Curso Intro Big Data – MongoDB

Intro a BigData y NoSQL

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

DEFINIENDO BIG DATA

Big Data

Una Primer Definición

"*Volumen masivo de datos*, tanto *estructurados como no-estructurados*, los cuales son *demasiado grandes y difíciles de procesar* con las bases de datos y el software *tradicionales*." (ONU, 2012)

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

DEFINIENDO BIG DATA

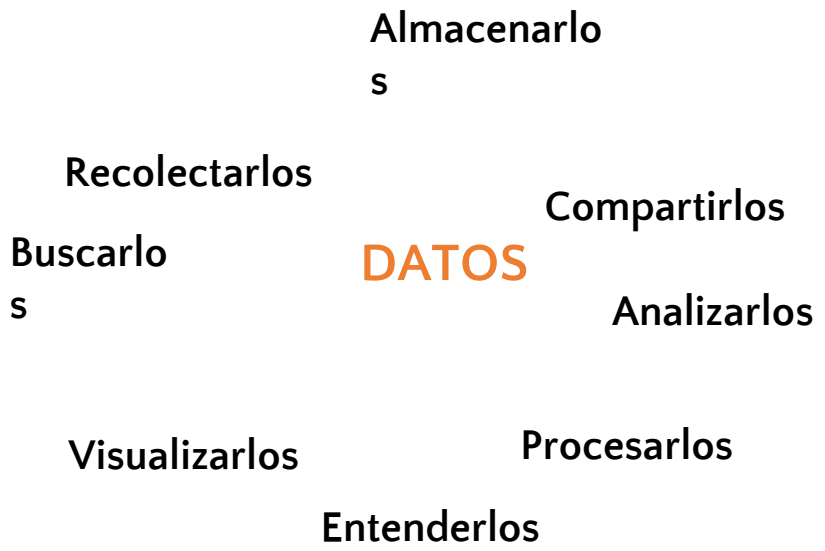
Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DEFINIENDO BIG DATA

Big Data es el sector de IT que hace referencia a ***grandes conjuntos de datos*** que por la ***velocidad*** a la que se generan, la capacidad para tratarlos y los ***múltiples formatos y fuentes***, es necesario procesarlos con mecanismos distintos a los tradicionales.

BIG DATA

- Volumen
- Velocidad
- Variedad
- ”Veracidad”

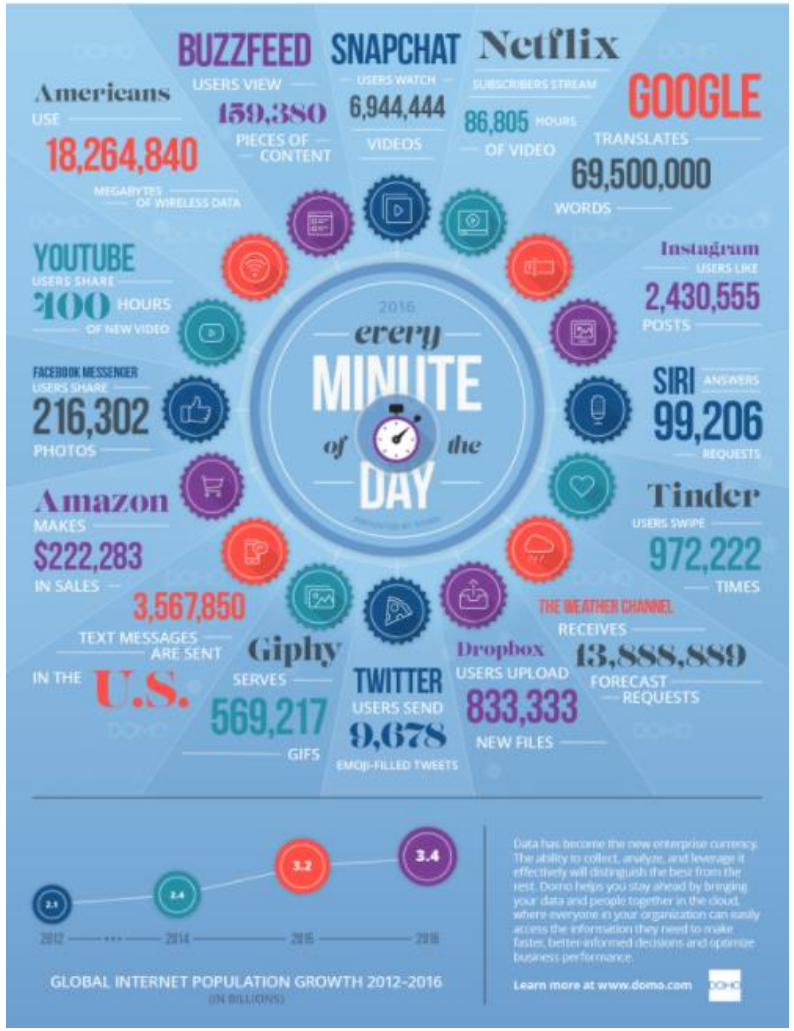


Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

DEFINIENDO BIG DATA

Año 2016



Por cada minuto del día

YouTube 400 hs. de Video
Google 69.500.000 palabras traducidas
Netflix 86,805 Horas de Video
Facebook 4,166.667 User share
Siri 99,206 Requests
Whatsapp 347,222 Photos
Tinder 972,222 Users Swipe
SnapChat 6,944,444 Videos Watched
TWC 13,888,889 Pronósticos

Población Total de Internet

3.400.000.000 de personas

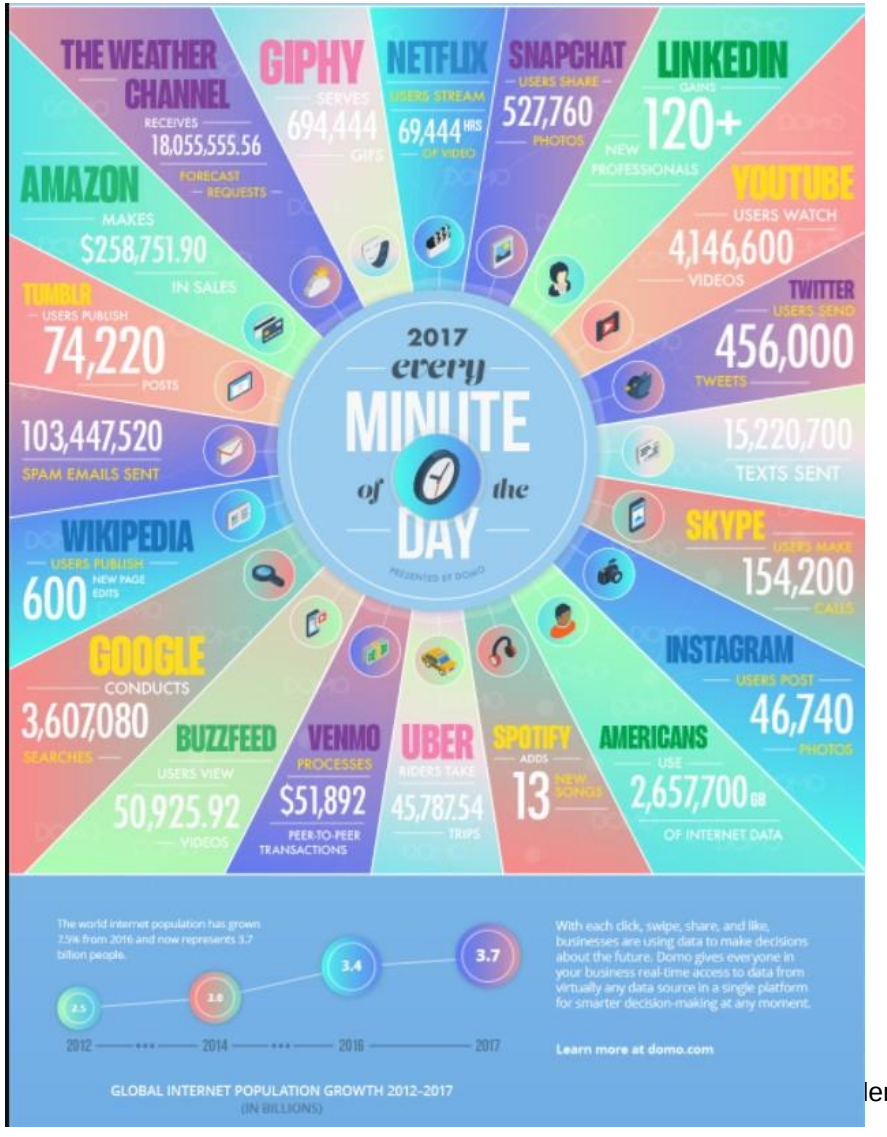
https://web-assets.domo.com/blog/wp-content/uploads/2016/06/16_domo_data-never-sleeps-4-2.png

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

DEFINIENDO BIG DATA

Año 2017



Por cada minuto del día

Spotify 13 nuevas canciones
Uber 45,787 viajes
Mails Spam 103,447,520 mails
Twitter 456,000 Tweets enviados
Youtube 4,146,600 Videos vistos
SnapChat 527,760 Fotos Compartidas
TWC 18,055,555 Pronósticos

Población Total de Internet

3.700.000.000 de personas

https://www.domo.com/learn/data-never-sleeps-5

lerechos reservados.

