



Semantic data & semantic data storages

2021.03.08

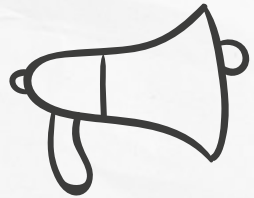
CONTENTS



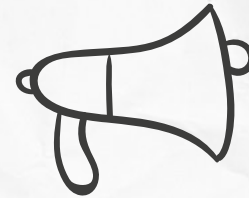
Semantic data



**Semantic data
storages**



Ontology



Knowledge bases



**NoSQL
Database**



PART 01

Semantic Data

Semantic Data

“I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web — the content, links, and transactions between people and computers. A Semantic Web, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The intelligent agents people have touted for ages will finally materialize.”

(Berners-Lee, 1999)

The **Semantic Web** is a ‘web of data’ that facilitates machines to understand the semantics, or meaning, of information on the WWW. It extends the network of hyperlinked human-readable web pages by inserting machine-readable metadata about pages and how they are related to each other, enabling automated agents to access the Web more intelligently and perform tasks on behalf of users

Berners-Lee is now the director of the World Wide Web Consortium (W3C), which oversees the development of Semantic Web standards. Since 2013, Semantic Web activities have been subsumed by Web of Data activities

Semantic Data

Knowledge of the specific task domain in which the program is to do its problem solving was more important as a source of power for competent problem solving than the reasoning method employed.

The subareas of **knowledge representation** and **knowledge acquisition**, which was later called **knowledge engineering**, were created to provide methods and techniques to represent human knowledge in a machine-understandable manner.

Semantic Data

Semantics, the Science of (Meaning)

Semantic technology provides machine-understandable (or better machine-processable) descriptions of data, programs, and infrastructure, enabling computers to reflect on these artifacts.

Now, what does [machine-processable semantics](#) really mean?

```
<person>  
  <name>Sir Tim</name>  
  <phone number>01-444444</phone number>  
</person>
```

Semantic Data

Logic

From an algorithmic perspective, implementing logical-reasoning systems demonstrates clearly how complex decidability and complexity are to manage.

- Propositional Logic
- First-Order Predicate Logic
- Second-Order Predicate Logic
- Description Logics

Semantic Data

Semantic Web Languages

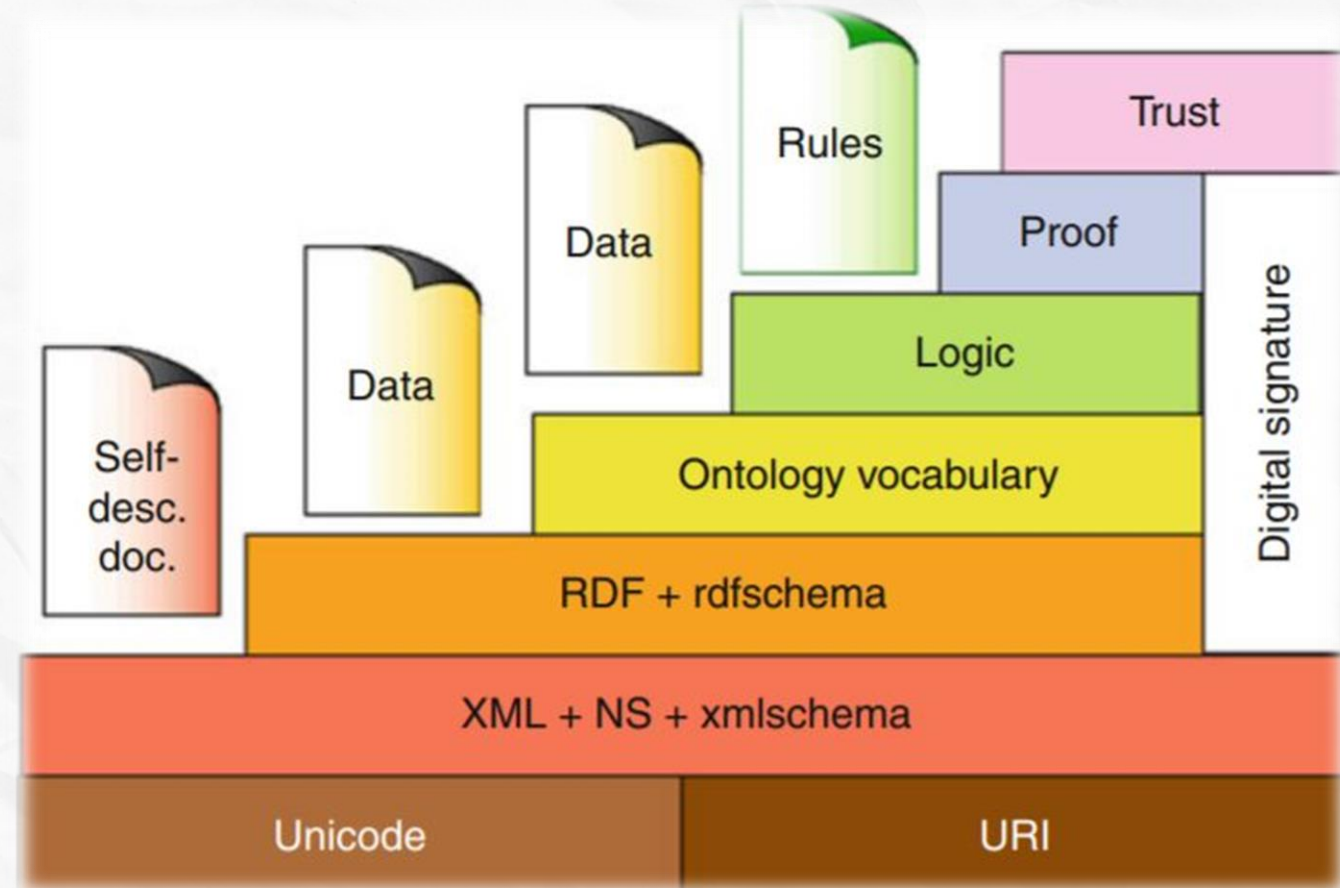
- HTML
- Extensible Markup Language (XML)
- Resource Description Framework (RDF)
- RDF schema (RDFS)
- Web Ontology Language OWL (OWL Lite, OWL DL, and OWL Full)

```
<META name = "Author" lang= "fr" content = "Arnaud Le Hors">
```


