

Multiprocesador Distribuidas

Victor Vladimir Quisbert Mendoza

Universidad Mayor de San Andrés

Facultad de Ciencias Puras y Naturales

Carrera de Informática

Simulación de Sistemas

victory5@hotmail.com

RESUMEN

Caracterizar las arquitecturas multiprocesador distribuidas enfocadas a cluster, grid y cloud computing, con énfasis en las que utilizan procesadores de múltiples núcleos ("multicores"), con el objetivo de modelizarlas, estudiar su escalabilidad, analizar y predecir performance de aplicaciones paralelas y desarrollar esquemas de tolerancia a fallas en las mismas.

También se trata de clasificar todas las computadoras de máximos recursos con énfasis en los sistemas de microprocesadores, con lo cual tratamos conceptos de procesadores de múltiples núcleos y software de base. Para mejorar y garantizar un máximo rendimiento eficaz

Palabras clave

Multicore ,GPGPU,GPU, HPC

1. Introducción

La investigación en Sistemas Distribuidos y Paralelos es una de las líneas de mayor desarrollo en la Ciencia Informática actual. En particular la utilización de arquitecturas multiprocesador configuradas en clusters, multiclusters, grids y clouds, soportadas por redes de diferentes características y topologías se ha generalizado, tanto para el desarrollo de algoritmos paralelos, la ejecución de procesos que requieren cómputo intensivo y la atención de servicios WEB concurrentes. El cambio tecnológico, fundamentalmente a partir de los procesadores multicore, ha impuesto la necesidad de investigar en paradigmas "híbridos", en los cuales coexisten esquemas de memoria compartida con mensajes. Asimismo la utilización de procesadores gráficos (GPGPUs) como arquitecturas paralelas presenta una alternativa para alcanzar un alto speed-up en determinadas aplicaciones.

Es importante en este contexto desarrollar nuevos paradigmas y herramientas para la programación eficiente de aplicaciones. A su vez el concepto de eficiencia se refiere tanto al aspecto computacional como el energético y el impacto del consumo sobre arquitecturas con miles de procesadores que trabajan concurrentemente debe tenerse en cuenta en la estructura de los sistemas paralelos y en la planificación de la utilización de recursos por las aplicaciones.

Asimismo, aparecen líneas de I/D tales como el scheduling dinámico basado en el consumo del sistema paralelo y de cada subsistema (llegando al nivel de cada núcleo), el control en tiempo real de la frecuencia de reloj de los procesadores para optimizar consumo, la detección en bajo nivel de errores de concurrencia, el estudio y desarrollo de lenguajes, compiladores y estructuras de datos adecuados a estas arquitecturas y la detección y tolerancia a fallos

tratando de minimizar el overhead de tiempo y aprovechando alguna redundancia en la misma arquitectura.

Por otra parte la heterogeneidad que caracteriza a los clusters y grids, así como a las redes de comunicaciones, se extiende a las nuevas arquitecturas multicore y GPGPU enfocando funcionalidades específicas para algunos núcleos, lo cual puede mejorar la performance pero al mismo tiempo complejiza el scheduling de los procesos paralelos.

La aparición de las arquitecturas tipo Cloud obliga a poner especial atención a los problemas de virtualización y predicción de performance (para la asignación dinámica de recursos).

Naturalmente a mayor potencia del Cloud, también crecen las complejidades al analizar la comunicación y el acceso a memoria en arquitecturas que están distribuidas y a su vez conformadas por placas con un número variable de procesadores multicore y/o GPGPU.

En el proyecto se ha abierto una línea específicamente dedicada a los problemas de configuración y administración eficiente de Cloud, incluyendo entre los parámetros el consumo estimado de las aplicaciones.

1.1. Ciclo de desarrollo y de investigación

- Arquitectura de procesadores multicore. Cluster de multicores. Software de base. Multithreading.
- Modelos de predicción de performance para arquitecturas tipo cluster de multicores, grids y clouds. Simulación de arquitecturas.
- Virtualización en clusters, grids y clouds. Predicción de performance aplicada a la virtualización.
- Nuevas estructuras de datos, orientadas a procesadores paralelos sobre clusters, grids y clouds.
- Nuevas técnicas de scheduling para sistemas paralelos, en particular en función del consumo de los procesadores.
- Detección de errores de concurrencia en tiempo de ejecución.
- Métricas de evaluación de performance y escalabilidad para las nuevas arquitecturas paralelas, a partir del uso de procesadores de múltiples núcleos y/o GPGPUs.
- Procesamiento paralelo basado en GPUs. Aplicación sobre clusters de multicores. Comparación de rendimiento con arquitecturas basadas en multicores "clásicos".
- Cloud computing. Software de base y overhead introducido por la administración de recursos en cloud.
- Análisis comparativo de performance en cluster y cloud para problemas de HPC.
- Detección y tolerancia a fallos (de hardware y software) en clusters, grids y clouds.