DBlandIT®

DEFINIENDO BIG DATA

PRINCIPALES CAMBIOS QUE SE PRODUJERON EN LA TECNOLOGÍA Y EN LOS ÚLTIMOS 15 AÑOS

- MASIFICACIÓN USO DE INTERNET
- SURGIMIENTO DE LAS REDES SOCIALES
- CRECIMIENTO EXPONENCIAL DE DISPOSITIVOS MÓVILES
- INTERFACES DE USUARIO MAS SIMPLES E INTUITIVAS
- CAMBIOS EN LAS FORMAS DE PROCESAMIENTO
- FUERTE BAJA EN LOS COSTOS DE ALMACENAMIENTO

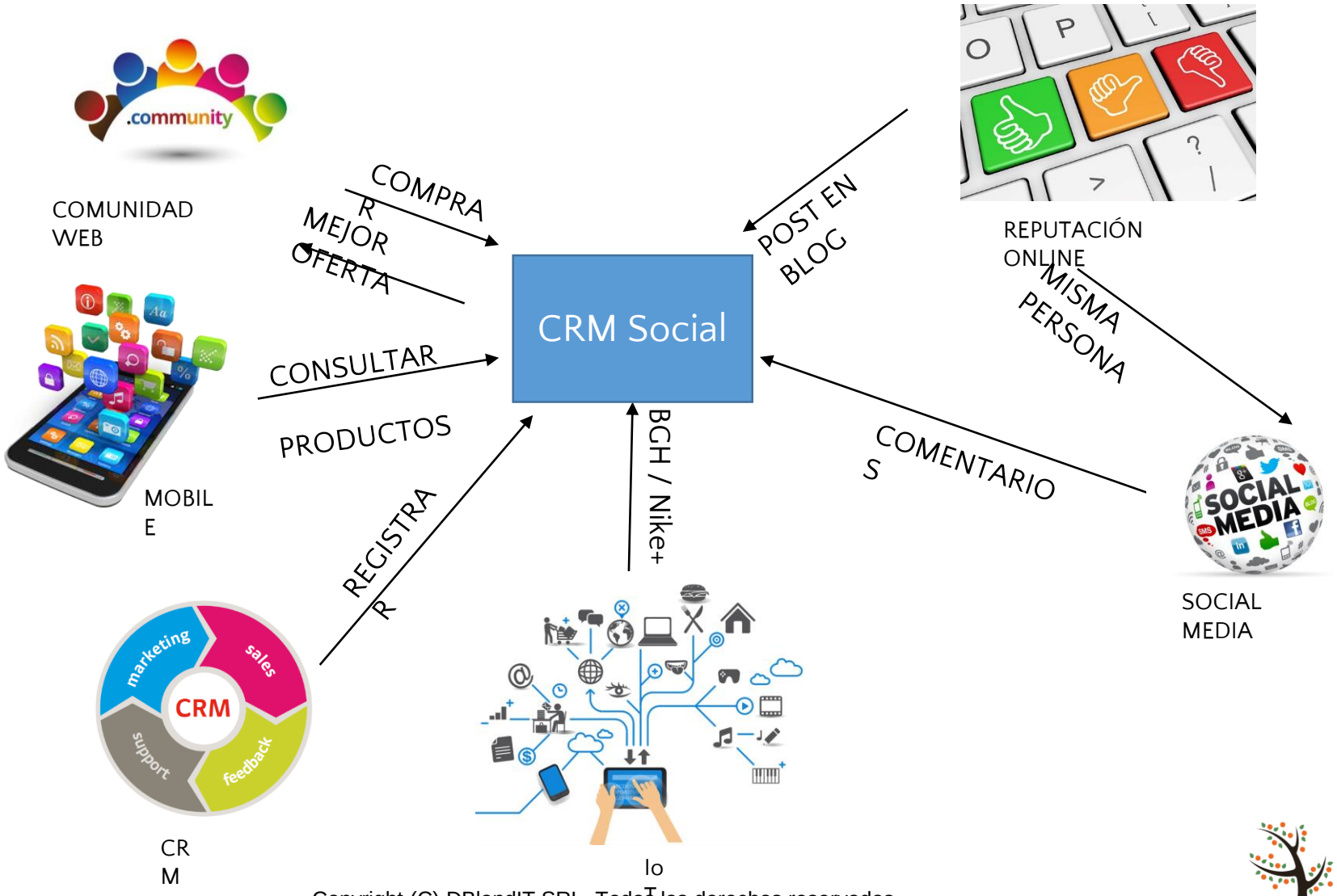
CADA DÍA CREAMOS 2,5 QUINTILLONES DE BYTES DE DATOS. (2,5 Exabytes)

EL 90% DE LOS DATOS DEL MUNDO DE HOY SE GENERARON EN LOS ÚLTIMOS 2 AÑOS

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

BIG DATA Y LOS NEGOCIOS



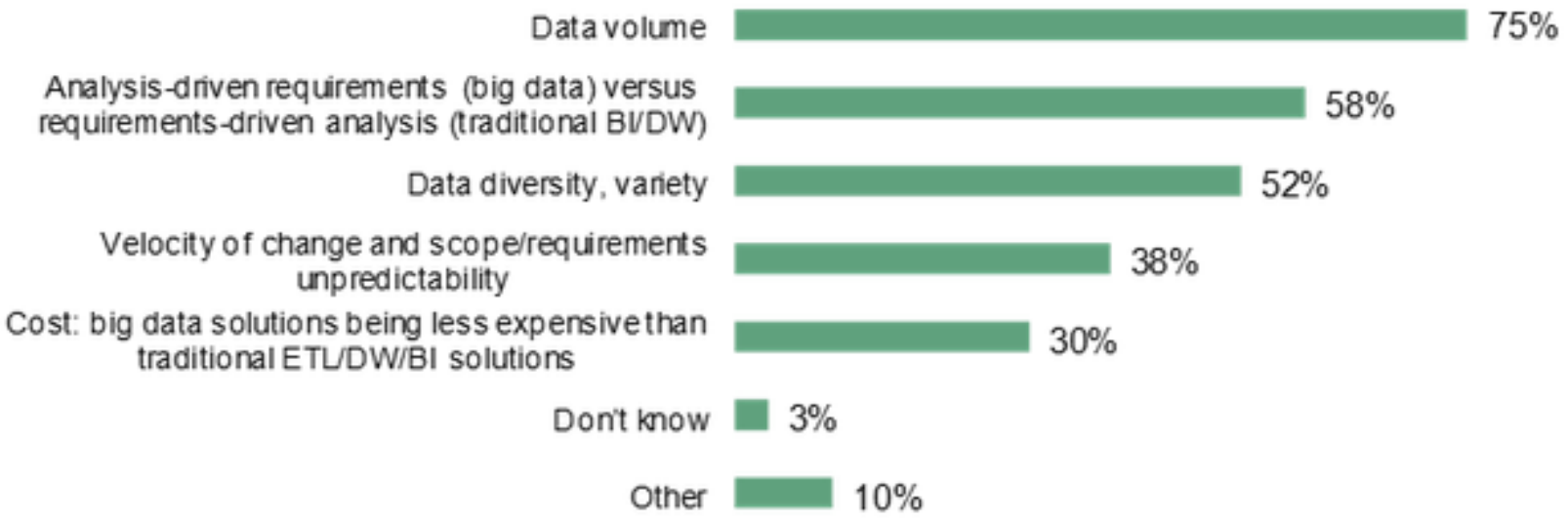
Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

DEFINIENDO BIG DATA

¿ CUÁLES DE LAS 4 Vs TIENEN MAYOR INFLUENCIA ?

