



# Intel® Cluster Ready

Simplificando la instalación y verificación de clústeres  
para Computación de Alto Desempeño

Ricardo Medel

Argentina Software Design Center

# Nota legal



INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

UNLESS OTHERWISE AGREED IN WRITING BY INTEL, THE INTEL PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE INTEL PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

This document contains information on products in the design phase of development.

All products, computer systems, dates, and figures specified are preliminary based on current expectations, and are subject to change without notice.

Intel product plans in this presentation do not constitute Intel plan of record product roadmaps. Please contact your Intel representative to obtain Intel's current plan of record product roadmaps.

Intel, VTune, Cilk, Xeon and the Intel logo are trademarks of Intel Corporation in the U.S. and other countries.

\*Other names and brands may be claimed as the property of others

Copyright© 2012 Intel Corporation. All rights reserved.

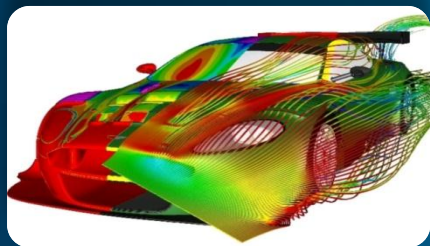
Software & Services Group, Developer Products Division

Copyright© 2012 Intel Corporation. All rights reserved. \*Other brands and names are the property of their respective owners.

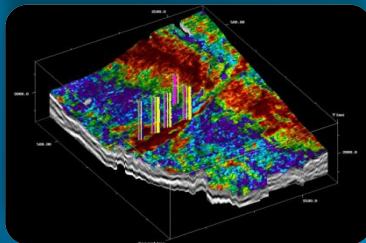
# Computación de Alto Desempeño (HPC)



Utilización de una colección de recursos computacionales cooperando para resolver problemas computacionalmente complejos, involucrando grandes cantidades de datos y cálculos.



CAD



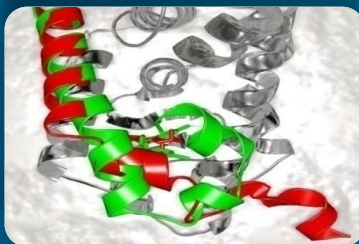
Energía



Climatología



Animación



Biotecnología



Finanzas

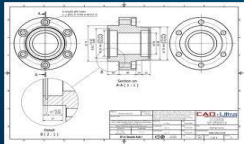
Software & Services Group, Developer Products Division

Copyright © 2012 Intel Corporation. All rights reserved. \*Other brands and names are the property of their respective owners.

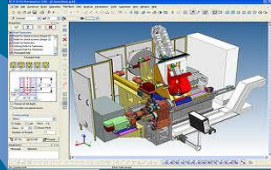
# Visión de Intel en HPC



Crear

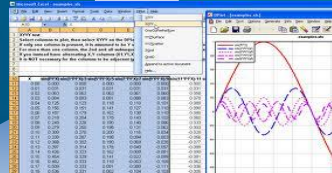


Simular



Visualizar

Analizar



- Creando e innovando con nuevas ideas
- Simulando y validando a menor costo
- Analizando e iterando a mayor velocidad
- Visualizando de maneras nunca antes posibles

**¡Transformando ideas en realidad!**

Software & Services Group, Developer Products Division

Copyright © 2012 Intel Corporation. All rights reserved. \*Other brands and names are the property of their respective owners.

# Soluciones de Intel para HPC



**¡Más potencia para ser más productivos!**

Software & Services Group, Developer Products Division

Copyright © 2012 Intel Corporation. All rights reserved. \*Other brands and names are the property of their respective owners.



# Pero hay algunos problemas...



Aplicaciones



Middleware de administración



Sistema operativo y de aprovisionamiento



Servidores/HW

Redes de conexión



- Usualmente comprado y enviado en partes
- Se necesitan expertos para armar, corregir y poner a punto
- Puede necesitar más retoques para asegurar que la aplicación deseada se ejecute

# Intel® Cluster Ready

Una arquitectura. Más oportunidades. Menor costo.



