



Introducción

Este proyecto tuvo como objetivo llevar a la práctica conceptos clave de sistemas distribuidos, paralelismo y concurrencia. A través de dos implementaciones un sistema de procesamiento distribuido (hilos vs procesos) y un sistema de almacenamiento distribuido usando MongoDB con sharding se buscó dejar de ver la teoría como algo abstracto y enfrentar los retos reales que aparecen al programar sistemas que trabajan en paralelo o distribuyen información entre nodos.

Parte 1: Sistema de Procesamiento Distribuido

El primer sistema consistió en ejecutar 20 tareas con distintos niveles de complejidad utilizando hilos y procesos, para comparar su desempeño y entender sus diferencias internas.

Implementación

Las tareas simulaban dos tipos de carga:

1. I/O, usando `time.sleep()`
2. Cálculo, con operaciones matemáticas intensivas

Se desarrolló la aplicación con dos enfoques:

- Multithreading (hilos)
- Multiprocessing (procesos)

Control de concurrencia y sincronización

Durante el desarrollo surgió el primer problema real: race conditions.

Al intentar contar las tareas finalizadas desde varios hilos sin control, el contador no llegaba siempre a 20.



Para solucionarlo, fue necesario aplicar mecanismos de sincronización:

Para hilos → Lock

Para procesos → multiprocessing.Value (memoria compartida + lock interno)

El aprendizaje fue claro: la concurrencia sin sincronización siempre conduce a errores sutiles.

Los hilos fueron más rápidos por varias razones:

1. Menor costo de creación comparado con procesos.
2. El GIL no afecta mucho cuando hay tiempo de espera por I/O.
3. Los hilos comparten memoria, los procesos requieren IPC para comunicarse.

¿Cuándo elegir cada uno?

Hilos (Threads)

Ideal para tareas I/O (APIs, DB, archivos)

Bajo consumo de recursos

Comunicación en memoria compartida

Procesos (Processes)

Ideal para cálculos pesados de CPU

Permite verdadero paralelismo en varias CPUs

Cada proceso tiene su propio espacio de memoria

Una conclusión clave fue que elegir entre hilos y procesos no depende del gusto del programador, sino del tipo de carga del sistema.

Parte 2: Sistema de Almacenamiento Distribuido (Sharding)

La segunda parte consistió en implementar un sistema de almacenamiento distribuido usando MongoDB en Docker, simulando un entorno con dos nodos que almacenen los documentos de manera distribuida.



Arquitectura implementada

- 2 nodos de MongoDB en Docker
- Distribución de documentos según hashing de la clave (md5(document_id))
- Función hash determinista para saber a qué nodo enviar cada documento
- Esto permitió simular de forma sencilla el comportamiento básico de bases de datos distribuidas reales.

Estrategia de distribución

Se aplicó hash-based sharding:

- Cada documento se almacena en un nodo dependiendo del valor del hash de su ID.
- La distribución resultó parejas: 51% / 49%

Este resultado confirmó que el hashing genera balance sin necesidad de una lógica compleja.

Proceso de búsqueda distribuida

Se utilizó una estrategia en dos fases:

1. Buscar en el nodo donde *debería* estar el documento según el hash.
2. Si no aparece, buscar en los demás nodos.

Esto permitió lograr un balance entre eficiencia y robustez.

Limitaciones detectadas

El sistema funciona, pero carece de características propias de un sistema de producción:

- No hay replicación: si un nodo cae, los datos se pierden
- No hay rebalanceo automático si se agrega un nodo



- No existe tolerancia a fallos (sin failover)

Esto evidenció uno de los principales aprendizajes:

En sistemas distribuidos, la complejidad no está en guardar datos, sino en mantenerlos disponibles y consistentes.

Lo aprendido (aspectos técnicos y profesionales)

Principales aprendizajes técnicos

La concurrencia sin sincronización genera race conditions

Hilos y procesos tienen usos diferentes Sharding distribuye carga y mejora escalabilidad Docker

facilita la simulación de entornos distribuidos reales

Habilidades desarrolladas

- Análisis de trade-offs: rendimiento vs robustez, simplicidad vs funcionalidad
- Diseño orientado a fallos
- Uso práctico de herramientas modernas: Docker, MongoDB