En ambientes tradicionales de BI y DW primero se generan los requerimientos y luego las aplicaciones. Dicho de otra forma, los requerimientos direccionan las aplicaciones. En Big Data es al revés, ya que se utiliza la exploración de datos libre para generar hipótesis para encontrar un patrón



El costo es un factor en muchos casos. Las tecnologías utilizadas en Big Data son más económicas que las tradicionales.

fuente: forrester research, global big data survey

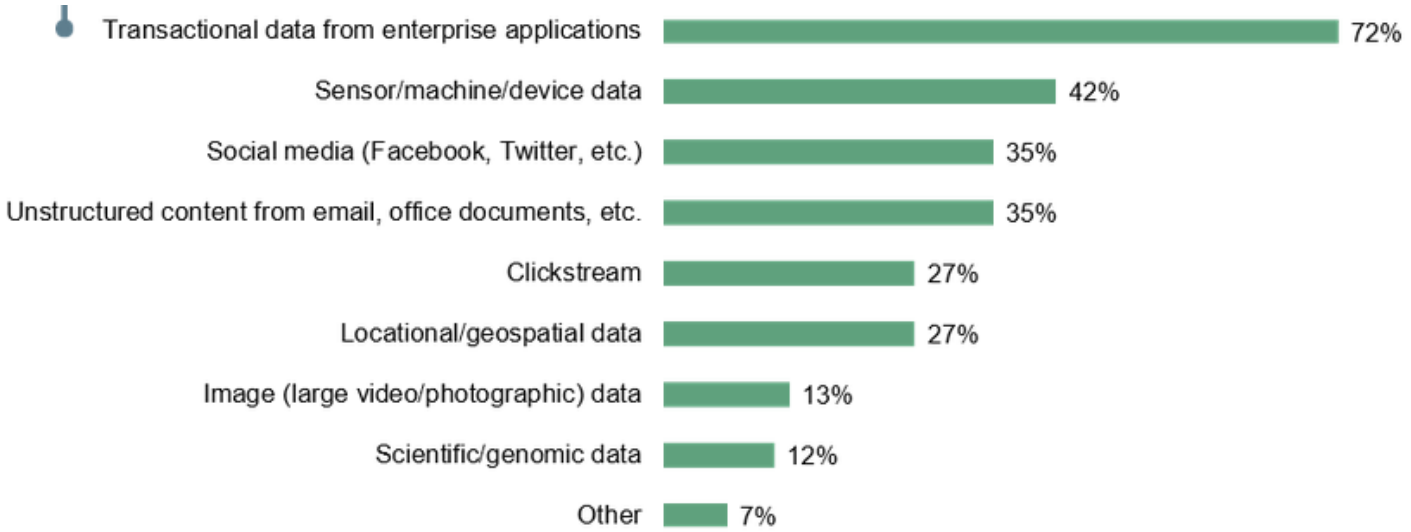
Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

DEFINIENDO BIG DATA

TIPOS DE DATOS QUE SE ANALIZAN A TRAVÉS DE BIG DATA

Si bien pareciera que Big Data se utiliza fundamentalmente para datos puros obtenidos de redes sociales, sensores, tráfico web la realidad es que más allá de lo que se supone, las empresas utilizan Big Data para operar en datos operacionales.



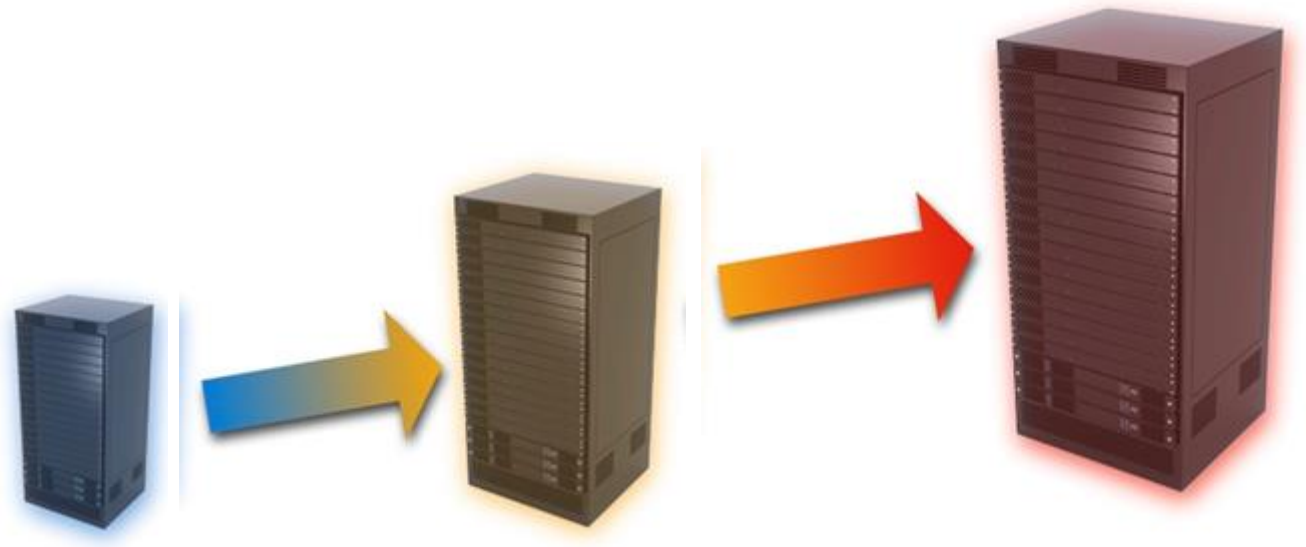
fuente: forrester research. global big data survey

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

IMPLEMENTANDO BIG DATA

ESCALAMIENTO



Escalamiento Vertical

Escalamiento Vertical

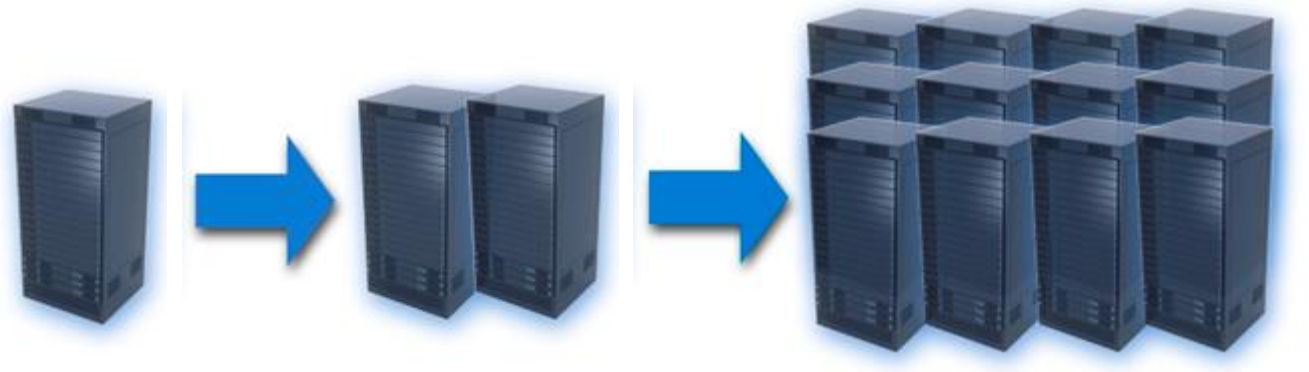
- Escalamiento dentro de un mismo servidor.
- Implica incrementar la capacidad de un Servidor agregando más recursos de CPU, memoria y de almacenamiento.

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

IMPLEMENTANDO BIG DATA

ESCALAMIENTO



Escalamiento Horizontal

- Escalamiento en varios servidores.
- Cluster de Servidores.
- Replicación de Datos.
- Particionamiento de Datos.
- Procesamiento Paralelo.

Escalamiento Horizontal

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

IMPLEMENTANDO BIG DATA

Cluster

Grupo de servidores independientes interconectados a través de una red dedicada que trabajan como un único recurso de procesamiento

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

ALTA DISPONIBILIDAD Y TOLERANCIA A FALLOS

- Aplicaciones 7 x 24.
- Aplicaciones de Misión Crítica.

