Introduction to Big Data

ECAP456

Dr. Rajni Bhalla

Associate Professor

Learning Outcomes

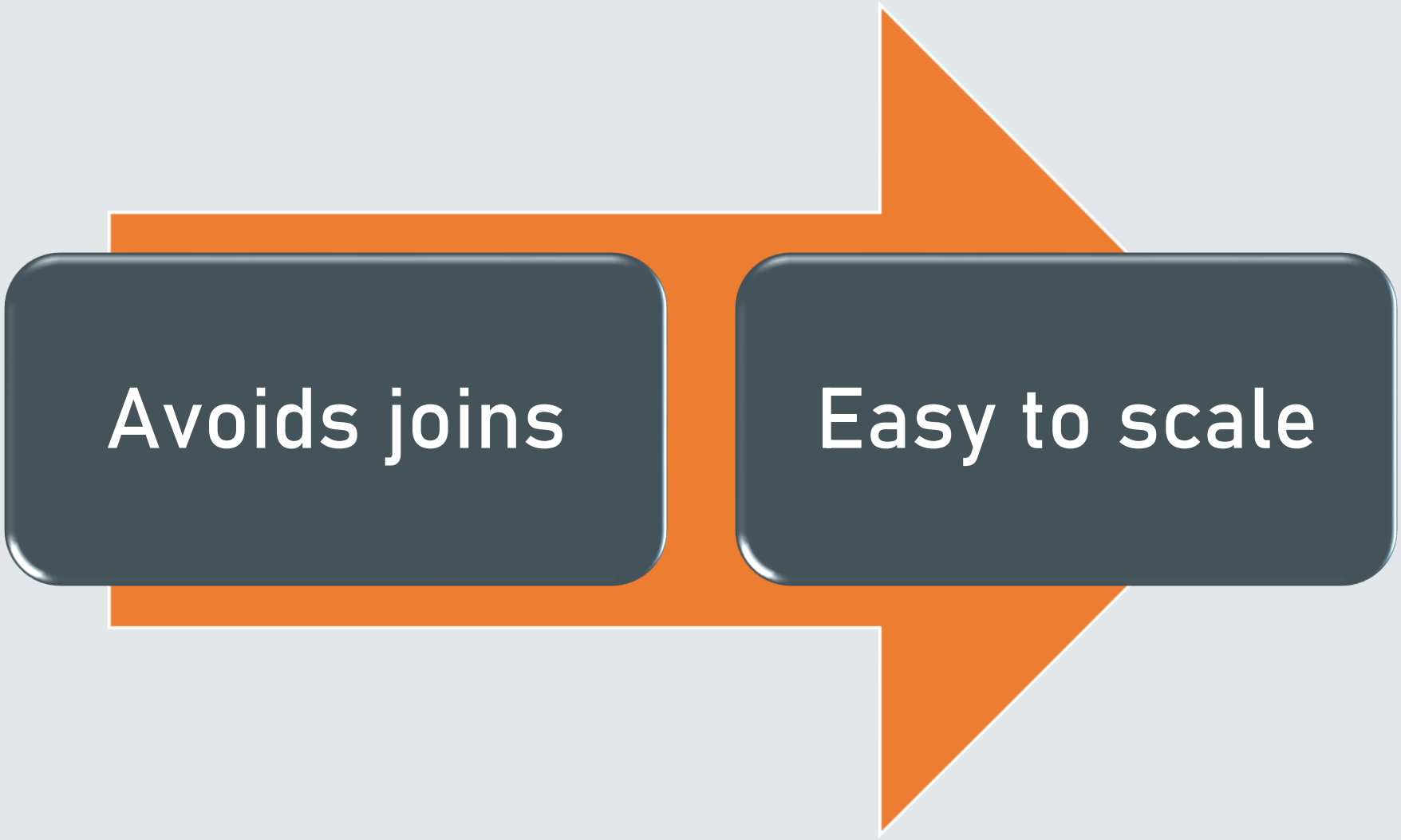
After this lecture, you will be able to

- exploring and understanding the concepts of nosql.
- learn history of nosql databases

Introduction to NoSQL

Non-relational Data Management System

Introduction to NoSQL



Introduction to NoSQL

What is the major purpose of using a NoSQL database?

- The major purpose of using a NoSQL database is for distributed data stores with huge data storage needs. NoSQL is used for Big data and real-time web apps.

Introduction to NoSQL

Introduction to NoSQL

Introduction to NoSQL

Introduction to NoSQL

- NoSQL database stands for
  - "Not Only SQL" or "Not SQL.“
  - Though a better term would be "NoREL", NoSQL caught on.
- Carl Strozz introduced the NoSQL concept in 1998.

