

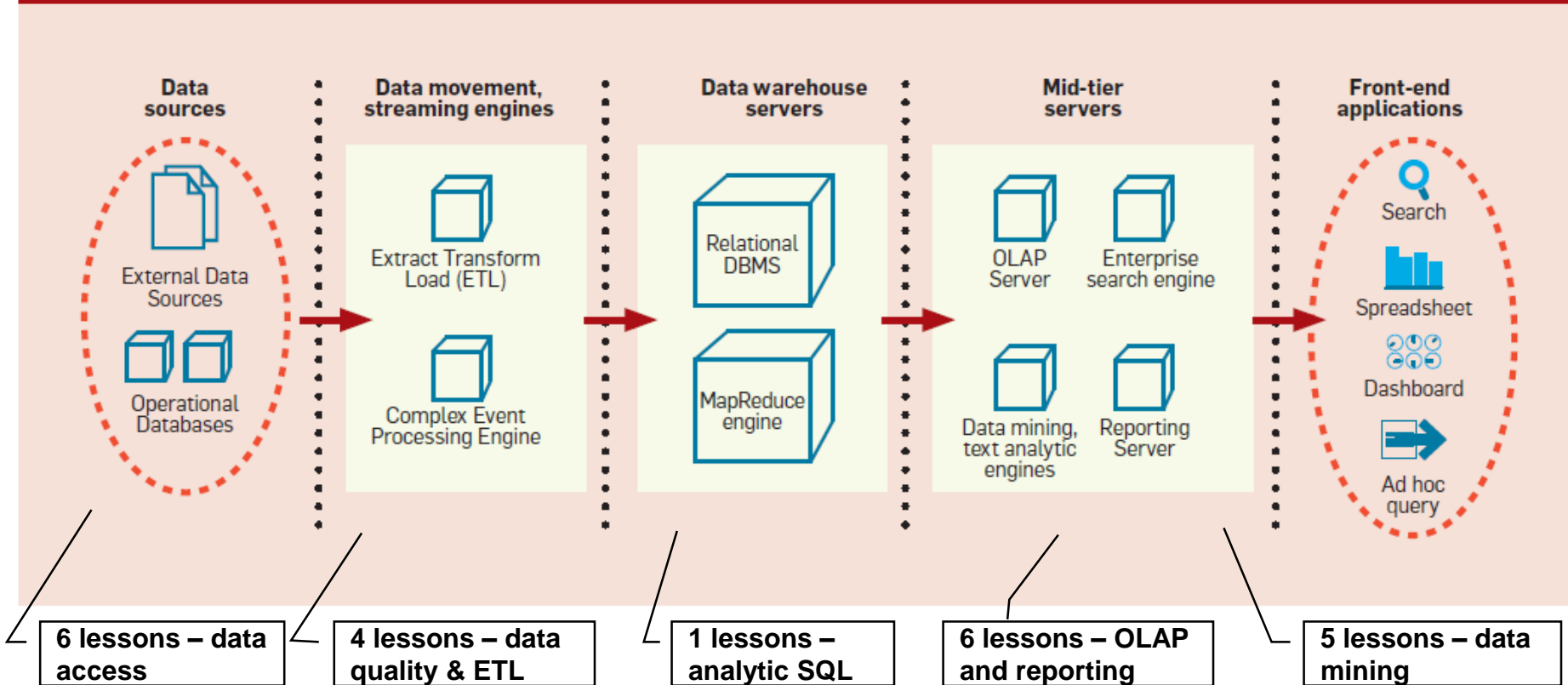
LABORATORY OF DATA SCIENCE

Business Intelligence Architectures

BI Architecture

2

Figure 1. Typical business intelligence architecture.



