IBM Business Continuity and Recovery Services

Resilient Data Centre Selection and Design

Contents

Introduction

With over 100,000 square metres of its own business-critical data centre space in the UK alone, IBM understands the importance of an integrated, end-to-end approach to complex data centre design. As the worldwide leader in the design and integration of computing technology, IBM is always a step ahead in understanding the environmental infrastructure requirements that are imperative to providing resilience and high availability in data centre environments.

IBM frequently advises clients on the selection of appropriate data centre locations to support Business Resilience and data centre consolidation. IBM's Site Enablement Services assists clients with data centre designs, new builds, upgrades or post-disaster remediation and has been providing this service to external clients for over 12 years. IBM's Business Continuity and Recovery Services assists clients in the design of resilient data centre IT solutions underpinned by practical and pragmatic Disaster Recovery (DR) arrangements.

The exposure of a data centre and the IT assets within it to the threat of interruption or unplanned downtime from a wide range of sources can be greatly reduced given careful choice of location and design. Over the course of numerous engagements a number of factors have been identified by IBM as good practice in the selection and design of data centres. The principle design considerations are documented here. This is not an exhaustive list and represents good practice at the time of publication – good practice is constantly evolving. It does not include resilient running considerations, which is another topic in its own right.

Data centre selection checklist

The following factors should be taken into consideration when selecting a backup data centre:

Primary selection criteria – location

- Distance from primary data centre location.
  - Within systems replication distance constraints (synchronous replication is generally technology-constrained with the distance depending upon the platform but generally sub-100km cable distance. Physical distance and cable distance can differ greatly!)
  - Sufficiently distanced from primary data centre to avoid concurrent loss (particular care is needed in the case of flood plains and earthquake fault lines).
  - Within reasonable travelling time for data centre staff.

- Natural environment risk exposure.
  - Flooding, landslip, subsidence, heave, effects of adverse and winter weather conditions and so on.

- Neighbourhood risk exposure.
  - Proximity to polluting sites, Control of Major Accident Hazards (COMAH) controlled sites (COMAH Regulations 1999 applies to sites such as bulk chemical manufacturing facilities, nuclear power stations, refineries and so on), and filling stations. Prevailing winds in the UK are generally from the South West. Locating upwind of potential airborne hazards such as nuclear power stations, chemical plants and so on is ideal.
  - Previous land use, such as underground mining and landfill.
  - Local crime rates and related insurance premiums.
  - Neighbours as sources of disruption, for example, targets for terrorists and demonstrators/protestors.
  - Proximity to the emergency services for response times.
  - Avoid a co-tenanted building and, if possible, a building with immediately adjoining premises.

- Local transport infrastructure and utility pinch points.
  - Is the local telephone exchange capable of delivering the required communications services? Ideally, is there a second exchange?
  - Is there only one route in and out of the site that could be blocked?
  - Is there more than one route between the primary data centre and this site that would be passable in adverse weather conditions?
  - Is there a history of disruptions to utility services, including power and water? Water disruption might include wastewater problems such as backflow due to flooding.

- Future location risk.
  - Could neighbourhood risk arise in the future, for example, local site for a new filling station?
  - Previous land use and potential clean-up costs?
  - Does the building fabric contain asbestos?
  - If selecting a company-owned site, what is the long-term future for the site? Will it be further developed/expanded or closed/sold? What is the lease period, break points and notice period?

Note: A number of resources are available to assist with the assessment of location risk. The following are suggested:

i The environment protection agency 'what's in your backyard' free Web site (www.environment-agency.gov.uk) for flood risk and pollution controlled sites.

ii The UpMyStreet free Web site (www.upmystreet.co.uk) for local crime rates and population demographics (Acorn Profile).

iii For more detailed risk mapping, a paid service offering much greater detail is required. An example is the Landmark Information Group (www.landmark-information.co.uk), which utilises 32 data sets including the Fuel Database (filling stations), COMAH, coal mining areas, shallow mining, natural subsidence hazard and registered landfill sites. Another alternative is Risk Management Solutions (www.rms.com) and its UK peril rating profiles examining windstorm, flood, fire, theft, freeze and subsidence risks by postcode for primary insurers. The client should explore if it has access to such information via its insurers or its Estates Department.

