Types of NoSQL Databases

Column-Oriented Databases

Column-oriented Databases

- A **column-oriented DBMS** is a database management system that stores data tables as sections of columns of data
- Useful if
  - aggregates are regularly computed over large numbers of similar data items
  - data is sparse, i.e. columns with many null values
- Often used in combination with in-memory databases

Column-oriented Databases

Example

Id	Genre	Title	Price	Audiobook price
1	fantasy	My first book	20	30
2	education	Beginners guide	10	null
3	education	SQL strikes back	40	null
4	fantasy	The rise of SQL	10	null

- Row based databases are not efficient at performing operations that apply to the entire data set
  - Need indexes which add overhead

Column-oriented Databases

- In a column-oriented database, all values of a column are placed together on disk

Genre:	fantasy:1,4	education:2,3		
Title:	My first book:1	Beginners guide:2	SQL strikes back:3	The rise of SQL:4
Price:	20:1	10:2,4	40:3	
Audiobook price:	30:1			

- A column matches the structure of a normal index in a row-based system
- Operations such as: find all records with price equal to 10 can now be executed directly
- Null values do not take up storage space anymore
- Aggregate operation (“GROUP BY like”) can be executed very quickly.
- The in-memory engine of Power BI is a column store database (called *xVelocity*)

Column-oriented Databases

- Disadvantages
  - Retrieving all attributes pertaining to a single entity becomes less efficient
  - Join operations will be slowed down
- Examples
  - Google BigTable, Cassandra, HBase, and Parquet

Source

- Principles of Database Management – Chapter 11.4