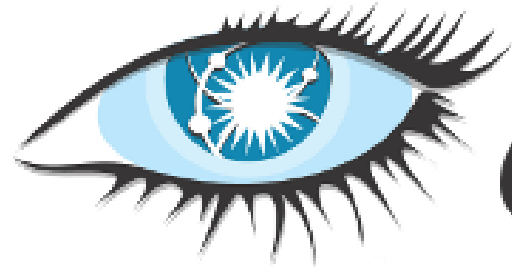




le cnam
Liban



Cassandra

Préparé par Maya Abi Karam

C'est quoi Cassandra?

- C'est une base de données NoSQL - open source
- Orientée colonnes.
- Décentralisée
- Fondée initialement par Facebook.
- En 2008, Facebook transmet ce projet à la fondation Apache

Pourquoi Cassandra?

- Ne nécessite pas un système de fichiers distribué comme HBase.
- Permet de stocker une large quantité de données.
- Très rapide pour manipuler un volume important de données
- Permet la scalabilité horizontale.
- Assure la disponibilité et la tolérance aux pannes.

Qui utilize Cassandra?



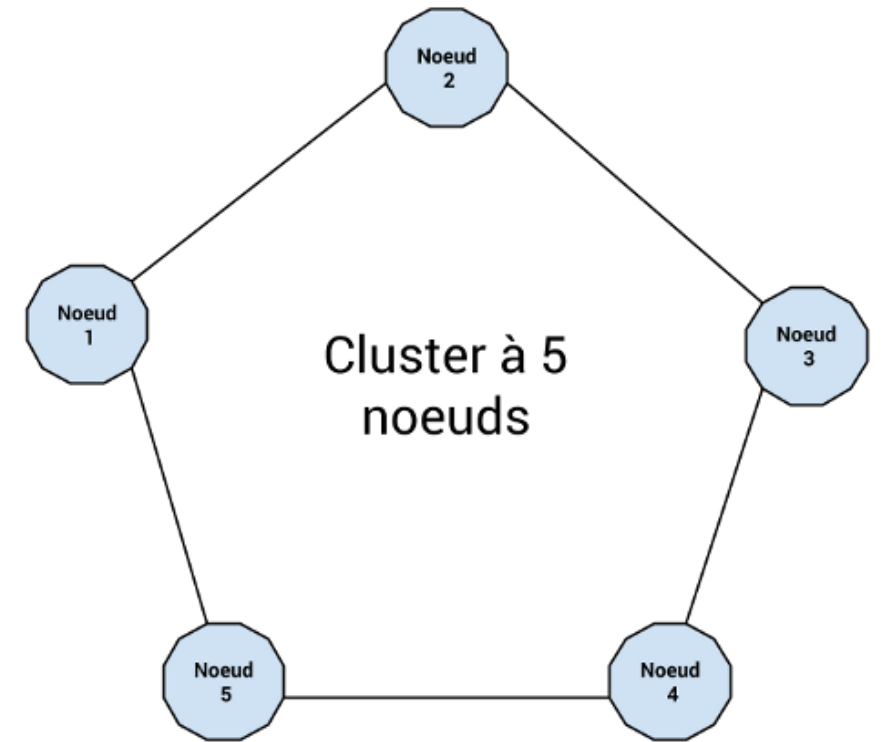


Théorème de CAP

- La cohérence (**C**onsistency)
- **La disponibilité (A**vailability)
- **La tolérance aux pannes (P**artition-tolerance)

Architecture de Cassandra

- Distribuée sur plusieurs serveurs appelés nœuds.
- Les nœuds sont égaux : pas de maître/esclave.
- communiquent ensemble pour la gestion des données.
- Le regroupement de nœuds est appelé un cluster.



Architecture de Cassandra

- Les nœuds communiquent périodiquement ensemble (l'état et la disponibilité de chacun).
- Ce processus s'exécute toutes les secondes.
- Chaque nœud possède un fichier de configuration où ils stockent les informations concernant les autres.

