

Presentación

Luis Llana Díaz

Departamento de Sistemas Informáticos y Computación
Universidad Complutense de Madrid

6 de septiembre de 2026

Contacto con el profesor

Nombre Luis Llana Díaz.

Despacho 115 (1ª planta).

Email llana@ucm.es (consultas personales, **NO** dudas).

Tutorías L,M,X 13:00-14:00 (previa petición).

Recursos digitales

Campus Virtual

NO se usa el oficial UCM:
<https://wild.ucm.es>

- Dudas y consultas.
- Anuncios
- Material



Recursos digitales

Campus Virtual

NO se usa el oficial UCM:

<https://wild.ucm.es>

- Dudas y consultas.
- Anuncios
- Material



Servidor con MongoDB

<https://isabelita.fdi.ucm.es/jupyter>

- Servidor MongoDB
- Ejercicios y material online



Objetivos

- MongoDB.

Objetivos

- MongoDB.
 - Realizar consultas en mongoDB.

Objetivos

- MongoDB.
 - Realizar consultas en mongoDB.
 - Realizar programas que usen MongoDB.

Objetivos

- MongoDB.
 - Realizar consultas en mongoDB.
 - Realizar programas que usen MongoDB.
 - Instalación de mongoDB en local

Objetivos

- MongoDB.
 - Realizar consultas en mongoDB.
 - Realizar programas que usen MongoDB.
 - Instalación de mongoDB en local
 - Uso de Atlas.

Objetivos

- MongoDB.
 - Realizar consultas en mongoDB.
 - Realizar programas que usen MongoDB.
 - Instalación de mongoDB en local
 - Uso de Atlas.
 - Instalación de servidor con Replicación (ReplicaSet) y Fragmentación (Sharding).

Evaluación

La evaluación continua

Evaluación

La evaluación continua

Participación. 20 %

La evaluación continua

Participación. 20 %

- Control de asistencia obligatoria.

La evaluación continua

Participación. 20 %

- Control de asistencia obligatoria.
- Cuestionarios.

La evaluación continua

Participación. 20 %

- Control de asistencia obligatoria.
- Cuestionarios.

3 exámenes parciales 80 %.

La evaluación continua

Participación. 20 %

- Control de asistencia obligatoria.
- Cuestionarios.

3 exámenes parciales 80 %.

- Pesos 1, 2, 3.

La evaluación continua

Participación. 20 %

- Control de asistencia obligatoria.
- Cuestionarios.

3 exámenes parciales 80 %.

- Pesos 1, 2, 3.
- Es necesario sacar un 4 en el 3er examen parcial.

Evaluación

La evaluación continua

Participación. 20 %

- Control de asistencia obligatoria.
- Cuestionarios.

3 exámenes parciales 80 %.

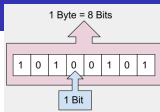
- Pesos 1, 2, 3.
- Es necesario sacar un 4 en el 3er examen parcial.

Examen final

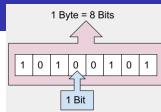
Se puede aprobar solamente con el examen final.

¿Qué es Big Data? Magnitudes de memoria

¿Qué es Big Data? Magnitudes de memoria

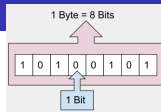


¿Qué es Big Data? Magnitudes de memoria



kilobyte $2^{10} \approx 10^3$.

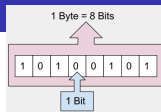
¿Qué es Big Data? Magnitudes de memoria



kilobyte $2^{10} \approx 10^3$.

megabyte $2^{20} \approx 10^6$.

¿Qué es Big Data? Magnitudes de memoria

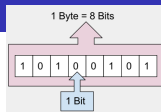


kilobyte $2^{10} \approx 10^3$.

megabyte $2^{20} \approx 10^6$.

gigabyte $2^{30} \approx 10^9$.

