

Proyecto

Johan Rodríguez

Mathew Carballo

Instituto Tecnológico de Costa Rica

IC4302 - Bases de Datos II

Ing. Efrén Jiménez

November 15, 2023

Tabla de contenidos

Resumen ejecutivo	1
Introducción	2
Objetivos	2
Objetivo general	2
Objetivos específicos	3
Descripción del problema	3
Historia de Home Credit	3
Solución a Construir	4
Desarrollo	5
Modelo E-R	5
Modelo Relacional	6
Migrador	7
Servidor de Auditoría	7
Funciones de CRUD	8
Triggers	8
Seguridad	9
Índices no Clúster	9
Consultas SARGABLES	9
Aplicación Web	9
Conclusiones	10
Recomendaciones	11
References	12

Resumen ejecutivo

El proyecto de optimización de bases de datos para Home Credit se destaca por su enfoque exhaustivo en la gestión eficiente de grandes volúmenes de datos en el sector financiero. Al adoptar SQL Server como el Sistema de Gestión de Bases de Datos principal, se garantiza robustez y compatibilidad con estándares comerciales. Desde la construcción de modelos E-R hasta la implementación de funciones avanzadas como migración, seguridad y una interfaz web intuitiva, cada fase del proyecto se ha abordado con precisión y atención al detalle.

La estrategia de migración, respaldada por un migrador eficiente en Go, ha demostrado su capacidad para transferir datos a una velocidad significativa, asegurando la coherencia y la integridad de la información. La inclusión de características como un servidor de auditoría, funciones de CRUD, triggers y niveles de acceso diferenciados refuerzan la seguridad de la base de datos, contribuyendo directamente a un análisis de riesgo crediticio más preciso y equitativo. La aplicación web desarrollada con Vue.js proporciona una interfaz visualmente atractiva para la presentación de datos, consolidando así un proyecto que no solo resuelve desafíos técnicos, sino que también sienta las bases para una gestión de bases de datos efectiva y adaptable a las demandas del mundo empresarial.

En términos de rendimiento, las pruebas han validado la capacidad de la base de datos optimizada para manejar solicitudes simultáneas, marcando el éxito del proyecto en cumplir con los estándares y requisitos detallados. En conclusión, esta iniciativa representa un paso importante hacia la aplicación práctica de habilidades técnicas para abordar problemas empresariales del mundo real, particularmente en el ámbito financiero donde la precisión y la eficiencia son cruciales.

Introducción

En un entorno empresarial cada vez más orientado a los datos, la eficiente gestión de bases de datos se convierte en un pilar fundamental para el éxito de las organizaciones. Este proyecto, concebido en el marco del segundo semestre de 2023, no solo busca ser un ejercicio académico, sino una oportunidad concreta para aplicar habilidades técnicas y solucionar problemas del mundo real.

La elección de SQL Server como SGBD principal no solo se basa en su robustez, sino en su amplio uso en el ámbito comercial, proporcionando un terreno fértil para la aplicación de conocimientos y técnicas avanzadas. El proyecto se sumerge en el sector financiero, utilizando datos proporcionados por Home Credit, una empresa que actualmente emplea métodos estadísticos y de aprendizaje automático para realizar predicciones en el ámbito del riesgo crediticio.

La importancia del análisis de riesgo crediticio radica en la capacidad de otorgar préstamos de manera justa y sostenible. El proyecto busca no solo aplicar metodologías de bases de datos avanzadas sino también contribuir a un proceso que asegure que los clientes capaces de reembolso no sean rechazados injustamente, y que los préstamos se otorguen con condiciones que maximicen el éxito de los clientes.

Las secciones siguientes detallarán la estrategia para abordar cada uno de los objetivos específicos, proporcionando una visión integral del proyecto desde su conceptualización hasta la implementación práctica. Este proyecto no solo es un desafío técnico; es un paso hacia la aplicación efectiva de habilidades para resolver problemas empresariales del mundo real.

