LABORATORY OF DATA SCIENCE
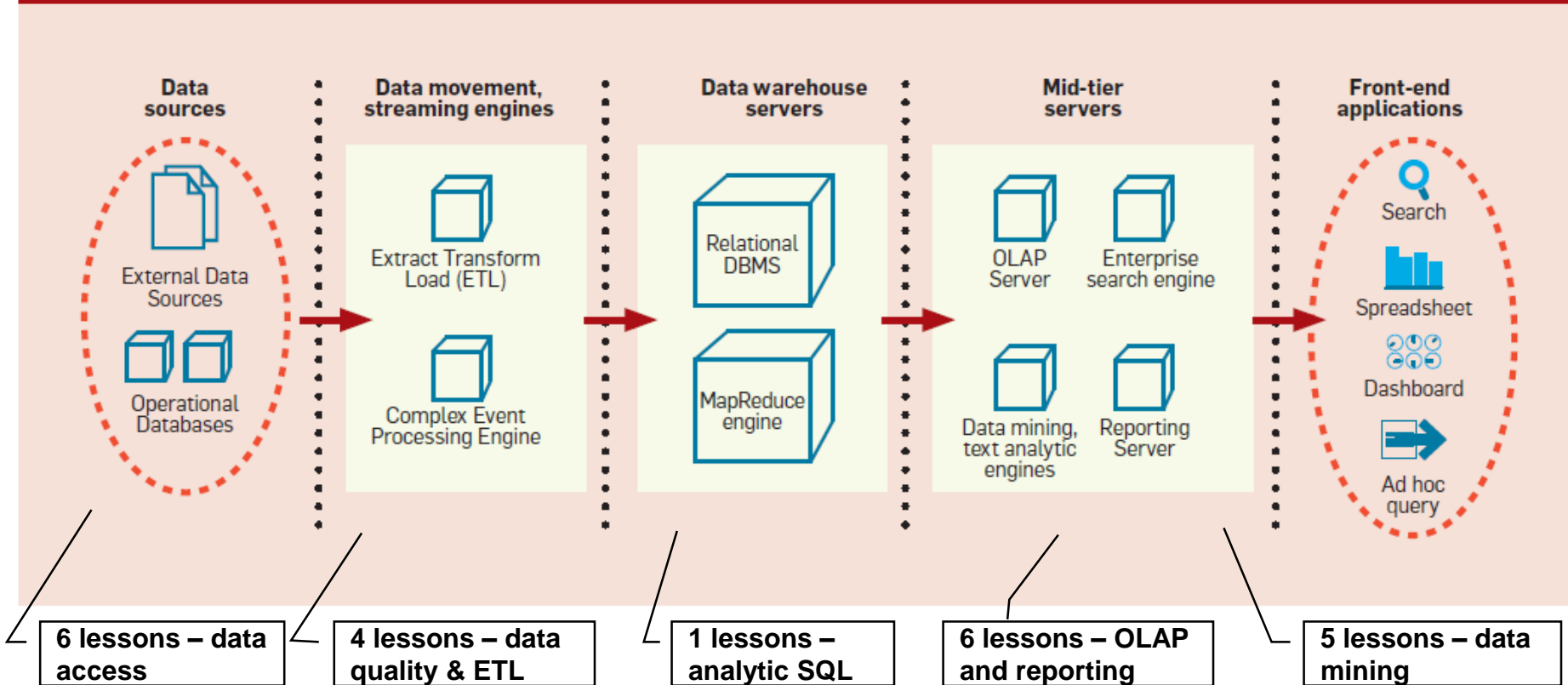
Business Intelligence Architectures

BI Architecture

Figure 1. Typical business intelligence architecture.