- Replicación

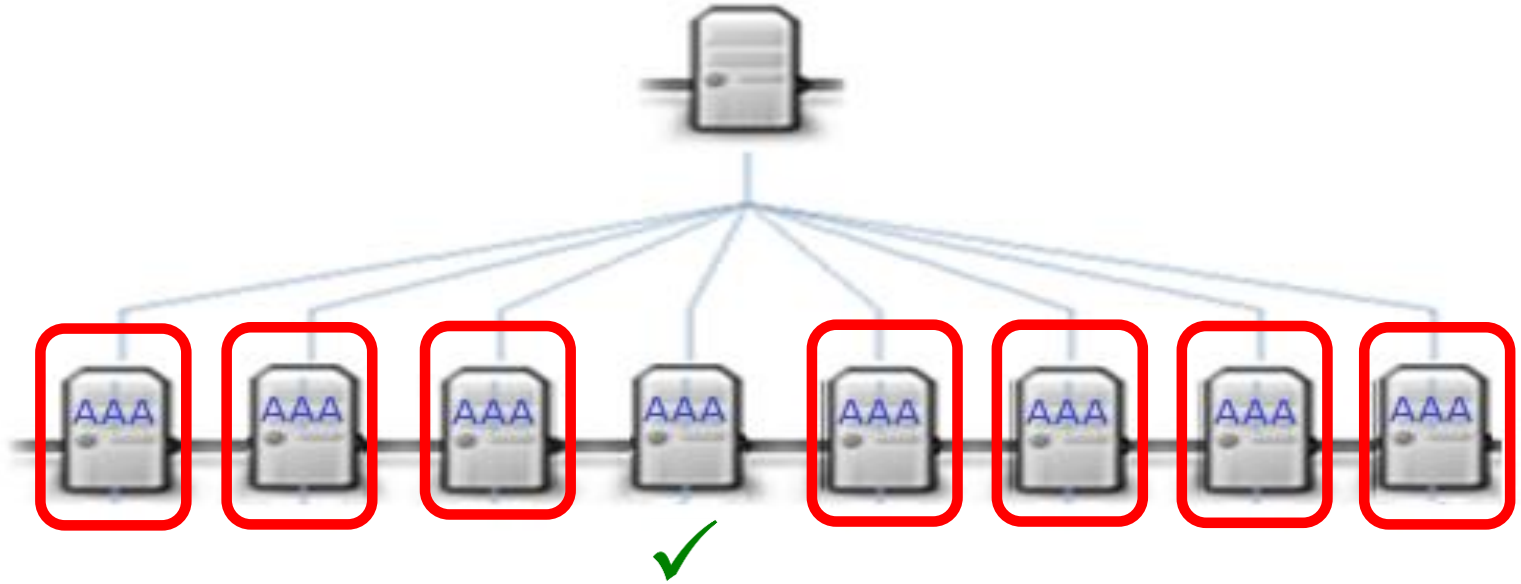
Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

ALTA DISPONIBILIDAD Y TOLERANCIA A FALLOS

- Aplicaciones 7 x 24.
- Aplicaciones de Misión Crítica.

- Replicación

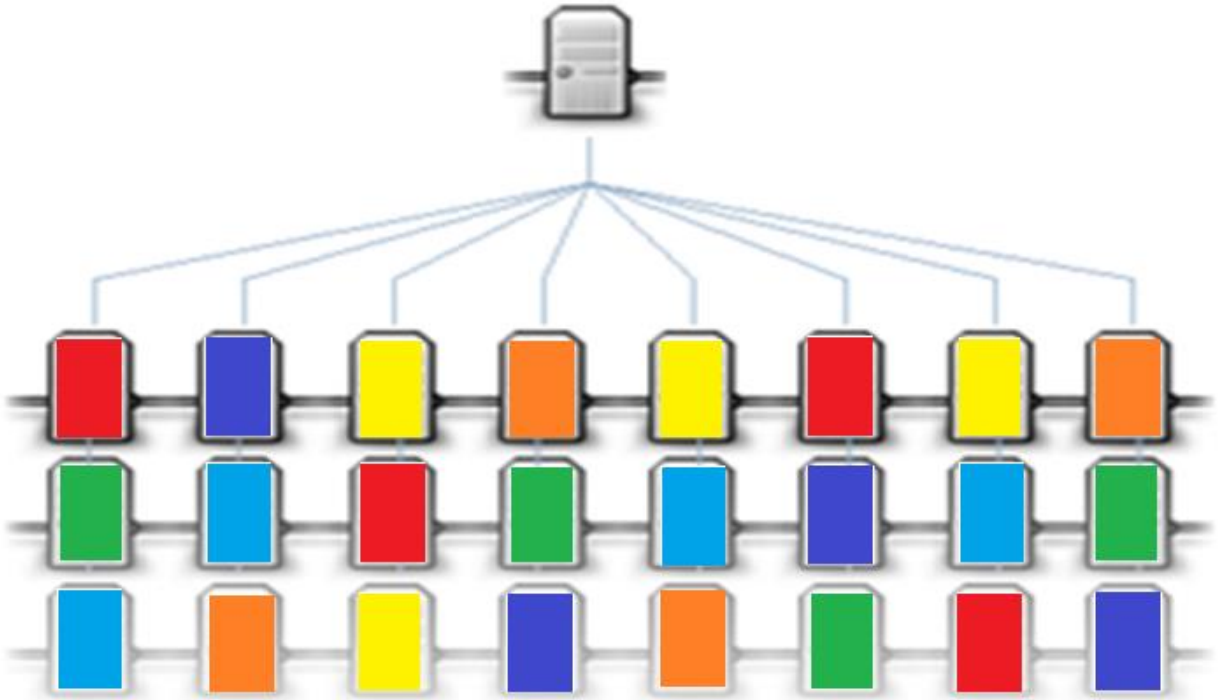


Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

IMPLEMENTANDO BIG DATA

PARTICIONAMIENTO DE DAT(



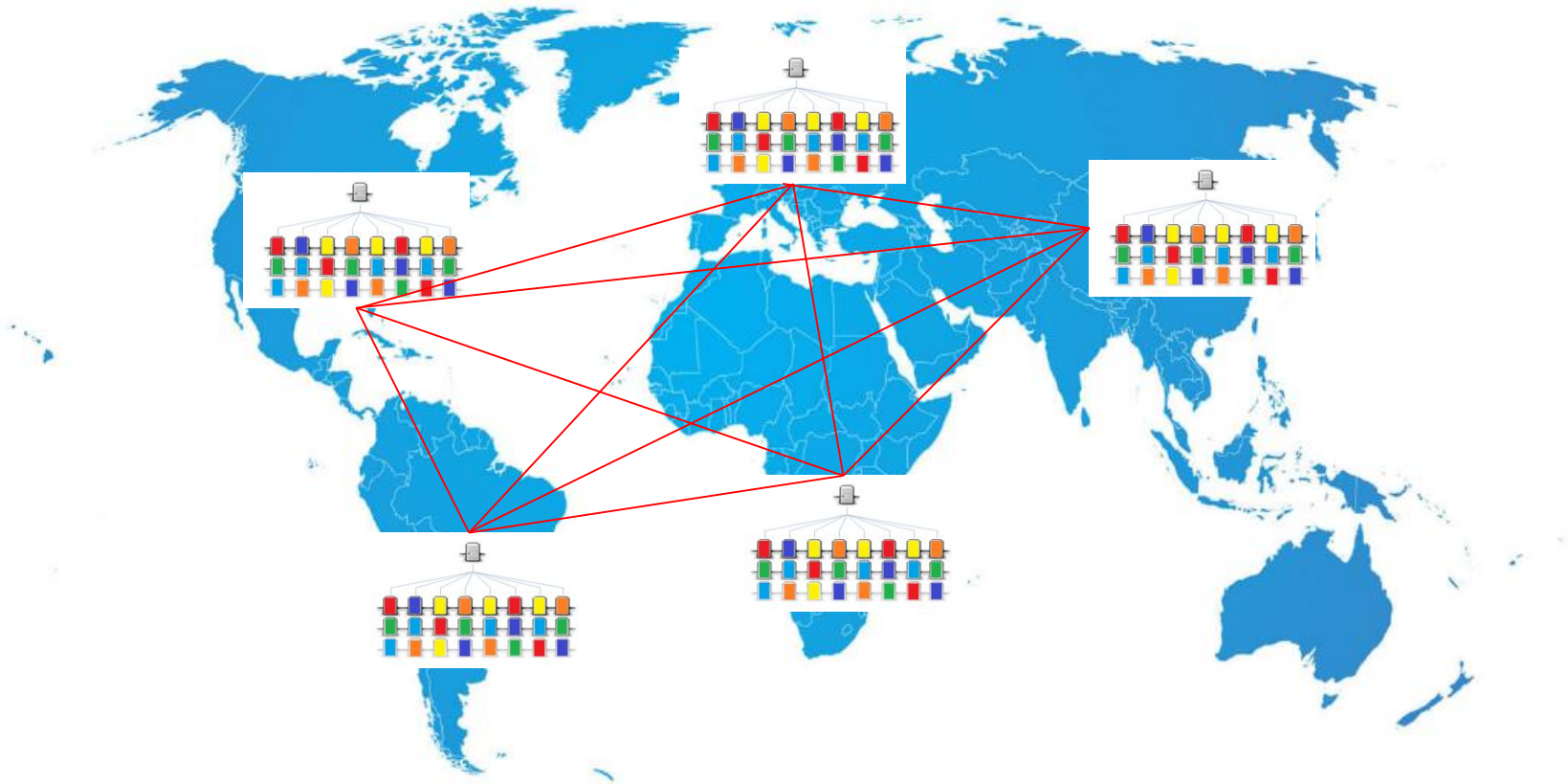
- Un solo Servidor no soporta almacenar la totalidad de los datos.
- Se deben particionar los datos en múltiples Servidores del Cluster.
- Además los datos se encuentran replicados.