Semantic Data

The Tower of Babel

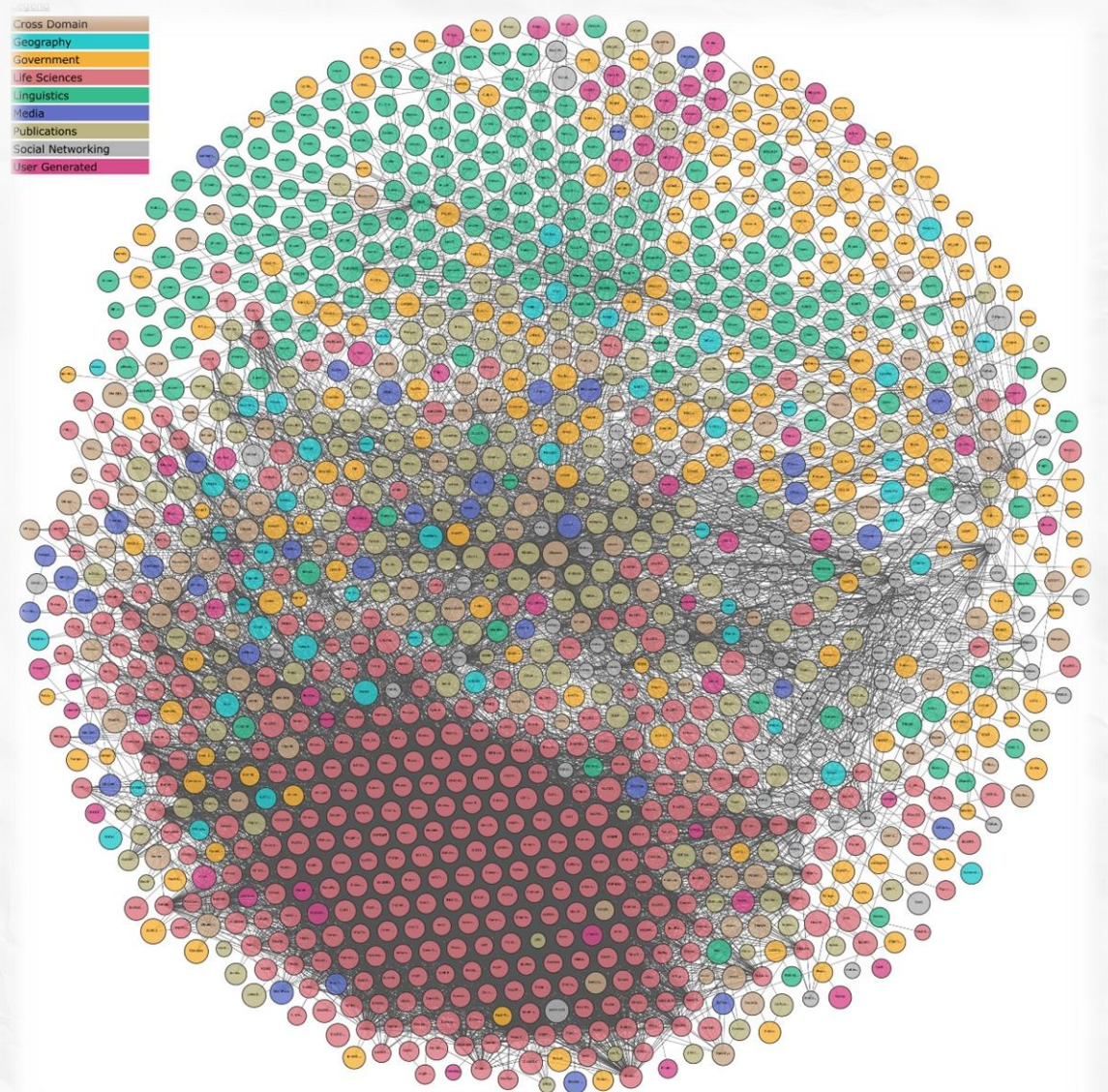
The Open Systems Interconnection model (OSI model) is a product of the Open Systems Interconnection effort at the International Organization for Standardization. It is a way of sub-dividing a communications system into smaller parts called layers. A layer is a collection of conceptually similar functions that provide services to the layer above it and receives services from the layer below.



