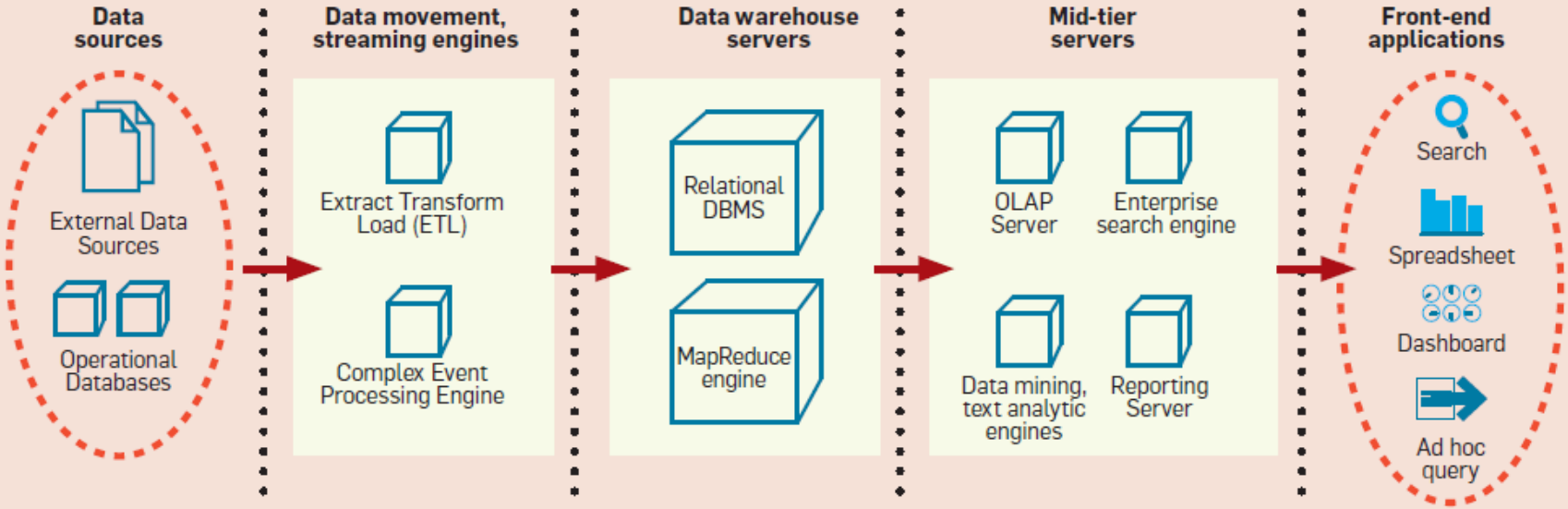
LABORATORY OF DATA SCIENCE

Business Intelligence Architectures

BI Architecture

Figure 1. Typical business intelligence architecture.



Data sources

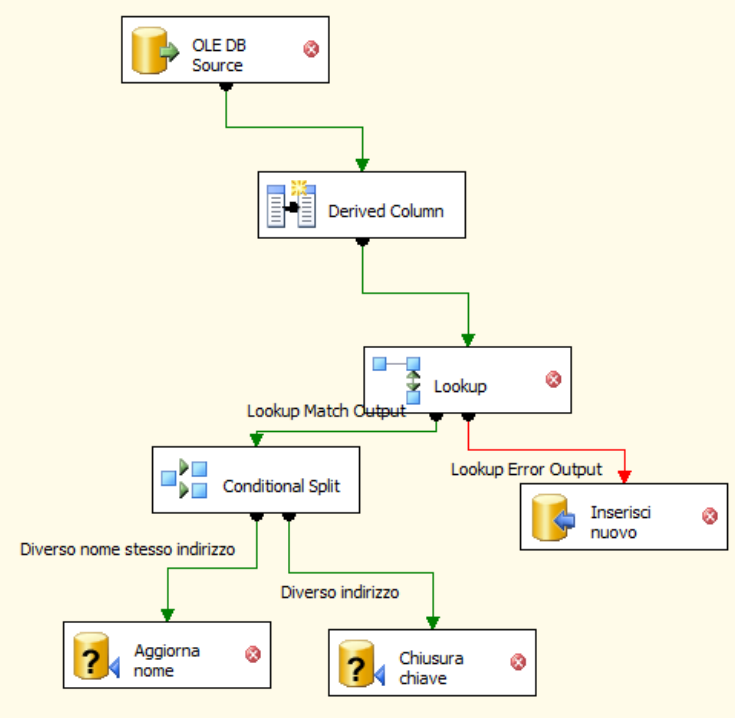
- Multiple operational data sources
  - Across departments of the organization, and external sources
  - Type and formats
    - Relational, multidimensional, time-seriers, spatial, text, multimedia, ..
  - Issues
    - Standards for representations, codes, formats of text files
    - Standards for querying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

