



DESCRIPCIÓN DEL PROYECTO
CI-0141 BASES DE DATOS AVANZADAS
I SEMESTRE 2024

Dr. Luis Gustavo Esquivel Quirós

(Versión original Dra. Elzbieta Malinowski G.)

(Esta actividad reemplaza la primera, segunda y tercera entrega del proyecto y tiene un valor de 30%)

Introducción

Una de las habilidades importantes del profesional en computación es el auto-aprendizaje y el compartir conocimiento con otros profesionales, es decir poner el conocimiento poseído por un individuo a disposición de otras personas. Este proceso se caracteriza por dos fases: externalización e internalización. La externalización permite que un individuo ceda su conocimiento a otros, mientras que, a través de la internalización, el receptor de este conocimiento le da sentido dentro de sus capacidades de absorción. La falta de compartir conocimiento afecta la creación de nuevo conocimiento y el crecimiento de los individuos. Sin embargo, los estudios demuestran que compartir conocimiento no es un comportamiento que fluya de forma espontánea y libre. Como consecuencia, se necesita alguna intervención gerencial para facilitar y motivar a las personas a compartir lo que saben. Al lograrlo se consigue un modelo activo de participación, que resulta enriquecedor (Camelo, García y Sousa, 2010).

En este proyecto se busca que usted ejercite el compartir conocimiento en su formación profesional ya que permite la adquisición de conocimiento en forma organizada, es decir, no es suficiente leer para para ampliar su conocimiento personal, al contrario, se necesita un elemento de selección de temas importantes, organización que permita “conectar” diferentes elementos entre sí y presentarlos a otros en forma clara, concisa e interesante.

Objetivo general

Introducirse a los conceptos relacionados con las bases de datos NoSQL por medio del estudio del modelo seleccionado y la implementación de un ejemplo de caso de estudio. Esto para poder enfrentar retos actuales y buscar los modelos de datos adecuados para solucionar diferentes tipos de problemas.

Objetivos específicos



1. Obtener los conocimientos teóricos sobre el modelo de datos seleccionado.
2. Construir un caso de estudio o experimento que permita la exploración de particularidades del modelo seleccionado.
3. Instalar y probar el funcionamiento de un motor de base de datos que permita realizar el caso de estudio o experimento planteado.
4. Implementar el caso de estudio o experimento.
5. Realizar las consultas de análisis sobre la base de datos implementada.

Descripción

Las bases de datos relacionales representan una alternativa de almacenamiento bien conocida. Sin embargo, el creciente volumen de datos y variedad de aplicaciones no tradicionales (por ejemplo, datos que por su naturaleza forman grafos, son semiestructurados o no estructurados) requieren un conocimiento nuevo sobre alternativas de modelado e implementación de bases de datos.

A veces estos modelos se llaman NoSQL o NewSQL y representan la perspectiva de buscar soluciones de almacenamiento y procesamiento diferentes a lo ofrecido por los DBMS relacionales. Existe una gran variedad de clasificaciones de las bases de datos NoSQL y en muchas ocasiones, estas bases de datos implementan varios enfoques diferentes, lo que les hace interesantes.

Como diferentes modelos se utilizan para situaciones particulares, deben obtener conocimientos básicos que les permitan solucionar un problema, para ello deben identificar, por ejemplo, el modelo requerido, cómo se implementa y en qué casos se recomienda. Como una segunda parte de esta etapa ustedes deben implementar el modelo identificado con el motor de base de datos seleccionado. Una vez realizada esta implementación o en paralelo, se deben explicar que estructuras utiliza para implementar el modelo seleccionado y porque las seleccionó para su trabajo, que forma de almacenamiento utiliza, cuál lenguaje de consultas implementa, entre otras características que consideren interesantes. Se espera que estudien la amplia literatura (incluyendo libros gratuitos) que se encuentra en Internet y asimilen el material estudiado de forma que les permita implementar su caso de estudio y compartir el conocimiento con otras personas.

