EQUITY RESEARCH | October 14, 2018 | 9:06PM EDT

Cloud Platforms, Volume 5

The Cutting 'Edge' of Computing

How edge computing will augment the cloud & unlock real-time, big data applications

As more devices generate more data from more locations, computing is facing a speed-versus-scale challenge. The public cloud remains unrivaled in its compute and storage resources, but getting data there and back takes time, and ultimately is limited by the speed of light and the size of internet "pipes." In cases where big data will be used to drive real-time decision-making, we see an opportunity for "edge computing" to become a key enabler and extension of the public cloud by putting compute and storage resources closer to the device or source of data generation. Edge computing could unlock a $40bn incremental market ($100bn in the bull scenario), including a range of new applications that can better direct operations—from "when to brake" for a self-driving truck to "when to change course" for an oil drill working miles underground. We see providers of a cohesive public cloud and edge solution as best positioned and highlight Microsoft in particular, in addition to Amazon, Pivotal, and VMware.

Heather Bellini, CFA
+1 212 357-7710
heather.bellini@gs.com
Goldman Sachs & Co. LLC

Ted Lin
+1 212 357-6107
ted.lin@gs.com
Goldman Sachs & Co. LLC

Mark Grant
+1 212 357-4475
mark.grant@gs.com
Goldman Sachs & Co. LLC

Caroline Liu
+1 212 357-9111
caroline.liu@gs.com
Goldman Sachs & Co. LLC

The Goldman Sachs Group, Inc.

Table of Contents

PM summary	3
Shift to the cloud continues in earnest	11
But computing is poised to shift back to a decentralized paradigm	13
What is edge computing?	17
Edge computing demand drivers	22
Sizing the potential market opportunity for virtualization and server operating systems	28
Edge computing vs. cloud computing performance	33
Killer apps for edge computing	44
Winners & losers: edge computing could sustain a renaissance in on-premise software	49
Valuation & risks	67
Disclosure Appendix	68

PM summary

Why is the edge important?

While the overarching theme in software will continue to be the centralization of compute (i.e. the moving of workloads from on-premises to public cloud), we believe that computing at the edge will play an increasingly important role, augmenting the capabilities of public cloud and bringing resources closer to the source of data generation. In edge computing, data is processed, analyzed, and acted upon at (or close to) the source of data generation, as opposed to raw data being sent directly to a public or private cloud to be acted upon. To accomplish this, edge computing adds the core building blocks of public cloud – including compute, networking, and storage – closer to the origin of the data, allowing insights to be generated and executed in real-time. In contrast with centrally-located traditional and purpose-built on-premise data centers or private clouds, edge servers can be placed far from centralized computing cores – in (or around) factories, airplanes, cars, oil rigs, or in conjunction with cell phone towers. In an edge + cloud world, processing is therefore divided between the edge and the cloud, and fundamentally, our view is that edge computing is complementary to (and not a substitute for) the public cloud – moving all compute to the edge would result in distributed and unmanageable clusters of chaos and forgo the scale benefits of public cloud.

Although public cloud has effectively limitless resources, edge computing has several advantages that cannot be effectively matched by the public cloud. For instance, latency (distance to the public cloud) and bandwidth (size of the pipe connected to the public cloud) remain issues in many instances. For use cases where reaction time is critical to the success of the overall system, the latency inherent with a round trip to the cloud via a hub-and-spoke model may be not be acceptable. Latency can be influenced by a plethora of uncontrollable factors, including the network connectivity of the location, the network provider, other network traffic, as well as the specific region, availability zone, and data center that the user connects to. Additionally, the speed of compute and data processing has far outclassed network bandwidth. Truly big data use cases will also create massive data generation, orders of magnitude above what could be transmitted back to the public cloud; in fact, these big data use cases will generate sufficient data that simply *storing* it, even with the resources of the public cloud (assuming that the data can be transmitted there), will be challenging; edge computing will enable the data to be processed immediately, and only relevant data needs to be sent back to the public cloud to be stored and further reasoned upon. Dependence on public cloud for all data processing and analytics may not be suitable for many use cases, particularly those that feature low or intermittent network connectivity, and we believe that even 5G may not be adequate bandwidth for many use cases. Finally, processing the data on the device or at edge, versus uploading raw data to the public cloud, can yield superior results for security and privacy, as there are inherent risks in transmission.

How big is this market?

