Introduction to Big Data

ECAP456

Dr. Rajni Bhalla

Associate Professor

Learning Outcomes

After this lecture, you will be able to

- Identify key differences between NoSQL and relational databases.
- Appreciate the architecture of NoSQL databases.
- Describe the major types of NoSQL databases and their features.

Introduction

- Organizing large amounts of heterogeneous data.
- Completely aligned with the nature of the problems being solved.
- Superfast in accomplishing that task.
- Relaxing many of the integrity.
- Redundancy constraints of storing data in relational databases.

Introduction

- Organizing large amounts of heterogeneous data.
- Completely aligned with the nature of the problems being solved.
- Superfast in accomplishing that task.
- Relaxing many of the integrity.
- Redundancy constraints of storing data in relational databases.

Introduction

- Organizing large amounts of heterogeneous data.
- Completely aligned with the nature of the problems being solved.
- Superfast in accomplishing that task.
- Relaxing many of the integrity.
- Redundancy constraints of storing data in relational databases.

Introduction

- Organizing large amounts of heterogeneous data.
- Completely aligned with the nature of the problems being solved.
- Superfast in accomplishing that task.
- Relaxing many of the integrity.
- Redundancy constraints of storing data in relational databases.

Introduction

- Organizing large amounts of heterogeneous data.
- Completely aligned with the nature of the problems being solved.
- Superfast in accomplishing that task.
- Relaxing many of the integrity.
- **Redundancy constraints of storing data in relational databases.**

What is Relational Database

- Powerful and universally used database technology.
- Structured and optimized to ensure accuracy and consistency of data.
- Stored on the largest and most reliable of computers.

What is Relational Database

- Powerful and universally used database technology.
- **Structured and optimized to ensure accuracy and consistency of data.**
- Stored on the largest and most reliable of computers.

What is Relational Database

- Powerful and universally used database technology.
- Structured and optimized to ensure accuracy and consistency of data.
- Stored on the largest and most reliable of computers.

Why NOSQL database emerge?

Unpredictable stream of data

Why NOSQL database emerge?

Unpredictable stream of data

Relational Database

X

Why NOSQL database emerge?

Unpredictable stream of data

Relational Database

X

Cost

Why NOSQL database emerge?

Unpredictable stream of data

Relational Database

X

Cost

Speed

Why NOSQL database emerge?

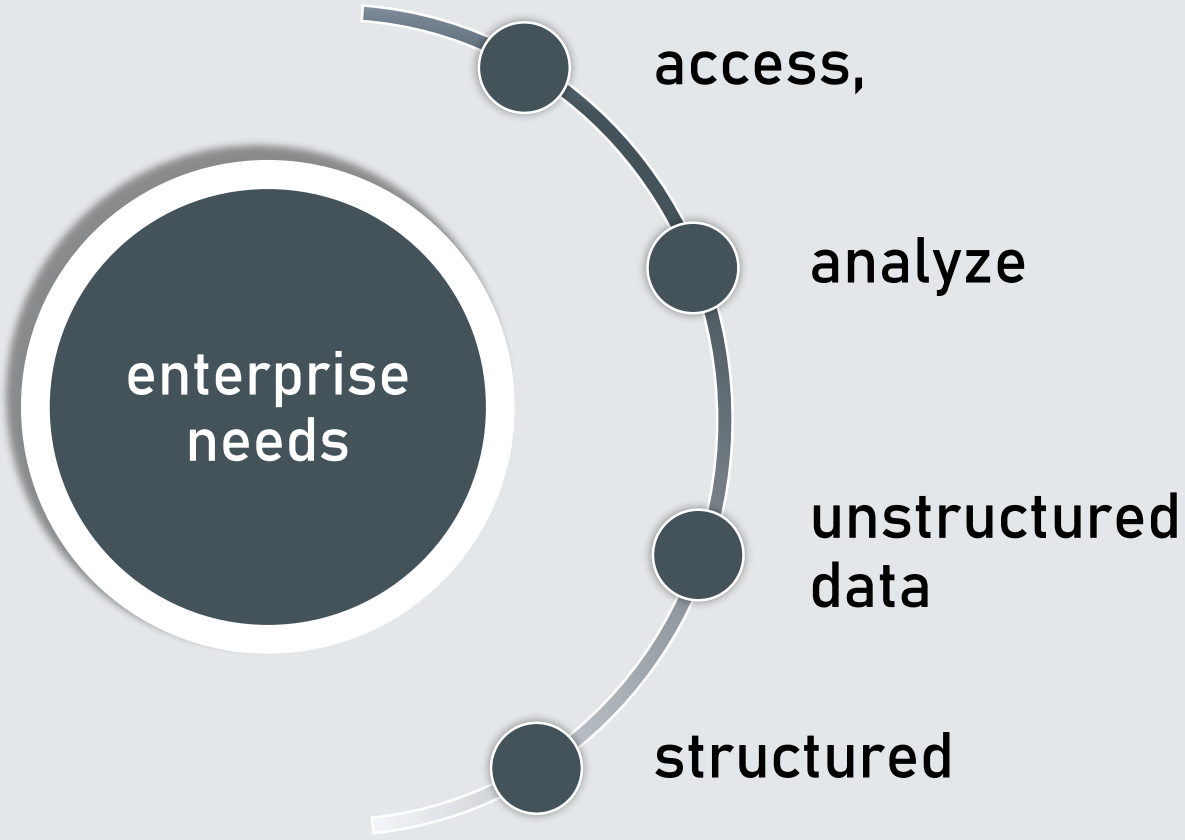
Depending upon which constraint(s) are relaxed, a different kind of database structure will emerge. These are called NoSQL databases

NOSQL Database

- Next-generation databases
- Differentiate it from antiquated, 'pre-relational' databases
- Organization uses NOSQL Database

