



GESTIÓN DE DATOS: COMPARACIÓN

RIAH



8 DE OCTUBRE DE 2024

GRUPO AIR
Manuel Perales Briones

0. CONCEPTOS BÁSICOS

IMPORTANTE: las tecnologías IBM, debido a sus limitaciones en su versión gratuita, no se tomarán en consideración.

Hemos decidido emplear un esquema relacional para la mayor parte de la BBDD, pero aún quedan dudas acerca de cómo manejar los archivos JSON que el proyecto necesita para gestionar grandes volúmenes de datos relativos al tracking. Por lo tanto, se buscarán diferentes SGDBs y tecnologías de análisis de datos para los cuales se determinará en orden de prioridad:

- Tipo de tecnología (SGDB, data analyzer...)
- Accesibilidad: qué tan sencillo es acceder a las funcionalidades de la herramienta (tarifas, disponibilidad al público, etc.).
- Compatibilidad: qué tan bien se integra esta herramienta con otras que pueden utilizarse en el proyecto.
- Usabilidad: qué tanta eficiencia tiene la interfaz tanto para el aprendizaje como para la realización de operaciones.
- Experiencia: nivel del desarrollador (yo) en dicha herramienta (las herramientas que no he utilizado previamente tienen la puntuación mínima).

Para las SGDB incluimos:

- Ajustes de seguridad: qué tan bien integra sistemas de seguridad el SGDB, dado que es un requisito indispensable.
- Volumen: cómo gestiona las grandes cantidades de datos necesarias.

Se valorará cada aspecto mediante un medidor personal en base a lo encontrado;

MUY MAL
No tiene ninguno de
los parámetros
deseados

MAL
Tiene muy pocos
parámetros deseados

NO ESTÁ MAL
Los parámetros más
importantes los cumple

BIEN
Cumple casi todos los
parámetros deseados

MUY BIEN
Cumple todos los
parámetros requeridos

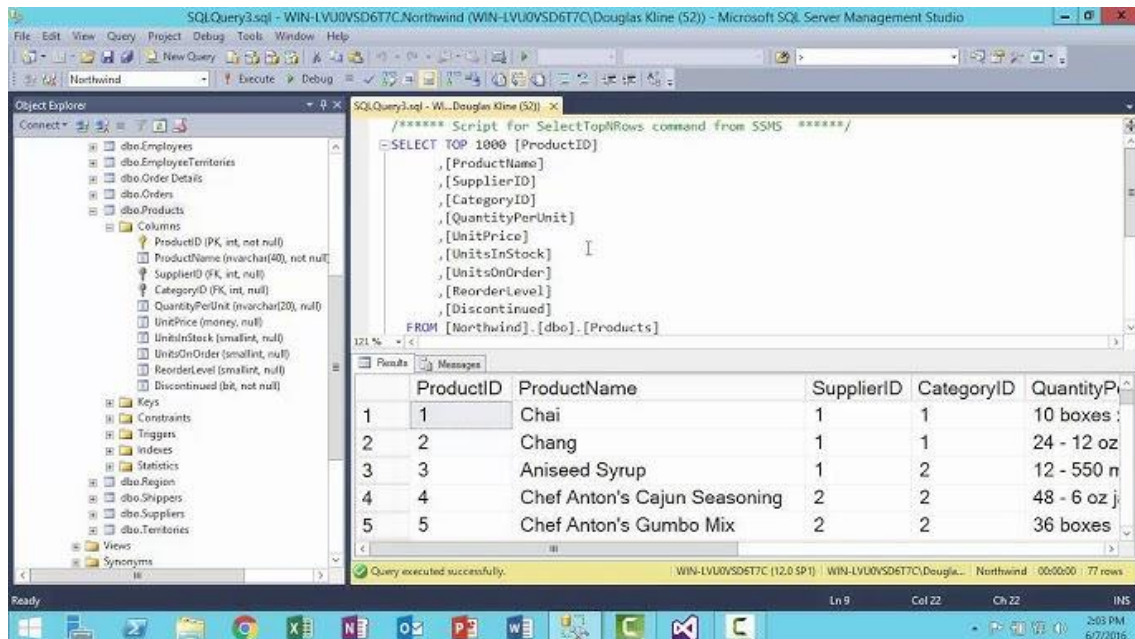


<https://www.ionos.es/digitalguide/hosting/cuestiones-tecnicas/sistema-gestor-de-base-de-datos-sgbd/>

