

High Availability & Disaster Tolerance

-

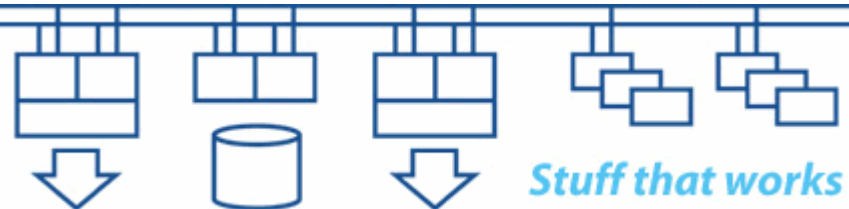
Systems Design

Colin Butcher

Technical Director, XDelta Limited

- **Occam's Razor:**
“Pluralitas non est ponenda sine neccesitate”
“Entities should not be multiplied unnecessarily”
“Keep it as simple as possible”
- **Hanlon's Razor:**
“Never attribute to malice that which can be adequately explained by stupidity”
- **Colin's Caveat:**
“Allow for failures – success is only one of many possible outcomes”

- **Business continuity encompasses everything you need to do in order for your business to continue to operate in the event of a problem or disaster.**
- **If any of your systems are mission critical then you need to invest wisely in appropriate high availability and disaster tolerant facilities.**
- **You need good information in order to make sensible decisions, especially when things go wrong. Put systems and processes in place to get you that information.**
- **Technology is only part of the solution.**



- **Mission critical systems need to be non-stop during the “operational window” – not necessarily 24x365.**
- **Need to understand the impact of failures on the business. What level of data loss can you live with if you have to?**
- **Requirements never remain static over an extended period of time, so we need to be able to make changes during the operational lifetime of the system.**
- **Circumstances change, so we often need to be able to extend the operational lifetime of a system with minimal risk. That may include further significant changes to both scale and functionality.**

Mission critical systems need to be able to:

- **Survive failures (resilience and failover)**
- **Survive changes (adapt and evolve)**
- **Survive people (simplify and automate)**
- **Never corrupt or lose critical data (data integrity)**

- **High Availability (HA)** – typically applies within a site and is primarily concerned with ensuring continuous service and data integrity during the “operational window”.
- **Disaster Tolerance (DT)** – typically encompasses multiple sites and is primarily concerned with the ability to survive major and potentially extended outages.
- **Many mission-critical applications need both HA and DT – which is a serious challenge!**

- **Reliability = Probability of failure at a given point in time, usually expressed as MTBF (Mean Time Between Failures)**
- **Availability = Probability of system being up and running at the instant when need it. Function of MTBF and MTTR (Mean Time To Repair), usually expressed as a % uptime , eg: 99.999%**
- **99.999% uptime (five nines) is equivalent to 5.26 minutes loss of service in a year for a 24x365 system. For a 12hour x 5day operational window it's only 1.87 minutes permissible outage in a year. That's difficult to achieve.**

- **Safety-critical systems (especially safety-critical real-time monitoring and control systems such as air traffic control) require exceedingly high levels of availability. They also have to be fail-safe in order not to endanger lives.**
- **True 24x365 mission-critical systems are fairly rare. With these there is no “downtime window” to take backups, fix faults or to make changes. So, whatever you do has to be done “live” – and very carefully!**
- **The closer you get to 100% uptime the more expensive a satisfactory solution will become.**

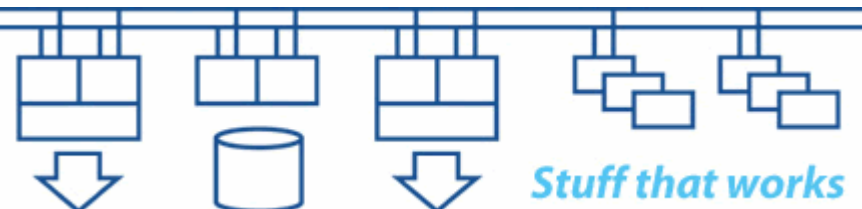
- **What is the maximum permissible pause in service (however brief) when a failure occurs?**
- **Implement automated monitoring and failure detection.**
- **How do you decide when to declare a failure or disaster?**
- **Understand the performance characteristics of the application and the implications of those characteristics on your infrastructure.**
- **Failover performance characteristics are determined by:**
 - **Latency** – **determines response times**
 - **Bandwidth** – **determines throughput**
 - **Sample time** – **determines “event detection time”**

- **What is the impact of a failure (lives, career, reputation, money etc.)?**
- **How long have we got in order to be able to continue without disruption, data loss or data corruption?**
- **How long have we got in order to be back and running normally ready for the next failure?**
- **Are there any “maintenance windows” in which we can work on the systems?**

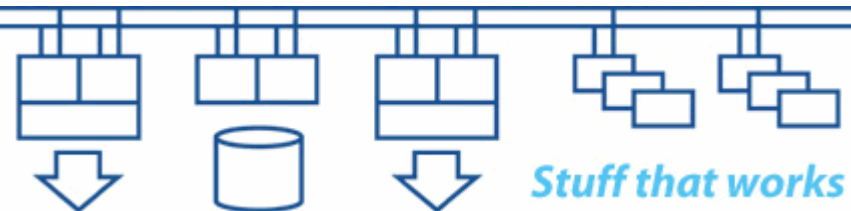
- **Can we design the system (or at least the essential core) to be inherently ‘fail safe’?**
- **Can we prove that we’ve considered all the components and the possible failure modes?**
- **Can we automate the critical parts of the failover process such that it behaves in a predictable manner?**

“Survivability test”

Cause of Outage:	Planned (Maintenance)	Unplanned (Failure)
Hardware	?	?
Operating System	?	?
Network Layer	?	?
Layered Products	?	?
Application Software	?	?
Application Data	?	?
Environment	?	?
Staff	?	?

