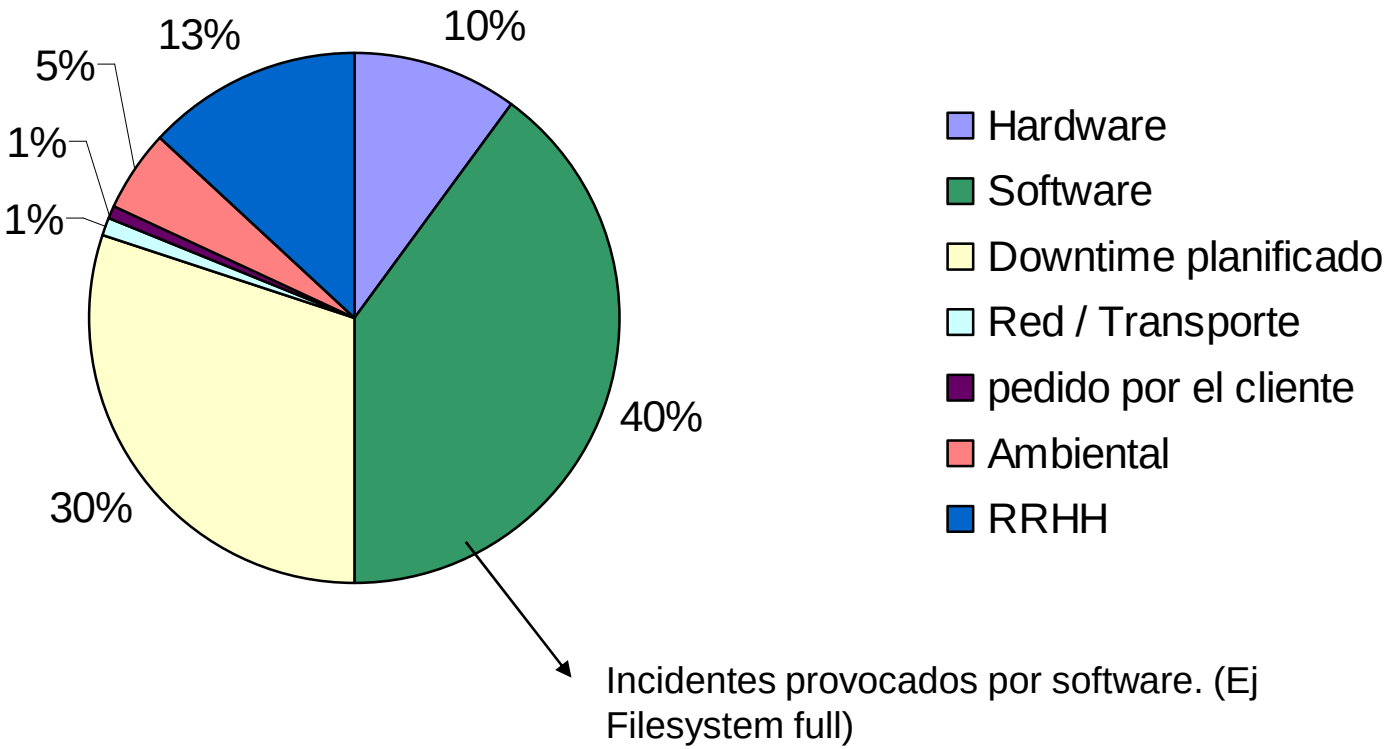
Computación de alta disponibilidad

Universidad Tecnológica Nacional - FRBA

Autor: Gustavo Nudelman

Necesidad de un sistema HA

Causas de downtime. (estudio realizado por IEEE)



Cluster

- ***Definición***: Sistema distribuido compuesto por un conjunto de computadoras autónomas, interconectadas (acoplamiento fuerte), trabajando juntas en forma cooperativa como un único recurso integrado.

- Se interconectan , por lo general, por LAN de alta velocidad.

- Se realizan para aumentar el rendimiento, disponibilidad y confiabilidad de un sistema

- Menor costo a una supercomputadora.

Tipos de cluster

- ***High Availability (HA):*** Provee monitoreo de ***recursos*** y desencadenamiento de proceso failover.
- ***Balanceo de carga:*** Un sistema externo de monitoreo va distribuyendo el workload en los diferentes sistemas utilizando una determinada métrica
- ***HPC:*** los programas dividen el procesamiento en diferentes nodos del cluster