Semantic Data

Semantic Web as a Database

- Which particular database techniques (e.g., partitioned hashes, column tables) are most applicable to high-performance RDF storage?
- How to structure benchmarks for large-scale repositories? Including what are the correct dimensions?
- When and where should reasoning be handled? For example, materialization (the precomputation and storage of inferred triples) is an expensive process which may not contribute to desired results.



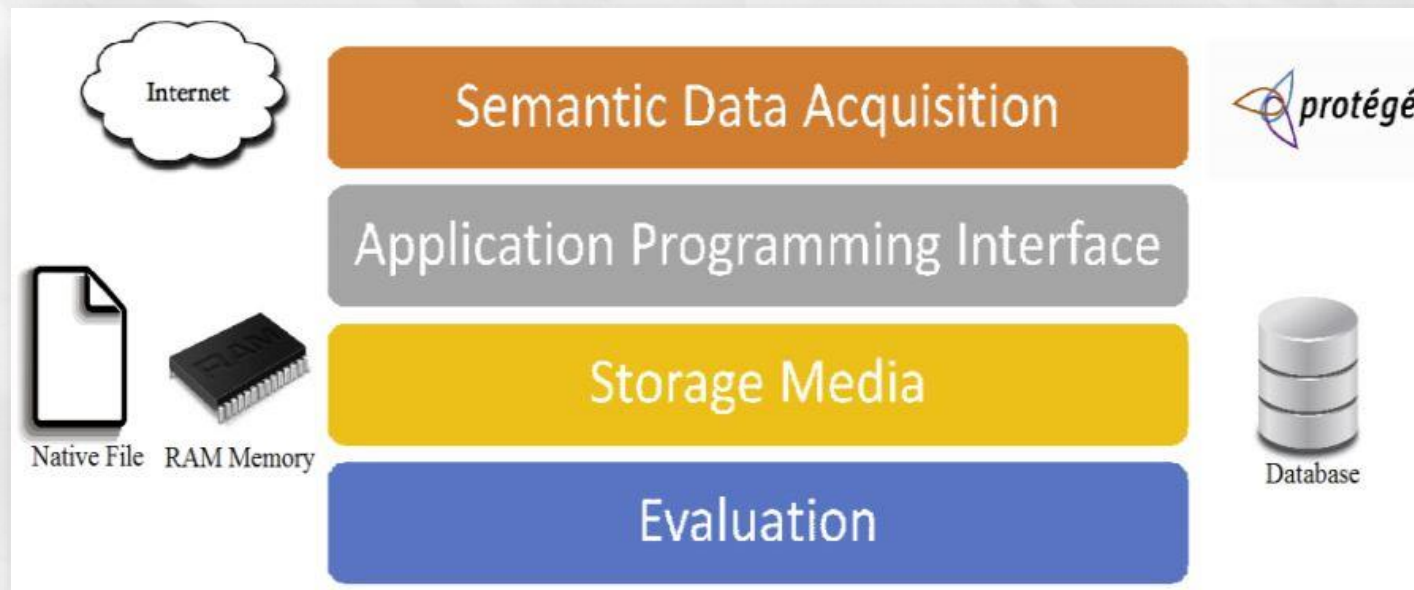


PART 02
Semantic data storages

Semantic data storages

Semantic Web as a Database

The storage and retrieval of information are important functions of information systems (IS). These IS functions have been realized for decades, due to the maturity of the relational database technology. In recent years, the concept of Semantic Information System (SIS) has emerged as IS in which information is represented with explicit semantic based on its meaning rather than its syntax to enable its automatic and intelligent processing by computers.



