UNIVERSIDAD DE BURGOS

# Microsoft HPC

V 1.0

José M. Cámara

(checam@ubu.es)

UNIVERSIDAD DE BURGOS

# Introducción

Microsoft HPC (High Performance Computing) es la solución de Microsoft a la computación de alto rendimiento.

Está enfocado principalmente a clusters de consumo.

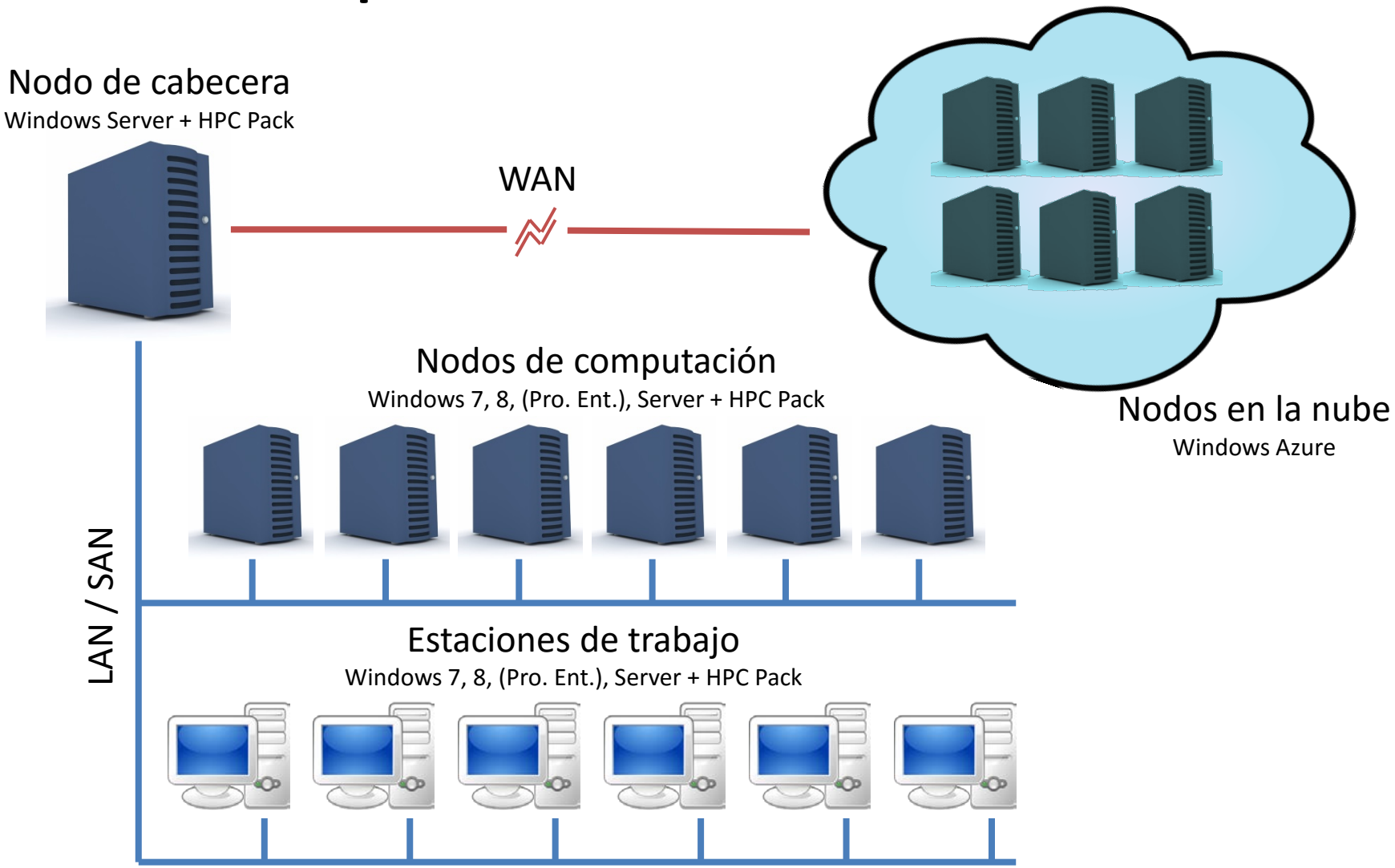
Se pretende obtener el máximo rendimiento de los recursos existentes para minimizar la inversión.

El sistema es altamente escalable debido a la incorporación de recursos internos y en la nube.

El rendimiento del conjunta se ve afectado por la inversión realizada en los distintos subsistemas: hardware de los nodos e interconexión.

