绿色低碳数据中心
有效实现灵活协同、系统融合与能源安全

丹佛斯[①]

引言

随着数字化的飞速推进，人工智能的快速普及正在重塑社会的各个层面。这场变革的核心基础设施正是数据中心，其构筑的数字骨干支撑着现代社会的发展。

对数据中心算力、存力以及由此带来的能源的需求正以前所未有的速度增长，在多个地区，该需求已超过传统基础设施的发展速度。精心规划数据中心并将其融入能源系统的重要性相较于以往更为紧迫，同时这也为构建更智能且更具韧性的能源系统提供了关键机遇。数据中心既是未来技术升级的必要前提，也是制定经济脱碳解决方案的关键所在，助力大幅提升各行业的能效，包括交通运输、建筑、工业等几乎所有领域。据国际能源署（IEA）预测，人工智能有望在 2035 年实现相当于全球能源相关排放量 5%的减排量。随着数据中心日益融入我们的日常生活乃至未来发展，实现数据中心可持续部署，无疑将在塑造下一代能源系统的进程中发挥关键作用。

我们深知，数据中心在承担支撑我们所依赖的数字化运行这一关键任务时消耗着大量资源。到 2030 年，数据中心消耗的电力将高达全球总耗电量的3%。从区域层面看，这种影响可能对各地电网造成集中冲击。80%的电力需求增长将来自中美两国。同时，数据中心目前的年耗水量约为 5,600 亿升，

[①] 本报告仅代表企业相关研究观点，不代表论坛主办单位和承办单位立场和观点。

到 2030 年可能增至 1.2 万亿升。若不能从整体上进行规划，保证数据中心部署的可持续性，其将使整个社会尤其是能源和水资源持续承压。

数据中心不能仅被视为资源消耗者。实际上，在优化全球能源系统方面，数据中心也是最具潜力的未开发领域之一。践行“以脱碳为竞争力”的战略，将数据中心绿色转型从单纯的环境责任转变为竞争力提升机遇。我们可以通过资源效率提升、灵活运营、余热回收利用、妥善融入本地能源生态，匹配可再生能源发电并支持电网灵活性等方式将这些高耗能设施转化为重要资产。若管理得当，数据中心可助力构建坚韧高效的能源系统。因此，政策制定者可采用全局规划策略来释放潜能并应对数据中心快速建设过程中所带来的挑战。提升设施设计与运营效率可显著减少预期的资源需求，同时确保数字化应用的增长匹配能源系统的承载力。通过跨行业协作能推动新技术开发与应用，促使能效提升至新水平。

我们对共同追求和建设的未来有着清晰而坚定的认识。数字化浪潮正在席卷全球经济与社会各领域，这一必然趋势将为创新与效率提升带来重大机遇。当下，我们必须在保持紧迫感的同时兼具远见卓识。我们必须将资源效率作为数据中心运营的坚实基础，通过规划让数据中心在未来成为灵活高效、坚韧可靠的能源系统的重要合作伙伴。

方行健（Kim Fausing）

丹佛斯集团总裁兼首席执行官

摘要

在人工智能与数字经济深度融合的当下，算力基础设施已成为支撑全球产业创新与社会数字化转型的核心底座，而作为算力核心载体的数据中心，也正从支撑传统信息化建设的计算与数据存储载体，向兼具数字赋能与绿色转型双重价值的新型基础设施演进。数字技术与人工智能（AI）应用的拓展，在催生算力需求指数级增长的同时，也让数据中心的能源、水资源消耗与碳排放问题受到全球范围的广泛关注。如何在保障算力稳定供给的同时，化解行业发展带来的资源环境压力，已成为全球产业共同面对的核心课题。结合国际能源署（IEA）等权威机构的专项研究数据与全球各地的标杆实践，探寻算力时代数据中心绿色低碳转型的发展方向与实践路径，助力行业在算力规模化增长与低碳发展目标之间寻找适配的平衡。

人工智能依托的智能算力的规模化落地，是驱动数据中心能耗快速攀升的核心动因。AI 大模型算力需求能耗远高于传统数字应用，伴随多模态场景的拓展，算力需求的增长节奏已远超传统半导体技术的演进周期。据国际能源署（IEA）2025 年相关报告测算，全球数据中心耗电量在全球终端能源消费中的占比，将从 2024 年的 1.5%翻倍至 2030 年的近 3%，同步增长的水资源消耗与碳排放，为行业发展带来了刚性挑战。面对算力增长与资源约束的平衡矛盾，对于占全球数据中心市场 85%的美国、中国和欧盟有着各自不同的政策引导与驱动。行业实践也同时印证，数据中心并非单纯的能耗主体，它既是数字经济高质量发展的核心算力底座，也是能源体系低碳转型的重要参与力量，随着未来低碳算力的广泛渗透，将赋能传统产业数字化与低碳化协同转型。

作为现代化产业体系的重要基础设施，数据中心可通过构建全链条低碳技术体系、深度融入能源系统，实现自身运营效率的全面跃升，更可转化为新型电力系统的优质灵活性资源。一、减量化（Reduce）既源头控耗，其核

心是算力设备的能效跃升，AI 芯片的代际技术迭代，是实现算力规模增长与能耗增速脱钩的核心支撑；而液冷等先进冷却技术的规模化应用，也为高密度智算中心的能效优化提供了成熟解决方案。二、数据中心运行过程中产生的大量低品位余热再利用（Reuse），通过热泵技术提升温度，接入区域供热系统实现能源梯级利用，北欧及中国已在多个项目实践中验证了其经济与环保双重价值。三、资源化（Resource）既价值重构可依托于数据中心与可再生能源体系的深度融合，通过源网储荷一体化布局、绿电市场化交易提升本地可再生能源消纳水平。通过算力负荷的时空可调特性，同时以跨区域算力调度、电网需求响应等方式参与电力系统灵活调节，让数据中心从单纯的能源消费者，转变低碳供热方以及为支撑新型电力系统建设、促进新能源消纳的重要载体。

面向未来，数据中心的绿色低碳发展，既需要以持续的技术创新筑牢能效底线，也需要以系统思维打通算力与热力、电力的协同壁垒，最终实现数字经济增长与低碳转型的适配，让数据中心发展真正成为推动社会绿色转型的中坚力量。

一、打破数据中心能源困局

近二十年全球数据中心产业取得了蓬勃发展，通过不断地新技术创新与融合，其形态从计算中心、信息中心、云中心加快向算力中心演变。当今世界正在快速拥抱新兴数字技术与人工智能，它们已成为人们日常生活的一部分。无论是工业界对新产品进行的虚拟测试，还是移动端各种流媒体的社交应用，亦或准确的天气预报以及我们正在目睹的，在严重疾病治疗领域不断取得的重大突破，均表明数字信息技术与人工智能正在改变人类的生活，人们对其需求的持续增长已是不争事实。现在，我们面临的重要挑战是如何在支持这种增长的同时，管理数据中心对关键资源的需求压力。政府和其他决

策者有责任实现这种增长，并管控数据中心对能源和水等资源的需求。防止产生能源困局的根本在于提升算力载体的用能效率，并将数据中心融入我们的能源系统，实现算力与能源系统的协同。

（一）算力时代的资源需求激增

从全球范围看，数据中心算力与存力已成为数字经济的“新石油”。近年来人工智能（AI）技术的应用加速，使 AI 服务器正迅速成为数据中心的算力设备能耗增长的核心驱动力。数据显示，AI 大模型以及 AI 大模型赋能的搜索引擎在单次请求中的能耗，显著高于传统搜索引擎单次请求的能耗水平。例如，标准 Google 搜索单次请求耗电量 0.3 瓦时，而 ChatGPT 大模型单次请求耗电则为 2.9 瓦时 [1]。随着 AI 大模型从文生文向文生图、文生视频等更复杂运算演进，算力与存力需求呈指数级增长，由此带来的数据中心能源需求也呈现显著提升趋势。

根据国际能源署（IEA）发布的 2025 年聚焦 AI 发展对全球数据中心电力需求驱动影响的专项报告 [2]，依据基准情景（Base Case）的预测，数据中心在全球终端能源消费中的占比将于 2030 年翻倍。从 2024 年的 1.5%（总耗电量达 415 太瓦时，与英国全年的用电量相当）上升至 2030 年近 3%（945 太瓦时，接近日本当前的全国用电规模）。自 2017 年以来，全球数据中心电力消耗年均增长约 12%，全球发达经济体数据中心用电增量将贡献整体电力需求增长的 20%以上。

2024 年，美国数据中心用电规模占全球最大的份额，达到了 180 太瓦时，占全球总量的 45%；中国和欧洲的数据中心耗电量分别占据了 25%和 15%的份额。美国、中国、欧洲仍为 2030 年前数据中心用电规模最大的三个区域。其中中美两国贡献了全球近 80%的用电增长：美国数据中心 2030 年较 2024 年用电增长约 240 太瓦时（增幅 130%），中国增长约 175 太瓦时

（增幅 170%），欧洲增长超 45 太瓦时（增幅 70%）；AI 驱动的服务器用电年均增长 30%，贡献了近一半的净用电增量；传统服务器年均增速仅 9%，贡献约 20%增量；冷却等配套基础设施、其他 IT 设备分别贡献 20%、10%的增量。

随着 AI 技术的不断发展，对能源消耗速度显著提升，其增长速度已打破半导体摩尔定律两年倍增周期。根据斯坦福大学研究资料及国际能源署（IEA）相关分析估算，GPT-4 模型使用高达 25,000 个 GPU 的算力集群训练长达 14 周，功耗超 22 MW，共消耗了 42.4 吉瓦时的电力（约是 GPT-3 模型训练耗电量的 30～40 倍），平均每天约 0.43 吉瓦时，相当于 2.85 万户发达经济体国家家庭的日用电总量。虽然中国的大语言模型 DeepSeek 带来了算法效率的提升，但这并未抑制对算力需求。反而因更多的用户和场景的加入，推动大模型普及与应用落地。从长期来看，算力设备的能耗仍将保持强劲上升态势。与此同时，数据中心的许多设施高度依赖水冷的温控方式，所以其水资源占用情况同样关键。全球数据中心年耗水量达到 5,600 亿升，到 2030 年可能升至 12,000 亿升，相当于欧盟 2022 年淡水总抽取量的六倍。若按照当前趋势，全球数据中心碳排放也将从 2024 年 1.8 亿吨跃升至 2035 年 3 亿吨，对全球针对气候变化进行的减碳努力带来挑战。

（二）政策引导和驱动数据中心的可持续发展

随着人工智能应用的加速和对数据中心需求的不断扩大，全球数据中心向大型数据中心迁移趋势明显，绿色低碳也进一步推进。截至 2024 年底，全球数据中心平均电能利用效率（PUE—衡量数据中心能源使用效率的指标）为 1.56[3]，比 2017 年的约 1.6 的平均 PUE 略有改善。于此同时，各国政府均面临着双重挑战，既要确保数据中心所在地电力及水资源的稳定可靠且价格合理，又要推动数据中心建设的增长以满足经济和社会发展的数字化、智能

化需求。在政策和行业监管层面，特别是数据中心的绿色低碳发展，各国政府有着不同的路径。

放眼全球，美国凭借其在半导体技术和算法研究方面的领先优势，长期占据全球算力市场的主导地位。过去十多年间，联邦政府曾通过一系列法规推动数据中心整合、信息技术（IT）行业能效提升与高性能绿色计算基础设施建设。但自 2023 年起，美国政府鲜有新的具体政策出台或更新相关目标路径；而相关数据中心行业的企业在能效提升以及绿色低碳发展方面却保持活跃。亚马逊、谷歌、微软等主流科技公司主导的绿电采购协议规模领先全球，也各自承诺了企业及数据中心的碳中和目标路径；美国的学术界也在服务器节能和能源的最优利用等方面做出了相应的贡献。总体来说，美国数据中心行业的绿色发展更多依赖于市场机制、行业责任以及学术研究的努力，而非联邦政府的政策主导。

对于欧盟，数据中心行业政策正历经从自愿引导到强制约束、从局部优化到系统治理的演进过程。2020 年欧盟率先提出包括数据中心在内的信息通信技术（ICT）行业的降碳目标，通过《欧盟数据中心能源效率行为准则的最佳实践指南》以及行业自身的《气候中和数据中心契约（CNDCC）》约束性目标文件等细化和规范节能标准、实施措施，引导建立绿色数据中心推进机制。2023 年《修订版欧盟能效指令（EED）》也标志着进入强制立法、技术指标和监管机制三位一体的硬约束阶段；结合其他相关可再生能源、资源循环及强制披露等标准指令，共同推动欧盟 ICT 基础设施及算力中心在 2030 年之前达到气候中性（Climate Neutral）。2023 年 5 月欧盟碳边境调节机制（简称“CBAM”）也将数据中心相关的 IT 设备进口、跨境云服务、数据处理服务纳入管控范围，要求非欧盟企业进口算力服务时缴纳碳差价，仅 2024 年就推动全球 15 家云服务商建立全生命周期碳足迹追踪系统。