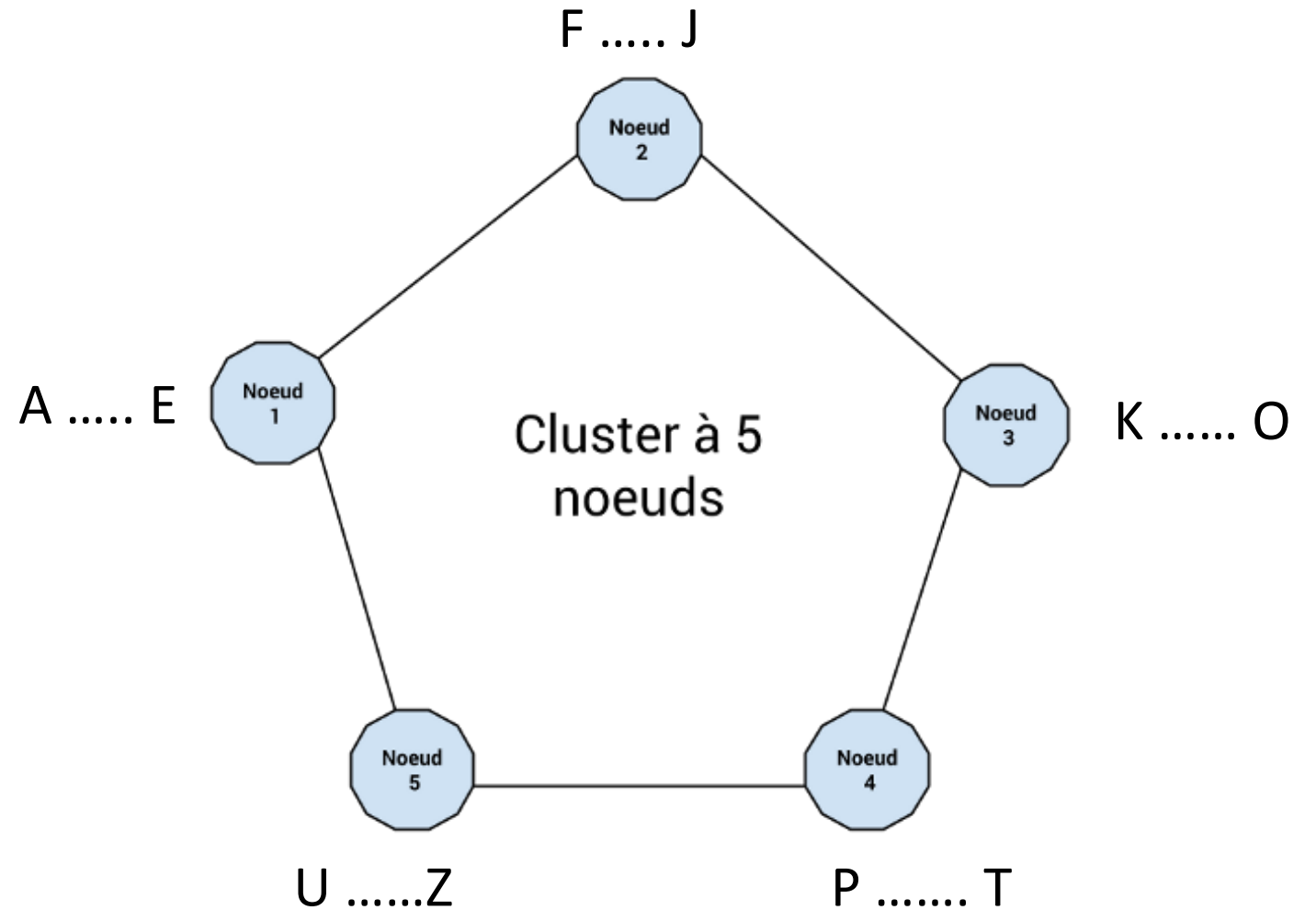
Distribution des données

1. De manière ordonnée
2. De manière aléatoire

Chaque opération de modification ou de lecture → une clé de partition

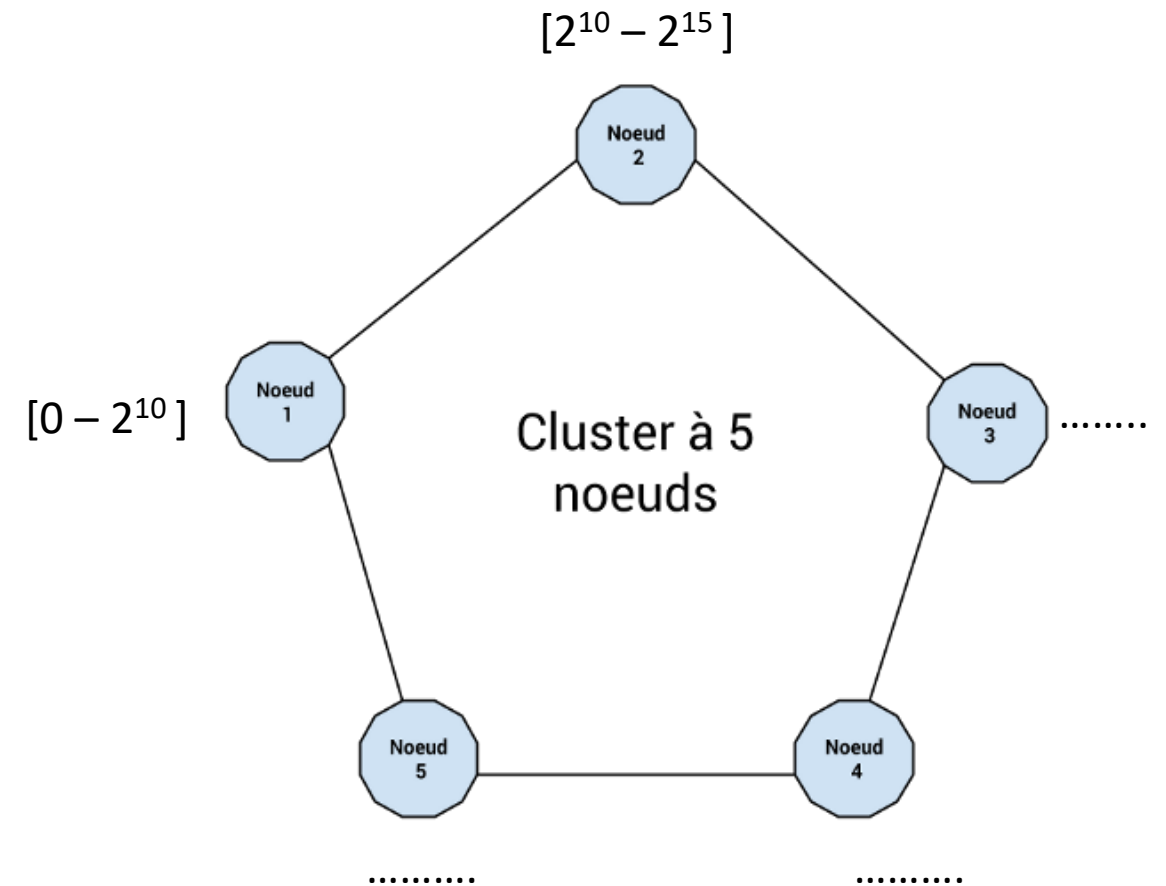
1. Manière ordonnée

- Prendre en charge une plage de clé de partition
- trié par ordre croissant

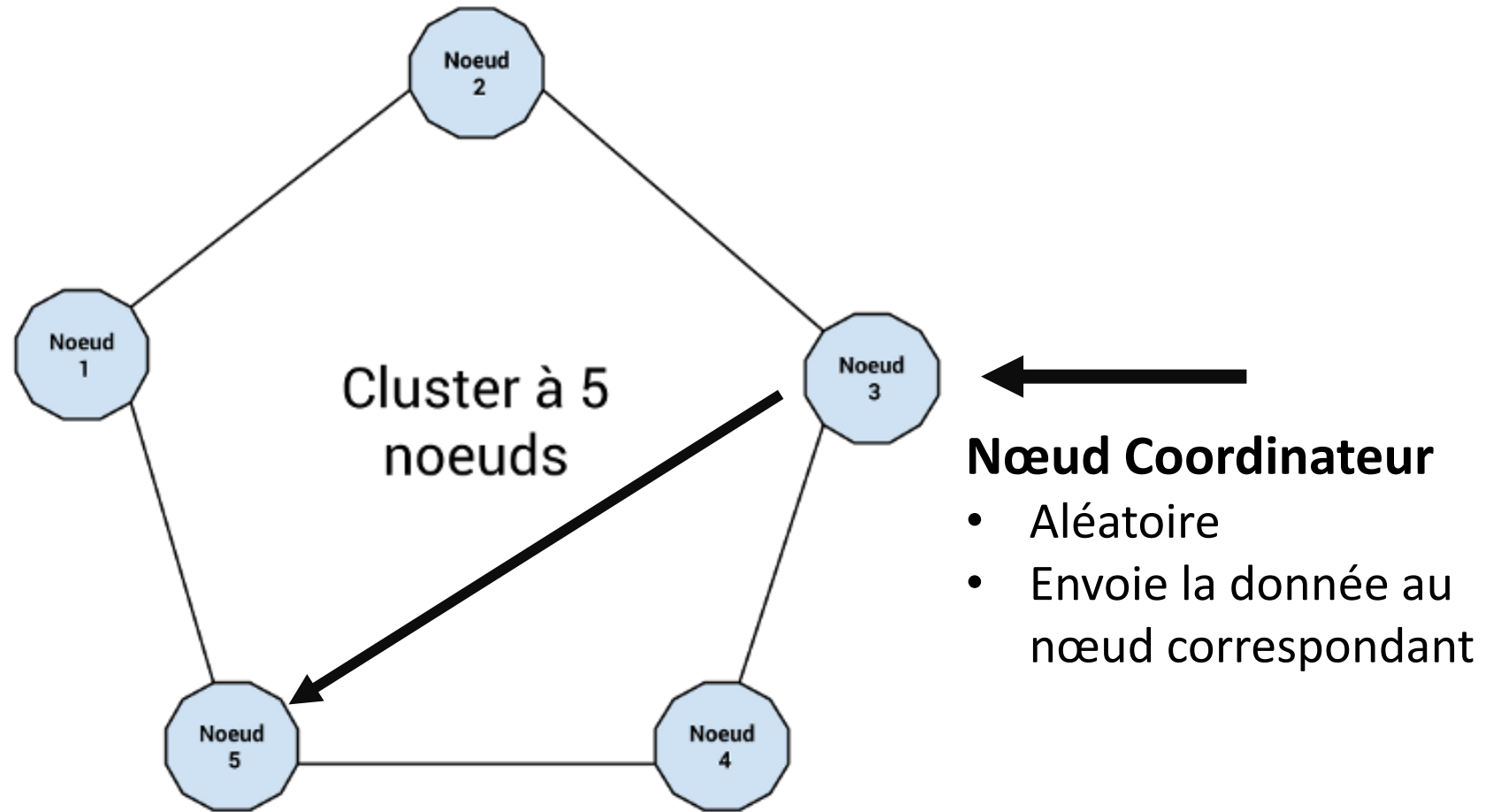


2. Manière aléatoire

- Clé de partition = K
- Fonction de hachage (ex: MD5): $F()$
- $F(K) = \text{token}$
- $\text{MD5} \rightarrow [0 - (2^{127} - 1)]$
- $\text{token} \in [0 - (2^{127} - 1)]$



Distribution des données



Réplication