UNIVERSIDAD DE BURGOS

# Arquitectura del cluster



UNIVERSIDAD DE BURGOS

# Arquitectura del cluster II

## Tipos de nodos

- Nodo de cabecera: no tiene especiales características hardware (doble tarjeta de red).
- Nodo de computación: dedicado a HPC.
- Estación de trabajo: computador de propósito general que ocasionalmente se une al cluster para apoyo a la computación. Se pueden incorporar manualmente o de acuerdo con un horario preconfigurado.
- Nodo en la nube: computador virtual contratado con Microsoft.

## Sistema operativo

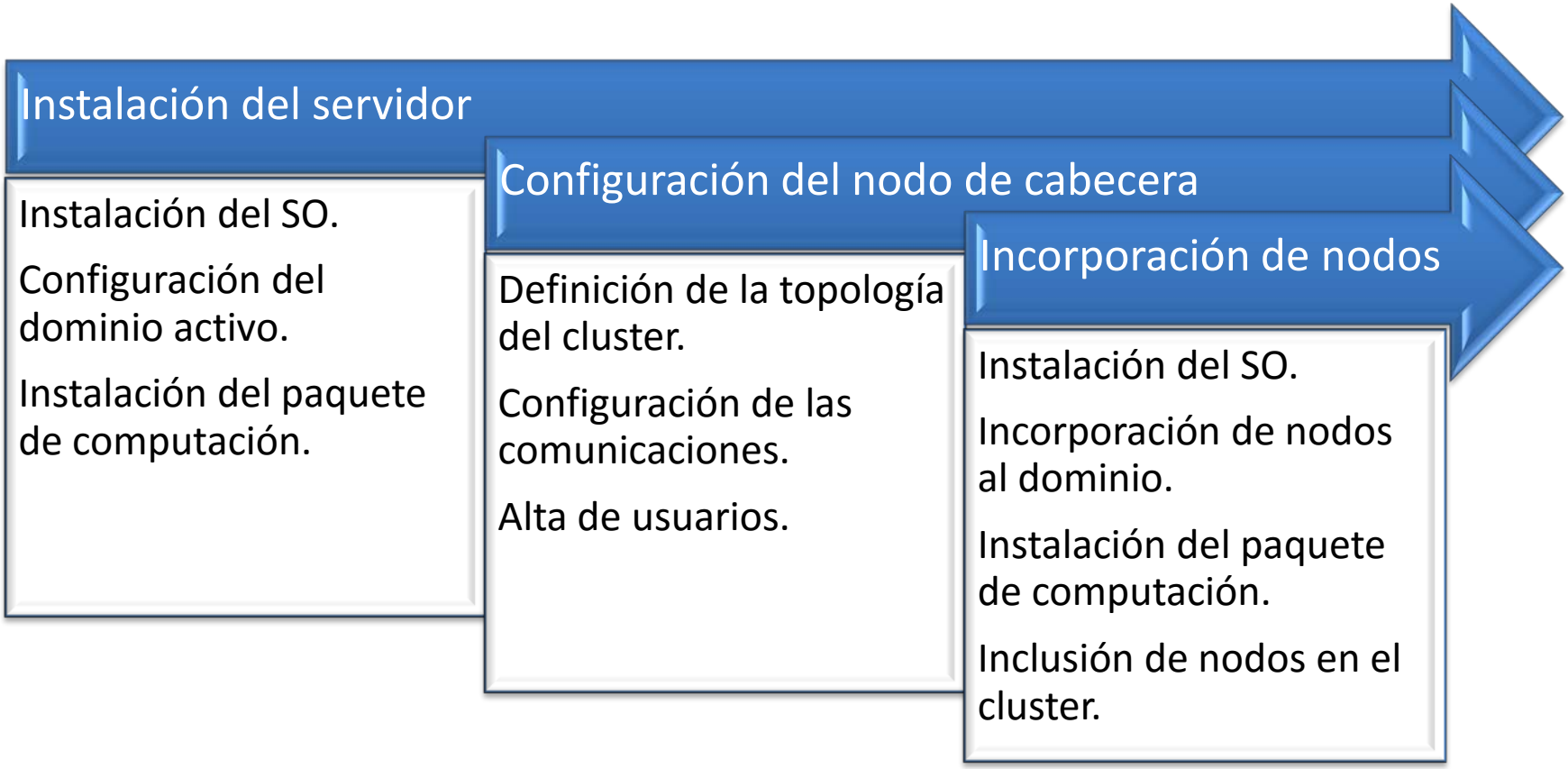
- El nodo de cabecera debe está equipado con Windows Server
- El resto de nodos admiten la versión de servidor de Windows pero también sistemas operativos convencionales, habitualmente restringidos a versiones: profesional, enterprise o ultimate.
- Los nodos en la nube se integran bajo Windows Azure.