Semantic data storages

Approaches for Storing Semantic Data

Three approaches are used to store ontology or semantic data, namely:

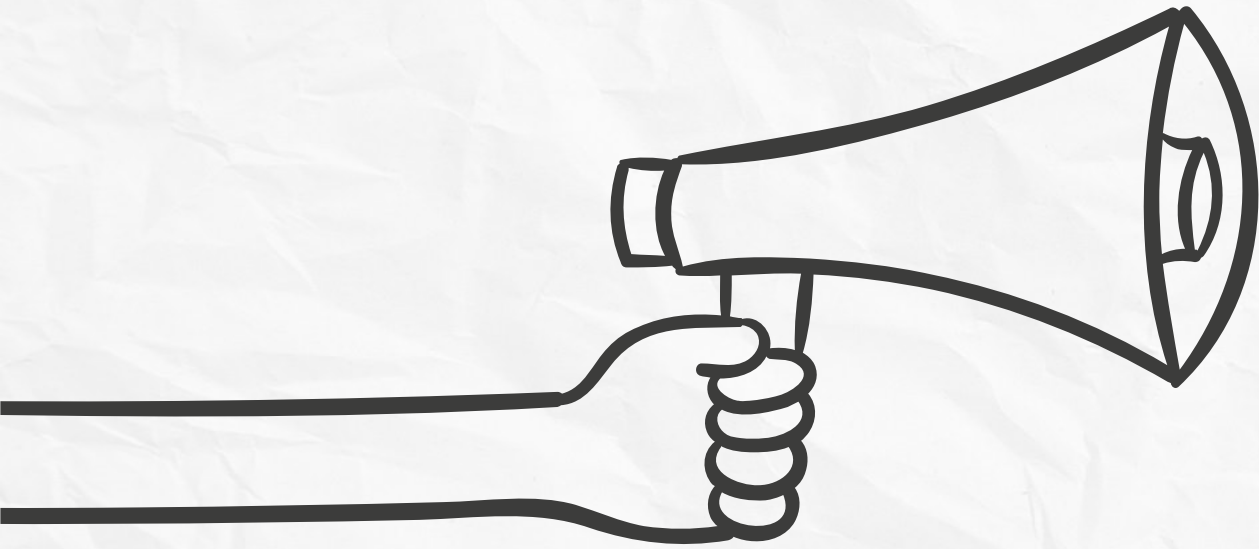
- in-memory
- native or file systems
- databases

Semantic data storages

Software Platforms for Semantic Data Storage

To enable the storage and query of semantic data, several platforms have been developed. The most popular of these platforms are: AllegroGraph, Jena, Open Anzo, Minerva and Sesame.

Platform	License	Operating system	Type of Storage
AllegroGraph	Commercial/Free	Linux	Native
Jena	Free/Open Source	Windows/Linux	Memory, Native, RDB
Sesame	Free/ Open Source	Windows/Linux	Memory, Native, RDB
Open Anzo	Free/ Open Source	Windows Linux	RDB
Minerva	Free	Windows/Linux	RDB



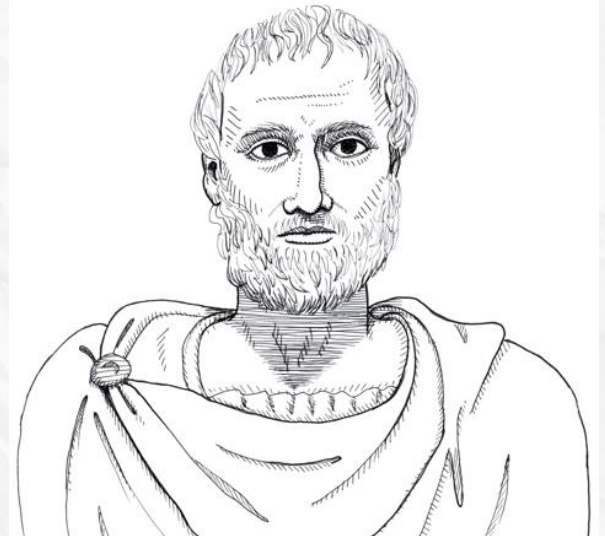
PART 03
Ontology

Ontology

Ontology in Philosophy

a philosophical discipline — a branch of philosophy that deals with the nature and the organization of reality

- Science of Being (Aristotle, *Metaphysics*, IV, 1)
- Tries to answer the questions:
 - What characterizes being?
 - Eventually, what is being?
- How should things be classified?



Ontology

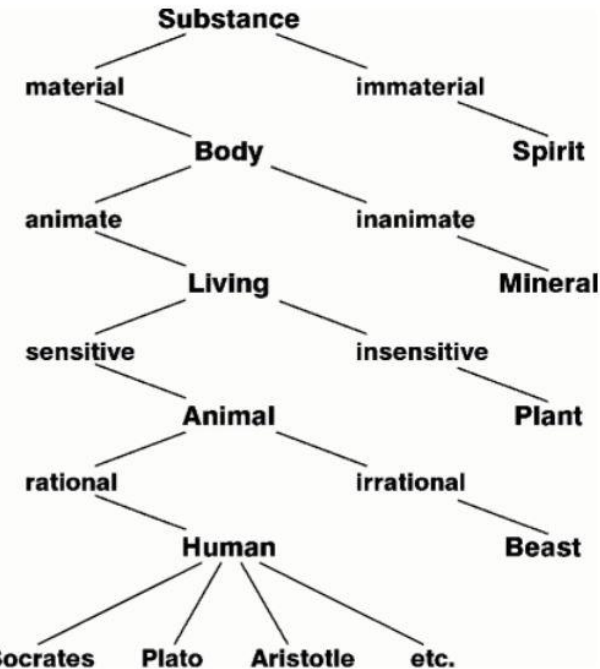
Ontology in Philosophy

In philosophy, *ontology* is the study of being or existence. It aims to find out what entities and types of entities exist:

- What exists?
- Is existence a property?
- What is an object?
- Do non-physical (abstract) objects exist?
- How things should be classified?

