



# Arcati Data Centre Infrastructure survey 2006



by Mark Lillycrop, Arcati Research

The following data has been compiled from Arcati's Data Centre Infrastructure survey, conducted in May/June 2006.

The objective of the research was to establish what level of availability is currently being achieved by enterprise data centres, and what challenges they face in delivering a 24/7 service to their businesses and customers.

The survey is based on responses from data centre managers at 47 large user organizations, split roughly 40/40/20 between USA/Canada, EMEA, and other regions. 48% of respondents

were running internal data centres and 7% external hosting services, while 45% reported a mixture of internal and external customers.

As can be seen from Charts 1 and 2, our respondents were primarily from large organizations, spanning a broad range of business sectors – with a particularly large representation from the banking, insurance and transport sectors (as well as a substantial number of dedicated IT service providers).

We asked respondents how they would classify their facility on the Uptime Institute Tier Scale,

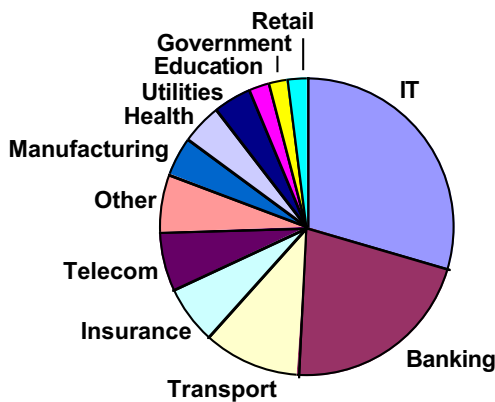


Chart 1: Business sectors of respondents

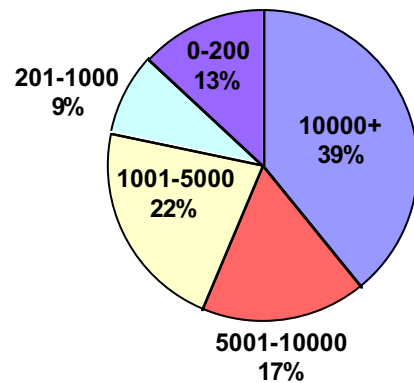


Chart 2: Number of users served by data centre

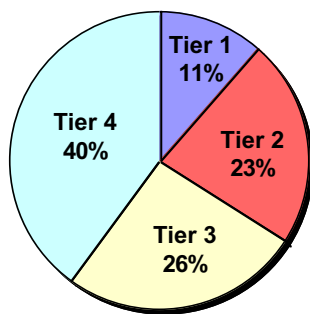


Chart 3: Tier level of surveyed data centres

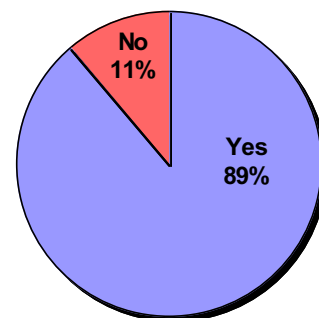


Chart 4: Is your data centre committed to providing a minimum uptime/service level to your organization/customer?

ranging from Tier 1 data centres (with the lowest levels of availability) to Tier 4 (with a highly resilient infrastructure and no single points of failure). As shown in Chart 3, most respondents put their data centres into the higher categories, as we would expect from the high proportion of large mission-critical sites represented.

Most data centres managers had a formal agreement in place to provide a minimum level of service to their internal or external customers. Of these, nearly two-thirds said that they “usually” meet their service-level objectives, and only one in three claimed to reach the target level consistently. A very small number felt that they rarely met service levels (Charts 4 and 5).

63% of respondents (Chart 6) believed that their data centre availability level has improved over the last three years, with 35% reporting a constant

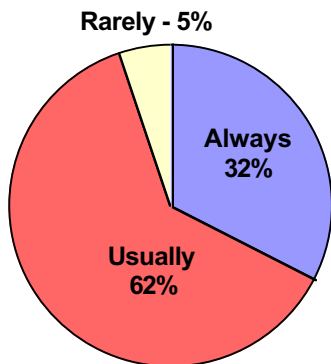
level of availability. Thankfully only one respondent noted an overall decline in uptime!