In this report, we evaluate the potential incremental infrastructure software spend that could be attributed to an increase in edge servers, driven by the need to perform processing closer to the source of data generation. With 2.72bn IoT endpoints (i.e. the connected "things" themselves) shipments in 2021, we estimate that in the most conservative scenario, the incremental annual value (i.e. license, maintenance, and subscription revenue) would be $14bn for virtualization and $7bn for server operating systems; in the most aggressive scenario, the incremental annual spend would be $69bn for virtualization and $34bn for server operating systems. We note, however, that these estimates likely skew conservative, as it does not account for other infrastructure software like NoSQL databases, which could potentially be a lightweight option for edge computing; nor does it account for analytics and application software, which will depend heavily on the types of use cases leveraged for edge computing resources. We also believe that container adoption could serve as a multiplier for spending, as Red Hat has commented that OpenShift is "almost 20x the price of RHEL on the same two-socket server." Finally, we highlight that these forecasts do not include any hardware or incremental storage capacity, just to name a few, that would also be directly impacted by the build out of edge networks.

"Killer apps" enabled by the edge

Based on the unique advantages of edge servers relative to public cloud and small IoT endpoints, we believe that edge computing enables a broad spectrum of use cases that leverages edge servers' ability to perform advanced computational tasks at the source of data generation. We believe use cases like autonomous cars/trucks, digital oilfields, and video analytics have the ability to revolutionize business processes; however, we believe that until infrastructure to enable inference at the edge is in place, these markets will fall short of their full potential. We highlight some potential edge computing use cases below; we note that these use cases are not an exhaustive list:

Autonomous cars & trucks: Real-time processing via an onboard edge server is critical to the safe operation of an autonomous vehicle, for both the passengers as well as the general public; an autonomous vehicle cannot afford the latency required to access the public cloud, as any delays in reaction speed could be potentially catastrophic. For this use case, analyzing the data in real-time – a task that can only be accomplished by an edge server – is critical to maintaining the vehicle's safety, efficiency, and performance.

AR/VR: Augmented and virtual reality use cases require large amounts of processing power; however, users are heavily sensitive to latency, precluding AR/VR from leveraging public cloud given the networking capabilities available today. While we would expect PCs remain the primary mode of compute for the time being, we could see use cases develop for the use of edge servers if this latency can be improved over time (i.e. through 5G), particularly where device-level compute is too difficult to achieve in a form factor that meets the needs of the user.

Digital oilfields: Edge computing is slated to play an increasingly vital role in oil and gas exploration, given the remote locations in which the industry operates. For instance, using real-time processing can help to maximize drills' output while minimizing energy consumption by analyzing drill data in real-time to make instant decisions about the drill's next best course of action.

IoT enterprises: As increasing amounts of compute, storage, and analytics capabilities are integrated into ever-smaller devices, we expect IoT devices to continue to proliferate, and as noted previously, Gartner expects IoT endpoints to grow at a 33% CAGR through 2021. In cases where reaction time is the raison d'être of the IoT system, the latency associated with sending data to the cloud for processing would eliminate the value of the system, necessitating processing at the edge; public cloud could still be leveraged where processing is less time sensitive or in instances where the scale and sophistication of public cloud need to be brought to bear.

Public safety (Amber Alerts): Video analytics is an example where bandwidth limitations, long latency, and privacy concerns converge to favor edge computing over leveraging public cloud. For instance, locating a lost child in a city is one potential real-world application of video analytics where public cloud limitations would prevent successful deployment. With an edge computing paradigm, the request to locate the missing child can instead be pushed out to all of the relevant devices: each camera would perform the search independently using nearby compute resources. If, and only if, the camera registers a positive match would it then upload data to the cloud: by distributing the analytics to the small-but-numerous devices in the edge (where the data resides), tasks can be quickly and efficiently processed.

Reiterating our Buy on Microsoft

We reiterate our Buy on Microsoft and our $123, 12-month price target based on the company's strong positioning in all three tiers of public cloud, as well as its emerging leadership in edge computing. We note that our EPS estimates are above FactSet consensus for CY19 and CY20:

Exhibit 1: Our EPS estimates are above consensus in CY19 and CY20
GS vs. consensus (in $ mn except per share data)

	CY19 (E)			CY20 (E)			CY21 (E)
	GS	Consensus	% Δ	GS	Consensus	% Δ	GS
Revenue	127,878	128,543	(0.5%)	140,660	142,834	(1.5%)	154,210
EPS	$4.67	$4.53	3.1%	$5.48	$5.25	4.5%	$6.42
FCF	40,676	36,778	10.6%	44,652	44,910	(0.6%)	48,174

Source: Goldman Sachs Global Investment Research, FactSet

One technical analogy often cited for public cloud is its similarity to a utility. Prior to the 1880s and the advent of central power plants, electricity was typically generated on-site and therefore limited to factories, hotels, and wealthy residences. These generators were typically located in the basement, or in close proximity (e.g. a nearby river or waterfall). However, due to variety of reasons, including scale benefits (i.e. volatility in demand, R&D, purchasing), the ability to shift capital expenditure to operating expenses, and the ability to offload non-core operations, electricity generation quickly moved to centralized power plants, with consumers and businesses alike purchasing electricity as a service.

