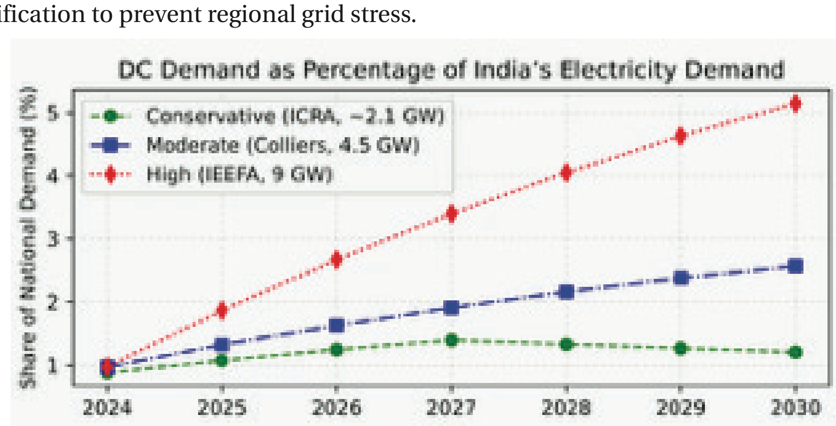
Powering the AI Era

Preparing the grid for the data centre boom

Parikshit Pareek, Assistant Professor, Department of Electrical Engineering, IIT Roorkee, and S.K. Soonee, Senior Adviser, Independent Consultant, Founding CEO of POSOCO (now Grid-India)

By the end of this decade, India's data centres could consume as much power as an entire state, such as Chhattisgarh did at its 2024 summer peak (about 6.5 GW). Standing at the cusp of a massive digital build-out, data centres are expanding at unprecedented speed. They are the backbone of cloud computing, artificial intelligence (AI) and high-performance computing. Colliers India estimates operational capacity at 1,263 MW (April 2025), projected to exceed 4,500 MW by 2030, backed by $20 billion-$25 billion of investment. ICRA forecasts 2,000-2,100 MW by FY 2027, while JLL projects about 1,825 MW in the same period, driven by 5G and AI adoption. The Institute for Energy Economics and Financial Analysis' high-growth case reaches 9 GW by 2030, about 3 per cent of national demand. At this pace, data centre load alone could soon rival a mid-sized state's demand, underscoring the need for anticipatory grid planning, operational flexibility and spatial diversification to prevent regional grid stress.



Note: Projected share of data centre annual energy demand as a percentage of India's total electricity demand (2024-30) under three scenarios: Conservative, moderate and high. Data centre facility energy calculation assumes that for 1 MW of IT load, 0.4 MW of auxiliary load will be required for data centres to function (power usage effectiveness [PUE] of 1.4), and national electricity demand is projected from a 2024 baseline of 1,600 TWh with about 5 per cent annual growth. Note that this auxiliary load will depend on atmospheric conditions, particularly temperature, and will not be constant across the year.

All roads lead to metros

Market assessments indicate that nearly three-fourths of India's data centre growth is concentrated within key metropolitan corridors: Mumbai, Hyderabad, Chennai-Vizag, and Delhi-National Capital Region. Mumbai alone hosts a multi-GW pipeline and ranks among global leaders in under-construction capacity, while Hyderabad and Chennai continue to attract hyperscale and colocation projects due to strong connectivity and land availability. Recently, Google announced a gigawatt-scale data centre cluster at Vizag. This clustering creates high local nameplate capacities, often hundreds of MW per campus and turns select substations into systemic power hubs. As these sites align with major internet exchanges and fibre trunks, their electrical demand becomes geographically concentrated and temporally correlated with urban business cycles, amplifying stress on metropolitan networks.

Taking things to the edge

As most large data centres remain clustered in metros, latency and regional access gaps persist. To bridge this, edge data centres, which are smaller facilities built closer to users, are expanding rapidly, with capacity expected to triple from 60-70 MW in 2024 to about 200 MW by 2027. While individually modest, their spread into Tier II and III cities will create new data centre load hotspots far from established metro corridors, posing new challenges for local distribution networks, reliability and cybersecurity, but also offering opportunities for localised flexibility and digital inclusion.

Why data centre load matters for the grid

Unlike industrial plants whose patterns follow production cycles, data centre demand depends on algorithms, not human activity. AI training workloads can run for days without pause, drawing nearly constant power from GPU clusters. Each rack can consume 40-60 kW, which is 5-10 times the traditional server intensity, and any voltage disturbance can trigger automated protection, instantly shifting load to backup. These "silent exits", observed in the US and ERCOT studies, can remove gigawatts of load in seconds, producing frequency overshoot and reactive power swings.

