

DATA WAREHOUSES (DWH/DW)

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.

February 2018

Su	Mo	Tu	We	Th	Fr	Sa
28	29	30	31	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	1	2	3
4	5	6	7	8	9	10



2:40:09 PM

[Change date and time settings...](#)

March 2018

Su	Mo	Tu	We	Th	Fr	Sa
25	26	27	28	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31
1	2	3	4	5	6	7



2:40:35 PM

[Change date and time settings...](#)

March 2018

Su	Mo	Tu	We	Th	Fr	Sa
25	26	27	28	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31
1	2	3	4	5	6	7



2:40:50 PM

[Change date and time settings...](#)

April 2018

Su	Mo	Tu	We	Th	Fr	Sa
25	26	27	28	29	30	31
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	1	2	3	4	5



2:41:41 PM

[Change date and time settings...](#)

April 2018

Su	Mo	Tu	We	Th	Fr	Sa
25	26	27	28	29	30	31
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	1	2	3	4	5



8:51:46 AM

[Change date and time settings...](#)

May 2018

Su	Mo	Tu	We	Th	Fr	Sa
29	30	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9



8:52:28 AM

[Change date and time settings...](#)

May 2018

Su	Mo	Tu	We	Th	Fr	Sa
29	30	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9

Semestrul al II-lea

19 februarie – 3 iunie

4 iunie – 17 iunie

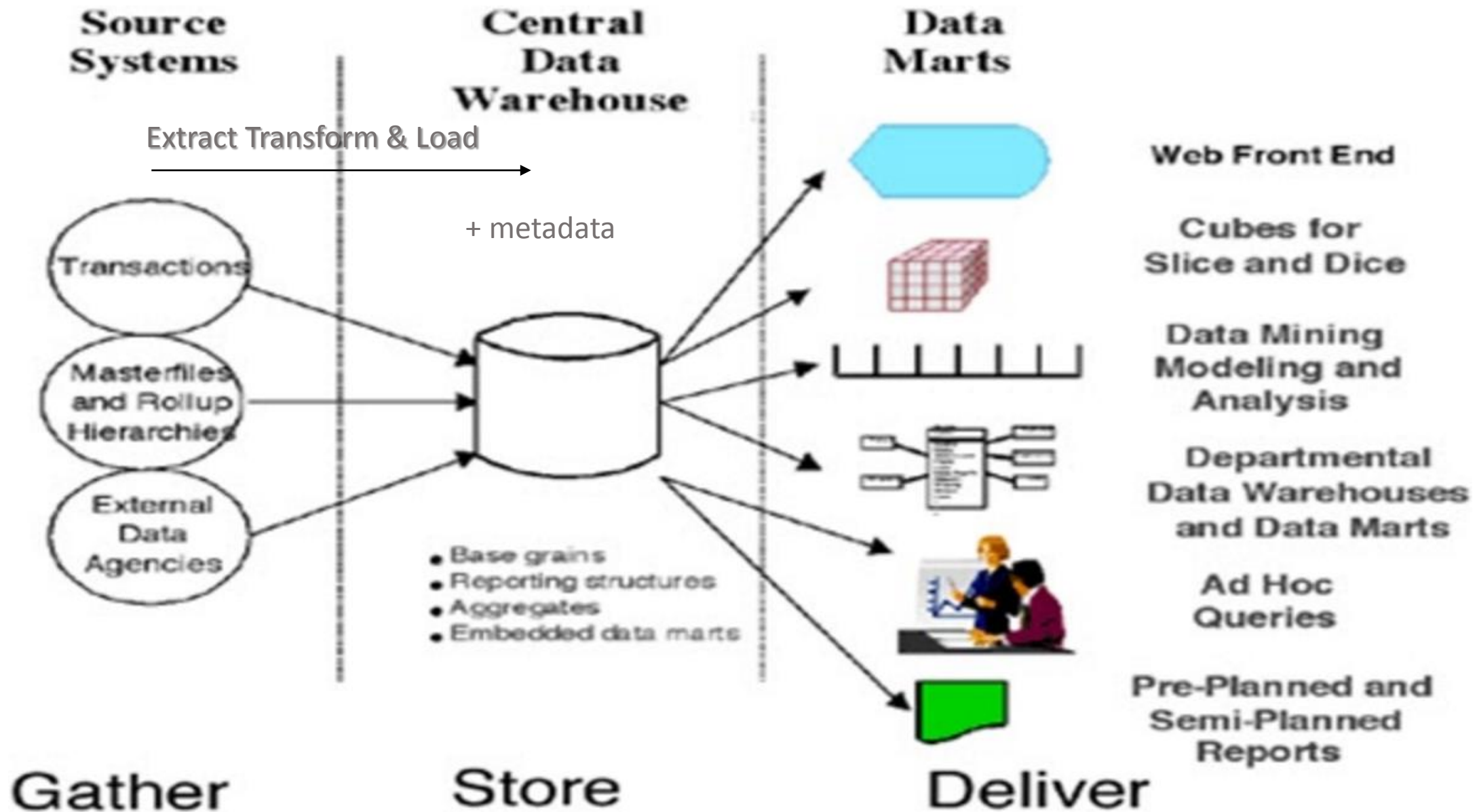
[Change date and time settings...](#)