Registre aplicaciones HPC



Certifique su diseño de  
clúster HPC

Intel®  
Cluster  
Ready

Verifique su solución HPC  
utilizando Intel® Cluster Checker



Software & Services Group, Developer Products Division

Copyright © 2012 Intel Corporation. All rights reserved. \*Other brands and names are the property of their respective owners.

# ¿Qué es Intel® Cluster Ready?



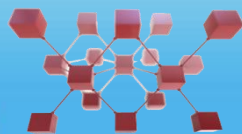
Es un **programa** estableciendo una **arquitectura estándar** para clústeres de alto desempeño basados en sistemas Intel sobre **plataformas Linux\***, que permite al usuario elegir cualquier aplicación registrada como Intel Cluster Ready, con la plena seguridad de que funcionará en cualquier clúster certificado como Intel Cluster Ready. Incluye:

- Una especificación definiendo las plataformas Intel Cluster Ready.
- Procesos de certificación de clústeres y registro de aplicaciones validando sistemas que cumplen con la especificación.
- El software Intel® Cluster Checker para validar la configuración y la funcionalidad del clúster
- Diseños de Referencia (recetas) que permiten construir clústeres siguiendo instrucciones paso a paso.
- Canales de comunicación (sitio web, foros) entre fabricantes de hardware, integradores de clústeres, desarrolladores de software y usuarios



# Simplificando el proceso de decisión

Los usuarios pueden elegir cualquiera de las aplicaciones registradas con la seguridad de que funcionará en cualquier clúster certificado *Intel Cluster Ready*



Especificación Intel  
Cluster Ready



Hardware verificado con  
Intel Cluster Checker



Soporta aplicaciones  
registradas



Elija una aplicación  
registrada  
Intel® Cluster Ready



Elija un cluster  
Intel® Cluster Ready  
que cumpla con los  
requisitos de su  
aplicación y sus  
posibilidades



¡Instalado listo  
para ejecutar!

Los diseños y los clústeres son  
verificados y validados...

...de modo que usted puede elegir, instalar,  
usar y mantener clústeres con confianza.

**Enfóquese en su trabajo, no en su clúster**

# Arquitectura Intel® Cluster Ready



Simplifica la Computación Técnica asegurando interoperabilidad mientras permite elegir opciones.

Gran cantidad de aplicaciones líderes en computación técnica

Análisis Mecánico

Dinámica de Fluidos

Simulación Climatológica

...muchas más

Diferentes opciones de middleware y software de base que funcionan correctamente juntos

## ARQUITECTURA INTEL® CLUSTER READY

Middleware de Admin. de Recursos (opcional)

LSF\*, PBSPro\*, Grid Engine\*, Torque\*, SLURM\*, Nagios\*, Ganglia\*, ...

Sist. Aprovisionamiento

PCM\*, BCM\*, Warewulf\*, Rocks\*, ...

Distribución Linux\*

RedHat Enterprise Linux\*, SUSE Linux Enterprise Solution\*, ...

Protocolos de Admin. de Redes y Clústeres

OFED\*, TCP/IP, ...

Redes de interconexión

1Gb - 10Gb Ethernet, SDR - FDR InfiniBand\*, ...

Plataformas basadas en procesadores Intel® Xeon®

Solución Certificada

Amplia selección de redes y almacenamiento compatibles

La arquitectura de servidores líder del mercado

Software & Services Group, Developer Products Division

Copyright © 2012 Intel Corporation. All rights reserved. \*Other brands and names are the property of their respective owners.

# ¿Cómo se obtiene el sello Intel® Cluster Ready?

Certificando un sistema como Intel Cluster Ready



## ¿Qué significa certificar una solución?

Certificar significa diseñar una solución que cumpla con la especificación Intel Cluster Ready y que sea repetible. También significa que se pueden ejecutar múltiples aplicaciones registradas en el mismo clúster certificado.