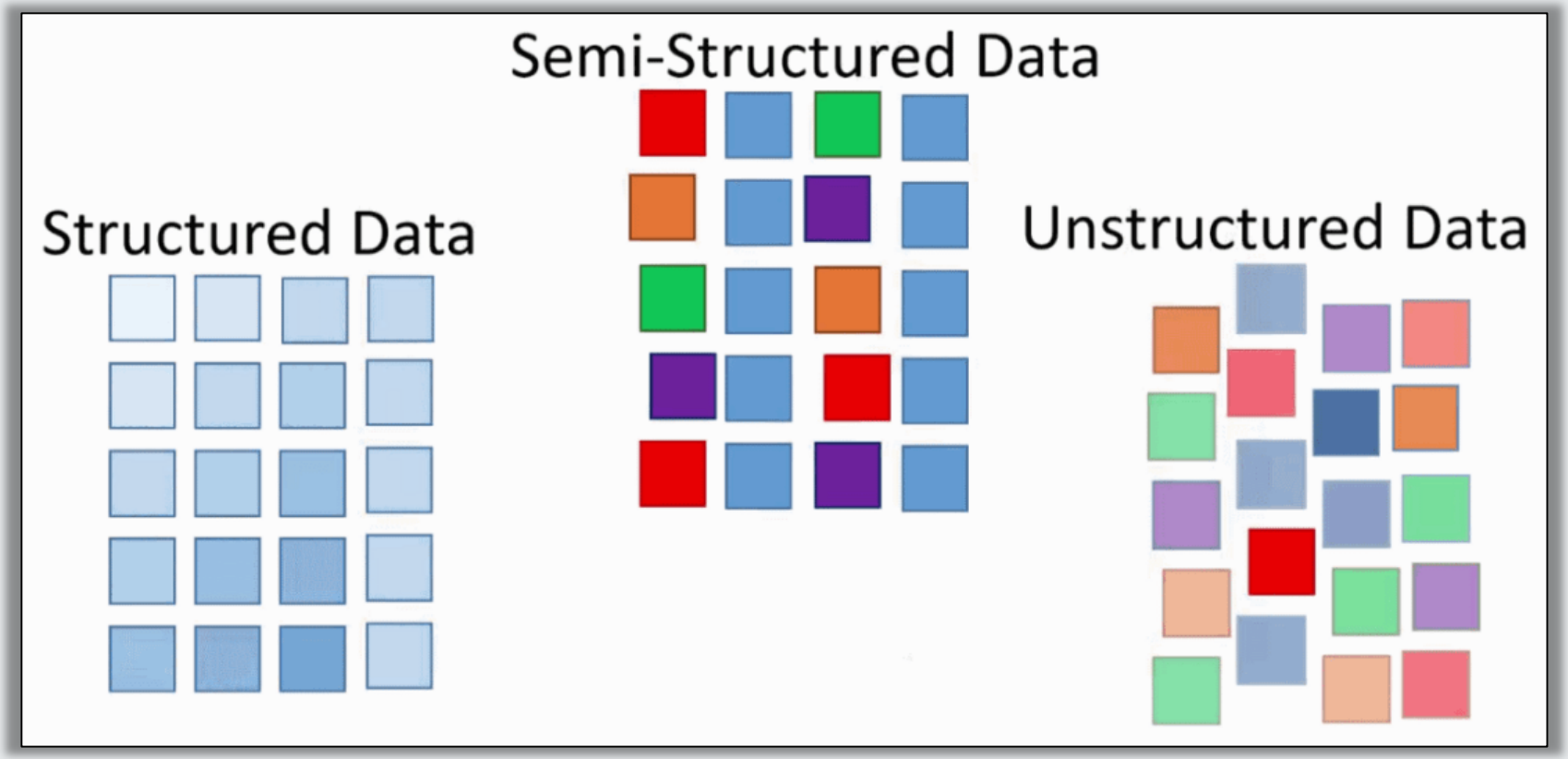
Introduction to NoSQL

Traditional RDBMS uses

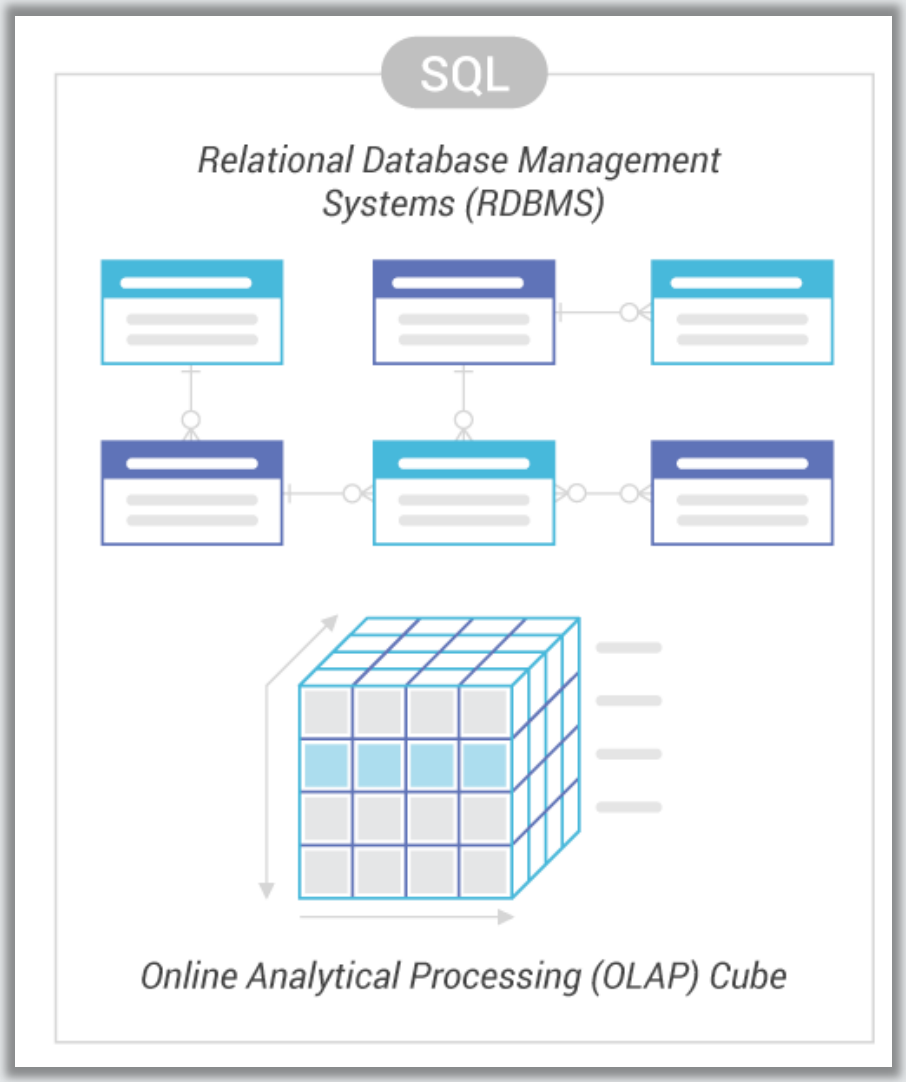
SQL Syntax

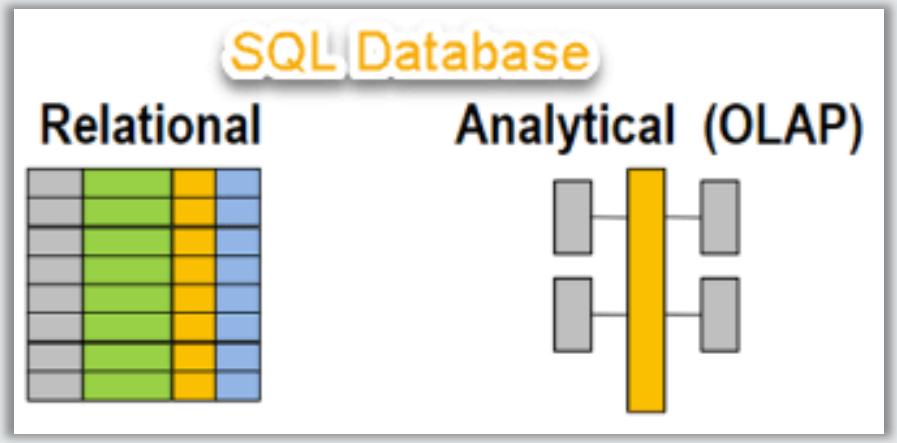
Introduction to NoSQL

Instead, a NoSQL database system covers a wide range of database technologies that can store



Example of SQL Database





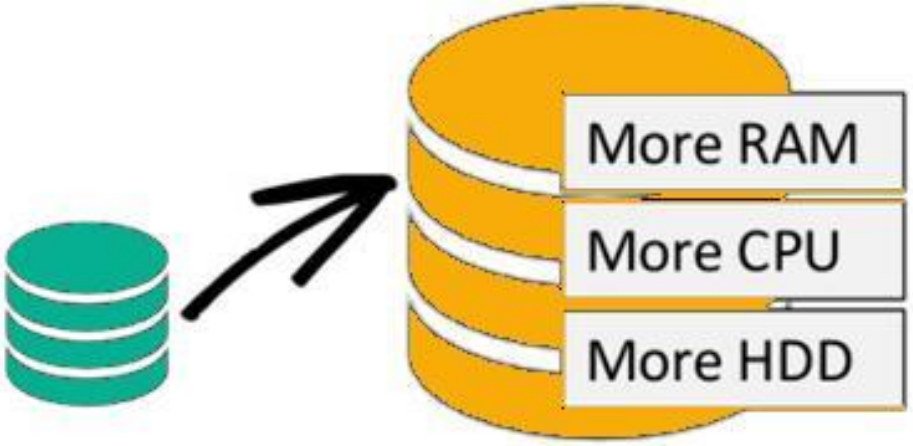
Why NoSQL?

Why NoSQL?

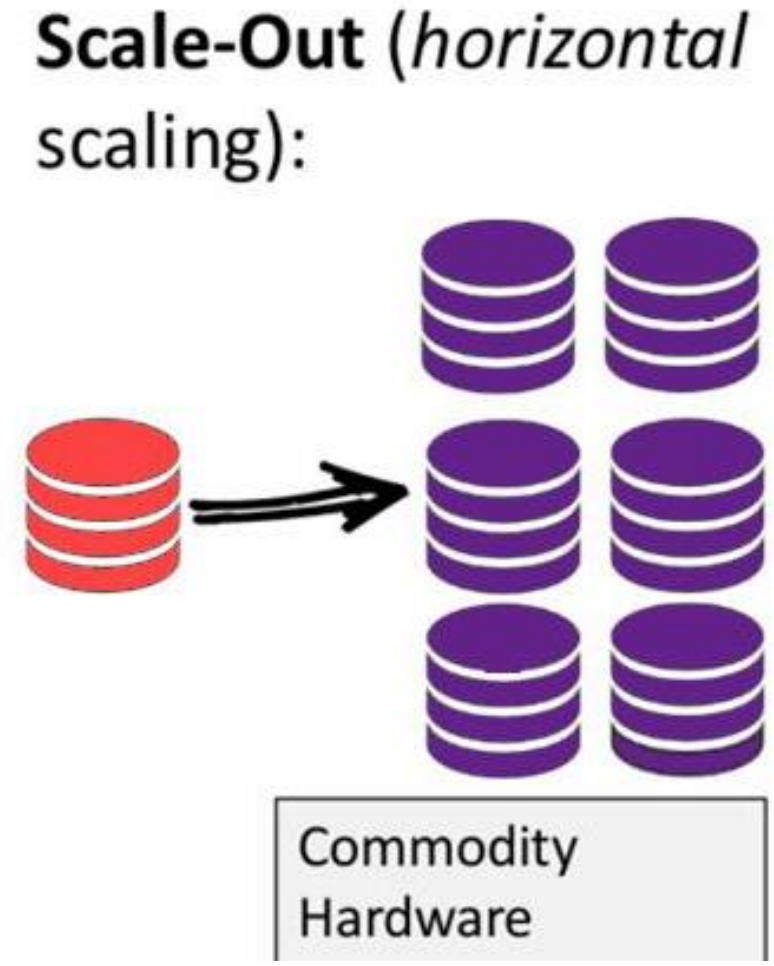
Slow Response Time

Why NoSQL?

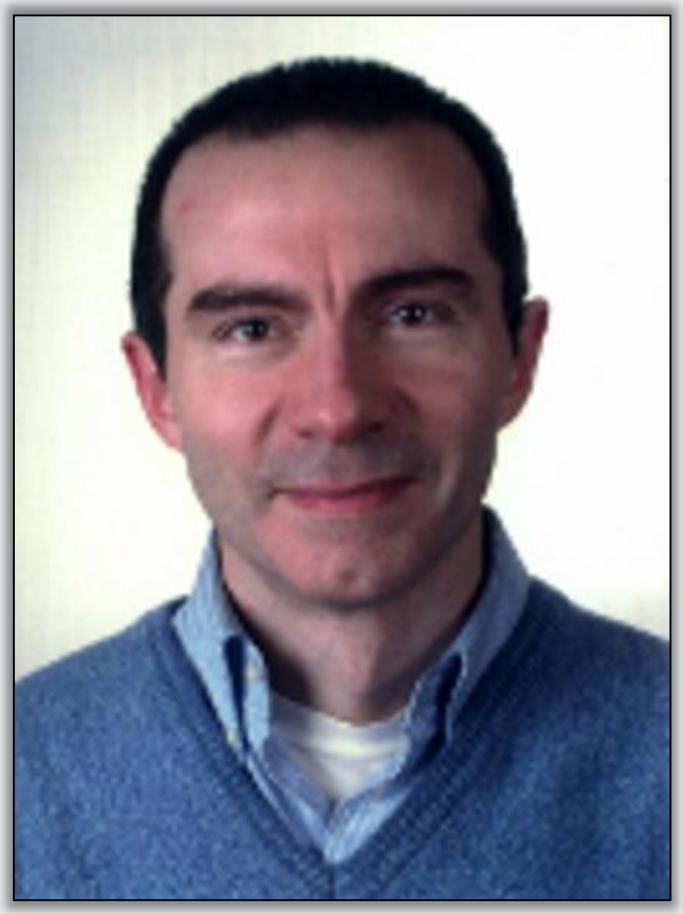
Scale-Up (*vertical* scaling):



Why NoSQL?