Comerciales

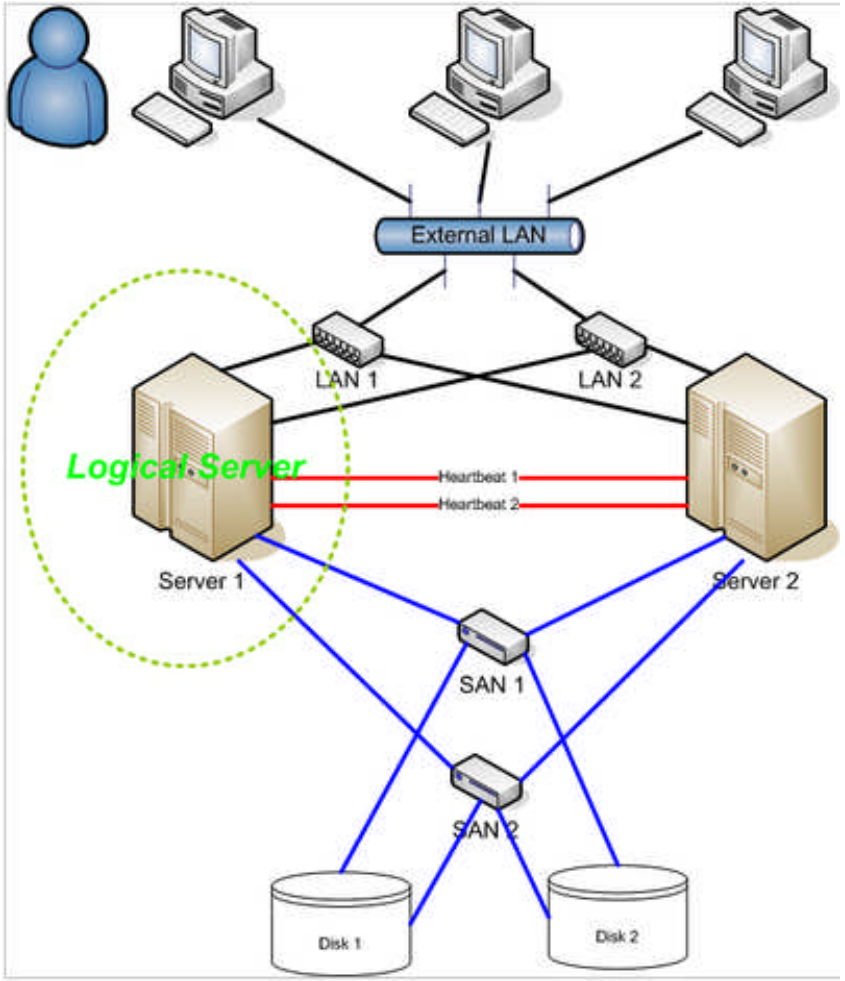
Recurso: Unidad de hardware o software que opera a nivel de nodo y es responsable de proveer un servicio.

ServiceGroup: Grupo de recursos que permiten que uno o mas nodos brinden un servicio

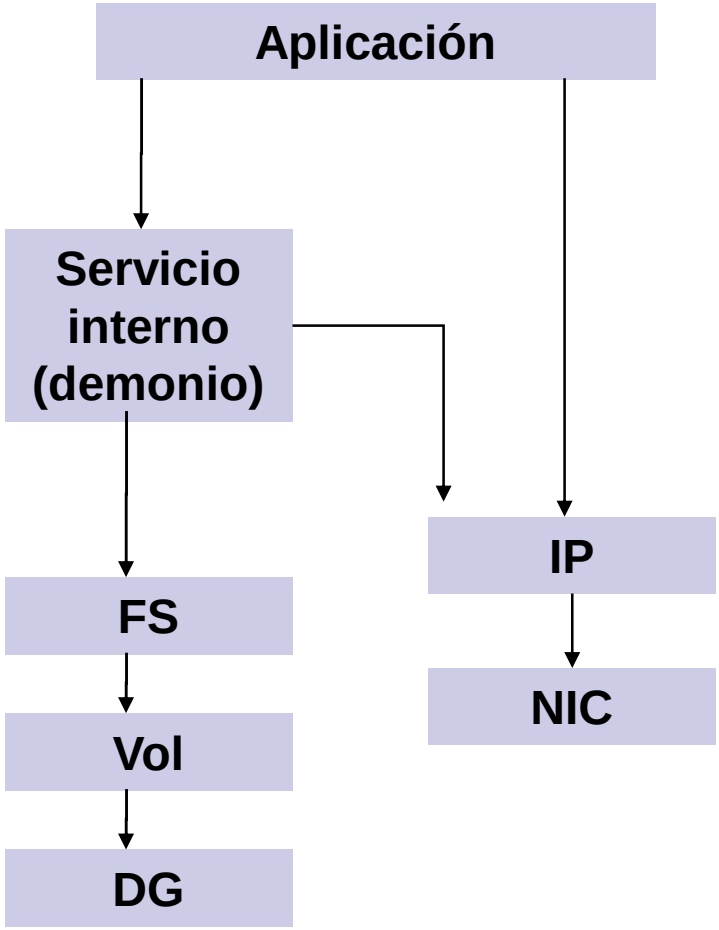
Failover: Proceso por el cual un nodo asume la responsabilidad de otro nodo importando los recursos

High Availability (HA):

- Tolerantes a fallas (failover).
- Disponibilidad de los servicios que el cluster provee.
- Redundancia de nodos. Implica redundancia electrica y de networking
- En general se utiliza la configuración activo-pasivo
- Cualquier servicio puede ser clusterizable
- LinuxHA (FREE!!)



