|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| logo berri CIFP  CIFP Ciudad Jardín LHII | | | **Curso de Especialización:** Inteligencia artificial y Big Data | | | |
| **Modulo:**  Sistemas de Big Data  Big Data Aplicado | | | **Cód. Modulo:**  SBIG  BIGD |
| **Curso:** IABD | | | |
| **1ª Evaluación** | | | **Fecha**: 21/12/2022 | | | |
| **Cód. Prof:** PTI9  PTI3 | | **Profesor:** Andoni Inza  Ainara Montoya | |  | | |
|  | | | | | | |
| **Datos del Alumno** | | | | | | |
| **Nº** | **Nombre: JUAN ANTONIO** | | | | **Firma** | |
| **Apellido 1: RUIZ** | | | | |
| **Apellido 2: LARRUBIA** | | | | |

**Calificación:**

Para realizar este examen se aconseja utilizar los documentos de teoría y ejercicios realizados durante el trimestre actual, que están en Moodle. Está prohibido utilizar conexión a internet durante la prueba. Para aprobación del examen, habrá que puntuar mínimo en 2 de los ejercicios propuestos y que la puntuación total sea como mínimo 5.

Todos los pasos del examen tienen que estar documentados en este documento, todas las capturas de pantalla deberán enseñar el nombre del alumno, las modificaciones de archivos, el comando realizado y su salida.

## EJERCICIO 1

## HADOOP

### Creación de clúster Hadoop y yarn

Partiendo de la máquina virtual maestra proporcionada, monta un clúster de 2 nodos. Conexión NAT:

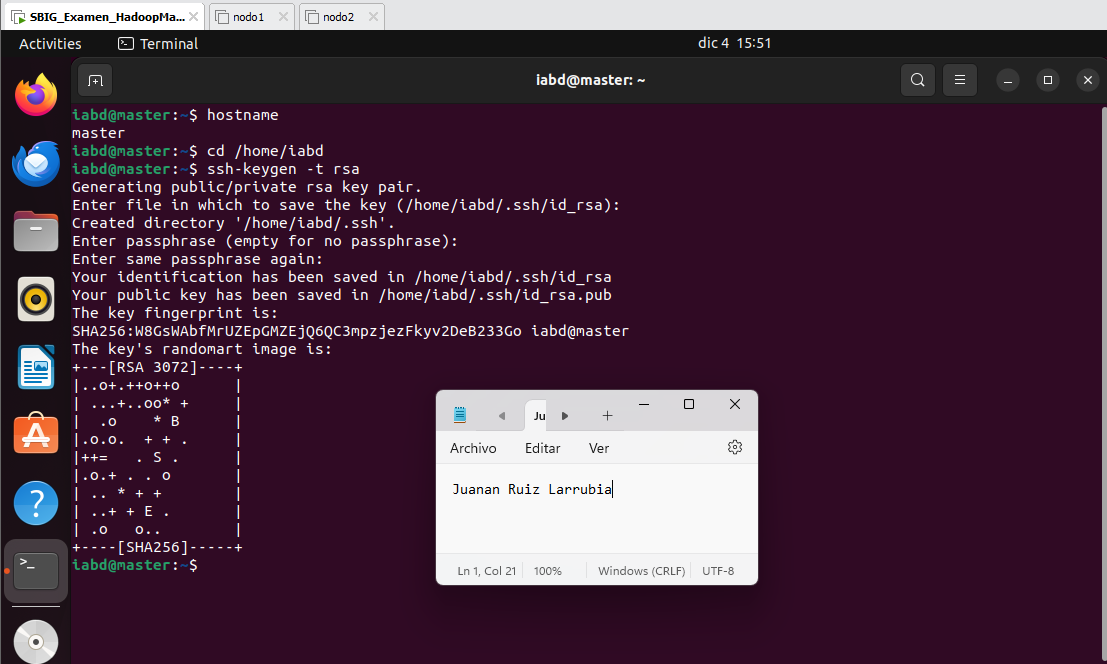
IP Maestro: 192.0.2.130

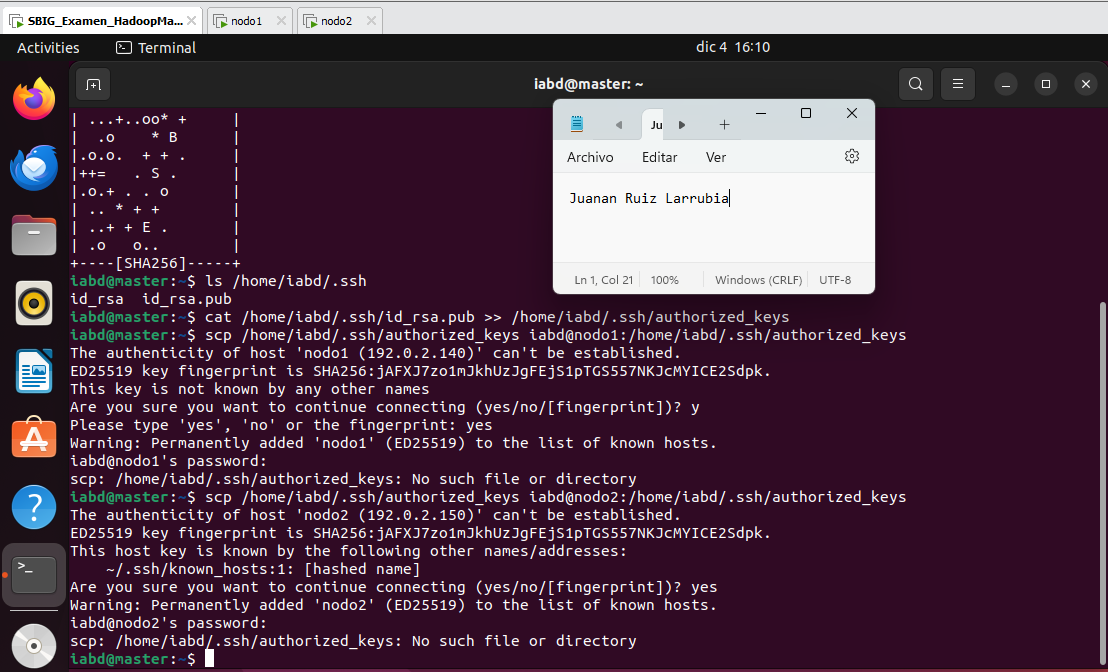
IP Nodo 1: 192.0.2.140

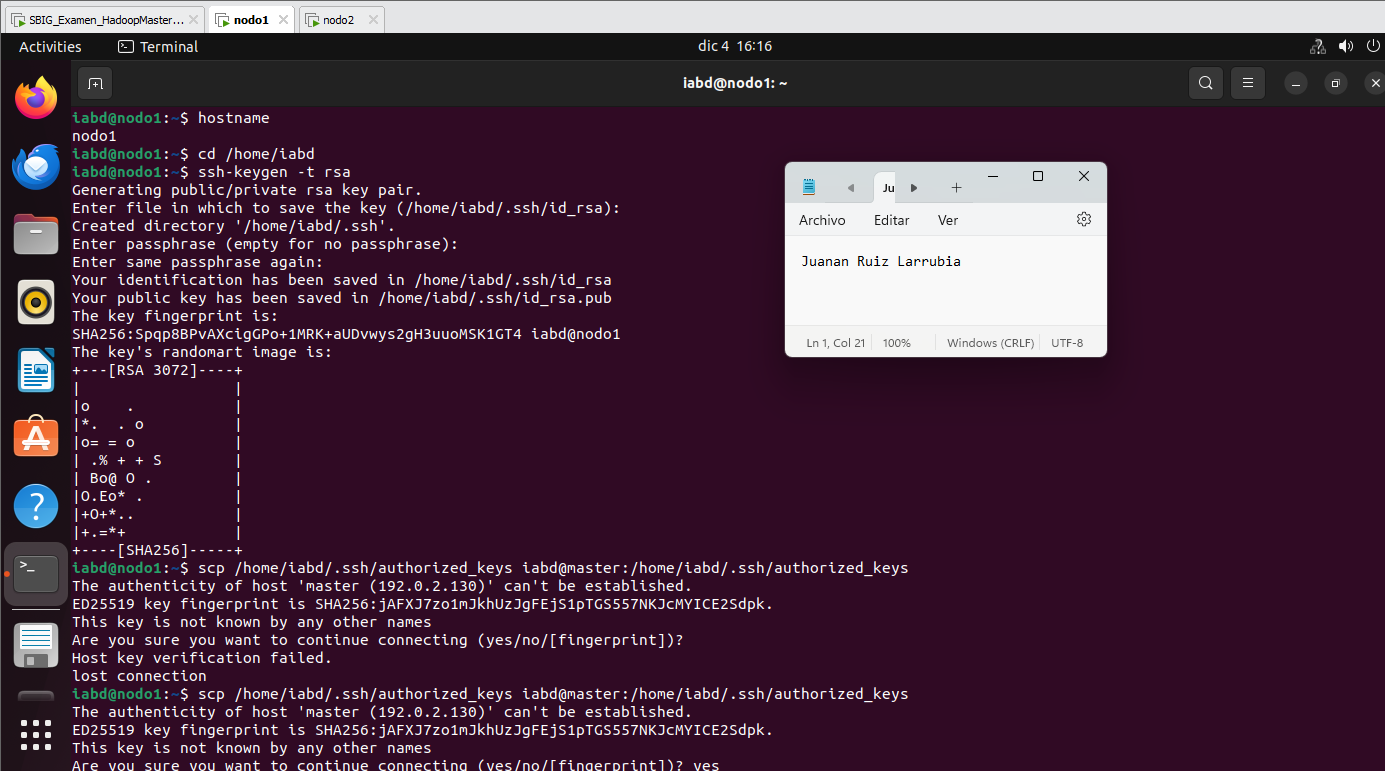
IP Nodo 2: 192.0.2.150

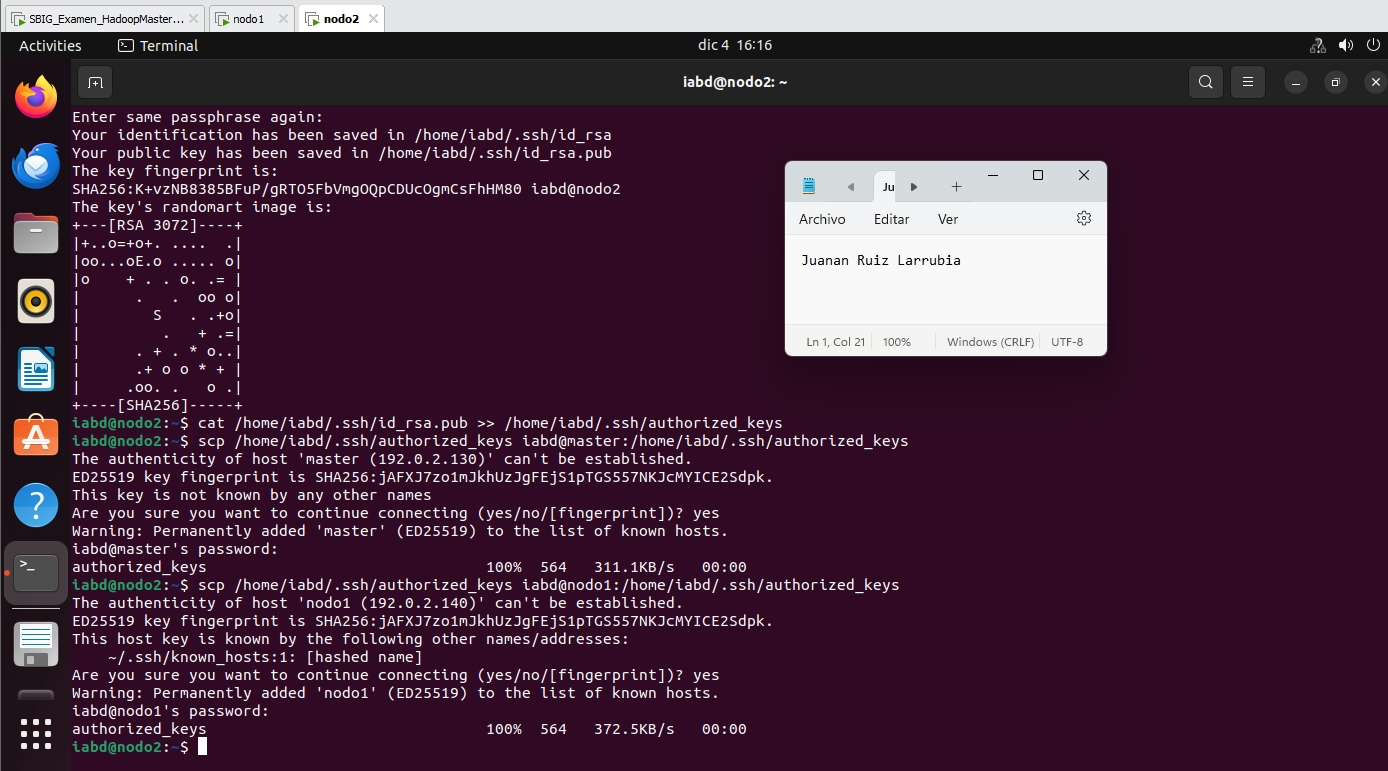
### Comprobacion DE FUNCIONAMIENTO clúster HADOOP

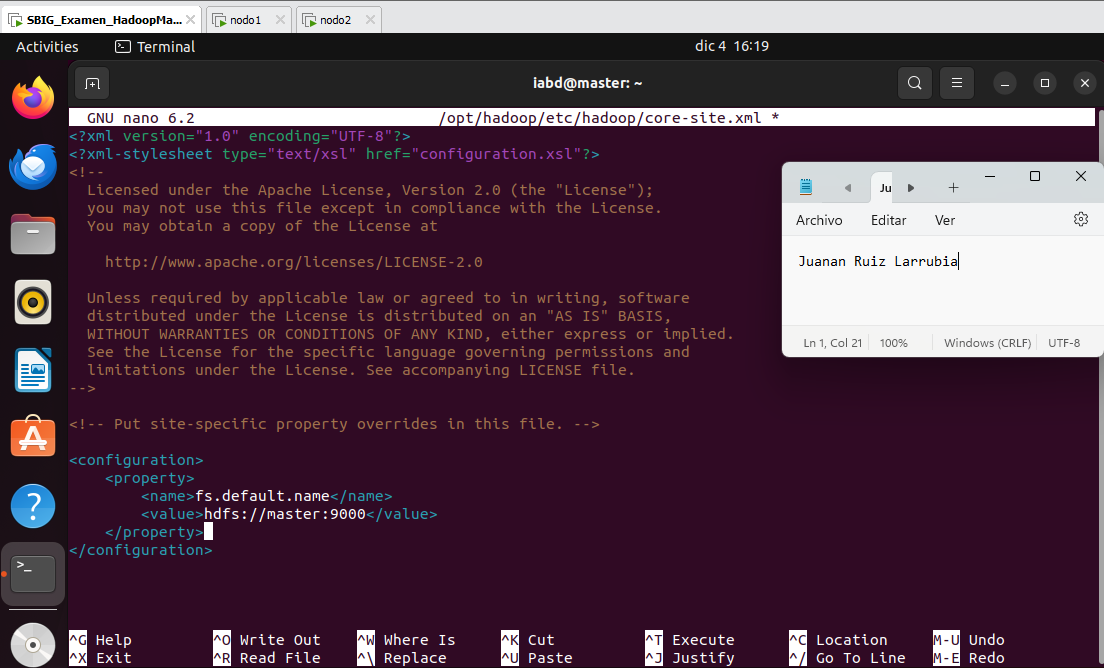
1. Pon en marcha el sistema hadoop, y verifica que están todos los nodos activos y en marcha.

