Fault Tolerant Solutions Overview

Prepared by:

Chris Robinson
Sales & Technical Director
24/7 Uptime Ltd

October 2013

This document defines and highlights the high level features, benefits and limitations of hardware and software based fault tolerant solutions compared to those available using non-fault tolerant solutions.

1. Table of Contents

2. HA versus FT – A Description .......................................................................................................................... 2

3. Fault Tolerant Hardware Solutions .................................................................................................................... 3

4. Software Fault Tolerant Solutions ..................................................................................................................... 4

5. High Availability Solutions .................................................................................................................................. 5

6. Citrix Xen server / VMware / HyperV Based Resilience .................................................................................. 6

7. Clustering Solutions ............................................................................................................................................. 6

8. Summary Table .................................................................................................................................................... 7
2. HA versus FT – A Description

The need for computer system resilience and high availability of critical applications has been a requirement that has evolved and refined since computing became commonly used in business processing (during the 1970s with wider adoption after the Microsoft Windows revolution started in 1983). Until relatively recently a system that may endure levels of unscheduled downtime of seconds, minutes, hours or even days has typically been acceptable (depending on the nature of the application) but we are now in an age of ‘always on’ computing and a zero downtime requirement is now needed for many critical applications. As this need has become more prevalent, so the IT industry has constantly strived to provide business with the applications and levels of availability required.

High availability solutions previously implemented in environments such as 24/7 manufacturing plants, travel, CCTV & access security, military applications, healthcare, banking, high volume retail transaction organisations, online ecommerce and booking systems are no longer good enough.

The term High Availability (HA) itself is a self-descriptive example of the problem in our 21st century processing environment: systems will be highly available but not ALWAYS available. To date HA solutions are reactive – typically with some active / passive arrangement of multiple physical servers. In other words, an active production server is running a critical application in normal processing conditions with a standby passive server ready to react to a failure, enabling processing to continue on the standby machine. The problem is that the passive server is waiting for the main server to fail before this reaction takes place; hence a period of application downtime while the passive machine takes over has to be accepted and tolerated. This can be hours, minutes or even just a matter of seconds but, to the most critical environments, this is not an acceptable level of downtime. Not only is there downtime when a failure strikes, but the very real risk of the loss of in-flight data when the fault occurs. (Typically – there is also a period of downtime after the underlying fault such as disk array crash, NIC failure or full server failure has been resolved and processing is rolled back to a protected status). System downtime and loss of data simply isn’t good enough for today’s critical “always on” applications.

To achieve true zero downtime we need to look beyond a High Availability solution. We need to tolerate faults at any level; we need to eliminate any single point of failure; we need to get rid of the period of downtime while a standby server reacts to a failure (and while it is restored to a re-protected state after a failure), and we need to eliminate the chance of any data loss when a fault occurs. Just as High Availability is a self-descriptive term, so is Fault Tolerance. By eliminating ANY single point of failure and getting two (or more) physical servers working together as a single unit, we are creating a Fault Tolerant (FT) environment.

In other words, faults are tolerated and computing of the critical application continues throughout the period of component or full system failure. In a FT environment, processing also continues throughout the period of restore to a protected status. With critical applications protected by a Fault Tolerant solution, critical environments suffer no downtime or data loss whatsoever when an underlying component or full system failure strikes.

The analysis of acceptable levels of downtime is the starting point for an organisation considering HA versus FT protection. Quite simply – if downtime and/or data loss is acceptable, then High Availability solutions can be considered and deployed; if no unscheduled downtime or data loss is acceptable, then Fault Tolerant solutions are the only ones that should be considered for deployment.
3. **Fault Tolerant Hardware Solutions**

Stratus ftServer hardware based fault tolerant servers are the only solutions to deliver complete component and full server fault tolerance in a single unit. Two identical server “slices” are built within a single robust chassis with all components managed and synchronised by the patented Automated Uptime Layer. This manages Lockstep Hardware Technology provided within the complete single unit package and achieves higher than 99.999% levels of availability. These very high performance servers are fully scalable in terms of processing power, RAM and storage. By adding on additional Stratus ft storage array/s up to 80Tb of storage is available, along with up to 256Gb of RAM and 32 processors (with hyperthreading) in the main 2 x slice chassis.

**Lockstep Hardware Technology**

Replicated, fault-tolerant hardware components process the same instructions at the same time. In the event of a component malfunction, processing doesn’t miss a beat. The redundant component acts as an active spare that continues normal operations without system downtime or data loss. But that’s just one of the major differences between Stratus ftServer systems and conventional servers.

The ftServer architecture separates PCI I/O from the rest of the motherboard and adds hardware logic in the form of custom Stratus chipsets. These chipsets provide the essential foundation for lockstep processing and the ability to detect, isolate, and withstand faults. Lockstep operation allows the ftServer system to isolate any hardware failure without any degradation in performance.