83% of respondents (Chart 7) said they have a back-up data centre to which they can switch their workload in the event of a failure, and of these the majority were internally managed facilities. An encouraging 91% said they have a formal disaster recovery in place (Chart 8), though our experience suggests that this proportion declines significantly among smaller SME-level data centres.

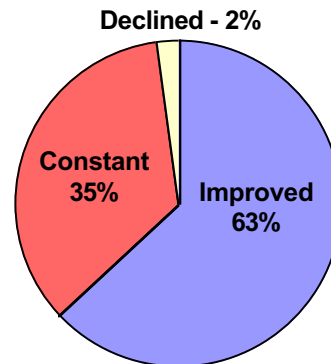
### Challenges and causes of failure

The survey delved a little deeper into the extent of unplanned downtime, the reasons for these failures, and the specific challenges facing data centre managers.

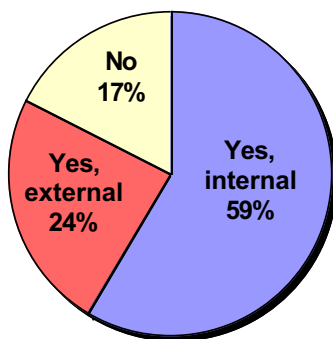
30% of respondents had made it through the last



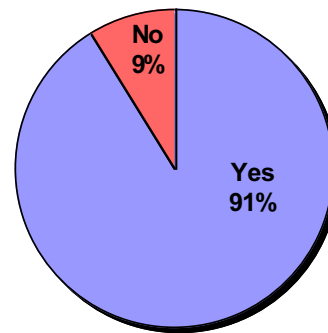
**Chart 5: How often do you meet the agreed service level to customers?**



**Chart 6: Has your data centre availability improved over the last three years?**



**Chart 7: Do you have a back-up data centre?**



**Chart 8: Do you have a formal disaster recovery plan?**

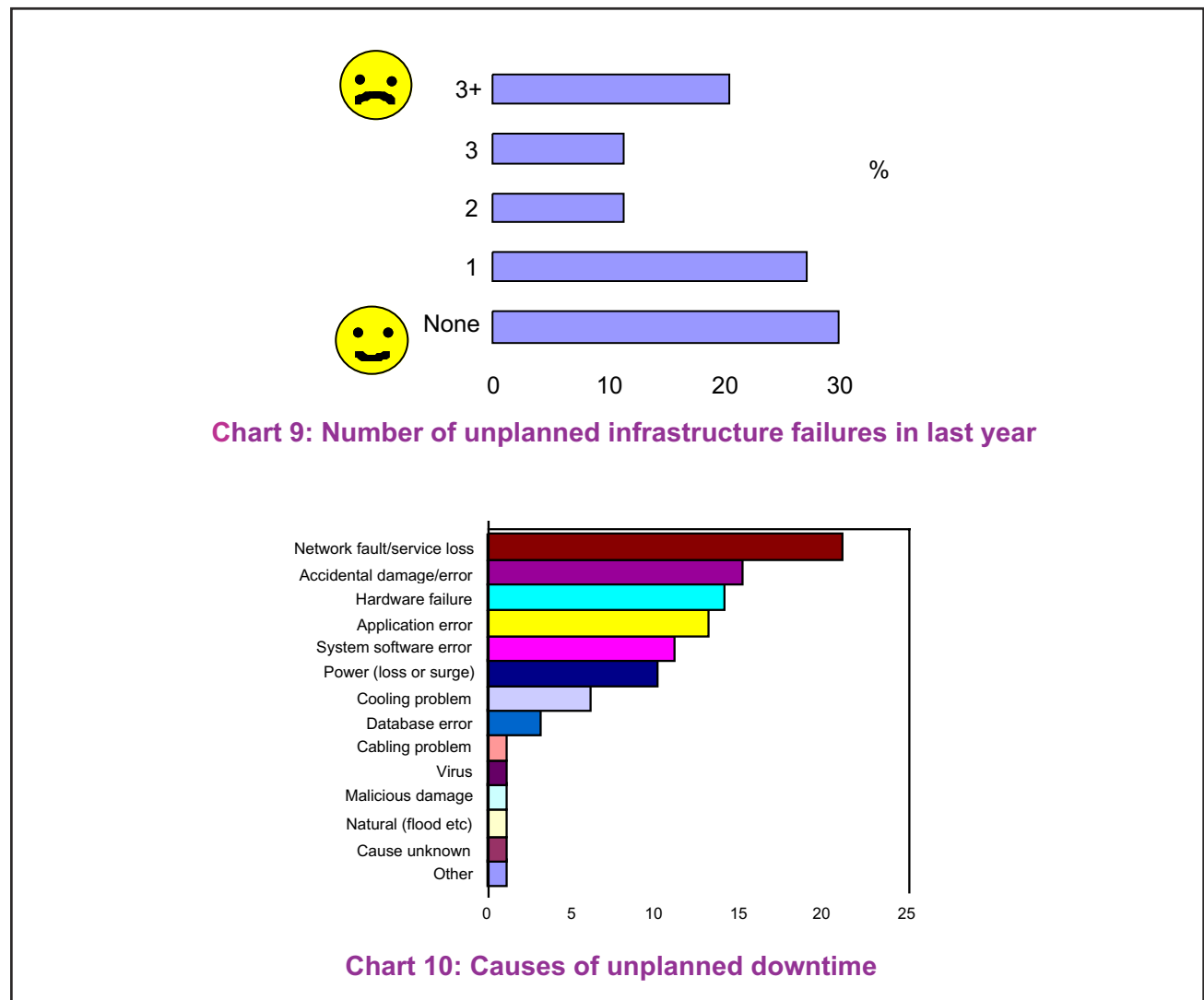
year with no unplanned downtime (Chart 9), but a slightly greater proportion reported three or more infrastructure failures. Interestingly, these high failure rates were just as prevalent among Tier 4 sites as elsewhere, and there was no discernible correlation among these sites between availability levels and the selected Uptime Institute classification!