相比海外数据中心政策以激励性与自愿性为主，部分或逐步转向强制性约束为辅的实践相比，中国在数据中心绿色低碳发展方面则更具国家层面的宏观顶层设计和规划。2020 年提出并后续实施的中国一体化大数据中心协同创新体系的“东数西算”工程，是依据中国东西部资源不均衡的局面，规划算力资源跨域布局与调配的顶层设计。布局在清洁能源丰富的西部地区，建设大型的数据中心算力设施，实现绿色减碳并进行集约化管理，为东部大量的算力需求提供计算服务。由此建立的八大算力网络枢纽节点及 10 个国家算力中心集群，有利推动政策和资源的集中，引导数据中心集约化、规模化和绿色低碳发展。

我们注意到近几年中国在国家和区域层面也不断推出数据中心和算力高质量发展的促进政策，从电能及水资源利用效率、可再生能源利用、整体上架率、单位算力能效和碳效等维度提升数据中心审计和准入门槛。截至 2024 年底，全国在用数据中心标准机架数超过 900 万，算力规模近 5 年平均增速近 30%，达到 280 EFLOPS（每秒百亿亿次浮点运算，FP32）[4]，位居全球占比第二，其中智能算力占比显著提升至 32%；全国数据中心平均电能利用效率（PUE）达到 1.46 [4]。国家标准《绿色数据中心评价》也于 2025 年 6 月 1 日起正式实施，结合中国数据中心建设特点与国际先进指标水平，明确了在资源高效利用、绿色设计、绿色运维等多维度应具备的各项指标要求。截至 2026 年初，国家级绿色算力设施达到 306 家，平均电能利用效率（PUE）降至 1.25 [4]，较全国平均值低 14.4%。目前中国数据中心总体算效（Computational Efficiency，CE）达到 21.95 GFLOPS/W（每瓦千兆浮点运算 FP32）[5]，总体存效（Storage Efficiency，SE）达到 102.2 GB/W（千兆字节每瓦），较 2022 年的 88.4 GB/W 同比增长 15.68% [5]。得益于中国一体化数据中心的协同规划、创新建设，可以对于中国数据中心整体能源效率进行如数据中心算效、存效等更多元维度的衡量。

（三）打造卓越的数据中心：效率与机遇的交汇点

算力设施是建设现代化产业体系的重要基础设施，也是中国能源消费增长较快的领域之一。中国工业和信息化部统计数据显示，2024 年中国算力设施用电量达到 1,660 亿千瓦时，约占全社会用电总量的 1.7%。中国“十四五”以来，全国算力基础设施用电量年均增速超过 10%，约为全社会用电量的年均增速的 1.5 倍。随着数字经济的快速发展，数据中心向算力中心演变，其综合能耗将呈现更快速的增长趋势。也有学术研究机构预测到 2030 年，中国数据中心算力的电力负荷约为 1.05 亿千瓦（数据中心耗电在全社会用电占比超 3.5%），水耗约为 2,629 亿升，碳排约为 3.1 亿吨（超过美国数据中心的碳排放量）[6]。

在数据中心未来对于电力及水资源需求不断增长的确定性之下，必须构建全链条算力设备及载体的低碳技术体系，建设高效数据中心，并将其彻底融入能源系统，促进其绿色低碳发展。通过优化的冷却技术、数字化工具以及人工智能（AI）本身来提高运营效率，增加整体透明度并提高生产力。数据中心也将成为新型电力系统一个良好的灵活性资源。其内部的算力、电力和热力灵活性资源，在确保数据中心稳定高效运行的同时，将其融入更广泛的能源系统，实现算—电、算—热协同，数据中心就有可能转化为一种机遇。如果规划得当，这些措施便可以助力逐步淘汰化石能源，减少电网投资，推进数据中心行业的持续节能降碳。同时低碳绿色算力也可以深度融入生产制造、交通运输、服务等各行业，有力驱动传统产业能耗结构提质优化，同步实现数字化转型与节能降碳的双重价值，推动全社会节能降碳效能的协同跃升。国际能源署（IEA）报告也指出，人工智能将成为构建韧性能源系统的基石，到 2035 年，仅在工业领域人工智能便可节省约 2,200 太瓦时能源，这一数字大致相当于欧盟整个工业部门的能源需求，且远高于届时全球数据中心预计总用电量[7]。

二、提高算力载体效率实现可持续增长（减量化 Reduce）

电能使用效率（PUE）是迄今为止立法规定最多的衡量数据中心能源使用效率的核心指标，既数据中心 IT 设备、制冷和供电等辅助设施的总能耗与 IT 计算设备的能耗之比。数据中心的电能使用效率 PUE 数值越低，数据中心整体的能效越高。以 2024 年中国数据中心平均电能利用效率（PUE）1.46 为例，IT 设备电能消耗（服务器、存储、网络设备等算力核心负载）占数据中心总能耗的 69%；基于 IEA/ENERGY STAR/LBNL 等权威资料估算制冷系统能耗平均占比约为 17% [2]；供配电系统平均损耗占总能耗约 10% [8]；其他辅助系统（照明、消防、监控、办公配套等）平均能耗约占总能耗的 4% [9]。对于新建高效超大型数据中心（PUE≤1.25）：IT 设备能耗占比可达 80%以上，制冷系统占比可降至 10%以内。因此数据中心能源消耗的主体还是 IT 算力设备(未来更是集中在 AI 算力服务器)。

（一）算力设备的能效跃升

如前所述，2019～2024 年中国总算力年均增长 30%，达到 280EFLOPS、而算力基础设施耗电量年均仅增长 10%。这种能耗和算力非同步增长的主导因素，正是作为数据中心能耗核心的算力设备本身的算力能效（单位能耗的算力输出）的大幅跃升。AI 加速芯片（GPU/NPU）是算力增长的核心载体，以英伟达（NVIDIA）GPU 产品为例，2020 年发布的 A100 芯片单位功耗的算力输出为（FP16 精度）1.56 TFLOPS/W（万亿次浮点运算每瓦）；2022 年落地的 H100 芯片单位功耗算力输出（FP16 精度）升至 2.83 TFLOPS/W [10]，算力能效提升超 80%。同样对于华为推出的昇腾 AI 处理器，2019 年发布的 Ascend910 在 FP16 精度下算力约 256 TFLOPS、功耗约 310 W；后续 910B 和 910C 通过架构与工艺优化，将算力提升至约 320～640 TFLOPS，同时功耗降低至约 250 ~ 310 W [11]，使单位能耗算力较初代产品提升约 1.5 ~ 3 倍。因此在从通用算力到智能算力的迭代过程种，依托 AI 芯片、存储、网络等算力设

备更先进的工艺制程与架构优化等创新，驱动在算力跃升的同时减少能源的消耗。这也更依赖于半导体行业的创新发展，有其自身的演进规律。

（二）算力载体基础设施提效降碳潜力

在数据中心 IT 算力设备配置架构确定后，这部分算力设备的能耗水平也基本固定（随上架率及业务场景而变化）。而此时数据中心能效提升的潜力更多地来自于占数据中心能耗 20% ~ 30%（对应 PUE 1.25 ~ 1.46）的数据中心制冷、供电等辅助系统。结合不同区域气候特征与数据中心业务属性，通过适配现有的先进节能降碳技术与装备，覆盖数据中心冷却热管理、供配电、系统集成与智能化运维管理等关键环节，实现数据中心运行时更少的能源和水资源消耗，降低碳排放。这包括最大化地利用自然冷源，减少水资源消耗的氟泵相变制冷，针对高功率密度机柜的液冷等更高效的热管理方式；也包括优化数据中心供电架构，减少供电转换环节，降低电力传输与转换过程中的线路损耗，推广高压直流（HVDC）、模块化 UPS、市电直供等技术应用；另外通过数字化解决方案（现场数据实时感知与依负载变化的变频驱动等）以及人工智能运维工具，实现更精准高效的热管理与能源供应，并以预防性维护实现数据中心的安全可靠运行。

（三）释放算力潜力的先进冷却系统

随着芯片性能不断增强，其在相同物理空间内产生的热量急剧增加。数据中心单机柜功耗正逐渐从传统通算中心 4 ~ 6 kW 跃升至当前智算中心的 20 ~ 40 kW，未来部分场景将会突破 100 kW。上升的热密度也逐渐超出传统利用空气做热交换的风冷系统所能应对的范围。为满足增长的高密算力需求，越来越多的数据中心转而采用直接芯片液冷（Direct-to-Chip Cooling）系统，这种闭环系统将冷却液通过冷量分配单元经过二次侧冷却液回路分配输送至处理器芯片，冷却液在服务器端通过冷板直接与处理器芯片做热交换带走芯

片运行时产生的热量；再通过冷量分配单元一次侧循环连接外部冷源（如冷水机组、冷却塔、干冷器）实现最终的热交换。相较于传统冷却方式，这种闭环直接芯片液冷技术的散热效率更高，耗水量大幅减少，通常可实现至少15%的节能效果。在电网容量和供水量受限的地区，这类能效提升手段的重要性日益凸显，其有助于数据中心在需求增长的情况下减少本地资源占用，实现更可靠地运营。液冷技术越来越成为智算中心的标配，从“可选”变为“必选”。同时液冷也将更有利于对数据中心余热的回收再利用。

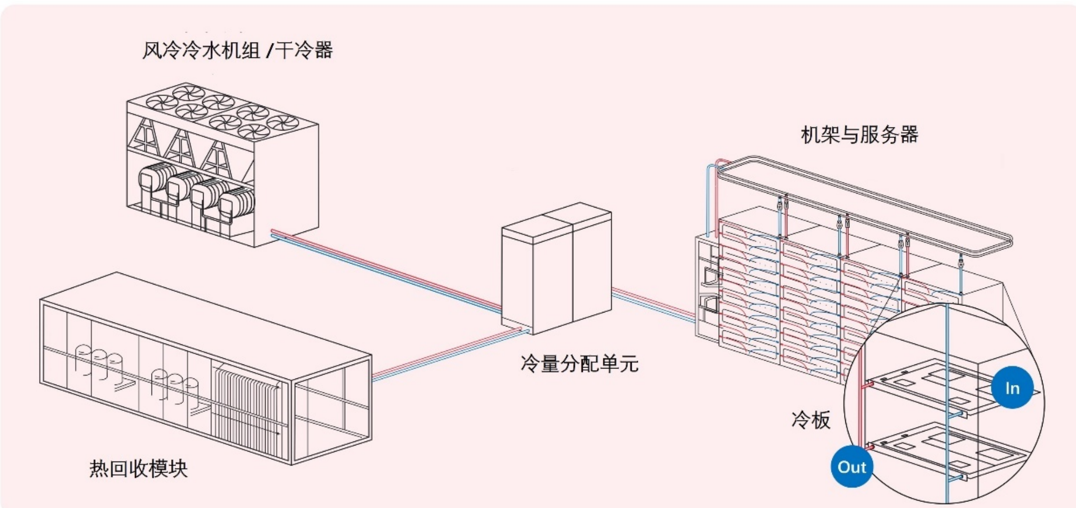

图 1：直接芯片液冷系统架构

近年来，全球人工智能算力需求快速增长，针对更高功率密度的服务器部署，以谷歌、亚马逊和微软等为代表的国际云计算与科技企业，较早开始探索并规模化应用液冷等新型数据中心散热技术。以微软为例，其在数据中心冷却与供能技术方面持续创新。2024 年，公司加速部署“芯片级液冷”系统（chip-level liquid cooling），全面取代传统空气冷却方式。这不仅可以大幅提升服务器密度，减少新建数据中心所需资源，也响应了公司 2030 年“水资

源正值”目标，消除对蒸发冷却的依赖。同时，公司应用生命周期评估（LCA）工具对冷却方案进行碳排与水耗评估，并通过开放平台（如 Open Compute Project）公开分享研究成果，推动产业协同绿色转型。

在中国，随着数字经济的快速发展，我们也看到作为 AI 产业链核心底座的算力中心也在不断打造绿色低碳的先进标杆。作为中国三大移动运营公司的中国移动在内蒙古呼和浩特建设的数据中心园区已经成为中国“东数西算”算力基础设施的重要节点之一。能够为用户提供异构算力、分布式存储、网络等 IaaS 基础设施服务，以及基于第三方平台赋能的 AI 开发框架和 AI 使能的 PaaS 服务，并构建了通用大模型和行业大模型的 MaaS 服务。在此园区中就有目前全球运营商最大单体液冷智算中心，部署约 2 万张 AI 加速卡，拥有 700 多个机架、2300 多台智算服务器，智能算力规模高达 6.7 EFLOPS（每秒 670 亿亿次浮点运算）。通过采用间接蒸发、磁悬浮相变、氟泵冷却、浸没式及冷板式液冷等多项节能技术、以及小母线技术和智慧化能耗管理平台等手段，实现数据中心电能使用效率（PUE）达到国内领先的 1.15[12]。在冬季，园区还充分利用余热回收技术，变数据中心运行产生的“废热”为“优能”，满足冬季采暖需求，降低冷却水系统能耗。