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

IMPLEMENTANDO BIG DATA

PARTICIONAMIENTO DE



- Necesidad de distribución geográfica de datos.

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

IMPLEMENTANDO BIG DATA

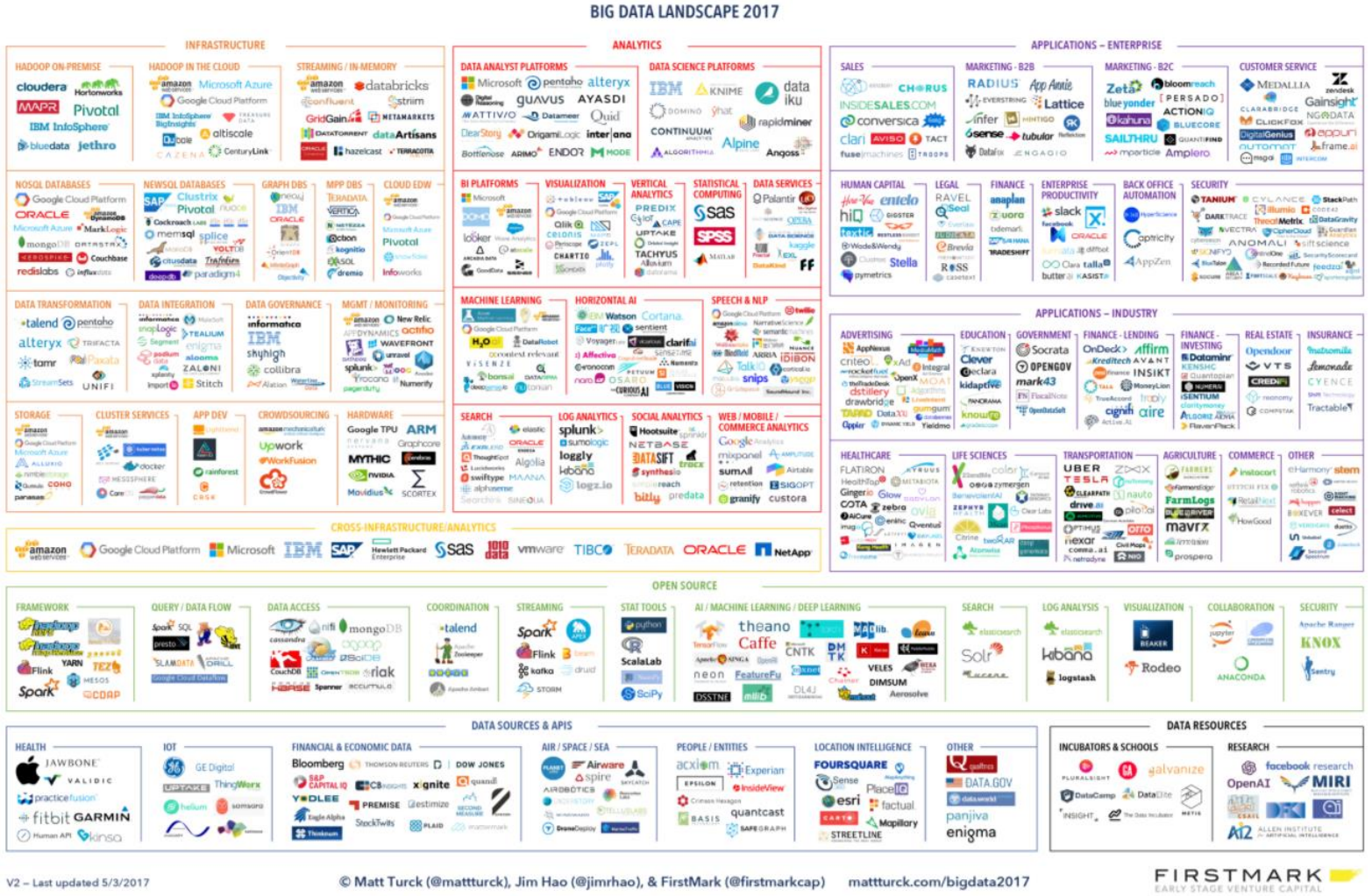
PROCESAMIENTO PARALELO

- Varios servidores procesan un mismo programa de forma simultánea para resolver un determinado problema.

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

Big Data Landscape

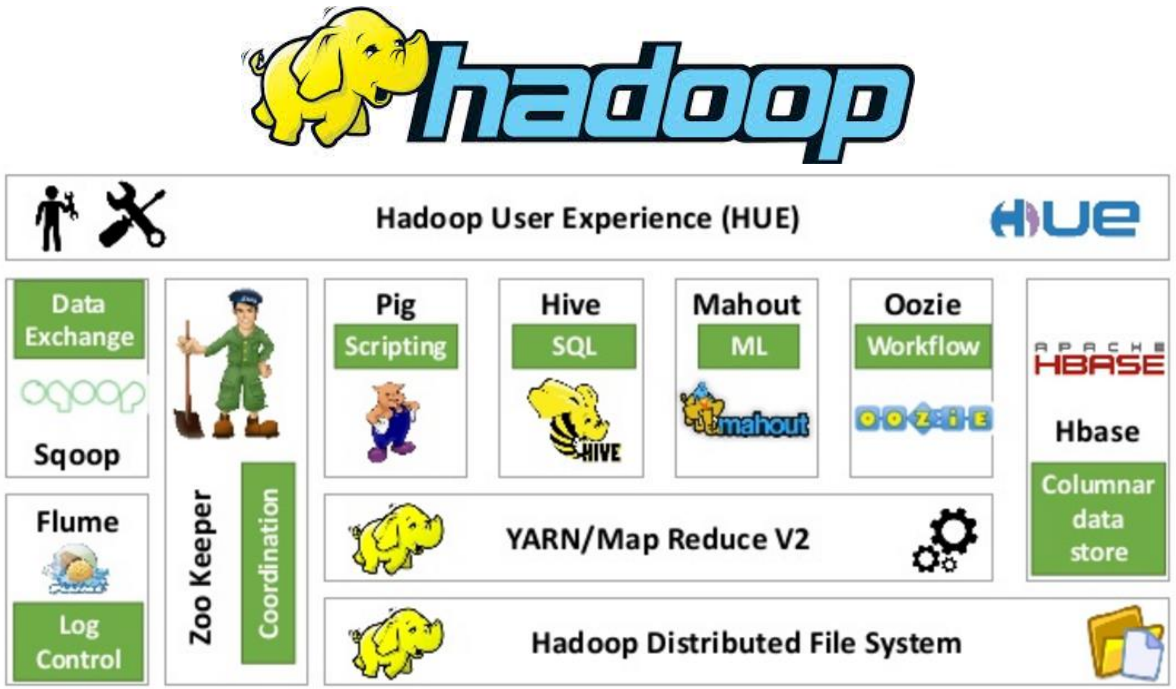


Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

IMPLEMENTANDO BIG DATA

HADOOP



¿ Qué es Hadoop ?

