MaximumOn™

Bringing High Availability to a New Level

Introducing the Comm100 Live Chat Patent Pending MaximumOn™ Technology

Comm100

Introduction

While businesses have become increasingly dependent on computer-based systems, those that serve global markets must keep their systems running 24x7. High availability is therefore a top priority.

As a real time customer-facing communication tool, live chat software must exhibit higher availability than most other business applications. Customers often make their buying decisions while engaged in a live chat session, and thus any interruption in the chat session will directly affect purchases.

Despite heavy testing, computer hardware and software can experience failures that result from power supply issues, computer network outages, security breaches or domain name resolution problems. Additionally, there are external elements, such as fires, earthquakes, floods and storms that can severely damage or destroy entire data centers.

Comm100 Live Chat *MaximumOn*^{TM} is a patent-pending technology that provides data center level redundancy for Comm100's live chat solution to achieve unprecedented high availability. When deployment in one data center fails, the redundant deployment in a different data center will automatically take over. Web site visitors will not even notice the switch, and on-going chats remain intact. Comprising industry-leading component and system level redundancies, Comm100's *MaximumOn*^{TM} technology can sustain the live chat service during almost all kinds of component, system, and data center level failures, including planned downtime as well as regular system maintenance.

In this white paper, we will discuss concepts related to high availability and then explain how Comm100's *MaximumOn*^{TM} technology implements these concepts to provide ultimate reliability in live chat.

High Availability Concepts

High Availability can be defined as the ability of a system to maintain continuous operation over an extended period of time. The concept of availability is expressed using several different terms, as explained in the subsections below.

The number of 9s

Availability is expressed as a percentage according to the equation below:
Availability = (total time – downtime) / total time * 100%

When the result of the above equation is greater than 99%, we refer to the total number of consecutive 9s as an indication of the availability. For example, 99.916% has three consecutive 9s, whereas 99.997% has four consecutive 9s. The more consecutive 9s, the better.

Planned downtime and unplanned downtime

There are two kinds of downtime, planned and unplanned. Planned downtime is anticipated and scheduled while unplanned downtime is unexpected due to system failure, human error or a process problem.

The table shows the total downtime in a year corresponding to a different number of 9s:

The number of 9s	Availability %	Total Annual Downtime
2	99%	3 days, 15 hours, 36 minutes
3	99.9%	8 hours, 45 minutes, 36 seconds
4	99.99%	52 minutes, 33 seconds
5	99.999%	5 minutes, 15 seconds

Table 1 - The Number of 9s

Recovery Time Objective (RTO)

Recovery Time Objective is the duration of time within which the system must recover from failure. RTO is widely used by businesses to set goals that indicate how much downtime a business can tolerate.

Failover

Failover refers to the act of switching to a different, redundant system upon the failure of – or abnormal conditions within – the currently active system. Manual failover may take minutes or even hours, while automatic failover usually takes seconds or even milliseconds. For high availability solutions, automatic failover is a must-have.

Three Levels of Redundancy

Redundancy refers to the duplication of critical components or functions of a system. High availability solutions depend on redundancy to eliminate single points of failure and maintain service availability. Different levels of redundancy provide different levels of availability. High Availability essentially is a business

tradeoff between the cost of downtime and that of avoiding or reducing the downtime. Generally speaking, there are 3 levels of redundancy.
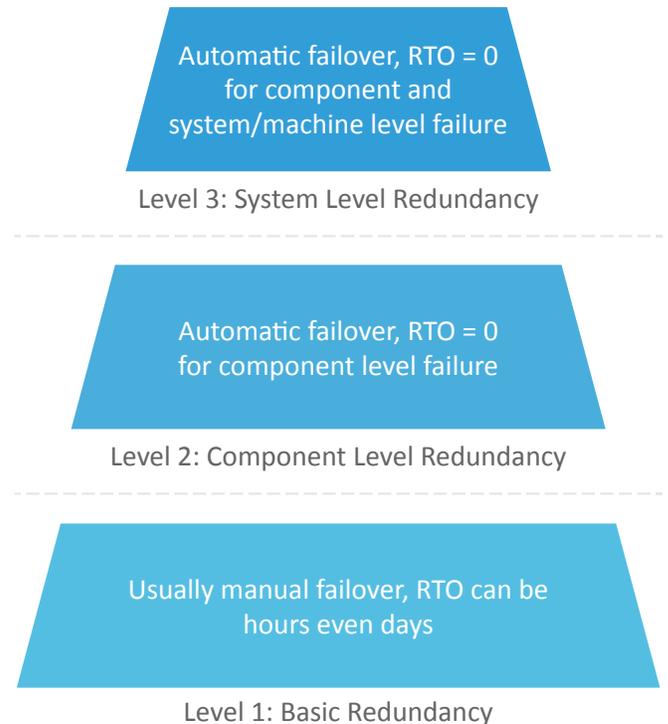
Automatic failover, RTO = 0 for component and system/machine level failure