6 lessons – data access

4 lessons – data quality & ETL

1 lessons – analytic SQL

6 lessons – OLAP and reporting

5 lessons – data mining

Lab of Data Science

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
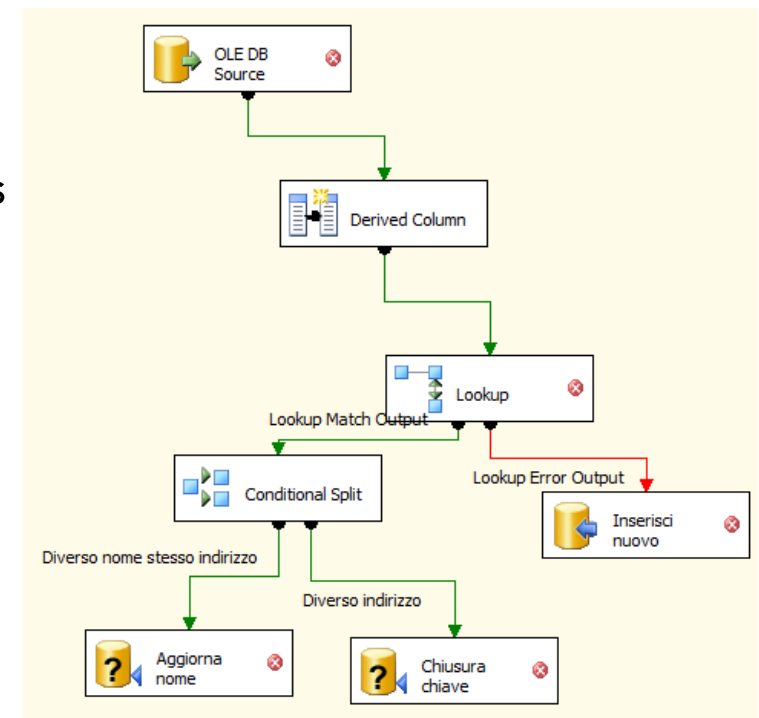 - Python access to RDBMS

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:
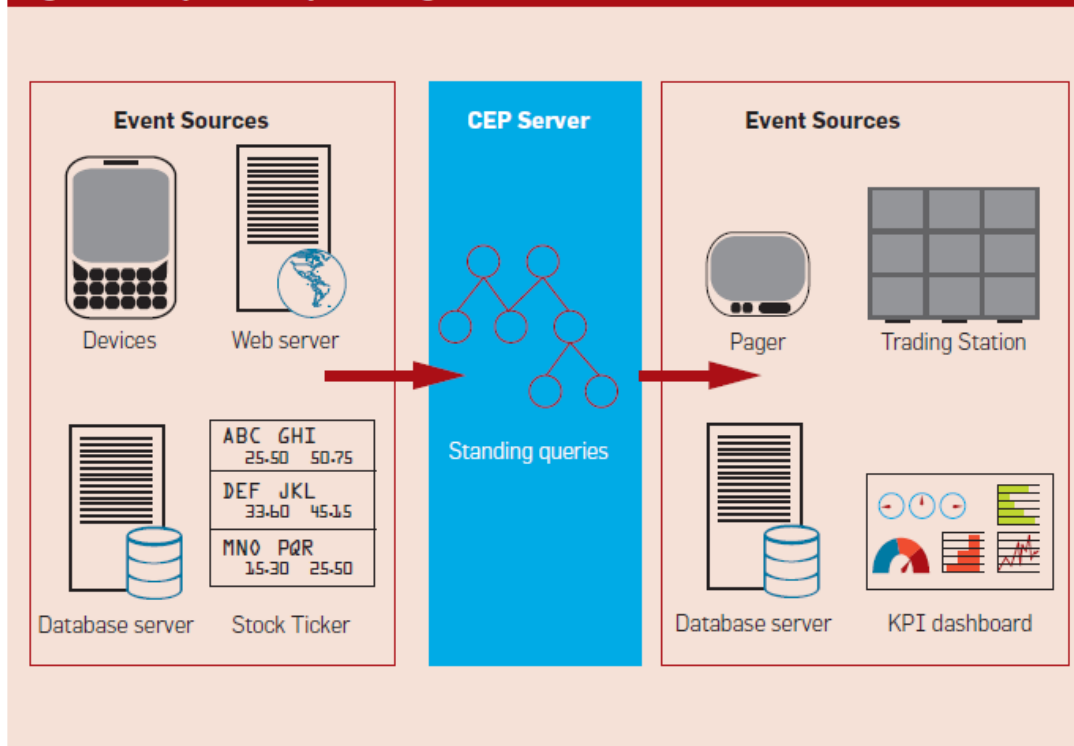 - SQL Server 2016 Integration Services

Lab of Data Science

Extract, Transform and Load

Incremental and real-time ETL

- ◻ Complex Event Processing (CEP)

Figure 4. Complex event processing server architecture.