¿Qué es Big Data? Magnitudes de memoria



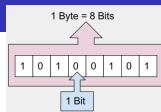
kilobyte $2^{10} \approx 10^3$.

megabyte $2^{20} \approx 10^6$.

gigabyte $2^{30} \approx 10^9$.

terabyte $2^{40} \approx 10^{12}$.

¿Qué es Big Data? Magnitudes de memoria



kilobyte $2^{10} \approx 10^3$.

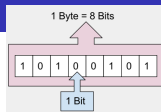
megabyte $2^{20} \approx 10^6$.

gigabyte $2^{30} \approx 10^9$.

terabyte $2^{40} \approx 10^{12}$.

petabyte $2^{50} \approx 10^{15}$.

¿Qué es Big Data? Magnitudes de memoria



kilobyte $2^{10} \approx 10^3$.

megabyte $2^{20} \approx 10^6$.

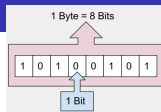
gigabyte $2^{30} \approx 10^9$.

terabyte $2^{40} \approx 10^{12}$.

petabyte $2^{50} \approx 10^{15}$.

exabyte $2^{60} \approx 10^{18}$.

¿Qué es Big Data? Magnitudes de memoria



kilobyte $2^{10} \approx 10^3$. El Quijote 850kb

megabyte $2^{20} \approx 10^6$.

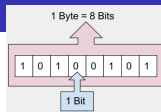
gigabyte $2^{30} \approx 10^9$.

terabyte $2^{40} \approx 10^{12}$.

petabyte $2^{50} \approx 10^{15}$.

exabyte $2^{60} \approx 10^{18}$.

¿Qué es Big Data? Magnitudes de memoria



kilobyte $2^{10} \approx 10^3$. El Quijote 850kb

megabyte $2^{20} \approx 10^6$. Foto 3Mb (8Mpx). 1CD 800Mb

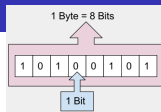
gigabyte $2^{30} \approx 10^9$.

terabyte $2^{40} \approx 10^{12}$.

petabyte $2^{50} \approx 10^{15}$.

exabyte $2^{60} \approx 10^{18}$.

¿Qué es Big Data? Magnitudes de memoria



kilobyte $2^{10} \approx 10^3$. El Quijote 850kb

megabyte $2^{20} \approx 10^6$. Foto 3Mb (8Mpx). 1CD 800Mb

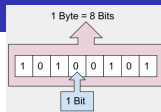
gigabyte $2^{30} \approx 10^9$. 4.7Gb DVD. 120Gb Wikipedia

terabyte $2^{40} \approx 10^{12}$.

petabyte $2^{50} \approx 10^{15}$.

exabyte $2^{60} \approx 10^{18}$.

¿Qué es Big Data? Magnitudes de memoria



kilobyte $2^{10} \approx 10^3$. El Quijote 850kb

megabyte $2^{20} \approx 10^6$. Foto 3Mb (8Mpx). 1CD 800Mb

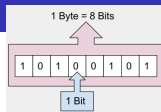
gigabyte $2^{30} \approx 10^9$. 4.7Gb DVD. 120Gb Wikipedia

terabyte $2^{40} \approx 10^{12}$. +200 DVD. 750Tb/día Facebook

petabyte $2^{50} \approx 10^{15}$.

exabyte $2^{60} \approx 10^{18}$.

¿Qué es Big Data? Magnitudes de memoria



kilobyte $2^{10} \approx 10^3$. El Quijote 850kb

megabyte $2^{20} \approx 10^6$. Foto 3Mb (8Mpx). 1CD 800Mb

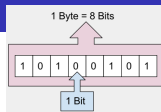
gigabyte $2^{30} \approx 10^9$. 4.7Gb DVD. 120Gb Wikipedia

terabyte $2^{40} \approx 10^{12}$. +200 DVD. 750Tb/día Facebook

petabyte $2^{50} \approx 10^{15}$. Netflix 3.14 Pb. Google Maps 20 Pb

exabyte $2^{60} \approx 10^{18}$.