Level 3: System Level Redundancy

Automatic failover, RTO = 0 for component level failure

Level 2: Component Level Redundancy

Usually manual failover, RTO can be hours even days

Level 1: Basic Redundancy

Figure 1 - 3 Levels of Redundancy

The table below further explains the different levels of redundancy:

	Level 1 – Basic redundancy	Level 2 – Component Level Redundancy	Level 3 – System Level Redundancy
Type of applications	No need for immediate recovery. Manual intervention is acceptable.	Important business applications tolerating little downtime.	Mission critical applications accepting virtually no downtime.
Cost and Complexity	Cost and complexity are significantly lower than the other levels of redundancy.	Cost is higher than that of level 1. Since commodity fault tolerant systems are widely available, the complexity is not very high.	Cost is high as usually both the component and the system level redundancies are provided. Managing the clusters and automatic failovers requires in-depth expertise.

	Level 1 – Basic redundancy	Level 2 – Component Level Redundancy	Level 3 – System Level Redundancy
Technologies	Basic data backup and use of commodity hardware.	Fault-tolerant operating system, hot swap hard disk array, redundant fan, redundant power unit, network card coupling, etc.	Server clusters, load balancing, redundant network switches, data mirroring, multiple network providers, multiple electric power supplies and UPS, etc.
RTO (Recovery Time Objective)	Usually hours or even days. Services are unavailable during the downtime.	RTO = 0 for component level failure. No interruption of service when a single component is down. Recovery time could be hours or days when there is a system level failure.	RTO = 0 for system/machine level failure. No interruption of service when a single server or network node is down.
Failover	Manual failover usually required.	Automatic failover for component level failures. Manual failover required for higher level failure.	Automatic failover for component and system level failures.

Table 2 - 3 Levels of Redundancy Comparison

The Limitations of the Level 3 Redundancy

A software solution with the above mentioned 3rd level redundancy is able to withstand most component level and system level failures. However, there are still scenarios that involve service downtime, which could be:

Data Center Level Failure

Modern advanced data centers usually deploy redundancy to ensure system availability in the event of node failure. But we all know data center failures do happen. The human error, planned upgrade or replacing of devices, manual failover of redundant devices, fire in the building, nature disaster, loss of city electricity supply for a long time, upstream network problem, DDOS attack or cooling system failure can all contribute the downtime of a data center.

Human Error

The higher the level of availability required, the more complex the system becomes, and the more complexity in the system design, the more likely human error creates downtime.

In the book entitled *High Availability Network Fundamentals*, CISCO states that the downtime caused by human error is "more than the downtime for all other reasons combined." The book goes on to state, "When they do have problems, it is just as likely to be a result of someone changing something as it is for any other reason."

In a high availability environment, there are many possible human errors, such as incorrect firewall settings, flawed security settings, accidental file deletion, incorrect application version deployment, and misconfiguration of network elements. As the list goes on, any human error could leave the component level and system level redundancy solution useless.

Planned Downtime for Upgrade or Maintenance

Businesses often need to upgrade hardware to accommodate more user requests, upgrade software to roll out new features or apply operating system patches to improve system security and stability.

Well managed planned downtime is usually considered as an investment in preventing or mitigating unplanned downtime. The fact is, no matter planned or unplanned, when the system is down, it is down. There is no difference to the users.

DNS failure

Web applications rely on DNS to find the IP address of the servers. If there is a problem with the DNS server for any reason, the server is not accessible for the client even though the server applications are still running.