Secondary selection criteria – the building

- Building structure supports physical security requirements of good perimeter security and secondary security to sensitive areas including the proposed computer room, utility entry points, plant room and so on. Ideally, a free-standing building with a boundary perimeter and vehicle access control.
- Building structure offers a suitably sized room to house the computer room that is not in the basement and ideally with no walls shared with the building exterior. Furthermore, the room should be distanced in both the horizontal and vertical plains from internal water pipes and related services, such as kitchen and toilet areas.
- Building floors will accept the equipment weight loads (including the weight of the raised floor itself!). Ceiling height will allow for raised floors, suspended ceilings or similar.
- If not already in place then there must be the capacity to site a standby generator and fuel tank on-site.
- If not already in place, the facility must support the addition of high voltage power supplies, transformers, switchgear and UPSs.
- Dual utility entry points with diverse physical routing to separate substations for electricity is desirable (communications is discussed in section 2).
- Additional rooms to provide for UPSs, fire suppressant cylinders, goods inwards and equipment configuration prior to installation, Operations Bridge, staff office and meeting room areas.
- Secure reception and site perimeter access infrastructure with visitors' parking at least 25m away from the building and ideally no car parking underneath the building.
- Appropriate width and load capacity access route from exterior to equipment rooms.
- Large data centres should be segmented into smaller equipment rooms with firebreak separation between and comms/power distribution, air-conditioning, fire suppression that is resilient to the loss of another equipment room. Wherever possible, all necessary water-bearing pipe-work should be outside equipment rooms.

Note: Refer to the data centre design and construction recommendations on page 6 for further selection criteria. In case of evaluating a third party DR provider some additional considerations must be included. An advice paper entitled 'What to look for in a DR Supplier' is separately available.

Data centre design and construction

The following guidelines indicate current good-practice advice in the design of a resilient and protected data centre environment:

Communications requirements

- A single data centre should be supported by more than one public exchange, ideally via different carriers. A cloud Wide Area Network (WAN) architecture with two data centres on different exchanges may render this unnecessary, although still desirable in case of future data centre consolidation.
- High bandwidth communications links with the public exchange are required.
- Dual physical and logical routes to the exchange(s) with adequate physical separation.
- Separate comms termination rooms to allow providers access without necessitating entry in to the machine rooms.
- Trend to increasing number of servers per cabinet increasing communications cabling requirements to each rack position.
- Servers with multiple network interface cards for dual (high speed) Local Area Network (LAN) connection resilience and the potential for trunking technology, for example, Cisco's EtherChannel, to increase bandwidth speeds.
- Dual LAN and, if applicable, Storage Area Network (SAN) switches, patching and cabling. Each switch with dual power supplies. Switches and routers that support hot swap of redundant power supplies and network interface cards are desirable.
- If applicable, the firewall should support automatic fail over to a standby or load-balancing firewall.
- Comprehensive network monitoring and alert facilities, if applicable including external monitoring, for example, of Web site visibility to clients from Internet and network delays imposed by the carrier.

Electrical requirements

- Trend to higher electrical loads; up from 1.0 to 1.5 – 2.0 KVA per rack. Some three-phase loads, for example, IBM storage. Some equipment suppliers require higher ratings, for example, two 32 amp supplies (Compaq). Design should include suitable capacity and/or expansion capability to meet future equipment and cooling power requirements.
- Dual high voltage feeds, transformers and low voltage switchgear.
- Standby power generator with adequate on-site fuel and generator capacity sized to meet future growth requirements.

Note: Trend to N+1 or 2N diesel generation unnecessary if replicating between data centres.

- Redundant UPS protection and mains maintenance bypass circuit. Trend to N+1 UPS to allow UPS withdrawal for maintenance without reliance upon mains.
- Redundant power distribution units (PDUs) and power cabling, that is power feeds from two UPSs to each cabinet. PDU types should be used that allow live working to avoid loss of resilience, for example whilst adding additional supplies.
- Redundant equipment power packs and connection to separate UPSs via separate PDUs.
- Emergency power-off switch positioned near the entrance(s) and protected from accidental actuation.
- Electromagnetic field shielding may also be necessary, for example if high voltage supplies and transformer equipment are close to the data centre.

