Business Intelligence

A smooth introduction …

Business Intelligence

•Contents

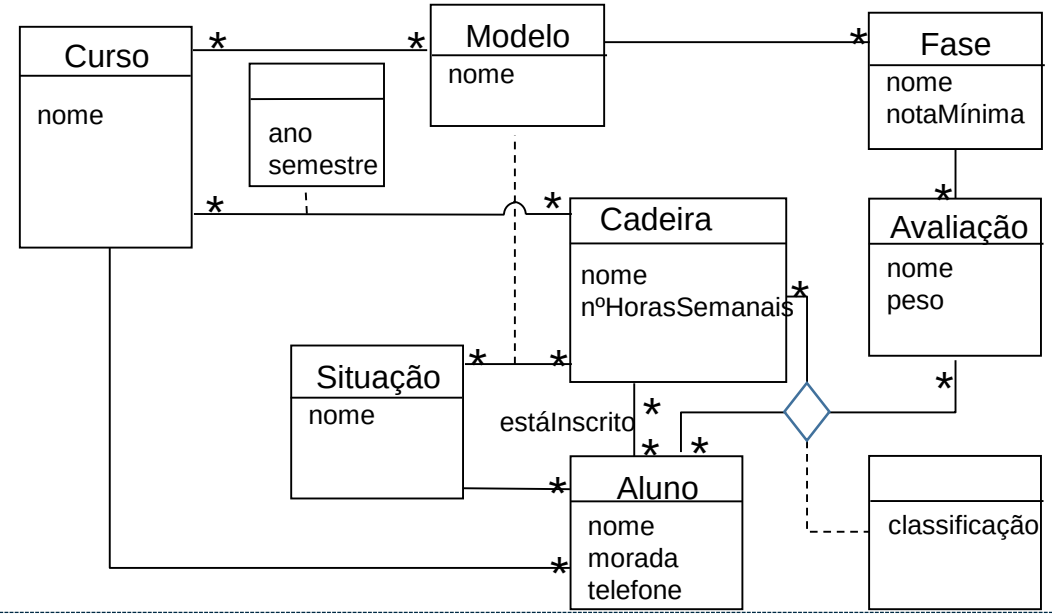
- Context
- 3 typical cases
- Concepts
- Business Intelligence (BI)
- References

Imagem retirada de www.demandsolutions.com a 11/03/2015

Context

Context

Since the 70/80 decade, the dominant model of databases has been the "normalized" relational model. An example of its representation:



Context

This way of designing databases is characterized by:-

- Do not store redundant information
- The design of the database is made according to the structure of the data
- Be optimized for INSERT, delete, and update records operations
- The space occupied by the data is minimized

Context

•Thus, to extract information from databases, we use the SQL language – Structured Query language

Examples

Students with a mean greater than or equal to 14

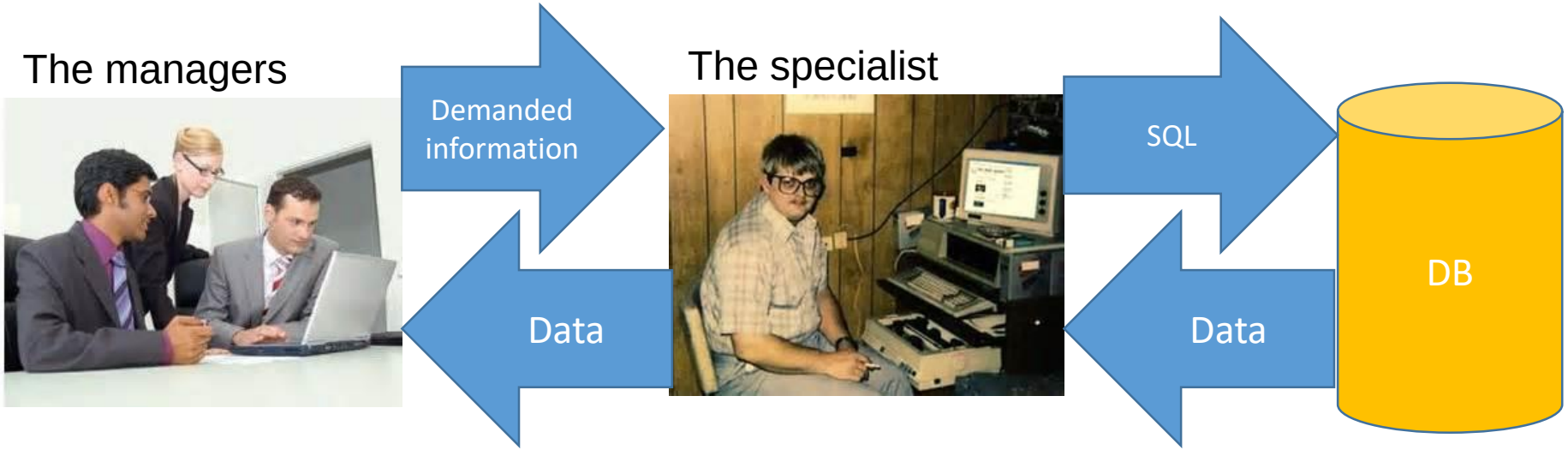
```
SELECT * FROM Aluno WHERE media >=14;
```

Students enrolled in Business Intelligence

```
SELECT Aluno.nome
FROM Aluno, Cadeira, AlunoCadeira
WHERE Aluno.idAluno=AlunoCadeira.idAluno
AND Cadeira.idCadeira=AlunoCadeira.idCadeira
AND Cadeira.nome='Business Intelligence';
```

Context

- **With so-designed databases, extract information from them only for specialists!!**



Context

- With this mode of organization the data analysis process has too many stakeholders, it becomes lengthy and, above all, **misaligned with the management practices**
- During the decade of 90 the data storage devices (notably the hard disks of computers) were getting:
  - Cheaper and cheaper
  - With increasing capacity
- Having redundant information has become cheaper