<https://blog.hubspot.es/sales/sistemas-gestores-bases-de-datos>

1. LISTA

SQL Server (SGDB)

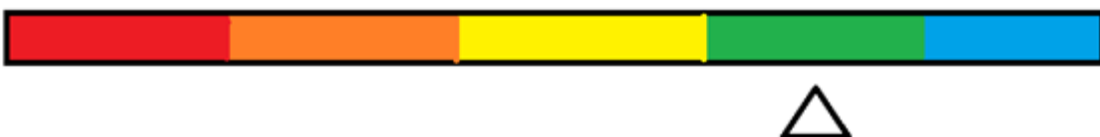


- Accesibilidad



Tanto la versión Developer como la versión Express son gratuitas.

- Compatibilidad



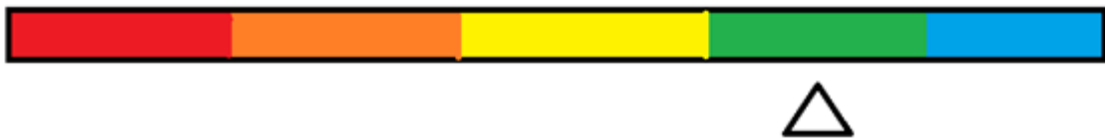
Mediante el JDBC Driver que integra la herramienta, se puede realizar cualquier consulta a la BBDD desde cualquier Backend que soporte JDBC (pero hay que descargarlo adrede). Además, permite almacenar JSONs, operaciones relacionadas con ellos y gestión de roles.

- Usabilidad



Posee una interfaz gráfica similar a otros SGDB, aunque la numerosa cantidad de opciones al inicio puede ser abrumadora de la manera que se presenta, además de que algunas herramientas como las auditorías no sean fácilmente configurables y accesibles.

- Experiencia



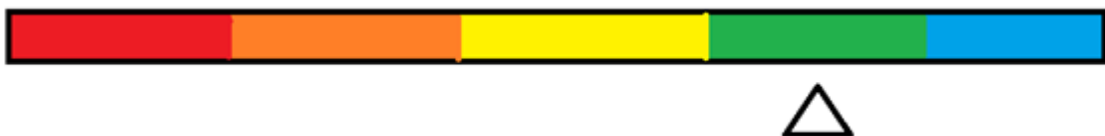
Ya se han realizado previamente multitud de configuraciones de BBDDs en ella, incontables queries, configuración de conexión, etc. Faltaría aprender algunos detalles como la activación de auditorías y el dominio de sistemas de seguridad.

- Ajustes de seguridad



Máscaras, hashes, conexiones cifradas, auditorías, encriptados, etc. Posee todas las funcionalidades de seguridad deseadas.

