



UNIVERSIDAD TECNOLÓGICA DE SAN LUIS RIO COLORADO

GUIA 3 PARCIAL

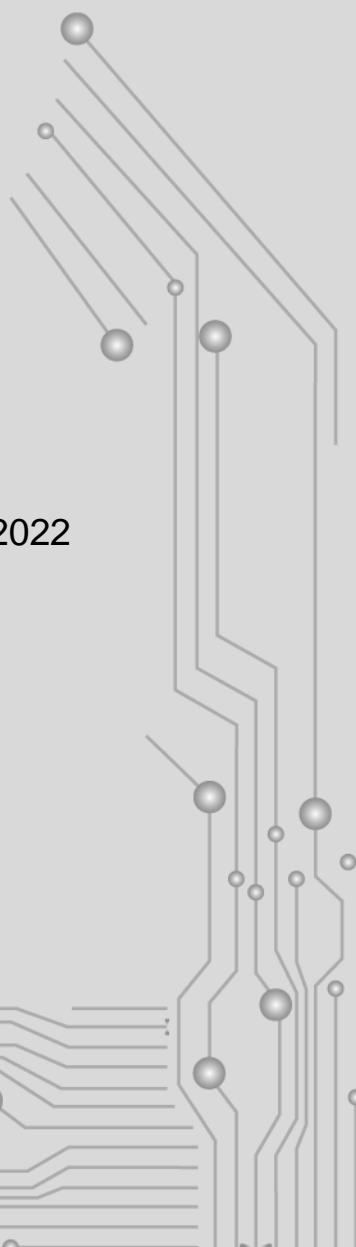
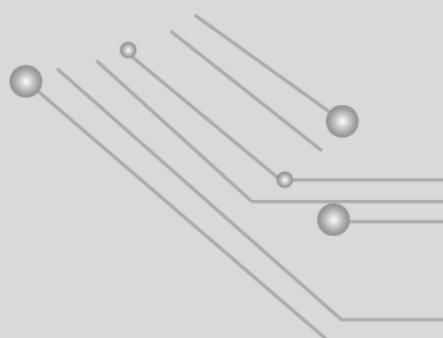
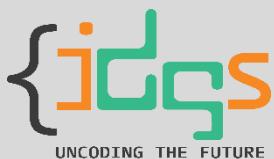
MTRA. JULIA HELIZABETH GARCIA HERRERA

ALUMNO: VICTOR MANUEL GALVAN COVARRUBIAS

ING. EN DESARROLLO Y GESTIÓN DE SOFTWARE

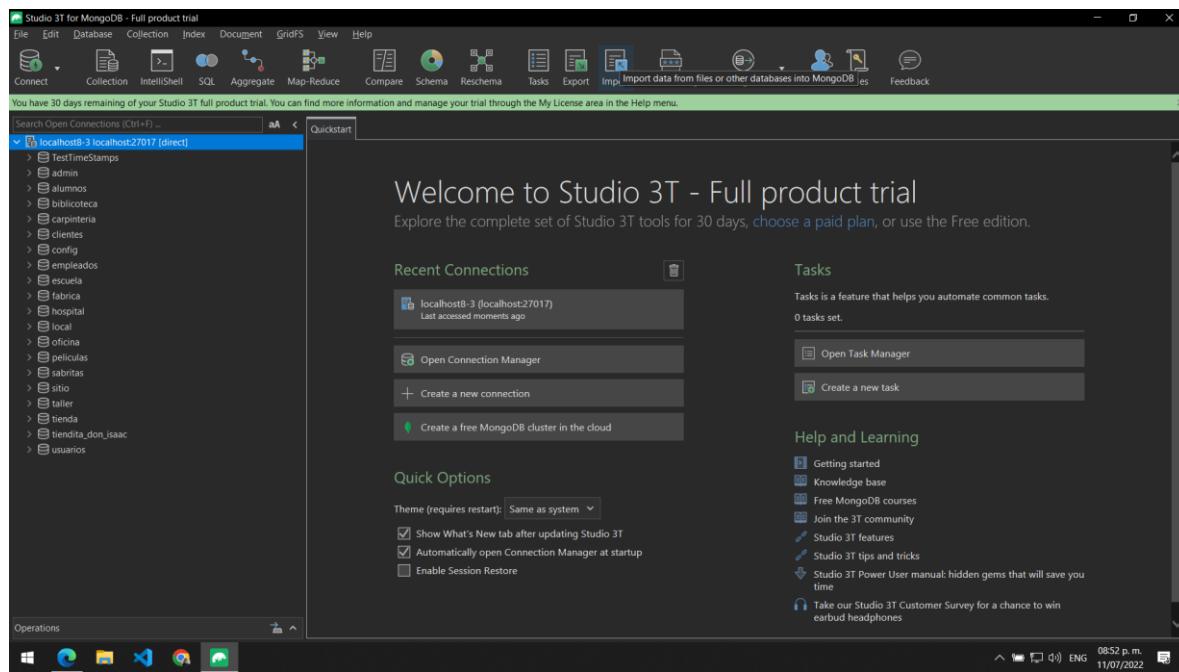
San Luis Rio Colorado, Sonora

Agosto, 2022

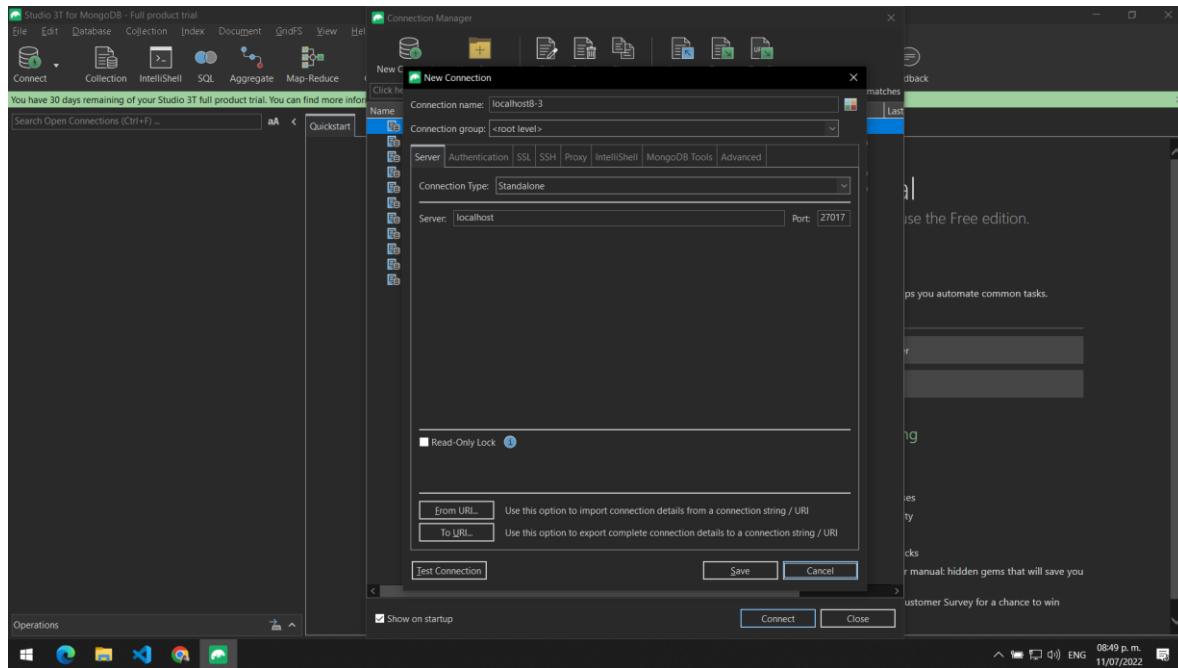




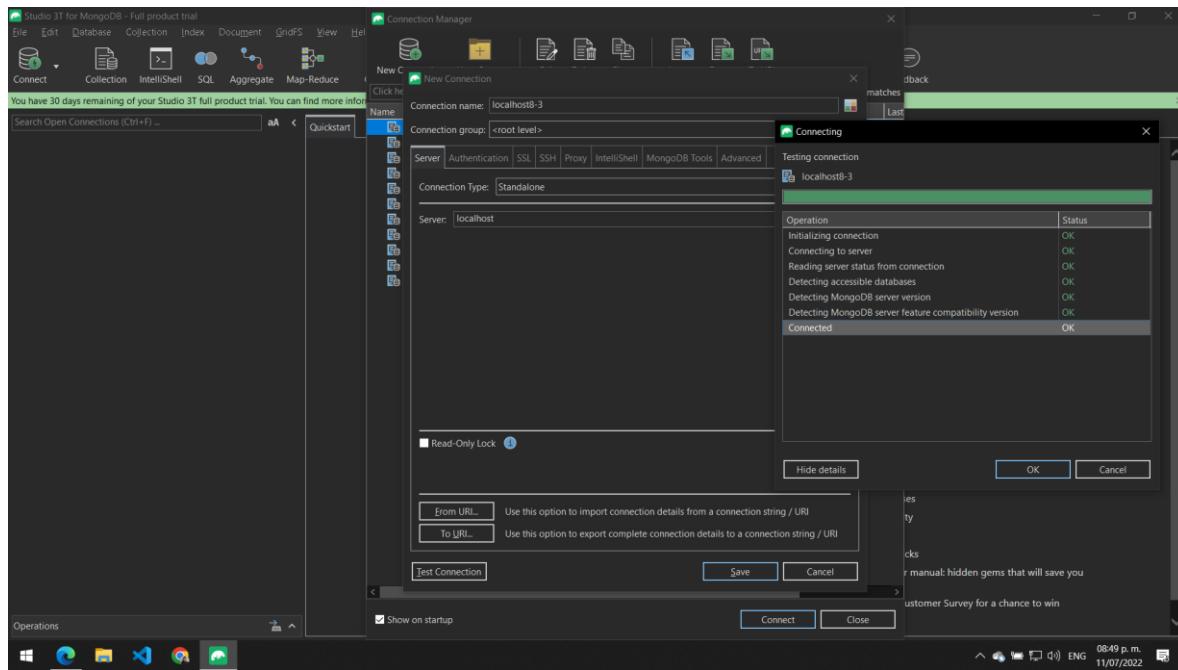
Studio 3T está diseñado para equipos profesionales en crecimiento y ofrece una variedad de formas de ver y consultar colecciones de datos, incluidas agregaciones sofisticadas, extensiones JSON nativas de Mongo, consultas SQL tradicionales y un generador de consultas de arrastrar y soltar.