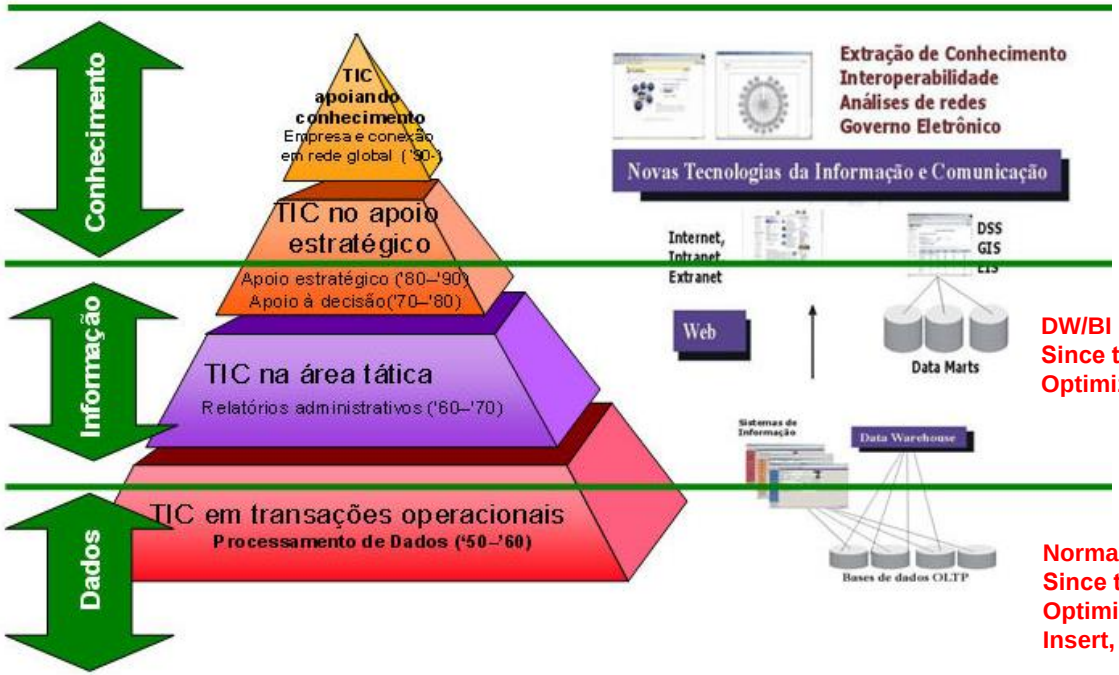
Context

•And it was at that time, mid-90, that data Warehousing/Business Intelligence systems emerged

•Its main characteristic is:

- The data is saved in order to facilitate consultations without needing intermediaries. Consequently:
  - Redundant information is stored
  - The database design is based on the queries you want to make to the database
  - It’s optimized for database query operations
  - The space occupied by the data is much higher than in a normalized relational database (it is estimated that 10 to 20 times greater)

Context



DW/BI
Since the 90's
Optimized for queries

Normalized relational databases
Since the 70/80's
Optimized for the operations:
Insert, update and delete

Image taken from http://revista.ibict.br/index.php/ciinf/article/viewFile/795/644/2174 in 15/02/2010

Context

In the perspective of organisations:

- The normalized relational databases exist and will continue to exist: they store the "base" data. The way data is saved depends on the characteristics of this data. Once created they endure in time... and endure... and endure...
- Business Intelligence systems depend on management processes. They depend on the information you want to get... much more subject to change...

3 typical cases

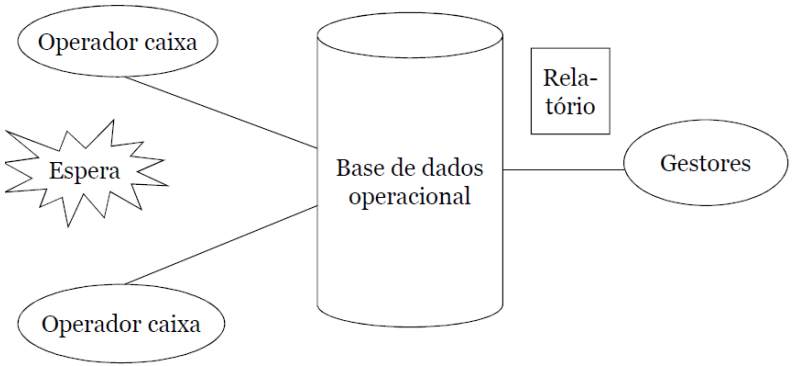
1st case: Hipermarket

Image taken from www.dreamstime.com in 11/03/2015

1st case: Hipermarket

Problem

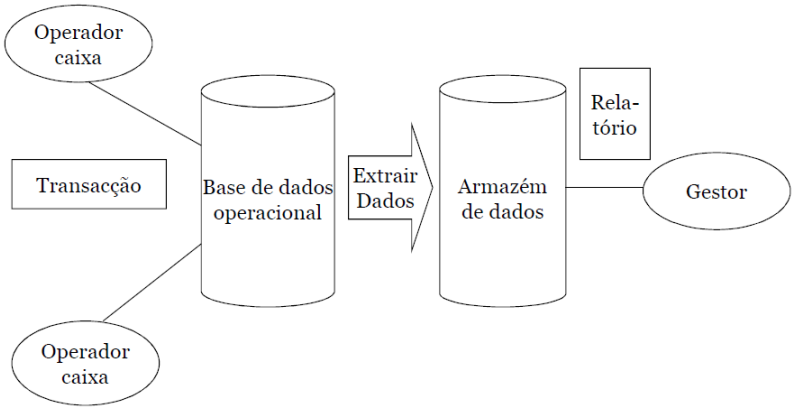
- Huge operational database
- Many boxes running at hours of greater movement
- Query queries about the database at these times create response delays in the operation of the boxes



1st case: Hipermarket

•Solution

- Extract the data needed for analysis of the operational database
- Store them in a data warehouse
- Refresh the Data warehouse at regular time intervals in order to always have data updated for analysis
- The Data warehouse will contain data in a historical perspective



