CONTINUITY SOFTWARE

Five Critical Recovery Flaws Your Last DR Test Missed

A look at why traditional DR testing fails to uncover serious disaster recovery gaps and what it can mean to your business

Introduction

Businesses today are spending millions of dollars to develop and maintain disaster recovery (DR) infrastructures that will ensure business continuity. Despite such huge investment of time and resources, most IT professionals are still not confident in their ability to recover in an emergency.

With industry analysts citing DR failure rates of at least 60 percent, there's a good reason to be concerned.

Operating in an environment that is increasingly complex and dynamic, IT managers at large organizations are challenged to keep up with disaster recovery goals.

These challenges are further compounded by the limitations of traditional DR testing.

At a time when businesses are under mounting pressure to ensure continuity and minimize data loss, IT organizations have no way to accurately verify and measure whether their DR plans will actually work when they need them.

The Failure of Disaster Recovery Testing

The Theory

A DR test should emulate how well business operations could be transferred to a remote facility to get the organization back online within a specified Recovery Time Objective (RTO) and Recovery Point Objective (RPO).

A good DR test requires considerable advance planning, along with a sizable investment in time and resources. Large numbers of people in the IT organization need to be involved. Network and storage resource mappings must be reconfigured; not just once but twice, first for the test and then again to restore normal operations. And to simulate a real disaster – which is the only way to truly determine how well the DR strategy works – mission-critical applications must be shut down during the test, a step which most businesses are loathe to take.

If any of the tests fail, the team has to pinpoint the problem, fix it, and repeat the process.

The Reality

DR tests are difficult, costly and complicated. Most IT organizations run lean and don't have the time or resources to execute a complete, by-the-book DR test. More so, simulating a disaster is a dangerous proposition; upon completion of the test, IT professionals often hold their breath, hoping that production can be flawlessly resumed. Facing such concerns and limitations, it's no wonder the scope of DR tests is frequently minimized. Common DR testing shortcuts include:

- Testing just portions of the infrastructure rather than testing the full DR environment
- Keeping key production components (storage, database, application management, domain name, or file servers) online while performing the test
- Conducting an orderly shutdown to protect production systems, rather than simulating the abrupt cessation of operations that would occur in a disaster
- Testing failover servers but not applications
- Testing applications but not simulating the actual load the application must bear following a full site recovery
- Neglecting to test dependencies, data inconsistencies, and mapping errors that may exist between SAN devices and hosts, or any other errors that can cause a recovery failure.

Given how common such shortcuts are, in the end, most test results are at best incomplete and at worst worthless.

Configuration Changes: The Monster in the Closet

Even the best DR test can only evaluate recoverability at a certain point in time. However, configuration changes are a daily occurrence in today's complex data center environment. Even a small change can create a configuration gap between the production and DR environments that will cause a recovery failure. Since there is no way to easily assess the impact a change may have on other components of the environment, any test results are thrown into question the moment a change is made.

Even when a test is conducted according to standard best practices, the number of gaps and errors it can miss is significant enough to pose a business risk that is further compounded given that most organizations have neither the time nor the resources to perform complete DR tests.

These risks can be classified into two categories:

Data Protection Risks: Application data, metadata, and data links can be jeopardized by gaps in replication, setup, sequence of procedures, accessibility, mapping, zoning, and other elements, resulting in data loss and potential RPO violations.

Availability Risks: Standby hosts, DR servers, and cluster members may be unable to fulfill their role because of erroneous configuration, extending recovery time and potentially resulting in RTO violation.

In this guide, we take a closer look at five common errors that often go undetected. We'll explore why they occur, why a DR test fails to catch them, and their potential business impact.

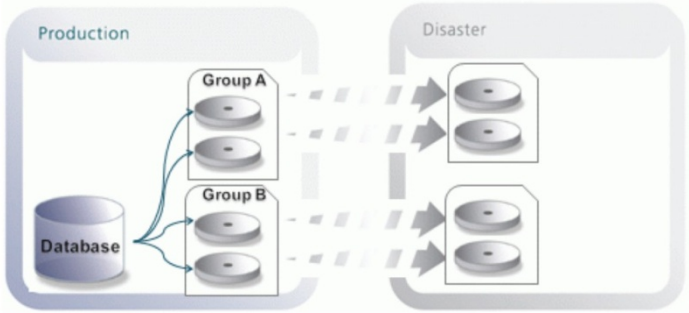
#1: Replication Inconsistencies

Risk

Data loss and increased time to recover (RPO and RTO violations)

How does it happen?

