

Sustainable and Resilient Data Center Design for Saudi Arabia's Emerging Digital Economy under Vision 2030.

Abstract:

Saudi Arabia's digital transformation is driving up the demand for cloud, AI, sovereign data and high availability computing infrastructure. The country's national energy transition is also leading to greater scrutiny of the environmental and resilience performance of new data centres. This paper reviews how the principles of sustainable and resilient data centre design can be applied to the Saudi context, specifically in relation to its hot arid climate and its Vision 2030 development path. A structured integrative review of 30 peer-reviewed studies published between 2020 and 2025 was conducted, covering the full range of relevant topics including energy efficiency, cooling, water use, renewable electricity, storage, waste heat recovery, thermal modelling, digital twins and infrastructure resilience. The findings demonstrate that there is no one-size-fits-all solution in terms of energy efficiency or cooling technology for the Saudi environment. A focus on low power usage effectiveness can in fact lead to increased water demand, while a very aggressive approach to water conservation can lead to increased electrical cooling loads. High density computing also undermines the long term viability of conventional air cooling and increases the value of direct liquid cooling, advanced heat rejection, workload aware controls and life cycle assessment. The paper therefore proposes a climate responsive design framework which integrates site selection, envelope and airflow design, liquid ready thermal architecture, low water heat rejection, renewable energy integration, storage, redundant power systems, digital controls and multi metric performance governance. Resilience is treated as a design attribute rather than a feature to be added on as a backup system, requiring modularity, fault isolation, thermal ride through, diversified energy supply and adaptation to extreme heat. The framework is positioned to support the growth of digital infrastructure in line with Vision 2030 by treating energy productivity, water security, carbon reduction and service continuity as co-equal design objectives. The paper concludes that Saudi data centres should be assessed on a whole system, climate specific and life cycle basis rather than simply against imported efficiency benchmarks.

1. Introduction

Data centres are clearly becoming strategic physical infrastructure for digital economies as cloud platforms, AI, financial services, public sector systems and data intensive industries all rely on a constant supply of computing capacity. The environmental footprint of data centres is not just in their electricity demand though. Recent work has demonstrated that the growth of data centres creates associated carbon, water and land pressures meaning that sustainability cannot be measured by a single facility efficiency metric [1-5]. This systems approach is particularly

relevant in the case of Saudi Arabia where the country is undergoing a period of rapid digitalisation as part of its broader programme of economic diversification, energy transition and infrastructure development. Vision 2030 has significantly increased the strategic importance of local digital capacity and at the same time has raised the expectation that any new infrastructure will be designed to deliver on the wider goals of resource efficiency, low carbon growth and resilience [27-30].

The Saudi scenario is a very particular design challenge. The ambient temperatures are very high so there's a significant requirement for mechanical cooling throughout the year. The aridity of the environment also makes water use for heat rejection a very significant environmental issue. There is evidence from operation of data centres in hot arid climates that you can have a very low PUE and at the same time have a very high water consumption so energy efficiency and water efficiency can be moving in opposite directions [5]. The increasing rack density is also pushing against the limits of what conventional room air cooling can achieve. Work on fan wall systems, thermal environments and CFD has shown that airflow management can make a huge difference in terms of reducing hot spots and overcooling, but the same work has also demonstrated that as heat density increases the ability to manage the airflow becomes progressively more difficult [6-9]. The reviews of next gen cooling all point to direct liquid cooling becoming increasingly important for high performance and AI workloads [10-13].

Resilience is another dimension. A data centre that is deemed to be sustainable and yet fails during an extreme heat event, a grid disturbance, a piece of equipment failing or a water constraint has not met the functional requirements of digital infrastructure. Reliability has to be built in from the outset alongside efficiency. Work on renewable energy scheduling, battery integration, demand response and zero carbon data centre portfolios has shown that computing workloads can in fact become active participants in the energy system rather than just passive consumers of electricity [18-24]. The incorporation of renewables does however introduce variability which has to be managed via storage, controllable loads, redundant supply paths and intelligent orchestration. The design challenge is therefore multi-objective: you have to achieve

the lowest possible resource intensity without compromising on service continuity and you have to build resilience without effectively locking the facility into inefficient redundancy.

Another issue is that a lot of the data centre literature is based on temperate, cold or humid climates. You can't just take results on free cooling, heat recovery and district energy integration and apply them to Saudi Arabia without some form of climatic interpretation. The availability of free cooling, the evaporative potential, the level of water scarcity, the quality of the solar resource, the level of dust and the peak summer temperatures all alter the value of individual technologies. The real question therefore isn't which technology has the lowest reported energy use in isolation, it's which combination of architectural, thermal, electrical and digital strategies is going to be robust in the context of the actual operating conditions in Saudi Arabia. That's what this review sets out to do by taking the available evidence and putting it together into a design framework which is specifically structured around the requirements of sustainability, climate adaptation and operational resilience.