Data sources

3

- Multiple operational data sources
 - ▣ Across departments of the organization, and external sources
 - ▣ Type and formats
 - Relational, multidimensional, time-series, spatial, text, multimedia, ..
 - ▣ Issues
 - Standards for representations, codes, formats of text files
 - Standards for querying relational data sources
 - Basic programs for data manipulation
- We will study:
 - ▣ Python access to text files
 - ▣ Python access to RDBMS

Extract, Transform and Load

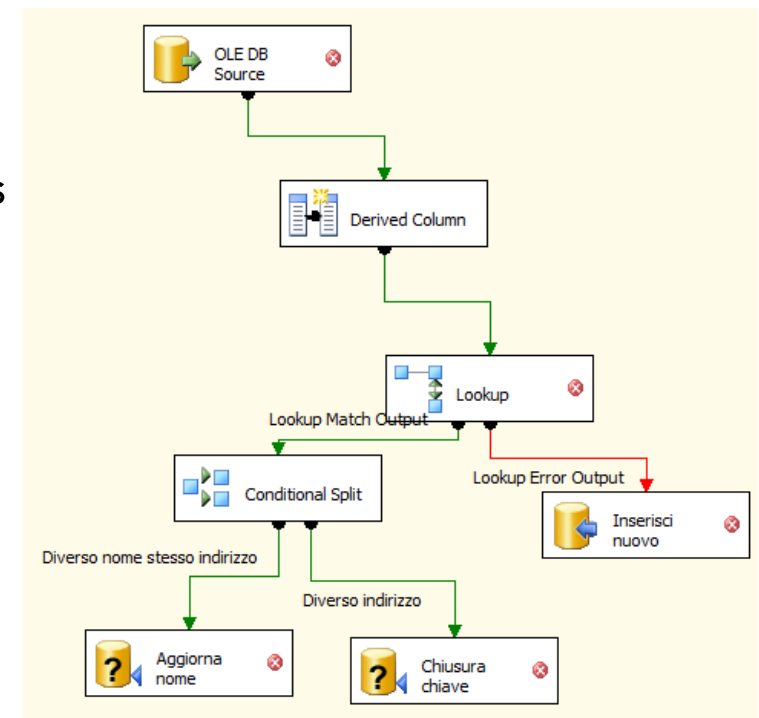
4

ETL (extract transform and load) is the process of extracting, transforming and loading data from heterogeneous sources in a data base/warehouse.

- Typically supported by (**visual**) tools

- We will study:

- SQL Server 2016 Integration Services



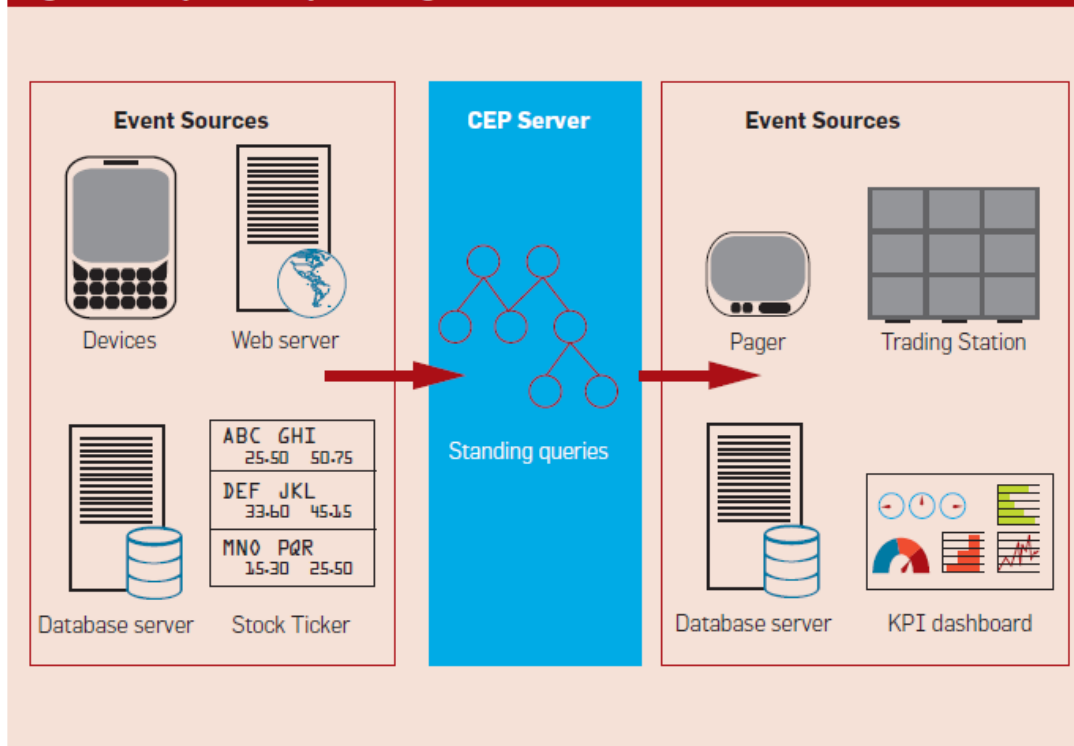
Extract, Transform and Load

5

Incremental and real-time ETL

- Complex Event Processing (CEP)

Figure 4. Complex event processing server architecture.



Data warehouse

6

“A **data warehouse** is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management’s decision-making process.”

W.H. Inmon

- Data warehousing: the process of building and using a datawarehouse

Data warehouse servers

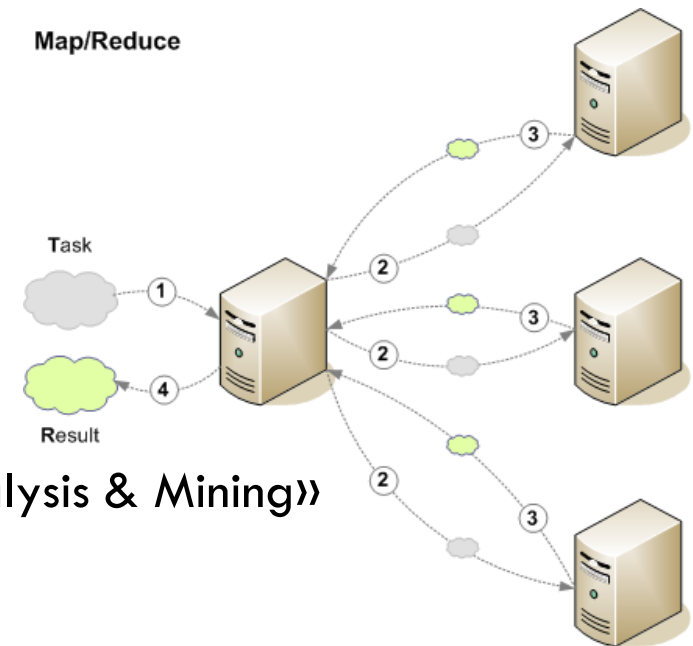
7

□ Relational DBMS (**RDBMS**)

- With specialized index and optimizations
 - star-join query, bitmap index, partitioning, materialized views
- We will study:
 - SQL Server 2016 with analytic SQL

□ MapReduce engine

- Big data challenge
 - Architect (low-cost) data platform
- Covered by 687AA «Distributed Data Analysis & Mining»



Which DBMS for DW?

