

Damm them alligators!

By

Paul Courry

HP certified swamp drainer

Hughes Network Systems

11717 Exploration Lane

Germantown, MD. 20876

(301) 428-5939

fax (301) 601-6303

pcourry@hns.com

paul@courry.com

The old joke is that “Employees should anticipate problems, be prepared with solutions and implement them at the request of management. However, when one is up to one’s allegories in alligators it is awful difficult to remember that your original purpose was to drain the swamp”. Today’s talk is a case study in how to drain the swamp or in managementese A Case Study in Proactive Error Detection and Management.

What I want you to walk away with today is new ideas and perspectives on how to increase your uptime and make your downtime less expensive based on what one company is actually doing day in and day out 24x7.

Each of you should have a CD (subject to approval from the legal department) that contains this paper, the graphics and most importantly, the actual code we use. If you file off the serial numbers and adapt the code to your environments you should be able to implement the ideas presented here today. Take them home and make them your own.

1. The Swamp

- 24x 7 environment

Who here has the luxury of shutting down the systems anytime they want? Remind me to give you my resume.

Many, if not most of us, now operate in a 24x7 environment where not only are our systems *defined* as “mission critical”, but are in *use* 24 hours a day, 7 days a week. The cost of downtime can be measured in thousands of dollars per minute. Some of you measure the cost of downtime using a few more zeros; 10’s of thousands, 100’s of thousands of dollars per minute.

The pace has also picked up. We no longer have the luxury of plenty of unused wall time in which to solve problems and get things fixed. It seems as if we are on a treadmill where we are trying to squeeze every drop of productivity from our machines. Businesses no longer close at 5:00 p.m., the world wide web has seen to that. Hughes is open 24x7, providing information, taking orders, processing batch jobs.

In our particular case we have plants in two states, spread across two countries and a few communications satellites to distribute information the people who install the satellite dishes we

make. By the way, we are Hughes Network Systems. Makers of DirecTV, DirecPC and a lot of gee-whiz phone systems scattered around the world.

- **Home grown, highly customized software**

Who here runs plain vanilla software, just as it came from the manufacturer? Let me when you are hiring. Just about all of us run software that has got about 10 years of customizations layered on top of the plain vanilla package, *minimum*. Some of you can't upgrade your software because it is no longer compatible with what the manufacturer is putting out and it would cost too much to make it so. Yes, my company is guilty of that. In fact my company is guilty of every darn thing I mention in my case study today. Hey, it wouldn't be a case study if this was not *our* swamp, *our* alligators and *our* attempt to drain it, but I'll bet dollars to donuts that most of you are grinning inwardly because it's your swamp too!

Once your packages get "customized" the alterations take on a life of their own, literally, and require care and feeding lest they bring everything to a halt. Special jobs, special programs, special interfaces grow like toadstools or should I say magnolia trees, choking up what was once a nice, clean, fully documented system. Documentation? Please don't get me started..... Well since you insist.

Got documentation? I don't. Most of us in this room don't. Most of the vendors out on the exhibit floor don't. WHY DON'T WE? Money, time, the other project we put off is overdue, as many reasons as there are people at HPWorld. There never seems to be enough time to document what we do, but hey it's just another mangrove tree in the swamp we create for ourselves. We always seem to be creating swamps for ourselves or for the person who follows us and others are busy creating *our* swamp at our next job.

- **Lack of staff**

Who has enough staff? Consistently, throughout the year, no holes due to people who quit, get fired or laid off? We have the same problem. We always seem to be behind the curve on manpower, always looking for just the right systems administrator or operator to fill that spare slot. It's always late nights and weekends for us. Wearing these electronic leashes so we are always reachable (they hope) for when the system goes down. And we know it always goes down right? I mean nobody has a perfect system that never hiccups, goes down or freezes during year end close? It's nice to know we *are* needed, but every night? At 2:33 a.m.?

