakes)

Some benefits of DWH

Inmon & Linstedt, 2015

Some of the many benefits of having a data warehouse include:

- Data is immediately available to the analyst. The data exists in a data warehouse and awaits analysis. There is no integration work required in order to begin analysis.
- Data is integrated consistently for all analysts. One analyst does not integrate data one way and another analyst does not integrate data another way.
- Data reconciliation is a real possibility. If there is a discrepancy in the results achieved by any two analysts, it is a simple matter to do a reconciliation.
- If an entirely new analysis needs to be built, there is a foundation of data on which to build.
- If it becomes necessary to do compliance or an audit, there is a believable foundation that stands ready for analysis.

Some benefits of DWH: support for OLAP & dashboards (with KPIs)

[store_id] - X ✓ f total_gen_and_adm_amount:=CALCULATE(SUM(expense_fact[amount]), expense_fact[account_id]=4100)									
store_id	account_id	exp_date	tim...	category_id	currency_id	amount	Add Col		
24	4200	9/1/1998 12...	975	ACTUAL	1	\$220.00			
24	4200	10/1/1998 1...	1005	ACTUAL	1	\$220.00			
24	4200	11/1/1998 1...	1036	ACTUAL	1	\$220.00			
24	4200	12/1/1998 1...	1066	ACTUAL	1	\$220.00			
0	4100	1/1/1997 12...	367	ACTUAL	1	\$942.00			
0	4100	2/1/1997 12...	398	ACTUAL	1	\$942.00			
total_contextual_amount: 1204299.96									
total_gen_and_adm_amount: 136237.32									
10_percent_of_total: 120429.996									
account	expense_fact	time_by_day							
Record: 14	4	1 of 2,400							

account_id	account_parent	account_description	account_type	account_rollup
1000		Assets	Asset	~
2000		Liabilities	Liability	~
3000	5000	Net Sales	Income	+
3100	3000	Gross Sales	Income	+
3200	3000	Cost of Goods Sold	Income	-
4000	5000	Total Expense	Expense	-
4100	4000	General & Administrat...	Expense	+
4200	4000	Information Systems	Expense	+
4300	4000	Marketing	Expense	+
4400	4000	Lease	Expense	+

account	expense_fact	time_by_day
---------	--------------	-------------

Figure 6. Example of using DAX Calculate function together with Sum & a condition similarly to SumIf / DSum in the traditional Excel and Select Sum(field_name) From table_name Where condition_or_conditions in SQL

Some benefits of DWH: support for OLAP & dashboards (with KPIs)