Conexión

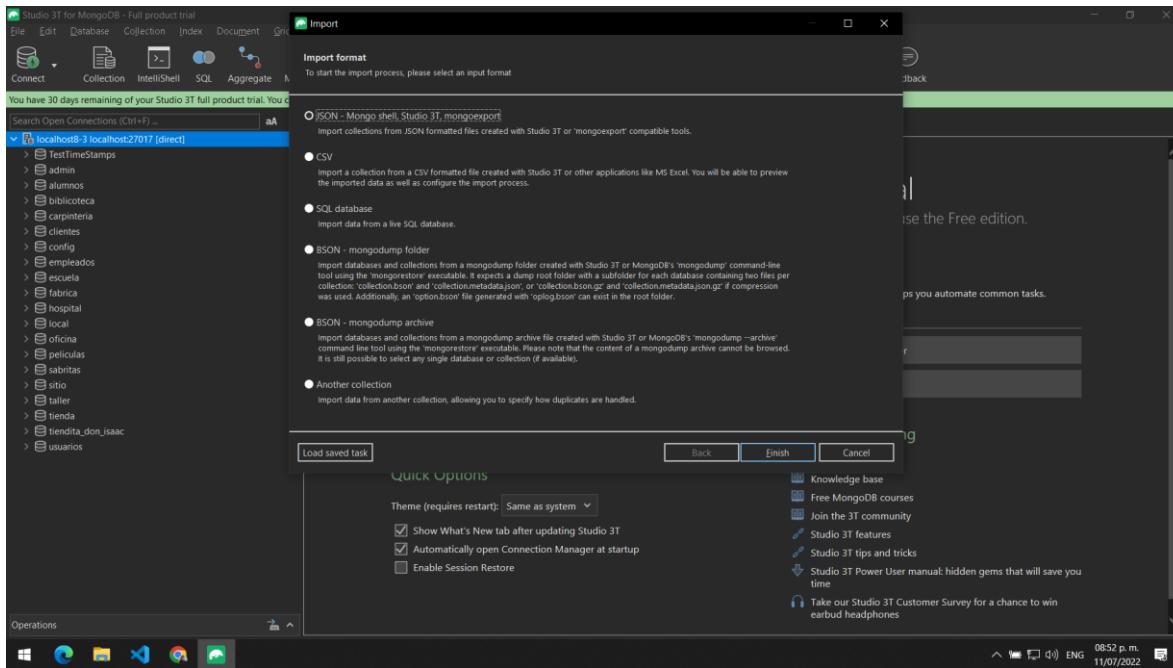


Damos click en agregar y establecemos los parámetros para nuestra conexión.

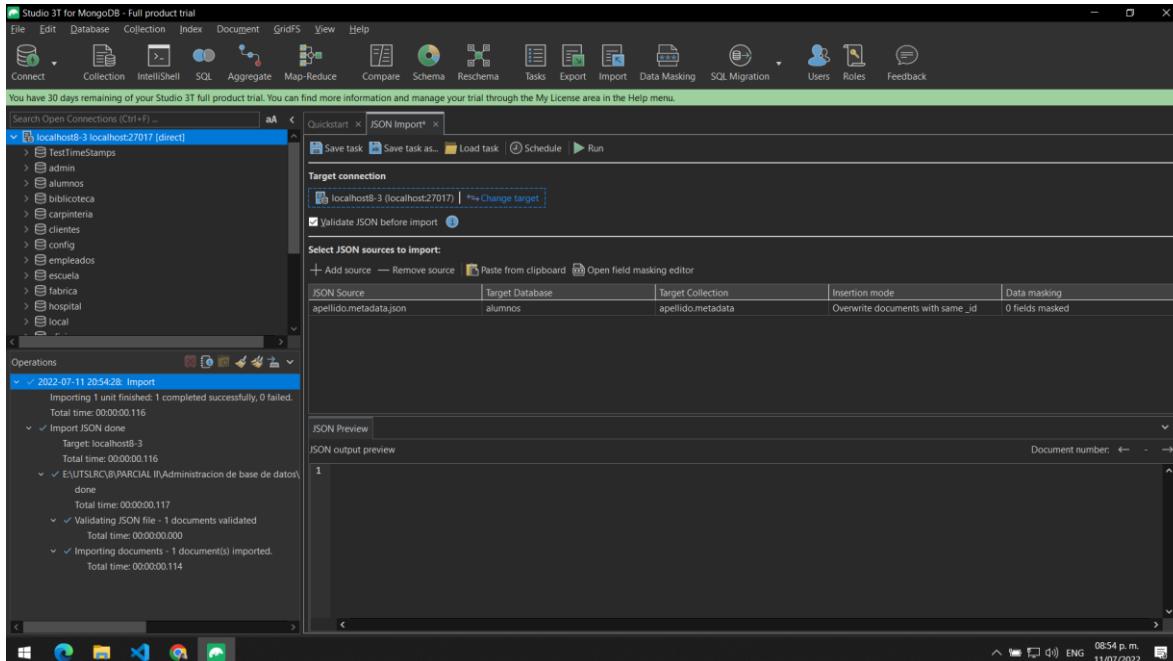


Probamos que la conexión este funcionando y la agregamos una vez nos dice que todo está OK.

Importación

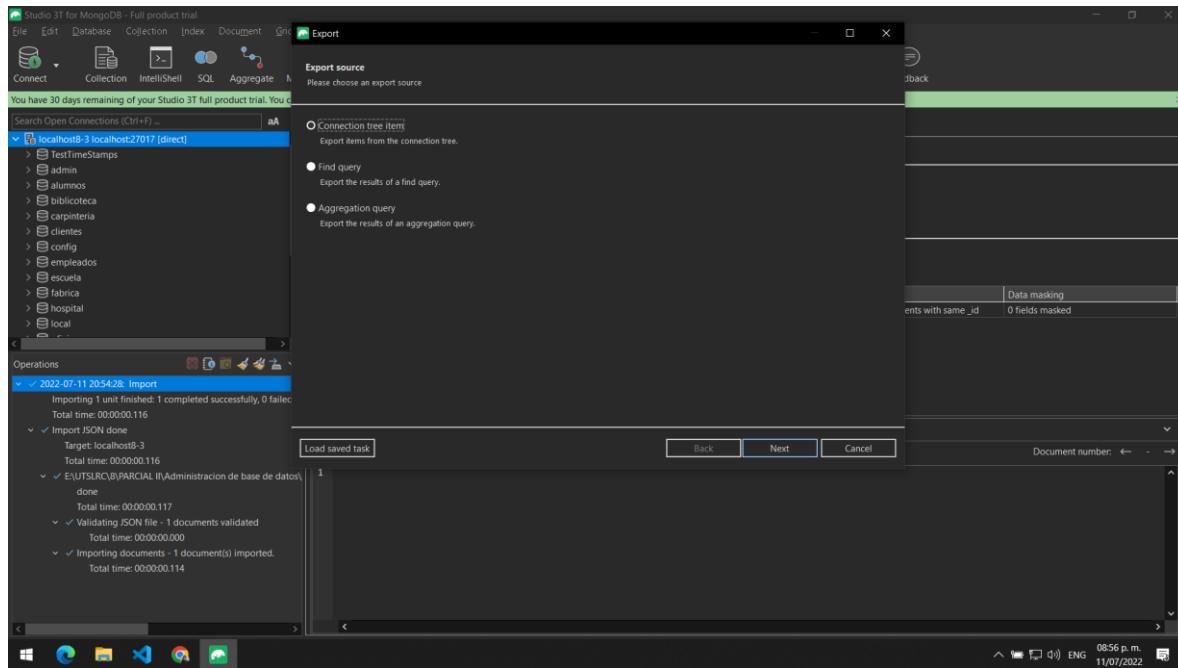


Damos click en importar y seleccionamos la base de datos donde lo queremos importar así como también el archivo a importar y su formato.