1.1 Aim of the Study

The objective of this review is to establish an evidence-based framework for sustainable and resilient data centre design in Saudi Arabia, bringing together climate appropriate thermal design, energy and water efficiency, use of renewable power, operational reliability and digital control within the context of Vision 2030. Rather than simply ranking technologies on the basis of one performance metric, the review looks at how the various design decisions interact with each other and over the life of the facility.

1.2 Research Objectives

The review is structured around five objectives: (1) to identify the key energy, thermal, water and resilience pressures which will shape datacentre development in hot arid environments; (2) to critically evaluate the options of air cooling, free cooling, liquid cooling and heat recovery for high density computing; (3) to assess the extent to which renewable generation, storage and demand flexibility can enable low carbon operation without impacting availability; (4) to establish a multi-metric design approach that is appropriate to the Saudi context; and (5) to derive research and policy priorities that will enable datacentre development to be aligned with Vision 2030 and long term climate resilience.

2. Review Methodology

We've used a structured integrative review as the methodology as the research question cuts across building engineering, thermal management, electrical systems, energy policy and digital operations. The intention has not been to statistically pool a single outcome, but to compare the disparate evidence and identify the design relationships that are still valid across the different technical approaches. The review is structured in the way that you'd expect in a rigorous piece of built environment/sustainability research: a clear aim, explicit inclusion/exclusion criteria, clearly defined analytical categories and a synthesis which makes the difference between what the evidence actually says and what we interpret it to mean. It also enables the same design framework to be informed by experimental studies, field data, optimisation models and critical reviews without implying that the results from each of those are in any way directly comparable.

The search was restricted to peer-reviewed literature published between January 2020 and December 2025. Searches were carried out in Scopus, Web of Science Core Collection, ScienceDirect, IEEE Xplore and the ACM Digital Library using combinations of the terms "data centre" or "data centre" with "sustainable", "resilient", "cooling", "liquid cooling", "free cooling", "water", "PUE", "WUE", "renewable energy", "battery storage", "waste heat", "digital twin", "hot arid", "Saudi Arabia" and "Vision 2030". Citation chaining was used where a more recent high relevance paper pointed to an earlier 2020–2025 study which directly addressed a

design mechanism. Only English language journal articles with a verifiable DOI were included in the final evidence base. Conference papers, vendor reports, non-peer-reviewed commentary, studies published prior to 2020 or after 2025 and sources without sufficient methodological detail were all excluded.

Relevance was judged on four criteria. Was the study looking at a physical or operational characteristic which in some way impacts on the sustainability or resilience of a datacentre? Did it present measured, modelled or critically synthesised evidence rather than just general advocacy? Could the finding be of use in the context of hot climate design even if the study in question was not specifically in that climate? Was the result specific enough to allow for a design implication rather than just a broad statement about the benefits of digitalisation? The final set of papers was 30 peer reviewed papers. The DOIs, titles, years of publication and journal details were all checked against the publisher or bibliographic database prior to inclusion.

Methodological quality was assessed on a fit-for-purpose basis rather than via a single score. Experimental and field studies were assessed on the basis of measurement transparency, validation and boundary conditions. Computational studies were assessed on the basis of model validation and sensitivity. Optimisation studies were assessed on the basis of the presence of explicit constraints and operational assumptions. Review papers were assessed on the basis of search transparency, technical coverage and critical comparison. The Saudi-specific studies were used to inform the policy and digital development context but were not used to make any inference about technical performance where they did not provide data centre engineering evidence [27–30]. The separation of the national strategy from the technical performance is thus not a substitution.

The synthesis employed five analytical domains: resource efficiency, thermal architecture, water stewardship, energy system integration and resilience/control. The findings were then assessed in terms of complementarities and conflicts. Particular focus was placed on the trade-offs between PUE and WUE, between capital redundancy and operating efficiency and between the variability

of renewables and service continuity. Figure 1 shows the review logic. The framework in Figure 2 was then developed by taking the recurring evidence and expressing it as a layered design sequence from site and climate through to thermal and power systems and finally to digital control and performance governance. There is no numerical screening flow reported beyond the final 30 paper corpus as the review was designed as a focused integrative synthesis rather than a quantitative evidence census.

Structured integrative review and synthesis logic

Evidence is filtered by relevance and methodological fit, then translated into climate-specific design decisions.



Figure 1. Structured integrative review and synthesis logic used to translate recent evidence into Saudi-specific design implications.

3. Thematic Literature Review and Critical Analysis

3.1 Energy, Carbon and Whole-System Performance

Energy efficiency is still very much the cornerstone of data centre design, but the literature is increasingly critical of the utility of facility level electricity ratios in isolation. Reviews of data centre energy metrics have shown that PUE is useful in identifying the level of facility overhead

but is not in any way a measure of carbon intensity, water use, IT productivity or resilience [3]. More general sustainability studies point to the fact that data storage and digital infrastructure have environmental impacts across the electricity supply chain, the water system and the land system [4]. For Saudi Arabia this means that a low PUE is essential but is not in itself sufficient to demonstrate a sustainable design. A data centre with a low PUE could be running on carbon intensive electricity and could be entirely dependent on scarce water and still be transferring the environmental burden to other parts of the system.