Aristotle's ontology:

Supreme genus:



Individuals:

Ontology

Ontology in Computer Science

An ontology is an **engineering artefact**

- It is constituted by a specific vocabulary used to describe a certain reality;
- And a set of explicit assumptions regarding the intended meaning of the vocabulary.
 - (Almost always including how concepts should be classified.)

Ontology

Ontology in Computer Science

Thus, an ontology describes a formal specification of a certain domain:

- Shared understanding of a domain of interest
- Formal and machine manipulable model of a domain of interest

“An explicit specification of a conceptualisation”

[Tom Gruber 1993]

Ontology

Ontologies in sciences

- Bioinformatics
 - The Gene Ontology, The Protein Ontology MGED, etc.
- Medicine
 - The Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) Ontology: a Core terminology of over 364,000 health care concepts; more than 984,000 descriptions; ≈ 1.45 million semantic relationships.
- • Linguistics
- Database integration
- User interface design
- Fractal Indexing
- ...

*Pericardium is-a Tissue and containedIn . Heart
Pericarditis is-a Inflammation and hasLocation . Pericardium
Inflammation is-a Disease and actsOn . Tissue
Disease and hasLocation . containedIn . Heart is-a
HeartDisease and NeedsTreatment*

Ontology

Semantic Technologies at the BBC

Launched in the mid 1990s, the BBC website was focused on supporting broadcast brands such as Top Gear as well as domain-specific sites: news, food, gardening, etc.

BBC Web-based service is one of the most visited websites and the world's largest news website. As of 2007, it contained over two million pages

Focus has been on separate, standalone HTML microsites that are not linked together and to other data sources on the Web

difficult to find everything BBC has published about any given object

cannot navigate from a page about a musician to a page with all the programmes that have played that artist, to their biography, etc.

Ontology

Semantic Technologies at the BBC – [BBC Ontologies](#)

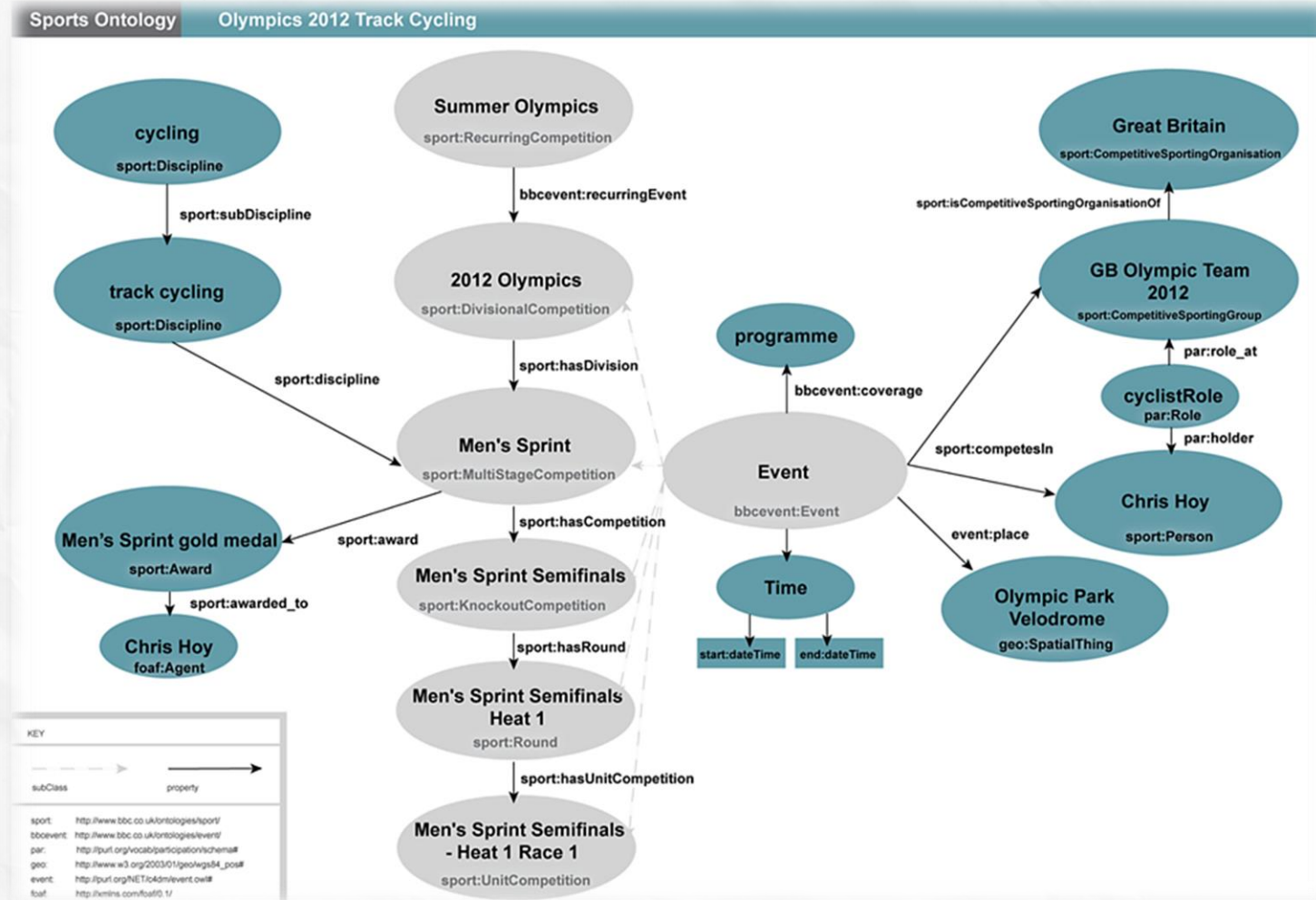
provides access to the ontologies the BBC is using to support its audience facing applications such as [BBC Sport](#), [BBC Education](#), [BBC Music](#), [News projects](#) and more. These ontologies form the basis of our Linked Data Platform.