High Availability (HA) : Proceso de failover



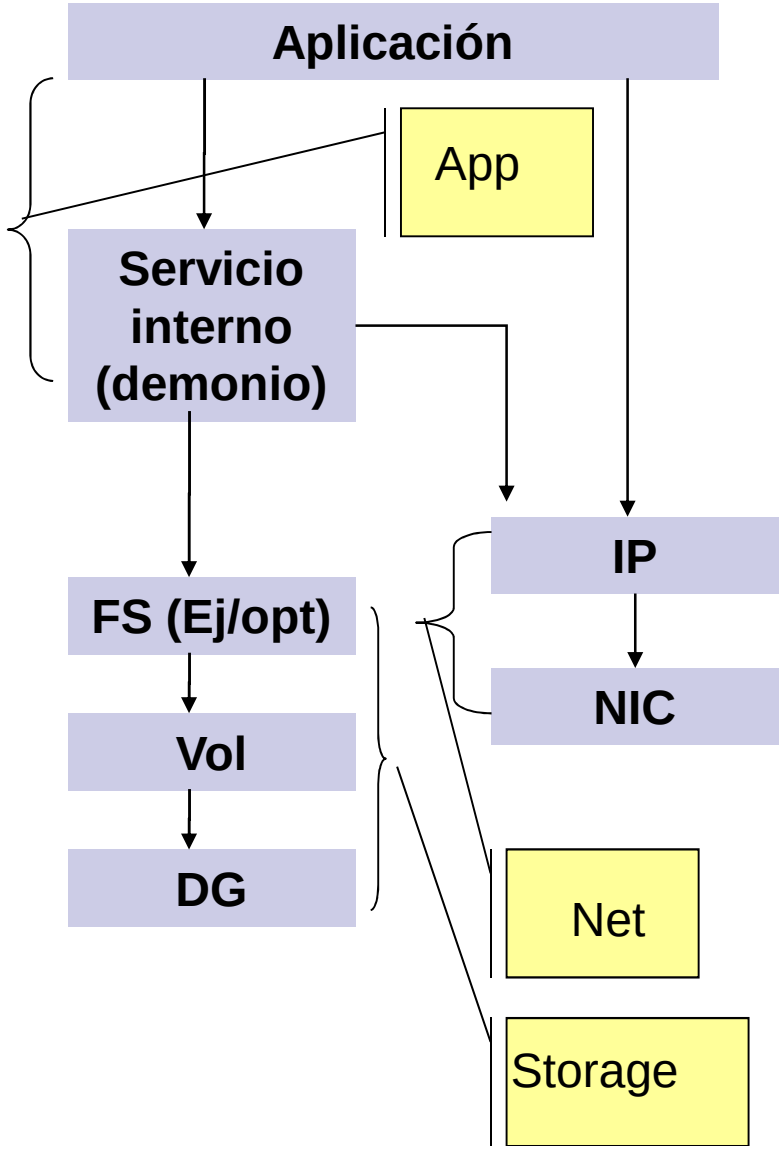
•En un proceso de failover es necesario movilizar a todo el servicegroup como muestra la figura.

•Los componentes de un servicegroup pueden clasificarse en 3 tipos

- •Programas
- •Dispositivos de Red
- •Storage

•En caso de failover, deben migrarse todos los elementos de un service group. Y esto se hace comenzando por los niveles mas bajos hacia los mas altos

High Availability (HA) : Agentes



- Cada sistema corre un agente para monitorear cada recurso
- Los agentes tienen carácter de deamons y se ejecutan en background
- Cada Agente actualiza continuamente un recurso compartido por el sistema (IPC) con el status de su recurso que este controlando
- Los agentes poseen puntos de entrada (funciones) para operar el recurso. Entre ellas las comunes entre los diferentes tipos de agentes son:
  - Online
  - Offline
  - Test
  - Clean (Kill -9)

Una vez que los recursos pasaron a ser controlados por el cluster no es de buena conducta la operación directa sobre los mismos.

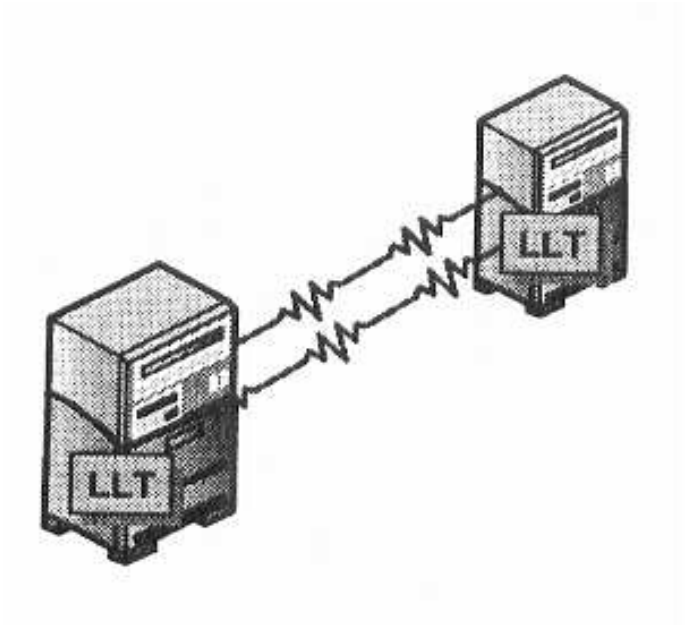
High Availability (HA) : Calculo de duración del proceso de failover

Si el proceso de failover se debe a un incidente. El tiempo de downtime comienza con el dicho incidente

- **Tiempo de detección de la falla:** esta relacionado con la frecuencia de monitoreo (entre 60s y 300s) y debemos tomar el peor caso.
- **Up Local**: Cantidad de reintentos configurados para el agente (usado en las NIC).
- **Bajar recursos:** Llevar al resto de los recursos del service group offline. Se configura cuanto tiempo se asigna para llevar a cada recurso offline de modo no forzado. (se debe sumar cada uno de estos tiempos mas el tiempo forzado)
- **Selección de nodo destino:** Se trata de las pruebas necesarias para decidir cual es el nodo al que se hace failover. Si dicho nodo esta predeterminado este tiempo es despreciable.
- **Up de servicios en el nodo destino:** Muchas veces se deben hacer test de los recursos compartidos que van a ser tomados por el otro nodo: Ej Integridad del Filesystem

High Availability (HA) : Comunicaciones entre nodos

- Se trata de una conexión LAN de alta velocidad (generalmente FC configurada en baja latencia con control de errores)
- Mantiene el membership en cada nodo mediante “heartbeats” (mensajes UDP del tipo keepalive con información de status obtenido de los agentes)
- Mantiene la configuración del cluster. Los cambios en la configuración se almacenan en cada nodo ya que cualquiera puede estar activo y solo en algún momento.
- Se recomienda doble enlace (como muestra la figura) de manera de evitar el estado “Split Brain” en la mayoría de los productos, aumentando la confiabilidad.



