

Sede Regional Chorotega

CAMPUS Nicoya

CARRERA	:	Ingeniería en Sistemas de Información
NOMBRE DEL CURSO	:	Administración de Bases de Datos
CODIGO	:	EIF402
PROFESOR	:	Carlos Espinoza Chavarria
PERIODO LECTIVO	:	II Ciclo del 2024

Práctica Preparación - Examen Final de Administración de Base de Datos

Duración: 3 horas

Modalidad: Individual

Entregables: 5 archivos

Formato: Subida digital (En el aula virtual)

Archivos esperados (solo como guía): cambiar numero cedula por el numero de identificación del estudiante

- numero cedula_ERD_Articles.pdf
- numero cedula_ModeloRelacional_Articles.pdf
- numero cedula_DDL_Articles.sql
- numero cedula_CargaDatos_Articles.sql
- numero cedula_Consultas_Articles.sql

Objetivos

Poner en práctica el ciclo completo de trabajo de un administrador de base de datos ante un caso realista:

- Analizar un conjunto de datos en formato JSON,
- Diseñar la estructura de almacenamiento adecuada,
- Implementarla en SQL Server,
- Importar la información usando solo T-SQL,
- Realizar consultas de análisis que permitan validar la correcta integración de los datos.

Caso: DataScienceHub – Repositorio de Publicaciones Científicas

La empresa DataScienceHub gestiona una plataforma digital dedicada a recopilar y organizar artículos científicos de diferentes áreas tecnológicas. Cada artículo almacenado en la plataforma contiene información clave, como su título, resumen, palabras clave, autores, fecha de publicación, instituciones participantes y el medio en el que fue presentado o publicado (revista, conferencia, etc.).

Con el crecimiento del volumen de información, la empresa ha decidido migrar sus registros experimentales, actualmente guardados en archivos JSON, hacia una base de datos en SQL Server. Esta migración busca mejorar la eficiencia en las búsquedas, así como facilitar la generación de estadísticas sobre autores, temas y fechas de publicación.

Tu función, como Administrador de Base de Datos Junior, consiste en diseñar la estructura de la base de datos que almacenará esta información y cargar los datos iniciales a partir del archivo JSON proporcionado. Este proceso te permitirá aplicar conocimientos sobre modelado relacional, tipos de datos en SQL Server y migración de datos, contribuyendo al manejo estructurado y analítico de la información científica dentro de la organización.

Parte 1 – Diseño Conceptual (ERD)

Archivo: ERD_DataScienceHub.pdf (agregar numerocedula_)

Con base en el JSON proporcionado, diseñe un modelo entidad-relación (ER) que refleje adecuadamente los datos.

Debe considerar al menos las siguientes entidades:

- Article (id, title, abstract, date, etc.)
- Keyword (lista de temas o etiquetas asociadas)
- Author (nombre de cada autor participante)
- Venue (nombre del medio o conferencia)
- Team (institución o grupo investigador)

Cada artículo puede tener múltiples autores, múltiples palabras clave, múltiples equipos y un solo venue.

Entrega: diagrama ER legible (usar draw.io)

Parte 2 – Modelo Relacional

Archivo: ModeloRelacional_DataScienceHub.pdf (agregar numerocedula_)

Transforme su modelo ER a un modelo relacional. Debe indicar:

- Nombre de las tablas.
- Claves primarias (PK).
- Claves foráneas (FK).
- Tipos de datos propuestos (ej. VARCHAR, NVARCHAR, INT, etc.).

Parte 3 – Implementación en SQL Server

Archivo: DDL_DataScienceHub.sql (agregar numerocedula_)

Cree la base de datos en SQL Server incluyendo:

- Base de datos con nombre: data_science_hub_db
- Filegroups: (unidades lógicas)
 - o PRIMARY para metadatos y tablas maestras.
 - o DATA_FG para tablas de detalle.
- Instrucciones de creación de tablas, claves primarias y foráneas, índices y constraints.

Puede agregar una tabla auxiliar de auditoría si lo considera pertinente.

Parte 4 – Carga de Datos desde JSON

Archivo: CargaDatos_DataScienceHub.sql (agregar numerocedula_)

Usando solo T-SQL, importe la información del archivo JSON (articles.json)

Debe utilizar las funciones integradas de SQL Server, como:

- OPENROWSET(BULK...) para leer el archivo JSON.
- OPENJSON() para descomponer los arrays (por ejemplo, keywords, authors, teams).

Parte 5 – Consultas SQL

Archivo: Consultas_DataScienceHub.sql (agregar numerocedula_)

Crear un script con las siguientes consultas.

Cada una debe usar los alias de columnas indicados (entre paréntesis)

- Mostrar el título del artículo, la fecha de publicación y la cantidad de autores asociados. (Columnas esperadas: ArticleTitle, PublicationDate, AuthorCount)
- Listar las 5 palabras clave más frecuentes entre todos los artículos.

(Columnas esperadas: Keyword, ArticleCount)

- Mostrar los artículos publicados en el medio "info.info-ai". (Columnas esperadas: ArticleID, ArticleTitle, VenueName)
- Mostrar los artículos que contengan la palabra "privacy" en su resumen. (Columnas esperadas: ArticleID, ArticleTitle, Snippet) Donde Snippet puede ser un substring del abstract.
- Contar cuántos artículos tienen más de 3 autores y al menos una palabra clave relacionada con "audio". (Columnas esperadas: ArticleCount)
- Bonus Crear una vista llamada vw_ArticleSummary que muestre el título del artículo, su fecha y el número total de autores y palabras clave. (Columnas esperadas: ArticleID, ArticleTitle, PublicationDate, AuthorCount, KeywordCount)

Parte 6 – Administración y Respaldo de la Base de Datos

Archivo: Administracion_DataScienceHub.sql

Realice las siguientes tareas relacionadas con la administración básica de la base de datos creada.

Cada script debe incluir comentarios breves que describan lo que realiza.

Mostrar el espacio total ocupado por la base de datos data_science_hub_bd, desglosado por cada uno de sus archivos físicos (data y log).

(Columnas esperadas: FileName, FileType, SizeMB, FilePath)