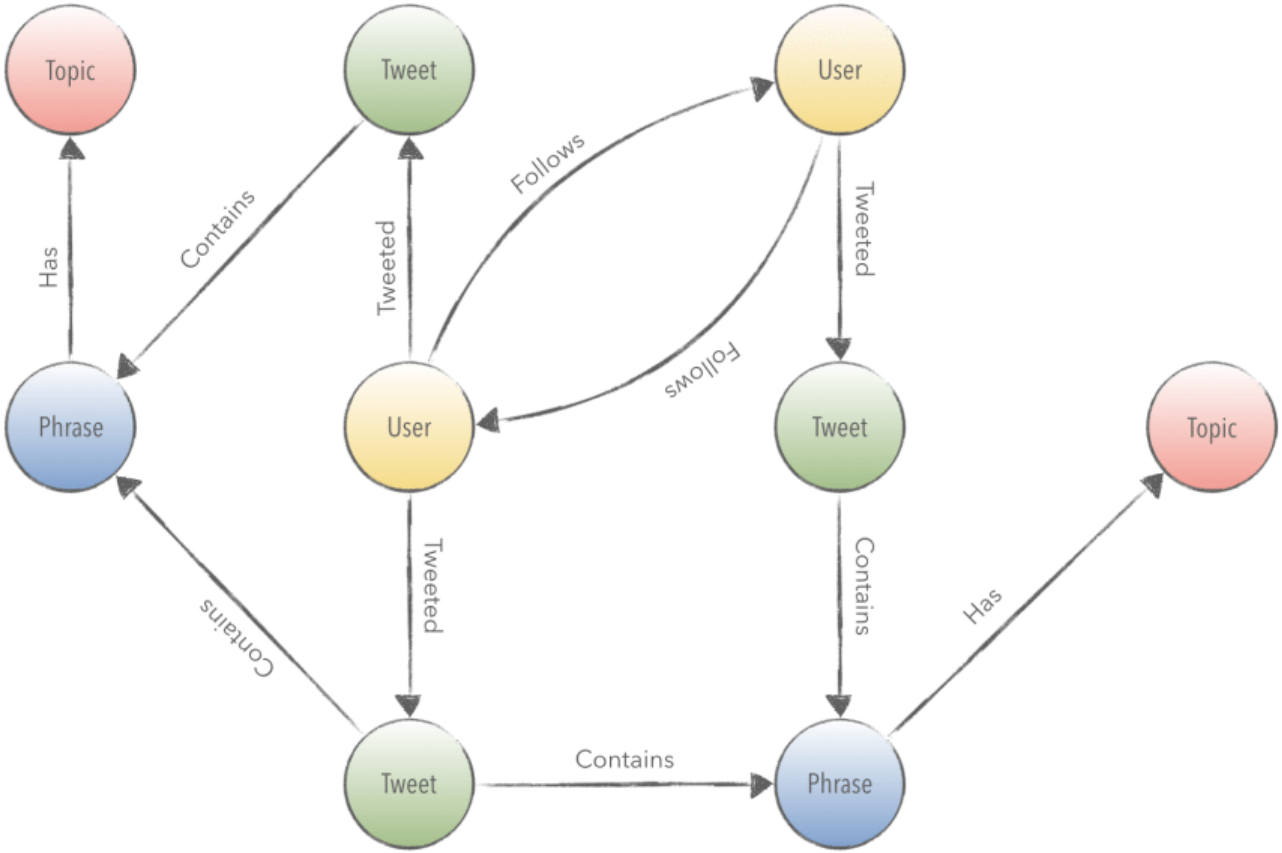


History of NoSQL Databases



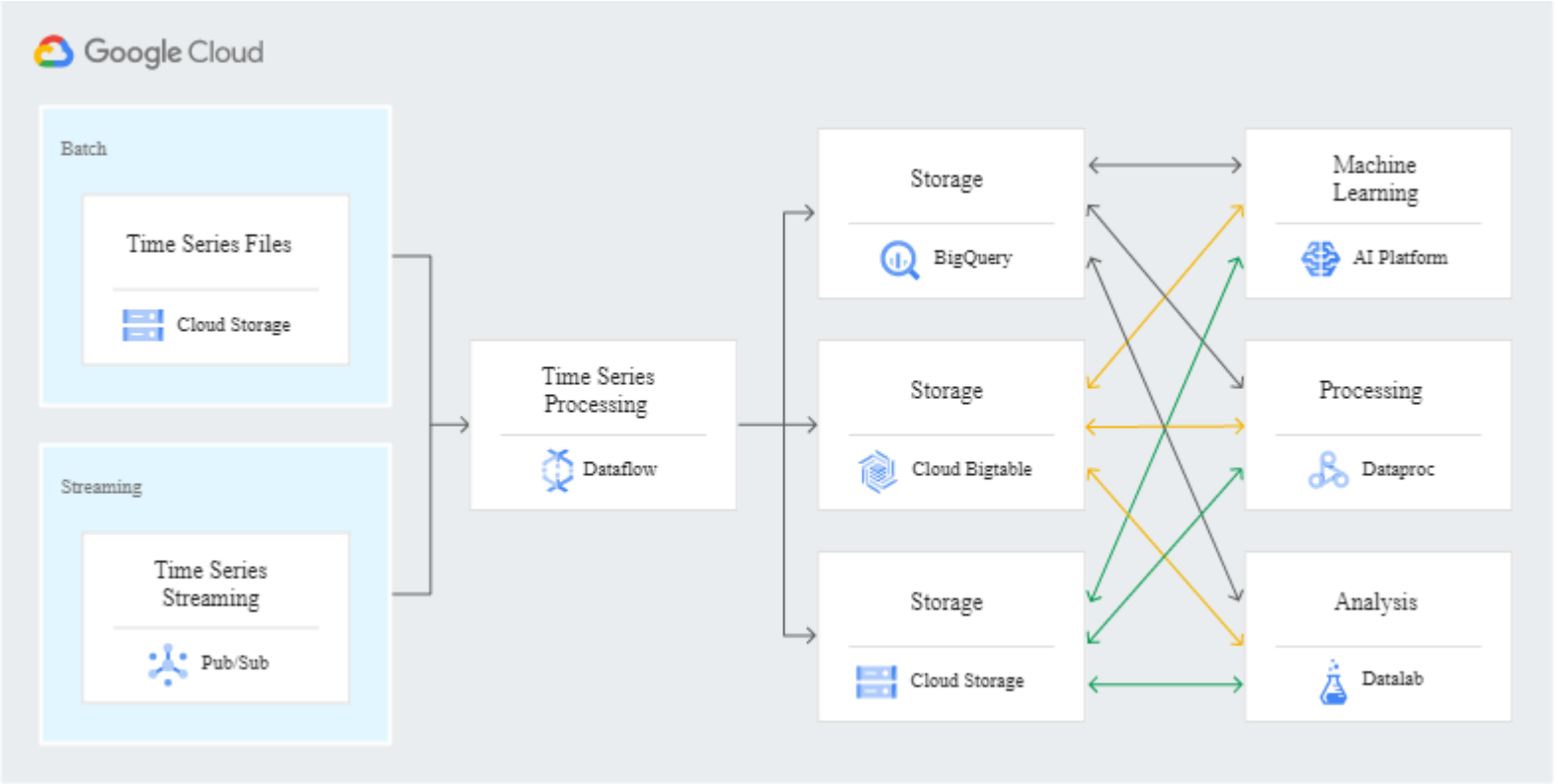
1998- Carlo Strozzi use the term NoSQL for his lightweight, open-source relational database

History of NoSQL Databases



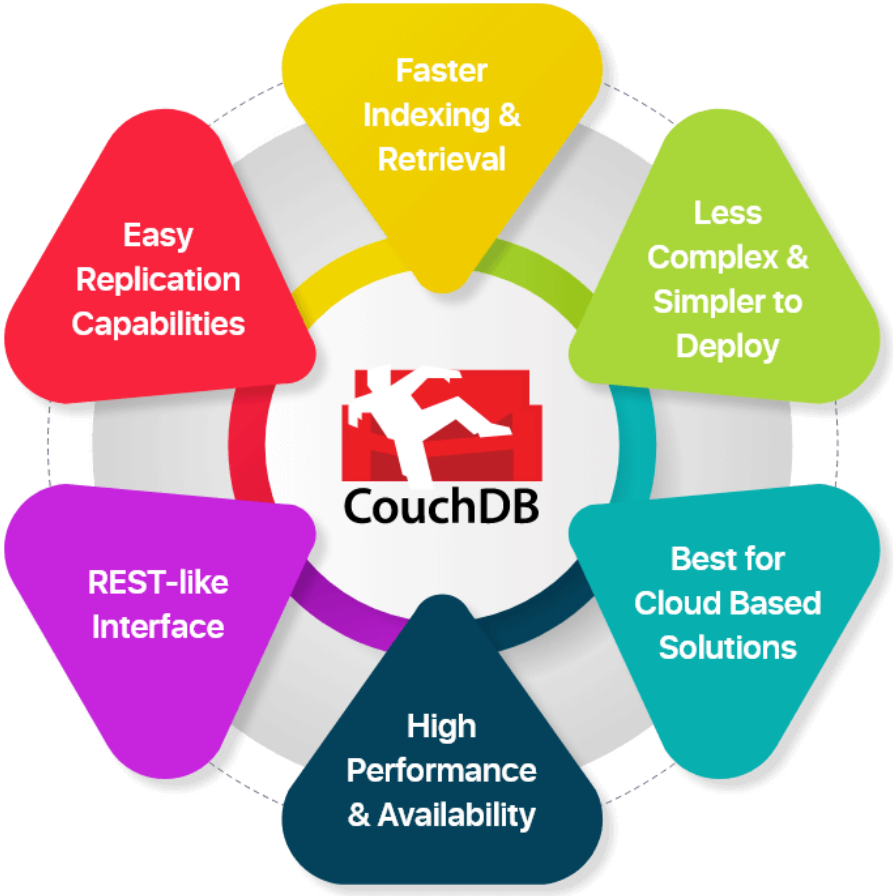
2004- Google BigTable is launched

History of NoSQL Databases



2000- Graph database Neo4j is launched

History of NoSQL Databases



2005- CouchDB is launched

History of NoSQL Databases

=

- Predictable performance
- Highly available
- Massively scalable
- Fully managed
- Low cost