Objetivos

Objetivo general

Desarrollar un sistema de bases de datos integral utilizando SQL Server como plataforma principal, abarcando desde la conceptualización de modelos (E-R), modelos

relacionales, esquemas, migradores, servidores de auditoría, funciones de inserción, triggers, consultas SARGABLES, índices no clúster, hasta la implementación de usuarios con diferentes niveles de acceso. El proyecto incluirá la creación de una interfaz de usuario eficiente, ya sea web, móvil o de escritorio, con un dashboard integrado.

Objetivos específicos

- Desarrollar un modelo E-R preciso que sirva como base para la construcción de un modelo relacional coherente, asegurando la integridad y estructura adecuada de la base de datos propuesta.
- Crear esquemas organizativos para dividir las tablas de manera eficiente, junto con un migrador que permita actualizaciones sin comprometer la integridad de los datos. Implementar un servidor de auditoría para mejorar la trazabilidad y seguridad de las transacciones.
- Diseñar funciones y triggers que garanticen la consistencia y validez de la información ingresada. Además, desarrollar consultas SARGABLES y construir índices no clúster para optimizar el rendimiento de las consultas. Establecer usuarios con diferentes niveles de permisos y crear una interfaz de usuario eficiente, como un sistema web, móvil o de escritorio, con un dashboard capaz de cargar en un máximo de un minuto y manejar un millón de registros por componente, respaldado por una documentación completa que abarque todo el proceso de desarrollo.

Descripción del problema

Historia de Home Credit

Home Credit, fundada en 1997 en la República Checa, ha evolucionado desde sus inicios como un proveedor local de préstamos al consumo hasta convertirse en una empresa financiera global líder. Desde su fundación, la empresa ha mantenido un enfoque singular en proporcionar servicios financieros a aquellas poblaciones que tradicionalmente han

tenido un acceso limitado a servicios bancarios. El modelo de negocio de Home Credit se ha destacado por su capacidad para llegar a consumidores que no cuentan con historial crediticio o que carecen de acceso a formas convencionales de financiamiento.

A lo largo de los años, Home Credit ha expandido su presencia a nivel mundial, estableciendo operaciones en múltiples países de Asia, Europa y América. Su presencia global le ha permitido adaptarse a diversas realidades económicas y necesidades financieras locales, consolidándose como una fuerza significativa en el sector de servicios financieros no bancarios.

La clave del éxito de Home Credit radica en su capacidad para evaluar el riesgo crediticio de manera innovadora. La empresa ha adoptado tecnologías avanzadas, como el análisis de datos y el aprendizaje automático, para realizar evaluaciones precisas de la solvencia crediticia de los clientes. Este enfoque ha permitido a Home Credit ampliar su alcance a segmentos de la población que a menudo son pasados por alto por las instituciones financieras tradicionales.

Solución a Construir

En la búsqueda constante de mejorar sus servicios y mantenerse a la vanguardia de la industria, Home Credit ha identificado la necesidad de optimizar su sistema de gestión de bases de datos. La solución propuesta para este proyecto tiene como objetivo abordar los desafíos específicos asociados con la gestión eficiente de grandes volúmenes de datos y garantizar la integridad y seguridad de la información.

La implementación de un Sistema de Gestión de Bases de Datos (SGBD) robusto es esencial para optimizar la normalización y seguridad de la información. Se busca construir un modelo E-R (Entidad-Relación) que refleje los contenidos de HomeCredit, a partir de este modelo se busca construir un modelo relacional que refleje los contenidos que serán almacenados en SQL Server.

Para abordar la necesidad de alta disponibilidad y eficiencia en la gestión de datos,

se desarrollará un migrador que permita la hacer una migración de los datos almacenados en los archivos CSV a la base de datos. Además, se implementará un servidor de auditoría para almacenar todo aquel registro que sea modificado, eliminado o insertado en la base de datos.

La solución incluirá la creación de funciones específicas para la inserción de datos, asegurando la coherencia y validez de la información introducida en la base de datos. La aplicación de triggers en todas las tablas garantizará la verificación de la validez de los valores ingresados, contribuyendo a mantener la integridad de los datos.

