

High Dependability through High Availability

There are dozens of potential reasons for email delivery delays and errors. With complex business-grade email systems, like Microsoft® Exchange and Lotus Notes, there are many ways to build and configure a system. Some conform to the highest standards, while others barely meet the minimum requirements. With very little effort, you can determine whether your system had been adequately configured to suit your needs

January 2009

For further information on this and other publications please contact:

E-mail: sales@Ceryx.com

Facsimile: 416-946-1484

Toronto: 416-482-9659 ext.513

New York: 212-764-2455 ext.513

Toll Free: 1-800-663-6245 ext.513

©2009 Ceryx Inc. All rights reserved. Ceryx® is a registered trademark of Ceryx Inc., in Canada and other countries. All other trademarks and registered trademarks are the property of their respective owners.

A Pragmatic Example: The Missing Email

As a professional, your clients know they can depend on you, and that you are there for them. But what happens when your technology is not there for you?

Imagine you are on the phone with your client, telling him with great pride that at long last the document he has been waiting for is complete. While you chat, you attach the document to an email, send it off...and wait.

And wait. The email is no longer in your outbox and not in his inbox. You double-check the email address, and as the conversation becomes more and more awkward, you assure the client that the document really does exist and there really was an email sent. He is frustrated and you are at a complete loss. You decide you cannot wait any longer and switch to your personal Hotmail account to resend the document, which may or may not work immediately either. The result is wasted time, and if you switched email systems, you no longer have a cohesive record of the exchange.

Your system administrator will list a number of reasons why this incident happened, reasons that are beyond his control:

- Junk mail folders
- Anti-virus systems
- User errors
- Bottlenecks in the Internet that day

The real reasons may actually run much deeper, which means that this lost email will not be an isolated incident.

Lost Emails Mean Lost Opportunities

Consumers, accustomed to fast responses through websites, are quickly losing their tolerance for slow email responses. Evidence of this can be seen through a recent study by the WAV Group, who claims in the real estate industry, the first agent to respond to a customer email inquiry has a 73% chance of securing the business. The third to respond has virtually no chance.¹ Similarly, a study conducted by Vodafone released in December of 2008 estimates that lost opportunities due to failures to reply promptly to email messages cost businesses approximately \$27,000 a year.²

What about email that goes missing entirely? This summer Apple Corp admitted to “losing” 10% of their MobileMe subscribers email between July 16th and July 18th.³ A wave of blog posts have popped up as a result of this problem, with subscribers making claims of everything from lost job opportunities to lost business. We tend to view email as a low cost service, yet clearly when it fails, the costs can be startling.

Apple would only remark that the problem was due to a “serious issue.” So what exactly does happen to these wayward email messages, delayed or lost?

The truth is there are dozens of potential reasons for email delivery delays and errors. With complex business-grade email systems, like Microsoft® Exchange and Lotus Notes, there are many ways to build and configure a system. Some conform to the highest standards, while others barely meet the minimum requirements. With very little effort, you can determine whether your system had been adequately configured to suit your needs.

High Availability

As the name implies, High Availability is a system or network that is operational, or available, with a high degree of certainty and frequency. Ceryx, a Hosted Exchange provider, has made high-availability one of their primary focuses. With data centers in New York and Toronto, they have developed various technologies to replicate all messaging data in real-time and can fail over to the secondary facility in the event of catastrophe, allowing them to provide a real 100% SLA. While most providers tend to choose between high availability and high performance, all of Ceryx's Microsoft Exchange deployments have been built to meet the highest standard of availability uptime without sacrificing performance.

One benchmark for testing an application provider's performance is to log in to the webmail application and switch from the email view to the calendar view. Pick a folder with lots of messages in it and try sorting by "sender" and then clicking through the various pages. If there is a noticeable delay between these actions then your application provider's performance may not be adequate.

A system's availability is determined by numerous factors, each of which is examined below.