8

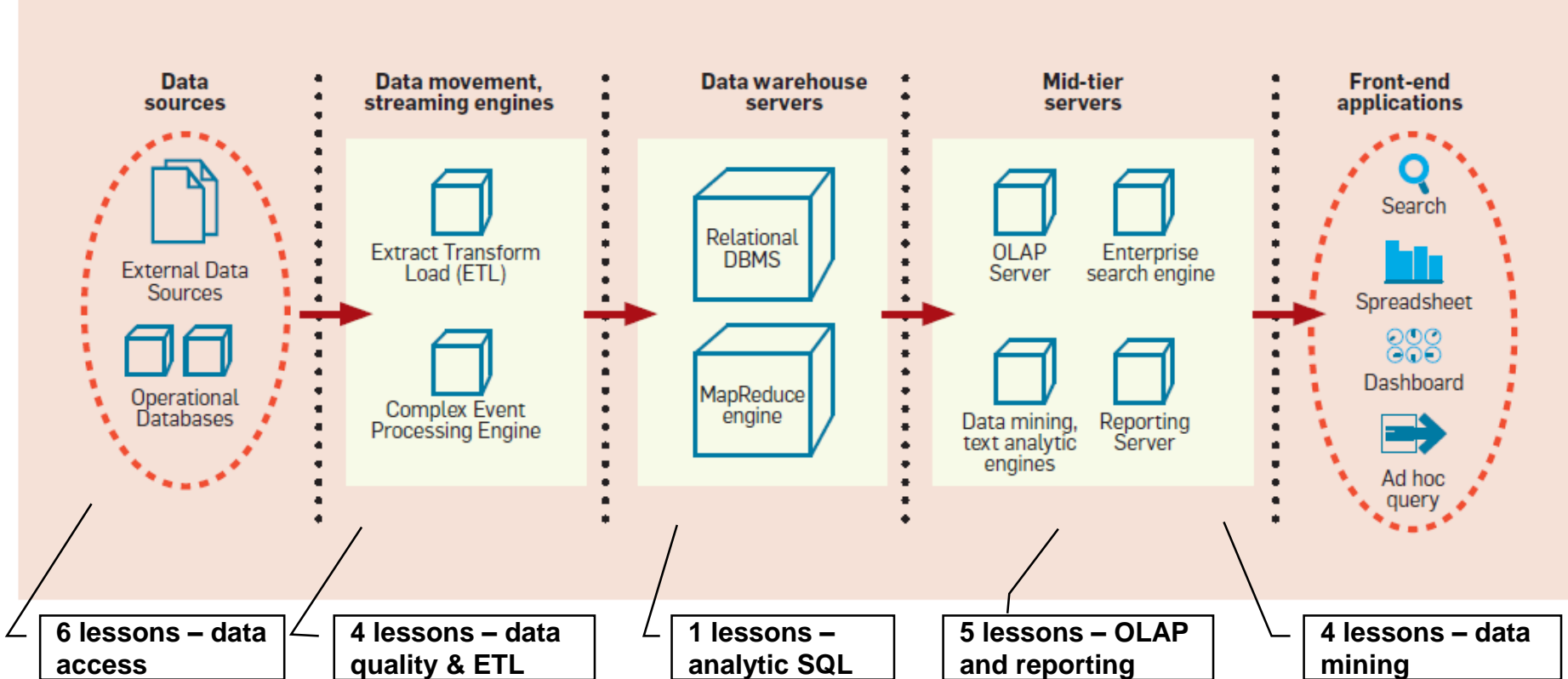
Gartner Magic Quadrant – February 2016



BI Architecture

9

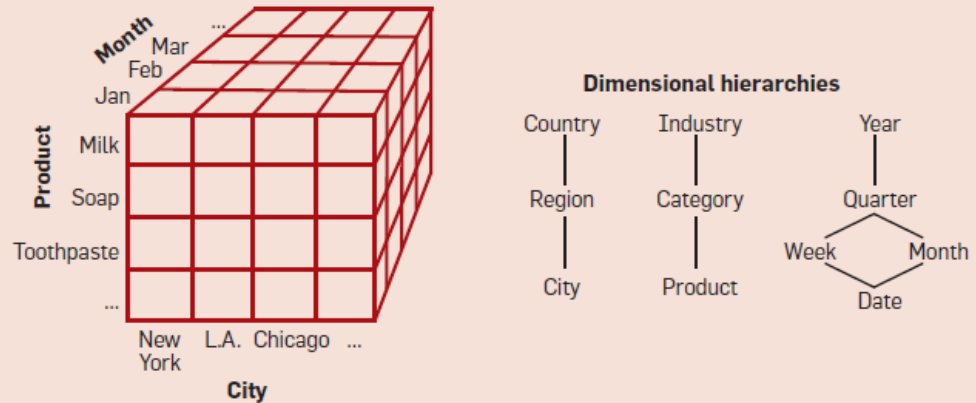
Figure 1. Typical business intelligence architecture.



