

Proyecto Programado 1

Diseño e implementación de una base de datos distribuida

1. Introducción

En el emocionante camino hacia la innovación y la eficiencia, nos encontramos ante la tarea de desarrollar un sistema distribuido para la biblioteca de una universidad. Este proyecto no solo representa un paso adelante en la vanguardia tecnológica, sino que también tiene el poder de transformar la forma en que estudiantes, profesores y personal interactúan con el vasto mundo del conocimiento.

Imaginemos un futuro en el que cada usuario puede acceder a la inmensa colección de recursos de nuestra biblioteca desde cualquier lugar del campus, e incluso desde la comodidad de sus hogares. Un sistema distribuido nos permitirá alcanzar esta visión, brindando acceso rápido y confiable a recursos, catálogos y servicios, enriqueciendo así la experiencia académica de todos los miembros de nuestra comunidad universitaria.

Un sistema distribuido no solo mejorará la accesibilidad a la información, sino que también permitirá una gestión más eficiente de los recursos bibliotecarios y reducirá la carga administrativa, liberando tiempo y energía para actividades más valiosas.

En el corazón de este proyecto se encuentra la **colaboración**. Nuestro equipo de desarrollo tendrá la oportunidad de trabajar codo a codo con expertos en tecnología, bibliotecarios y usuarios finales para diseñar y crear un sistema que satisfaga verdaderamente las necesidades de nuestra comunidad. A medida que enfrentamos los desafíos técnicos y estratégicos, también crearemos lazos más fuertes entre diferentes departamentos, fortaleciendo la cohesión y el sentido de propósito en toda la institución.

Usted y su equipo han sido seleccionados por la universidad para desarrollar un servicio que permita registrar la información para las operaciones de la biblioteca.

2. ¿Qué se busca con este proyecto?

- Mejorar las habilidades de resolución de problemas.
- Aumentar el conocimiento del estudiante sobre bases de datos distribuidas.
- Practicar la experimentación y la resolución de problemas (divide y vencerás).
- Ejercitar la toma de decisiones.
- Fomentar el trabajo grupal.
- Fomentar la investigación por parte del estudiante sobre cómo:
 - Implementar una base de datos distribuida en DBMS en particular.

- Aplicar los conocimientos teóricos de fragmentación y asignación en un entorno controlado.
- Formato de paper propuesto utilizando el formato IEEE

3. Contexto y las funcionalidades esperadas

Se le solicita diseñar una implementación para un sistema distribuido en el TEC, el cuál será utilizado por la biblioteca para gestionar las nuevas adquisiciones y generar indicadores en tiempo real.

En este momento, no se tiene ninguna experiencia previa con sistemas distribuidos en la institución; por lo que se le solicita a usted y su equipo de consultores en tecnología una demostración de la propuesta de implementación antes de aprobar el pago del saldo pendiente de la licitación.

Uno de los requisitos de la contratación, estipula que se requiere que el sistema funcione como un cluster ya que se planea agregar nuevos equipos al sistema distribuido en un futuro cercano con el fin de aumentar la capacidad del mismo y se sabe que también aumentarán los registros almacenados en el sistema.

El siguiente enunciado desarrolla las características solicitadas por parte del TEC para el sistema contratado.

Infraestructura

El sistema tendrá inicialmente un equipo para la fragmentación en el Campus Tecnológico de San Carlos, otro en el Campus Tecnológico de Limón y un tercero en el Campus Tecnológico de Alajuela. Un servidor de configuración y uno de los enrutadores serán administrados en el Campus Central de Cartago. El segundo servidor de configuración y el segundo enrutador estarán en el Campus Tecnológico de San José. El tercer router y el tercer servidor de configuración se encontrarán en la UCR y estarán conectados mediante una VPN.