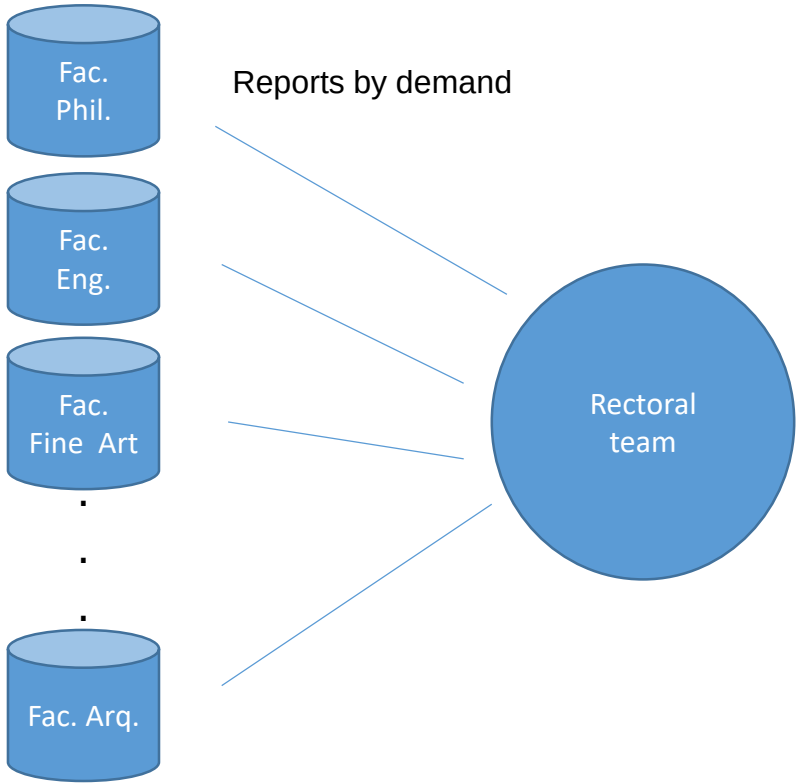
2nd case: University

Imagem retirada de www.tunghai.org a 11/03/2015

2nd case: University

•Problem

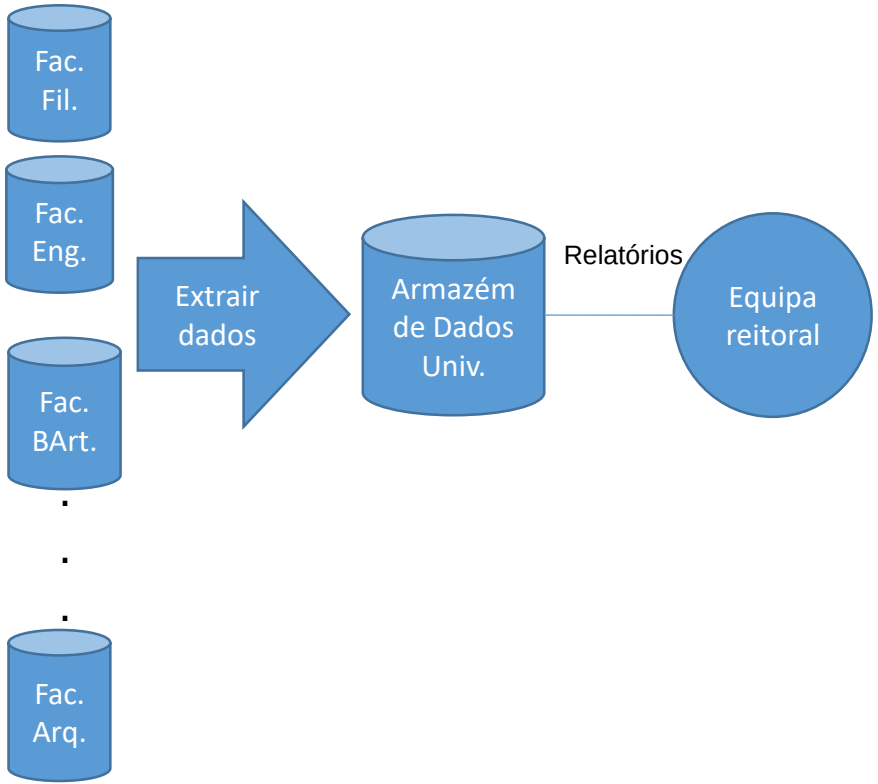
- Every college has its operating system
- Each college has its own management mode but...
- You must report your management actions to the Rectoral team who complains of....
- Not having adequate information to be able to compare management in different faculties



2nd case: University

•Solution

- Extract the relevant information for the Rectoral team from each of the colleges
- Storing the information in a single repository



3rd case: Cakes & Cookies

- **Problem**

- Cakes & Cookies is a small new company
- The president of the company wants her to grow
- To do this, you need information so you can make the best decisions