Extract, Transform and Load

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

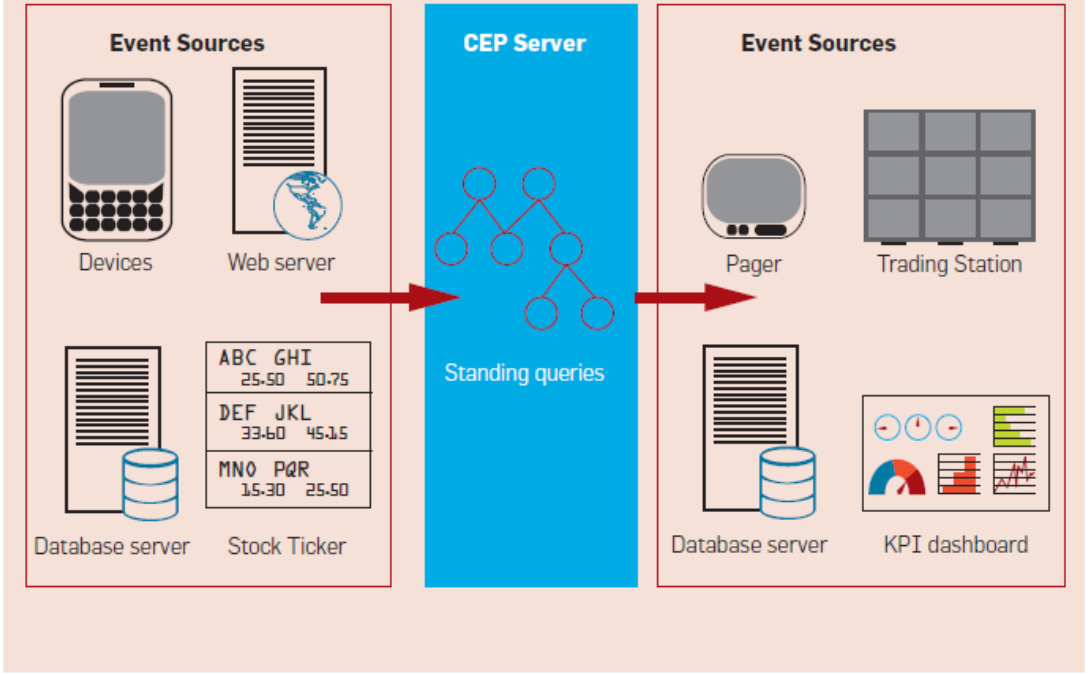


Extract, Transform and Load

Incremental and real-time ETL

- Complex Event Processing (CEP)

Figure 4. Complex event processing server architecture.



