![cognizant | nVIDIA]

Navigating the AI era

Accelerating innovation with advanced AI, digital twins, and omniverse solutions

Unlocking limitless possibilities with Omniverse enterprise and AI-driven solutions

Unlock the AI value with Cognizant

At Cognizant, we partner with global clients to help them scale the AI value efficiently. Leveraging our deep domain experience and a comprehensive AI ecosystem, we provide a robust infrastructure, advanced data models, and agent development powered by proprietary platforms and accelerators. We have active client engagements across various industries, enabling growth and driving business transformation through cutting-edge AI solutions.
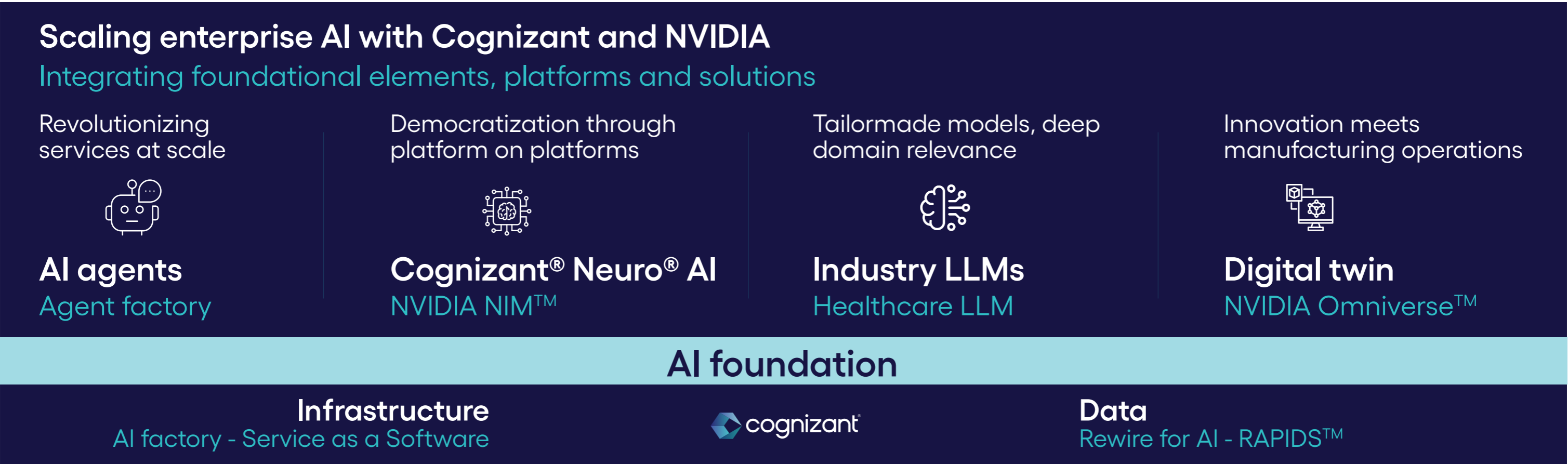
Our mission is to accelerate the cross-industry adoption of AI technology, driving innovation and efficiency across key areas. NVIDIA technology plays a pivotal role in our AI offerings.

Comprehensive AI solutions
As a full stack player, our solutions portfolio spans across

- Enterprise AI agents
- Tailored industry large language models
- Infrastructure for AI
- Physical AI

Our offerings portfolio

Scaling enterprise AI with Cognizant and NVIDIA
Integrating foundational elements, platforms and solutions

Revolutionizing services at scale	Democratization through platform on platforms	Tailormade models, deep domain relevance	Innovation meets manufacturing operations
AI agents	**Cognizant® Neuro® AI**	**Industry LLMs**	**Digital twin**
Agent factory	NVIDIA NIM™	Healthcare LLM	NVIDIA Omniverse™