Cada uno de los diferentes tipos de bases de datos NoSQL puede requerir diferentes acercamientos para la exitosa terminación de esta etapa de proyecto. Sin embargo, para facilitar la realización de esta etapa, se sugiere los siguientes planteamientos que no necesariamente deben ser ejecutados en forma lineal; más bien se propone su ejecución paralela:

1. Selección de aspectos de análisis sobre un problema particular donde se justifica el uso del modelo o modelos seleccionados: en esta etapa se debe utilizar el conocimiento sobre los típicos escenarios de análisis, en los cuales los modelos propuestos se usan con éxito. Es importante considerar las limitaciones que el modelo relacional puede representar si se usara en la aplicación propuesta, por ejemplo, en función de escalabilidad, rendimiento o modelado.



2. Selección de datos apropiados: en diferentes sitios de Internet existe una variedad de repositorios de datos que pueden ser usados para la implementación del proyecto, por ejemplo:
 - Conjuntos de datos públicos provisto por Amazon (<https://aws.amazon.com/es/public-data-sets/>)
 - Colección de conjuntos de datos de redes de Stanford (<http://snap.stanford.edu/data/index.html>)
 - Otros conjuntos de datos accesibles desde academia (<http://www.padjo.org/datasets/>, <http://kdd.ics.uci.edu/>, <http://archive.ics.uci.edu/ml/>, <https://www.kaggle.com/datasets>)
 - Colección de datos de sensores inalámbricos provistos por la Universidad de North Carolina: <https://home.uncg.edu/cmp/downloads/lwsndr.html>
 - Datos disponibles en el proyecto “City Plus” de Unión Europea: <http://iot.ee.surrey.ac.uk:8080/datasets.html>
 - Algunos sitios no-oficiales: http://hadoopilluminated.com/hadoop_illuminated/Public_Bigdata_Sets.html, <https://www.datasciencecentral.com/a-plethora-of-data-set-repositories/>
 - Datos de astronomía: https://fits.gsfc.nasa.gov/fits_home.html
 - Referencia a las compañías grandes y sus casos de estudios: https://www.bernardmarr.com/img/bigdata-case-studybook_final.pdf
3. “Modelaje” de estructuras: debido a que las bases de datos NoSQL se llaman schemaless, en general en la práctica encontrarán poco uso de los modelos conceptuales y lógicos para su implementación. Por lo general se comienza con la parte de programación que podemos llamar físico (actualmente, esta es la práctica común). Sin embargo, para este curso usted debe hacer uso de algún modelado de su selección con su respectiva justificación bibliográfica. La finalidad de esta representación es poder expresar en forma gráfica los datos de interés y así entender mejor los detalles de la implementación que se desarrollarán. Al desarrollar este “esquema” deben poner las referencias bibliográficas relacionadas con el modelado utilizado.
4. Caso de estudio o experimento, con el respectivo listado de escenarios de análisis con las consultas que se pretenden desarrollar o pruebas que se pretenden realizar. Deben especificarse al menos tres mediciones que se contrasten, por ejemplo, dificultad o facilidad de consulta, eficiencia o exactitud. En caso de duda, por favor consulte con su profesor.

Presentación del caso de estudio o elaboración del experimento
Descripción del problema o aspectos de análisis
Descripción y justificación de los datos
Modelado de los datos



Escenarios de análisis
Descripción de cómo se efectuará el caso de estudio o experimento
Preparación de la presentación

Entregable:

- Una especie manual técnico que permita replicar lo realizado.
- Scripts o cualquier otro material que considere pertinente.
- El conjunto de datos seleccionado (o enlace de donde se obtuvieron) para el caso de estudio o experimento.
- Presentación de caso de estudio o experimento.

Evaluación:

- Contenido del manual técnico y desarrollo del caso de estudio o experimento: 60%
- Organización y redacción del documento: 10%
- Presentación: 30% (Deben participar todas las personas miembros del grupo, la NO participación implica la pérdida del todo el puntaje en el proyecto).

Fecha de entrega:

- 27/06/2024

Bibliografía

Camelo C., García J. y Sousa E. "Facilitadores de los procesos de compartir conocimiento y su influencia sobre la innovación". Cuadernos de Economía y Dirección de la Empresa. Número 42. 2010.