2007- The research paper on Amazon Dynamo is released

History of NoSQL Databases

cassandra

2008- Facebooks open sources the Cassandra project

History of NoSQL Databases

2009- The term NoSQL was reintroduced

Features of NoSQL

- Non-relational
- Schema-free
- Simple API
- Distributed

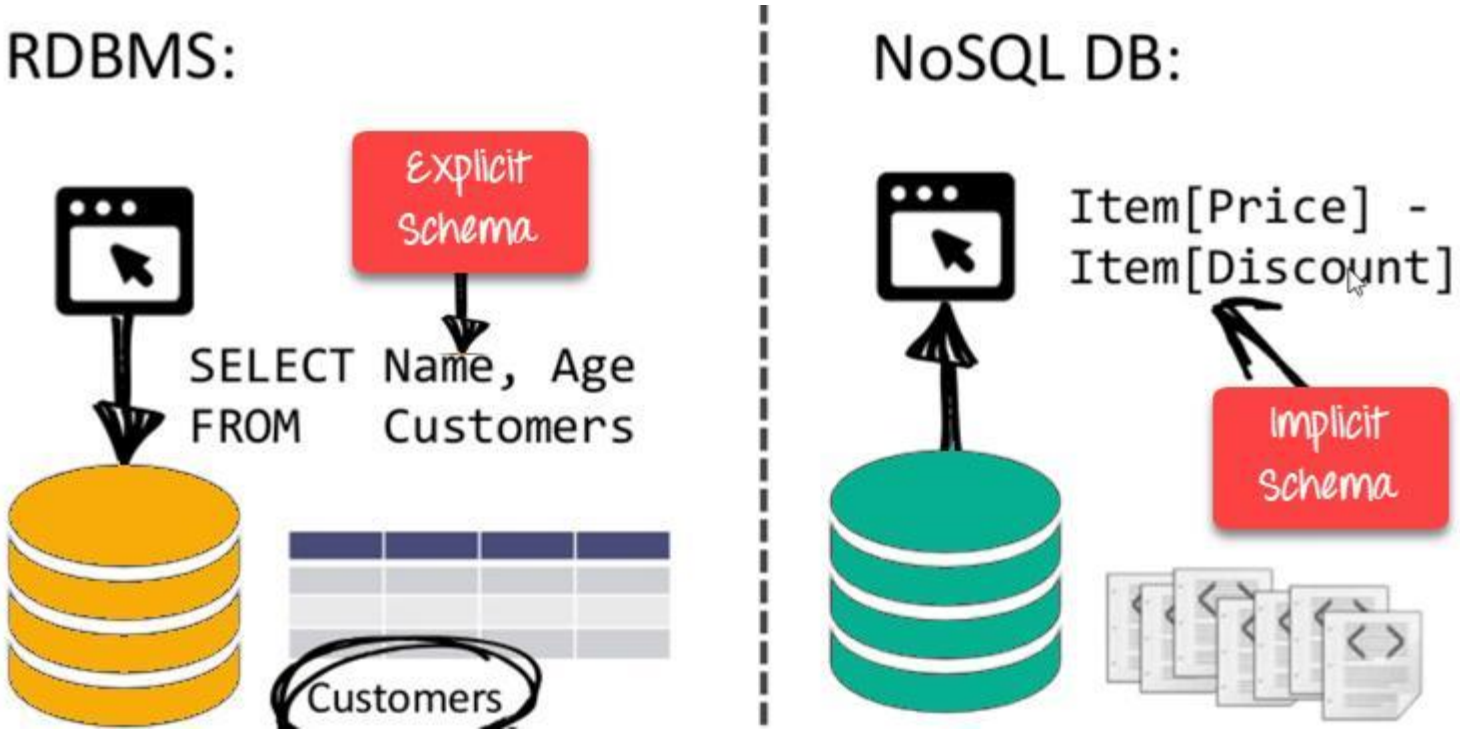
Features of NoSQL

Non-relational

- Never follow the relational model
- Never provide tables with flat fixed-column records
- Work with self-contained aggregates or BLOBs
- Doesn't require object-relational mapping and data normalization
- No complex features like query languages, query planners, referential integrity joins, ACID