Reasons for failure (Chart 10) were varied, but network problems or loss of network service were clearly the most widespread, with accidental damage and hardware and application problems close behind. Infrastructure issues such as power loss and cooling problems caused an appreciable number of failures, but other 'high profile' concerns like viruses and malicious damage accounted for a very small number of downtime cases.

Respondents were asked to list the greatest

challenges facing their data centre (Chart 11). The result closely mirrored the causes of downtime discussed in the previous question, as might be expected. Cost management came top of the list: although cost itself is rarely the cause of infrastructure failure, insufficient funds are often cited as a challenge and are regularly blamed for inadequate safeguards which in turn lead to unexpected data centre availability problems. Operator error and hardware/software and network reliability came next in the list of challenges; power, space and cooling once again occupied the middle ground; while viruses, natural disasters and physical security caused relatively little concern.

Most of the sites covered in our survey were heterogeneous in nature, with a broad range of enterprise servers in place. We asked respondents how the various platforms compared in terms of



overall reliability and availability. As can be seen in Chart 12, the mainframe rated far higher than other systems, meriting an availability 'score' of over 98%. Unix and Linux servers both scored over 77% and Windows just under 60%.

We also asked how respondents rated the service they received from power and telecom providers (Charts 13 and 14). There were no apparent problems with continuity of power supply, but telecoms presented a more mixed picture, with 16% rating their service as 'just acceptable' and 2% 'poor'.

### Future directions

The survey also explored trends and directions in

the data centre, and specifically asked respondents whether they were using grid computing or blade technology as a way of managing rising costs or reducing complexity (Charts 15 and 16). Grid applications were in use among just 15% of the sample, and these were distributed among various industry sectors. Grid has become well established in the scientific world, but among financially-oriented businesses its use tends to be restricted to data mining and processor-intensive applications.

