

Instituto Tecnológico de Costa Rica

Escuela de computación



Profesor

EFREN JIMENEZ DELGADO

Proyecto

Integrantes

Kelvin Núñez Barrantes

Kenny Vega Obando

Fecha: 24/11/2021

Tabla de contenidos	#
Objetivos.....	3
Introducción.....	4
Descripción del problema.....	4
Desarrollo.....	5
Conclusiones	7
Recomendaciones.....	7

Problema

¿Cuál es el proceso por el que un ingeniero en computación pasa al realizar la migración de una base de datos extensa a un sistema de gestión de bases de datos?

Objetivo general

Replicar el proceso por el un ingeniero en computación pasa al migrar una base de datos extensa a un sistema de gestión de bases de datos.

Objetivos específicos

- Analizar mediante la práctica el manejo de bases de datos en SQL SERVER.
- Descubrir métodos para la importación de archivos CSV a un sistema de gestión de bases de datos.
- Identificar problemas al realizar la migración de una base de datos.

Introducción

Las bases de datos son un mecanismo de suma importancia para la vida tecnológica actual y estas son unos de los componentes de más valor cuando se habla con respecto a una empresa. Por lo que ser capaz de migrar una base de datos de mucha información es un proceso que lleva tiempo y puede ser hasta complicado evitar problemas al manejar una gran cantidad de datos. Por eso es de mucha importancia para un ingeniero de computación que busque especializarse en el manejo de bases de datos, saber qué opciones se tienen para la migración y cuál es una manera que se adapte a las necesidades del caso.

En el siguiente proyecto se explicará cómo se hizo la migración de una base de datos con millones de datos, desde el proceso de creación del código para el migrar hasta la creación de usuarios, triggers y procedimientos almacenados para trabajar los datos. Todo esto con el objetivo de simular cuál es el trabajo que se tiene que hacer si se desea especializar en bases de datos.

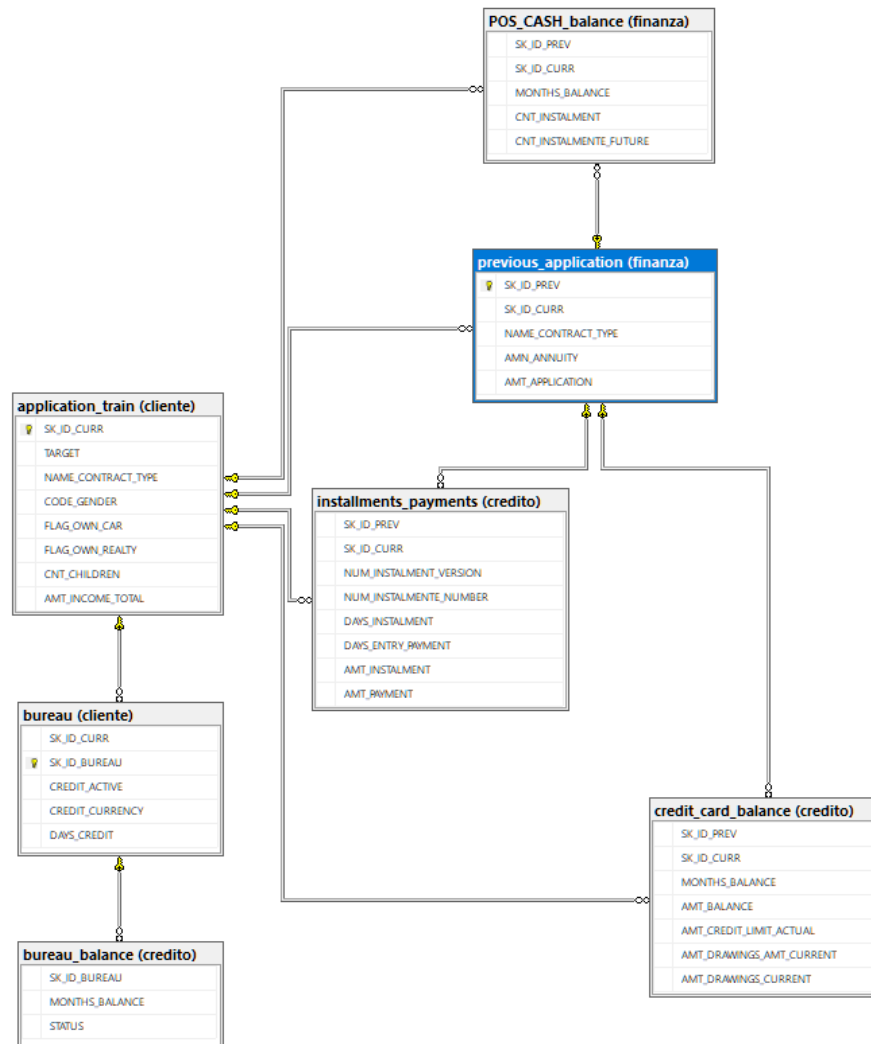
Descripción del problema

La base de datos de home credit busca almacenar la información de los clientes y conocer mediante una predicción qué tan probable y rentable es que un cliente que solicite un préstamo, vaya a pagarlo. Para lograr esto la base está llena de información con respecto a los bienes de los clientes y sus antecedentes financieros.

La solución debe de ser capaz de tener un sistema de orden para tener trazabilidad y que los datos mantengan jerarquía para así que su utilización se facilite

Desarrollo

Diagrama de la base de datos



La base de datos se compone de 3 a 8 atributos por tabla, esto para buscar una facilidad a la hora de manejarlos. La tabla Main(application_train) se combinó con application_test para tener una mayor cantidad de IDs en la muestra. Luego de ahí son pasados como llaves foráneas a otras tablas como se ve en la imagen.

