

Tipo de artículo: Artículo original

Subsistema para la Creación, Control y Desarrollo Automatizado de Scripts de simulaciones en LAMMPS (SimBox)

Subsystem for the Creation, Control and Automated Development of Simulation Scripts in LAMMPS (SimBox)

Gianni Martínez Padrón ^{1*} , <https://orcid.org/0000-0001-6810-578X>

Frank Alberto Broche Gómez ¹ , <https://orcid.org/0000-0003-2484-418X>

Edisel Navas Conyedo ¹ , <https://orcid.org/0000-0002-4648-7622>

Jorge Gulín González ¹ , <https://orcid.org/0000-0001-7912-2665>

¹ Centro de Estudios de Matemática Computacional (CEMC), Facultad de Ciencias y Tecnologías Computacionales (CiTEC), Universidad de las Ciencias Informáticas (UCI), Carretera a San A. de los Baños, Km 2½, Torrens, La Lisa, La Habana, Cuba.

* Autor para correspondencia: gulinj@uci.cu

Resumen

En el presente es cada vez más evidente la aparición de patógenos (virus, hongos, bacterias) que afectan la estabilidad y existencia de la vida humana, esto ha llevado a los científicos a la búsqueda de herramientas que permitan encontrar soluciones con mayor rapidez y confiabilidad. Los procesos químico-biológicos involucrados en la interacción entre un patógeno y el organismo humano son de alta complejidad y en muchos casos no es posible realizar experimentos in vivo sobre el estado presente de la infección y su posible evolución. En este contexto, las simulaciones por computadora son de gran utilidad pues permiten estudiar fenómenos y procesos a escalas micro y mesoscópica y realizar experimentos virtuales de procesos que son difíciles de estudiar a partir de técnicas experimentales. LAMMPS es un paquete informático que permite la modelación y simulación de procesos físico-químicos en diferentes escalas espacio-temporales. LAMMPS es particularmente útil en el estudio de sistemas en la mesoescala y, por ende, para investigar procesos en microfluidos de interés biológico como el plasma sanguíneo. En este trabajo se propone una herramienta nombrada Subsistema para la Creación, Control y Desarrollo Automatizado de Scripts de simulaciones en LAMMPS (SIMBOX) que tiene la función de asegurar la comunicación con el simulador LAMMPS para administrar los parámetros y controlar las simulaciones que ejecute el usuario, así como llevar a cabo la automatización del proceso de desarrollo de Scripts de simulación. Para ello se estudió el proceso de negocio asociado con la gestión de simulaciones y ya sistemas existentes.

Palabras clave: Simulaciones moleculares; LAMMPS; Subsistema para la Creación, Control y Desarrollo Automatizado de Scripts en LAMMPS.

Abstract

In the present is increasingly the appearance of pathogens (viruses, fungi, bacteria) that affect the stability and existence of human life. This has led scientists to search for tools that allow solutions to be found more quickly and more reliably. The chemical-biological processes involved in the interaction between a pathogen and the human body are highly complex and, in many cases, it is not possible to carry out in vivo experiments on the current state of the infection and its possible evolution. In this context, computer simulations are very useful because they allow the study of phenomena and processes at micro and mesoscopic scales and to carry out virtual experiments on processes that are difficult to study using experimental techniques. LAMMPS is a software package that allows the modeling and simulation of physical-chemical processes on different space-time



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional** (CC BY 4.0)

scales. LAMMPS is particularly useful in the study of mesoscale systems and, therefore, to investigate microfluidic processes of biological interest such as blood plasma. In this work, a tool named Subsystem for the Creation, Control and Automated Development of Simulation Scripts in LAMMPS (SIMBOX) is proposed, which has the function of ensuring communication with the LAMMPS simulator to manage the parameters and control the simulations executed by the user, as well as to carry out the automation of the development process of simulation Scripts. For this, the business process associated with the management of simulations and existing systems was studied.

Keywords: molecular simulations; LAMMPS; Subsystem for the Creation, Control and Automated Development of Scripts in LAMMPS.