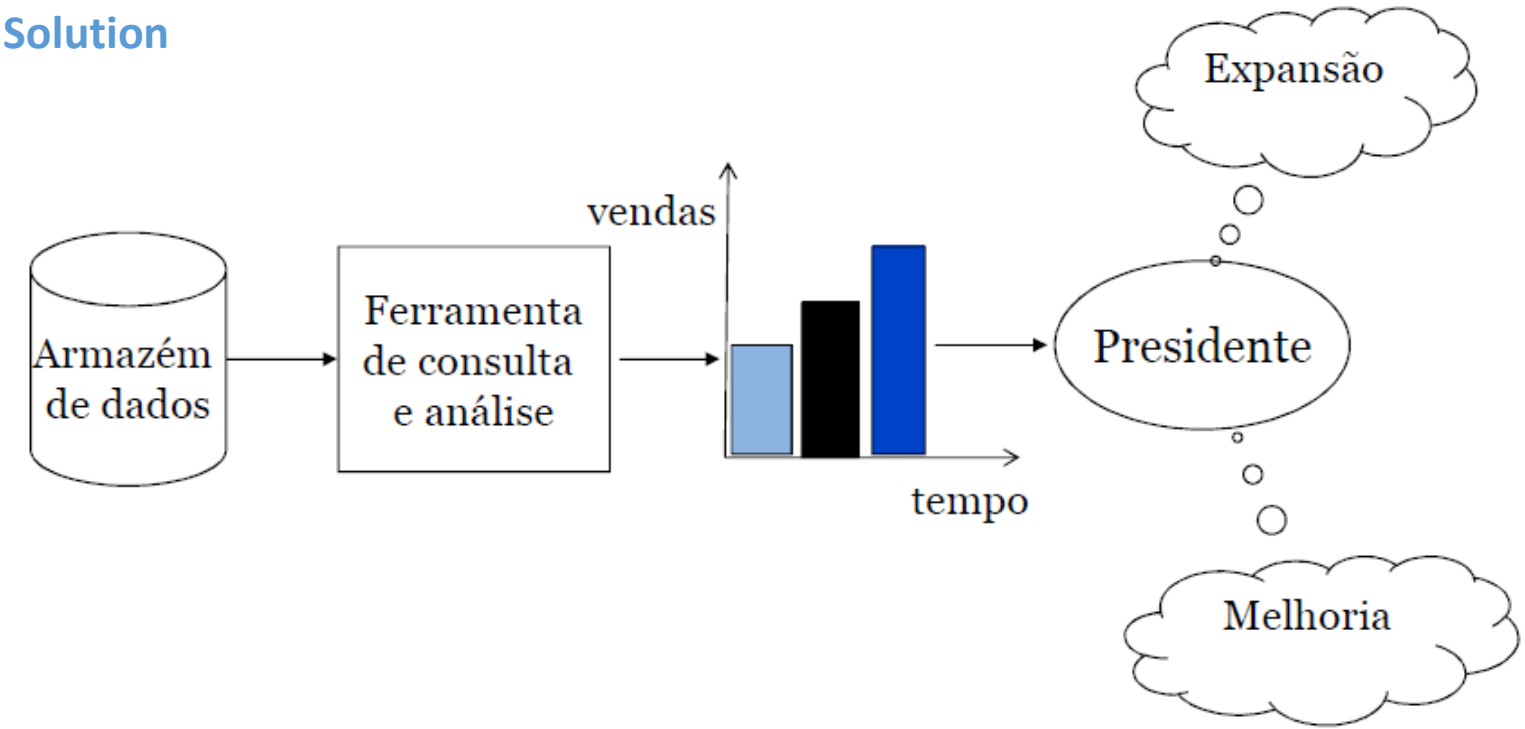
3rd case: Cakes & Cookies

•Solution

- Improve data quality by cleaning and processing before uploading to data Warehouse
- Using tools to query data that allows ad-hoc consultations to be performed

3rd case: Cakes & Cookies

- **Solution**



Concepts

Concepts

•*Data Warehouse*

•**Definition**: Collection of data to support the decision-making process by managers.

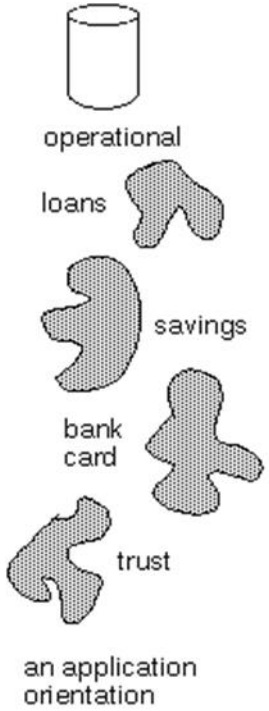
Characteristics of a Data warehouse:

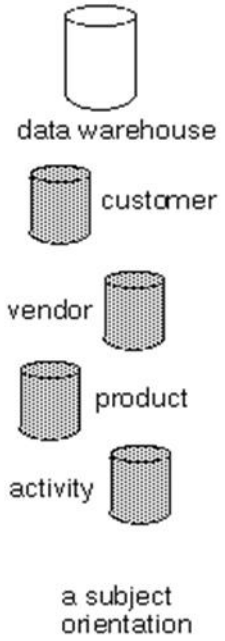
- Oriented to the subjects
- Integrated
- It makes the temporal record of the information
- Non volatile

Concepts

- ***Data Warehouse***

- The Data Warehouse is organized by subjects such as sales, products, customers, etc.
- It is oriented to the modelling and analysis of data for decision-making
- Excludes data that is not useful in the decision-making process

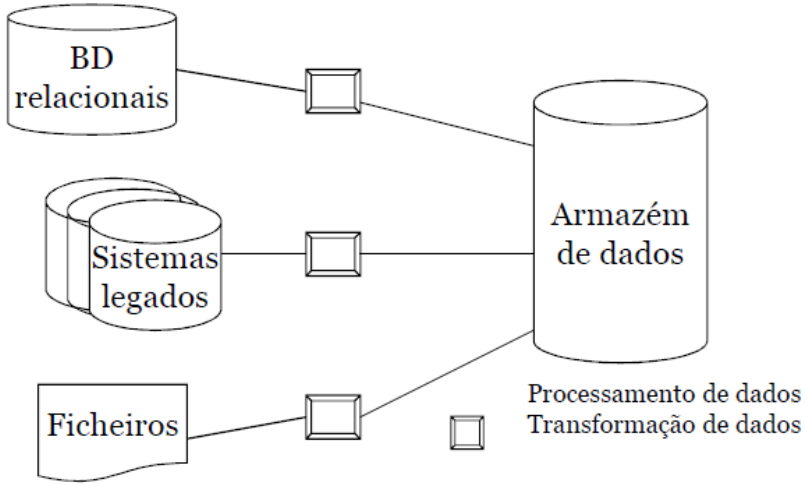




Concepts

•*Data Warehouse*

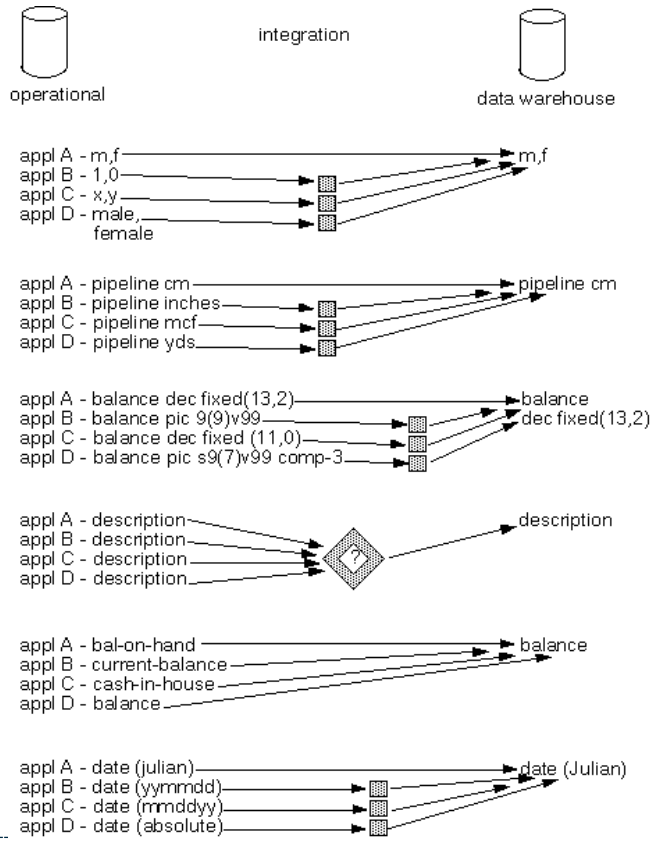
- A data warehouse is built by integrating data from more than one source, and these are typically heterogeneous sources.
- Data processing is carried out to ensure consistency of the same data.