**Automated Uptime Layer**

The Automated Uptime™ layer presents and manages the replicated ftServer components as a single system. This dramatically reduces complexity and operator error. Conventional technologies like clusters require you to synchronize state information between the nodes and between all the layers of multi-tiered applications such as the web layer, middleware, and back-end database.

Working in concert with lockstep technology, the Automated Uptime Layer prevents many errors from escalating into outages. Even in-memory data is constantly protected and maintained. Other issues are captured, analysed and reported to Stratus. This allows support personnel to take a proactive approach to correcting software problems before they recur.

**Supported Processing Platforms**

Stratus ftServers are delivered fully configured and ready for applications/s to be loaded on Microsoft Windows Server or Red Hat Enterprise Linux O/S with VMware and HyperV virtualisation hypervisors both supported.

**Monitoring, Support and Automated Slice Replacement**

Stratus provide a 24 x 7 real time monitoring service, with individual slices, components and subcomponents constantly monitored for faults or faults likely to strike in the near future. On-site maintenance is virtually zero – as problems of any nature are automatically detected and, if required, a replacement slice is delivered to site. The faulty slice is simply unlocked from the chassis and removed, and then replaced by the supplied complete server slice with no interruption to processing at all. No on site IT specialist support is required.
Performance Impact
Fault Tolerance is achieved at hardware level so there is no significant performance degradation compromise when protecting critical applications with Stratus ftServer technology. No additional processor cores or RAM need to be factored in to the hardware specification analysis over and above processing requirement of the application/s protected.

4. Software Fault Tolerant Solutions
Stratus also provide the only software based fault tolerant solution capable of 99.999% resilience for multi-core applications on standard off the shelf server hardware. Also uniquely, the solution is storage agnostic (local or attached storage can be used) – and is not reliant on the deployment of shared storage (itself a single point of failure). everRun MX achieves fault tolerance by pooling together the components of two (or more) standard HP, DELL, IBM (etc) servers through dedicated NIC to NIC availability links. The underlying operating environment at physical host level is hypervisor based, with the application/s running within Protected Virtual Machines (PVMs). In configuration, a master host is selected as being the primary processing host and all transactions, disk I/O and memory processing is taking place on both hosts at the same time, managed by a software based solution similar to the Lock Step feature used by the Stratus ftServer hardware solution. The resulting environment is protected against component failure, multiple component failure and full physical host failure. For example, an everRun PVM will continue to process uninterrupted should a full raid array fail on server one, at the same time as the production NIC fail on server two (in a simple 2 x host pool) – i.e. component fault tolerance. Similarly should a full physical host fail, the PVM continues to compute through the failure uninterrupted (system fault tolerance).

Level 2 and Level 3 Protection
There are two versions of the everRun solution – Express and Enterprise. Component Fault Tolerance as described above is termed Level 2 fault tolerance. System Fault Tolerance as described above is termed Level 3 protection. everRun MX Express can protect multiple PVMs to level 2 only; therefore it is actually a HA solution rather than a FT solution because there is up to approximately 2 minutes of system downtime in the event of a full server host failure. everRun MX Enterprise can protect PVMs to either level 2 or level 3, so less critical VMs can be protected to component level only and VMs requiring the highest level of resilience can be protected to level 3 full system fault tolerance in the same 2 x host pool.

Geographic Dispersion of Hosts and Off Site Disaster Recovery
An everRun 2 x host pool can be split across a LAN environment providing there are suitably fast comms links. By installing either Quorum services or everRun Split Site, a third device is then deployed to act as referee to ensure the two host servers do not continue to process independently (creating what is termed as a split brain environment) should the link between the two hosts fail (switches fail, cable severed during construction or maintenance, etc).
An optional add on module, everRun MX Extend, provides off site DR failover from the two hosts in the event of a full site disaster. This D/R module does incur a period of downtime during failover; however, as it would only be used in the event of both hosts failing in the everRun 2 x host pool, (i.e. a true site disaster such as fire / flood etc.) this is usually an acceptable level of downtime for the majority of deployments.

**Supported Processing Platforms**

everRun MX can protect Windows server environments only.

**Support**
everRun allows an IT department to resolve underlying failures (component or full server failure) without any downtime, hence likely avoiding errors by being rushed / pressured due to loss of productivity, revenue, reputation, pressure from FD / MD / Ops director etc. The solution does require specialist support and maintenance and is not a simple component “slice” replacement; nor is it monitored 24 x 7 by Stratus. SNMP alerting can be configured – and round the clock support is provided by Stratus partners and backbone support services, but not on an automated basis.

**Performance Impact**
There is little performance impact at level 2 protection; however level 3 protection requires additional processing cores and RAM. As every transaction is interrupted and checkpointed to achieve full synchronisation, high I/O and processing intensive environments may experience a noticeable level of performance degradation.

5. **High Availability Solutions**
Several products are available that achieve levels of High Availability to various levels of sophistication and complexity. Widely deployed solutions such as DoubleTake and NeverFail are examples of HA solutions achieving levels of 99.99% availability but, with these solutions, there are periods of downtime as they work on a reactive basis rather than proactively. A standby server is waiting for the main master server to fail and will then assume the role of the master server, allowing processing to continue.