Architecture

Delays are often a result of high server RPC latency. RPC, or Remote Procedure Call, is how the Outlook client or the Outlook Web Access client (OWA) communicates with the Exchanger servers. RPC latency refers to the delay between initiating a request and its completion. RPC latency, as seen from the client, is a combination of networking latency and server latency. For good performance, the Microsoft guideline is 50ms average latency on the mailbox servers.⁴ If mailbox server RPC latency averages much higher than this, the desktop user can experience "pop-ups" and warnings about problems with the connection from Outlook to the Exchange server. In the background, inbound and outbound email is not being processed as the system tries to catch up with other requests. Excessive "pop-ups" can become more than just an irritant; they can slow down a PC to the point of being unusable while Outlook tries to establish a connection to the server.

Ceryx aims to maintain an RPC rate 20ms or less. To do this, they have avoided the tendency of many providers who built their systems with less expensive virtualized environments and network attached storage. Instead Ceryx has invested in server clusters and high-end SANs (Storage Area Networks) at each of their data centers. SANs provide disk access performance, as well as redundancy through RAID configurations and fibre channel connectivity to the mailbox servers.

Many providers, in an effort to reduce costs, combine the Client Access Server and the Hub Transport Server (servers used to facilitate email delivery) on a single physical machine. This dual role can introduce latency at times of peak usage and throttle the ability to handle large outbound mail queues. Ceryx has engineered the Hub Transport role to ensure message queues can be quickly cleared even during periods of heavy load. By combining multiple, dedicated physical Hub Transport servers with the built-in round-robin load balancing capabilities of Exchange 2007, messages are quickly distributed to their destination on the internet. For legitimate email, hardware load balancing is used to ensure optimal performance of the critical Hub Transport role, which processes every single message that passes through the system.

Monitoring

Most hardware load balancing configurations can *achieve* both high performance and high availability; however, to *sustain* high performance and high availability - with a dynamic system like a Hosted Exchange deployment, where usage and load can vary dramatically - you must have an advanced monitoring system and the processes in place to scale the system in response to constantly changing variables.

Some providers install generic monitoring packages that tend to monitor every single metric available, whether the metrics provide meaningful insight to system performance or not. With Ceryx, every aspect of the environment from key metrics around SAN queue length through to standard metrics like CPU utilization and memory usage are closely monitored, trended, and understood. This information is used to develop highly accurate forecasting and scheduled system scaling. Any provider who has not invested in such monitoring tools and resources will just *react* to spikes in usage and load. Ceryx *plans* for them.

Breathing Room

Another major reason for email delays and poor performance can again be attributed to basic economics. A “store” is the unique databases for storing messaging data. With Microsoft Exchange 2007 Enterprise, there is a limit of 50 stores per server, with a Microsoft recommended limit of 200 GB for each store. Some providers will try to maximize the number of customers they can support by pushing these limits. Ceryx does the opposite. Ceryx maintains a maximum of 50 GB per store - one quarter of the maximum - in order to deliver the best raw database performance and protect against data corruption, common with larger stores. With these hard restrictions in place, it's difficult to imagine how some providers - who offer mailboxes up to 4GB - are able to run a sustainable business and maintain high performance and availability.

IOPS

Economics may also drive another cause of poor performance: IOPS, a measurement of the number of times data can be written and read to disk per second, can vary widely per system. With Microsoft Exchange, IOPS capacity is directly related to performance and a heavy user can use up to 1 IOPS under certain circumstances. Each disk has a maximum number of IOPS it can support, so a careful ratio of users per array of disks is essential to maintain decent performance.

Ceryx is careful to maintain a generous ratio of IOPS per user based on real-world metrics monitored from its large user base, taking into consideration peak activities times and not just daily averages.

Routing Issues

To be fair, email delays and errors may occur outside your environment. In the scenario where your colleague was eagerly sitting and waiting for an email to arrive, it may have got stuck somewhere “in the cloud” - somewhere in between both of your email environments.

On a pre-sales and support level, Ceryx has developed a number of troubleshooting tools to help identify potential routing issues. As with any application “hosted in the cloud,” bandwidth is an important consideration when trying to ensure a positive end-users experience. Microsoft Exchange, which consumes 3 KB of a client's internet connection per active user, is no different. Ceryx has developed a tool that simulates a series of network connections to determine if the client has adequate bandwidth to support their user base.