14 săptămâni: activitate didactică

1 săptămână liberă în perioada Sărbătorilor de Paști (9 – 15 aprilie)

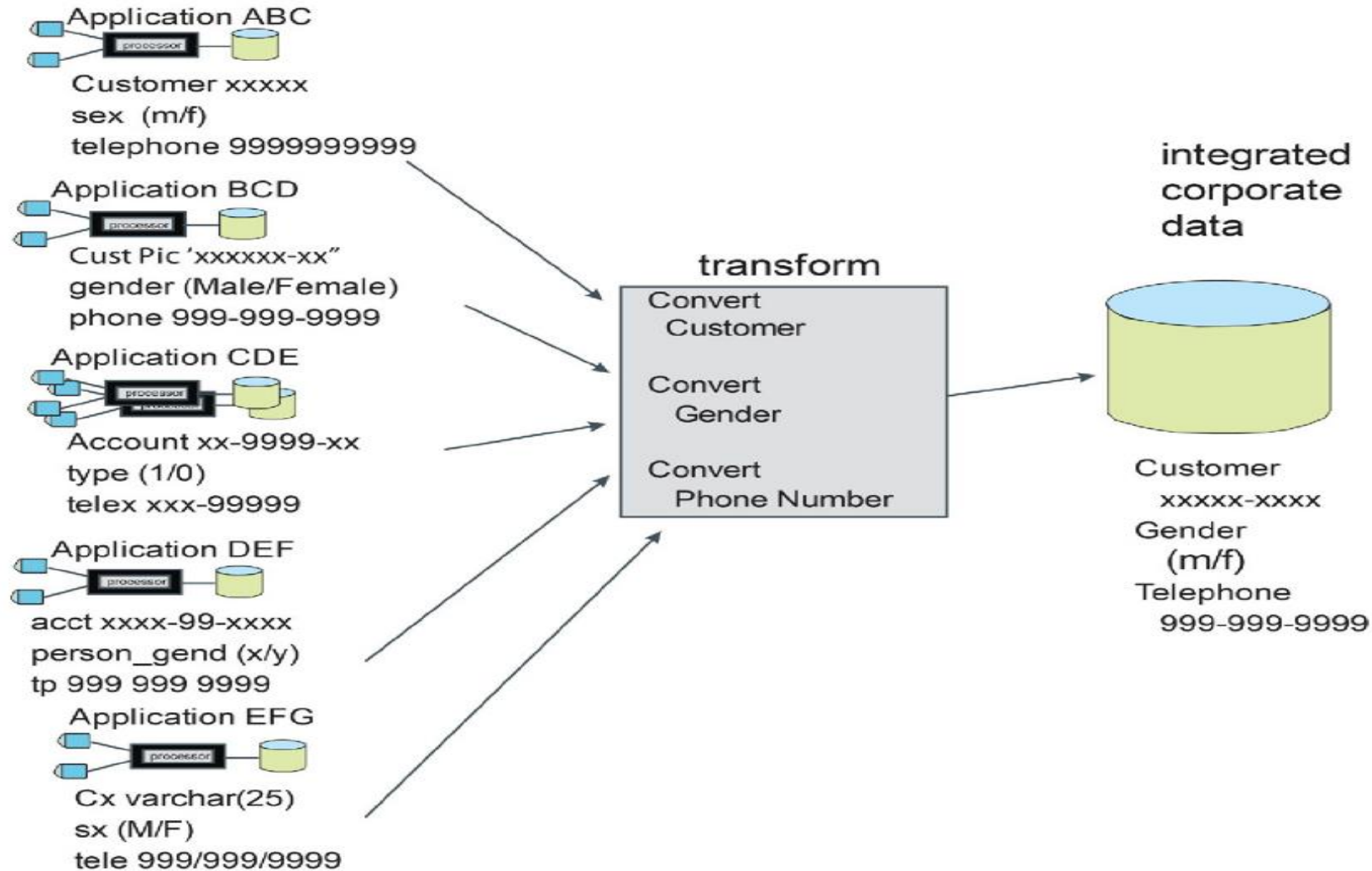
2 săptămâni: evaluare

DWH – overview



Extract Transform & Load (ETL)