将数据中心从高能耗设施转变为环境与经济的机遇，需将其视为能源系统的重要组成部分，既算力与热力，算力与电力的协同。如果布局前瞻合理、且能与发电侧和用户侧高效衔接的地点，数据中心便可通过余热回收实现低排放供热的重要资产，同时也可成为提升电力系统灵活性的关键资源，不仅可降低自身用能成本、提升产业竞争力，还能强化能源安全。

三、从能源消费者到低碳热力供给方（再利用 Re-use）

我们必须以全新视角审视数据中心的定位，这一点至关重要。将数据中心定位为能源系统的组成部分而非单纯能源消费者，可将其从挑战转化为机遇。

数据中心全年无间断运行，IT 设备因此会产生大量的热量。为保持服务器在合理温度范围稳定工作，数据中心的冷却系统至关重要，其运行产生的热量需经冷却介质（空气或液体）传递后排出。在传统配置中，这些热量通过冷却塔等设施直接排放到大气中，既浪费已支付成本的能源，又导致这些热源未能充分发挥其潜力。与此相反，可将载有服务器余热的冷却介质输送至热交换器，将热量转移至连接热能用户的供热二次循环回路再利用。在此环节中，会通过热泵给数据中心冷却介质中的相对低品位余热升温，以满足供热管网的温度要求。该流程对采用液冷的数据中心余热利用更为高效。

目前数据中心余热已可实现不同规模的再利用，如为数据中心周边的农业温室、养鱼场、游泳池等设施供暖，在更大规模应用中，余热可接入区域供热管网为城市供暖，增加供能的多元韧性，并减少碳排放。预计到 2030 年，全球数据中心耗电量将翻倍，其余热利用潜力极为可观。目前这类热量大多直接排放至大气中，造成能源浪费；若能对这种免费热源加以利用，可为提供经济且低排放的供热创造绝佳的机会。国际能源署估计，全球范围内靠近区域供热管网的数据中心可有效回收的热能已达 35 ~ 85 太瓦时，到 2030 年欧洲空间供暖需求的 10%可由邻近区域供热管网的数据中心余热回收利用满足，且成本远低于天然气[2]。

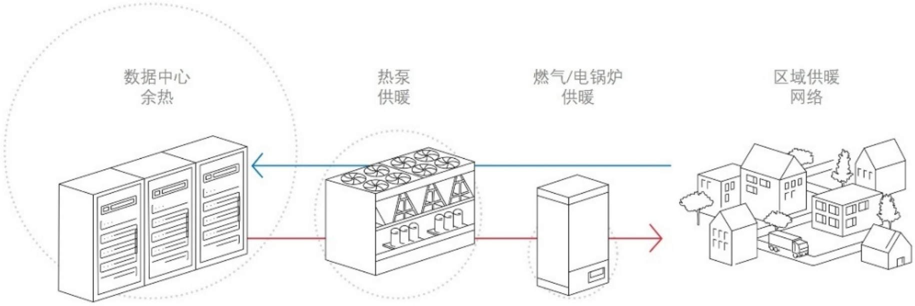

图 2：区域供暖的多元韧性

在欧洲低碳城市建设实践中，芬兰首都圈的数据中心余热利用项目已成为全球代表性案例之一。该项目由微软（Microsoft）与北欧能源企业富腾（Fortum）合作建设，通过回收微软当地两座大型数据中心运行冷却水系统中的余热，再由富腾公司建设的大型热泵站进行温度提升，并输送到城市区域供热管网中，为首都赫尔辛基周边埃斯波市（Espoo）和基尔科努米市（Kirkkonummi）地区居民住宅、公共建筑和商业设施提供清洁供暖。整个系统结合大型热泵、蓄热设施和电气化能源体系，是现代低碳区域供热的重要应用实践。该项目目前被普遍认为是全球规模最大的城市级数据中心余热利用项目之一。

从规模上看，该系统建成后预计可满足当地约 40%的区域供热需求，并有望每年减少约 40 万吨二氧化碳排放 [13]。这意味着数据中心不仅是数字经济基础设施，同时也成为城市能源系统的重要组成部分。与传统依赖天然气或化石燃料的供热方式相比，该模式显著提高了能源利用效率，同时降低了城市碳排放和能源成本。因此，该模式为全球城市提供了新的发展思路—将数据中心与城市能源系统协同规划，使算力基础设施转化为稳定的低碳热源。

图 3：基尔科努米市（Kirkkonummi）数据中心热泵站（2024 年 10 月）

在中国，位于天津武清区的清树科技园（总建筑面积 10.3 万平方米，为混合功能综合体）毗邻高村科技创新园数据中心集聚区，直线距离 400 米。该数据中心集群规划建设 9 座数据中心，运营服务器机架规模预计超过 10 万台。丹佛斯作为深耕数据中心制冷和区域供热的企业，积极参与此区域绿色数字基础设施的协同示范的建设试点。园区通过一套双模式磁悬浮热泵系统，实现对数据中心余热的规模化回收与综合利用。系统根据季节进行模式切换：冬季从数据中心循环冷却水中提取余热，为清树科技园的建筑综合体提供采暖及生活热水；夏季则转换为利用屋顶冷却塔为园区供冷。

即便在园区运营的首个供暖季入驻率尚未达到满负荷的情况下，采用数据中心余热的平均供热成本已较当地商业供暖成本格降低 60%（相较传统独立空气源热泵系统亦可节约 30%～40%）。若园区实现设计入住率运营后，该项目预计每年可为邻近数据中心节电约 110 万千瓦时，同时也降低了数据中

心传统通过冷却塔制冷对于水资源的消耗。全年有望实现二氧化碳减排约 1,659 吨[14]。因数据中心余热利用所增加的初始投资，预期投资回收期不超过 3 年。

经测算高村科技创新园数据中心集群按规划完全建成后，可回收的数据中心余热超 230 MW，理论上能满足四五百万平方米建筑的供热。数据中心这种低碳的余热热源经过热泵提温后，完全可以作为周边现有区域热力站的替代热源，减少其传统用于供热的燃气消耗以及相应的碳排放，并降低其运营成本。距数据中心集群最近的首创国际半岛区域，拥有超过 300 万平方米的居民住宅、学校以及商业设施等，我们了解到当地政府正积极会同数据中心所在开发区以及当地建筑规划设计咨询机构，进行数据中心余热资源再利用的盘点，以进一步复制清树科技园数据中心余热再利用的落地实践。

四、算—能灵活协同、增强电网韧性（资源化 Re-source）

数据中心的加速部署对电力的供给和电网稳定带来挑战。现行政策与规划框架将数据中心视为孤立的电力用户，而非能源系统中的积极参与者。数据中心参与需求响应或灵活性市场，与能源系统整合的技术潜力正在逐步得到验证，但仍未被系统性开发。这种脱节现象限制了数据中心在可再生能源并网、脱碳和电网稳定性方面所能发挥的作用。如何将数据中心统筹纳入更广泛的能源系统规划之中，将对电网、环境和经济产生将巨大影响。

（一）区域电力规划和能源供应的新挑战

如前文所述，根据根据国际能源署（IEA）的最新研究，随着全球数据算力基础设施快速扩张，2030 年全球数据中心的电力需求在全球总用电量中占比预计将翻倍达到 3%。当下在具体区域层面，数据中心的累积负荷甚至对局部电网产生了显著影响。如中国上海已投运 160 多家互联网数据中

心，2025 年用电量超过全社会用电量的 3%，这一比例显著高于全国平均水平 1.7%[15]。部分“东数西算”节点城市和区域的数据中心用电占比更超过 20%。在美国北弗吉尼亚，这个被称作“数据中心走廊”的区域拥有着 200 多个数据中心，是全球最密集、规模最大的商业数据中心集群之一，其耗电与波士顿这一百万人口级城市的峰值用电接近。行业与研究报告也多次指出该区域数据中心电力负荷占该区域电网用电份额约 20%。这一数据中心集群曾有约 60 个数据中心，因由行业内普遍采用的一项标准安全机制引发而几乎同时切换至备用电源，导致区域内约 1,500MW 的负荷瞬间脱离主网[16]。该事件对电网稳定带来风险，迫使电网运营商迅速调整发电输出以维持供需平衡。这一事件说明，在高密度数据中心集群背景下，电网运行弹性与应急调度面临新的挑战。

（二）算力与可再生能源的高效协同

随着人工智能浪潮席卷全球，数据中心及 AI 产业浪潮最终要落实到能源成本。算力的尽头是电力，也逐渐成为行业的共识。在基础物理层的电力供应上，作为全球数据中心发展主要区域的中国和美国，其电力市场却呈现完全不同的格局。在美国，人工智能与数据中心的爆发式增长将电网推向极限，由于电网承载能力不足，接入电网排队过长将延缓数据中心的部署和建设。目前新建数据中心的并网排队时间动辄长达一年以上：美国平均排队时间为 1 至 3 年，部分州甚至长达 7 年[2]。电网容量已严重饱和，电力市场的供需矛盾达到临界点之际，“缺电”危机也频现，在数据中心聚焦的区域电价也屡创纪录，甚至触及监管价格上限。因此，美国政府近期也表示，建设这些高耗能数据中心基础设施的大型科技公司有义务为自己的电力需求提供保障，并承担由此带来的电力成本的增长，而非将成本转嫁给普通消费者。

而在中国，2025 年，全社会用电量首次突破 10 万亿千瓦时。这一数字相当于美国全年用电量的两倍多，超过欧盟、俄罗斯、印度、日本四个经济体的年用电量总和。截至 2025 年底，中国总发电装机容量约 3,890 GW（千兆瓦），年同比增长约 16%。其中新增风电、光伏装机容量占比超 80%，达到创纪录的 430GW。风光合计累计装机容量也达到全国累计发电装机容量的 47%[17]。中国快速扩张的电力供给能力，尤其是可再生能源装机和发电量的迅猛增长，让中国的电价可以保持在相对低位。在风光装机大幅增长的背景下，数据中心应把“算力作为可调资源”与可再生出力深度耦合，从而提高可再生能源的消纳并降低碳排放。

中国“西数东算”战略工程正是结合东西部地区可再生能源资源禀赋的不同，推动算力向西部可再生能源富集区布局，实现算力扩张与低碳发展的平衡。目前在中国青海也正在建设绿电与算力协同的示范项目—柴达木绿色微电网算力中心。该项目是面向荒漠化土地与光伏资源、同步建设青海移动算力中心与配套清洁能源供电系统，包括光伏发电、智能微电网、储能等配套系统。算力中心规划部署高功率机架约 1000 个，采用冷板式液冷、智能母线、间接蒸发冷却等节能技术；配套光伏装机 122 MW、储能 75 MW/300 MWh，并构建微电网与能量管理系统，实现白天光伏发电供电并为储能充电，夜间由储能出力供算力中心。预期建成后年可向算力中心供应清洁电力约 1.26 亿千瓦时，预计每年可节约标煤约 4.21 万吨，减少 CO_2 排放约 11.55 万吨[18]。该项目作为中国首个规模化清洁能源微电网算力中心，同步探索离网/微电网在偏远算力节点的经济性与运行机制，为西部新能源富集区解决绿电就地消纳和算力承载提供了实践样本。

（三）数据中心作为负荷侧的灵活协同潜力

正如达木绿色微电网算力中心项目所展现的，中国西北地区风光资源富集，但也面临外送通道受限与局部弃风弃光压力，算力中心作为高负荷、可扩展的新型用电主体，通过微电网实现绿电就地消纳，把“波动的电力”转化为“可交易的算力”，提升新能源利用效率，减少弃电。与传统工业负荷不同，数据中心负荷具有时间和空间的灵活性，是一种可以不依赖电网而实现电力负荷瞬时转移的，可作为电网“需求响应”资源的新型柔性负荷。电力系统用以平衡供需、提高弹性的需求响应（Demand Response, DR）机制是通过价格信号或合同激励，鼓励电力用户在电网紧张或可再生出力高峰时段有序改变用电行为（比如削峰、填谷、短时降载或延迟可调负荷），从而减少对昂贵或高碳调峰机组的依赖、提升可再生能源消纳能力并降低系统运行成本。数据中心除了算力任务负荷的可调节潜力之外，内部也存在大量的电力灵活性资源可参与新型电力系统的灵活性调节，包括不间断电源（UPS）、备用的柴油发电机、电池储能以及空调制冷、蓄冷等热力灵活性资源等。

数据中心提供电网灵活性协同的方式包括：

- 跨区域任务调度：将电网紧张区域的数据中心所承担的算力任务转移至可再生能源电力充足的区域的数据中心执行。
- 时间维度上的任务调度：数据中心部分算力任务（如训练、批处理、非实时计算）具有时间可调性。通过负荷管理与储能协同，算力中心可在新能源出力高发时段提升负载，在低谷时段降低用电，发挥“电力系统调节器”作用，增强电网稳定性。
- 本地发电设施：可用作数据中心的备用电源或基本负荷电源。发电量可以根据电网供应情况进行灵活调节，从而降低数据中心对电力生产波动的敏感性。如果通过氢燃料电池，绿氢更可提供零噪音、高可用

