



ANEXO 1

FORMATO PARA LA PRESENTACIÓN DE PROYECTOS DE INVESTIGACIÓN
CON EL FINANCIAMIENTO DEL FEDU

1. Título del proyecto

Grafos e hipergrafos para representar tareas en entornos simulados de computación de alto rendimiento

2. Área de Investigación

Área de investigación	Línea de Investigación	Disciplina OCDE
Computación	Ciencias de la Computación	1.02.01

3. Duración del proyecto (meses)

12 meses

4. Tipo de proyecto

Individual	<input type="radio"/>
Multidisciplinario	<input checked="" type="radio"/>
Director de tesis pregrado	<input type="radio"/>

4. Datos de los integrantes del proyecto

Apellidos y Nombres	Ingaluque Arapa Marga Isabel
Escuela Profesional	Ingeniería de Sistemas
Celular	988557025
Correo Electrónico	miingaluque@unap.edu.pe

Apellidos y Nombres	Arcaya Coaquira William Eusebio
Escuela Profesional	Ingeniería de Sistemas
Celular	951680806
Correo Electrónico	wearcaya@gmail.com

Apellidos y Nombres	Flores Velásquez Edelfré
Escuela Profesional	Ingeniería de Sistemas
Celular	951680705
Correo Electrónico	edelfre.flores@gmail.com

- I. Título (El proyecto de tesis debe llevar un título que exprese en forma sintética su contenido, haciendo referencia en lo posible, al resultado final que se pretende lograr. Máx. palabras 25)

Grafos e hipergrafos para representar tareas en entornos simulados de computación de alto rendimiento



- II. Resumen del Proyecto de Tesis (Debe ser suficientemente informativo, presentando -igual que un trabajo científico- una descripción de los principales puntos que se abordarán, objetivos, metodología y resultados que se esperan)

En el sitio web top500.org se puede apreciar el ranking de las supercomputadoras que se tiene en el mundo actualmente, este tipo de infraestructura computacional está orientada a ejecutar aplicaciones con una demanda de alto poder computacional. En este contexto los recursos deben ser óptimamente utilizados, y uno de esos recursos es el tiempo.

La planificación de tareas de la aplicación que se ejecutará en ese entorno consiste en asignar las tareas a los procesadores considerando el tiempo de ejecución estimado así como la transferencia de datos que se da entre las tareas. Las tareas pueden modelarse utilizando grafos dirigidos no cíclicos que a su vez pueden tratarse con algoritmos que minimicen el camino de ingreso a la salida del grafo "makespan". Los grafos pueden ser de muy alta complejidad para lo cual existen bibliotecas con algoritmos estándar tal como **metis**.

En la presente investigación se hará experimentos comparativos aplicando los algoritmos estándar en datasets convencionales. A su vez se revisará el uso de hipergrafos.

- III. Palabras claves (Keywords) (Colocadas en orden de importancia. Máx. palabras: cinco)

Grafos, hipergrafos, computación de alto rendimiento

- IV. Justificación del proyecto (Describa el problema y su relevancia como objeto de investigación. Es importante una clara definición y delimitación del problema que abordará la investigación, ya que temas cuya definición es difusa o amplísima son difíciles de evaluar y desarrollar)

La programación de tareas siempre es requerida para los sistemas computacionales para asignar tareas a los procesadores luego de analizar el desempeño usando modelos para representarlos en los entornos de simulación apropiados. Una buena programación de tareas que utilice todo el potencial del entorno heterogéneo con los diferentes recursos que se tenga es de valiosa importancia. El planificador de tareas usa los costos de computación y comunicación de la aplicación para asignar las tareas en los procesadores. Esto se logra evaluando el costo total en tiempo del grafo de la aplicación en el entorno involucrado.