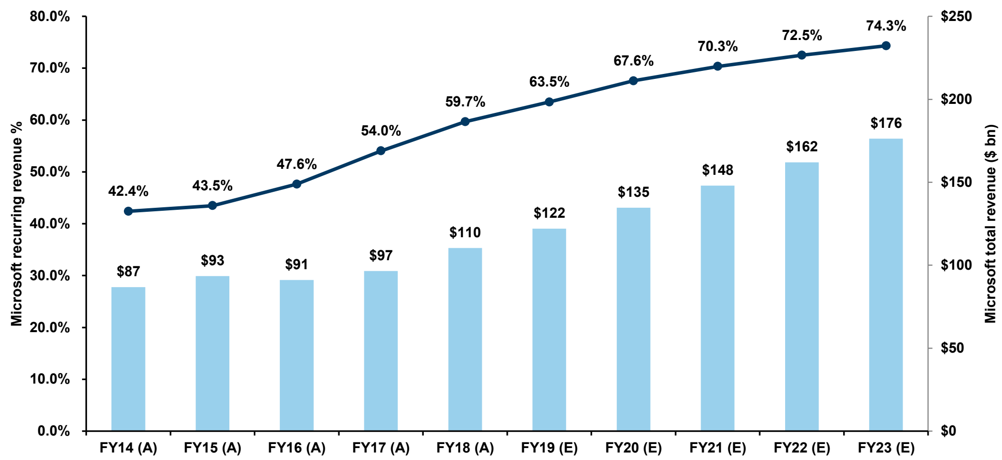
We believe that cloud computing will follow a similar trajectory, with servers and computing platforms increasingly delivered as a service, due to the same benefits that existed for electricity to become delivered as a service: scale, capex -to-opex, and offloading non-core operations. As such, as public cloud becomes increasingly central to enterprises' IT stacks, we believe the key components of servers (compute, networking, and storage) will increasingly resemble utilities like electricity and water, where resources are generated centrally, then delivered and consumed as needed by customers.

We would caveat, however, that there are important core differences in the comparison of public cloud business models and utilities business models. Importantly, utilities are a natural monopoly, and as a result, it is functionally impossible for a company to churn off (as there are no competitors and going off the grid would be clearly infeasible). For public cloud, we would foresee at least three major competitors moving forward (AWS, Azure, and GCP), and while we continue to believe in the increasing stickiness of the platforms, particularly as customers adopt PaaS features, it is clearly possible to migrate workloads from one platform to a competitor (and partners have noted that this indeed occasionally occurs). Additionally, utilities are guaranteed an ROE, and while they may overearn or underearn in certain years, they can generally apply to regulators to increase revenue in the event of underearning. By contrast, public cloud services are determined by market-clearing rates, and we note that in some instances, services may, in fact, be priced below cost. As a result, we would expect the ROE of public cloud to continue to be more volatile than that of utilities'. Finally, we note that while Microsoft pays a consistent dividend (yield of ~1.7%), this is approximately half that of the average of the utilities that we evaluated (~3.4%).

For the major public cloud vendors, revenue derived from supplying these resources is therefore recurring and sticky. Enterprise applications (e.g. enterprise resource planning applications, customer relationship management systems, human resources management systems, specialized industry applications) and data are typically fundamental to the operation of the business; without this infrastructure, the business ceases to operate effectively. As a result, even in the face of economic headwinds, the spending impact on this core infrastructure will be relatively muted to other areas that may be more susceptible to spending reductions. In the traditional enterprise software perpetual license + maintenance model, customers could choose to churn off maintenance *and still retain the usage of the software*; this is not possible with subscription-type models (e.g. public cloud, SaaS), where the churning off the platform means that the customer is no longer entitled (legally, and typically technically as well) to use the software.

We believe that Microsoft continues to be one of the best-positioned enterprise vendors to take advantage of the secular shift towards cloud, with strong offerings in all three major cloud categories, with Azure for IaaS and PaaS, as well as Office 365, Dynamics 365, and LinkedIn in SaaS. As a result of the company's pivot towards public cloud, Microsoft's base of recurring revenue has continued to push upwards over time, growing in both absolute dollars and as a proportion of the company's overall revenue.

