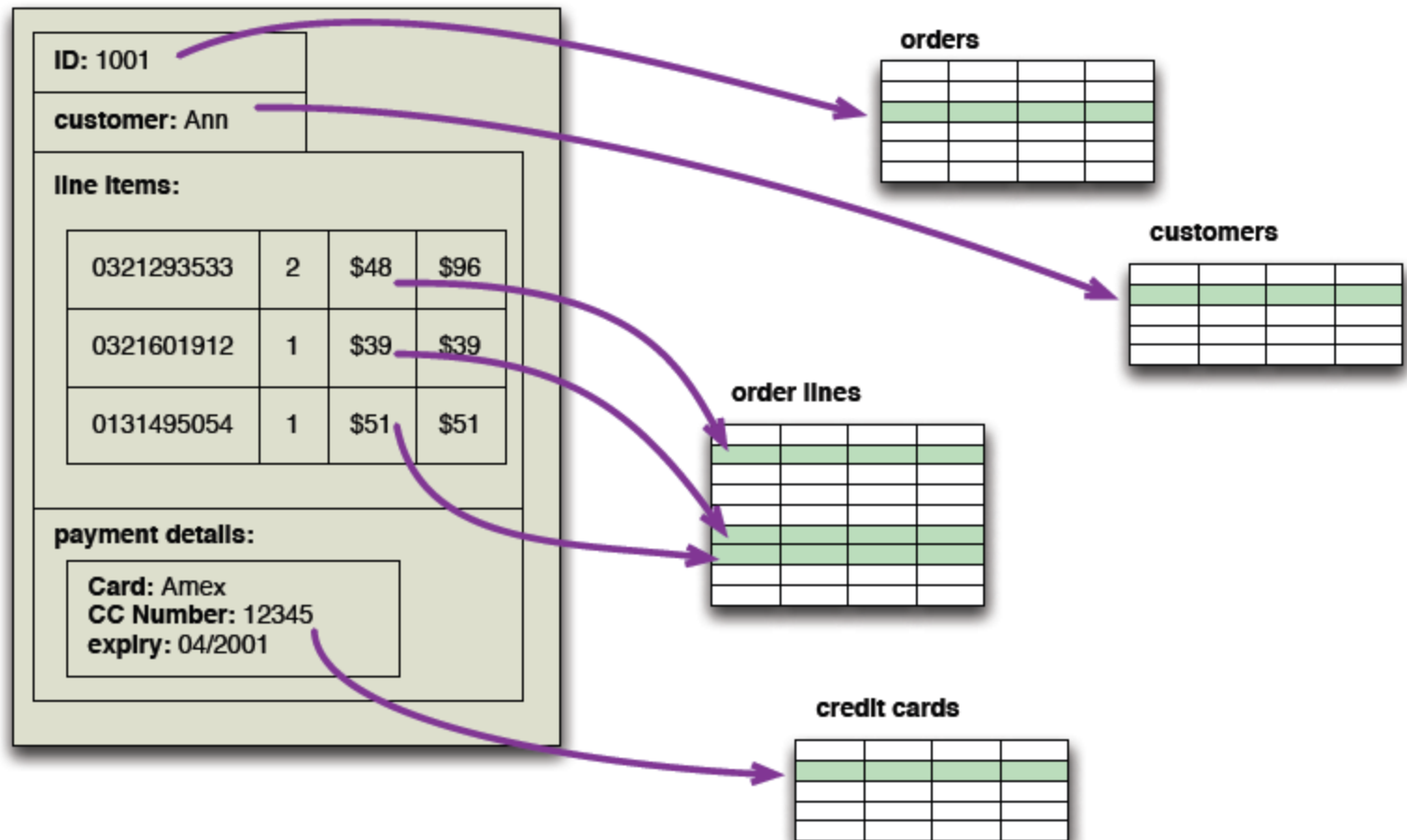




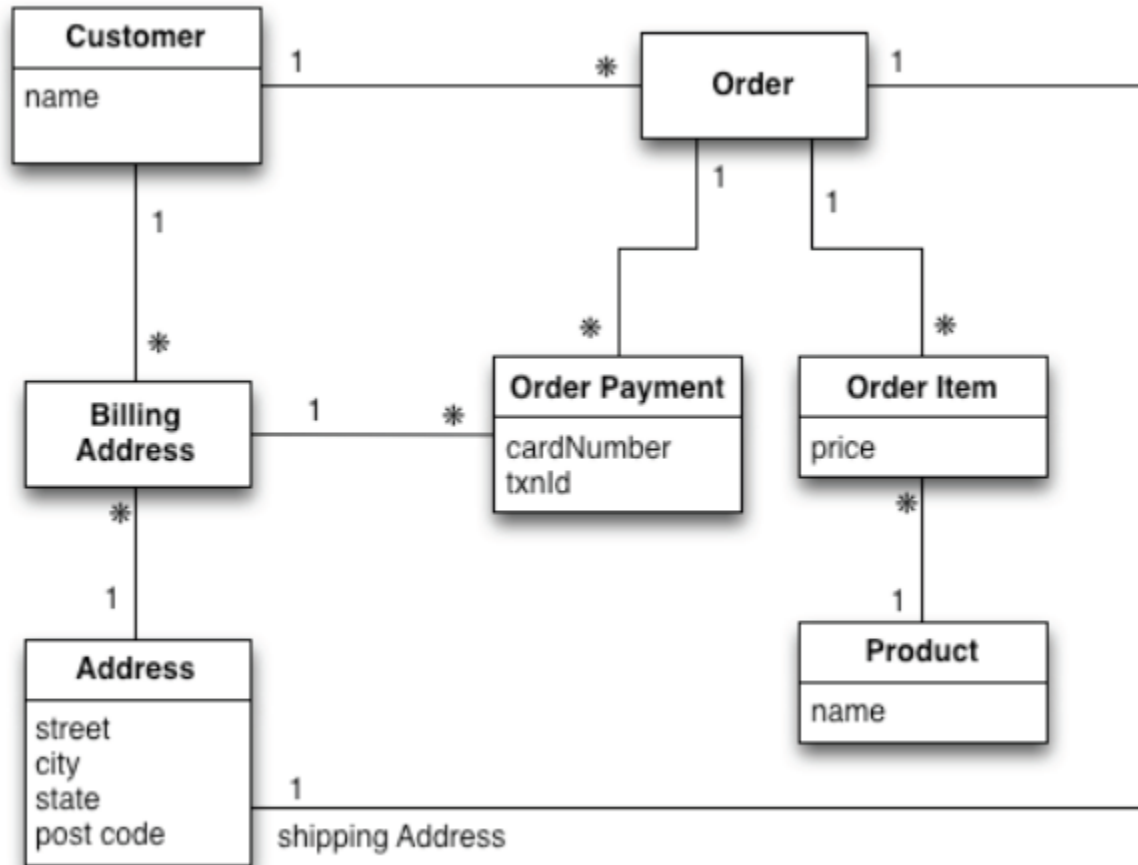
BDD - Introduction

Le monde Relationnel : Impedance mismatch

Difficulté à représenter un objet dans le monde relationnel



Domain driven data models



RDBMS data

Customer	
Id	Name
1	Martin

Product	
Id	Name
27	NoSQL Distilled

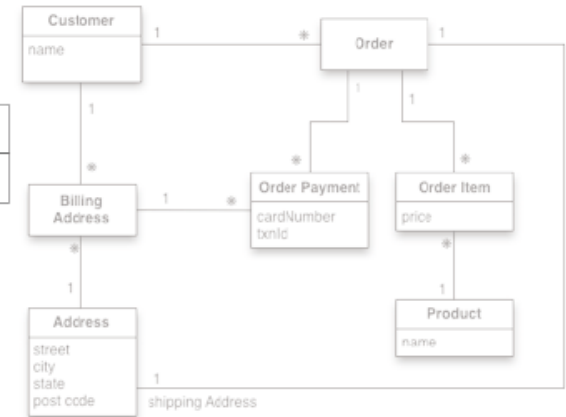
OrderItem			
Id	OrderId	ProductId	Price
100	99	27	32.45

OrderPayment				
Id	OrderId	CardNumber	BillingAddressId	txnId
33	99	1000-1000	55	abelif879rft