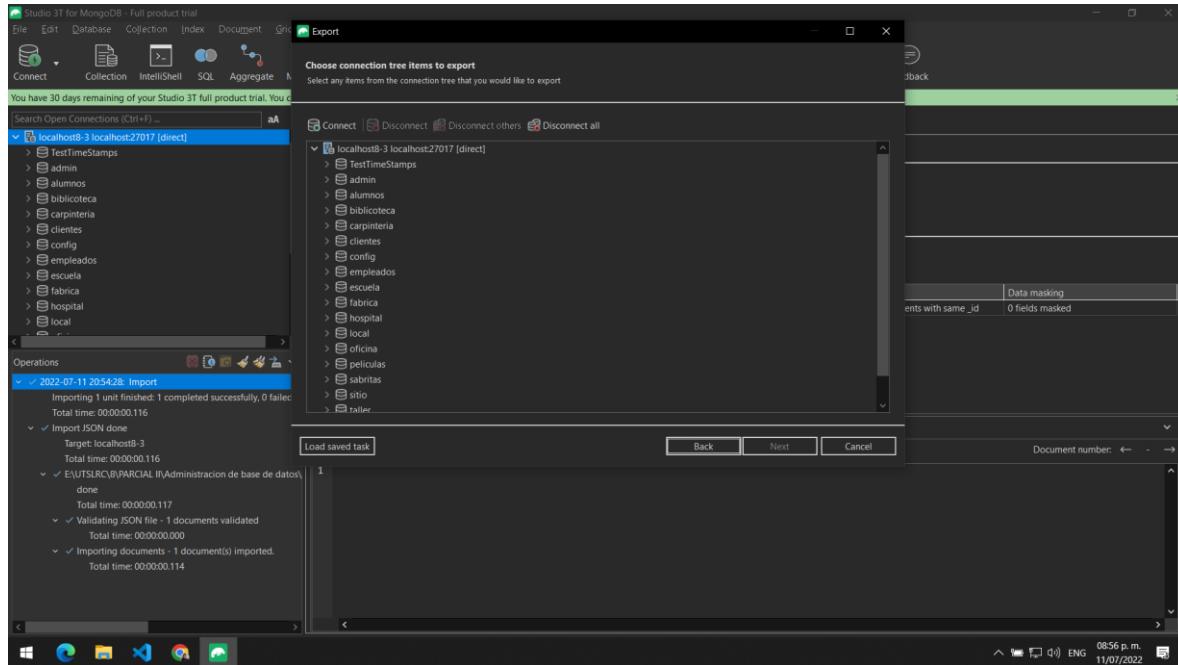


Dentro de la importación podemos correrla, probarlas y programas para que se haga periódicamente.

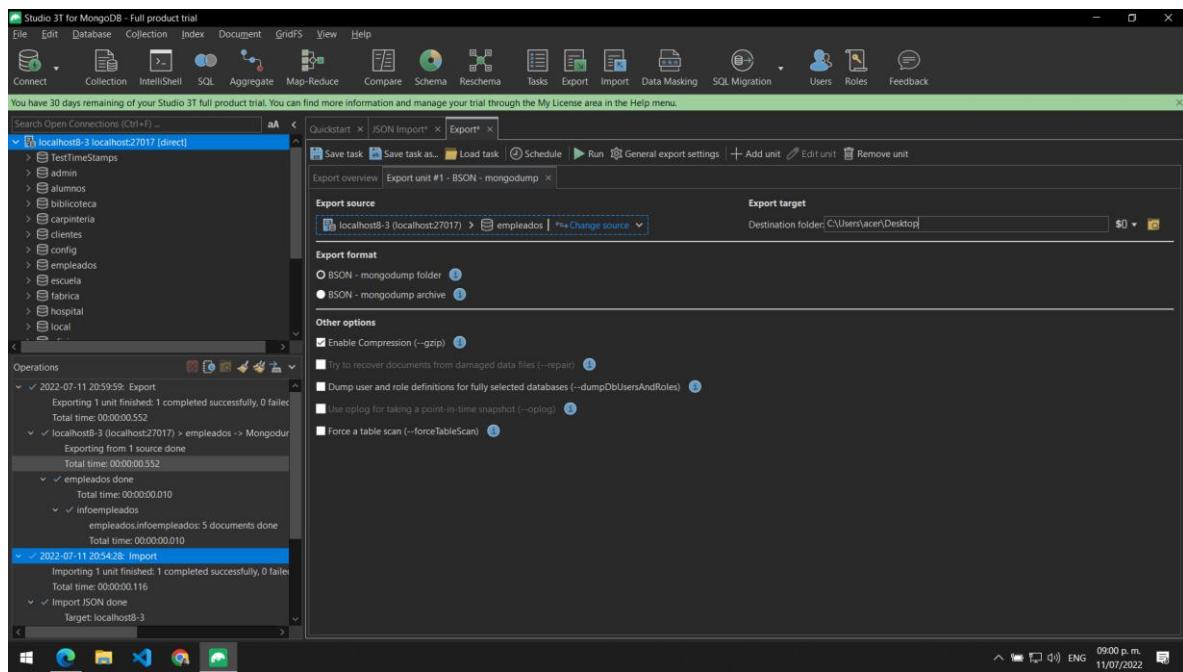
Exportación



Damos click en exportar y seleccionamos la primera opción.

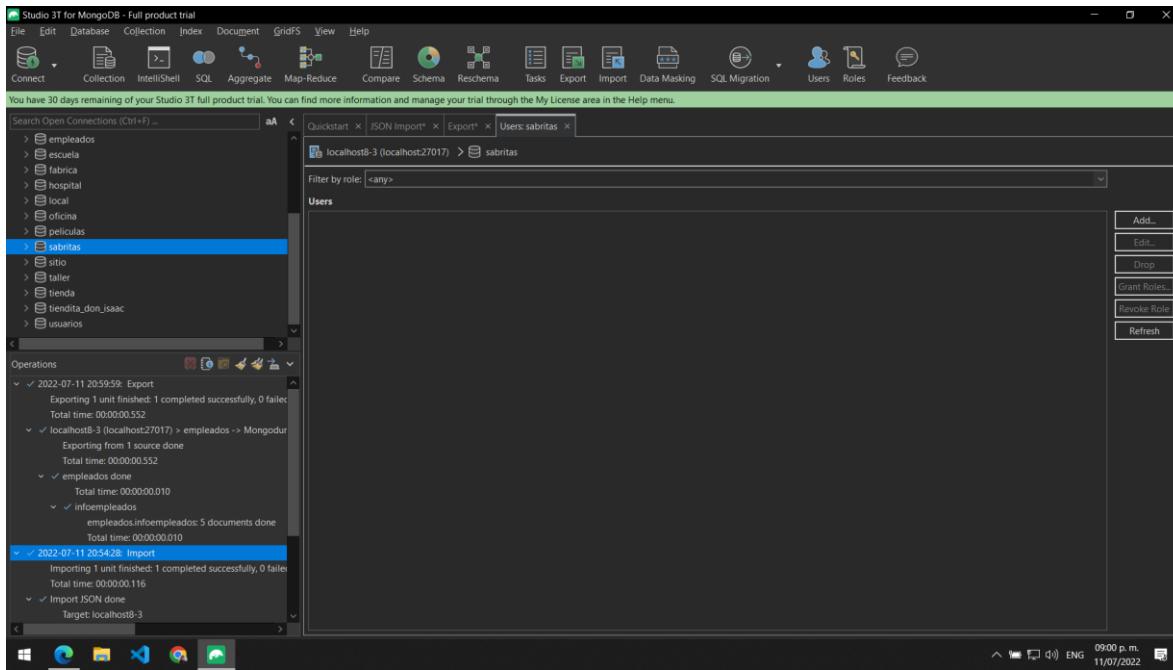


Elegimos una base de datos de donde queramos hacer la copia.

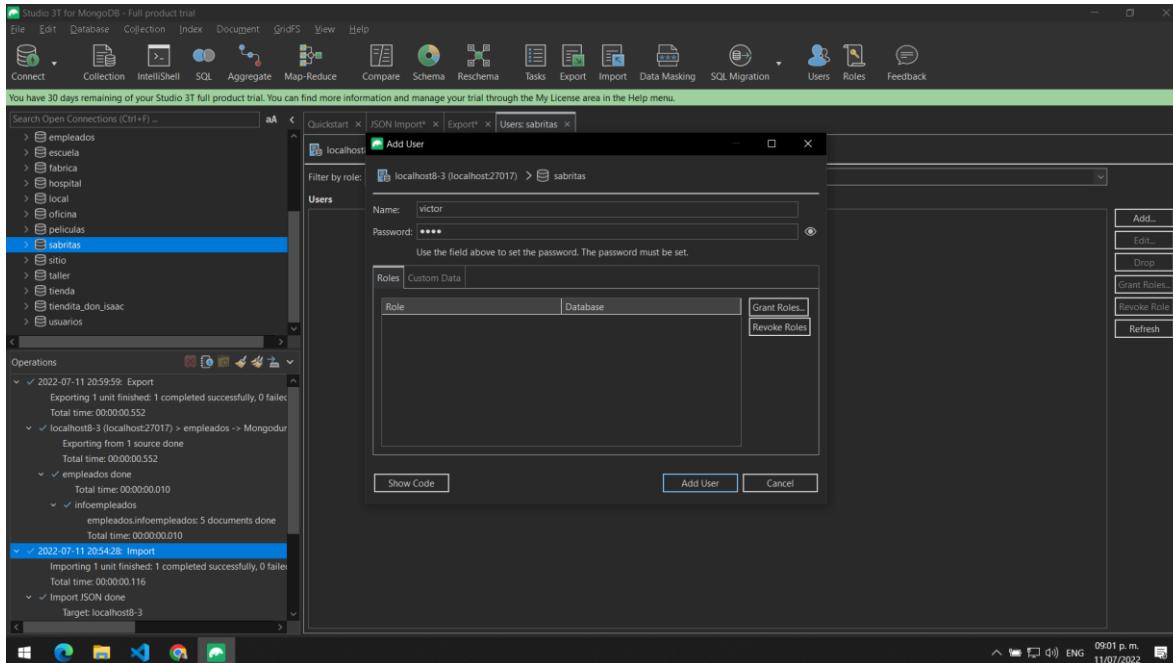


Dentro de la pestaña para exportar podemos probar nuestra copia con los parámetros que establecimos y también programar una tarea para que se realice el mismo proceso periódicamente.