AI foundation

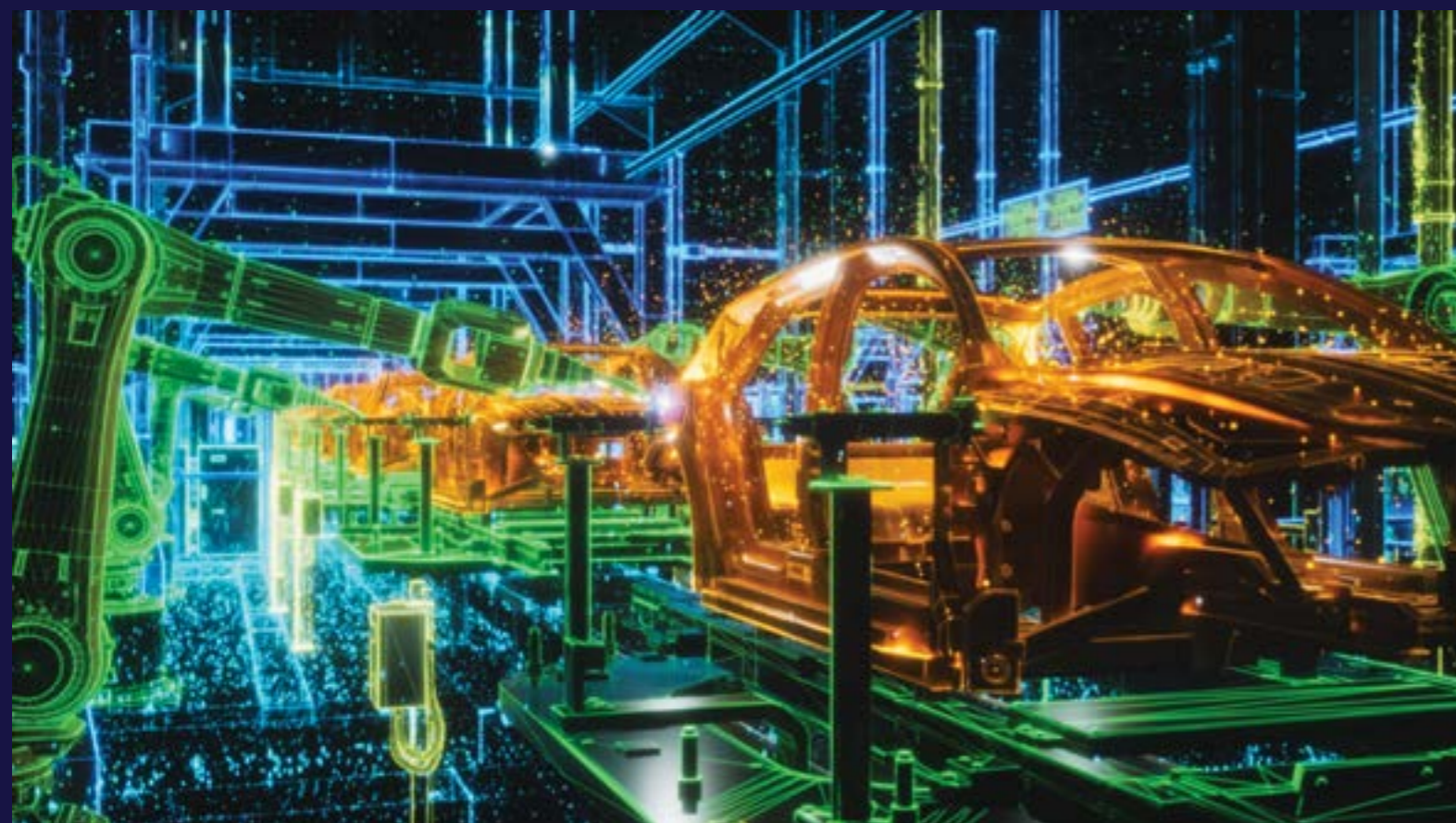| **Infrastructure**		**Data**
AI factory - Service as a Software		Rewire for AI - RAPIDS™

Digital twin

Discover the transformative potential of our digital twin solutions, integrating agentic AI technology into manufacturing operations through the power of NVIDIA Omniverse.

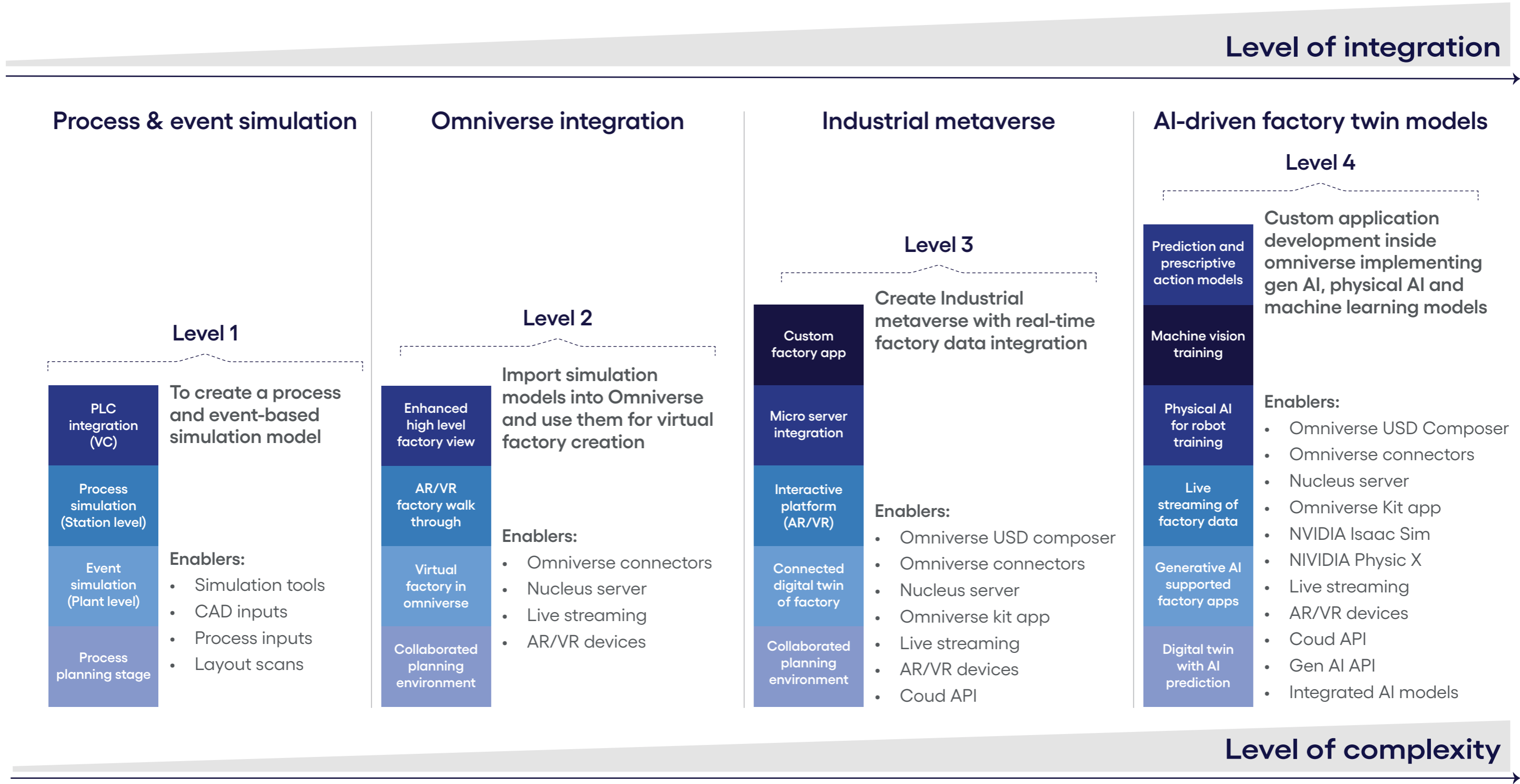## Our offering tackles critical business challenges and drives constant innovation.

