Ark Continuity - Continuously Delivering high integrity data centres

Maximising the Grid Capacity of a Data Centre

The historical perspective and the latest trends in how to maximise the ROI in the grid connection

Dr Ian F Bitterlin

PhD BSc(Hons) BA DipDesInn MCIBSE MIET MBCS MBIFM MIEEE

Publication: April 2011 | **Reference** 2011/04/08

WHITE PAPER 2011

Ark Continuity delivers high integrity data centres for Government and Corporate occupiers. Ark is now building data centre campus locations at Spring Park, Wiltshire and Cody Park, Hampshire. Data Centre SQ17, module one, is now operational. All facilities are designed to be the most secure, available and sustainable in Europe.

This document is one in a series of white papers prepared by Ark Continuity. If you would like further information, please contact Ark Continuity on 0845 389 3355 or visit www.arkcontinuity.co.uk

Head office: Hartham Park, Corsham, Wiltshire, SN13 0RP

Introduction

The provision of a power grid connection for a modern large data centre is problematical and expensive. In the UK the load expansion pressure on the grid (albeit slightly eased by the 2009-10 recession) makes 'power supply' a key item on the site selection agenda. Having contracted the supply, there follow-on punitive charges for retaining the power capacity but not having the load demand, although it is the lost opportunity for profit if the available power is not used for computing that remains a problem. This paper reviews the historical perspective and the latest trends in how to maximise the return on investment in the grid connection, but what is not addressed is the simple problem of low load in a completed data centre and the slightly more complex issue of partial load efficiency in mechanical cooling systems.

Historical Landscape

Although we never used the term PUE (Power Usage Effectiveness - recently innovated by The Green Grid) data centre power system designers have, for more than four decades, worked 'backwards' from the connected IT load to establish the mains capacity required.

A very simple example is shown in Fig 1, (right). This would have applied to a typical static-UPS fed non-free-cooling chilled water CRAC facility. In today's terms this does not represent a PUE of 2.47 for several reasons[1] but the concept of PUE can be adapted to this type of 'planning' purpose. Given this calculation it could be envisaged that a grid connection of 2.5MVA or, with a safety margin, 3MVA would be sourced. In this way it can be seen that a given IT load would often require a grid connection power capacity of 2.5-3 times its size.

Legacy Data Centre c1990	kVA
IT Load	**1,000**
UPS losses	120
Mechanical load	850
Battery Charging	250
Distribution losses	50
Lighting, small power	150
NOC	50
Peak grid capacity	**2,470**

Fig 1

Legacy Data Centre c1990	kVA
IT Load	**600**
UPS losses	90
Mechanical load	663
Battery Charging	-
Distribution losses	28
Lighting, small power	150
NOC	39
Grid load	**1,570**
PUE	**2.62**

Fig 2

For completeness we can consider the same facility in 'normal' operation at a typical 60% load and predict a PUE of around 2.6, see Fig 2, (left). Given that free-cooling was rarely used c1990 this PUE would have hardly varied by the season.

If the grid power system was installed at 3MVA capacity then the 60% IT load would correspond to a grid load of c50%. Although not common at that time if the power system was 2(N)[2] or 2(N+1)[3] then each grid supply would only be running at c25% capacity, albeit the overall power demand would be only slightly higher.

Even with all of the best air-management (cold-aisle containment, blanking plates, floor-hole stopping etc) these legacy facilities are limited to an annualised PUE of around 2.0.

---

[1] It is kVA not kW (or kWh), it is a 'snapshot' not an annualised average and it represents a planned 'maximum' consumption not actual consumption divided by actual load with the allowance for battery charging which is a very infrequent and short-term event
[2] What we would now refer to as Tier IV, according to The Uptime Institute latest definition
[3] Tier IV according to the latest version of TIA942 and the 'original' TUI definition

Modern Trends

Since the 'dot.bomb' disaster of the early 2000's, data centre developers have been careful to invest in modular and scalable designs rather than monolithic deployments. In this way, smaller development phases ensure that spaces are occupied faster and, within those smaller spaces, a degree of flexibility of power-density is possible to ensure high load versus capacity. At the same time the mainstream facilities have aspired to Tier III rather than the more capital and redundancy intensive (and marginally less energy efficient) Tier IV topology, again maximising the system load percentage.