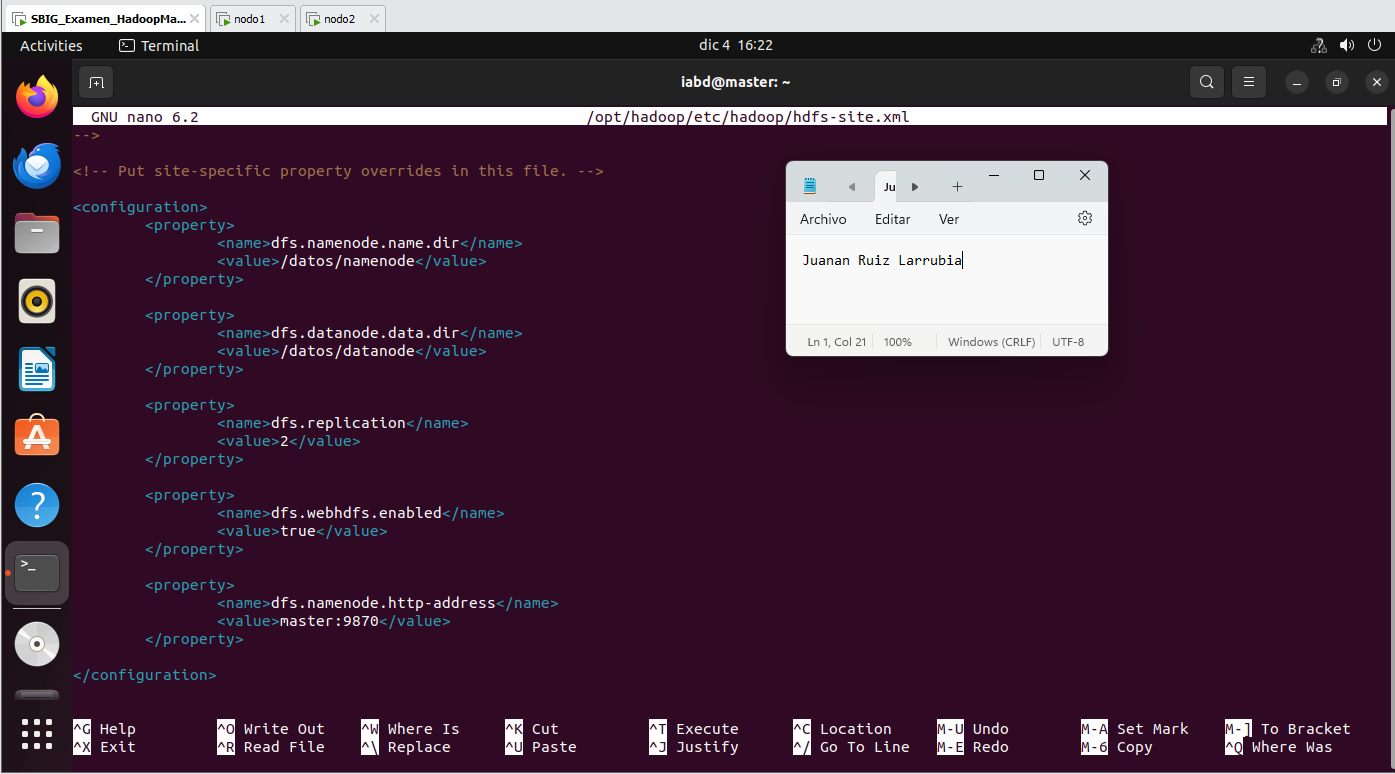




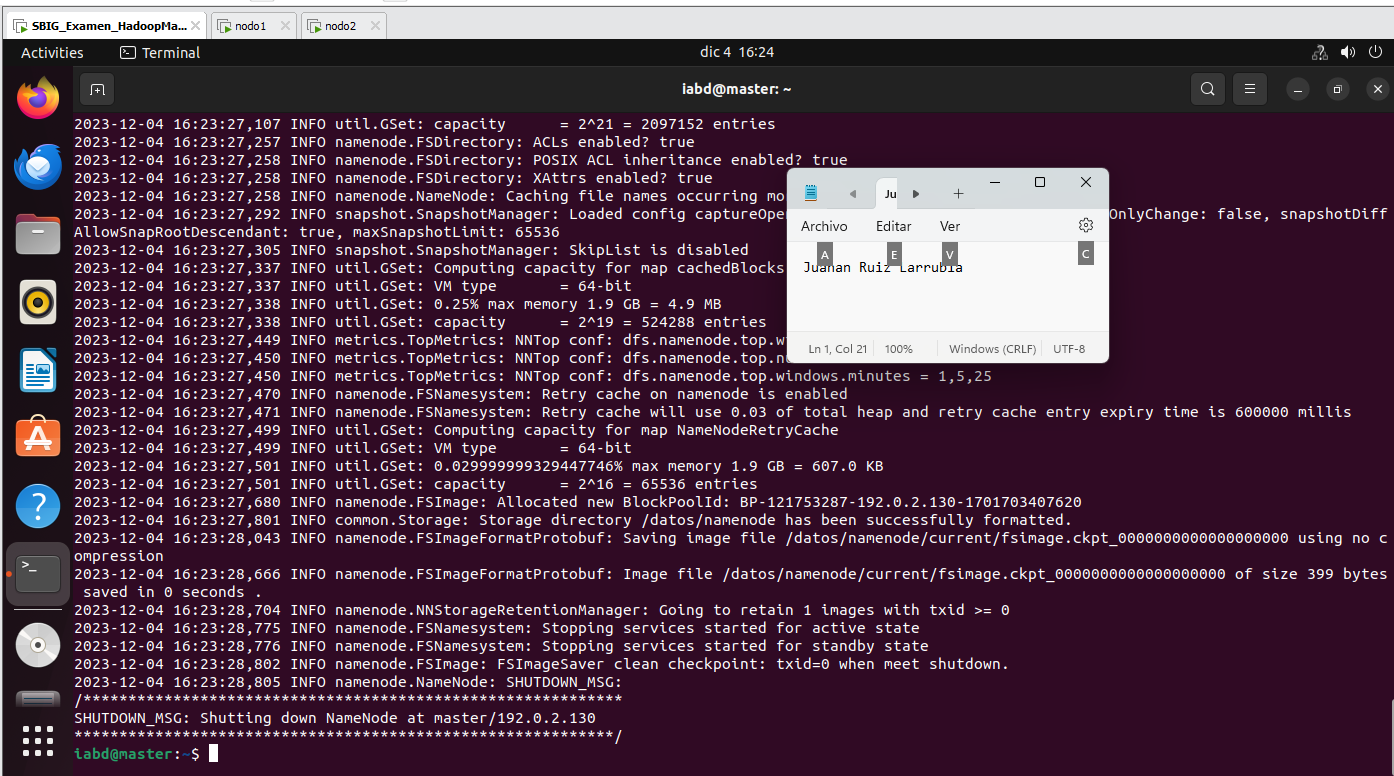


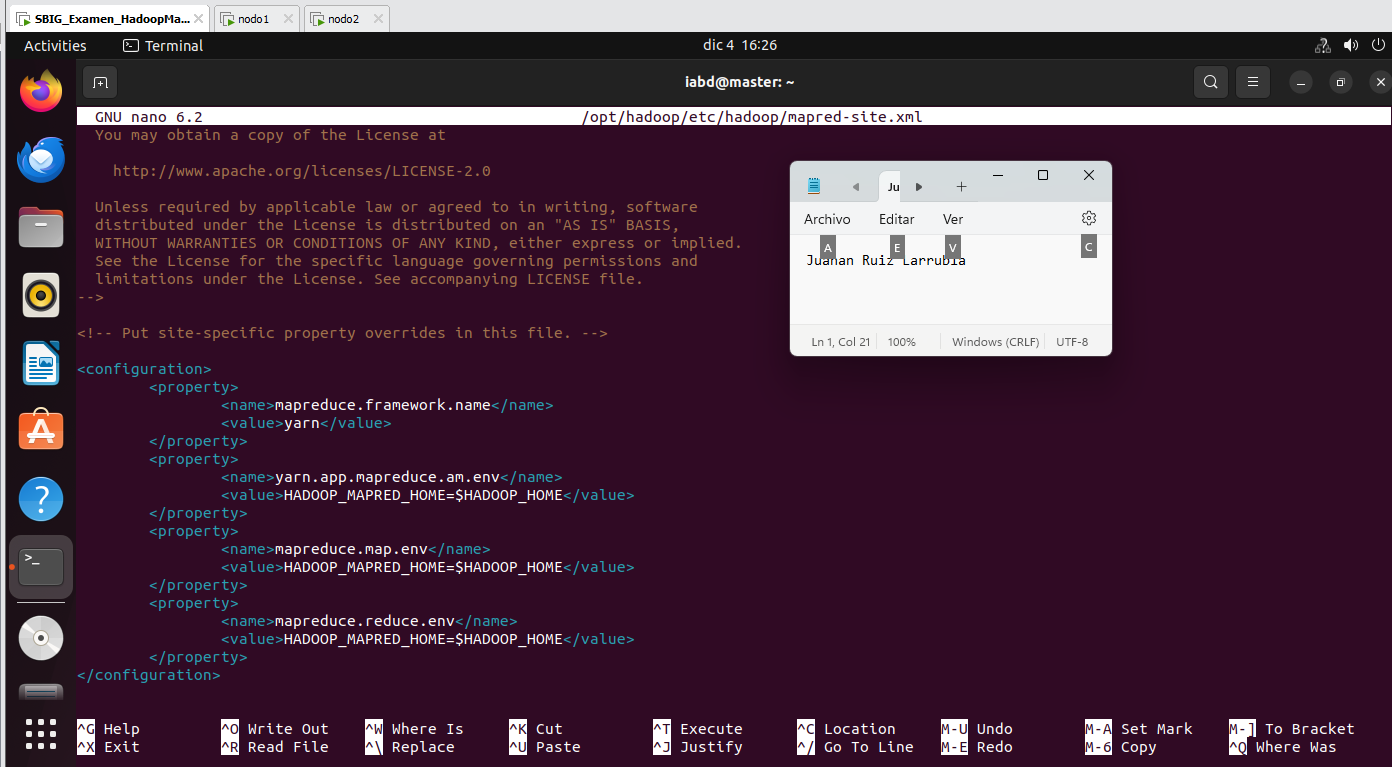


