



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI



comillas.edu

Universidad Pontificia Comillas ICAI

Analítica Social y de la Web *Social and Web Analytics*

2.3 Almacenamiento en MongoDB

Javier Ruiz de Ojeda

Curso 2023-24
Segundo semestre





COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

MongoDB



MongoDB

Ventajas de MongoDB



- **Diseño no-rígido (*schema-less*)**: A diferencia de las bases de datos relacionales tradicionales, MongoDB proporciona un modelo basado en documentos (estructuras JSON) y colecciones (grupos de documentos).
- **Alto rendimiento**: Las funciones de indexado permiten acceso y almacenamiento muy rápidos.
- **Alta disponibilidad**: Las funciones de respaldo permiten crear copias en nodos distintos para garantizar la disponibilidad.
- **Escalado automático**: La función de “Sharding” permite escalar grandes conjuntos de datos automáticamente (<https://docs.mongodb.com/v3.0/sharding/>).

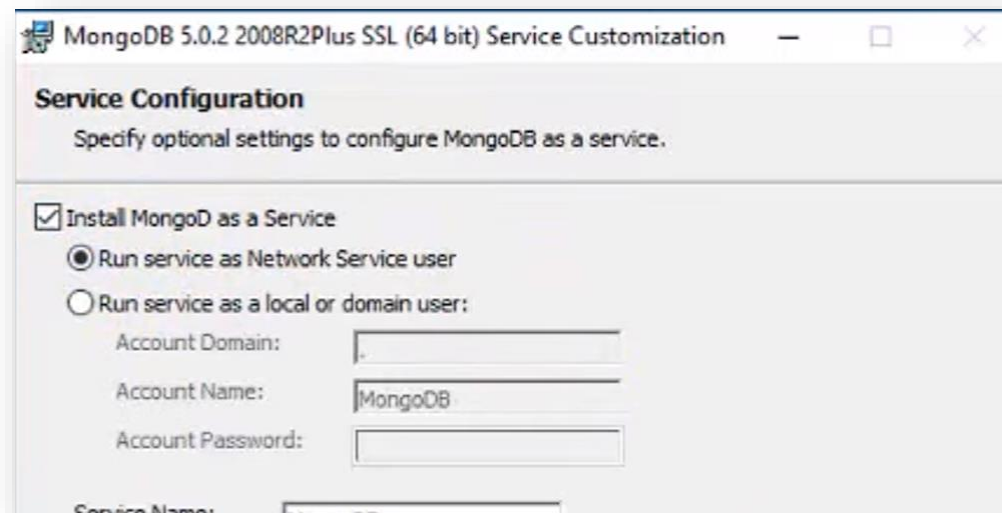
MongoDB

Instalación



- MongoDB puede descargarse aquí:
<https://www.mongodb.com/try/download/community>