NOSQL Database

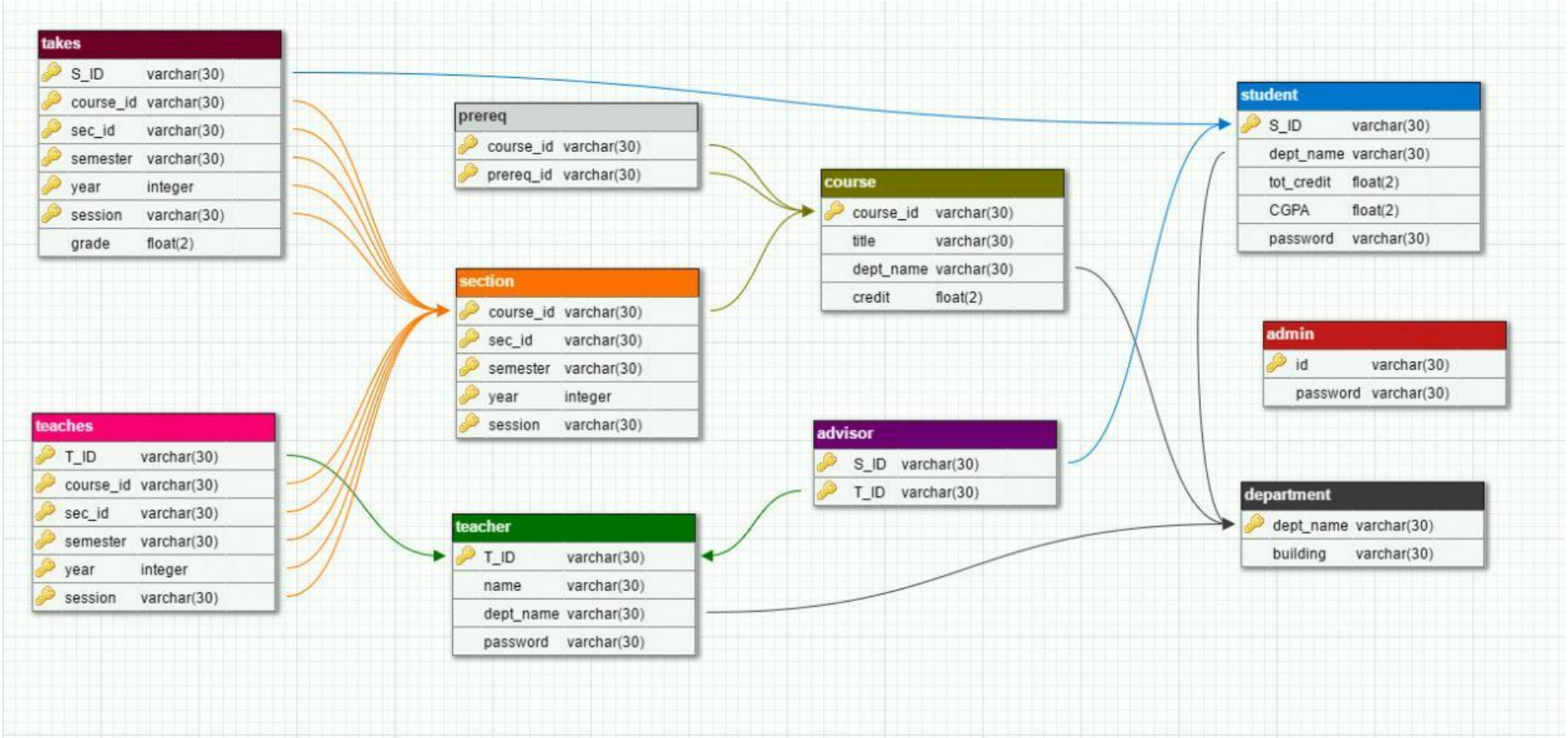
- NoSQL database is useful when



NOSQL Database

- Relational database require
  - any data element could be randomly accessed
  - value could be updated in that same physical location
- NOSQL databases files
  - written once and almost never updated in place
  - to link the appended data to the original file.