Recibido: 13/06/2022
Aceptado: 28/07/2022
En línea: 01/08/2022

Introducción

Desde un punto de vista científico, se considera a la simulación como el artificio contextual que referencia la investigación de una única o un conjunto de hipótesis de trabajo utilizando modelos. Una definición bastante acertada establecida por Robert. E. Shannon plantea que: “La simulación es el proceso de diseñar un modelo de un sistema real y llevar a términos de experiencia con él, con la finalidad de comprender el comportamiento del sistema o evaluar nuevas estrategias- dentro de los límites impuestos por un cierto criterio o conjunto de ellos- para el funcionamiento del sistema” (Shannon, 1998).

Las simulaciones por computadora se desarrollaron a la par del vertiginoso progreso tecnológico de los últimos 30 años. Un primer despliegue a gran escala de estas simulaciones fue en un proyecto de investigación y desarrollo conocido como “Proyecto Manhattan”, el cual tuvo lugar durante la Segunda Guerra Mundial para recrear una detonación nuclear, con la finalidad de diseñar una bomba nuclear. Este proyecto fue liderado por los Estados Unidos, contando con el apoyo del Reino Unido y Canadá (BÚ, 1994).

Se define como modelado computacional como el uso de las computadoras para simular y estudiar sistemas complejos haciendo uso de técnicas de las matemáticas, la física y la informática. Un modelo computacional está contenido por numerosas variables que caracterizan el sistema a estudiar (Reimann, 2017). Para realizar la simulación es necesario ajustar las variables, solas o combinadas, y observar y analizar los resultados finales. El modelado computacional permite a los científicos realizar miles de experimentos simulados por computadora. Los miles de experimentos por computadora identifican unos pocos casos experimentales que tienen más probabilidades de resolver el problema bajo estudio, facilitando en la mayoría de los casos, el trabajo manual en los laboratorios y ahorrando recursos energéticos y materiales.



Actualmente los modelos computacionales pueden estudiar un sistema biológico en múltiples niveles, lo cual se conoce como “modelado multiescala”. Este tipo de modelos son muy útiles porque permiten estudiar una variedad de fenómenos y procesos que transcurren en diferentes escala espacio-temporales, entre ellos los que ocurren en el torrente sanguíneo, como pueden ser la creación de trombos o la propagación de virus y células cancerígenas.

Entre los programas disponibles para la comunidad científica para simular sistemas atómico-moleculares multiescalas, se destaca Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS); concebido para el desarrollo y análisis de simulaciones a nivel atómico/molecular. LAMMPS es un código de dinámica molecular clásica que permite modelar conjuntos de partículas en los diferentes estados (líquido, sólido o gaseoso). Además, ofrece la modelación de sistemas atómicos, poliméricos, biológicos, sistemas metálicos, granulados, y de grano grueso utilizando una variada serie de campos de fuerzas y condiciones de contorno. El programa integra las ecuaciones de movimiento de Newton para las interacciones entre los átomos, moléculas y/o partículas macroscópicas que se encuentren en el espacio de simulación definido. La modelación de los sistemas de partículas puede ser tanto en dos dimensiones (2D) como en 3 (3D), dependiendo de la finalidad de la investigación (Laboratories, 2021).

Con la intención de asegurar eficiencia computacional, LAMMPS utiliza listas de elementos vecinos permitiendo realizar un seguimiento de las partículas cercanas. En términos de trabajo en paralelo, utiliza varias técnicas de descomposición espacial de la partición del dominio de simulación 3D en pequeños sub-dominios, asignando un sub-dominio a cada procesador. Los procesadores son los encargados de comunicar y almacenar la información referente a los llamados átomos “fantasma” que rodean a sus subdominios. Este sistema cuenta con varias ventajas, pero evidentemente las más impresionantes son la capacidad de trabajar con simulaciones de manera simultánea, lo que se conoce en términos informáticos como trabajo en paralelo, y por otra parte que este sistema es libre de pago, de código abierto y posee un API que permite incluir nuevas funcionalidades acorde a las necesidades del investigador. Su desventaja más notable es la dificultad de trabajo dado que todo el proceso se realiza de manera manual por consola, lo cual se hace bastante tedioso cuando se enfoca en una investigación complicada o extensa, debido a que el estudio de sistemas complejos requiere de analizar conjuntos grandes de moléculas, lo cual no solo dificulta la definición del script de trabajo, sino que también trae graves consecuencias a nivel de tiempo de procesamiento. Por todo lo expresado anteriormente se han definido los siguientes elementos de investigación.

