

Les Bases de données

« NoSQL »

Part 3

Présenté par

Marie Esther BOUA & Marc-Dexter ZEGBOLOU
Data Scientist



Modèle de données

Rappel

Un modèle de base de données illustre la **structure logique** d'une base de données, y compris les **relations** et les **contraintes** qui déterminent comment les données peuvent être stockées et accessibles.



Modèle de données

Dans les SGBDR les données sont enregistrées sous **forme Tabulaire**

Exemple : Représentation de ventes en relationnel

Table Sales

#ticket	#date	#book
1	01/01/16	2212121504
1	01/01/16	2212141556
2	01/01/16	2212141556

Table Book

#isbn	#title	#author
2212121504	Scenari	1
2212141556	NoSQL	2

Table Author

#id	surname	firstname
1		
2	Bruchez	Rudi



Ranking Modèle Relationnel

<https://db-engines.com/en/ranking/relational+dbms>

☐ include secondary database models

146 systems in ranking, June 2021

Rank			DBMS	Database Model	Score		
Jun 2021	May 2021	Jun 2020			Jun 2021	May 2021	Jun 2020
1.	1.	1.	Oracle +	Relational, Multi-model ⓘ	1270.94	+1.00	-72.65
2.	2.	2.	MySQL +	Relational, Multi-model ⓘ	1227.86	-8.52	-50.03
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model ⓘ	991.07	-1.59	-76.24
4.	4.	4.	PostgreSQL +	Relational, Multi-model ⓘ	568.51	+9.26	+45.53
5.	5.	5.	IBM Db2 +	Relational, Multi-model ⓘ	167.03	+0.37	+5.23
6.	6.	6.	SQLite +	Relational	130.54	+3.84	+5.72
7.	7.	7.	Microsoft Access	Relational	114.94	-0.46	-2.24
8.	8.	8.	MariaDB +	Relational, Multi-model ⓘ	96.79	+0.10	+7.00
9.	9.	9.	Hive	Relational	79.69	+3.51	+1.04
10.	10.	↑ 14.	Microsoft Azure SQL Database	Relational, Multi-model ⓘ	74.79	+4.33	+27.01
11.	11.	↓ 10.	Teradata	Relational, Multi-model ⓘ	69.34	-0.64	-3.95
12.	12.	12.	SAP HANA +	Relational, Multi-model ⓘ	54.11	+1.36	+3.29
13.	↑ 14.	13.	FileMaker	Relational	49.13	+2.40	-1.02
14.	↓ 13.	↓ 11.	SAP Adaptive Server	Relational, Multi-model ⓘ	48.99	-0.98	-4.11
15.	15.	15.	Google BigQuery +	Relational	39.63	+2.00	+11.33
16.	16.	↑ 49.	Snowflake +	Relational	34.75	+4.71	+32.22



Modèle de données NoSql

**Modèle Clé-
Valeur**

A

B

**Modèle
Document**

C

**Modèle
Colonne**

D

**Modèle
Graphe**

Modèle de données NoSQL orienté

Clé - valeur

An orange key icon with a circular head and a notched blade, positioned at the end of the 'Clé - valeur' text.



Modèle de données: Clé - valeur

- Stocke les informations sous forme de ***paires clé-valeur*** dans lesquelles une clé sert ***d'identifiant unique***.
- la **valeur associée** à une clé peut être ***une simple chaîne de caractère comme un document, ou encore un objet beaucoup plus complexe*** pouvant contenir une multitude d'information.





Modèle de données: Clé - valeur

- Exemple:

CISSE



Num : 09xxxxxx
Adresse: Abidjan
Rue : Lumière

AMADOU





Modèle de données: Clé - valeur

Dans quel but l'utiliser?

➤ **Applications sollicitant un groupe d'info selon une entrée**

➤ Absence d'écritures

➤ Pas d'accès à une partie du contenu

→ E-commerce

→ Fichiers Log

→ chat





Modèle de données: Clé - valeur

Avantages

- Les clés et les valeurs peuvent se présenter sous plusieurs formes, textes, objets simples ou objets composés
- hautement divisibles et permettent une mise à l'échelle horizontale à des échelles que d'autres types de bases de données ne peuvent pas atteindre.





Modèle de données: Clé - valeur

Dans quel cas utiliser le modèle clé valeur?

- Applications manipulant des collection d'informations
- Index primaire sur la clé, absence de clé secondaire
- Impossibilité d'accès via le contenu.





