IT Recovery Testing

***Written by Mel Gosling in 2007**: "You may have what looks like an excellent recovery plan for your IT systems, but have you actually tested them to make sure that they work as expected? How long would it really take to recover your most critical systems?"*

Helen Jones, the CIO of XYZ Industries Inc., was feeling pretty pleased with herself. She had been recruited after the previous CIO had been forced to resign following a poor audit report which revealed the organisation's IT recovery plans were inadequate and poorly documented. She had just completed developing the last of the IT recovery plans for the company's critical systems, and she was looking forward to this year's audit, and maybe a pay rise.

It was just then, at 10am on that Monday morning, that Helen's Operations Manager came rushing into her office with the news that their sales order processing, materials control, and manufacturing system, EMan, was down. Apparently their SunFire server, running EMan, Solaris and Oracle, had failed, and an engineer had been called out. Helen was confident; she knew that the recovery plan for EMan was well documented, and fully expected that the system would be recovered well within the 48 hours that had been established as its Recovery Time Objective (RTO).

The call out time for the engineer was 2 hours, and in case the system and its data needed to be recovered from the backup tapes once the server was fixed, the recovery plan stated that these tapes were to be brought back to site as soon as the engineer was called. The delivery time from the company that stored the backup tapes was 2 hours. In the meantime, the users had been advised of the failure and had reverted to manual processing until EMan was back up and running.

The engineer and backup tapes both arrived, as planned, at 12 noon, and the engineer immediately started investigating the cause of the failure, and within 2 hours had replaced a number of parts in the server and was ready to re-start it. All seemed to go well, and the system came back up as expected. However, just as an integrity check on the system was being made the server failed again.

The engineer spent a further hour investigating this latest failure, replaced some more parts, and brought the system back up again only to see it fail during the integrity check. It was now 4pm, the system had been down for 6 hours, and the users wanted to know when they could expect it to be back up and running. Although they had agreed to the RTO of 48 hours, they didn't expect the system to be out of action for more than a few hours unless there had been a major incident, and to them, a server failure didn't sound like a major incident.

By 5pm the engineer announced that the server could not be fixed and would have to be replaced. XYZ Industries had a maintenance contract with its server supplier that provided a replacement on site within 24 hours, and a full system recovery on a new server had been estimated at 8 hours. If all went well this would mean that EMan would be up and running again by 1am on Wednesday morning, ready for the early shift to use the system when the

staff arrived at 6am. This would mean an outage of 44 hours compared to the RTO of 48 hours – so far so good.

The replacement server was ordered, but unfortunately the maintenance contract had not been kept up to date with XYZ Industries' requirements. The contract specified a replacement server with six 146 GB SAS drives, whereas the data held in EMan had grown to require eight 146 GB SAS drives. It was not a major problem, but acquiring the additional drives would take an additional 12 hours. It was then that Helen decided to contact her CEO to advise him that the RTO for EMan of 48 hours could not be met, and that she expected it to be 51 hours. This meant that EMan was going to be available by 1pm on Wednesday. The CEO was not pleased, but it could have been worse.

At 2 pm the next day, on the Tuesday, Helen's Operations Manager came into her office to break some bad news. Apparently, the new server had fallen off a pallet when it was being loaded into a vehicle and had been badly damaged. A replacement had been ordered, which would arrive in 20 hours, by 10am on the Wednesday, putting the recovery of EMan back to 6pm on Wednesday. On hearing the news the CEO made the decision that staff would be asked to work through the night on Wednesday to clear the processing backlog ready for a clean start on the Thursday.

The new server arrived at 10am on the Wednesday as planned, and Helen's support staff started to load Solaris. Unfortunately, her support staff did not have a copy of the latest release of Solaris, which was not discovered until it had been loaded. Upgrading to the latest release took a couple of hours, and by 1pm the server was running with the correct release of Solaris. Helen advised her CEO that the system would not now be available for staff to start clearing the backlog until 8 pm.

The data load from the backup tapes commenced at just after 6 pm, but after several attempts it was discovered that the first tape was unreadable. This meant that none of the backup could be used, and the previous night's backup would have to be obtained and loaded, such that XYZ Industries would lose a full day's processing in addition to the 4 hours that had been lost from the start of the shift at 6am on the Monday morning. It was now 7 pm, and with a two-hour wait for the backup tapes plus a four-hour data load, EMan was not going to be available until 3 am on Thursday. Helen contacted the CEO again and the staff that had been put on standby to work through the night were told that they would not be required. To say that the CEO and the users were angry would be an understatement. Helen was going to be lucky to keep her job after what had happened.

The previous night's backup tapes duly arrived at 9 pm, and the data load commenced. After waiting for an hour to make sure that the tapes were readable and the data was being loaded, Helen decided to go home and get some well earned sleep before coming in early to tell the users and her CEO that EMan was back up and running again.

Helen was fast asleep when her Operations Manager called her at home at 3 am to tell her that only 25% of the data had been recovered so far, and that after taking advice from the suppliers of EMan had estimated that the total data recovery time would be 24 hours because of all the checks that the application was making and the large quantity of data that was being restored. Helen decided to get up and make her way to work ready to greet the 6 am shift with the bad news – and to call her CEO at home before he started for work.

The data were finally recovered by 10 pm that evening, Thursday, slightly later than estimated, and the RTO of 48 hours had, in reality, turned into a downtime of 84 hours. However, because of all the failures to recover the system when promised, the CEO had not organised any overnight working and so the first time that EMan was going to be used would be at 6am on the Friday – an effective downtime of 92 hours.

Is this a familiar story? To me, it is. In my work as a business continuity consultant I am constantly coming across organisations that have developed and documented IT recovery plans for their critical systems – but have never tested the plans. I have also lost count of the number of times that I tell CEOs and CIOs that their IT recovery plans should be tested, only to receive a grudging "Yes, of course you're right, but we haven't got the time right now, and it's going to cost quite a bit of money".

So, while accepting the principle that testing is a good thing, many organisations tend to carry on in the naïve belief that their systems will be recovered as planned. They will probably find a need to recover their systems from minor failures from time to time, and find that this can be achieved well inside the agreed RTO. This gives a false sense of security. Typically, these recoveries will be from such things as power failure, and will not entail the need for a full recovery of a system from scratch.

In the past few years I have had three examples of clients that have ignored my advice on recovery tests only to find that in reality it took a lot longer than the publicised RTO to recover a critical system:

- An insurance broker that lost their main line of business system only to discover that all their backup tapes were unreadable. Instead of the RTO of 72 hours it took 3 weeks to recover the system after sending the tapes to a company specialising in recovering data from unreadable tapes.
- A local government authority that took 5 days to recover its Email system that had an RTO of 24 hours.
- A manufacturing company that lost a critical system for a week when the IT people had stated that they could recover it on to a spare server within 24 hours.

The message is clear, simple, and unambiguous. Make sure that you test the recovery plans of your critical systems. The first time you attempt to recover them, it will invariably take longer than expected, and it's not a good idea to find this out when the system is being recovered for real.

Did Helen lose her job? Fortunately, before she went in to see the CEO first thing on the Friday morning she remembered that the consultant that she'd used in helping to draw up the recovery plans had advised her that they should be tested, and six months before she had sent an email to the CEO proposing a recovery test for EMan. He had rejected the idea for the time being as it was going to be too expensive, and Helen had found his email reply just before she was summoned into his office!