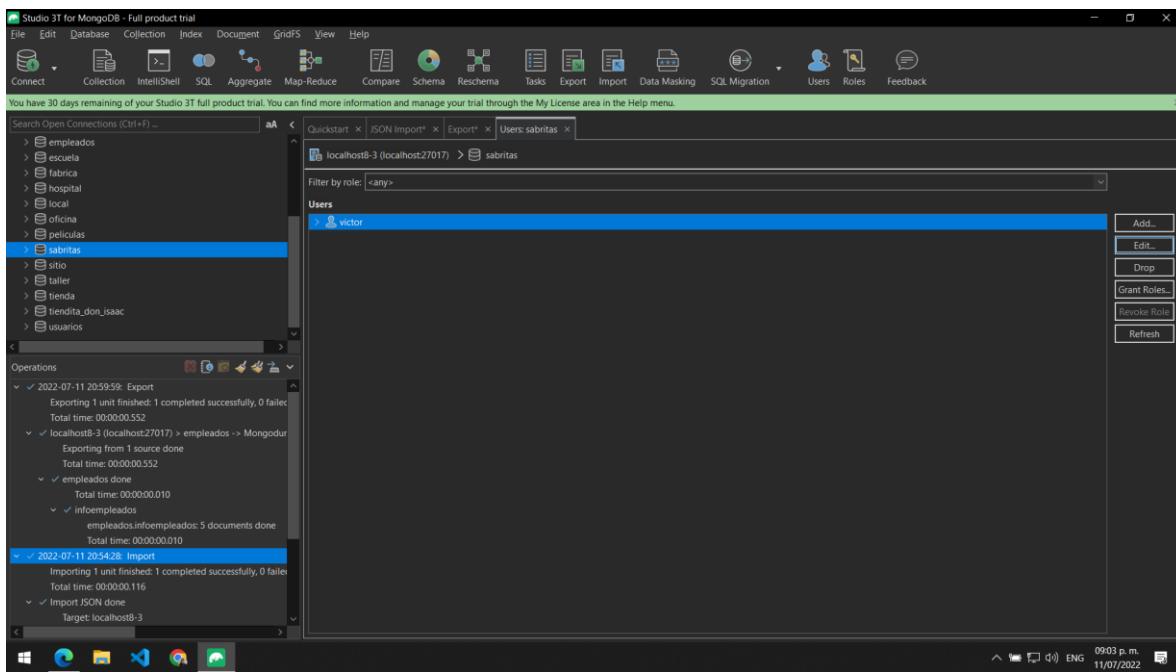
Usuarios



Damos click en usuario y entraremos a apartado donde se encuentran todos los usuarios de nuestras bases de datos realizadas.

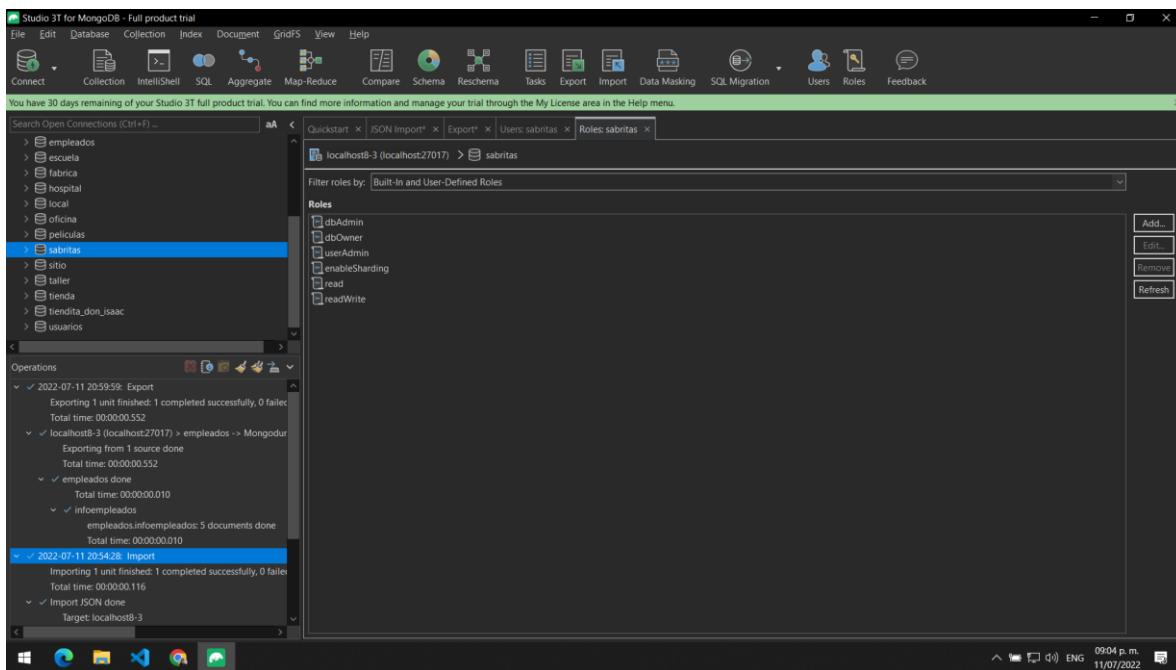


Damos click en 'Add' y podremos agregar un usuario para la base de datos seleccionada.

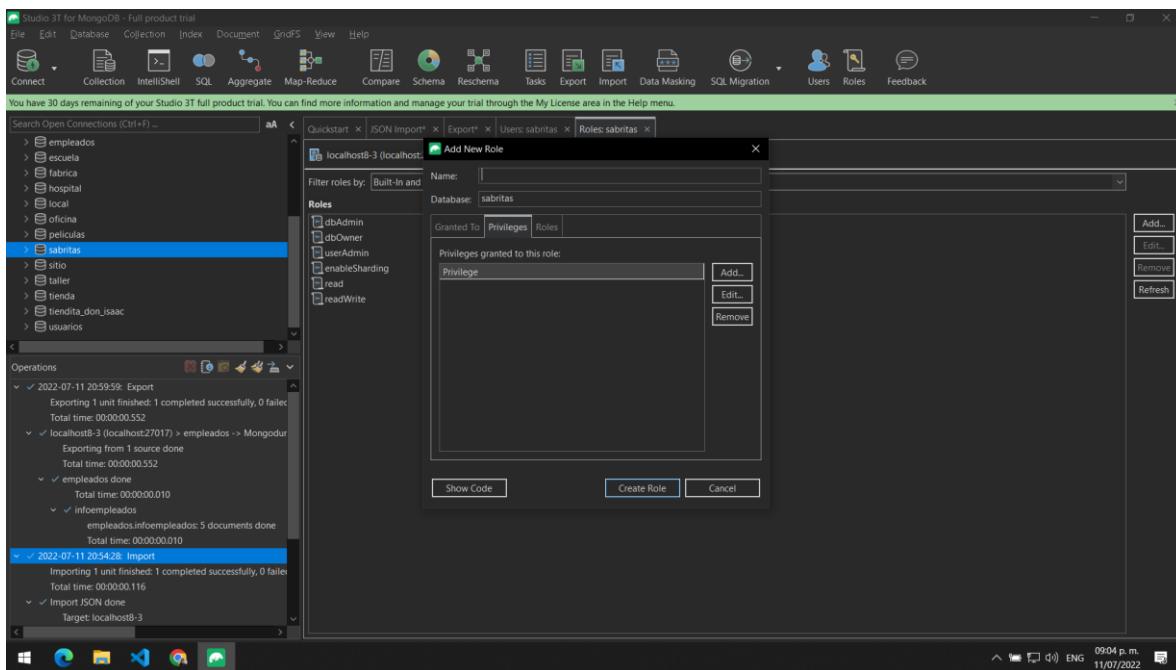


Una vez realizado podremos visualizar nuestro usuario creado.