En este artículo se presenta el Subsistema para la Creación, Control y Desarrollo Automatizado de Scripts de simulaciones en LAMMPS (SIMBOX) que tiene la función de asegurar la comunicación con el simulador LAMMPS para administrar los parámetros y controlar las simulaciones que ejecute el usuario, así como llevar a cabo la automatización del proceso de desarrollo de Scripts de simulación.



Materiales y métodos

El Proyecto GESES-PETRI (con nombre de producto GETRIX), consiste en el desarrollo de un sistema web cuya única y exclusiva función es facilitar el trabajo de simulaciones con LAMMPS para todos los científicos de su comunidad. Para esto se ha diseñado un sistema que consta de 3 componentes fundamentales y un almacén de datos.

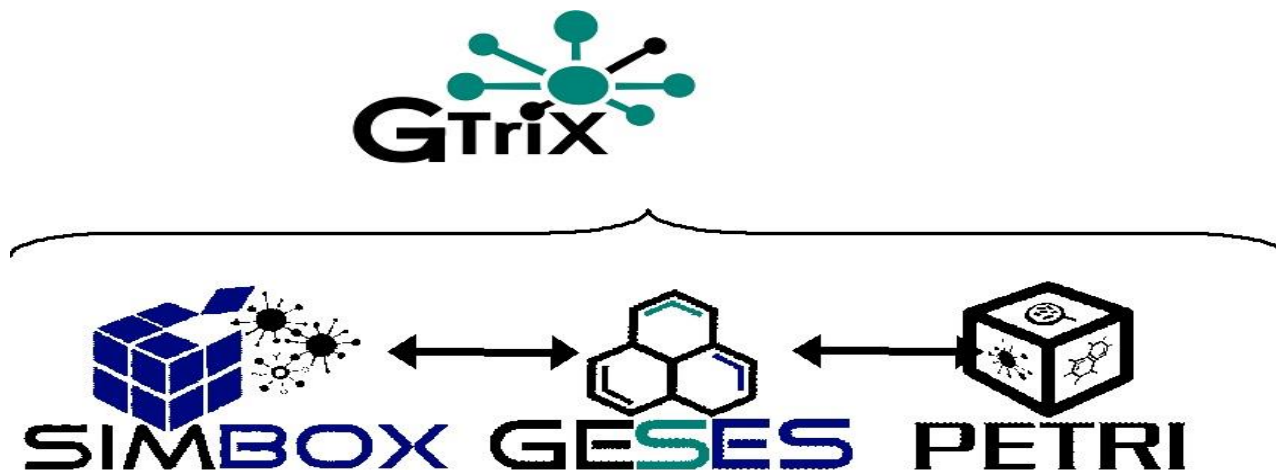


Figura 1: componentes fundamentales del sistema.

- *Sistema Gestor de Simulaciones y Espacios de Simulación (GESES)*- este componente es el sistema central del proyecto, el cual sería accedido vía web con autenticación única para cada usuario. Entre sus principales funciones están gestionar varias instancias de simulaciones LAMMPS desde una misma ventana matriz, ajustar sus parámetros y observar la consola de progreso de la simulación, crear espacios de simulación con moléculas previamente seleccionadas o de su preferencia, ajustar el tamaño del espacio y definir las cantidades de moléculas individuales a insertar. Para garantizar la seguridad, los datos serán alojados en un servidor o HPC que permita ejecutar las instancias de simulaciones LAMMPS.

- *Subsistema para la Creación de Espacios de Simulación (PETRI)*- Este microsistema estará solamente relacionado con GESES y tendrá como función principal la creación de espacios de simulación. Tendrá acceso al almacén de datos del Sistema GESES-PETRI.

- *Subsistema para la Creación, Control y Desarrollo Automatizado de Scripts de simulaciones en LAMMPS (SIMBOX)*- Este microsistema tendrá como función principal la comunicación con el simulador LAMMPS



para administrar los parámetros y controlar las simulaciones. Solamente tendrá comunicación con el sistema GESES, el simulador LAMPPS y con el almacén de datos.