One More Level of High Availability

The Comm100 Live Chat Patent Pending *MaximumOn™* technology introduces a new level of redundancy into the live chat market, which brings redundancy beyond the data center boundary and can withstand not only component and system level failures, but also data center level failures.

By having redundancy at the data center level, the downtime can be significantly reduced. For a deployment in one data center with an availability percentage of 99.99%, the annual downtime is 52 minutes, 33 seconds. If the two deployments are spread across two separate data centers in different areas, each with the same 99.99% reliability, the annual downtime can be near to zero, as the probability that these two failure times overlap is very small.

Note that human errors and planned downtime that do not cross the data center boundary will not bring the service down.
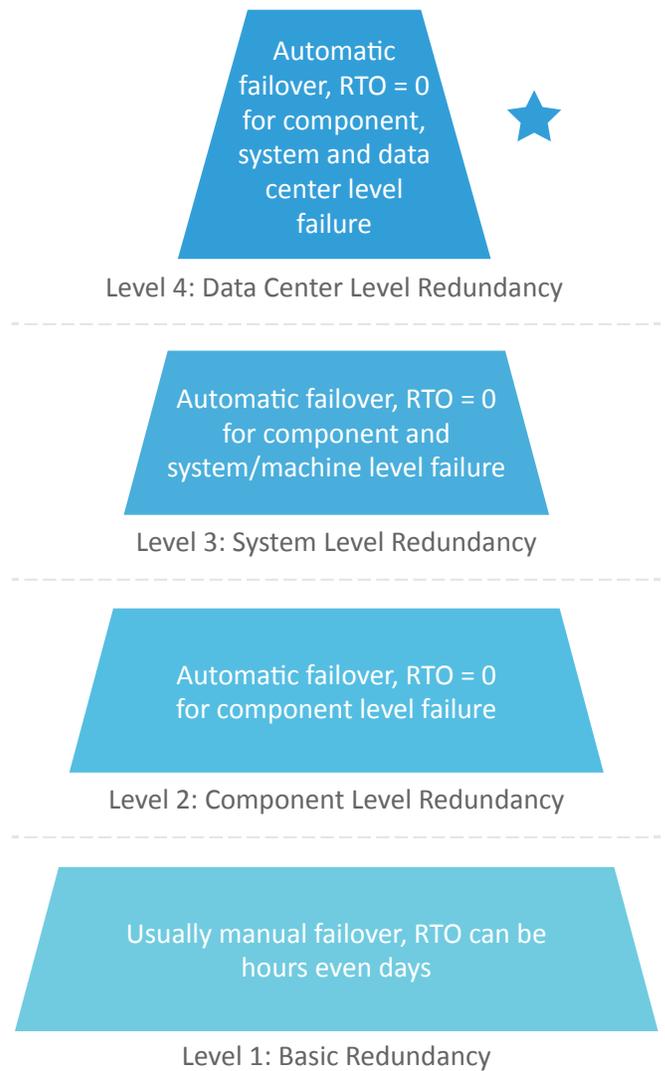
Automatic failover, RTO = 0 for component, system and data center level failure

Level 4: Data Center Level Redundancy

Automatic failover, RTO = 0 for component and system/machine level failure

Level 3: System Level Redundancy

Automatic failover, RTO = 0 for component level failure

Level 2: Component Level Redundancy

Usually manual failover, RTO can be hours even days

Level 1: Basic Redundancy

Figure 2 - 4 Levels of Redundancy

Comm100 Live Chat *MaximumOn™* Technology

Comm100 Live Chat *MaximumOn™* is an advanced technology developed by Comm100, based on years of research and experience in understanding businesses' live chat needs. The technology can protect Comm100's live chat solution from both component/system level failures as well as data center level failures, which means chat conversations can remain intact even when a whole data center is down.

Why It Matters?

1. Data center level failure happens and it's expensive.
A study conducted by Ponemon Institute in 2013 shows that 91% of the data centers surveyed have experienced an unplanned data center outage in the past 24 months. Taking the planned downtime into account, you know how often it happens. The study further indicates that the average cost per minute of data center downtime has increased to $7,908 in 2013. Remember that the live chat solution is for businesses to have real time communication with their customers, which means the cost can be even higher.