It is important to get data centre power requirements right without over-sizing. This is because as more power is needed, more cooling is necessary, which in turn requires more power, all of which increases costs in a non-linear manner. Similarly, it is expensive to provision a large excess of data centre space in advance as there are significant fixed costs and potential running costs to maintain a controlled environment (albeit a nearly empty one).

Cooling requirements

- Redundant air-conditioning with adequate capacity, or data centre space for additional units, to meet future growth in cooling requirements. As power requirements increase, so do cooling requirements.
- Careful attention paid to airflow to eliminate hot spots. As the room is populated the air flow and hot spots may change and this should be taken into account. Such is the growth in power requirements that new cooling strategies may be called for to more closely manage and direct air circulation around, through and away from heat-generating equipment.
- As the trend for racks to be densely packed with equipment such as thin servers that take up the entire shelf footprint, air-conditioning must blow through racks from front to back rather than bottom to top. As a consequence, such racks will require metal mesh doors front and rear to support cooling requirements. Rack and air-conditioning placement therefore to be in context of front-to-back blowing through racks. Some humidification for staff comfort and anti static may be required. Close temperature control is not necessary (+/- 10°C is acceptable).
- A positive/over pressure to minimise dust ingress.

Automatic fire suppression

- Certain gas and powder-based fire suppression systems leave a residue on equipment, which then requires specialist cleaning before it can be used again without risk of subsequent failure. Furthermore, the room must be sealed to prevent gas escape, it must maintain its integrity under the increased pressure during a discharge (possibly through an over-pressure vent) and there must be a means of evacuating the gas afterwards. Some discharges are quite aggressive and have been known to dislodge floor tiles and turn any loose objects into dangerous projectiles. As the computer room is filled with equipment the overpressure problem is exacerbated. Most gas-based systems are breathable based upon the air to gas ratio at the time of design and hazard to health is then more likely from the effects of aggressive discharge. One advantage of a gas-based system is that discharge heads can be positioned to drench the floor void. A standard sprinkler system is not advisable, however a distilled water mist system, provided power is automatically isolated and the water used is pure, is surprisingly effective and equipment is often unaffected once thoroughly dried. It offers the advantages of localised activation and gentle discharge. Water-based systems should not hold water in the pipe work unless activated and activation should be on a per-head basis rather than across an entire zone.
- For gas-based fire suppression systems, the air to gas ratio established at the time of design allows breath-ability whilst inhibiting combustion. As a room fills with equipment and the level of gas remains the same, the ratio of air diminishes. It is prudent to advise staff to evacuate as quickly as possible closing the door afterwards and then to minimise time working in the room until there is no risk of a fire restarting and the gas has been evacuated. Cylinders should be recharged as quickly as possible. Note that FM200 is increasingly difficult to permission in some countries.

- Systems are available to reduce the extraordinarily damaging effects of smoke, for example by circulating the air through a pipe containing a water mist head to scrub out the smoke particles and a means of recalculating the small volumes of water used within a closed system.
- A further fire prevention option is a system that reduces the oxygen content of the air within the data centre from 21 percent to 15 percent, at which there is sufficient air to breathe (equivalent to working 3,000m above sea level) but insufficient oxygen to support combustion. The Oxyreduct system by Wagner, for example, uses nitrogen generated from atmospheric air to maintain an environment in which a fire will not start without the need for gas bottles.
- It is prudent to divide a data centre into separate machine rooms with fire breaks between and running off separate infrastructure to facilitate containment of disaster effects.
- Consideration should be given to a fire suppression system that can be triggered more than once without the need for re-priming.

