

WHITE PAPER

SUBJECT: High Availability for Mission-Critical Applications

TITLE: Force Majeure: Understanding and Managing the Risk of Data Center Failure

RELEASE DATE: October, 2008

SPONSORED BY: Ceryx Inc.

ABSTRACT:

In the last year some of North America's top data centers experienced failures and outages that impacted thousands of businesses and compromised the delivery of mission-critical applications. In many cases these failures were beyond the control of data center operators and fell under the jurisdiction of 'Force Majeure' – a common caveat in Service Level Agreements that refers to an 'Act of God' or 'Superior Force'.

This paper examines these causes, ranging from environmental to software-related issues, and the resulting financial and productivity impact it can have on today's businesses.

From the perspective of the delivery of mission-critical, business grade-email, Ceryx shares its insight and speaks of their development efforts to deliver the highest availability possible. Accepting the short-comings of modern data centers, Ceryx advocates the use of data replication across two physically remote data center to mitigate the risks and exposure that are inherent with a single data-center strategy.

The year 2008 began with a dire prediction from Subodh Bapat, a vice president in the eco-computing team at Sun Microsystems, when he declared, "You'll see a massive failure in a year." He went on to say, "We are going to see a data center failure of that scale," referring to the worm that took down 5% of worldwide UNIX boxes in 1988.¹

This time he isn't citing security lapses as the root cause but rather failure caused by the massive computing power required to run today's applications.

Though certainly an extreme position, the past year has seen a rash of data center failures that brings into question how reliable single data centers are for the delivery of mission-critical applications.

Vulnerabilities ranging from the most common, like natural disasters and infrastructure failure (data center power outage, burst pipes, construction work damaging fibre lines,) to hardware failure, storage or database failure and common software problems, have been causing regular disruptions to businesses and come with a high price tag.

Recent events in the news support the fact that even with good planning, resourcing and design, some of the most sophisticated facilities can still experience catastrophic failure.

Last summer, the state-of-the-art 365 Data Center, in San Francisco – built with more than \$125 million - was offline for hours due to a power grid outage by Pacific Gas & Electric that put a significant portion of San Francisco in the dark. Subsequently, the backup generators at the facility also experienced failure and had to be manually started.²

"When researching data centers, new facilities often boast N+2 levels of redundancy," says Roger Smith, V.P. of Operations at Ceryx Inc. "However, as these same facilities fill up and age, that often becomes N+1, or in some areas no redundancy at all."

According to Sun Microsystems Executives, the typical life span of a data center is only about 10 to 12 years and many data centers - built at the beginning of the dot-com era – now need to be rebuilt.

"As the person who is accountable for uptime I have to balance which applications are considered critical by upper-management and clearly communicate the cost and investment required to provide high-availability," says Roger Smith. "When you present the facts, it becomes clear to everyone that an in-house data center couldn't possibly provide the levels of redundancy required and even on a co-location level we would need redundancy."

In many cases, no contingency plan could avoid the issues that plague individual data centers. On July 14th of this year, the Peer 1 data center in downtown Vancouver – one of the largest facilities in Canada – was offline for almost an entire day. An underground fire caused massive power outages throughout downtown Vancouver. While backup generators at Peer 1 started without issue, the water-based cooling system failed as firefighters – in their attempt to douse the fire - depleted the water pressure required to keep the cooling systems operational. This caused the backup generators to overheat and any failover to UPS was limited to a short battery life.³

In a similar event this summer, The Planet, a prominent hosting provider in Houston, experienced a major explosion in their data center, taking more than 9000 customer servers offline for several days. Backup generators worked perfectly, but again the fire department would not allow the facility to resume power until it was deemed safe. In some cases servers were physically migrated to a new facility.⁴

In the aftermath of this disaster the Planet was applauded for their response to the crisis; allocating every resource they could to address the problem and proactively communicating status reports and issuing SLA credits.⁵

Google, whose Enterprise App customers experienced multiple outages on August 6th, 11th and 15th of this year, took a more reactive stance, promising to build a communication dashboard and issuing a blanket credit for all customers, regardless of whether they were impacted by the outage. ⁶

The real question remains, what is the cost of data center failure and the resulting downtime for organizations? Is it covered by SLA credits? Most SLA credits reflect the cost of the services rendered and almost never provide for business losses.