2. Live chat is one of the main customer communication tools for increasingly more businesses.
According to a Forrester report in 2013, a 24% rise in chat usage for customer service has been witnessed in the past three years. The most beautiful part in live chat communication is that it happens in real time. Having a chat button on the website saying "I'm online to chat" can give customers a sense of reliability and by getting customers' questions answered instantly, businesses can further boost customers' confidence in

the brand. The disruption of live chat can harm not only the sales or customer satisfaction, but also the brand reputation of a business.

Architecture

Figure 3 is a high level architectural description of the Comm100 Live Chat *MaximumOn*^{TM} technology.

In the Primary Deployment , Comm100 employs level 2 and level 3 redundancies to sustain the system in the event of component or system level failures. A high end N+1 data center is selected to host the servers. Rigorous disaster recovery plans are implemented. Data are continually backed up both locally and to a remote site.

In addition to the Primary Deployment, Comm100 implements a Redundant Deployment in a separate data center, which synchronizes with the Primary Deployment in both software applications and databases.
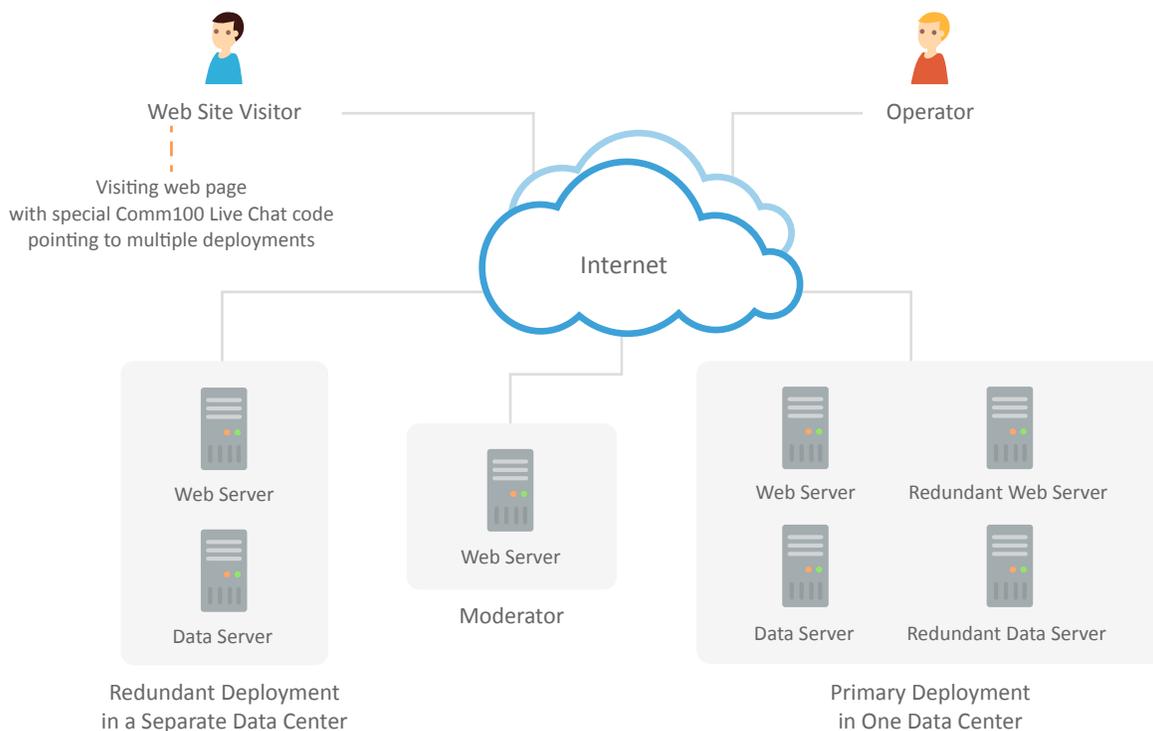

Figure 3 - High Level Architectural Description of The Comm100 Live Chat MaximumOn^{TM} *Technology*

The Moderator monitors the status of the two deployments and indicates the currently active one for Operators and Website Visitors. Failover is automatic by default. Normally, Operators and Website Visitors are connected to the Primary Deployment. When it fails, the Moderator will direct them to the stand-by Redundant Deployment. When the Primary Deployment is back to normal, the Moderator will set it as active again with all the data synchronized. When the Moderator fails, Operators and Website Visitors will bypass the Moderator to directly connect with the Primary Deployment or the Redundant Deployment. Manual failover is also provided in the event of planned downtime.