Lab of Data Science

Data warehouse

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

W.H. Inmon

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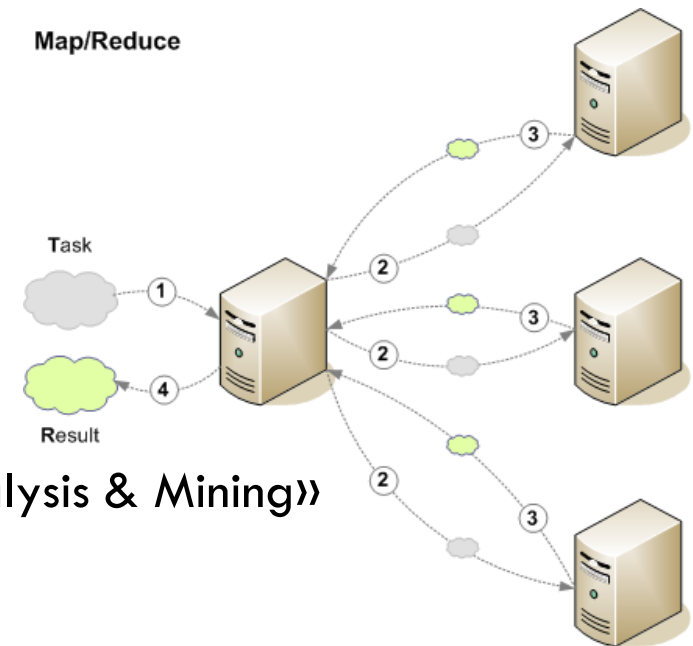
Data warehouse servers

- Relational DBMS (**RDBMS**)
  - With specialized index and optmizations
    - star-join query, bitmap index, partitioning, materialized views
  - We will study:
    - SQL Server 2016 with analytic SQL

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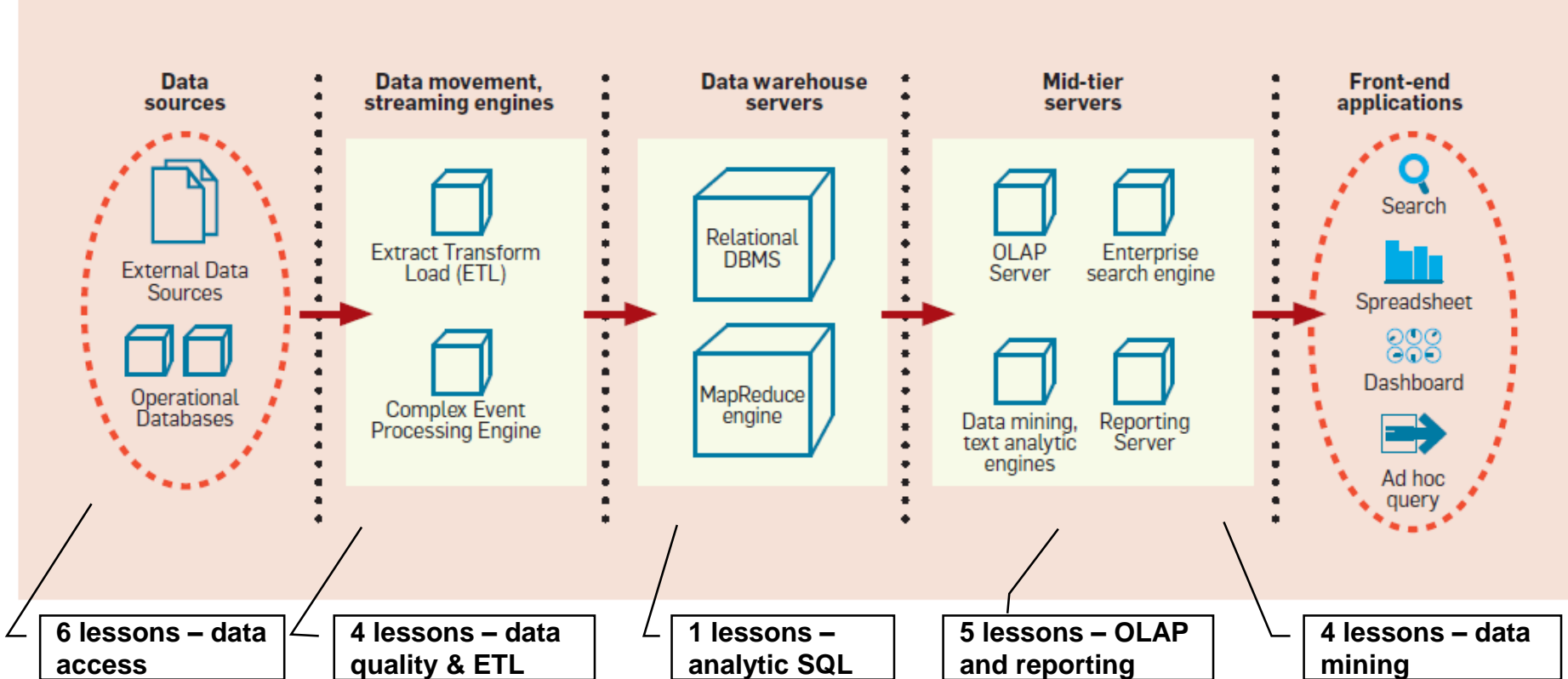 - Covered by 687AA «Distributed Data Analysis & Mining»