- [BBC Ontology](#)
- [Business News Ontology](#)
- [CMS Ontology](#)
- [Food Ontology](#)
- ...

Ontology

Semantic Technologies at the BBC – [Sport Ontology](#)

The Sport Ontology is a simple lightweight ontology for publishing data about competitive sports events.





PART 04
Knowledge base

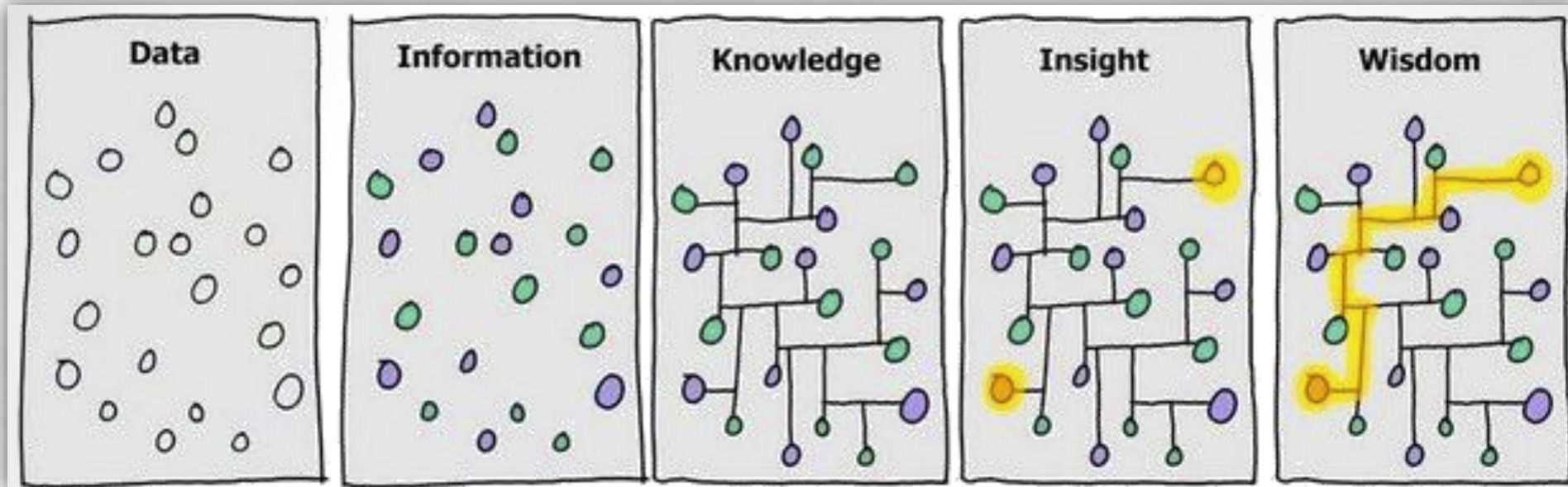
Knowledge bases

*A **knowledge base (KB)** is a technology used to store complex structured and unstructured information used by a computer system. The initial use of the term was in connection with expert systems; which were the first knowledge-based systems.*

A comprehensive semantically organized machine-readable collection of universally relevant or domain-specific entities, classes, and facts (attributes, relations)

- Plus spatial and temporal dimensions
- Plus commonsense properties and rules
- Plus contexts of entities and facts, e.g., textual/visual evidences
- Plus...

