

Bases de données  
documentaires et distribuées,  
<http://b3d.bdpedia.fr>

Cassandra et la rÃplication

# Réplication

La réplication est fixée au niveau du *keyspace* : 3 par défaut.

On indique également une *replication strategy*.

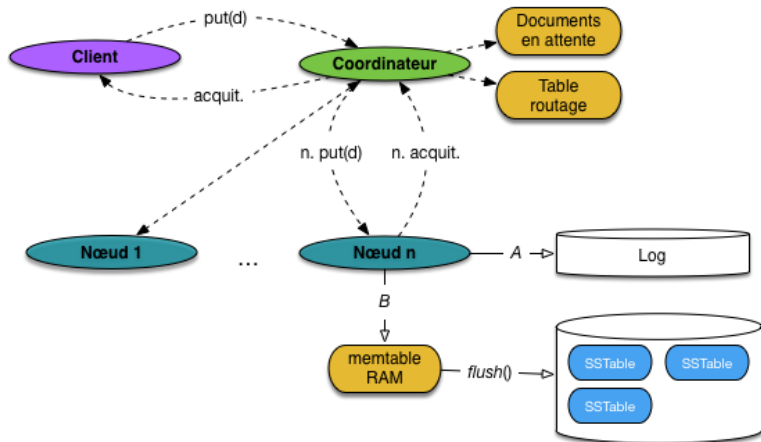
```
CREATE keyspace repli
  with replication = {'class': 'SimpleStrategy',
                    'replication_factor': 3};
```

La stratégie de réplication "simple" copie sur les successeurs du nœud-cible de l'insertion.

Une stratégie plus complexe tient compte de la topologie du réseau (niveaux *racks* et *data centers*).

# Écriture et cohérence des données

Cassandra propose un mécanisme avancé de gestion du compromis entre cohérence et latence.



# Paramétrage de la cohérence en écriture

Le niveau indique le nombre d'acquittements que le coordinateur doit recevoir des nœuds de stockage avant d'acquitter à son tour le client.

**ANY** : on accepte 0 acquittements ! Le document est conservé dans une zone temporaire, en attente d'une nouvelle tentative d'écriture...

**ONE** : Au moins un acquittement

**QUORUM** : Au moins  $\lfloor \text{replication}/2 \rfloor + 1$  acquittements, soit 2 pour un facteur de réplication 3.

**ALL** : La réponse au client sera assurée lorsque la ressource aura été écrite dans tous les réplicas.

Rappel : plus on demande d'acquittements, plus on privilégie la cohérence sur la latence.

Cassandra, système réputé très rapide en insertions.

# Les stratégies en lecture

Même type de paramétrage en lecture. La version avec l'estampille la plus récente est renvoyée au client.

ONE : le coordinateur transmet la première réponse reçue.

QUORUM : Le coordinateur reçoit la réponse de au moins  $\lfloor replication/2 \rfloor + 1$  réplicas, et renvoie au client la ressource avec l'estampille la plus récente.

ALL : Le coordinateur attend d'avoir reçu la réponse de tous les réplicas.

Même remarque : compromis latence/cohérence

# Comment régler les paramètres ?

Soit

**N** : Taux de réplication

**W** : Nb minimum d'écritures devant accuser de réception

**R** : Nb copies d'une donnée à consulter pour une requête

**Alors, si  $W + R > N$ , on assure la cohérence des lectures.**

Intuition : il existe un recouvrement entre les réplicas lus et les derniers réplicas écrits, de sorte qu'au moins une lecture va accéder à la dernière version.

# Illustration

## Exemple

Soit  $N=3$ ,  $W=ALL$ ,  $R=1$  : toute lecture lit un document à jour

Soit  $N=3$ ,  $W=1$ ,  $R=ALL$  : une des lectures lit le document à jour, et c'est celui-ci qui est renvoyé

Soit  $N=3$ ,  $W=QUORUM$ ,  $R=QUORUM$  : une des lectures lit forcément lune des versions à jour, et c'est celui-ci qui est renvoyé

Simple et élégant !