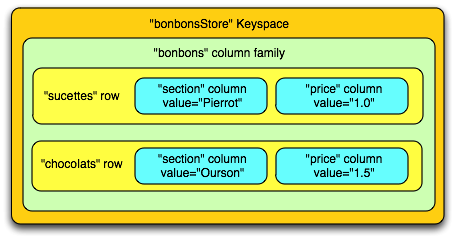
**Modèle des données**

Apache Cassandra est un système permettant de gérer une grande quantité de données de manière distribuée. Ces dernières peuvent être structurées, semi-structurées ou pas structurées du tout. Cassandra a été conçue pour être hautement scalable sur un grand nombre de serveurs tout en ne présentant pas de *Single Point Of Failure (SPOF)*. Elle a fournit un schéma de données dynamique afin d'offrir un maximum de flexibilité et de performance.

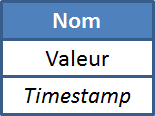
Mais pour bien comprendre cet outil, il faut tout d'abord bien assimiler le vocabulaire de base.

* **Keyspace** : c'est l'équivalent d'une database dans le monde des bases de données relationnelles. À noter qu'il est possible d'avoir plusieurs « Keyspaces » sur un même serveur.
* **Colonne (Column)**: une colonne est composée d'un nom, d'une valeur et d'un timestamp (instant de création).
* **Ligne (Row)**: les colonnes sont regroupées en Rows. Une Row est représentée par une clé et une valeur. Il existe deux types de rows : les « wide row » permettant de stocker énormément de données, avec beaucoup de colonnes et les « skinny row » permettant de stocker peu de données.
* **Une famille de colonnes (Column family**) : c'est l'objet principal de données et peut être assimilé à une table dans le monde des bases de données relationnelles. Toutes les rows sont regroupées dans les column family.

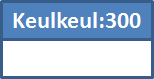


**Colonne**

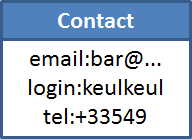
Une colonne est la plus petite unité du modèle de données de Cassandra. C'est un triplet contenant un nom, une valeur et un timestamp. Ce dernier sert à déterminer la mise à jour la plus récente. La taille du nom peut être contenue jusqu'à 64 KO. La valeur quant à elle peut contenir 2 GO de données.



La valeur n'est pas obligatoire. Son omission peut conduire à une amélioration de performance. Le nom de la colonne peut être considéré comme une valeur. Par exemple sur la figure ci-dessous, le nom de la colonne contient le nom de l'utilisateur et un score

  
Colonne sans valeur

La case valeur peut contenir plusieurs valeurs. C'est le cas par exemple d'une collection de chaînes de caractères.

  
Colonne avec plusieurs entrées dans la case valeur

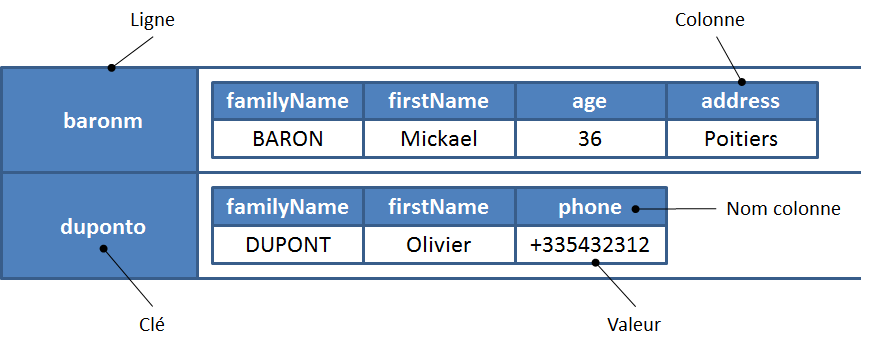
Une colonne a un type appelé *comparator*. Une valeur a également un type appelé *validator*. Nous listons sur le tableau ci-dessous les différents types proposés. Nous montrons deux colonnes. Une première relative au type natif utilisé en interne et une seconde plus propice à une utilisation pour le langage CQL (Cassandra Query Language).

|  |  |
| --- | --- |
| Type Interne | Nom CQL |
| BytesType | blob |
| AsciiType | ascii |
| UTF8Type | text, varchar |
| IntegerType | varint |
| Int32Type | int |
| LongType | bigint |
| UUIDType | uuid |
| TimeUUIDType | timeuuid |
| DateType | timestamp |
| BooleanType | boolean |
| FloatType | float |
| DoubleType | double |
| DecimalType | decimal |
| CounterColumnType | counter |

**Ligne**

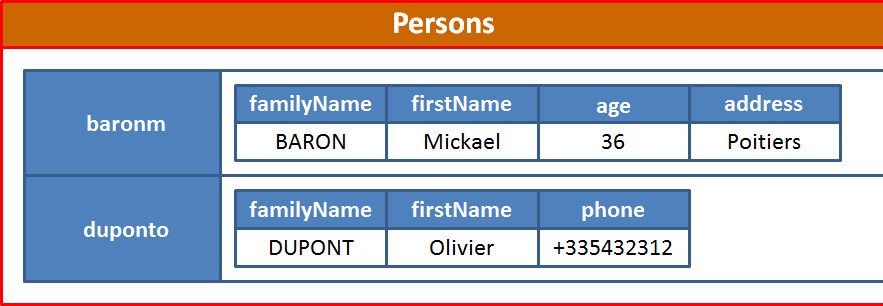
Une ligne est composée d'un ensemble de colonnes. Elle est identifiée par une clé. Une clé peut contenir jusqu'à 64 Ko de données et peut contenir jusqu'à deux milliards de colonnes. Il est possible d'utiliser des colonnes comme clé primaire.

Nous montrons sur la figure ci-dessous un exemple de deux lignes. Notez 1) que le nombre de colonnes n'est pas identique et 2) que les colonnes ne sont pas forcément les mêmes.

[](http://mbaron.developpez.com/tutoriels/nosql/cassandra/installation-outils-administration/images/ligne.png)  
Description d'une ligne

**Famille de colonnes (Column Family)**

Une famille de colonnes ou *column family* en anglais est un regroupement logique de lignes. Pour faire le parallèle avec le monde des bases de données relationnelles, une famille de colonnes est en quelque sorte une table.

[](http://mbaron.developpez.com/tutoriels/nosql/cassandra/installation-outils-administration/images/famillecolonnes.png)  
Description d'une famille de colonnes

Lors de la définition d'une famille de colonnes, vous pouvez y ajouter des informations concernant les métadonnées des colonnes. Ces informations peuvent être utilisées pour renseigner le nom et le type des colonnes. Toutefois c'est au moment de l'ajout d'une ligne que vous choisirez quelles sont les colonnes à exploiter.

C'est ainsi que Cassandra distingue deux types de famille de colonnes :

* statique : les colonnes sont définies lors de la création ou modification de la famille de colonnes ;
* dynamique : les colonnes sont définies lors de la création ou modification d'une ligne.