Modèle de données: Clé - valeur

Exemple et caractéristiques

<https://db-engines.com/en/ranking/key-value+store>

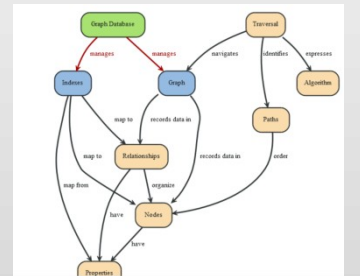


☐ include secondary database models

63 systems in ranking, June 2021

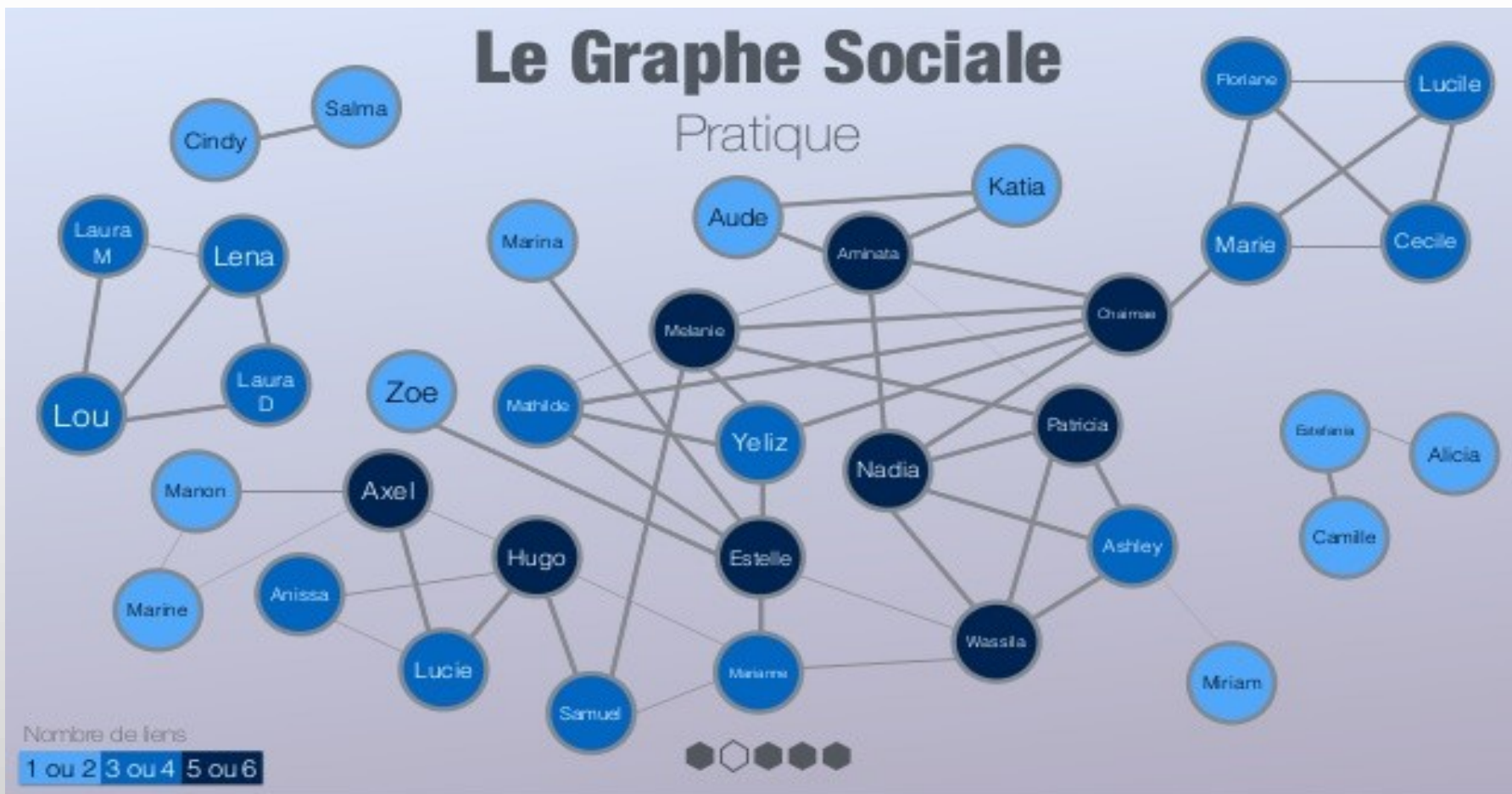
Rank			DBMS	Database Model	Score		
Jun 2021	May 2021	Jun 2020			Jun 2021	May 2021	Jun 2020
1.	1.	1.	Redis +	Key-value, Multi-model ⓘ	165.25	+3.08	+19.61
2.	2.	2.	Amazon DynamoDB +	Multi-model ⓘ	73.76	+3.69	+8.90
3.	3.	3.	Microsoft Azure Cosmos DB +	Multi-model ⓘ	36.47	+1.76	+5.67
4.	4.	4.	Memcached	Key-value	25.18	+0.68	+0.37
5.	5.	↑ 6.	etcd	Key-value	10.22	+0.80	+2.17
6.	6.	↓ 5.	Hazelcast +	Key-value, Multi-model ⓘ	9.37	+0.19	+0.96
7.	7.	↑ 8.	Ehcache	Key-value	7.49	+0.26	+1.21
8.	8.	↓ 7.	Aerospike +	Key-value, Multi-model ⓘ	5.77	+0.86	-0.89
9.	9.	↑ 10.	Riak KV	Key-value	5.40	+0.82	+0.40
10.	10.	↑ 11.	Ignite	Multi-model ⓘ	4.93	+0.54	+0.06
11.	11.	↓ 9.	ArangoDB +	Multi-model ⓘ	4.92	+0.53	-0.47
12.	12.	12.	OrientDB	Multi-model ⓘ	4.45	+0.26	-0.37
13.	13.	13.	Oracle NoSQL	Multi-model ⓘ	4.31	+0.61	+0.09
14.	14.	↑ 17.	RocksDB	Key-value	3.58	+0.49	+0.72
15.	15.	↓ 14.	InterSystems Caché	Multi-model ⓘ	3.24	+0.34	-0.22
16.	↑ 17.	↓ 15.	Oracle Berkeley DB	Multi-model ⓘ	3.08	+0.44	-0.12

Modèle de données NoSQL orienté Graphe



Modèle de données: Graphe

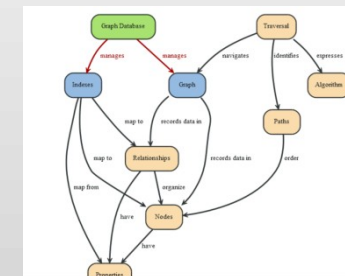
Comment stocker et manager ces données ?





