

# La Gaceta

## ÓRGANO OFICIAL

### DE LA UNIVERSIDAD NACIONAL DE INGENIERÍA

""""""""""  
AÑO LVII LIMA 10 DE FEBRERO DE 2022 NÚMERO 020  
""""""""""

#### UNIVERSIDAD NACIONAL DE INGENIERÍA Escuela Central de Posgrado

Se invita a la comunidad universitaria a participar de la videoconferencia de la defensa pública virtual de la Tesis de **DOCTORADO EN INGENIERÍA DE SISTEMAS**, del Mag. José Luis Ponce Vergara, a realizarse el día lunes 14 de febrero, a las 09 h 00.

#### TÍTULO DE LA TESIS:

**“METODOLOGÍA PARA DETERMINAR LOS FACTORES QUE INFLUYEN EN EL RENDIMIENTO DE SISTEMAS DE GESTIÓN DE BASES DE DATOS RELACIONALES CON APOYO DE APRENDIZAJE DE MÁQUINA ESTADÍSTICO”.**

#### ASESOR:

**Dr. TINO EDUARDO REYNA MONTEVERDE**  
Docente de la Facultad de Ingeniería Industrial y de Sistemas  
Universidad Nacional de Ingeniería

#### RESUMEN

Los procesos de negocios que soportan sus operaciones con aplicaciones que interactúan con Sistemas de Gestión de Bases de Datos Relacionales (*Relational Database Management Systems* o abreviadamente *RDBMS*) pueden incrementar su productividad a través de la identificación de los factores que afectan el rendimiento de las ejecuciones de las sentencias SQL (*Structured Query Language*) que conforman su carga de trabajo (*workload*), especialmente la carga soportada por las aplicaciones del ambiente de Producción, que es recurrente en el tiempo. Entiéndase al verbo “afectar” (referido en su tiempo presente indicativo “afectan”) como sinónimo de “influir”, y pueden ser intercambiados libremente para la mejor comprensión de esta tesis. Los sistemas e infraestructuras de computación que soportan a los gestores de Base de Datos Relacionales son complejos en esencia, ya sea que estos se desplieguen en ambientes administrados directamente por el dueño (cliente) de los sistemas (plataformas *on-premise*), o sean administrados por un tercero (plataformas en modalidad *housing, hosting, cloud computing* o *híbrida*). La complejidad radica en que las infraestructuras están conformadas por una diversidad de componentes y plataformas interconectadas, donde se pueden mencionar a servidores, procesadores, memorias, discos, redes, interfaces, repositorios y software.

Los sistemas de gestión de Bases de Datos relacionales modernos registran la información del procesamiento de una sentencia SQL a lo largo de su ciclo de vida, en la forma de planes, estadísticas y métricas, desde que se solicita su ejecución (fase de análisis sintáctico y semántico) hasta que se

obtienen los resultados de la misma (fases de ejecución y retorno). Estos conjuntos de datos son insumos importantes para este emprendimiento, tal como se apreciará en el desarrollo de la tesis.

La tesis que se presenta propone los fundamentos y lineamientos para construir un modelo de identificación de factores que afectan el rendimiento de las ejecuciones de las sentencias SQL que se ejecutan en un RDBMS, introduciendo al modelo propuesto la aplicación de algoritmos de aprendizaje de máquina estadístico (*Principal Component Analysis* y *Canonical Correlation Analysis*), que explotan los planes, estadísticas y métricas generadas por las ejecuciones de las sentencias SQL que conforman una carga de trabajo recurrente. Considérese para este efecto que las sentencias SQL son generadas por aplicaciones de negocios comerciales operando en ambientes de Producción.

El propósito de la tesis es presentar un cuerpo de conocimientos, plasmado en un modelo, que favorezca la identificación de los factores que influyen en el rendimiento de las operaciones que realiza un sistema de gestión de Base de Datos Relacional.

Los resultados del modelo propuesto en la investigación son refrendados, y a su vez puestos en evaluación, por los hallazgos obtenidos a través de la identificación teórica de los factores considerados a priori como relevantes por un RDBMS en el rendimiento de las ejecuciones de la carga de trabajo SQL, la cual se fundamenta en un razonamiento analítico-deductivo del proceso generación de planes de ejecución realizado por el optimizador de un sistema de gestión de Base de Datos Relacional.

Se estima que, gracias al modelo propuesto, podrían mejorar los niveles de rendimiento de los sistemas de información basados en este tipo de soluciones, especialmente en plataformas de producción por lo expuesto anteriormente. El modelo tomaría como entrada la información del ciclo de vida del procesamiento SQL registrada en la misma plataforma de producción o similares. La salida del modelo son los factores identificados, que a su vez son origen de los cuellos de botella de rendimiento que se observan, u observarían, durante la operación de los sistemas de gestión de Bases de Datos relacionales.

### **ENLACE**

**La Escuela Central Posgrado UNI le está invitando a una reunión de Zoom programada.**

**Tema: SUSTENTACIÓN DE TESIS DOCTORAL**

**FECHA: 14 febrero 2022**

**HORA: 09:00 a. m. Lima**

**Unirse a la reunión Zoom <https://us02web.zoom.us/j/85981039712>**

**ID de reunión: 859 8103 9712**



**EDITOR: SECRETARIA GENERAL UNI  
IMPRESA DE LA EDUNI**