layout and operation called a digital twin. To fully appreciate the scope of this new tool, consider the varieties of data centers and their functions.

HIGH-LEVEL DESCRIPTION OF DATA CENTERS

Data centers provide space, power, cooling, and network access for IT infrastructure that executes a wide variety of workloads, from cloud computing to traditional enterprise, to HPC and AI. They can range from small shipping container form factors to hundreds of thousands of square feet while consuming over 100 MW of power in the largest facilities. Figure 2 displays the major subsystems in a typical data center. Electrical power is delivered from the public utility grid or local microgrid to IT and cooling resources in the data center where it is eventually dissipated as heat to the external environment via the data center's cooling infrastructure. The power and cooling subsystems are generally operated independently of the IT infrastructure, though integrated management of IT and facility subsystems has shown the potential for efficiency improvements.[2]

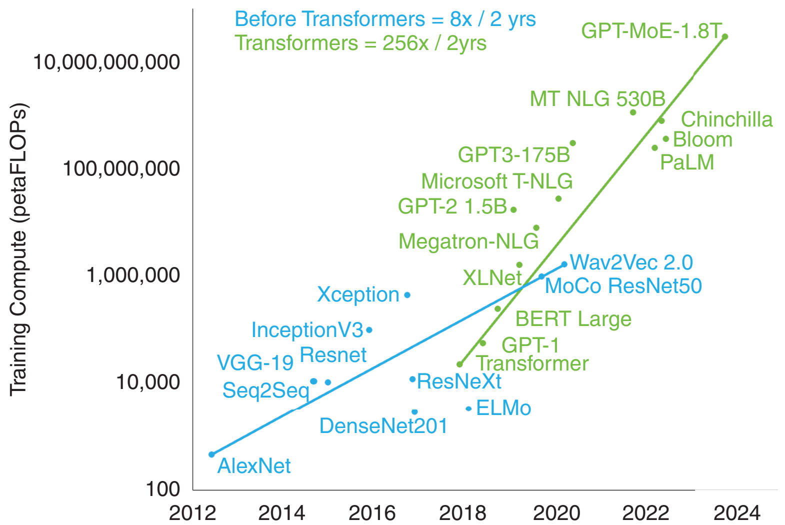

FIGURE 1. Explosive growth in AI computational requirements drives data center upgrades (source: NVIDIA analysis: reproduced with NVIDIA permission).

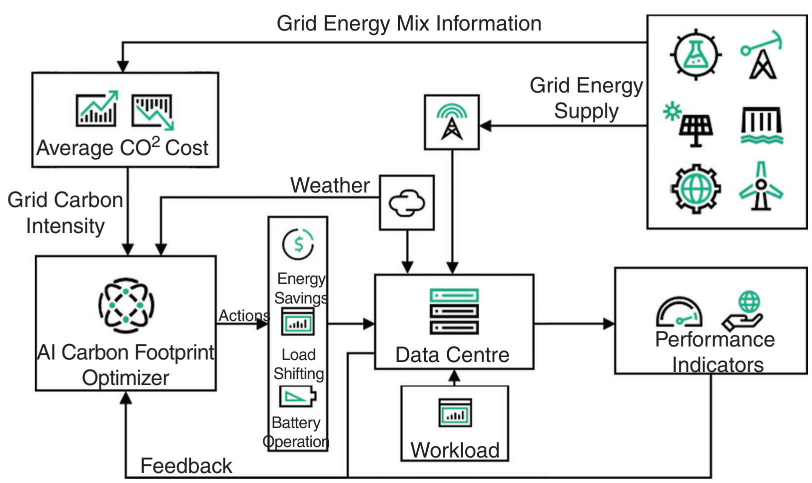

FIGURE 2. Data center architecture: IT equipment and workload, grid energy supply, carbon intensity, and so on.

DIGITAL TWINS

Adapting NASEM's definition of a digital twin[3] to data centers, we define a data center digital twin (DCDT) as a virtual model that replicates the structure, context, and behavior of a data center. Continuously updated with data from its physical counterpart, a DCDT possesses predictive capabilities and aids in informed decision-making to optimize operations, extend operating life, and realize value. A key aspect of a DCDT is the bidirectional interaction between the virtual model and the physical data center.

Replicating the state of the physical data centers (see Figure 3) enables users to complement existing data center management functionality with complex functionality not feasible or practical with runtime management tools. The replicated state can be combined with an external state and with simulation tools resulting in superior prediction capabilities.

It is important to point out that digital twins replace neither human oversight nor data center management tools, but rather complement both by providing advisory and optimization functionalities. Digital twins operate within the granularity range of minutes to hours, sufficiently fast to enable an interactive mode of operations with users, yet slower than management tools to enable practical consistency with the physical instances (see Figure 4).

Some key use cases for digital twins in data centers include the following:

1. *Energy optimization*: DCDT can model and predict energy consumption patterns, enabling real-time adjustments to cooling systems, power distribution, and workload allocation to maximize energy efficiency.
2. *Capacity planning*: By simulating different scenarios, DCDT can help data center managers optimize space utilization, predict