- **Multi-national, multi-timezone operation**

Do I need to say much about this one? Everyone here knows how multiple time zones complicate our jobs. Some of us here have operations spread across more than one country with the problems that creates such as different retention periods, varying privacy laws, legal safe harbors, etc. etc.? The legal environment alone is a nightmare. Remember, you too could have a SWAT team come crashing into your computer room.

- **Highly integrated redundant networking requirements**

We use satellites for communications between the various parts of our organizations. We also use a MAN, are spread over two countries and 2 states. Even the sneakernet has two pairs of sneakers. I'm not sure that the redundant lines of communication we have are more of a hindrance than a help, but there sure do add a layer of complexity.

- Erratic maintenance windows

How many of you have nice regular maintenance windows that give you all the time you want? I envy you, tell me how you managed it. Around our place it's like herding cats to get everyone to agree to a maintenance window.

Not having regular maintenance windows or sometimes not having maintenance windows at all complicates things greatly. It's like watching your radiator boil over, your gas gauge hover on empty, driving through a bad neighborhood and thinking "it'll hold until I reach the freeway". Only if you're lucky.

When you need to increase database sizes, reboot the net, apply patches or reboot the computer nothing is so necessary as LIMIT 0,0 and the last thing you need is a shop floor manager who says "weelll I don't know, how about 30 minutes today and 30 minutes tomorrow?" Sometimes there is nothing I need so much as a big club to beat some sense into people who otherwise seem so intelligent.

Does anybody want to add anything else to this description of the swamp we all share? Surely, *someone* has a good story to tell? Okay, we'll move on.

2. The Alligators

- How to manage for uptime
 - Commitment from upper management

First off, you are doomed unless upper management backs you. Without their backing you will constantly be fighting a rear guard action on time, money and personnel. It is those things that will drive what you are able to do. If they can be made to understand the cost, the true cost, of downtime they will see the business reasons for minimizing it. At our place of business there is visibility to the highest levels with written escalation policies that wake managers out of bed. This is an example of their commitment. On the downside, our bonuses are based in part on how well we keep the systems up and running.

- The importance of record keeping

It is said you can't know where you are going unless you know where you have been. This is extremely true in computer operations. You need a program running 24x7 on the computers to capture data on how the machine is being used. You need to keep careful records of uptime and downtime. You also need to spend time on massaging this data so you know what you are looking at. I would also personally suggest that this information be available to everyone in the company, ours is. In fact, it is published every month in easy to read format.

- How to minimize downtime

Minimizing downtime consists of a number of tricks that together get the *cost* of outages down as well as the *number*. The cost of an outage is just as important as how long it is, in fact more so.

- How to turn unscheduled downtime into scheduled downtime

Turning unscheduled downtime into scheduled downtime consists of catching problems far enough in advance so that notification can be given to involved parties, negotiations held and preparations made so that the business impact is minimal.

- Why this transformation is important

This transformation is important because scheduled downtime costs less per minute than unscheduled downtime. As an example, if the shipping manager knows that the computer is going down at say 3:00 p.m. he tells his people to get their computer work done early and schedule their non-computer work for after 3:00 p.m.. Everybody keeps working throughout the outage, the outage does not result in people sitting on their hands and the shipping manager does not have packages stacked up that can't go out until the next day.

In other words, sufficient notice allows shifting of tasks to minimize the impact of an outage and therefore it's impact on the business.

3. Draining the Swamp

- Management defined requirements

- Defining key risk areas

Where are your single points of failure? Where are your historical points of failure? What failures affect your business and which don't? You need to do some spade work in order to avoid being surprised some day.

- Promulgating a strategy to maximize detection and minimize risk

- Hardware

If you can afford it extra hardware on site as well as the experience necessary to install it, works wonders for uptime. A good hardware support contract either with HP or a 3rd party is also critical. We have a complete duplicate machine, down to the patch levels, to cannibalize from if we need to. Most times we simply call the Response Center and tell what board to bring out, meet them in the parking lot and swap boards. The machine has long since been repaired and is back up and running by the HP has arrived on site.

