IRIS FailSafe™ 2.0

High-Availability Solution

Silicon Graphics IRIS FailSafe 2.0 is the next-generation high-availability solution for your business-critical applications. Using ccNUMA-based Origin™ servers, IRIS FailSafe 2.0 provides a highly available application platform at a fraction of the cost of specialized fault-tolerant systems. IRIS FailSafe runs in a cluster environment. In the event of a failure IRIS FailSafe automatically fails over applications from one system in the cluster to the other. In combination with a RAID or mirrored disk configuration, an IRIS FailSafe cluster provides resilience from any single point of failure and acts as insurance against unplanned outages. You don't even need to modify your applications in order to realize the high level of availability provided by IRIS FailSafe.

High Availability at a Low Cost

IRIS FailSafe 2.0 uses advanced distributed software technology and standard off-the-shelf hardware to provide high availability at a low cost. In normal operation all systems in a cluster can be active, working as if they were independent servers. In the event of a failure, one of the other systems will take over the services of the failed system, transparently fulfilling requests from clients on the network. IRIS FailSafe supports a mix of CHALLENGE® and Origin servers within one cluster, preserving your investment in your existing computing infrastructure.

Easy Setup and Administration

The IRIS FailSafe Cluster Manager GUI is a Java™ language-based interface that will support you through the entire process of configuring and administering an IRIS FailSafe 2.0 cluster. Designed with substantial input from system administrators experienced in working with clusters, the IRIS FailSafe Task Manager presents you with a task-oriented, step-by-step interface to configure your cluster and manage all servers from a single desktop. The IRIS FailSafe Cluster View tool provides an up-to-the-minute graphical representation of your cluster, servers, and resources and gives you access to the tools you need to manage them. The IRIS FailSafe Cluster Manager CLI, a text-based interface, allows administration over slow connections and supports writing scripts for automated control of administrative activities.

IRISconsole™ provides a single view for console services on all systems in the cluster, simplifying system administration. You can run the IRIS FailSafe Cluster Manager GUI and IRISconsole from the same workstation.

Performance Co-Pilot™ monitoring tool helps locate and visualize trouble spots, preventing failures before they affect the system. A special-purpose Performance Co-Pilot agent for IRIS FailSafe provides integrated cluster-wide performance management.