At the Continuity Insights Management Conference in 2006, Agility Recovery Solutions stated that 78% of businesses who suffer a catastrophe without a contingency plan are out of business within 2 years. And 90% of companies unable to resume business operations within 5 days of a disaster are out of business within 1 year.⁷

Clearly some applications are considered more critical and have more visibility than others. Large companies feel the impact immediately when their ERP, CRM (*SaleForce.com is still plagued by a prime-time outage more than two years ago caused by a failure with an Oracle Database Cluster* ⁸), Business Intelligence or E-mail systems become unavailable.

However, with the proliferation of mobile devices and 'everywhere access', e-mail clearly stands out as the premier mission-critical application of today. Systems like Lotus Notes® and Microsoft® Exchange maintain a living record of a company's existence, storing every activity, process and thought an organization and its employees have. It's no surprise public companies are now required to maintain a record of e-mail activity for compliance purposes.

While the vast majority of businesses rely on e-mail everyday to send contracts, proposals, quotes and the majority of correspondence, most e-mail systems have not yet reached the point of reliability that phone service provides (99.999% or 5.2 minutes of downtime per year)⁹.

According to Osterman Research, most North American businesses experience more than one e-mail outage every month -- and many indicate that they could lose more than \$100,000 as the result of a single major e-mail outage.¹⁰ Osterman also found that the average business experiences nearly seven hours of e-mail downtime every year and that outages can bring many workers to a virtual standstill, who on average are 25% less productive during e-mail downtime.

"Forget the fact my billing rate gets impacted if I can't access my email system," says a partner at a major North American law firm who prefers to remain anonymous. "My company image gets tarnished immeasurably when I am working on a multi-million dollar, highly-confidential deal and I have to send out a set of documents using my Hotmail account because my email system is down. Somebody gets fired for that."

Michael Osterman, goes on to say, "Organizations are not meeting their targets for messaging system availability," and adds that the average e-mail system experiences about 70 minutes of downtime during a typical month, which translates to 99.84% uptime. To this he poses the question, "Is this good enough?" ¹¹

Ceryx Inc., a Hosted Microsoft Exchange provider with data center facilities in Canada and the United States, doesn't think so. They were the first in the industry to offer a real 100% SLA based on their multi-data center architecture and software design. Customers' data is replicated in real-time and resides in both data centers – more than 500 miles apart – so that even in the event of catastrophic failure, the primary system would fail over with almost no impact to the end-user.

"We operate on the premise that even the best data center can and will experience failure due to circumstances beyond anyone's control," says Dr. David Penny, CIO at Ceryx. "We focus our R&D on keeping the application highly available and rely on our replication technology to mitigate the vulnerabilities that exist on the data center level. And then we make the operating and capital investments necessary to execute daily."

For the past 4 years Penny and his team have worked with Enterprise Messaging systems, like Lotus Notes and Microsoft Exchange, developing technology to deliver high availability. Since 2004 they have been providing a geo-replicated Microsoft® Exchange 2003 service to medium and large-sized companies who see the cost and performance benefits of the Ceryx solution.

Most recently Dr. Penny and his team have been working with Geographic Clustering in Server 2008 and native Microsoft Exchange 2007 CCR (Cluster Continuous Replication) technology. What this allows for is clustering over a wide area network. Traditional clusters, which rely on the same RAID system in order to continue to function properly, are susceptible to logical corruption and certain physical corruptions that can propagate across an entire RAID array causing complete failure. Geo-Clustering eliminates the reliance of redundant servers on the same set of disks thereby eliminating a very common single point of failure.

“Even with WAN replication we need to ensure that the corruption itself isn’t replicated,” says Dr. Penny. For this they are utilizing log-shipping with delayed application rather than block-level replication, thereby avoiding the replication of corruptions caused by application defects. By monitoring performance on the primary system closely they can stop bad changes from being committed to the secondary system.