A whole systems perspective is also required as a design change in one area can have knock-on effects in others (e.g. shifting the load between IT, cooling and power conditioning). Thermal optimisation may lead to a reduction in chiller energy but an increase in fan energy. Server consolidation may improve IT utilisation but at the same time create local heat density problems. Life cycle evidence is another factor. A 2025 assessment of advanced cloud cooling found that cold plate and immersion solutions had the potential to reduce GHG emissions, energy demand and blue water consumption compared to conventional solutions (although the extent of the reduction was dependent on the interplay between the servers, the facility systems and the electricity supply) [25]. This underlines the value of a life cycle comparison at the early stages rather than making a decision based solely on operational cooling efficiency.

3.2 Thermal Design for Hot-Arid Conditions

Cooling is by far the most climate sensitive element of the Saudi data centre design. Air cooling is still viable for moderate rack densities (particularly if you take a proactive approach to airflow rather than just treating it as a uniform room condition). Fan wall studies have shown that rack orientation and supply configuration can have a significant effect on airflow and can be used to improve the distribution of airflow and reduce peak server temperatures [6]. Field and CFD work has also shown that relatively small changes to the airflow can go a long way to eliminating thermal non-uniformity and reducing the amount of overcooling [7]. CFD has therefore become an important tool in both the design and commissioning of data centres and there have been

recent advances in both the accessibility of CFD and the speed of the calculations through the use of open source workflows and super resolution techniques [8,9].

These findings do support containment, short air paths, leakage control and validated airflow modelling as the baseline measures, they don't remove the density constraint. Critical reviews of data centre cooling point to the increasing impracticality of room scale air systems as the chip and rack heat fluxes increase [10]. Direct liquid cooling reduces the number of thermal resistances between the electronic components and the heat transfer medium and is therefore becoming increasingly attractive for the very high density AI and HPC applications [11,12]. Recent reviews also highlight that liquid cooling can lead to higher coolant return temperatures which in turn improves the quality of the recoverable heat and reduces the need for large volumes of conditioned air [13]. For the Saudi facilities that are going to be hosting AI workloads "liquid ready" design should be considered at the building and white space stage even if the initial deployment is still a mix of air cooling.

The design implication is a cooling hierarchy rather than a single technology prescription. The lower and medium density areas should start with containment, fan efficiency, wider allowable temperature bands and real time control. The higher density areas can then move to direct to chip or immersion where appropriate life cycle, maintenance and material compatibility studies support it. Heat rejection should be considered separately from heat capture. In the Saudi context the thermal system will have to be judged on annual energy, annual water use, peak electrical demand and performance in the extreme summer rather than on nominal CoP alone.

3.3 Water Stewardship and Heat Rejection

Water is a strategic constraint as evaporative cooling can reduce electricity use at the expense of increased direct water demand. The hot-arid case study by Karimi et al. clearly illustrates this trade-off with one facility having a much lower average PUE but significantly higher source

water use [5]. This is very much the case for Saudi Arabia where water scarcity gives WUE the same order of importance as PUE. Any design should therefore incorporate both the direct facility water and the upstream water associated with the electricity where data are available.

Free cooling research is of value but needs to be interpreted in a climate specific way. Water side free cooling can be very effective in locations where there are sufficient hours of the year when the ambient temperature is below the required approach temperature [15]. Integrated free cooling and heat recovery systems have also been shown to deliver significant benefits in a cool climate [16]. These results would not be transferable in a straight line to the Saudi design as the free cooling window is much smaller there. Radiative cooling is more promising as it allows you to reject heat to the sky without having to evaporate any water; modelling has shown considerable energy and water savings in certain configurations [14]. For Saudi Arabia the technology is worth investigating locally with a test programme that would include the effects of dust, sky conditions, surface degradation and the high levels of solar load during the day.

The preferred heat rejection strategy should be adaptive. Air cooled dry heat rejection protects water resources but can be very inefficient in the heat. Evaporative or hybrid systems will reduce peak electricity but will require water risk controls. Liquid cooling doesn't get rid of the need for heat rejection, it just shifts the temperature at which heat can be rejected. A higher return water temperature may in fact improve the performance of a dry cooler and may lead to reduced compressor running, making direct liquid cooling potentially beneficial for both energy and water resilience. Water treatment, leak detection, alternative water sources and drought operating modes should all be design requirements rather than operational afterthoughts.

3.4 Renewable Power, Storage and Grid Interaction

Saudi Arabia's solar resource provides a very strong case for linking data centre expansion to renewable generation. The challenge of course is that the supply is intermittent and the demand

is continuous. Work on renewable powered data centres has shown that high levels of renewable utilisation require a coordinated approach to generation, batteries, server provision and workload scheduling [18-21]. Kwon has shown that high renewable utilisation and good service quality can be achieved simultaneously rather than being seen as conflicting objectives [20] and there is research into the co-design of battery storage and data centre loads within the context of smart grids [21].