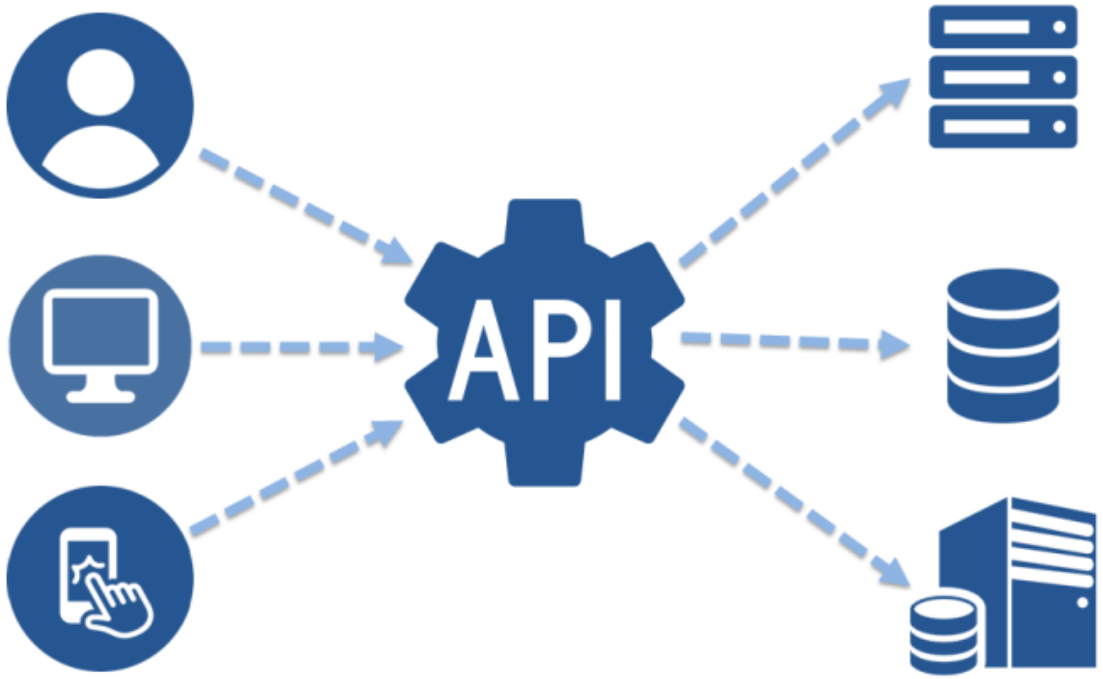
Features of NoSQL

Schema-free



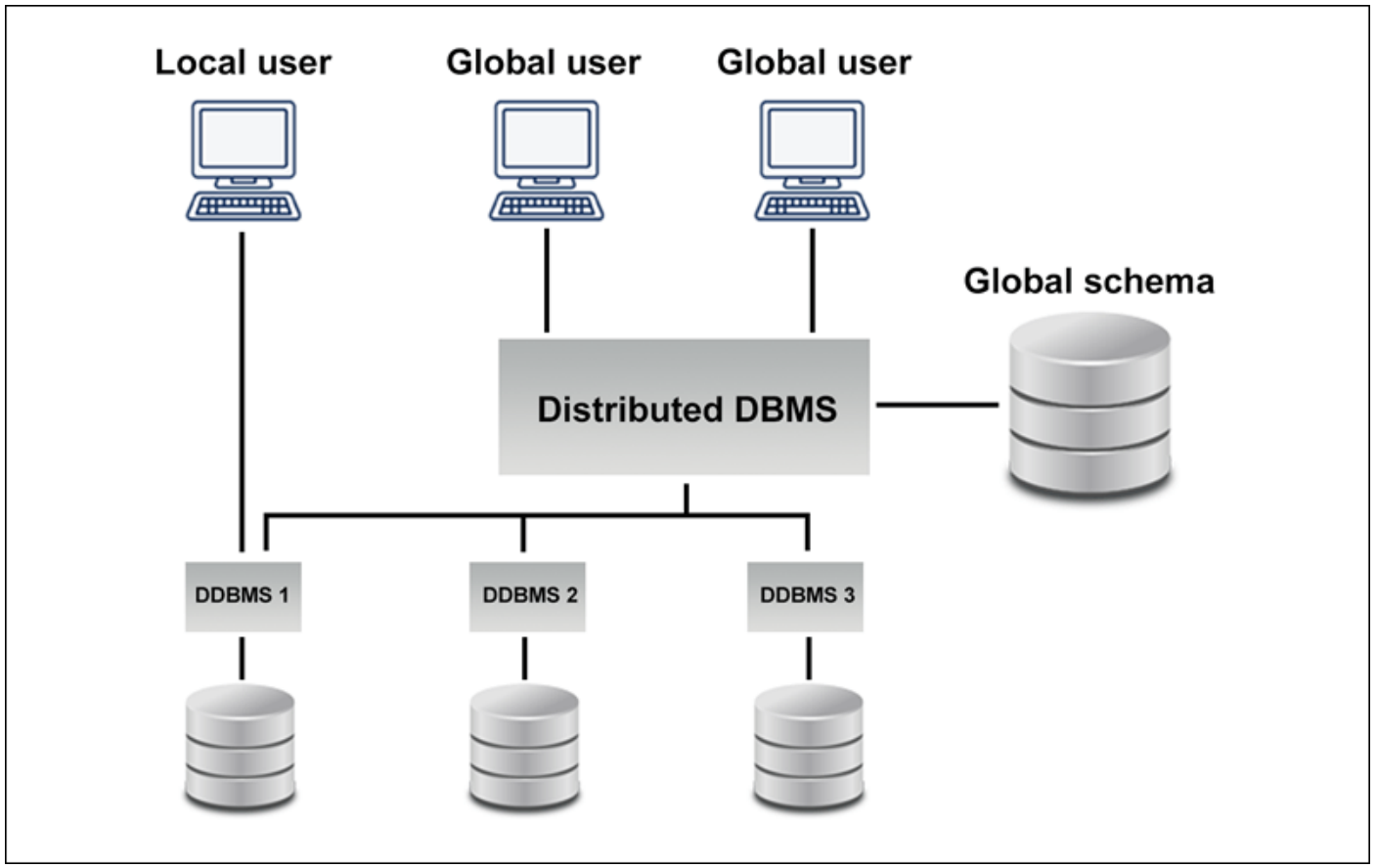
Features of NoSQL

Simple API



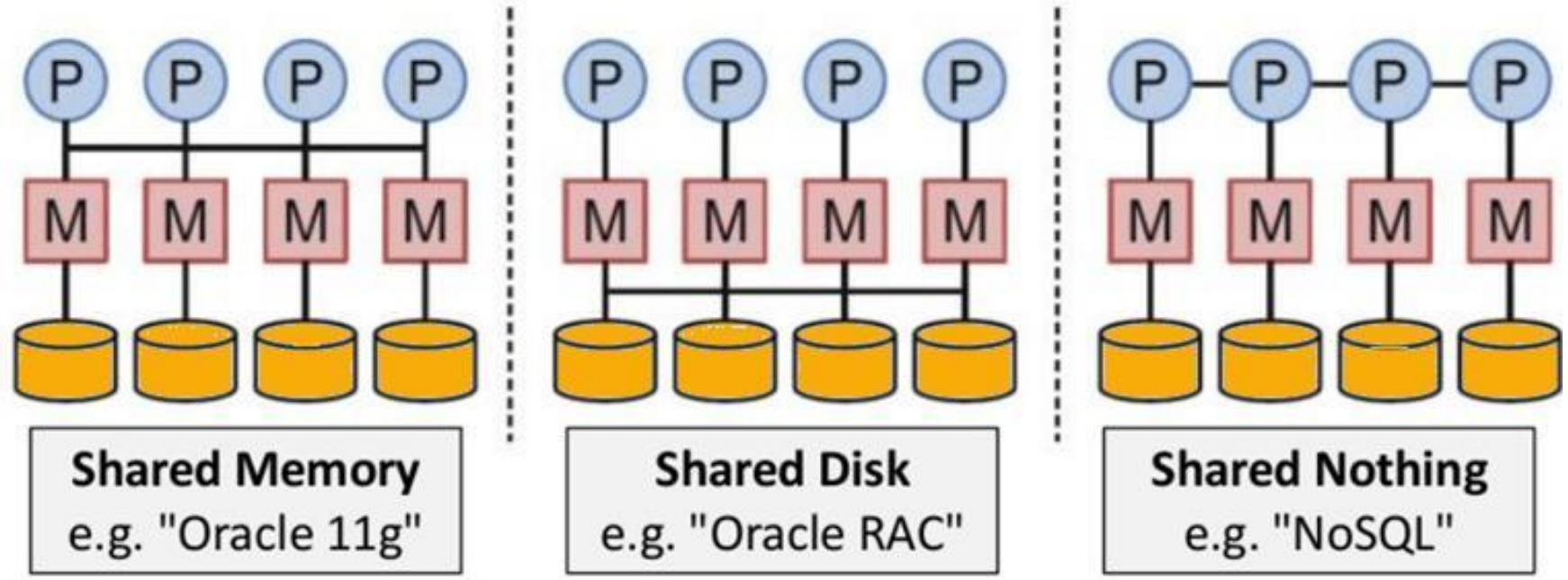
Features of NoSQL

Distributed

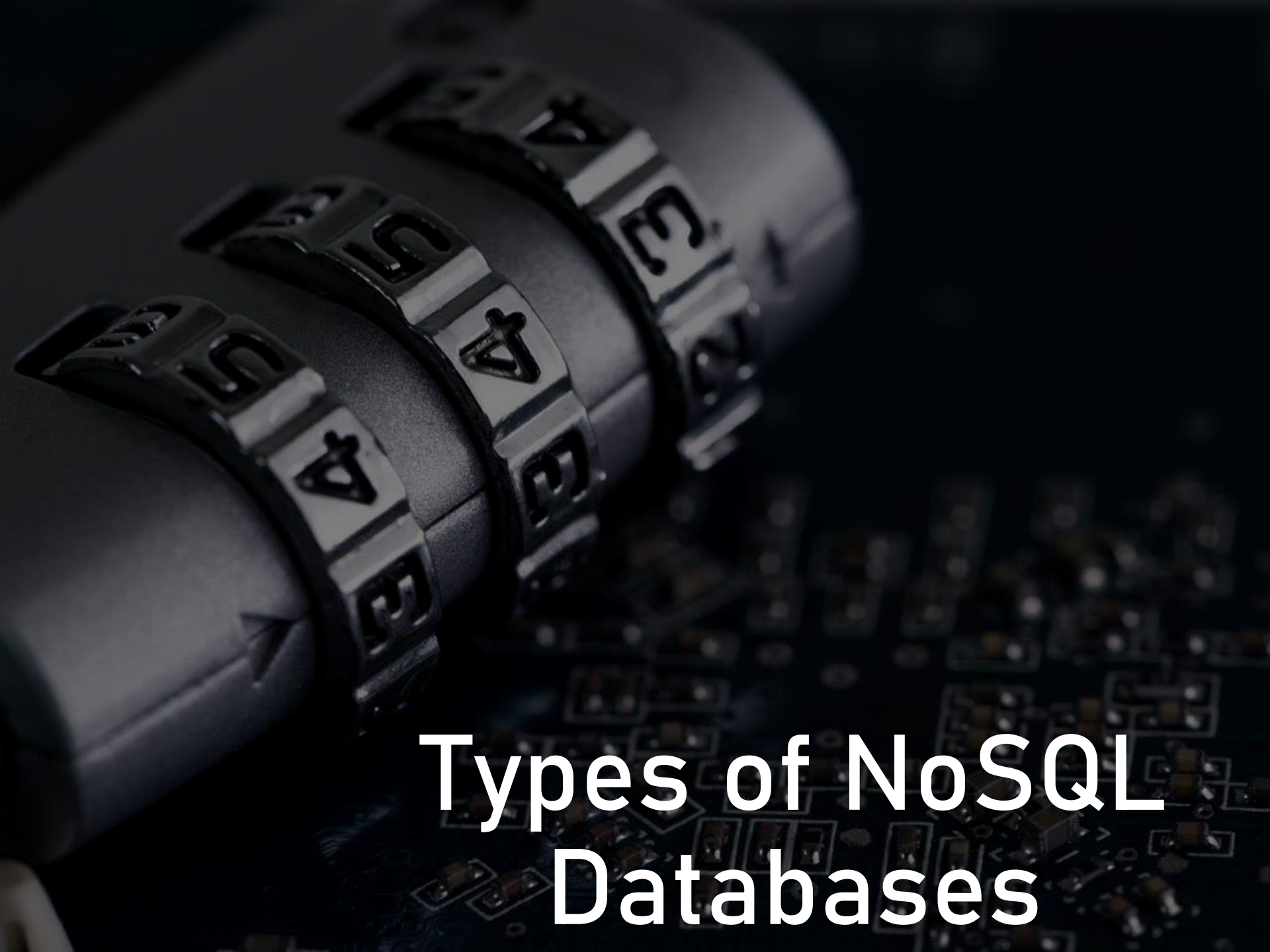


Features of NoSQL

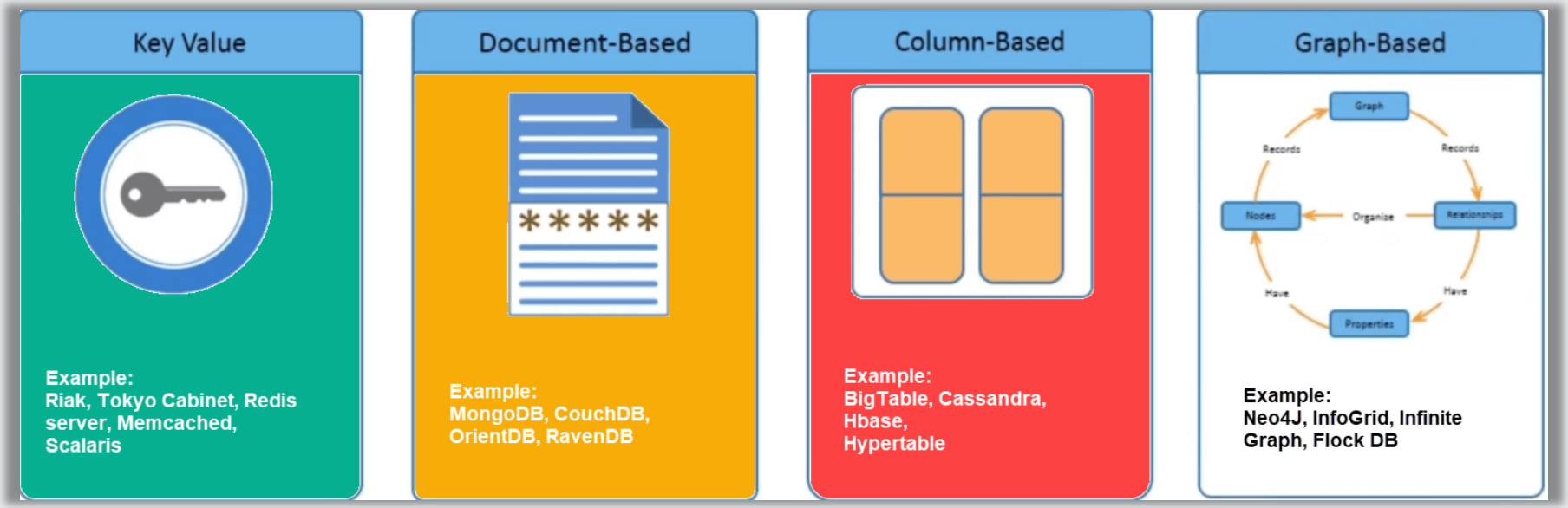
Distributed



NoSQL is Shared Nothing.


Types of NoSQL
Databases

Types of NoSQL Databases



Key Value Pair Based

Key	Value
Name	Joe Bloggs
Age	42
Occupation	Stunt Double
Height	175cm
Weight	77kg

Key Value Pair Based

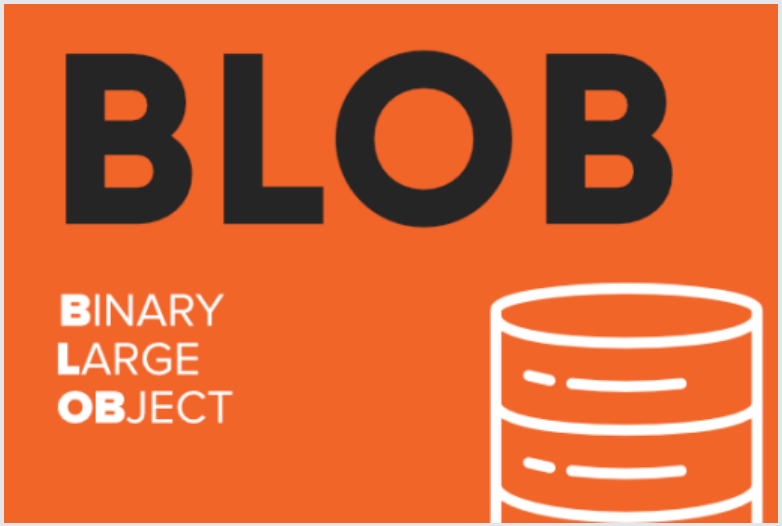
```
{
    "d": {
        "results": [
            {
                "__metadata": {
                    "type": "EmployeeDetails.Employee"
                },
                "UserID": "E12012",
                "RoleCode": "35"
            }
        ]