Tolérance aux pannes \rightarrow Réplication des données sur les nœuds de cluster:

- le nombre réplication (= nombre de nœuds) \rightarrow **facteur de réplication (RF)**
- RF = 1 \rightarrow une seule réplication, RF = 2 \rightarrow 2 réplications

Structure des données

- Colonnes
- Lignes
- Famille de colonnes
- Keyspace

Colonne

- La plus petite unité de modèle
- constituée d'un nom, d'une valeur et d'un timestamp
- Le timestamp → la valeur la plus récente
- Le nom peut atteindre **64 Koctets**
- la valeur **2 Goctets**

Nom
Valeur
<i>Timestamp</i>

Lignes

- Un ensemble de colonnes
- Peut contenir jusqu'à deux milliards de colonnes
- Possède une clé qui peut atteindre aussi **64 Koctets**

baronm	familyName	firstName	age	address
	BARON	Mickael	36	Poitiers
duponto	familyName	firstName	phone	
	DUPONT	Olivier	+335432312	

Famille de colonnes

- Un ensemble de lignes
- Correspond à peu près à une table dans une base de données ordinaire.

Persons				
baronm	familyName	firstName	age	address
	BARON	Mickael	36	Poitiers
duponto	familyName	firstName	phone	
	DUPONT	Olivier	+335432312	

Keyspace

CassandraDEMO				
Persons				
baronm	familyName	firstName	age	address
	BARON	Mickael	36	Poitiers
duponto	familyName	firstName	phone	
	DUPONT	Olivier	+335432312	
...				