Mid-tier server

10

Figure 2. Multidimensional data.



- OnLine Analytical Processing (**OLAP**)
 - Provides a multidimensional view of data warehouses
 - Pre-compute aggregates
 - in ad-hoc structures (multidimensional OLAP - MOLAP)
 - in relational DB (relational OLAP - ROLAP)
 - in-memory OLAP
- We will study:
 - SQL Server 2016 Analysis Services and MDX Query Language

Mid-tier servers

11

Reporting Servers

- Enable definition, efficient execution, and rendering of reports
- Data is retrieved from datawarehouse or OLAP servers

We will study:

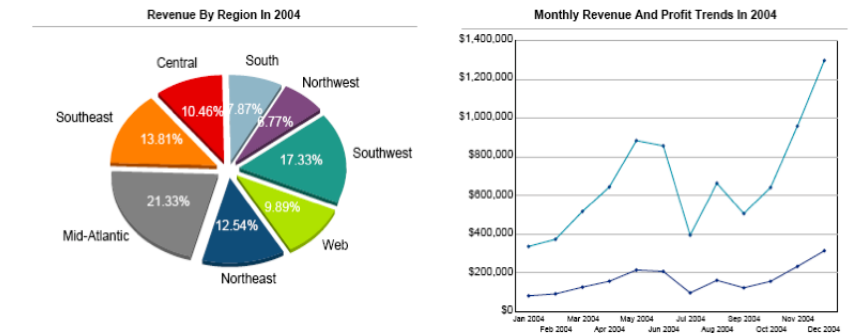
- Microsoft Power BI



Operational Performance Scorecard

Status	Trend	Metrics	Target	This Month	Last Month	%Δ From LM	This Month LY	%Δ From TM LY
▲	▲	Revenue	\$ 1,076,234	\$ 1,296,667	\$957,865	35%	\$1,445,116	-10%
▲	▲	Profit	\$246,777	\$312,376	\$231,740	35%	\$352,001	-11%
◆	▶	Margin	25.78%	24.09%	24.19%	-0.4%	24.36%	-1.1%
▲	▲	Units Sold	26,661	32,122	22,800	41%	34,047	-6%
◆	▲	Order Count	22,919	21,420	13,020	65%	17,000	26%
◆	▼	Avg Revenue per Order	\$139	\$61	\$74	-18%	\$85	-29%
◆	▲	Customer Count	8,300	10,000	8,091	24%	9,380	7%
◆	▲	Avg Revenue per Customer	108	\$130	\$118	10%	\$154	-16%

This Month: Dec 04 Trend ▲ From Prior Month: ▲ More than 10% ▶ Between -5% and 10% ▼ Less than -5% Abbreviations: LM = Last Month LY = Last Year