Modos de trabajo para HA

- Activo – Pasivo
  - La totalidad de los recursos se encuentran concentrados en un solo nodo llamado “master”
  - Es económicamente mas costosa porque se tiene un nodo totalmente pasivado

- Avtivo – Activo
  - Hay diferentes servicegroups distribuidos en los nodos
  - Se aprovecha la posibilidad de trabajar con mayor rendimeinto
  - En caso de producirse un failover, un solo nodo debe ser capáz de asumir toda la carga

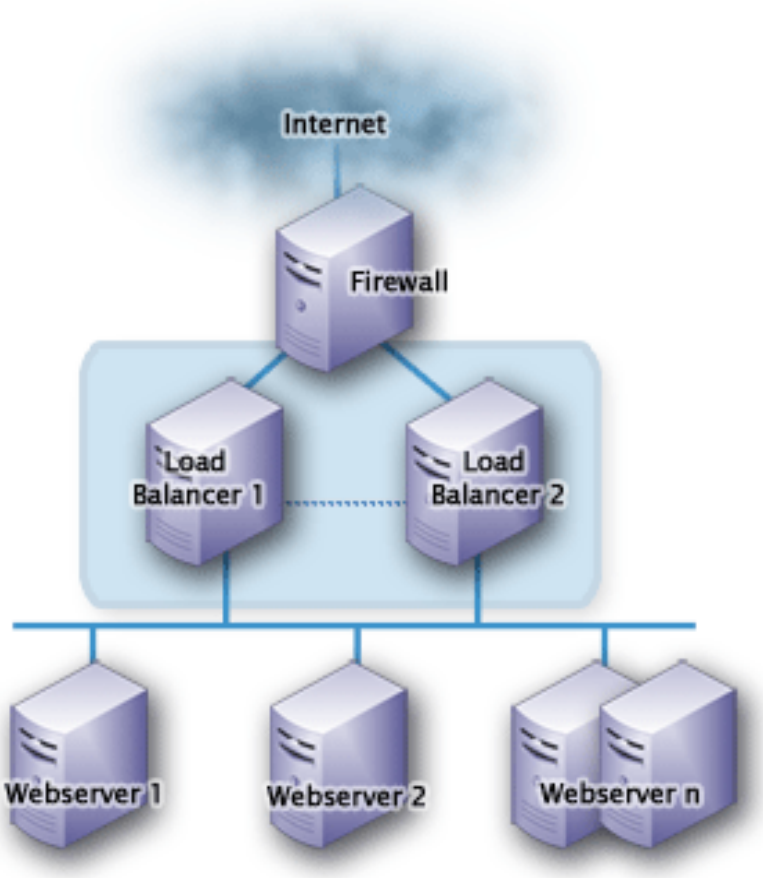
Balanceo de carga

•Es una técnica mediante la cual se distribuye un cierto trabajo, entre varias partes todas capaces de realizar las mismas tareas

•Se lo conoce como “server farm”

•Permite tener redundancia, baja de equipos para mantenimiento programado, y también escalar horizontalmente ampliando recursos según necesidad sin demandar migraciones de datos o interrupciones de servicio.

•Debe existir independencia de datos entre las
tareas individuales. (Esto puede constituir una desventaja)



Métricas para balanceo de carga

- **Round-robin**: Se proveen conexiones equitativas a cada servidor y el LB va rotando desde al primero hacia el ultimo a medida que entran conexiones. Analogo al funcionamiento del scheduler.

- **Weighted round-robin**: Es similar a Round Robin, pero se puede administrar según las diferentes capacidades de los servidores. La secuencia comenzara con los de mayor prioridad asignada.

- **Least-connection:** Se redirigen las conexiones a los servidores en base a cual tiene menos conexiones concurrentes al momento.