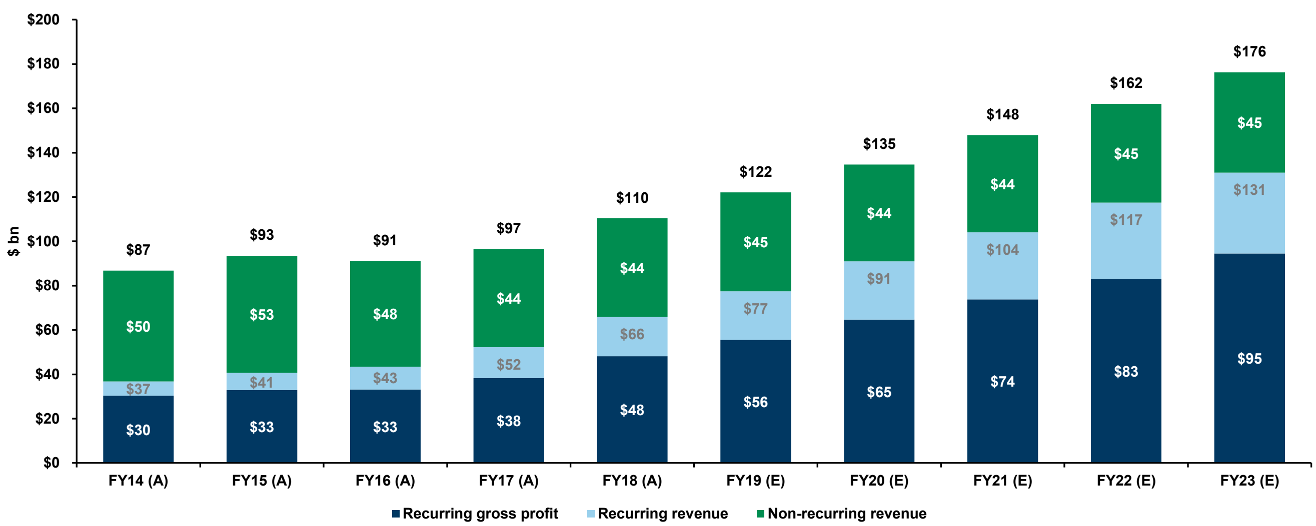
Exhibit 2: Microsoft's recurring revenue (as a % of total revenue) is inflecting, and revenue growth continues to be steady…
Microsoft recurring revenue % and total revenue



Microsoft recurring revenue is defined as commercial Office 365, commercial Office annuity, consumer Office 365, Dynamics 365, LinkedIn, Azure, server & tools annuity, Windows volume licensing, and Xbox Live

Source: Company data, Goldman Sachs Global Investment Research

Exhibit 3: …and an increasing proportion of gross profit dollars will also be recurring
Microsoft recurring gross profit dollars, recurring revenue, and total revenue

