

# From Supply Chain to Operations: A Portfolio Approach to Capacity Expansion in Colocation Data Centers

#Generative AI   #IoT/Data Utilization   #Research & Development

## Author

**Malarvizhi Sankaranarayananamy, Ph.D.**

Big Data Analytics Solutions Lab, R&D., Hitachi America Ltd.  
*Current work and research:* R&D of data center and industrial solutions using reinforcement learning, first principles simulation, and generative AI technologies. *Society memberships:* The Association for the Advancement of Artificial Intelligence (AAAI) and the Society of Manufacturing Engineers (SME).

**Neda Zarayeneh, Ph.D.**

Big Data Analytics Solutions Lab, R&D., Hitachi America Ltd.  
*Current work and research:* Forecasting models for time series industrial data to achieve long-term prediction, anomaly detection, and optimized operational decisions. *Society memberships:* The Institute for Operations Research and Management Sciences (INFORMS).

**Omanshu Thapliyal, Ph.D.**

Big Data Analytics Solutions Lab, R&D., Hitachi America Ltd.  
*Current work and research:* Urban Air Mobility R&D. *Society memberships:* The Society of Automotive Engineers (SAE) and IEEE Control Systems Society (IEEECSS).

**Ravigopal Vennelakanti**

Big Data Analytics Solutions Lab, R&D., Hitachi America Ltd.  
*Current work and research:* Accelerating technologies from concept to market realization, harnessing the power of data-centric strategies to tackle complex challenges across diverse industry landscapes. Leading initiatives relating to data center ecosystems at Hitachi America, Ltd.'s R&D center.

**Pegah Mavaie, Ph.D.**

Big Data Analytics Solutions Lab, R&D., Hitachi America Ltd.  
*Current work and research:* R&D of data center and industrial solutions using deep learning, and generative AI technologies.

**Prasun Singh**

Big Data Analytics Solutions Lab, R&D., Hitachi America Ltd.  
*Current work and research:* Research and development of solutions applying blockchain, AI, and sensor fusion. *Society memberships:* The Association for Computing Machinery (ACM) and IEEE.

## Highlight

The colocation data center market, representing 38% of the industry, faces significant challenges in managing power capacity portfolios and optimizing resource allocation. This article explores a novel technology-driven portfolio approach that leverages agentic AI for data extraction and system abstraction, facility digitization with metaverse, and microclimate platforms to address the challenges across data center capacity segments. The proposed approach, enabled by large language models and agentic AI, generates demand insights, assesses power consumption patterns, identifies SLA violations, and explores supply constraints along spatial, resource, and power axes. This solution is devised to autonomously extract, analyze and interpret complex data from diverse sources, providing a clear path toward optimizing resource allocation, maximizing capacity utilization, and driving sustainable growth in the rapidly expanding colocation data center market.

## 1. Introduction

In today's digital economy, data centers have become critical infrastructure, often referred to as "AI Factories" for their role in powering advanced computing. Colocation data centers, which allow multiple businesses to rent space, power, and cooling for their servers in a shared facility, are a cornerstone of digital infrastructure. These facilities provide essential resources for securely housing IT infrastructure. As digitization accelerates, the demand for data center services is surging, presenting significant economic opportunities. However, the industry faces challenges such as energy constraints and equipment shortages, threatening sustainable growth. To address these challenges, colocation providers require a paradigm shift in their operational strategies. We propose a portfolio

management approach that focuses on optimizing current, new, and total capacity segments. This strategy deconstructs capacity management into its core components: space, resources, and power, enabling providers to tackle the complex issues of energy efficiency and equipment utilization head-on.

Our analysis reveals that optimizing colocation portfolio management requires innovative technologies grounded in fundamental principles of physics, economics, and operational dynamics. Hitachi America R&D has developed a suite of advanced solutions to support this requirement, including Metaverse for Facility Digitization, Microclimate Platforms, and Large Language Models with Agentic AI. This technology-driven portfolio approach supports demand-driven growth and optimizes resource utilization across the entire portfolio of colocation facilities. By leveraging these solutions, providers can make decisions anchored in data-driven insights rather than historical practices, directly addressing the challenges of energy constraints and equipment shortages. This technology-driven strategy not only supports demand-driven growth, but also positions colocation providers for long-term success in an increasingly competitive and resource-constrained market.

