



ON AIR

**A**s we approach the pervasive-computing era in which users will have round-the-clock access to information and services from any location, embedded-system designers are under growing pressure to boost the availability of servers, remote devices, and the data-transport infrastructure. Due to their application, embedded devices have a much higher reliability expectation than most other computing systems. You cannot stop or reboot some of these systems, such as those in critical applications, without risking loss of life, property, or essential information. In meeting these requirements, embedded-system designers use an arsenal of clever hardware- and software-redundancy techniques to routinely achieve availabilities as high as 99.999%, or less than six minutes of downtime per year.

“High availability” describes the characteristics of a system that allow it to sustain continuous operation in the event of hardware or software failures. With built-in monitoring and duplicate datapaths, a highly available system survives failures by transparently substituting alternative hardware or software components to reproduce normal functions. In general, a high-availability system also includes provisions for replacing failed components or upgrading performance without disrupting operation. With the advent of universal connectivity, data security also becomes an element of availability, because expected functions can see interruption due to an unauthorized hacker, malicious

# ALWAYS **O**N: EMBEDDING HIGH AVAILABILITY

BY WARREN WEBB • TECHNICAL EDITOR

DESIGNERS ARE TUNING HIGH-AVAILABILITY ARCHITECTURES TO MEET CUSTOMER DEMANDS FOR A PERSISTENT DATA INTERFACE FOR THE NEXT GENERATION OF ULTRARELIABLE EMBEDDED-SYSTEM APPLICATIONS.



**Figure 1**  
Adlink's aTCA-6900 features two quad-core Intel Xeon processors and dual mezzanine-card bays to extend the performance of high-availability AdvancedTCA systems.

software, or an external denial-of-service attack. Availability is normally defined as  $MTBF/(MTBF+MTTR)$ , where MTBF is the mean time between failures and MTTR is the mean time to repair.

Although high availability has become essential to a growing list of embedded-system applications, escalating technology trends make the system-design task increasingly difficult. For example, as customers demand more functions in embedded devices, the added hardware and software components create new failure modes to anticipate. Obviously, added components work against higher availability and even generate other redundancies that continue to increase system complexity. The current movement toward ubiquitous connectivity also creates a host of data-security and communications-reliability problems for the high-availability-embedded-system designer. Although the most reliable system is probably a simple, stand-alone device with limited resources, designers must adapt a strategy that extends the availability of any embedded configuration.

### CAN YOU HEAR ME NOW?

Most of the current tricks and techniques for extending service availability originated in the telecommunications industry. Over the years, telecommunication manufacturers devised multiple schemes to provide uninterrupted service despite hardware and software failures. Unfortunately, most of these

#### AT A GLANCE

- ▣ High-speed serial-data links and switched-fabric technology enable dynamic paths so that you can reroute information around inoperable subsystems.
- ▣ Management software automatically monitors system operation and substitutes redundant components in the event of a failure or degraded performance.
- ▣ Hot-swap features enable repairs and upgrades with no downtime and pave the way for fault-tolerant, self-healing systems.
- ▣ Clusters of blade computers enable scalable, high-density, highly available server systems at reduced acquisition and operating cost.

schemes were proprietary, expensive to maintain, and difficult to update as requirements evolved. They also required long development cycles. Equipment designers were unable to use COTS (commercial off-the-shelf) building blocks, because there were no common built-in provisions for extending service availability. To tackle the availability conundrum, board manufacturers created a series of hardware and software specifications that could match the performance of proprietary systems.

One of the earliest standards to address availability was the IPMI (Intelligent Platform Management Interface) specification, which Dell, Intel, Hewlett-

Packard, and NEC defined to allow local and remote monitoring of equipment for power management, cooling, electronic keying, and hot-swap transactions. IPMI interacts with a management controller that works on its own if the host processor is defective. With platform management, operators can monitor equipment for marginal operation or potential problems and correct them before they become system failures. PICMG (Peripheral Component Interconnect Industrial Computer Manufacturers Group) incorporated variations of IPMI into both the CompactPCI and ATCA (Advanced Telecom Computing Architecture) board-level specifications.

To derive the maximum benefit from IPMI, equipment customers needed a hot-swap capability to replace defective boards without shutting down their systems. Hot-swap systems require hardware and software that can dynamically route signals around defective components while waiting for repairs. One of the essential technologies of hot swapping is the physical connection between the board and the backplane. A simple direct connection can disrupt the other boards on the bus without controlling power-supply inrush current and backplane-signal connections. For example, CompactPCI uses staged pins of different lengths to control the physical connection to the backplane. Card guides ensure that board insertion is perpendicular to the backplane. The longer pins are the first to mate and supply power and ground to precharge the PCI-bus signals. Series resistance limits the power-supply current surge. The medium-length pins connect to the PCI-bus signals that are in a precharged, high-impedance, or disabled state. The shortest pins enable bus communications.