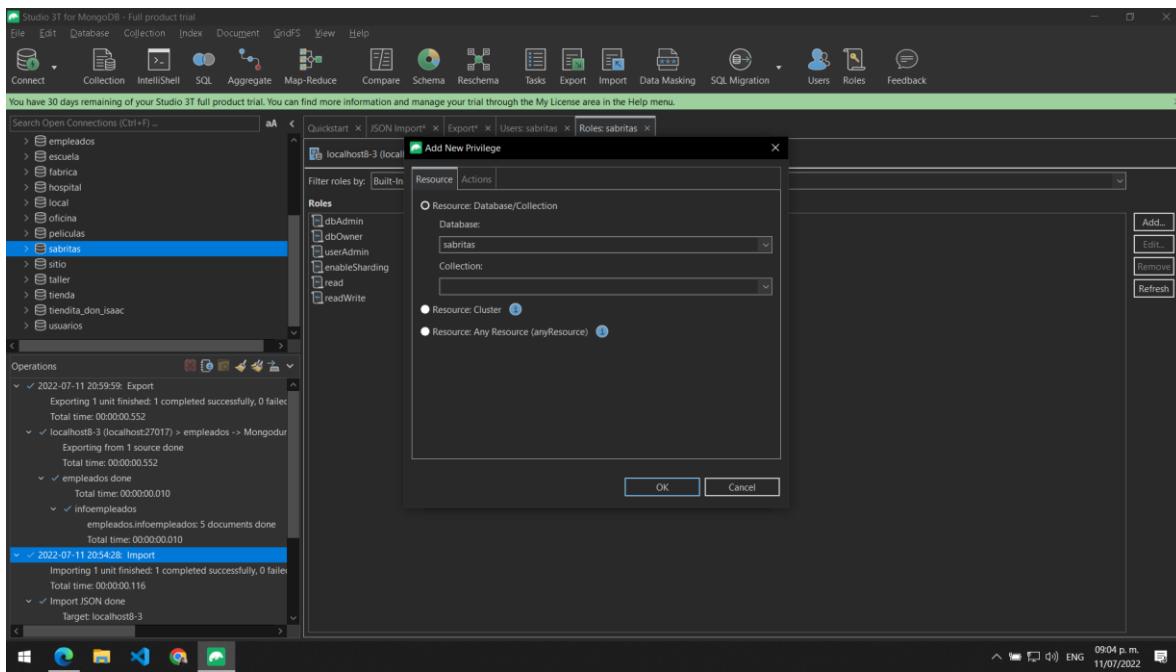
Roles



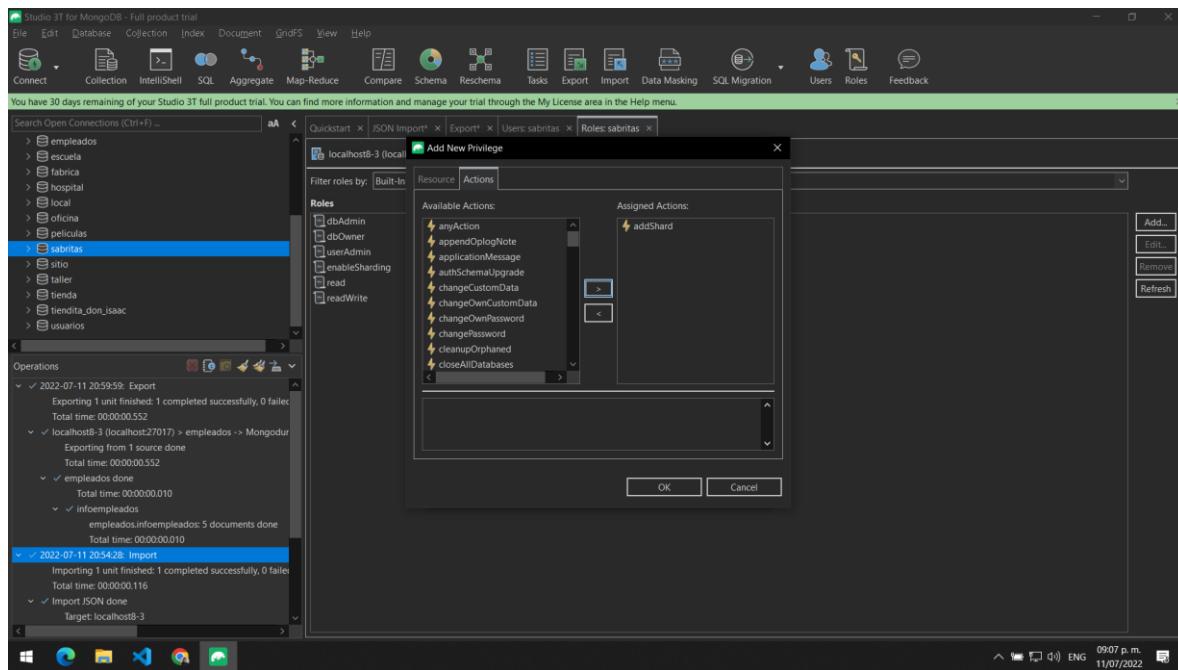
Entramos al apartado de roles y podremos ver todos lo que se encuentran por defecto.



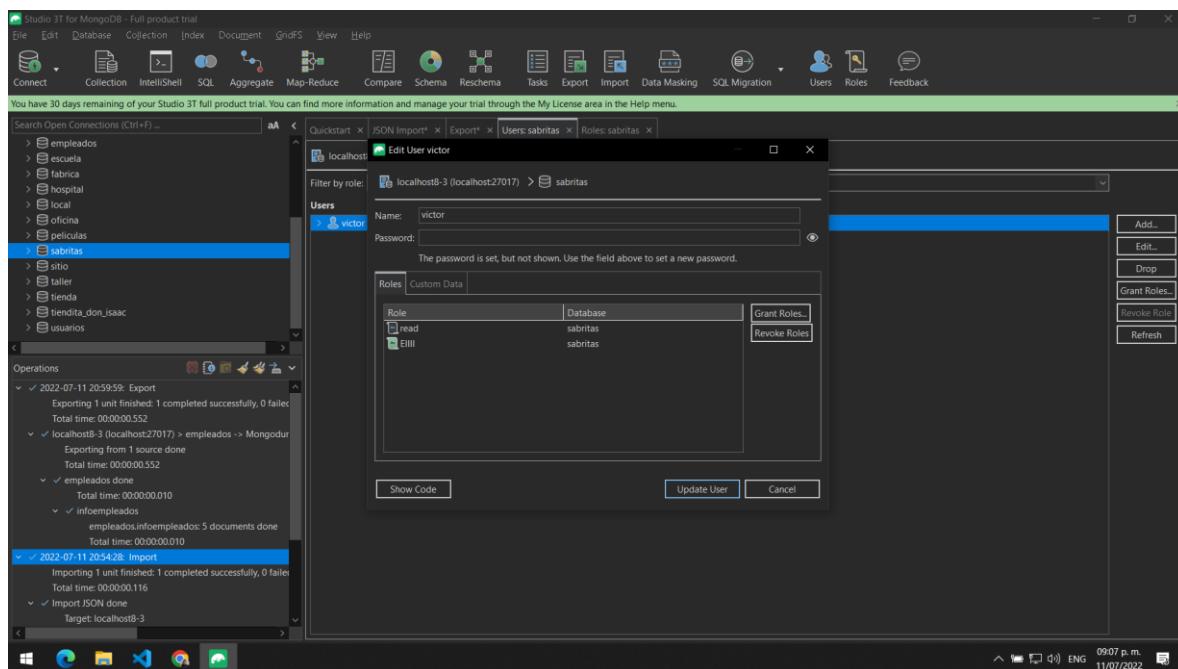
Para agregar uno damos click en 'Add' y nos muestra los parámetros que podemos establecer para el nuevo rol.



Definimos para qué base de datos se establecerán los privilegios.

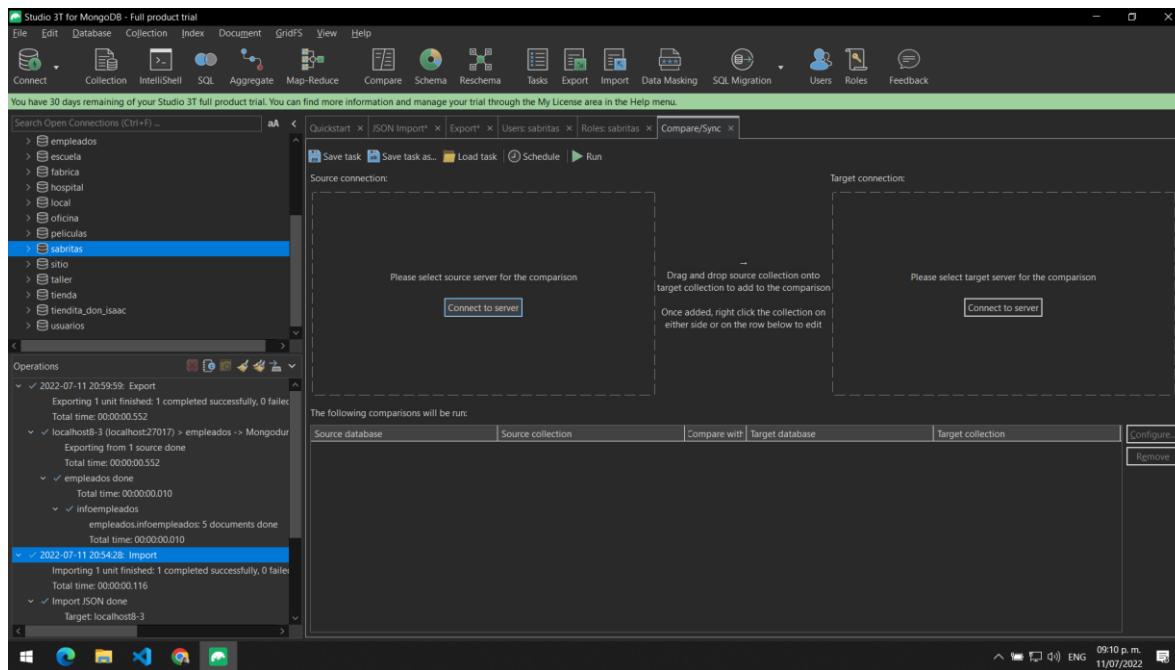


Establecemos que privilegios se le otorgaran a este rol.