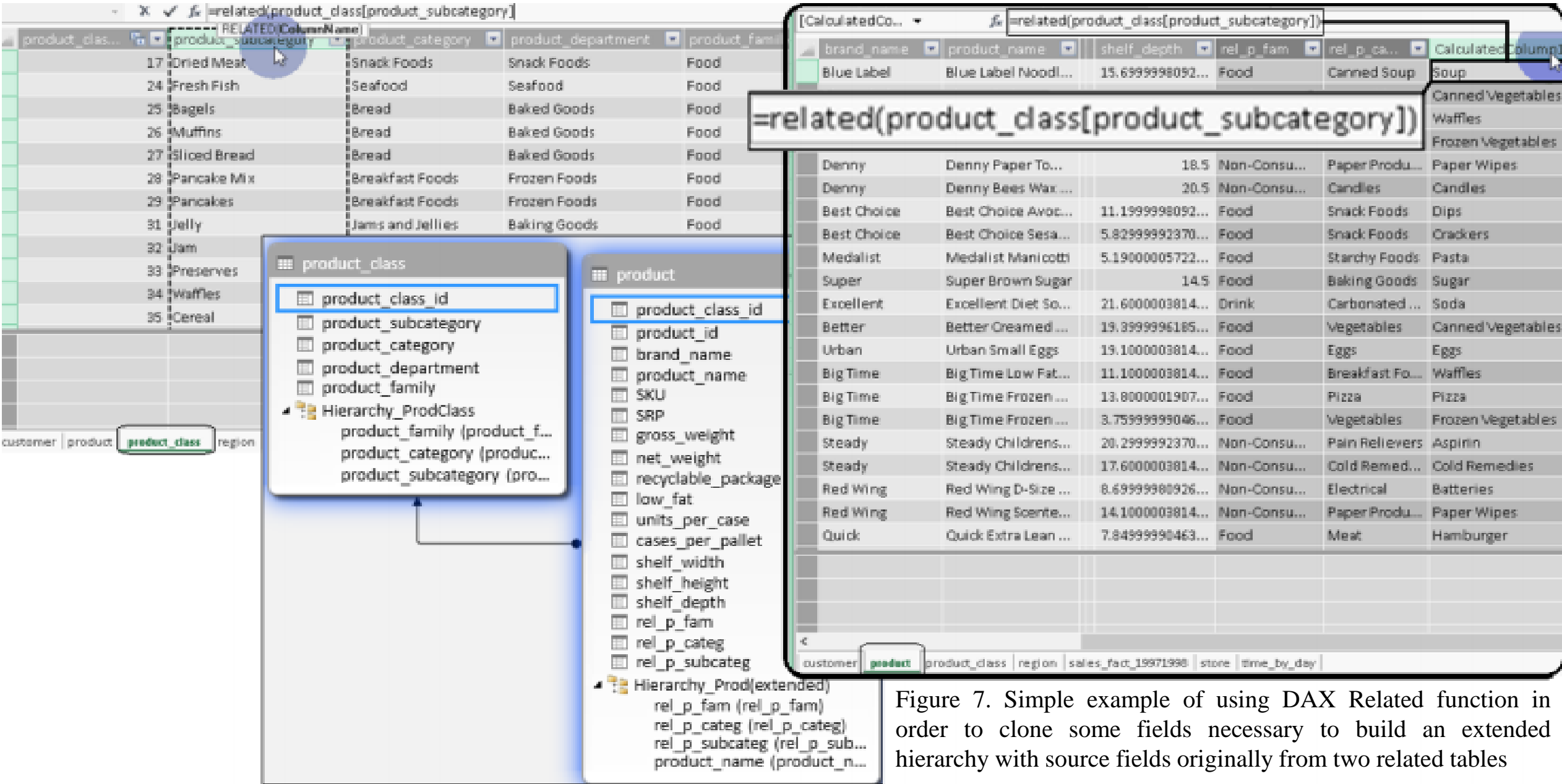


Figure 7. Simple example of using DAX Related function in order to clone some fields necessary to build an extended hierarchy with source fields originally from two related tables

Some benefits of DWH: support for OLAP & dashboards (with KPIs)