Orders		
Id	CustomerId	ShippingAddressId
99	1	77

BillingAddress		
Id	CustomerId	AddressId
55	1	77

Address	
Id	City
77	Chicago



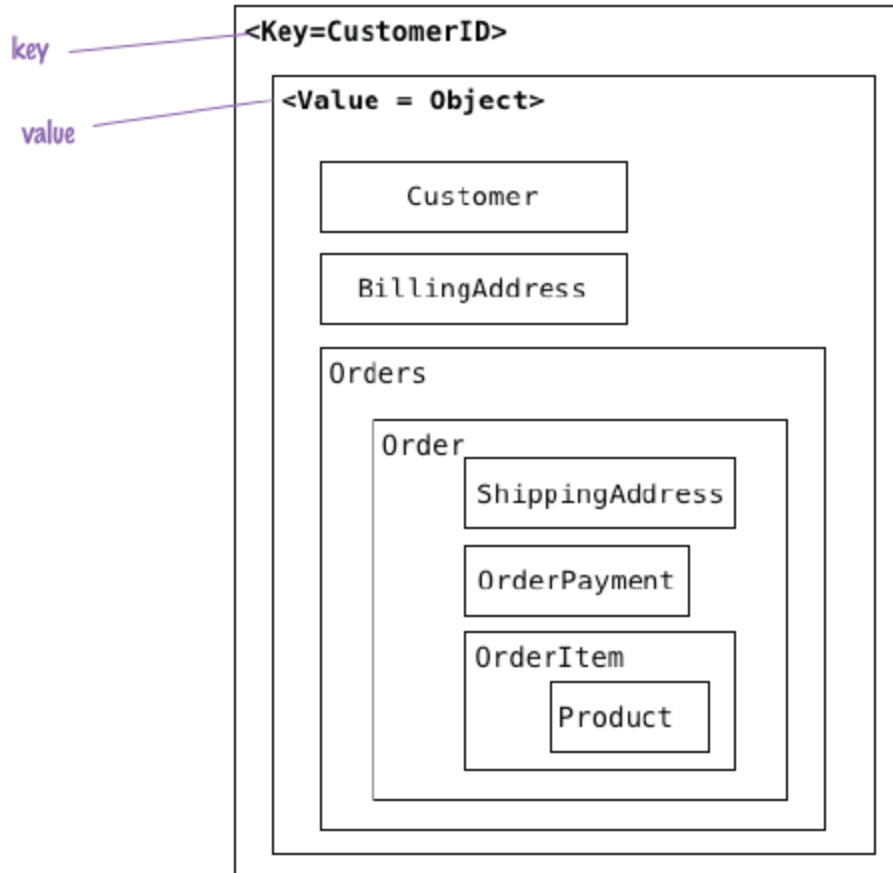


Le monde Relationnel

Caractéristiques:

- Données structurées (Table/Schéma)
- Données Normalisées (Formes Normales)
- Standard (SQL)
- Transactionnel (ACID)
- Requête Complexe (Jointure)
- Des contraintes(Intégrité des données)
- ...