    }
}
```

JSON

Key Value Pair Based

```json
{
    "d": {
        "results": [
            {
                "__metadata": {
                    "type": "EmployeeDetails.Employee"
                },
                "UserID": "E12012",
                "RoleCode": "35"
            }
        ]
    }
}
```

JSON

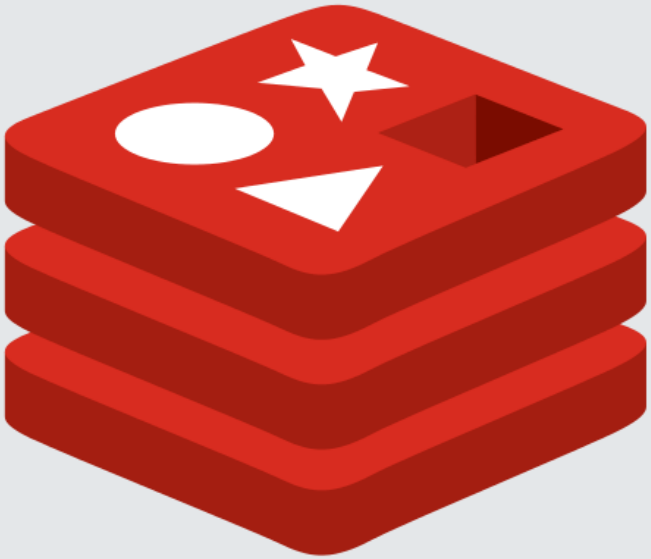


BLOB(Binary Large Objects)

Key Value Pair Based

- NoSQL database is used as a collection, dictionaries, associative arrays, etc.
- Key value stores help the developer to store schema-less data.
- They work best for shopping cart contents.

Key Value Pair Based



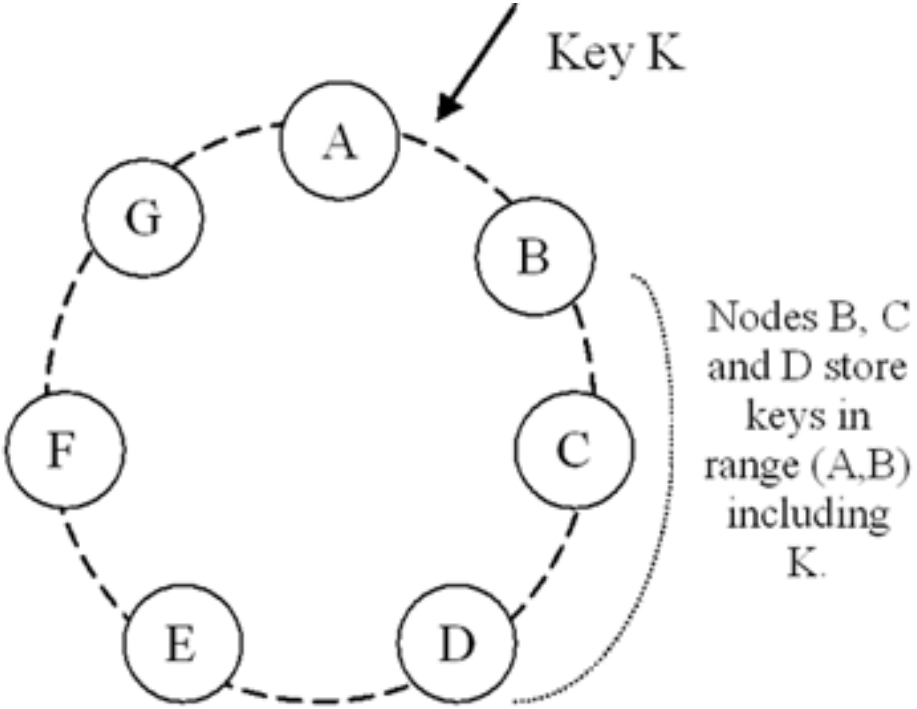
redis

Key Value Pair Based

Amazon DynamoDB

Key Value Pair Based

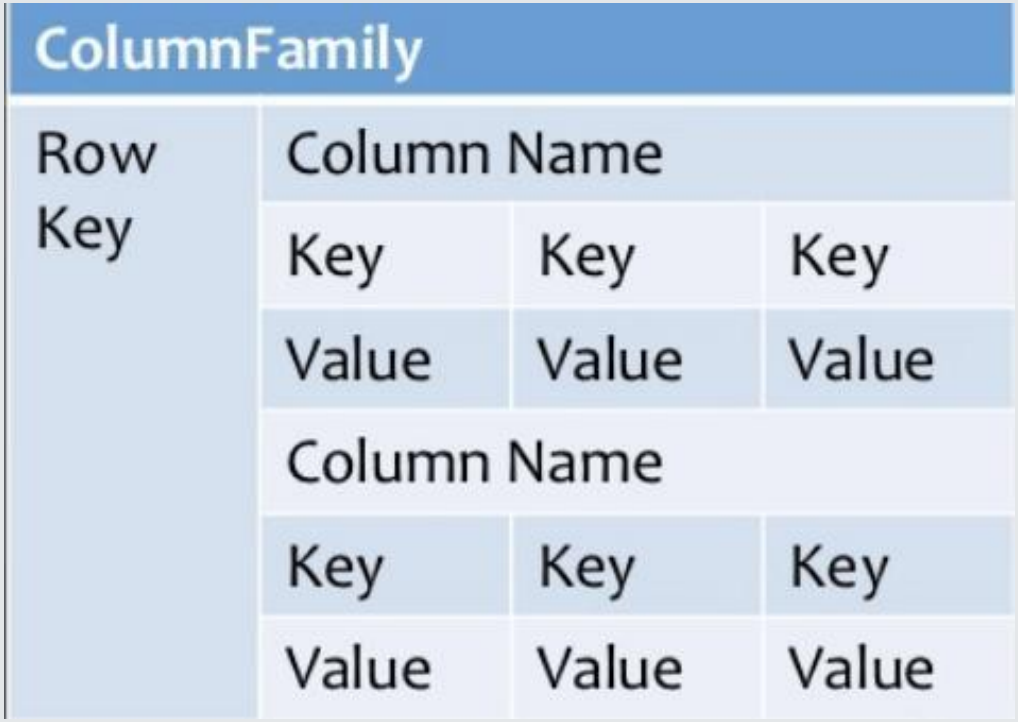
Key Value Pair Based



Based on Amazon's Dynamo paper.

