

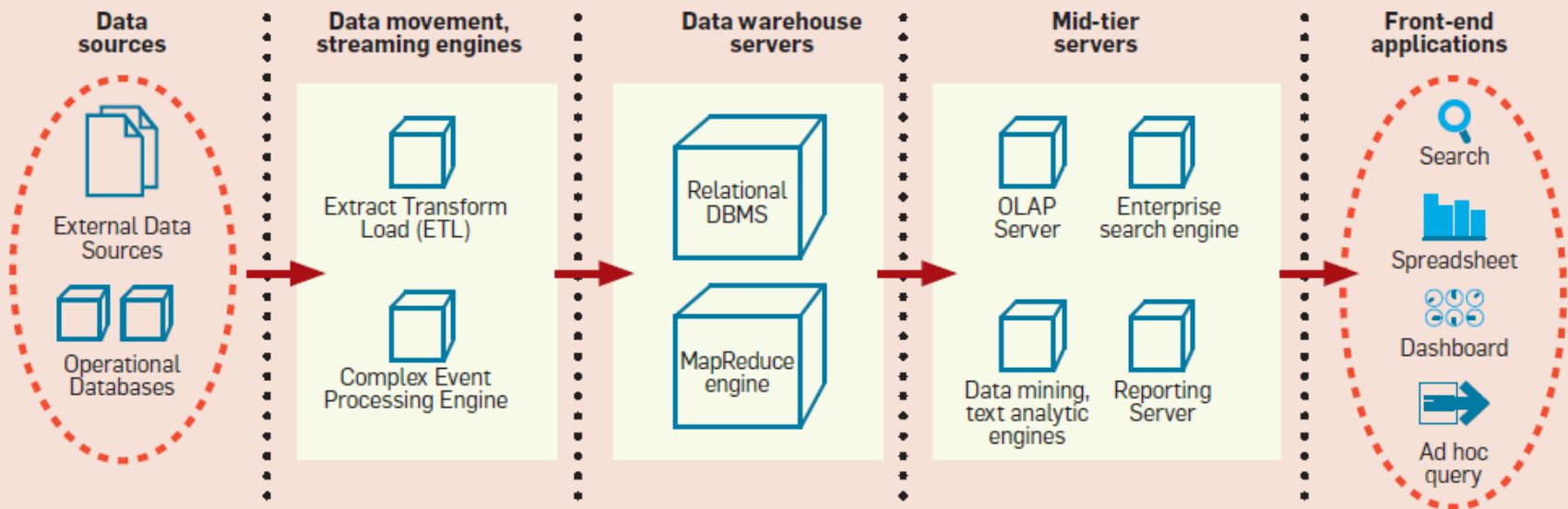
LABORATORY OF DATA SCIENCE

Business Intelligence Architectures

BI Architecture

2

Figure 1. Typical business intelligence architecture.



Data sources

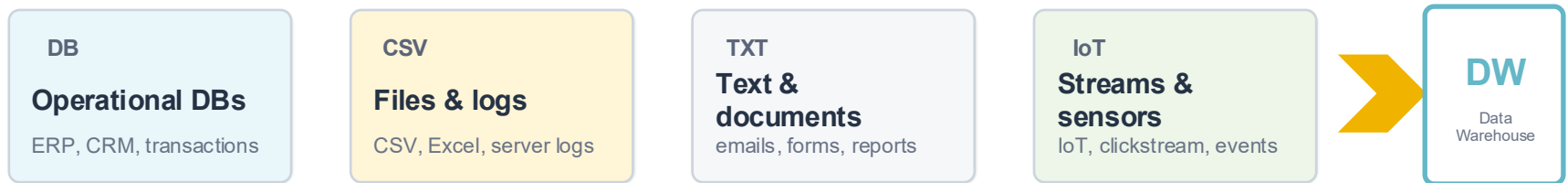
3

- 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

Data sources in BI architectures

4

A BI system rarely starts from one clean database. It integrates heterogeneous operational and external sources into an analytical environment.



Key message: data sources are designed for different purposes; not all of them are immediately suitable for analytical decision support.

Internal vs external data sources

5

Business Intelligence combines data generated inside the organization with external context that helps explain, benchmark, or enrich internal indicators.

Internal data

Data produced by the organization's own processes.