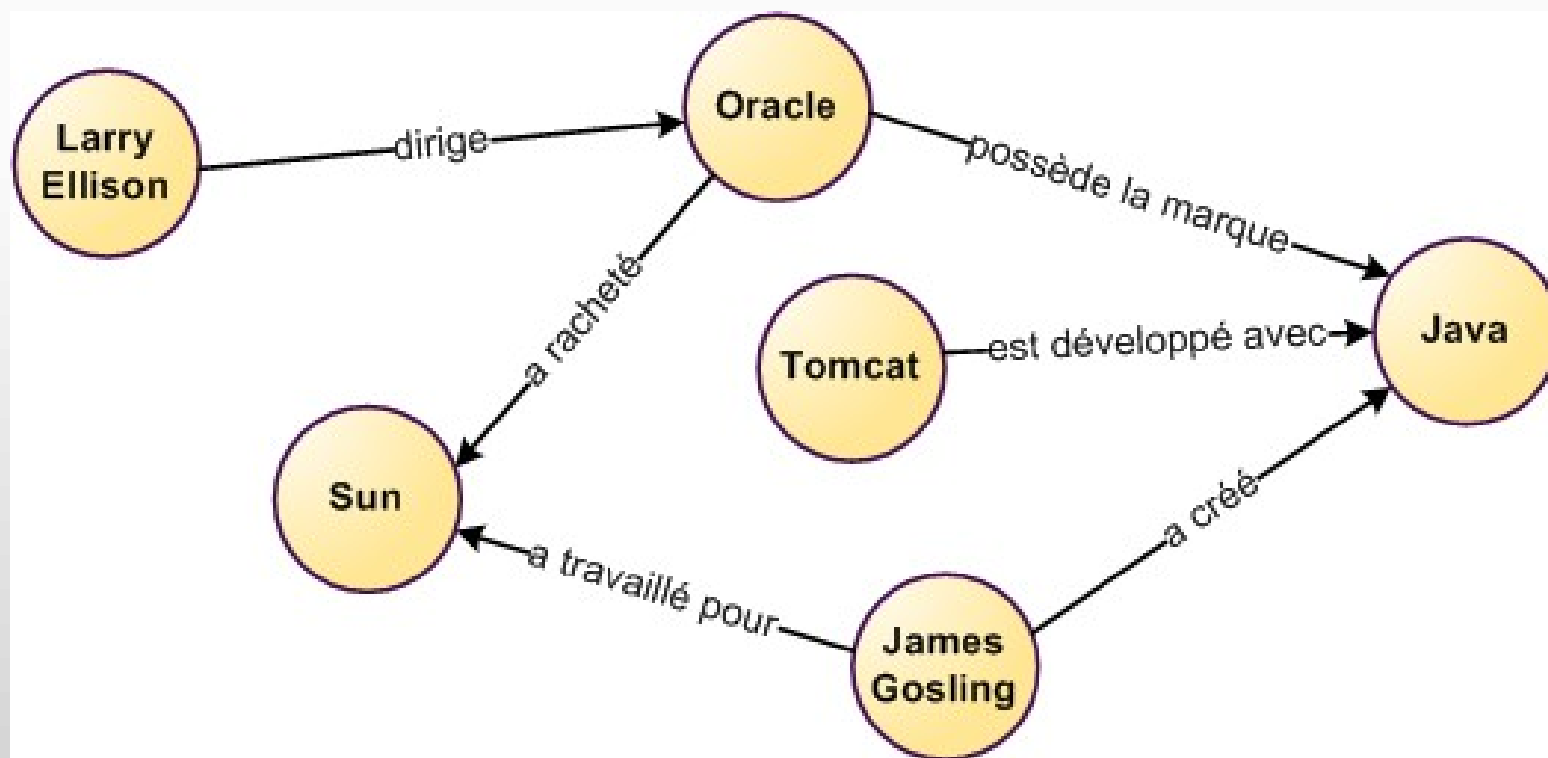
Modèle de données: Graphe

- Analyser des informations interconnectées,
- Comprendre, Evaluer et Exploiter les processus et les connexions présentent entre les données



Modèle de données: Graphe

- Exemple de graphe

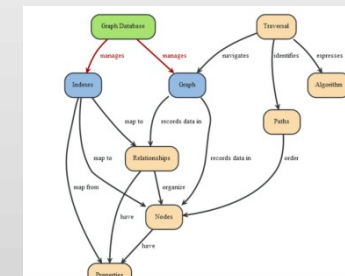




Modèle de données: Graphe

Avantages:

- Opérations de la théorie du graphe
 - requêtes basées sur la gestion de chemins

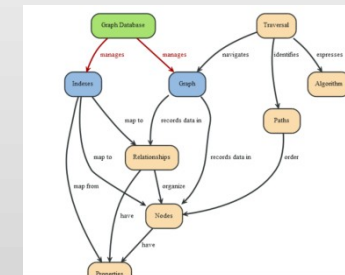




Modèle de données: Graphe

Inconvénients:

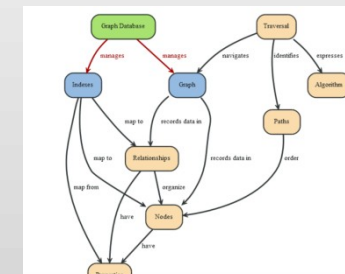
- Complexité de gestion des graphes complexes





Dans quel cas utiliser le modèle graphe?