## Arquitectura de red

- La conexión a los recursos en la nube y/o de gestión remota se realiza a través de una conexión WAN convencional. Obviamente un elevado ancho de banda es deseable.
- La conexión a los nodos internos se realiza mediante redes LAN o SAN; la mayoría de las soluciones comerciales son integrables. Los nodos pueden estar interconectados mediante redes privadas, red corporativa o ambas.

UNIVERSIDAD DE BURGOS

# Despliegue del cluster



UNIVERSIDAD DE BURGOS

# Dominio activo

- Para que todas las operaciones a realizar en el cluster se puedan llevar a cabo de manera segura, todos los nodos deben pertenecer a un dominio común o a dominios con relaciones de confianza establecidas.
- El dominio se configura como un nuevo bosque y habitualmente el servidor es promovido a controlador del dominio; a no ser que se emplee otro servidor para esta función.
- Todos los nodos y usuarios deben pertenecer al dominio para poder ser añadidos al cluster.

UNIVERSIDAD DE BURGOS

# Paquete de computación

El paquete de computación (HPC Pack) es proporcionado por Microsoft de forma gratuita.

El mismo paquete permite configurar el nodo de cabecera y las estaciones de trabajo.

Al seleccionar el tipo de nodo a incorporar, se determinan las herramientas necesarias.

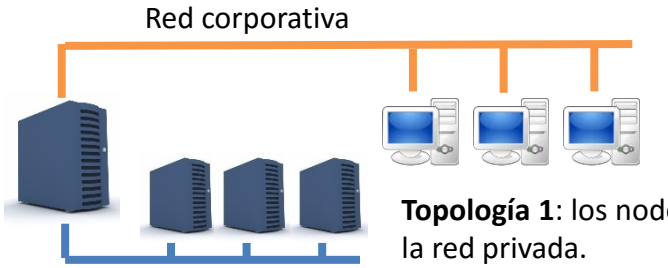
Tanto en el nodo de cabecera como en el resto, deja instalado un grupo de herramientas para gestión del cluster.

UNIVERSIDAD DE BURGOS

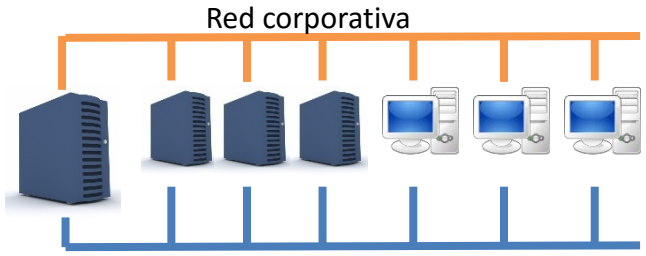
# Topologías

Microsoft utiliza este concepto para definir una serie de alternativas de interconexión. No tiene relación con el grafo de la red.

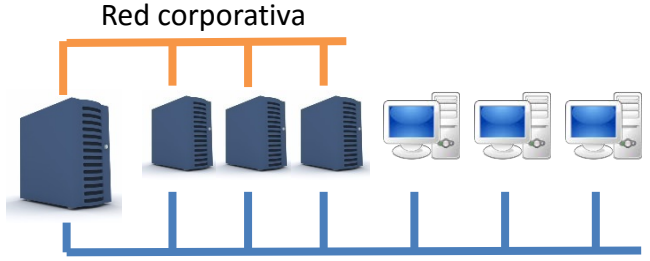

**Topología 1**: los nodos de computación y las estaciones de trabajo se conectan solo a la red privada.

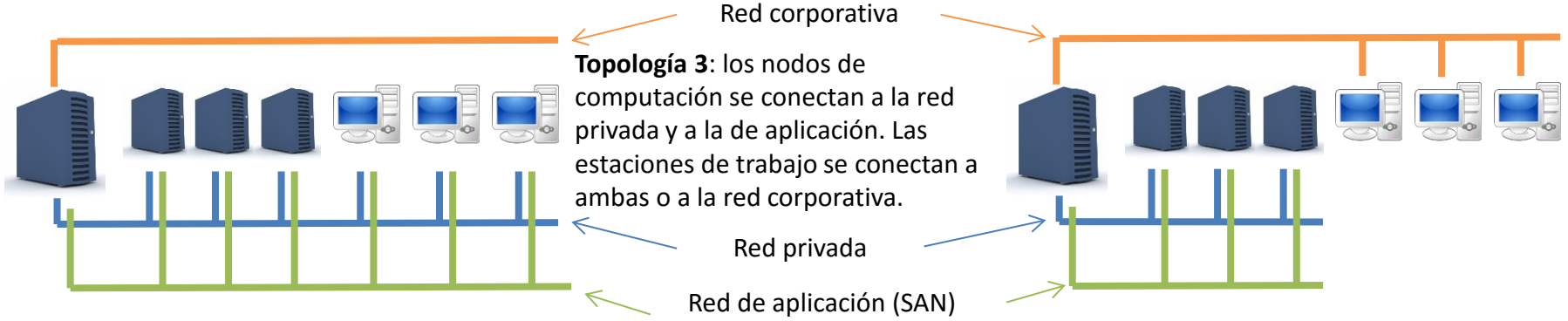



**Topología 1**: los nodos de computación se conectan a la red privada.
Las estaciones de trabajo se conectan a la red corporativa.




