Instituto Tecnológico de Costa Rica

Unidad de Computación

Proyecto 1

Carlos Akion Garro Campos

Sede San Carlos

14/06/2022

Tabla de contenidos

Contenido

Resumen Ejecutivo

Home Credit es una entidad financiera encargada de brindar créditos a clientes que no cuentan con un historial, para hacer esto debe de hacer uno de una gran cantidad de datos para determinar si este cuenta con capacidad de pago.

Por lo que para indicar cuales clientes cumplen es necesario la visualización de la información presente en la base de datos, por lo que en el presente proyecto se desarrolla un sistema gestor de bases de datos que permita visualizar los datos de los usuarios para ver si pueden optar por un crédito.

Objetivo General

- Crear un sistema gestor de bases de datos robusto para mostrar la información de Home Credit.

Objetivos Específicos

- Normalizar los datos de Home Credit para hacer el modelo relacional del sistema.
- Hacer el proceso de migración de Home Credit hacia la base de datos.
- Utilizar técnicas de optimización para la ejecución de consultas de manera rápida.
- Crear un sistema para la visualización de los datos financieros de Home Credit.

Introducción

Los créditos son muy importante a la hora de optar por alguna mejora o adquisición de bienes, pero las entidades financieras tienen que asegurarse que los clientes tengan la capacidad económica para pagar, por lo que Home Credit no es la excepción, pero muchas veces por historiales de crédito insuficientes o inexistentes los clientes no pueden optar por uno.

Home Credit utiliza varios métodos estadísticos y de aprendizaje automático para hacer predicciones sobre si es factible o no brindar un crédito a una persona por lo que, en el presente proyecto se va a hacer uso de los datos brindados para desarrollar un sistema capaz de mostrar información importante acerca de Home Credit.

Por otra parte la información brindada por el dataset de Home Credit es información desnormalizada en formato CSV por lo que en el presente trabajo se hará énfasis en el proceso de normalización, migración, seguridad y visualización de datos de manera eficiente haciendo uso de las buenas prácticas de SQL Server.

Descripción del problema

Home Credit es una institución financiera que se encarga de realizar estudios para ampliar los créditos a personas que no se encuentran asociadas a historiales crediticios o personas que cuentan con pocos créditos en una institución financiera. Por lo que esta empresa va a tratar de brindar la mejor experiencia a las personas que van a optar por realizar un crédito, pero a su vez también determinar por medio de una gran cantidad de datos alternativos brindados por los usuarios para predecir las capacidades de pago de sus clientes.

De esta manera, Home Credit utiliza todos los datos mencionados anteriormente para ingresarlos en varios métodos estadísticos y de aprendizaje automático para hacer estas predicciones, ya asegurará que los clientes capaces de pagar no sean rechazados y que los préstamos se otorguen con un calendario de capital, vencimiento y reembolso que permita a sus clientes tener éxito.

Para la utilización del dataset de Home Credit, primeramente se analiza que se encuentran almacenados los datos en archivos CSV, por lo que se debe de hacer la obtención de los títulos de cada columna que estos se van a convertir en los atributos de cada entidad de la base de datos. Una vez obtenidos estos títulos se tiene que empezar a analizar de manera muy cuidadosa, tratando de identificar posibles anomalías para proceder a aplicar la normalización de estas y empezar con la creación del modelo relacional y E/R.

Después una vez normalizados se empieza la creación de todas las tablas, restricciones y procedimientos almacenados que se van a encargar de ingresar los datos provenientes del CSV hacia la base de datos.

Posteriormente se presenta el problema de que los datos se encuentran en archivos CSV, por lo que se debe de crear toda la lógica del migrador, donde un diagrama que nos puede facilitar este proceso se visualiza en la figura 1.