Las tablas externas en el diagrama tuvieron el problema de que en la información que contenían habían llaves que no se encontraban en la tabla principal por lo que para resolver este problema se crearon 4 tablas auxiliares que son las que poseen los datos para evitar el error. Estas tablas no tienen relación con la tabla principal.

Se crearon procedimientos almacenados y triggers para la inserción y la validación de nuevos datos a la tabla. Además se crearon 3 usuarios para administrar la base de datos, uno que puede realizar cualquier acción, otro que solo es capaz de ver tablas y el último que solo es capaz de crear backups. Y también se crearon índices no clúster para facilitar las consultas específicas en ciertas tablas.

Para la creación del dashboard se optó por utilizar una conexión con Excel Power Pivot en donde ya los datos migrados de SQL SERVER fueron leídos y se pudo realizar una filtración en 3 tablas.

Filtros										Application Train										Installments Payments										Bureau										
Cantidad de ...										SK_ID_CURR	ENT_CHILDRE	AMT_ANCLN_TOTA	NAME_CONTRACT_TYR	TARGE	CODE_GENDER											SK_ID_CURR	CREDIT_ACTIVE	PREDI_CURR	SK_ID_CURR	Sum of SK_ID_CURR										
0	103596	103596	103596	103596	103596	103596	103596	103596	103596	7	65500	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	Closed	currency1	103596	103596	103596	Closed	currency1	103596	103596				
1	103596	103596	103596	103596	103596	103596	103596	103596	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	Closed	currency1	103596	103596	103596	Closed	currency1	103596	103596				
2	103596	103596	103596	103596	103596	103596	103596	103596	103596	7	203500	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	Closed	currency1	103596	103596	103596	Closed	currency1	103596	103596				
3	103596	103596	103596	103596	103596	103596	103596	103596	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	Closed	currency1	103596	103596	103596	Closed	currency1	103596	103596				
4	103596	103596	103596	103596	103596	103596	103596	103596	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	Closed	currency1	103596	103596	103596	Closed	currency1	103596	103596				
5	103596	103596	103596	103596	103596	103596	103596	103596	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	Closed	currency1	103596	103596	103596	Closed	currency1	103596	103596				
6	103596	103596	103596	103596	103596	103596	103596	103596	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	Closed	currency1	103596	103596	103596	Closed	currency1	103596	103596				
7	103596	103596	103596	103596	103596	103596	103596	103596	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	Closed	currency1	103596	103596	103596	Closed	currency1	103596	103596				

Tipos de préstamo

Cash loans
Revolving loans

Difficultad

0
2
1

Genero

F
M
XNA

SK_ID_CURR	ENT_CHILDRE	AMT_ANCLN_TOTA	NAME_CONTRACT_TYR	TARGE	CODE_GENDER	SK_ID_CURR	ENT_CHILDRE	AMT_ANCLN_TOTA	NAME_CONTRACT_TYR	TARGE	CODE_GENDER	SK_ID_CURR	ENT_CHILDRE	AMT_ANCLN_TOTA	NAME_CONTRACT_TYR	TARGE	CODE_GENDER	SK_ID_CURR	ENT_CHILDRE	AMT_ANCLN_TOTA	NAME_CONTRACT_TYR	TARGE	CODE_GENDER	SK_ID_CURR	ENT_CHILDRE	AMT_ANCLN_TOTA	NAME_CONTRACT_TYR	TARGE	CODE_GENDER	SK_ID_CURR	ENT_CHILDRE	AMT_ANCLN_TOTA	NAME_CONTRACT_TYR	TARGE	CODE_GENDER
103596	7	65500	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	203500	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F
103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000	Cash loan	0	F	103596	7	100000									

Para el proceso de migración se realizó el código en Python aprovechando la librería Pandas para el manejo del dataframe y Pyodbc para la conexión con el servidor de SQL SERVER. Se leyó el archivo CSV y se escribió el insert respectivo a cada tabla para que se hiciera la migración. Se creó un archivo para cada tabla y un main que se encargara de correr todos los archivos.

Conclusiones

Se concluye que el proceso de migración de bases de datos es un proceso complejo y que requiere mucho tiempo. Es importante definir los métodos por el cual se desea realizar la migración ya que esto va a requerir la mayor cantidad del tiempo. También se concluye que el proceso de normalización es donde se van a generar la mayor cantidad de problemas ya que al ser tantos datos es frecuente que no todos estén estandarizados por lo que es de suma importancia que todos sigan un mismo patrón para evitar problemas.

El dashboard es un método muy útil para poder buscar información pero requiere que la base de datos ya esté migrada y que los datos tengan su relación para así encontrar coincidencias y que puede de ser de mucha utilidad para la visualización de información y más con información tan extensa.

Por último, se concluye que el todo de una base de datos está compuesto por varias partes, desde tablas hasta esquemas, procedimientos almacenados y funciones y que es muy importante tener un plan de acción ya que lo mejor es empezar desde lo más simple para así ir relacionando los objetos y que ya se forme la base de datos como se requiere

Recomendaciones

Recomendaciones para futuros proyectos

- Se recomienda tener una breve explicación del contenido de las tablas y de los datos para así entender qué partes son de mayor utilidad y trabajar con respecto a eso.
- Realizar una normalización general de los datos ya que esto puede generar muchos problemas al realizar la migración.
- Tomar en cuenta que el proceso de migración tomará mucho tiempo ya que este tiene que recorrer todos los archivos y luego escribirlos en el sistema de gestión de la base de datos por lo que es normal que la espera pueda extenderse si se trata de migrar millones de datos.