Hadoop es un **sistema de código abierto** que se utiliza para **almacenar, procesar en paralelo y analizar grandes volúmenes de datos**.

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

IMPLEMENTANDO BIG DATA

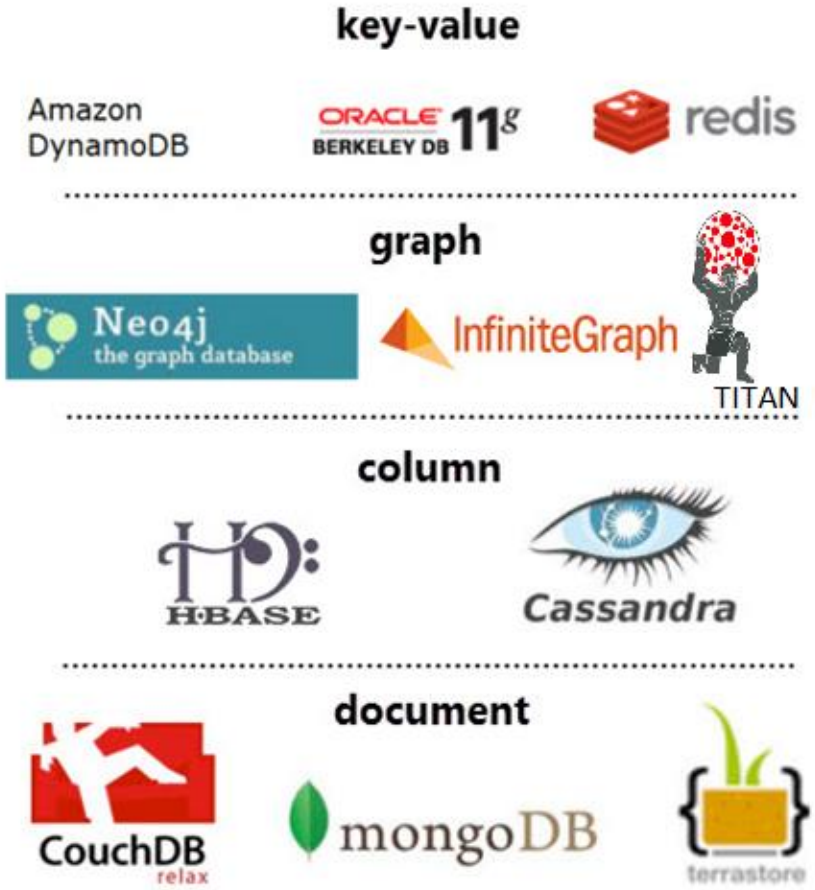
HADOOP –

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

IMPLEMENTANDO BIG DATA

BASES DE DATOS NO-SQL



¿ Qué es NoSQL ?

Sistemas de gestión de bases de datos que difieren del modelo clásico de bases de datos relacionales: no usan SQL como lenguaje de consulta, los datos almacenados no requieren estructuras fijas como tablas, no garantizan consistencia plena y escalan horizontalmente.

¿ Qué es la Pesistencia Políglota ?

Utilizar dentro de un mismo ambiente o aplicación un conjunto de bases de datos, que colabora, cada una en lo que es más importante.

reservados.

DBlandIT®

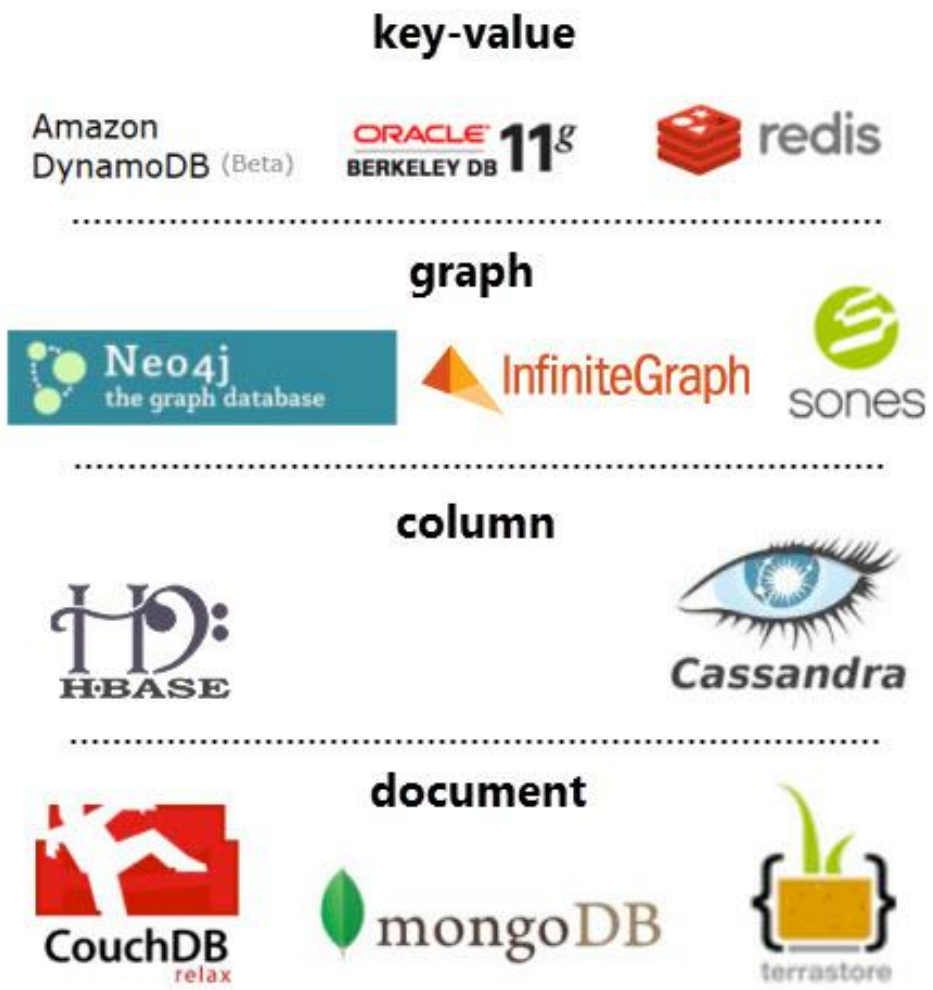
Introducción a

Bases de Datos NoSQL

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

NOSQL DATABASE TYPES



¿ Qué es NoSQL ?

Sistemas de gestión de bases de datos que difieren del modelo clásico de bases de datos relacionales: no usan SQL como lenguaje de consulta, los datos almacenados no requieren estructuras fijas como tablas, no garantizan consistencia y escalan horizontalmente.

¿ Qué es la Pesistencia Políglota ?

Utilizar dentro de un mismo ambiente o aplicación un conjunto de bases de datos, que colabora, cada una en lo que es más importante.

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

Persistencia Políglota

Diferentes tecnologías de bases de datos para resolver diferentes problemas desde una misma aplicación.

Búsquedas Performantes sobre Catálogo de Productos

Información Distribuida Geográficamente Profile de usuarios y Documentación de Productos con Info no estructurada

Caché de Sesiones Lockeos Distribuidos

Transacciones Económicas.