Business challenges addressed

1 **Integration of multiple systems and data sources**
Our digital twins seamlessly combine simulation results with T, OT and IoT systems with photo-realistic visualizations

2 **Performance impact**
Our digital twin offers a new way to deliver impactful use cases in asset performance, plant layout optimizations, digital gemba, planning automation, etc.

3 **Scaling physical models**
Our digital twins deliver physical AI at mega-scale through synthetic data generation at speed

4 **Enhanced collaboration, security and governance**
Our digital twins enrich collaboration across teams to drive faster decision-making

Solution framework

Our comprehensive framework integrates state-of-the-art technology to enhance planning, collaboration, and monitoring capabilities within your operations.

Level of integration

Process & event simulation

Omniverse integration

Industrial metaverse

AI-driven factory twin models

Level 4

Custom application development inside omniverse implementing gen AI, physical AI and machine learning models

Level 3

Create Industrial metaverse with real-time factory data integration

Level 2

Import simulation models into Omniverse and use them for virtual factory creation

Level 1

To create a process and event-based simulation model

Process & event simulation column:
- PLC integration (VC)
- Process simulation (Station level)
- Event simulation (Plant level)
- Process planning stage

Enablers:
- Simulation tools
- CAD inputs
- Process inputs
- Layout scans

Omniverse integration column:
- Enhanced high level factory view
- AR/VR factory walk through
- Virtual factory in omniverse
- Collaborated planning environment

Enablers:
- Omniverse connectors
- Nucleus server
- Live streaming
- AR/VR devices

Industrial metaverse column:
- Custom factory app
- Micro server integration
- Interactive platform (AR/VR)
- Connected digital twin of factory
- Collaborated planning environment

Enablers:
- Omniverse USD composer
- Omniverse connectors
- Nucleus server
- Omniverse kit app
- Live streaming
- AR/VR devices
- Coud API

AI-driven factory twin models column:
- Prediction and prescriptive action models
- Machine vision training
- Physical AI for robot training
- Live streaming of factory data
- Generative AI supported factory apps
- Digital twin with AI prediction

Enablers:
- Omniverse USD Composer
- Omniverse connectors
- Nucleus server
- Omniverse Kit app
- NVIDIA Isaac Sim
- NIVIDIA Physic X
- Live streaming
- AR/VR devices
- Coud API
- Gen AI API
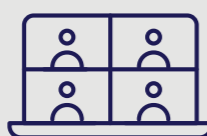- Integrated AI models

Level of complexity

Use cases

1. Enhancing planning and visualization
Integrate 3D plant and scan data with detailed digital representations of facilities

2. Facilitating collaboration and training
Use AR/VR for remote collaboration, expert consultations, and immersive training experiences
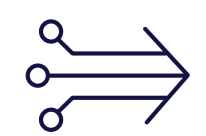
3. Shopfloor data monitoring
Real-time process KPI monitoring through integrated shopfloor databases within Omniverse simulation models

4. Optimized operations
AI-driven simulations to enhance workflows, improve efficiency, and reduce operational costs

5. Enabling advanced use cases
Simulate the impact of AGVs, co-bots, computer vision, and other technologies in real-world scenarios

6. Promoting sustainability
Identify energy-saving opportunities and track performance to support sustainability goals

Business impact

Innovative integrations
Leverage NVIDIA Omniverse and advanced analytic tools for asset health modeling, robotics, and AMRs, driving cutting-edge manufacturing solutions

Collaborative labs
Utilize Omniverse to enhance digital twin capabilities, augmented training, and AI-driven decision-making, applicable in manufacturing plants and smart product development

Scalability and security
Architect scalable solutions with robust security guardrails for advanced analytics in manufacturing and smart products

Optimizing productivity
Use LLMs, RAG, and synthetic data generation to boost operational efficiency, reduce downtime, and enhance sustainability in manufacturing and client building management

Case study: Large HVAC manufacturer

Our successful implementation of digital twin solutions with a large HVAC manufacturer, showcases the tangible benefits and innovations achieved.

Connect with our subject matter experts

Sharath Prasad:

Daniel Mcgrath:

Enterprise AI powered by Cognizant® Neuro® AI on NVIDIA NIMS

Platform on platforms

Enhance your business capabilities with the Cognizant Neuro® AI, powered by NVIDIA NIMS. This advanced multi-agent orchestration platform enables seamless identification of opportunities, creation of prototypes, and development of AI-driven decision-making processes across siloed operations.

By leveraging specialized AI agents, Neuro® AI not only enhances performance but also generates new revenue streams, optimizing decisions for measurable real-world impact.

Introducing the Neuro® AI multi-agent accelerator
a revolutionary no-code/low-code development framework

This platform includes an array of pre-built agent networks that allow enterprises to rapidly create, customize, and scale multi-agent systems. These pre-built networks provide a robust starting point for addressing various enterprise functions and industry-specific processes such as sales, marketing, and supply chain management. Additional agent networks can be swiftly developed using natural language descriptions to match different scenarios and client use cases.

The framework offers the flexibility to integrate third-party agent networks and is built on NVIDIA NIM models and endpoints, incorporating LangChain-based substitute agents from NVIDIA blueprints.