¿Qué es Big Data? Magnitudes de memoria



kilobyte $2^{10} \approx 10^3$. El Quijote 850kb

megabyte $2^{20} \approx 10^6$. Foto 3Mb (8Mpx). 1CD 800Mb

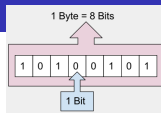
gigabyte $2^{30} \approx 10^9$. 4.7Gb DVD. 120Gb Wikipedia

terabyte $2^{40} \approx 10^{12}$. +200 DVD. 750Tb/día Facebook

petabyte $2^{50} \approx 10^{15}$. Netflix 3.14 Pb. Google Maps 20 Pb

exabyte $2^{60} \approx 10^{18}$. Google 10 Eb on disk

¿Qué es Big Data? Magnitudes de memoria



kilobyte $2^{10} \approx 10^3$. El Quijote 850kb

megabyte $2^{20} \approx 10^6$. Foto 3Mb (8Mpx). 1CD 800Mb

gigabyte $2^{30} \approx 10^9$. 4.7Gb DVD. 120Gb Wikipedia

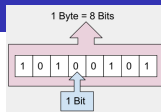
terabyte $2^{40} \approx 10^{12}$. +200 DVD. 750Tb/día Facebook

petabyte $2^{50} \approx 10^{15}$. Netflix 3.14 Pb. Google Maps 20 Pb

exabyte $2^{60} \approx 10^{18}$. Google 10 Eb on disk

zettabytes (zb), yottabytes (yb)...

¿Qué es Big Data? Magnitudes de memoria



kilobyte $2^{10} \approx 10^3$. El Quijote 850kb

megabyte $2^{20} \approx 10^6$. Foto 3Mb (8Mpx). 1CD 800Mb

gigabyte $2^{30} \approx 10^9$. 4.7Gb DVD. 120Gb Wikipedia

terabyte $2^{40} \approx 10^{12}$. +200 DVD. 750Tb/día Facebook

petabyte $2^{50} \approx 10^{15}$. Netflix 3.14 Pb. Google Maps 20 Pb

exabyte $2^{60} \approx 10^{18}$. Google 10 Eb on disk

zettabytes (zb), yottabytes (yb)...

<http://www.bluebulbprojects.com/MeasureOfThings/>

<http://www.internetlivestats.com/>

Tiempo. Complejidad Algorítmica

¿De qué depende el tiempo de ejecución de un programa?

Tiempo. Complejidad Algorítmica

¿De qué depende el tiempo de ejecución de un programa?

- Máquina sobre la que se ejecuta

Tiempo. Complejidad Algorítmica

¿De qué depende el tiempo de ejecución de un programa?

- Máquina sobre la que se ejecuta
- Lenguaje de implementación

Tiempo. Complejidad Algorítmica

¿De qué depende el tiempo de ejecución de un programa?

- Máquina sobre la que se ejecuta
- Lenguaje de implementación
- Algoritmo

Tiempo. Complejidad Algorítmica

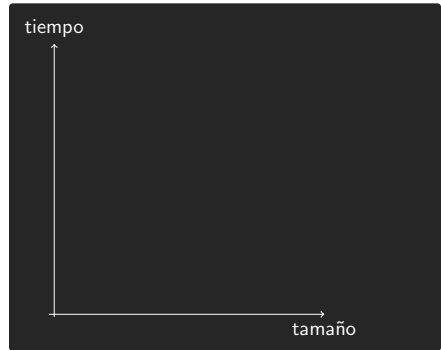
¿De qué depende el tiempo de ejecución de un programa?

- Máquina sobre la que se ejecuta
- Lenguaje de implementación
- Algoritmo
- Datos de entrada

Tiempo. Complejidad Algorítmica

Comportamiento de un programa según crece el tamaño de los datos.