性的备用电力。同样，高效离网式燃气发电机也可作为摆脱石油和煤炭发电厂的可行过渡途径。

- 电池储能系统（BESS）：是管理大型数据中心电力需求波动、且避免电网失稳的必要手段。在可再生能源发电过剩时段，电池可以用于储存来自本地或电网侧的绿电。通过这种方式，数据中心就可以在电力供应紧张或者价格高的时段使用可再生能源电力。
- 液冷系统的灵活性：液冷技术利用水冷的方式为 IT 设备降温，冷却过程中水温升高，需再次冷却后才能循环使用。可以在电价低廉时举行蓄冷，这种冷能可实现数据中心制冷及时响应的灵活调节，有时甚至能跨季节储存。

一项针对三个美国产电区（中大西洋 Mid-Atlantic、德州 Texas、以及西部互联 WECC）模拟不同层级的数据中心灵活性方案的研究表明，此举最多可减少 40%的排放，加速淘汰化石能源，并将整个电力系统成本降低 5%。即使在灵活性措施未直接带来脱碳效果的情况下，其仍能降低系统总成本。因为这使电网得以优先调用成本最低的电力资源并减少对新建输配电设施的需求，最终降低所有用户的能源支出[19]。Google 已实践在其数据中心部署的一种按电力碳强度动态优化算力负载时间与地点的调度机制—碳感知调度系统。通过实时识别电网碳强度结合 AI 分析对可延迟的计算任务进行时间与空间迁移，使数据中心运行与低碳电力供给同步，这是实现 Google 2030 年 24/7 零碳电力目标的关键技术手段，也验证了数据中心负荷灵活性在高可再生电网环境下的显著减排潜力[20]。

中国上海以超大城市算力需求快速增长为背景，也正把数据中心从高耗能主体转为电网的灵活调节资源（2025 年上海数据中心用电占全市用电约 3%，远高于全国 1.7%的平均水平）。国家电网与中国电信联合开展了跨省算

力迁移实测（“上海—十堰”、“上海—福建”），实现百卡集群、分钟级、多任务的跨省调度；“上海—福建”实验将响应规模扩大到 50 kW，标志着技术由验证走向可商用。上海虚拟电厂聚合规模达 2,568 MW，其中数据中心聚合 182 MW（占比 7.1%）；夏季用电高峰期间数据中心最大响应能力达 77 MW，较 2024 年增长超 5 倍[21]。有学术研究认为中国数据中心 2030 年的灵活性资源约在 2.3 ~ 4.0 万兆瓦之间，数据中心算力—电力协同将成为信息和能源协同发展的重要场景[6]。

五、政策建议

数据中心正成为数字经济与人工智能发展的核心基础设施，其能源与水资源消耗正在引起广泛的关注。亟需考虑如何将数据中心从高耗能主体，转为兼具约束与赋能能源系统的关键基础设施，以缓解区域性能源与水资源压力，并把算力发展转化为低碳机遇。以更多维度的顶层设计推动‘算力—电力—热力’的灵活协同，既为产业增长提供确定性保障，也通过制度与市场机制激励数据中心参与低碳、弹性的能源体系转型。共同塑造数据中心绿色低碳的未来发展路径。

（一）深化能效技术革新应用，提升系统能源利用效率

聚焦数据中心基础设施全生命周期能效升级，针对制冷、供电两大核心能耗板块，因地制宜推进技术革新与规模化应用，优化系统整体架构，持续降低数据中心 PUE、WUE（水资源利用效率）等核心能效指标，构建全系统高效节能的绿色算力载体。全域推广先进制冷节能技术，结合不同区域气候特征、数据中心功率密度及业务场景属性，建立分场景、差异化的先进制冷技术推广体系；推动制冷系统与 IT 负载智能联动，基于算力负载动态变化精

准调整制冷运行策略，实现“按需供冷、精准匹配”，持续压缩制冷系统能耗占比。

同步加快高效供电系统技术迭代升级，优化数据中心供电架构设计，精简供电转换环节，降低电力传输与转换过程中的电能损耗；提升供电系统转换效率与动态响应能力，在保障供电安全可靠的前提下，合理降低配套设备冗余度，优化建设与运营成本。

强化数据中心能效准入与运行全流程监管，完善统一的能耗计量标准与信息披露机制，逐步构建行业可对标、可监督的能效监管体系。通过政策引导与市场机制协同，全面提升行业整体能源利用效率，防范粗放式发展对区域电力、水资源系统造成的结构性压力，为算力低碳发展筑牢能效基础。

（二）余热纳入城市能源布局，促进多方共赢的再利用模式

数据中心不仅可以承载算力需求，其在运行过程中产生的大量低品位热能，还可以与区域供热系统结合，形成“算力+热力”的综合基础设施模式，在降低碳排放的同时提升城市能源效率与可持续发展水平。目前数据中心余热利用技术均已具备，但实际余热利用落地规模还比较有限，主要以部分企业单个项目试点为主。要真正实现广泛的数据中心余热利用，还有待系统统筹与规划以及各相关方的有效协同。对于城市区域供热系统利用数据中心这种低品位热能，也可以借鉴欧洲“第四代区域供热系统”的发展经验。其强调低温供热网络、分布式热源接入与高比例可再生能源整合，其核心理念是通过降低供热温度与提升系统灵活性，实现多源热能的高效接入。

政策层面应在城市能源规划和园区建设方面，加强数据中心与区域供热系统、工业用热系统之间的统筹布局，把余热利用纳入城市基础设施整体规划之中。对于具备条件的项目，应在规划阶段开展余热回收可行性论证，推动形成标准化接口和技术规范，降低接入成本与协作门槛。同时，可通过市

场化机制与适度政策支持，鼓励热力企业与数据中心运营方开展长期合作，支持示范项目验证低温管网、热泵提升与多热源调度的技术可行性，探索多元化的热能利用模式，降低一次能源消耗。类似天津武清区的清树科技园数据中心余热利用的模式，可以进一步落地推广。在稳妥推进的前提下，逐步提升数据中心在城市低碳供热体系中的功能角色，使其从单一用能主体转变为可参与城市能源循环利用的重要资源。

（三）推动算力—电力的系统耦合与灵活性协同

在中国可再生能源装机规模持续提升、新型电力系统加快构建的背景下，数据中心作为高耗能关键基础设施，其负荷的可调节特性为提升电力系统灵活性、促进新能源消纳提供了重要潜在空间，更是实现“算力低碳化、电力高效化”双向赋能的核心纽带。当前，中国正在构建一体化算力网，建立跨区域、多层次算力调度体系，这为也推动算力的时空调度与电力系统深度耦合、灵活协同的良性循环奠定坚实基础。

在新能源资源富集地区，应鼓励数据中心与本地可再生能源协同布局，规划建设集约化、高效化的绿色数据中心集群，通过市场化方式提升绿电消纳水平。推广“源网储荷”一体化模式，支持数据中心配套建设新型储能设施，实现绿电生产、存储、消纳全链条闭环运营。同时，完善绿色电力交易机制，简化流程、降低交易成本，扩大数据中心绿电使用比例。为落实中国“东数西算”工程对国家枢纽节点新建数据中心绿电占比超 80%的刚性约束，提供有效实施路径。

引导数据中心参与电网需求响应，鼓励具备负荷调节能力的数据中心参与电网削峰填谷、应急备用等辅助服务，推动多个数据中心负荷聚合联动，提升需求响应规模化能力。在保障数据中心核心业务稳定性的前提下，逐步探索与电力系统运行节奏相适应的负荷优化模式，进一步提升电力系统灵活

性。同时考虑通过价格杠杆引导数据中心主动调整用电行为，推动实时电力价格与碳强度信息的公开，促进其与算力调度系统深度对接，完善需求响应和灵活性市场的接入规则，为数据中心开展更精细化的能耗与碳排放管理创造条件。

（四）总结

总体而言，数据中心作为新型基础设施，其发展既承载数字经济增长需求，也直接关联能源结构转型与城市可持续发展。未来政策应以能效提升为底线，以余热再利用为延伸，以电网灵活性协同为方向，在保障数据中心产业绿色低碳发展的同时，增强能源系统的整体韧性与低碳属性。

主要参考文献：

本报告以《丹佛斯影响力系列发展报告第八册—数据中心》为框架蓝本撰写

1. De Vries, A.（2023）《人工智能日益增长的能源足迹》《Joule》期刊，第 7 卷第 10 期
2. 国际能源署《能源与人工智能》（2025）
3. Uptime Institute（国际数据中心标准与认证机构）年度调研
4. 中国工业和信息化部
5. 中国信通院《绿色算力发展研究报告》（2025）
6. 《中国数据中心综合能耗及其灵活性预测》（北京理工大学学报 2025 年 3 月）
7. 国际能源署《能效报告 2025》
8. 美国 Energy Star 指南《降低配电单元（PDU）电能损耗》
9. 劳伦斯伯克利国家实验室（LBNL）《电力与照明系统（Power & Lighting）》
10. 英伟达（NVIDIA）产品数据分析
11. 华为（HUAWEI）产品数据分析
12. 中国移动公司
13. William Holley，《向低碳供热转型》，2025 年 7 月
14. 人民网《用低品位余热“温暖”城市》
15. 澎湃研究所
16. Datacenterdynamics.com
17. 国家能源局、中电联统计
18. 中国能源网
19. 美国国家经济研究局：《灵活型数据中心与电网：更低成本，还是更高排放？》
20. Google Blog：《我们正在将更多计算任务部署到清洁能源更充足的地区》
21. 澎湃新闻网

Green and Low-Carbon Data Center
A new lever of flexibility, integration, and security

Paper Presented by Danfoss [1]

Introduction

Digitalization is advancing at an extraordinary pace, and the rapid adoption of artificial intelligence (AI) is reshaping every part of society. The core infrastructure of this transformation is data center, providing the digital backbone that enables modern society to progress.

The demand for data center's computing power, storage capacity, and the resulting energy need is accelerating far beyond historical rate, and in many regions, it has outpaced infrastructure developments. Careful planning and integration of data centers into our energy system is more urgent than ever and presents a critical opportunity to build a smarter, more resilient energy system. Data centers are a prerequisite to future technological advancements and for developing the solutions needed to decarbonize our economies. They hold the power to significantly improve energy efficiency in virtually all sectors, from transport to heavy industry. According to the IEA, AI has the potential to deliver emission reductions equivalent to around 5% of global energy-related emissions in 2035. As data centers become ever more embedded in our daily lives, and our future, it's clear that their expansion will play a key role in shaping the next generation energy system.

We know that data centers consume significant resources as they carry out the essential work of delivering the digitalization that we're reliant on. Data centers will consume as much as 3% of global electricity by 2030. Regionally, the impact can be even more concentrated and harmful to the electricity grid. 80% of electricity demand growth will be in China and the US. In the EU, demand is expected to triple.

[1] The views expressed in this report are those of the enterprise research and do not represent the official stance or opinions of the forum host and organiser.

Meanwhile, data centers use around 560 billion liters of water per year, potentially increasing to 1,200 billion by 2030. Their impact on the energy system is already significant and will intensify further unless we redefine the way we handle their rapid expansion.

Seeing data centers solely as resource consumers only tells part of the story. In reality, data centers represent one of the greatest untapped opportunities to strengthen global energy systems. Embracing "Competitive decarbonization" turns the green transition from a purely environmental responsibility into an economic opportunity - a principle increasingly relevant to data centers. Through improved resource efficiency, flexible operations, waste heat recovery and reuse, proper integration into local energy ecosystems, alignment with renewable energy generation, and supporting grid flexibility, we can turn these high energy-consuming facilities into great assets. When managed effectively, data centers can contribute to a resilient and efficient energy system. Therefore, policymakers can help unlock the potential and address the challenges of rapid expansion by adopting a holistic approach. Driving efficiency across both facility design and operations can significantly lower the projected resource demand and ensure that digital application growth does not outpace energy system capacity. A cross-sector, collaborative approach can further the development and deployment of new technologies and next-level efficiency.

It's clear and resolute on the future we should pursue and build together. Digitalization is accelerating across our economy and society, and its inevitability can bring significant opportunities for innovation and efficiency. Now, we must act with both urgency and vision. We must make resource efficiency a non-negotiable foundation of data center operations and plan for data centers to become key partners in the flexible, efficient, and resilient energy system of the future.

Kim Fausing

President & CEO, Danfoss

Abstract

In the current era of deep integration between artificial intelligence and the digital economy, computing infrastructure has become a foundational pillar supporting global industrial innovation and societal digital transformation. As the core carriers of computing power, data centers are evolving from traditional facilities focused on computation and data storage for legacy IT systems into new infrastructure that delivers both digital enablement and green transformation.