This is a common gap found in large SAN environments (for example EMC SRDF/S and SRDF/A with multiple RDF groups. It occurs most often when storage volumes from different consistency groups are provisioned to the host and used by the same database or file system (provisioning tools do not alert or prevent this configuration). With most high-end storage arrays, each consistency group is typically associated with different replication adapters and potentially different network infrastructures. Rolling disaster scenarios can result in corrupted replicas at the disaster recovery site as different groups fail one after the other



Synchronous Replication Inconsistency—RDF Group

Why will the DR test miss this?

During a DR test, an orderly shutdown of applications, databases, and hosts will leave data in a consistent state. However, most real-life disasters are characterized by rolling, gradual failure that bring systems or network elements down one by one, as opposed to abrupt and immediate cessation. Such rolling disasters are extremely difficult to emulate in a DR test.

What is the impact?

In a rolling disaster, network components will not fail at exactly the same time, resulting in one RDF group being out of sync with other RDF groups. This will irreversibly corrupt the database at the disaster recovery site. Data will need to be restored from a recent backup, putting both RTO and RPO at risk.

Note: Many companies experience this problem but incorrectly assume it is the result of some network abnormality. However, unless the issue is properly diagnosed and corrected, it will reoccur.

#2: Missing Network Resources

Risk

Extended recovery time, potential data loss (RTO and RPO violations)

How does it happen?

This risk can typically be traced to a configuration mistake when DR considerations are not thoroughly considered during the configuration process. For example, a production host is accessing network file systems (using CIFS or NFS) stored on a production file server or NAS device, while its DR peer also accesses the same resource, instead of a replicated copy that should exist on the recovery site. If during the DR test the production file server is not brought offline, the test would succeed. In a real disaster, however, the file server will not be available.

Why will the DR test miss this?

When running the DR test for a specific business service or application, most companies do not shut down the entire production datacenter. The DR test will result in a false positive because production assets are still accessible and responding,

What is the impact?

If the network file systems were not replicated to a DR site, data loss will result. Even if the systems are replicated, recovery time will be extended while the administrator locates corresponding file systems on the DR site and mounts them on the DR standby server.

Note: For simplicity sake, the discussion here is limited to network file systems; however in reality this risk can manifest itself in any network service.

#3: Tampering Risk

Risk

DR failure and data corruption (RPO violation)

How does it happen?

This hidden risk is the result of an unauthorized host at the DR site erroneously configured with access to one or more storage devices. This is a very common error, and, much to the surprise of many organizations, there are dozens of reasons why it can happen. In each case, however, it can remain dormant during normal operations and is only revealed during an actual full-blown disaster. Here are just a few reasons why this error may occur:

- When performing a storage migration, the storage administrator may forget to remove old device mappings to the host. After repurposing the old devices to new hosts, some of these devices may still visible by the original, now unauthorized host.
- Incorrect zoning and masking configuration could easily result in the wrong host getting access to other's data. Zoning and masking is a complex process and in the large data canter it's not uncommon to find devices that were mapped incorrectly due to typos or a slip of the finger when using Storage Resource Management tools (e.g., EMC ECC, HiCommand).
- Sometimes HBAs are replaced not because they are faulty but because greater bandwidth is required. If soft-zoning is used and not updated accordingly, an old HBA still retains permission to access the original storage devices. Once the HBA is reused on a different host (which can occur months after the upgrade), this host will erroneously be granted access rights to the SAN devices belonging to the original host.

Why will the DR test miss this?

Many organizations choose to test only a subset of the environment at a time. During the test, the original and unauthorized servers may not be started at the same time. In a real disaster they would, wreaking havoc on the data.

What is the impact?

During a disaster, a racing condition will develop, with several potential scenarios:

Scenario A: The unauthorized host might gain exclusive access to the erroneously mapped disk. In this case, the designated standby will be unable to mount and use the locked devices, and it could take some time to isolate and fix the problem..

Scenario B: Both the standby and the unauthorized hosts get concurrent access to the disk. If the unauthorized host attempts to use the erroneously mapped disk, not only will the data be corrupted, but the now-active standby may unexpectedly crash.

#4: Point-in-Time Copies Never Tested

Risk

Data loss and increased time to recover (RTO violation)

How does it happen?

Point-in-time copies such as snapshots and BCVs are the second line of defense against human errors, viruses, and outages. The DR configuration for applications typically includes:

- Multiple local point-in-time copies such as EMC TimeFinder, HDS ShadowImage/Snapshot, NetApp FlexClone/Snapshot, or CLARiiON SnapView;
- Remote synchronous replication such as EMC SRDF, Hitachi TrueCopy, CLARiiON MirrorView, and NetApp SnapMirror; and
- Local point-in-time copies on the remote site.