Business challenges addressed

Cumbersome processes

The current methods for extracting appropriate medical codes are labor-intensive and prone to errors. Manual review of clinical documentation by coding analysts is both time-consuming and susceptible to human error.

Inadequate solutions

Existing general-purpose language models do not meet the specific needs of the healthcare industry, often failing to capture the complexity and specificity of medical terminology, resulting in inconsistent and inaccurate code extraction.

Manual interventions

The need for manual verification and correction of suggested codes further complicates the process, increasing the workload for coding analysts and introducing additional potential points of failure.

Business impact

Augmented decision-making

The Neuro AI Decisioning platform enables quick prototyping and creation of decision-making use cases using a multi-agent system and Gen AI, running on NIM endpoints.

Rapid development

The Neuro AI Multi-agent Accelerator Platform, powered by NVIDIA NIMS/NEMO, facilitates the quick production and interoperability of multiple agents across various enterprise functions.

No-code/low-code framework

This framework allows for the building of prototype agent networks in hours rather than days.

Pre-built networks

Utilize tried and tested use case networks to reduce implementation time and technical risk.

Interoperability

Seamlessly integrates with third-party networks using APIs and natural language, such as Salesforce, Agentforce, and ServiceNow.

Connect with our subject matter expert

Babak Hodjat:

AI agent factory: accelerating enterprise scale agentification

Cognizant WorkNEXT™ AI proudly presents its advanced genAI portfolio of solutions, meticulously crafted to deliver a comprehensive "**Total Experience**" for end users.

This innovative portfolio features an AI platform with an Enterprise AI Bus, seamlessly integrating cloud, infrastructure, and workplace components, ready for deployment on any cloud and platform. Leveraging cutting-edge AI capabilities, the platform enables multilingual conversational experiences across diverse channels, integrated with underlying datasets and systems of engagement. This integration ensures observability, correlation, diagnosis, integration, and auto-remediation.

Moreover, the platform adeptly identifies knowledge gaps, offers intelligent recommendations, and provides active listening capabilities to filter noise from infrastructure and workplace environments.

This facilitates effective root cause analysis, real-time solution recommendations, and automated execution, resulting

End-user productivity enhanced by more than

30%

Automation efficiency boosted by over

45%

Use cases

Customer service enhancement through AI-driven support systems

1

Automated incident management and resolution

2

Intelligent data analysis and insights generation

3

Personalized user experiences through advanced AI capabilities

4

Business challenges addressed

Ensuring high-quality AI Models
Tailored to specific enterprise needs, capable of handling complex tasks effectively

Seamless integration
Integrates into existing systems and workflows to minimize disruption

Enhancing operational efficiency
Through optimal performance and reliability of AI agents

Continuous improvement
Based on metrics to ensure agents evolve and improve over time

Business impact

High-quality AI models

Tailored to specific enterprise needs, ensuring they can handle complex tasks effectively

Seamless integration

Integrates into existing systems and workflows to minimize disruption

Operational efficiency

Ensures optimal performance and reliability of AI agents

Continuous improvement

Based on metrics, ensuring agents improve and evolve over time

Connect with our subject matter expert,
Saswata Kundu:
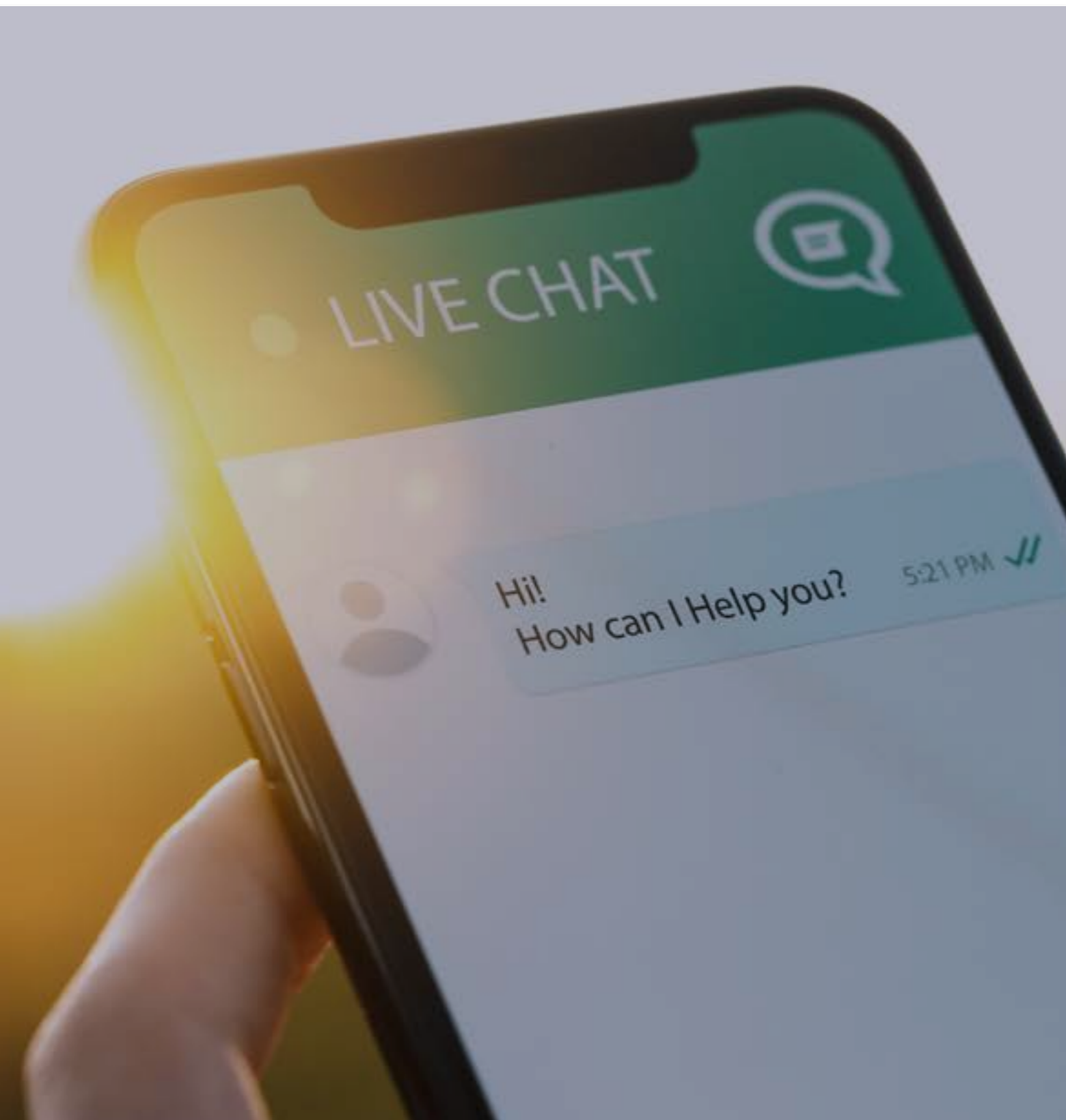
Transforming healthcare administration

