

The Anatomy of a High-Availability Rack

November 2009



Online Tech, Inc.
220 E. Huron
Ann Arbor, MI 48104
(734) 213-2020

www.onlinetech.com

1. Audience

This paper is intended for small and medium-sized business executives in need of an IT infrastructure that provides always-on power and always-connected Internet 24 x 7 x 365.

2. Abstract

With its redundant power and network connectivity, High-Availability (HA) racks are uniquely designed to withstand single points of failure. Each server can have two completely separate and redundant power circuits and two completely separate and redundant paths to an Internet connection, making the high-availability architecture essential for businesses that depend on their servers to be always-on and always connected to the Internet.

3. Introduction

Simply put, business today has evolved to rely on computers and the Internet. More and more frequently, companies are turning to and relying on the Internet to generate revenue. But what happens in the event power is lost or Internet connectivity is disabled? The short answer for many is that their business will lose money. And the longer the business is offline, the more difficult it becomes to rebound from the loss of revenue. This is especially true for small businesses, particularly those that rely solely on Internet sales and commerce.

Using high-availability architecture can greatly reduce the risk of losing business due to a disaster or other anomaly in a leg of the power infrastructure or network connection to the Internet. How does a high-availability server provide this type of protection? The answer lies in the anatomical design of the server rack.

The purpose of this paper is to describe the anatomy of a high-availability rack. This paper illustrates how redundant power and network connectivity help ensure a greater level of protection against any single point of failure in the power delivery or network connectivity to a company's servers, resulting in a higher uptime service level. A high-availability rack helps businesses survive network or power failures, keeping computers online and maximizing the potential to generate revenue.

4. Anatomical Breakdown

Figure 1 shows the anatomy of a high-availability rack with dual home run power (two separate power runs from the server to the utility power and back-up generators) and dual home run networks (two separate networks connecting the server to the Internet service provider).