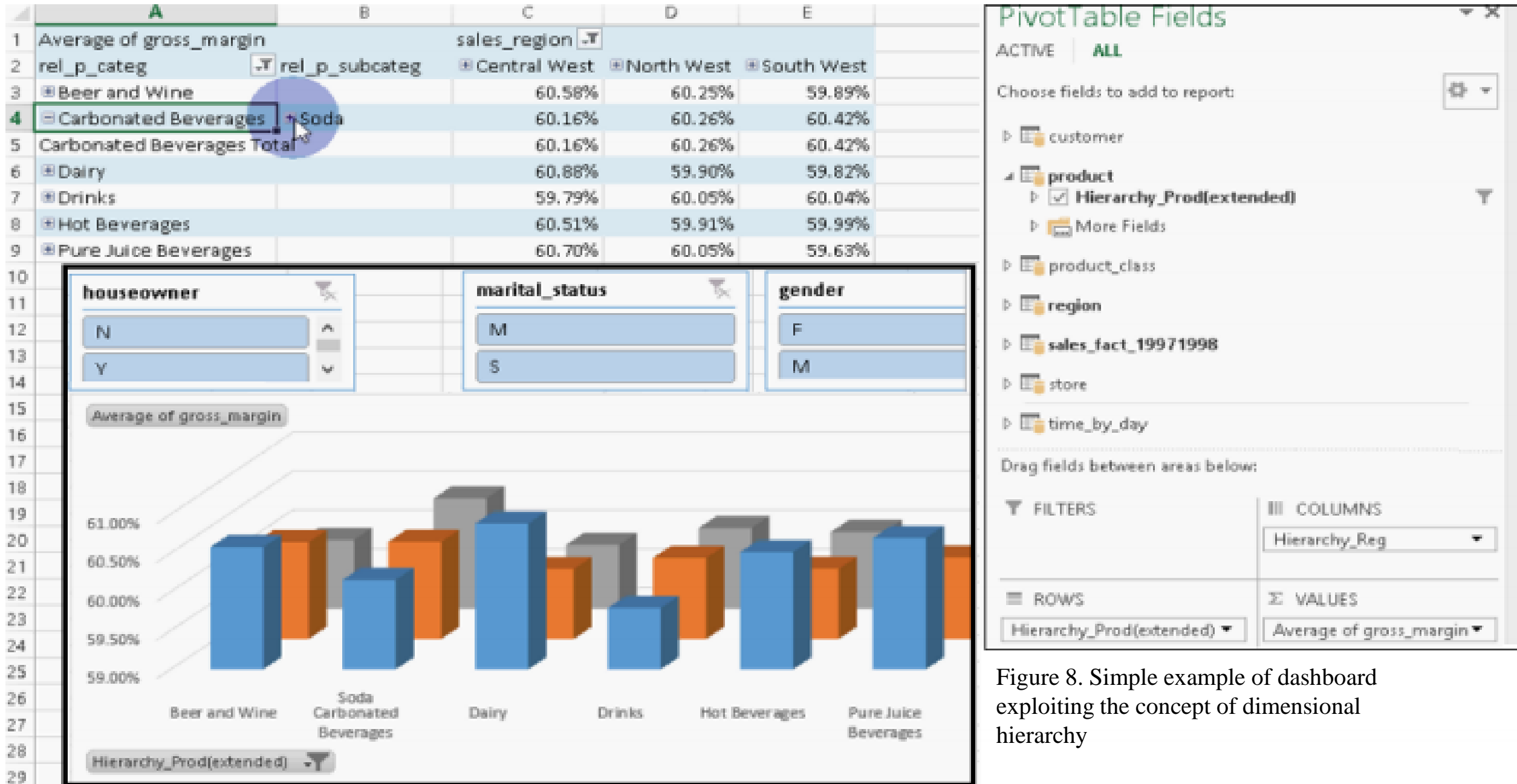
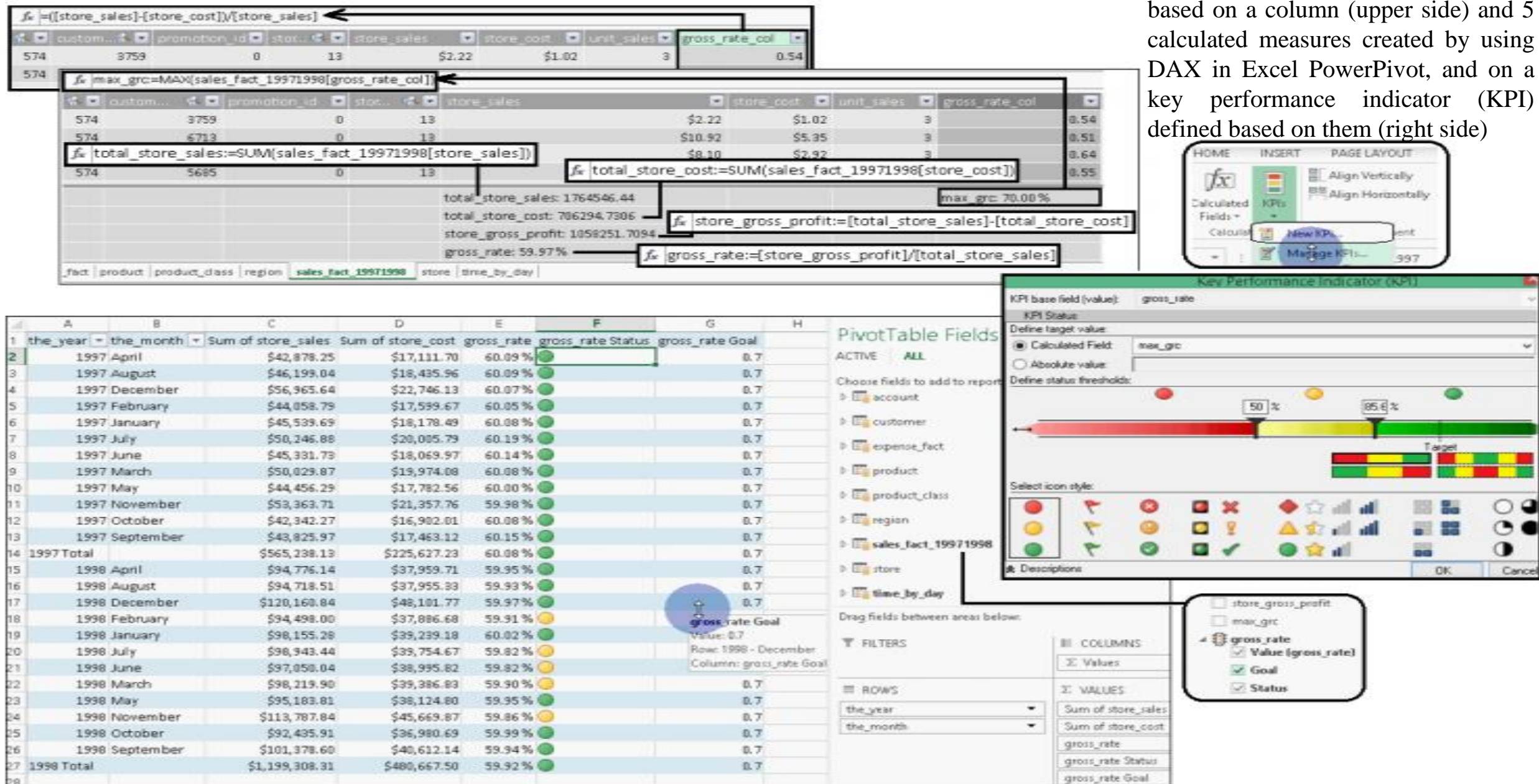