Evidencias

```
joshua@LQ:~$ mkdir exam-sistemas-distribuidos
joshua@LQ:~$ cd exam-sistemas-distribuidos
joshua@LQ:~/exam-sistemas-distribuidos$ mkdir -p part1-processes-threads
joshua@LQ:~/exam-sistemas-distribuidos$ mkdir -p part2-distributed-storage
joshua@LQ:~/exam-sistemas-distribuidos$ mkdir -p docs
joshua@LQ:~/exam-sistemas-distribuidos$ git init
hint: Using 'master' as the name for the initial branch. This default branch name
hint: is subject to change. To configure the initial branch name to use in all
hint: of your new repositories, which will suppress this warning, call:
hint:
hint:   git config --global init.defaultBranch <name>
hint:
hint: Names commonly chosen instead of 'master' are 'main', 'trunk' and
hint: 'development'. The just-created branch can be renamed via this command:
hint:
hint:   git branch -m <name>
Initialized empty Git repository in /home/joshua/exam-sistemas-distribuidos/.git/
joshua@LQ:~/exam-sistemas-distribuidos$ touch README.md
joshua@LQ:~/exam-sistemas-distribuidos$ touch docker-compose.yml
joshua@LQ:~/exam-sistemas-distribuidos$ touch part1-processes-threads/task_processor.py
joshua@LQ:~/exam-sistemas-distribuidos$ touch part1-processes-threads/results_analysis.md
joshua@LQ:~/exam-sistemas-distribuidos$ touch part2-distributed-storage/storage_system.py
joshua@LQ:~/exam-sistemas-distribuidos$ touch part2-distributed-storage/mongo_setup.md
joshua@LQ:~/exam-sistemas-distribuidos$ touch docs/final_report.pdf
joshua@LQ:~/exam-sistemas-distribuidos$ ls
README.md      part1-processes-threads
docker-compose.yml  part2-distributed-storage
docs
joshua@LQ:~/exam-sistemas-distribuidos$ cd part1-processes-threads
joshua@LQ:~/exam-sistemas-distribuidos/part1-processes-threads$ ls
results_analysis.md  task_processor.py
joshua@LQ:~/exam-sistemas-distribuidos/part1-processes-threads$ cd task_processor.py
-bash: cd: task_processor.py: Not a directory
joshua@LQ:~/exam-sistemas-distribuidos/part1-processes-threads$ nano task_processor.py
joshua@LQ:~/exam-sistemas-distribuidos/part1-processes-threads$ nano task_processor.py
joshua@LQ:~/exam-sistemas-distribuidos/part1-processes-threads$ nano task_processor.py
joshua@LQ:~/exam-sistemas-distribuidos/part1-processes-threads$ nano task_processor.py
joshua@LQ:~/exam-sistemas-distribuidos/part1-processes-threads$ ls
```



```
joshua@L0Q:~$ cd exam-sistemas-distribuidos
joshua@L0Q:~/exam-sistemas-distribuidos$ ls
README.md      part1-processes-threads
docker-compose.yml  part2-distributed-storage
docs
joshua@L0Q:~/exam-sistemas-distribuidos$ cd docker-compose.yml
-bash: cd: docker-compose.yml: Not a directory
joshua@L0Q:~/exam-sistemas-distribuidos$ nano docker-compose.yml
joshua@L0Q:~/exam-sistemas-distribuidos$ ls
README.md      part1-processes-threads
docker-compose.yml  part2-distributed-storage
docs
joshua@L0Q:~/exam-sistemas-distribuidos$ cd part2-distributed-storage
joshua@L0Q:~/exam-sistemas-distribuidos/part2-distributed-storage$ ls
mongo_setup.md  storage_system.py
joshua@L0Q:~/exam-sistemas-distribuidos/part2-distributed-storage$ nano storage_system.py
joshua@L0Q:~/exam-sistemas-distribuidos/part2-distributed-storage$ nano mongo_setup.md
joshua@L0Q:~/exam-sistemas-distribuidos/part2-distributed-storage$ docker --version
Docker version 28.5.1, build e180ab8
joshua@L0Q:~/exam-sistemas-distribuidos/part2-distributed-storage$ docker-compose --version
Docker Compose version v2.40.0-desktop.1
joshua@L0Q:~/exam-sistemas-distribuidos/part2-distributed-storage$ pipinstall pymongo
pipinstall: command not found
joshua@L0Q:~/exam-sistemas-distribuidos/part2-distributed-storage$ pip3 install pymongo
Command 'pip3' not found, but can be installed with:
sudo apt install python3-pip
joshua@L0Q:~/exam-sistemas-distribuidos/part2-distributed-storage$ pip3 --version
Command 'pip3' not found, but can be installed with:
sudo apt install python3-pip
joshua@L0Q:~/exam-sistemas-distribuidos/part2-distributed-storage$ cd
joshua@L0Q:~$ cd exam-sistemas-distribuidos
joshua@L0Q:~/exam-sistemas-distribuidos$ ls
README.md      docs      part2-distributed-storage
docker-compose.yml  part1-processes-threads
```



```
docker-compose.yml part2-processes-storage
joshua@LQ:~/exam-sistemas-distribuidos$ docker-compose up -d
WARN[0000] /home/joshua/exam-sistemas-distribuidos/docker-compose.yml: the attribute 'version' is obsolete, it will be ignored
tential confusion
[+] Running 9/9
  ✓ mongodb1 Pulled                                23.6s
  ✓ mongodb2 Pulled                                23.6s
  ✓ 4e898e1375f7 Pull complete                      4.0s
  ✓ 02613098a4eb Pull complete                      6.8s
  ✓ f8a6564ff91f Pull complete                      3.8s
  ✓ d91d51de2f11 Pull complete                      6.7s
  ✓ 907d967ea28b Pull complete                     19.6s
  ✓ 9d9d86f60485 Pull complete                      5.7s
  ✓ 0f1c32d68638 Pull complete                      6.4s
[+] Running 5/5
  ✓ Network exam-sistemas-distribuidos_distributed_network Created 0.1s
  ✓ Volume exam-sistemas-distribuidos_mongodb1_data Created 0.0s
  ✓ Volume exam-sistemas-distribuidos_mongodb2_data Created 0.0s
  ✓ Container mongodb_node1 Started 1.5s
  ✓ Container mongodb_node2 Started 1.5s
joshua@LQ:~/exam-sistemas-distribuidos$ docker ps
CONTAINER ID   IMAGE      COMMAND                  CREATED        STATUS        PORTS
df9dbe09ccd2   mongo:7.0  "docker-entrypoint.s..." 19 seconds ago Up 18 seconds (healthy) 0.0.0.0:27018->27017/tcp, [::]:2
e2
18c4fd4c8ebf   mongo:7.0  "docker-entrypoint.s..." 19 seconds ago Up 18 seconds (healthy) 0.0.0.0:27017->27017/tcp, [::]:2
e1
joshua@LQ:~/exam-sistemas-distribuidos$ docker exec mongodb_node1 mongosh --eval "db.adminCommand('ping')"
{ ok: 1 }
joshua@LQ:~/exam-sistemas-distribuidos$ docker exec mongodb_node2 mongosh --eval "db.adminCommand('ping')"
{ ok: 1 }
```

```
joshua@LQ:~/exam-sistemas-distribuidos/part2-distributed-storage$ docker exec -it mongodb_node1 mongosh
Current Mongosh Log ID: 6913f71f25716f0ad7ce5f46
Connecting to:      mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000&appName=mongosh
Using MongoDB:      7.0.25
Using Mongosh:      2.5.8

For mongosh info see: https://www.mongodb.com/docs/mongosh-shell/

To help improve our products, anonymous usage data is collected and sent to MongoDB periodically (https://www.mongodb.com/telemetry-policy).
You can opt-out by running the disableTelemetry() command.

-----
The server generated these startup warnings when booting
2025-11-12T02:51:19.558+00:00: Using the XFS filesystem is strongly recommended with the WiredTiger storage engine.
    See https://www.mongodb.com/docs/mongodb-7.0/notes/xfs-fs/#xfs-fs for more information.
2025-11-12T02:51:19.945+00:00: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted.
2025-11-12T02:51:19.945+00:00: vm.max_map_count is too low
-----
```



```
joshua@LQ:~/exam-sistemas-distribuidos/part2-distributed-storage$ docker exec -it mongodb_node2 mongosh
Current Mongosh Log ID: 6913f72e628edfdc32ce5f46
Connecting to: mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000&appName=mongosh+2.5.8
Using MongoDB: 7.0.25
Using Mongosh: 2.5.8

For mongosh info see: https://www.mongodb.com/docs/mongodb-shell/

To help improve our products, anonymous usage data is collected and sent to MongoDB periodically (https://www.mongodb.com/legal/privacy-policy).
You can opt-out by running the disableTelemetry() command.

-----
The server generated these startup warnings when booting
2025-11-12T02:51:19.557+00:00: Using the XFS filesystem is strongly recommended with the WiredTiger storage engine. See http://dochub.mongodb.org/core/prodnotes-filesystem
2025-11-12T02:51:19.953+00:00: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted
2025-11-12T02:51:19.953+00:00: vm.max_map_count is too low
-----
```

```
joshua@LQ:~/exam-sistemas-distribuidos/part1-processes-threads$ python3 tas
k_processor.py
=====
SISTEMA DE PROCESAMIENTO DISTRIBUIDO
Comparación: Procesos vs Hilos
=====

Tareas generadas:
Task 1: Dificultad 1
Task 2: Dificultad 1
Task 3: Dificultad 3
Task 4: Dificultad 2
Task 5: Dificultad 2
Task 6: Dificultad 2
Task 7: Dificultad 1
Task 8: Dificultad 5
Task 9: Dificultad 1
Task 10: Dificultad 5
Task 11: Dificultad 4
Task 12: Dificultad 1
Task 13: Dificultad 1
Task 14: Dificultad 1
Task 15: Dificultad 2
Task 16: Dificultad 2
Task 17: Dificultad 5
Task 18: Dificultad 5
Task 19: Dificultad 1
Task 20: Dificultad 5
```