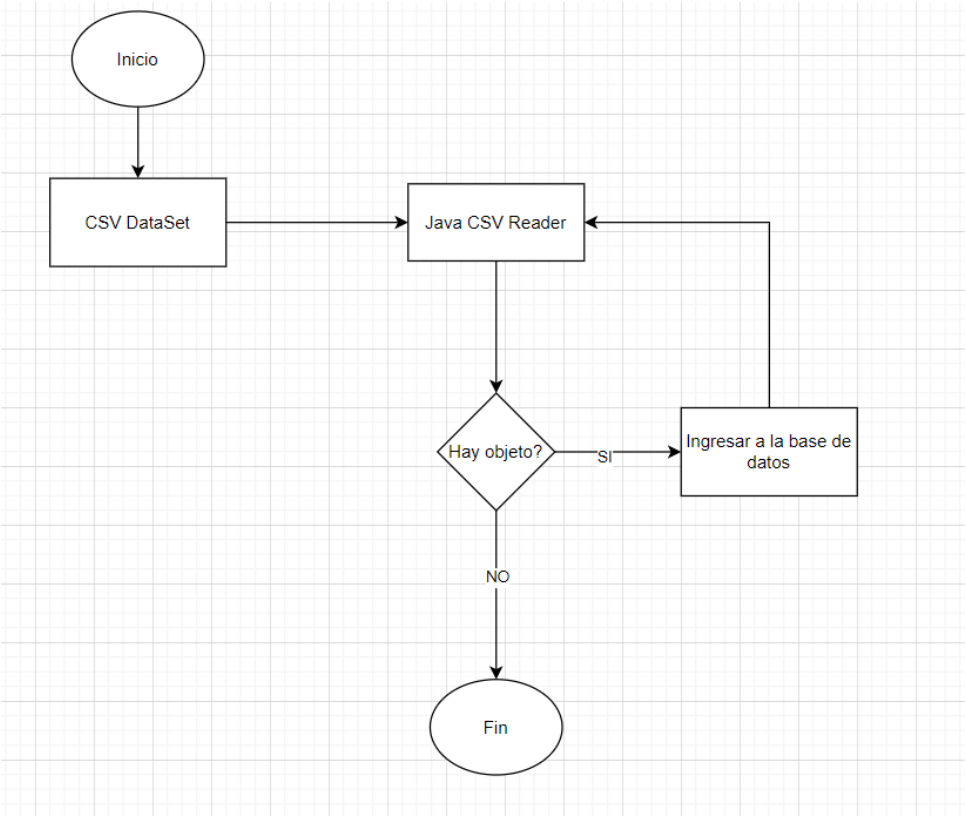


Figura 1. Diagrama que nos ayuda a comprender el ingreso de los datos a la base.

Por último se debe de crear una aplicación para mostrar todos los datos ingresados en la base de datos, esta debe de contener filtros y componentes que nos demuestran la información de los clientes de Home Credit.

Desarrollo

Primeramente para la creación del proyecto se planteó y se dividieron varias secciones de importancia para poder modular el problema, esta división está enfocada en el método de divide y vencerás donde se resuelve tareas más pequeñas y luego se agrupan estas soluciones para la creación de la solución general. En la figura 2 se puede ejemplificar la manera en que se dividieron los módulos para posteriormente empezar a trabajarlos de manera individual.