Función	Provincia	Cantón	Nombre del equipo	Nombre del grupo	Puerto de publicación
Shard	San Carlos	Santa Clara	Betelgeuse	shard-orion	27018
Shard	Limón	Limón	Rigel		27018
Shard	Alajuela	Alajuela	Alnilam		27018
router	Cartago	Oriental	Antares	router-scorpio	27017
router	San José	San José	Shaula		27017
router	San José	Montes de Oca	Sargas		27017
configuration	Cartago	Oriental	Sirio	config-CanMayor	27019

configuration	San José	San José	Aldebarán		27019
configuration	San José	Montes de Oca	Elnath		27019

De momento, se cuenta con autorización del Departamento de Tecnologías de Información y Comunicación para utilizar los puertos 27017, 27018 y 27019 para ser utilizados a nivel interno. Cualquier otro puerto, no podrá ser utilizado para la transmisión de datos.

Si tras el análisis del sistema solicitado, consideran que hay una distribución mejor para los nodos; la puede aplicar siempre y cuando aparezca en la documentación la justificación del cambio. Es decir, la función de un nodo específico puede cambiar, pero no la ubicación de los equipos “entregados”.

Además, los equipos pueden ser virtualizados mediante dockers o máquinas virtuales, o bien, pueden optar por conectar 9 nodos independientes en una red para la revisión.

Licencias

Por los recortes presupuestarios del gobierno central, no es posible incurrir en nuevos costos de licencias. De modo que se requiere un sistema escalable a largo plazo que minimice los costos en software principal.

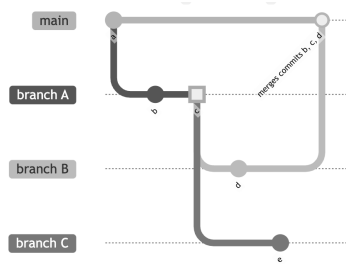
Dado que la implementación se realizará principalmente mediante servidores virtuales geográficamente distribuidos, se tiene libertad en la selección del sistema operativo siempre y cuando no se incurra en gastos adicionales por el rubro de licencias.

Por políticas institucionales, los datos no pueden salir de la organización, así que tampoco se pueden utilizar servicios en la nube para la instalación y se debe utilizar la versión de la base de datos en mongodb de la comunidad apropiadamente configurada para trabajar en un clúster con los nueve nodos solicitados.

Mecanismos de acceso y control de versiones

Para introducir los nuevos registros, se debe implementar la GUI que permita interactuar con la BD y realizar operaciones CRUD, lo cuál puede ser vía web o aplicación de escritorio.

El equipo de consultoría tiene libertad para seleccionar el software a utilizar y deberá utilizar un sistema de versionamiento de código (GitHub, GitLab o algún otro similar) con el fin de dar un seguimiento apropiado de las versiones utilizadas en el sistema y poder trazar los distintos cambios aplicados. Se recomienda trabajar al menos con ramas de producción y desarrollo, con un manejo independiente de los features desarrollados:



Fuente: <https://docs.gitlab.com/ee/user/project/repository/branches/>

Descripción de la base de datos

Para el sistema se licitó una base de datos llamada bibliotecaTEC donde actualmente se tienen los datos de algunos libros vendidos en Amazon.com, pero se ingresarán posteriormente todos los registros disponibles en la biblioteca de la institución de forma manual.

Para cada libro se desean registrar al menos los siguientes atributos, pero si su análisis refleja que se requiere más información, lo pueden ampliar a su gusto:

- Name: título del libro
- Author: autor(a) del libro
- User rating: evaluación dada por los lectores o clientes
- Reviews: cantidad de comentarios sobre el libro
- Price: precio del libro, en dólares
- Year: año en que se publicó
- Genre: género literario al que pertenece el libro

Una muestra de estos datos se encuentran en el archivo bsamazon.json, el cual se puede importar dentro de bibliotecaTEC para utilizarlo dentro de su demostración o para pruebas de las consultas.

Por entrevistas realizadas con los expertos del negocio, se sabe que históricamente la biblioteca ha tenido un proceso de adquisiciones y ubicaciones un poco desorganizado, por ejemplo:

- Los libros de ciencia ficción se consumen principalmente en Cartago.
- Los libros de arquitectura solo se utilizan en San José, pero se encuentran distribuidos físicamente por todos los campus.
- En Alajuela hay problemas para entregar los libros porque estos solo pueden llegar una vez por semana con los mensajeros de los otros campus.
- Los usuarios de los sistemas, casi siempre realizan las búsquedas por tema, más que por autor o título del libro.
- En Alajuela solo se tienen estudiantes del año 2019 en adelante.

