Data Bases

Introduction

Sergio Peignier sergio.peignier@insa-lyon.fr

Associate Professor
INSA Lyon
Biosciences department

Definitions

- System of Information (SI)
- Database (DB)
- Database Management System (DBMS)

Definitions to fill

- System $\mapsto$?
- Data $\mapsto$?
- Information $\mapsto$?

Definition

- **System** $\mapsto$ Set of **interacting elements**, according to given **principles**
  (nature of the elements, kinds of interactions ...)

- **Data** $\mapsto$ Set of **observation, measures or facts**. Set of **symbols** and **potential messages**.

- **Information** $\mapsto$ Set of **organized data**, allowing the **extraction of knowledge**.

Definition | System of Information (SI)

Set of **resources** (i.e., people, hardware, software, methods) to:

- **Collect** information
- **Store** information
- **Analyze** and **query** information
- **Disseminate** information

There are different kinds of organization methods.

Definition | Database

- Set of **data accessible** and **usable** by means of a **computational system**.

- **Data modeling** real world **objects** (in the broad sense) and being the **basis** of a **computational application.**

Database Management System

Software $\rightarrow$ management of a database

Objectives of a DBMS

Which are the **(sub)-objectives** of a **DBMS**?

Objective of a DBMS | Basic goals

- **Save** data - **persistence**
- **Structure** data
- **Query** information
- **Modify** the elements of a Database.

Objectives of a DBMS | ACID

ACID: Database transactions' properties that guarantee data validity.

- **Atomicity**
  $\Rightarrow$ Treat each multi-statement transactions as a unit.
- **Consistency**
  $\Rightarrow$ Transactions lead the database to valid states.
- **Isolation**
  $\Rightarrow$ Concurrency control.
- **Durability**
  $\Rightarrow$ Completed transactions persistence.

Objectives of a DBMS | The Big Data era

- Deal with **large volumes of data** (Big Data)
  $\Rightarrow$ **Efficient data structures**
  $\Rightarrow$ terabytes/petabytes

- Deal with **High throughput** applications (Big Data)
  $\Rightarrow$ **fast access methods**.
  $\Rightarrow$ thousands/millions transactions/hour

Objectives of a DBMS | Users

- Deal with **sensitive data**
  $\Rightarrow$ Software **security** and **trustability**.
- Exploitation by **different kinds of users**
  $\Rightarrow$ Allow different **user-interfaces**.
- Deal with **multi-users**
  $\Rightarrow$ **Concurrency control**.

Objective of a DBMS | Independence

- Ensure the **physical independence**:
  $\Rightarrow$ Change inner DBMS mechanisms without changing conceptual schemes of applications.

- Ensure the **logical independence**.
  $\Rightarrow$ Adapt the conceptual schemes of applications without changing the inner structure of the DBMS.

Levels of Abstraction

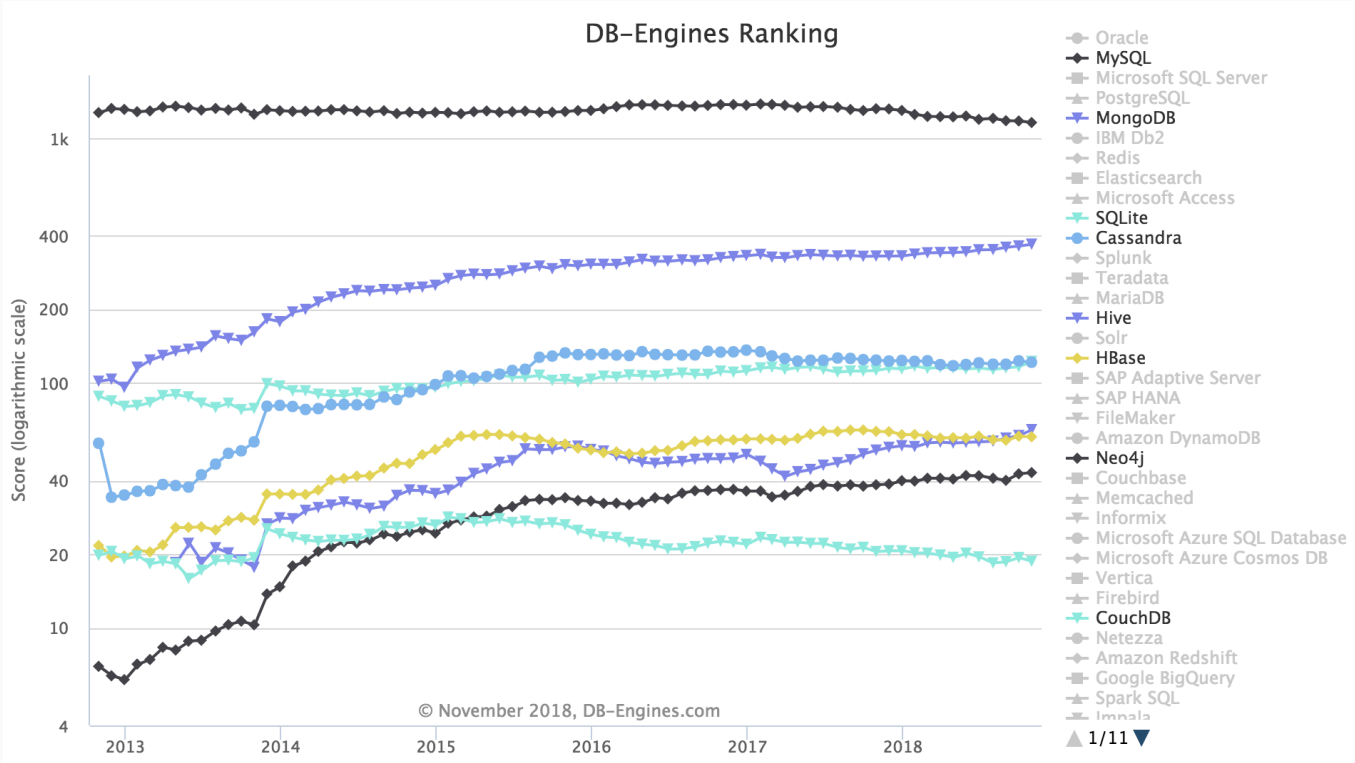
- **External** level (users) $\mapsto$ read, query.
- **Conceptual** level (designer, administrator) $\mapsto$ Database models.
- **Internal** level (store) $\mapsto$ Data structure (files, indexing).