Through this exercise they have discovered that - with any client moving from an on-premise environment to a hosted environment - any increase in bandwidth required is partially offset by decommissioning and offloading SMTP traffic, external connections spam traffic and attacks.

Spam/Anti-Virus

Of course the most notorious place a message gets “stuck” or delayed is in a provider's anti-virus and spam filtering system. Ceryx runs seven different anti-virus products in its environment to ensure system health and email hygiene. That level of protection could potentially cripple an ordinary system that wasn't built for high performance; however, Ceryx has not only sized their solution with the performance hit associated with anti-virus scanning in mind, but also closely monitors its environment to ensure email delivery is never compromised by spam and anti-virus filtering.

Beware of any provider who either ignores this potential performance hit or, worse yet, removes backend anti-virus all together. Removing anti-virus may appear to be a good way to control costs, as high-end solutions can be very expensive, but removing them allows viruses in, and once in, are often hard to then find, never mind remove. Ceryx's system scans incoming and outgoing messages as well as messages in the store on a scheduled basis. A regular scan of mailbox databases is just as important as gateway anti-virus to reduce the instance of catastrophic failure.

Mobility

The benefits and usability of the Ceryx architecture can be measured not only on the desktop level but for mobile users as well, who share the same highly optimized and redundant environment. Ceryx also maintains fully-replicated BES (BlackBerry Enterprise Servers) in both its environments to maintain maximum mobility uptime as well. Forum Oilfield Services, out of Houston Texas, currently has 600 employees using the Ceryx Hosted Exchange solution, with almost 100 mobile users.

Beware of “Bargains”

Unfortunately the market is crowded with Hosted Exchange providers who are not as concerned about delivering a dependable and resilient solution as they are about keeping their costs low and selling high volumes of mailboxes. Providers that offer 3 or 4 GB mailboxes are ignoring the commonly held understanding that mailboxes of this size do not perform well and are highly susceptible to data corruption. That corruption is not limited to just the user in question, but can impact all users on that store, potentially even the entire server.

These providers can only offer mailboxes of this size at a bargain price if they compromise on some other expense - dedicated servers, SANs, anti-virus solutions or a fully redundant architecture. Of course the much greater expense is the lost productivity and opportunity that results from using a solution that was designed with the primary goal of meeting the lowest possible price point to compete with free solutions in the market, and not delivering Enterprise performance and availability.

Ask all potential vendors for the following:

- Customer references
Quiz the references about any down-time they may have experienced)
- The SLA

Have a close look at the financial penalties if high availability is not maintained. Professional hosting providers, confident in their systems ability to maintain high availability, will include clearly worded conditions in their SLA around the exact fees that will be paid to a customer should they not maintain an acceptable level of up-time.

How Do You Measure Up?

A lost or delayed email could be viewed as a nuisance or as a warning. Purchasing decisions need to be fiscally responsible, but often, seemingly bargain solutions can have a devastating and costly effect in the long run.

So how does your system measure up? Using Ceryx as a control, measure the speed, versatility, and resilience of your system.

	Ceryx	Your System
SLA	100%	
Latency	20ms	
GB/Store	50GB/Store	
IOPS : User	High	
Bandwidth Sizing Tools	Yes	
Spam/Antivirus	Gateway + Mailbox Server, Integrated	
Mobile Device Access	BlackBerry, ActiveSync	

Ceryx customers share the belief that their data is their companies most valued asset. As the most used collaboration tool in business, an individual Microsoft Exchange account typically contains a blueprint of an employee's tenure: their schedule, correspondence (and commitments), contacts, contracts and much more. Ceryx customers trust that this real-time journal is safe, available and always accessible.

¹ The Wav Group: Gaining an Edge in Real Estate with Smartphones <http://wavgroup.com>

² The Open Press: <http://www.theopenpress.com/index.php?a=press&id=42004>

³ Apple Support Forum: <http://support.apple.com/kb/TS1953>

⁴ Microsoft Exchange Team Blog: <http://msexchangeteam.com/archive/2005/09/28/411674.aspx>

⁵ Microsoft TechNet: <http://technet.microsoft.com/en-us/library/bb738147.aspx>