Data warehouse

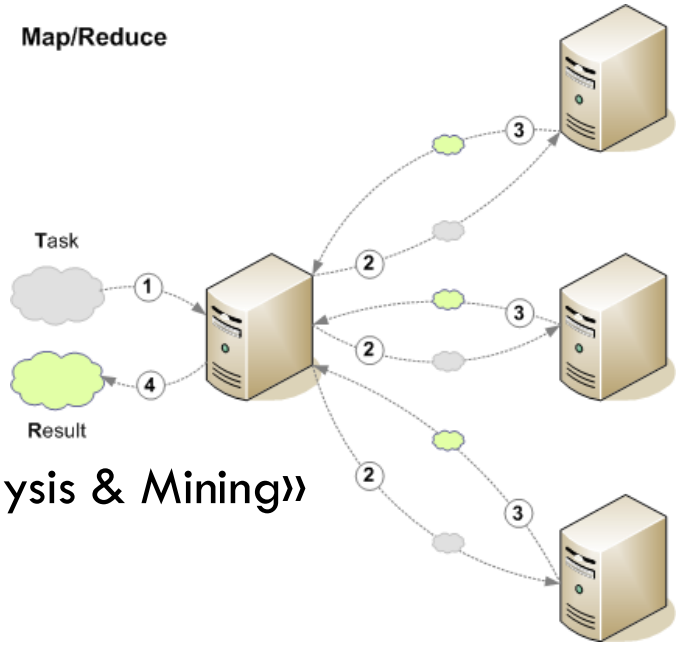
"A **data warehouse** is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision-making process."

W.H. Inmon

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

Data warehouse servers

- Relational DBMS (**RDBMS**)
  - With specialized index and optmizations
    - 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?

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

Which BI platform?

2024 Magic Quadrant **for Analytics and Business Intelligence Platforms**

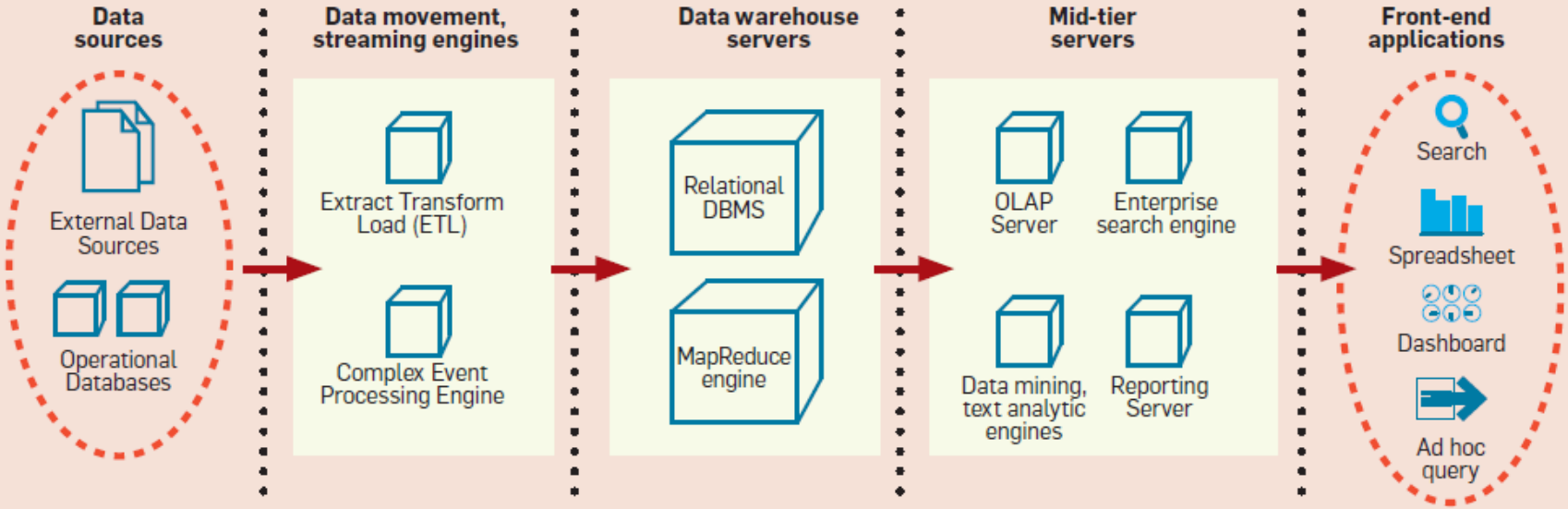
Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms

Gartner (June 2024)