Concepts

- ***Data Warehouse***

- The data **integration** can be done in terms of:
  - Coded structures
  - Attribute measures
  - Physical attribute of data
  - Naming convention
  - Data type format



Concepts

- ***Data Warehouse***

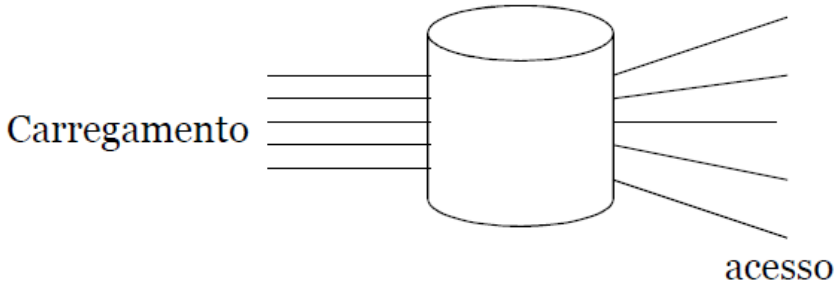
- It makes the **temporal record of information**, i.e. it has information with historical perspective, e.g., last 5-10 years

- All key structures implicitly or explicitly contain a time element

Concepts

•***Data Warehouse***

- Data once entered into the Data warehouse cannot be updated
- The Data Warehouse allows you to perform two types of operations on the data:
  - Initial data load
  - Data access



Concepts

- **Comum terms**

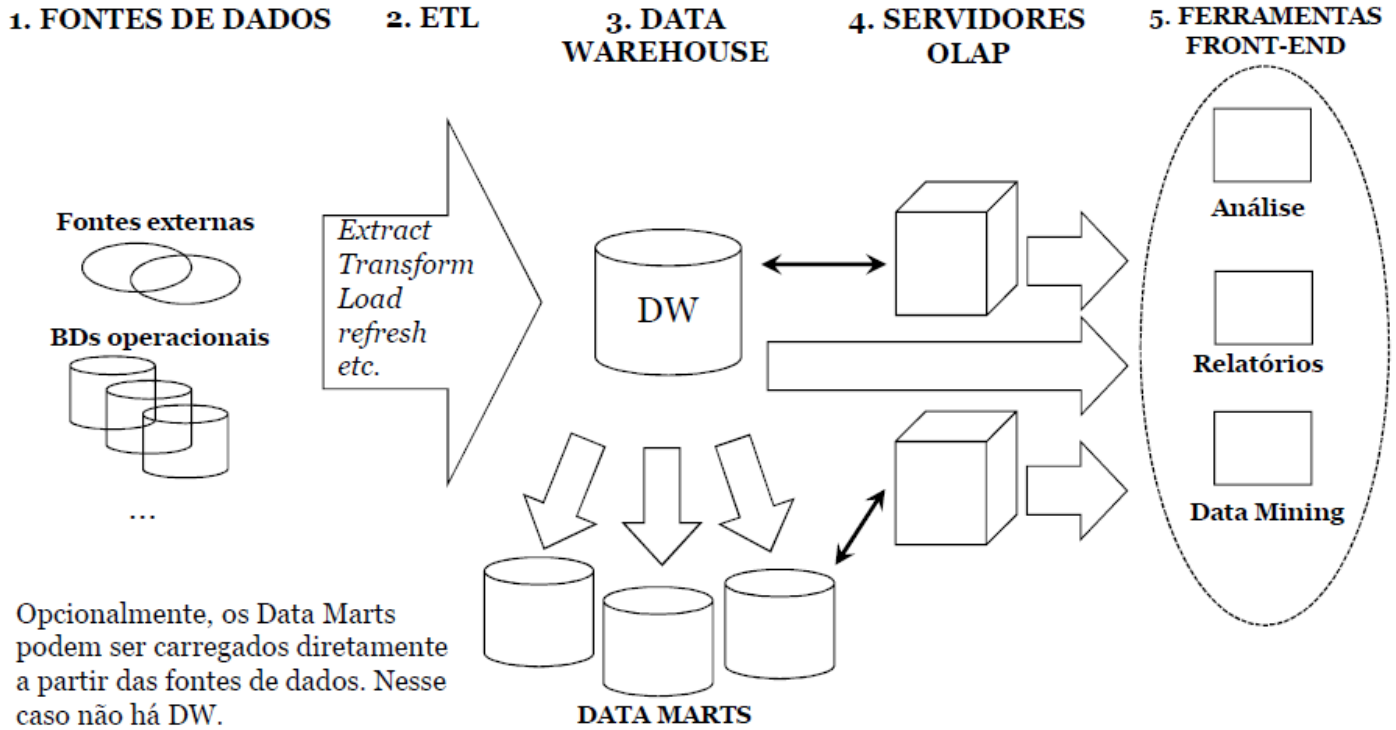
OLTP: OnLine Transactional Processing
Operational
Agregated
SQL
Data Warehouse
By subjects
OLAP: OnLine Analytical Processing
DB normalized relational
Redundant data
Space minimization
Temporal analysis
Slow queries