For critical Saudi facilities, on-site PV generation should be viewed as part of a layered energy strategy rather than as a substitute for grid connection or for standby capacity. Storage can provide short duration ride through, demand management and renewable shifting. UPS will still be required for fault isolation and power quality. Longer duration resilience may require multiple grid feeds, storage of various discharge durations and a progressively decarbonised dispatchable backup. The most resilient architecture is not simply one that is "renewable powered" it is one that can maintain computing service through the variability of the renewables, grid disturbances and equipment failures.

Flexibility can be a major benefit to this architecture. Workloads vary in their latency and deferrability requirements so non-critical computing can be shifted to times of higher renewable availability or lower cooling intensity. Zero carbon portfolio studies have demonstrated that multi-energy systems and proactive demand response can lead to a reduction in system cost whilst increasing the amount of renewables that can be accommodated [24]. The potential for such flexibility is particularly strong in the case of training workloads, batch analytics and non-urgent data processing. It would be wrong to assume that this kind of flexibility can be applied to real time government, financial or safety critical services however. The Saudi design framework therefore segregates hard availability loads from the schedulable loads and only applies energy flexibility where the service level agreements allow it.

3.5 Waste Heat, Circularity and Climate Suitability

282

283 Waste heat recovery is regularly pitched as a route to improved data centre resource efficiency.
284 Data centres are often described as potential prosumers of energy with the electricity that goes
285 into the IT equipment in the end becoming waste heat which in principle could be re-used [1].
286 Experimental systems integrating free cooling with heat recovery have performed well in cold
287 climates [16] and adsorption based recovery has been shown to reduce the cooling energy
288 demand under appropriate temperature conditions [17]. Net zero feasibility studies have also
289 demonstrated that waste heat can contribute to the building level energy balance where there is
290 an adjacent heat demand [18].

291

292 The Saudi case is more complicated in that the space heating demand is relatively low. The
293 conventional district heating pathways therefore may not be as valuable as they would be in, say,
294 northern Europe or northern China. Higher temperature liquid cooling does open up a number of
295 useful options (absorption, industrial use of low grade heat, desalination preheat, local thermal
296 applications etc) but each of those requires a matched and economically viable heat sink. A
297 resilient design shouldn't be predicated on the continuous export of heat unless that external
298 demand is both contractually and technically secure. Heat recovery is best viewed as an optional
299 circularity layer which is a performance enhancer if there is a secure heat sink, rather than as a
300 panacea which justifies the siting of a datacentre in all circumstances.

301

302 3.6 Digital Control, Predictive Operation and Resilience

303

304 The most significant trend at the moment is the integration of the physical infrastructure with
305 digital control. Model free RL has been shown to be able to optimise cooling setpoints in a
306 manner that still respects the operational constraints [22]. Digital twin work has taken this further
307 with a system that links in real time the monitoring, simulation, diagnosis and optimisation. This
308 has resulted in a significant reduction in cooling energy in a real data centre heat pipe system

[23]. This is particularly relevant in the context of Saudi Arabia where static setpoints are not well suited to the large seasonal and diurnal temperature swings.

Digital control can also be resilient if it's designed with fault awareness rather than efficiency per se. A thermal digital twin can spot abnormal temperature profiles before they lead to equipment failure. Predictive models can detect a loss of cooling capacity. Workload orchestration can shed or migrate non-critical workloads when power or cooling margins are getting tight. The same control layer can manage battery SoC, renewable availability, chiller staging and rack level thermal conditions. The more sophisticated the control, the greater the cyber and model risks. Manual override, fail safe local control and independent protection are all still required. Resilience is about the optimisation software enhancing the physical redundancy rather than being the single point of failure.

4. Integrated Design Framework for Saudi Arabia

The evidence points to a multi-layered design approach in which sustainability and resilience are established prior to any equipment selection. The first layer is the site and climate. Site assessment should cover grid capacity and redundancy, solar availability, exposure to flooding and extreme heat, water security, dust conditions, diversity of telecoms and proximity to opportunities for heat or water reuse. The Saudi digital infrastructure studies highlight the increasing strategic significance of domestic data capacity [29] and the national sustainability work shows that digital growth must be planned in such a way as to be in line with environmental objectives rather than simply assumed to be 'green' [30].

Integrated sustainable-resilient data-centre design framework

A layered design sequence couples climate adaptation, resource efficiency and service continuity.



Figure 2. Layered framework for coupling climate adaptation, resource performance and operational resilience in Saudi data centres.

The second layer is passive and spatial design. Data centres are of course equipment intensive buildings but orientation, solar exposure, envelope performance, service yard layout and internal zoning all have an impact on cooling loads and maintainability. The white space layout should be designed to minimise recirculation and to support containment. Structural and service routes should be considered in such a way as to allow for future liquid cooling distribution without the need for major reconstruction. Modular electrical and thermal plants will improve fault isolation and enable staged capacity growth, thereby reducing the energy penalty of having oversized equipment running at low loads.