Industry language model for medical code extraction

We proudly present our cutting-edge healthcare language model, meticulously fine-tuned on specialized data to revolutionize accuracy in medical code extraction.

This model leverages an extensive dataset of medical records, including clinical notes and discharge summaries, to ensure precise and efficient code extraction.

Business challenges addressed

Cumbersome processes

Traditional methods for extracting medical codes are labor-intensive and fraught with errors. Manual review of clinical documentation by coding analysts is time-consuming and prone to human error

Inadequate solutions

Current solutions, including general-purpose language models, fall short in meeting the unique demands of the healthcare sector. They often miss the complexity and specificity of medical terminology, leading to inconsistent and inaccurate code extraction
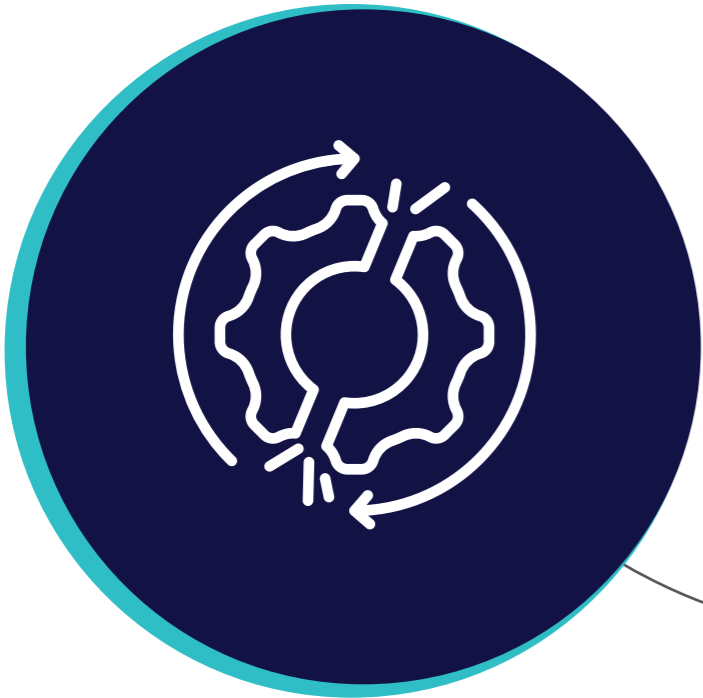
Manual interventions

The need for manual verification and correction of suggested codes complicates the process further. This increases the workload for coding analysts and introduces additional potential for errors

Solution framework

Our fine-tuned healthcare language model offers a robust framework for automating and optimizing the medical code extraction process, ensuring accuracy and efficiency at every step.

1 Data preparation

- Collection & injection of medical domain taxonomy, guidelines, and ICD code descriptions
- Data curation using NeMo curator
- Prepared redacted 2K+ clinical notes & generated 10K+ synthetic data using NeMo curator

2 Benchmarking

- Selection of medical domain and open-source models
- Base-lining using different approaches such as RAG, CoT, and few-shot learning
- Define and evaluate models

3 Model fine-tuning

- Medical domain adaptation using medical terminologies, guidelines, and ICD-10 coding
- Fine-tuning model with medical code extraction tasks utilizing synthetic data
- Leveraged Parameter-Efficient Fine-Tuning (PEFT); Low-Rank Adaptation (LoRA) technique

4 Model inference

- Model deployment, provisioning & configuration using NIM container
- Medical code extraction business app build and deployment
- Model monitoring, performance & scalability

Cognizant's value

• Experienced SME base – 300+ AAPC certified coders with a proven track record on sensitive submissions (e.g., to CMS) • Access to rich and diverse datasets – 2000+ datasets and ontologies	• TriZetto product and healthcare domain expertise in model validation • Leverage Cognizant's benchmarking framework on NVIDIA Evaluator	• Strong technical expertise in fine-tuning and customizing language models across domains • 1000+ resources trained and certified across NVIDIA, GCP, and generative AI	• Agentic modular architecture for easy scalability • Seamless integration with enterprise apps and TriZetto product ecosystem

NVIDIA tech stack

NeMo Curator	NeMo Evaluator	NeMo Customizer, Evaluator	NeMo Retriever, Guardrails, NIM (LoRA)