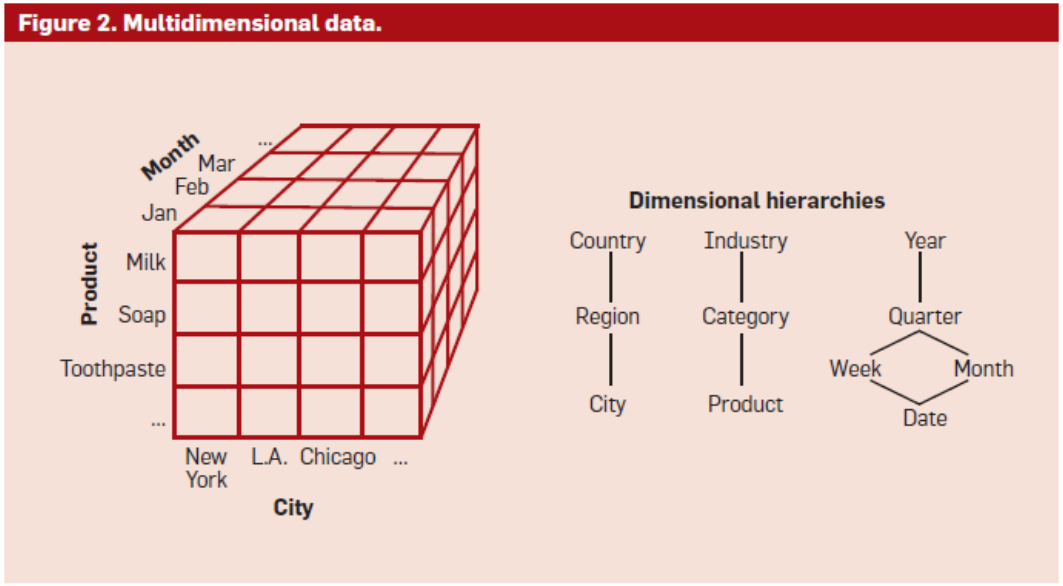
Gartner

BI Architecture

Figure 1. Typical business intelligence architecture.