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

NOSQL – KEY VALUE DB

Key Value / Caching

RDS amazon web services

memcached
high-performance, distributed memory object caching system

redis

riak

Simple Key-Value Reads/Writes

No Analytics

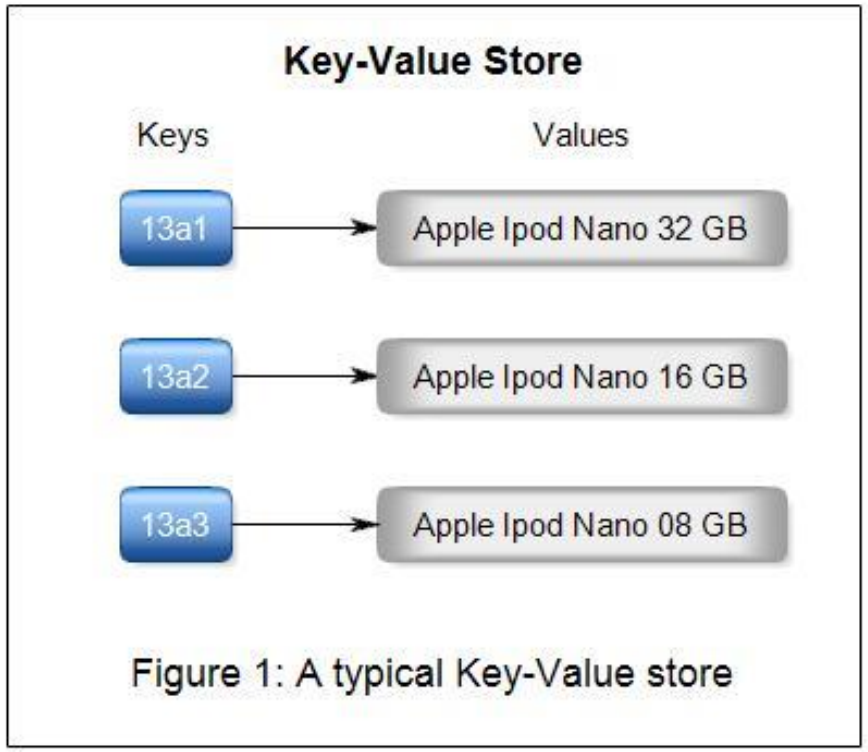


Figure 1: A typical Key-Value store

¿ Cuándo se Usan ?

- Almacenar información de sesiones
- Perfiles de Usuarios
- Información de carros de compras

github SOCIAL CODING

Pinterest

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

NOSQL – COLUMNAR DB

Tabular

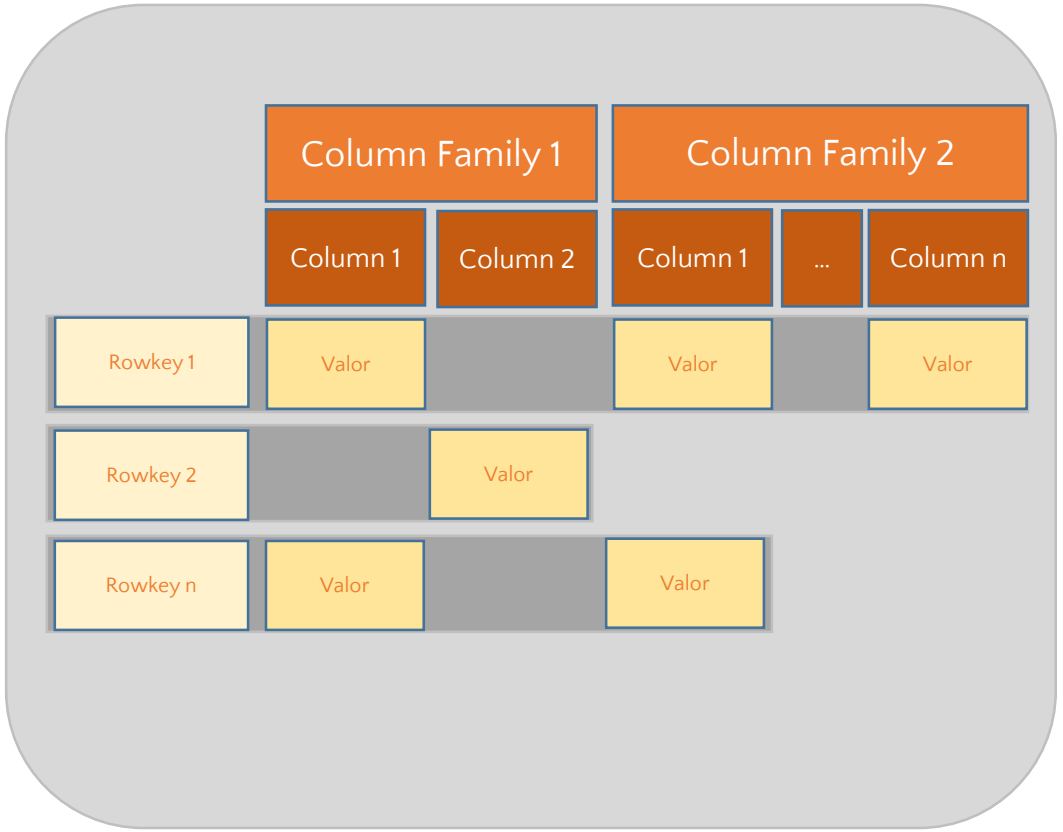
Google Bigtable

APACHE HBASE

cassandra

Sparse Column Families(~Tables)

No Analytics



¿ Cuándo se Usan ?

CMS, blogging

Web-analytics / Real-Time analytics

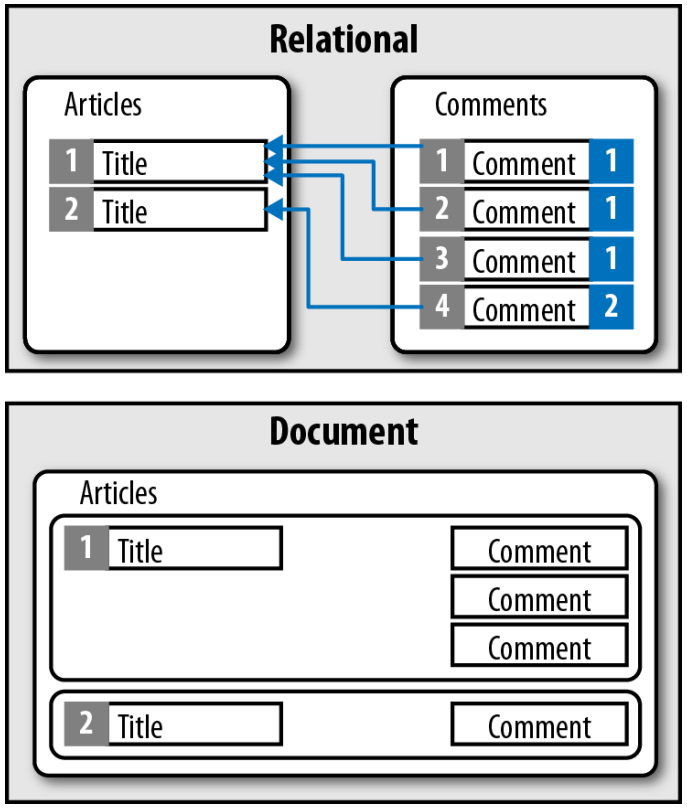
Expiring

Time series

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

NOSQL – DOCUMENT BASED DB



¿ Cuándo se Usan ?