**Topología 2**: los nodos de computación se conectan a la red privada y corporativa. Las estaciones de trabajo a ambas o solo a la privada.

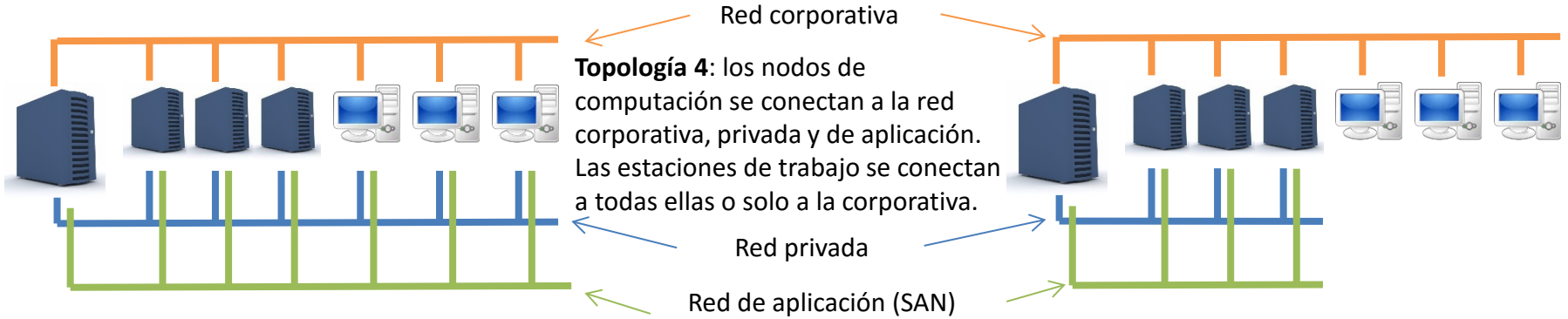


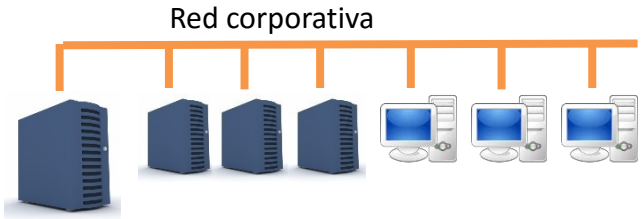

**Topología 3**: los nodos de computación se conectan a la red privada y a la de aplicación. Las estaciones de trabajo se conectan a ambas o a la red corporativa.

UNIVERSIDAD DE BURGOS

# Topologías II




**Topología 4**: los nodos de computación se conectan a la red corporativa, privada y de aplicación. Las estaciones de trabajo se conectan a todas ellas o solo a la corporativa.




**Topología 5**: no hay red privada ni de aplicación. Todos los nodos se conectan a la red corporativa.

El usuario no está obligado a ajustarse a la configuración real de la red. Puede existir una conexión a la red corporativa pero no ser interesante que las estaciones de trabajo la utilicen.

UNIVERSIDAD DE BURGOS

# Gestión del cluster

El administrador del sistema debe realizar una serie de tareas de configuración y supervisión del cluster.

El Gestor del Cluster (Cluster Manager) es la herramienta que le permite realizar estas operaciones:

- Gestión de nodos.
- Gestión de usaurios.
- Configuración del planificador.

UNIVERSIDAD DE BURGOS

# Gestión de nodos

- Dar de alta/baja nodos.
- Traer online/offline estaciones de trabajo: en el caso de que se haya optado por realizar esta tarea de forma manual y no de acuerdo a un horario/calendario.
- Formar grupos de nodos para realizar una gestión conjunta. Por ejemplo definir que determinados tipos de trabajos recaigan sobre grupos concretos.
- Monitorizar los nodos: “estado de salud”, propiedades, red, métricas…

UNIVERSIDAD DE BURGOS

# Gestión de usuarios

Dar de alta/baja nuevos usuarios.

Gestionar grupos de usuario.

Asignar roles de usuarios:

- **Usuario**: puede administrar sus propios trabajos.
- **Administrador**: puede administrar trabajos y recursos.
- **Administrador de trabajo**: puede administrar trabajos , pero no recursos.
- **Operador de trabajo**: puede administrar trabajos de una manera restringida (ver, cancelar, finalizar, re-encolar).