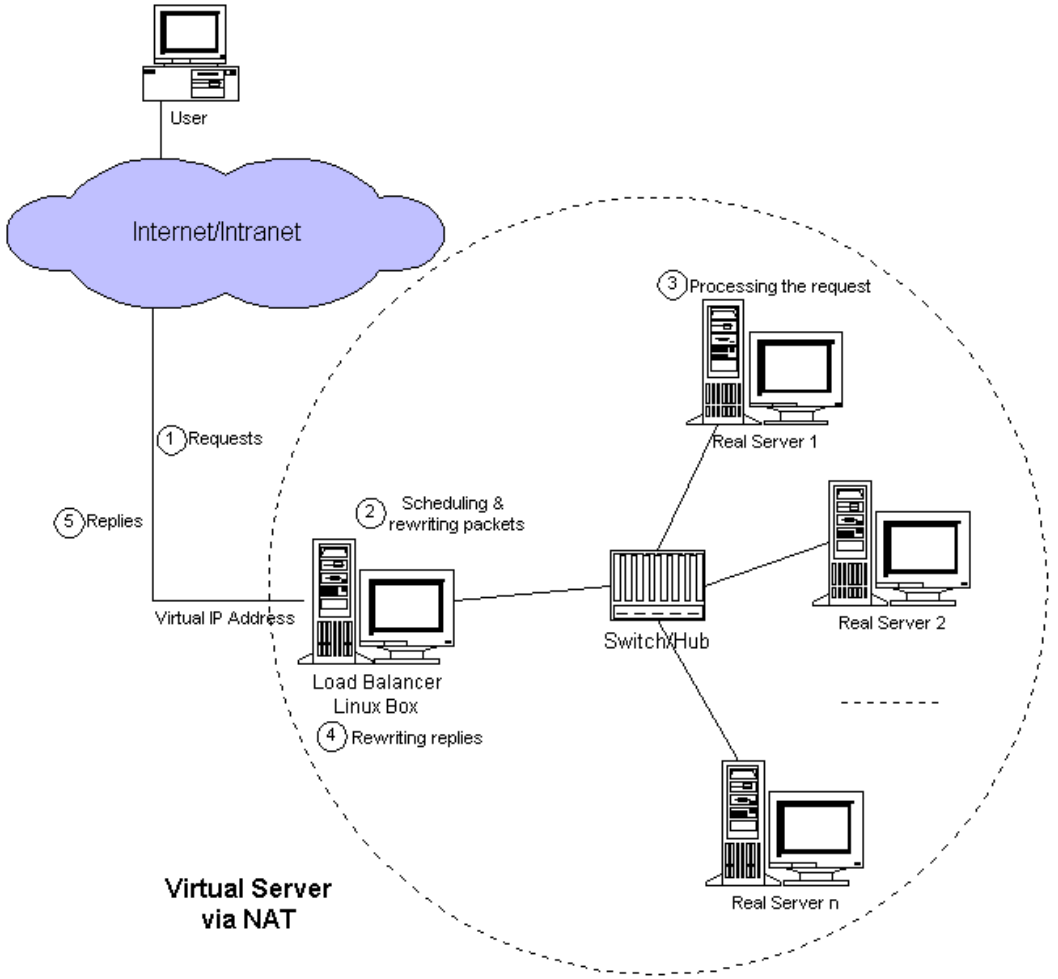
- **Load-based:** Se redirigen conexiones al servidor que se encuentre con menos Workload.

Topologías para balanceo de carga

- NAT
- Tuneling
- Direct Routing

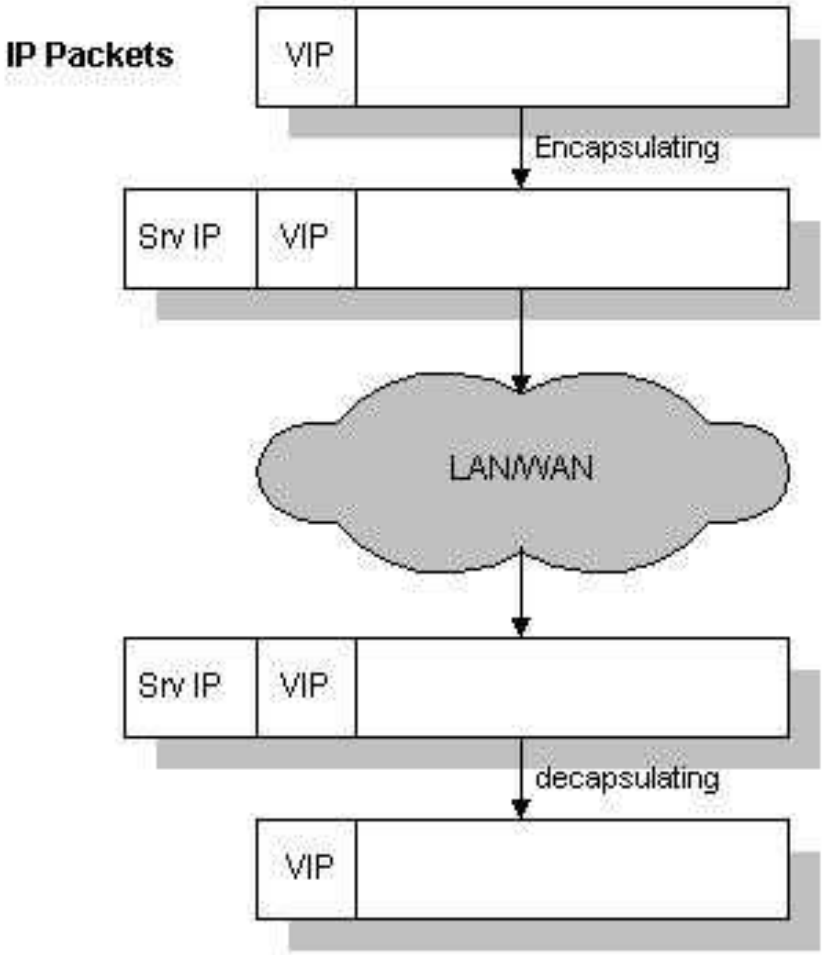
Balanceo de Carga NAT

- El Load Balancer recibe la petición del cliente
- El paquete es reescrito (Nuevo socket) y es reenviado a uno de los servers de la granja
- El servidor procesa la petición y devuelve el resultado al Load Balancer
- El Load Balancer reescribe la respuesta al socket inicial que mantiene o no con el cliente.
- Desventaja: Sobrecarga de trafico en el Load balancer y de CPU del mismo