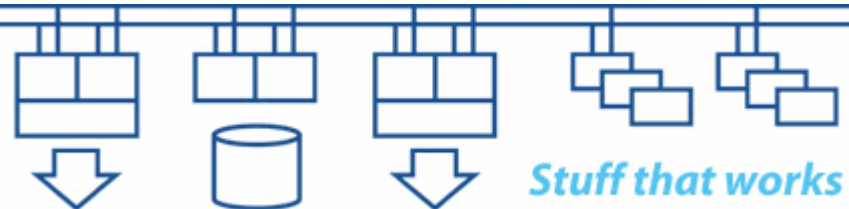


- **Use multiple systems within a site – and have appropriate physical separation between them.**
- **Use multiple sites with appropriate geographical separation.**
- **Understand what happens when a failure occurs and how the overall system configuration is likely to respond.**
- **Avoid the risk of more than one system thinking that it's in charge when a failure has happened.**
- **Ensure that the surrounding infrastructure and environment is appropriate to the needs of your systems.**
- **Configure the individual components of the systems in a manner that maximises interchangeability.**



- **Losing data is a disaster.**
- **Availability is more important than performance.**
- **Size storage subsystem based on minimum components and maximum estimated throughput.**
- **Segment storage subsystem to provide gradual degradation rather than wholesale failure.**
- **Need adequate backup capacity and throughput in order to meet permissible backup windows.**
- **Understand application behaviour and storage performance requirements (bandwidth and latency).**

- **Scale network so that overall performance is based on minimum essential number of paths and maximum estimated traffic.**
- **May wish to take advantage of installed bandwidth capacity to provide additional functionality when everything is working.**
- **There is no such thing as a single protocol network. Understand the behaviours of the different protocols under failure conditions.**
- **Duplicate devices such as network printers on different paths.**
- **Segment the network to provide gradual degradation rather than wholesale failure.**



- **Constant temperature environment**
- **UPS and generators**
- **Emergency lighting**
- **Physical security & access control**
- **Documented procedures & training**
- **Wall-board diagrams**
- **Remote monitoring & diagnostics (webcams etc.)**
- **Separate data paths for monitoring equipment**
- **Cable labelling, colour coding and tracing**

Design in the ability to make changes – ‘backwards compatibility’ is essential, so:

- **Establish the minimum requirements that have to be met**
- **Determine system boundaries**
- **Decouple system components**
- **Use standard interfaces wherever possible**
- **Define and enforce interface definitions**
- **Establish and enforce conventions**
- **Keep everything as simple as possible**

Automate the processes using documented mechanisms wherever possible:

- **Software build**
- **Software installation**
- **System configuration**
- **System startup**
- **System administration (errors, logs, queues, users, printing, backups, archive etc.)**
- **Application monitoring**

Comprehensive, well planned and thorough testing is crucial, especially for safety critical systems:

- **Representative test environment and data**
 - **Functionality**
 - **Performance**
 - **Scalability**
- **Upgrade / regression procedures**
- **Equipment replacement planning and procedures**
- **Data validation and verification**
- **Allow for regular testing – you're only as good as your last successful test.**

Invest extensively in:

- **Good people**
- **Training and appropriate test / simulation facilities**
- **Documentation and revision control**
- **Procedures and scenario drills – not only doing the right things, but doing them in the right order**
- **Planning for future changes**
- **Problem investigation and resolution**

- **Understand the problems and know your objectives**
- **Sound technology is important – proven technology is generally better than the latest available technology**
- **Thorough analysis and good design are important**
- **Implementation is critical:**
 - **A poorly implemented HA / DT solution increases business risk and lowers availability**
 - **A well implemented HA / DT solution is an asset to the business and can be a significant competitive advantage**
- **Good people are essential.**

- **You have to invest serious time and effort into making systems HA / DT capable. You cannot buy it off the shelf.**
- **The key problem is being able to continue to deliver service to the users – and understanding exactly what has to be done in order to achieve that.**
- **Continued awareness, training, planning and testing are mandatory – you never know when disaster will strike.**
- **Don't assume that it will work just because you've spent money on a technology solution. That's only part of solving the Business Continuity problem.**

Colin Butcher, XDelta Limited

Office: +44 117 904 8209
Cellphone: +44 7768 857615
E-mail: colin.butcher@xdelta.co.uk
Web: <http://www.xdelta.co.uk/>