- Réseaux sociaux
- Routes
- Réseaux électriques
- Web social



Modèle de données: Graphe

Exemple : Représentation de ventes en relationnel

Table Sales

#ticket	#date	#book
1	01/01/16	2212121504
1	01/01/16	2212141556
2	01/01/16	2212141556

Table Book

#isbn	#title	#author
2212121504	Scenari	1
2212141556	NoSQL	2

Table Author

#id	surname	firstname
1		
2	Bruchez	Rudi



Classe Sales

#oid	
4d040766076 6b236450b45 a3	property ticket : 1
	property date : 01/01/16
	relation book : 4d0407660766b236450b45a5
	relation book : 4d0407660766b236450b45a6
4d040766076 6b236450b45 a4	property ticket : 2
	property date : 01/01/16
	relation book : 4d0407660766b236450b45a6

Classe Book

#oid	
4d040766076 6b236450b45 a5	property title : Scenari
4d040766076 6b236450b45 a6	property title : NoSQL
	relation author : 4d0407660766b236450b45a8

Classe Author

#oid	
4d040766076 6b236450b45 a8	property surname : Bruchez
	property firstnam : Rudi



Modèle de données: Graphe

<https://db-engines.com/en/ranking/graph+dbms>



☐ include secondary database models

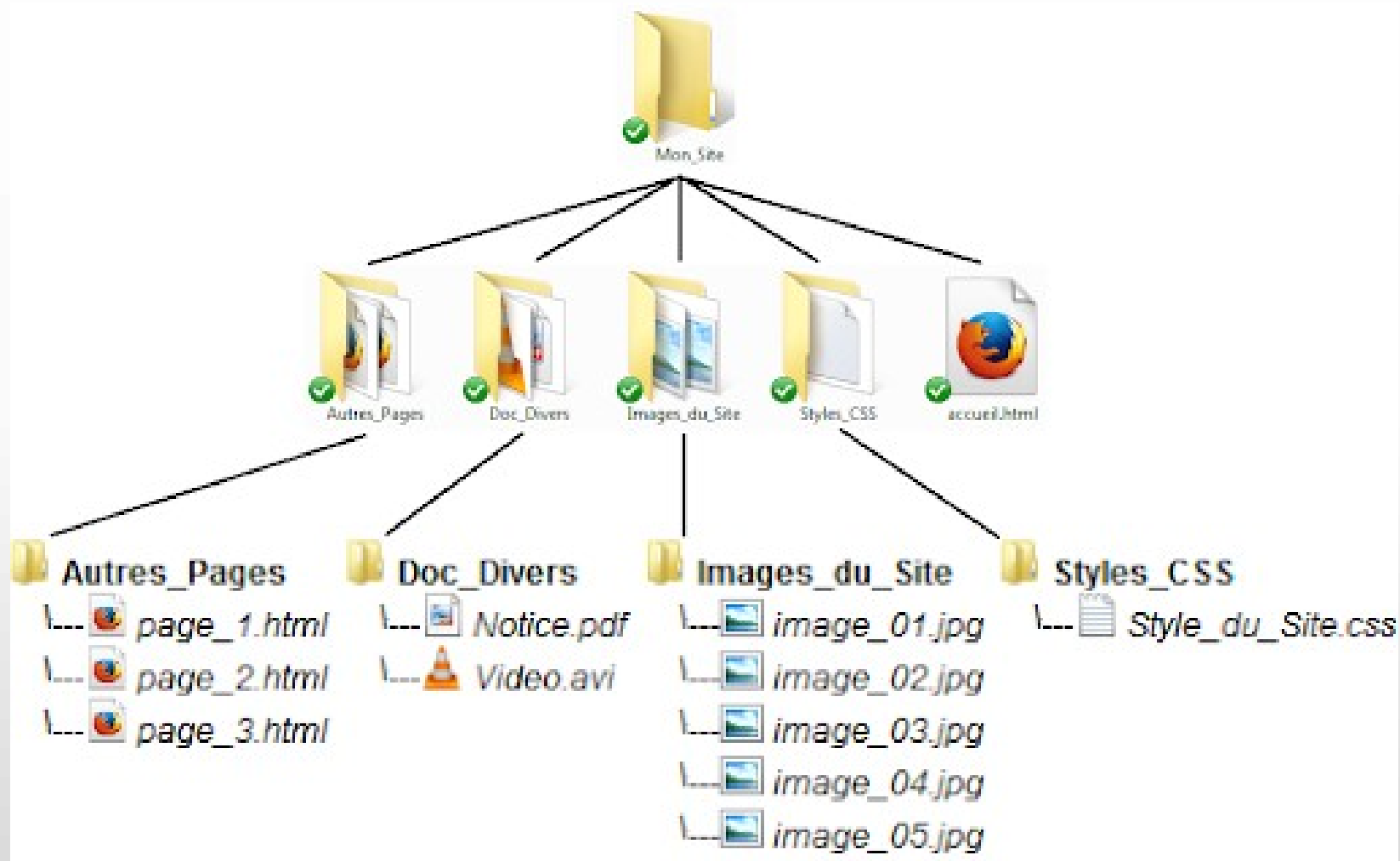
32 systems in ranking, June 2021

Rank			DBMS	Database Model	Score		
Jun 2021	May 2021	Jun 2020			Jun 2021	May 2021	Jun 2020
1.	1.	1.	Neo4j	Graph	55.75	+3.52	+7.48
2.	2.	2.	Microsoft Azure Cosmos DB	Multi-model	36.47	+1.76	+5.67
3.	3.	3.	ArangoDB	Multi-model	4.92	+0.53	-0.47
4.	4.	4.	OrientDB	Multi-model	4.45	+0.26	-0.37
5.	5.	5.	Virtuoso	Multi-model	3.69	+0.25	+1.41
6.	7.	7.	JanusGraph	Graph	2.49	+0.32	+0.48
7.	6.	9.	GraphDB	Multi-model	2.31	+0.06	+1.06
8.	8.	6.	Amazon Neptune	Multi-model	2.20	+0.26	+0.03
9.	9.	13.	TigerGraph	Graph	1.88	+0.14	+0.99
10.	10.	11.	Stardog	Multi-model	1.73	+0.02	+0.57
11.	12.	8.	Dgraph	Graph	1.71	+0.35	+0.31
12.	11.	10.	Fauna	Multi-model	1.67	+0.18	+0.48
13.	13.	14.	AllegroGraph	Multi-model	1.34	+0.01	+0.45
14.	14.	12.	Giraph	Graph	1.20	+0.07	+0.23
15.	15.	21.	Nebula Graph	Graph	1.04	+0.01	+0.79
16.	17.	15.	Blazegraph	Multi-model	0.82	+0.01	+0.15