We also standardize on one type of drive and keep a spare drive installed on the machine, configured into sysgen and set up as a dummy volume. This means we always have ready a spare drive that has been burned in. In a true emergency we can even forgo the replacement of a bad drive and simply use Volutil to remove the bad drive, add the hot spare and restore the data.

Another example is mirroring of your MPEXL_SYSTEM_VOLUME_SET. Can't be done? Ask EMC. Apparently they can and we are busy, with HP's blessing, putting our system volume set over on the EMC. While there may be other ways we own several EMC frames so we choose the easiest path.

Don't forget those small items that can stop you cold like terminators, cables, particularly odd cables that HP might not stock locally, a spare printer you can use for parts, power

cords, power strips, spare SLTs, tools for god sake, get yourself a good expensive tool kit. An analog phone for testing modem lines, a copy of all the names and numbers of your vendors should be kept at home along with the serial numbers of the machines.

- **Training**

Training is extremely important. It is also the first thing that gets cut from the budget. I would urge those of you in the audience that have responsibility for budgets to think long and hard before cutting training budgets. In my opinion undertrained system administrators and operators, particularly operators, do more to raise the cost of doing business than anything else.

Examples from our site would include the system administrator who did not attend a particular 3rd parties training for their software package. We had the money for the package, but not for training. They accidentally blew away all copies of a production database. How about the operator who did not understand daily the backup processes and allowed a tape to be overwritten? Time could never be found to train them on the ins and outs of our backup process. How about the system engineer who did not understand the approval process for installing demonstration packages? For 5 months we were vulnerable to attack from the Internet. Thank god that no one did.

Moving on, it is a pity that HP does not offer Customer Engineer training anymore. A good system administrator armed with just a little knowledge can cut down the amount of downtime just by telling the HP CE what boards to bring out with them, thus saving a trip back to the office for parts.

Also, you have a set of free, repeat FREE, training tools available to you. They are called HP3000-L and the HP ITRC forums. If you can free up a little time each day for your personnel to peruse these sites they can learn a lot and from the very best in the industry too! In these forums people from all over the world ask questions and some of the best minds in the HP community answer them, it is fascinating what you can learn.

- **Network**

As far as networking goes we have duplicate NIC cards installed in the machine, but not yet configured. Cold swaps as it were. When a Network Interface card fails we do not have to reconfigure using SYSGEN, shutdown the machine, replace the card, and reboot. A long process as you know. We simply stream a job and have everybody back up in 10 minutes. It can even be done by an operator in your absence. How to we do that? We have 2 cold swap Network Interface cards installed in the machine. If a card goes down we simply stream a job that shuts down the network, copies a pre-built NMCONFIG file into place and restarts the network. Total time is 10 minutes. We schedule the removal of the bad card for a maintenance window.

- **Software**

- **Open View/ ITO**

HP's Open View/ ITO package is what we use to pass emails to our sys admins. There are other packages out there which I am sure will perform just as nicely, but HP's offering is what we use. It's ability to capture console messages and send emails to our pagers and cell phones is at the core of our strategy. Without it our all

programs fall apart. Capturing console messages specifically is the key to our strategy.

- **Hire a food taster**

Around our place, nothing goes on the production machine without first being tested on the development box. No patches, no software upgrades, no nuthin' goes on the production box until we are sure it won't affect production. In addition, all major changes that affect many users, such as changes to the security database, are done incrementally. This way if something is radically wrong I have the luxury of dealing with one hundred unhappy users rather 3,000 unhappy users. I may vote Democratic, but my production box is a conservative Republican.