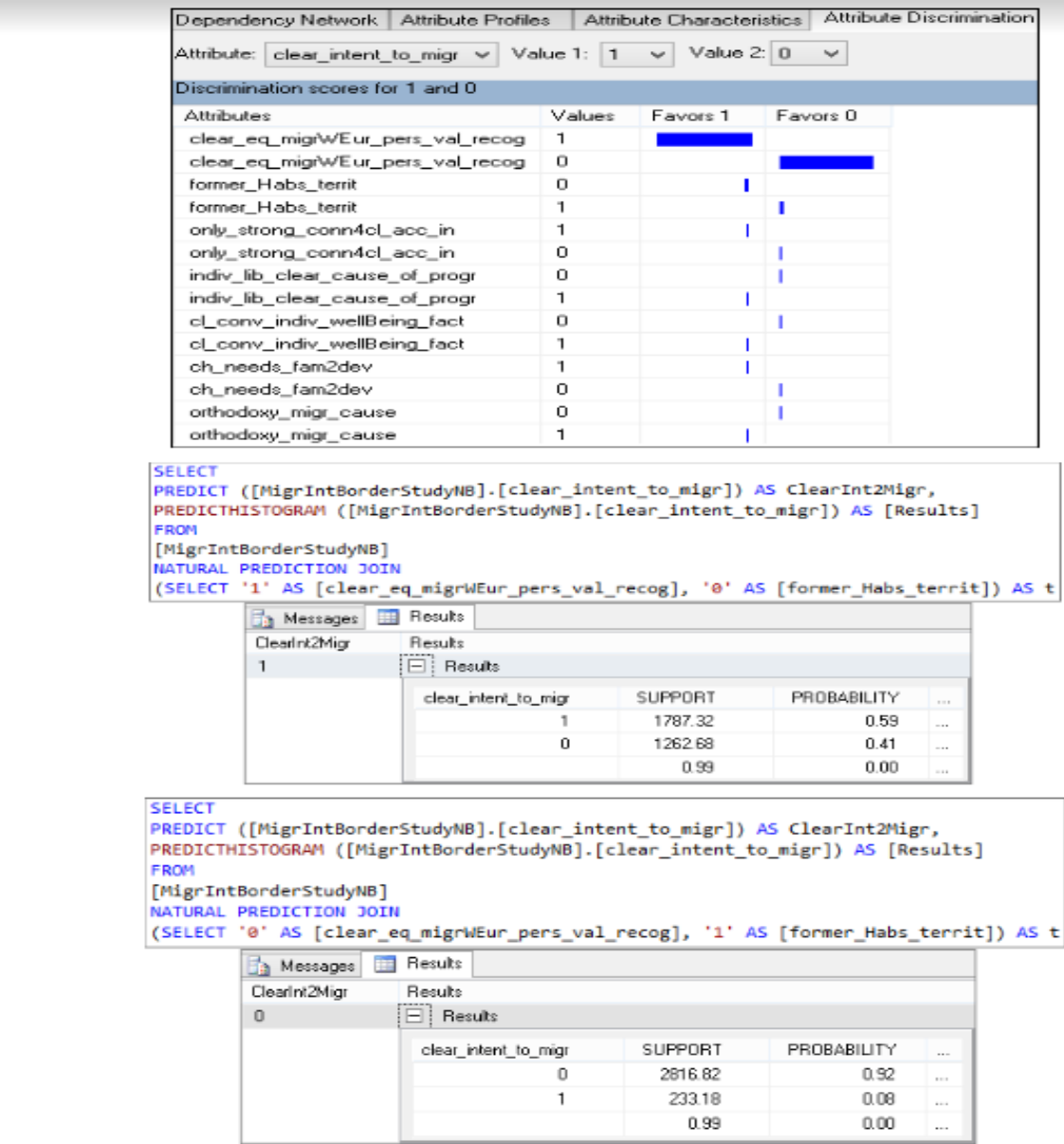
Figure 8. Simple example of dashboard exploiting the concept of dimensional hierarchy