Concepts

	Operational	Informative
Caracteristics	Operational Processing	Informative processing
Orientation	Transaction	Analysis
Userr	Clerk, DB admin, DB professionals	Knowledge workers
Function	Daily operations	Decision support
Data	Actual	Historical
View	Detailed, table format	Agreggated, multidimensional
DB design	Oriented towards applications	By subjects
Working unit	Short, simple transaction	Complex queries
Acess	Read/write	Mostly reading
Focus	Entrance data	Intended information
Nr of accessed records	Dozens	Milions
Nr of users	Thousands	Hundreds
DB size	From 100MB to GB	DFrom 100 GB to TB
Priority	High performance, high availability	High flexibility, End-users authonomy

Business Intelligence

Business Intelligence



Business intelligence

1. Data sources
   - Operational databases, Internet, files,...
2. Etl
   - Data extraction, transformation and loading
3. Data Warehouse server: Data Warehouse (DW)
   - Almost always a relational DBMS, rarely table type files
4. OLAP servers
   - To manage and operate dimensional data structures
5. Front-end tools
   - Tools for miscellaneous purposes (reporting, analysis, Data Mining,...) to be used by the end user

Business Intelligence

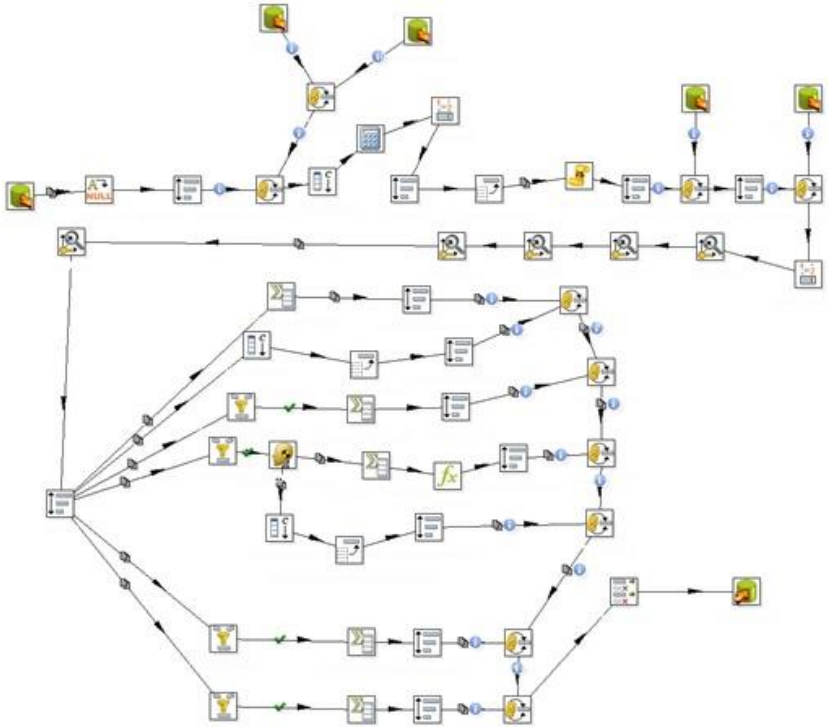
- **1. Data sources**
- The databases that the company/institution has
  - Typically, relational databases, Enterprise Resource Planning (ERP),...
- Internet data
  - Weather, stock quotes,...
- Files
  - Excel file data,...
  - But it is strongly inadvisable to be difficult to ensure that the structure of the data source does not change...

Business Intelligence

- **2. ETL**
- It uses its own tools to accomplish this task
  - E.g.: kettle, Oracle Data Integrator, SQL Server Integration Services, etc.
- This task is the one that consumes the most time in a development project of a Data Warehousing system
  - 70% to 80% of the total development time
- The process, once created, runs the number of times necessary in order to ensure the previously defined historical perspectiveanual, mensal, daily, …

Business Intelligence

- **2. ETL**
- Example:



Business Intelligence

- **2. ETL**
- It understands the following tasks:
  - Data selection
  - Data pre-processing
    - Filling missing values
    - removing inconsistencies
  - Transformation & Data Integration
  - Data loading

- Data from a data warehouse is typically stored in the form of fact tables and dimension tables

Business intelligence

- **3. DW**
- Change of paradigm:
  - In the traditional relational model, it seeks to avoid repetitions of information
  - To do this, relational databases are typically normalized to the 3rd normal form
  - Result: information (almost) without repetitions, occupying little disk space, resulting in complex and slow database readings
  - In DW we can use the relational model but the database is poorly normalized: there is repetition of the information, achieving with this database readings simpler and faster

Business Intelligence

- **3. DW**

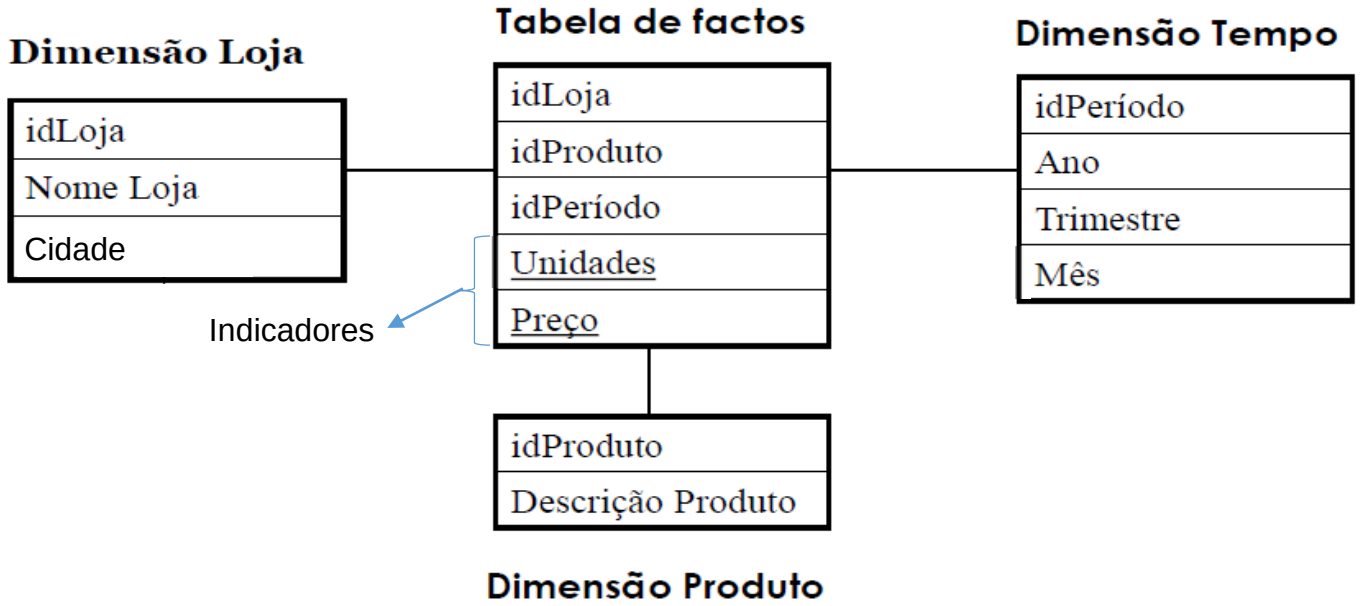
- **Data Warehouse:** Global repository for the entire company/institution
- **Data Mart:** Repository for a department or function of the company/institution
- The development of a Data Warehouse is typically very time-consuming
- It is common to choose to develop Data marts, that is, to do the development in a phased manner by department/function