Knowledge bases



Database: flat data representation (tabular)

Knowledge base: “The ideal representation for a knowledge base is an object model (ontology) with classes, subclasses, and instances.”

--Wikipedia

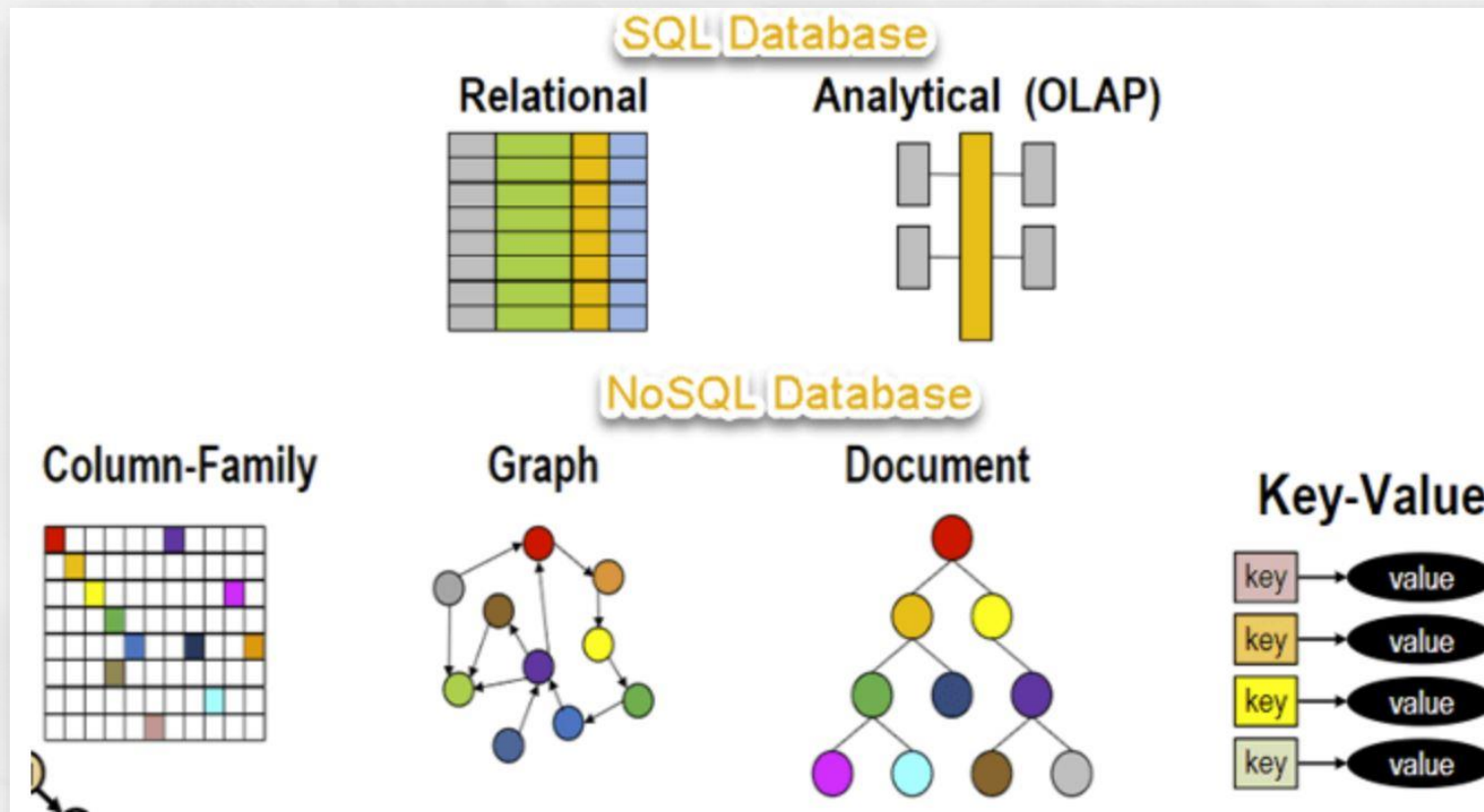


PART 05

NoSQL Database

NoSQL Database

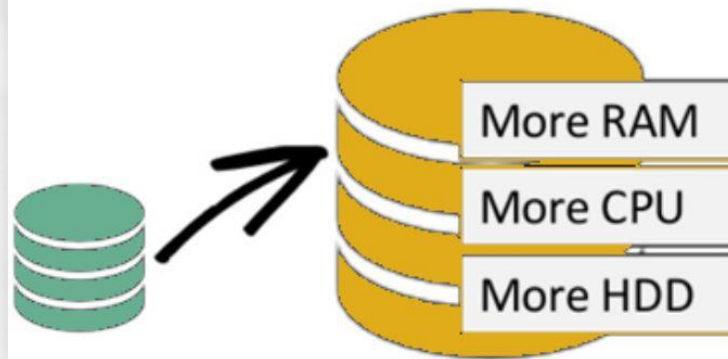
A *NoSQL database* provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations used in relational databases.