## 2. Vision for Portfolio Management

The portfolio management approach, grounded in the first principles thinking outlined in the introduction, offers a strategic solution to the challenges faced by colocation providers. This approach involves the oversight and optimization of multiple data center facilities, resources, and services offered by a colocation provider to meet customer demands efficiently while maximizing revenue and minimizing risks. The portfolio management strategy addresses power constraints, supply chain issues, and resource limitations by focusing on three key segments:

- (1) New Capacity  
Planning for expansion and upgrades based on projected demand.
- (2) Current Capacity  
Optimizing the utilization of existing resources and infrastructure.
- (3) Total Capacity  
Managing the overall portfolio to ensure long-term sustainability and profitability.

Each segment requires distinct operational strategies tailored to its unique demands. The vision for this approach is to create a network of interconnected, highly efficient data centers that dynamically adapt to market demands. This is achieved by leveraging cutting-edge technologies - including advanced analytics and AI-driven optimization - to implement data-driven strategies across three critical capacity segments: space, resources, and power. Unlike traditional colocation management, which often treats facilities as individual silos, the portfolio-centric approach enables providers to:

- (1) Dynamically allocate resources across facilities to meet fluctuating demand
- (2) Optimize utilization rates by balancing workloads and capacity across the entire portfolio
- (3) Implement proactive SLA management through predictive analytics and real-time monitoring
- (4) Scale capacity efficiently by leveraging underutilized resources before expanding

By adopting this portfolio approach, colocation providers can address individual challenges within each capacity segment while leveraging synergies across their entire operation. The following sections will demonstrate how specific in-house technologies, such as metaverse and agentic-ai driven space optimization and advanced power management systems, can be employed within this portfolio approach to develop innovative capacity planning and management strategies. These technologies enable colocation providers to optimize space utilization, enhance equipment efficiency, and manage power effectively, ultimately supporting scalable growth and sustainable operations in an increasingly competitive market.

## 3. Achieving Maximum Utility Across Capacity Segments

To ensure maximum utility across current, new, and total capacity segments, a suite of in-house technologies has been developed. These technologies are designed to address the unique challenges of each capacity segment by integrating digital solutions into the portfolio management approach.

- (1) Metaverse for Facility Digitization  
This technology creates high-fidelity digital twins of data center assets and facilities, enabling:
  - (a) Comprehensive facility digitization through practical capture and reconstruction
  - (b) Asset Digital Twins: High-fidelity capture and reconstruction for optimized layout and synchronized operations
  - (c) Enhanced Capacity Management and Customer Engagement: Collaborative platform for improved communication and resource allocation while minimizing resource fragmentation.  
By integrating quota management policies and simplifying customer interactions, this technology significantly enhances the customer experience, particularly for new capacity segments.
- (2) Microclimate Platform  
This solution combines sensor fusion and advanced metering systems for real-time environmental monitoring and control:

- (a) Granular Environmental Monitoring: Precise tracking of temperature, humidity, and airflow at rack and server levels
  - (b) HVAC multi-setpoint control automation: AI-driven system for optimal, localized environmental management
- The Microclimate Platform integrates with the metaverse, enhancing environmental management across all capacity segments, with benefits for current capacity optimization.
- (3) Large Language Model and Agentic AI Architectures
- The architectures automate strategic operational plans through:
- (a) Alarms Management: Event profiling and summarization for improved SLA management and issue resolution
  - (b) Demand-Supply Forecast and Optimization: By abstracting and contextualizing external and internal signals, large language models generate demand insights, assess power consumption, and explore supply constraints across spatial, resource, and power axes.
- This technology has demonstrated a 15% improvement in long-range demand forecasting accuracy and a 20% reduction in supply delays, benefitting total capacity management.

By leveraging these technologies within the portfolio management framework, providers can address individual challenges while capitalizing on synergies across their entire operation, ensuring sustainable growth across all capacity segments.

### 3.1. New Capacity Visibility

In managing new capacity, colocation providers face the challenge of seamlessly integrating new infrastructure while optimizing existing resources. The facility digital twin in metaverse offers an interactive platform that enhances customer engagement and accelerates decision-making by providing virtual walk-throughs of data center spaces before physical deployment. The metaverse, built on a croquet architecture, enables real-time collaboration in a 3D shared space where users can interact and view each other's work, emphasizing context-based cooperation<sup>1)</sup>. This approach allows customers to visualize the layout, power, and cooling infrastructure, ensuring that their requirements are met in real time (see Figure 1).