En términos de seguridad, se establecerán diferentes niveles de acceso a través de la creación de usuarios con permisos específicos, como administrador, usuario normal y usuario de respaldo. Además, se implementarán índices no clúster para optimizar el rendimiento de las consultas y consultas SARGABLES que funcionen como filtros eficientes en un posible sitio web.

La solución propuesta no solo aborda los desafíos técnicos de la gestión de bases de datos, sino que también se alinea con la misión de Home Credit de proporcionar servicios financieros eficientes y accesibles a segmentos de la población que históricamente han enfrentado barreras en este aspecto.

En resumen, la solución a construir se centra en la modernización y optimización del sistema de gestión de bases de datos de Home Credit, utilizando herramientas avanzadas para garantizar la eficiencia operativa, la seguridad de la información y, en última instancia, la capacidad de seguir ofreciendo servicios financieros innovadores y accesibles a nivel global.

Desarrollo

Modelo E-R

Para la construcción del modelo E-R se utilizó la herramienta de modelado *diagrams.net*. El modelo E-R se puede observar en la figura 1.

Modelo Relacional

[illegible]

Figure 2

Modelo Relacional

Migrador

Para la construcción del migrador se utilizó el lenguaje de programación Go, debido a que es un lenguaje de programación compilado, concurrente y de tipo estático, que tiene como objetivo ser sencillo, eficiente y escalable. La principal razón de utilizar este lenguaje de programación es debido a que ofrece una gran cantidad de librerías en adición a que tiene concurrencia nativa, lo cual permite que el proceso de migración sea más rápido.

El migrador utiliza un archivo de configuración en formato JSON, el cual contiene la información de la base de datos de origen, la base de datos de destino y la ruta de los archivos CSV. El migrador se encarga de leer el archivo de configuración, conectarse a la base de datos de origen (CSV) y a la base de datos de destino (SQL Server), para luego realizar la migración de los datos.

En una computadora con las siguientes características:

- Procesador: Intel Core i5-13600K
- Memoria RAM: 32 GB 3600 MHz

El proceso de migración de los datos tomó aproximadamente 2 horas en total, lo cual es un tiempo aceptable para la cantidad de datos que se están migrando. Logrando una transmisión de aproximadamente 30,000 registros por segundo.

Servidor de Auditoría

Para la construcción del servidor de auditoría se utilizó una tabla en la base de datos de destino, la cual contiene los siguientes campos:

- **Id:** Identificador único de la auditoría.
- **Fecha:** Fecha en la que se realizó la operación.
- **Usuario:** Usuario que realizó la operación.

- **Tipo:** Tipo de operación que se realizó sobre la tabla.
- **Descripción:** Descripción de la operación que se realizó y en donde se realizó.

De esta manera, cada vez que se realice una operación sobre la base de datos de destino, se insertará un registro en la tabla de auditoría, el cual contendrá la información de la operación que se realizó. Para que esto sea posible, se utilizó un trigger en cada tabla de la base de datos de destino, el cual se encarga de insertar un registro en la tabla de auditoría cada vez que se realice una operación sobre la tabla.

Funciones de CRUD

Para la generación de las funciones de inserción se utilizó el plugin *SQL Complete* de Devart (2023), el cual permite generar funciones de inserción, actualización y eliminación de manera automática. El plugin genera las funciones basado en la tabla que se seleccione, generando una función por cada tabla de la base de datos. De esta manera, cada función de CRUD tiene un formato regular y se puede utilizar para cualquier tabla de la base de datos.

Triggers

Para la generación de los triggers se estableció un formato regular, el cual luego fue enviado a ChatGPT OpenAI (2023), una herramienta de inteligencia artificial que permite generar texto basado en un texto de entrada. De esta manera, se generaron los triggers para cada tabla de la base de datos de destino. Cada uno de estos triggers hacen validaciones correspondientes a la tabla en la que se encuentran, por ejemplo, si se realiza una operación de inserción en la tabla de *application_test*, el trigger se encargará de validar que los datos que se están insertando sean válidos, en caso de que no lo sean, el trigger no permitirá la inserción de los datos, generando un rollback.