The third layer is thermal architecture. A Saudi facility should be designed around a density map rather than a uniform cooling assumption. The air cooled areas can be designed with containment and variable airflow; the high density areas should be ready for direct liquid cooling and heat rejection should be selected via an hourly climate simulation comparing dry, hybrid and other low water options. Table 1 takes the literature and turns it into Saudi design priorities and Table 2 compares the principal thermal and energy strategies. The point is not to prescribe one

configuration but to ensure that every option is evaluated in the context of energy, water, resilience and maintainability.

Table 1. Evidence-to-design translation for Saudi data-centre projects.

Domain	Evidence from 2020–2025 literature	Saudi design implication	Core verification metric
Facility efficiency	PUE alone does not capture water, carbon, IT productivity or resilience [3,4].	Use a balanced performance dashboard rather than a single target.	PUE; WUE; carbon intensity; availability
Thermal density	Containment and airflow optimisation improve air cooling, but density increasingly favours liquid approaches [6–13].	Zone by rack density and provide liquid-ready routes from initial build.	Rack kW; inlet compliance; cooling kW/IT kW
Heat rejection & water	Hot-arid operation creates direct energy-water trade-offs [5,14–16].	Compare dry, hybrid and low-water options using hourly climate simulation.	Annual/peak WUE; peak cooling demand
Renewables & storage	Generation, batteries and workloads can be coordinated while maintaining service quality [19–24].	Separate critical loads from schedulable loads and size storage by service role.	Renewable match; storage duration; ride-through

Domain	Evidence from 2020–2025 literature	Saudi design implication	Core verification metric
Life cycle & resilience	Advanced cooling benefits depend on equipment, infrastructure and electricity context [25,26].	Assess embodied impacts, service life, repairability and climate failure modes.	Life-cycle GHG; component life; capacity margin

Table 2. Comparative assessment of principal cooling and energy strategies for hot-arid application.

Strategy	Principal strength	Main Saudi constraint	Best-fit application	Priority indicators
Contained air cooling	Mature, serviceable and effective at moderate density.	Fan energy and diminishing control at high heat flux.	General enterprise racks and moderate density.	Inlet temperature; fan power; recirculation
Dry / hybrid heat rejection	Reduces dependence on evaporative water.	Dry performance degrades during extreme heat; hybrid modes need water governance.	Water-constrained sites with robust design-day capacity.	Peak kW; annual water; approach temperature
Water-side / free cooling	Can substantially reduce compressor operation in suitable weather.	Limited annual hours under hot conditions.	Seasonal operation or cooler Saudi microclimates.	Free-cooling hours; compressor kWh

Strategy	Principal strength	Main Saudi constraint	Best-fit application	Priority indicators
Direct-to-chip liquid cooling	Efficient heat capture and higher coolant temperature.	Facility plumbing, leak management and maintenance transition.	AI/HPC and high-density clusters.	Coolant temperatures; pump power; rack density
Immersion cooling	Very high heat-transfer capability with reduced air movement.	Fluid compatibility, service procedures and supply-chain maturity.	Selected ultra-high-density deployments.	Fluid life; auxiliary power; maintainability
PV + storage + workload flexibility	Reduces carbon intensity and can support grid interaction.	Solar mismatch with 24/7 critical demand; storage cost/duration.	Flexible batch loads plus resilient grid-backed operation.	Renewable match; state of charge; service-level compliance

359

360 The 4th layer is the electrical energy system. There should be dual power paths, efficient
361 conversion, energy storage and the integration of appropriate levels of onsite renewables. The
362 Saudi renewable energy policy and the development of the resource base provide a very
363 favourable strategic context for solar integration [28]. Reliability will of course require a
364 diversified supply and realistic storage sizing. Carbon aware or renewable aware workload
365 scheduling should be applied only to the flexible computing. The essential loads must maintain
366 their deterministic power and cooling margins irrespective of any optimisation that may be
367 possible.

368

369 The fifth layer is digital governance. Sensors should be providing us with very fine grained data
370 on IT load, temperature, flow, humidity, electrical losses, water consumption and equipment
371 state. That data should be feeding into validated models or digital twins which in turn can be

used for predictive control, anomaly detection and scenario testing. The operational dashboards should be showing PUE as well as WUE, carbon intensity, renewable matching, thermal compliance, capacity margin and availability. The performance targets should be annual and peak condition based so that a design which looks very efficient in the mild months doesn't suddenly become a water or power hog in the summer.

5. Discussion

The key finding of this review is that sustainable data centre design in Saudi Arabia is fundamentally a matter of coupled optimisation. There are a number of very good technologies in the literature but their benefits are highly dependent on climate, density, energy source and operational context. Free cooling can be very effective but is climate constrained; evaporative cooling can reduce the electricity requirement at the expense of increased water use; liquid cooling can improve the thermal efficiency but does so at the cost of increased maintenance and fluid management requirements; and renewable generation can reduce the operational carbon but introduces a temporal mismatch. The right design solution is therefore a portfolio of measures all linked by a common set of performance criteria.