Some benefits of DWH: support for OLAP & dashboards (with KPIs)

Figure 9. Dynamic dashboard built based on a column (upper side) and 5 calculated measures created by using DAX in Excel PowerPivot, and on a key performance indicator (KPI) defined based on them (right side)



Some benefits of DWH: availability for Data Mining and support for modeling



Variables	(a)	(b)	(c)	(d)
clear_eq_migrWEur_pers_val_recog	2.343*** (0.105)	2.337*** (0.106)	2.286*** (0.107)	2.275*** (0.108)
former_Habs_territ	-0.516*** (0.104)	-0.526*** (0.106)	-0.572*** (0.107)	-0.579*** (0.108)
interpers_trust_lv_under_or_eq3			0.447** (0.184)	0.447** (0.186)
orthodoxy_migr_cause			0.591*** (0.161)	0.548*** (0.163)
accept_or_doubts_notpaying_pub_t			0.251* (0.142)	0.269* (0.143)
only_strong_conn4cl_acc_in			0.241** (0.111)	0.245** (0.111)
indiv_lib_clear_cause_of_progr			0.494*** (0.128)	0.494*** (0.128)
cl_conv_st_cr_wellBeing			0.238** (0.110)	0.252** (0.111)
cl_conv_st_inv_moreEc			-0.411*** (0.121)	-0.357*** (0.122)
min_dist2fHt_over_or_eq75		0.312*** (0.105)		0.307*** (0.107)
inh_val_resp4tradit		-0.318*** (0.110)		-0.285** (0.112)
m_edu_over_or_eq12		0.233** (0.112)		0.199* (0.113)
both_p_gone_abroad		0.258* (0.150)		0.265* (0.152)
CONSTANT	-1.987*** (0.084)	-2.212*** (0.130)	-2.724*** (0.233)	-2.981*** (0.259)
Observations	3051	3051	3051	3051
LR chi2(df)	563.43	588.59	620.46	641.62
df	2	6	9	13
Prob > chi2	0.0000	0.0000	0.0000	0.0000
Pseudo R square	0.1893	0.1977	0.2085	0.2156
AUC(ROC)	0.7707	0.7897	0.8002	0.808