Modèle de données
NoSQL orienté

DOCUMENT

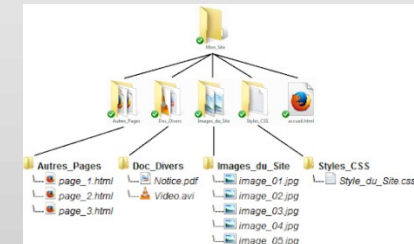
Illustration





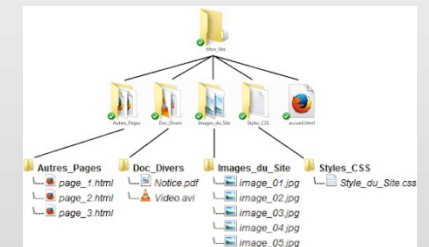
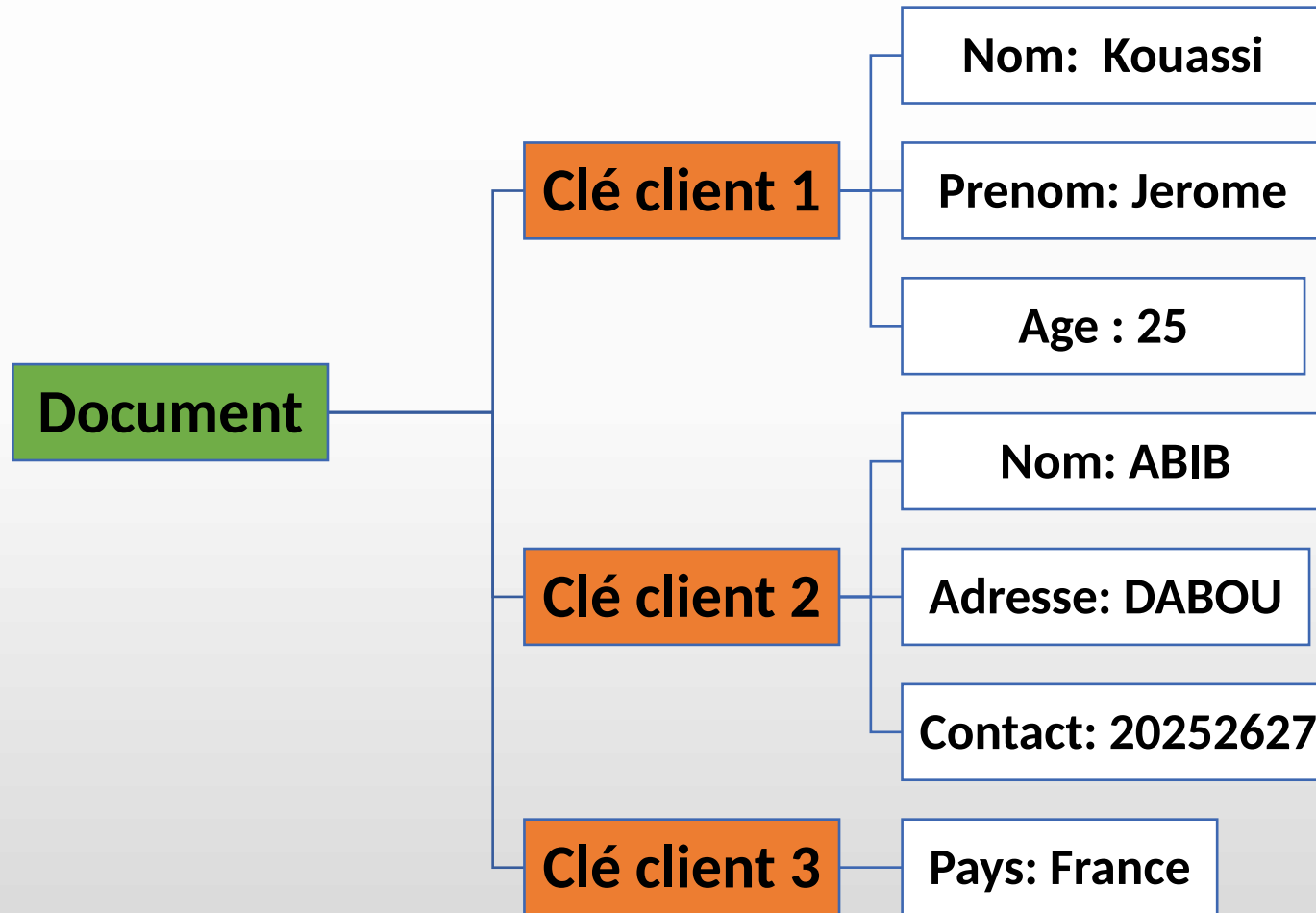
Modèle de données: Document

- Stocke et manipule des données de structures complexes
- Similaire au modèle clé-valeur
- Structure arborescente
- Un champ peut être structuré en sous champs
- Langages d'interrogations riches comme dans les BD classiques
- Index secondaires





Modèle de données: Document



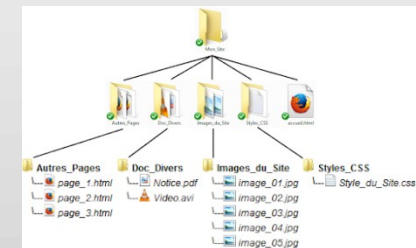


Modèle de données: Document

Les documents sont principalement de type JSON ou XML.