Modelo Conceptual:

Para el desarrollo de cualquier sistema es necesario conocer, con alta prioridad, cómo funciona el negocio sobre el que se va a plantear el producto como solución. De tal manera, tras un análisis del negocio en cuestión se obtuvo el siguiente Modelo Conceptual donde se muestran los principales conceptos que interactúan en el negocio, así como las relaciones entre ellos.

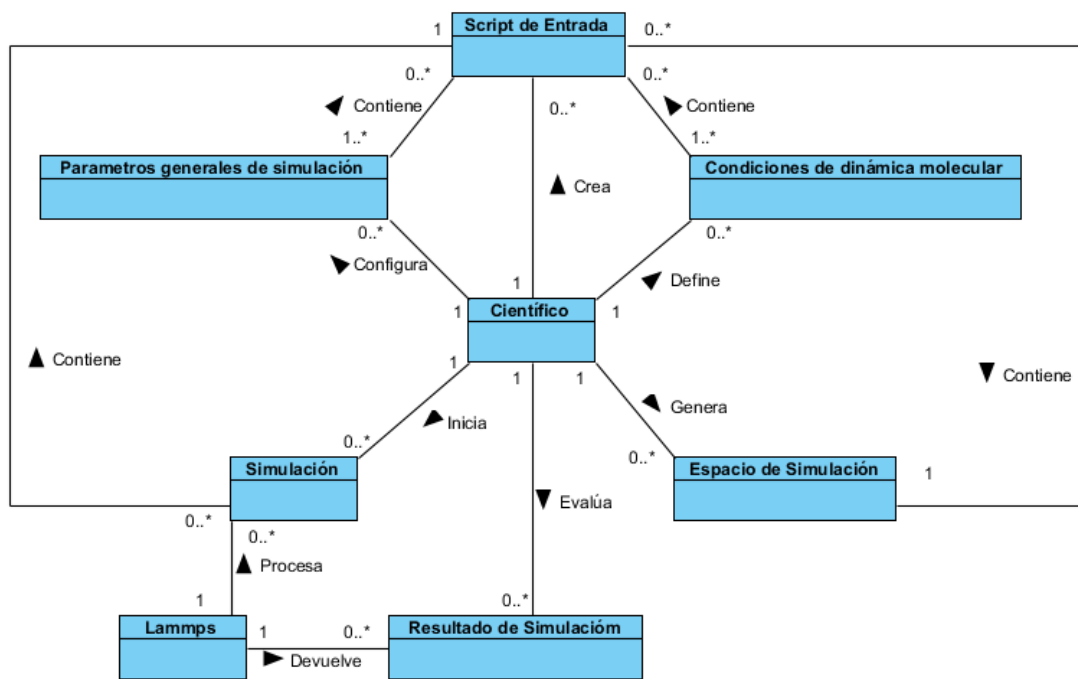


Figura 2: Modelo Conceptual.

Descripción del Sistema

El Subsistema para la Creación, Control y Desarrollo Automatizado de Scripts de simulaciones en LAMMPS (SIMBOX) tendrá como función principal la comunicación con el simulador LAMMPS para administrar los parámetros y controlar las simulaciones del usuario, así como llevar a cabo una automatización del proceso de desarrollo de Scripts de simulación. Contará con una interfaz de comunicación encargada de vincular el subsistema



SIMBOX con el sistema GESES. Incluye una vista de administración a través de la cual se configurarán los recursos disponibles para ejecutar simulaciones, así como las configuraciones pertinentes a (SIMBOX). Consumirá del almacén de datos de GESES-PETRI los datos necesarios para ejecutar cada simulación.

Ingeniería de requisitos

A continuación, se recogen los requerimientos candidatos y sus derivaciones, tanto a nivel de usuario como de sistema, obtenidos tras el análisis realizado durante la etapa de ingeniería de requisitos.

➤ **Requisitos candidatos:**

Creación y control de simulaciones de forma asíncrona.

Guardar en base de datos centralizada los resultados de la simulación.