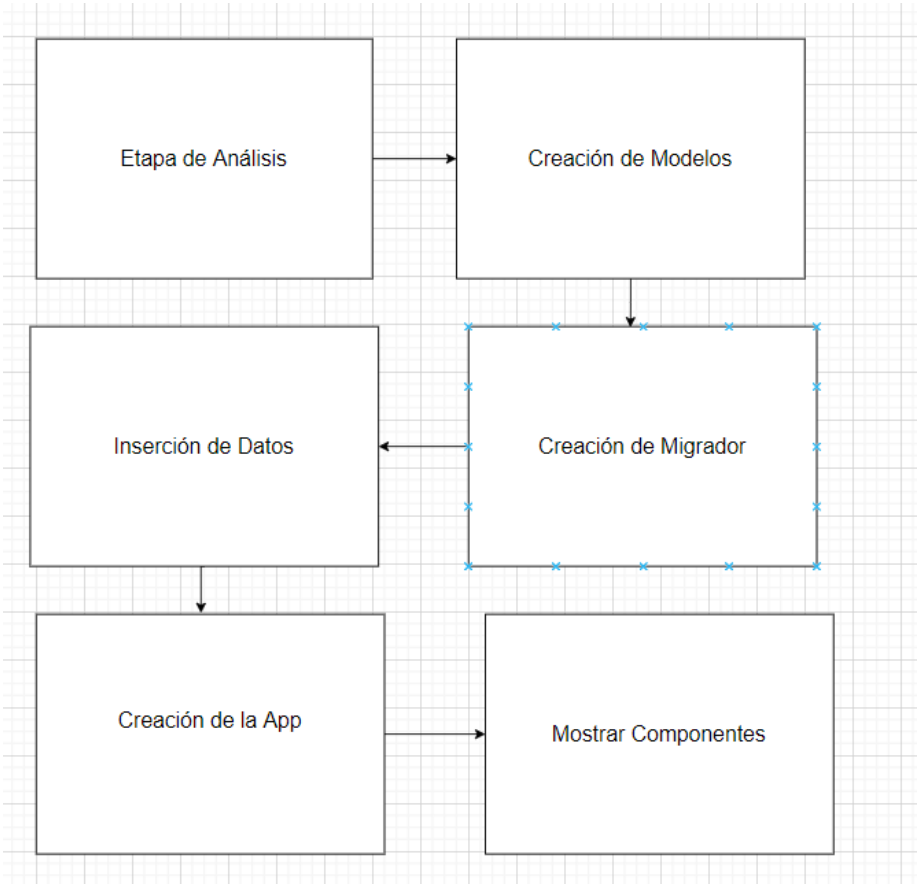
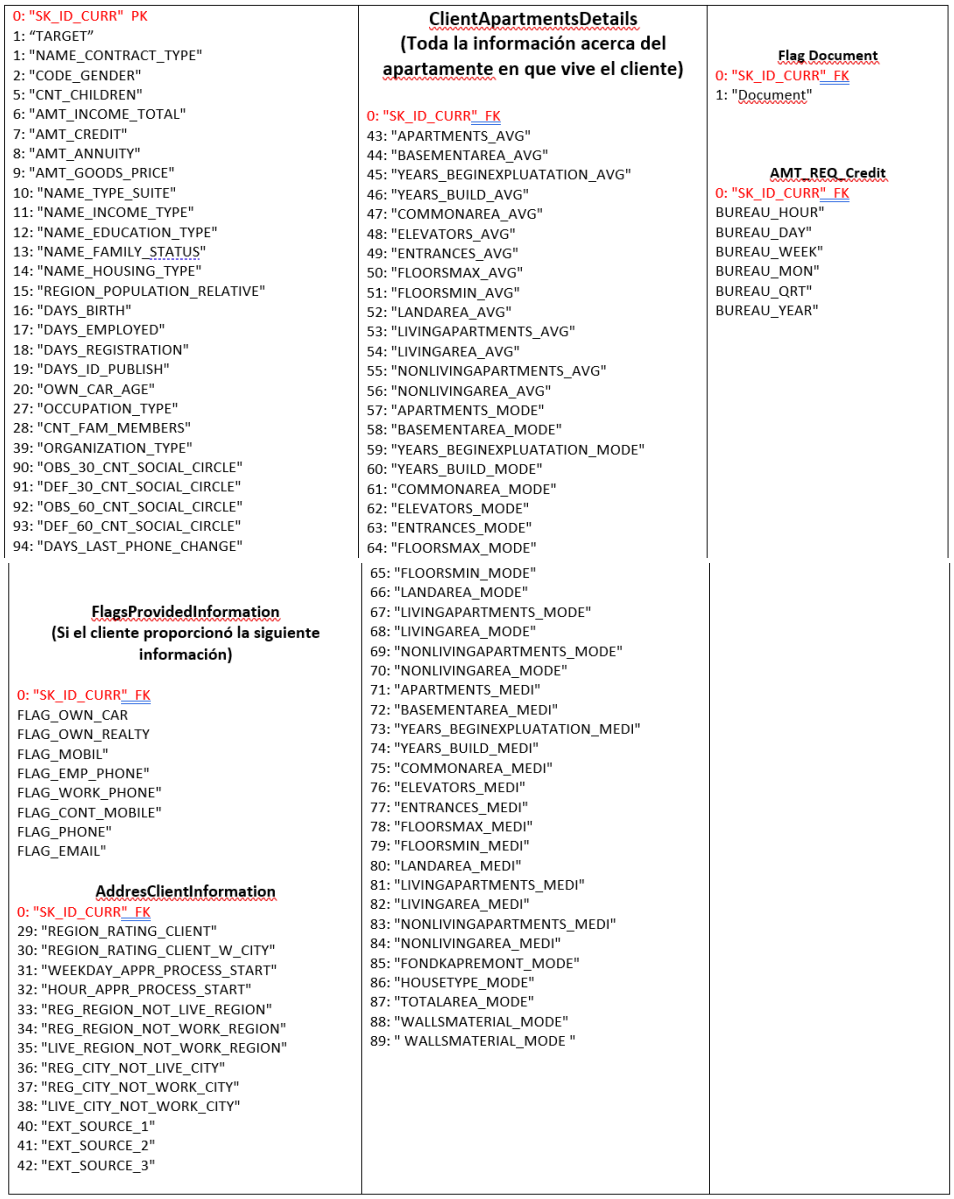