- **Tweak your SLT tape**

We found the hard way once that our strategy of using two tape drives in parallel to shorten the backup time had a gotcha in it. The SLT we so faithfully cut every month did not have the two modules on it that were required by our backup software to do parallel backups and restores. When we went to restore the system volume set we found ourselves spending double the time to restore the data because we could only use one drive. The fix for this was to go into sysgen and add those two modules to the files that went onto the SLT. This meant that an install or update of the operating system from an SLT restored those two modules as well. Some of you may not be aware that you can modify what goes onto an SLT, instructions will be found in the appendix along with a short list of what you should add. I've tried to list what each backup software vendor considers critical for doing parallel backups.

- **The DEW Line (this is important)**

Just like the old DEW line, Distant Early Warning, which was run by the military to give early warning of enemy attack, so to is our DEW line designed to warn us and give us as much time as possible to plan and respond to potential crises. We don't believe that predictive support is the be all and end all for proactive management and monitoring of our systems. We feel there is a lot we as system administrators can do to clean our crystal balls and peer a few hours into the future. To give ourselves time to warn our management and our users of impending problems while we can still manage the consequences.

Examples. How many of you use Predictive Support? Well when it runs at 3: a.m. we parse the results and email the sys admins of any critical problems. This way we get a few additional hours of warning of full volumes, flaky hardware and other problems.

Every morning we all get a wakeup call from our pagers. Our computers run a test job which flashes a message on the console, which is in turn picked up by ITO and a page is sent out. It's nice to know that the early warning system is up and running.

Anytime we get a Bad UFID (Unique File Identifier) message, an indication that a file can't be accessed because the file system is going south, we are immediately paged so we can contain the problem while it is small. Bad UFID's can be caused by a bad drive, bad SCSI card, a problem with MPE, it is the canary in the coal mine, when the canary keels over you know you are in trouble and the same with seeing bad UFID messages on your console. Once the underlying cause has been fixed you can use FSCHECK in the Telesup account to clean up the bad UFID's. However, let me warn you that this is a truly dangerous program which has the capability of

corrupting MPE if you try to run it while any sessions or jobs other than yours are on the system. I urge you to only run it with one hand on the keyboard, the other hand should be firmly grasping a phone that has the Response Center on the other end.

If certain HP utilities are kicked off by MPE to take snapshots of problem areas we are paged to warn us of flaky hardware. Yes, HP does have some utilities out there that run automatically should the system hiccup. When these utilities fire up we get paged and we are immediately on the line to the Response Center asking for an analysis. Some of these utilities may run time after time, but you won't know about it. The only evidence will be the files the utility leaves behind. Again, see the appendix for details and yes, I'm adding to this list as I pry information out of HP. In fact I would like to ask your help. If any of you who happen to be here today or are reading this paper off the conference CD, know of any of these utilities and the files they create, please let me know.

Certain key words and phrases from many different sources kick off messages and that list is growing every day. The appendix contains that list and a description of what program generates the message under what condition.

By the way, everything I've discussed here is quite separate from our Job Rescue package from Nobix. Job Rescue is the package we use to catch software errors in our jobs. In many instances we are alerted via pager should certain jobs fail. Likewise, jobs that are the responsibility of certain members of the application team page these team members when they fail.

The whole thrust of the DEW line is to give us as much warning as possible of impending problems. To allow us the time to find way to solve a problem at the least possible cost to the company and with the minimum disruption. A bad stick of memory that is giving correctable single-bit errors is a different problem than a bad stick of memory that has deteriorated to the point to where it is giving double bit errors and crashing programs. A whiny disc drive that has exceeded the soft error threshold is a different problem than a disc drive where the motor finally conked out. One type of problem can be examined, discussed and replacement scheduled at minimum disruption, the other is a crisis that results in workers sitting on their hands waiting for the computer to be fixed.

Attached to your copy of today's talk is a complete list of what we look for currently. This list is always being expanded. Depending on the opinion the user community has of this approach I'll be throwing up a website detailing what we have done, how to implement it and how we are improving it. I would be most appreciative to hear if any of you are doing similar things with perhaps different packages.

Appendix

For further information contact Paul Courry at PAUL@COURRY.COM

For a complete listing of all items mentioned in this talk go to WWW.COURRY.COM