El sistema que su equipo de consultores desarrollará, no va a poder solucionar todos los problemas de logística de la institución; pero buscará que al menos los sistemas funcionen lo mejor posible para los usuarios. El segundo problema que debe afrontar es la disponibilidad de libros por sede.

Por ejemplo, el siguiente es un fragmento del estado de la biblioteca:

```
{ "usuarios": [ {
  "id": 1, "nombre": "Juan Pérez",
  "libros_alquilados": [101, 203],
  "fecha_alquiler": "20/08/2023"
},
{
  "id": 2,
  "nombre": "María García",
  "libros_alquilados": [105, 307, 401] ,
  "fecha_alquiler": "15/01/2023"
}
],
"libros": [
  { "id": 101,
    "titulo": "Cien años de soledad",
    "autor": "Gabriel García Márquez",
    "disponible": false ... },
  { "id": 203,
    "titulo": "1984",
    "autor": "George Orwell",
    "disponible": false ... },
  { "id": 105,
    "titulo": "El principito",
    "autor": "Antoine de Saint-Exupéry",
    "disponible": false ... },
  ...
]
}
```

En este caso, aplican las siguientes reglas:

- Existen tantos registros, como alquileres existan por usuarios (1 registro por cada alquiler).
- El id del estudiante es su carnet.
- La fecha del alquiler es importante porque los libros se devuelven 15 días después de su préstamo, en caso de pasarse de la fecha establecida, se procede a aplicar un cobro de 2000 colones por cada libro alquilado por cada semana de retraso.
 - Los libros con calificación de 4 o más, se cobran a 3500 cuando incurre en retraso porque son libros de alta demanda y afectan a los demás usuarios de la biblioteca.
- Se necesita llevar el control del número de libros disponibles para no intentar alquilar libros sin existencia.
- Cuando se devuelva un libro, se debe cambiar el estado a disponible.
- Se considera que un libro está extraviado cuando ha pasado más de un mes desde que se considera vencida su fecha de devolución.
- El registro de los alquileres podría incluir la carrera de los estudiantes y el campus donde se encuentran matriculados.

Perfiles de uso del sistema:

Los administradores del sistema, necesitan hacer un seguimiento de los principales factores que contribuyen al funcionamiento y éxito del negocio, en específico, necesitan un dashboard y una serie de aplicativos para el seguimiento de las adquisiciones.

En el cartel de licitación se solicitaron tres perfiles de usuario para el sistema:

- Bibliotecarios nivel 1:
 - Accederán al sistema mediante una interfaz gráfica sencilla a criterio de los consultores que les permitirá realizar las operaciones CRUD en las diferentes tablas de la base de datos.
 - Debe permitir consultar la información de los libros o de los usuarios registrados para la biblioteca.
 - El aplicativo puede ser en el lenguaje de programación deseado y estará conectado a Antares.
- Bibliotecarios nivel 2:
 - Requieren un dashboard en PowerBI, donde se pueda observar lo siguiente:
 - Resumen General:
 - Número total de libros en la biblioteca.
 - Número de libros disponibles para préstamo.
 - Número de libros en préstamo.
 - Número de usuarios registrados.
 - Estado de los Libros:
 - Gráficos de barras o pastel que muestren la distribución de libros por género.
 - Gráfico de barras que muestre el número de libros disponibles y en préstamo por categoría.
 - Actividad de Préstamos:
 - Gráfico de líneas que muestre la tendencia histórica de préstamos en el tiempo.
 - Gráfico de barras que muestre los libros más populares en función de la cantidad de préstamos.
 - Gráfico de barras que muestre los usuarios con más préstamos.
 - Gráfico con la cantidad máxima de Reviews recibida para cada género literario.
 - Top 10 de libros mejor evaluados por los lectores.
 - Historial de Préstamos:
 - Tabla que muestre los préstamos actuales y su estado (libro, usuario, fecha de préstamo, fecha de devolución esperada).
 - Gráfico de barras que muestre los libros con menor stock y que necesitan ser reabastecidos.
 - Notificaciones y Alertas:
 - Sección que muestre notificaciones sobre libros próximos a su fecha de devolución, libros vencidos, etc.
 - El dashboard estará conectado a Shaula.
- Servicio de cobros:
 - Requiere de una interfaz gráfica sencilla, a criterio del equipo de consultores.
 - Estará conectado a Sargas.
 - Este paquete o software, debe permitir generar:

- Reporte de morosos y monto a recaudar unificado por cada persona que alquiló libros.
 - Este reporte debe incluir el lugar donde fue alquilado el libro y la fecha más antigua de la que se tengan pendientes.
- Reporte de libros en condición de “extraviados”
 - Debe emitir un registro por cada libro extraviado, la cantidad disponible, la cantidad extraviada y los campus donde han desaparecido.
- Los reportes se deben poder exportar en formato PDF con la fecha de emisión.

Cada uno de los perfiles es excluyente en sus funciones, por lo que es necesario tener distintos roles de autenticación para la aplicación.

El sistema debe hacer validación de usuarios para permitir que solo usuarios registrados utilicen el sistema y que las funcionalidades utilizadas por los mismos, sean las que corresponden según su rol.

4. Por Entregar

- Diseñar el sistema (las aplicaciones y la base de datos) de manera distribuida:
 - a. Diseñar la fragmentación y la asignación. Debe justificar claramente el porqué de la fragmentación que realiza y el porqué de la asignación.
 - b. La aplicación debe tener un módulo para que los empleados de las sucursales puedan registrar la entrada y salida de paquetes; un módulo para los clientes para consultar sus paquetes y otro para los administradores o gerentes (los administradores están asociados a las sucursales y los gerentes a las oficinas centrales). El sistema debe hacer validación de usuarios para permitir que solo usuarios registrados utilicen el sistema y que las funcionalidades utilizadas por los mismos, sean las que corresponden según su rol.
- Se debe implementar el sistema de manera distribuida según el diseño propuesto en el punto 1.
- Generar un paper donde se describa la metodología utilizada para generar la base de datos distribuida en el DBMS seleccionado, el paper debe contener las siguientes secciones: introducción, descripción del problema, un resumen de la investigación realizada para implementar una base de datos distribuida en el DBMS seleccionado, un ejemplo de la implementación de la base de datos distribuida para el contexto de este proyecto y referencias bibliográficas; conclusión. Se debe utilizar el formato de IEEE para papers

5. Aspectos administrativos

- Entregar un documento con al menos los siguientes apartados. Los apartados DEBEN entregarse en ese orden:
 - Portada.
 - Aspectos generales: indicar en este apartado cuáles herramientas se utilizaron para implementar el sistema y el DBMS con el cual se implementó la base de datos distribuida.
 - Debe incluir también el estado actual del proyecto: si es 100% funcional o bien indicar qué no se pudo implementar y por qué.

- Diseño de la base de datos (describir tablas y sus campos). Incluya el modelo (diagrama de la base de datos, antes de la fragmentación)
 - Diseño de la fragmentación: Indique claramente sobre cuáles tablas hace fragmentación y cuáles predicados utiliza. Debe explicar el diseño.
 - Diseño de la asignación, debe indicar claramente qué tablas y fragmentos quedaron en cada nodo y justificar por qué.
 - Para cada consulta solicitada se debe mostrar un screenshot donde se observe claramente la consulta realizada y su resultado.
 - Debe incluir screenshots para las consultas realizadas por los distintos perfiles: Bibliotecarios nivel 1, Bibliotecarios nivel 2 y servicio de cobros:
 - Incluye la captura del dashboard y los aplicativos diseñados para interactuar con la BD.
 - Si deciden optar por los puntos extra de la reasignación de nodos, se debe incluir una sección pequeña que se respalde en el análisis realizado sobre la aplicación para justificar su cambio en la infraestructura.
- Si alguna funcionalidad queda sin documentar, no será tomada en cuenta durante la evaluación y el grupo de trabajo perderá el puntaje respectivo.