Collection Sales

#oid	
4d0407660766b236450b45a3	<pre>"ticket" : 1 "date" : "01/01/16" "books" : [{ "isbn" : 2212121504 "title" : "Scenari" } { "isbn" : 2212141556 "title" : "NoSQL" "author" : { "surname" : Bruchez "firstname" : Rudi } }]</pre>
4d0407660766b236450b45a4	<pre>"ticket" : 2 "date" : "01/01/16" "books" : [{ "isbn" : 2212141556 "title" : "NoSQL" "author" : { "surname" : Bruchez "firstname" : Rudi } }]</pre>

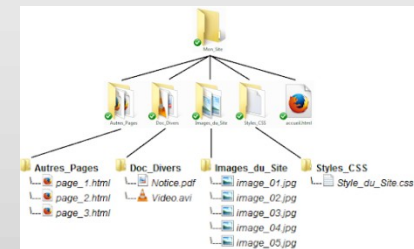




Modèle de données: Document

Avantages:

- faible besoin de jointures
- Lecture rapide si document entier sollicité
- Besoin faible de transactions

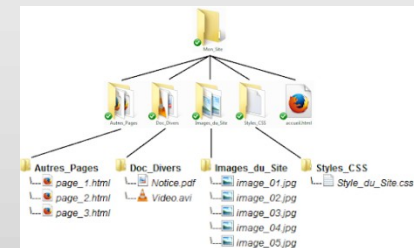




Modèle de données: Document

Inconvénients:

- Imbrication nécessite beaucoup d'itérations pour trouver une information
- Perte d'autonomie des entités





Modèle de données: Document

Dans quel cas utiliser le modèle document?

- Applications manipulant des collection d'informations
 - Bibliothèques numériques, collections de produits
 - Gestions des historiques d'utilisateurs sur les réseaux sociaux
 - Gestion de catalogues



Modèle de données: Document

<https://db-engines.com/en/ranking/document+store>

☐ include secondary database models

51 systems in ranking, June 2021

Rank			DBMS	Database Model	Score		
Jun 2021	May 2021	Jun 2020			Jun 2021	May 2021	Jun 2020
1.	1.	1.	MongoDB	Document, Multi-model	488.22	+7.20	+51.14
2.	2.	2.	Amazon DynamoDB	Multi-model	73.76	+3.69	+8.90
3.	3.	3.	Microsoft Azure Cosmos DB	Multi-model	36.47	+1.76	+5.67
4.	4.	4.	Couchbase	Document, Multi-model	29.07	-1.16	-0.07
5.	5.	6.	Firebase Realtime Database	Document	17.13	+0.22	+3.14
6.	6.	5.	CouchDB	Document, Multi-model	16.12	+0.15	+0.05
7.	7.	8.	Realm	Document	9.68	-0.11	+1.33
8.	8.	7.	MarkLogic	Multi-model	9.38	-0.15	-1.87
9.	9.	9.	Google Cloud Firestore	Document	7.42	+0.46	-0.10
10.	10.	10.	ArangoDB	Multi-model	4.92	+0.53	-0.47
11.	12.	11.	Google Cloud Datastore	Document	4.68	+0.67	-0.47
12.	11.	12.	OrientDB	Multi-model	4.45	+0.26	-0.37
13.	13.	13.	Oracle NoSQL	Multi-model	4.31	+0.61	+0.09
14.	14.	20.	Virtuoso	Multi-model	3.69	+0.25	+1.41
15.	17.	14.	RethinkDB	Document, Multi-model	3.51	+0.42	-0.35
16.	15.	15.	RavenDB	Document, Multi-model	3.50	+0.25	-0.32



Modèle de données NoSQL orienté **COLONNE**



Modèle de données: Colonne

- Stocke et utilise de gros volumes de données avec un schéma rigide
- Structure « hybride »
 - Clé valeur à 2 dimensions
 - Stockage par rowkey identique au modèle document

ColumnFamily			
Key		Value	
AddressBook	SuperColumns	Key	
		Value	
	person1	Column	
		Name	Value
		firstName	John
		lastName	Calagan
	person2	Column	
		Name	Value
		firstName	George
		lastName	Truffe



Modèle de données: Colonne

ColumnFamily		
Key	Value	
AddressBook	SuperColumns	
	Key	Value
	person1	Column
		NameValue
		firstNameJohn
		lastNameCalagan
	person2	Column
		NameValue
		firstNameGeorge
		lastNameTruffe



Modèle de données: Colonne

- Structuration
 - Colonne : couple clé/valeur associé à une version
 - Super colonne: imbrication de colonne
 - **Rowkey** : identifiant unique d'une ensemble de colonne
 - **Keyspace** : ensemble de familles de colonnes



Modèle de données: Colonne

- Exemple :