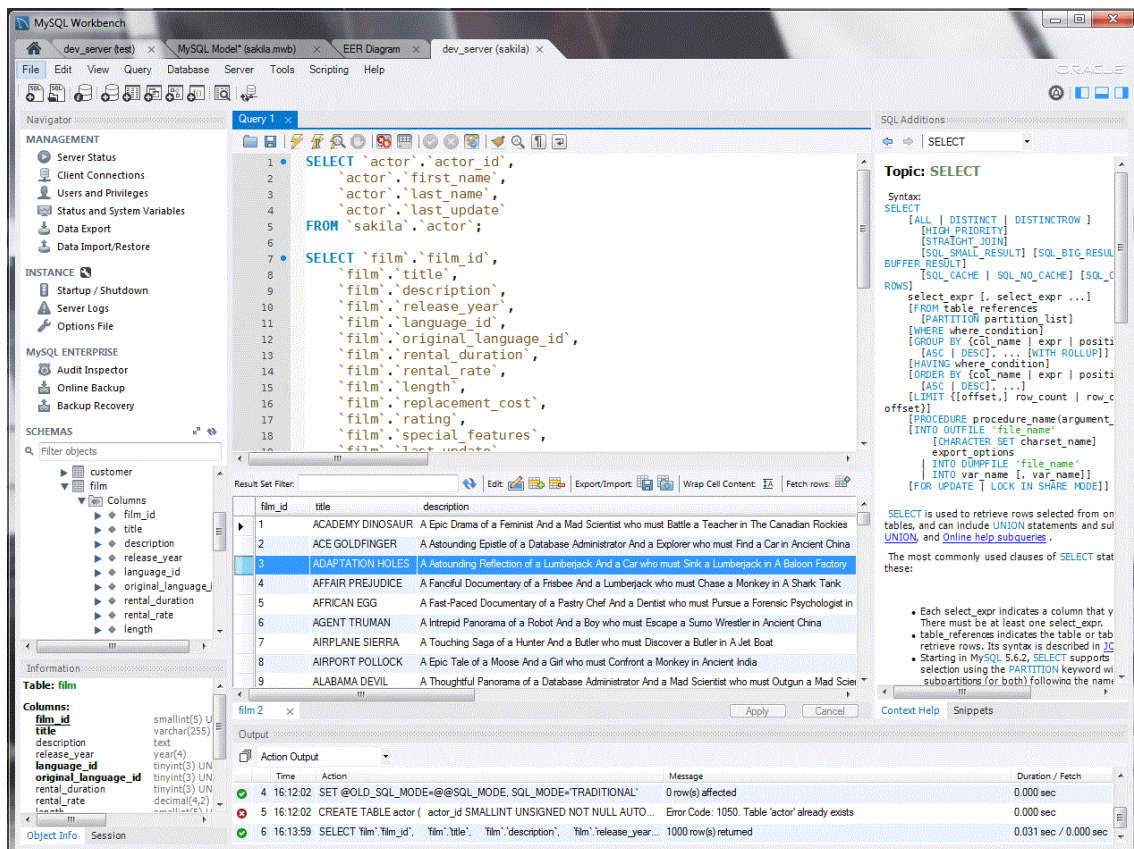
- Volumen



Puede gestionar grandes volúmenes de forma relacional, pero puede dar problemas al trabajar con ficheros JSON grandes. Integra funciones de tuning.

MUY BUENA OPCIÓN

MySQL



● Accesibilidad



Descarga gratuita desde Windows.

● Compatibilidad



Se puede realizar cualquier consulta a la BBDD desde cualquier Backend que soporte JDBC. Además, permite almacenar JSONs, operaciones relacionadas con ellos y gestión de roles.

● Usabilidad



La presentación, tanto de conexión como de gestión, son completamente intuitivas. Además, maneja a primera vista aspectos como jerarquía de BBDD y estado de la misma (auditoría).

- Experiencia



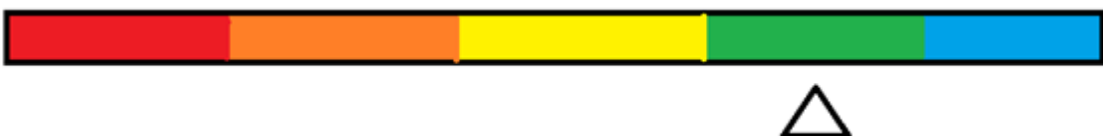
Aunque no se ha trabajado con ninguna de las herramientas que integra, se han realizado conexiones, queries y creación de BBDD con anterioridad en ella.

- Ajustes de seguridad



Máscaras, hashes, conexiones cifradas, auditorías, encriptados, etc. Posee todas las funcionalidades de seguridad deseadas.