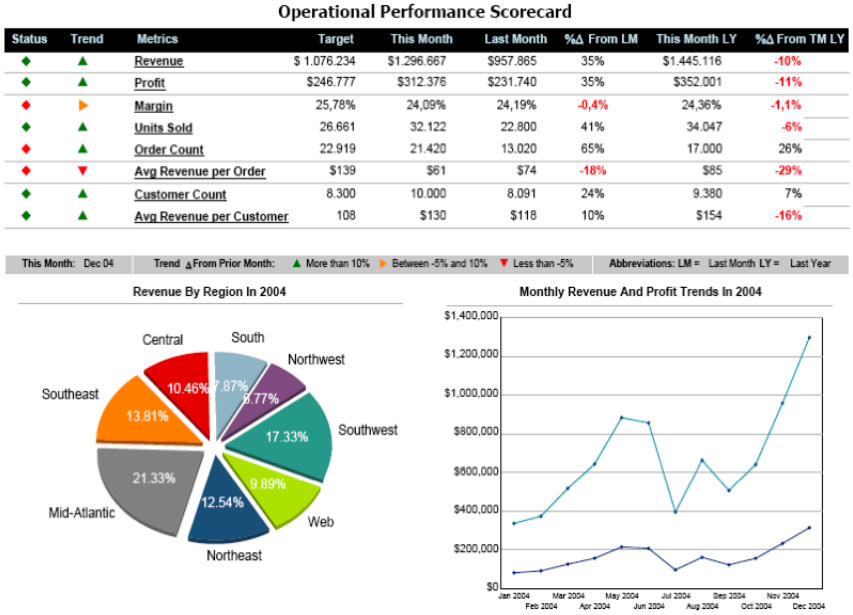
Mid-tier server



- 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

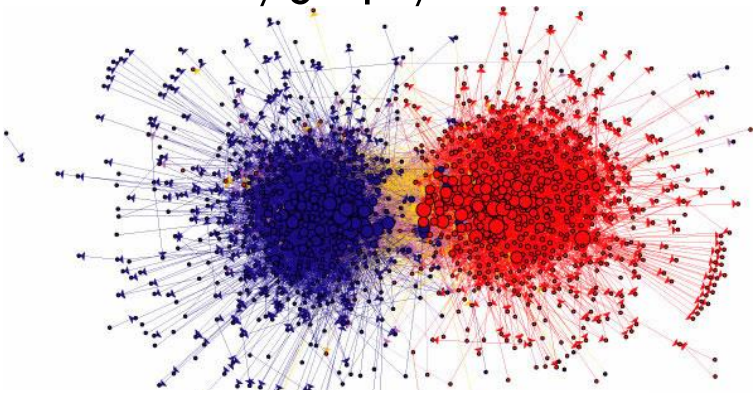
Mid-tier servers

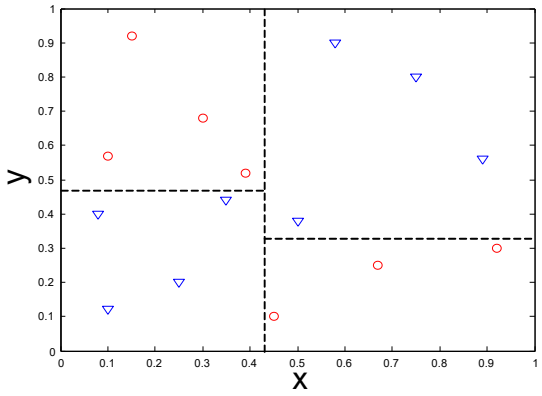
- **Reporting** Servers
  - Enable definition, efficient execution, and rendering of reports
  - Data is retrieved from datawarehouse or OLAP servers
- We will study:
  - Microsoft Power BI

