

BD Orientée Colonnes

APACHE HBASE

Définition

- Apache HBase est une BD open-source, distribuée, versionnée, non relationnelle et orientée colonnes
- les bases de données orientées colonnes sont organisées en familles de colonnes
- les données sont stockées par colonne dans des tables, non par ligne, on peut facilement ajouter des colonnes aux tables.
- c'est un modèle proche d'une table dans un SGBDR mais le nombre de colonnes:
 - a. est dynamique
 - b. peut varier d'un enregistrement à un autre, ce qui évite de retrouver des colonnes ayant des valeurs NULL

HBase VS SGBDR

HBase	SGBDR
N'a pas de schema fixe.definit seulement les familles de colonnes	il a un schema fixe qui decrit la structure de la table
Fonctionne bien avec les données structurées et semi-structurées	fonctionne bien avec les données structureés
il peut avoir des données de-normalisées(valeurs manquantes)	Stock seulement les données normalisées
Concu pour les tables larges pouvant evoluer horizontalement	concu pour des «petites» tables difficiles a evoluer

Modele de données

le modele se base sur six concepts,qui sont :

- 1.Table :dans HBase les données sont organisées dans des tables.
- 2.Row: Une table HBase est un ensemble de lignes.une ligne est identifiée par une clé unique(RowKey)
- 3.Column Family:
 - Les données au sein d'une ligne sont regroupées par column family.
 - Les familles de colonnes sont definies a la creation de la table dans HBase.
- 4.Column qualifier:
 - L'accès aux données au sein d'une column family se fait via le column qualifier .
 - Ce dernier n'est pas specifié a la creation de la table mais plutot a l'insertion de la donnée.

The diagram illustrates an HBase table structure. It features a table with columns grouped into two column families: 'SOCIAL NETWORK' and 'ACTIVITY'. The first column, 'ROW KEY', is used for row identification. Annotations with arrows point to these components: 'Row key' points to the 'ROW KEY' column, 'Column family' points to the 'SOCIAL NETWORK' and 'ACTIVITY' headers, and 'Column qualifier' points to the 'TEXT' column within the 'ACTIVITY' family.

ROW KEY	SOCIAL NETWORK			ACTIVITY		
	SN NAME	USER ID	PAGE ADDRESS	TYPE	DATE	TEXT
Boboo% a_banana	Boboo	a_banana	boboo.com/ a_peach	Like	6/3/2018	
Boboo% a_pineapple	Boboo	a_pineapple	boboo.com/ a_banana	Comment	6/13/2018	Nice pic, bro!

4.cell :

- La combinaison du RowKey,de la Column Family ainsi que la Column qualifier identifié d'une manière unique une cellule
- les données stockées dans une cellule sont appelées les valeurs de cette cellule.les valeurs n'ont pas de type,ils sont toujours considérés comme tableaux d'octet

5.version:

- Les valeurs au sein d'une cellule ont plusieurs versions.
- Les versions sont identifiées par leur timestamp.
- par défaut ,le nombre de versions est egale a trois.

The diagram illustrates a table structure with the following components:

- Row key:** Points to the first column, 'ROW KEY'.
- Column family:** Points to the two main sections, 'SOCIAL NETWORK' and 'ACTIVITY'.
- Column qualifier:** Points to the specific columns within the 'ACTIVITY' family: 'TYPE', 'DATE', and 'TEXT'.
- Cell:** Points to a specific data entry in the 'TEXT' column.

ROW KEY	SOCIAL NETWORK			ACTIVITY		
	SN NAME	USER ID	PAGE ADDRESS	TYPE	DATE	TEXT
Boboo% a_banana	Boboo	a_banana	boboo.com/ a_peach	Like	6/3/2018	
Boboo% a_pineapple	Boboo	a_pineapple	boboo.com/ a_banana	Comment	6/13/2018	Nice pic, bro!
Boboo% a_watermelon	Boboo	a_watermelon	boboo.com/ a_pineapple	Comment	6/10/2018	Hey, that's my jacket! I've been looking all over!
Chiching% a_cucumber	Chiching	a_cucumber	chiching.com/ a_kohlrabi	Comment	5/25/2018	Wow! What a bike!
Chiching% a_kohlrabi	Chiching	a_kohlrabi	chiching.com/ a_kohlrabi	Comment	5/25/2018	Yeah, I know :D