Business impact
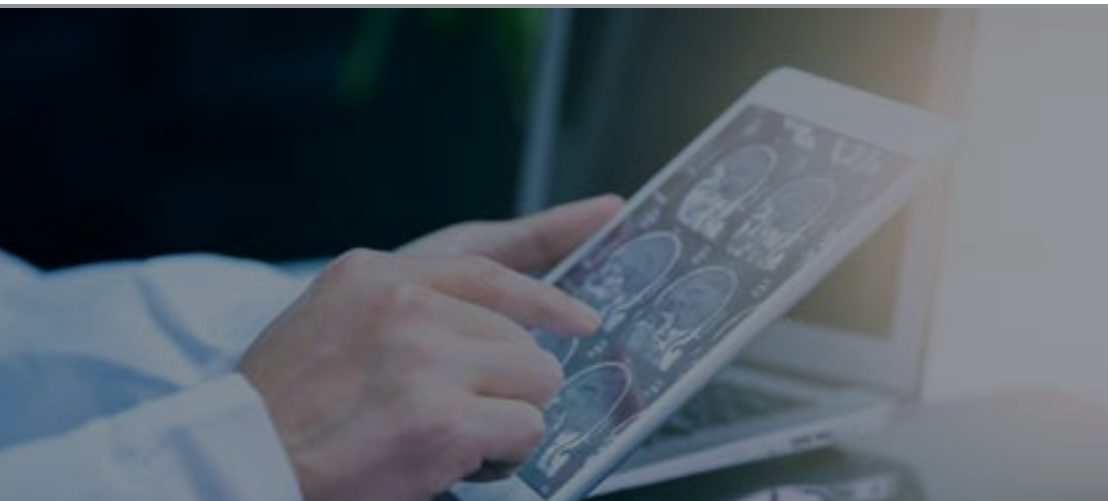
Experience a transformative leap in healthcare efficiency with our fine-tuned language model, designed to meet the highest standards of precision and performance. Join us in revolutionizing medical code extraction and setting new benchmarks in the healthcare industry.

Improved accuracy

Our fine-tuned model significantly enhances the accuracy of medical code extraction, minimizing errors and ensuring reliable billing and reporting

Efficiency gains

By automating code extraction, we streamline workflows, reduce the burden on coding analysts, and expedite the overall process

Enhanced patient care

Accurate coding is crucial for optimal patient care and treatment planning. Our solution ensures that patients receive appropriate care and services

Cost savings

Reducing errors and manual interventions translates to substantial cost savings for healthcare organizations, both in operational efficiency and by avoiding financial penalties from incorrect billing

Standard-ization

Our model fosters standardization, reducing subjective interpretations and ensuring consistency and reliability in medical coding

Connect with our subject matter expert,
Rachit Gupta:

AI Infrastructure -
AI Factory -
Service as a Software

Welcome to the future of business innovation with Cognizant's AI Foundation. We are dedicated to building accessible, trustworthy, and future-ready infrastructure foundations designed to help businesses become data-centric, driving growth through AI-powered decision-making.

Our expertise leverages generative AI capabilities, large-scale model migration, and robust data quality frameworks to transform your business outcomes.

Partnering with NVIDIA for unmatched AI solutions

In collaboration with Nvidia, Cognizant offers secure and fully managed infrastructure, ensuring that AI models can be run wherever they are most impactful - whether in the cloud, data center, or at the edge. This collaboration ensures a seamless experience, enabling your business to harness the full potential of AI.

Overcoming business challenges

Lack of internal expertise

Our solutions bridge the gap, providing the necessary AI expertise to seamlessly integrate transformative technologies for impactful business outcomes

Initial investment

Despite the substantial initial investments, we ensure that the resources allocated deliver meaningful business impact through cost-effective cloud-based AI platforms

Time consuming

We streamline the adoption process, minimizing delays and bottlenecks to ensure timely achievement of desired outcomes

Unrealized business value

We help businesses explore opportunities across the organization, ensuring successful integration and scalability of AI projects to realize true value

Data sovereignty

Complying with diverse regulations, we ensure data remains within national borders and is managed securely, maintaining control and security

Use cases

Generative AI
Curate content, enhance personalization, realign market models, and build use case-specific applications

Accelerated data processing
Optimize data processing, ETL, machine learning, deep learning, CNN, graph analytics, time series analysis, and visualization

HPC, digital twin, and digital assistance
Run any business or vertical use case on our robust platform assistance

Business impact

Cost savings
Running AI workloads on-premises is **50-60% cheaper t**han **cloud** (TCO for 5 years) and delivers **up to 30% faster processing times** compared to cloud-based GPU services

Flexible & scalable platform
Our market-leading NVIDIA AI Enterprise solution provides a scalable platform deployed as per client requirements across cloud, on-premises, and edge

Accelerating AI
Providing clients the option to bring their own model or leverage pre-built LLMs, custom models developed by Cognizant, and industry solutions and use cases built across the AI lifecycle, data analytics, and high-performance computing (HPC)

Ensuring security
Implementing robust security measures to protect sensitive data, ensuring businesses can use cloud resources without compromising security

Improved experience and simplified operations
Transforming the way clients consume services by building a single pane of glass view for customers with a common set of tools to simplify management and visibility

Connect with our subject matter expert,
Hemant Patade:

Cognizant NVIDIA RAPIDS™ powered by GPU-based infrastructure

Drive **3X** performance at **one-third** of the cost

By transitioning clients to GPU-based infrastructure, we will lay the groundwork for implementing exciting new generative AI use cases in critical business functions.

NVIDIA RAPIDS™ is a part of NVIDIA CUDA-X, an open-source suite of GPU-accelerated data engineering, data science and AI libraries with APIs that match the most popular open-source data tools.