EJECUTANDO CON HILOS (THREADS)

```
[THREAD] Task 1 completed with difficulty 1, result: 332833500 | Total completed: 1
[THREAD] Task 2 completed with difficulty 1, result: 332833500 | Total completed: 2
[THREAD] Task 7 completed with difficulty 1, result: 332833500 | Total completed: 3
[THREAD] Task 12 completed with difficulty 1, result: 332833500 | Total completed: 4
[THREAD] Task 9 completed with difficulty 1, result: 332833500 | Total completed: 5
[THREAD] Task 14 completed with difficulty 1, result: 332833500 | Total completed: 6
[THREAD] Task 13 completed with difficulty 1, result: 332833500 | Total completed: 7
[THREAD] Task 19 completed with difficulty 1, result: 332833500 | Total completed: 8
[THREAD] Task 4 completed with difficulty 2, result: 2664667000 | Total completed: 9
[THREAD] Task 6 completed with difficulty 2, result: 2664667000 | Total completed: 10
[THREAD] Task 5 completed with difficulty 2, result: 2664667000 | Total completed: 11
[THREAD] Task 15 completed with difficulty 2, result: 2664667000 | Total completed: 12
[THREAD] Task 16 completed with difficulty 2, result: 2664667000 | Total completed: 13
[THREAD] Task 3 completed with difficulty 3, result: 8995500500 | Total completed: 14
[THREAD] Task 11 completed with difficulty 4, result: 21325334000 | Total completed: 15
[THREAD] Task 8 completed with difficulty 5, result: 41654167500 | Total completed: 16
[THREAD] Task 10 completed with difficulty 5, result: 41654167500 | Total completed: 17
[THREAD] Task 18 completed with difficulty 5, result: 41654167500 | Total completed: 18
[THREAD] Task 20 completed with difficulty 5, result: 41654167500 | Total completed: 19
[THREAD] Task 17 completed with difficulty 5, result: 41654167500 | Total completed: 20
```