Map/Reduce

Task

Result

Which DBMS for DW?

Gartner Magic Quadrant – February 2016

Lab of Data Science

BI Architecture

Figure 1. Typical business intelligence architecture.

Data sources	Data movement, streaming engines	Data warehouse servers	Mid-tier servers	Front-end applications
External Data Sources / Operational Databases	Extract Transform Load (ETL) / Complex Event Processing Engine	Relational DBMS / MapReduce engine	OLAP Server / Enterprise search engine / Data mining, text analytic engines / Reporting Server	Search / Spreadsheet / Dashboard / Ad hoc query

6 lessons – data access

4 lessons – data quality & ETL

1 lessons – analytic SQL

5 lessons – OLAP and reporting

4 lessons – data mining

Lab of Data Science

Mid-tier server

Figure 2. Multidimensional data.
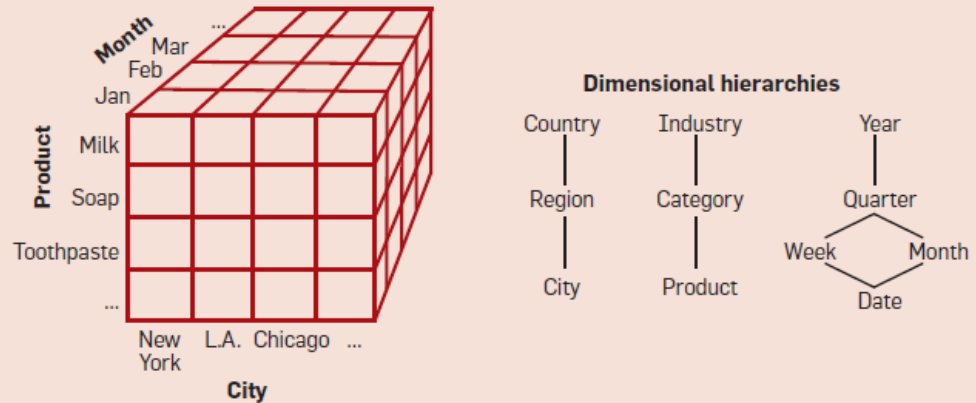
- ☐ OnLine Analytical Processing (**OLAP**)
  - ▫ Provides a multidimensional view of data warehouses
  - ▫ Pre-compute aggregates
    - ■ in ad-hoc structures (multidimensional OLAP - MOLAP)
    - ■ in relational DB (relational OLAP - ROLAP)
    - ■ in-memory OLAP
- ☐ We will study:
  - ▫ SQL Server 2016 Analysis Services and MDX Query Language

Lab of Data Science

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

Lab of Data Science
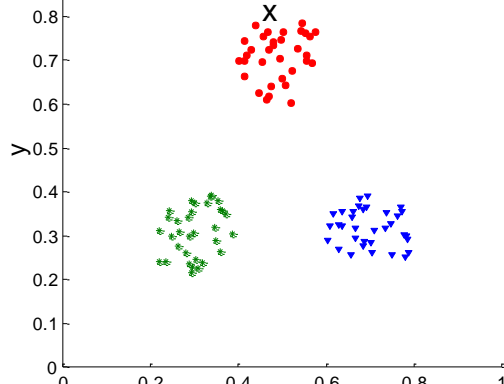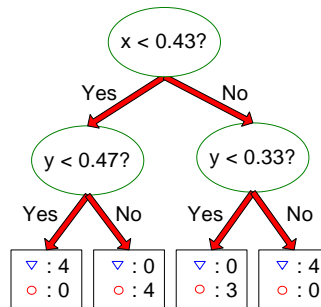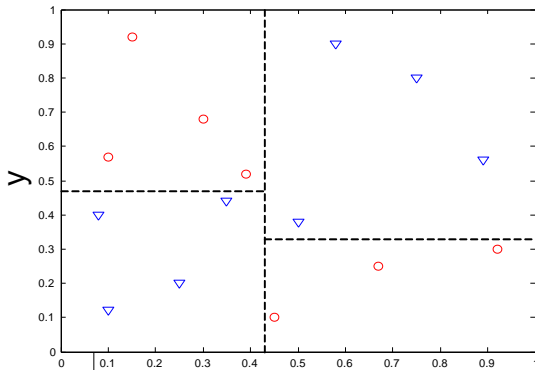
Mid-tier servers

