

Parallel Programming of High Performance Computers for Mining

5º Seminario de Acercamiento Tecnológico
Codelco Digital: Minería del Futuro

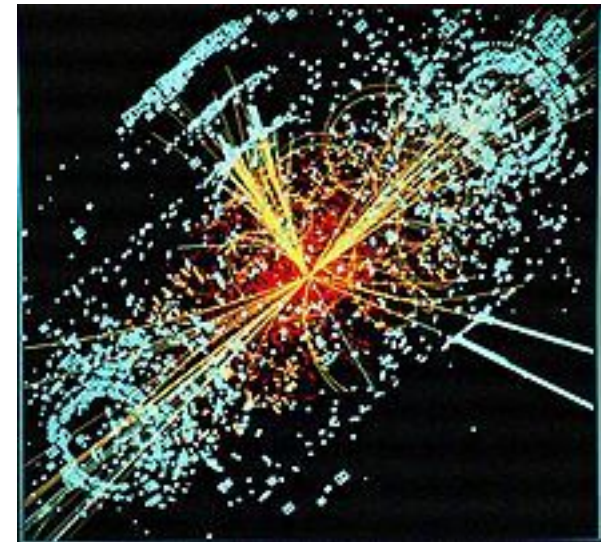
Mario Leyton¹, Nelson Morales²

¹NIC Labs, mleyton@niclabs.cl

²Delphos Mine Planning Lab, AMTC
nmorales@ing.uchile.cl

Desafíos Complejos

- Large Hadron Collider
- Datos producidos
 - 15 petabytes/year
 - 500Mb/s
 - **2 Enciclopedias Britanicas x segundo!**



Desafíos Complejos



- BOINC
- La mas grande red voluntarios de cálculo
- SETI@HOME -> Búsqueda de ET
- FightAIDS@Home -> Drogas para SIDA
- Docking@Home -> *Folding* de Proteinas
- PrimeGrid -> numeros primos gigantes
- etc....

- *Problema del Vendedor Viajero:*
 - *Recorrido mas corto que pasa por todas las ciudades*

- Alemania - 2001
- 15.112 ciudades
- 66 000 km
- 22.6 años CPU
- 110 computadores



High Performance Computing

- Tres razones para usar HPC
 - **Cálculos intensivos**
 - **Datos intensivos**
- **Restricción temporal:** “*resultados para ayer*”

Complejidades de HPC

- Programa HPC es ~10 veces más costoso que un programa normal
- ¿Por qué?
 - Redefinir los algoritmos
 - Mayor complejidad del hardware
 - Errores no determinísticos
 - Data races,
 - Deadlocks,
 - Livelocks, etc..

Complejidades de HPC

- ¿Qué hardware debo usar?
 - Comprar Clusters Ad-hoc
 - Alta inversion
 - Altos costos operacionales
 - (2 años de uso equivalen a inversion)
 - Arrendar Clusters
 - Amazon EC2: \$0.10 x hora
 - Universidades / Research Labs
 - Reutilizar Desktops
 - Voluntarios: BOINC



Complejidades de HPC

- ¿Qué herramientas debo usar?
- Algunos Modelos de Programación Paralela
 - MPI (Message Passing Interface)
 - Objetos distribuidos
 - Workflows / Web Services
 - Componentes
 - BSP
 - **Patrones de Paralelismo (esqueletos)**

Skandium: Patrones de Paralelismo

- Patrones de Paralelismo
- Biblioteca en Java
- Open Source (GPL)
- <http://skandium.niclabs.cl/>

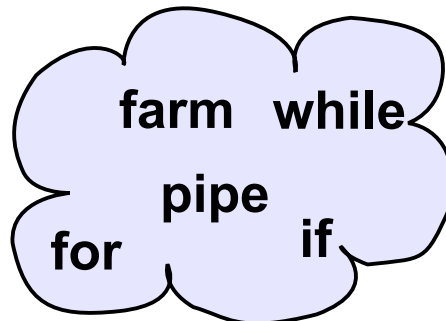


Skandium: Patrones de Paralelismo

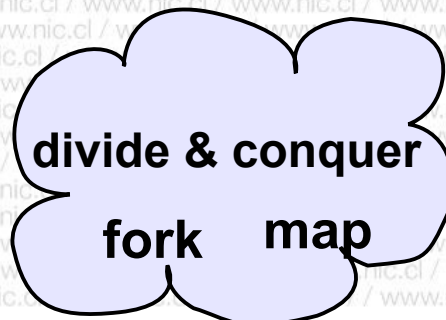
- Modelo de Programación de **Alto Nivel** [Cole89]
- Esconde la complejidad de la programación paralela
- Explota **patrones** anidables de paralelismo