Environmental monitoring

- To include over temperature, under temperature, humidity, moisture detection, fire detection, mains power failure, air conditioning unit failure, UPS failure/low battery, intruder alarm activation, generator running, engine faults and low fuel. Intrusion detection and systems monitoring is also required for networks, servers, storage and perhaps also for critical applications, databases, interfaces and batch processes, although this is not discussed here in detail. Environmental monitoring alarms should be monitored around-the-clock and provide granular information, for example, which moisture detection sensor has been triggered and when. Sensor numbers and positions should be carefully considered and include sensors in floor and ceiling voids, for example, VESDA/ smoke detectors should not be positioned in front of air-conditioning unit blowers. Consideration might be given to an under-floor low point sump and drainage pump to assist in removal of any water that penetrates the data centre.

Physical security requirements

- Walls, ceiling and door built to provide protection from intrusion as well as a fire outside the computer room. These are typically a lacquered double steel skin with a special-purpose filler in between.
- Access control system requiring both a swipe in and out (with emergency break glass override) to provide an audit trial.
- Automatic door closure to avoid compromise of fire suppression and security.
- Security of peripheral infrastructure – consider the generator, fuel tank, UPS, fire suppressant cylinders, utility entry points, chillers and air intakes.
- Security around delivery areas and fire escape routes.
- Double layer security system, for example, intruder alarms on doors and PIR detectors covering the door from the inside. If the data centre is to be unmanned then consider additional security to deter and delay intruders, once detected, until a response can be effected. Ideally there should be a security presence on-site at all times.
- Consider monitored CCTV externally and within the computer room – particularly if lone working in the computer room is commonplace. Motion-activated CCTV can be helpful in ensuring the CCTV operator's attention. CCTV can now be transmitted in IP over a WAN and monitored remotely either in-house or by an external service provider.
- The computer room will be more secure if it is located away from the building perimeter. The effect sought is a secure enclosure within the confines of the building's own secure perimeter. If there is no reason for unauthorised staff to be in the vicinity of the computer room then intruders are easier for staff to detect and challenge.
- Windows are unnecessary and present a source of structural weakness, temperature fluctuations and can disclose the room's contents to prying eyes.

- Security controls should be appropriate to the perceived threat and vulnerabilities, for example, biometric access controls and external ram raider defensive structures are probably unnecessary in the case of many organisations. Some organisations may require greater physical security – international settlement agency data centres might literally be bunkers on military-style compounds with regular physical penetration testing by special forces personnel.
- There should be a physical perimeter and sterile zone between building and the exterior with barrier to free entry. Ram-raid defensive structures around the building exterior may be appropriate.
- Some form of around-the-clock on-site security presence is recommended, with regular patrols in addition to (potentially off-site) CCTV monitoring.

Other factors

- Doors and access routes within the building of sufficient size and capacity, for example, in the case of a lift, to support installation of equipment.
- Raised floor with removable anti-static floor tiles to support delivery of power and data cables to equipment racks. Consider under-floor cabling in raised mesh trays to improve protection from water ingress. Note that the potential floor loads have increased in recent years. Disc is heavy and it is increasingly common to find racks entirely and densely populated with disc storage arrays.
- Separate area for accepting equipment deliveries, unpacking and assembly prior to installation in the data centre. This dramatically reduces dust and clutter within the computer room.
- Post and packages should ideally be opened in another building before delivery to the data centre.
- Separate environmentally controlled and secure room for UPS and possibly for fire suppressant cylinders.
- Tack mats at entry points and slight positive pressure to minimise dust and dirt ingress.

- Easy-clean and non-static surfaces that do not release dust or moisure (like a plaster finish on walls will do).
- Fire alarm linked to automatically summon the fire brigade in the event of an alarm.
- Adequate emergency lighting to support rapid and safe egress; dense packing of racks will necessitate additional emergency lighting, possibly at floor level.
- Sealed cable/service entry ducts, sealed with intumescent material that can be easily reopened and resealed for future changes in cabling requirements (illustration 2).
- Around-the-clock Mechanical and Electrical support on-site is desirable.
- Around-the-clock manned on-site Operations Bridge. Consider a fallback Operations Bridge in case of loss of primary.