UNIVERSIDAD DE BURGOS

# Gestión del planificador

El planificador decide qué trabajos lanzar en cada momento y contra qué recursos.

Lo hace de acuerdo a una serie de políticas:

- Encolada.
- Balanceada.

Permite la utilización de “backfilling”: adelantar trabajos sobre uno más prioritario que está en espera de obtener los recursos que necesita, siempre que éste no se vea retrasado por ello.

UNIVERSIDAD DE BURGOS

# Políticas de gestión de trabajos

## Subpolíticas

**Precedencia:** trabajos más prioritarios pueden detraer recursos de trabajos menos prioritarios ya en curso.

- **Elegante:** detraer recursos de tareas que terminan.
- **Inmediata:** cancelar todas las tareas en ejecución.
- A nivel de tarea: cancelar tareas individualmente.

**Asignación dinámica de recursos:** los recursos asignados a un trabajo se pueden alterar en tiempo de ejecución.

- **Incremento automático:** priorizar la asignación de nuevos recursos sobre el inicio de trabajos menos prioritarios.
- **Decremento automático:** detraer recursos no usados por trabajos que no tienen tareas pendientes.

## Políticas

### Encolada

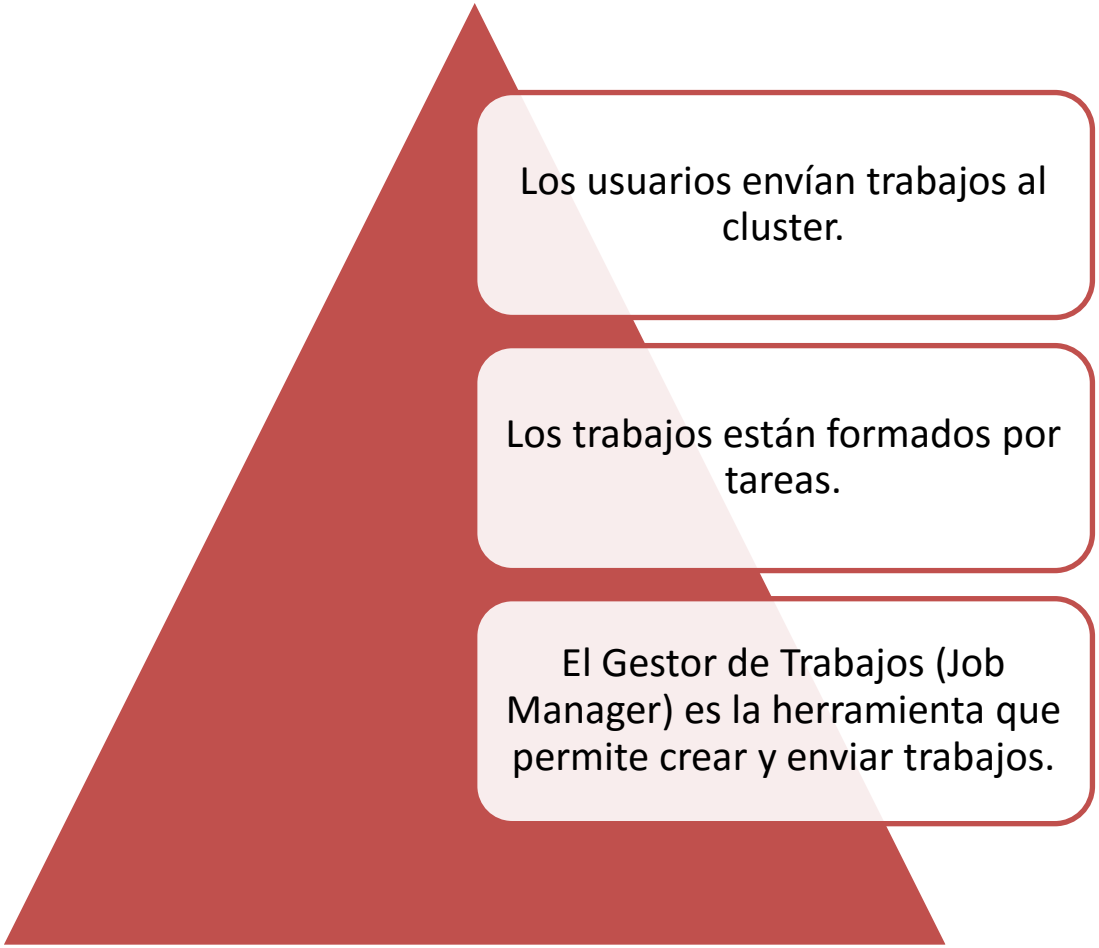
- Intenta arrancar los trabajos en orden de llegada.
- Optimiza trabajos voluminosos.
- Precedencia elegante por defecto.
- Habilitados incremento y decremento automáticos por defecto.