This also fundamentally changes the definition of resilience. The traditional model of datacentre resilience is all about redundancy and backup capacity. That's still important, but climate resilience requires us to ask a whole different set of questions. How does the facility operate during a period of sustained extreme heat, reduced water availability, poor outdoor air quality, renewable shortfall or grid stress? A design with redundant chillers but a single point of failure in the heat rejection might still be a common mode failure. Likewise a facility with a massive amount of diesel backup might be able to achieve its availability targets in the short term, but that's entirely at odds with the long term decarbonisation agenda. Resilience should be viewed as the ability to maintain the critical digital service by being able to switch between viable energy, cooling and operating modes.

400

401 The framework also fits better with Vision 2030 than the rather limited 'green datacentre'
402 concept. The Saudi research on the low carbon transition is very much focused on the need for
403 integrated policy and institutional arrangements [27] and the renewable energy studies point to a
404 continued expansion of non-fossil generation as part of the national transformation agenda [28].
405 Digital infrastructure can help with diversification but there's recent Saudi evidence that ICT
406 diffusion can actually exacerbate environmental pressures in the early stages before the benefits
407 kick in at a more mature stage [30]. That again points to the need to build sustainability into the
408 growth of the infrastructure.

409

410 There's another issue relating to life cycle and urban system boundaries. Efficiency in operation
411 can mask the environmental impacts of the equipment itself (in terms of manufacture), the
412 cooling hardware, the water supply and the electricity generation. There's now life cycle
413 evidence that advanced cooling can in fact lead to reductions in GHG emissions, energy demand
414 and blue water use when the technology is appropriately matched to the server density and the
415 infrastructure conditions [25]. At the urban system level the data centres are competing for land,
416 grid capacity and water and at the same time they create opportunities for flexible demand,
417 infrastructure investment and selective heat reuse [26]. The Saudi projects should therefore
418 establish a design baseline that records the embodied and operational impacts separately and then
419 test how each of the major interventions affects both. This is particularly important when
420 considering the replacement of air cooling with liquid cooling, the addition of batteries, the
421 expansion of the PV capacity or the choice of heat rejection equipment as a lower annual
422 operating load may be offset by additional materials, refrigerants and replacement cycles. Life
423 cycle assessment should therefore be integrated with resilience assessment. The equipment is
424 expected to perform reliably under the conditions of extreme heat, dust and power disturbance
425 and it should be assessed not just on its nominal efficiency but on its service life, repairability,
426 component availability and degradation. This boundary ensures that short term efficiency gains
427 are not erroneously classified as sustainable in the long term and makes the trade-offs clear to the
428 designers, owners and regulators prior to any capital being committed. It also enables the

comparison of competing design options under the same functional, climatic and availability assumptions throughout the project life.

A practical priority is the generation of Saudi specific design data. Hourly weather files are available locally but the research base is lacking in measured data on PUE, WUE, rack density, cooling system performance, water quality, dust effects, failure events and the matching of renewables across the various operational facilities. Designers will have to extrapolate from Phoenix, Singapore, China and Europe in the absence of such data. The data from those locations is of course useful, but the transferability is very limited. A national benchmarking programme using anonymised operational data would allow for direct comparison of climate zones, facility types and cooling architectures on a common basis and would go a long way towards improving future design decisions.

6. Research Gaps and Future Research

Four research gaps are particularly significant. The first is the lack of empirical data on liquid cooling under actual Saudi dust, water quality and maintenance conditions. The second is that very few studies have actually attempted to jointly optimise PUE, WUE, carbon intensity, reliability and life cycle impacts for a hot arid data centre environment (the majority of work to date has only focused on one or two of these). The third is that the interaction between AI workloads and national grid planning is something that will require location specific modelling as the high density of the data centres will create very concentrated electrical demand. The fourth is that heat reuse research should look at heat sinks that are relevant to Saudi Arabia rather than simply import the district heating assumptions. Future work should combine instrumented pilot facilities, validated digital twins, life cycle assessment and multi-objective optimisation so that the design decisions can be tested under both normal and extreme operating conditions.

7. Practical and Policy Implications

For developers procurement specs should be asking for multi-metric performance guarantees rather than just a single PUE. Cooling tenders should be providing annual electricity, peak demand, direct water use, expected performance at design day and upgrade paths for higher rack density. The electrical design should separately quantify the role of renewable matching and storage from the role of backup reliability. Commissioning should include failure mode testing, thermal mapping and verification of control sequences under partial load and extreme heat conditions.

For policymakers the incentives for data centre growth need to be tied to transparent energy and water reporting, climate resilience planning and the integration of renewables without compromising grid reliability. Standards can be used to promote liquid ready infrastructure, reclaimed water compatibility where appropriate and metering at the subsystem level. A national performance database would help establish Saudi benchmarks and ensure that imported targets don't in effect become the design rules. This would underpin the wider economic value of the digital infrastructure whilst protecting our energy and water resources.