The expansion of digital technologies and AI applications is driving exponential growth in computing demand, while simultaneously raising global concerns regarding data centers' energy and water consumption, as well as carbon emissions. Balancing the stable supply of computing power with the mitigation of environmental and resource pressures has therefore become a critical challenge shared by industries worldwide.

Drawing on specialized research from authoritative organizations such as the International Energy Agency (IEA), as well as global best-practice case studies, this report explores pathways and practical strategies for the green and low-carbon transformation of data centers in the computing era, helping the industry achieve a sustainable balance between scalable computing capacity and low-carbon development goals.

The large-scale deployment of intelligent computing power underpinning artificial intelligence is the primary driver of rapidly increasing energy consumption in data centers. The energy demand of AI large models is far higher than that of traditional digital applications. With the expansion of multimodal AI scenarios, the pace of computing demand growth now far outstrips the evolution cycle of conventional semiconductor technologies.

According to a 2025 report by the International Energy Agency (IEA), the share of global data center electricity consumption in total final energy use is projected to double from 1.5% in 2024 to nearly 3% by 2030. This growth is accompanied by

corresponding increases in water consumption and carbon emissions, posing significant challenges to industry.

In response to the tension between surging computing demand and resource constraints, the United States, China, and the European Union, which collectively account for 85% of the global data center market, have implemented distinct policy frameworks to guide and incentivize sustainable development.

Industry experience further demonstrates that data centers are not merely energy-consuming facilities. They serve as the core computing foundation for high-quality digital economic development and are an increasingly important participant in the low-carbon transformation of the energy system. As low-carbon computing becomes more widely deployed, data centers will play a critical role in enabling the coordinated digitalization and decarbonization of traditional industries.

As critical infrastructure for a modern industrial system, data centers can achieve substantial improvements in operational efficiency by establishing a full-chain low-carbon technology system and deeply integrating with the energy system. Moreover, they can be transformed into high-quality flexibility resources for emerging power systems.

First, Reduce – Minimizing Energy at the Source

The core of reduction lies in improving the energy efficiency of computing equipment. Generational upgrades of AI chips provide the foundational support for decoupling the growth of computing capacity from the rise in energy consumption. In parallel, the large-scale deployment of advanced cooling technologies, such as liquid cooling, offers mature solutions for enhancing energy efficiency in high-density intelligent computing centers.

Second, Reuse – Recovery and Utilization of Waste Heat

Data centers generate substantial amounts of low-grade waste heat during operation. By employing heat pump technologies to upgrade this heat and integrating it into district heating systems, energy can be utilized in cascaded stages. Multiple projects in Nordic countries and China have demonstrated the dual economic and environmental benefits of this approach.

Third, Resource – Leveraging Energy and Carbon Value

Value reconstruction can be achieved through deep integration of data centers with renewable energy systems. By deploying an integrated layout of generation-grid-storage-load, and through market-based renewable electricity transactions, data centers can enhance local renewable energy absorption. Their temporal and spatial load flexibility, combined with cross-regional computing dispatch and demand response participation, enables data centers to evolve from mere energy consumers into low-carbon heat providers and critical enablers of renewable energy integration and new power system development.

Looking ahead, the green and low-carbon development of data centers requires not only continuous technological innovation to secure a baseline of energy efficiency, but also a systems-thinking approach to integrate and coordinate computing, thermal, and electrical resources. This holistic approach ensures that the growth of the digital economy is compatible with low-carbon transformation, positioning data centers as a key driver of societal green transition.

1. Breaking the energy gridlock of data centers

Over the past two decades, the global data center industry has experienced rapid growth. Driven by continuous technological innovation and integration, data centers have evolved from traditional computing centers and information centers to cloud centers and are increasingly transforming into computing power hubs. The world is quickly embracing new digital technologies and AI. Their use is becoming part of everyday life: industries test new products virtually, various mobile social and streaming applications, accurate weather forecasts we are witnessing major breakthroughs in the treatment of serious diseases. It’s already clear that Digital information technologies and artificial intelligence are changing how we live and the appetite for it will only continue to grow. Now the important challenge is how to support this growth while simultaneously managing the demands data centers place on critical resources. Governments and other decision-makers are responsible for enabling this growth while also managing data centers' demand for resources such as energy and water. Fundamental to preventing energy gridlock is energy efficiency of computing infrastructure and integration data centers into our energy systems to enable coordination between computing power and the energy system.

1.1 The Computing Era Is Driving A Surge In Resource Demand

From a global perspective, data center compute and storage capacity have become the "new oil" of the digital economy. In recent years, the accelerated adoption of artificial intelligence (AI) technologies has made AI servers a rapidly emerging core driver of energy consumption growth in data center computing infrastructure. Data shows that the energy consumption per request for large AI models, as well as AI-enabled search engines, is significantly higher than that of traditional search engines. For example, a standard Google search consumes approximately 0.3 watt-hours per request, while a single request processed by the ChatGPT large model consumes about 2.9 watt-hours.[1] As large AI models evolve from text generation to more computationally intensive applications - such as text-to-image and text-to-video - the demand for computing and storage capacity is growing exponentially. Consequently, the associated energy demand from data centers is also showing a marked upward trend.

According to the International Energy Agency (IEA) in its 2025 special report examining the impact of AI development on global data center electricity demand,[2] projections under the Base Case scenario indicate that the share of data centers in global final electricity consumption will double by 2030. This share is expected to rise from 1.5% in 2024 - equivalent to 415 terawatt-hours (TWh), roughly comparable to the United Kingdom's total annual electricity consumption - to nearly 3% by 2030, reaching approximately 945 TWh, close to Japan's current national electricity demand. Since 2017, global data center electricity consumption has grown at an average annual rate of around 12%. In advanced economies, the increase in electricity use by data centers is projected to account for more than 20% of overall electricity demand growth.

In 2024, the United States accounted for the largest share of global data center electricity consumption, reaching 180 terawatt-hours (TWh) and representing approximately 45% of the global total. China and Europe followed, accounting for about 25% and 15% of global data center electricity use, respectively. The U.S., China, and Europe are expected to remain the three largest regions in terms of data center electricity consumption through 2030.

China and U.S. together are projected to contribute nearly 80% of the global increase in data center electricity demand. By 2030, electricity consumption by U.S. data centers is expected to increase by approximately 240 TWh compared with 2024 levels (a 130% increase), while China's data center electricity use is projected to rise by about 175 TWh (a 170% increase). Europe is expected to see an increase of more than 45 TWh (around 70%).

Electricity consumption from AI-driven servers is projected to grow at an average annual rate of about 30%, contributing nearly half of the net increase in data center electricity demand. In contrast, traditional servers are expected to grow at an average annual rate of only 9%, accounting for roughly 20% of the increase. Supporting infrastructure such as cooling systems and other IT equipment are projected to contribute approximately 20% and 10% of the additional electricity demand, respectively.

As AI technologies continue to advance, the pace of energy consumption growth has increased markedly, exceeding the two-year doubling cycle described by Moore's Law in the semiconductor industry. According to research from Stanford University and estimates based on analyses by the International Energy Agency (IEA), training the GPT-4 model required a computing cluster of up to 25,000 GPUs operating continuously for approximately 14 weeks. The process consumed more than 22 MW of power and a total of 42.4 gigawatt-hours (GWh) of electricity - around 30 to 40 times the energy required to train the GPT-3 model. This corresponds to an average daily electricity consumption of about 0.43 GWh, roughly equivalent to the total daily household electricity use of 28,500 homes in advanced economies.

Although China's large language model DeepSeek has improved algorithmic efficiency, this has not curbed the demand for computing power. On the contrary, the expansion of users and application scenarios has accelerated the adoption and deployment of large models. Over the long term, energy consumption from computing infrastructure is therefore expected to continue rising strongly.

At the same time, many data center facilities rely heavily on water-based cooling systems for temperature control, making water usage another critical consideration. Global data centers currently consume approximately 560 billion liters of water annually, a figure that could increase to 1.2 trillion liters by 2030 - equivalent to six times the European Union's total freshwater withdrawals in 2022. If current trends persist, global data center carbon emissions could rise from 180 million tons in 2024 to 300 million tons by 2035, posing additional challenges to global efforts aimed at reducing emissions and addressing climate change.

1.2 Policy Guidance and Enablers for Sustainable Data Center Development

With the accelerated adoption of artificial intelligence and the growing demand for data centers, there is a clear global trend toward larger-scale facilities, alongside an intensified push for green and low-carbon development. By the end of 2024, the average Power Usage Effectiveness (PUE) - a metric measuring data center energy efficiency - of global data centers was 1.56 [3], showing a modest improvement compared to the average PUE of around 1.6 in 2017. Governments worldwide face a dual challenge: ensuring that electricity and water resources at data center locations are stable, reliable, and reasonably priced, while promoting data center construction to meet the economies and society's increasing digital and intelligent development needs. At the policy and industry regulation level, particularly regarding green and low-carbon development, countries have adopted different approaches.

The United States has long maintained a dominant position in the global computing power market, leveraging its leadership in semiconductor technology and algorithm research. Over the past decade, the federal government promoted data center consolidation, IT industry energy efficiency improvements, and the development of high-performance green computing infrastructure through a series of regulations. However, since 2023, few new policies or updated targets have been introduced. Meanwhile, companies in the data center sector have remained active in improving energy efficiency and advancing green and low-carbon initiatives. Major technology companies such as Amazon, Google, and Microsoft lead the world in renewable energy procurement agreements, and each have committed to corporate and data center carbon-neutrality pathways. U.S. academia has also contributed to server energy-saving technologies and optimal energy utilization. Overall, the green development of the U.S. data center industry relies more on market mechanisms, corporate responsibility, and academic research, rather than federal policy leadership.

In the European Union, data center policies have been evolving from voluntary guidance to mandatory regulation and from localized optimization to systemic governance. In 2020, the EU led the way by setting decarbonization targets for the information and communication technology (ICT) sector, including data centers, through initiatives such as the Best Practice Guidelines for the EU Code of Conduct on Data Centre Energy Efficiency and the industry-led Climate Neutral Data Centre Pact (CNDCC). These documents define energy-saving standards and implementation measures, establishing mechanisms to promote green data centers.

The 2023 revised EU Energy Efficiency Directive (EED) marked a transition to hard constraints, integrating mandatory legislation, technical standards, and regulatory mechanisms. Together with other directives on renewable energy, resource circularity, and mandatory disclosures, these policies aim to achieve climate neutrality for EU ICT infrastructure and computing centers by 2030. In May 2023, the EU's Carbon Border Adjustment Mechanism (CBAM) extended regulatory coverage to include IT equipment imports, cross-border cloud services, and data processing services for data centers. Under CBAM, non-EU companies importing computing services are required to pay carbon-related tariffs; in 2024 alone, this prompted 15 global cloud service providers to establish full lifecycle carbon footprint tracking systems.

In contrast to overseas practices that rely largely on incentives and voluntary measures, China's green and low-carbon development of data centers is guided by top-level national planning. The "East Data, West Computing" initiative, proposed in 2020 and implemented under China's Integrated Big Data Center Collaborative Innovation System, reflects this macro-level strategy. It addresses regional resource imbalances by strategically allocating and managing computing resources across the country. Large-scale data center facilities are deployed in western regions with abundant clean energy to achieve green carbon reduction and centralized management, providing computing services for the high demand in eastern regions. The resulting infrastructure, comprising eight major computing network hub nodes and ten national computing center clusters, promotes policy and resource centralization, guiding the development of data centers toward intensive, large-scale, and green low-carbon operations.

In recent years, China has continuously introduced national- and regional-level policies to promote high-quality development of data centers and computing power. These policies have progressively raised audit and access standards across multiple dimensions, including electricity and water use efficiency, renewable energy utilization, overall rack occupancy, energy efficiency per unit of computing power, and carbon efficiency.

By the end of 2024, the total number of standard racks in operation across China exceeded 9 million, with computing capacity growing at an average annual rate of nearly 30% over the past five years, reaching 280 EFLOPS (exaflops, FP32) [4], ranking second globally. Notably, intelligent computing accounted for 32% of total capacity. The national average Power Usage Effectiveness (PUE) of data centers declined to 1.46 [4]. China's national standard, <*Green Data Center Evaluation*>, officially came into effect on June 1, 2025. It aligns with both the characteristics of Chinese data center construction and advanced international benchmarks, defining multi-dimensional requirements for resource-efficient utilization, green design, and green operations. By early 2026, the number of national-level green computing facilities reached 306, with an average PUE of 1.25 [4] - 14.4% lower than the national average.

The overall Computational Efficiency (CE) of Chinese data centers reached 21.95 GFLOPS/W (FP32) [5], and overall Storage Efficiency (SE) reached 102.2 GB/W, representing a 15.68% [5] increase compared to 88.4 GB/W in 2022. Thanks to the integrated planning and innovative construction of China's data centers, the country can now evaluate energy efficiency across multiple dimensions, including computing efficiency and storage efficiency, providing a more comprehensive assessment of the overall performance and sustainability of its data center infrastructure.