### Balanceada

- Inicia los trabajos tan pronto como disponen del mínimo de recursos solicitados.
- Nuevos recursos disponibles son asignados a los trabajos en curso de acuerdo con su prioridad.
- Optimiza tareas cortas e interactivas.
- Precedencia inmediata por defecto.

UNIVERSIDAD DE BURGOS

# Gestor de Trabajos



UNIVERSIDAD DE BURGOS

# Tipos de tareas

| | |
|---|---|
| Básica | • Ejecuta una única instancia de una aplicación serie o paralela. |
| Barrido paramétrico | • Ejecuta múltiples instancias de una aplicación.<br>• El número exacto lo determina un parámetro que toma valores diferentes.<br>• El valor del parámetro tiene un significado para cada instancia. |
| Preparación de nodo | • Un comando o “script” para ser ejecutado en cada nodo.<br>• Se ejecuta antes que cualquier otra tarea del trabajo. |
| Liberación de nodo | • Un comando o “script” para ser ejecutado en cada nodo.<br>• Se ejecuta cuando el nodo es desasignado al trabajo. |
| Servicio | • Ejecuta un comando o servicio en cada recurso asignado al trabajo.<br>• Si una instancia del comando finaliza y el recurso sigue asignado, una nueva instancia se inicia. |

UNIVERSIDAD DE BURGOS

# Tipos de trabajos

| | |
|---|---|
| Trabajo | • Es el tipo más general.<br>• Los otros dos pueden ser definidos a partir de él también.<br>• Estos trabajos pueden incluir múltiples tareas y barridos. |
| Single-task job | • Pensado para facilitar la configuración de trabajos simples.<br>• Se utiliza para trabajos de una única tarea. |
| Parametric sweep job | • Pensado para facilitar la configuración de trabajos simples.<br>• Se utiliza para trabajos con una única tarea paramétrica. |

UNIVERSIDAD DE BURGOS

# Propiedades de los trabajos

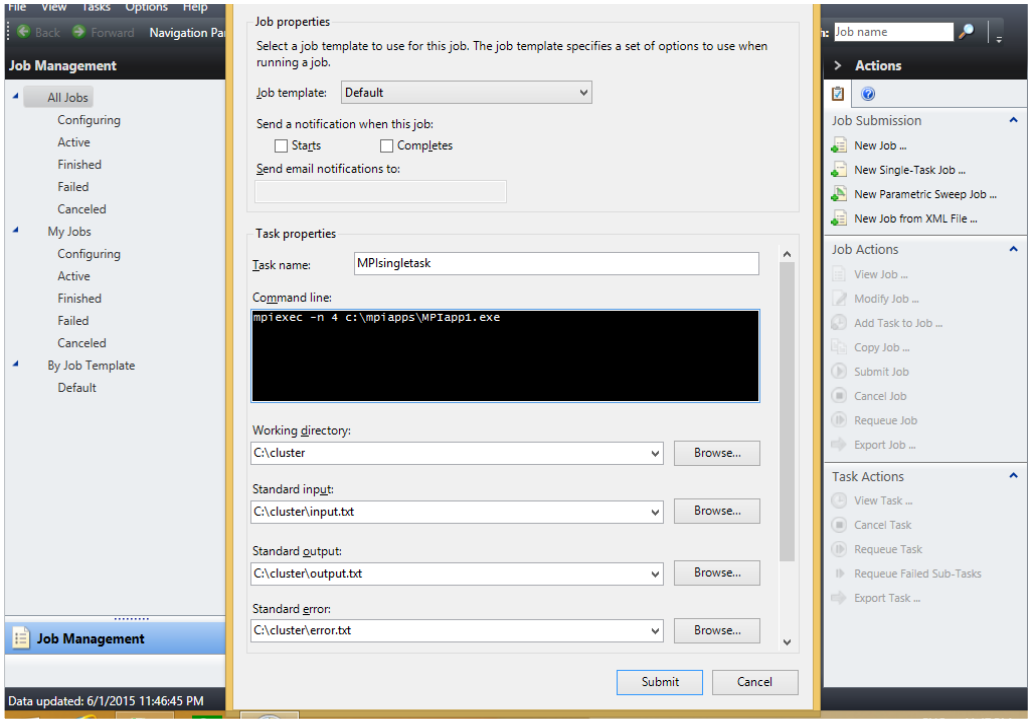
| | |
|---|---|
| ID del trabajo | • Identificación numérica del trabajo.<br>• Asignada por el Job Manager. |
| Nombre del trabajo | • Identificación textual del trabajo.<br>• Asignada por el usuario. |
| Plantilla de trabajo | • Nombre de la plantilla bajo la que se quiere configurar el trabajo.<br>• Las plantillas establecen valores y restricciones por defecto y son creadas por el administrador. |
| Prioridad | • Valor numérico que establece la prioridad del trabajo.<br>• Varía de 0 a 4000, siendo 4000 la prioridad más alta. |
| Tiempo de ejecución | • Tiempo máximo para completar el trabajo.<br>• Si se supera el trabajo es cancelado. |
| Memoria | • Mínima cantidad de memoria necesaria para que un nodo pueda ser asignado. |
| Licencias | • Licencias solicitadas por el trabajo. |

Aparte del “nombre”, no es obligatorio establecer el resto de propiedades.