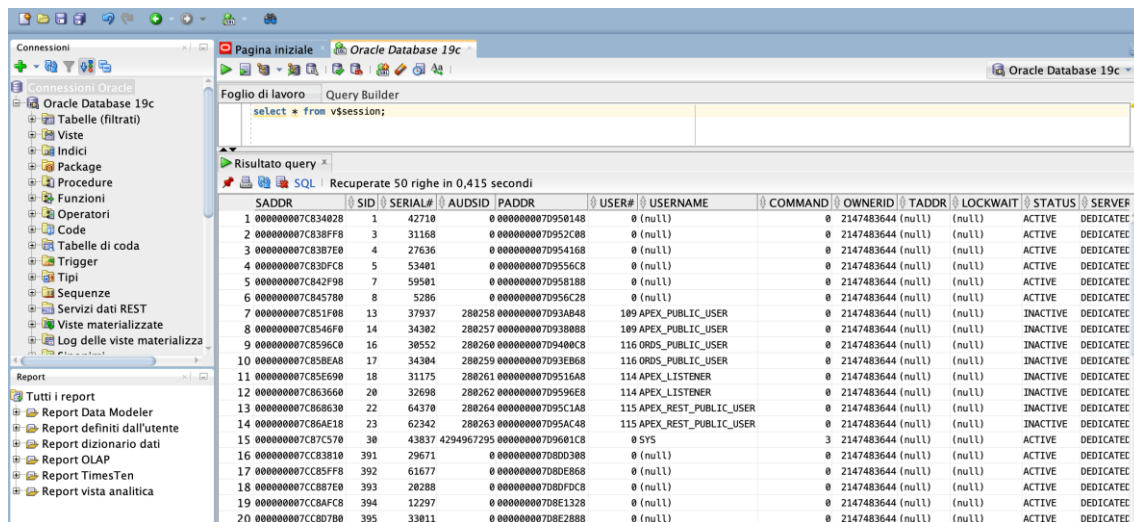
- Volumen



Puede gestionar grandes volúmenes de forma relacional, pero puede dar problemas al trabajar con ficheros JSON grandes. Integra funciones de tuning.

MUY BUENA OPCIÓN

Oracle Database (SGDB)



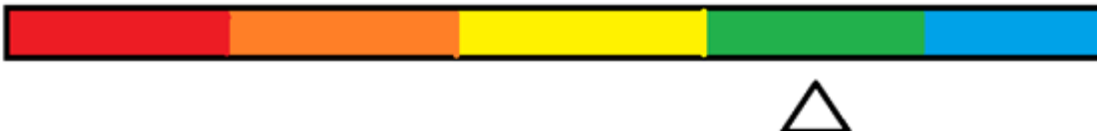
SADDR	SID	SERIAL#	AUDSID	PADDR	USER#	USERNAME	COMMAND	OWNERID	TADDR	LOCKWAIT	STATUS	SERVER
1 000000007C834028	1	42710	0 000000007D950148	0 (null)	0	(null)	0	2147483644 (null)	(null)	(null)	ACTIVE	DEDICATEE
2 000000007C838FF8	3	31168	0 000000007D952C08	0 (null)	0	(null)	0	2147483644 (null)	(null)	(null)	ACTIVE	DEDICATEE
3 000000007C838FE0	4	27636	0 000000007D954168	0 (null)	0	(null)	0	2147483644 (null)	(null)	(null)	ACTIVE	DEDICATEE
4 000000007C83DFC8	5	53401	0 000000007D9556C8	0 (null)	0	(null)	0	2147483644 (null)	(null)	(null)	ACTIVE	DEDICATEE
5 000000007C842F98	7	59581	0 000000007D958188	0 (null)	0	(null)	0	2147483644 (null)	(null)	(null)	ACTIVE	DEDICATEE
6 000000007C845780	8	5286	0 000000007D956C28	0 (null)	0	(null)	0	2147483644 (null)	(null)	(null)	ACTIVE	DEDICATEE
7 000000007C851F08	13	37937	280258 000000007D93A048	109 APEX_PUBLIC_USER	109	APEX_PUBLIC_USER	0	2147483644 (null)	(null)	(null)	INACTIVE	DEDICATEE
8 000000007C8546F0	14	34302	280257 000000007D930808	109 APEX_PUBLIC_USER	109	APEX_PUBLIC_USER	0	2147483644 (null)	(null)	(null)	INACTIVE	DEDICATEE
9 000000007C8596C0	16	38552	280260 000000007D9400C8	116 ORDS_PUBLIC_USER	116	ORDS_PUBLIC_USER	0	2147483644 (null)	(null)	(null)	INACTIVE	DEDICATEE
10 000000007C858EA8	17	34304	280259 000000007D93E868	116 ORDS_PUBLIC_USER	116	ORDS_PUBLIC_USER	0	2147483644 (null)	(null)	(null)	INACTIVE	DEDICATEE
11 000000007C85E690	18	31175	280261 000000007D9316A8	114 APEX_LISTENER	0	(null)	0	2147483644 (null)	(null)	(null)	INACTIVE	DEDICATEE
12 000000007C863660	20	32698	280262 000000007D9596E8	114 APEX_LISTENER	0	(null)	0	2147483644 (null)	(null)	(null)	INACTIVE	DEDICATEE
13 000000007C868630	22	64370	280263 000000007D95C1A8	115 APEX_REST_PUBLIC_USER	115	APEX_REST_PUBLIC_USER	0	2147483644 (null)	(null)	(null)	INACTIVE	DEDICATEE
14 000000007C86AE18	23	62342	280263 000000007D95AC48	115 APEX_REST_PUBLIC_USER	115	APEX_REST_PUBLIC_USER	0	2147483644 (null)	(null)	(null)	INACTIVE	DEDICATEE
15 000000007C87C570	30	43837	4294967295 000000007D9601C8	0 SYS	3	(null)	3	2147483644 (null)	(null)	(null)	ACTIVE	DEDICATEE
16 000000007C838110	391	29671	0 000000007D8BD308	0 (null)	0	(null)	0	2147483644 (null)	(null)	(null)	ACTIVE	DEDICATEE
17 000000007C85FF8	392	61677	0 000000007D8DE868	0 (null)	0	(null)	0	2147483644 (null)	(null)	(null)	ACTIVE	DEDICATEE
18 000000007C887E0	393	28288	0 000000007D8DFDC8	0 (null)	0	(null)	0	2147483644 (null)	(null)	(null)	ACTIVE	DEDICATEE
19 000000007C88AFC8	394	12297	0 000000007D8E1328	0 (null)	0	(null)	0	2147483644 (null)	(null)	(null)	ACTIVE	DEDICATEE
20 000000007C8D7B0	395	33811	0 000000007D8E2888	0 (null)	0	(null)	0	2147483644 (null)	(null)	(null)	ACTIVE	DEDICATEE