✓ Todas las tareas completadas con hilos

🕒 Tiempo total: 0.52 segundos

📊 Tareas completadas: 20



EJECUTANDO CON PROCESOS (PROCESSES)

```
[PROCESS] Task 1 completed with difficulty 1, result: 332833500 | Total completed: 1
[PROCESS] Task 2 completed with difficulty 1, result: 332833500 | Total completed: 2
[PROCESS] Task 7 completed with difficulty 1, result: 332833500 | Total completed: 3
[PROCESS] Task 9 completed with difficulty 1, result: 332833500 | Total completed: 4
[PROCESS] Task 12 completed with difficulty 1, result: 332833500 | Total completed: 5
[PROCESS] Task 14 completed with difficulty 1, result: 332833500 | Total completed: 6
[PROCESS] Task 13 completed with difficulty 1, result: 332833500 | Total completed: 7
[PROCESS] Task 19 completed with difficulty 1, result: 332833500 | Total completed: 8
[PROCESS] Task 4 completed with difficulty 2, result: 2664667000 | Total completed: 9
[PROCESS] Task 5 completed with difficulty 2, result: 2664667000 | Total completed: 10
[PROCESS] Task 6 completed with difficulty 2, result: 2664667000 | Total completed: 11
[PROCESS] Task 15 completed with difficulty 2, result: 2664667000 | Total completed: 12
[PROCESS] Task 16 completed with difficulty 2, result: 2664667000 | Total completed: 13
[PROCESS] Task 3 completed with difficulty 3, result: 8995500500 | Total completed: 14
[PROCESS] Task 11 completed with difficulty 4, result: 21325334000 | Total completed: 15
[PROCESS] Task 8 completed with difficulty 5, result: 41654167500 | Total completed: 16
[PROCESS] Task 10 completed with difficulty 5, result: 41654167500 | Total completed: 17
[PROCESS] Task 17 completed with difficulty 5, result: 41654167500 | Total completed: 18
[PROCESS] Task 18 completed with difficulty 5, result: 41654167500 | Total completed: 19
[PROCESS] Task 20 completed with difficulty 5, result: 41654167500 | Total completed: 20
```