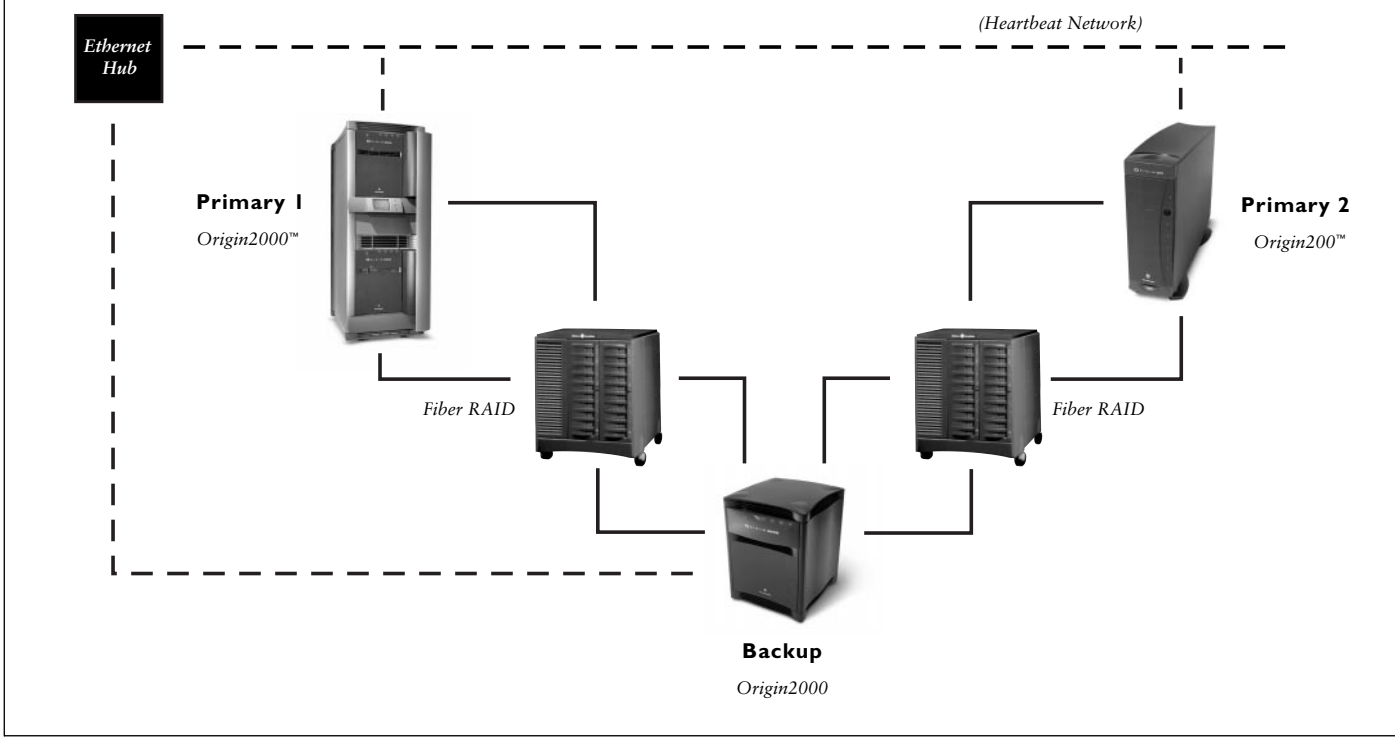
IRIS FailSafe Architecture

In an IRIS FailSafe 2.0 cluster, up to eight servers are connected to both a public and a private network. Clients use the public network to access services from the cluster. The IRIS FailSafe software uses the private network to exchange heartbeat and other control messages. In the event of a server or application failure, one of the surviving systems will assume the public network address of the failed system and respond to client requests on the network. Clients perceive the failover process as a rapid reboot of the system to which they were connected.

IRIS FailSafe 2.0 is built on top of the Cluster Administration and Membership Services (CAMS) layer. The CAMS layer is highly sophisticated distributed software that efficiently controls all the

IRIS FailSafe 2.0

2 + 1 Configuration Diagram



resources and applications in a cluster. This layer enables efficient addition and deletion of systems and applications from a cluster and also intelligently recovers the cluster from network partitioning.

The clustered servers share storage either on RAID or mirrored disks. A shared storage subsystem allows multiple servers to assume control of the data in the event of a failure. In the event of a failure, filesystems are automatically made available on the system(s) where their corresponding applications are resumed.

An IRIS FailSafe 2.0 cluster can be configured either in N x 1 or in N x N mode. In an N x 1 configuration N machines run various mission-critical applications while one machine is in standby mode, prepared to take over if any of the primary N servers fails. This configuration ensures that your environment sees no performance degradation even after failure of a server. In N x N configurations all servers are running business applications while also acting as backup servers. In the event of a failure of a server, its workload is failed over to one or more of the surviving servers in the cluster. The workload of the failed server can be divided among multiple surviving servers, preventing overloading of any server.

Application Recovery Kits

IRIS FailSafe 2.0 is flexible enough to meet the needs of any high-availability application. Since all customer requirements are not the same, IRIS FailSafe 2.0 utilizes scripts that can be modified to control the behavior of the cluster in the event of a failure or recovery. A suite of optional IRIS FailSafe recovery kits implements high-availability solutions for NFS™, Web and e-mail serving, databases (Oracle, Sybase, and Informix), and other common services. Additionally, customers can develop simple scripts to integrate their own crash-tolerant applications with IRIS FailSafe, extending the high-availability functionality to a wide range of additional applications.

SiliconGraphics Computer Systems

Corporate Office
2011 N. Shoreline Boulevard
Mountain View, CA 94043
(650) 960-1980
www.sgi.com

U.S. 1(800) 800-7441
Europe (44) 118-925.75.00
Asia Pacific (81) 3-54.88.18.11
Latin America 1(650) 933.46.37

Canada 1(905) 625-4747
Australia/New Zealand (61) 2.9879.95.00
SAARC/India (91) 11.621.13.55
Sub-Saharan Africa (27) 11.884.41.47

1748 (12/98)