- **ERP / accounting:** orders, invoices, inventory
- **CRM:** customers, campaigns, support
- **Operational DBs:** transactions and master data
- **Logs:** application usage and process traces

External data

Data acquired from outside the organization.

- **APIs:** payments, maps, social platforms
- **Open data:** public statistics and registries
- **Web data:** prices, reviews, competitor signals
- **IoT / sensors:** environmental and operational context

Class question: which external source could improve a sales dashboard, and what data-quality risks would it introduce?

Extract, Transform and Load

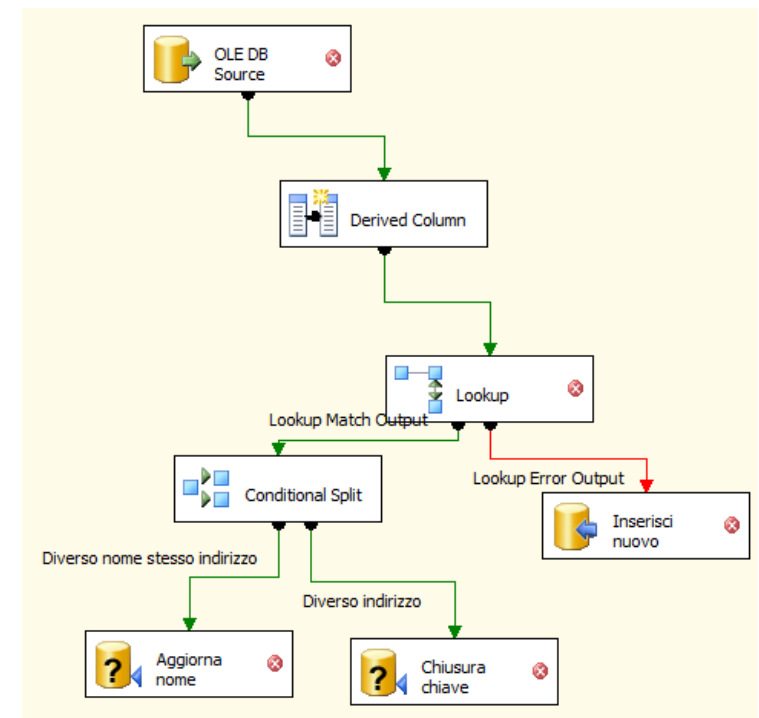
6

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 Integration Services



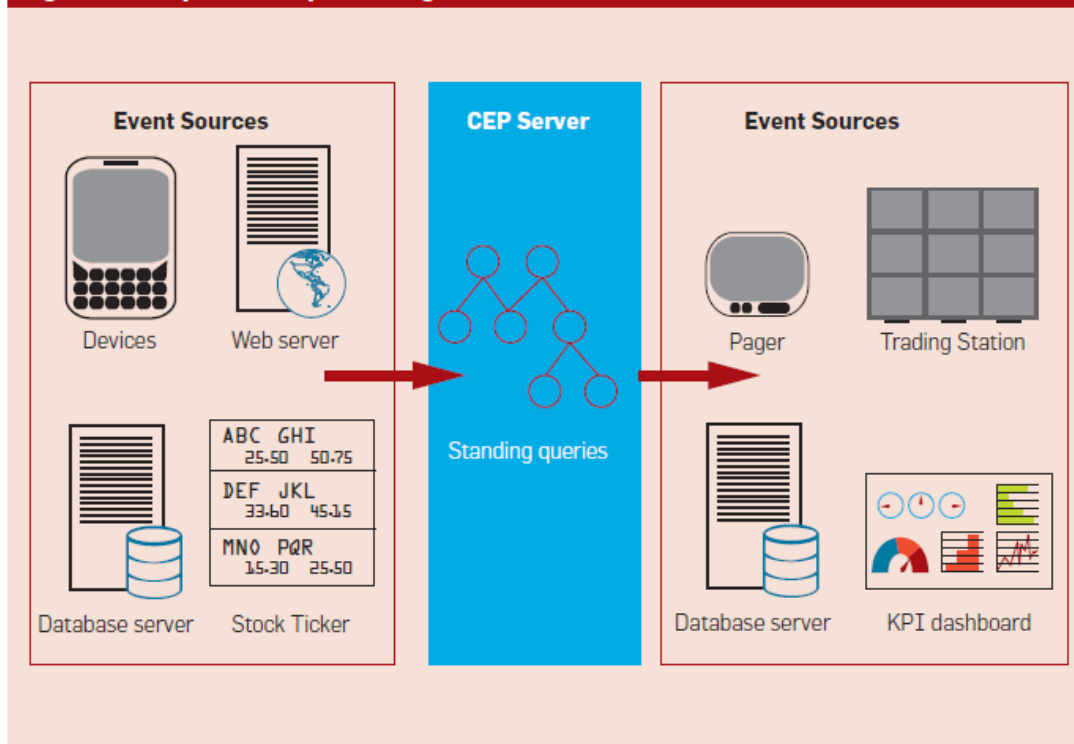
Extract, Transform and Load

7

Incremental and real-time ETL

- Complex Event Processing (CEP)

Figure 4. Complex event processing server architecture.



Data warehouse

8

“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