It must be stated that there are known instances of data loss and IP clash problems with both of these solutions as well as scripting requirements when restoring back to a protected state. They definitely cannot be classed as Fault Tolerant solutions and extremely critical environments, where downtime is unacceptable, cannot rely on this class of HA solution.

As detailed in the Software Fault Tolerance section of this document, level 2 protection that can be deployed by everRun MX Express is a HA solution that does provide component fault tolerance and downtime of up to approx 2 minutes should a full server host fail. This is without data loss or any IP clash / recovery scripting requirements.

Specialist training and knowledge, typically alongside outsourced support, is definitely required to manage and maintain these environments.
6. **Citrix Xen server / VMware / HyperV Based Resilience**

All three of the main hypervisor providers, Citrix, VMware and HyperV provide HA resilience. Citrix Xen server (without the add-on of everRun which incorporates an OEM version of Citrix Xen server), can provide what is termed as a basic level 1 failover solution. Any failure at component or full server level will incur downtime as the passive server reacts to the failure. This solution is dependent on the use of shared storage and of course shared storage (in the absence of costly and complex SAN to SAN real time failover) is itself a single point of failure and hence definitely not a fault tolerant solution.

VMware FT can provide true software based fault tolerance, and failure at component or full host level will be tolerated with zero downtime and data loss. VMware FT is limited to the protection of single core applications only though. Also this solution requires the deployment of shared storage – and again as above, shared storage (in the absence of costly and complex SAN to SAN real time failover) is a single point of failure.

VMware can provide the basis of a very high level of resilience for multi core applications, but is not really deemed Fault Tolerant. Two very high end servers with VMware standard licences or above provide HA levels of resilience, with failover to a passive server that reacts to failures as described above. By deploying VMotion, VM environments can move instantly from one physical host to another; however this relies on BOTH hosts being available and therefore is not automated fault tolerance. This model also relies on an independent PC or server to control the whole environment which is itself a single point of failure.

Microsoft HyperV provides levels of high availability automatically as part of the solution and is available to push out up to 4 x VMs when purchasing MX server enterprise licences. A standby server can be nominated as a failover machine in the event of the main HyperV server failing (component or server failure would count as a server failure in this model). There is downtime and manual intervention is required, therefore defining this as a HA solution and not Fault Tolerant.

Specialist knowledge is required to manage and maintain these type of implementations.

7. **Clustering Solutions**

Clustering solutions provide very high levels of availability, typically only relevant to the environment that is being protected (i.e. SQL clustering protects only SQL). Several host servers can be pooled in a cluster but all rely on the deployment of shared storage which, as detailed above, is a single point of failure and hence not a fault tolerant solution. They are typically complex to deploy, manage and maintain. Environments requiring zero downtime in the event of component and / or full host failure cannot rely on a clustering solution to protect the most critical of environments. Clustering typically provides network resilience but there will be some downtime if storage or full server fails without deploying other complementary solutions - managed switch layers, resilient storage units that can auto failover and complex multiple lines of connectivity between all components in the resulting architecture. Even then, complexity inevitably results in some level of downtime should a component or full host failure occur, especially at recovery to a protected status. Typically, considerable custom scripting is also required relevant to the applications and
the individual configuration of the environment. In short, it is very complex and downtime is still likely to be incurred.

Additionally, considerable specialist knowledge is required typically with outsourced support.

8. Summary Table

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>Hardware FT</th>
<th>Software FT</th>
<th>HA solutions</th>
<th>Hypervisor HA</th>
<th>Hypervisor FT (VMware FT)</th>
<th>Clustering solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single application O/S environment</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Runs applications without modification</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>No scripting required</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Automated fault handling and failback to protected state</td>
<td>YES</td>
<td>YES</td>
<td>Varied</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Geographic separation</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>Limited</td>
<td>Limited</td>
<td>Limited</td>
</tr>
<tr>
<td>No downtime for storage failures</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>No downtime for server failures</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Zero loss window for disk data</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>No SAN required</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Multi-processing (Virtual CPUs)</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Support for Linux</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Support for VMware</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>VMware only</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Support for Hyper-V</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>Hyper-V only</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Works with commodity servers</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>No downtime for most planned events</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

1Typically, applications not addressed by the HA add ons require customised configuration work
2Scripting needed for many non-cluster-aware applications
3Some HA solutions do automate fault handling and restore (eg Neverfail), while others do not (eg DoubleTake)
4SAN distances only (VMware specific); Site Recovery Manager (SRM) requires asynchronous data replication that introduces the chance of data loss. SRM is a D/R solution rather than a local availability solution
5SAN distances only; Multi-site clustering requires asynchronous data replication that introduces the chance of data loss. Multi-site clustering is a D/R solution rather than a local availability solution
6At time of writing (October 2013), not possible – but Linux supported software Fault Tolerance using everRun is included in the next release

Chris Robinson
Sales & Technical Director
24/7 Uptime Ltd