Table 2. Raw coefficients for logit regressions (Stata 13) considering the entire dataset

Figure 2. Using a classification technique (Naive Bayes algorithm) in Microsoft Excel Data Mining and exploring the model's most powerful influences using SQL Server Analysis Services

Some benefits of DWH: availability for Data Mining

Do you see **the problem** with the time-series based Excel DM prediction model below???

! If so, tell me / send me an e-mail with details !

!!! It will help increase your final score !!!

The screenshot shows an Excel spreadsheet with a time-series dataset. The 'Forecast' ribbon is active, and a 'Browse' dialog box is open, showing a line chart of 'AvgSales' over time. The chart shows a downward trend with a forecast line. The 'Mining Legend' shows 'AvgSales' as the target variable. The 'Prediction steps' are set to 6. The 'Show historic predictions' and 'Show Deviations' checkboxes are checked. The 'Copy to Excel' button is visible. The 'SQLQuery1.sql' window shows the following query:

```
SELECT LEFT(CONVERT(varchar(10), OrderDate, 112), 6) AS OrdYM, AVG(SalesAmount) As AvgSales
FROM [AdventureWorksDW2012].[dbo].[FactResellerSales]
GROUP BY LEFT(CONVERT(varchar(10), OrderDate, 112), 6)
ORDER BY LEFT(CONVERT(varchar(10), OrderDate, 112), 6)
```

The 'Results' window shows the following data:

OrdYM	AvgSales
200712	1159.1114
200801	1203.0009
200802	1300.321
200803	1256.5921
200804	1228.2024
200805	1164.4801
200806	1137.7345

Query executed successfully. MV-W81-32BITS (11.0 SP1) mv-w81-32bits\admin (54) master 00:00:00 36 rows

	A	B
1	OrdYM	AvgSales
2	200507	1390.138
3	200508	1959.756
4	200509	1966.1
5	200510	1692.828
6	200511	2101.389
7	200512	2120.728
8	200601	1891.556
9	200602	2194.906

Bibliography

- Airinei, D., Depozite de date, Polirom, Iași, 2002
portal.feaa.uaic.ro/Master/sia/Pages/default.aspx
- Airinei, D., Dospinescu, O., Huiban, A., Aplicatii practice cu sisteme OLAP si Depozite de date, Editura Sedcom Libris, Iași, 2008
- Homocianu, D., Sistemele de asistare a deciziilor in contextual societatii cunoasterii, Editura UAIC, Iasi, 2009 (ssrn.com/abstract=2384380)
- Inmon, W.H., Linstedt, D., Data Architecture: A primer for the data scientist, MK, MA, 2015 (tinyurl.com/h7dcq66)
- Kimball, R., Ross, M., The Data Warehouse Toolkit Third Edition. The Definitive Guide to Dimensional Modeling, John Wiley & Sons, New York, 2013 (tinyurl.com/jogo5uy)
- Kimball, R., Ross, M., The Kimball Group Reader. Relentlessly Practical Tools for Data Warehousing and Business Intelligence – Remastered Collection, Second Edition, Wiley, New York, 2016 (tinyurl.com/johox4v)
- Krishnan, K., Data Warehousing in The Age of Big Data, Morgan Kaufmann (MK), MA, 2013 (tinyurl.com/zrjo75j)
- Reeves, L.L., A Manager's Guide to Data Warehousing, Wiley, New York, 2009 (tinyurl.com/htsslk8)
- Sarka, D., et. al., Implementing a Data Warehouse with Microsoft SQL Server 2012. Training Kit, O'Reilly Media, Sebastopol, 2012 (tinyurl.com/jk6lclk)
- Sheldon, B., et. al., Professional Visual Basic 2012 and .NET 4.5 Programming, John Wiley & Sons, Indianapolis, 2013 (tinyurl.com/hrb6j9u)