Comunicación con el simulador LAMMPS.

Interfaz de comunicación con GESES.

Vista administración donde se configuran los recursos del sistema.

➤ **Requisitos a nivel de Usuario:**

Requisitos funcionales:

Gestionar simulaciones utilizando el simulador LAMMPS.

Almacenar los resultados de simulación en la base de datos centralizada de GESES.

Vincular con el sistema GESES para la recibir las peticiones de simulación.

Presentar una interfaz de usuario administrador para gestionar los parámetros del sistema.

Requisitos no funcionales:

Las simulaciones deben crearse de manera simultánea.

La interfaz de usuario administrador debe ser simple e intuitiva.

Las transferencias de información deben realizarse de manera segura.

El sistema debe ser capaz de recuperarse de posibles fallos.

➤ **Requisitos a nivel de Sistema:**

Requisitos funcionales:



Gestionar simulaciones utilizando el simulador LAMMPS.

- ◆ Elaborar script de entrada.
- ◆ Archivar script de entrada.
- ◆ Enviar script de entrada a LAMMPS.
- ◆ Notificar posibles errores durante el procesamiento del script de entrada.
- ◆ Iniciar simulación.
- ◆ Detener simulación.

Almacenar los resultados de simulación en la base de datos centralizada de GESES.

- ◆ Archivar resultados de la simulación.

Vincular con el sistema GESES para la recibir las peticiones de simulación.

- ◆ Implementar una interfaz de comunicación con el sistema central GESES para atender las peticiones de gestión de simulación y enviar información de monitoreo.

Presentar una interfaz de usuario administrador para gestionar los parámetros del sistema.

- ◆ Desarrollar una interfaz de usuario para la administración local de los parámetros del servidor.

Requisitos no funcionales:

Usabilidad: El sistema será un subsistema de GESES-PETRI.

Usabilidad: El subsistema se ubicará en un entorno HPC diferente al del resto del sistema.

Soporte: Empleará la base de datos centralizada de GESES-PETRI para las consultas de espacios de simulación y el alojamiento de los resultados del sistema.

Soporte: Versión de LAMMPS actualizada (todos los meses como mínimo, la comunidad lanza un nuevo parche, pero no es problema trabajar con una versión no tan desactualizada y además no es para nada complicado actualizar el software).

Soporte: La base de datos será diseñada en Postgres SQL 10.5 y el gestor de base de datos será PgAdmin 4.

Diseño: La interfaz de administración será implementada con el framework Django 2.0.

Diseño: El sistema será implementado utilizando el lenguaje de programación Python en su versión 3.7.



Resultados y discusión

Diagrama de Casos de Uso del Sistema:

Como se explicó anteriormente, esta investigación tributa a un sistema mayor, por lo tanto, a continuación, se muestra el Diagrama de Casos de Uso del Sistema GESES-PETRI.