8. Limitations

This review is constrained by the paucity of peer-reviewed operational data from Saudi data centres and so part of the evidence base is drawn from other climates. The literature selected is limited to journal papers from 2020 to 2025 with verifiable DOIs (to ensure currency) and so excludes the more foundational engineering work and non-peer reviewed industry data. The range of study methodologies used also means that direct numerical comparison across all the

technologies is not possible. The framework proposed is therefore to be viewed as an evidence based design logic which will need to be underpinned by project specific modelling rather than as a prescriptive standard applicable in all cases.

9. Conclusion

Saudi Arabia's nascent digital economy will require data centres that are efficient, resource aware, climate appropriate and operationally resilient. The literature makes it clear that you can't achieve this through one cooling technology, one renewable energy target or one efficiency figure. In the hot arid environment you're inevitably going to have energy-water trade-offs and the increasing computing density will make the case for liquid ready thermal systems, sophisticated modelling and adaptive control even more compelling. Renewable generation and storage can help to reduce the carbon intensity and improve grid interaction, but only if you're able to properly coordinate that with the criticality of the workload and the need for deterministic reliability.

The most defensible design approach is a whole-system, layered one: climate informed siting, modular and upgradeable spatial planning, density matched cooling, low water heat rejection, resilient power architecture, renewable integration, digital twin enabled operation and governance via a range of performance metrics. If you're going to build out the data centre estate in Saudi Arabia then the right way to do it is to embed these principles into the planning process from the outset. That way you can support Vision 2030 and at the same time reduce your exposure to resources and improve your continuity under the future climate and future workload scenarios. The key design issue is not how to optimise one piece of the system, it's how to maintain the digital service whilst continuously reducing the total energy, water, carbon and resilience cost of computation.

References

- [1] Huang P, Copertaro B, Zhang X, Shen J, Löfgren I, Rönnelid M, Fahlen J, Andersson D, Svanfeldt M. A review of data centers as prosumers in district energy systems: Renewable energy integration and waste heat reuse for district heating. *Applied Energy*. 2020;258:114109. doi:10.1016/j.apenergy.2019.114109.
- [2] Liu L, Zhang Q, Zhai ZJ, Yue C, Ma X. State-of-the-art on thermal energy storage technologies in data center. *Energy and Buildings*. 2020;226:110345. doi:10.1016/j.enbuild.2020.110345.
- [3] Shao X, Zhang Z, Song P, Feng Y, Wang X. A review of energy efficiency evaluation metrics for data centers. *Energy and Buildings*. 2022;271:112308. doi:10.1016/j.enbuild.2022.112308.
- [4] Al Kez D, Foley AM, Lavery D, Furszyfer Del Rio D, Sovacool BK. Exploring the sustainability challenges facing digitalization and internet data centers. *Journal of Cleaner Production*. 2022;371:133633. doi:10.1016/j.jclepro.2022.133633.
- [5] Karimi L, Yacuel L, Degraft-Johnson J, Ashby J, Green M, Renner M, Bergman A, Norwood R, Hickenbottom KL. Water-energy tradeoffs in data centers: A case study in hot-arid climates. *Resources, Conservation and Recycling*. 2022;181:106194. doi:10.1016/j.resconrec.2022.106194.
- [6] Xiong X, Fulpagare Y, Lee PS. A numerical investigation of fan wall cooling system for modular air-cooled data center. *Building and Environment*. 2021;205:108287. doi:10.1016/j.buildenv.2021.108287.
- [7] Tang Y, Bai X, Jin C, Fu Y, Mao W, Xu X. Experimental and numerical study on the thermal environment of a data center. *Building and Environment*. 2022;212:108741. doi:10.1016/j.buildenv.2021.108741.