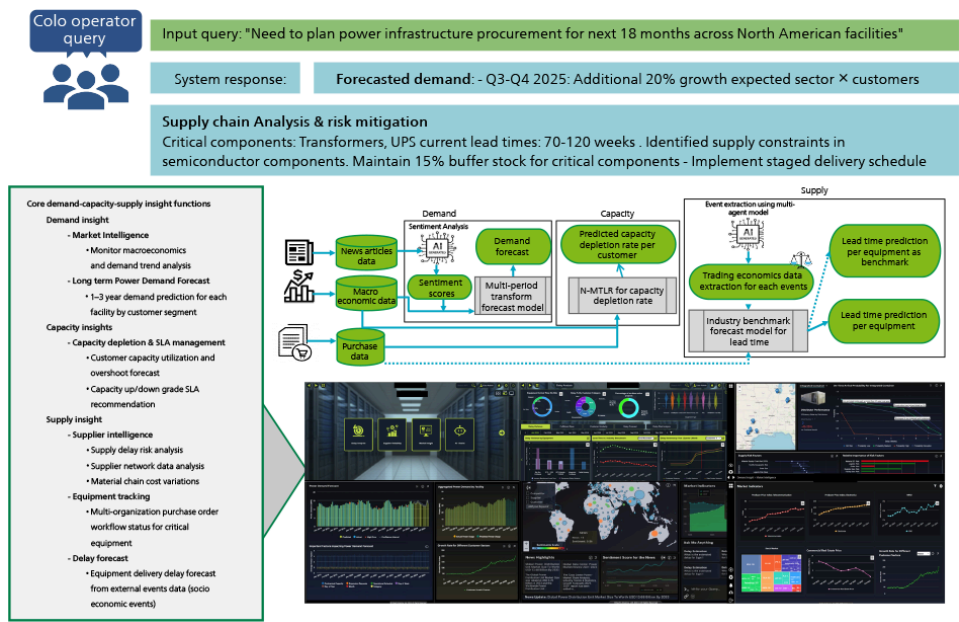


LLM: large language model, RoI: return on investment, SLA: service level agreement

Figure 1 illustrates the process from initial customer engagement to design validation. Customers can interact with 3D visualizations of critical infrastructure and collaborate on layout modifications using large language models for real-time optimization. The system captures the spatial-physical distribution of the current setup, maps temporal utility constraints and available space, and identifies an adjacent cage setup location for expansion. This virtual environment minimizes resource fragmentation, streamlines approval processes, and reduces planning cycles.

New data center construction and expansion require substantial capital investment, necessitating access to diverse debt capital markets. In addition to rising development costs and extended project timelines, regulatory hurdles, supply chain disruptions, and labor shortages also contribute to these challenges. The Company's long-range forecasting and supply risk management solution compresses time-to-market for new customer deployments, highlighting the benefits of its portfolio management approach (see Figure 2)<sup>2)</sup>.

Figure 2—Procurement Portfolio Investment Risk Minimization. Long-term Power Demand Forecast for Supply Chain



AI: artificial intelligence, N-MTLR: neural multi-task logistic regression

In Figure 2, the focus shifts to long-term procurement planning. The transformer-based forecasting with agentic AI captures market signals to optimize deal closure. The system analyzes current usage patterns and market trends to predict capacity needs, enabling proactive equipment procurement and minimizing supply chain delays. By integrating demand forecasting with supply chain management, colocation providers can effectively plan future capacity while reducing investment risks through data-driven decision making.