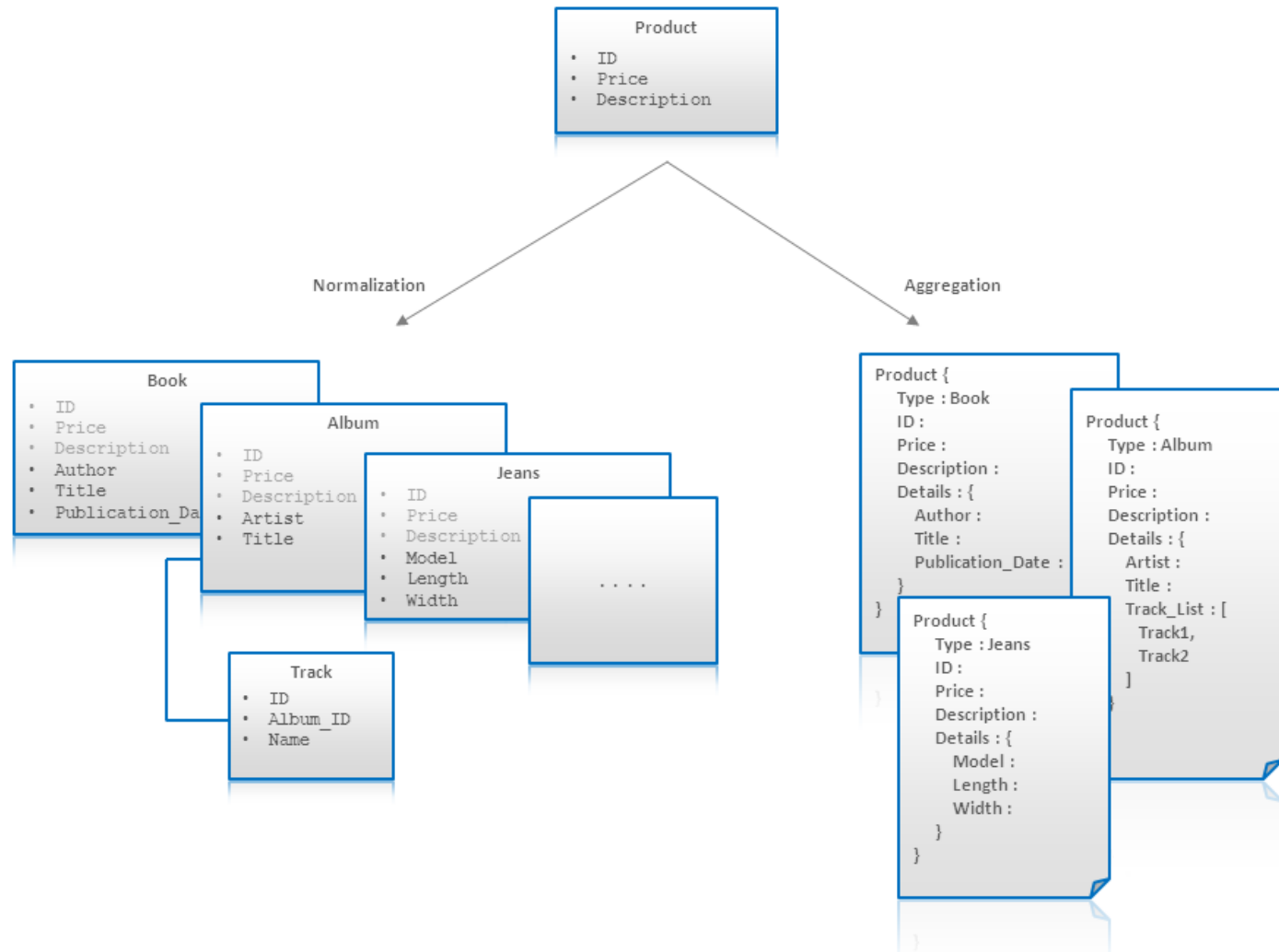
Aggregate model



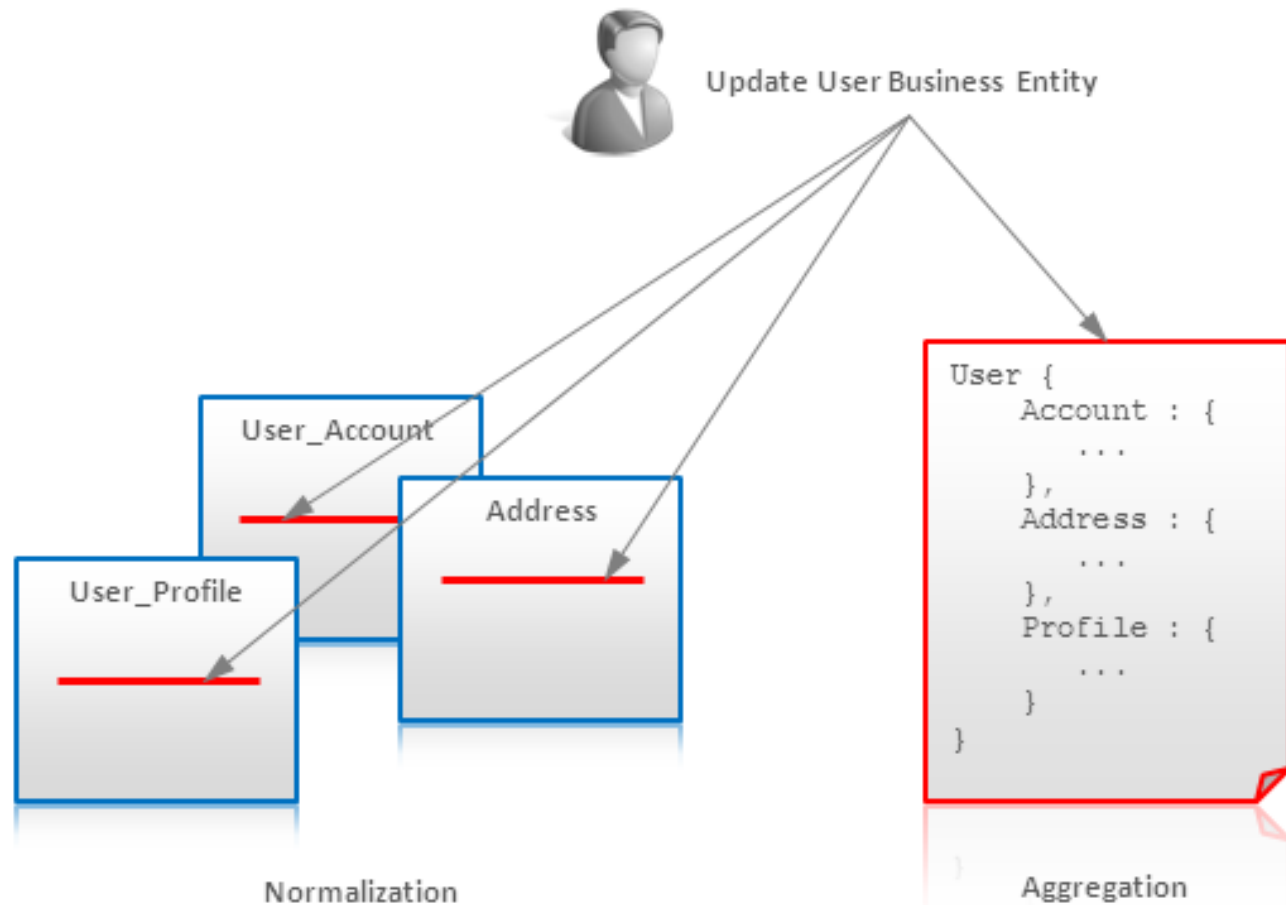
Aggregate Data

```
// in customers
{
  "customer": {
    "id": 1,
    "name": "Martin",
    "billingAddress": [{"city": "Chicago"}],
    "orders": [
      {
        "id": 99,
        "orderItems": [
          {
            "productId": 27,
            "price": 32.45,
            "productName": "NoSQL Distilled"
          }
        ],
        "shippingAddress": [{"city": "Chicago"}],
        "orderPayment": [
          {
            "ccinfo": "1000-1000-1000-1000",
            "txnId": "abelif879rft",
            "billingAddress": {"city": "Chicago"}
          }
        ]
      }
    ]
  }
}
```

Normalisation (Monde Relationnel) vs Agrégation