With GPU-driven NVIDIA RAPIDS™ libraries, Cognizant can help businesses significantly reduce computing costs and accelerate workloads to extract more value from their cloud investments, revolutionize the way data engineers and data scientists process and analyze large datasets, delivering unprecedented speed and efficiency.

Cognizant aims to offer end-to-end services to resolve our clients' today's data modernization problems, from modernization to optimization.

Overcoming business challenges
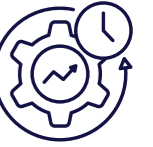
As enterprises increasingly migrate their systems to the cloud, they encounter significant challenges in controlling and optimizing their cloud expenditures. The rapid growth of data and the need for real-time analytics have led to escalating costs and performance

Traditional CPU-based methods can fall short in handling large-scale data processing efficiently, leading to higher potential operational costs and slower organizational decision-making.

Cognizant believes that our clients need to rewire their data ecosystem for AI to embark on next wave of data modernization which is fixing biggest challenges that clients face today with their data, where it's data quality, data useability, data discoverability, data interoperability, data accessibility, faster data analysis and insight generation, and all of these have to be done while managing cost & performance.

Solution framework

Let Cognizant accelerate your Spark workloads with NVIDIA RAPIDS™

Qualification
Check how many of your Spark logs can be accelerated with NVIDIA RAPIDS to achieve cost savings

Bootstrap
Optimize RAPIDS Spark configs based on GPU cluster shape

Profiling
Tune RAPIDS Spark congigs based on initial job run leveraging Spark event logs

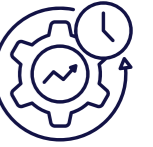## No code change required
(only compute & pipeline reconfiguration)

RAPIDS

Getting everything together   cognizant®   Driving acceleration, deriving value

Use cases

Accelerate your Spark workloads with NVIDIA RAPIDS™

Accelerate through 4-6 weeks of proof of concept. Select Spark logs, run them on NVIDIA RAPIDS™, and further optimize them with finetuning on profiling toolcase-specific applications

Databricks Spend Management powered by NVIDIA RAPIDS™

Define business case, decision framework, deploy spend management in 4-5 weeks POC with an aim of CONSULT + ENGAGE+ ACCELERATE

ETL Modernization with Cognizant Data and Intelligence Toolkit & NVIDIA RAPIDS™

Try a proof of concept for accelerating Spark codes after Modernizing your ETL ecosystem

Connect with our subject matter expert,
Diptesh Singh

Business impact

Drive **3X performance** at **one-third of the cost** with Cognizant NVIDIA RAPIDS™ offerings

Cognizant proprietary benchmarks show

| **2.5x** acceleration in data engineering and data science workloads | **66%** reduction in computing costs compared with CPU processing |

When combined with the 50-80 percent automation in cloud modernization processes driven by Cognizant's accelerators, Cognizant and NVIDIA together help businesses

- **Accelerate their data processing capabilities**

- **Optimize costs**

- **Drive innovation**

AI-native CPE

Transforming the connected living experience through NVIDIA accelerated edge computing and Cognizant device engineering

Cognizant's device engineering offers end-to-end services for cable operators, telcos, ODMs and OEMs, providing accelerated value creations, simplified operations and transformations through innovations. Our solutions, accelerators and rich capabilities span across design & development, maintenance & migrations, automations & validations, release management, deployments, triage & support services, providing comprehensive coverage in the Networking, CPU/NPU/SoC, embedded-Linux, firmware, middleware and application stacks.

Cognizant device engineering offerings are now further expanded and amplified by integrating NVIDIA's groundbreaking accelerated edge computing, AI platforms and technologies.

In collaboration with NVIDIA, Cognizant device engineering presents a new class of broadband-CPE devices – industry's first 'AI-native RDK CPE' for an AI-driven connected living and experiences of tomorrow. These AI-native CPEs leverage NVIDIA's Jetson-Orin Embedded class GPU+CPU based edge computing platform, speech AI, generative AI, RDK & open-source technologies.

We envision these transformative AI-native CPEs to be the steppingstone for operator's journey towards the NextGen managed CPE fleets, effectively delivering versatile network-edge workloads and range of AI-powered operator & customer-experience use cases.

Business challenges addressed

Redefining and transforming customer experience

Through Speech AI combined with generative AI, the 'AI-native CPE' democratizes the technical nuances currently mandated by the traditional CPEs. **Through multilinguistic speech and personalized voice-assistant on-demand, subscribers are empowered to express intents and self-serve their fulfilments via natural languages.**