- Accesibilidad



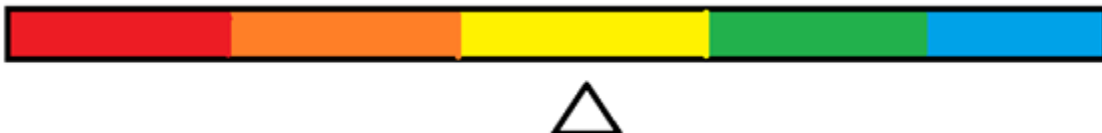
Descarga gratuita.

- Compatibilidad



Se puede realizar cualquier consulta a la BBDD desde cualquier Backend que soporte JDBC. Además, permite almacenar JSONs, operaciones relacionadas con ellos (aunque limitadas o poco acordes a lo necesitado) y gestión de roles.

- Usabilidad



Posee una interfaz similar a otros SGDB, pero el acceso a herramientas está poco visible. Además, la muestra de errores de queries suele ser visualmente poco atractiva.

- Experiencia



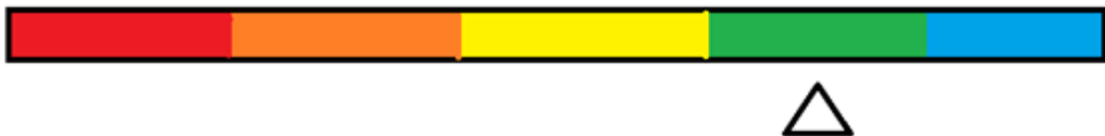
Experiencia desarrollando bases de datos, triggers, queries, ciertos sistemas de seguridad y roles y otras herramientas.

- Ajustes de seguridad



Máscaras, hashes, conexiones cifradas, auditorías, encriptados, etc. Posee todas las funcionalidades de seguridad deseadas.