Quitar para
optimizar
rendimiento









MongoDB

Configuración y arranque

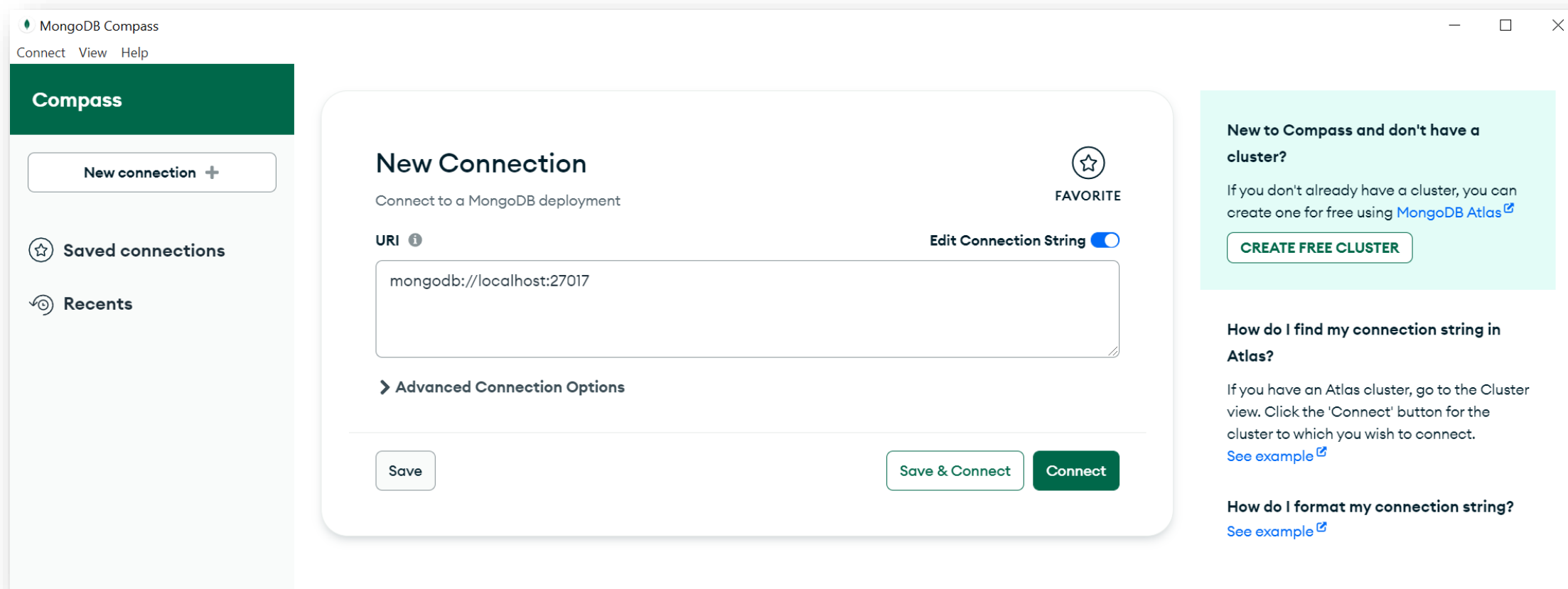


- Se encuentra en:  > Este equipo > Disco local (C:) > Archivos de programa > MongoDB > Server > 6.0 > bin
- Crear:  > Este equipo > Disco local (C:) > data > db

Nombre	Fecha de modificación	Tipo	Tamaño
 InstallCompass.ps1	17/01/2023 15:22	Script de Windows P...	2 KB
 mongod.cfg	17/01/2023 15:22	Archivo CFG	1 KB
 mongod.exe	17/01/2023 16:43	Aplicación	54.979 KB
 mongod.pdb	17/01/2023 16:43	Documento PDB	824.532 KB
 mongos.exe	17/01/2023 16:08	Aplicación	33.976 KB
 mongos.pdb	17/01/2023 16:08	Documento PDB	460.212 KB

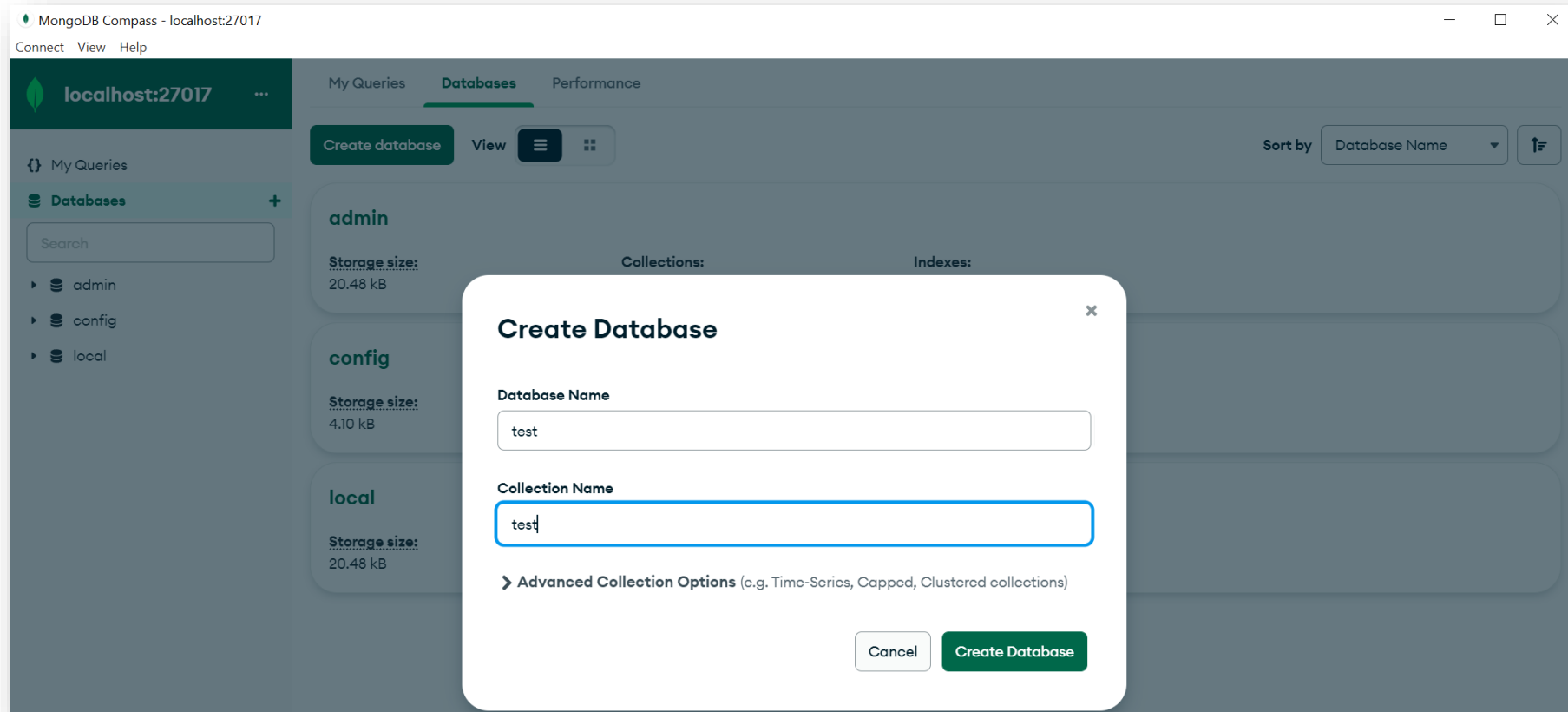
MongoDB

Conexión desde Compass



MongoDB

Conexión desde Compass



MongoDB

Conexión con Python



- Instalamos el módulo PyMongo: `pip install pymongo`
- Importamos y conectamos al servidor (levantado):

```
from pymongo import MongoClient  
client = MongoClient ('localhost:27017')
```

