451 Research
Pathfinder Paper

July 2024

Industry-specific needs call for differentiated edge deployments

Commissioned by

Red Hat

S&P Global
Market Intelligence

Table of contents

Introduction

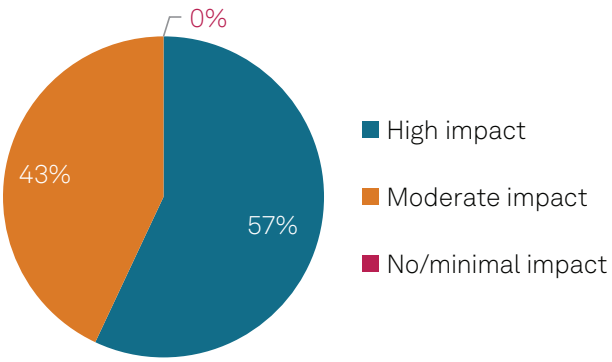
Organizations today encounter great challenges, but also significant opportunities, from the increased pace of IT innovation — driven by the torrents of data generated by digital business processes and unleashed from industry endpoints. Centralized datacenter or cloud computing can address many of those challenges, but not all of them. Edge computing helps by bringing compute resources closer to data sources, reducing latency and enabling organizations to respond swiftly to changing business conditions. Additionally, edge computing supports AI models serving at the edge, improving accuracy and efficiency. By leveraging the capabilities of edge computing, organizations can harness the power of data to drive innovation and optimize operations and resources to deliver exceptional customer experiences.

Crafting an impactful edge strategy presents its own challenges. As digital transformation increasingly becomes a business imperative, edge compute deployments — and the industry use-case requirements that drive them — are proliferating rapidly as well. Yet no two organizations have the same needs, challenges, environments, workloads or staff, and because of that, edge implementations are not one-size-fits-all. For technology teams — IT and, in many cases, OT as well — job one is fully understanding the use cases and business outcomes that need supporting. Just as important for technology decision-makers is selecting the right edge infrastructure and topology to get the job done properly — right-sized and agile today, and accounting for long-term platform management and scaling needs. Choose incorrectly, and edge infrastructure becomes a problem in its own right: unnecessarily expensive; difficult to deploy, secure and operate; and limited in its ability to support distributed, modern, AI-driven applications.

Edge computing arrives

While many new technologies arrive via the hype-train, “modern” edge computing has evolved organically. Most general-purpose, on-premises computing, even in carpeted environments, could rightly be considered edge. Industry-specific operational edges such as retail point-of-sale machines, bank ATMs and manufacturing assembly lines also generate data that, when analyzed, can deliver new insights. Finally, the emergence of today’s edge technologies may be best viewed as a response to the limitations and expense of cloud computing and the need for digitally driven organizations to become as responsive as possible. In short, investment in edge today is driven by innovation and change: 60% of surveyed organizations invest in edge in response to changing IT architectures, while 57% do so to build new capabilities, according to 451 Research’s Voice of the Enterprise: Edge Infrastructure & Services, Economic & Environmental Impact 2024. From the IT department to the C-suite, edge is viewed as a significant enabler of digital transformation (see Figure 1).

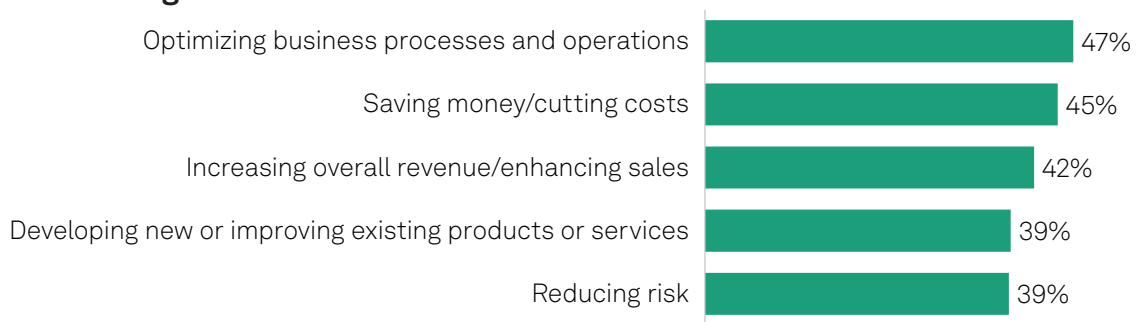
Figure 1: Organizations see edge driving their operational transformation



Q. Overall, how much of an impact do you expect edge computing to have on your organization's IoT and/or operational technology (OT) initiatives and ongoing operations?
Base: Organizations that have adopted or plan to adopt edge computing.
Source: 451 Research's Voice of the Enterprise, Internet of Things, The OT Perspective, Use Cases & Outcomes 2024.