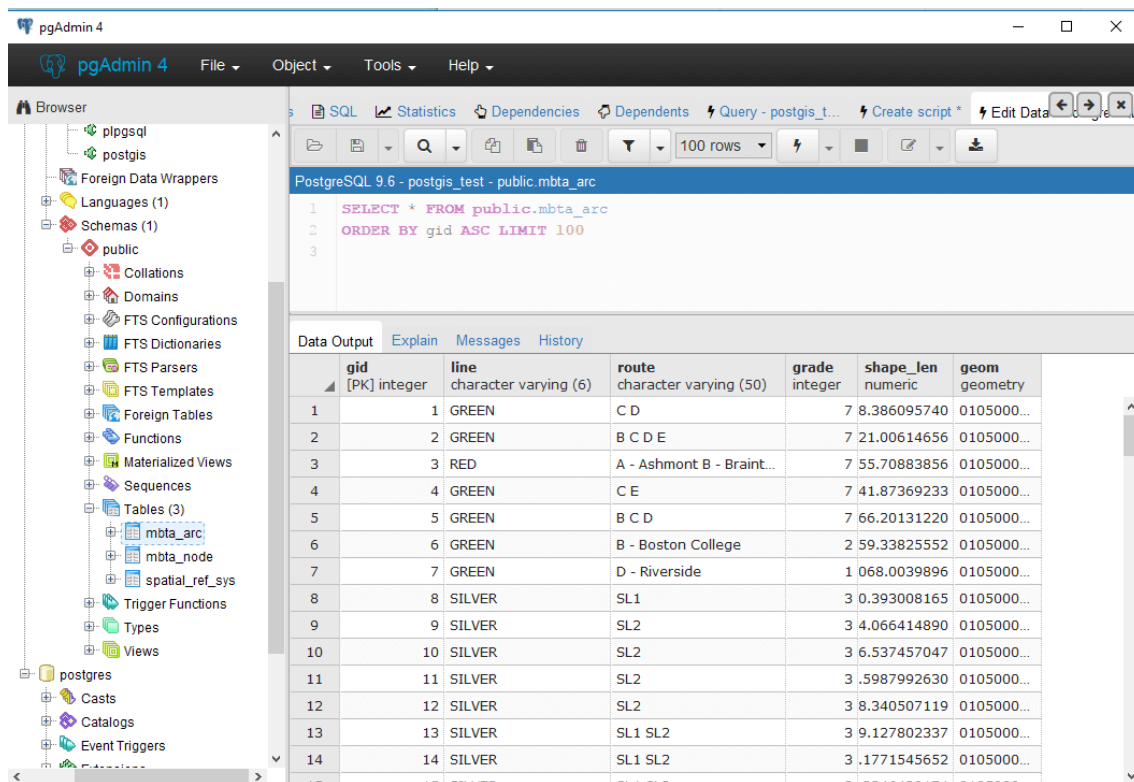
- Volumen



Al igual que las anteriores, no se especializa en volumen pero sí soporta y gestiona grandes tamaños de datos.

BUENA OPCIÓN

PostgreSQL (SGDB)



The screenshot shows the pgAdmin 4 interface. On the left is the 'Browser' pane with a tree view of the database structure, including 'public' schema and 'mbta_arc' table. The main pane shows a SQL query: `SELECT * FROM public.mbta_arc ORDER BY gid ASC LIMIT 100`. Below the query is a 'Data Output' tab displaying a table with 15 rows. The columns are: gid [PK] integer, line character varying (6), route character varying (50), grade integer, shape_len numeric, and geom geometry. The data shows various bus routes with their respective line numbers, colors, and shapes.