Tamaño del problema n :

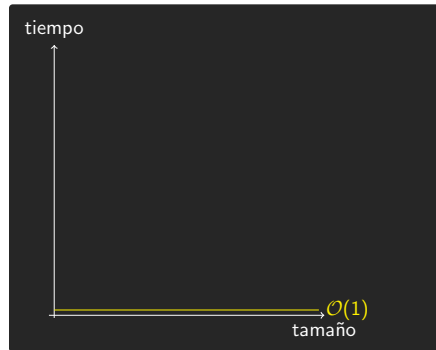


Tiempo. Complejidad Algorítmica

Comportamiento de un programa según crece el tamaño de los datos.

Tamaño del problema n :

- $\mathcal{O}(1)$: tiempo constante.

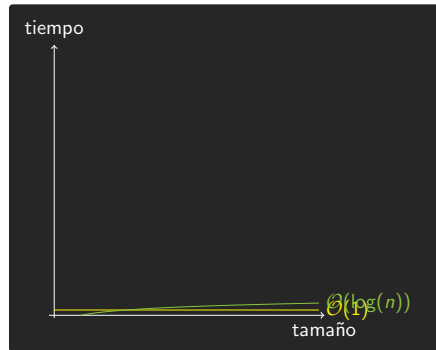


Tiempo. Complejidad Algorítmica

Comportamiento de un programa según crece el tamaño de los datos.

Tamaño del problema n :

- $\mathcal{O}(1)$: tiempo constante.
- $\mathcal{O}(\log(n))$: tiempo logaritmico.

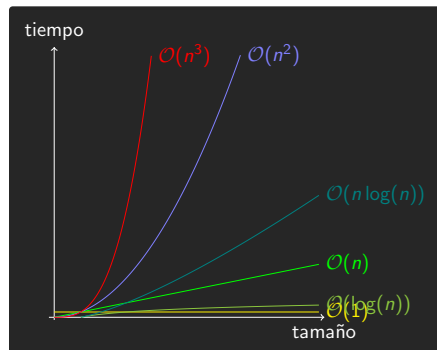


Tiempo. Complejidad Algorítmica

Comportamiento de un programa según crece el tamaño de los datos.

Tamaño del problema n :

- $\mathcal{O}(1)$: tiempo constante.
- $\mathcal{O}(\log(n))$: tiempo logaritmico.
- $\mathcal{O}(n)$: tiempo lineal.
- $\mathcal{O}(n \log(n))$.
- $\mathcal{O}(n^2)$.
- $\mathcal{O}(n^3)$.



Si los datos crecen... ¿cómo reducir tiempos?

- ¿Lenguajes más eficientes?

Si los datos crecen... ¿cómo reducir tiempos?

- ¿Lenguajes más eficientes?
- ¿Mejorar Máquinas?

Si los datos crecen... ¿cómo reducir tiempos?

- ¿Lenguajes más eficientes?

Factor constante 🙄

- ¿Mejorar Máquinas?

Si los datos crecen... ¿cómo reducir tiempos?

- ¿Lenguajes más eficientes?

Factor constante 😞

- ¿Mejorar Máquinas?

Factor constante 😞

Si los datos crecen... ¿cómo reducir tiempos?

- ¿Lenguajes más eficientes?

Factor constante 😞

- ¿Mejorar Máquinas?

Factor constante 😞

- ¿Mejorar Algoritmos?

Si los datos crecen... ¿cómo reducir tiempos?

- ¿Lenguajes más eficientes?
- ¿Mejorar Máquinas?
- ¿Mejorar Algoritmos?