### Proceso de Certificación de Sistemas Intel® Cluster Ready

Diseñar un clúster  
basado en la  
especificación

Certificar que el  
clúster cumple con  
la especificación

Integrar Intel  
Cluster Ready en el  
proceso de  
manufactura

Preparar equipo de  
mkt & ventas para  
Intel Cluster Ready



Software & Services Group, Developer Products Division

Copyright © 2012 Intel Corporation. All rights reserved. \*Other brands and names are the property of their respective owners.

# Intel® Cluster Checker

Validando su sistema en todo momento...



**Paso 1 – Entrada**  
Toma el archivo de configuración del cluster



**Paso 2 – Verificación**  
Ejecuta un conjunto de tests en todo el cluster



**Paso 3 – Resultados**  
Devuelve resultado de pasó/falló con información de diagnóstico

**Paso 4 – Corridas regulares**  
Permite asegurar la salud del cluster durante su vida útil. Avisa de potenciales problemas, permite mantenimiento preventivo, etc.



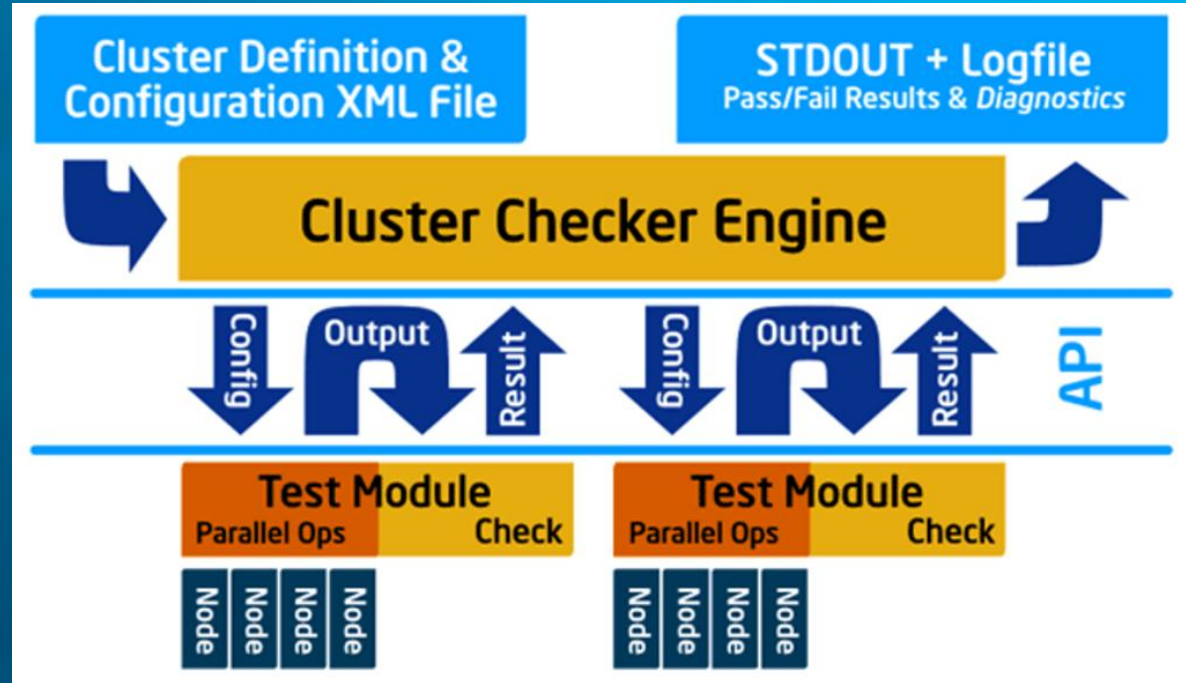
# Arquitectura de Intel Cluster Checker



Configurable

Diseño modular

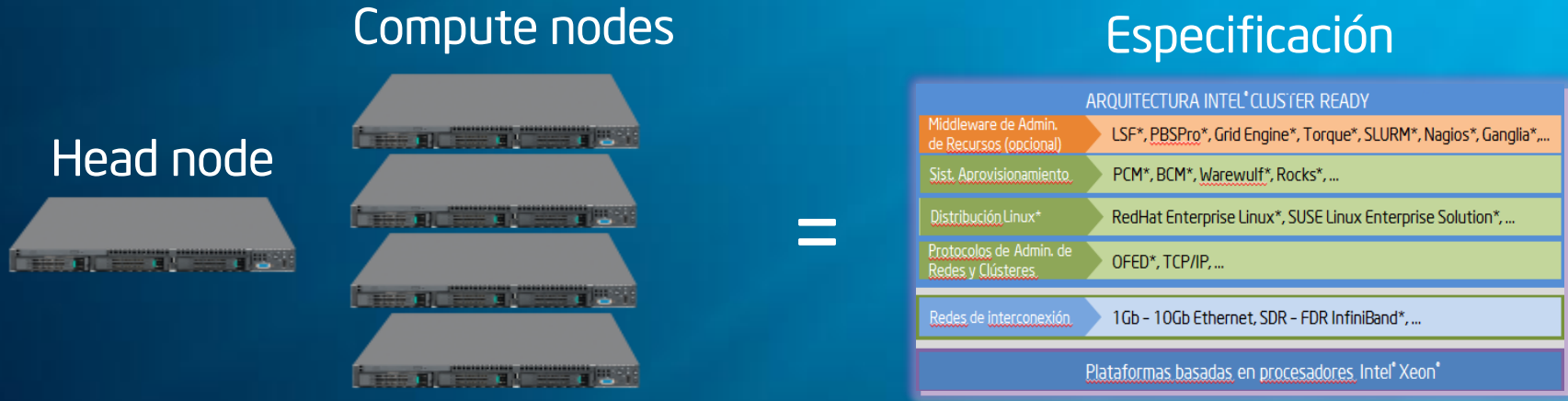
Informació de diagnòstic



# Verificación de Arquitectura



El modo compliance verifica si el cluster cumple con la especificación de la arquitectura Intel Cluster Ready