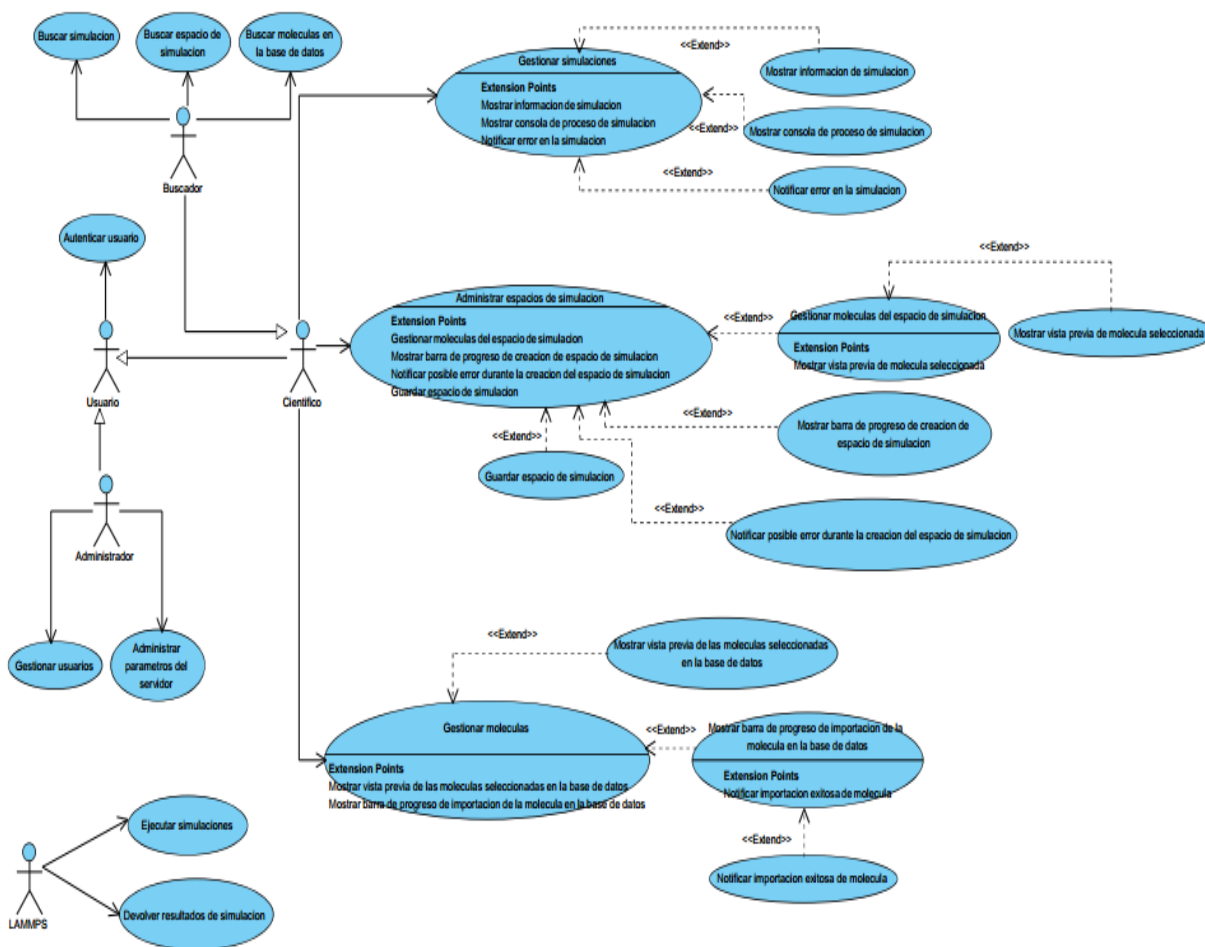


Figura 3: Diagrama de Casos de Uso del Sistema.

Arquitectura del Sistema:

El sistema de manera general se ha concebido utilizando una arquitectura por componentes donde nuestro Subsistema en cuestión constituye uno de ellos. Para la implementación del Subsistema Simbox se utilizará una arquitectura



Maestro-Esclavo. Este patrón de diseño es muy ventajoso al crear aplicaciones multitarea, pues brinda un enfoque más modular para el desarrollo de aplicaciones gracias a su funcionalidad de bucle múltiple, además de ofrecer un mayor control de la administración del tiempo de su aplicación. Cada ciclo paralelo se trata como una tarea o hilo independiente, donde un hilo es definido como una parte de un programa que se puede ejecutar de manera independiente de otras partes del sistema (Sommerville, 2009).

Si la aplicación no utiliza subprocesos separados, el sistema la como un solo subproceso. Cuando divide en varios subprocesos, cada uno de ellos comparte el tiempo de procesamiento por igual. Esto permite más control sobre cómo se cronometra su aplicación, y ofrece al usuario un mayor control sobre su aplicación.

Modelo de Datos:

El modelo de datos tiene la función de describir los elementos de la realidad que intervienen en un problema dado, así como las relaciones entre ellos. El modelado de los datos constituye un aspecto fundamental para el desarrollo de cualquier aplicación donde se almacenen datos. A continuación, se evidencian los diagramas correspondientes al modelo de clases persistentes de base de datos y al modelo entidad-relación.

