

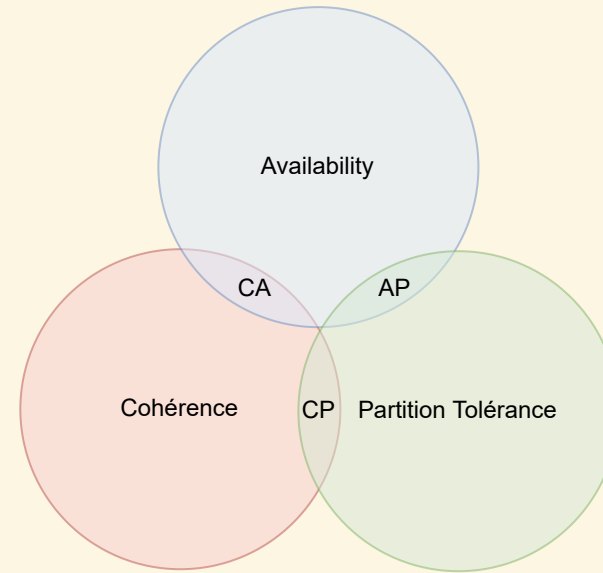
PRINCIPE ET THÉORÈMES

Pour les bases de données.

THÉORÈME DU CAP

*Un système distribué ne peut fournir que deux des trois caractéristiques souhaitées : cohérence (**Consistency**), disponibilité (**Availability**) et tolérance au partitionnement (**Partition tolerance**)*

THÉORÈME DU CAP



COHÉRENCE OU CONSISTENCY

Tous les noeuds du système voient exactement les mêmes données au même moment.

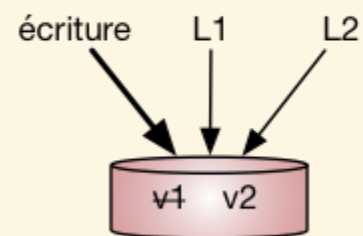
DISPONIBILITÉ OU AVAILABILITY

Garantie que toutes les requêtes reçoivent une réponse

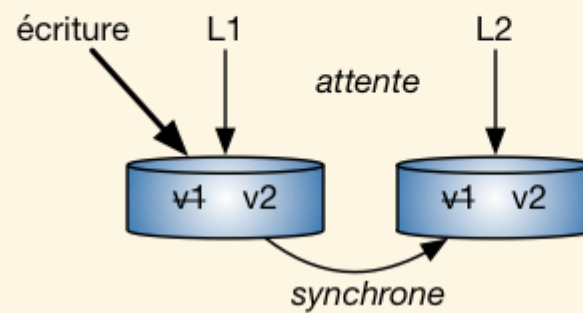
TOLÉRANCE AU PARTITIONNEMENT OU PARTITION TOLERANCE

Aucune panne moins importante qu'une coupure totale du réseau ne doit empêcher le système de répondre correctement.

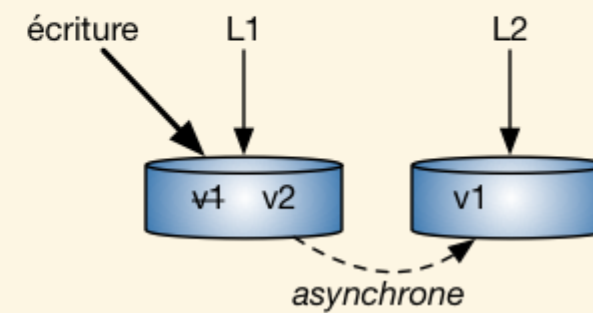
CA
Cohérence + Disponibilité



CP
Cohérence + Distribution



AP
Disponibilité + Distribution



BASE DE DONNÉES CA

Principalement les bases de données relationnelles

- Oracle
- My Sql
- SQLServer
- PostgreSQL
- DB2

BASE DE DONNÉES AP

- Redis
- Spark
- Elasticsearch
- Neo4j

BASE DE DONNÉES CP

- H. Base
- Big Table

BASE DE DONNÉES QUI PEUVENT ÊTRE AP OU CP

Le théorème définit qu'un système ne peut pas être CAP en même temps.

Certaines bases de données peuvent choisir entre AP et CP

BASE DE DONNÉES QUI PEUVENT ÊTRE AP OU CP

- Mongo Db (par défaut CP)
- Cosmos Db (par défaut CP)
- Cassandra (par défaut AP)
- Dynamo Db (par défaut AP)

ACID

Principes pour les transactions

ATOMICITÉ

Une transaction s'effectue entièrement ou pas du tout

COHÉRENCE

Le contenu d'une base doit être cohérent au début et à la fin d'une transaction

ISOLATION

Les modifications d'une transaction ne sont visibles/modifiables que quand celle-ci a été validée

DURABILITÉ

Une fois la transaction validée, l'état de la base est permanent (non affecté par les pannes ou autres)

LES BASES NON ACID

Ces propriétés ne sont pas applicables dans un contexte distribué tel que les bases de données No sql.

BASE

Caractéristiques des bases de données No Sql

BASICALLY AVAILABLE

Quelle que soit la charge de la base de données, le système garantit un taux de disponibilité de la donnée.

SOFT-STATE

La base peut changer lors des mises à jour ou lors d'ajout/suppressions de serveurs.

La base NoSQL n'a pas à être cohérente à tout instant.

EVENTUALLY CONSISTENT

À terme, la base atteindra un état cohérent.

BASE

Les bases de données NoSQL favorisent l'efficacité en sacrifiant certaines contraintes (dont celles ACID).

ACID/BASE

- ACID: Fortes contraintes, mais temps de réponse plus faible.
- BASE: Faibles contraintes, mais meilleur temps de réponse.

LES 3 OU 5 V DU BIGDATA

*Les ensembles de données traités correspondant à la définition du big data répondent à trois caractéristiques principales :
volume, vitesse et variété.*

Définition du bigdata par la [CNIL](#)

LES TROIS PREMIERS V

En 2001, Douglas Laney a défini les 3 premiers V du big Data.

VOLUME

Les systèmes se disant du **BigData** peuvent gérer un grand nombre de données et supporter son accroissement.

VÉLOCITÉ

Les données doivent être collectées, traitées rapidement, voire même en temps réel.

VARIÉTÉ

Les données peuvent prendre des formes très variées et très hétérogènes.

LES DEUX AUTRES VS

Par la suite, deux autres Vs ont pu être identifiés.

VALEUR

Les données peuvent apporter de la valeur.

VÉRACITÉ

Les données doivent être fiables, valides et de qualité.