Clearly blade technology has caught on much more rapidly, and the consolidation of distributed systems onto industry-standard blade servers addresses many of the system management issues that have plagued enterprise data centres

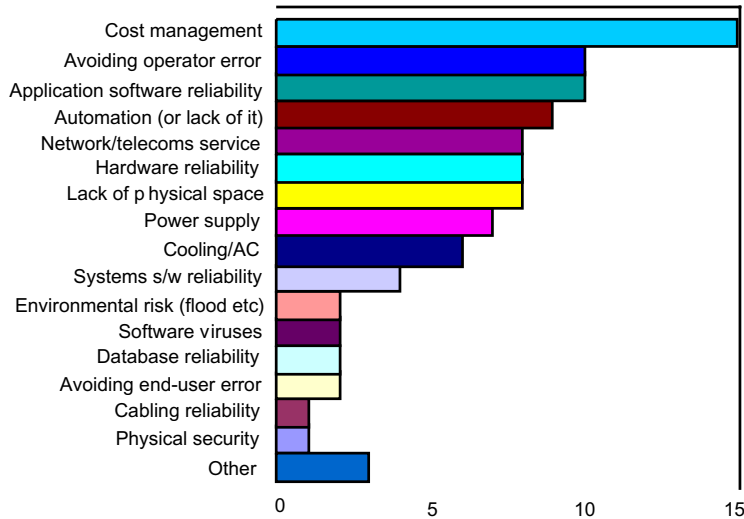


Chart 11: The greatest challenges facing the data centre

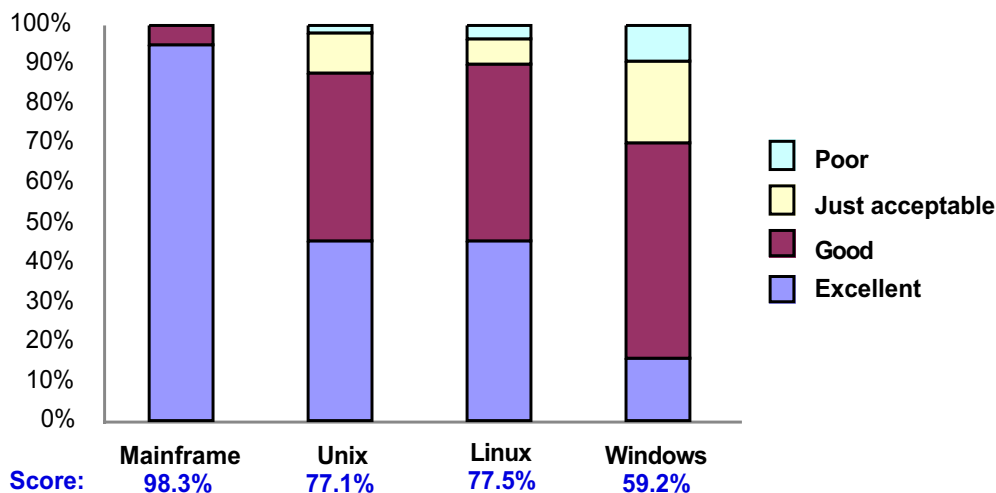


Chart 12: Overall reliability of data centre servers

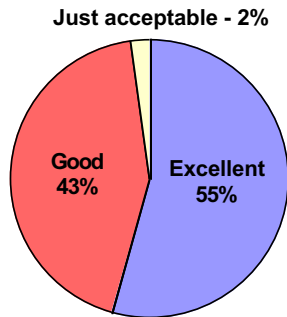


Chart 13: How would you rate the service you receive from power providers?

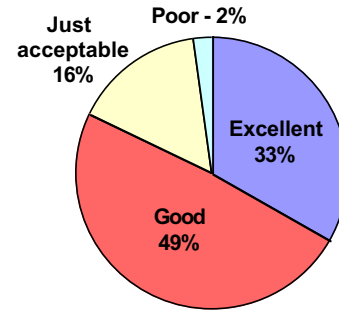


Chart 14: How would you rate the service you receive from telecom service providers?