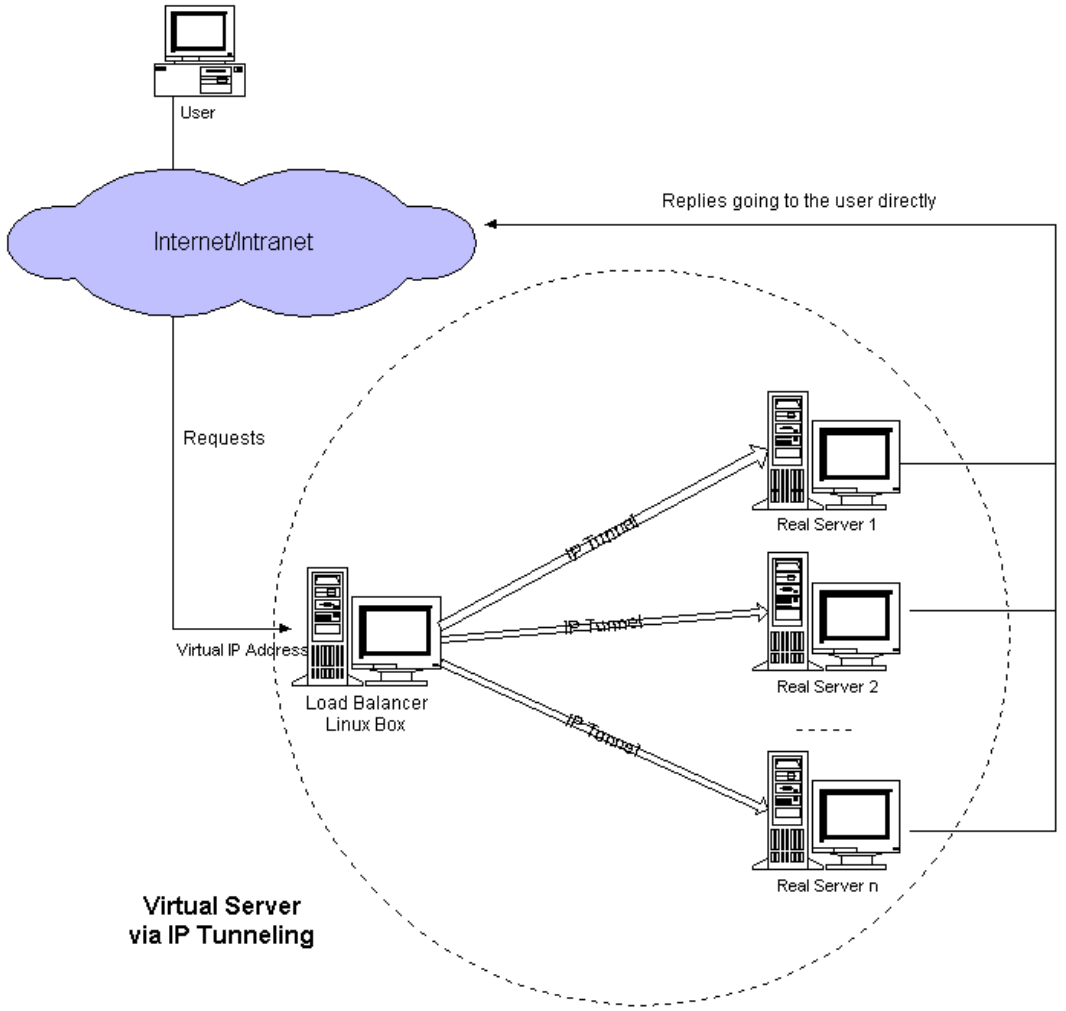
IP tunneling - Concepto

- IP Tunneling consiste en encapsular un datagrama IP dentro de otro y redirigirlo a otra máquina.
- La máquina receptora debe desencapsular el paquete.