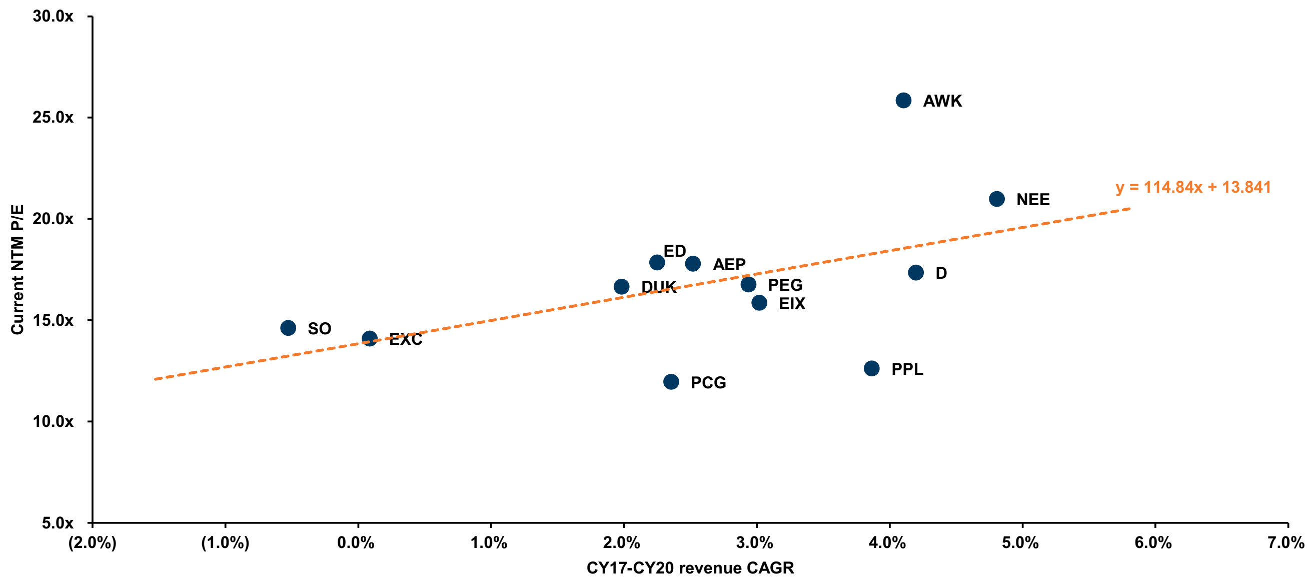


Source: Company data, Goldman Sachs Global Investment Research

As Microsoft's business increasingly resembles that of a utility's, we examine utilities' valuations as a framework for where Microsoft's valuation could trend, as both are businesses that have steady revenue streams that are relatively well-insulated from macroeconomic shocks. Additionally, like utilities, the public cloud is increasingly viewed as a staple of business, and we believe that the public cloud vendors should have pricing power (within reason) over time as it becomes increasingly integral to day-to-day business operations. We note that *unlike* utilities, public cloud vendors are *not* highly regulated, which we view as imparting a positive bias to the valuation.

In Exhibit 4, we plot utility stocks' valuations (power and water utilities covered by GS, given the aforementioned similarities to public cloud) as measured by NTM P/E) and three-year revenue CAGRs (CY17 through CY20), noting that in general, there is a positive correlation between revenue growth and valuation.

Exhibit 4: There is a positive correlation between revenue growth and valuation in utilities
Utilities valuations vs. revenue growth

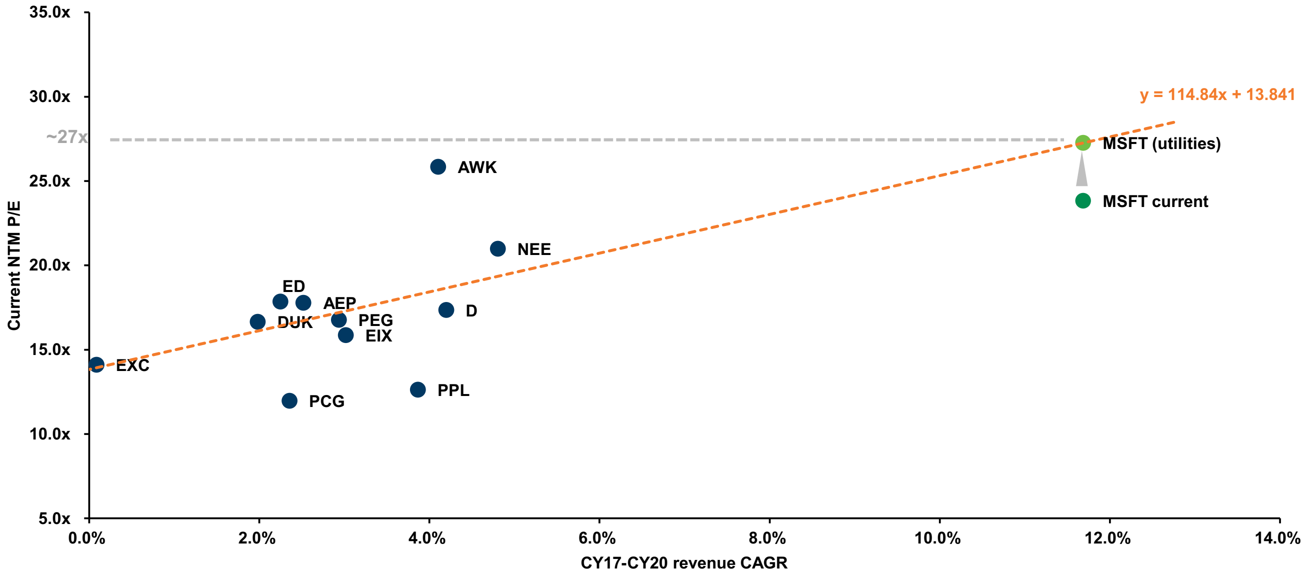


Source: Goldman Sachs Global Investment Research, FactSet

Microsoft, we believe, resembles a utility in terms of the stability of its business model; however, unlike most utilities, Microsoft is growing at a much faster pace, as the company benefits from a secular tailwind to the cloud and has much fewer regulatory constraints.

Microsoft currently trades at ~25x consensus CY19 EPS of $4.53 which compares to the S&P500 at ~16x. We believe that as it becomes increasingly apparent that Microsoft has the recurring revenue and defensibility characteristics of utilities, the multiple could continue to re-rate upwards. We note that granting Microsoft the valuation of a utility stock at ~12% NTM revenue growth would yield a multiple of ~27x NTM EPS (Exhibit 5). For a full comp table, please see Exhibit 42 at the end of this report.