While some form of "edge" computing has existed for decades, the increasingly critical concept of hybrid IT points to a new world in which applications are run both in the cloud, for elasticity and cost reasons, as well as at the edge closer to data sources and users, for latency, performance and security reasons. Organizations need to effectively deploy and manage these distributed environments and orchestrate applications across them. The edge is increasingly where innovation and digital transformation come to life — even more so with increasing AI adoption. The edge allows organizations to make mission-critical use of operational machine and IoT data. It also enables rapid-fire AI inferencing in support of real-time use cases such as computer vision-enabled quality control or cashier-less retail checkout, and it brings compute to disconnected or remote locations, such as remote healthcare facilities, software-defined cars or isolated industrial sites. These capabilities directly address the most important goals of digital transformation, from optimizing operations to cutting costs to reducing risk (see Figure 2).

Figure 2: Edge-driven digital enablement addresses key organizational goals and challenges



Q. Which of the following are drivers of IoT initiatives at your organization? Please select all that apply.
Base: All respondents (n=739).
Source: 451 Research's Voice of the Enterprise: Internet of Things, The OT Perspective, Use Cases & Outcomes 2024.

That said, the edge creates new and unique challenges as well. Connected endpoints and machines generate huge amounts of data that need to be processed at the edge. Because they are outside of a centralized, standardized, easily accessible datacenter, distributed edge fleets can be difficult to manage, protect and update. Because of the unique challenges outside of a datacenter (scarcity of power, connectivity, cooling or space, to name a few), they need to interoperate with a range of legacy systems and on a wide range of hardware, increasing integration challenges. For consistency's sake, application development and deployment at the edge must leverage the same cloud-native approaches as at the core and in the cloud. Many times, there is no on-site expertise or staff at the edge, particularly at companies with large, highly distributed operations such as brick-and-mortar retail. Add to that compounding challenges with connectivity, scalability, data privacy and compliance — the list goes on.

All of this points to a need for organizations to make well-thought-out decisions about their edge infrastructure, matched closely to their IT and business needs (see Figure 3).

Figure 3: Edge computing decision checklist

Organizations deploying edge infrastructure must:

- **Assess workloads and constraints:** Understand the specific needs and limitations of the workloads and devices that will be deployed at the edge.
- **Choose the right deployment topology:** Select the appropriate edge deployment topology based on factors such as device size, connectivity and security requirements.
- **Automate deployment and management:** Utilize tools and practices for automated deployment, scaling and management of containerized applications at the edge to streamline operations.
- **Implement life cycle management:** Develop processes for provisioning, updating and decommissioning edge devices throughout their life cycle.
- **Monitor and manage performance:** Implement monitoring, observability and management tools to track the performance and health of edge devices and applications.

Source: S&P Global Market Intelligence 451 Research.

Industry decision points: Choosing the right edge approach

Organizations must view the edge through the lens of the specific industry and use case to harness its full potential. Edge computing, with its localized processing and storage, offers significant benefits, particularly for industries with unique and demanding requirements. To illustrate this requirement, we'll consider three vertical-focused deployment approaches, showcasing how different industries can leverage edge computing to drive innovation and optimize operations:

- **Remote, distributed edge in the energy sector (single-node):** This section examines how energy providers use edge computing to optimize renewable energy sources, using single-node cluster edge infrastructure.
- **In-store use cases for retail (multi-node):** Edge computing empowers retailers with real-time data analysis and improved customer experiences through multi-node cluster deployments in the brick-and-mortar back office.
- **Multifaceted, distributed edge infrastructure in the industrial sector (mix of single- and multi-node):** Manufacturers deploy multiple edge instances across their facilities, leveraging and holistically managing different node types to enhance efficiency, safety and predictive maintenance.