The high level concept of the Comm100 *MaximumOn™* technology may look simple, but the actual patent pending technology to keep the on-going chats intact requires sophisticated algorithms, strong attention to details and years of experience in developing live chat applications.

The Comm100 Live Chat *MaximumOn™* technology is activated through a special piece of JavaScript code placed on customer-facing web pages, which points to both the Primary Deployment and the Redundant Deployment, and switches between the deployments when one fails.

Failures and Scenarios

Table 3 itemizes failure scenarios and describes how the Comm100 Live Chat *MaximumOn™* technology responds.

Failure Scenario	What Happens	System keeps running?
Redundant component failure, such as hard disk, fan, power unit and network card failure	The component level redundancy protects the deployment.	Yes
Non-redundant component failure, such as mother board, CPU failure	The system level redundancy protects the deployment.	Yes
Server failure	The system level redundancy protects the deployment.	Yes
Network device failure	If the network device is redundant, the system level redundancy protects the deployment.	Yes
	If the network device is not redundant, the data center level redundancy protects the service.	Yes
Data center failure	The data center level redundancy protects the service.	Yes
Human error	If the human error causes system failure and the error is limited in the data center boundary, the data center level redundancy protects the service.	Yes
	If the human error causes system failure and the error is across the data centers, the whole service is down.	No

Failure Scenario	What Happens	System keeps running?
Hardware upgrade	Do it only in one data center at a time. Shut down the system, upgrade the hardware and start the system. When one deployment is down, the redundant deployment in the other data center sustains the service.	Yes
Software upgrade and system patch	Do it only in one data center at a time. Shutdown the system, upgrade the software or apply the patches and start the system. When one deployment is down, the redundant deployment in the other data center sustains the service.	Yes
The active deployment failure	The standby deployment will take over.	Yes
The standby deployment failure	Not affected. But the failure need to be fixed.	Yes
The active deployment failure while chatting	The standby deployment will take over. The ongoing chat content will be pushed from the Operator's side to the server. The Operator and the Web Site Visitor do not notice the failover.	Yes
One DNS server failure	The other server will be used automatically.	Yes
Moderator failure	The Operator and Web Site Visitor sides will just bypass the Moderator. The Primary Deployment will be used as the active deployment by default. If the Primary Deployment is not working properly, the Redundant Deployment will be used as the active one.	Yes
Both the Primary Deployment and the Redundant Deployment are down	In this event, the system is down.	No
The Primary Deployment is back online	The chat transcripts and offline messages in the Redundant Deployment will be synchronized back to the Primary Deployment.	Yes

Table 3 - Failure Scenarios & How MaximumOn^{TM} Responds

Getting Started

By pasting the special JavaScript code provided by Comm100 onto customer-facing web pages, businesses can start enjoying the benefits of the new technology immediately and begin the journey of seamless chatting. An active Comm100 Live Chat account is required, for sure.

Summary

Given the importance of high availability, it is essential to understand the different levels of redundancy and how they affect availability. The Comm100 Live Chat *MaximumOn™* technology offers data center level redundancy to live chat, dramatically increasing availability, and maintaining chat conversations despite system-level failures. Global businesses can depend on this reliability for maintaining critical customer chat conversations.

About Comm100

Comm100 is the global provider of live chat, email marketing and help desk software. With "100% communication, 100% success" as our motto, we are committed to making customer communication simple and effective for you.

Comm100 Live Chat is an enterprise-grade live support solution that enables businesses or organizations to engage their website visitors in real time so as to improve conversions, sales and customer satisfaction. Our clients include Fisheries and Oceans Canada, Toronto Public Health, Stanford University, Whirlpool and many more.

Contact Us

TEL | (778) 785-0464
FAX | (888) 837-2011
E-Mail | support@comm100.com
Web | www. comm100.com

Suite 238 - 1027 Davie Street
Vancouver, British Columbia
V6E 4L2 Canada