1.3 Exceptional Data Centers: The Intersection of Efficiency and Opportunity

Computing facilities are a critical component of a modern industrial infrastructure and represent one of the fastest-growing sectors of energy consumption in China. According to statistics from the Ministry of Industry and Information Technology (MIIT), electricity consumption by China's computing facilities reached 166 billion kWh in 2024, accounting for approximately 1.7% of total national electricity usage. Since the 14^{th} Five-Year Plan, the annual growth rate of electricity consumption for national computing infrastructure has exceeded 10%, about 1.5 times the annual growth rate of total national electricity consumption.

With the rapid development of the digital economy and the evolution of data centers into comprehensive computing power centers, overall energy demand is expected to increase at an accelerated pace. Academic research projects that by 2030, China's data center computing power load could reach approximately 105 GW, with water consumption of around 262.9 billion liters and carbon emissions of roughly 310 million tons - surpassing the carbon emissions of data centers in the United States [6].

Given the growing and inevitable demand for electricity and water by data centers, it is essential to develop a full-chain low-carbon technology system for computing equipment and infrastructure, build highly efficient data centers, and fully integrate them into the energy system to promote green and low-carbon development. Operational efficiency can be enhanced through optimized cooling technologies, digital tools, and artificial intelligence (AI) itself, increasing overall transparency and productivity.

Data centers can also serve as flexible resources within emerging power systems. Their internal computing, electrical, and thermal flexibility, when managed to ensure stable and efficient operation, can be integrated into broader energy systems to enable computing-electricity and computing-heat synergy, turning data centers into a strategic opportunity. When properly planned, these measures can help phase out fossil fuels, reduce grid investments, and advance sustained energy savings and carbon reductions in the data center sector.

Moreover, low-carbon and green computing can be deeply integrated into industries such as manufacturing, transportation, and services, driving improvements in energy efficiency across traditional sectors while simultaneously achieving digital transformation and carbon reduction. This dual benefit enhances overall societal energy efficiency and carbon reduction effectiveness.

The International Energy Agency (IEA) has also highlighted that AI will become a cornerstone for building resilient energy systems. By 2035, AI alone could save approximately 2,200 TWh of energy in the industrial sector - roughly equivalent to the entire energy demand of the EU's industrial sector and far exceeding the projected total electricity consumption of global data centers at that time [7].

2. Enhancing Computing Infrastructure Efficiency to Achieve Sustainable Growth (Reduction)

Power Usage Effectiveness (PUE) is the most widely legislated and recognized metric for measuring data center energy efficiency. It is defined as the ratio of total energy consumption of a data center - including IT equipment, cooling, and power delivery systems - to the energy consumed by IT computing equipment alone. The lower the PUE value, the higher the overall energy efficiency of the data center.

For example, in 2024, China's data centers had an average PUE of 1.46. Of the total energy consumption, IT equipment - including servers, storage, and networking devices that form the core computing load - accounted for 69%. Based on authoritative sources such as IEA, ENERGY STAR, and LBNL, cooling systems contributed an average of approximately 17% [2], power distribution systems about 10% [8], and other auxiliary systems - such as lighting, fire protection, monitoring, and office facilities - around 4% [9].

For newly constructed, highly efficient, ultra-large data centers (PUE $\leq$ 1.25), the share of IT equipment energy consumption can exceed 80%, while cooling system consumption can be reduced to below 10%. Therefore, IT computing equipment remains the primary driver of data center energy use, and in the future, this will be increasingly concentrated in AI computing servers.

2.1 Leap in Computing Equipment Energy Efficiency

As previously noted, from 2019 to 2024, China's total computing capacity grew at an average annual rate of 30%, reaching 280 EFLOPS, while the electricity consumption of computing infrastructure increased by only 10% per year. The key factor behind this decoupling of energy consumption and computing growth is the substantial improvement in the energy efficiency of computing equipment itself—that is, the computing output per unit of energy consumed, which lies at the core of data center energy use.

AI acceleration chips, such as GPUs and NPUs, are the primary drivers of computing growth. For example, NVIDIA's A100 GPU, released in 2020, delivered 1.56 TFLOPS/W of FP16 performance per watt. By 2022, the H100 GPU increased FP16 performance per watt to 2.83 TFLOPS/W [10], achieving an energy efficiency improvement of over 80%. Similarly, Huawei's Ascend AI processors demonstrate significant gains: the Ascend 910, launched in 2019, delivered approximately 256 TFLOPS at FP16 with a power consumption of around 310 W. Subsequent iterations, the 910B and 910C, leveraged architectural and process optimizations to increase computing performance to roughly 320–640 TFLOPS while reducing power consumption to 250–310 W [11], resulting in a 1.5–3 times improvement in performance per watt compared with the original generation.

Thus, during the transition from general-purpose computing to intelligent computing, innovations in AI chips, storage, networking, and other computing equipment - including advanced process technologies and architectural optimizations - drive substantial leaps in computing capacity while simultaneously reducing energy consumption. This progress depends heavily on continued innovation in the semiconductor industry and follows its inherent evolutionary trajectory.

2.2 Efficiency Improvements and Carbon Reduction Potential of Computing Infrastructure

Once the IT computing equipment configuration and architecture of a data center are established, the energy consumption of this equipment is largely fixed, varying mainly with rack occupancy and business workload. At this stage, the greatest potential for improving data center energy efficiency lies in auxiliary systems - such as cooling and power delivery - that account for 20%–30% of total energy use (corresponding to PUE values of 1.25–1.46).

By aligning these systems with regional climate conditions and specific data center workloads, and by deploying advanced energy-saving and carbon-reduction technologies, operators can optimize key areas including thermal management, power distribution, system integration, and intelligent operations. This approach reduces energy and water consumption during operation and lowers carbon emissions. Key measures include maximizing the use of natural cooling sources, employing fluorine-pump phase-change cooling to minimize water use, and implementing high-efficiency thermal management such as liquid cooling for high-power-density racks.

Power delivery systems can also be optimized by minimizing conversion stages, reducing losses in electricity transmission and conversion, and adopting technologies such as high-voltage direct current (HVDC), modular UPS, and direct utility supply. In addition, digital solutions - such as real-time field data monitoring and variable-frequency drives responsive to load changes - combined with AI-driven operations and maintenance tools, enable precise, efficient thermal management and energy supply. Preventive maintenance further ensures the safe and reliable operation of the data center.

2.3 Advanced Cooling Systems to Unlock Computing Potential

As chip performance continues to increase, the heat generated within the same physical space rises sharply. Power density per rack in data centers is gradually increasing from the traditional general-purpose center range of 4–6 kW to 20–40 kW in current intelligent computing centers, with some scenarios expected to exceed 100 kW in the future. This rising heat density increasingly exceeds the capacity of conventional air-based cooling systems to dissipate heat effectively.

To meet the growing demand for high-density computing, an increasing number of data centers are adopting Direct-to-Chip (D2C) liquid cooling systems. In this closed-loop system, coolant is distributed via a secondary-side loop from a cooling distribution unit to the processor chips. At the server level, cold plates directly contact the processors, removing the heat generated during operation. The coolant then returns to the primary-side loop of the distribution unit, which connects to external cooling sources - such as chillers, cooling towers, or dry coolers - for final heat exchange.

Compared with traditional air-cooling methods, closed-loop direct-to-chip liquid cooling provides higher heat dissipation efficiency and significantly reduces water consumption, typically achieving at least 15% energy savings. In regions with constrained grid capacity or limited water availability, this energy efficiency measure becomes increasingly critical, enabling data centers to meet rising demand while reducing local resource usage and ensuring more reliable operation. Liquid cooling is increasingly becoming standard for intelligent computing centers, evolving from an “optional” to a “mandatory” solution. Additionally, liquid cooling facilitates the recovery and reuse of waste heat from data centers.

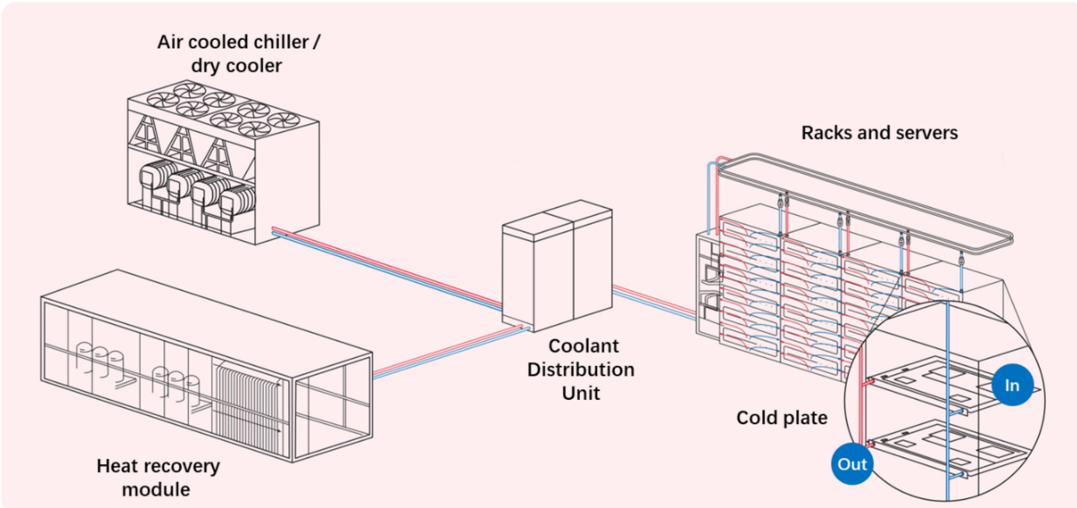

Figure 1: Structure of Direct-to-Chip (D2C) liquid cooling system

In recent years, global demand for artificial intelligence computing power has grown rapidly. To support the deployment of higher power-density servers, leading international cloud and technology companies - such as Amazon, Google, and Microsoft - have been early adopters and large-scale implementers of advanced data center cooling technologies, including liquid cooling.

Microsoft, for example, has continued to innovate in data center cooling and power infrastructure. In 2024, the company accelerated the deployment of chip-level liquid cooling systems, gradually replacing traditional air-cooling approaches. This transition not only significantly increases server density and reduces the resources required for constructing new data centers but also supports the company’s 2030 “water positive” commitment by eliminating reliance on evaporative cooling systems.

At the same time, Microsoft applies Life Cycle Assessment (LCA) tools to evaluate the carbon emissions and water consumption associated with different cooling solutions. Through open platforms such as the Open Compute Project, the company publicly shares its research findings and technical practices, helping to promote collaborative innovation and green transformation across the broader data center industry.

In China, alongside the rapid development of the digital economy, computing centers - the core infrastructure underpinning the AI industry chain - are increasingly establishing benchmarks for green and low-carbon development. One notable example is the data center campus built by China Mobile, one of the China's three major mobile network operators, in Hohhot, Inner Mongolia. The campus has become a key computing infrastructure node within China's "*East Data, West Computing*" initiative.

The facility provides users with Infrastructure-as-a-Service (IaaS) capabilities, including heterogeneous computing power, distributed storage, and network resources. It also supports AI development frameworks through third-party platforms and offers AI-enabled Platform-as-a-Service (PaaS). In addition, the campus has established Model-as-a-Service (MaaS) capabilities that support both general large language models and industry-specific foundation models.

Within the campus is currently the largest single liquid-cooled intelligent computing center operated by a telecom provider globally. The facility deploys approximately 20,000 AI accelerator cards, with more than 700 racks and over 2,300 intelligent computing servers, delivering a total intelligent computing capacity of 6.7 EFLOPS (6.7×10^{18} floating-point operations per second).

Through the deployment of multiple energy-saving technologies - including indirect evaporative cooling, magnetic-levitation phase-change cooling, fluorine-pump cooling, immersion cooling, and cold-plate liquid cooling - as well as busbar power distribution technology and an intelligent energy management platform, the data center has achieved a Power Usage Effectiveness (PUE) of 1.15 [12], among the leading levels in China.

During the winter season, the campus also utilizes waste heat recovery technologies, transforming the “waste heat” generated during data center operations into a valuable energy resource to support local heating demand. This approach reduces the energy consumption of the cooling water system while improving overall energy efficiency.

Transforming data centers from high-energy-consuming facilities into opportunities for both environmental and economic value requires recognizing them as an integral part of the broader energy system. This involves enabling synergies between computing and heat, as well as between computing and electricity.

With forward-looking planning and sitting in locations that allow effective integration with both power generation sources and end users, data centers can become valuable assets for low-emission district heating through waste heat recovery. At the same time, they can serve as important resources for enhancing the flexibility of the power system. Such integration not only reduces data centers’ own energy costs and strengthens the competitiveness of the industry but also contributes to improving overall energy security.