Installation

Installer Cassandra

Prérequis:

- Au moins 2 serveurs (=nœuds) Ubuntu
- Configurer chaque serveur et les connecter ensemble
- Chaque serveur doit avoir Cassandra installé



Pour un tutorial sur les autres étapes voir:

<https://www.digitalocean.com/community/tutorials/how-to-run-a-multi-node-cluster-database-with-cassandra-on-ubuntu-14-04>

Installer Cassandra sur un nœud

Etapes:

- Oracle Java SE Runtime Environment (JRE)
- Installer Cassandra

Installer Oracle Java SE Runtime Environment (JRE)

1. Rendre le JRE package disponible

\$ sudo add-apt-repository ppa:webupd8team/java

2. Mise à jour par la commande suivante:

\$ sudo apt-get update

3. Installer JRE avec cette commande:

\$ sudo apt-get install oracle-java8-set-default

4. Une vérification pour savoir que tout est correcte:

\$ java -version

5. Vous devez recevoir un message pareil :

java version "1.8.0_60"

Java(TM) SE Runtime Environment (build 1.8.0_60-b27)

Java HotSpot(TM) 64-Bit Server VM (build 25.60-b23, mixed mode)

Installer Cassandra

1. Rendre disponible le package Cassandra

```
$ echo "deb http://www.apache.org/dist/cassandra/debian 39x main" | sudo tee -a  
/etc/apt/sources.list.d/cassandra.sources.list
```

2. Préciser la source à l'aide de ces 2 commandes

```
$ echo "deb-src http://www.apache.org/dist/cassandra/debian 39x main" | sudo tee -a  
/etc/apt/sources.list.d/cassandra.sources.list
```

Installer Cassandra

3. Ajouter 3 clés publiques :

A. La première, à l'aide de ces 2 commandes :

```
$ gpg --keyserver pgp.mit.edu --recv-keys F758CE318D77295D
```

Puis,

```
$ gpg --export --armor F758CE318D77295D | sudo apt-key add -
```

B. La seconde, à l'aide de ces 2 commandes :

```
$ gpg --keyserver pgp.mit.edu --recv-keys 2B5C1B00
```

Puis,

```
$ gpg --export --armor 2B5C1B00 | sudo apt-key add -
```

C. La troisième, à l'aide de ces 2 commandes:

```
$ gpg --keyserver pgp.mit.edu --recv-keys 0353B12C
```

Puis,

```
$ gpg --export --armor 0353B12C | sudo apt-key add -
```


Installer Cassandra

4. Mettre à jour package avant installation:

\$ sudo apt-get update

5. Installer Cassandra

\$ sudo apt-get install Cassandra

- Pour savoir le statut de Cassandra, si tout va bien et Cassandra est disponible on utilise cette commande:
\$ sudo service cassandra status
- Si Cassandra est disponible on verra le message suivant :

```
● cassandra.service - LSB: distributed storage system for structured data
   Loaded: loaded (/etc/init.d/cassandra)
   Active: active (running) since Mon 2017-03-06 10:05:25 PST; 4min 54s ago
     Docs: man:systemd-sysv-generator(8)
  Process: 1111 ExecStart=/etc/init.d/cassandra start (code=exited, status=0/SUCCESS)
    CGroup: /system.slice/cassandra.service
            └─1728 java -Xloggc:/var/log/cassandra/gc.log -ea -XX:+UseThreadPr...

Mar 06 10:05:25 ubuntu systemd[1]: Starting LSB: distributed storage system.....
Mar 06 10:05:25 ubuntu systemd[1]: Started LSB: distributed storage system ...a.
```

- Pour démarrer Cassandra on utilise cette commande :
\$ sudo service Cassandra start
- Pour l'arrêter :
\$ sudo service Cassandra stop

Pour manipuler les données et utiliser le langage CQL on exécute d'abord la commande suivante:

\$ cqlsh

References

<https://www.infoq.com/fr/articles/modele-stockage-physique-cassandra>

http://blog.xebia.fr/wp-content/uploads/2015/07/Data_Xebia_Cassandra_Programmez.pdf

<http://soat.developpez.com/articles/cassandra/#LI>

<http://mbaron.developpez.com/tutoriels/nosql/cassandra/installation-outils-administration/>

FIN