Business Intelligence

- **3. DW**
- Physical Schema:
  - Star scheme
  - Snowflake Scheme
  - Fact Constellation Scheme

Business Intelligence

- **3. DW – Star schemma**



Business Intelligence

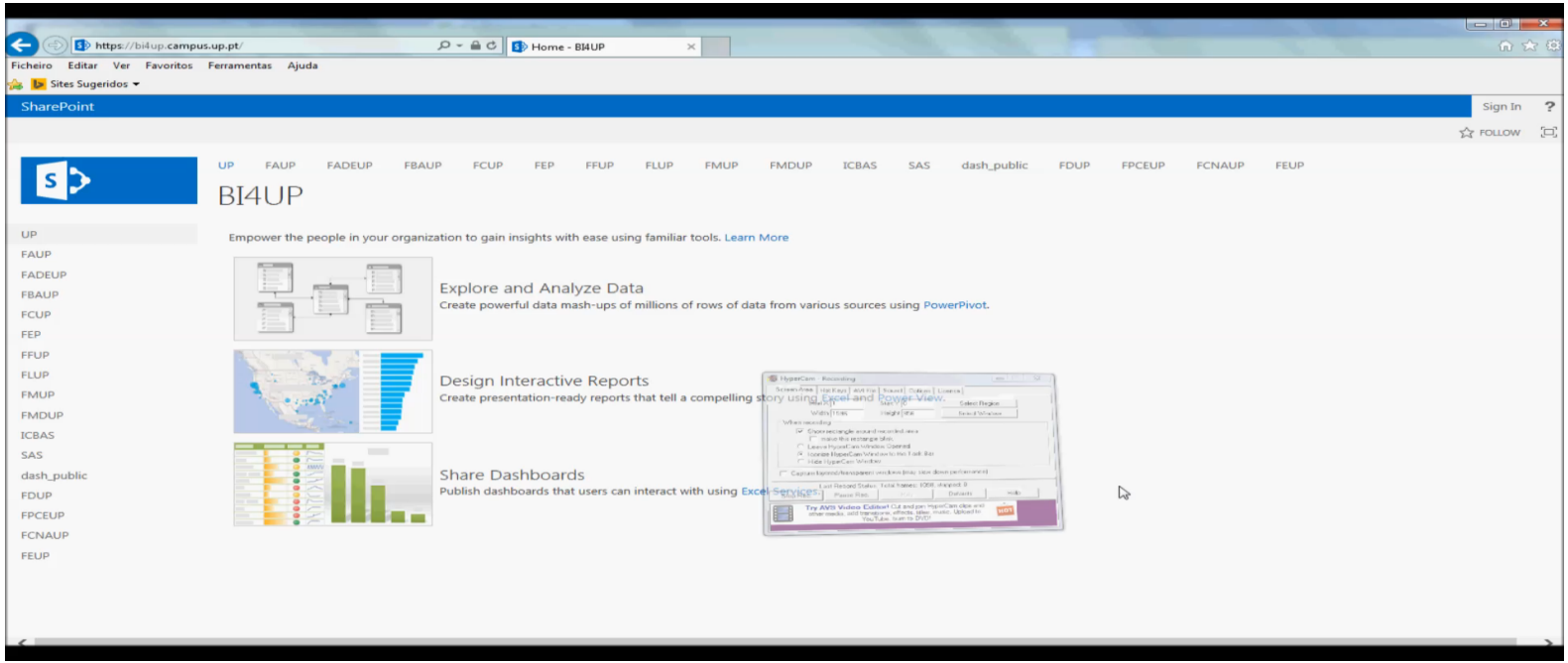
- **4. OLAP server**
- OLAP: OnLine Analytical Processing
- An OLAP is a high-capacity, multi-user data manipulation device designed specifically to support and operate on multi-dimensional data structures

Business Intelligence

- **5. front-end tools**
- Querying tools
- Reporting tool
- Spreadsheets
- WEB interfaces / dashboards
- Data Mining