Major types of DBMS

- **Relational** DBMS - Mostly **SQL**
- **noSQL - Not Only SQL**
  - **Key-value** stores
  - **Wide column** stores
  - **Document** stores
  - **Graph** DBMS
  - **Object oriented** DBMS
  - **Time Series** DBMS
  - ...

Ranking of DBMS

More than 300 pieces of software.



Rankings and more information:

`https://db-engines.com`

`https://www.kdnuggets.com/2016/06/`

DBMS

Rank Nov 2018	Rank Oct 2018	Rank Nov 2017	DBMS	Database Model	Score Nov 2018	Score Oct 2018	Score Nov 2017
1.	1.	1.	Oracle	Relational DBMS	1301.11	-18.16	-58.94
2.	2.	2.	MySQL	Relational DBMS	1159.89	-18.22	-162.14
3.	3.	3.	Microsoft SQL Server	Relational DBMS	1051.55	-6.78	-163.53
4.	4.	4.	PostgreSQL	Relational DBMS	440.24	+20.85	+60.33
5.	5.	5.	MongoDB	Document store	369.48	+6.30	+39.01
6.	6.	6.	IBM Db2	Relational DBMS	179.87	+0.19	-14.19
7.	7.	↑ 9.	Redis	Key-value store	144.17	-1.12	+22.99
8.	8.	↑ 10.	Elasticsearch	Search engine	143.46	+1.13	+24.05
9.	9.	↓ 7.	Microsoft Access	Relational DBMS	138.44	+1.64	+5.12
10.	↑ 11.	↑ 11.	SQLite	Relational DBMS	122.71	+5.96	+9.95
11.	↓ 10.	↓ 8.	Cassandra	Wide column store	121.74	-1.64	-2.47
12.	↑ 13.	↑ 15.	Splunk	Search engine	80.37	+3.48	+15.50
13.	↓ 12.	↓ 12.	Teradata	Relational DBMS	79.31	+0.67	+1.07
14.	14.	↑ 18.	MariaDB	Relational DBMS	73.25	+0.12	+17.96
15.	↑ 16.	↑ 19.	Hive	Relational DBMS	64.57	+3.47	+11.32
16.	↓ 15.	↓ 13.	Solr	Search engine	60.87	-0.44	-8.28
17.	17.	↓ 16.	HBase	Wide column store	60.41	-0.26	-3.15
18.	18.	↓ 14.	SAP Adaptive Server	Relational DBMS	56.57	-2.00	-10.47
19.	↑ 21.	↑ 20.	SAP HANA	Relational DBMS	55.88	+1.50	+6.70
20.	↓ 19.	↓ 17.	FileMaker	Relational DBMS	55.75	-0.29	-3.09
21.	↓ 20.	↑ 22.	Amazon DynamoDB	Multi-model	53.81	-0.65	+16.69
22.	22.	↓ 21.	Neo4j	Graph DBMS	43.12	+0.47	+4.67
23.	23.	23.	Couchbase	Document store	34.85	-1.06	+2.54
24.	24.	24.	Memcached	Key-value store	29.75	-0.80	+1.77
25.	↑ 26.	25.	Informix	Relational DBMS	26.45	+0.21	-1.26
26.	↓ 25.	26.	Microsoft Azure SQL Database	Relational DBMS	26.19	-0.08	+4.08

Database Life Cycle

- Definition of **specifications**.
- **Conceptual model**.
- **Logic Model**.
- **DBMS choice**.
- **Maintenance** and **use**.