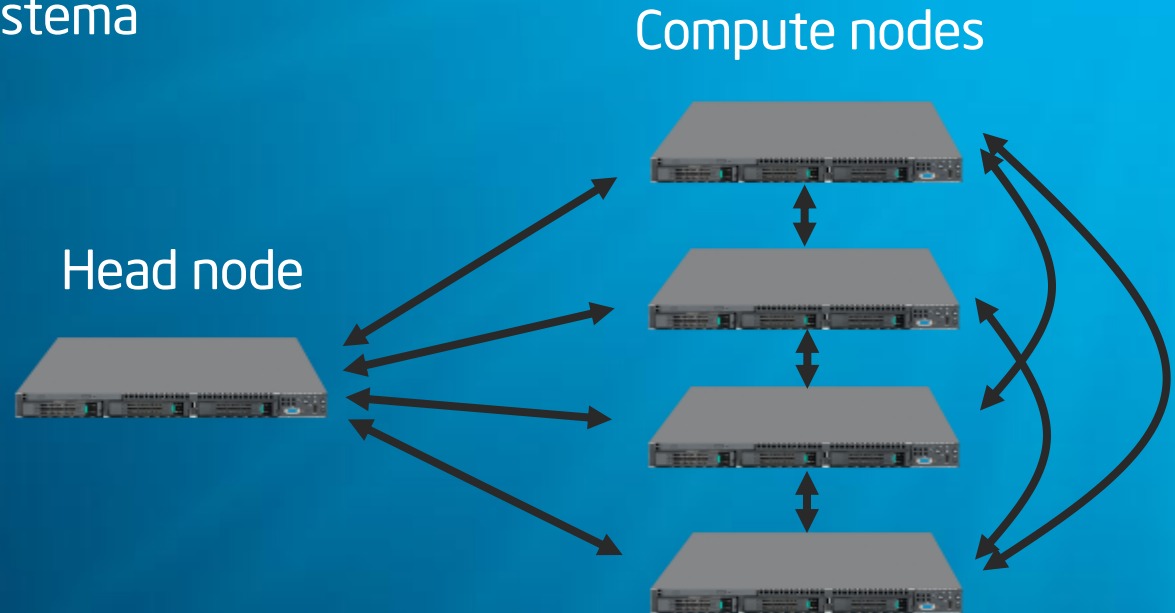


# Metodologías de Testeo



Las distintas metodologías de testeo aseguran una total cobertura del sistema

- Funcionalidad todos con todos
- Consistencia del sistema
- Benchmarks de desempeño



# Metodologías de Testeo



Las distintas metodologías de testeo aseguran una total cobertura del sistema

- Funcionalidad todos con todos

Head node



=/+

- Consistencia del sistema

Compute nodes



=



=



=



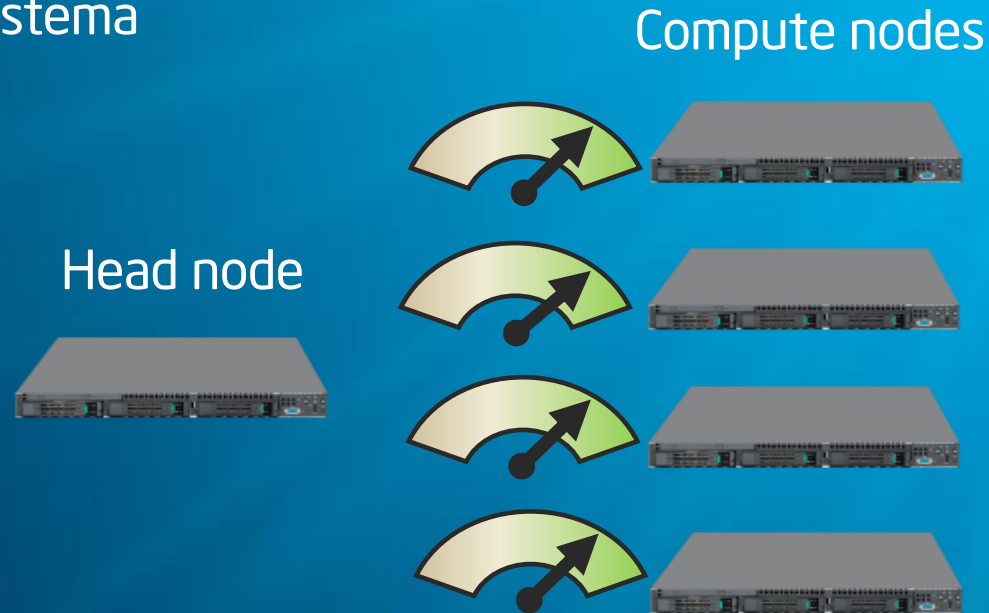


# Metodologías de Testeo



Las distintas metodologías de testeo aseguran una total cobertura del sistema

- Funcionalidad todos con todos
- Consistencia del sistema
- Benchmarks de desempeño