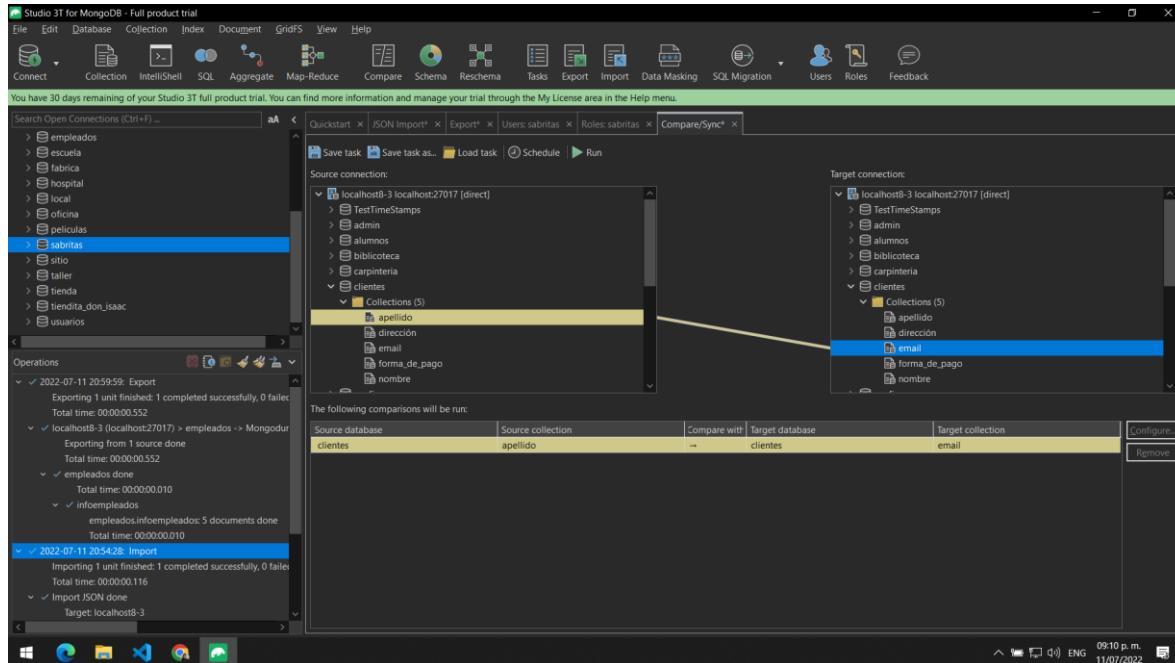


Una vez creado el rol nos lo podremos agregar a nuestros usuarios creados.

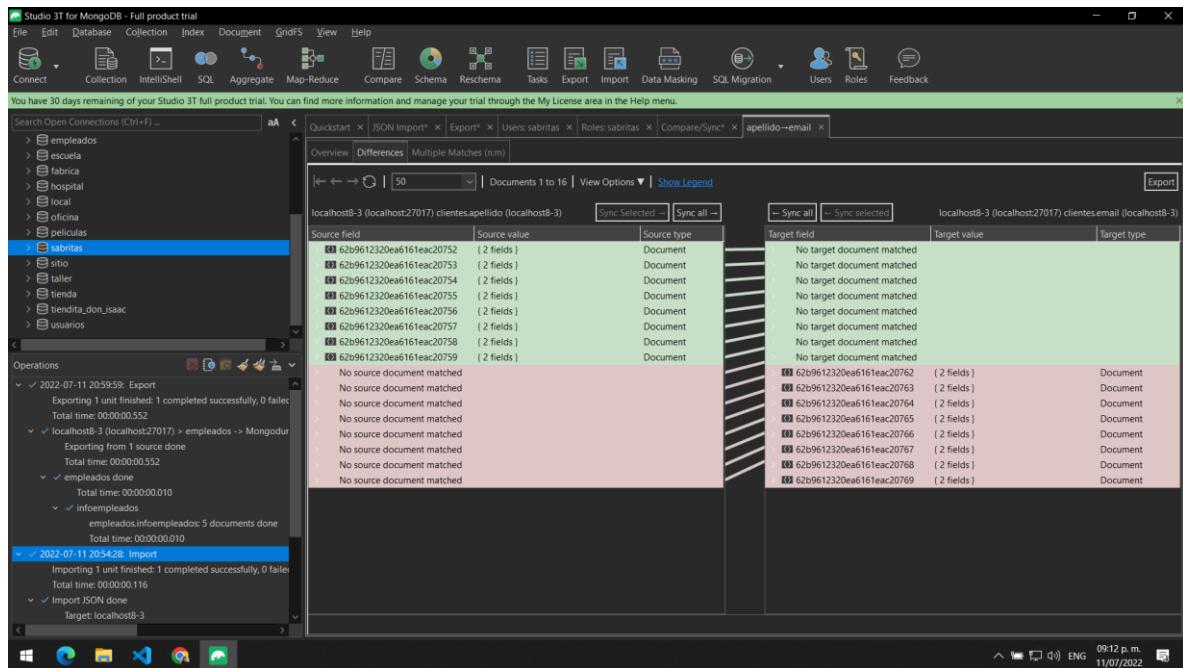
Comparación



Damos click en el apartado para comparar y seleccionamos las colecciones que queremos comparar.

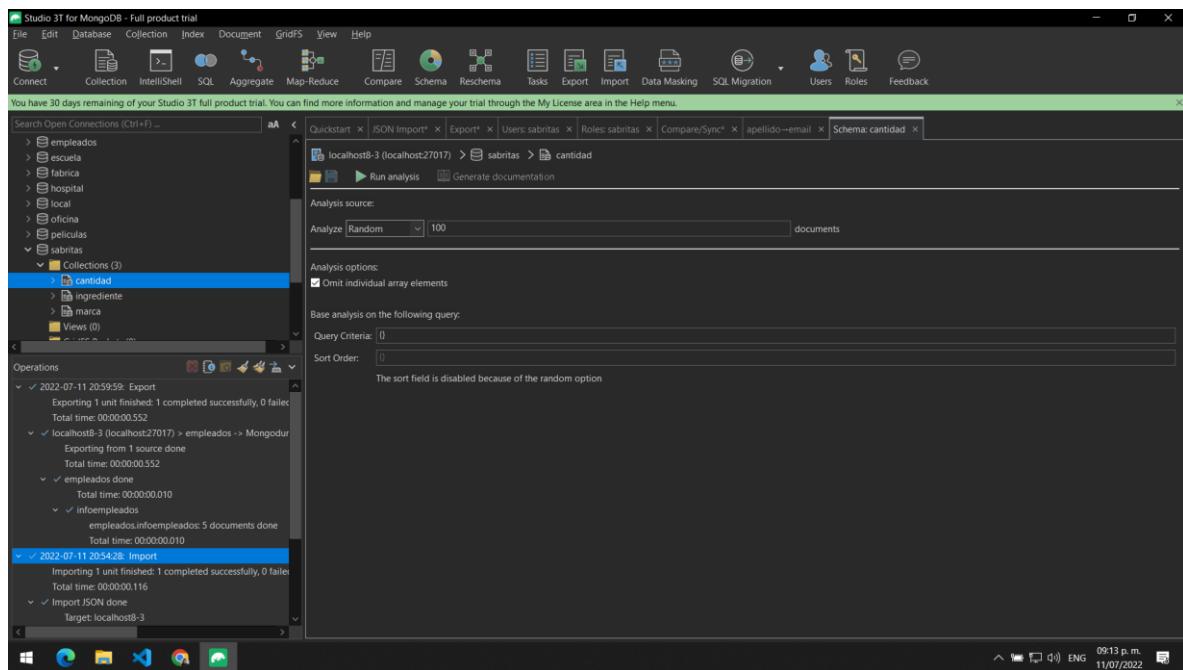


Arrastramos de izquierdas a derechas para poder comparar las colecciones.