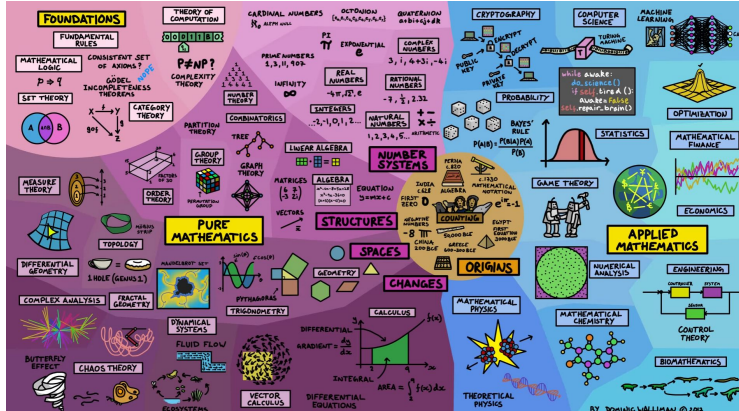
Factor constante



Factor constante

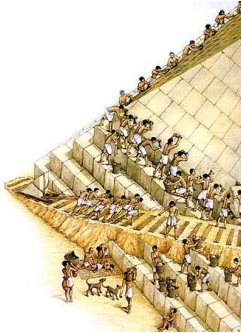


No es fácil

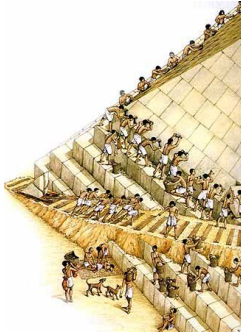


Si los datos crecen... ¿cómo reducir tiempos?

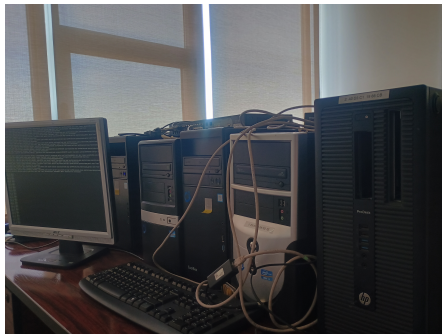
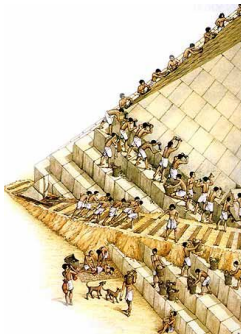
Si los datos crecen... ¿cómo reducir tiempos?



Si los datos crecen... ¿cómo reducir tiempos?



Si los datos crecen... ¿cómo reducir tiempos?



Muchas máquinas 😊 !!

¿Cuánto se puede ganar añadiendo más máquinas?

¿Cuánto se puede ganar añadiendo más máquinas?

- Ley de Amdahl:

La mejora está siempre limitada por la parte no paralelizable.

https://es.wikipedia.org/wiki/Ley_de_Amdahl

¿Cuánto se puede ganar añadiendo más máquinas?

- Ley de Amdahl:

La mejora está siempre limitada por la parte no paralelizable.

https://es.wikipedia.org/wiki/Ley_de_Amdahl

- Ley de Gustafson:

Más recursos permiten siempre resolver problemas más grandes.

https://es.wikipedia.org/wiki/Ley_de_Gustafson

¿Cuántos servidores tiene Google?

- 15 centros de datos reconocidos

¿Cuántos servidores tiene Google?

- 15 centros de datos reconocidos
- Entre 120000 y 200000 servidores en cada centro de datos

¿Cuántos servidores tiene Google?

- 15 centros de datos reconocidos
- Entre 120000 y 200000 servidores en cada centro de datos
- Asumiendo que cada servidor tiene 2Tb de memoria

¿Cuántos servidores tiene Google?

- 15 centros de datos reconocidos
- Entre 120000 y 200000 servidores en cada centro de datos
- Asumiendo que cada servidor tiene 2Tb de memoria
- Cada datacenter tendría entre 240 petabytes y 300 petabytes.

¿Cuántos servidores tiene Google?