- [8] Liu W, Lian S, Fang X, Shang Z, Wu H, Zhu H, Hu S. An open-source and experimentally guided CFD strategy for predicting air distribution in data centers with air-cooling. *Building and Environment*. 2023;242:110542. doi:10.1016/j.buildenv.2023.110542.
- [9] Hu B, Yin Z, Hamrani A, Leon A, McDaniel D. Super-resolution-assisted rapid high-fidelity CFD modeling of data centers. *Building and Environment*. 2024;247:111036. doi:10.1016/j.buildenv.2023.111036.
- [10] Alkrush AA, Salem MS, Abdelrehim O, Hegazi AA. Data centers cooling: A critical review of techniques, challenges, and energy saving solutions. *International Journal of Refrigeration*. 2024;160:246-262. doi:10.1016/j.ijrefrig.2024.02.007.
- [11] Azarifar M, Arik M, Chang JY. Liquid cooling of data centers: A necessity facing challenges. *Applied Thermal Engineering*. 2024;247:123112. doi:10.1016/j.applthermaleng.2024.123112.
- [12] Kong R, Zhang H, Tang M, Zou H, Tian C, Ding T. Enhancing data center cooling efficiency and ability: A comprehensive review of direct liquid cooling technologies. *Energy*. 2024;308:132846. doi:10.1016/j.energy.2024.132846.
- [13] Wang N, Guo Y, Huang C, Shao S. Advances in direct liquid cooling technology and waste heat recovery for data center: A state-of-the-art review. *Journal of Cleaner Production*. 2024;477:143872. doi:10.1016/j.jclepro.2024.143872.
- [14] Aili A, Long W, Cao Z, Wen Y. Radiative free cooling for energy and water saving in data centers. *Applied Energy*. 2024;359:122672. doi:10.1016/j.apenergy.2024.122672.
- [15] Mi R, Bai X, Xu X, Ren F. Energy performance evaluation in a data center with water-side free cooling. *Energy and Buildings*. 2023;295:113278. doi:10.1016/j.enbuild.2023.113278.
- [16] Ding J, Zhang H, Leng D, Xu H, Tian C, Zhai Z. Experimental investigation and application analysis on an integrated system of free cooling and heat recovery for data centers. *International Journal of Refrigeration*. 2022;136:142-151. doi:10.1016/j.ijrefrig.2022.01.003.
- [17] Gupta R, Puri IK. Waste heat recovery in a data center with an adsorption chiller: Technical and economic analysis. *Energy Conversion and Management*. 2021;245:114576. doi:10.1016/j.enconman.2021.114576.

- [18] Güğül GN, Gökçül F, Eicker U. Sustainability analysis of zero energy consumption data centers with free cooling, waste heat reuse and renewable energy systems: A feasibility study. *Energy*. 2023;262:125495. doi:10.1016/j.energy.2022.125495.
- [19] Rostirolla G, Grange L, Minh-Thuyen T, Stolf P, Pierson JM, Da Costa G, Baudic G, Haddad M, Kassab A, Nicod JM, Philippe L, Rehn-Sonigo V, Roche R, Celik B, Caux S, Lecuivre J. A survey of challenges and solutions for the integration of renewable energy in datacenters. *Renewable and Sustainable Energy Reviews*. 2022;155:111787. doi:10.1016/j.rser.2021.111787.
- [20] Kwon S. Ensuring renewable energy utilization with quality of service guarantee for energy-efficient data center operations. *Applied Energy*. 2020;276:115424. doi:10.1016/j.apenergy.2020.115424.
- [21] Guo C, Luo F, Cai Z, Dong ZY, Zhang R. Integrated planning of internet data centers and battery energy storage systems in smart grids. *Applied Energy*. 2021;281:116093. doi:10.1016/j.apenergy.2020.116093.
- [22] Mahbod MHB, Chng CB, Lee PS, Chui CK. Energy saving evaluation of an energy efficient data center using a model-free reinforcement learning approach. *Applied Energy*. 2022;322:119392. doi:10.1016/j.apenergy.2022.119392.
- [23] Zhu H, Lin B. Digital twin-driven energy consumption management of integrated heat pipe cooling system for a data center. *Applied Energy*. 2024;373:123840. doi:10.1016/j.apenergy.2024.123840.
- [24] Xu D, Xiang S, Bai Z, Wei J, Gao M. Optimal multi-energy portfolio towards zero carbon data center buildings in the presence of proactive demand response programs. *Applied Energy*. 2023;350:121806. doi:10.1016/j.apenergy.2023.121806.
- [25] Alissa H, Nick T, Raniwala A, Arribas Herranz A, Frost K, Manousakis I, et al. Using life cycle assessment to drive innovation for sustainable cool clouds. *Nature*. 2025;641:331-338. doi:10.1038/s41586-025-08832-3.

- [26] Liu FHM, Lai KPY, Seah B, Chow WTL. Decarbonising digital infrastructure and urban sustainability in the case of data centres. *npj Urban Sustainability*. 2025;5:15. doi:10.1038/s42949-025-00203-1.
- [27] Alomari MA, Heffron RJ. Utilising law in the transition of the Kingdom of Saudi Arabia to a low-carbon economy. *Environmental Innovation and Societal Transitions*. 2021;39:107-118. doi:10.1016/j.eist.2021.03.003.
- [28] Samargandi N, Islam MM, Sohag K. Towards realizing vision 2030: Input demand for renewable energy production in Saudi Arabia. *Gondwana Research*. 2024;127:47-64. doi:10.1016/j.gr.2023.05.019.
- [29] Al-Nefaiee AH, Alhaif AM, Algoblan N. Data Centers in the Kingdom of Saudi Arabia and their Role in Big Data Processing for Decision Support. *Journal of Educational and Social Research*. 2022;12(2):237-248. doi:10.36941/jesr-2022-0048.
- [30] Alfehaid F, Omri A, Altwaijri A. Impact of ICT diffusion and opportunity entrepreneurship on environmental sustainability in Saudi Arabia. *Heliyon*. 2024;10(19):e39009. doi:10.1016/j.heliyon.2024.e39009.