

## **Failure mode analysis in data centres and the benefits of operational training**

**B J Elliott**

Most of the data centre standards we have at present such as the TIA 942, EN 50174-5 and the forthcoming BICSI 002 and the Dutch NEN proposal 38188 focus on getting the initial design correct. This is of course vital but once the data centre is built and up and running it must also be operated in the correct manner.

There are currently no standards dedicated to this although the UpTime Institute is preparing a Management and Operations review document that approaches operational management in a manner that does not degrade the original Tier rating of the data centre.

Other standards such as ISO 27000, Information Security Management Systems, touch on the subject as do many other standards focussing on Business Continuity but the approach tends to be very broad.

We at Capitoline are developing our own data centre management and operations methods largely based on our work with the Amsterdam Internet exchange and other data centres.

We thought a good place to start would be to try and analyse why data centres do go wrong. Information on this has been published before but usually by manufacturers who have a specific interest in justifying a demand for their own products or sometimes by users such as Google who are not in a hurry to give much away about their own shortcomings. As a result information tends to be varied and with no common reporting terminology.

Our approach was to roundup up all press articles published over the last thirty months in on-line trade journals all over the world. Their sources in turn were very often the 'status dashboards' published by the data centre operators for their own customers.

Over a thirty month period 32 major failures were identified. A 'major failure' is here defined by us as something that took down the entire data centre or at least rendered its main operational status as unusable for one or more major customers.

With 32 failures in 30 months this is just under once a month. But this is only for those incidents publically announced. If we take an assumption that this is only half of all incidents then it means there is a major operational incident at a data centre about every two weeks.

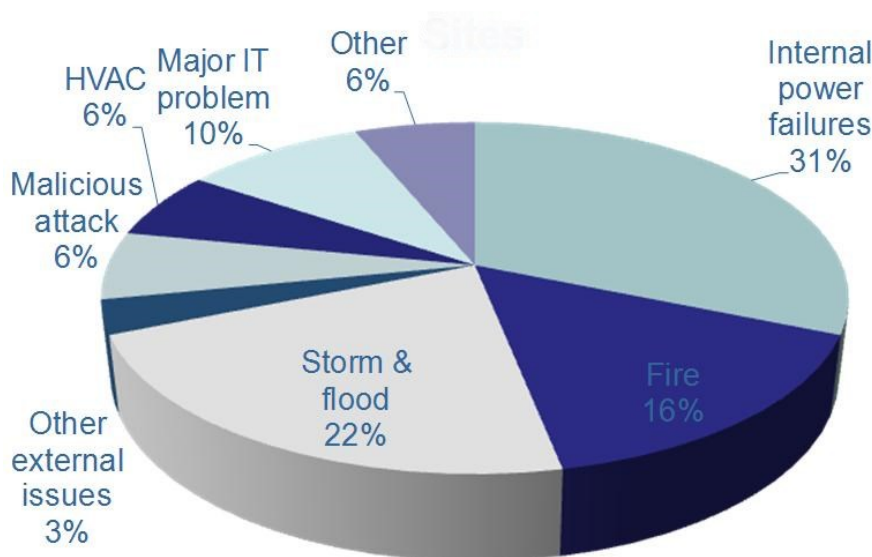
What does this represent in terms of failure rates? To know this we would need to know how many data centres there are in the world. This is probably an impossible question because at what point does a small server room become defined as a data centre? Do we include completely private data centres as well as all colocation centres? Do we define a data centre with eight 1000 m<sup>2</sup> data halls as one or 8 data centres?

CBRE put the amount of first class data centre space in Europe at about half a million square metres of computer room floor space. If we accept this represents about 25% of the world total we would then have about two million square metres in total. In our experience a major data centre has an

average of 2000 m<sup>2</sup> of floor space which gives us a figure of about 1000 major data centres in the world. Failures at a rate of 24 per year then give a chance of major failure at about 2.4% or one in forty, per year. As a statistic figures like this have to be seen in perspective of course. A well designed and run Tier 4 data centre may last 40 years without major failure whereas a poorly operated data centre will fail every year.

Another interesting figure to be gleaned from our survey shows the average length of downtime after a major incident has occurred. Of the 32 failures we investigated 24 of them reported actual outage times. The figures varied from 48 hours to 30 minutes with an average downtime of 14.7 hours per major incident. Not surprisingly major incidents such as fire, flooding and total power loss take some time to put right.

The final area we looked at was the cause of the outage.



(c) Capitoline 2010

Source: Capitoline from published sources 2008-2010

From our analysis we can see that power problems are the main cause of failures at 31%. This includes power failures that started with the utility supply but where the internal data centre back-up system failed to respond correctly. This includes generators failing to start for numerous reasons and multiple generators failing to synchronise.

Storm and flood damage came second at 22% and must make data centre operators think carefully about location, building design and lightning protection.

Fire was the next major event. This includes fires within the data centre and centres also taken out by buildings on fire in close proximity.

We were surprised that HVAC problems only came in at 6%, about the same as external malicious denial of service attacks and less than major IT equipment problems.

Other external issues included events such as the evacuation of a data centre due to a leaking gas main outside the building.

At least three quarters of the problems could have been avoided through the use of better design and operational practices, not least a proper testing, under load, of all the back-up power supply equipment.

Our approach to data centre operational management thus takes in the following key headings:

- Understanding failure mechanisms
- Good Housekeeping within the data centre
- Optimised equipment layouts
- Understanding the power train
- BMS and status monitoring
- Policies and procedures to manage a data centre
- Essential maintenance
- Health and safety requirements
- Fire policy
- Physical Security policy

We at Capitoline continue to develop a strategy for good data centre operational management based on real-world statistics and feedback from data centre auditing.

See our web site for details on our Data Centre Design (DCD) and Data Centre Operational Management Course (DCOM).