Seguridad

Para la seguridad de la base de datos se crearon tres usuarios, cada uno con diferentes permisos. El primer usuario es el usuario de respaldo, el cual tiene permisos de lectura y generación de respaldos de la base de datos de destino. El segundo usuario es el usuario normal, el cual tiene permisos de lectura y escritura sobre la base de datos de destino, pero no tiene permisos para realizar modificaciones sobre la estructura de la base de datos. El tercer usuario es el usuario administrador, el cual tiene permisos de super usuario sobre la base de datos de destino, lo cual le permite realizar cualquier operación sobre la base de datos.

Índices no Clúster

Para la creación de los índices no clúster primero se definieron las consultas que se realizarían sobre la base de datos, para luego crear los índices no clúster basados en estas consultas. De esta manera, las consultas que son mostradas en la aplicación web, son consultas que utilizan los índices no clúster, lo cual permite que las consultas sean más rápidas y eficientes.

Consultas SARGABLES

Para la creación de las consultas SARGABLES se aprovechan los índices no clúster que se crearon anteriormente. De esta manera, las consultas que se realizan en la aplicación web son consultas SARGABLES, lo cual permite que las consultas sean más rápidas y eficientes, de modo que el usuario pueda visualizar la información de manera más rápida.

Aplicación Web

Para la construcción de la aplicación web se utilizó el framework de desarrollo web *Vue.js* You (2023), el cual es un framework progresivo para la creación de interfaces de

usuario. El framework se encarga de realizar las peticiones a un backend, el cual se encuentra desarrollado en el lenguaje de programación Go.

El backend se encarga de realizar las consultas a la base de datos de destino, para luego enviar la información a la aplicación web. De esta manera, la aplicación web se encarga de mostrar la información de manera ordenada y estructurada, para que el usuario pueda visualizar la información de manera más sencilla.

Para la visualización de la información se utilizó la librería *Google charts* Google (2023), la cual permite generar gráficos de manera sencilla. De esta manera, la aplicación web se encarga de generar los gráficos basados en la información que se encuentra en la base de datos de destino, para luego mostrarlos al usuario en el dashboard de la aplicación web, tal como se puede observar en la figura 3.



Figure 3
Aplicación Web

Conclusiones

En resumen, el proyecto de optimización de bases de datos para Home Credit se destaca por su enfoque completo, desde la construcción de modelos E-R hasta la implementación de funciones de migración y seguridad. La aplicación web resultante, desarrollada en Vue.js, ofrece una interfaz intuitiva y visualmente atractiva para acceder y

analizar datos.

De acuerdo con los resultados de las pruebas de rendimiento, la base de datos optimizada es capaz de manejar un gran número de solicitudes simultáneas sin problemas, lo cual se considera un éxito en el rango de aplicación del proyecto y los estándares detallados en el documento de requerimientos.

Recomendaciones

Si bien el caso de Home Credit es un ejemplo de éxito de la implementación de una base de datos relacional, sin embargo existen algunas excepciones en las cuales Home Credit no es muy adaptable, principalmente en el caso de que al ser un dataset inicialmente pensado para un modelo de aprendizaje automático, no se cuenta con la información necesaria para realizar un análisis de datos más profundo el cual pueda contar con fechas y datos históricos de los clientes, lo cual es un factor importante para la implementación de un sistema que busca generar estadísticas en una vista de dashboard.

References

- Devart. (2023). *Dbforge sql complete - powerful t-sql formatting tool*. Retrieved from <https://www.devart.com/dbforge/sql/sqlcomplete/>
- Google. (2023). *Charts / google for developers*. Author. Retrieved from <https://developers.google.com/chart>
- OpenAI. (2023). *Chatgpt*. Retrieved from <https://openai.com/chatgpt>
- You, E. (2023). *The progressive javascript framework*. Retrieved from <https://vuejs.org/>