Beyond the physical vulnerabilities of a single data center, Ceryx is protected against a number of other vulnerabilities anyone using a single data center is exposed to. “When negotiating our contract, our provider knows how easy it is for us to move facilities,” says Roger Smith. “The data is already replicated and we don’t need to physically migrate servers. Migration to a new facility can occur without any impact to our customers. We can’t be held hostage to a bad contract or radical increases in pricing or continued poor performance.”

Ceryx also has a lot of flexibility where routing is concerned and should a backbone be down or congested, Ceryx with front-end servers operating at both facilities, has the flexibility to route traffic through a separate facility and bypass potential network congestion that can plague operators running out of a single data center.

While there are a number of solutions in the market that provide continuity through an interim e-mail system in the event of downtime, the Ceryx system is different in that it doesn’t require the user to even change settings when the e-mail system fails over to the secondary facility. Moreover, things like e-mail history, sent items and calendar entries all remain intact.

In this respect the Ceryx solution is not a continuity solution but rather a high-availability solution that provides layers of redundancy, from the software level up to the facility level.

Hosted archiving solutions – a good plan for any company facing regulatory and legal compliance - also provides a layer of assurance and access to e-mail records, should the primary facility suffer complete failure. However, these solutions will not provide business continuity or availability.

Moreover, if the primary e-mail provider experiences failure due to data corruption, the data being archived may be corrupt as well. Large data stores, even at the mailbox level, lead to corruption and the current trend of Hosted Exchange vendors selling e-mail accounts with massive storage allowances is introducing a higher probability of data corruption and subsequent failure. A good archiving strategy can be used to keep mailbox sizes manageable and subsequently reduce the likelihood of corruption.

So while extremely valuable in today’s world of mission-critical e-mail, archiving to an external hosted facility should not be mistaken for a multi-data center strategy. Instead archiving is a good backup plan and will not provide the protection businesses today need against the inevitable vulnerabilities that exist with a single-data center strategy.

These vulnerabilities are typically covered in the fine print of a facility’s SLAs, under the term ‘Force Majeure’; a phrase often translated as an ‘Act of God’ or the literal French translation, ‘Superior Force’

and is included as a clause to excuse interruptions in services caused by extraordinary circumstances beyond the control of the provider. Circumstances that - as demonstrated over the past year - are becoming more and more common.

Michael Osterman concludes, in his presentation on the Importance of E-mail Continuity, that the only solution to the inevitable problems that plague mission-critical service delivery is with a geo-replicated, multi-data center solution, like the one being offered by Ceryx.

Footnotes:

¹ CNET News: http://news.cnet.com/8301-10784_3-9828570-7.html

² Data Center Knowledge: <http://www.datacenterknowledge.com/archives/2007/07/24/generator-failures-caused-365-main-outage>

³ Data Center Knowledge: <http://www.datacenterknowledge.com/archives/2008/07/15/vancouver-power-outage-kos-plenty-of-fish/>

⁴ Data Center Knowledge: <http://www.datacenterknowledge.com/archives/2008/06/01/explosion-at-the-planet-causes-major-outage/>

⁵ Center Networks: <http://www.centernetworks.com/the-planet-data-center-fire>

⁶ CIO WebBlog: http://www.cio-weblog.com/50226711/google_manning_up_for_august_outages.php

⁷ London Chamber of Commerce Study, 2006

⁸ The Importance of Messaging in the Enterprise: A survey of email application continuity, Applicationcontinuity.org, 2006

⁹ CIO WebBlog: http://www.cio-weblog.com/50226711/salesforcecom_outage_root_cause_oracle.php

¹⁰ BNET Business Network: http://findarticles.com/p/articles/mi_m4PRN/is_2008_July_8/ai_n27893385

¹¹ Appcon 2007: Application Continuity Conference, 'The importance of Email Continuity' - (webinar available at <http://www.teneros.com/infocenter/>)

Ceryx Inc.

65 St. Clair Avenue East
Toronto, Ontario
Canada, M4T 2Y3

244 Madison Avenue, #720
New York, New York,
USA 10016-2819

Phone: 1-800-663-6245 ext.513
E-mail: info@ceryx.com