	A	B	C	D	E
1	foo	bar	hello		
2		Tom			
3			java	scala	cobol

Organisation d'une table dans
une BDD relationnelle

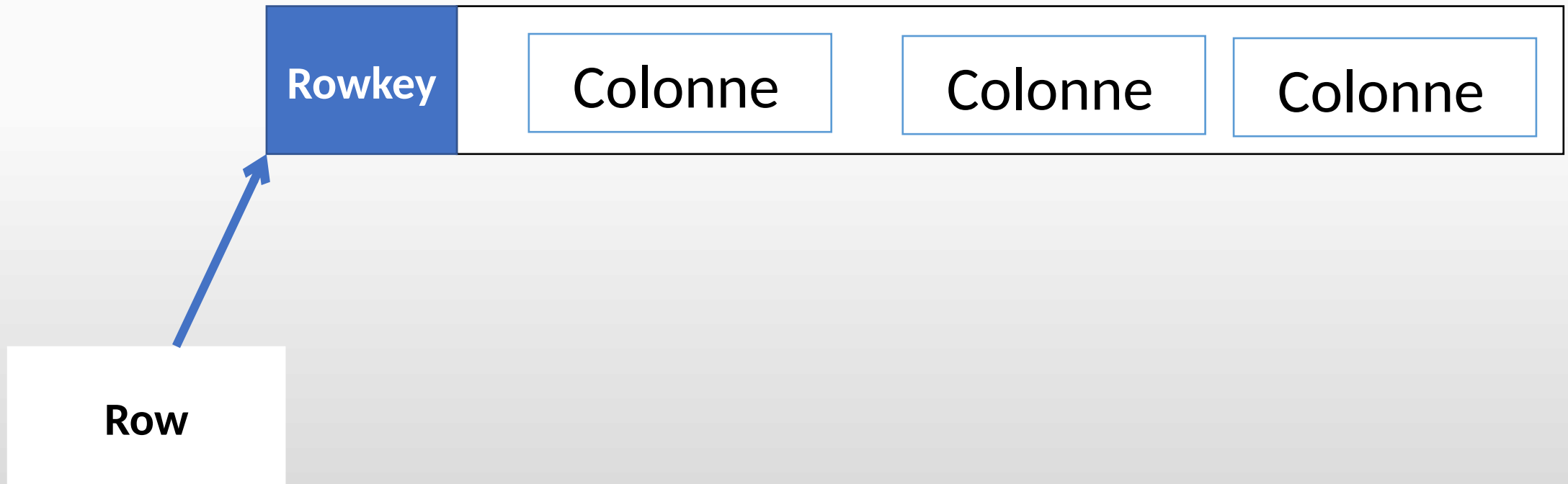
1	A foo B bar C hello				
2	B Tom				
3	C java D scala E cobol				

Organisation d'une table dans
une BDD orientée colonnes



Modèle de données: Colonne

- Exemple :





Modèle de données: Colonne

- Exemple:

Famille de
colonne
(keyspace)



1	A foo B bar C hello
2	B Tom
3	C java D scala E cobol

Organisation d'une table dans
une BDD orientée colonnes



Modèle de données: Colonne

Avantages:

- Colonnes dynamiques
 - valeurs non atomiques et diversité des types
 - valeur complexes structurées
- Facilité de distribution des données et de la gestion de leur historique



Modèle de données: Colonne

Inconvénients:

- Difficulté des opérations de jointures
- Difficulté de définition d'un schéma de qualité



Modèle de données: Colonne

Dans quel cas utiliser le modèle colonne?

- Applications manipulant des collections d'informations
- e-commerce
- gestion catalogue
- Gestion de logs
- Gestion de commande
- Gestion des profils d'utilisateurs
- Google Plus, Google Analytics, Google Finance, etc.



Modèle de données: Colonne

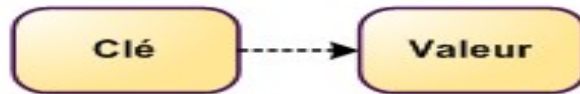
<https://db-engines.com/en/ranking/wide+column+store>

☐ include secondary database models

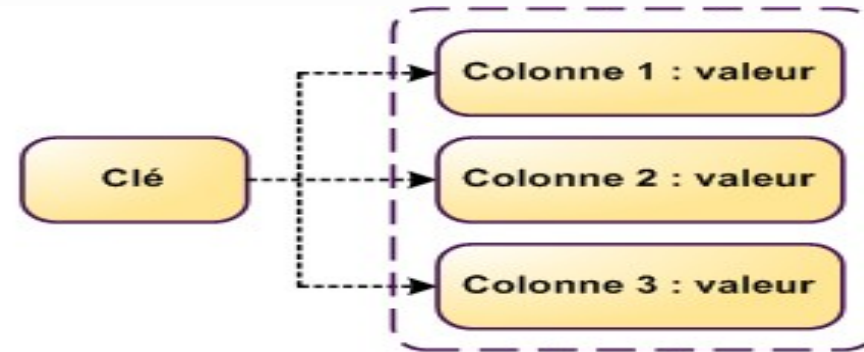
13 systems in ranking, June 2021

Rank			DBMS	Database Model	Score		
Jun 2021	May 2021	Jun 2020			Jun 2021	May 2021	Jun 2020
1.	1.	1.	Cassandra +	Wide column	114.11	+3.18	-4.90
2.	2.	2.	HBase +	Wide column	43.52	+0.27	-5.22
3.	3.	3.	Microsoft Azure Cosmos DB +	Multi-model i	36.47	+1.76	+5.67
4.	4.	4.	Datastax Enterprise +	Wide column, Multi-model i	7.76	+0.21	-0.60
5.	5.	5.	Microsoft Azure Table Storage	Wide column	4.75	+0.46	-0.52
6.	6.	6.	Accumulo	Wide column	3.94	+0.05	+0.08
7.	7.	7.	Google Cloud Bigtable	Wide column	3.58	+0.47	+0.65
8.	8.	8.	ScyllaDB +	Multi-model i	2.71	+0.57	0.00
9.	9.	9.	HPE Ezmeral Data Fabric	Multi-model i	0.90	+0.04	+0.27
10.	10. ↑	11.	Elassandra	Wide column, Multi-model i	0.59	+0.04	+0.32
11.	11.		Amazon Keyspaces	Wide column	0.44	0.00	
12.	12. ↓	10.	Alibaba Cloud Table Store	Wide column	0.38	+0.02	+0.03
13.	13.		SWC-DB	Wide column, Multi-model i	0.00	±0.00	