COLUMN-BASED

Column-based



Column-based

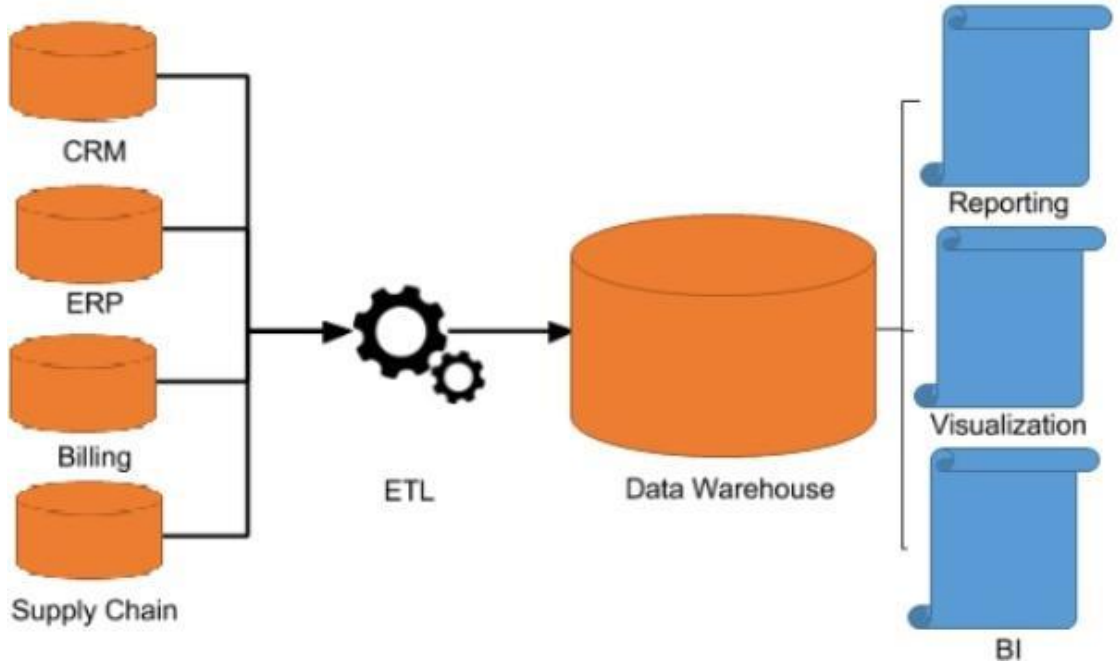
They deliver high performance on aggregation queries

- COUNT( ) returns the number of rows
- SUM( ) returns the sum
- AVG( ) returns the average value
- MIN( ) returns the smallest value
- MAX( ) returns the largest value
- FIRST( ) returns the first value
- LAST( ) returns the last value

ColumnFamily			
Row Key	Column Name		
	Key	Key	Key
	Value	Value	Value
	Column Name		
	Key	Key	Key
	Value	Value	Value

Column-based

Column-based NoSQL databases are widely used to manage



Dataware house

Column-based

Column-based NoSQL databases are widely used to manage

Column-based

Column-based NoSQL databases are widely used to manage

Library card catalogs

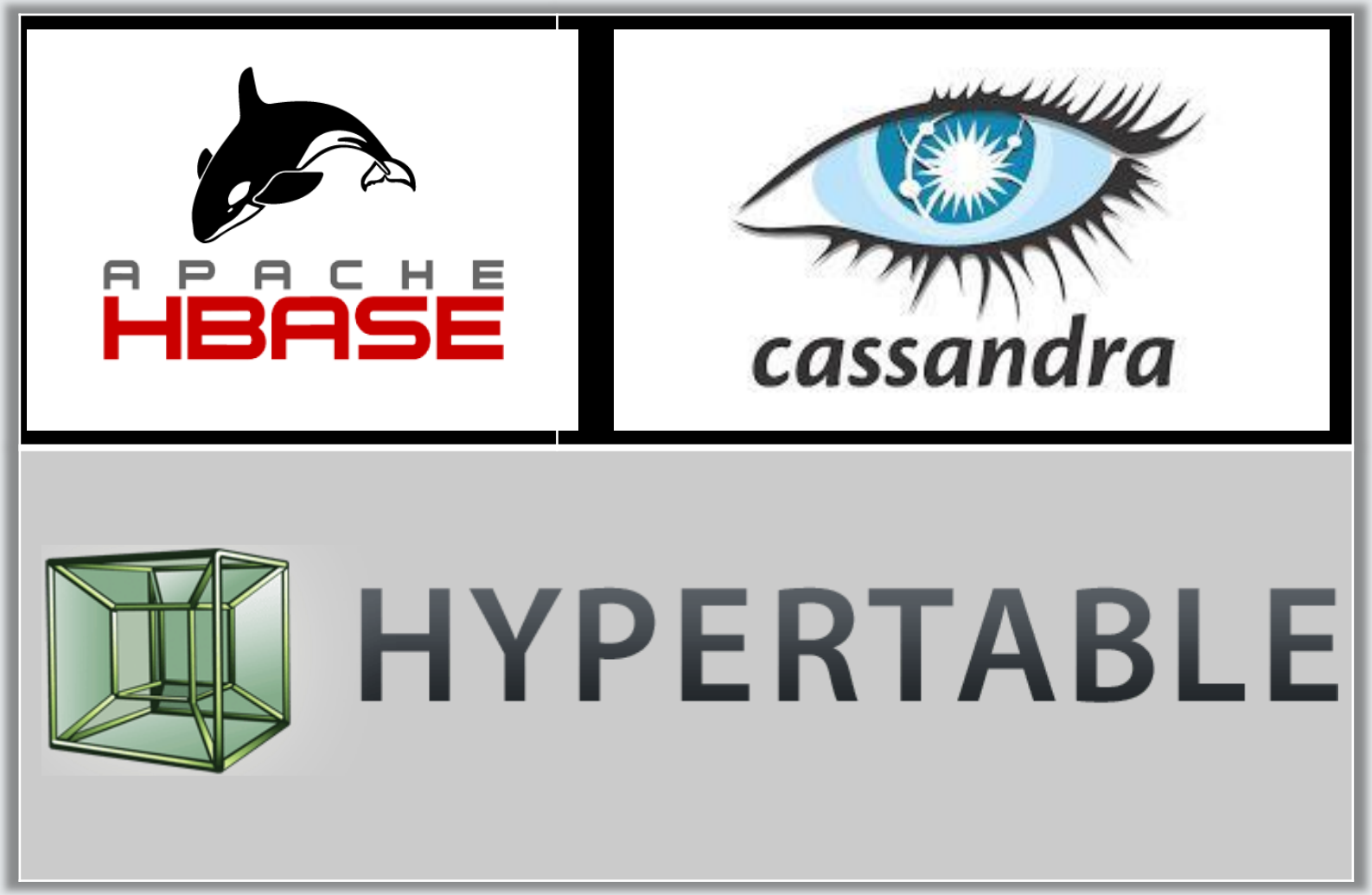
Column-based

Column-based NoSQL databases are widely used to manage

Column-based

Column-based

Column-based



Document-Oriented

Col1	Col2	Col3	Col4
Data	Data	Data	Data
Data	Data	Data	Data
Data	Data	Data	Data

Document 1

```
{
  "prop1": data,
  "prop2": data,
  "prop3": data,
  "prop4": data
}
```