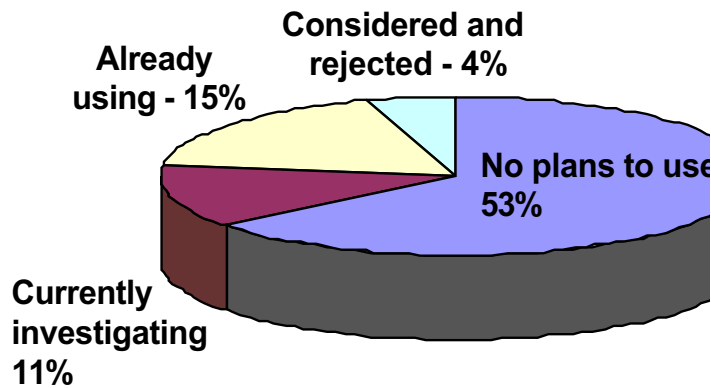


Chart 15: Are you considering grid computing as a way of reducing cost and complexity in your data centre?

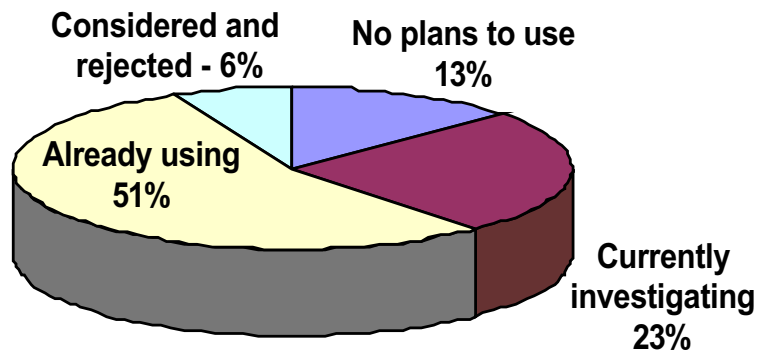


Chart 16: Are you considering blade technology as a way of reducing cost and complexity in your data centre?

in recent years. Over half of our respondents said that they already had blades in use, with a further 23% investigating the idea.

Other future trends and technologies listed by respondents are shown in Chart 17.

Our final questions concerned standards and best practices in the data centre (Chart 18). A massive 61% of respondents said that they adhere to best practices defined within the IT Infrastructure Library (ITIL). Relatively unknown five years ago outside Europe, ITIL is now setting the scene for data centre operational standards worldwide. ITIL now embraces most aspects of operations procedure, but respondents felt that problem/change management and also data centre automation needed more formal standardization.

**Bottom line**

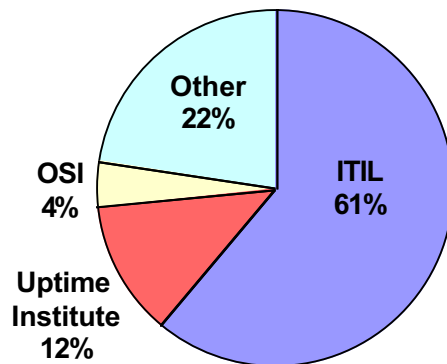
Overall, our research suggests that today's enterprise data centres display a well managed infrastructure with strong procedures in place to cope with unexpected failures.

However, there were some surprises in the results, such as the high level of unplanned failures among Tier 4 sites and the number of sites that fall below agreed service levels.

Clearly there is much to do in improving network service performance and in eliminating application, hardware and operator errors, and it's to be hoped that maturing operational standards across the board will go a long way towards addressing these issues over time.

- \* Water cooled racks
- \* Logical server partitioning
- \* Consolidation to an enterprise model (we have 15 state wide universities in our system)
- \* De-centralisation, followed by centralisation, followed by de-centralisation etc.
- \* Just expanding on the abilities of Gigabit Ethernet and SAN solutions
- \* Kill the mainframe mentality
- \* Linux on System z, SOA, Unix systems fail-over
- \* Linux, VoIP
- \* Moving web applications to the mainframe.
- \* Server virtualisation
- \* SOA and the need to avoid lowest common denominator of different 'services' woven together
- \* Terminal servers and remote access
- \* Virtualisation
- \* VTL and D2D backups, tiered storage
- \* XML Integration appliance

**Chart 17: Other trends and technologies shaping the future of the data centre**



**Chart 18: What standards/best practices do you follow in the data centre?**