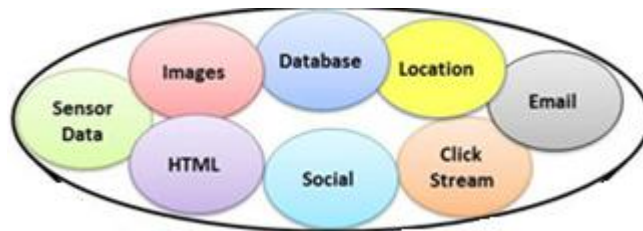
Normalisation vs Agrégation



Le monde aujourd'hui



Volumétrie



Variété de données

Vers le Big Data





Les évolutions...

- **Nouvelles Données :**
 - Web 2.0 : Facebook, Twitter, news, blogs,
 - Flux : capteurs, GPS, ...
- **Nouveaux Traitements :**
 - Moteurs de recherche , Extraction, analyse, ...
 - Recommandation, filtrage collaboratif, ...
- **Nouvelles Infrastructures :**
 - Cluster, réseaux mobiles, microprocesseurs multi-coeurs, ...
- → très gros volumes, données pas ou faiblement structurées
- → transformation, agrégation, indexation
- → distribution, parallélisation, redondance

Big Data

The Google logo, featuring the word "Google" in its characteristic multi-colored font.