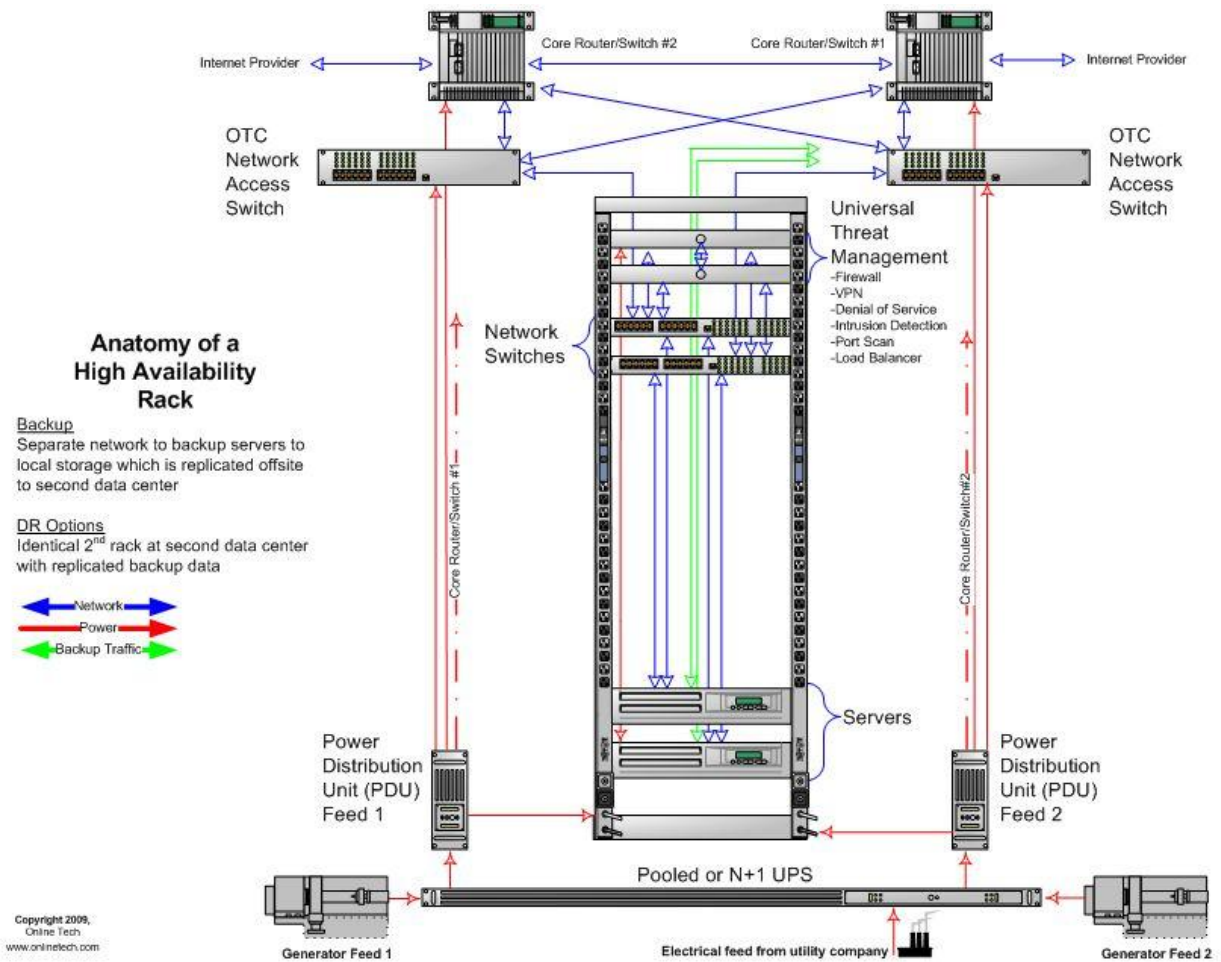


Figure 1 - The Anatomy of a High Availability Rack

4.1 Power Distribution

The high-availability rack is designed to withstand any single point of power failure. A server mounted in the rack with two separate power supplies can withstand a failure anywhere in the power infrastructure - from a power supply to generator failure - and continue to operate without an outage. This is accomplished with “dual home run power,” meaning two separate, independent power runs from the server to the utility power source and backup generators are in place to deliver power to two separate power supplies in the server.

Starting from the server itself, let’s explore the anatomy of high-availability power:

- The server power supply is the most common point of server hardware failure - a mechanical failure in the power supply fan can cause the power supply to overheat and shut down. To avoid this type of failure, high-availability servers should have redundant power supplies.

- Each power supply on a server is plugged into separate power strips in the rack. Power strips with digital amp load current readouts aid in monitoring power levels and help avoid tripping a circuit breaker, which would shut down the entire power strip.
- Each power strip in the rack is connected to different PDUs (Power Distribution Units), which are distributed across the data center. PDUs are central points where uninterruptable power is protected with circuit breakers avoiding overload and distributing power to racks across the data center. By delivering power to an HA rack from two separate PDUs, another level of physical redundancy is built into the power infrastructure.
- Each PDU is connected to separate Uninterruptable Power Supplies (UPS) or banks of pooled N+1 Uninterruptable Power Supplies, which cleans and distributes the power and provides backup power through a bank of batteries in the event of a power outage. The clean power from the UPS is stable; therefore, any fluctuation in power, both power surge and brown out, is regulated by the UPS.
- Each UPS is fed with one or more power feeds from the utility company. The utility power feed is wed to two generators that run on either diesel or natural gas. If utility power is lost, the UPS maintain stable power to the racks while the generators kick on to provide backup power.

With dual home run power, the high availability design delivers a completely redundant power infrastructure from the utility service and backup generators to the servers. If one point of power distribution should fail, a backup source exists to power servers, leaving no single point of power failure that can take the servers offline.

4.2 Network Distribution

The network connectivity in a high-availability rack is designed to replicate the same redundancy as the power distribution so the network and Internet connectivity offer no single source of failure.

Starting again from the server, let's look at the anatomy of high-availability network connectivity:

- Each server in an HA application should have two separate Network Interface Cards (NICs) that allows the server connection to the redundant HA network infrastructure.
- Each NIC in the server is connected to different network switches, which disperse the network connectivity to all servers contained within the rack.
- Each network connection is connected to a firewall or more advanced Universal Threat Management (UTM) device, which protects traffic on each segment of the network from intruders and security threats.
- Each firewall or UTM connection is connected to separate Cisco routers and network access switches.
- The two core routers are fed from multiple Internet Service Providers (ISPs) and cross-meshed between both routers and Online Tech's network access switches. This provides two diverse network paths from multiple ISPs to each HA rack.

Dual home run networks, much like high availability power, delivers a completely redundant network connection to multiple Tier 1 Internet service providers. Should any point in the network fail, a redundant network is in place to automatically deliver the network traffic from the server to the Internet.

5. Increased Redundancy and Availability Beyond a Single Data Center

High-availability rack architecture, with redundant power and network connectivity, eliminates the majority of single failure points within a data center. In fact, when Online Tech's clients select an HA rack design, Online Tech enhances its power Service Level Agreement (SLA) from 99.9% to 99.99% uptime guarantee.