Hay varias propiedades más. Para una lista completa consultar:
https://technet.microsoft.com/es-es/library/ff919649

UNIVERSIDAD DE BURGOS

# Envío y monitorización.




Envío de trabajos desde el Job Manager.

Los trabajos pueden ser creados y enviados desde un nodo a través del Job Manager.

Elige el tipo de trabajo y rellena los campos.

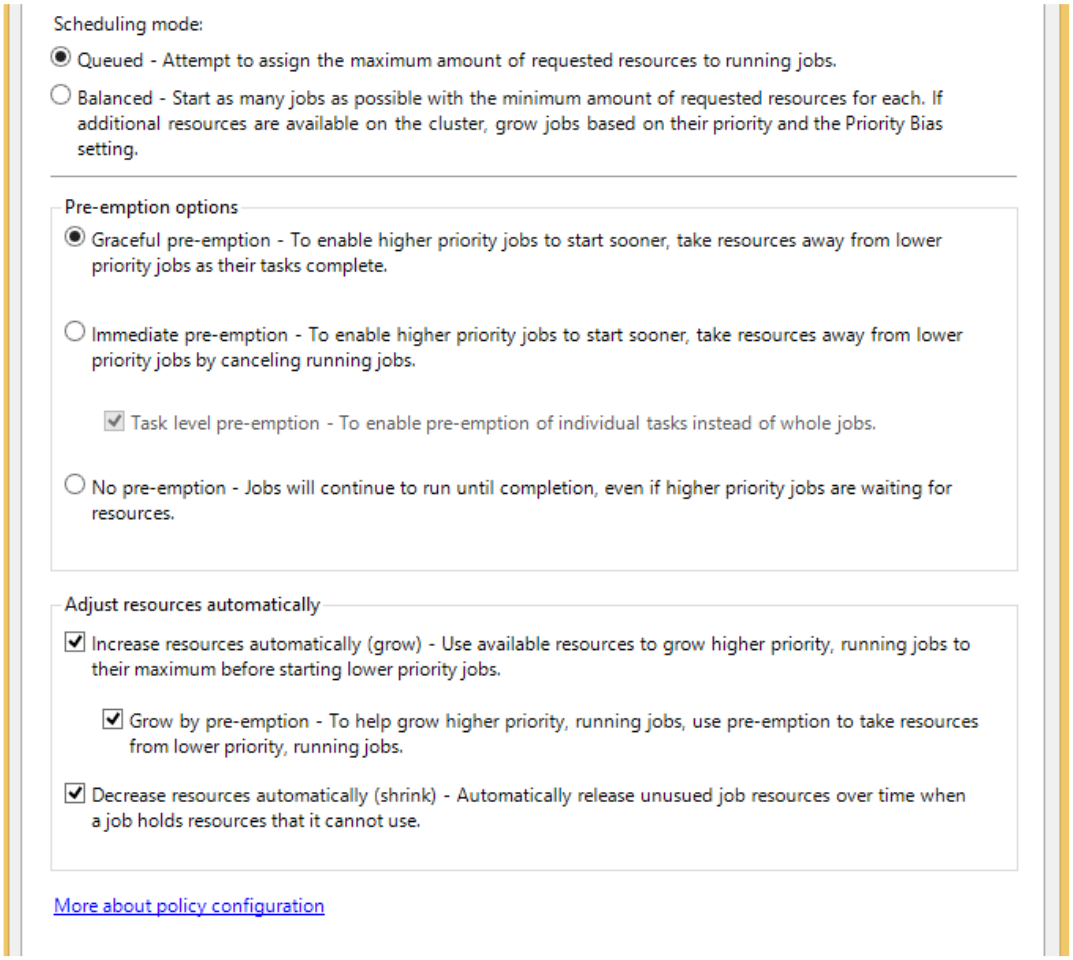
Define la línea de comandos y los ficheros de salida.

Los valores por defecto se pueden respetar si no son necesarios otros.

Los trabajos se pueden enviar remotamente. Para ello se debe establecer una conexión de escritorio remoto a uno de los nodos.
Una referencia completa de cómo hacer esto está disponible en:
https://technet.microsoft.com/en-us/library/gg315415%28v=ws.10%29.aspx

UNIVERSIDAD DE BURGOS

# Planificación encolada.




Principales opciones de planificación encolada.