# Resultado de Intel® Cluster Checker



```
System V Interprocess Communication, (clean_ipc).....Succeeded
gettimeofday() clock granularity, (clock_granularity).....Succeeded
Clock Synchronization, (clock_sync).....Succeeded
Minimum cluster size for ICR 1.2, (cluster_size).....Succeeded
Core Count (Multi-core & Intel(R) Hyper-Threading Technology), (core_count).....Succeeded
Core frequency uniformity, (core_frequency).....Succeeded
/proc/cpuinfo Uniformity, (cpuinfo).....Succeeded
C Shell, (csh).....Succeeded
Single-node Disk Bandwidth, (disk_bandwidth).....Succeeded
Uniform environment variables, (environment).....Succeeded
File Existence, Ownership, and Permissions, (file_permissions).....Succeeded
Korn Shell, (ksh).....Succeeded
Bourne Shell, (sh).....Succeeded
Enhanced C Shell, (tcsh).....Succeeded
Intel(R) C++ Compiler runtime libraries, (intel_cce_rtl).....Succeeded
Intel(R) Fortran Compiler runtime libraries, (intel_fce_rtl).....Succeeded
Intel(R) MPI Library Runtime Environment (Single-node), (intel_mpi_rt).....Succeeded
Intel(R) MPI Library Runtime Environment (All nodes), (intel_mpi_rt_internode).....Succeeded
Linux Kernel Runtime Parameters, (kernel_parameters).....Succeeded
Single-node Memory Bandwidth (STREAM), (memory_bandwidth_stream).....Succeeded
Single-node floating point performance (Intel(R) Math Kernel Library),
(mflops_intel_mkl).....Succeeded
Installed Packages, (packages).....Succeeded
PCI Device Consistency, (pci).....Succeeded
HPC Challenge Benchmark (Intel(R) C++ Compiler, Intel(R) MPI Library, Intel(R) Math
Kernel Library), (hpcc)
Attention: this check may take a long time to complete.....Succeeded
MPI Collectives (Intel(R) MPI Benchmarks; Intel(R) MPI Library),
(imb_collective_intel_mpi).....Succeeded
MPI Message Integrity Check (Intel(R) MPI Benchmarks; Intel(R) MPI Library),
(imb_message_integrity_intel_mpi).....Succeeded
Network Performance (Intel(R) MPI Benchmarks; Intel(R) MPI Library),
(imb_pingpong_intel_mpi).....Succeeded
```

Check has **Succeeded**.

Total Elapsed Time: 0:05:15

[182] icr@jerry [~]

>

# Formas de aislar problemas



1. Ejecute un único módulo

```
>$cluster-check --include-only clock_frequency
```

2. Incremente la verbosidad

```
>$cluster-check --include-only clock_frequency --verbose 5
```

3. Ejecute el módulo en modo “debug”

```
<cpuinfo>  
  <debug/>  
</cpuinfo>
```

4. Ejecute sobre menos nodos

```
>$cluster-check --nodefile custom_nodefile.txt
```

5. Setee el ambiente para expandir los resultados obtenidos

```
<imb_collective_intel_mpi>  
  <env>export I_MPI_DEBUG=50<env>  
  <fabric>  
    <device>rdssm</device>  
  </fabric>  
</imb_collective_intel_mpi>
```

# Usando <debug> para indentificar un problema



```
Network Performance (Intel(R) MPI Benchmarks; Intel(R) MPI
Library), (imb_pingpong_intel_mpi).....FAILED
[ERROR]
subtest 'bandwidth (I_MPI_DEVICE = rdssm)' failed
- failing All hosts returned: 'undefined value'
subtest 'latency (I_MPI_DEVICE = rdssm)' failed
- failing All hosts returned: 'undefined value'
```

- El módulo  
imb\_pingpong\_intel\_mpi  
falla con "undefined value"