Single-node edge topology: Utilities/renewable energy

Snapshot	
Edge use cases	Optimizing renewable energy sources, enabling remote monitoring and control, and improving grid resilience.
Business outcomes	Increased energy efficiency, reduced operational costs and improved customer service.
Performance requirements	Real-time data processing, low latency and high reliability.
Location considerations	Distributed, often disconnected existing sites, such as substations, with a goal of minimizing environmental impact and maximizing efficiency.
Preferred edge deployment	Single-node edge infrastructure to support remote and distributed operations.

The energy industry is rapidly embracing renewable energy sources such as solar, wind and geothermal to address climate change concerns and meet sustainability goals. These renewables present unique challenges and opportunities for digital solutions, including the role of edge computing. As utilities and their partners prepare for the greater adoption of renewables, key use cases include remote monitoring and management of renewable assets, optimization of energy generation and distribution, and predictive maintenance to improve uptime and reduce costs.

Edge computing plays a crucial role in supporting the digital use cases associated with renewable energy. Edge devices placed near renewable assets enable real-time data collection, processing and analysis, reducing latency and bandwidth requirements. That's especially important when decisions, such as mapping power demand vs. supply, require quick responses to optimize resources. Edge compute infrastructure can also address issues of performance, security, data sovereignty and compliance, which is essential in the power and utilities sector where there is significant regulatory oversight, coupled with concern about the impact of moving away from traditional energy sources. By leveraging edge computing, energy companies can enhance operational efficiency, improve decision-making and facilitate the seamless integration of renewables into the energy grid.

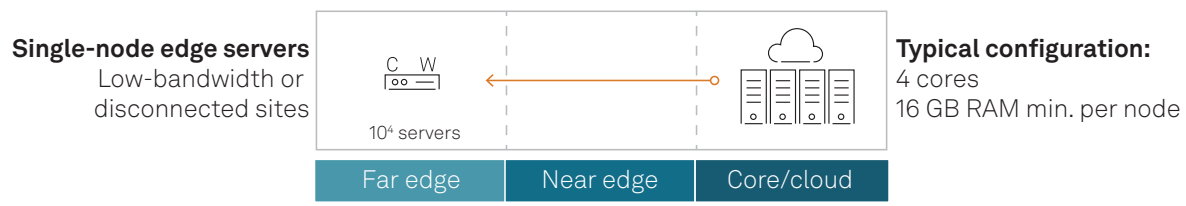
What drives strategic investment for utilities?

Supporting load growth (cited by 38% of firms), followed by decarbonization (26%), integrating renewables (21%) and improving resiliency (15%).

Source: S&P Global Market Intelligence 451 Research survey of global utilities organizations.

The best topology for edge computing in the energy industry involves a combination of multiple single-node deployments. This approach aligns with the distributed nature of renewable energy assets and the need for centralized control and operations. Single-node topologies offer scalability, flexibility and cost-effectiveness, while centralized control ensures efficient management of distributed edge devices. Best of all, single nodes can be extremely small, fitting into the smallest remote sites where space is at a premium. There are also multiple rugged hardware options available for locations where heat, sand, dust or liquids would cause issues for standard server hardware. This topology allows energy companies to optimize their edge infrastructure, balancing the need for local autonomy with the benefits of centralized oversight and coordination — accelerating the operational and economic viability and time frame for renewables adoption.

Figure 4: Single-node edge topology



Source: S&P Global Market Intelligence 451 Research.

Additional industries that can make use of a single-node edge topology approach:

- **Transportation and logistics:** Single-node topologies can support edge computing within vehicles or along transportation routes, enabling real-time data collection, vehicle diagnostics, traffic management and fleet optimization, enhancing safety, efficiency and cost savings.
- **Shipping/cargo:** Ships traverse vast oceans, and cargo facilities often operate in remote or disconnected areas, all while transporting high-value freight. A single-node edge server deployed in such locations can support critical use cases such as vessel performance monitoring, cargo tracking and crew management without requiring network access. The compact size and low power requirements of a single-node edge server are well-suited to the tight spaces and power-constrained environments typical of maritime vessels.
- **Agriculture:** Precision agriculture leverages IoT and edge computing for real-time monitoring and management of crops and livestock. Given the scale of large farms, single-node, distributed edge compute makes a lot of sense. Edge-enabled sensors and drones can collect data on soil conditions, crop health and weather patterns, which can be processed locally to provide immediate insights. This information helps in optimizing irrigation, pesticide application and harvest timing, improving yields and sustainability.