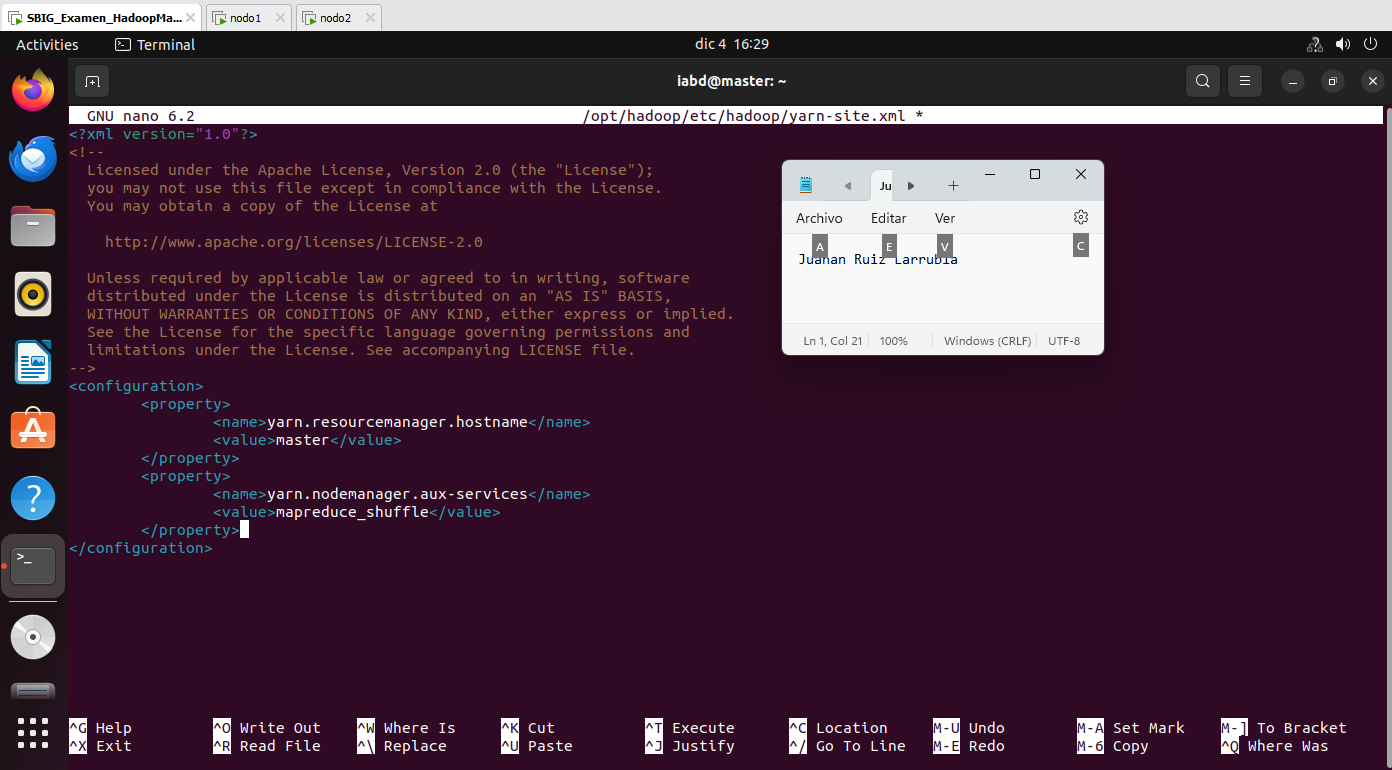


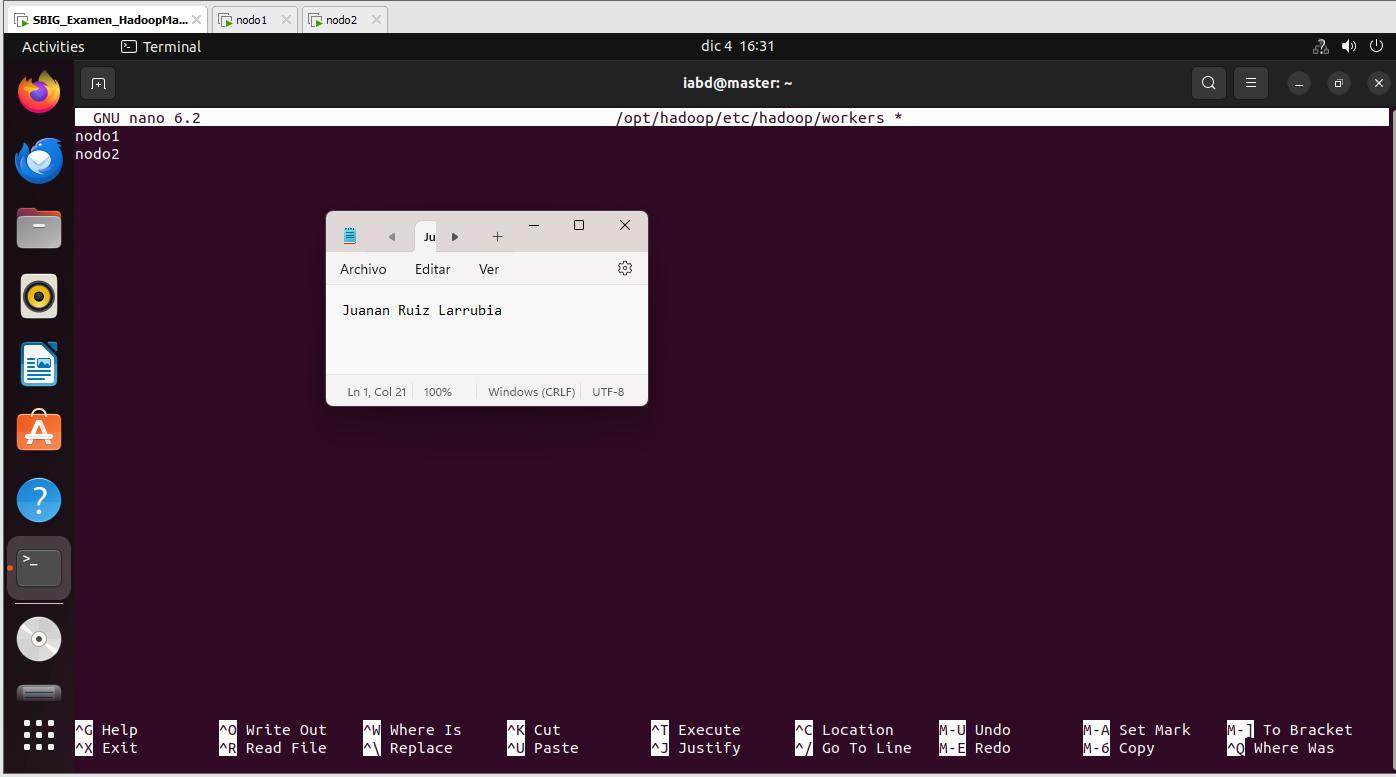


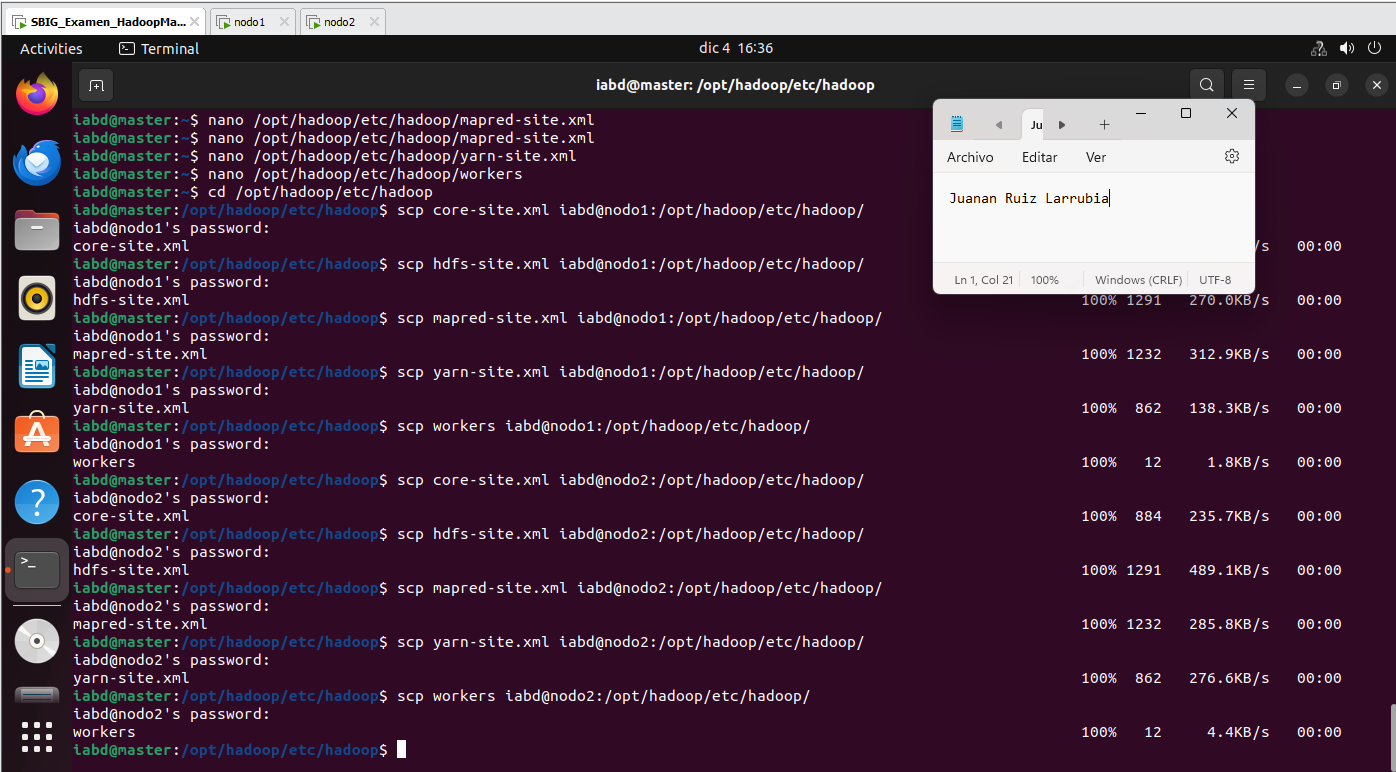
Verifica los servidores web de HDFS y YARN.