However, despite the best data center design, any single data center is still at risk of a disaster (such as fire, earthquake or tornado) that goes beyond the redundancy built in the infrastructure.

Connectivity across multiple data centers can add an additional level of availability when a disaster strikes. An example of this can be seen below with the architecture that Online Technologies deploys across its multiple data centers.

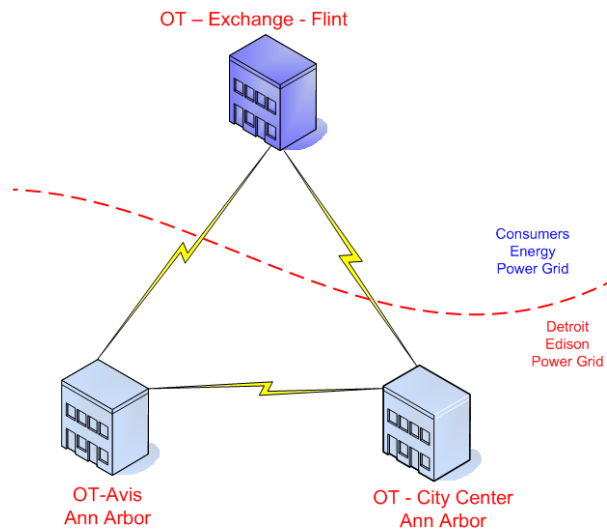


Figure 2 - Online Tech's Cross-Data Center Infrastructure

The infrastructure across multiple data centers should be examined by power, network connectivity, and geographic protection.

- Placing multiple data centers across different power grids provides an additional layer of protection between the primary (production) and disaster recovery (DR) deployments. If utility power is lost to the production data center, utility power is most likely unaffected at the DR site. In Online Tech's example, the Online Tech Flint Data Center was completely unaffected by the August 2003 power outage that struck

the entire east coast, while its other two data centers went to generator backup to keep all three data centers online.

- High speed dedicated gigabit fiber between each of the data centers provides an additional layer of redundancy for Internet connectivity. With multiple Tier 1 Internet Service Providers connected across multiple data centers, the Internet connection automatically fails-over to another provider should any ISP or its connection fail.
- A distance of over 50 miles between data centers provides good geographic separation between the production and DR sites should a catastrophe strike one of the data centers.

6. Preventative Maintenance and High Availability

For a business that relies on online connectivity, it cannot afford to be down, not even for preventative maintenance. However, if preventative maintenance is not performed, the risk for a failure increases.

Preventative maintenance should be done on servers to avoid future downtime. A high-availability design allows the necessary preventative maintenance on your equipment, while still maintaining customer access to your site.

With two servers, properly configured, one server can be shut down for preventative maintenance, while customer traffic is maintained by the other server. For example, if a business has two network switches in their rack and all their servers are connected to it, they can unplug one server, do preventative maintenance on it, plug it back in then do the same with the other switch without impacting servers' uptime. That's high availability.

7. Conclusion

High-availability architecture can greatly reduce the risk of losing business due to a disaster or other anomaly in a leg of the power infrastructure or network connection to the Internet. With the proper design, single failures in the power or network infrastructure will not affect your server uptime.

Redundant power and network connectivity help ensure a greater level of protection against any single point of failure in the power delivery or network connectivity to your servers resulting in a higher uptime service level. A high-availability rack can help businesses survive network or power failures, keeping computers online and maximizing the potential to generate revenue.

8. Always-On, Always-Online with Online Tech

Online Tech offers high-availability solutions to its clients to increase their uptime availability. In fact, our high-availability service-level agreement (SLA) provides a 99.99%¹ warranty on power uptime and 99.99% uptime warranty on the core network.

1

The anatomy of the rack is designed to provide businesses, like yours, that rely on the Internet and their network to drive their business. These businesses cannot afford to be offline. Therefore, a high-availability package with Online Tech provides the peace-of-mind to know your business will be always-on and always connected to the Internet.

In addition to high-availability solutions, Online Tech offers backup and disaster recovery (DR) options that span across our multiple data centers. A replicated server can be made available at a different data center than the production server, which resides in a different data center on a different power grid, providing an even higher level of redundancy.

Online Tech (www.onlinetech.com) is one of the Midwest's Premier Managed Data Center Operators. Online Tech helps companies manage their growing demand for data and computing capacity through its SAS 70 Type II audited secure and reliable multi-tenant data centers. With a full range of managed colocation and managed dedicated server offerings, industry leaders trust Online Tech to insure their servers are always on, always online, and always safe.

To find out how we can help you deliver a high availability application, call us at today at (734) 213-2020.