1.2. Modelo de sistema

Un procesador multicore integra dos o más núcleos computacionales dentro de un mismo “chip”. La motivación de su desarrollo se basa en incrementar el rendimiento, reduciendo el consumo de energía en cada núcleo. Una GPU (Graphics Processing Unit) es una arquitectura multicore dedicada a procesamiento gráfico, con un gran número de cores simples. En los últimos años, estas arquitecturas, fueron utilizadas para aprovechar su potencia de cómputo en aplicaciones de propósito general logrando un alto rendimiento y dando lugar al concepto de GPGPU (General-Purpose Computing on Graphics Processing Units).

Un cluster es un sistema de procesamiento paralelo compuesto por un conjunto de computadoras interconectadas vía algún tipo de red, las cuales cooperan configurando un recurso que se ve como “único e integrado”, más allá de la distribución física de sus componentes.

Cada “procesador” puede tener diferente hardware y sistema operativo, e incluso puede ser un “multiprocesador”. Cuando se conectan dos o más clusters sobre una red tipo LAN o WAN, se tiene un multicluster. La configuración más simple a considerar es la conexión de clusters homogéneos sobre una red LAN o WAN, utilizando un sistema operativo común.

Las arquitecturas tipo “Cloud” se presentan como una evolución natural del concepto de Clusters y Grids, integrando grandes conjuntos de recursos virtuales (hardware, plataformas de desarrollo y/o servicios), fácilmente accesibles y utilizables por usuarios distribuidos, vía WEB.

Estos recursos pueden ser dinámicamente reconfigurados para adaptarse a una carga variable, permitiendo optimizar su uso.

1.2.1. Aspectos de Interés

El incremento en el número de procesadores disponibles en clusters, grids y clouds obliga a poner énfasis en el desarrollo de los algoritmos de virtualización de modo de explotar la arquitectura con más de una aplicación concurrente.

La heterogeneidad es inevitable en estos sistemas paralelos complejos. A su vez es un factor que condiciona la predicción de performance y consumo.

A partir de la complejidad creciente del hardware, se hace más desafiante el desarrollo de capas de software eficiente, desde el middleware hasta los lenguajes de aplicación.

Los problemas clásicos de scheduling y mapeo de procesos a procesadores tienen nuevos objetivos (en particular los relacionados con el consumo) y deben considerar la migración

dinámica de datos y procesos en función de performance y consumo.

Los modelos de predicción de performance resultan especialmente complejos. Resulta de interés el estudio de esquemas sintéticos

(“firmas”) propios de la aplicación para estimar tiempos y consumo, ejecutando un código mínimo frente al de la aplicación real.

El tema de la detección y tolerancia a fallos de hardware y software se vuelve un punto crítico al operar sobre arquitecturas con gran número de procesadores, los cuales pueden reconfigurarse dinámicamente.

2. Investigación Experimental

- Desarrollo y evaluación de aplicaciones sobre cluster de multicores y multicluster heterogéneo basado en multicores (total 192 procesadores). Incluyendo el estudio de las alternativas de lenguajes de programación y bibliotecas de comunicación/sincronización.
- Desarrollo de la capa de software para virtualización del cluster de multicores para emular servicios de cloud computing.
- Simulación de variantes en las arquitecturas de multicores.
- Pruebas de consumo en cluster de multicores y PGPU, analizando eficiencia computacional, escalabilidad y eficiencia energética.

2.1. Formación de Recursos

En cooperación con Universidades iberoamericanas se ha implementado la Maestría en Cómputo de Altas Prestaciones y se continúa dictando la Especialización en Cómputo de altas Prestaciones y Tecnología GRID.

En esta línea de I/D existe cooperación a nivel nacional e internacional. Hay 8 Investigadores realizando su Doctorado, 4 realizando la Maestría y 4 alumnos avanzados están trabajando en su Tesina de Grado de Licenciatura. En 2011 se aprobó 1 Tesis Doctoral, 1 de Maestría, 3 de Especialista y 3 Tesinas de Grado en temas del proyecto.

3. CONCLUSIONES

Esta línea de Investigación está dentro del proyecto “Arquitectura Multiprocesador Distribuidas. Modelos, Software de Base y Aplicaciones” acreditado por el Ministerio de Educación y de proyectos específicos apoyados por organismos nacionales e internacionales.

4. REFERENCIAS

- Mc. Cool M. “Programming models for scalable multicore programming”. 2007.
<http://www.hpcwire.com/features/17902939.html>
- General-Purpose Computation on Graphics Processing Units.
<http://ggpu.org>.
- Foster I. “There's Grid in them thar Clouds”. 2 de Enero, 2008.
<http://ianfoster.typepad.com/blog/2008/01/theresgrid-in.html>.
 Noviembre, 2010.
- Burger T. W. “Intel Multi-Core Processors: Quick Reference Guide”.
http://sedici.unlp.edu.ar/bitstream/handle/10915/19893/Documento_completo.pdf?sequence=1