3. From Energy Consumer to Low-Carbon Heat Provider (Re-use)

It is essential to adopt a new perspective on the role of data centers. By positioning data centers as integral components of the energy system rather than merely energy consumers, they can be transformed from a challenge into an opportunity.

Data centers operate continuously throughout the year, and the IT equipment within them generates significant amounts of heat as a result. To maintain servers within an appropriate operating temperature range, cooling systems are essential to data center operations. The heat generated during operation must be transferred through a cooling medium—either air or liquid—and subsequently discharged.

Under traditional configurations, this heat is typically released directly into the atmosphere through facilities such as cooling towers. This approach not only wastes energy that has already been paid for but also prevents these thermal resources from realizing their full potential. By contrast, the cooling medium carrying waste heat from servers can be routed to heat exchangers, where the heat is transferred to a secondary heating loop connected to end users, enabling its reuse.

During this process, heat pumps are used to upgrade the relatively low-grade waste heat contained in the data center cooling medium to temperatures suitable for district heating networks. This approach is particularly effective for data centers employing liquid cooling technologies, where waste heat recovery can be achieved more efficiently.

At present, waste heat from data centers can already be reused at various scales. For example, it can provide heating for nearby facilities such as agricultural greenhouses, aquaculture farms, and swimming pools. At a larger scale, recovered heat can be integrated into district heating networks to supply urban heating, enhancing the resilience and diversification of energy supply while reducing carbon emissions.

Global electricity consumption by data centers is expected to double by 2030, suggesting substantial potential for waste heat recovery. Currently, most of this heat is discharged directly into the atmosphere, resulting in significant energy waste. If effectively harnessed, this otherwise free heat source could create valuable opportunities for providing cost-effective and low-emission heating.

According to estimates from the International Energy Agency (IEA), data centers located near district heating networks worldwide could potentially recover between 35 and 85 TWh of heat. By 2030, up to 10% of Europe's space heating demand could be met through the recovery of waste heat from nearby data centers connected to district heating systems, at costs significantly lower than those associated with natural gas heating [2].

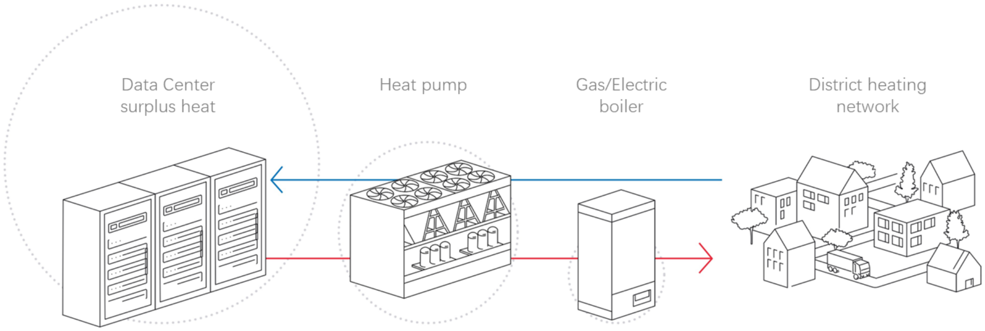

Figure 2: Enhancing the Diversity and Resilience of District Heating Systems

In the context of low-carbon urban development in Europe, a data center waste heat recovery project in the Helsinki metropolitan area of Finland has emerged as one of the most representative global case studies. The project is a collaboration between Microsoft and the Nordic energy company Fortum.

The system recovers waste heat from the cooling water systems of two large Microsoft data centers located in the region. The recovered heat is then upgraded through a large-scale heat pump facility constructed and operated by Fortum and subsequently delivered into the urban district heating network. This clean heat supply provides heating for residential buildings, public facilities, and commercial properties in the surrounding municipalities of Espoo and Kirkkonummi, near Helsinki.

The overall system integrates large-scale heat pumps, thermal storage infrastructure, and an electrified energy system, representing a significant practical application of modern low-carbon district heating. The project is widely regarded as one of the largest city-scale data center waste heat recovery initiatives in the world.

In terms of scale, once fully operational, the system is expected to meet approximately 40% of local district heating demand and reduce carbon dioxide emissions by around 400,000 tons annually [13]. This demonstrates that data centers are not only critical infrastructure for the digital economy but can also become an integral component of urban energy systems.

Compared with traditional heating methods that rely on natural gas or other fossil fuels, this model significantly improves energy utilization efficiency while lowering both urban carbon emissions and energy costs. As such, it offers a new development pathway for cities worldwide - one in which data centers and urban energy systems are planned in coordination, enabling computing infrastructure to be transformed into a stable source of low-carbon heat.

Figure 3: Heat pump station at Kirkkonummi data center (October 2024)

In China, the Qingshu Technology Park in Wuqing District, Tianjin (with a total built-up area of 103,000 m² as a mixed-use complex) is located adjacent to the Gaocun Technology Innovation Park data center cluster, at a straight-line distance of 400 meters. The data center cluster is planned to include nine data centers, with an expected operational scale of over 100,000 server racks. Danfoss, as a company with extensive experience in data center cooling and district heating, is actively participating in the pilot demonstration of collaborative green digital infrastructure in this area.

The park employs a dual-mode magnetic-levitation heat pump system to enable large-scale recovery and integrated utilization of waste heat from the data centers. The system switches operating modes seasonally: during winter, it extracts waste heat from the data centers’ circulating cooling water to provide heating and domestic hot water for the technology park buildings; in summer, the system is converted to provide cooling for the park using rooftop cooling towers.

Even during the first heating season of the park’s operation, when occupancy had not yet reached full capacity, the average heating cost using data center waste heat was already approximately 60% lower than local commercial heating costs, and 30%–40% lower compared with conventional standalone air-source heat pump systems.

Once the park reaches full design occupancy, the project is expected to save around 1.1 million kWh of electricity annually for the adjacent data centers, while also reducing water consumption associated with traditional cooling tower operations. Over the course of a year, the system is projected to reduce carbon dioxide emissions by approximately 1,659 tons [14]. Despite the additional upfront investment required for waste heat recovery infrastructure, the expected payback period is less than three years.

Based on calculations, once the Gaocun Technology Innovation Park data center cluster is fully completed according to plan, the recoverable waste heat from the data centers is expected to exceed 230 MW, theoretically sufficient to meet the heating demand of 4–5 million square meters of buildings.

This low-carbon heat source from data centers, once upgraded via heat pumps, can fully serve as a substitute heating resource for existing regional heating stations, reducing their reliance on natural gas for heating, lowering associated carbon emissions, and cutting operational costs.

The nearest district, Shouchuang International Peninsula, comprises over 3 million square meters of residential, educational, and commercial facilities. Local authorities, in coordination with the development zone housing the data center cluster and regional building planning and design consultants, are actively assessing opportunities to reuse data center waste heat, aiming to replicate the successful implementation demonstrated at the Qingshu Technology Park.

4. Flexible Computing–Energy Coordination to Enhance Grid Resilience (Resource)

The accelerated deployment of data centers poses challenges to electricity supply and grid stability. Current policies and planning frameworks tend to treat data centers as isolated electricity consumers rather than active participants within the broader energy system.

While the technical potential for data centers to engage in demand response programs or participate in flexibility markets is increasingly being demonstrated, it has yet to be systematically developed. This disconnect limits the contribution that data centers could make toward renewable energy integration, decarbonization, and grid stability. Integrating data centers into comprehensive energy system planning could have profound implications for the power grid, the environment, and the economy.

4.1 New challenges in regional power planning and energy supply

As noted earlier, according to the latest research by the International Energy Agency (IEA), the rapid global expansion of computing infrastructure is expected to double the share of electricity consumption by data centers, reaching 3% of total global electricity demand by 2030.

At the regional level, the cumulative load of data centers has already had a significant impact on local power grids. For example, in Shanghai, China, more than 160 internet data centers are currently in operation. By 2025, their electricity consumption is projected to exceed 3% of the city’s total electricity demand, significantly higher than the national average of 1.7% [15]. In some “East Data, West Computing” node cities and regions, the share of electricity consumed by data centers exceeds 20%.

In Northern Virginia, USA, a region known as the “Data Center Corridor,” over 200 data centers form one of the densest and largest commercial data center clusters in the world. The electricity consumption of this cluster is roughly comparable to the peak demand of a city the size of Boston. Industry and research reports indicate that data centers account for approximately 20% of the regional electricity demand in this area.

This cluster has previously experienced an incident in which around 60 data centers simultaneously switched to backup power due to a widely adopted standard safety mechanism, resulting in an instantaneous 1,500 MW load drop from the main grid [16]. The event posed a risk to grid stability and forced operators to rapidly adjust generation output to maintain supply-demand balance. This incident underscores that, in regions with high-density data center clusters, grid operational flexibility and emergency dispatch face new and heightened challenges.

4.2 Efficient Coordination of Computing Power and Renewable Energy

As the global wave of artificial intelligence continues to accelerate, the growth of data centers and the AI industry ultimately translate into energy costs. The reality that computing power is fundamentally dependent on electricity is increasingly recognized across industry.

At the level of physical electricity supply, the two primary regions for global data center development - China and the United States - exhibit markedly different power market dynamics. In the United States, the explosive growth of AI and data centers is pushing the grid to its limits. Due to insufficient grid capacity, new data centers face long queues for grid connection, which can delay their deployment and construction. Currently, the average grid connection waiting time for newly built data centers ranges from 1 to 3 years, with some states experiencing delays of up to 7 years [2].

Grid capacity has become severely constrained, and the supply-demand imbalance in the electricity market has reached a critical point. “Power shortages” are increasingly common, and electricity prices in regions with concentrated data centers frequently reach record highs, sometimes even hitting regulatory price caps. In response, the U.S. government has stated that large technology companies building these high-energy data center infrastructures are responsible for securing their own electricity supply and bearing the resulting cost increases, rather than passing these costs onto ordinary consumers.

In China, total electricity consumption exceeded 10,000 TWh for the first time in 2025, a figure more than double the annual electricity consumption of the United States and surpassing the combined annual electricity use of the European Union, Russia, India, and Japan.

By the end of 2025, China’s total installed power generation capacity reached approximately 3,890 GW, representing an annual increase of about 16%. Of this growth, over 80% comes from newly installed wind and solar capacity, totaling a record 430 GW. Combined, wind and solar capacity will account for 47% of the nation’s cumulative installed generation capacity [17].

China’s rapidly expanding electricity supply, particularly the accelerated deployment and generation from renewable energy, allows electricity prices to remain relatively low. Against this backdrop of significant wind and solar capacity growth, data centers should treat computing power as a flexible resource, closely coupled with renewable output to enhance renewable energy utilization and reduce carbon emissions.

China’s “West Data, East Computing” strategic initiative leverages the differences in renewable energy endowments between the eastern and western regions, directing computing capacity toward renewable-energy-rich areas in the west to achieve a balance between computing expansion and low-carbon development.

In Qinghai, China, a demonstration project integrating green power and computing is currently under construction - the Qaidam Green Microgrid Computing Center.

This project is designed for decertified land with abundant photovoltaic resources, simultaneously developing the Qinghai Mobile computing center and a supporting clean energy supply system, including photovoltaic generation, an intelligent microgrid, and energy storage systems.

The computing center is planned to deploy approximately 1,000 high-power server racks, employing cold-plate liquid cooling, intelligent busbars, and indirect evaporative cooling to optimize energy efficiency. The supporting renewable infrastructure includes 122 MW of installed photovoltaic capacity and 75 MW / 300 MWh of energy storage, integrated with a microgrid and energy management system. During the day, the microgrid supplies power from PV generation and charges the energy storage, at night, stored energy powers the computing center. Once completed, the project is expected to supply approximately 126 million kWh of clean electricity annually to the computing center, saving around 42,100 tons of standard coal and reducing CO_2 emissions by roughly 115,500 tons per year [18].

As China's first large-scale clean-energy microgrid computing center, this project simultaneously explores the economic viability and operational mechanisms of off-grid and microgrid systems at remote computing nodes, providing a practical model for local renewable energy utilization and computing capacity deployment in western renewable-energy-rich regions.

4.3 Flexible Demand-Side Potential of Data Centers

As demonstrated by the Qaidam Green Microgrid Computing Center, the northwestern region of China is rich in wind and solar resources but faces constraints in power transmission and local curtailment of wind and solar generation. As a high-load, scalable new electricity user, the computing center leverages a microgrid to enable local consumption of renewable power, effectively converting variable electricity into tradeable computing capacity, improving renewable energy utilization and reducing curtailment.

Unlike traditional industrial loads, data center demand exhibits temporal and spatial flexibility. It represents a type of new flexible load capable of instantaneous load shifting independent of the main grid, making it a viable participant in demand response (DR) programs. Demand response mechanisms in power systems encourage users, through price signals or contractual incentives, to adjust electricity consumption during periods of grid stress or peak renewable output - such as peak shaving, valley filling, short-term load reduction, or deferred flexible loads - thereby reducing reliance on expensive or high-carbon peaking units, enhancing renewable integration, and lowering system operating costs.