### FAILPROOF FABRIC

Serial-switched-fabric technology is another design innovation that has multiple benefits for high-availability systems. These architectures allow dynamic data-paths between computing nodes and support multiple simultaneous data transfers. A major benefit of a switched fabric is that each connection is a direct point-to-

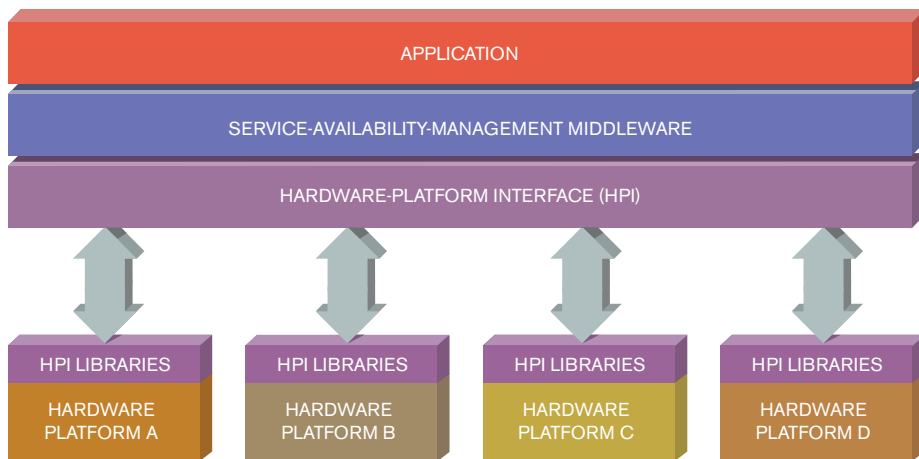


Figure 2 The hardware-platform-interface specification from the Service Availability Forum defines an interface between COTS hardware and management middleware.

## THE HIGH AVAILABILITY LINUX PROJECT HOSTS AN OPEN-SOURCE-DEVELOPMENT EFFORT TO PROVIDE A CLUSTERING ARCHITECTURE FOR THE LINUX OPERATING SYSTEM TO PROMOTE RELIABILITY, AVAILABILITY, AND SERVICEABILITY.

point datapath and yields better electrical characteristics, allowing higher frequencies and bandwidth than bus architectures. A typical switching fabric uses multiple stages of switches to route transactions between a source and a target. These dynamic paths are also valuable to high-availability designs, allowing you to route data around inoperable subsystems. Most of the major board standards now provide for switched fabrics, although they do not call out a specific fabric technology for data transport. Instead, a series of subsidiary specifications defines backplane details for the various fabrics, such as Ethernet, InfiniBand, StarFabric, PCI Express, and RapidIO. Although this approach satisfies differing views within the industry, it can also create interoperability issues within the same standard.

The VITA (VMEbus International Trade Association) 41 VXS (VITA switched serial extensions) add some of the high-availability benefits of fabric technology to the popular VMEbus (Ver-

sa-module eurocard bus). The VXS specification defines a payload card, a switch card, and a new high-bandwidth-backplane connector and retains the standard parallel-VMEbus connectors. Each new fabric port consists of two sets of four ganged serial-bit channels—one set for input data and the other set for output data supporting data rates of 10 Gbps for each serial channel. Switch cards contain the fabric switching necessary to route serial data between payload cards and around failures. To remain fabric-agnostic, VITA 41 subspecifications define switch and payload-card definitions for InfiniBand, serial RapidIO, GbE (gigabit Ethernet), and PCI Express.

Targeting the unique requirements of telecom equipment, the PICMG (PCI Industrial Computer Manufacturers Group) released the ATCA specification to provide an alternative to open architectures such as VME and CompactPCI. With emphasis on high-availability features, ATCA uses the high-speed serial-data links and switch-fabric technology. The extra-large board area supports the complex telecom circuitry and provides input power and cooling for as much as 200W per slot. The ATCA specification features hot-swap capability for all boards and active modules, allowing systems minimum downtimes. A shelf-management element, which the specification based on IPMI, monitors the health, power, cooling, and even keying of plug-in modules to ensure that subsystems are operating efficiently. Modules get power from redundant -48V-dc power feeds and data from redundant control and data planes to prevent a single failure from bringing down an entire chassis.