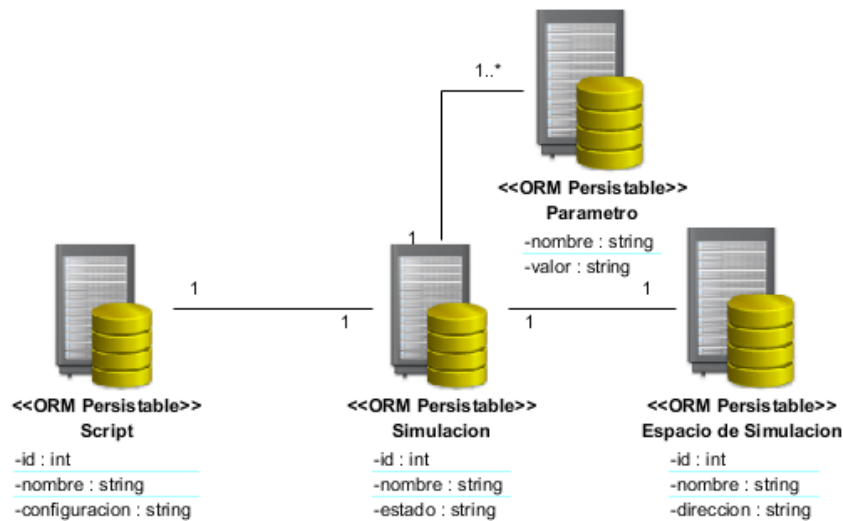


Figura 4: Modelo de Datos.



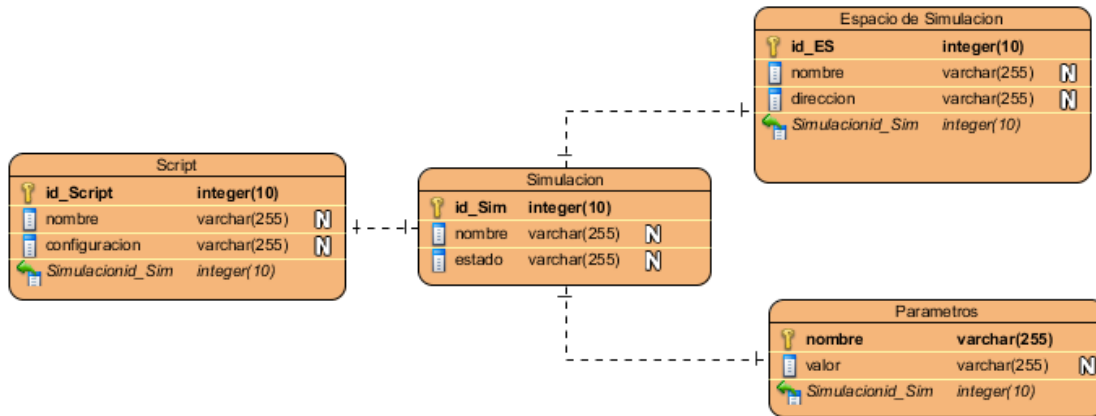


Figura 5: Modelo Entidad-Relación.

Modelo de Despliegue:

El modelo de despliegue es aquel que describe la distribución física del sistema en términos de cómo se distribuye la funcionalidad entre los nodos de cómputo. Se muestra como un conjunto de nodos unidos por conexiones de comunicación, de forma que se describan las relaciones físicas de los distintos nodos que componen el sistema.

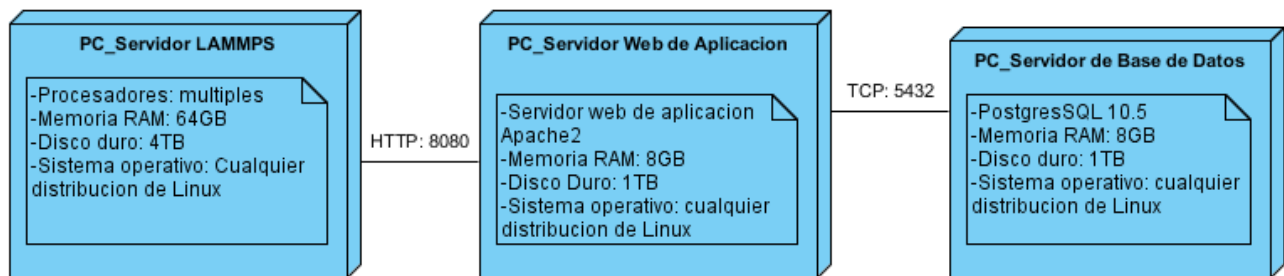


Figura 6: Modelo de Despliegue.