Patrones de Paralelismo

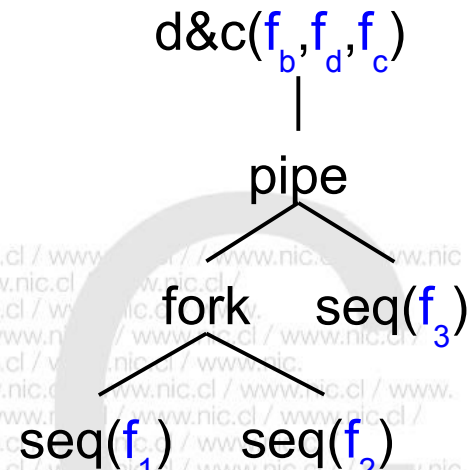
Tareas



Datos



BLAST Programa Ejemplo



Skandium: Detonación de Paralelismo

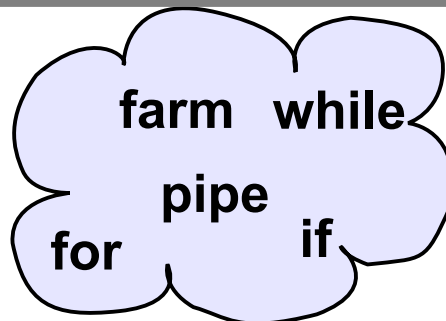
```
public boolean condition(BlastParams param)
{
    File file = param.dbFile;

    return file.length() > param.maxDBSize;
}
```

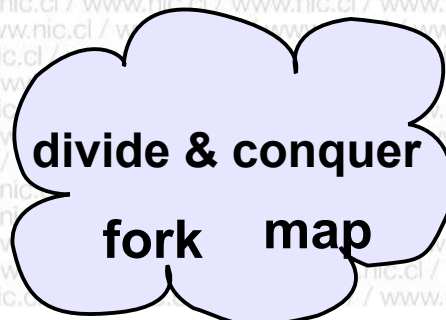
alela

T
rama Ejemplo

Tareas



Datos



d&c(f_b, f_d, f_c)

