COST-BENEFIT ANALYSIS: HIGH AVAILABILITY IN THE CLOUD

AVI FREEDMAN, TECHNICAL ADVISOR

a white paper by servercentral®

COST-BENEFIT ANALYSIS: HIGH AVAILABILITY IN THE CLOUD

Is the first question you ask when evaluating IT infrastructure, "how much is this going to cost?"

This white paper presents a cost-benefit analysis to help you make choices involving adding redundancy and high availability (HA) to your IT infrastructure. In it, we cover options for deploying fault-tolerant applications and servers via cloud and dedicated hardware, in single- and multi-site configurations.

Dedicated infrastructure can perform up to 20% better than cloud-based solutions. This performance improvement is typically attributed to the overhead of running a virtualization layer.

Hybrid infrastructures help minimize costs by utilizing dedicated servers for some components (e.g. databases requiring high IOPS and shared compute resources for less critical workloads).

TABLE OF CONTENTS

THE COST OF FAILURE

Costs incurred in making IT infrastructure decisions are based on your willingness to tolerate risk. We often see that the customer cost of a failure far exceeds the price tag of implementing High Availabilty (HA) in the application architectures and underlying infrastructure. Lack of HA, especially the lack of redundancy from the application layer down, can result in lost data and decreased availability of critical services.

Infrastructure reliability determines the amount of downtime that can be expected in an infrastructure over a given time period. Over the last 15 years, we've consistently seen that infrastructures engineered for higher reliability incur lower overall costs associated with downtime in production.

This seems intuitive, and it is. However, we still work diligently to help companies recognize the long-term costs associated with a non-HA decision.

WHAT, EXACTLY, IS THE COST OF FAILURE?

The matrix below illustrates the risks and costs associated with various HA implementations for an enterprise that generates $100,000 in revenue per day. We're using "cost risk" numbers below that model an average of the conversations we've had with our customers running e-commerce sites.

In some cases, these numbers are optimistic—especially in the case where single points of failure lead to prolonged periods of mini outages or data loss.

	MTBF	MTTR	COST	RELIABILITY	EXPECTED DOWNTIME RISK
DEDICATED, NON-REDUNDANT	2 YEARS	DAYS	MEDIUM	99.0%	$365,000
REDUNDANT, DEDICATED	5-10 YEARS	MINUTES-HOURS	HIGH	99.9%	$3,000
NON-REDUNDANT CLOUD	6 MONTHS	MINUTES	LOW	99.0%	$365,000
REDUNDANT CLOUD	5-10 YEARS	MINUTES	MEDIUM	99.9%	$3,000
REDUNDANT, CLOUD, MULTI-SITE	5-10 YEARS	MINUTES	HIGH	99.999%	$400

MEAN EXPECTED UPTIME VS. DOWNTIME COSTS

The costs attributed to failure not only affect revenue, they impact productivity. Infrastructure remediation strains internal resources requiring staff to address technical problems on top of day-to-day operations. These critical productivity costs are rarely calculated into the TCO for infrastructure Outage durations will also have a direct effect on the organization's current and future revenue stream.

THE HIGH AVAILABILITY DECISION PROCESS

Determining the right level of redundancy or HA for the infrastructure supporting your applications is a calculation unique to your organization. The critical decision factor is the cost of implementing redundancy or HA versus lost revenue, reputation, and internal staff time associated with downtime.

When determining the infrastructure solution that is right for your enterprise, two key factors to account for are are Mean Time Between Failures (MTBF)—how likely it is that components will fail—and Mean Time To Resolution (MTTR)—how long outages are likely to last. HA solutions can increase reliability while reducing overall risk—typically at very little additional cost.

LEVELS OF HIGH AVAILABILITY IMPLEMENTATION

Single site redundancy solutions can provide for always-available infrastructure - as long as the power and network to the single site remain active. However, best practices for most enterprise applications are:

- Engage partial offsite backups with a high MTTR (e.g., full offsite DR capability with access to data and enough resources to continue in or scale to production traffic)
- A fully active-active, multi-site configuration
- A properly architected multi-site infrastructure shields supported applications and users from catastrophic failure such as hurricane, earthquake, regional power failure, or loss of an entire data center.

Hurricane Sandy caused several of New York City's largest data centers to fail. The application downtime caused by this event didn't have to happen.

CONFIGURATION MANAGEMENT

One other leading cause of outage is human error. The best HA implementations involve, at a minimum, a basic level of change and configuration management.

Over the last 15 years, tools have emerged that allow any company to take advantage of the same processes and software that enable large web-scale companies to operate with always-on delivery of content and services.

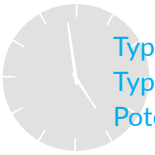
The leaders in the configuration management space are Chef and Puppet. CFEngine, and more recently, Salt and Ansible are also viable alternatives.

The key benefits of using configuration management tools are that all changes are tracked, systems are configured according to an authoritative state and software database, and any mistakes can be quickly identified and rolled back.

An investment in configuration management can have a very high ROI and will often cut the time to recover from human-induced availability issues by orders of magnitude.

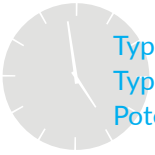
HA COSTS FOR DEDICATED INFRASTRUCTURE

Non-redundant, dedicated infrastructures result in low MTBF and MTTR. In the event of a component failure, the infrastructure becomes completely unavailable. MTTR increases significantly due to the requirements of physical hardware provisioning for recovery.