Bigtable

The Amazon logo, featuring the text "amazon.com" with a yellow curved arrow underneath the word "amazon".

Dynamo



Le NoSQL

Les bases NoSQL (comprendre Not Only SQL) répondant aux problématiques:

- de hautes disponibilités,
- grandes performances en lecture et/ou écriture
- le traitement de grands volumes de données.



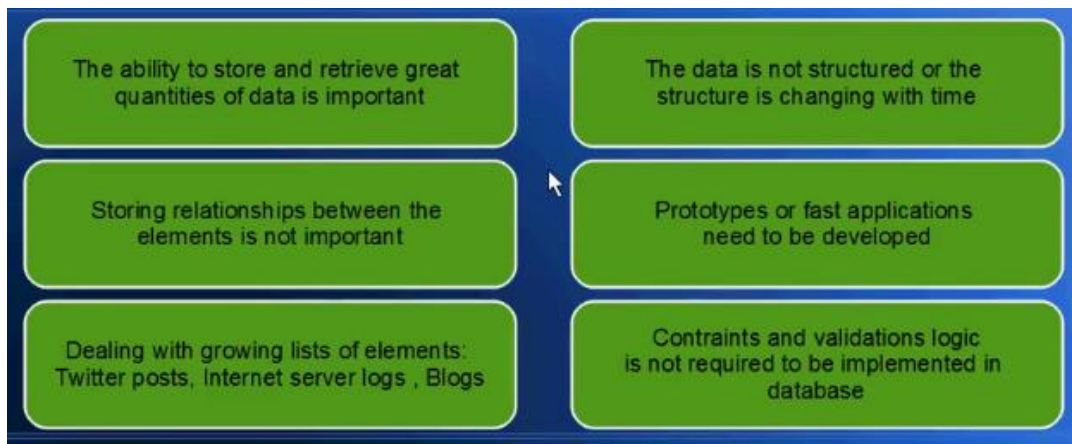
Le NoSQL

Un moteur NoSQL c'est...

- Une structure de données
- Une absence de contraintes
- Une méthode de modélisation
- Un schéma de données... flexible
- Une relation avec un modèle objet
- Du requêtage
- Une absence de transactions et des compromis

Le NoSQL

Alors, SQL versus NoSQL ?



← NoSQL

SQL →





Théorème CAP

Théorème CAP (proposé par Brewer, 2000 et démontré par Gilbert et Lynch 2002)

- Dans un environnement distribué, il n'est pas possible de respecter simultanément :
 - **C** : Cohérence
 - **A** : Disponibilité
 - **P** : Résistance au morcellement
- On peut, en revanche, respecter deux contraintes.