	gid [PK] integer	line character varying (6)	route character varying (50)	grade integer	shape_len numeric	geom geometry
1	1	GREEN	C D	7	8.386095740	0105000...
2	2	GREEN	B C D E	7	21.00614656	0105000...
3	3	RED	A - Ashmont B - Braint...	7	55.70883856	0105000...
4	4	GREEN	C E	7	41.87369233	0105000...
5	5	GREEN	B C D	7	66.20131220	0105000...
6	6	GREEN	B - Boston College	2	59.33825552	0105000...
7	7	GREEN	D - Riverside	1	068.0039896	0105000...
8	8	SILVER	SL1	3	0.393008165	0105000...
9	9	SILVER	SL2	3	4.066414890	0105000...
10	10	SILVER	SL2	3	6.537457047	0105000...
11	11	SILVER	SL2	3	.5987992630	0105000...
12	12	SILVER	SL2	3	8.340507119	0105000...
13	13	SILVER	SL1 SL2	3	9.127802337	0105000...
14	14	SILVER	SL1 SL2	3	.1771545652	0105000...
15	15	SILVER	SL1 SL2	3	.3216428174	0105000...

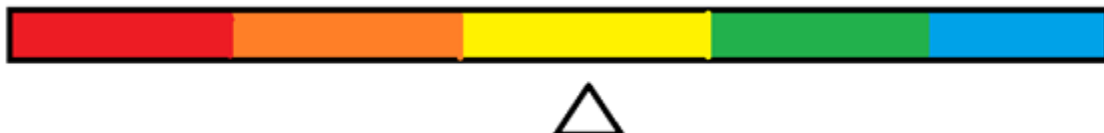
Curiosidad de esta SGDB: está pensada para ser orientada a objetos, cumplir principios ACID y gestionar concurrencia.

- Accesibilidad



Herramienta gratuita y deja a disposición todas sus versiones.

- Compatibilidad



Aunque permite gestionar JSONs, no tiene comandos específicos. Además, para conectar la BBDD con el Backend es necesario un Driver (al menos en el caso de Java).

- Usabilidad



Hay muchísimas herramientas: es difícil diferenciar las básicas de las más específicas. Para gestionarla adecuadamente se debe tirar de comandos de terminal, aunque las queries básicas están disponibles en su interfaz.

- Experiencia



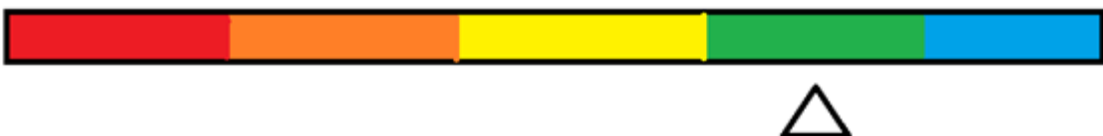
Aunque moderna, no tengo experiencia usando esta BBDD, además de que la curva de aprendizaje por experiencia de internautas ajenos es más larga.

- Ajustes de seguridad



Máscaras, hashes, conexiones cifradas, auditorías, encriptados, backups, roles, etc. Posee todas las funcionalidades de seguridad deseadas.