pipe

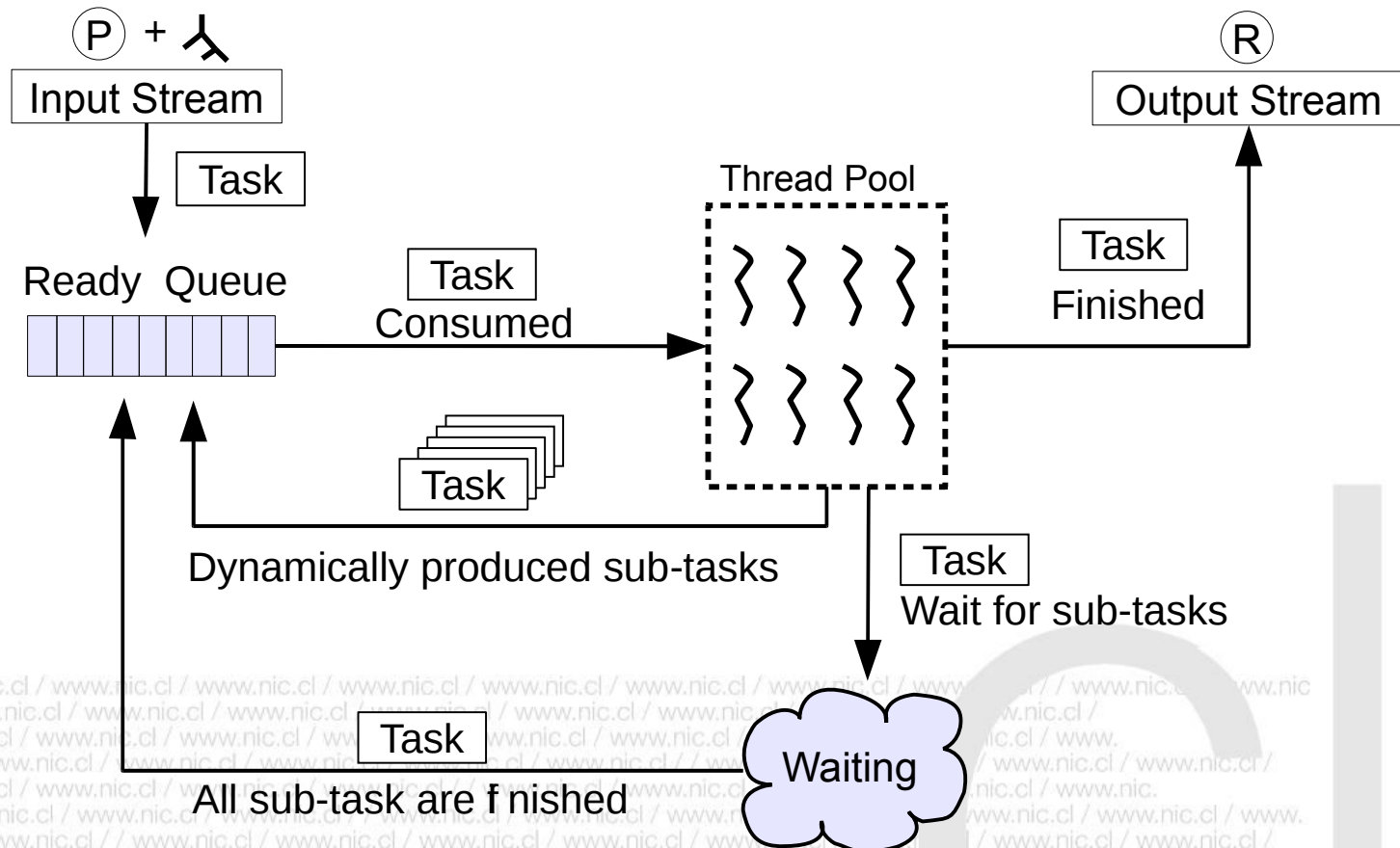
fork

seq(f_3)

seq(f_1)

seq(f_2)

Skandium: Patrones de Paralelismo

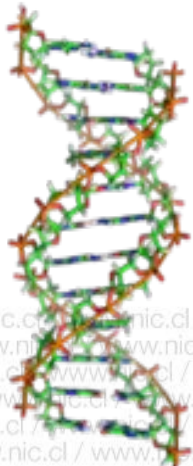


Skandium: Patrones de Paralelismo

- Resultados

- N-Reinas: Como colocar N reinas en un tablero de ajedrez

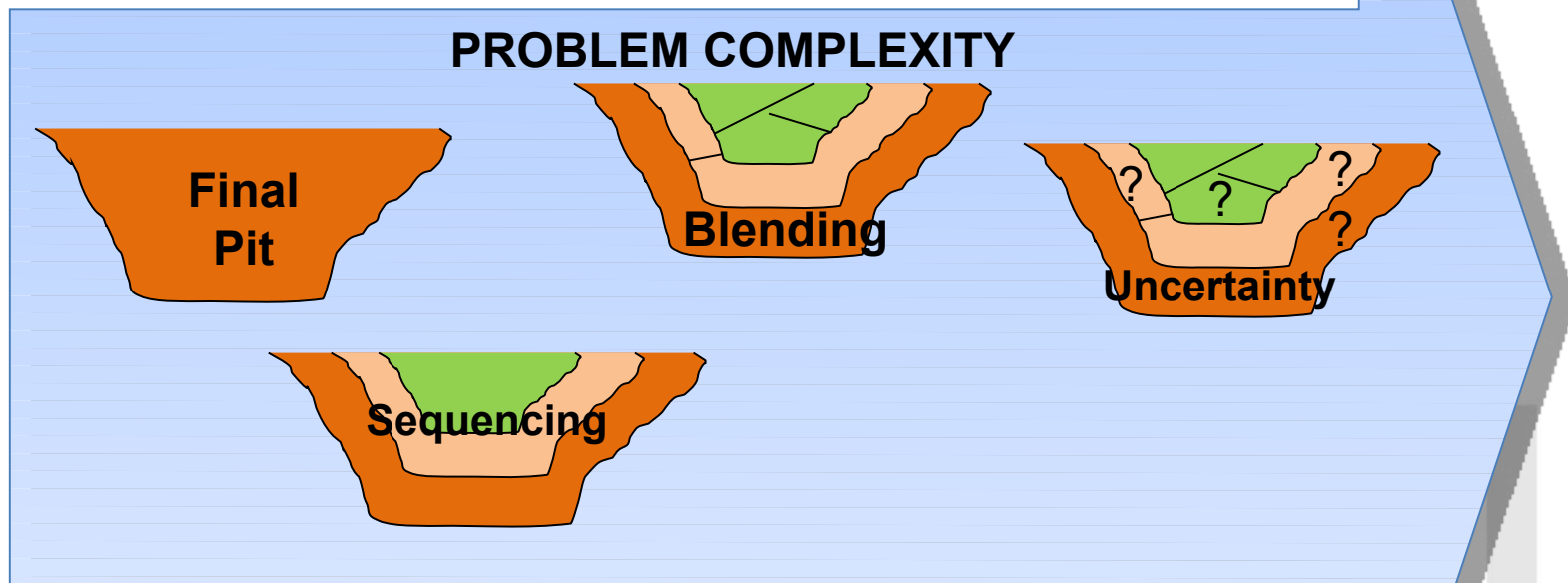
- Pi encontrar los X decimales:
 $3.1415926535897932384626433...$