Exhibit 5: Applying utilities' valuation to Microsoft, adjusted for 11% NTM revenue growth yield an NTM P/E of ~27x
Utilities valuations vs. revenue growth & Microsoft



Source: Goldman Sachs Global Investment Research, FactSet

We also note that in addition to Microsoft's strong cloud portfolio, the company continues to believe that hybrid cloud will continue to be the IT paradigm of choice for the foreseeable future. Although we view all three of the major public cloud vendors (Amazon Web Services, Microsoft Azure, and Google Cloud Platform) as well-positioned moving forward, we believe that Microsoft is uniquely positioned for a hybrid cloud and edge computing architecture, given its strength in on-premise software (i.e. Windows Server, SQL Server, Systems Center, Windows, and Office) coupled with Azure and offerings like Azure Stack; neither AWS nor GCP has the on-premise and hybrid legacy possessed by Microsoft.

In the utility analogy, we note that although centralized power generation is clearly the dominant form of electricity production today, electricity continues to be generated locally in many instances. For instance, every modern automobile has an alternator, used to generate electricity to power the car's electronics and charge the car's battery. Every airplane also has at least one alternator; the Boeing 787 has six generators – two per engine and two on the auxiliary power unit. Remote locations like oil rigs also require generators, as they are too geographically isolated to hook up to the electrical grid. Critical infrastructure like hospitals, government buildings, banks, and ironically, public cloud data centers, also typically have generators that can function as backup for the electrical grid in case of a failure. Even with all the benefits of large central power plants, there is clearly still a need for small-scale power generation; we believe this is analogous to the need for edge computing even with all the benefits of large public cloud data centers.

Similarly, Microsoft CEO Satya Nadella has noted that as generation of data continues to increase exponentially, the "edge of the cloud," or on-premise servers, will become increasingly important, as it will be impractical (or impossible due to latency) to shift petabytes of data generated from on-premise sensors to the cloud for analysis.

Who else stands to benefit?

In the near-term, we would expect that edge servers leverage very similar architectures as on-premise data centers today, to ensure maximum compatibility between the edge server and data center. As such, we would expect that **VMware**'s vSphere, which has ~90%+ market share in what has emerged as a winner-takes-all virtualization market, would clearly be the vendor to benefit from the need for virtualization in edge servers, in our view. **Red Hat**, as the preeminent enterprise Linux provider (Red Hat Enterprise Linux, or RHEL), would be also be a beneficiary of edge server computing, as we envision that enterprises would strive to maintain a consistent environment between the edge server, their data centers, and the public cloud, which is dominated by Linux distributions. We would also expect that containers play an increasing role in edge computing, given the necessity of wringing out every possible bit of performance from a finite and constrained resource like an edge server, and with the rise of containers in edge computing, we believe that infrastructure agnostic container platforms like Red Hat OpenShift and **Pivotal** Cloud Foundry would benefit.

As a result, we reiterate our Buy rating on Pivotal, and we continue to believe that the company can help bridge software development for public cloud, on-premise, and edge computing. Pivotal Cloud Foundry leads to large gains in developer productivity, regardless of where the workload is deployed, and our view is that this enormously strategic product will lead to significant expansion opportunities with customers. For Red Hat, while recent performance has reflected early traction in OpenShift, we continue to see the investor debate focused primarily on whether RHEL, the company's core "bread-and-butter" product, remains competitive in public cloud against lower-cost cloud-native solutions (e.g. AWS Linux); RHEL's growth rate has decelerated over the past several quarters. As such, we maintain our Sell rating on Red Hat, given that we believe that the moving of workloads to the public cloud represents a net headwind to Red Hat Enterprise Linux (RHEL), even after accounting for the impact of edge computing.