Multi-node edge topology: Big-box retail

Snapshot	
Edge use cases	Digital points of sale; cameras for footfall tracking and shopper surveillance; self-checkout; plus support for legacy back-office store applications.
Business outcomes	Improve the customer experience; enable more personalized products and services; maximize profitability via more automated operations from warehouse to shelf.
Performance requirements	Multiple applications require significant processing power; edge-to-core data exchange between stores and the cloud; AI inferencing/computer vision for advanced applications.
Location considerations	Variety of in-store venue types; somewhat controlled/conditioned but also a public environment that must be secured and monitored.
Edge deployment	Multi-node edge cluster enabling low-latency performance; local processing to reduce network and cloud costs; enhanced reliability through redundancy; ability to manage multiple clusters remotely; opportunity to scale up if more compute resources are needed.

The retail industry is undergoing digital transformation driven by the increasing adoption of new technologies such as IoT, AI, and edge and cloud computing. These technologies enable retailers to improve their operations, customer experience and profitability. Top edge-enabled use cases include self-checkout, which can allow customers to scan and pay for items using their smartphones, eliminating the need for traditional checkout lines; proximity marketing, which uses Bluetooth beacons to send targeted marketing messages to customers based on their location in a store; and enhanced customer experience using digital signage, interactive displays and personalized recommendations to create a more engaging and immersive shopping experience.

Edge compute infrastructure is essential for supporting the digital use cases in the retail industry. By bringing computing resources closer to retail data sources, edge enables new

retail applications that require real-time data processing, such as cashier-less checkout and proximity marketing. By processing more data locally, edge reduces costs and enables applications that require high bandwidth, such as delivering video surveillance streams to help minimize retail shrinkage. Finally, edge computing can help large chain national or multinational retailers that must comply with data sovereignty regulations by keeping data within the country or region where it is collected.

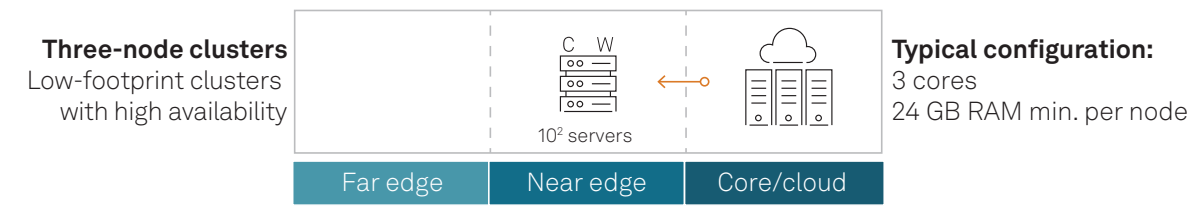
In many cases, the best topology for supporting digital use cases in the retail industry is an in-store multi-node cluster topology. This topology provides a number of critical advantages, including high availability, essential for mission-critical applications supporting both retail employees and customers; the ability to scale up or down to meet the fast-changing demands of today's retailers; and the flexibility to deploy in a variety of locations, including retail store back offices, warehouse distribution centers and corporate offices.

What are the top tech-enabled initiatives for retailers?

Improving customer fulfillment options (cited by 40% of firms), followed by improving data security and governance (38%) and optimizing the customer experience across channels (37%).

Source: 451 Research's Voice of the Enterprise: Customer Experience & Commerce, Merchant Study 2023.

Figure 5: Multi-node edge topology



Source: S&P Global Market Intelligence 451 Research.

Additional industries that can make use of a multi-node topology and approach:

- **Telecommunications:** A multi-node cluster topology can be used to support demanding, yet increasingly distributed, telecommunications use cases such as network function virtualization, operational and billing system application modernization, and mobile edge computing.
- **Entertainment and media:** In stadiums, theaters and amusement parks, edge computing supports real-time content delivery, augmented reality applications and crowd management, ensuring safety and an immersive experience. For instance, multi-node edge clusters can process high-bandwidth streaming of live events to provide instant replays on mobile devices or manage AR overlays during performances.
- **Healthcare:** The healthcare industry is embracing edge computing to support critical real-time applications such as patient monitoring, telemedicine and real-time diagnostics. Such use cases require significant processing — albeit often within a single, networked building — making a multi-node edge topology ideal.