- V. Antecedentes del proyecto (Incluya el estado actual del conocimiento en el ámbito nacional e internacional. La revisión bibliográfica debe incluir en lo posible artículos científicos actuales, para evidenciar el conocimiento existente y el aporte de la Tesis propuesta. Esto es importante para el futuro artículo que resultará como producto de este trabajo)

Los grafos son estructuras de datos que consisten en nodos y arcos que



representan relaciones entre ellos, se pueden representar usando matrices dispersas. Los grafos son ampliamente utilizados en diversas áreas como las redes sociales, sistemas de transportación y biología computacional. Para analizar de manera efectiva los datos de grafos de larga escala se requiere algoritmos eficientes de programación de tareas. La programación de tareas mediante grafos computacionales se refiere al proceso de mapear las tareas computacionales en los recursos computacionales disponibles tales como los procesadores CPU cores, GPUs y clusters de manera eficiente y balanceada.

Investigaciones actuales en esta área se han enfocado en desarrollar algoritmos que puedan lograr un efectivo balance de las cargas de trabajo y la minimización de la sobrecarga de comunicación. Elegir un algoritmo de planificación depende de las características del grafo tal como su tamaño, su estructura y complejidad así como de los recursos disponibles.

La computación de alto rendimiento se enfoca en acelerar y optimizar el procesamiento de datos utilizando técnicas como la paralelización y la distribución de tareas. La combinación de grafos y la computación de alto rendimiento puede ser útil en diversas aplicaciones, como la minería de datos, la inteligencia artificial y la simulación de sistemas complejos.

Los grafos dirigidos acíclicos DAG son grafos que tienen direcciones en sus arcos y no tienen ciclos. Esto significa que no hay camino que permita volver al mismo nodo siguiendo los arcos. Los DAGs se utilizan en una amplia gama de aplicaciones como la programación de tareas, la gestión de proyectos, el procesamiento de lenguaje natural y la modelación de dependencias en sistemas complejos.

Los hipergrafos son la generalización de los tradicionales grafos mediante el cual también se puede modelar complejas relaciones entre más de dos entidades. En los últimos años los hipergrafos alcanzaron un inusitado interés por los investigadores de diferentes áreas incluidos el aprendizaje automático, la minería de datos así como la computación científica.

Sin embargo el procesamiento de hipergrafos de larga escala es todavía un campo de investigación especialmente en el campo de la computación de alto rendimiento HPC. La computación de alto rendimiento provee una potente plataforma para procesar grafos, pero los algoritmos convencionales para tratar grafos así como sus estructuras son insuficientes para tratar los hipergrafos, siendo un campo de investigación actual.

VI. Hipótesis del trabajo (Es el aporte proyectado de la investigación en la solución del problema)

Los grafos y los hipergrafos permiten optimizar la programación de tareas en las aplicaciones de computación de alto rendimiento

VII. Objetivo general

Desarrollar experimentos comparativos de utilizar grafos e hipergrafos en la programación de tareas en aplicaciones de computación de alto rendimiento.

VIII. Objetivos específicos

Utilizar grafos para modelar la programación de tareas y aplicar algoritmos de bibliotecas estándar (metis y otros)



Utilizar hipergrafos para modelar la programación de tareas y aplicar algoritmos de bibliotecas tal como hMetis

- IX.** Metodología de investigación (Describir el(los) método(s) científico(s) que se empleará(n) para alcanzar los objetivos específicos, en forma coherente a la hipótesis de la investigación. Sustentar, con base bibliográfica, la pertinencia del(los) método(s) en términos de la representatividad de la muestra y de los resultados que se esperan alcanzar. Incluir los análisis estadísticos a utilizar)

Al ser la computación de alto rendimiento un área de permanente innovación, la presente investigación iniciará con un alcance exploratorio. Luego al profundizar la aplicación de los conceptos y técnicas de particionamiento de grafos e hipergrafos, la investigación se tornará en experimental; tomando en cuenta que al utilizar datasets estándar, lo que se ejecutan son experimentos repetibles cuyos resultados sean replicables.