Shift to the cloud continues in earnest

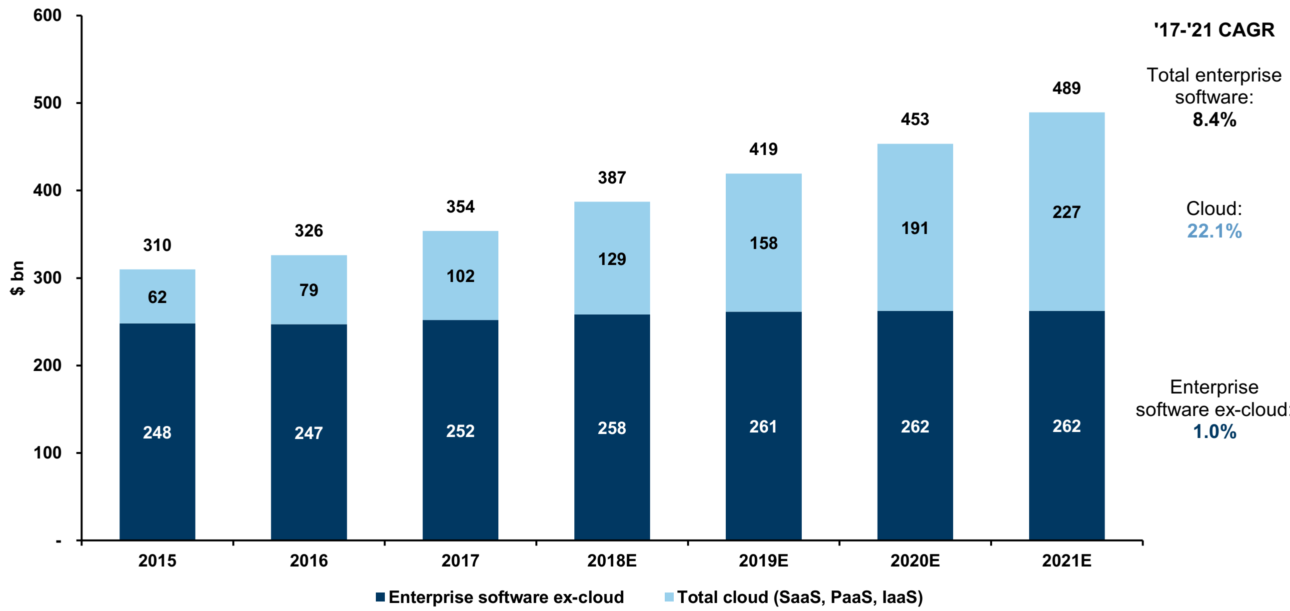
Key economic and technology drivers of public cloud remain intact

We continue to believe that the moving of workloads to the public cloud remains the overarching secular trend in software. This thesis remains intact, as the public cloud continues to enjoy a multitude of advantages over on-premise data centers:

- **Economies of scale (volatility):** With public cloud, companies can "burst" workloads – or send workloads to the public cloud during times of peak utilization (essentially using the public cloud as excess spillover capacity). For these customers, bursting offers efficiencies, as they do not pay for excess capacity on an ongoing basis; they pay for the extra compute resources only when they are required. Because different industries may burst at different times (i.e. financial services firms may begin their batch processing after the close, while another industry may wait until the middle of the night), demand levels for a public cloud vendor are much less volatile than demand levels for a single company's data center. As a result, public cloud vendors can service their base of customers with dramatically lower total capacity than if each customer were to build out their own infrastructure.
- **Economies of scale (R&D):** Because public cloud vendors have thousands of customers, they can afford to spend billions of dollars on research and development of new public cloud services (Microsoft FY19E capex: $15bn). For instance, Microsoft has developed Cognitive Services, which are artificial intelligence APIs that can be leveraged by any application to do vision (image-processing and recognition), language (natural language processing), speech (audio-to-text, voice recognition), and search (webpages, images, videos, news). These sophisticated models are pre-built and pre-trained by Microsoft and have the effect of democratizing new technologies that were previously available to only the largest companies.
- **Economies of scale (purchasing):** One element of scale that the public cloud providers benefit from is the ability to purchase and deploy infrastructure at huge volumes (Microsoft FY19E capex: $15bn). AWS, for example, spans 55 Availability Zones (AZs) within 18 Regions around the world, with announced plans for 12 more AZs and 4 more Regions. In 2017, Amazon spent $10.2bn in cash capex (primarily for AWS), as well as $9.6bn in additions to capital leases, for a total of ~$19.8bn in total capex.
- **Capex to opex:** Public cloud allows companies to avoid large capital expenditures for data center buildouts and infrastructure refreshes. Instead, leveraging public cloud enables companies to shift their lumpy capex requirements to smoother operating expenses, paying for only what they use.
- **Offload non-core operations**: For most non-technology companies, building, running, and maintaining computing infrastructure is not within their core competency. In the same way that companies pay utilities for electricity, paying

public cloud vendors for compute and storage enables companies to offload non-critical back-office functions to focus on the core business.

Exhibit 6: The shift to cloud continues in earnest
Enterprise software spend ($ bn)