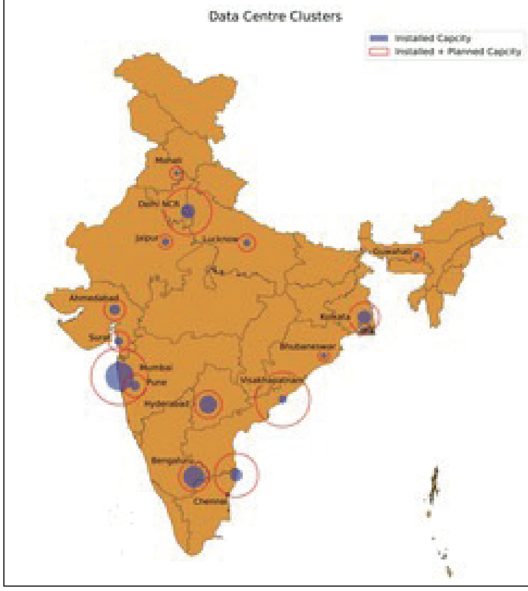
Data centre load is high-density, high-velocity

Even a single hyperscale data centre can impose a demand comparable to a major industrial cluster, creating highly concentrated and fast-varying loads within already stressed urban grids. In cities such as Hyderabad, Delhi and Mumbai, where peaks range from a few to about 8 GW, an individual 100-500 MW facility may represent several per cent of total demand.

From a power-engineering perspective, these centres represent a new class of localised, power-electronic-dominated loads whose dynamics directly influence grid stability. Their rapid, autonomous ramps and disconnections can cause instantaneous power imbalances, steep frequency excursions and high rates of change of frequency, quickly exhausting reserves and degrading area control error performance. At substations, converter and UPS systems draw and inject reactive power dynamically, distorting voltage profiles and occasionally triggering protection operations or voltage instability. Since they are dominated by non-linear electronics, data centres also generate harmonics that propagate through the network, affecting nearby sensitive equipment. As multiple campuses cluster around the same 220/440 kV nodes, transformer loading, short-term balancing and protection coordination all face new stress regimes, underscoring the need for predictive local planning and dynamic control strategies.

Unlike conventional industrial demand that follows daily or seasonal cycles, data centre load shows weak correlation with time-of-day or consumer activity. High-performance computing and AI training tasks can run continuously for days, creating near-constant or burst-type consumption that may peak during off-peak hours. As India begins hosting large AI clusters, such workloads could trigger sudden gigawatt-scale power draws without prior disclosure. Since real-time utilisation data are rarely shared due to commercial and cybersecurity concerns, planners face a "black box" in forecasting intraday ramps. This opacity complicates scheduling and reserve management, underscoring the need for mandatory telemetry and anonymised operational data-sharing to help operators anticipate computational surges in real time.

Data centre clusters



Modelling data centre load: Beyond static assumptions

Behind the steady hum of a data centre lies a dynamic system of servers, cooling units, UPS banks and power-electronic converters – all interacting within milliseconds. Traditional ZIP (constant impedance, current and power) models cannot capture these rapid, non-linear shifts in consumption. Recent work now moves toward component-based and dynamic models that simulate how IT loads, cooling and auxiliary systems respond during voltage dips or fault events. Such models, endorsed in IEEE 2781, are vital, as AI-driven workloads introduce bursty, unpredictable demand patterns. For India's planners and grid modellers, adopting these advanced representations will be essential to anticipate how hyperscale and edge facilities influence system stability, transient behaviour and reliability margins across an increasingly digital grid.

Industrial vs data centre load characteristics

Attribute	Industrial load	Data centre load
Load profile	Predictable	Continuous/bursty
Disturbance response	Operator-controlled	Automated UPS transfer
Power factor	Steady lagging	Dynamic/inverter-linked
Ramp rate	5-10 % per min	Up to 100 % per second

From Virginia to Dublin: Global lessons

Across the world, rapid data centre growth is reshaping power systems and market dynamics. In the US, PJM's capacity market for 2026/27 cleared at its maximum price of $329.17 per MW-day and thus, signalling tight supply and localised surges. Utilities such as Dominion Energy have revised multi-year investment plans to meet Virginia's soaring data centre-driven load. Ireland offers a sharp warning: with data centres consuming nearly one-fifth of national electricity in 2023, the government imposed a temporary moratorium on new Dublin projects to avoid grid stress. Such concentrated growth introduces regulatory and planning risks, including high augmentation costs, potential stranded assets and dependence on a handful of large consumers.

Operational incidents in the US and Europe reveal that sudden multi-gigawatt disconnections of voltage-sensitive loads can cause frequency overshoot and instability. Studies document extreme load ramps, forced oscillations and harmonic distortions from dense power electronic deployments. Beyond electricity, water use and emissions are emerging flashpoints: a 100 MW hyperscale data centre may consume about two million litres per day, while Microsoft reported a 29 per cent rise in emissions since 2020.

From burden to flexibility

While data centres impose challenges, they also offer potential as flexibility assets. Global practice highlights three pathways: demand response, self-generation and workload shifting. Companies such as Google have demonstrated demand modulation to aid grid stability. Texas Senate Bill 6 (2025) requires controllable shutdown capability for large consumers, while Europe's Market Model 3.0 integrates data centre flexibility into ancillary markets. India could adopt analogous frameworks linking tariff in-