Operational systems vs analytical systems

9

The Data Warehouse exists because operational systems are optimized to run the business, while analytical systems are optimized to understand the business.

Operational systems

Run the business

Current, detailed data

Many small transactions

INSERT / UPDATE intensive

Normalized schemas

Analytical systems — OLAP

Understand the business

Historical, aggregated data

Complex analytical queries

READ intensive

Dimensional schemas

Typical flow



Data warehouse servers

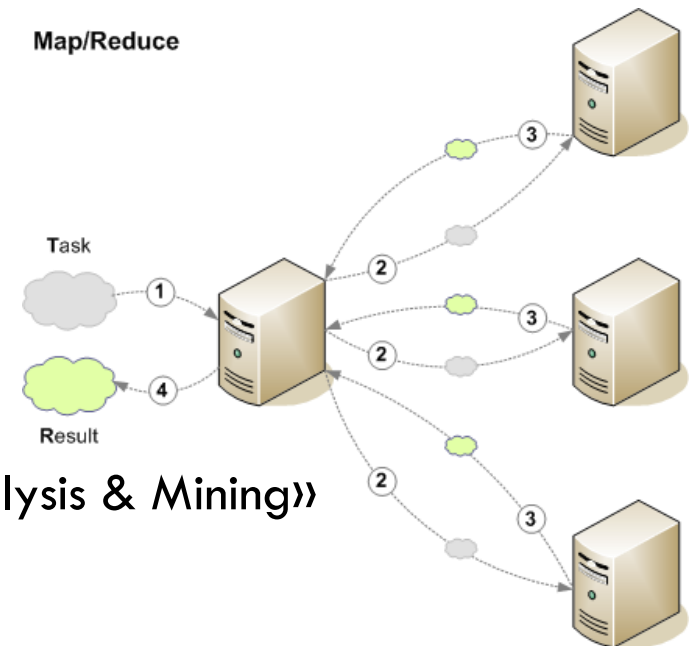
10

□ Relational DBMS (**RDBMS**)

- With specialized index and optimizations
 - star-join query, bitmap index, partitioning, materialized views
- You will see it in the DSD part:
 - SQL Server with analytic SQL

□ MapReduce engine

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



Which DBMS for DW?

11

Gartner names Microsoft a leader in the Magic Quadrant for Data Management Solutions for Analytics



Which BI platform?

