

In-Class Discussion: Clusters

In the space below, list some questions you think would be important to answer when developing a cluster.

Types of Clusters

High-availability (HA) clusters:

- Implemented primarily to improve the availability of services
- Rely on redundant nodes, data storage, and network connections which take over when a system component fails or if system maintenance is required
- HA operating system must be able to:
 - Detect hardware/software faults
 - Automatically (without human intervention) configure new component and shift failed component's responsibilities to it as seamlessly as possible
- Primary use is for databases, file sharing, business applications, and customer service applications that cannot afford to be down
- Monitoring of health split across the nodes:
 - Each node must have a heartbeat to indicate to other nodes its status
 - Must be some way to monitor health of network, i.e., if network goes down, all nodes may think that they are alone and try to bring up all services on their own. (Split-brain)

Load-balancing clusters:

- Used to provide increased capacity over a uni-processor system by spreading work over multiple nodes.
- Examples include hosting popular web or download sites or DNS servers
- Like HA, allows for continued service in the event of component failure or maintenance
- In some cases, the nodes may be widely separated so as to provide better reliability and balance of service
- Externally, the system appears to be a single device with single IP/network address
- Internally, system is a number of services running on multiple physical devices
- A load balancing operating system must be able to:
 - manage connections between clients and virtual server based on load or specified service
 - allocate requests to the virtual server to be handled by a specific physical server
 - detect hardware/software faults so as not to assign further requests to failed device
 - maintain a client's persistent connection to a specific physical device if so configured
 - monitor health of nodes through probing polls rather than listening for heartbeat

High performance computing (HPC) clusters:

- Also viewed as a single computing system, but with an eye toward improving the performance of a single, demanding application
- Typically associated with scientific research toward the solutions of complex problems
- More concerned with performance rather than reliability and automatic failsafe features
- HPC operating system must be able to:
 - support the user's application parallel computing software
 - create a unified interface for the user
 - monitor performance and load of individual nodes and network traffic

Grid computing:

- Grid computing is not necessarily a different classification of clusters, rather it is a different physical arrangement
- Physically dispersed to the point that interactivity is not encouraged due to network traffic restrictions and a lack of trust between devices
- The results of one job should have no effect on any other jobs, i.e., jobs are independent
- Grid computing middleware must be able to:
 - support a more diverse collection of devices
 - divide workload so that there is very little or no interaction between nodes

What system details would you use to classify a cluster as high-availability, load balancing, high performance computing or grid computing?

Applications

Using the questions you came up with on the first part of this worksheet, try to define the characteristics of a cluster system to support the following applications.

- <http://www.desire2learn.com/> (Hopefully you know what this is.)
- <http://www.jamstec.go.jp/es/en/index.html> (Earth simulator project)
- http://www.oprah.com/slideshow/oprahshow/slideshow1_ss_gifts_of_350 (Assume you are a manufacturer whose product was just listed as one of Oprah's favorite things.)
- <http://www.inrix.com/> (Traffic information – Google Maps provides a similar service)
- http://en.wikipedia.org/wiki/Thin_client (Thin client model)
- <http://setiathome.berkeley.edu/> (Search for Extraterrestrial Intelligence using Internet –connected computers)