RDBMS vs NOSQL

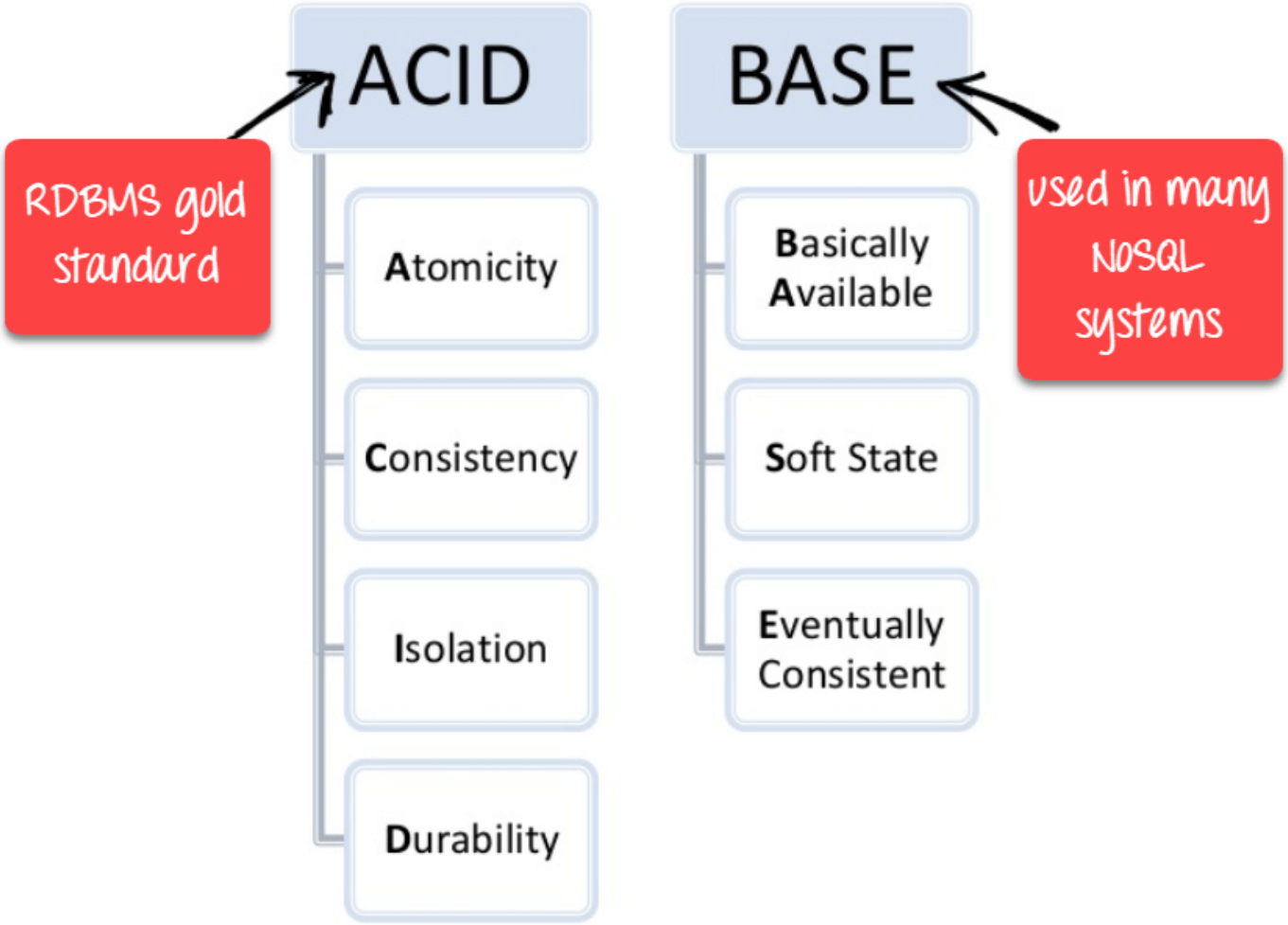


Do not Support Relational Schema

RDBMS vs NOSQL

- NoSQL stands mostly for “Not only SQL”.
- transaction processing capabilities are fast but weak
- do not support the ACID (Atomicity, Consistency, Isolation, Durability) properties

RDBMS vs NOSQL



RDBMS vs NOSQL

- NoSQL stands mostly for “Not only SQL”.
- transaction processing capabilities are fast but weak
- do not support the ACID (Atomicity, Consistency, Isolation, Durability) properties
- approximately accurate at any point in time

RDBMS vs NOSQL

- Third, these databases are also distributed and horizontally scalable to manage web-scale databases using Hadoop clusters of storage.
- Thus, they work well with the write-once and read-many storage mechanism of Hadoop clusters.

Comparative features of RDBMS and NoSQL.

Feature	RDBMS	NOSQL
Applications	Mostly centralized Applications (e.g. ERP)	Mostly designed for the decentralized applications (e.g. Web, mobile, sensors)
Rigor	Support ACID properties for Transaction Processing	Support BASE properties for approximate reporting
Availability	Moderate to high	Continuous availability to receive and serve data

Comparative features of RDBMS and NoSQL.

Feature	RDBMS	NOSQL
Velocity	Moderate velocity of data	High velocity of data (devices, sensors, social media, etc.). Low latency of access
Data Volume	Moderate size; archived after for a certain period	Huge volume of data, stored mostly for a long time or forever; Linearly scalable DB.

Comparative features of RDBMS and NoSQL.

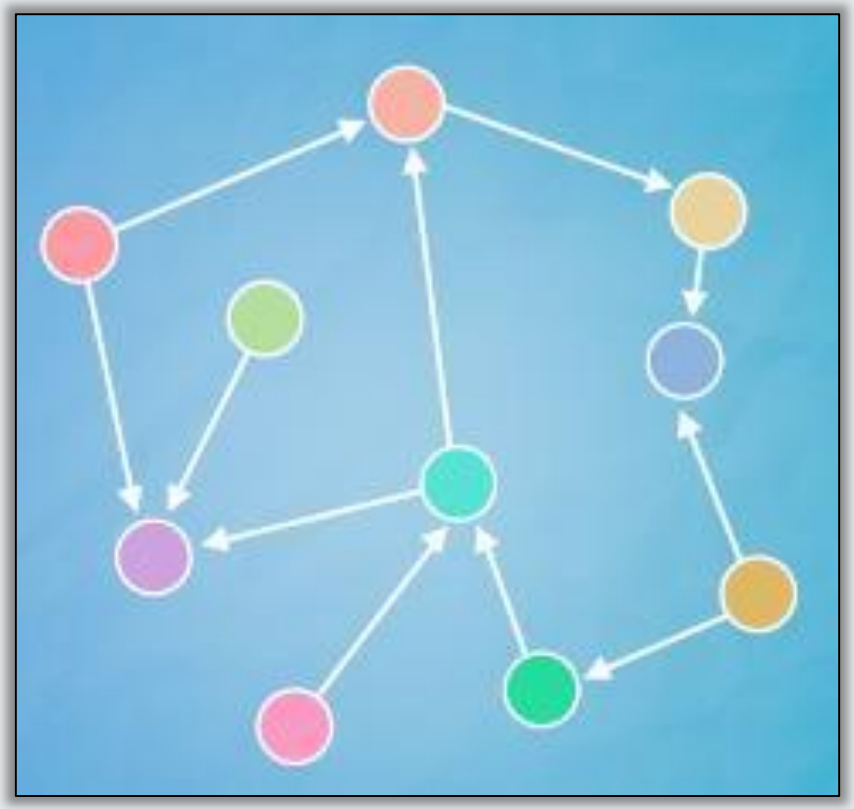
Feature	RDBMS	NOSQL
Data Sources	Data arrives from one or few, mostly predictable sources	Data arrives from multiple locations and are of unpredictable nature
Data type	Data are mostly structured	Structured or unstructured data
Data Access	Primary concern is reading the data	Concern is both read and write

Comparative features of RDBMS and NoSQL.