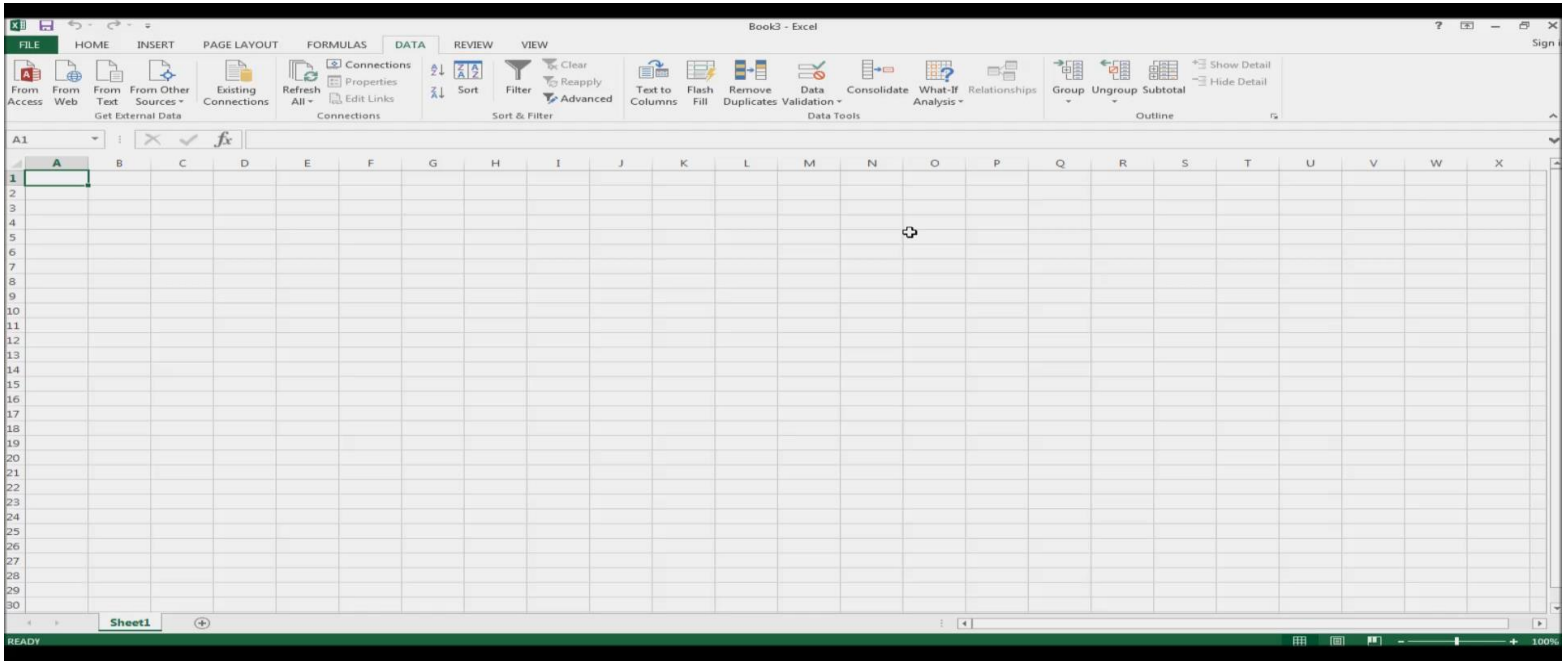
Business Intelligence

- **5. Front-end tools: Web interfaces / dashboards**



Business Intelligence

- **5. Front-end tools: spreadsheets**



Classes Plan

Classes plan

Plan	Feb/Mar	Mar	Mar	Mar	Mar/Apr	Apr	Apr	Apr	Apr	May	May	May	May	May/Jun	Jun
Monday	28	7	14	21	28	4	11	18	25	2	9	16	23	30	6
Tuesday	1	8	15	22	29	5	12	19	26	3	10	17	24	31	7
Wednesday	2	9	16	23	30	6	13	20	27	4	11	18	25	1	8
Thursday	3	10	17	24	31	7	14	21	28	5	12	19	26	2	9
Friday	4	11	18	25 1	1 4-5	8 8-9	15	22 12	29	6	13	20	27	3	10
Saturday	5	12	19	26 2-3	2 6-7	9 10-11	16	23	30	7	14	21	28	4	11
					30 Delivery 1		16 Delivery 2			15					

Class	Summary	WORK	DOCUMENTS	
TP01	Introduction		1-Business Intelligence-ENG	
TP02	MultiDim and dimensions		3-MultidimensionalModel & Conceptual DW design	Book 7
TP03	KPI & OLAP operations	Works presentation	2-PerformanceMeasures, BI-project1 & BI-project2	Books 1 and 7
TP04	Kimball Data Marts: mySQL		Development methodology Kimball & BI-project2	Book 9
TP05	EDW Inmon and project		Building the DataWarehouse - project	Book 8
TP06	ETL: Kettle	Delivery of MultiDim and DW (25%)	BI-project3	Book 7
TP07	ETL: project			
TP08	OLAP server: mondrian		OLAP-operations & BI-project4	Book 7
TP09	MDX		MDX	Book 7
TP10	Dashboards		BI-project4	Book 4
TP11	Project			
TP12	Test & work presentations			

Book 1	Harold Kerzner, Project Management Metrics, KPIs, and Dashboards: A Guide to Measuring and Monitoring Project Performance Paperback – 2017
Book 2	Alan Simon, Enterprise Business Intelligence and Data Warehousing: Program Management Essentials Paperback – 1 Dec 2014
Book 3	Ralph Hughes, Agile Data Warehousing for the Enterprise: A Guide for Solution Architects and Project Leaders Paperback – 15 Jul 2015
Book 4	Lawrence Corr and Jim Stagnitto, Agile Data Warehouse Design: Collaborative Dimensional Modeling, from Whiteboard to Star Schema - 24 Nov 2011
Book 5	Brian Larson, Delivering Business Intelligence with Microsoft SQL Server 2016, Fourth Edition Paperback – 16 Nov 2016
Book 6	Rick Sherman, Business Intelligence Guidebook: From Data Integration to Analytics Paperback – 7 Nov 2014
Book 7	Alejandro Vaisman and Esteban Zimanyi, Data Warehouse Systems: Design and Implementation (Data-Centric Systems and Applications) - 2014
Book 8	Ralph Kimball and Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker, The Data Warehouse Lifecycle Toolkit, 2nd Edition - 2007
Book 9	Wiliam H. Inmon, Building the data warehouse - fourth edition, Wiley, 2005

References

References

Main reference

- Vaisman Alejandro; Data warehouse systems. ISBN: 978-3-642-54655-6

Complementar references

- Harold Kerzner; Project Management Metrics, KPIs, and Dashboards: A Guide to Measuring and Monitoring Project Performance, Wiley, 2017. ISBN: 978-1119427285
- Inmon, W. H.; Building the data warehouse. ISBN: 0-471-08130-2
- Kimball Ralph 070; The data warehouse lifecycle toolkit. ISBN: 9781118075043
- Rick Sherman; Business Intelligence Guidebook: From Data Integration to Analytics, Morgan Kaufmann, 2014. ISBN: 978-0124114616
- Lawrence Corr, Jim Stagnitto; Agile Data Warehouse Design: Collaborative Dimensional Modeling, from Whiteboard to Star Schema, DecisionOne Press, 2011. ISBN: 978-0956817204