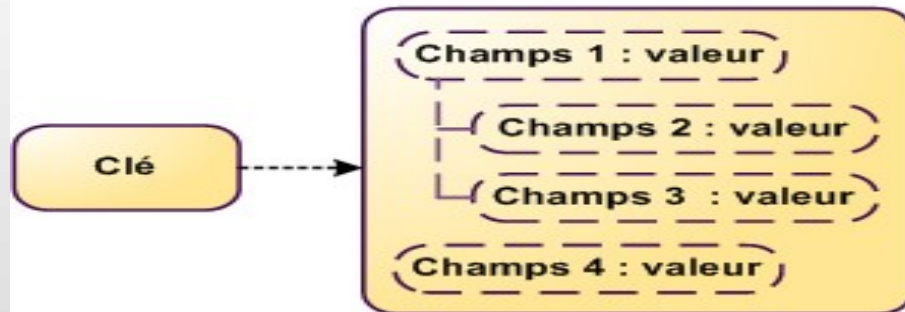
Modèle de données



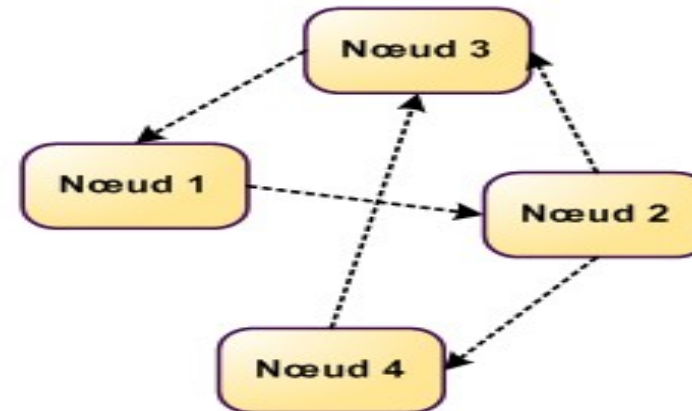
BDD Clé-Valeur



BDD Orientée colonnes



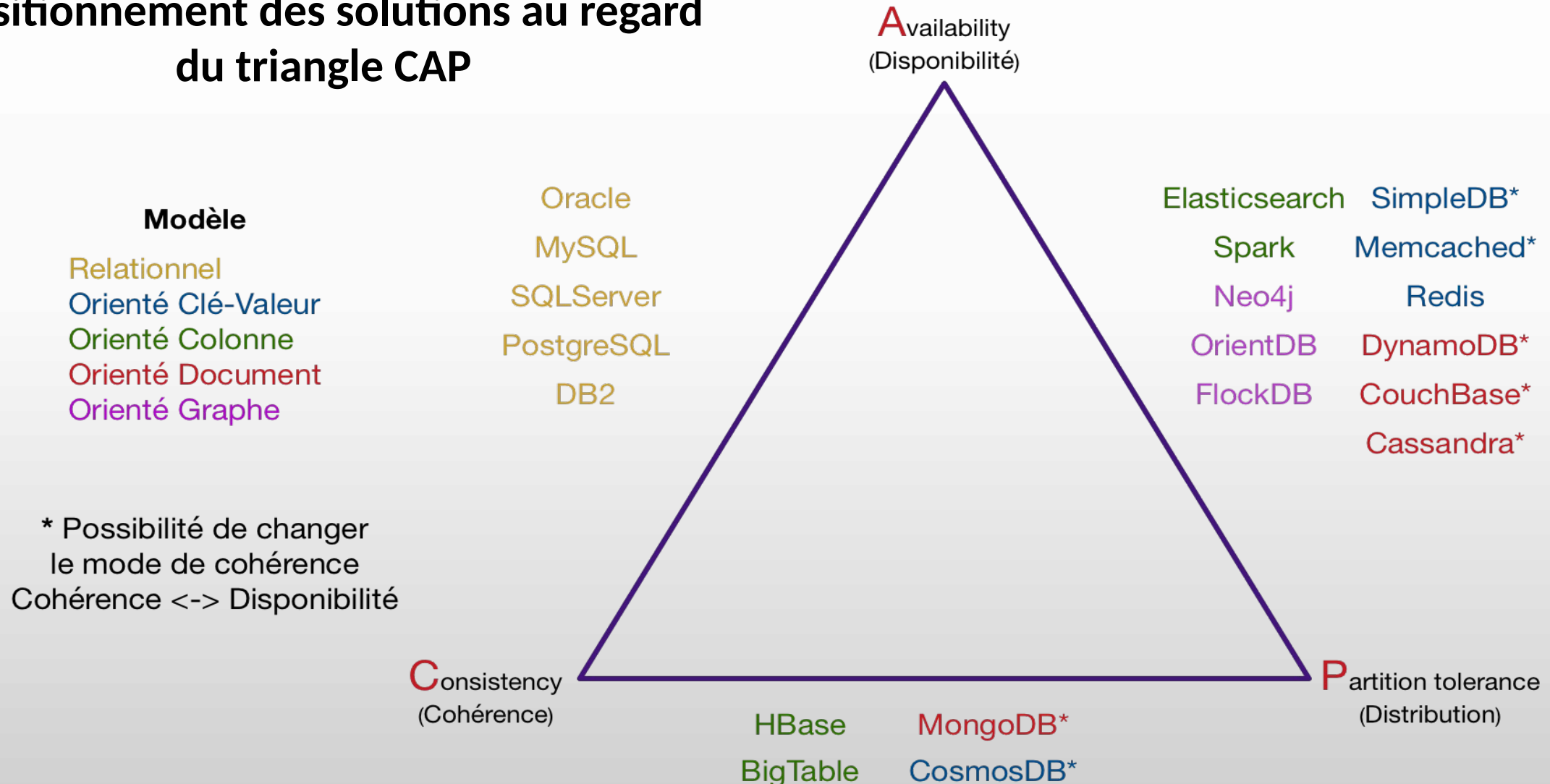
BDD Orientée document



BDD Orientée graphe

Modèle de cohérence: Théorème CAP

Positionnement des solutions au regard du triangle CAP



MERCI