Feature	RDBMS	NOSQL
Technology	Standardized relational schemas; SQL language	Many designs with many implementations of data structures and access languages
Cost	Expensive; commercial	Low; open-source software

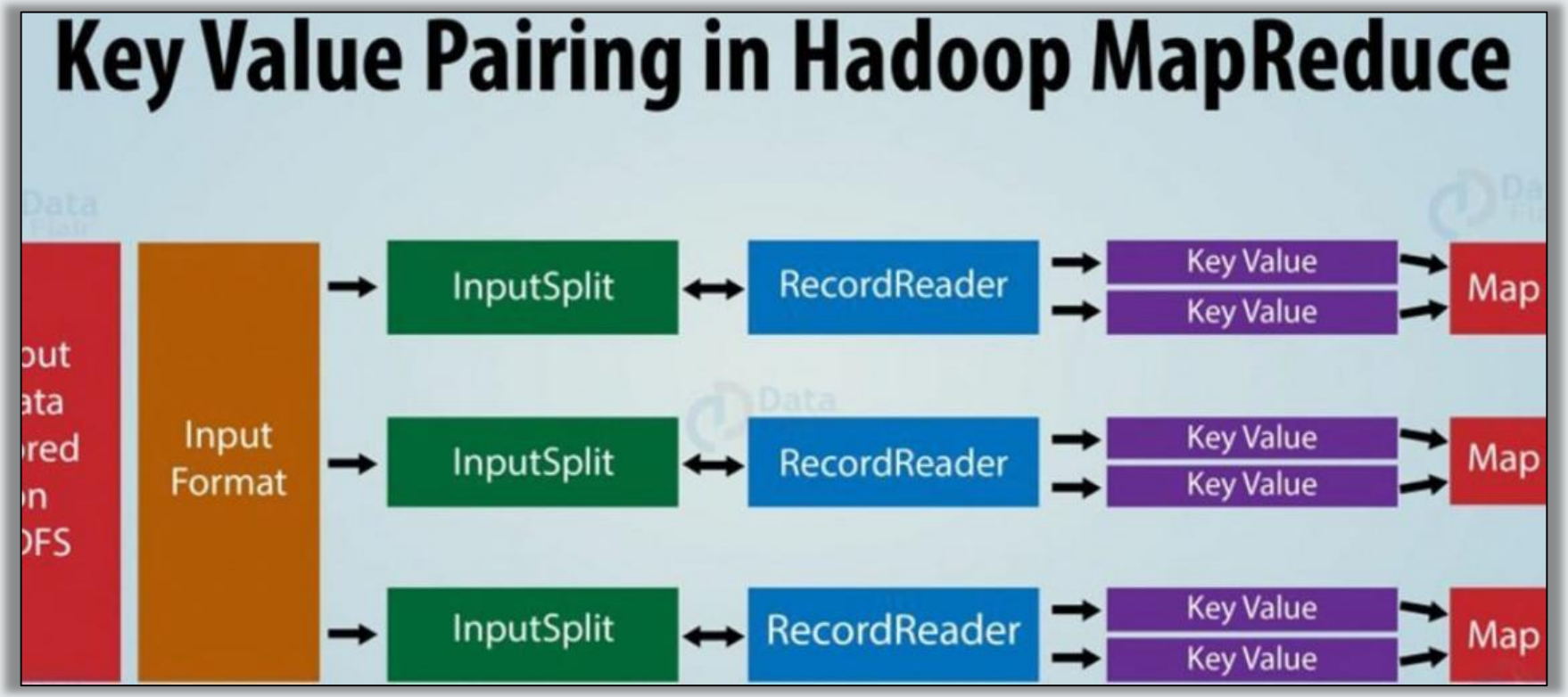
TYPES OF NOSQL DATABASES

File size and types will vary enormously.