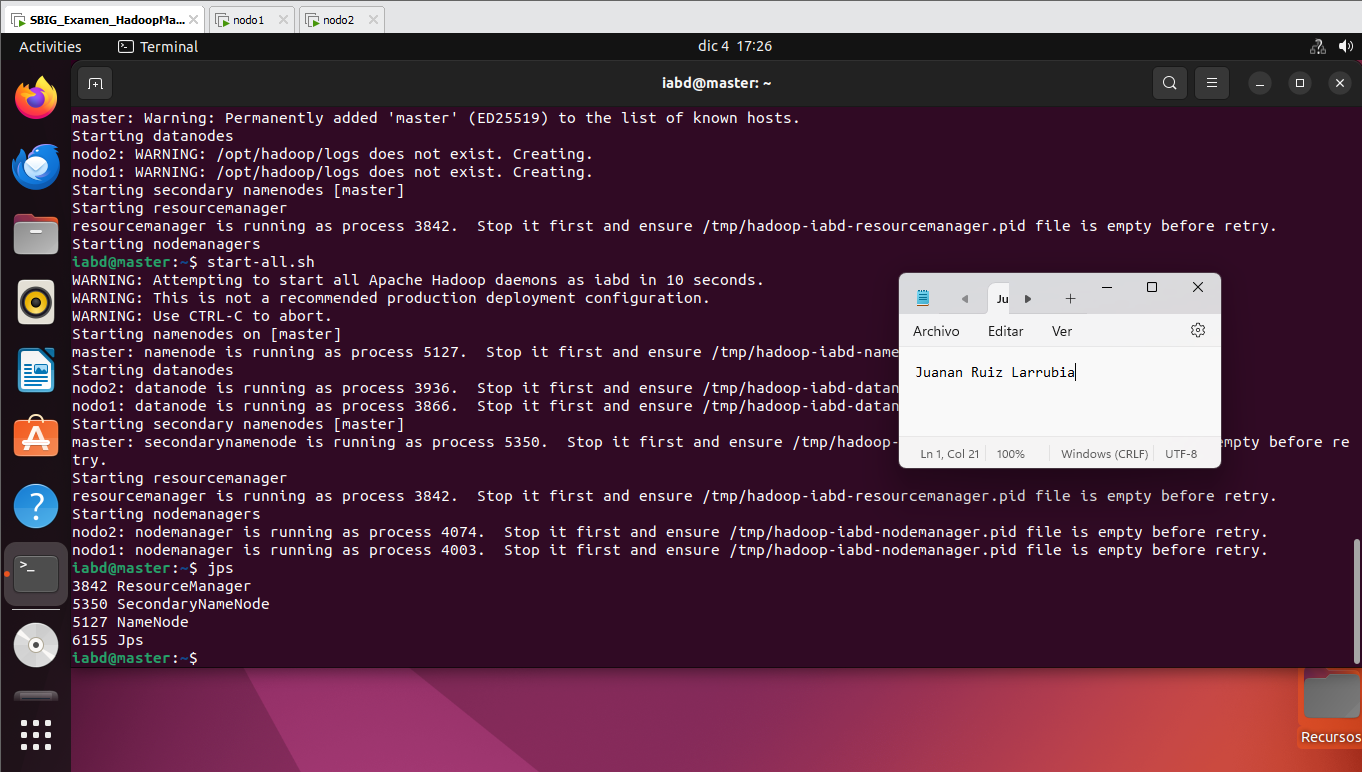




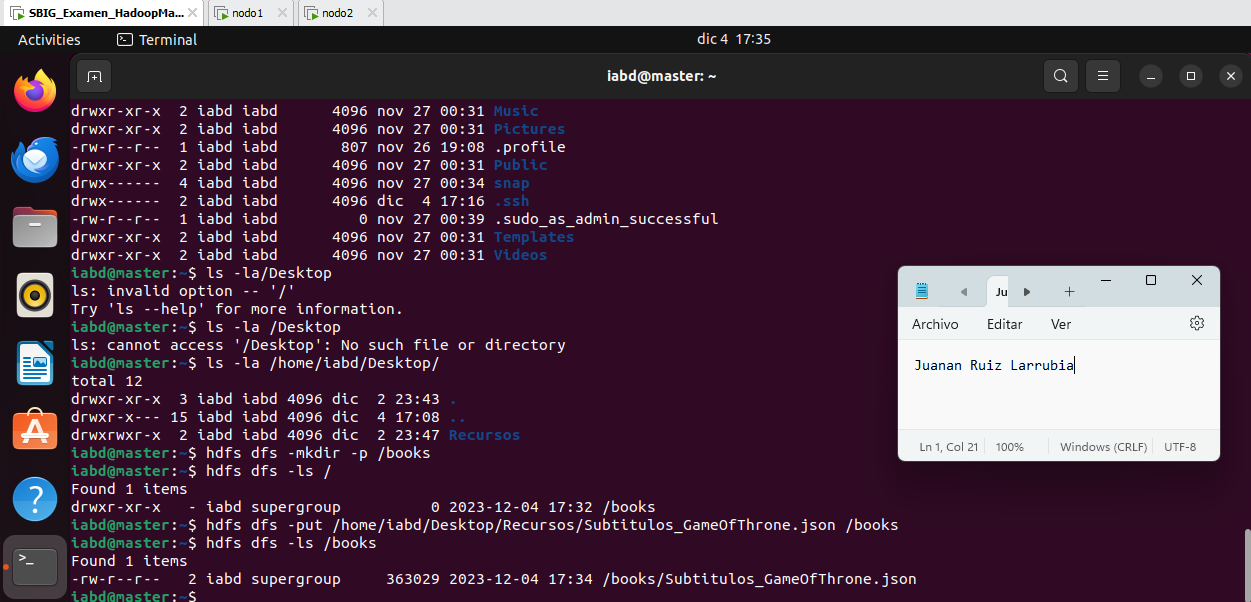




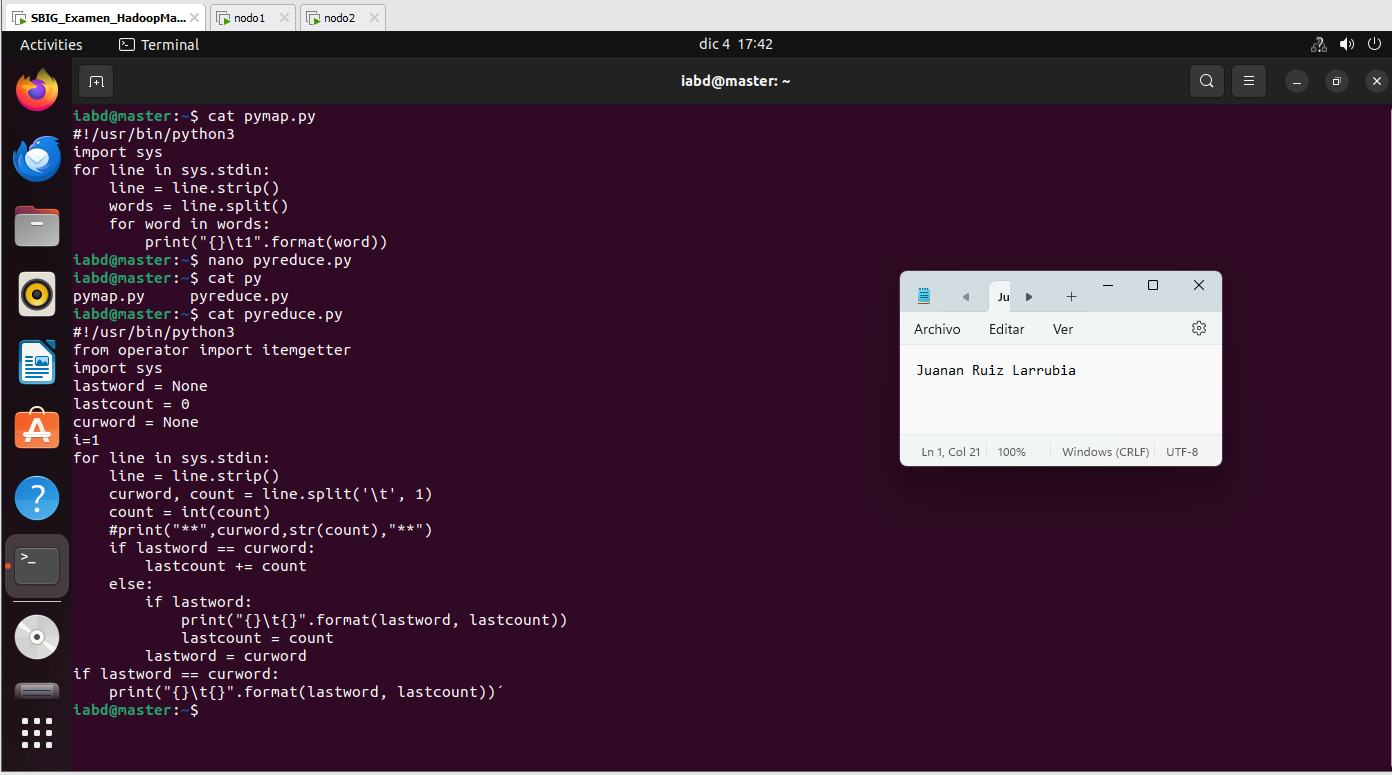


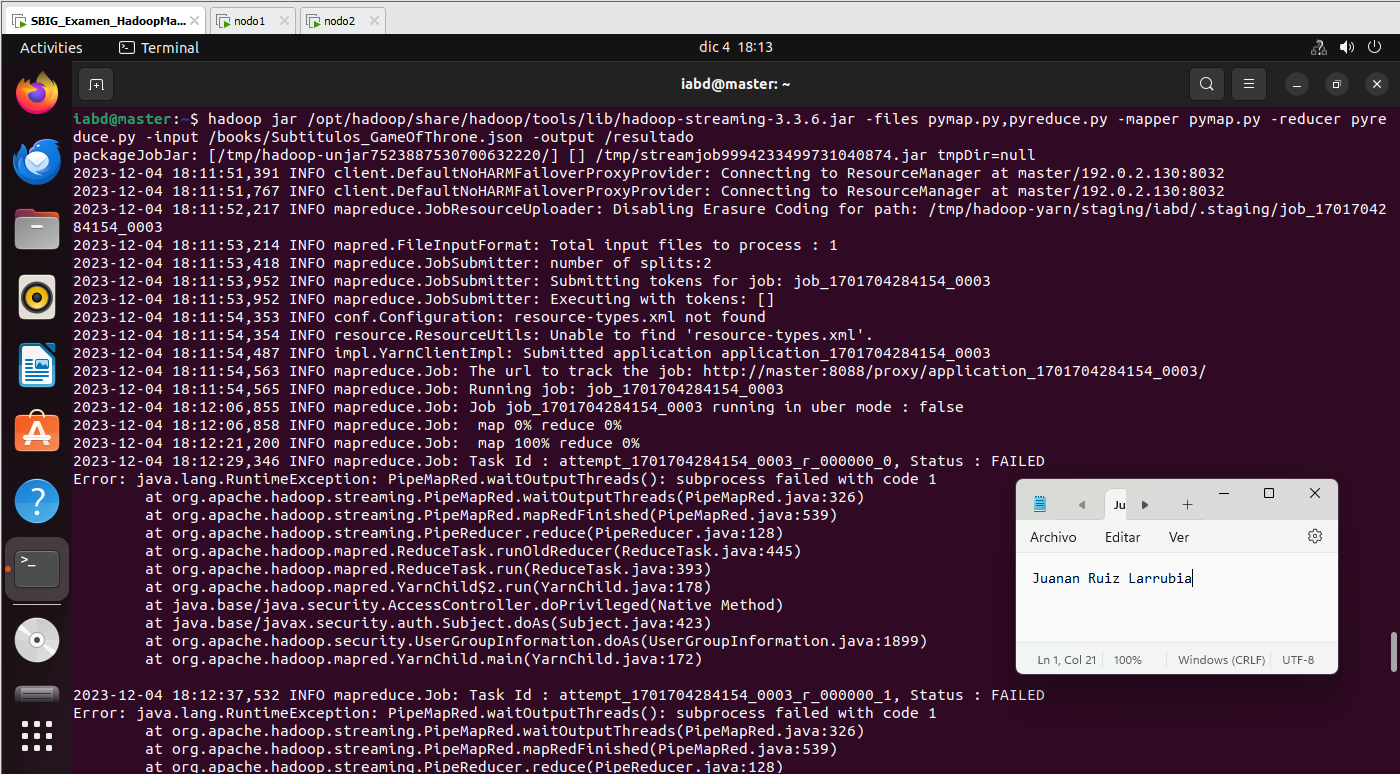


1. Crea una carpeta y sube un archivo. (usar fichero “Subtitulos\_GameOfThrone.json”)



1. Crear un “Wordcount” del archivo subido y verifica el archivo de salida.





Comprueba el tiempo de ejecución del wordcount. Muestra los parámetros más relevantes.

1. Apaga uno de los nodos, y verifica que los archivos son accesibles, y si la ejecución del “Wordcount” es afectado.

## EJERCICIO 2

## MONGODB

### Procesamiento de datos con MongoDB

**Creación de las colecciones y documentos de las Base de datos**

Considera el escenario de una aplicación de gestión de una biblioteca. Tu tarea es diseñar e implementar las colecciones necesarias en MongoDB para gestionar la información de libros, autores y préstamos. Utiliza las teorías de relaciones (1:1, 1:pocos, 1:muchos, muchos:muchos) para modelar de manera eficiente la base de datos.

Entidades Principales:

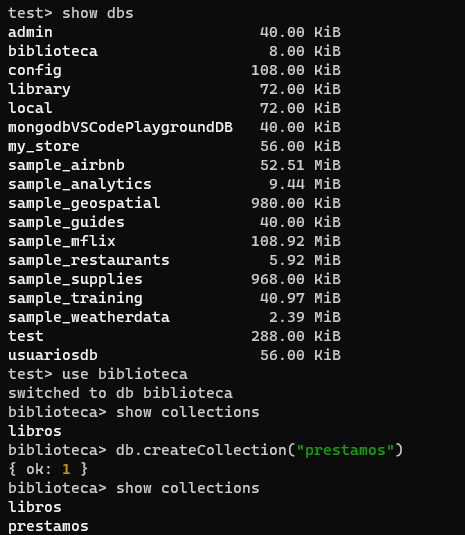
1. Cada libro tiene un ISBN único, título, año de publicación y género. Pueden tener uno o varios autores. 3 reseñas con calificación como máximo, y creadas por usuarios del apartado préstamo.
2. Autores: Numero único, nombre y nacionalidad;
3. Prestamos de los libros asociados a un libro y usuario, y puede tener varios prestamos asociados a lo largo del tiempo.
4. Un libro puede tener varios géneros.