However it is the gradual acceptance and introduction of lower energy mechanical cooling systems that has started to make the largest impact on the utilisation of the grid power capacity:

Alternative cooling strategies

Enabled by the 2008-9 relaxation of temperature and humidity set-points[4] by ASHREA's Technical Committee 9.9 and the growing realisation that the largest single consumer of energy in legacy data centres was the mechanical cooling system, two trends are strongly growing:

- **Free cooling[5]**
  - Where the outside air temperature (also with adiabatic water application) is used for cooling the heated transfer fluid for recirculation, thus minimising the use of a refrigeration cycle and associated electric motors driven compressors.
  - Not using compressors (just fans and pumps) saves at least 30% of the usual cooling power demand.
  - In northern latitudes (for example Southern England) the compressor motors may be required for only 30-40% of the year and annualised PUE's of c1.5 are possible.

- **Fresh Air cooling[6]**
  - Outside air is brought into the critical space after being filtered, often blended with heated exhaust air and (where required) humidified/de-humidified and used for directly cooling the critical load.
  - The only power used in normal operation is forced ventilation fans (optimally variable speed).
  - In northern latitudes (for example Southern England) the fresh air cooling may be sufficient 98-99% of the year and annualised PUE's of <1.2 are possible.
  - For the remaining 1-2% (100-150h) of the year a standby mechanical cooling system is required to cover the 'hottest-day-of-the-year' events.
  - It is the provision of the emergency cooling system in fresh air cooling solutions that we will now explore the impact of.

[4] In very broad terms from 21°C±1°C and 50%±10%RH to 18-27°C and 20-80%RH
[5] Described as 'water-side economization' in US terminology
[6] Described as 'air-side economisation' in US terminology

Emergency Cooling Systems in Fresh-air Solutions

Even in the UK the external ambient occasionally strays above the maximum allowable temperature of ASHREA, 27°C, and the 'fan-only' fresh-air system will have to be supplemented or substituted. Air contamination (smoke, fumes/odours, particulates, pollen etc) is another reason why a fully rated substitution system would be required. When these events occur (on average for less than 100h/year) the data centre infrastructure has to respond by shutting off the external vents and switching over to an alternative mechanical cooling system.

To minimise the capital expenditure, and to cope with elevated external temperatures, the emergency system is rarely engineered for maximum power efficiency – indeed that could often equate to a waste of resources given the infrequent duty-cycle of 1-2%. The most common solution is an array of cooling coils with a compressor driven Direct Expansion gas circuit feeding an externally mounted series of fan-cooled condensers.

When this type of system is running at full load in an ambient of >30°C the anticipated spot PUE will be c1.8 – and here is the problem for the grid capacity.

Divergence of Spot PUE between 'Normal' and 'Emergency' Operation

Taking a well-engineered fresh-air system running for most of the year at PUE<1.2 (and sometimes reaching 1.12) with an emergency alternative system of DX offering a spot PUE of c1.8 produces a problem when designing the grid connection:

For example take a 1MW IT load - for >98% of the year the grid demand will be 1.2MW but, on the hottest days of the year and <2% of the time the grid load will be 1.8MW. Clearly a difficult situation for planning and utilisation of (a valuable resource) the grid supply.

Solution for Divergent Spot PUE Applications

There is a practical solution available to all 'data centre' designs that minimises capital expenditure and maximises the utilisation of the available grid supply.

- Size the capacity of the grid supply based upon the design PUE, e.g. 1.2 (plus any growth or reserve as would apply to a conventional design)
- Size the capacity of the emergency diesel generation on the peak PUE, e.g. 1.8
- Automatically start the generators when the external vents close AND the load is higher than the grid capacity

Given that few data centres ever run at full load the chances of the generator sets being required in this mode 'in anger' are quite low. The only downside to this approach is the blurring of any classical Tier classification due to the theoretical 'under-sizing' of the grid connection.

End

To read more white papers from Dr Ian Bitterlin, please visit www.arkcontinuity.co.uk