Mix of single- and multi-node edges: Discrete product manufacturing

Snapshot	
Use cases	Assembly line automation via robotics; automated quality control; predictive maintenance; asset tracking; inventory supply chain visibility.
Business outcomes	Supporting real-time, data-driven, automated decision-making for enhanced operational efficiency; improved product quality; reduced downtime; improved utilization of capital; competitive advantage; AI/ML insights and operational control.
Performance requirements	Super-high performance, security and reliability; ability to run sophisticated AI models/applications on-site; manage multiple server nodes across a large plant.
Location considerations	Varied and distributed over typically large locations or campuses. Many edge devices will be placed on harsh and busy factory floors, while some plants will also have more IT-conditioned environments such as server rooms or micro-datacenters. Multiple, high-performance systems and processes each require their own compute instance.
Preferred edge deployment	Distributed (at times temporarily disconnected) server nodes, typically a mix of single nodes and multi-node clusters, small clusters with centralized management, security and support for cloud-native application development, deployment and orchestration.

Manufacturers are no newcomers to digital transformation. The term Industry 4.0 itself is a nod to the many generations of technology-driven reengineering that have brought the benefits of intelligent automation and efficient operations to the factory plant. The sector's ongoing digital transformation will be fueled by large doses of edge computing, deployed near and attached to industrial machinery, controls and sensors. Many of the most impactful manufacturing digital use cases — such as predictive maintenance, real-time quality control, and remote monitoring and control — require edge computing as an enabler of more real-time, fully automated operations.

Critical benefits of edge computing in manufacturing include reduced latency for better performance, high bandwidth for data-intensive applications, enhanced security to protect sensitive data and data sovereignty compliance. Edge computing devices strategically placed close to the data source optimize these capabilities. In large plants, there are likely to be a multitude of edge capabilities deployed: small-form compute attached directly to machinery for ultra-low-latency performance; single-node edge servers for more compute-intensive on-the-floor operations; and multi-node edge servers running in server rooms or as micro-datacenters handling larger jobs and helping to remotely manage the distributed edge instances.

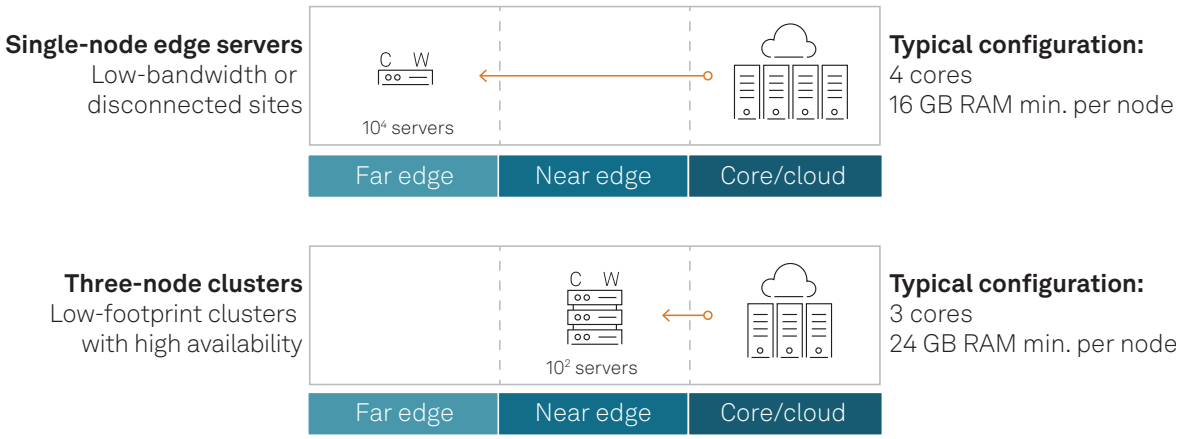
What are the top edge-enabled use cases for manufacturers?

Production monitoring (in use by 54% of respondent firms), followed by quality assurance (45%), inventory monitoring and management (43%) and predictive maintenance (40%).

Source: 451 Research's Voice of the Enterprise: Internet of Things, The OT Perspective, Use Cases & Outcomes 2024.

To support that more complex environment, manufacturers should consider a mix of edge topologies. At times, single-node servers will make sense to support a single function, while at other times a larger cluster will be needed to enable high performance and run multiple critical applications. This topology offers flexibility, scalability, reliability and security advantages. Each node can operate independently if needed, ensuring uninterrupted operations. At the same time, centralized security, control and operations facilitate efficient management and protection of the distributed nodes.

Figure 6: Mix of single-node and multi-node edge topologies



Source: S&P Global Market Intelligence 451 Research.