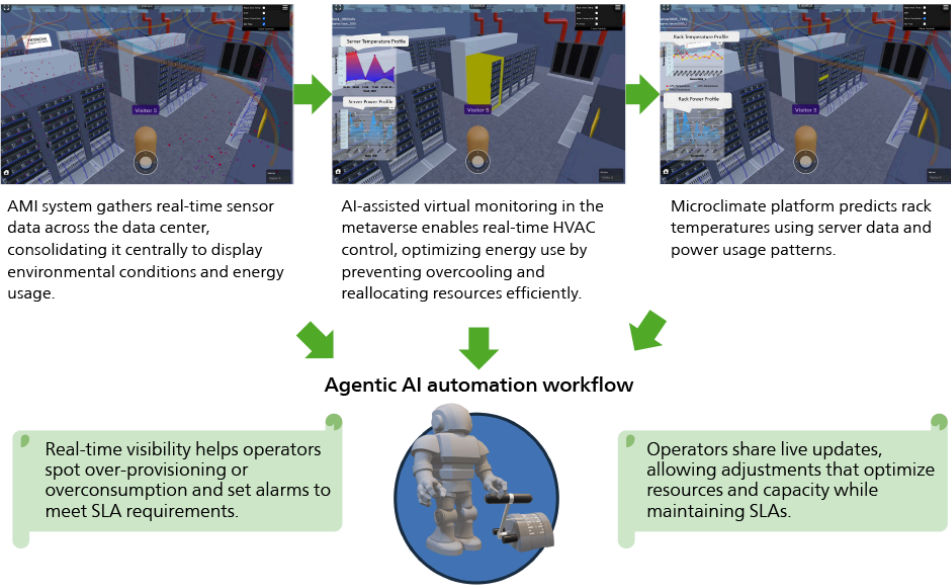
3.2. Current Capacity Depletion Rate

Colocation providers face significant challenges in managing current capacity, including blind utility control, overconsumption, stranded capacity, fragmented resource allocation, and inflexibility due to SLA constraints. To address these issues, a solution integrating Advanced Metering Infrastructure (AMI) and AI-assisted virtual monitoring into a Microclimate platform has been developed (see Figure 3).

The AMI system gathers real-time sensor data across the data center, consolidating environmental conditions and energy usage information centrally. This addresses the blind utility control problem by providing comprehensive visibility. AI-assisted virtual monitoring in the metaverse enables real-time HVAC control, optimizing energy use by preventing overcooling and efficiently reallocating resources. This tackles overconsumption and improves resource allocation. The Microclimate platform uses dynamic energy modeling and real-time sensor data to analyze localized environmental conditions and power usage patterns, enabling optimization of server operations and cooling systems while identifying underutilized capacity.

These components are interconnected through an Agentic AI Automation Workflow, which enables operators to spot over-provisioning or overconsumption in real time and set alarms to meet SLA requirements. The system allows operators to share live updates and make adjustments that optimize resources and capacity while maintaining SLAs. This integrated approach provides a powerful solution for managing current capacity challenges in colocation data centers, ensuring efficient operations and customer satisfaction.

Figure 3—Advance Metering Infrastructure with Microclimate Platform and Intelligent HVAC Controls



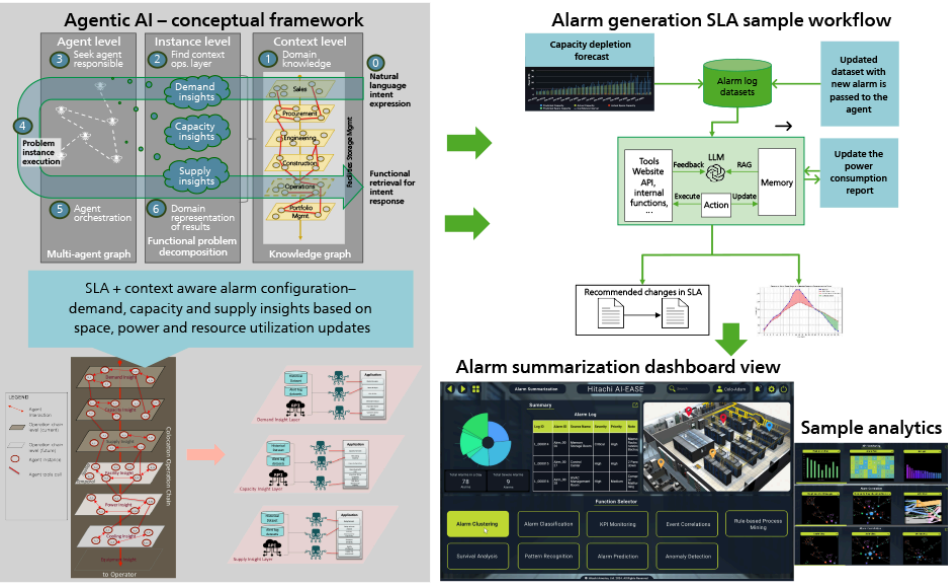
AMI: advances metering infrastructure

The illustration shows an integrated data center management system with three components connected by the Agentic AI Automation Workflow: Advanced Metering Infrastructure (AMI) for real-time sensor data, AI-assisted virtual monitoring for intelligent HVAC control, and the Microclimate platform for predictive analytics. Together, they provide operators with real-time visibility, resource optimization, and automated management while ensuring SLA compliance.