✓ Todas las tareas completadas con procesos
⌚ Tiempo total: 0.53 segundos
📊 Tareas completadas: 20

✓ Todas las tareas completadas con procesos
⌚ Tiempo total: 0.53 segundos
📊 Tareas completadas: 20

ANÁLISIS DE RESULTADOS

```
⌚ Tiempo con hilos: 0.52 segundos
⌚ Tiempo con procesos: 0.53 segundos
📊 Diferencia: 0.02 segundos
🏆 Los hilos fueron 1.03x más rápidos
💡 Razón: Menor overhead de creación y comunicación
```



```
joshua@LOQ:~/exam-sistemas-distribuidos/part2-distributed-storage$ python3 s
torage_system.py
=====
SISTEMA DE ALMACENAMIENTO DISTRIBUIDO
MongoDB con 2 nodos
=====

🔑 Conectando a nodos MongoDB...
✅ Nodo 1 conectado: mongodb://localhost:27017/
✅ Nodo 2 conectado: mongodb://localhost:27018/

✅ Sistema inicializado con 2 nodos

🗑 Limpiando todos los datos...
Nodo 1: 0 documentos eliminados
Nodo 2: 0 documentos eliminados
✅ Datos limpiados

📄 Generando 100 documentos de ejemplo...

📁 Insertando documentos en el sistema distribuido...
📄 Documento doc_0000 → Nodo 2
📄 Documento doc_0001 → Nodo 1
📄 Documento doc_0002 → Nodo 1
📄 Documento doc_0003 → Nodo 2
📄 Documento doc_0004 → Nodo 2
📄 Documento doc_0005 → Nodo 1
📄 Documento doc_0006 → Nodo 1
📄 Documento doc_0007 → Nodo 1
📄 Documento doc_0008 → Nodo 1
📄 Documento doc_0009 → Nodo 2
📄 Documento doc_0010 → Nodo 2
📄 Documento doc_0011 → Nodo 2
📄 Documento doc_0012 → Nodo 1
📄 Documento doc_0013 → Nodo 1
📄 Documento doc_0014 → Nodo 1
📄 Documento doc_0015 → Nodo 2
```



✓ 100 documentos insertados

=====

🇩🇪 ESTADÍSTICAS DE DISTRIBUCIÓN

=====

📦 Total de documentos: 100
💻 Número de nodos: 2

Nodo 1: mongodb://localhost:27017/

📄 Documentos: 48
🇩🇪 Porcentaje: 48.0%



Nodo 2: mongodb://localhost:27018/

📄 Documentos: 52
🇩🇪 Porcentaje: 52.0%



⚖ Balance de distribución:
✓ Excelente (diferencia: 4.00%)

=====



```
=====
🔍 PRUEBA DE BÚSQUEDA DISTRIBUIDA
=====

🔍 Buscando documento: doc_0000
  → Buscando en nodo esperado 2...
  ✅ Encontrado en nodo 2 (como se esperaba)

  📄 Documento encontrado:
    ID: doc_0000
    Título: Documento 0
    Categoría: arte
    Nodo: 2

🔍 Buscando documento: doc_0050
  → Buscando en nodo esperado 2...
  ✅ Encontrado en nodo 2 (como se esperaba)

  📄 Documento encontrado:
    ID: doc_0050
    Título: Documento 50
    Categoría: arte
    Nodo: 2

🔍 Buscando documento: doc_0099
  → Buscando en nodo esperado 1...
  ✅ Encontrado en nodo 1 (como se esperaba)

  📄 Documento encontrado:
    ID: doc_0099
    Título: Documento 99
    Categoría: ciencia
    Nodo: 1
```