



Optimizing AI infrastructure costs without compromising performance

Turn innovation into measurable ROI with the Cognizant AI Factory with Dell Technologies infrastructure and NVIDIA-accelerated computing. Cost is the biggest barrier to AI

AI has become a defining priority for organizations across every industry. It is reshaping customer experiences, transforming operations and opening new paths to competitive advantage. Yet even as enthusiasm grows, cost remains a major barrier to adoption and scale. While AI is a critical innovation goal for both IT and business leaders, most leaders cite infrastructure as the main constraint, with 50% saying lack of budget is a top AI adoption challenge.¹

As organizations move from pilots to production, they need a way to increase performance without allowing spending to spiral. This is where the Cognizant® AI Factory comes in. By utilizing Dell Technologies infrastructure and NVIDIA-accelerated computing, the platform provides a high-performance, enterprise-ready AI solution that improves efficiency, accelerates outcomes and lowers the total cost of ownership (TCO).

Reducing costs through flexible consumption models

Flexible consumption is one of the most effective ways to control AI infrastructure spending. Unlike a traditional CapEx model, pay-per-use allows you to align investments with demand to improve ROI. This flexibility is especially important as AI workloads fluctuate and mature.

The Cognizant AI Factory offers cloudlike agility with a pay-as-you-go consumption model, delivering two to three times the cost efficiency of traditional cloud AI-managed services.² By paying only for the GPU or token consumption required, you can avoid overprovisioning, reduce idle capacity and maintain predictable spending while still benefiting from high-performance on-premises infrastructure.

Improving financial efficiency with intelligent operations

Cognizant's autonomous IT operations and FinOps capabilities help optimize resources, plan capacity intelligently and maintain predictable costs as AI usage grows. These capabilities support seamless operations, cost transparency and strong governance.

Key benefits include visibility and cost control through real-time dashboards for budgeting, cost attribution, forecasting and predictive analytics, supported by show-back models and dynamic resource optimization through model tiering, batching and serverless refactoring.

Together, these capabilities can help you reduce waste, improve utilization and maintain financial discipline as AI scales.

Eliminating the tradeoff between performance and cost

With the Cognizant AI Factory, you don't have to choose between performance and cost efficiency. The platform combines high-density CPU and GPU configurations with optimized software and expert tuning to deliver stronger performance at lower cost.

Combining all the above benefits, organizations have reduced host count by up to 85% through higher-density systems, which also lowers power consumption by 56%. Overall costs can be reduced by more than 25% using infrastructure that's more cost-effective than on-demand cloud infrastructure as a service (IaaS). Parallel processing reduces spend by 30%, while improved efficiency lowers operating expenses and accelerates time to market.²

¹Gartner, [Gartner Survey Finds 54% of Infrastructure & Operations Leaders Are Adopting AI to Cut Costs](#), October 2025.

²Source: Internal Cognizant testing results.

At the same time, the solution delivers significant performance improvements, including 30% faster AI processing speeds, reduced latency and up to 86% faster job runs.² This combination of efficiency and performance delivers better ROI and faster results for AI pilots and production workloads.

A more cost-effective future for enterprise AI

By utilizing Dell Technologies infrastructure, advanced NVIDIA-accelerated computing and Cognizant expertise in operationalizing AI, the Cognizant AI Factory makes deployment simpler, management more efficient and scaling less complex. The joint solution is purpose-built to tackle the most pressing barriers enterprises face, such as siloed IT environments, escalating infrastructure expenses and heightened security demands.

The Cognizant AI Factory provides a practical path to scaling AI workloads with consistent high-performance and enterprise-grade governance, enabling organizations to reliably move from isolated experiments to production-ready AI solutions across the business. By uniting robust infrastructure, adaptable consumption models, enterprise-grade governance and security, and Cognizant's deep domain knowledge, the solution establishes a resilient platform for enterprise AI. This foundation enables organizations to accelerate innovation, improve reliability in model deployment and translate AI initiatives into tangible business outcomes.

[Learn more](#) 

²Source: Internal Cognizant testing results.

Author

- Hemant Patade**

AVP, CIS

Hemant.Patade@cognizant.com



Cognizant (Nasdaq: CTSH) is an AI Builder and technology services provider, building the bridge between AI investment and enterprise value by building full-stack AI solutions for our clients. Our deep industry, process and engineering expertise enables us to build an organization's unique context into technology systems that amplify human potential, realize tangible returns and keep global enterprises ahead in a fast-changing world. See how at www.cognizant.com or [@cognizant](https://twitter.com/cognizant).

World Headquarters

300 Frank W Burr Blvd
Suite 36, 6th Floor
Teaneck, NJ 07666, USA
Tel: (201) 801-2333

European Headquarters

280 Bishopsgate
London
EC2M 4AG
England
Tel: +44 (0)1 020 7297 7600

India Corporate Office

Siruseri-Software Technology Park of India (STPI)
SDB Block – Ground floor north wing
Plot No H4, SIPCOT IT Park
Chengalpattu District
Chennai 603103, Tamil Nadu
Tel: 1800 208 6999

APAC Headquarters

1 Fusionopolis Link, Level 5
NEXUS@One-North, North Tower,
Singapore 138542
Tel: + 65 6812 4000

© Copyright 2025–2027, Cognizant. All rights reserved. No part of this document may be reproduced, stored in a retrieval system, transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the express written permission of Cognizant. The information contained herein is subject to change without notice. All other trademarks mentioned herein are the property of their respective owners.