Document 2

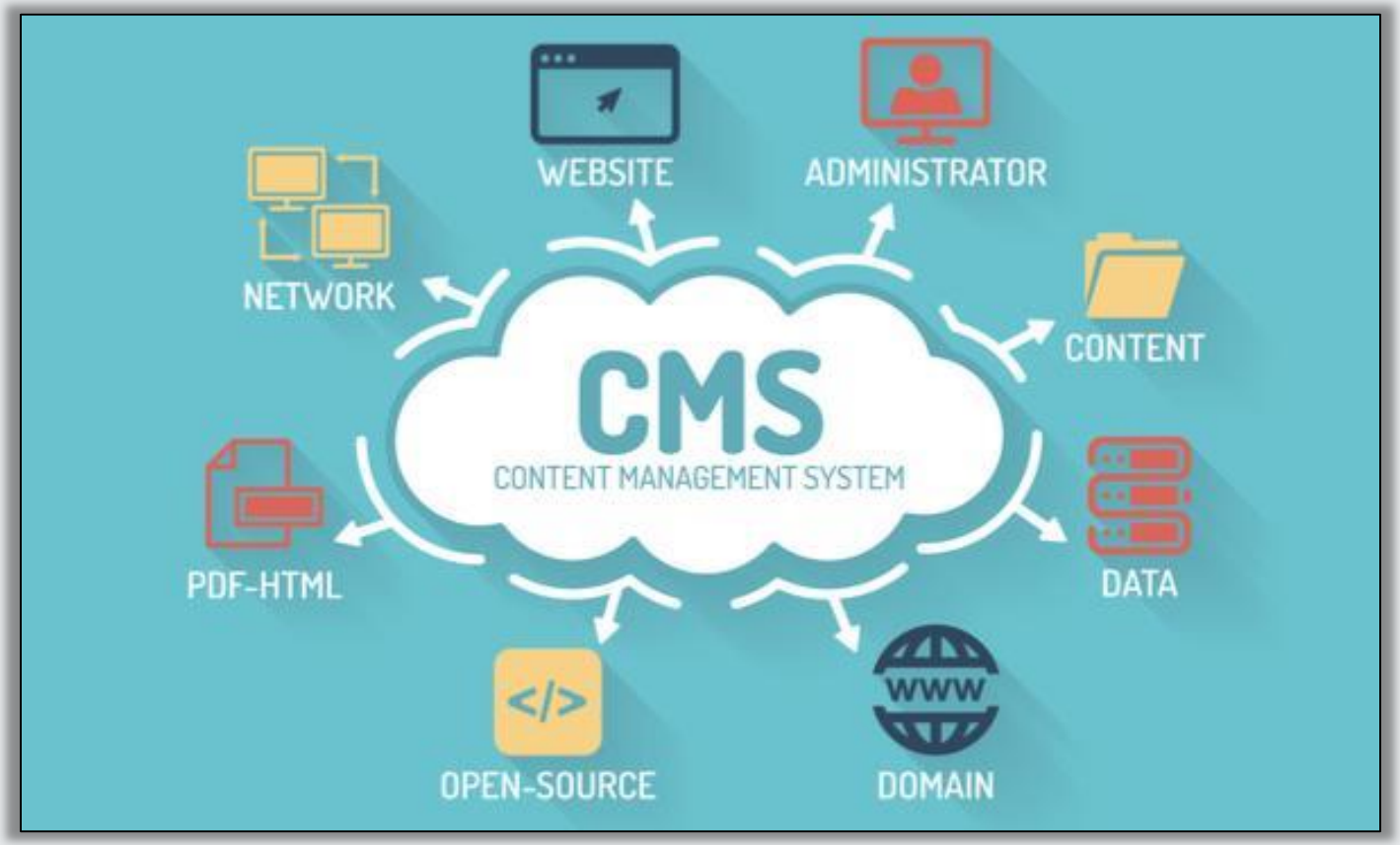
```
{
  "prop1": data,
  "prop2": data,
  "prop3": data,
  "prop4": data
}
```

Document 3

```
{
  "prop1": data,
  "prop2": data,
  "prop3": data,
  "prop4": data
}
```

Document-Oriented

The document type is mostly used for



Document-Oriented

The document type is mostly used for

Document-Oriented

The document type is mostly used for

Document-Oriented

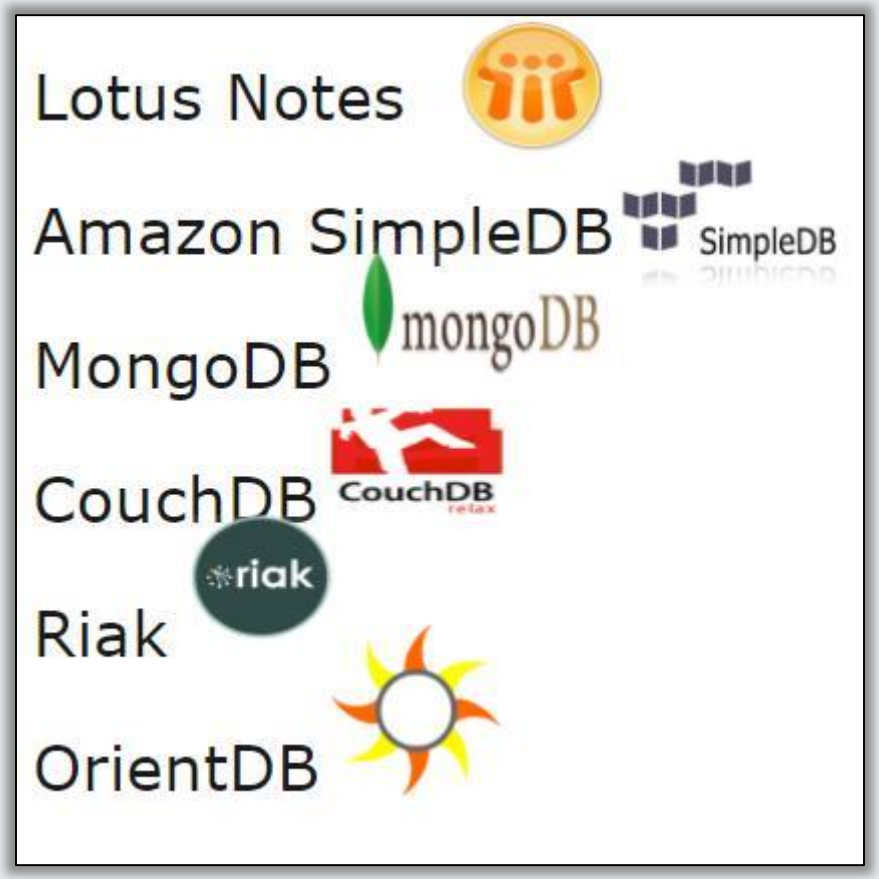
The document type is mostly used for

Document-Oriented

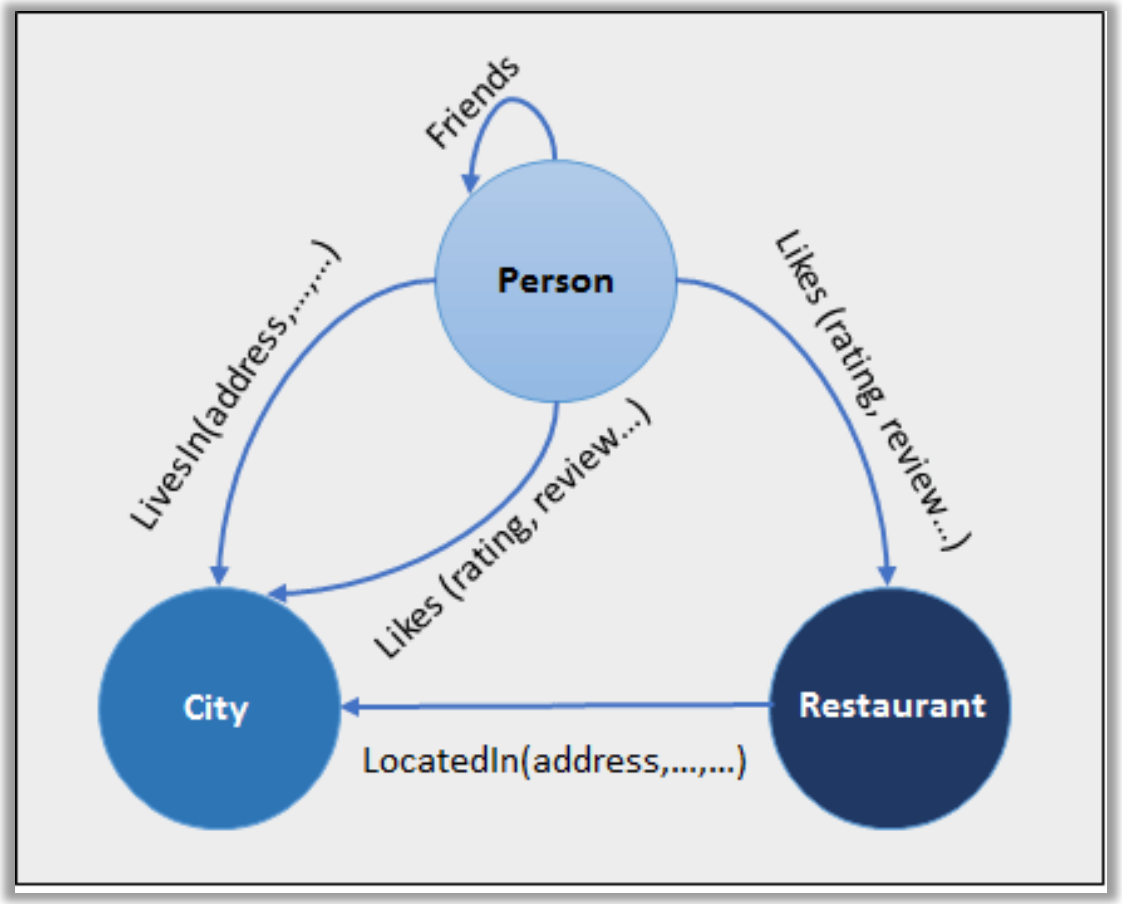
It should not use for complex transactions which require multiple operations or queries against varying aggregate structures.

Document-Oriented

Popular Document originated DBMS systems



Graph-Based


Friends
Person
LivesIn(address,...,...)
Likes (rating, review...)
Likes (rating, review...)
City
Restaurant
LocatedIn(address,...,...)

Graph-Based

Graph base database mostly used for

- social networks,
- logistics,
- spatial data.

Graph-Based

Some popular graph-based databases.

- Neo4J,
- Infinite Graph,
- OrientDB,
- FlockDB

Query Mechanism tools for NoSQL

- The most common data retrieval mechanism is the REST-based retrieval of a value based on its key/ID with GET resource
- Document store Database offers more difficult queries as they understand the value in a key-value pair.
  - For example, CouchDB allows defining views with MapReduce


That's all for now...