Mid-tier servers

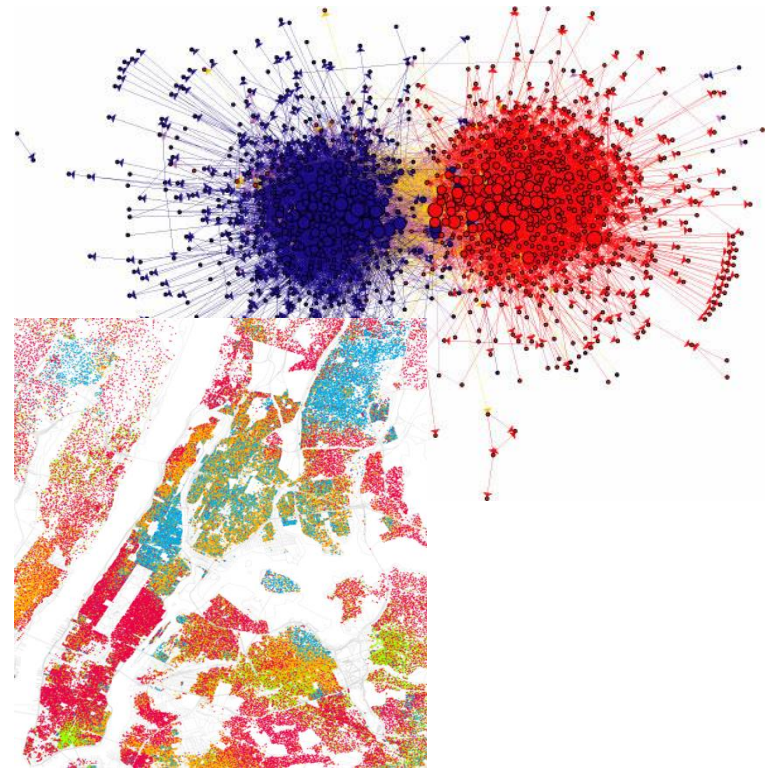
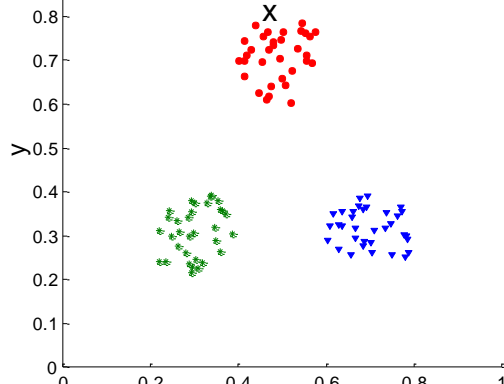
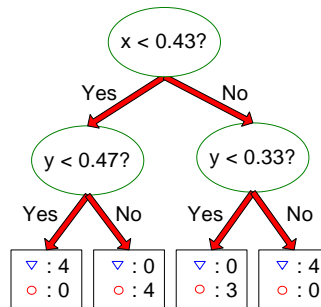
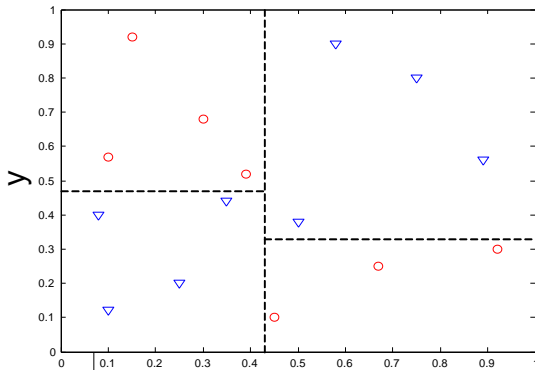
12

□ Data/web/text mining servers

▣ Extract descriptive & predictive models from structured/graph/textual data

□ We will study:

▣ WEKA: a data mining software in Python

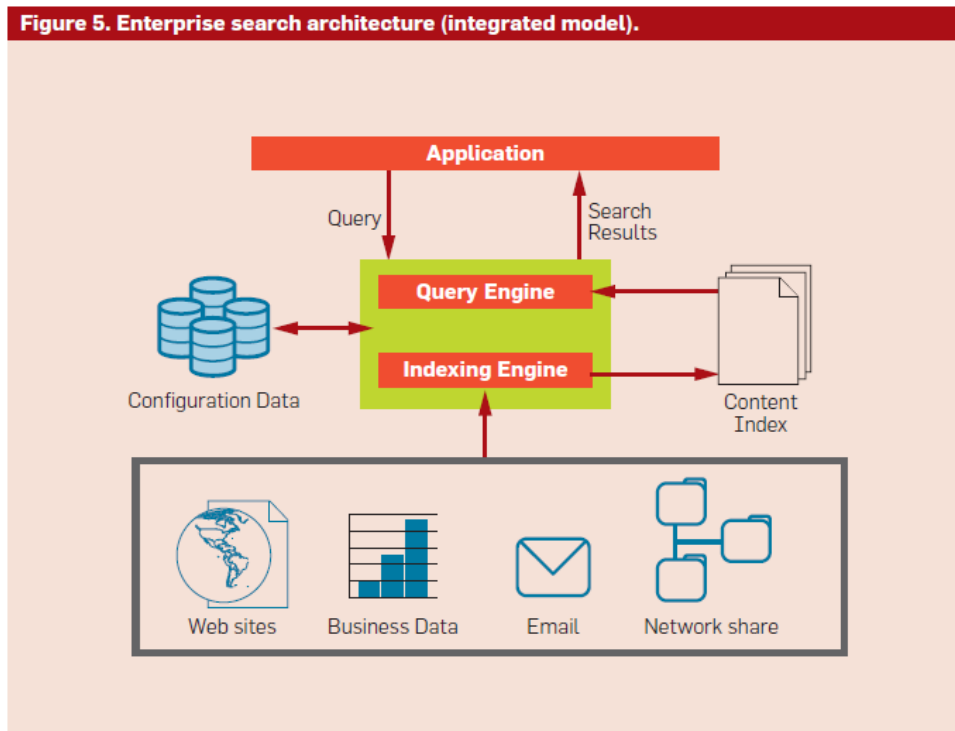


Mid-tier servers

13

□ Enterprise Search Engine

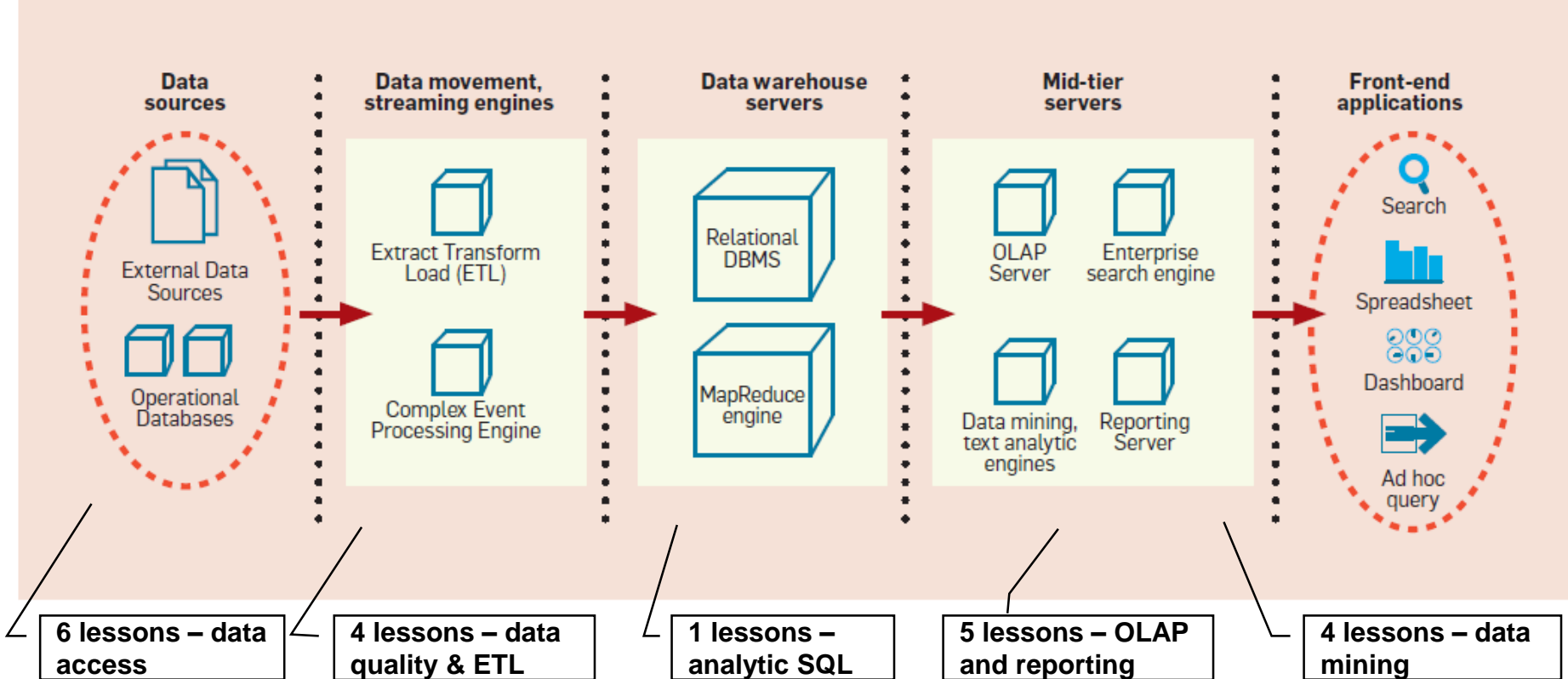
- ▣ Crawl, index and search by keywords over different types of data
- ▣ Covered by 289AA «Information Retrieval»



