Web site load-balancing solutions

Understanding and evaluating availability systems for Internet presence

PRELIMINARY DRAFT.
This content is preliminary and is still under validation. Check with Networkshop before using this information in budgetary or technical evaluations. Please report any inconsistencies to info@networkshop.ca.

Executive summary

As the Internet grows, e-businesses face two main challenges: robust availability and affordable scalability. Load-balancing devices deliver both of these by making clusters of servers look like a single server. This mitigates the impact of a single server's failure, and allows managers to roll out more, slightly slower, servers at great savings.

But load-balancing systems are complex. They include a WAN and a local bind phase that must perform sophisticated health checks in order to ensure that visitors aren't sent to failed systems. They include scheduling algorithms that factor in proximity, responsiveness, load levels, and administrative thresholds. And they can be deployed in a vast array of network architectures.

This report will help you navigate the complex—but absolutely essential—technology of load balancing. Over a period of six months, Networkshop analyzed offerings from the twelve leading vendors in this space. This involved hands-on testing in our Montreal lab, vendor interviews, research, and detailed questionnaires. The report presents a high-level model of load balancing, and then looks at ways in which online systems can go awry. Finally, we look at all twelve vendors in detail.

Contents

Introduction

This report takes a detailed look at systems that balance the load of many visitors across all components of a service. A service might be an online store, a travel reservation system, or an electronic bank. At the same time, these systems ensure that the service remains operational even when some of its functions stop working.

Fundamental feature set

In a TCP/IP network, the relationship between client and server is fundamentally one-to-one. This makes it hard to scale a network service beyond a single node. Load-balancing systems share a load equitably across more than one device. This capability in turn makes it possible to deploy many cheaper systems rather than a single, high-priced, monolithic system.

At the same time, the one-to-one relationship makes it difficult to ensure a highly available service. In order to ensure that a service is always available to visitors, single points of failure must be identified and eliminated. In addition to distributing load, a load-balancer must therefore also offer high availability by detecting failures and correcting or working around them.

These two goals—equitable load sharing and granular per-user determinism—are fundamentally at odds. Distributing personalized, one-to-one connections that ensure consistent, reliable behavior across a session is hard to do when a load is smoothly distributed across a service at a number of locations on a number of servers. At a high level, then, load-balancing systems must balance the granular needs of individual users against the macroscopic, coarse needs of an even load distribution.

To achieve these goals, load-balancing systems must track the health of all possible locations for the service. Knowing this health, they need to select the optimal location from which to fulfill a particular query, and bind the visitor to that server farm.

Within the farm, the system must select the optimal server and bind the visitor to the specific server. For the duration of the session, the system must maintain this binding while overcoming failures in the services, software, and network.

How this report is organized

This report is a detailed look at load-balancing, high-availability systems for e-business deployment. In preparing the report, Networkshop studied and met with leading vendors of load-balancing technology and put their devices through rigorous in-house testing using our own testing environments and equipment.

We begin with a discussion of availability technologies in general and the evolution of web sites in particular. Armed with this understanding, we consider what load-balancers do by looking at the life cycle of an individual connection. We then discuss in detail the ways in which a service can fail, and techniques that mitigate such failure.

In the next section of the report, we look at ways of evaluating the various high-availability load-balancing appliances on the market. We present background information on our testing procedures. We then discuss the various ways in which load-balancing and high-availability systems might be deployed; this is critical because different vendors' solutions may be better suited to specific environments. We then look in detail at the results and characteristics of each vendor's offering.

We also present each vendor's response to their results in a detailed appendix, so the reader can consider Networkshop's findings in the context of the vendor's own position.

Disclaimers

The information contained in this report is believed to be accurate at the time of publication. In preparing this document, we have taken reasonable care to present each vendor with their evaluation and ensure that our information was correct. In the event of an error, we have made corrections; in the event of a disagreement, the complete text of their response is available at the end of the report.

Networkshop, Inc. believes that the sources cited herein to be reliable, but we do not guarantee their correctness. Should our views on the subjects or statistics contained herein change, we do not promise to notify you of such changes; subscriptions are available for this purpose. Networkshop, Inc., its employees, affiliates, or partners—or members of their families—may perform services on behalf of, and/or hold equity positions in, and/or engage in business with, one or more of the companies referred to in this document, or their competitors. Networkshop, Inc., shall not be liable for errors contained herein or for incidental or consequential damages in connection with the publication, interpretation, or use of this material.

Credit and thanks

This report would not be possible without the active participation and support of the vendors covered herein. It takes a good deal of trust and commitment to allow one's products to be evaluated by an independent lab, and to have those results published. We would like to thank ArrowPoint, Cisco Systems, F5 Networks, Holontech, Ipivot, RADWARE, Phobos, and Resonate for the loan of equipment and time, and Alteon, Coyote Point, Hydraweb and Foundry for responding to product queries. The testing harness used in the report also relies heavily on code from the Open Source community, which deserves our respect and admiration.

About Networkshop, Inc.

Networkshop is a research consultancy focused on e-business infrastructure. We specialize in hands-on, real-world deployment of the various components that make e-business run—from security and site design to back-end systems and global load balancing.

Based in Montreal and Ottawa, Networkshop's staff of engineers, technicians, and researchers advises Global 2000 companies, innovative start-ups, and major networking vendors on emerging trends in networking technology. Our labs offer e-commerce readiness testing services, from load-generation to availability and performance testing. We also provide detailed consulting on market positioning and product creation.

Alistair Croll and Eric Packman founded Networkshop in 1997.