Figura 2. División inicial del problema a resolver

En la etapa de análisis se obtuvieron todos los títulos de las columnas de los archivos CSV, ya que estos son los atributos de las entidades en la base de datos. En este proceso se verificaron que se encuentra con varias anomalías ya que algunos de estos datos no se encontraban ni en la primera,segunda ni tercera forma normal qué es lo que se recomienda. En las siguientes figuras se muestra de manera inicial el proceso de análisis y especificación de algunas entidades.



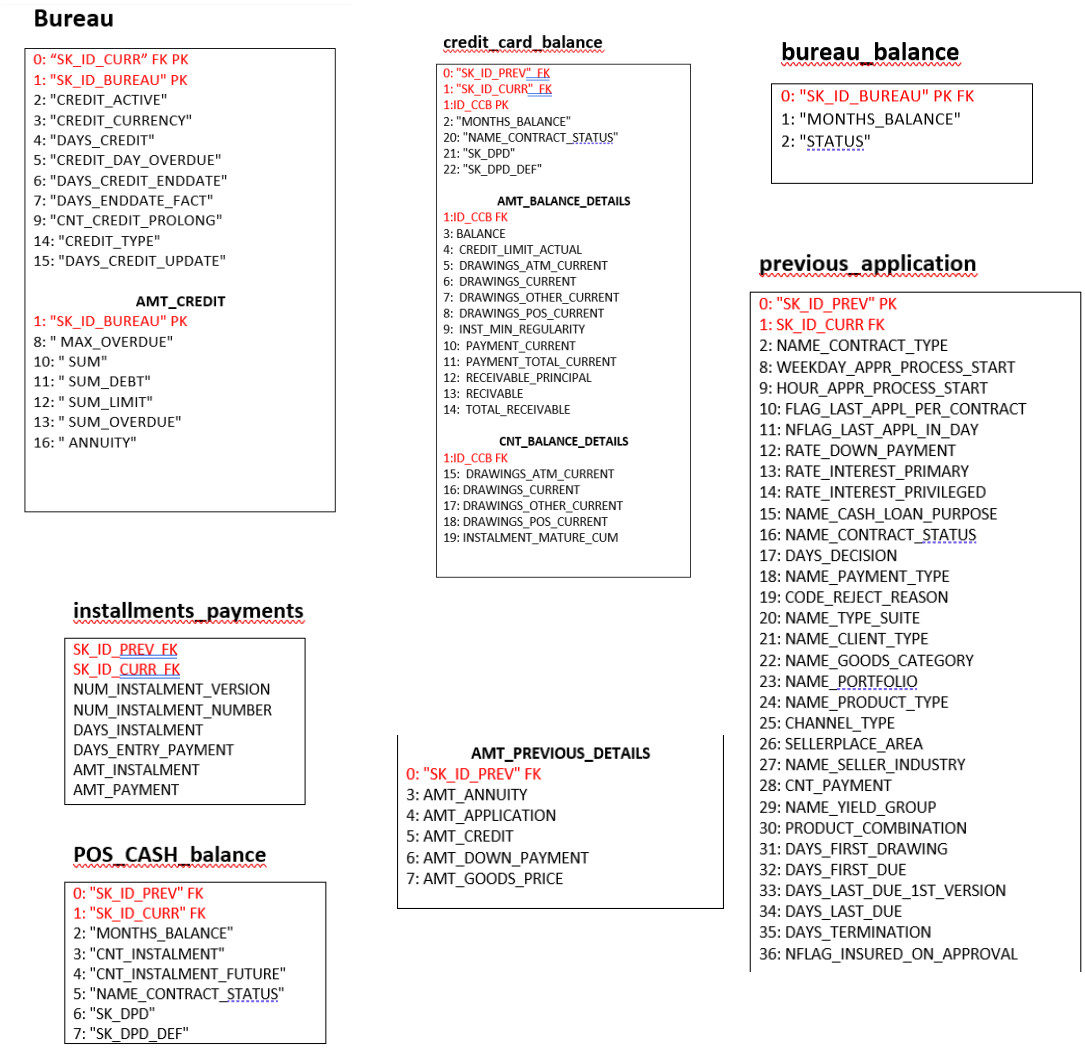


Figura 3. División de las entidades

Una vez realizado el proceso de análisis y normalización se inicia con la creación de los modelos de E/R y modelo relacional, que son de suma importancia para la creación de la base de datos, donde en estos se termina de arreglar los errores del proceso de anterior y se define de la manera en que va a quedar la base de datos. En la siguiente figura se muestra un ejemplo de algunas entidades del modelo relacional de Home Credit.