Typical MTBF: 2 Years
Typical MTTR: Days
Potential Risk: $365,000

Dedicated infrastructures with redundant components enjoy increased MTBF through availability of applications and services in the event of a discrete component failure. The application infrastructure will continue to service the end-user while repairs or maintenance are performed.



Typical MTBF: 2 Years
Typical MTTR: Days
Potential Risk: $3,000

HOW MUCH REVENUE AND CUSTOMER LOYALTY IS LOST IN TWO DAYS OF DOWNTIME?

In a dedicated hardware configuration, an e-commerce site realizes maximum performance due to the nature of the dedicated hardware. The site will realize both scheduled and unscheduled downtime during all component and maintenance windows. Site upgrades will result in a direct loss of revenue, as services will be unavailable.

Adding redundancy to dedicated infrastructure incurs a cost equal to the initial, non-redundant deployment. This is a simple calculation as identical physical systems are purchased and deployed. Additionally, because of the fully redundant systems, this architecture will require the addition of a dedicated network switch to manage bandwidth and routing.

A dedicated-redundant configuration provides for maximum performance and uptime leveraging direct attached storage (DAS) while reducing network latency and storage area network (SAN) overhead. A DAS solution increases performance of the overall application through decreased latency and increased bandwidth to the local data. However, the capital and operational expenditures necessary for this configuration directly impact the Cost of Goods Sold (COGS). IOPS-intensive applications can perform significantly faster in a dedicated SAN environment through increased IOPS and transfer rate and lower latency.

In a dedicated-redundant configuration, an e-commerce site performs optimally, supporting uptime during component and maintenance windows, as well as failover support. The critical success factor becomes the cost per transaction, as there is over a 2x multiplier for this configuration. Infrastructure cost must be monitored closely for continued profitability on products and services sold.

To put this in perspective, a common 99% SLA will result in an average of 41 minutes of downtime per month.

You can learn more about the actual value of SLAs's here: *http://www.servercentral.com/the-9-print/*

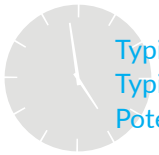
HA COSTS FOR CLOUD INFRASTRUCTURE

Redundant cloud infrastructures eliminate single points of failure (SPOF) within the infrastructure at significant cost savings.

HA cloud infrastructures with CMS integration enjoy the benefits of dynamic elasticity, low latency, and more efficient use of resources.

Typical MTBF: 5-10 Years
Typical MTTR: Minutes
Potential Risk: $3000.00

A cloud infrastructure without redundancy will affect application and service availability due to the SPOF in each tier of the infrastructure. In the event of a failure on a discrete component, application availability is lost. A replacement resource can be redeployed on the cloud in less time when compared to a dedicated environment, but application availability is still affected.



Typical MTBF: 6 Months
Typical MTTR: Minutes
Potential Risk: $365,000.00

HOW MUCH REVENUE AND CUSTOMER LOYALTY IS LOST IN MINUTES OF DOWNTIME?

In a single-site cloud deployment, redundancy is achieved through dual servers and multi-path network connections. This configuration provides high availability in the event of an underlying hardware failure.

This solution provides reliability and redundancy at reasonable cost. The e-commerce site will perform well with throughput that can handle high traffic with little latency. Cloud infrastructures access centralized block storage on a SAN to provide high-availability of data and workload isolation in the storage pool. Cloud infrastructures deployed with best practices will also enjoy the ability to dynamically provision additional webservers in minutes during a traffic surge.

Dual-site redundant cloud infrastructures offer extremely high reliability (99.999%). This is currently considered the best in class architectural practice. It is not uncommon for this configuration to provide near zero downtime in a given year. This service level is the result of a duplicate configuration deployment at the secondary data center that is capable of operating independently of the primary data center. This dual-site configuration provides high availability in the event of application performance and/or connectivity loss to either of the two data centers.

The e-commerce site architecture in this solution provides the highest level of reliability and is able to tolerate a catastrophic outage. During natural disasters and loss of an entire data center, business continuity is insured. Redundant cloud infrastructures provide increased MTBF and decreased MTTR while decreasing overall cost.

Redundant cloud architectures offer near zero downtime and can decrease or eliminating site availability during maintenance windows.

Redundant cloud architectures offer near zero downtime and can decrease or eliminate site availability during maintenance windows.

CONCLUSION

High availability infrastructure implementations increase MTBF and reduce MTTR, reducing the overall risk of downtime and outages while maximizing business continuity by eliminate SPOF in the infrastructure and providing the benefits of ease of maintenance, scalability, and protection from catastrophic failures.

Determination of the proper level of redundancy for your infrastructure will help your organization mitigate risk and opportunity cost.

The actual business continuity impact is measured in dollars as opposed to uptime. So ultimately, key decisions are financial ones that then drive technical implementations. We have seen that in e-commerce, downtime makes a real difference, especially if customers see repeated inability to access sites or applications.

MOVING FORWARD

If you would like to discuss this topic further, or to get a quick understanding of the costs involved in deploying high availability and redundant IT architectures, give us a call. We welcome the opportunity to learn about your business and IT environment, and will be happy to share our experience and help you evaluate the trade-offs involved.

servercentral®

COST-BENEFIT ANALYSIS:
HA IN THE CLOUD

MEET AVI FREEDMAN, SERVERCENTRAL TECHNICAL ADVISOR

Avi Freedman advises on ServerCentral's cloud and VMware platforms. He founded the first ISP in Philadelphia and was previously featured in Forbes magazine.