The data centre should not be positioned below ground level in a building where it is at risk from flooding, storm water ingress, drains backing up or similar problems and any gas or smoke could prove difficult to vent. Furthermore, the data centre should be positioned away from the building core, which is typically used to run services vertically throughout the building as these services inevitably include water. The problem is frequently exacerbated through the positioning of toilets and kitchen areas close to the building core for this reason. Internal plumbing faults are a major cause of disasters in data centres. For this reason and wherever possible, water bearing pipes should be removed from the data centre. Where this is not possible, moisture detection sensors should be positioned, for example at the low point in angled drip trays beneath the pipe work to offer an early alarm and direct water away from equipment.

Build strategy

Suppliers of data centres now design and build what amounts to a building within a building using a modular construction technique (illustration 3). This allows the data centre to be expanded easily in the future should additional space be required simply by repositioning walls. It may also permit the dismantling and rebuild of the data centre in the event of a data centre move, protecting the organisation's investment to some degree. A number of suppliers will manage the entire data centre package including fire suppression, air-conditioning and power distribution. IBM's Site Enablement Services can assist with design and project management if either a traditional or modular data centre construction approach is required.

The modular data centre design approach has a number of benefits:

- Data centre space can be added as it is required, which contains costs.
- Rapid construction once all of the components and permissions are in place.
- Capability to build inside 'difficult' existing buildings, for example, listed buildings.

The effect of a modular approach on data centre costs is most marked when the data centre concerned is expected to grow significantly over time. A traditional approach to data centre construction would necessitate the provision of environment and infrastructure for data centre space from the start. The build costs are incurred at the start and running costs commence for the cooling, power and so on immediately. Where data centre requirements are growing significantly over time a modular approach can prove more cost-effective than an apparently cheaper traditional build.

European standard EN1047

A British and European standard, EN1047, was recently introduced to define a standard for computer room physical security and environment protection. This dictates that the room environment under external fire testing should not exceed a maximum permissible 85 percent humidity or 500°C after a minimum of 60 minutes of heating to avoid damage to magnetic media. Some manufacturers offer tighter limits. The standard also defines guidelines regarding physical security. A copy of the standard can be obtained from the BSI (www.BSI.org.uk).

A fire immediately outside a conventional build data centre will have the effect of raising the temperature and humidity within the data centre as moisture is released from the brick/plaster walls and heat is transferred, which could result in equipment failures/damage even if the fire does not penetrate the room. A computer room complying with EN1047 will resist these damaging effects for much longer.

Salutary tales and reminders

One FTSE 100 company decided to build a new, state-of-the-art data centre and went to great lengths to ensure that it was located in a relatively risk-free position and built with redundancy and resilience in all aspects.

Unfortunately once completed, the large, low, flat building out in the middle of nowhere was adopted as a way-point for the US Air Force Fairchild A10 Thunderbolt (Warthog) tank-buster squadron based nearby, which took to making high speed, low altitude turns over the building on a regular basis presenting a previously unforeseen risk of interruption!

A UK public sector organisation that sited its backup data centre some distance from its primary data centre but on the banks of the same river, which flooded demonstrating both sites were concurrently at risk.

Location risk profile does not necessarily vary in a linear fashion with distance. Note, this example is concerned with river flood plains, but could also apply to coastal flood plains, prevailing wind direction, earthquake fault lines, and so on.

A FTSE Telco with its main data centre at its head office location suffered a failure of its single data centre UPS, which also disabled the bypass on fail circuit. Power was consequently lost to the data centre whilst power to the headquarters office areas was unaffected. Power was narrowly restored before regional billing information buffers overflowed, which would have resulted in lost billing information.

It is surprising how many data centre designers apparently do not understand how to include resilient power provision in their designs. The Telco example above is a good one; the designer clearly thought that adding a UPS was creating resilience whereas in fact it was simply changing the single point of failure. Furthermore, resilience is often a tactical point solution rather than an end-to-end provision. For example, resilience from the supply side to the UPS but not from the UPS to the servers through the internal power distribution equipment and cabling.

An online share brokerage's main data centre was nearly flooded following poor plumbing work installing a drinks machine for a building co-tenant two floors above. Statistically, plumbers and plumbing problems present a major hazard to continuity of operations.

A law firm discovered its new offices and data centre at the same location were subject to interruption as all power was distributed from the low point in the building's basement, which was also prone to flooding during heavy rainfall.