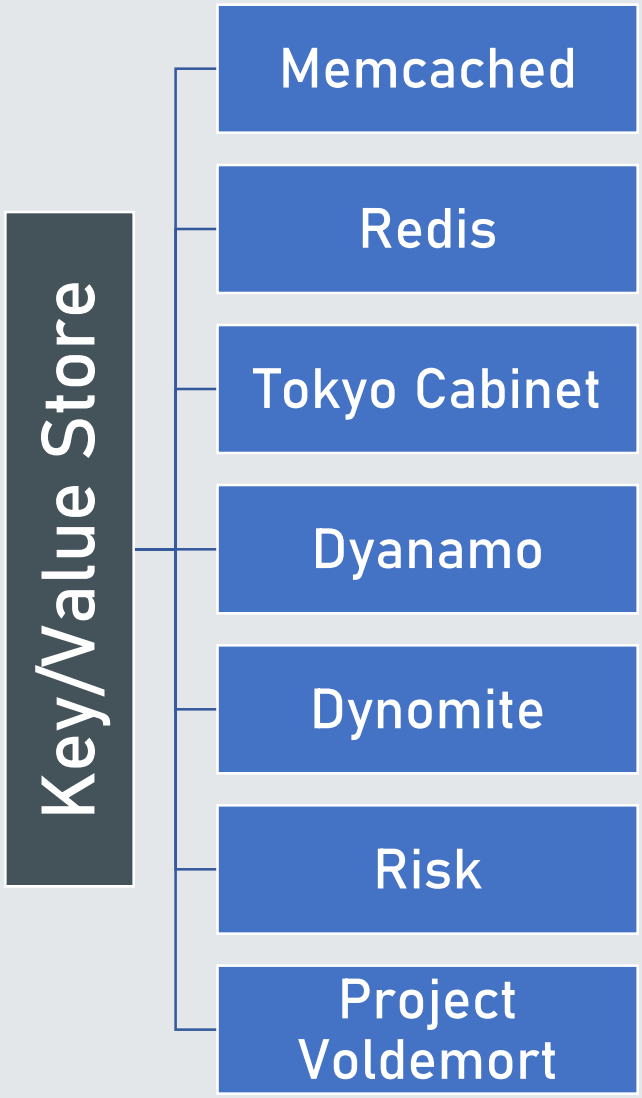
Uses no structured query language

TYPES OF NOSQL DATABASES



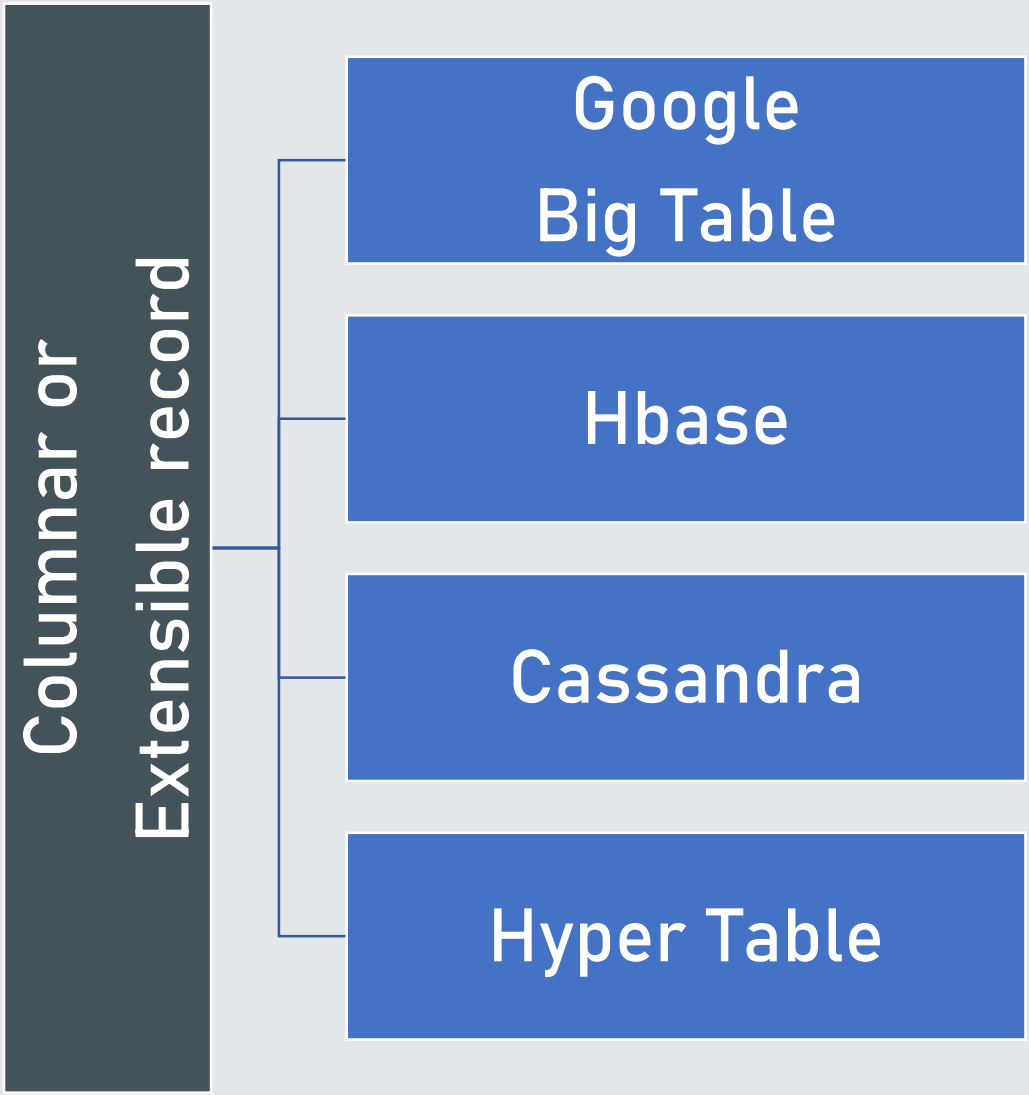
Key value/Pair

TYPES OF NOSQL DATABASES



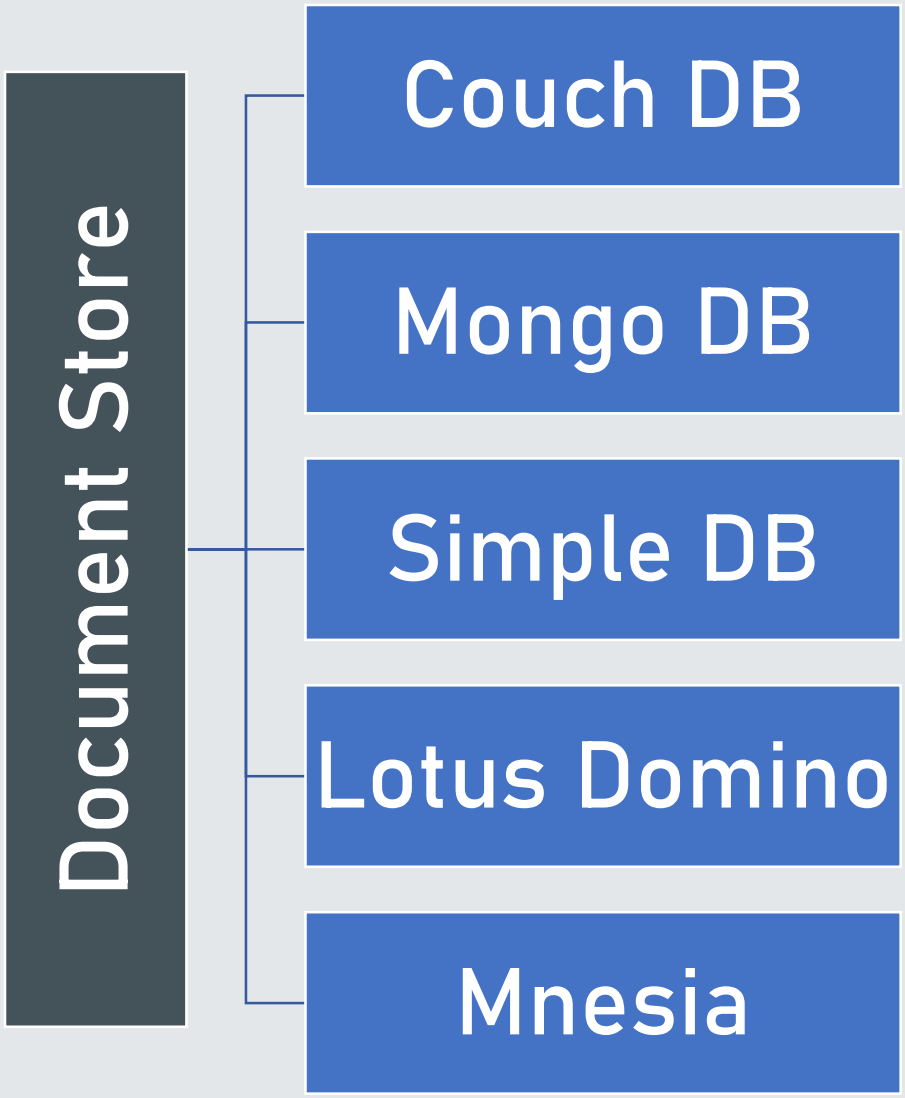
Offerings in NoSQL databases

TYPES OF NOSQL DATABASES



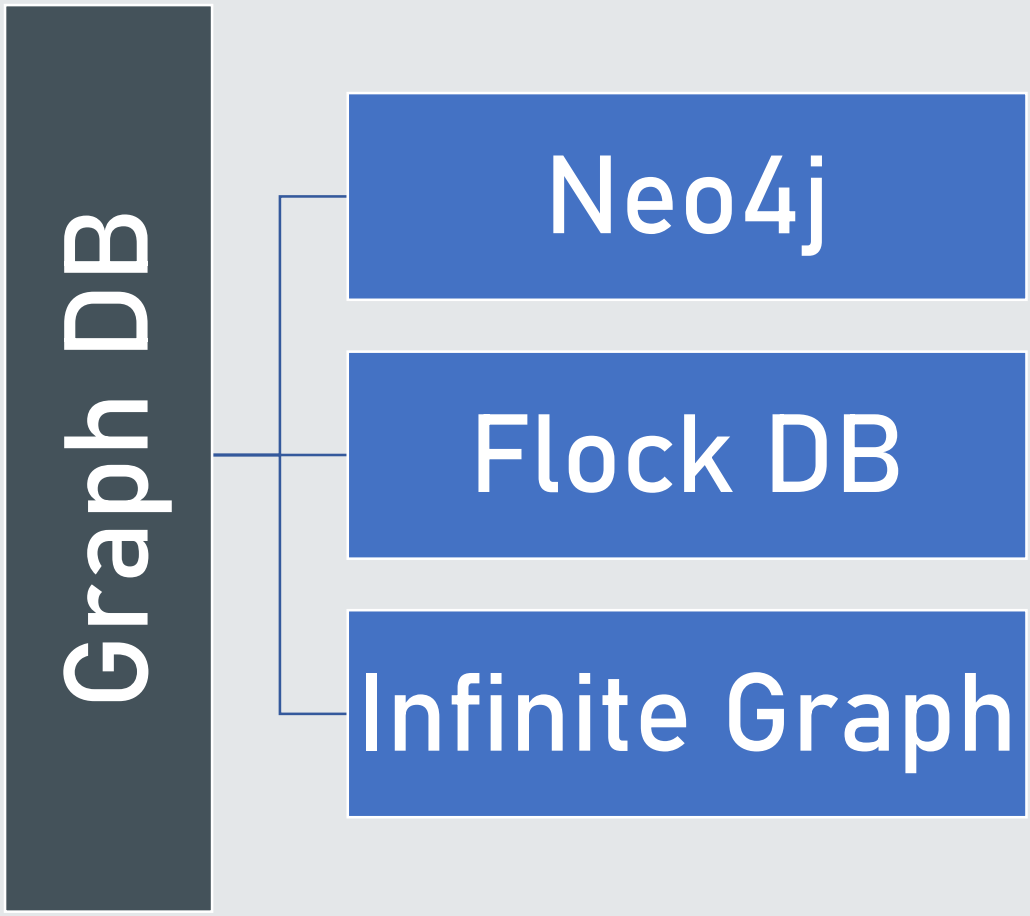
Offerings in NoSQL databases

TYPES OF NOSQL DATABASES



Offerings in NoSQL databases

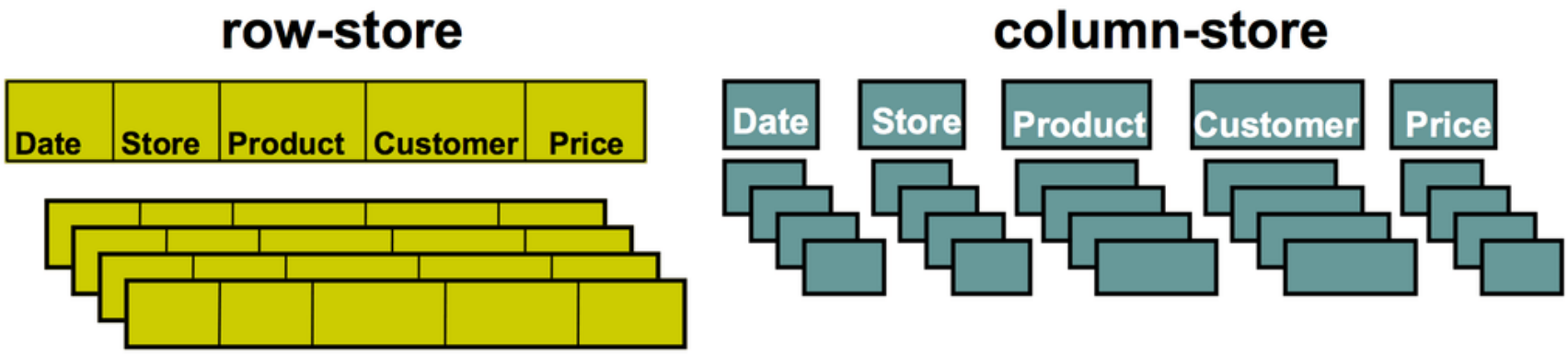
TYPES OF NOSQL DATABASES



Offerings in NoSQL databases

TYPES OF NOSQL DATABASES

1. Columnar Databases:



TYPES OF NOSQL DATABASES

1. Columnar Databases:

Content Management System

Blogging Platform

TYPES OF NOSQL DATABASES

1. Columnar Databases:

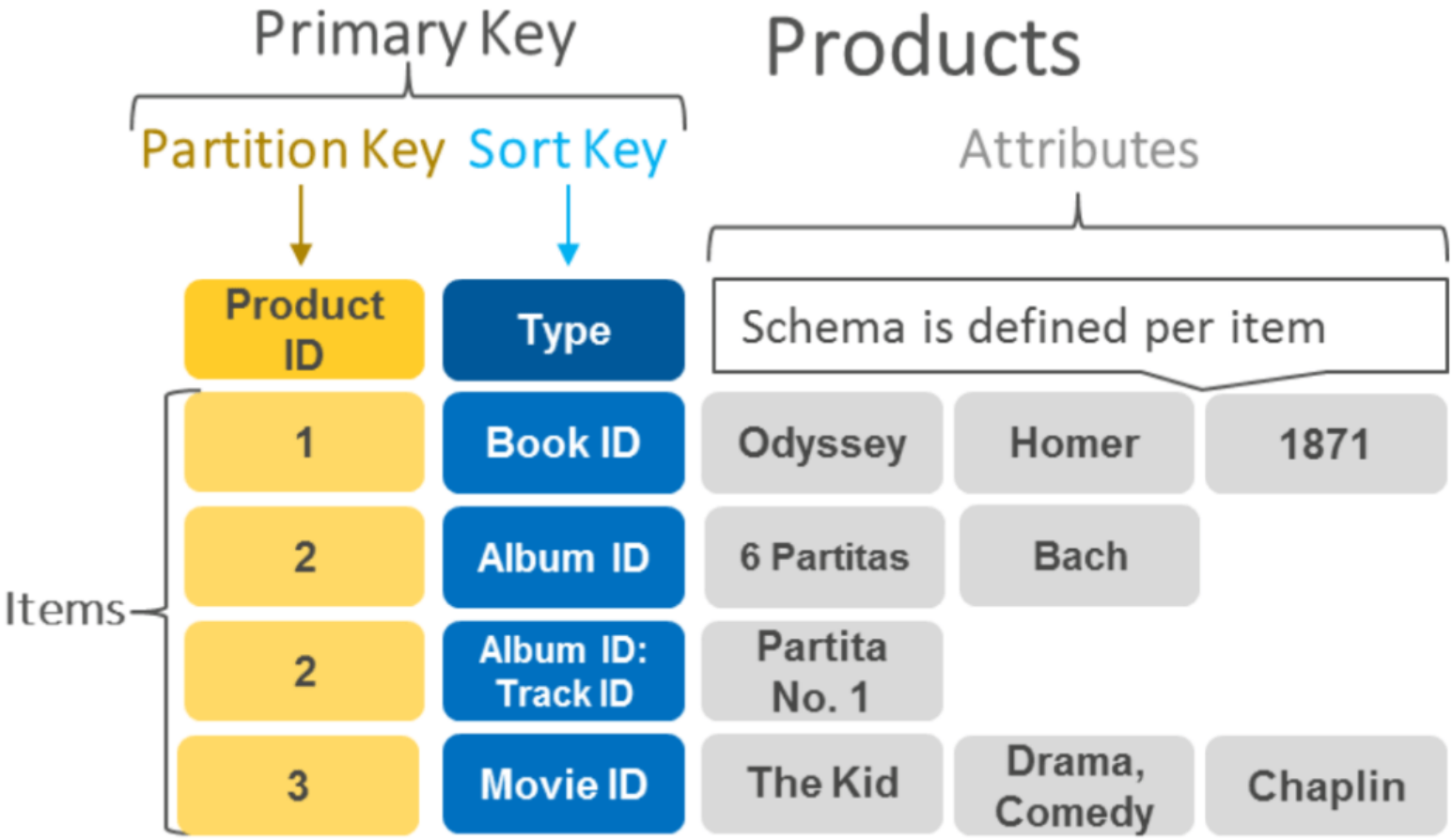
- Maintaining counter
- Expiring usage
- Heavy write volume such as log aggregation.

TYPES OF NOSQL DATABASES

1. Columnar Databases:

TYPES OF NOSQL DATABASES

2. Key-Value Pair Databases



TYPES OF NOSQL DATABASES

2. Key-Value Pair Databases

TYPES OF NOSQL DATABASES

3. Document Database

Relational

ID	first_name	last_name	cell	city	year_of_birth	location_x	location_y
1	'Mary'	'Jones'	'516-555-2048'	'Long Island'	1986	'-73.9876'	'40.7574'

ID	user_id	profession
10	1	'Developer'
11	1	'Engineer'

ID	user_id	name	version
20	1	'MyApp'	1.0.4
21	1	'DocFinder'	2.5.7

ID	user_id	make	year
30	1	'Bentley'	1973
31	1	'Rolls Royce'	1965

MongoDB

```
{
first_name: "Mary",
last_name: "Jones",
cell: "516-555-2048",
city: "Long Island",
year_of_birth: 1986,
location: {
        type: "Point",
        coordinates: [-73.9876, 40.7574]
},
profession: ["Developer", "Engineer"],
apps: [
 { name: "MyApp",
   version: 1.0.4 },
 { name: "DocFinder",
   version: 2.5.7 }
 ],
cars: [
   { make: "Bentley",
     year: 1973 },
   { make: "Rolls Royce",
     year: 1965 }
  ]
}
```