In addition, the copies could be mapped to the target DR servers, configured with multi-path software such as EMC PowerPath, Veritas DMP and MPIO, and defined in logical volumes such as Veritas VxVM.

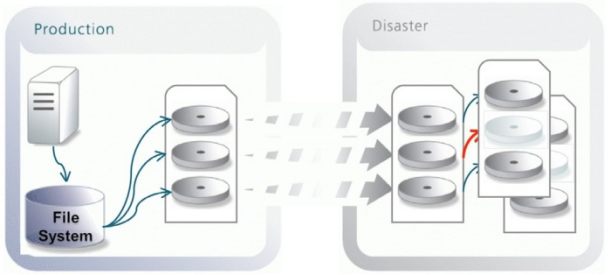
Corruption of point-in-time copies will remain undetected unless the application is fully started and data integrity is thoroughly tested. The diagram below illustrates a file system replica corruption caused by data-age inconsistency among replica devices. There are numerous scenarios that can lead to corruption; for example, when the replica devices do not all belong to the same consistency group.

Why will the DR test miss this?

Few organizations would go beyond testing the primary (synchronous or asynchronous) copy, as validating each additional copy would double testing time and wrongfully assuming that if the primary copy is valid, all other point-in-time copies must also be valid.

What is the impact?

With the replica corrupt and unusable, the file system will need to be recreated at the disaster recovery site and data restored from a recent backup, increasing time to recovery. Worst of all, data created since the last backup will be lost. In many cases, corrupted file systems may still be accessible, and only a close inspection of the content will reveal the data is invalid.



Replication Age Inconsistency

#5: Insufficient DR Resources

Risk

Extended recovery time (RTO violation)

How does it happen?

When building a DR data center, organizations tend to assign fewer resources than their production environments have. If the DR configuration includes significantly fewer resources than the production environment, it may be unable to properly assume production upon failover. For example, it is not unusual to find a production environment that has multiple paths to storage or software while the DR environment has fewer, or even just a single path. DR sites with insufficient memory or CPU to support full production load are also a common occurrence.

Why will the DR test miss this?

Most DR tests do not simulate full production load, so these errors remain undetected. Since DR is mostly offline, this issue never comes to light until an emergency occurs.

What is the impact?

When the DR site cannot assume production as planned, business operations cannot resume in accordance with the company's established SLA.

Uncovering the Hidden Monsters with Automated DR Testing and Monitoring

Periodic DR tests and manual audits will always be an important part of any DR strategy. They can uncover important flaws in processes, procedures or technology that could impact readiness. However, it would be foolhardy to ignore the serious drawbacks of traditional testing that leave critical applications and data unprotected.

That's why many leading organizations today are augmenting their DR strategies with automated auditing and risk detection tools that can identify vulnerabilities and allow correcting them before they impact business operations, ensuring continuous protection and readiness that goes beyond point-in-time testing.

Automated DR monitoring technology can penetrate deeply into the environment to ensure the infrastructure is consistently aligned with protection goals. DR Management software can perform tasks that are too time-consuming or complex for humans to carry out.

This is the equivalent of performing millions of manual test points, daily.

While traditional DR testing provides the IT organization with valuable insights, only automated DR auditing and monitoring can enable true DR readiness.

How RecoverGuard Detects Hidden DR Vulnerabilities

RecoverGuard automated DR and HA testing and monitoring help you ensure 24x7 business continuity by verifying that your production and DR environments are always in sync and detecting errors before they impact your operations. When a risk is identified, RecoverGuard provides a detailed description and suggested remediation, empowering your business continuity and IT teams to proactively and collaboratively resolve the issue.

Leading organizations use RecoverGuard to validate and measure disaster recovery SLAs, RPOs and RTOs across their entire IT infrastructure, transitioning from point-in-time DR and HA testing to automated, uninterrupted business continuity assurance.

For more information

Website: www.continuitysoftware.com
Email: info@continuitysoftware.com
Tel: 1-888-782-8170 or +1-646.216.8628

About the Author

Doron Pinhas is a DR expert with over 20 years of experience in data and storage management, high availability, real-time applications, operating system design and development, open systems, and networking architecture engineering. He has served as Continuity Software's CTO since joining the company in 2005. Previously, Doron was a driving force at Xpert Integrated Systems, a leading Israeli system integrator, first as its Chief Operating Officer and later forming and heading its Business Continuity Solutions division. Prior to joining Xpert, Doron served in the Israeli Defense Force for 10 years as a system architect for mission critical information systems, retiring with the rank of Major.

CONTINUITY SOFTWARE