Taking advantage of the hot-swap and shelf-management features of ATCA and extending the performance envelope, Adlink Technology recently announced the aTCA-6900 CPU blade featuring two quad-core Intel Xeon processors and two AdvancedMC (mezzanine-card) bays for design flexibility (Figure 1). The aTCA-6900 CPU blade supports as many as eight CPU cores plus a fabric architecture that includes dual 10-Gbit Ethernet, dual PCI Express, and dual Fibre Channel interfaces. Onboard storage includes 4-Gbyte USB flash and a variety of hard-disk-mounting options. Front-panel I/O includes

### MORE AT EDN.COM

+

For more on designing for reliability, go to [www.edn.com/article/CA6535342](http://www.edn.com/article/CA6535342).

+

For more on hot-swapping and high availability, go to [www.edn.com/article/CA46756](http://www.edn.com/article/CA46756).

+

Go to [www.edn.com/080515df](http://www.edn.com/080515df) and click on Feedback Loop to post a comment on this article.

video, three USB 2.0 ports, two RJ-45 Ethernet ports, and an RJ-45 serial port. Prices for the aTCA-6900 start at less than \$5000.

### SURE-FIRE STREAMS

As the number of networked embedded devices grows, the need for a dedicated and reliable data source becomes a major consideration in any new-product development. If you deploy multiple devices and each requests a different but simultaneous data stream, the data-server-processing requirements become critical. Many embedded-system applications, such as file sharing, security surveillance, and entertainment, require an independent and always-on data stream from a dedicated server. To meet the availability expectations of these data-centric projects, designers are turning to high-density computer arrays with hundreds of CPUs per rack and multiple CPUs per board. A system with multiple computer boards is typically called a blade server and features system management, load balancing, hot-swap capability, and shared peripherals to provide the high-reliability data for Web access and data services. Individual blade computers generally have no local peripherals, and you manage them remotely. Cluster-type servers run management software to balance the computing load, report failures, provide blade-configuration information, and oversee hot-swap transactions. Blade servers are basically high-availability systems that require special software to manage the system for maximum uptime. A separate management network increases server security by keeping critical operating-system information and updates from passing over public networks or the Internet.

Several open-source and commercial software organizations are dedicated to improving the reliability of operating systems and embedded firmware. For example, the High Availability Linux Project hosts an open-source-development effort to provide a clustering architecture for the Linux operating system to promote reliability, availability, and serviceability. Heartbeat, the most well-known component of the project, sends periodic packets across the network to the other instances of Heartbeat to verify performance. When the system

no longer receives packets, it assumes a node failure and automatically reroutes services to an alternative node, according to a user-supplied formula.

Similarly, the Service Availability Forum comprises communications and computing companies working to develop high-availability and management-software-interface specifications. These specifications target the developers of telecommunications systems and services built with COTS building blocks, such as CompactPCI and ATCA. The objective is to allow for greater hardware and software reuse plus shorter product-development cycles. The hardware-platform-interface specification defines the interface between the COTS hardware and the high-availability management middleware (**Figure 2**). Applications can then independently discover, monitor, and manage the hardware without proprietary software interfaces.

Thanks to the latest generation of board standards plus a dedicated community of software developers, designers now have the tools to configure high-availability embedded systems using off-the-shelf products. Despite the trend toward multifunction and complex embedded products, designers can combine components from a variety of suppliers to match their performance requirements and still attain or even exceed the elusive “five-nines” (99.999%) availability target. **EDN**

### FOR MORE INFORMATION

**Adlink Technology**  
[www.adlinktech.com](http://www.adlinktech.com)

**Dell**  
[www.dell.com](http://www.dell.com)

**Hewlett-Packard**  
[www.hp.com](http://www.hp.com)

**High Availability Linux Project**  
[www.linux-ha.org](http://www.linux-ha.org)

**Intel**  
[www.intel.com](http://www.intel.com)

**NEC Electronics**  
[www.necel.com](http://www.necel.com)

**PICMG (PCI Industrial Computer Manufacturers Group)**  
[www.picmg.com](http://www.picmg.com)

**Service Availability Forum**  
[www.saforum.org](http://www.saforum.org)

**VITA (VMEbus International Trade Association)**  
[www.vita.com](http://www.vita.com)

You can reach  
Technical Editor  
**Warren Webb**  
at 1-858-513-3713  
and [wwebb@edn.com](mailto:wwebb@edn.com).