- Accedemos a la colección `articles` en el scrapper de BD:

```
db_scrapper = db.scrapper  
collection_articles = db_scrapper.articles
```


MongoDB

Insertar datos



#Insertar documento en colección

```
docs = []
```

```
for _ in range (0,10):
```

```
    docs.append({
```

```
        "author" = "...",
```

```
        "content" = "...",
```

```
        "comment" = ["...", ... ]
```

```
    })
```

```
db.collection.insert_many(docs) #todos a la vez
```

```
for doc in docs: #de uno en uno
```

```
    db.collection.insert_one(doc)
```

MongoDB

Encontrar datos



#La función `find()` devuelve un cursor, así que iteramos sobre él para obtener los propios datos

```
docs = [d for d in db.collection.find()]
```

#para cogerlos de 100 en 100

```
batch_size = 100
```

```
iteration = 0
```

```
count = db.collection.count() #contamos el número de documentos
```

```
while iteration * batch_size < count:
```

```
    docs = [d for d in db.collection.find().skip(batch_size *  
iteration).limit(batch_size)]
```

```
    iteration += 1
```

```
query = { 'author': 'Tolkien' }
```

```
docs = [d for d in db.collection.find(query)]
```

MongoDB

Actualizar datos



#Para publicar los documentos cuyo autor sea Tolkien

```
query_search = { 'author': 'Tolkien' }
```

```
query_update = { '$set': { 'published': True } }
```

```
db.collection.update_many(query_search, query_update)
```

#O podemos actualizar solo el primero que encaje

```
db.collection.update_one(query_search, query_update)
```

MongoDB

Eliminar datos y colecciones



#Para eliminar los documentos cuyo autor sea Tolkien

```
query_search = { 'author': 'Tolkien' }
```

```
db.collection.delete_many(query_search)
```

#O podemos eliminar solo el primero que encaje

```
db.collection.delete_one(query_search)
```

#Y para eliminar la colección entera

```
db.collection.drop()
```

```
db.dropDatabase()
```

MongoDB

Ejemplo Facebook



```
import facebook_scraper as fb
from pymongo import MongoClient, errors

client = MongoClient ('localhost:27017')

db = client["scraping"]
db_fb = db["facebook"]
```

MongoDB

Ejemplo Facebook



Insert:

```
for post in fb.get_posts("google", pages=1):
    post_id = post.pop("post_id")
    record = {**post, "_id": post_id}
    try:
        db_fb.insert_one(record)
        print(f"Post inserted: {post_id}")
    except errors.DuplicateKeyError:
        db_fb.update_one(
            {"_id": post_id},
            {"$set": record}
        )
        print(f"{post_id} is already in our DB")
```

MongoDB

Ejemplo Facebook



Get:

```
db.facebook.find_one({"post_id": post_id})
```

Delete collection:

```
db.facebook.drop_collection()
```



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

Ejercicio por parejas:

*Extraer información de datos de
IMDB almacenados en MongoDB*



Ejercicio

Presentación para los Óscar



Tras la publicación de las nominaciones de los *Óscar*, *Oppenheimer* parece el favorito indiscutible. Los productores de cada competidora os han contratado como consultores para ayudarles a evitarlo:

1. Almacenar las 250 reseñas más relevantes de la película en IMDB en una base de datos
2. Proponer 3 estrategias de marketing basadas en datos obtenidos de esas reseñas para incrementar la popularidad de la película
3. Preparar una presentación ejecutiva con Proceso de Obtención, Análisis de Datos, Limpieza, Visualizaciones, y Estrategia de actuación (conclusiones)

Presentación en clase el miércoles 14 de febrero.

Los que se quedan
Mark Johnson
American Fiction
Cord Jefferson, Ben LeClair, Nikos Karamigios, ...
The Zone of Interest
James Wilson
Barbie
Margot Robbie, Tom Ackerley, David Heyman, ...
Oppenheimer
Christopher Nolan, Emma Thomas, Charles Roven
Pobres criaturas
Emma Stone, Giórgos Lánthimos, Ed Guiney, ...
Past Lives
Christine Vachon, David Hinojosa, Pamela Koffler
Anatomía de una caída
Marie-Ange Luciani, David Thion
Maestro
Bradley Cooper, Steven Spielberg, Amy Durning, ...
Los asesinos de la luna
Martin Scorsese, Dan Friedkin, Bradley Thomas, ...



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

¡Muchas gracias!