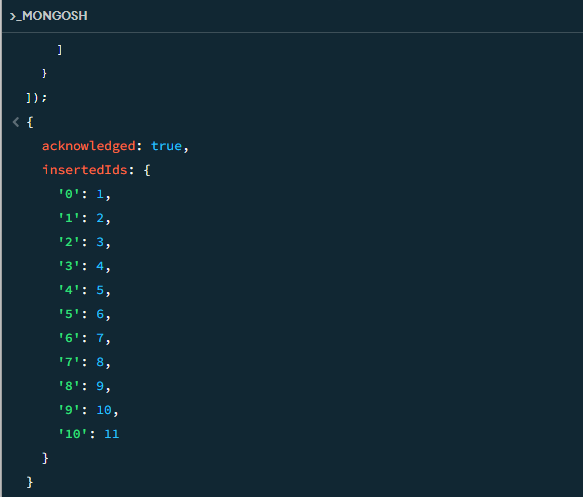
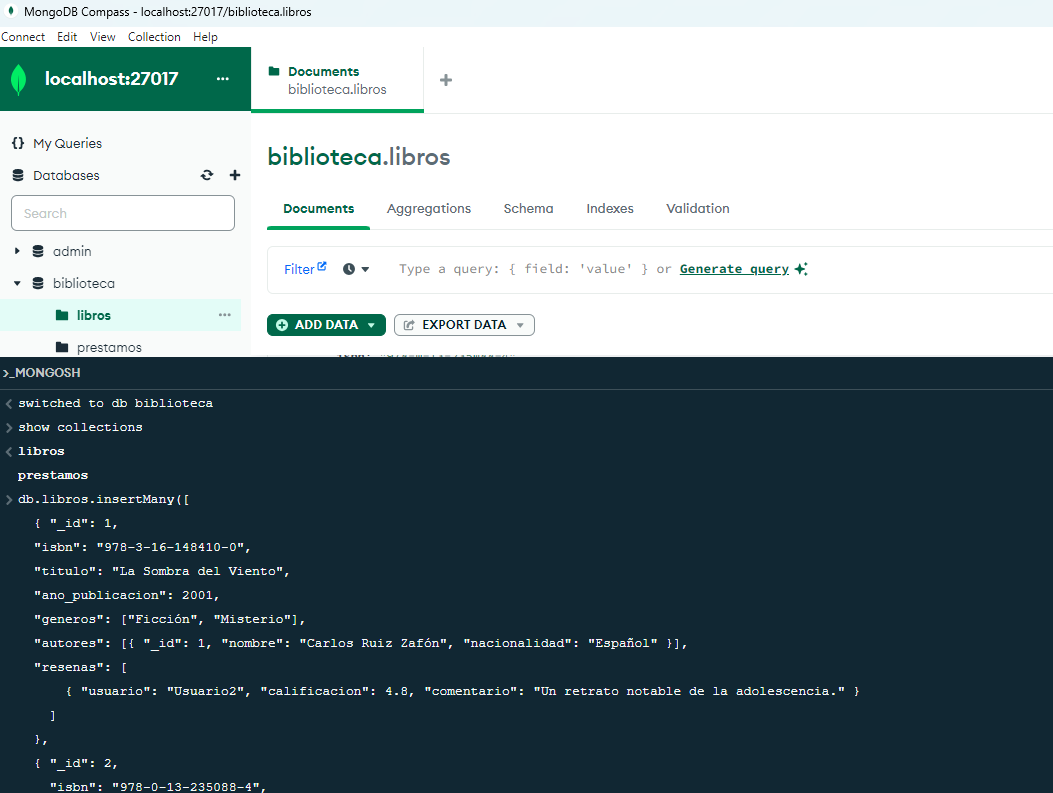
Asegúrate de incluir las referencias adecuadas en cada colección para establecer las relaciones. Utiliza identificadores únicos para enlazar los documentos entre colecciones.

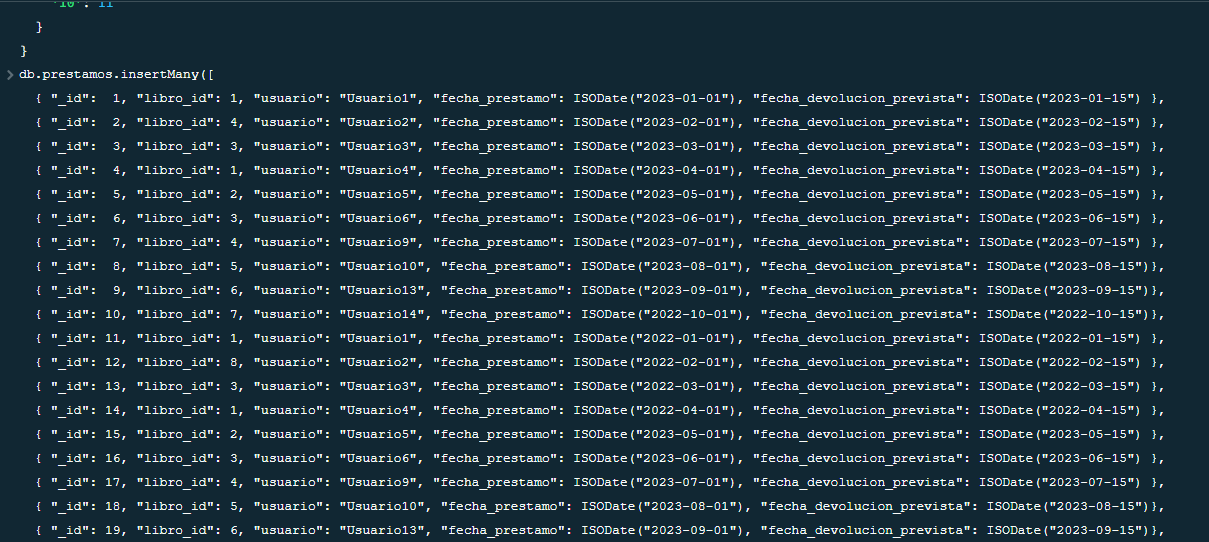
Diseña índices adecuados para optimizar las consultas relacionadas con las relaciones establecidas.

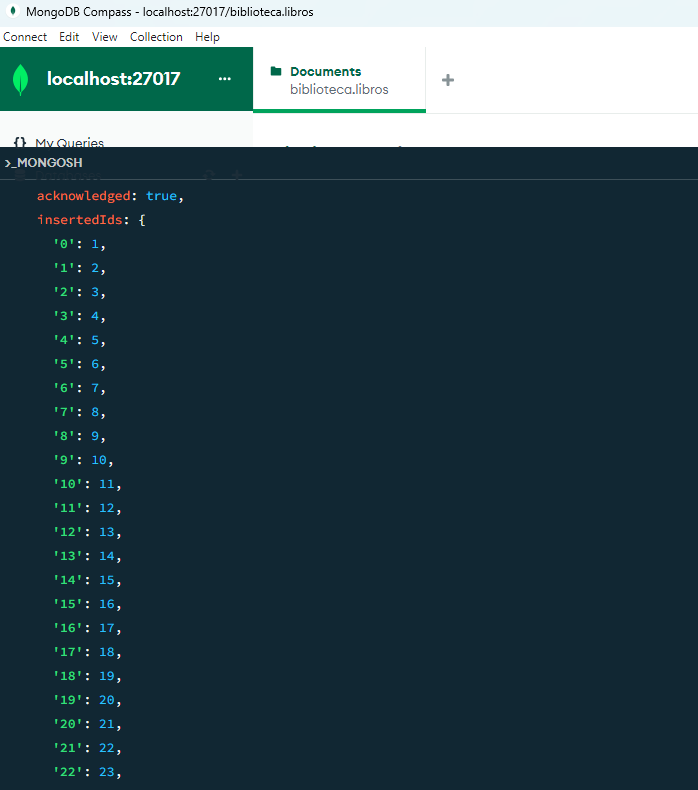
Crea ejemplos de documentos para cada colección que demuestren cómo se relacionan entre sí. Puedes incluir algunos datos ficticios para ilustrar el diseño de la base de datos.

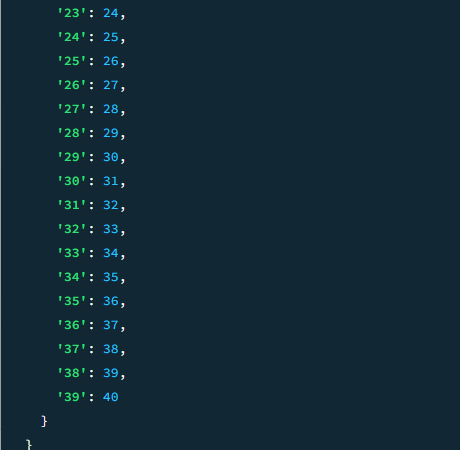
Crea una base de datos e inserta varios documentos en las diferentes colecciones para los ejercicios.