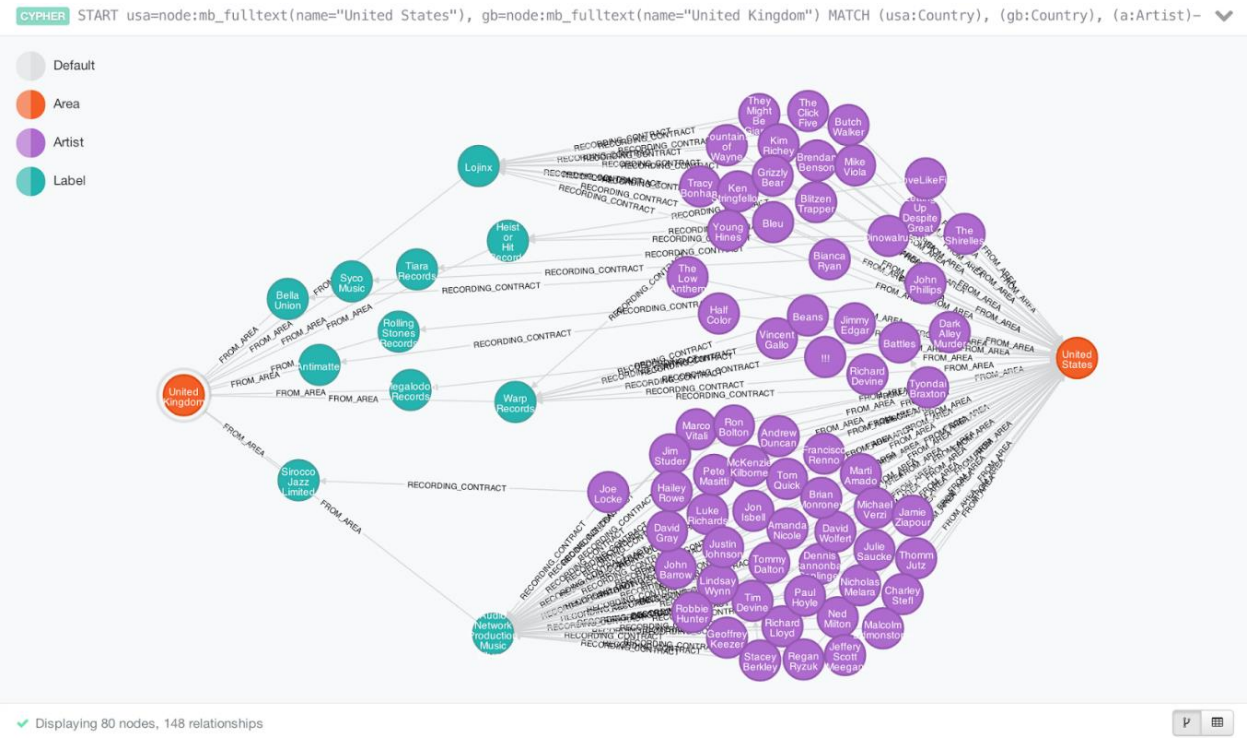
- Logging de Eventos
- CMS, blogging
- Web-analytics / Real-Time analytics
- E-Commerce

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

NOSQL – GRAPH DB

Neo4j
the graph database



¿ Cuándo se Usan ?

- Datos interconectados
- Servicios de Ruteo / Despachos
- Motores de recomendaciones

cisco ebay

cisco ebay

Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

NOSQL – DB-ENGINES.COM

381 systems in ranking, November 2021

Rank Nov 2021	Rank Oct 2021	Rank Nov 2020	DBMS	Database Model	Score Nov 2021	Score Oct 2021	Score Nov 2020
1.	1.	1.	Oracle	Relational, Multi-model	1272.73	+2.38	-72.27
2.	2.	2.	MySQL	Relational, Multi-model	1211.52	-8.25	-30.12
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	954.29	-16.32	-83.35
4.	4.	4.	PostgreSQL	Relational, Multi-model	597.27	+10.30	+42.22
5.	5.	5.	MongoDB	Document, Multi-model	487.35	-6.21	+33.52
6.	6.	↑ 7.	Redis	Key-value, Multi-model	171.50	+0.15	+16.08
7.	7.	↓ 6.	IBM Db2	Relational, Multi-model	167.52	+1.56	+5.90
8.	8.	8.	Elasticsearch	Search engine, Multi-model	159.09	+0.84	+7.54
9.	9.	9.	SQLite	Relational	129.80	+0.43	+6.48
10.	10.	10.	Cassandra	Wide column	120.88	+1.61	+2.13
11.	11.	11.	Microsoft Access	Relational	119.24	+2.86	+2.01
12.	12.	12.	MariaDB	Relational, Multi-model	102.19	-0.41	+9.90
13.	13.	13.	Splunk	Search engine	92.31	+1.69	+2.60
14.	14.	↑ 15.	Hive	Relational	83.31	-1.43	+13.05
15.	15.	↑ 17.	Microsoft Azure SQL Database	Relational, Multi-model	81.32	+1.60	+14.33
16.	16.	16.	Amazon DynamoDB	Multi-model	76.99	+0.43	+8.09
17.	17.	↓ 14.	Teradata	Relational, Multi-model	69.59	-0.24	-6.01
18.	18.	↑ 42.	Snowflake	Relational	64.19	+5.93	+54.09
19.	19.	↑ 20.	Neo4j	Graph	57.98	+0.11	+4.45
20.	20.	↓ 19.	SAP HANA	Relational, Multi-model	55.53	+0.26	+1.95
21.	21.	↑ 23.	FileMaker	Relational	54.22	+1.38	+7.56
22.	22.	↓ 21.	Solr	Search engine, Multi-model	53.85	+2.69	+2.04

http://db-engines.com/en/ranking
HOY

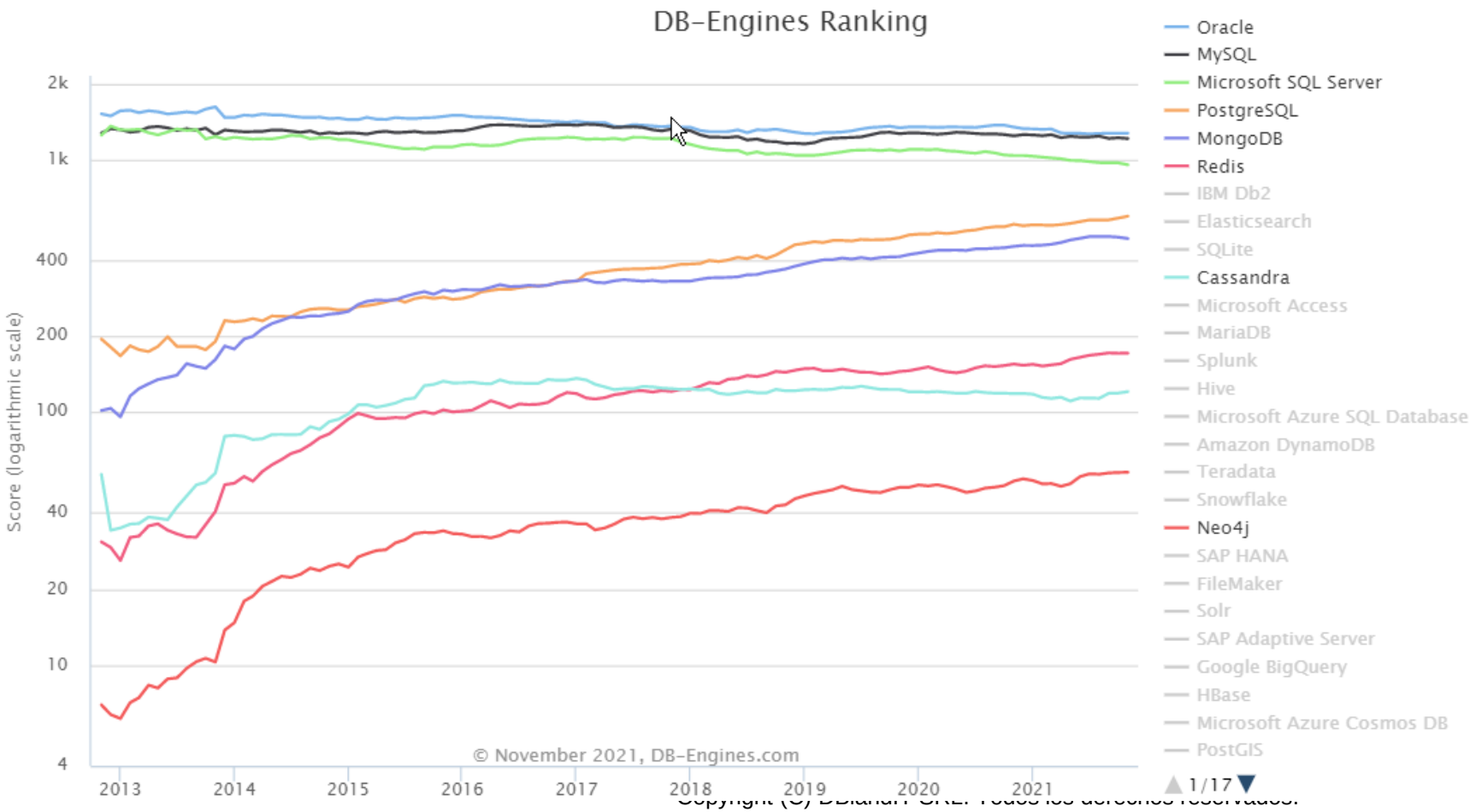
Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®

NOSQL – DB-ENGINES.COM

http://db-engines.com/en/ranking
HOY

November 2021



Copyright (C) DBlandIT SRL. Todos los derechos reservados.

DBlandIT®