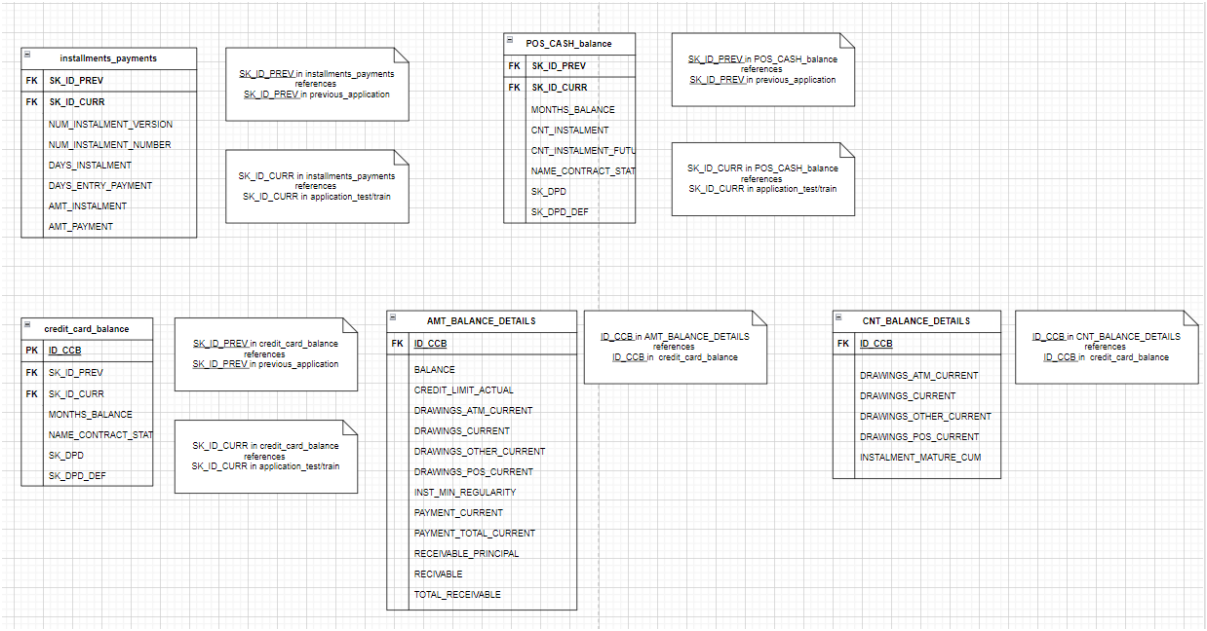


Figura 4: Entidades del modelo relacional de Home Credit

En la creación del migrador se hizo uso del lenguaje Java junto a una librería llamada Open CSV, que proporciona todos los métodos necesario la para lectura de los archivos CSV, a estos se le hizo un casteo para transformarlos en datos aceptados por los procedimientos almacenados creados con anterioridad. Se empieza a ejecutar el migrador de forma en que las entidades que son independientes se ingresen y posteriormente las que tienen llaves foráneas a estas otras entidades. En la figura 5 se muestra la manera en que se agregaron.

```
m1.migrationAplicationTrain(conex);
m1.migrationClientApartments(conex);
m1.migrationAddressClientInformation(conex);
m1.migrationFlagDocument(conex);
m1.migrationFlagsProvidedInfo(conex);
m1.migrationAMTREQCredit(conex);
m1.migrationBureau(conex);
m1.migrationBureauBalance(conex);
m1.migrationAMTCreditDetails(conex);
m1.migrationPreviousApplication(conex);
m1.migrationPosCashBalance(conex);
m1.migrationInstallmentsPayments(conex);
m1.migrationAMT_PREVIOUS_DETAILS(conex);
m1.migrationCreditCardBalance(conex);
m1.migrationAMT_BALANCE_DETAILS(conex);
m1.migrationCNTBalanceDetails(conex);
```

Figura 5. Ingreso de los datos a la base Home Credit.

Por último se utilizó Java con la parte gráfica de Swing, la cuál permitió crear una interfaz sencilla pero funcional para mostrar los datos de la base de datos Home Credit.

Análisis de resultados

Tarea/Requerimiento	Estado	Observaciones
Modelo E/R	Completo	
Modelo Relacional	Completo	
Esquemas	Completo	Se crearon 3 esquemas
Migrador	Completo	Se utilizó Open CSV de Java
Funciones para insertar los datos	Completo	
Consultas	Completo	Se crearon 3 consultas
Índices No Clúster	Completo	
Crear usuarios	Completo	
Componentes	Completo	Se crearon dos componentes con +1000 000 de datos. Y se muestran con JFreeChart

Conclusiones