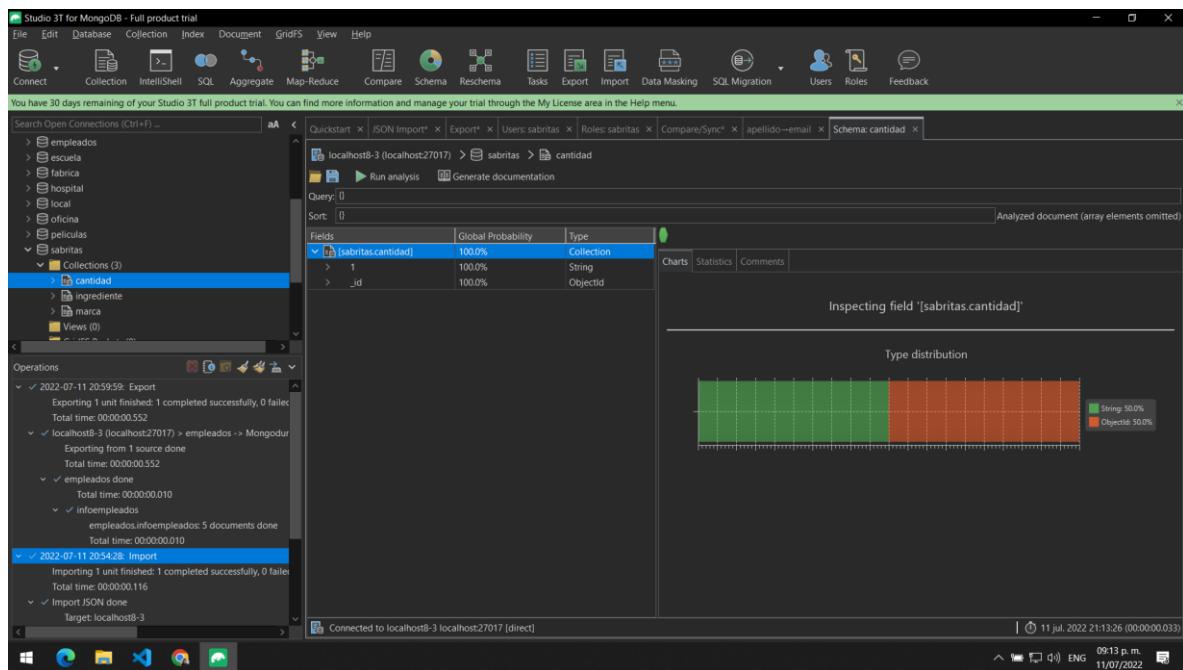


Visualizamos los registro y la relación que estos tienen.

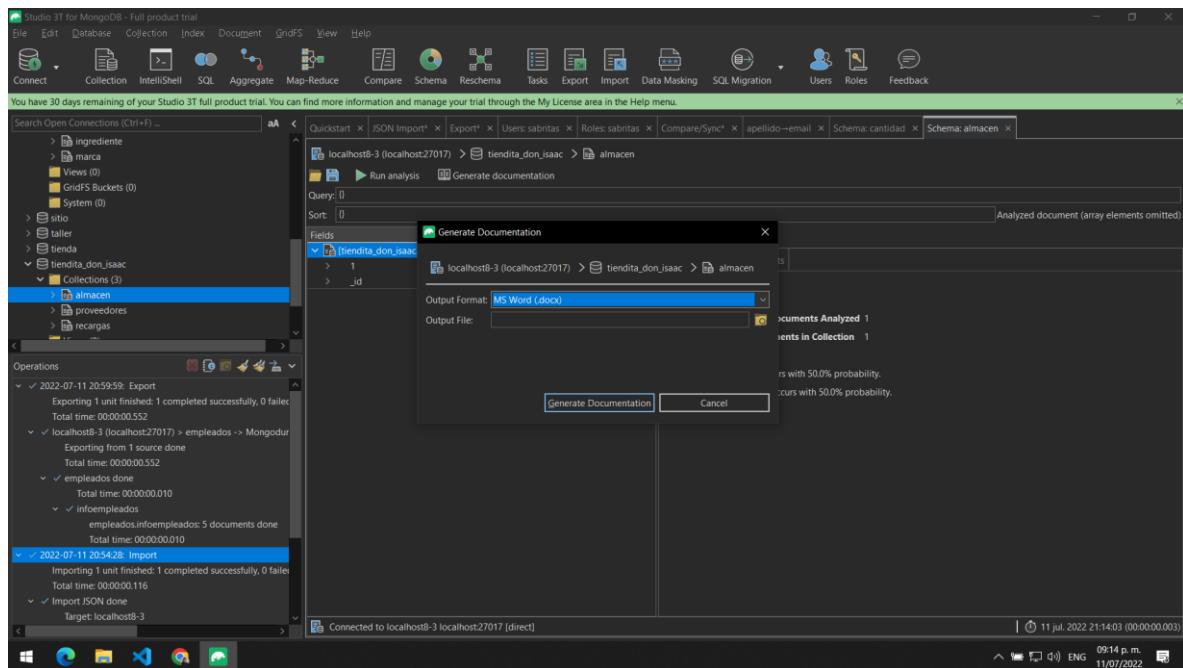
Esquema



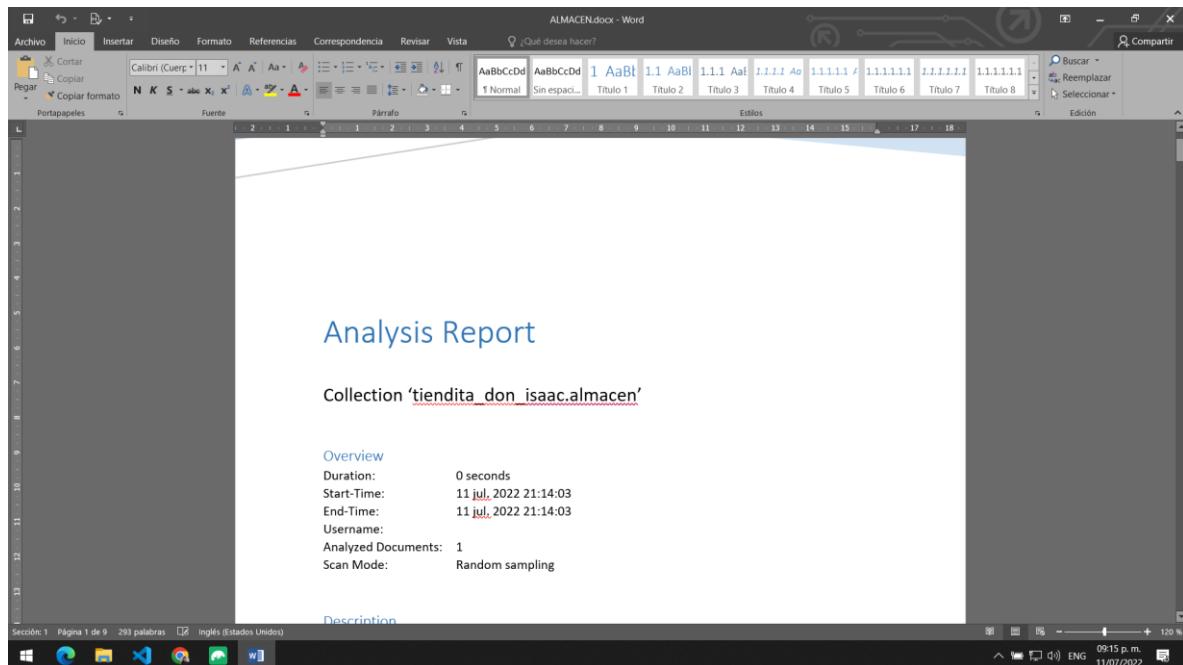
Con una colección seleccionada damos click en esquema y podremos realizar un análisis de los datos contenidos en esta.



Nos muestra los registros, así como también los datos que esta contiene.



Podremos incluso indicar que nos elabore un reporte de nuestra colección para poder analizar la información más a fondo.



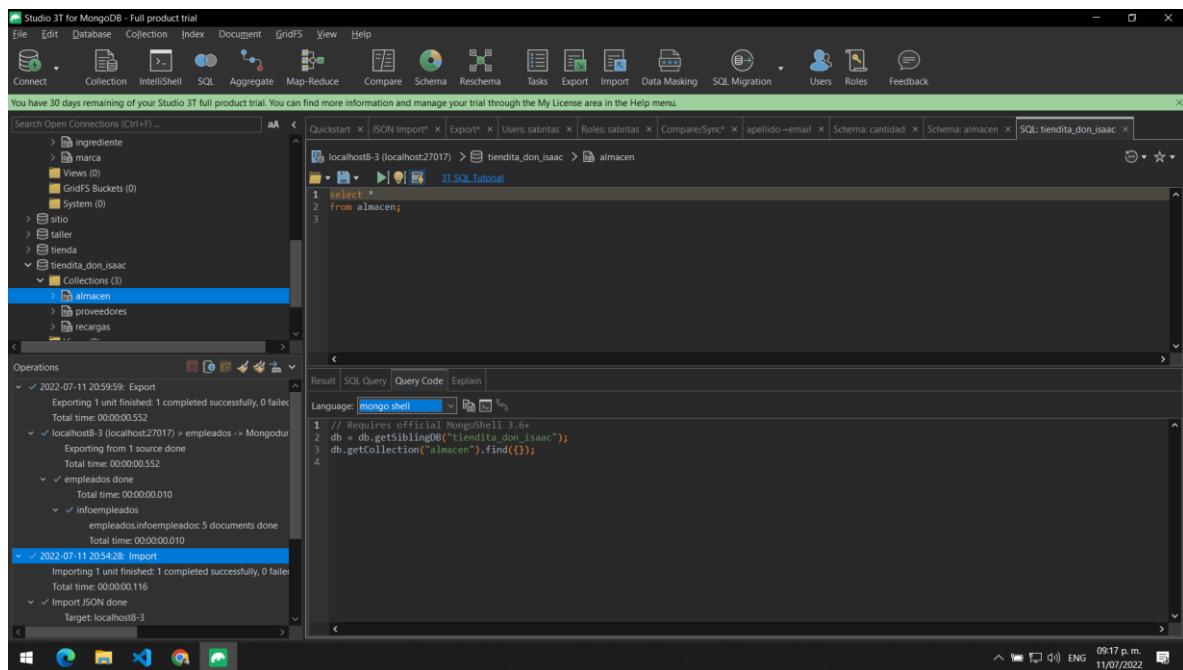
Una vez termina el proceso no abre el archivo de Word creado.