Inmon & Linstedt, 2015



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)

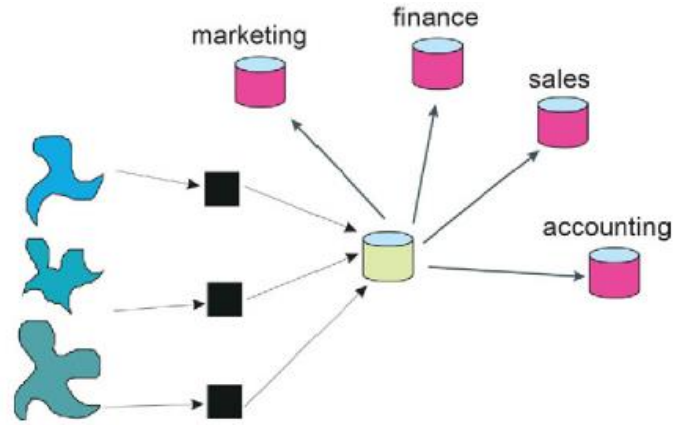


Figure 3.4.3

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.

(Inmon & Linstedt, 2015)

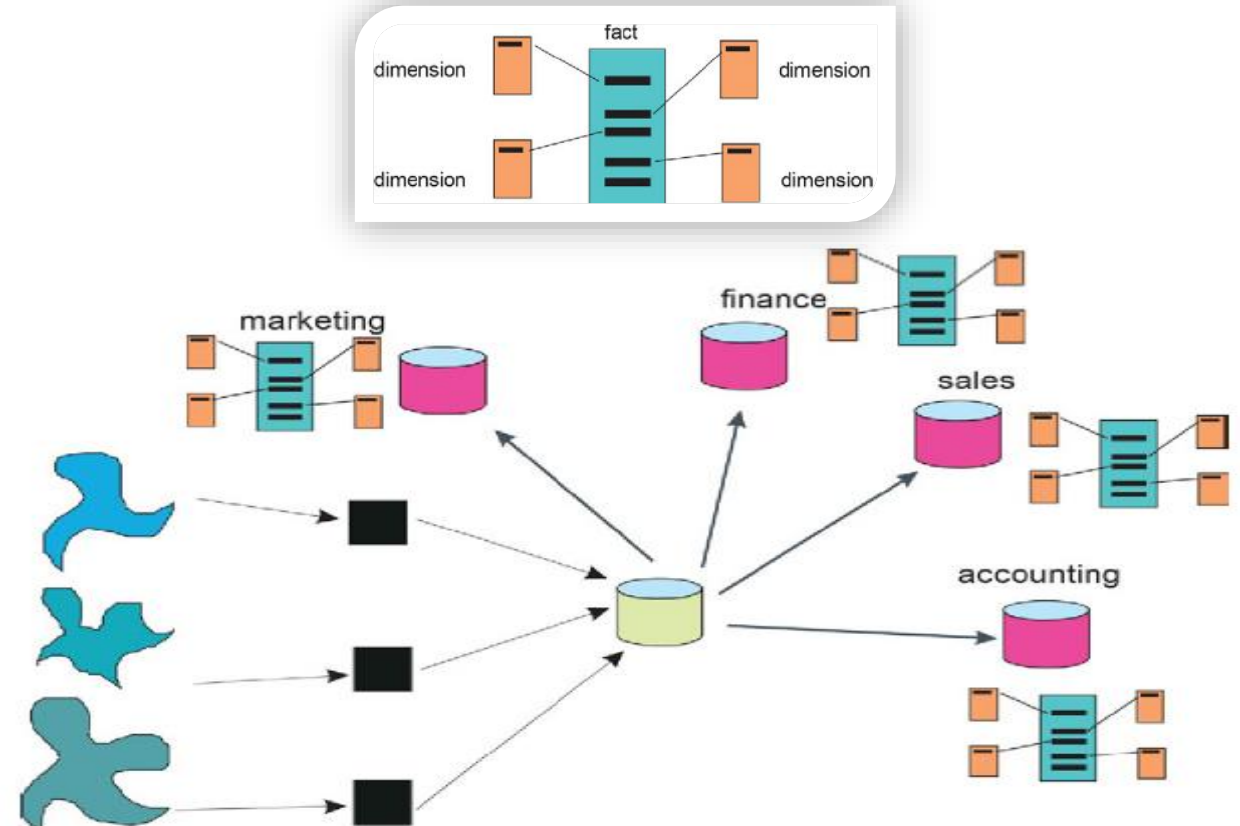


Figure 3.4.4

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		
Supplier Deliveries	X							X	X	X	X	X	X		

	<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.

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)