- X.** Referencias (Listar las citas bibliográficas con el estilo adecuado a su especialidad)

- [1] G. Wang, Y. Wang, H. Liu, and H. Guo, "HSIP: A Novel Task Scheduling Algorithm for Heterogeneous Computing," *Sci Program*, vol. 2016, 2016, doi: 10.1155/2016/3676149.
- [2] "Erich Strohmaier | TOP500." <https://top500.org/project/authors/erich-strohmaier/> (accessed Nov. 29, 2022).
- [3] S. Z. Koohi, N. A. W. A. Hamid, M. Othman, and G. Ibragimov, "MEMPHA: Model of Exascale Message-Passing Programs on Heterogeneous Architectures," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 11, pp. 2570–2581, Nov. 2020, doi: 10.1109/TPDS.2020.2995867.
- [4] G. Karypis, "METIS * A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-Reducing Orderings of Sparse Matrices Version 5.1.0," 2013. [Online]. Available: <http://www.cs.umn.edu/~karypis>
- [5] G. Karypis and V. Kumar, "hMETIS * A Hypergraph Partitioning Package," 1998. [Online]. Available: <http://www.cs.umn.edu/~karypis>
- [6] R. Hernández, C. Fernández, and P. Baptista, *Metodología de la investigación*, vol. 53, no. 9. 2014. doi: 10.1017/CBO9781107415324.004.
- [7] R. S. Wazlawick, *Metodologia de pesquisa para ciência da computação*.
- [8] "DIMACS Networks | Network Data Repository." <https://networkrepository.com/dimacs.php> (accessed Dec. 05, 2022).
- [9] "ISPD98/IBM Benchmarks." https://cseweb.ucsd.edu/~kastner/labyrinth_vault/benchmarks/index.html (accessed Dec. 05, 2022).
- [10] "A Benchmark Set for Multilevel Hypergraph Partitioning Algorithms | Zenodo." <https://zenodo.org/record/291466#.Y49ZA8vMLb0> (accessed Dec. 05, 2022).

- XI.** Uso de los resultados y contribuciones del proyecto (Señalar el posible uso de los resultados y la contribución de los mismos)

Un posible uso del proyecto a realizar es que cualquier investigador interesado en el área podrá tener referencia del estado de arte sobre el tema para así tener un completo y mejor panorama del tema científico que se trata en el proyecto.



XII. Impactos esperados

i. Impactos en Ciencia y Tecnología

La programación de tareas en computación de alto rendimiento es un tema vigente para la investigación científica actual en ciencias de la computación y sus afines.

ii. Impactos económicos

Con la demanda de procesamiento de la inmensa cantidad de datos que se producen en diferentes campos de investigación, así como en el normal desarrollo de las organizaciones que utilizan la computación, el mejor uso de los recursos de la computación de alto rendimiento tiene un impacto económico importante.

iii. Impactos sociales

El impacto social de este tipo de investigaciones se da en forma indirecta; en el normal desarrollo de la vida y al usar diferentes dispositivos computacionales que acceden a recursos de computación de alto rendimiento.

iv. Impactos ambientales

La computación de alto rendimiento consume enormes cantidades de energía, el estudio de su modelamiento implica optimizar el uso de recursos, entre ellos la energía, entonces una mejor utilización de recursos implica lograr menor uso de energía con el consecuente ahorro de recursos y su impacto en el medio ambiente.

XIII. Recursos necesarios (Infraestructura, equipos y principales tecnologías en uso relacionadas con la temática del proyecto, señale medios y recursos para realizar el proyecto)

Una computadora Lap Top de última generación con sistema operativo Windows 11 y Office 365.
Una computadora personal de última generación con SO Linux.
Una impresora a chorro de tinta de impresión en blanco y negro y a color.
Internet
Papel y otros útiles de escritorio

XIV. Localización del proyecto (indicar donde se llevará a cabo el proyecto)

Ciudad de Puno, distrito de Puno, provincia de Puno, Departamento de Puno



XV. Cronograma de actividades

Actividad	Trimestres			
	1	2	3	4
Presentación del Proyecto	X			
Revisión de bibliografía		X		
Desarrollo metodológico y obtención de datos		X	X	
Elaboración del informe fina				X

XVI. Presupuesto

Descripción	Unidad de medida	Costo Unitario (S/.)	Cantidad	Costo total (S/.)
Responsable del proyecto	persona	3600	03	10,800
Materiales	unidades	30	30	12,000
Servicio de Internet	meses	120	12	1440
total				24,240