12

2024 Magic Quadrant for Analytics and Business Intelligence Platforms

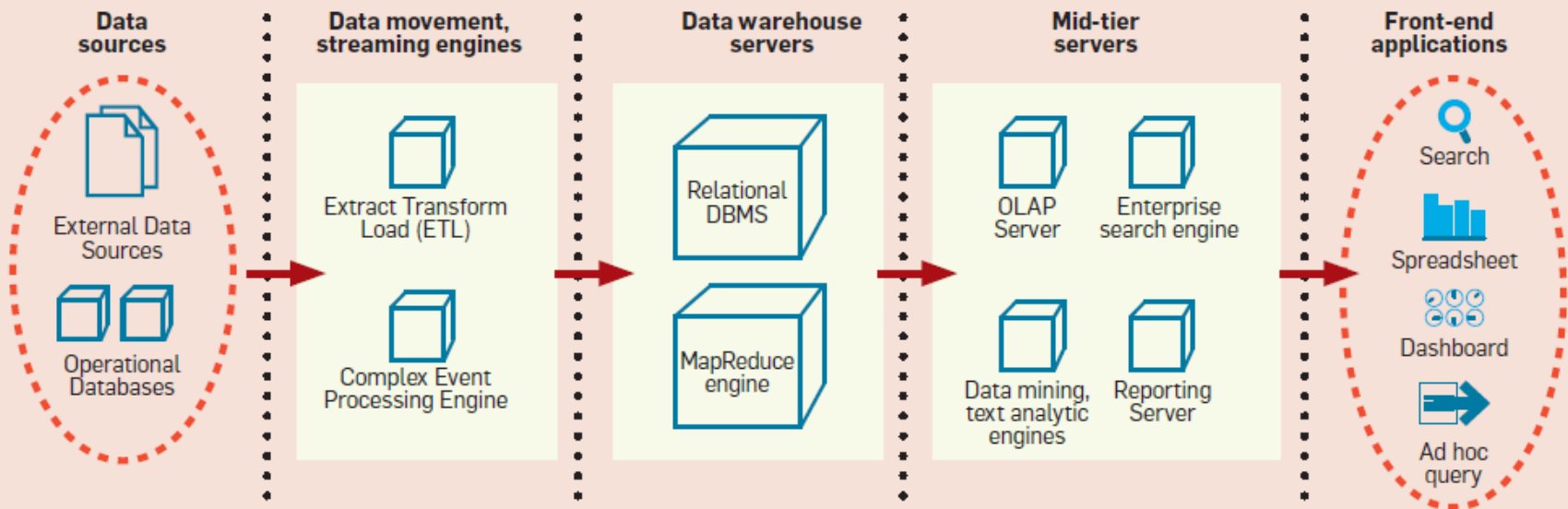
Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms



BI Architecture

13

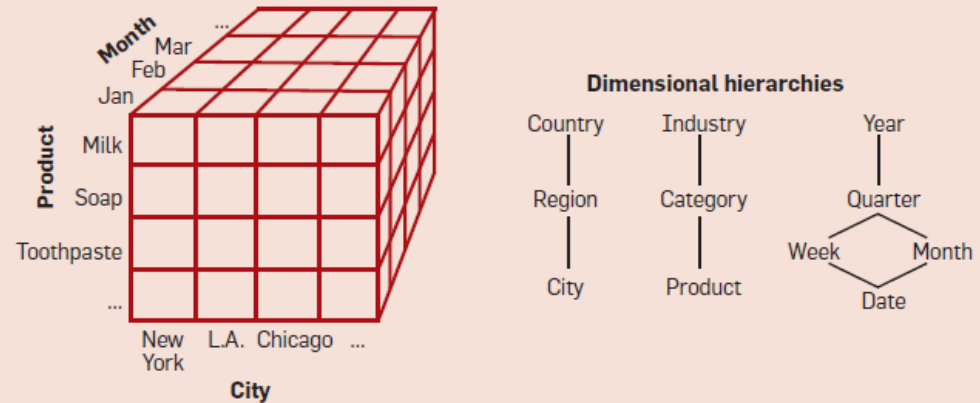
Figure 1. Typical business intelligence architecture.



Mid-tier server

14

Figure 2. Multidimensional data.



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

Mid-tier servers

15

□ Reporting Servers

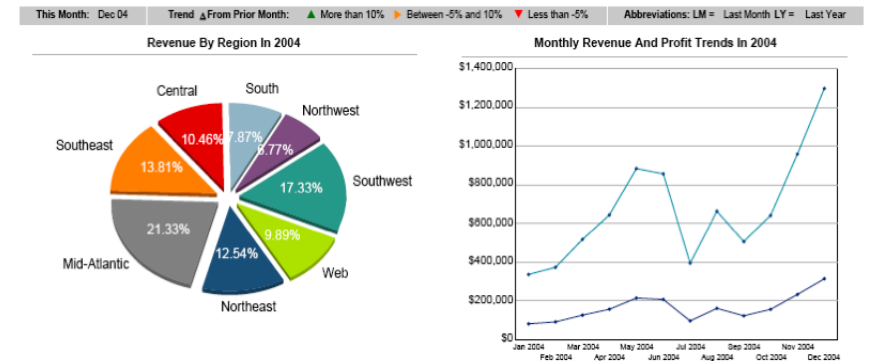
- Enable definition, efficient execution, and rendering of reports
- Data is retrieved from datawarehouse or OLAP servers

□ We will study:

- Microsoft Power BI



Operational Performance Scorecard							
Status	Trend	Metrics	Target	This Month	Last Month	%Δ From LM	%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%



Mid-tier servers

16

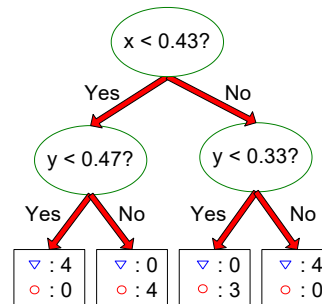
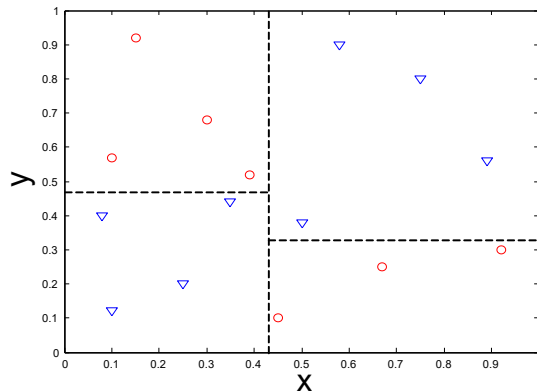
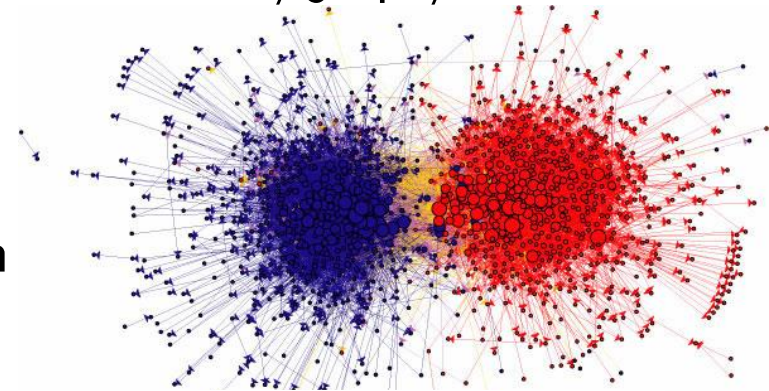
□ Data/web/text mining servers

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

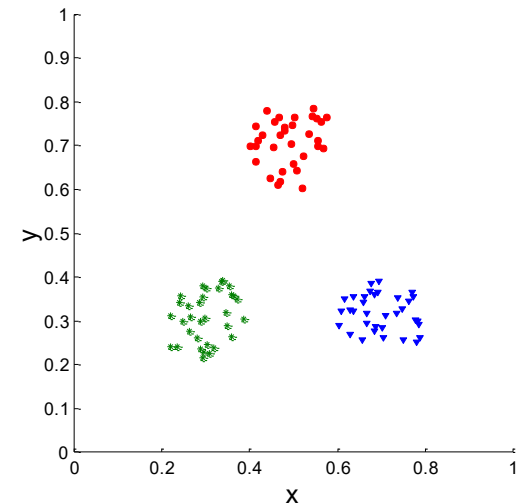
□ We will study:

▣ Azure Machine Learning

▣ How to model a DM & ML problem



Lab of Data Science



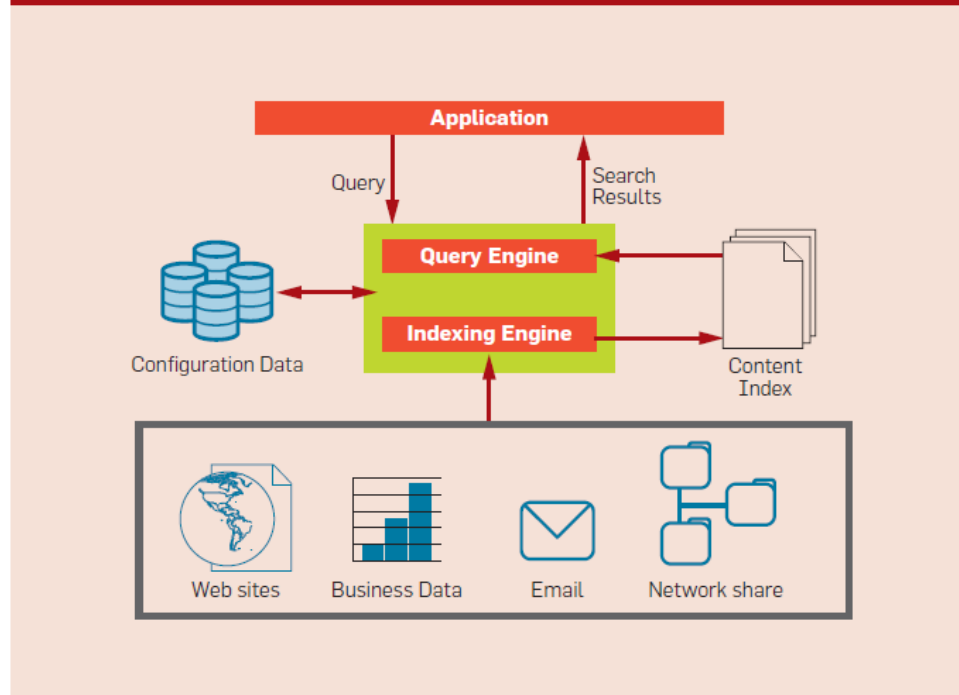
Mid-tier servers

17

□ Enterprise Search Engine

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

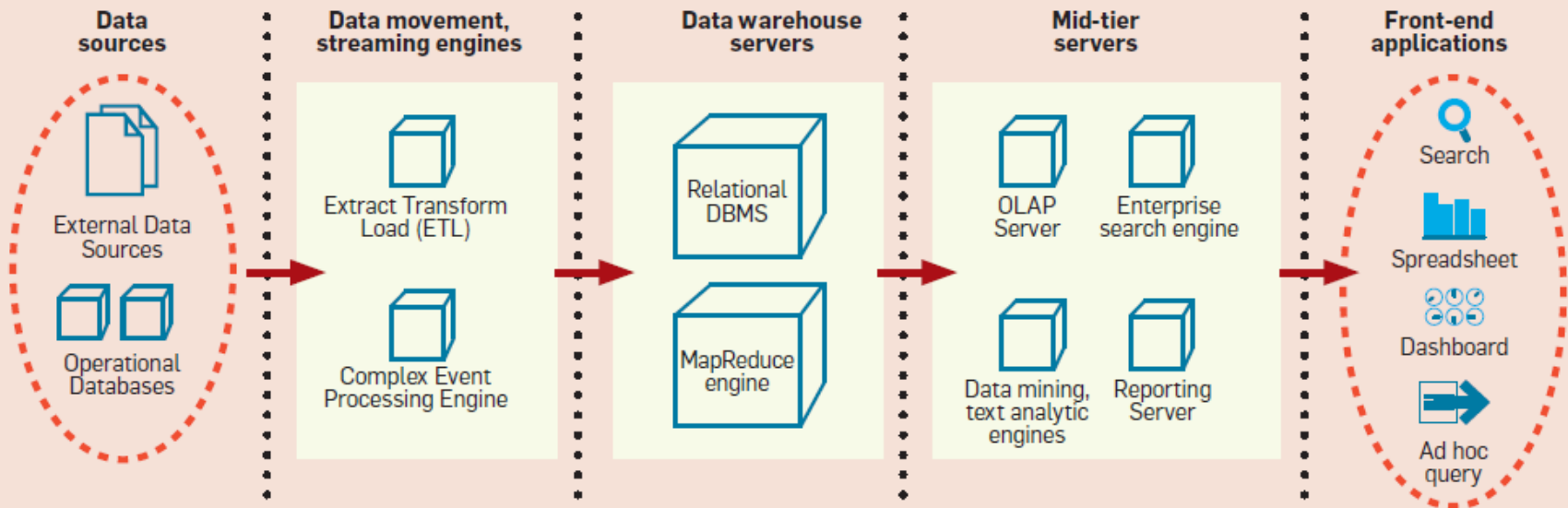
Figure 5. Enterprise search architecture (integrated model).



BI Architecture

18

Figure 1. Typical business intelligence architecture.



Front-end applications

19

- 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

20

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	\$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

Etichette di col...

Product Cate...

Geography

Etichette di riga

Valori

Date.Calendar

Internet Sale...

Rinvia aggiornament...

Aggiornamento

\$10.000.000,00

\$5.000.000,00

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...
Etichette di col...: Geography

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

Rinvia aggiornament...