Al trabajar con datasets tan grandes es importante darle énfasis en la optimización y utilizar las buenas prácticas de las bases de datos, ya que estas están fuertemente vinculadas al tiempo de espera de un usuario final, por lo que va a ser muy importante que cuando el usuario solicita una información tratar de brindarla en el menor tiempo posible.

El proceso de migración es sumamente importante ya que es un proceso que se realiza con frecuencia en el ámbito profesional ya que los sistemas se actualizan a tecnologías más eficientes y por lo que los datos deben de migrarse.

Recomendaciones

Para la creación de estos sistemas gestores de bases de datos se recomienda que se utilice Web en vez de una aplicación de escritorio, ya que hay muchas más información y además hay muchas más opciones que permiten dibujar los gráficos y mostrar los datos de una manera más profesional, ya que al adentrarse un poco en la investigación de herramientas para crear gráficos en java fueron pocas las opciones que se presentaron.

Bibliografía

How to add row of data to Jtable from values received from jtextfield and comboboxes. (2014, 15 enero). Stack Overflow. https://stackoverflow.com/questions/21135452/how-to-add-row-of-data-to-jtable-from-values-received-from-jtextfield-and-combob

Melgoza, J. (2019, 1 abril). Como Hacer Graficos con Java – Pastel, Barras, Lineas, 3D . . .. Jonathan Melgoza. https://jonathanmelgoza.com/blog/como-hacer-graficos-con-java/

Gigena, M., & Perfil, V. T. M. (s. f.). Todo Java. Todo Java. http://labojava.blogspot.com/2012/06/graficos-estadisicos-jfreechart

Elaboración de cuadros estadísticos en netbeans. (2018, 15 febrero). YouTube. https://www.youtube.com/watch?v=MBWZMQ-IneI&ab_channel=TICUTMACH

Home Credit Default Risk | Kaggle. (s. f.). Kaggle. https://www.kaggle.com/competitions/home-credit-default-risk/overview

M. (2020, 26 diciembre). How to read and parse CSV file in Java. Mkyong.Com. https://mkyong.com/java/how-to-read-and-parse-csv-file-in-java/