

Clúster Kabré + OpenMP

IC-3101 Arquitectura de Computadoras

Instituto Tecnológico de Costa Rica, Sede Central Cartago
Escuela de Computación, Ingeniería en Computación
I Semestre 2018
Prof. Ing. Esteban Arias Méndez

INTRODUCCIÓN

La base formal de la ingeniería informática cree que la metodología de proveer un medio de hacer trabajos paralelos de cualquier tipo fue inventada por Gene Amdahl de IBM, que en 1967 publicó lo que ha llegado a ser considerado como el libro blanco inicial de procesamiento paralelo: la ley de Amdahl que describe matemáticamente el aceleramiento que se puede esperar paralelizando cualquier otra serie de tareas realizadas en una arquitectura paralela.

El término clúster se aplica a los conjuntos o conglomerados de ordenadores unidos entre sí normalmente por una red de alta velocidad y que se comportan como si fuesen una única computadora.

La tecnología de clústeres ha evolucionado en apoyo de actividades que van desde aplicaciones de supercómputo y software para aplicaciones críticas, servidores web y comercio electrónico, hasta bases de datos de alto rendimiento, entre otros usos.

El cómputo con clústeres surge como resultado de la convergencia de varias tendencias actuales que incluyen la disponibilidad de microprocesadores económicos de alto rendimiento y redes de alta velocidad, el desarrollo de herramientas de software para cómputo distribuido de alto rendimiento, así como la creciente necesidad de potencia computacional para aplicaciones que la requieran.

Simplemente, un clúster es un grupo de múltiples ordenadores unidos mediante una red de alta velocidad, de tal forma que el conjunto es visto como un único ordenador, más potente que los comunes de escritorio.

Los clústeres son usualmente empleados para mejorar el rendimiento o la disponibilidad por encima de la que es provista por un solo computador típicamente siendo más económico que computadores individuales de rapidez y disponibilidad comparables. de interfaces gráficas que permitan interactuar de forma más natural al usuario del juego con el sistema.

Propósitos

- Hacer una introducción al estudio de las supercomputadoras, programación paralela y distribuida.
- Estudiar la estructura de una supercomputadora.
- Poner en práctica la ley de Amdahl.
- Evidenciar mediante un ejemplo práctico la programación paralela y distribuida.
- Poseer un conocimiento más complementado en cuanto a la arquitectura de computadores.
- Cohesionar los conceptos anteriores con la materia del curso.

PROYECTO

Usted debe usar 4 configuraciones de clúster usando 2, 3, 4 y 5 nodos del Clúster Kabré del Colaboratorio Nacional de Computación Avanzada, CNCA, del Centro Nacional de Alta Tecnología, CeNAT. Aquí usará el Sistema Operativo Linux que corre dicho Clúster.

Debe proponer e implementar un pequeño programa paralelizable en C/C++ y ejecutarlo en el clúster utilizando la plataforma vista en los talleres llamada OpenMP. Deberá correr dicho programa en cada una de sus configuraciones de 2, 3, 4 y 5 nodos en el Clúster. Para trabajar entre nodos podría requerir utilizar adicionalmente OpenMPI.

Deberá hacer una comparación entre las configuraciones de clúster por cantidad de nodos realizada y varias leyes entre ellas: Ley Gustafson, Ley Moore, Ley Amdahl. Y explicar cada una en referencia a los tiempos de ejecución obtenidos, los nodos y las leyes.

Deberá hacer múltiples pruebas (al menos 10 de cada una) y corridas del código en cada una de las posibles configuraciones tomando los tiempos en cada caso y tabulando dicha información, luego promediándola para hacer una tabla comparativa y justificar los resultados según las leyes descritas y sus propias observaciones.

Acá se aconseja algún programa que tarde varios minutos en ejecución en una máquina convencional como su computadora de trabajo, para que puedan observarse las diferencias en la duración de tiempos entre cada configuración. Indique en la documentación el tiempo que tarda el programa en su computadora personal y brinde las características técnicas de la misma. El programa que vaya a ejecutarse debe ser altamente paralelizable para que se pueda configurar su corrida en 2, 3, 4 y 5 nodos respectivamente para el clúster y analizar los cambios.

El programa implementar debe ser diferente y único para cada grupo de trabajo en el curso.

Deberán además mostrar pruebas de acceso, fechas y pruebas de ejecución con el usuario de cada miembro del grupo en el clúster y los datos propios de cada una de sus propias máquinas.

Deberá buscar un programa que le permita calcular la cantidad de flops que corren sus máquinas de trabajo, y cada una de las configuraciones de clúster que configurará. Para cada una deberá haber descrito la configuración física de cada uno de los componentes de las computadoras, incluidas las propias: con marca, procesadores, tipo, memoria RAM: tipo, cantidad y velocidad, tipo de disco duro, cantidad y tipos de memoria caché, etc.

Finalmente, el grupo debe de generar a partir del trabajo realizado un artículo científico, con los datos principales de los equipos de trabajo utilizados, el programa empleado, las tablas de las corridas y tomas de datos, sus gráficos, fórmulas, tablas, ecuaciones usadas, referencia al material de apoyo, evidencias científicas, etc... para generar toda clase de conclusiones a cerca de las configuraciones del clúster de alto rendimiento usado. Para tal fin use el formato IEEE para artículos en conferencias en hoja tamaño carta (<https://www.ieee.org/conferences/publishing/templates.html>). En el artículo no se incluyen las pruebas de acceso al clúster por parte de cada miembro, para esto agregará un archivo de reporte de apéndice separado en PDF de documentación con todos los aspectos indicados, fotos de trabajo, pruebas y evidencias de trabajo y demás datos no relevantes para el artículo científico, pero si para demostrar su trabajo.

En este link encontrará material previamente usado en la Escuela de Veranillo en HPC en el CeNAT, <https://github.com/CNCA-CeNAT/EV-HPC-2017>

Material adicional de referencia: <https://www.youtube.com/watch?v=nExN4Bf8XI>

Disposiciones generales

Los fraudes en cualquier actividad llevada a cabo durante el semestre implicarán que se perderá el curso y se reportará la nota mínima. Además, se enviará una carta al expediente del estudiante. Podría haber defensa del proyecto de forma individual. La entrega se hará el día asignado para lo cual tendrá que presentar la documentación al correo:

earias@ic-itcr.ac.cr, su artículo y su reporte ambos en PDF, deberá además copiar a su compañero de trabajo en el correo y toda la documentación.

El trabajo es en parejas de 2 máximo.

La fecha de entrega máxima será el Lunes 4 de Junio antes de las 12 media noche. Pueden entregar el trabajo en cualquier fecha anterior.

Enviar un PDF con el reporte de sus resultados, pruebas de todos los escenarios y un mini marco teórico sobre todas las herramientas empleadas.

Utilice Git durante el desarrollo de su proyecto y muestre pruebas de uso en su reporte (no en el artículo). Sus archivos de documentación también deben estar en su repositorio Git. Incluya el link al repositorio Git en su reporte de trabajo.