6. Condiciones finales

Este proyecto programado se rige por las siguientes condiciones:

- Debe cumplir con todo lo indicado en esta especificación.
- Deberá entregarse en tiempo y forma según el plazo establecido por la profesora al momento de la lectura de este documento.
- Las aplicaciones deben poder ejecutarse desde cualquiera de los tres nodos simultáneamente inclusive.
- No se recibirán proyectos fuera de la fecha/hora indicada, tampoco se recibirán por correo electrónico ni otro medio que no sea el TECDigital. Todo trabajo entregado después de la fecha y hora indicada tendrá una nota de cero.
- Si se detecta copia o plagio se procederá según lo indicado en el Reglamento del Régimen de Enseñanza Aprendizaje del ITCR.
- Las citas de revisión del proyecto se asignan después de la entrega del mismo.
- Para la revisión, el grupo de trabajo deberá llevar los equipos necesarios para poder mostrar el sistema distribuido en funcionamiento.
- No se harán revisiones previas de los diseños.
- Se debe implementar el sistema de manera distribuida según el diseño propuesto.
- El estudiante puede seleccionar el software que utilizará para hacer la interfaz gráfica del sistema, excepto para el dashboard, que se le solicitará trabajar con PowerBI.
- El estudiante deberá utilizar un sistema de versionamiento de código (GitHub, GitLab o algún otro similar).
- La implementación de los nodos **DEBE** realizarse con la versión de mongodb para la comunidad. Cualquier implementación realizada con mongoAtlas no recibirá puntos en este rubro.
- El Sistema Operativo seleccionado, queda a criterio del estudiante siempre y cuando no se incurra en pagos por servicios; por ejemplo, usar servicios en la nube ya preconfigurados.

7. Entrega:

La entrega de este proyecto es el domingo 3 de Setiembre del 2023 a las 11:55pm, deben subir un archivo .rar o .zip con el código fuente en el TEC Digital, este archivo además debe contener un archivo txt con la dirección del repositorio del software de versionamiento.

Posteriormente se asignarán citas de revisión para este proyecto.

8. Criterios de Evaluación

Rubro	Puntaje
Configuración del entorno distribuido	17
Implementación de 3 nodos para sharding	1
Al apagar un nodo del sharding, el resto de la información sigue disponible	2
Implementación de 3 nodos para configuración.	1
Al apagar el nodo primario de configuración, se realiza automáticamente la reasignación a otro nodo.	2
Implementación de 3 nodos para enrutamiento.	1
Al apagar un nodo de enrutamiento, los otros dos siguen funcionando.	2
Adecuada asignación de puertos para los grupos de shards, configs y routers	3
Adecuada asignación de nombres de grupos de equipos	1
Adecuada asignación de nombres de equipos	1
Configuración de tarjetas de red para acceder desde equipos externos a la MV o docker construido.	2
Configuración de IPs estáticas para direccionamiento de los equipos	1
Propuesta de mejora para la distribución de los nodos válida.	** 5 pts extra
Nota: La implementación de los nodos DEBE realizarse con la versión de mongodb para la comunidad. Cualquier implementación realizada con mongoAtlas o con un servicio sujeto a pagos para escalar el sistema no recibirá puntos en este rubro.	

El Sistema Operativo seleccionado, queda a criterio del estudiante siempre y cuando no se incurra en pagos por servicios; por ejemplo, usar servicios en la nube ya preconfigurados.

