

## How the Cluster Service Works

Windows clusters are built from identical (or nearly identical) server computers, referred to as *nodes*. Each node must contain its own internal hard drives, which are used to store the operating system and any applications that will run on that node, including any clustered applications.

Each node must be connected to an external storage device. Typically, this involves a physical connection to an external small computer system interface (SCSI) drive array or a storage area network (SAN) connection to a fibre channel drive array. The drive controllers connected to the external storage must specifically support Windows clustering, and they must use drivers that comply with Microsoft's requirements for cluster storage devices. The external storage array is used to store data for any clustered applications, as well as the cluster's configuration information, or *quorum*. Windows clustering supports multiple external storage arrays per cluster.

When the first node in the cluster is powered up, it performs a reset of the SCSI bus, which gives it immediate control of the external storage array. After Windows starts, the node starts all its clustered applications, which access their data on the external storage array. The node also begins sending a heartbeat signal to other nodes in the cluster. This signal can be carried by a typical network connection, but it's recommended that a separate network be used to carry the heartbeat signals. Typically, this separate network requires an independent network adapter in each cluster node, with a dedicated hub or switch connecting each adapter to the others.

When additional nodes are powered up, they also perform a SCSI bus reset. However, the first node detects this and immediately performs its own bus reset, retaining control of the external storage array. The new node sees this second reset and assumes a passive role in the cluster. When Windows starts, the new node leaves any clustered applications in a suspended state and begins monitoring the heartbeat signal from the passive node.

If the heartbeat signal stops for more than one second, the passive node performs a SCSI bus reset on the external storage array. It then waits to see whether the active node reasserts itself by resetting the bus; if it does, the passive node assumes a network failure in the heartbeat signal and waits for the heartbeat to resume. If, however, the active node fails to reset the bus, the passive node assumes that the active node has failed and immediately assumes the active role, starting its clustered applications and reading the cluster configuration status from the quorum.

Windows clustering supports active-active configurations, as well. In these configurations, each node is connected to multiple external storage arrays—at least one storage array per node in the cluster. Each node has active control over at least one array and is passive for the

other arrays. Each node monitors the heartbeat signal from the other nodes. So, if a node fails, another node immediately seizes control of the failed node's storage array and begins starting whichever clustered applications the failed node was running, based on the information in the quorum. Essentially, an active-active configuration represents multiple logical clusters running on multiple physical machines.