TYPES OF NOSQL DATABASES

3. Document Database

TYPES OF NOSQL DATABASES

4. Graph Database

Social networks

TYPES OF NOSQL DATABASES

4. Graph Database

Social networks

Spatial data

TYPES OF NOSQL DATABASES

4. Graph Database

Social networks

Spatial data

Routing information

TYPES OF NOSQL DATABASES

4. Graph Database

Social networks

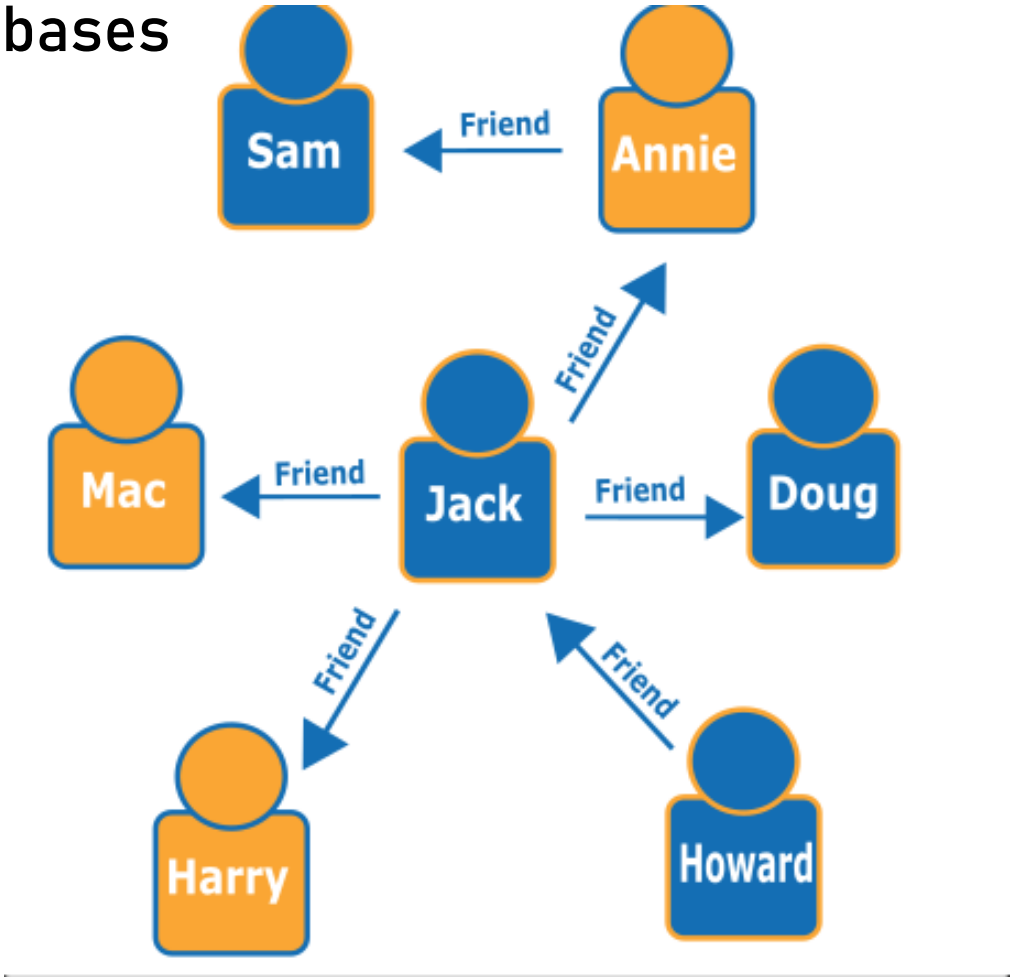
Spatial data

Routing information

Recommendation engines

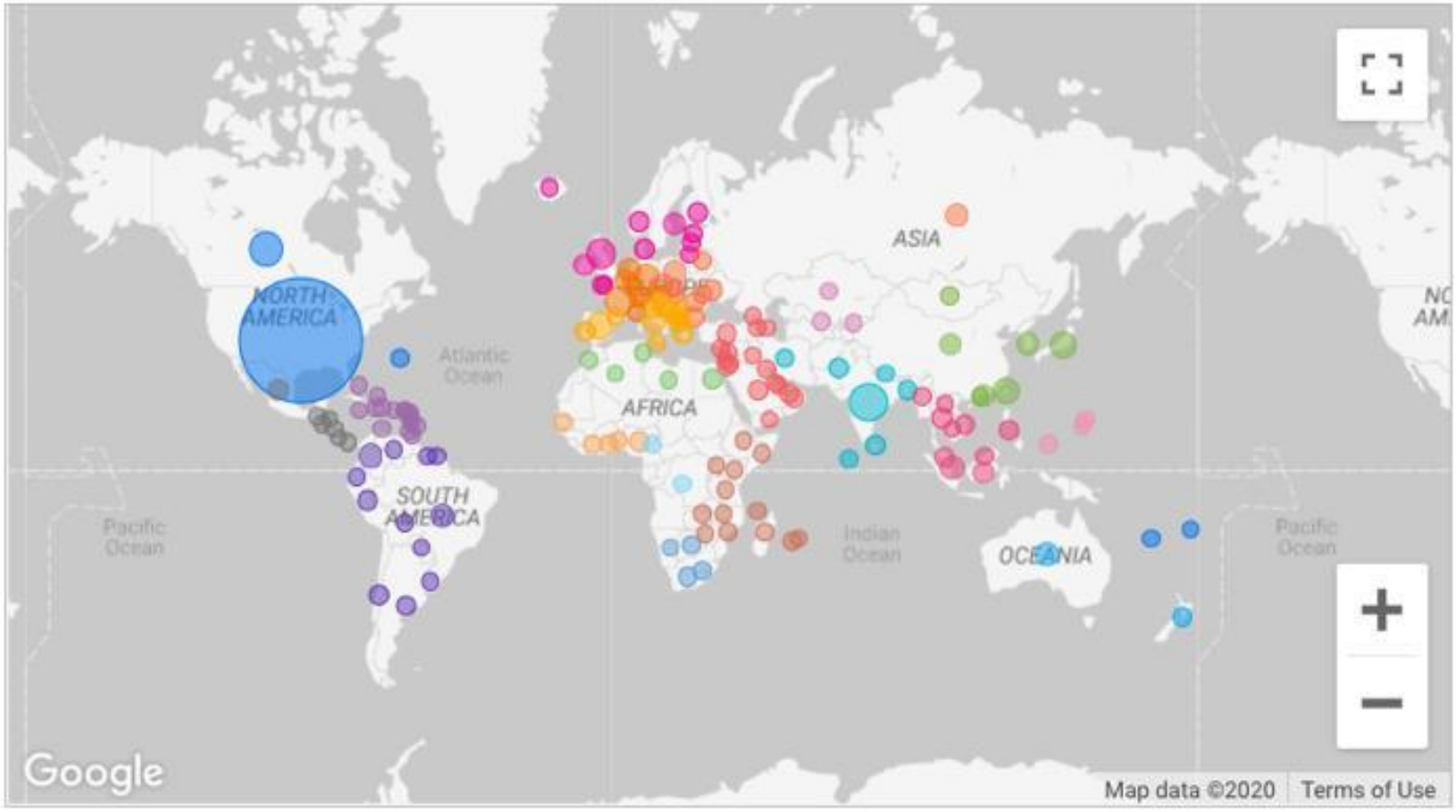
TYPES OF NOSQL DATABASES

4. Graph Databases



TYPES OF NOSQL DATABASES

4. Graph Databases



Geographic map data

TYPES OF NOSQL DATABASES

4. Graph Databases

TYPES OF NOSQL DATABASES

4. Graph Databases

TYPES OF NOSQL DATABASES

4. Graph Databases

TYPES OF NOSQL DATABASES

4. Graph Databases


That's all for now...