- 15 centros de datos reconocidos
- Entre 120000 y 200000 servidores en cada centro de datos
- Asumiendo que cada servidor tiene 2Tb de memoria
- Cada datacenter tendría entre 240 petabytes y 300 petabytes.

<http://what-if.xkcd.com/63/> <https://www.google.com/about/datacenters>

<https://www.youtube.com/watch?v=XZmGGAbHqa0>

<https://www.youtube.com/watch?v=zDAYZU4A3w0>

Retos del Big Data

- Sistema de archivos distribuido

Retos del Big Data

- Sistema de archivos distribuido
- Fácilmente escalable

Retos del Big Data

- Sistema de archivos distribuido
- Fácilmente escalable
- Tolerante a fallos

Retos del Big Data

- Sistema de archivos distribuido
- Fácilmente escalable
- Tolerante a fallos
- Programación distribuida



- Base de datos NoSQL.



- Base de datos NoSQL.
- Almacena *documentos JSON*, mas concretamente en BSON.

```
{  
  'name': {'first_name': 'Luis', 'last_name': 'Llana'},  
  'role': 'Teacher',  
  'email': 'llana@ucm.es',  
  'office': '450',  
  'teaches': ['Big_Data', 'Parallel_Programming']  
}
```



- Base de datos NoSQL.
- Almacena *documentos JSON*, mas concretamente en BSON.

```
{  
  'name': {'first_name': 'Luis', 'last_name': 'Llana'},  
  'role': 'Teacher',  
  'email': 'llana@ucm.es',  
  'office': '450',  
  'teaches': ['Big Data', 'Parallel Programming']  
}
```

- Admite replicación (replica sets).



- Base de datos NoSQL.
- Almacena *documentos JSON*, mas concretamente en BSON.

```
{  
  'name': {'first_name': 'Luis', 'last_name': 'Llana'},  
  'role': 'Teacher',  
  'email': 'llana@ucm.es',  
  'office': '450',  
  'teaches': ['Big Data', 'Parallel Programming']  
}
```

- Admite replicación (replica sets).
- Admite distribución de datos (sharding).



- Base de datos NoSQL.
- Almacena *documentos JSON*, mas concretamente en BSON.

```
{  
  'name': {'first_name': 'Luis', 'last_name': 'Llana'},  
  'role': 'Teacher',  
  'email': 'llana@ucm.es',  
  'office': '450',  
  'teaches': ['Big_Data', 'Parallel_Programming']  
}
```

- Admite replicación (replica sets).
- Admite distribución de datos (sharding).
- *Casi* libre.



- Base de datos NoSQL.
- Almacena *documentos JSON*, mas concretamente en BSON.

```
{  
  'name': {'first_name': 'Luis', 'last_name': 'Llana'},  
  'role': 'Teacher',  
  'email': 'llana@ucm.es',  
  'office': '450',  
  'teaches': ['Big_Data', 'Parallel_Programming']  
}
```

- Admite replicación (replica sets).
- Admite distribución de datos (sharding).
- *Casi* libre.
 - Community version.



- Base de datos NoSQL.
- Almacena *documentos JSON*, mas concretamente en BSON.

```
{  
  'name': {'first_name': 'Luis', 'last_name': 'Llana'},  
  'role': 'Teacher',  
  'email': 'llana@ucm.es',  
  'office': '450',  
  'teaches': ['Big_Data', 'Parallel_Programming']  
}
```

- Admite replicación (replica sets).
- Admite distribución de datos (sharding).
- *Casi* libre.

- Community version.

- Enterprise version.

Cluster en la Facultad



Cluster en la Facultad



Cluster en la Facultad



- Clúster de raspberry de ordenadores antiguos.

Cluster en la Facultad



- Clúster de raspberry de ordenadores antiguos.
- Sirve para acostumbrarse a clusteres reales.

Cluster en la Facultad



- Clúster de raspberry de ordenadores antiguos.
- Sirve para acostumbrarse a clusteres reales.
- Acceso remoto con ssh.