- La salida verbosa no ayuda a  
identificar el problema

```
Library), (imb_pingpong_intel_mpi, s6-rev31005-b823).....FAILED
[ERROR]
subtest 'bandwidth (I_MPI_DEVICE = rdssm)' failed
- failing hosts compute-0-1 -> compute-0-0, compute-0-1 -> compute-0-3,  
compute-0-3 -> compute-0-0, compute-0-2 -> compute-0-0, compute-0-2 ->  
compute-0-1, compute-0-2 -> compute-0-3 returned: 'undefined value'
subtest 'latency (I_MPI_DEVICE = rdssm)' failed
- failing hosts compute-0-1 -> compute-0-0, compute-0-1 -> compute-0-3,  
compute-0-3 -> compute-0-0, compute-0-2 -> compute-0-0, compute-0-2 ->  
compute-0-1, compute-0-2 -> compute-0-3 returned: 'undefined value'
```

```
output:
DAT: library load failure: libdaplcm.a.so.1: cannot open shared object file: No s
uch file or directory
DAT: library load failure: libdaplcm.a.so.1: cannot open shared object file: No s
uch file or directory
DAT: library load failure: libdaplfo.a.so.2: cannot open shared object file: No s
uch file or directory
DAT: library load failure: libdaplfo.a.so.2: cannot open shared object file: No s
uch file or directory
rank 0 in job 1 compute-0-3_56281 caused collective abort of all ranks
exit status of rank 0: return code 254
```

- La fuente del problema es  
rápidamente identificada en el  
archivo "debug"



# Intel Cluster Checker para mantenimiento



- Valida que el sistema continúa cumpliendo con la especificación
- Provee información de diagnóstico detallada
- Más de 100 tests, con la posibilidad de agregar sus propios tests



- Ayuda a diferenciar problemas del sistema de los problemas de la aplicación
- Muestra la salud del sistema
- Reduce el tiempo de resolución de problemas y minimiza el personal necesario para soporte

# Intel® Cluster Ready en resumen



- Arquitectura estándar para clústeres de HPC, con recetas para construir clústeres y una herramienta para verificar su correcto funcionamiento
- El usuario elige el hardware, el software y la aplicación de entre un gran número de socios del programa Intel® Cluster Ready
- El usuario sabe que el clúster funcionará al ser instalado
  - Independientemente del proveedor del clúster
  - Independientemente del tamaño del clúster
- El usuario puede verificar el estado del clúster
  - Al ser instalado
  - Durante su vida útil, regularmente para controlar el estado del cluster
- O más simplemente: Intel® Cluster Ready – hacemos su vida más fácil!

Intel®  
Cluster  
Ready



<http://www.intel.com/go/cluster>

Software & Services Group, Developer Products Division

Copyright © 2012 Intel Corporation. All rights reserved. \*Other brands and names are the property of their respective owners.

# Argentina Systems & Tools Team



- Ingeniería de clusters

- Especificación de plataforma
- Implementaciones de referencia (recetas)
- Certificación & Registración

- Herramienta de soporte

- Intel® Cluster Checker versión 2.0



- Desarrollo del ecosistema regional

- Simposio HPC LatAm (desde 2008)
- Charlas/Seminarios
- Intel Academic Community
- Cluster disponible para I+D via Internet

- Recursos

- 9 ingenieros, 2 pasantes
- ~500 nodos + 1 cluster portable

Software & Services Group, Developer Products Division

Copyright © 2012 Intel Corporation. All rights reserved. \*Other brands and names are the property of their respective owners.



Intel®  
Cluster  
Ready



Software & Services Group, Developer Products Division

Copyright © 2012, Intel Corporation. All rights reserved. \*Other brands and names are the property of their respective owners.



# Optimization Notice



## Optimization Notice

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Notice revision #20110804

# Socios del programa

Intel®  
Cluster  
Ready



Developers

Software & Services Group, Developer Products Division

Copyright © 2012, Intel Corporation. All rights reserved. \*Other brands and names are the property of their respective owners.

Rock your code.

# Proyectos de Pasantías



- Generador paralelo de esqueletos de proteínas
  - Convenio con CONICET y FuDePAN para el desarrollo de código paralelo (C++) para generar un repositorio de esqueletos de proteínas. Liberado con licencia GPL.
- API multi-lenguaje para Cluster Checker
- Intel Cluster Ready LiveCD

