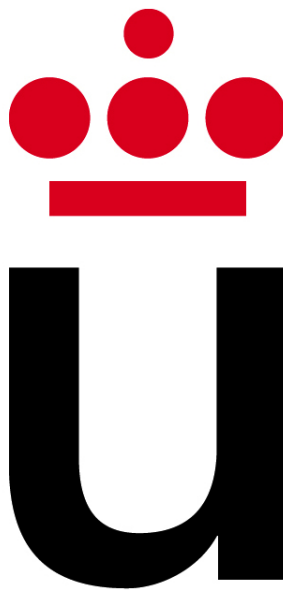


# FORO DE PREGUNTAS

## **Análisis de Big Data**

### PREGUNTAS TEMA 7

*José Ignacio Escribano*



MÓSTOLES, 9 DE ABRIL DE 2016

# Índice

1. Preguntas	1
1.1. . . . . .	1
1.2. . . . . .	1

# 1. Preguntas

## 1.1. **A pesar de su nombre, tomado del villano de la saga Harry Potter, Voldemort es una conocida base de datos NoSQL. Responda a estas preguntas: ¿En qué empresa surgió? ¿En cuál de las cuatro familiar NoSQL presentadas en teoría se puede encuadrar? ¿Cuáles son sus principales ventajas e inconvenientes?**

Voldemort es una base de datos NoSQL clave-valor, usada en LinkedIn, en numerosos sistemas críticos[1]. Sus principales características son:

- Replicación de datos automática en múltiples servidores.
- Particionado de datos automático en múltiples servidores.
- Fallos en el servidor de forma transparente.
- Los datos son versionados para maximizar la integridad en caso de fallo, sin comprometer la disponibilidad del sistema.
- Cada nodo es independiente de los demás.

Voldemort no es una base de datos relacional, ni intenta satisfacer relaciones arbitrarias, ni propiedades ACID. Es básicamente una gran tabla hash distribuida, persistente y tolerante a fallos.

Las principales ventajas de Voldemort son:

- Escalado horizontal
- Disponibilidad de información
- Código abierto
- Manejo de gran cantidad de información
- Evita cuellos de botella

Entre sus desventajas se encuentran:

- No cumple las reglas ACID.
- Poco utilizada [6].

Fuentes:

- [1] <http://www.project-voldemort.com/>
- [2] [http://static.usenix.org/events/fast/tech/full\\_papers/Sumbaly.pdf](http://static.usenix.org/events/fast/tech/full_papers/Sumbaly.pdf)
- [3] [https://en.wikipedia.org/wiki/Voldemort\\_\(distributed\\_data\\_store\)](https://en.wikipedia.org/wiki/Voldemort_(distributed_data_store))
- [4] <http://emanuelpeg.blogspot.com.es/2011/10/voldemort-una-base-de-datos-n>
- [5] [https://www.usenix.org/legacy/publications/login/2011-10/openpdfs/Burd.](https://www.usenix.org/legacy/publications/login/2011-10/openpdfs/Burd)
- [6] <http://db-engines.com/en/ranking>

## **1.2. Busque información sobre la primera base de datos (por fecha de lanzamiento) con arquitectura clave-valor de la familia NoSQL. ¿Dónde surgió, en una empresa o como proyecto de software libre? ¿Qué factores o necesidades determinaron su aparición?**

La primera base de datos fue dbm, acrónimo de DataBase Manager, lanzada en 1979 por AT&T y escrita por Ken Thompson. Propiamente no se trata de una base de datos sino de una librería.

dbm almacena datos arbitrarios mediante el uso de una clave única en buckets de longitud fija y utiliza técnicas de hashing para permitir una fácil recuperación de los datos por clave.

dbm surgió de la necesidad de crear y manipular una hashed database.

A partir de dbm surgieron varios sucesores:

- Ndbm: Berkeley introdujo en 1986 Ndbm (New DataBase Manager). Añadía soporte para múltiples bases de datos abiertas concurrentemente.
- GDBM (GNU dbm): versión libre escrita por Philip A. Nelson del proyecto GNU. Añadía soporte para datos de longitud arbitraria en la base de datos.
- tdbm: versión de ndbm con transacciones atómicas, bases de datos en memoria. Licencia BSD.

Fuentes:

- [1] <http://blog.knuthaugen.no/2010/03/a-brief-history-of-nosql.html>
- [2] <https://www.quora.com/Which-is-the-first-NoSQL-database>
- [3] <https://en.wikipedia.org/wiki/Dbm>
- [4] <https://en.wikipedia.org/wiki/Ndbm>
- [5] <http://www.gnu.org.ua/software/gdbm/>