Beyond the adjustable computing workload, data centers also house internal flexibility resources that can contribute to the operational flexibility of modern power systems. These include uninterruptible power supplies (UPS), standby diesel generators, battery energy storage, and thermal flexibility assets such as air conditioning, cooling, and thermal storage systems.

Ways in Which Data Centers Provide Grid Flexibility Coordination Include:

- Cross-regional workload scheduling: Computational workloads handled by data centers in regions experiencing grid constraints can be shifted to data centers located in areas with abundant renewable electricity.
- Time-based workload scheduling: Certain data center workloads - such as AI training, batch processing, and other non-real-time computing tasks - possess temporal flexibility. Through coordinated load management and energy storage, computing loads can be increased during periods of high renewable energy generation and reduced during peak grid demand, allowing data centers to function as a “power system stabilizer” that enhances grid reliability.
- On-site power generation facilities: Local generation assets can serve either as backup power sources or as baseload supply for data centers. Their output can be flexibly adjusted in response to grid conditions, reducing the sensitivity of data center operations to fluctuations in electricity supply. When powered by hydrogen fuel cells, green hydrogen can provide zero-noise, high-availability backup electricity. Similarly, high-efficiency off-grid gas generators can serve

as a viable transitional solution for phasing out oil- and coal-fired power generation.

- Battery Energy Storage Systems (BESS): Energy storage systems are essential for managing fluctuations in electricity demand from large-scale data centers while preventing grid instability. During periods of surplus renewable generation, batteries can store green electricity from on-site generation or the grid. This stored energy can then be used during times of supply constraints or high electricity prices, enabling data centers to maintain operations while supporting grid balancing.
- Flexibility of liquid cooling systems: Liquid cooling technologies use water-based cooling to dissipate heat from IT equipment. As the cooling water absorbs heat, it must be re-cooled before recycling. Cooling energy can be stored during periods of low electricity prices through thermal energy storage, allowing data center cooling systems to respond flexibly to changes in power supply. In some cases, cooling energy can even be stored across seasons, providing additional operational flexibility.

A study simulating different levels of data center flexibility across three U.S. power markets - the Mid-Atlantic, Texas, and the Western Interconnection (WECC) - shows that such measures could reduce power-sector emissions by up to 40%, accelerate the phase-out of fossil fuels, and lower total system costs by around 5%. Even in scenarios where flexibility measures do not directly lead to additional decarbonization, they can still reduce overall system costs. This is because enhanced flexibility allows grid operators to prioritize the lowest-cost electricity resources while reducing the need for new transmission and distribution infrastructure, ultimately lowering energy expenditures for all electricity consumers.[19]

In practice, Google has already implemented a carbon-aware scheduling system across its data centers. This mechanism dynamically optimizes the timing and geographic location of computing workloads based on the real-time carbon intensity of electricity. By continuously identifying grid carbon intensity and applying AI-driven analytics, the system shifts delay-tolerant computing tasks across both time and location, aligning data center operations with periods and regions of lower-carbon electricity supply. This approach represents a key technological pathway toward Google’s 24/7 carbon-free energy goal by 2030 and provides a practical demonstration of the significant emissions-reduction potential enabled by flexible data center loads in power systems with high shares of renewable energy.[20]

Against the backdrop of rapidly growing computing demand in mega-cities, Shanghai, China is also exploring ways to transform data centers from high-energy consumers into flexible grid resources. By 2025, electricity consumption by data centers in Shanghai accounted for around 3% of the city’s total power demand, significantly higher than the national average of 1.7%.

To explore the potential of computing–power coordination, the State Grid Corporation of China and China Telecom jointly conducted real-world demonstrations of cross-provincial computing workload migration. Tests were carried out along routes such as “Shanghai–Shiyan” and “Shanghai–Fujian,” enabling the coordinated dispatch of hundreds of GPU cards across clusters, with multi-task scheduling at the minute scale. In the Shanghai–Fujian experiment, the demand response capacity was expanded to 50 kW, marking an important step from technical validation toward commercial application. Shanghai has also established a virtual power plant (VPP) platform with an aggregated capacity of 2,568 MW, of which data centers contribute 182 MW, representing 7.1% of the total. During peak summer electricity demand, data centers have demonstrated a maximum response capability of 77 MW, representing more than a fivefold increase compared with 2024.[21]

Academic research further suggests that by 2030, the flexible capacity of China’s data centers could reach approximately 23–40 GW. In this context, computing–power coordination between data centers and electricity systems is expected to become a key scenario for the integrated development of digital infrastructure and energy systems, strengthening both grid flexibility and low-carbon energy integration.[6]

5. Policy Recommendations

Data centers are rapidly becoming core infrastructure underpinning the digital economy and the development of artificial intelligence. Their growing consumption of energy and water resources has therefore attracted increasing global attention. It is imperative to reconsider how data centers can evolve from high-energy-consuming facilities into key infrastructure that both constrains and empowers the energy system, thereby alleviating regional pressures on energy and water resources while transforming the expansion of computing power into a low-carbon opportunity.

Through more comprehensive top-level policy design, it is essential to promote flexible coordination among computing, electricity, and heat systems. Such an approach can not only provide certainty for industrial growth but also incentivize data centers - through regulatory frameworks and market mechanisms - to actively participate in the transition toward a low-carbon and resilient energy system, shaping a sustainable development pathway for the sector.

5.1 Deepen Energy-Efficiency Technology Innovation and Improve System Energy Utilization Efficiency

Efforts should focus on upgrading energy efficiency across the entire lifecycle of data center infrastructure. Targeting the two major energy-consuming subsystems - cooling and power supply - technological innovation and large-scale deployment should be promoted in accordance with regional conditions. By optimizing overall system architecture and continuously reducing key efficiency indicators such as Power Usage Effectiveness (PUE) and Water Usage Effectiveness (WUE), a highly efficient and energy-saving green computing infrastructure can be established.

Advanced energy-efficient cooling technologies should be promoted nationwide. Considering differences in regional climate conditions, data center power density, and business scenarios, a scenario-based and differentiated deployment framework for advanced cooling technologies should be established. At the same time, intelligent linkage between cooling systems and IT workloads should be strengthened. By dynamically adjusting cooling strategies according to computing load variations, data centers can achieve “cooling on demand” and precise system matching, thereby continuously reducing the proportion of energy consumed by cooling systems.

In parallel, the iterative upgrading of high-efficiency power supply systems should be accelerated. Data center power architecture should be optimized by simplifying power conversion stages and reducing electricity losses during transmission and conversion. Enhancing conversion efficiency and dynamic response capabilities will allow for a reduction in unnecessary equipment redundancy while maintaining power reliability, thereby lowering construction and operational costs.

Furthermore, comprehensive regulatory oversight of data center energy efficiency should be strengthened across both entry and operational stages. Establishing unified energy measurement standards and information disclosure mechanisms will help create an industry-wide benchmarking and supervision system. Through coordinated policy guidance and market mechanisms, the sector’s overall energy utilization efficiency can be improved, preventing extensive development from placing structural pressure on regional electricity and water systems, and laying a solid efficiency foundation for the low-carbon development of computing infrastructure.

5.2 Integrate Waste Heat into Urban Energy Planning to Promote Multi-Stakeholder Reuse Models

Data centers not only provide computing capacity but also generate large quantities of low-grade waste heat during operation. When integrated with district heating systems, this heat can support a “computing + thermal energy” integrated infrastructure model, which improves urban energy efficiency while reducing carbon emissions.

Although the technical solutions for utilizing data center waste heat are already mature, large-scale deployment remains limited, with most projects still confined to individual pilot initiatives. Achieving widespread adoption will require systematic planning, coordinated policy support, and effective collaboration among stakeholders.

Cities seeking to incorporate data center waste heat into district heating systems may draw lessons from Europe's Fourth-Generation District Heating (4GDH) concept. This approach emphasizes low-temperature heating networks, distributed heat sources, and high levels of renewable energy integration, with the core principle being to enable efficient integration of diverse heat sources by lowering supply temperatures and increasing system flexibility.

At the policy level, urban energy planning and industrial park development should strengthen the integrated layout of data centers with district heating and industrial heat systems, incorporating waste heat recovery into overall urban infrastructure planning. For projects with suitable conditions, feasibility studies for waste heat recovery should be conducted at the planning stage, and standardized technical interfaces and specifications should be developed to reduce integration costs and collaboration barriers.

Market-based mechanisms and appropriate policy incentives can further encourage long-term partnerships between district heating operators and data center operators, supporting demonstration projects that validate low-temperature networks, heat-pump upgrading technologies, and multi-source heat dispatching systems. Diverse heat utilization models can thus be explored to reduce primary energy consumption.

Practical examples - such as the waste heat recovery project at the Qingshu Technology Park in Wuqing District, Tianjin - demonstrate the feasibility of this model and provide a replicable pathway. With prudent expansion, data centers can gradually evolve from single energy consumers into important contributors to urban energy recycling systems, strengthening their role in low-carbon urban heating networks.

5.3 Promote System Coupling and Flexibility Coordination Between Computing and Power Systems

As China's renewable energy capacity continues to expand and a new type of power system takes shape, data centers - despite their high electricity consumption - also possess significant load flexibility that can help enhance grid resilience and facilitate renewable energy integration. They therefore represent a crucial link in enabling the dual transition toward low-carbon computing and high-efficiency power systems. China is currently constructing an integrated national computing network, establishing a multi-layer, cross-regional scheduling framework for computing resources. This development provides an important foundation for enabling spatiotemporal coordination between computing workloads and electricity systems. In regions rich in renewable resources, data centers should be encouraged to co-locate with renewable energy generation, forming large-scale green data center clusters. Market-based mechanisms should be used to increase renewable energy consumption. The "generation-grid-storage-load" integrated model should be promoted, supporting the deployment of new energy storage facilities alongside data centers to create a closed-loop system covering renewable generation, storage, and consumption.

At the same time, improvements should be made to green electricity trading mechanisms, simplifying transaction procedures, reducing costs, and increasing the share of renewable electricity used by data centers. These measures can help implement China's "East Data, West Computing" strategy, which requires that over 80% of electricity used by new data centers in national hub nodes come from renewable sources.

Data centers should also be guided to participate in demand response programs, enabling those with load-adjustment capabilities to contribute to peak shaving, valley filling, and emergency reserve services. Aggregating loads across multiple data centers can enhance the scale and effectiveness of demand response.

While ensuring the stability of core computing services, new load optimization models aligned with the operational rhythm of power systems should be explored to enhance grid flexibility. Price signals can also be used to encourage data centers to actively adjust electricity consumption patterns. Greater transparency in real-time electricity prices and carbon intensity information would enable closer integration with computing workload scheduling systems, improving demand response participation and flexibility market rules while supporting more refined energy and carbon management.

5.4 Conclusion

Overall, as a form of new infrastructure, data centers simultaneously support the growth of the digital economy while being closely linked to the transformation of energy systems and the sustainability of urban development.

Future policy frameworks should therefore take energy efficiency improvement as the baseline, waste heat reuse as an extension, and grid flexibility coordination as the strategic direction. By doing so, the green and low-carbon development of the data center industry can be ensured while enhancing the overall resilience and decarbonization of the energy system.

Key References

This report is primarily developed with reference to the framework of <Danfoss Impact Series Report No. 8 - Data Centers>.

1. De Vries, A. (2023). The Growing Energy Footprint of Artificial Intelligence. Joule, Vol. 7, No. 10.
2. International Energy Agency (IEA) (2025). Energy and Artificial Intelligence.
3. Uptime Institute - Annual Survey (International Data Center Standards and Certification Organization).
4. Ministry of Industry and Information Technology of China (MIIT)
5. China Academy of Information and Communications Technology (CAICT) (2025). <Research Report on the Development of Green Computing Power>.
6. <Comprehensive Energy Consumption of Data Centers in China and Its Flexibility Forecast> Journal of Beijing Institute of Technology, March 2025.
7. International Energy Agency (IEA) (2025) <Energy Efficiency Report 2025>.
8. U.S. Energy Star Program. Guide <Reducing Power Distribution Unit (PDU) Energy Losses>.
9. Lawrence Berkeley National Laboratory (LBNL). <Power & Lighting Systems>.
10. NVIDIA - Product technical data and analysis.
11. Huawei - Product technical data and analysis.
12. China Mobile.
13. William Holley (July 2025). <Transitioning to Low-Carbon Heating>.
14. People’s Daily Online. <Using Low-Grade Waste Heat to Warm Cities>.
15. The Paper Institute.
16. DataCenterDynamics.com - Industry analysis and reports.
17. National Energy Administration of China (NEA) and China Electricity Council (CEC).
18. China Energy Network.
19. National Bureau of Economic Research (NBER). <Flexible Data Centers and the Grid: Lower Costs or Higher Emissions?>.
20. Google Blog. <Shifting More Computing to Regions with Abundant Clean Energy>.
21. The Paper (Pengpai News).