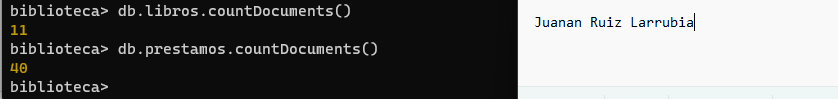


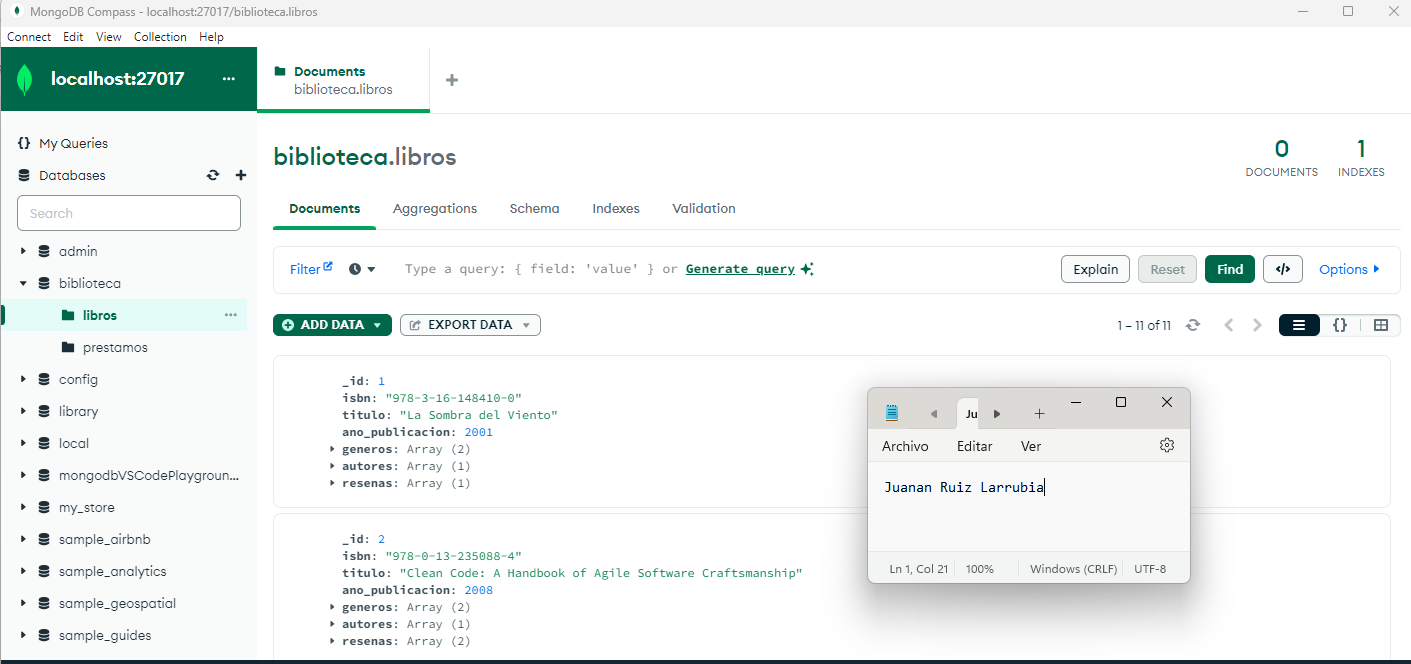


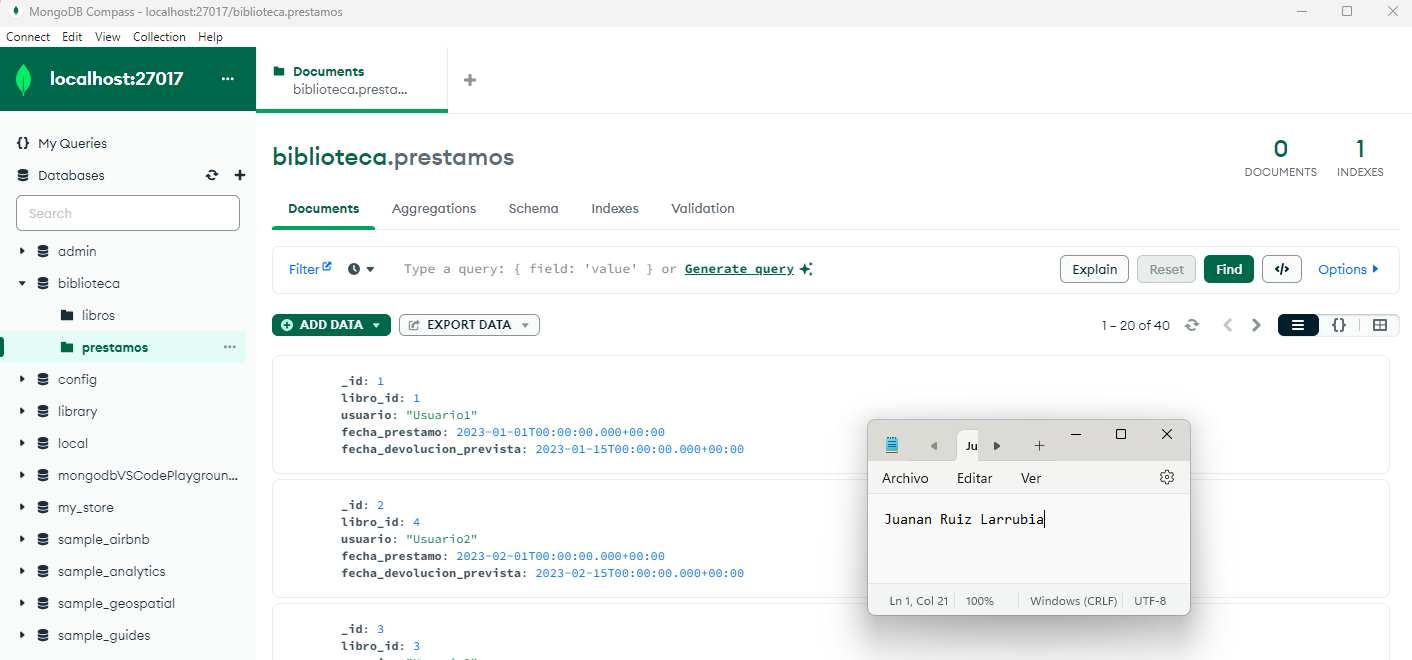






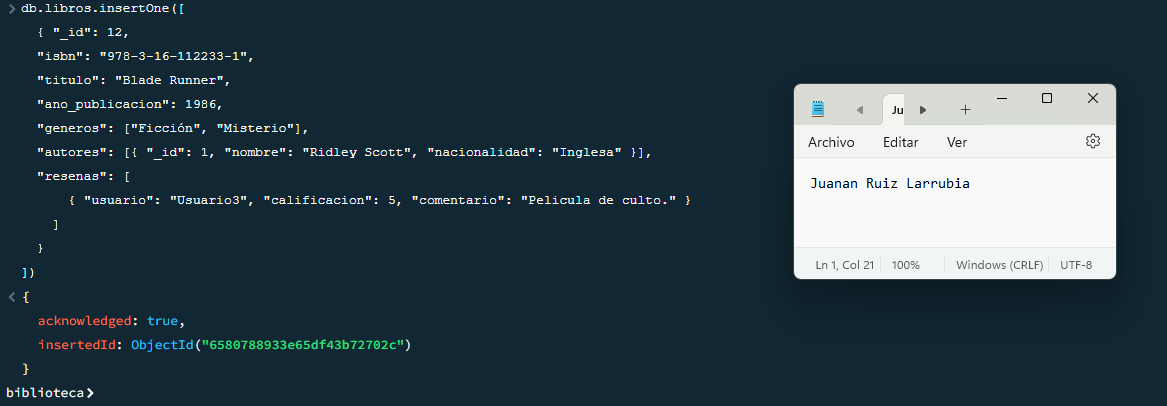


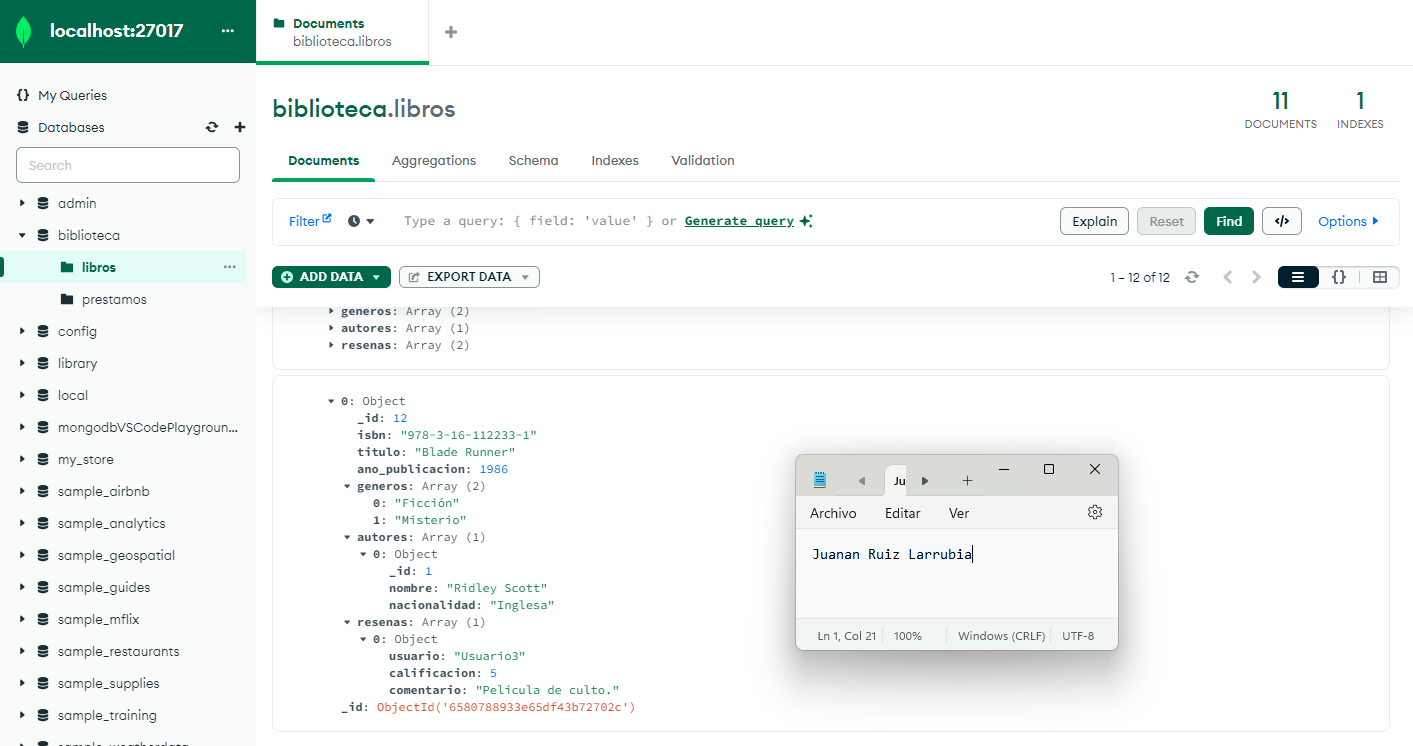
****

****

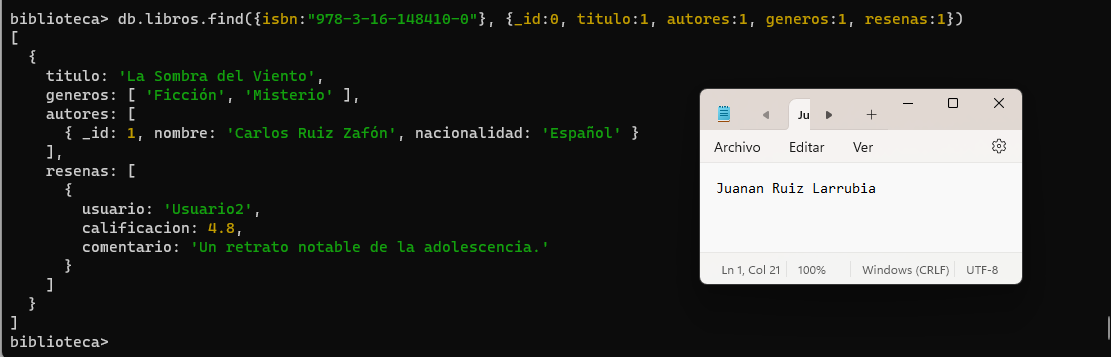
**Ejercicios CRUD**

1. Inserta un nuevo libro en la colección "Libros". Asegúrate de proporcionar valores para todos los campos necesarios, como ISBN, título, año de publicación, géneros, autores, etc.





