

Clusters. Programación en Clusters
Algebra Lineal en Paralelo sobre Clusters

Fernando G. Tinetti
fernando@info.unlp.edu.ar
Curso de Postgrado
Facultad de Informática, UNLP
50 y 115, 1900, La Plata
Argentina

2009

Clusters. Programación en Clusters Algebra Lineal en Paralelo sobre Clusters

UNLP - 2009

1. Mencionar los tres problemas más importantes que resolvió utilizando computadoras.
2. Identificar el contexto en el cual resolvió cada problema (producción, proyecto de investigación, proyecto de cátedra, interés personal).
3. Indique qué aprendió de cada uno de los problemas.
4. Enumere los pasos que siguió para resolver cada problema.
5. Indique si está en un proyecto de investigación actualmente e indicar cuáles son los objetivos a corto, mediano y largo plazo desde su punto de vista (en caso de estar en un proyecto).

Nombre:

Cargo/Posición:

Universidad/Lugar de Trabajo:

1 Introducción

Contenido

- Programa/Detalles del Curso publicado
 - Objetivos
 - * Conceptos de Programación distribuida.
 - * Análisis de los modelos de comunicación entre procesos y procesadores.
 - * Soluciones basadas en bibliotecas tipo MPI y PVM.
 - * Clusters homogéneos y heterogéneos.
 - * Caracterización del modelo de arquitectura de cluster.
 - * Resolución de problemas numéricos y no numéricos sobre clusters.
 - * Identificación de las características de rendimiento paralelo en general y de las específicamente relacionadas con el rendimiento paralelo en los clusters.
 - * Identificación de las ventajas de utilizar clusters en general para cómputo paralelo.
 - * Identificación de las posibles penalizaciones de rendimiento en los clusters.
 - * Identificación de las herramientas básicas para análisis de rendimiento sobre clusters.
 - Programa
 - * Unidad I. Introducción.
 - * Unidad II. Ideas de procesamiento intensivo extraídas del área de aplicaciones de algebra lineal. Análisis de problemas no numéricos.
 - * Unidad III. Arquitecturas paralelas: desde las computadoras paralelas específicas hasta los clusters.
 - * Unidad IV. Bibliotecas de cómputo numérico/matricial secuenciales y paralelas.
 - * Unidad V. Problemas y soluciones para operaciones con altos requerimientos de cómputo. Análisis de problemas y soluciones posibles.
 - * Unidad VI. Extensiones a más de un cluster ¿Más problemas que soluciones?
 - * Unidad VII. Análisis posibles trabajos finales.
 - Duración
 - * 20 horas de clase presencial.
 - * 40 horas de trabajo fuera de clase.
 - * 2 horas para exponer en forma individual los proyectos.
 - Modo de Evaluación
 - * Proyectos de trabajo individual con 3/6 meses para presentarlos.

- Temas del Curso
 - Introducción.
 - Algebra lineal: características, aplicaciones, problemas clásicos.
 - Evaluación de rendimiento secuencial y paralelo.
 - Clusters: arquitectura de procesamiento paralelo.
 - Presentación y análisis de bibliotecas de software disponibles.
 - Análisis de los problemas y soluciones posibles para el balance de carga.
 - Ejemplo de modelización: desde un problema hasta un modelo a paralelizar (tiempo...).
 - Algunas ideas de instalación y administración de clusters para procesamiento paralelo.
 - Análisis de posibles trabajos.
 - Levantando el nivel de abstracción (si pudiéramos):
 - * Soporte de sistemas operativos
 - * Bibliotecas en general
 - * Abstracciones intercluster
 - * Internet computing
 - * Grid computing
 - * ...

Schedule

	Original	<i>Current</i>
Lectures	Every Monday	Every Day, 1 Week
Practice	Written homeworks. Individual. Every Moday. Conclusions.	Some <i>Little Work</i> . Individual. Every Day?
Grading	Project. Individual.	Project. Individual.

Objectives:

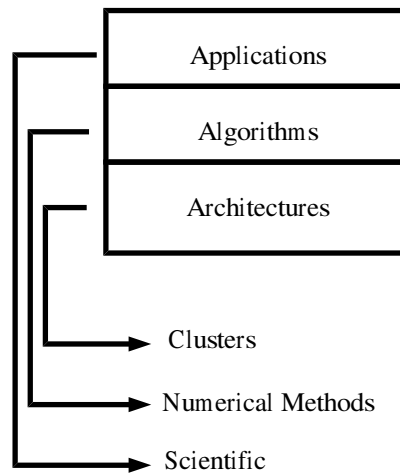
- Linear algebra.
- Parallel computing on clusters.
- Parallel algorithms analysis (*on clusters*).
- Heterogeneity and workload balance.
- Some practice with MPI.