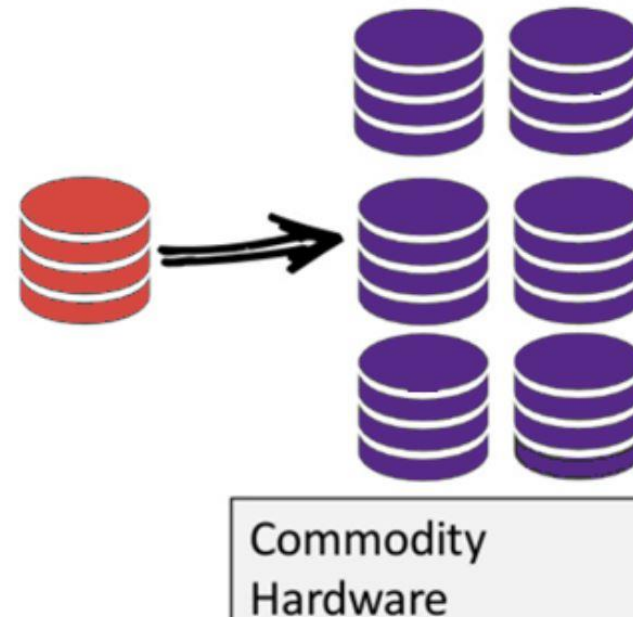
NoSQL Database

NoSQL database is non-relational, so it scales out better than relational databases as they are designed with web applications in mind.

Scale-Up (*vertical* scaling):

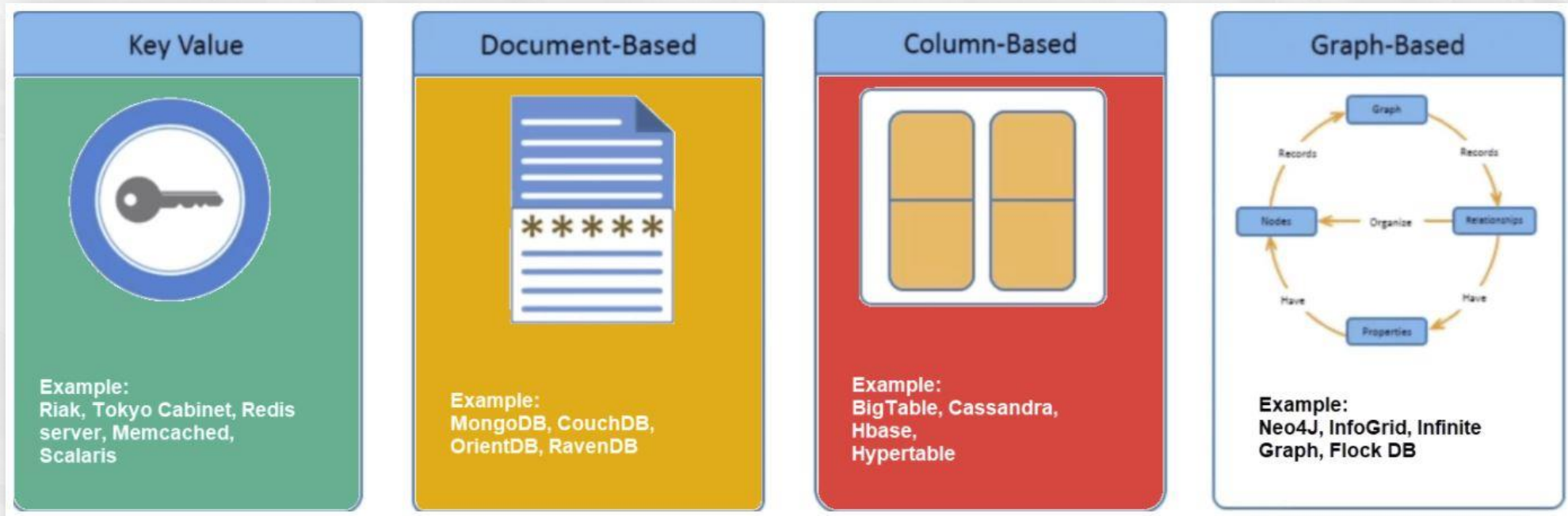


Scale-Out (*horizontal* scaling):



NoSQL Database

NoSQL Databases are mainly categorized into four types: Key-value pair, Column-oriented, Graph-based and Document-oriented.



NoSQL Database

CAP theorem is also called brewer's theorem. It states that is impossible for a distributed data store to offer more than two out of three guarantees

- Consistency
- Availability
- Partition Tolerance

NoSQL Database

Many NoSQL databases were designed by young technology companies like Google, Amazon, Yahoo, and Facebook to provide more effective ways to store content or process data for huge websites. Some of the most popular NoSQL databases include the following:

- Apache CouchDB
- Apache Cassandra
- MongoDB
- Redis
- Elasticsearch



THANK YOU

2021.03.08

mantx626@gmail.com