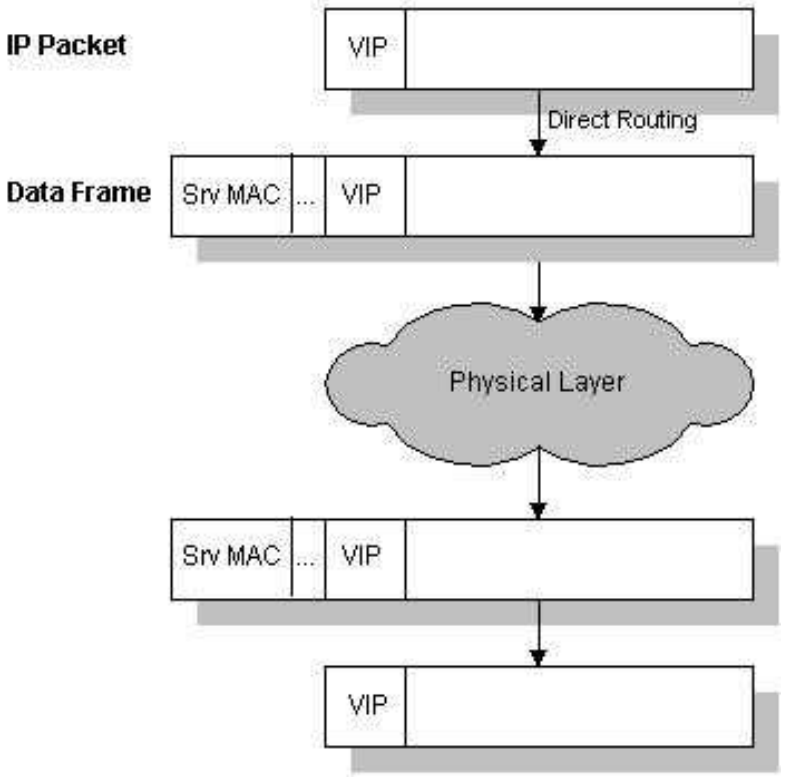
Balanceo de carga por IP tunneling

- El load balancer recibe la petición del cliente.
- El paquete es encapsulado y reenviado a uno de los servidores.
- El servidor desencapsula el paquete, procesa la petición y devuelve los resultados directamente al cliente.
- Pueden manejarse mas conexiones concurrentes por parte del balanceador y escalar mas servidores



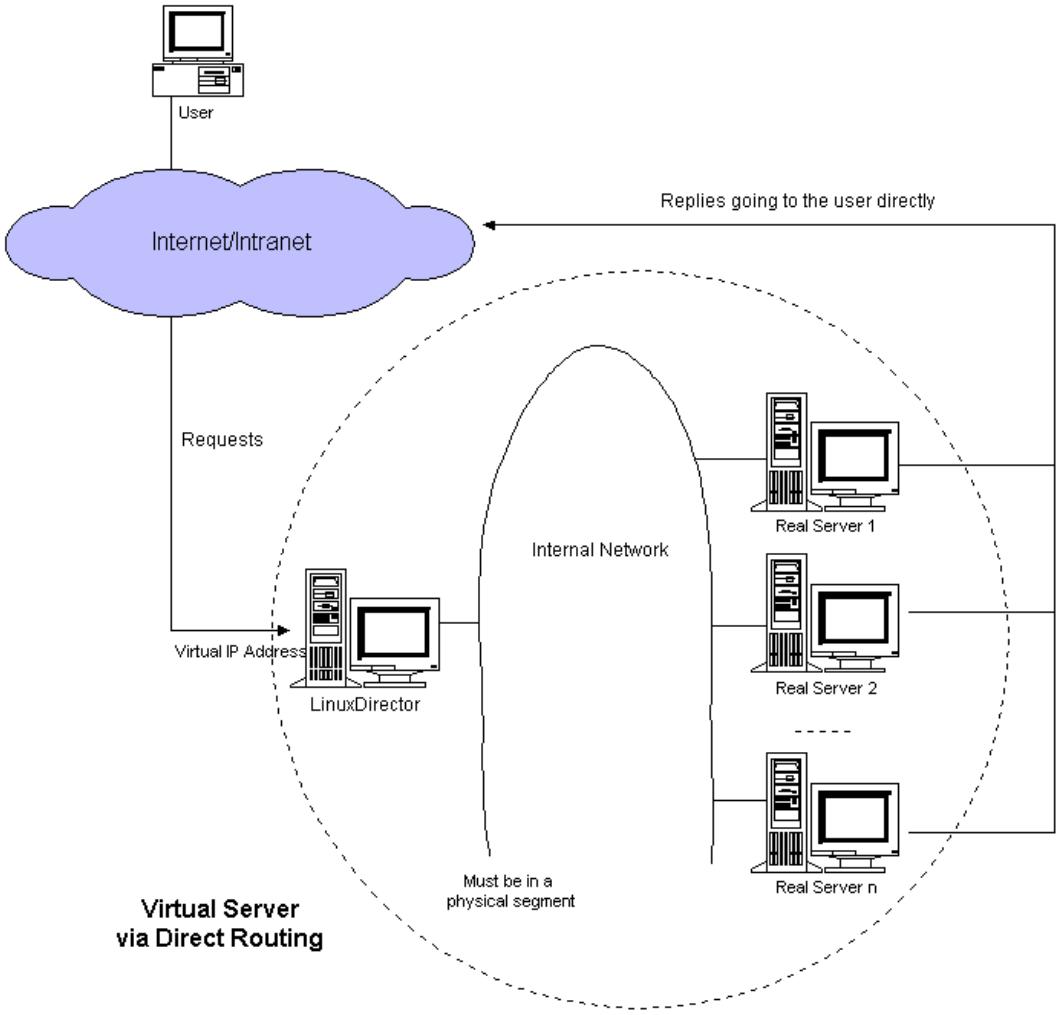
Direct Routing

- Todas las máquinas comparten la misma IP.
- El load balancer enruta el paquete del cliente al servidor elegido basándose en la dirección MAC.
- Los demás servidores, pese a tener la misma IP, rechazarán el paquete.