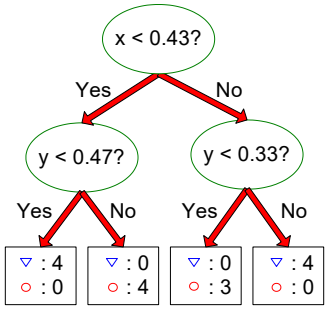


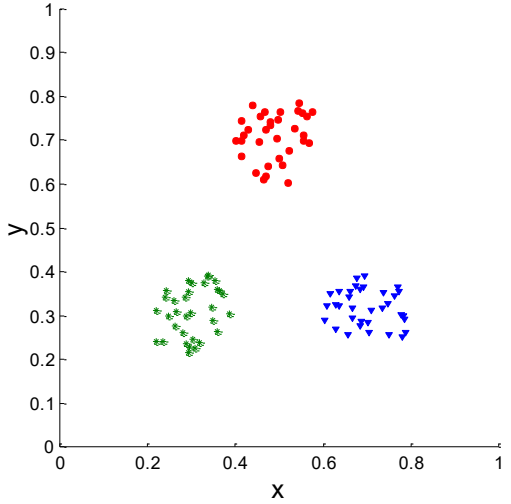
Mid-tier servers

- **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



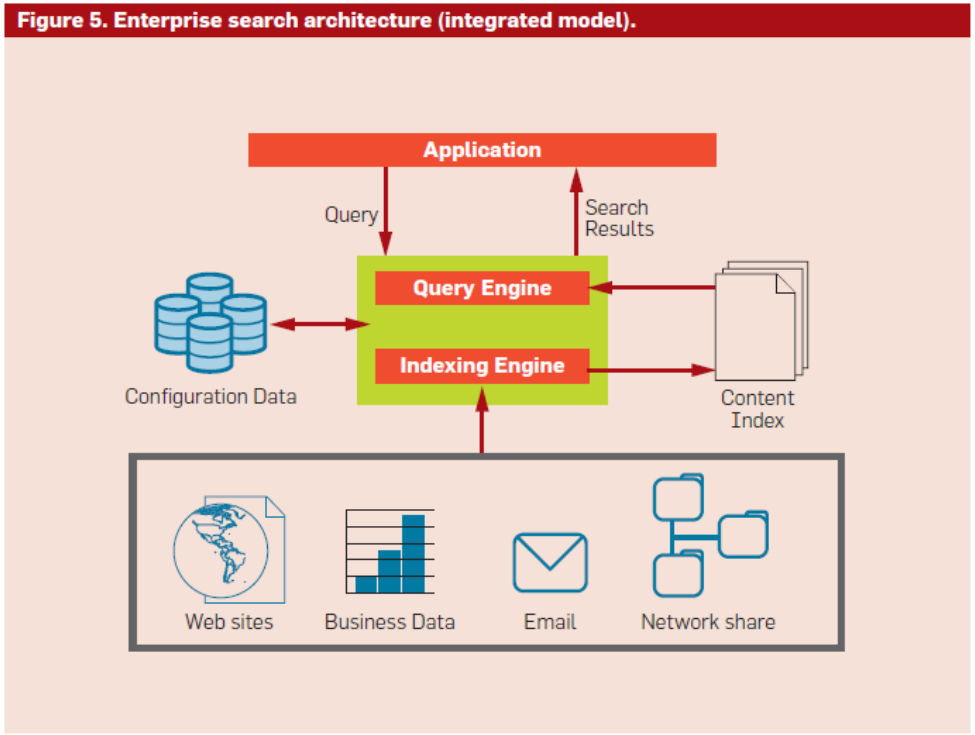






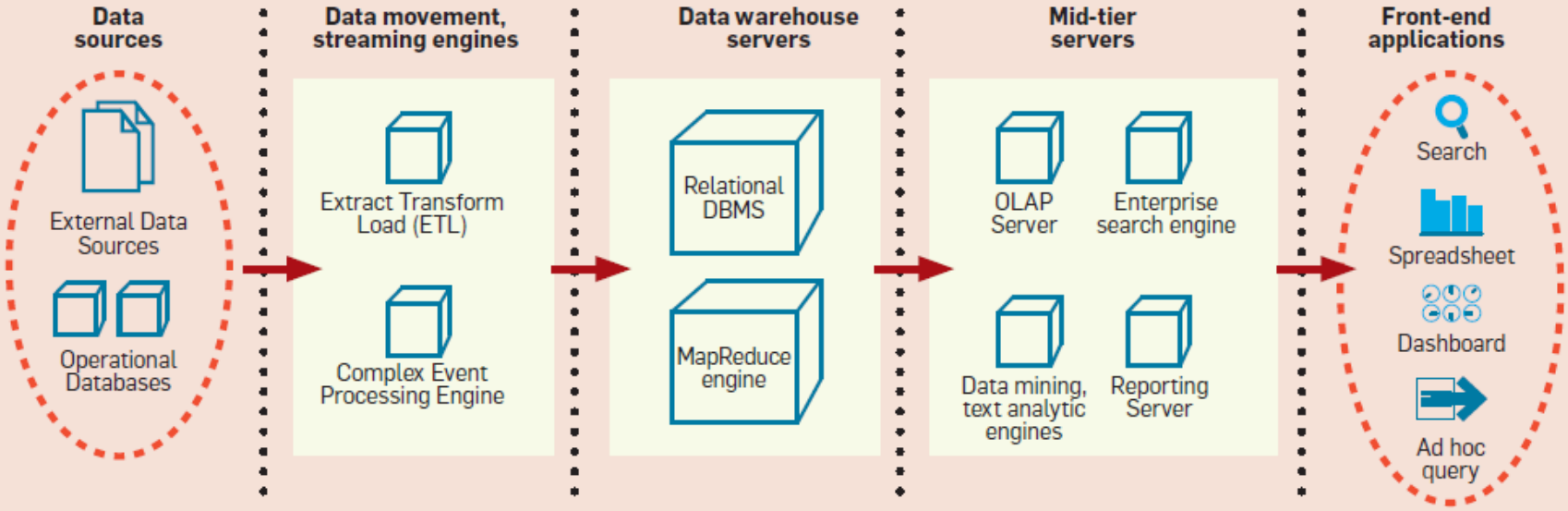
Mid-tier servers

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



BI Architecture

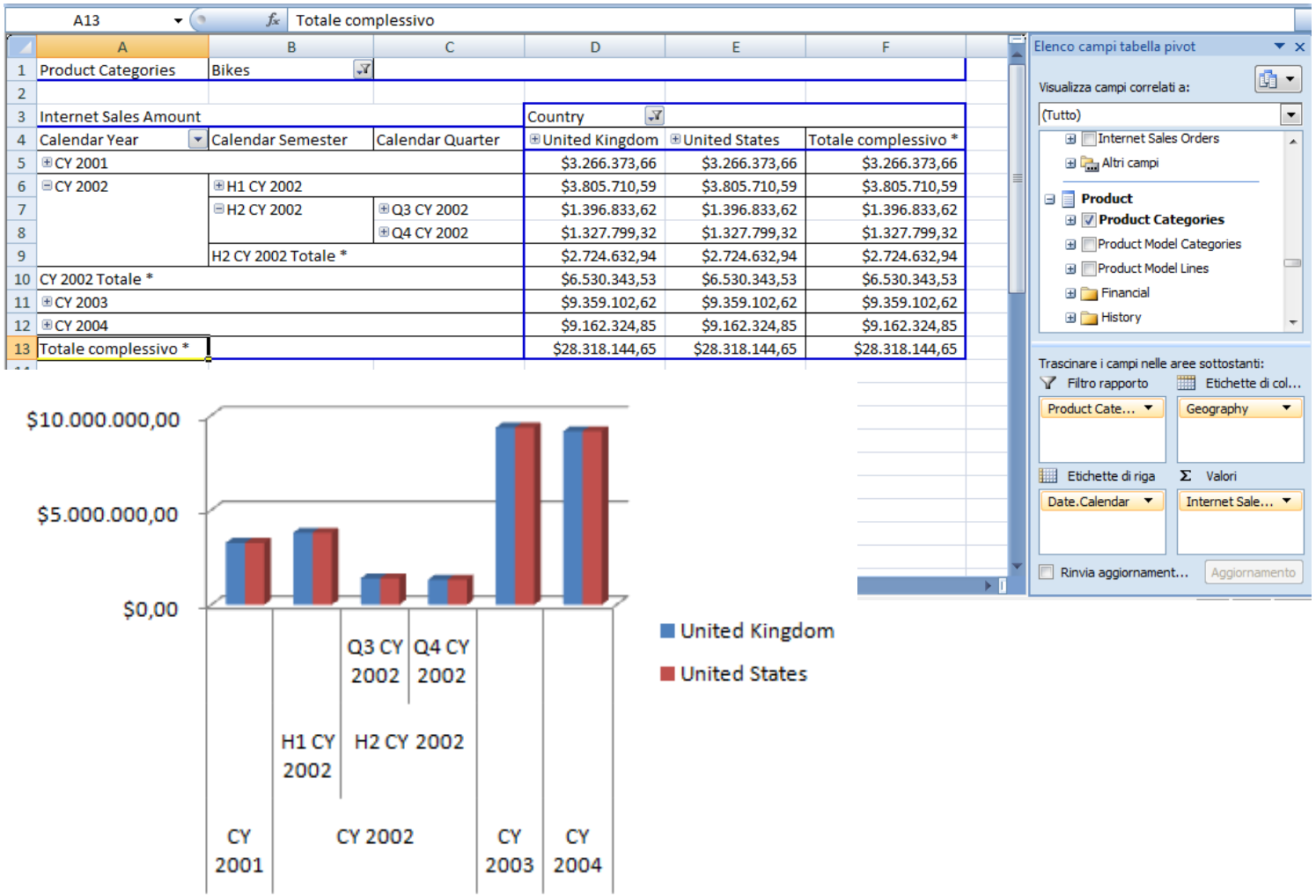
Figure 1. Typical business intelligence architecture.



Front-end applications

- 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



A13 | 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: Product Cate...
Etichette di col...: Geography
Etichette di riga: Date.Calendar
Valori: Internet Sale...

Rinvia aggiornament... | Aggiornamento