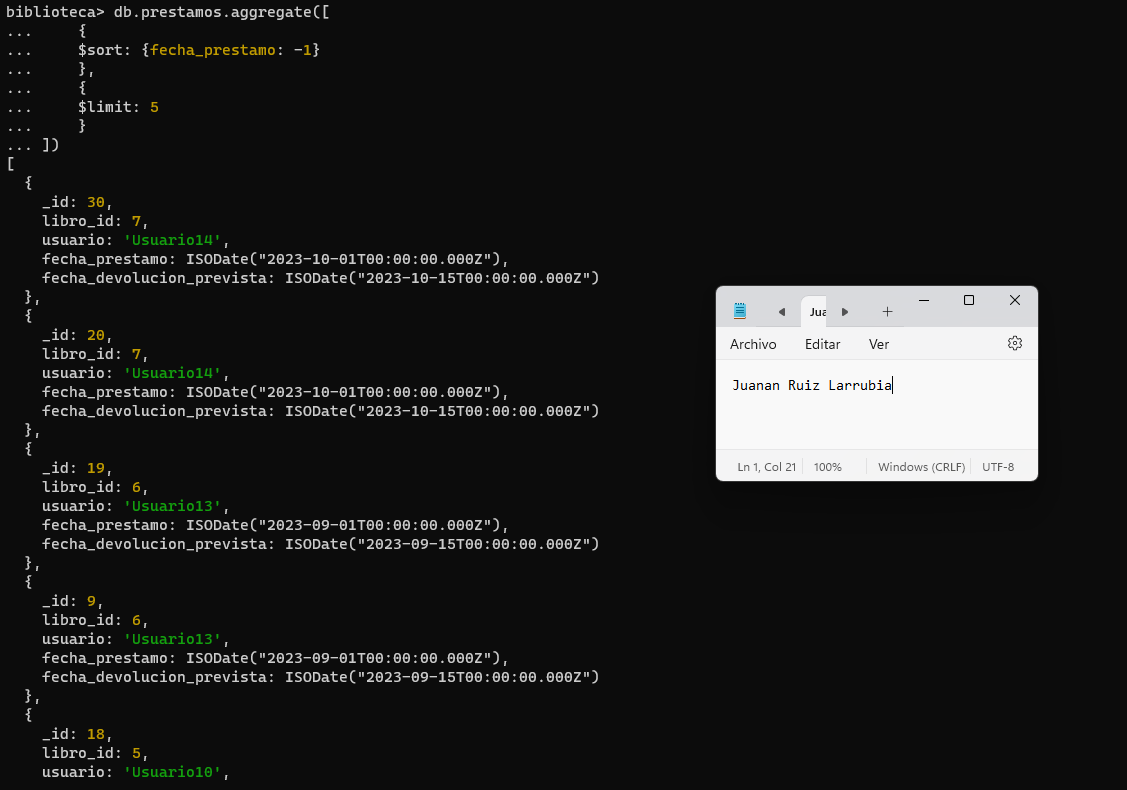
1. Recupera la información de un libro específico utilizando su ISBN. Asegúrate de visualizar solo detalles sobre el nombre del libro, los autores, géneros y reseñas asociadas.



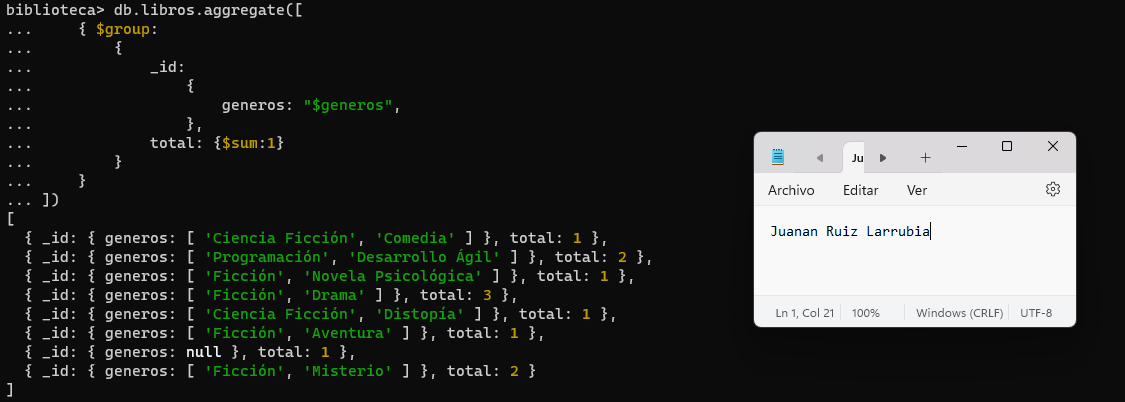
1. Inserta 2 préstamos nuevos asociados a algún usuario existente
2. Actualiza los datos de entrega de los prestamos anteriores

**Ejercicios AGREGACIONES.**

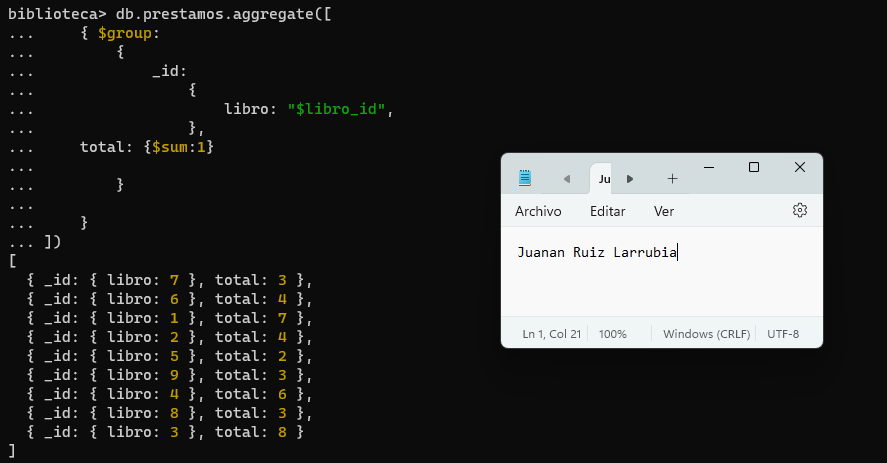
1. Obtiene los primeros 5 préstamos de la colección "Prestamos" ordenados por fecha de préstamo. (más nuevo a más antiguo)



1. Agrupa los libros por género y calcula el número total de libros en cada género.



1. Obtiene el libro con la mejor calificación, y visualiza su nombre, autor/es, reseña, usuario.
2. Agrupa los préstamos por libro y cuenta cuántas veces cada libro ha sido prestado. Ordena los resultados por el número de préstamos en orden descendente.



1. Lista de autores con la cantidad de libros escritos (filtrando aquellos con al menos 2 libros)

NOTA: Se podrá usar la terminal MongoSH, Python o la aplicación MongoCompass para realizar estos ejercicios. Se valorará usar diferentes vías de resolución.

### MONGODB REPLICAS

Crear dos conjuntos de réplicas de dos nodos cada uno, y a su vez, particionaremos los datos en dos y tres shaders respetivamente, usando docker-compose en una máquina virtual con Ubuntu 22.

Además, añadir un único router y un servidor de configuración.

Para el ejercicio, en la carpeta de recursos tenéis un fichero llamado images.json

1. Importar y particionar los datos por la id del image.json dataset.
2. Una vez cargado los datos, obtener el estado del sharding.
3. Si los datos no están particionados, forzar el split de los mismos, y obtener el estado del sharding.

## CRITERIOS DE EVALUACIÓN

|  |  |
| --- | --- |
| EJERCICIO 1 | PUNTUACIÓN |
| Correcta instalación y configuración de Hadoop y verificación de conectividad entre master y nodos | 0..5 |
| Creación exitosa de carpetas y correcta subida de archivos en el sistema de archivos distribuido HDFS | 0.5 |
| Ejecución exitosa de “Wordcount”, y verificación del resultado correctos | 0.5 |
| Extra: Realizar el “Wordcount” en python | 0.5 |
| Verificación de tiempo de ejecución en el entorno web | 0.5 |
| Apagado controlado de un nodo, y verificación de disponibilidad de archivos | 0.5 |

|  |  |
| --- | --- |
| EJERCICIO 2 | PUNTUACIÓN |
| Diseño y creación de la base de datos. Claridad y coherencia de los datos. Consideración adecuada de relaciones entre colecciones. | 2 |
| Implementación correcta de las consultas CRUD | 0.6 |
| Uso adecuado de operadores de agregación para obtener resultados deseados. | 2 |
| Extras: |  |
| Uso de un servidor en docker | 0.1 |
| Uso de diferentes medios para consultas | 0.3 |

|  |  |
| --- | --- |
| EJERCICIO 3 | PUNTUACIÓN |
| Correcta configuración de los ficheros Docker-compose.yml e init.sh de todos los elementos. | 1 |
| Importación de archivos y verificación de estado del sharding | 1 |