- BLAST: Alineamiento de DNA en paralelo

- DNSSEC: Paralelización de firmado de zona

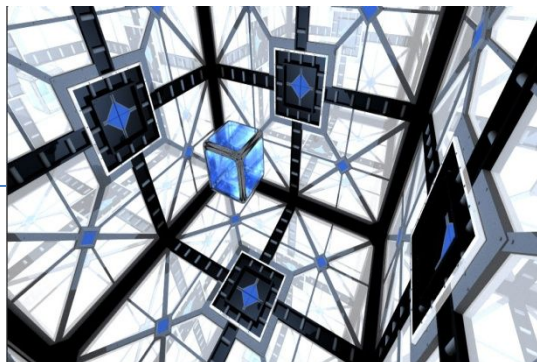
Desafíos HPC en Minería



Actualment, capacidad limitada de cálculo, obliga a dividir el problema en etapas, deasociadas entre ellas (planificación largo vs corto plazo: planificaciones tardías y sub-óptimas

Desafíos HPC en Minería

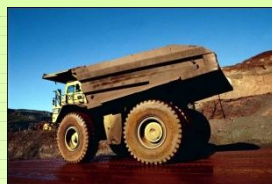
... y continuamente integrar planificación minera con resultados en-linea y la operación minera



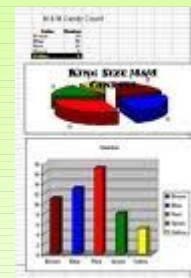
Drilling & Blasting on Demand



All-Term Planning

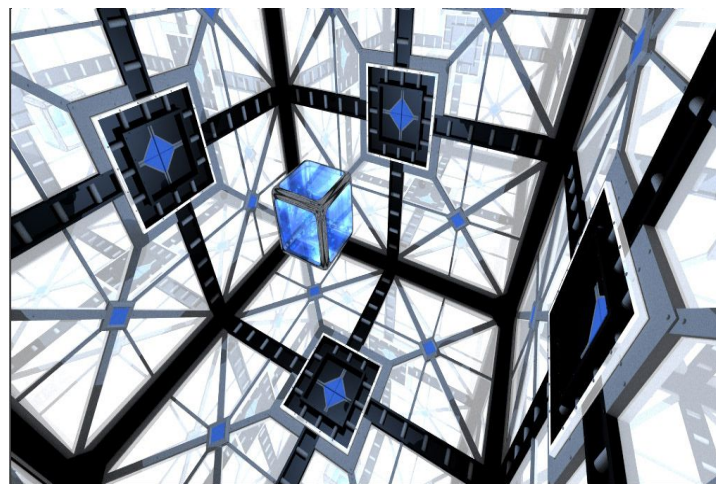


Dispatch
Routing
Maintenance
Prod. Sched.



Online Consolidation

Desafíos HPC en Minería



HPC permitirá

- resolver problemas integrados (reales)
- Obteniendo soluciones en-linea (actualmente el proceso diseminado toma semanas).

Conclusiones

- **HPC es necesario** para ciencia de frontera e industrias competitivas
- Existe una variedad de **oportunidades** donde aprovechar HPC en minería
- Múltiples **alternativas hardware**: comprar, arrendar, etc...
- Programar HPC es **costoso** y difícil
- **Es importante elegir el/los modelos de programación adecuados**

Muchas Gracias!



¿Preguntas?

<http://skandium.niclabs.cl>

mleyton@niclabs.cl

Muchas Gracias!



¿Preguntas?

<http://skandium.niclabs.cl>

mleyton@niclabs.cl