Base de datos distribuida	40
A través de la aplicación permite consultar información de cualquier nodo desde la sede central aún si el nodo no está disponible	5
Implementación correcta y coherente de la asignación	7.5
Implementación correcta y coherente de la fragmentación	7.5
Rediseño de la base de datos para soportar las operaciones contratadas	5
Paquetes registrados en los nodos se reflejan inmediatamente en el nodo central	10
Gestión de roles y autenticación para las aplicaciones	5
Requerimientos de la aplicación	30
Configuración e Implementación de algún software de control de versiones para el proyecto con evidencia del trabajo en equipo	2
Bibliotecarios nivel 1- CRUD de libros	2
Bibliotecarios nivel 1- CRUD de Préstamos de libros	2
Bibliotecarios nivel 1- Consulta información de libros	1
Bibliotecarios nivel 1- Consulta información de Préstamos de libros	1
Bibliotecarios nivel 1- Conectado a Antares	1
Bibliotecarios nivel 2 - Uso de PowerBI para integración de la solución solicitada	2
Bibliotecarios nivel 2 - Resumen General- Número total de libros en la biblioteca.	1
Bibliotecarios nivel 2 -Resumen General- Número de libros disponibles para préstamo.	1
Bibliotecarios nivel 2 -Resumen General- Número de libros en préstamo.	1

Bibliotecarios nivel 2 -Resumen General- Número de usuarios registrados.	1
Bibliotecarios nivel 2 -Estado de los Libros- Gráficos de barras o pastel que muestran la distribución de libros por género.	1
Bibliotecarios nivel 2 - Estado de los Libros - Gráfico de barras que muestre el número de libros disponibles y en préstamo por categoría.	1
Bibliotecarios nivel 2 - Actividad de Préstamos - Gráfico de líneas que muestre la tendencia histórica de préstamos en el tiempo.	1
Bibliotecarios nivel 2 - Actividad de Préstamos - Gráfico de barras que muestre los libros más populares en función de la cantidad de préstamos.	1
Bibliotecarios nivel 2 - Actividad de Préstamos - Gráfico de barras que muestre los usuarios con más préstamos.	1
Bibliotecarios nivel 2 - Actividad de Préstamos - Gráfico con la cantidad máxima de Reviews recibida para cada género literario.	1
Bibliotecarios nivel 2 - Actividad de Préstamos - Top 10 de libros mejor evaluados por los lectores.	1
Bibliotecarios nivel 2 - Historial de Préstamos - Tabla que muestre los préstamos actuales y su estado (libro, usuario, fecha de préstamo, fecha de devolución esperada).	1
Bibliotecarios nivel 2 - Historial de Préstamos - Gráfico de barras que muestre los libros con menor stock y que necesitan ser reabastecidos.	1
Bibliotecarios nivel 2 - Notificaciones y Alertas - Sección que muestre notificaciones sobre libros próximos a su fecha de devolución, libros vencidos, etc.	1
Bibliotecarios nivel 2 - Conectado a Shaula	1
Servicio de cobros - Reporte de morosos y monto a recaudar unificado por cada persona que alquiló libros	1
Servicio de cobros - Reporte de libros en condición de "extraviados"	1
Servicio de cobros - Exporta a PDF	1

Servicio de cobros - Conectado a Sargas	1
Documentación	13
Elaboración del artículo usando el formato de la IEEE.	1
Aspectos generales: indicar en este apartado cuáles herramientas se utilizaron para implementar el sistema y el DBMS con el cual se implementó la base de datos distribuida.	1
Diseño de la infraestructura implementada (si usaron la propuesta o alguna ajustada por ustedes)	1
Estado actual del proyecto entregado: si es 100% funcional o bien indicar qué no se pudo implementar y por qué.	1
Diseño de la base de datos (describir tablas y sus campos). Incluya el modelo (diagrama de la base de datos, antes de la fragmentación)	1
Diseño de la fragmentación: Indique claramente sobre cuáles tablas hace fragmentación y cuáles predicados utiliza. Debe explicar el diseño.	2
Diseño de la asignación, debe indicar claramente qué tablas y fragmentos quedaron en cada nodo y justificar por qué.	2
Para cada consulta solicitada (incluidas las de los reportes) se debe mostrar un screenshot donde se observe claramente la consulta realizada y su resultado obtenido.	2
Screenshots para las consultas realizadas por los distintos perfiles: Bibliotecarios nivel 1, Bibliotecarios nivel 2 y servicio de cobros: <ul style="list-style-type: none"> Incluye la captura del dashboard y los aplicativos diseñados para interactuar con la BD. 	2