BI Architecture

14

Figure 1. Typical business intelligence architecture.



Front-end applications

15

- Applications through which users perform BI tasks
 - Spreadsheets
 - for navigating multidimensional data
 - We will study: Excel
 - Enterprise portals
 - for accessing reports and dashboards
 - for searching through query
 - GUI
 - for accessing mining models
 - for exploratory data analysis
 - for ad-hoc queries
 - Vertical packaged applications for CRM, Supply-Chain, Finance, Opinion mining ...
 - More specialized tools for building **storytellings** to produce understandable stories to presents information to the users.
 - Covered by 602AA «Visual Analytics»

Front-end applications

16

A13 fx Totale complessivo						
A	B	C	D	E	F	
1	Product Categories	Bikes				
2						
3	Internet Sales Amount		Country			
4	Calendar Year	Calendar Semester	Calendar Quarter	United Kingdom	United States	Totale complessivo *
5	CY 2001			\$3.266.373,66	\$3.266.373,66	\$3.266.373,66
6	CY 2002	H1 CY 2002		\$3.805.710,59	\$3.805.710,59	\$3.805.710,59
7		H2 CY 2002	Q3 CY 2002	\$1.396.833,62	\$1.396.833,62	\$1.396.833,62
8			Q4 CY 2002	\$1.327.799,32	\$1.327.799,32	\$1.327.799,32
9		H2 CY 2002 Totale *			\$2.724.632,94	\$2.724.632,94
10	CY 2002 Totale *			\$6.530.343,53	\$6.530.343,53	\$6.530.343,53
11	CY 2003			\$9.359.102,62	\$9.359.102,62	\$9.359.102,62
12	CY 2004			\$9.162.324,85	\$9.162.324,85	\$9.162.324,85
13	Totale complessivo *			\$28.318.144,65	\$28.318.144,65	\$28.318.144,65

Elenco campi tabella pivot

Visualizza campi correlati a:

(Tutto)

- Internet Sales Orders
- Altri campi
- Product**
 - Product Categories
 - Product Model Categories
 - Product Model Lines
 - Financial
 - History

Trascinare i campi nelle aree sottostanti:

Filtro rapporto: Product Cate... Geography

Etichette di riga: Date.Calendar Internet Sale...

Etichette di col...: Valori

Rinvia aggiornament... Aggiornamento