Assumptions (should be):

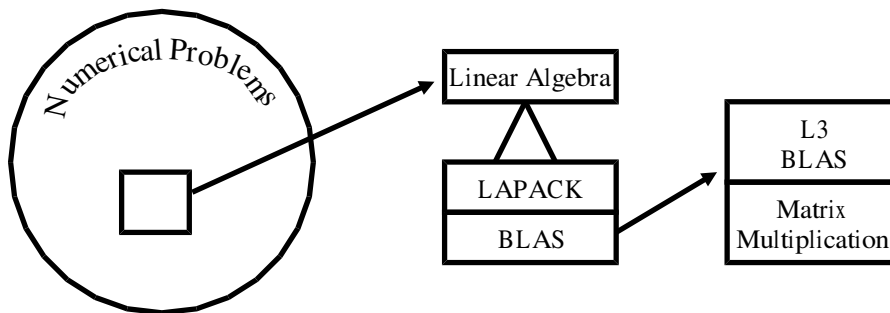
- *Minimum* algebra.
- Parallel computing.
- Parallel algorithms analysis.

Too many problems...

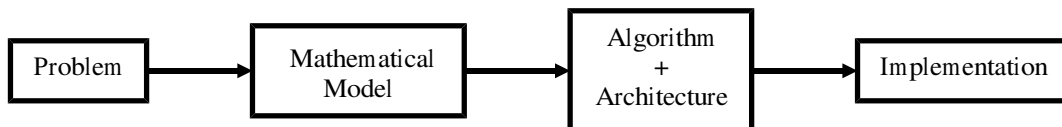
Production Software:



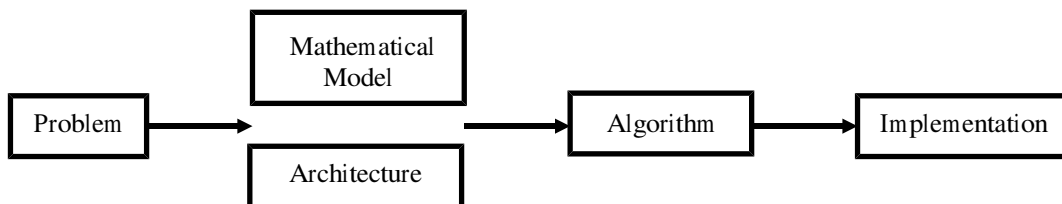
Algorithmic View:



Getting Production Code:



However, Getting Production Code (2):



Many algorithms proposed and used with strong emphasis on

- Mathematical models to real problems.
- Numerical stability, error analysis.
- Performance. Well, it is not possible to avoid this...
- Parallel approaches... (un)fortunately.

From a Call for Papers

“The use of supercomputing technology, parallel and distributed processing, and sophisticated algorithms is of major importance for computational scientists. Yet, the scientists’ goals are to solve the challenging problems, not the software engineering tasks associated with it. For that reason, computational science and engineering must be able to rely on dedicated support from program development and analysis tools. Focusing on this background, the following question must be investigated:

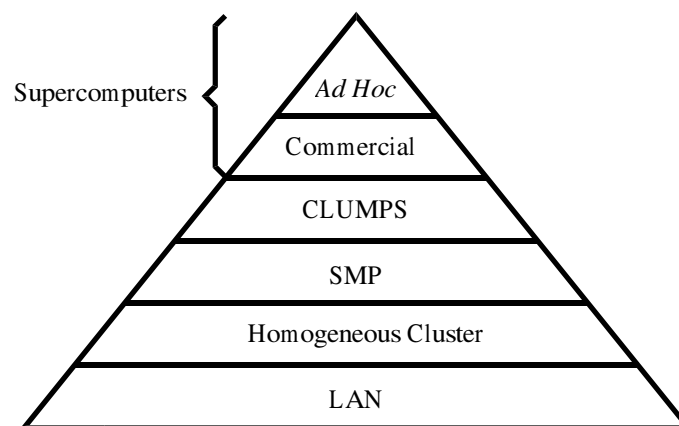
How to support users of computational science and engineering during program development and analysis?”

From the *Basic parallel cluster computing* ideas:

- Parallel algorithms are not initially proposed for clusters.
- Clusters have strong differences with parallel computers (?).
- Algorithms should be analyzed *in the context of* parallel cluster computing.

1.1 Current Computing Hardware

Just *another* classification (*to be corrected*):



Why not a new class? Clusters of Symmetric Multicores

Práctica con Multiplicación de Matrices

Evaluación de rendimiento secuencial

- a) Las tres iteraciones
- b) Las optimizaciones del compilador
- c) Qué es lo que se puede entender/explicar
- d) La idea del procesamiento por bloques
- e) El rendimiento del procesamiento por bloques (de *una* de las posibilidades)