In one of the UK's worst civil engineering disasters of recent times, a tunnel for the Heathrow Express Rail Link collapsed during construction.

"On Friday 22 October 1994 at around 01:00am, an initial collapse of a temporary concrete tunnel lining occurred some 20 metres below ground close to the main working shaft. The caving in then spread to two other tunnels producing a loss of ground and a huge hole to appear on the surface. Nearby buildings, temporary site cabins and offices were forced to lean and crack under the moving ground. With a possible collapse imminent, tunnelers working below ground were evacuated only 20 minutes before the series of events began. Fortunately nobody was injured."

(Source: Bristol University Civil Engineering Deptartment)

Excerpts from the BBC's coverage following the court case in 1999 follow:

"…the court was told that the collapse could have crushed to death people using the nearby Piccadilly tube line and the judge hearing the case, Justice Peter Cresswell, said it was 'luck more than judgment' that this did not happen.

"The engineering disaster caused a huge crater to appear between the airport's two main runways and caused damage to car parks and buildings. It took months to clear up the damage.

"Justice Cresswell said: 'This was one of the worst civil engineering disasters in the United Kingdom in the last quarter of a century. The tunnels were being built below part of the world's busiest international airport and there was considerable potential for harm.'

"The tunnel's collapse delayed the opening of the Heathrow Express by six months and caused long hold ups on part of the Jubilee line extension, where engineers were using similar building techniques."

(Source: BBC)

Not a disaster affecting a data centre, but tunnelling and tunnel collapse represents another often overlooked potential threat to data centres in some locations.

About the author

Robin Gaddum is a senior consultant with IBM Global Services in the UK and responsible for leading the UK consulting team for IBM's Business Continuity and Recovery Services. He can be reached at: *gaddumr@uk.ibm.com*.

Relevant IBM services

IBM Business Continuity and Recovery Services is a leading provider of business resilience, continuity and disaster recovery solutions. IBM is able to draw upon more than 40 years experience in assisting clients to develop and implement their business continuity strategies and plans. As part of this service, IBM has completed thousands of engagements, large and small, on behalf of over 5,000 clients across a range of industries around the world.

IBM's Crisis Response Team has responded to countless emergencies, including earthquakes in California, Colombia, Turkey, Greece, Taiwan, El Salvador, India and Peru and has managed, supported and restored operations following floods, hurricanes, hailstorms, tornadoes, ice storms, volcanoes, civil unrest, hazardous material situations, the Oklahoma City bombing and the World Trade Center disaster.

IBM's Site Enablement Services undertakes premises designs, new builds, upgrades or post-disaster remediation with particular expertise in data centre environments and supporting infrastructure.

Besides its expertise in the complementary areas described above, IBM has skills in security, high availability solutions, systems and data management, network design and implementation and desktop infrastructure as well as platform and application knowledge. Following its acquisition of the PricewaterhouseCoopers Consultancy, IBM also has industry-leading general business consulting skills. This ensures that IBM has a solution for any unforeseen issue likely to be encountered by its clients.

Solutions are tailored to a multi-vendor environment, geographical location(s) and availability/risk mitigation requirements.

For more information

To learn more about IBM Business Continuity and Recovery Services in the UK, visit:
www-5.ibm.com/services/uk/portfolios/bcrs.html

Or contact us at:
E-mail: BCRSHELP@uk.ibm.com
Tel: 01926 464103

IBM United Kingdom Limited
emea marketing and publishing services (emaps)
Normandy House
PO Box 32
Bunnian Place
Basingstoke
RG21 7EJ
United Kingdom

The IBM home page can be found at **ibm.com**

References in this publication to IBM products, programs or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program or service is not intended to imply that only IBM products, programs or services may be used. Any functionally equivalent product, program or service may be used instead.

IBM hardware products are manufactured from new parts, or new and used parts. In some cases, the hardware product may not be new and may have been previously installed. Regardless, IBM warranty terms apply.

This publication is for general guidance only. Information is subject to change without notice. Please contact your local IBM sales office or reseller for latest information on IBM products and services.

Photographs may show design models.

ITWEE01018-4 (05/04) JT