Additional industries that can make use of a distributed approach with a mix of single nodes and multi-node clusters:

- **Smart cities:** Urban centers utilize edge computing to manage IoT devices that monitor and control everything from traffic lights and public transport to air quality sensors and public safety systems. The sector is suited to a combination of highly distributed device- or single-node edge clusters for remote processing and centralized, larger edge clusters for aggregating and executing more data-processing-heavy workloads. Such an approach can enable use cases to optimize traffic flow, reduce congestion and enhance emergency response times.
- **Public transportation:** In the transportation sector, especially for public transport systems, edge computing plays a role in both highly distributed, at-times disconnected venues such as on a train or bus, while operations centers benefit from heftier edge compute to support larger-scale applications such as route planning, predictive maintenance of vehicles and management of traffic flows. A distributed topology with a variety of right-sized clusters that can also be centrally managed can help public transport systems significantly enhance operational efficiency and safety.
- **Financial services:** In the financial sector, edge computing is crucial for a range of use cases, from enabling real-time transaction processing to fraud detection to high-frequency trading. Distributed single-node edge servers deployed in bank branches or ATMs can swiftly process high volumes of transactions locally, reducing latency and enhancing security. Trading applications, meanwhile, require more compute capacity, as well as the low latency and high responsiveness that compute at the edge can provide. At the same time, distributed, single-pane-of-glass management of fleets of edge venues across a financial services IT environment is necessary for secure and agile operations.

Conclusion

To optimize edge infrastructure deployment, organizations should look to modern, cloud-native edge platforms and deployment approaches that best fit use-case requirements and on-the-ground realities. Such an approach fosters speed and agility, improves manageability and scaling, and enables maximum innovation. Flexible topologies and composable installations allow for customization based on specific use cases, enabling seamless scaling and adaptation to evolving application requirements.

Organizations must also carefully analyze their unique needs before deploying edge infrastructure. Factors such as latency, security and data processing capabilities should be considered to ensure the infrastructure aligns with application demands. By adopting these principles, industries can effectively leverage edge infrastructure to enhance operational efficiency and drive business value.

If you'd like to see how all of this fits into everyday technology, check out Hatville, the miniature city where edge computing comes to life.

About the author

Rich Karpinski

Principal Analyst, IoT

Rich Karpinski is principal analyst and channel lead for the 451 Research Internet of Things and Applied Infrastructure & DevOps channels within S&P Global Market Intelligence. Rich tracks, analyzes and anticipates the pace and direction of IoT adoption, as well as the use of IoT to enable vertical industry transformation. As part of that work, he oversees a quarterly survey of IoT adopters and a twice-annual survey of operations technology (OT) professionals.

Rich's recent areas of concentration include IoT connectivity and managed platform services, IoT edge computing, IT/OT collaboration, IoT market sizing and data flow analysis, IoT digital maturity analysis, and the adoption of IoT use cases across a variety of sectors.

Before joining 451 Research (acquired by S&P Global Market Intelligence in 2019), Rich was mobile services analyst for Yankee Group and a technology editor for a range of industry publications in both telecom and enterprise IT. Rich holds a holds a Bachelor of Arts degree from the University of Illinois and a Master of Science degree from Syracuse University.

About this paper

A Pathfinder paper navigates decision-makers through the issues surrounding a specific technology or business case, explores the business value of adoption, and recommends the range of considerations and concrete next steps in the decision-making process.

About S&P Global Market Intelligence

At S&P Global Market Intelligence, we understand the importance of accurate, deep and insightful information. Our team of experts delivers unrivaled insights and leading data and technology solutions, partnering with customers to expand their perspective, operate with confidence, and make decisions with conviction.

S&P Global Market Intelligence is a division of S&P Global (NYSE: SPGI). S&P Global is the world's foremost provider of credit ratings, benchmarks, analytics and workflow solutions in the global capital, commodity and automotive markets. With every one of our offerings, we help many of the world's leading organizations navigate the economic landscape so they can plan for tomorrow, today. For more information, visit www.spglobal.com/marketintelligence.

CONTACTS

Americas: +1 800 447 2273
Japan: +81 3 6262 1887
Asia-Pacific: +60 4 291 3600
Europe, Middle East, Africa: +44 (0) 134 432 8300

www.spglobal.com/marketintelligence
www.spglobal.com/en/enterprise/about/contact-us.html