OpEx savings and scaled operations for operators

Uptick in customer driven personalized self-services for CPE installs & service impacting triages to hugely benefit operational logistics and measures such as professional installs, expensive truck-rolls and customer-churns. This is also complemented by AI-driven SON (Self-Organizing-Network) and telemetry summarization related measures aboard the AI-native CPE.
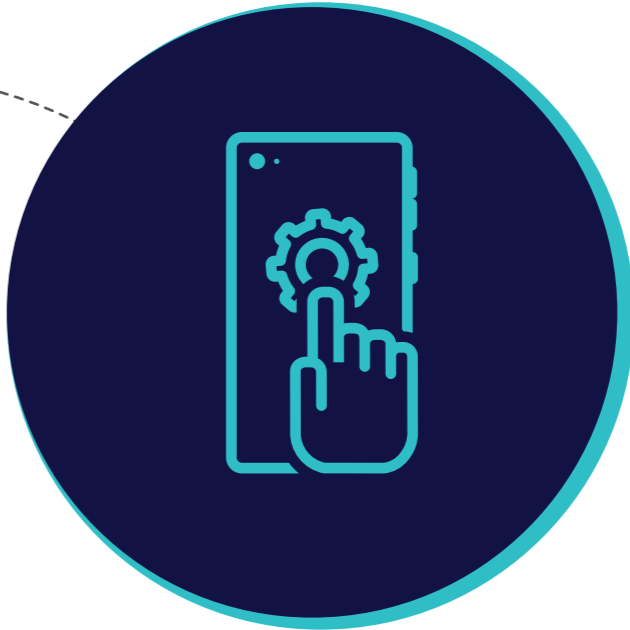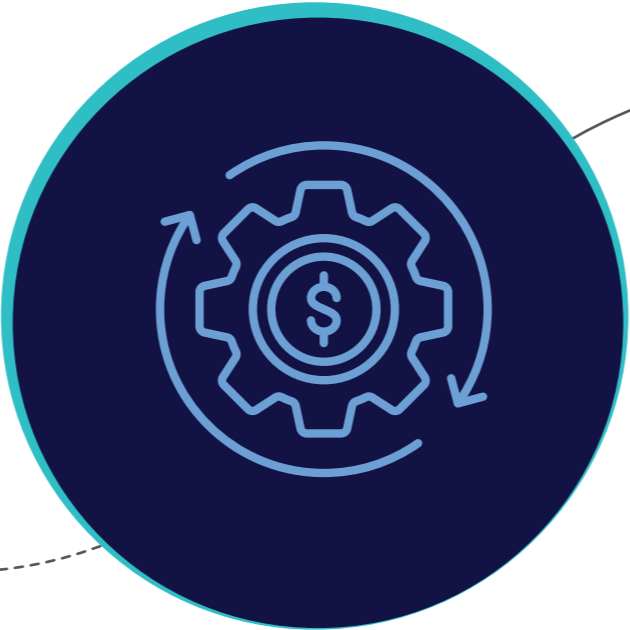
Future readiness, competitiveness, flexibility and cost efficiency

With rapid developments in AI, accelerated computing, networking and open-source technologies, higher intelligence & processing capabilities emerge at the edge leading to more offloading of workloads from CPU to GPU. **Beside first-mover advantage, operators can leverage AI-edge platform to drive new set of value-added services to retail/SMB/enterprise subscribers.**

Use cases
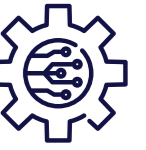
FieldOps & operational efficiency

with AI-native RDK-CPEs can now interact through natural speech-driven interface for new installs, log collections, trouble shooting, predictive monitoring, recovery

Network monitoring

An AI-powered mechanism that enhances the conventional network monitoring and aids in proactive detection of anomalies and self-healing

Self-triage by end user

Customers through natural language interface can now quickly perform the CPE & network setups and self-triages for many of the common root causes without Operator/Field-Ops involvements

Significant cost savings for operators

Reduced incoming call volumes, trouble-tickets & truck-rolls for operator's customer care

### Network optimization & efficiency

Reduction in central-computing loads and carbon footprints with Edge intelligence, processing with GPU powered CPEs

Solution framework

Cognizant's solution framework diligently integrates NVIDIA's Jetson-Orin embedded class GPU+CPU based edge computing platform with speech AI, generative-AI and RDK open-source technologies.

Leveraging edge-centric technologies and solutions, AI-native CPE addresses the security & privacy from get-go.

Further, to support operator usecases involving the new-installs and service-down scenarios, our modular solution also encompasses an innovative design that provides 100% offline speech AI (ASR, STT & TTS) availability to serve subscribers with personalized on-demand voice-assistant functions.
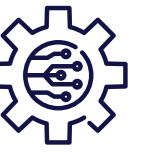
Business impact

Innovative integrations

NVIDIA's AI platforms & technologies and Cognizant's Device Engineering capabilities are leveraged to create cutting-edge solutions for operators & customers to enhance operational efficiency, cost reductions alongside rich customer experiences.

OpEx savings

The introduction of AI-native CPE devices in networks leads to significant cost savings in customer support and truck-roll dispatches.

Rich customer experience

Simplified interactions through speech-AI interfaces for intent fulfillment empower non-technical customers to quickly self-triage common issues, significantly enhancing customer satisfaction.

Competitive advantage, operational efficiency

Reduced central computing loads and carbon footprints through edge intelligence, while premium AI-powered services at the edge enhance efficiency and deliver a competitive

Connect with our subject matter expert,
Shanker Bhat Mangalore:
shanker.bhatmangalore@cognizant.com

Navigating the AI era

Accelerating innovation with advanced AI, digital twins, and omniverse solutions

cognizant®

Cognizant (Nasdaq-100: CTSH) engineers modern businesses. We help our clients modernize technology, reimagine processes and transform experiences so they can stay ahead in our fast-changing world. Together, we're improving everyday life. See how at www.cognizant.com or @Cognizant.

World Headquarters
300 Frank W. Burr Blvd.
Suite 36, 6th Floor
Teaneck, NJ 07666 USA
Phone: +1 201 801 0233
Fax: +1 201 801 0243
Toll Free: +1 888 937 3277

European Headquarters
1 Kingdom Street
Paddington Central
London W2 6BD England
Phone: +44 (0) 20 7297 7600
Fax: +44 (0) 20 7121 0102

India Operations Headquarters
#5/535 Old Mahabalipuram Road
Okkiyam Pettai, Thoraipakkam
Chennai, 600 096 India
Phone: +91 (0) 44 4209 6000
Fax: +91 (0) 44 4209 6060

APAC Headquarters
1 Changi Business Park Crescent
Plaza 8@CBP # 07-04/05/06
Tower A, Singapore 486025
Phone: + 65 6812 4051
Fax: + 65 6324 4051