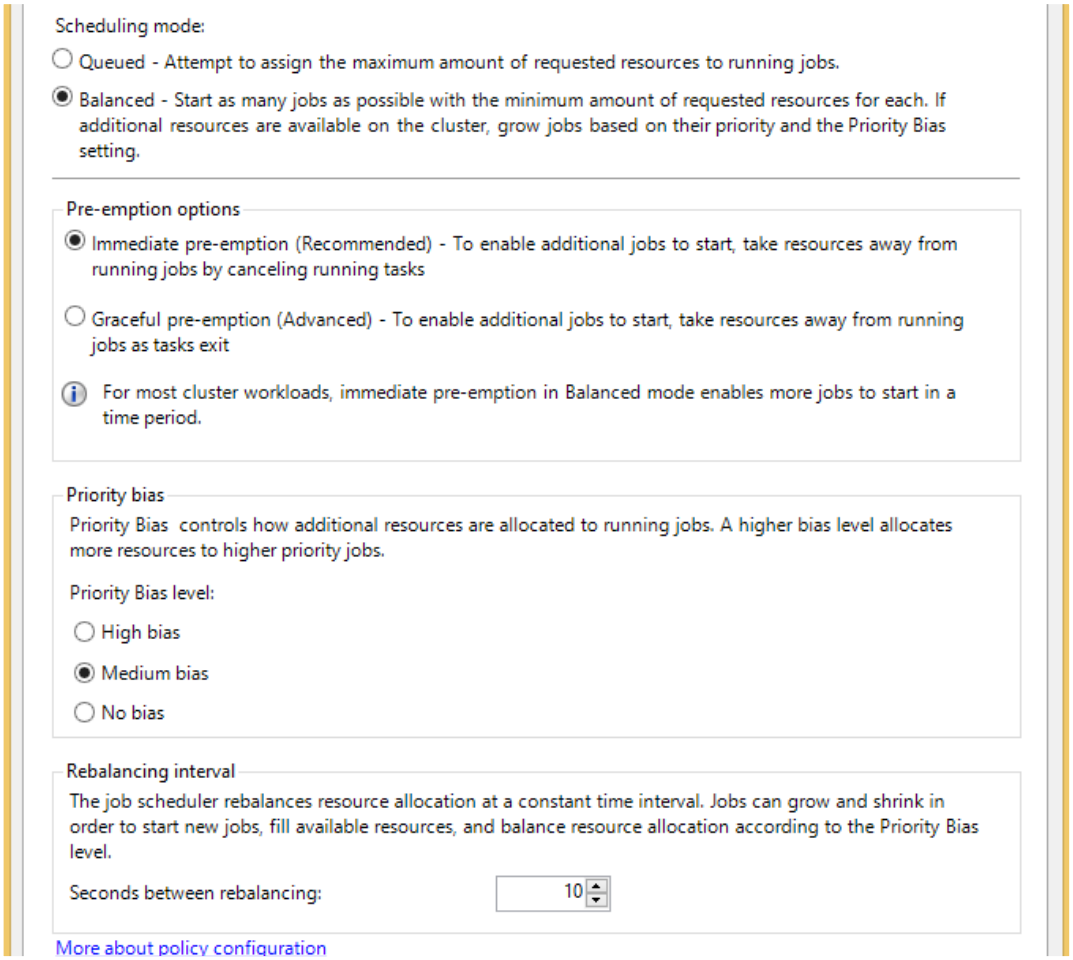
Intenta asignar el mayor número de recursos posibles a los trabajos.

Es posible configurar opciones de precedencia.

También se puede decidir cómo se ajustan los recursos asignados a los trabajos.

UNIVERSIDAD DE BURGOS

# Planificación balanceada.




Principales opciones de planificación blanceada.

Intenta arrancar tantos trabajos como sea posible aunque no pueda asignarles el máximo de recursos.

Es posible configurar opciones de precedencia.

También se puede ajustar el sesgo aplicado a la prioridad a la hora de asignar nuevos recursos disponibles.

El periodo de rebalanceo de recursos es también configurable.

# Referencias

- Cluster network topologies: https://technet.microsoft.com/en-us/library/gg145543.aspx
- Node management: https://technet.microsoft.com/es-es/library/ff919378
- Managing cluster users: https://msdn.microsoft.com/en-us/library/ff919335.aspx
- New cluster user roles: http://blogs.technet.com/b/windowshpc/archive/2013/09/04/hpc-pack-2012-sp1-available.aspx
- Job scheduler: https://technet.microsoft.com/es-es/library/ff919436
- Job manager: https://technet.microsoft.com/es-es/library/ff919691