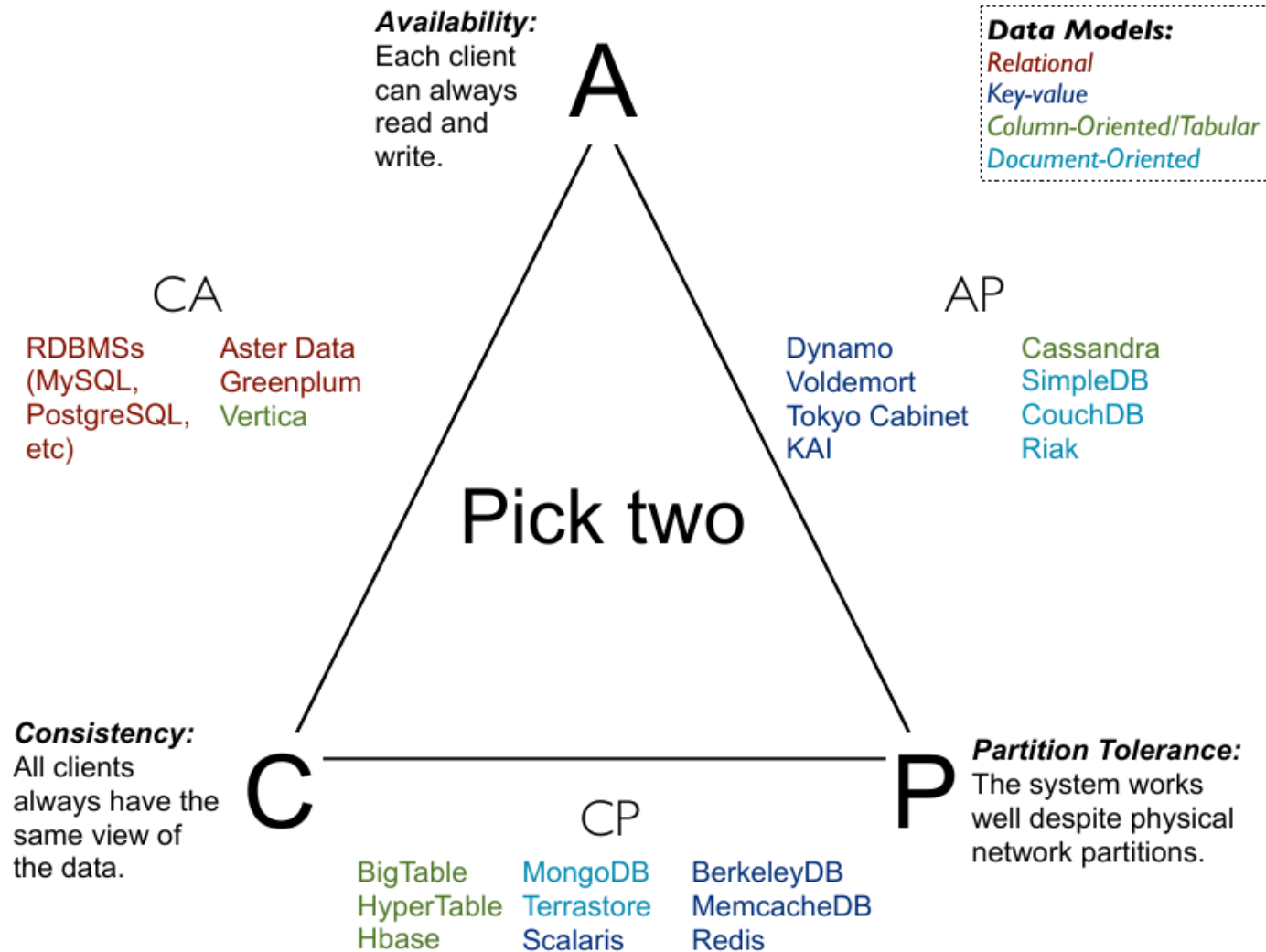
Illustration : CAP

Soit A et B deux utilisateurs du système, soit N1 et N2 deux nœuds du système.


















- **Si** A modifie une valeur sur N1, **alors** pour que B voie cette valeur sur N2 il faut attendre que N1 et N2 soient synchronisés.
- **Si** N1 et N2 doivent toujours servir des valeurs cohérentes, **alors** il y a un temps incompressible entre le début de l'écriture, la synchronisation et la lecture suivante.

Sur un système très chargé et très vaste, ce temps incompressible va considérablement influencer la disponibilité et la résistance au morcellement.

Guide visuel des solutions



Les solutions NoSQL

Document Database	Graph Databases
    Couchbase  MarkLogic  mongoDB	 Neo4j  InfiniteGraph The Distributed Graph Database
Wide Column Stores	Key-Value Databases
 redis  amazon DynamoDB  AEROSPIKE  riak	 accumulo  HYPERTABLE  Cassandra  APACHE HBASE  Amazon SimpleDB



Les solutions NoSQL

- **Clé/valeur.** à la manière des tableaux associatifs des langages de programmation. A une clé correspond une valeur.
- **Orienté colonne** : à une clé correspond un ensemble de colonne, chacune ayant une valeur.
- **Orienté document** : à une clé correspond des ensembles champs/valeurs qui peuvent être hiérarchisés
- **Orienté graphe** : les données sont modélisées sous forme de nœuds qui ont des liaisons entre eux.