3.3. Total Capacity Utilization and Management

Portfolio management for colocation providers requires a comprehensive view of capacity across current and new facilities, enabling strategic decisions about resource allocation and risk management. The Agentic AI Framework (see Figure 4) demonstrates how this portfolio approach automates and optimizes capacity management through a three-tiered architecture.

Figure 4—Agentic AI-powered Task-based Alarm Summarization and SLA recommendation for Portfolio Quota Management



API: application programming interface, RAG: retrieval augmented generation

Figure 4 illustrates the framework’s portfolio management capabilities through two key implementations: alarm management and quota management. At the Agent Level, the system orchestrates problem-solving tasks across facilities. The Instance Level extracts demand, capacity, and supply insights, while the Context Level builds a knowledge graph for intelligent decision-making. This hierarchical approach enables portfolio-wide optimization rather than facility-specific solutions.

The framework’s portfolio approach is demonstrated through its Alarm-to-SLA Sample Workflow, where it generates alarms based on SLA violations and capacity depletion forecasts, which are summarized in a dashboard that provides operators with real-time visibility into system performance. This allows for proactive adjustments to meet SLAs, significantly reducing resolution times and improving resource allocation. The Alarm Summarization Dashboard provides operators with portfolio-wide visibility through features like alarm clustering and pattern recognition. The bottom portion shows how the framework integrates demand, capacity, and supply insights across facilities based on space, power, and resource utilization updates.

This portfolio-based implementation enables colocation providers to optimize resource allocation across multiple facilities while maintaining SLA compliance, transforming traditional facility-specific management into a comprehensive, data-driven portfolio optimization approach.

## 4. Conclusions

The portfolio management vision, centered on new, current, and total capacity segments, forms the foundation for efficient colocation data center management. This approach enables providers to address challenges and opportunities within each segment, optimizing resource allocation, SLA management, and customer satisfaction. Our key technologies supporting this vision include:

- (1) The Metaverse for Facility Digitization provides a collaborative platform for colocation operators and customers to optimize current capacity and plan for new capacity.
- (2) The Microclimate Platform ensures optimal environmental conditions across all capacity segments, enhancing operational efficiency and sustainability.
- (3) Large Language Models and Agentic AI enable intelligent decision-making and automation, particularly in managing total capacity and forecasting future needs.

Integration of these technologies within the portfolio management framework has demonstrated significant improvements such as increased utilization rates, reduced energy costs, accelerated time-to-market, boosted revenue per square foot, and enhanced customer retention. Our demand forecasting system offers long-range visibility, enhancing operators' ability to plan proactively with utilities and customers through our collaborative metaverse platform. This proactive approach to capacity planning and resource optimization allows colocation providers to defer capital expenditures on new facilities, fostering sustainable growth and long-term financial success.

Looking ahead, Hitachi will continue developing R&D activities and testing these innovations through customer PoC projects. We aim to build expertise in digital solutions for data centers alongside our existing product line. As the digital landscape evolves, our portfolio management strategy positions colocation facilities as the cornerstone of next-generation digital infrastructure, ready to meet the demands of a data-driven world.

REFERENCES
<p>1) D. A. Smith et al., “Croquet-a collaboration system architecture,” First Conference on Creating, Connecting and Collaborating Through Computing, 2003. C5 2003. Proceedings, pp. 2–9, IEEE (Mar. 2003).</p> <p>2) N. Zarayeneh et al., “Colocation Datacenter Customer Power Usage Forecasting Using Synthetic Data and Integration of Macroeconomic Indicators,.” 2023 IEEE International Conference on Big Data (BigData), pp. 3453–3457, IEEE (Dec. 2023).</p>

# Hitachi Review

*Hitachi Review* is a technical medium that reports on Hitachi's use of innovation to address the challenges facing society.

The *Hitachi Review* website contains technical papers written by Hitachi engineers and researchers, special articles such as discussions or interviews, and back numbers.

*Hitachi Hyoron*  
(Japanese) website

<https://www.hitachihyoron.com/jp/>



*Hitachi Review*  
(English) website

<https://www.hitachihyoron.com/rev/>



## Hitachi Review Newsletter

Hitachi Review newsletter delivers the latest information about Hitachi Review when new articles are released.