□ **Data/web/text mining** servers

- ❑ Extract descriptive & predictive models from structured/graph/textual data

□ We will study:

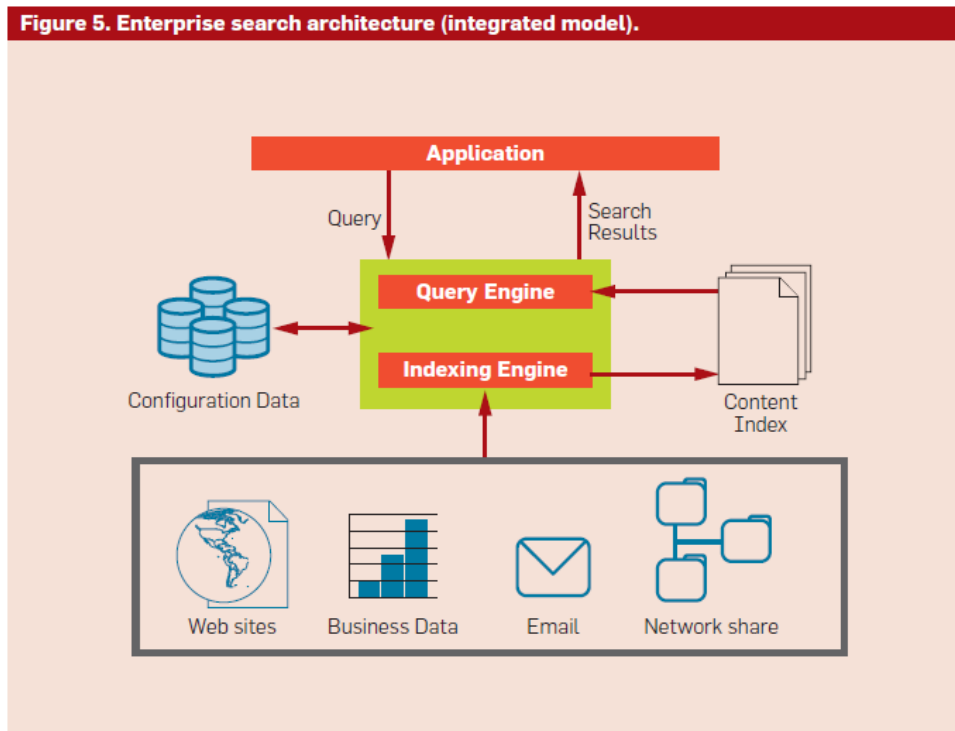- ❑ WEKA: a data mining software in Python
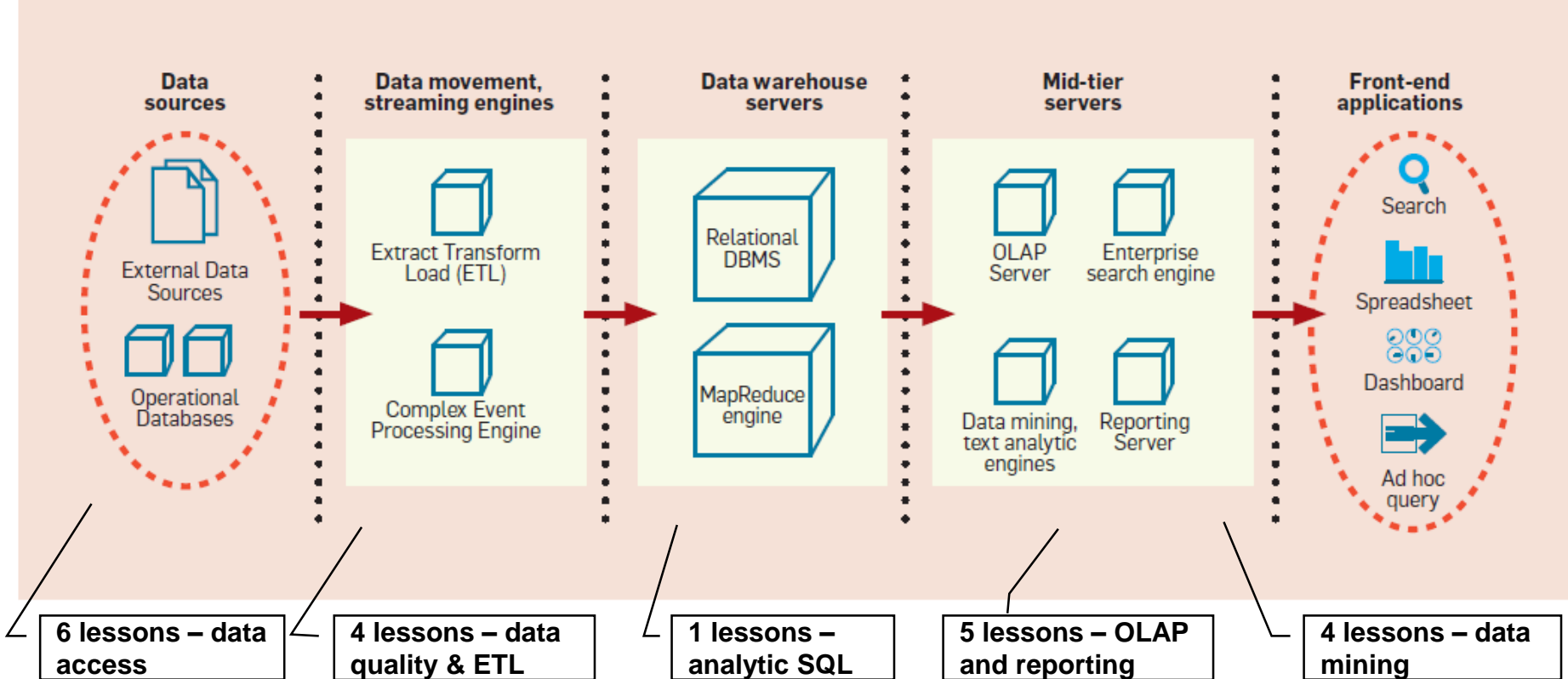
Lab of Data Science
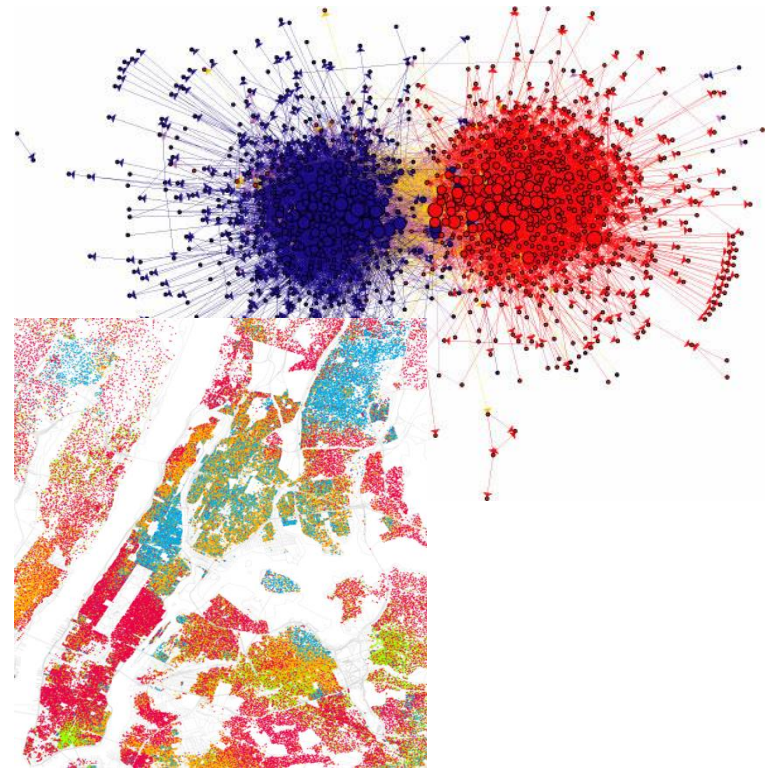
Mid-tier servers

- 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).

Lab of Data Science

BI Architecture

Figure 1. Typical business intelligence architecture.

6 lessons – data access

4 lessons – data quality & ETL

1 lessons – analytic SQL

5 lessons – OLAP and reporting

4 lessons – data mining

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»

Lab of Data Science

Front-end applications