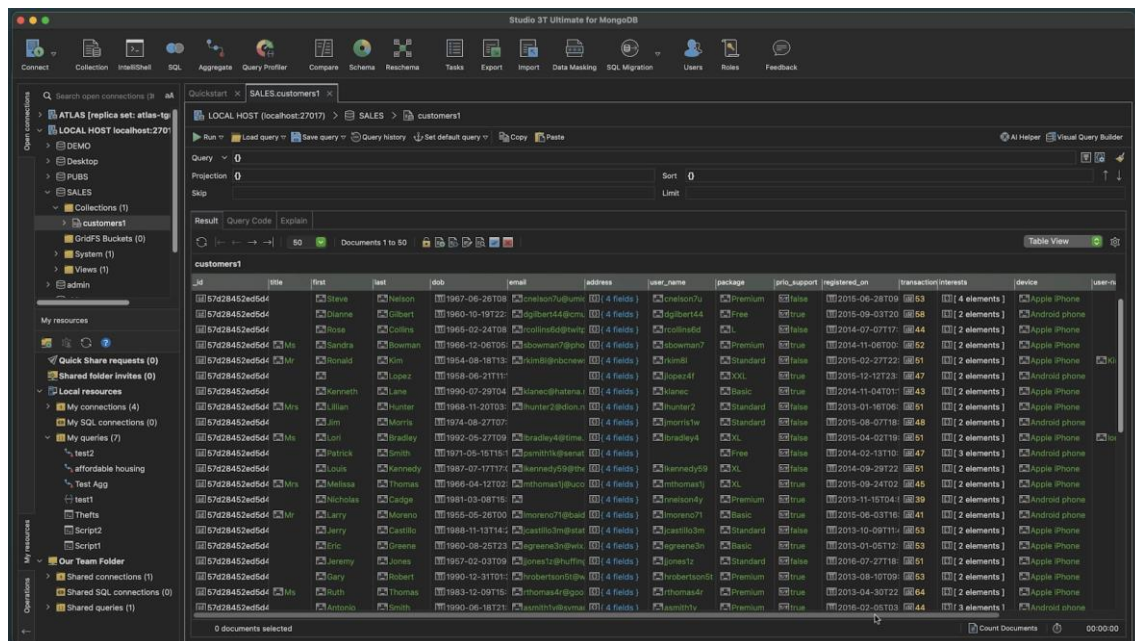
- Volumen



Al igual que las anteriores, no se especializa en volumen pero sí soporta y gestiona grandes tamaños de datos.

OPCIÓN NO RECOMENDABLE

MongoDB (SGDB)



No está pensada para modelos relacionales, por lo que solo veremos si tiene utilidad para los JSONs (sistema clave-valor).

- Accesibilidad



Herramienta gratuita y deja a disposición todas sus versiones.

- Compatibilidad



Aunque permite gestionar JSONs, no tiene comandos específicos. Además, para conectar la BBDD con el Backend es necesario un Driver (al menos en el caso de Java).

- Usabilidad



Gran flexibilidad y la curva de aprendizaje es corta.

- Experiencia



Aunque moderna, no tengo experiencia usando esta BBDD.

- Ajustes de seguridad



Máscaras, hashes, conexiones cifradas, auditorías, encriptados, roles, etc. Posee todas las funcionalidades de seguridad deseadas.

- Volumen



Su característica más representativa. Su nombre viene de “huMONGOus”, que es “enorme”. Las lecturas están diseñadas para ser más rápidas que nunca.

BUENA OPCIÓN (PARA ALMACENAR JSONs)

Spark Sql (data analyzer)

Esta tecnología permite crear estructuras llamadas DataFrames para operar con datos estructurados y semiestructurados de bases de datos externas. También permite la generación de archivos en otros formatos de texto.

<https://www.youtube.com/watch?v=M40WQzjGePw>

```
Spark SQL DataFrame

In [2]: from pyspark.sql import SparkSession

In [3]: spark = SparkSession.builder.appName('DataFrame').getOrCreate()

In [4]: df = spark.read.json('personas.json')

In [5]: df.show()

+----+-----+
|edad|nombre|
+----+-----+
| null|Miguel|
|  25 |Carlos|
|  19 | Juan|
+----+-----+

In [ ]: 
```

- Accesibilidad



Herramienta gratuita y fácilmente instalable.

- Compatibilidad



No solo puede conectar con multitud de datos a través de JDBC, sino que también permite generar CSVs a partir de ficheros JSON y trabajar con las dos estructuras de datos que manejamos y realizar operaciones SQL con ellas. Contempla todo.

- Usabilidad



Aunque no contempla una interfaz gráfica y todo se basa en comandos y arte ASCII, existe documentación en internet que ayuda a operar con ella.

- Experiencia



Partimos de cero.

2. CONCLUSIONESS

Salvo en el caso de PostgreSQL, cuya adaptabilidad y propiedades analizadas no se terminan de ajustar para este proyecto, todas las herramientas pueden ser de utilidad para el proyecto. Por preferencia final, podemos emplear las siguientes opciones:

- MySQL como SGDB para todos los datos y SparkSQL como procesador de datos JSON.
- MySQL como SGDB para los datos estructurados, MongoDB como SGDB para los JSON y SparkSQL como procesador de datos JSON.
- SQLServer como SGDB para todos los datos y SparkSQL como procesador de datos JSON.
- SQLServer como SGDB para los datos estructurados, MongoDB como SGDB para los JSON y SparkSQL como procesador de datos JSON.