Conclusiones

Con la culminación de la investigación se concluye que:

- Como resultado del uso de las técnicas referentes a la Metodología de la Investigación se construyó un marco teórico con el cual se puede establecer las bases del desarrollo de la investigación, garantizando una mejor comprensión del proceso de desarrollo del Sistema para la Creación y Control de Simulaciones.



- Durante el proceso de desarrollo de software, donde se usó de guía la metodología RUP, se logró definir un conjunto de artefactos ingenieriles necesarios para el diseño y desarrollo del Sistema para la Creación y Control de Simulaciones

Conflictos de intereses

Los autores no poseen conflictos de intereses.

Contribución de los autores

1. Conceptualización: Gianni Martínez Padrón, Frank Alberto Broche Gómez, Edisel Navas Conyedo, Jorge Gulín González.
2. Curación de datos: Gianni Martínez Padrón, Frank Alberto Broche Gómez.
3. Análisis formal: Gianni Martínez Padrón, Frank Alberto Broche Gómez.
4. Investigación: Gianni Martínez Padrón, Frank Alberto Broche Gómez.
5. Metodología: Gianni Martínez Padrón, Frank Alberto Broche Gómez.
6. Administración del proyecto: Jorge Gulín González.
7. Software: Gianni Martínez Padrón.
8. Supervisión: Edisel Navas Conyedo, Jorge Gulín González.
9. Validación: Gianni Martínez Padrón.
10. Visualización: Gianni Martínez Padrón.
11. Redacción – borrador original: Gianni Martínez Padrón, Frank Alberto Broche Gómez, Edisel Navas Conyedo, Jorge Gulín González.
12. Redacción – revisión y edición: Gianni Martínez Padrón, Frank Alberto Broche Gómez, Edisel Navas Conyedo, Jorge Gulín González.

Financiamiento

La investigación no requirió fuente de financiamiento externa.

Referencias



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional** (CC BY 4.0)

- Asencio, Jesús Martínez. 2018. Simulación Atomística de manipulación e irradiación de grafito y grafeno. 2018.
- BÚ, Raúl Coss. 1994. Simulación: Un enfoque práctico. s.l. : Editorial Limusa, 1994.
- David Alejandro Martínez Pérez, Edisel Navas Conyedo y Jorge Gulín Gonzalez. 2016. Herramientas de visualización dinámica de simulaciones del proceso de difusión en micro-fluidos con componentes biológicos. Cuba : s.n., 2016.
- Dror, Scott A. Hollingsworth and Ron O. 2018. Molecular Dynamics Simulation for All. 2018.
- Fritzon, Peter. 2011. Introducción al Modelado y Simulación de Sistemas Técnicos y Físicos con Modelica. s.l. : Wiley-IEEE Press, 2011.
- Kelton, Averill M. Law & David. 1991. Simulation Modeling & Analysis. 1991.
- Laboratories, Sandia National. 2021. LAMMPS Users Manual. [En línea] 2021. <http://lammps.sandia.gov>:Sandia National Laboratories.
- Paradigm, Visual. 1999. <http://www.visual-paradigm.com/product/vpuml>. [En línea] 1999.
- Pressman, Roger. 2010. Ingeniería de Software, un enfoque práctico. 2010.
- Reimann, Daniela. 2017. Teaching and Learning of Computational Modeling in Creative Shaping Processes. 2017.
- Rui Xu y Donald C. Wunsch, II. 2008. Clustering. 2008.
- Shannon, Robert E. 1998. Introduction to the art and science of simulation. 1998.
- Sommerville. 2009. Ingeniería de Software. Enfoque Práctico. 2009.
- Sommerville, Ian. 2009. Ingeniería de Software. Madrid : Pearson Educación, 2009.
- Valcárcel, Carlos Cruz. 2055. Simulación mediante híbridos clásicos-cuánticos de la relajación vibracional de moléculas en disolución. Murcia: s.n., 2055.