SQL

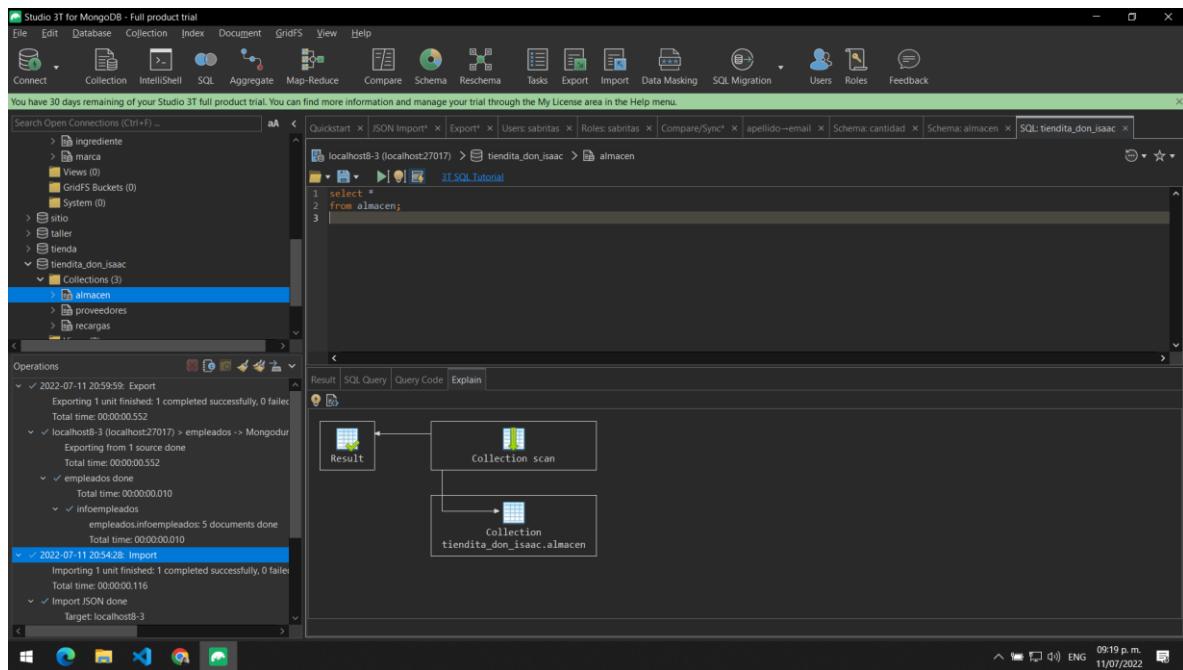
```
1 select *
2 From almacen;
```

| _id | Sabritas |
|-------------------|----------|
| 62aeeca775a205... | |

Damos click en el apartado de SQL y podremos utilizar lenguaje SQL y realizar consultas a nuestras base de datos en MongoDB como si se tratara de base de datos relacionales.



Nos muestra ejemplos de cómo se realizaría la consulta por línea de comandos.



Y también nos muestra un pequeño diagrama para visualizar nuestra colección.

Calidad en los datos

Importancia

Mala calidad se consideran la fuente de problemas. Ejemplo del daño que pueden causar los problemas de calidad de los datos incluyen gastos adicionales cuando los productos se envían a puntos incorrectos.

¿Como la determinamos?

Se realiza un inventario donde se inspecciona la precisión relativa, la singularidad y la validez de los datos se miden.

Evaluación

Crear un conjunto de reglas para la calidad de nuestros datos basándonos en los requisitos comerciales. Estas reglas nos dicen el nivel de calidad que requeriremos en los conjuntos de datos y no especifican los elementos que deben incluir para así nosotros poder evaluarlos.

Beneficios

1. Ahorro de tiempo y recursos.
2. Mejor capacidad de toma de decisiones y elaboración de estrategias más oportunas y eficientes basadas en el análisis de datos de calidad, confiables y validados.
3. Los riesgos en los proyectos se reducen.
4. Mejora la confianza de los clientes y la imagen de la empresa.
5. Ayuda a optimizar la captación de clientes y su fidelización.
6. Personalización de la oferta y el trato con el cliente.
7. Mayor rentabilidad de las campañas.
8. Conocer las expectativas y necesidades de clientes con más precisión.

Problemas

1. **Gran variedad de formatos y fuentes** de las que proviene la información, lo que dificulta el establecimiento de una norma o métrica estandarizada que sirva para la normalización de la información.
2. **La velocidad a la que se genera la información** es muy grande, lo que ocasiona problemas para su medición y gestión de manera fiable.

3. **El volumen de información y datos** que se generan a diario es inmenso, lo que puede provocar falta de precisión en los análisis y complejidad en los procesos.

Retos

La diversidad de fuentes de datos aporta abundantes tipos de datos y estructuras de datos complejas y aumenta la dificultad de la integración de datos.

El volumen de datos es tremendo, y es difícil juzgar la calidad de los datos dentro de un tiempo razonable.

Los datos cambian muy rápido y la "puntualidad" de los datos es muy corta, lo que requiere mayores requisitos para la tecnología de procesamiento.

No hay demasiados estándares de calidad de datos unificados y aprobados y la investigación sobre la calidad de datos de big data.

Estándares jerárquicos

1. Disponibilidad:

1.1. Accesibilidad:

1.1.1. Si se proporciona una interfaz de acceso a datos

1.1.2. Los datos pueden hacerse fácilmente públicos o fáciles de adquirir

1.2. Oportunidad:

1.2.1. Dentro de un tiempo dado, si los datos llegan a tiempo

1.2.2. Si los datos se actualizan regularmente

1.2.3. Si el intervalo de tiempo entre la recopilación y el procesamiento de los datos hasta la liberación cumple los requisitos

2. Usabilidad:

2.1. Credibilidad:

2.1.1. Los datos provienen de organizaciones especializadas de un país, campo o industria

2.1.2. Expertos o especialistas auditán regularmente y comprueban la exactitud del contenido de los datos

2.1.3. Los datos existen en el rango de valores conocidos o aceptables

3. Confiabilidad:

3.1. Exactitud

3.1.1. Los datos proporcionados son precisos

3.1.2. La representación de datos (o valor) refleja bien el estado real de la información de origen

3.1.3. La representación de información (datos) no causará ambigüedad

3.2. Consistencia:

3.2.1. Despues de procesar los datos, sus conceptos, dominios de valor y formatos todavía coinciden como antes de procesar

3.2.2. Durante un cierto tiempo, los datos permanecen consistentes y verificables

3.2.3. Todos los datos son consistentes o verificables

3.3. Integridad:

3.3.1. El formato de los datos es claro y cumple los criterios

3.3.2. Los datos son consistentes con la integridad estructural

3.3.3. Los datos son consistentes con la integridad del contenido

3.4. Completitud:

3.4.1. Si una deficiencia de un componente afectará el uso de los datos para datos con componentes múltiples

3.4.2. Si una deficiencia de un componente afectará la precisión y la integridad de los datos

4. Pertinencia:

4.1. Conveniencia:

4.1.1. Los datos recogidos no coinciden completamente con el tema, pero exponen un aspecto

4.1.2. La mayoría de los conjuntos de datos recuperados están dentro del tema de recuperación que los usuarios necesitan

4.1.3. El tema de la información proporciona coincidencias con el tema de recuperación de los usuarios

5. Calidad de presentación:

5.1. Legibilidad:

5.1.1. Los datos (contenido, formato, etc.) son claros y comprensibles

5.1.2. Es fácil juzgar que los datos facilitados satisfacen las necesidades

5.1.3. La descripción de los datos, la clasificación y el contenido de codificación satisfacen la especificación y son fáciles de entender

David Loshin



Presidente de Knowledge Integrity, Inc, es un líder intelectual reconocido y consultor experto en las áreas de calidad de datos, gestión de datos maestros e inteligencia comercial. Su libro, "Business Intelligence: The Savvy Manager's Guide" (junio de 2003) ha sido aclamado como un recurso que permite a los lectores "comprender la inteligencia empresarial, las disciplinas de gestión empresarial, el almacenamiento de datos y cómo todas las piezas funcionan juntas". Su libro, "Gestión de datos maestros", ha sido respaldado por líderes de la industria de gestión de datos.

El establece que la gestión de la calidad de los datos empieza con la identificación y la medición del efecto de los resultados empresariales. Se definen las reglas, se fijan los objetivos de rendimiento y se implementan los métodos de mejora de la calidad, así como la limpieza específica de datos y se implementan procesos de mejora.