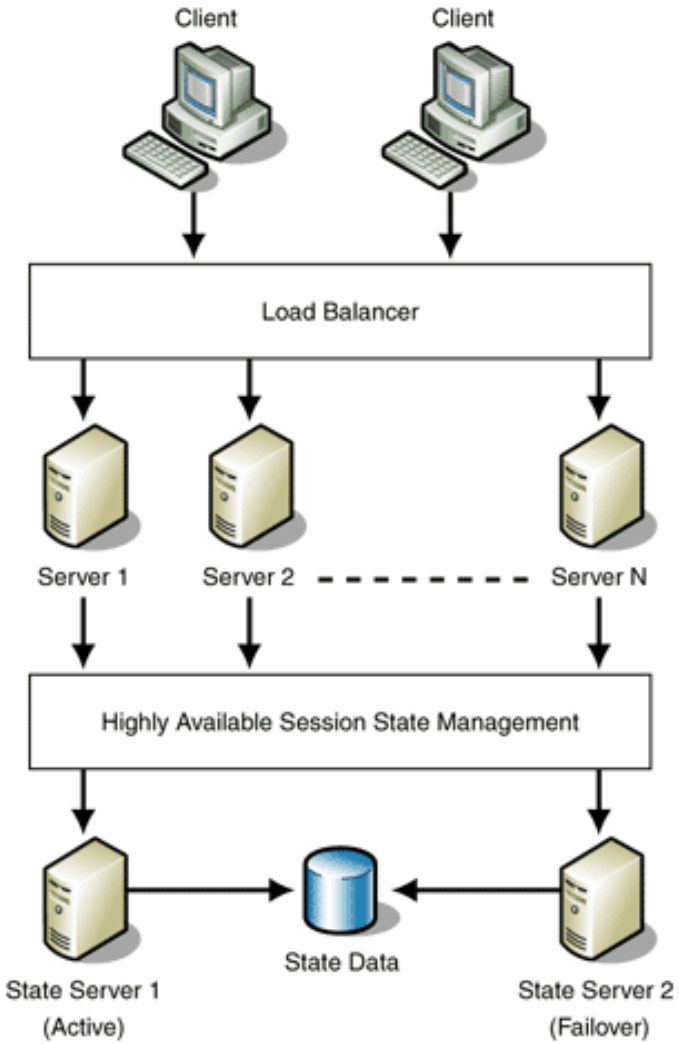
Balanceo de carga por Direct Routing

- El load balancer recibe la petición del cliente.
- Se elige el servidor adecuado y se enruta el paquete hacia él mediante su dirección MAC.
- El servidor procesa la petición y devuelve los datos al cliente directamente.



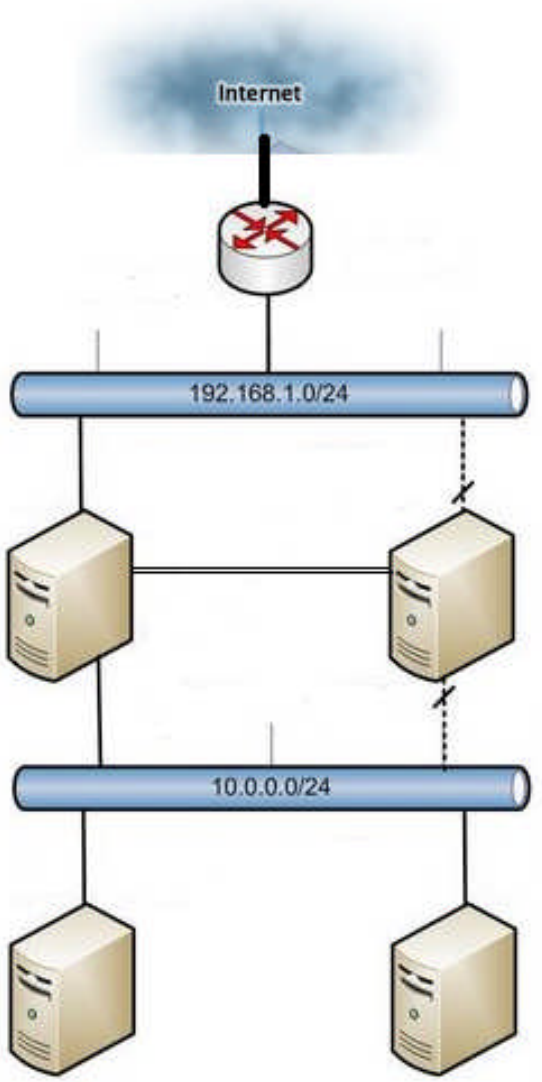
Combinación de LB Y HA

Los servidores de granja, por medio del load balancer dividen la atención de las peticiones y cuando es necesario acceden a un sistema HA (En general como base de datos)



Combinación de LB Y HA (2)

Load balancer formado por un cluster Activo-pasivo para firewall y distribuir sesiones TCP en farm servers.



Procesamiento paralelo (HPC)

•Con este tipo de cluster, se Incrementa el rendimiento de un sistema dividiendo una tarea computacional a través de diferentes nodos del cluster
•Aplicación creciente en Cómputo científico.
- •Genética
- •Sistemas de descifrado de claves de seguridad
- •Procesamiento de imágenes
- •Astronomía
- •Física y matemáticas

Este tipo de cluster no esta orientado a servicios comerciales. Por lo que no dispone de topologías físicas estándares.

En general se trata de redes LAN de alta velocidad donde un nodo funciona como master. Es decir se encarga de distribuir el trabajo en tareas mas pequeñas para los otros nodos.

Finalmente este mismo nodo se encarga de recibir los diferentes resultados, recomponerlos y generar la salida en cuestión

Linux HA

- Proyecto Open Source iniciado en 1999
- Sistema compuesto por procesos
- “deamons” que controlan los recursos y se comunican con sus pares en otros nodos
- Bajo consumo de procesamiento < 1% en latencias superiores a 1s