Source: Goldman Sachs Global Investment Research, Gartner

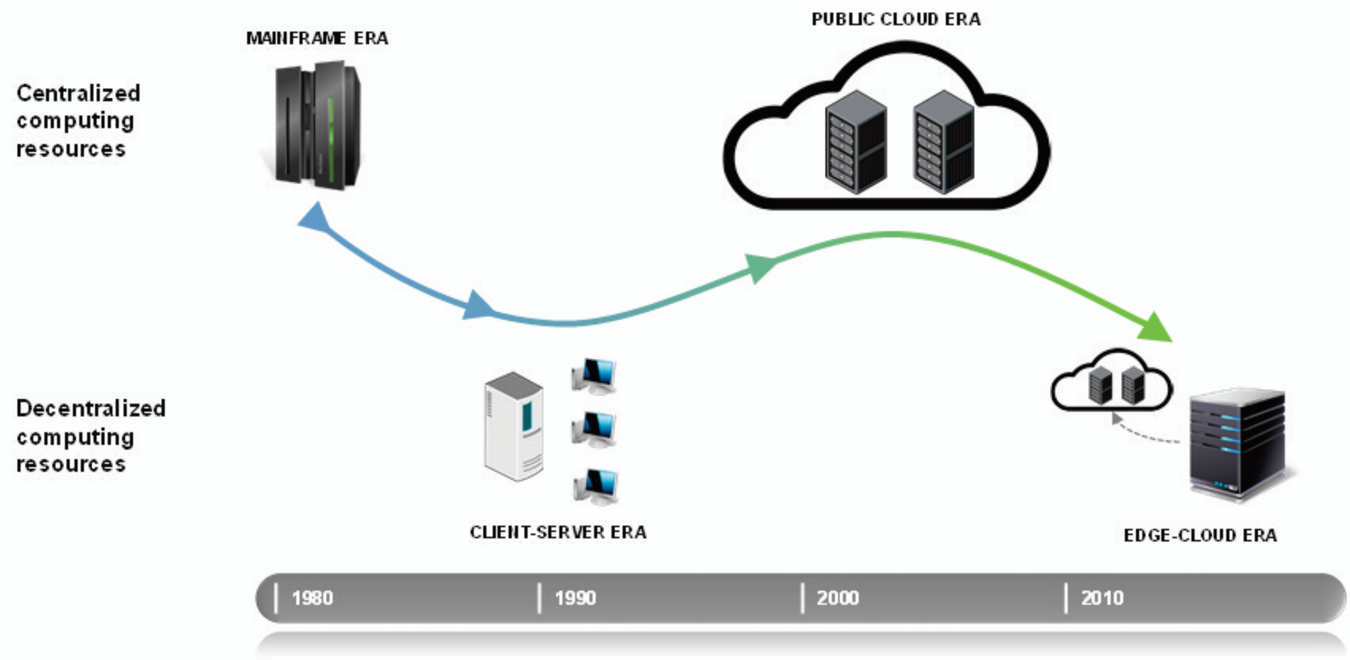
But computing is poised to shift back to a decentralized paradigm

Historically, computing has oscillated between centralized and decentralized paradigms. From the 1950s through the 1970s, mainframes were the dominant form of computing (although we note that given the fault-tolerant and resiliency of mainframes coupled with the mission criticality of mainframe workloads, a long tail of mainframe usage persists through today, particularly in government and financial services). Given the high costs of mainframe systems, in addition to the size and weight of these systems, mainframe computing was a highly centralized model, supported and controlled by large central IT organizations and specialized IT personnel. Access to the mainframe was provided through "dumb" terminals – machines with no processing power, serving simply as interfaces to the mainframe.

As technology progressed, however, components, and therefore computers, began to shrink in size. These smaller machines packed sufficient processing power to run business applications, and as a result, as PCs became increasingly prevalent, compute became decentralized, with compute resources primarily residing on PCs. Ultimately, these PCs evolved to be networked together, sharing files on communal systems that everyone could access (servers), ushering in the client-server era. Unlike mainframes, however, which have high utilization rates given their value, servers typically had lower utilization rates (5-10%); this inefficiency helped to drive the next era of computing.

The early 2000s saw the rise of cloud computing, enabled by technologies like the internet, automation, and virtualization, which allowed for the separation of computing resources from physical hardware. With cloud, large pools of configurable resources (compute, storage, and networking) are consolidated together and able to be quickly provisioned, delivered (over the internet), and scaled. Consolidating these resources together with a single vendor allowed for enormous efficiencies in terms of hardware purchases and scale benefits (similar to utilities), as well as the research and development of new services and offerings, helping to democratize cutting-edge services like big data analytics, AI, and machine learning. As the cost, scalability, and superior feature sets of the public cloud began to resonate with enterprises, coupled with the proliferation of mobile devices, the connectivity of which enabled perpetual access to cloud resources, the rise of the cloud pushed the pendulum back towards a centralized model of computing. As we detail in this note, our view is that it is time for the pendulum to begin swinging back – towards (more) decentralized computing, in an edge-cloud world, as this will enable a new set of computing use cases like autonomous cars, IoT, and AR/VR.

Exhibit 7: Computing, which has historically oscillated between centralized and decentralized paradigms, is swinging back from centralized (public cloud) to decentralized (edge computing)
Historical computing paradigms



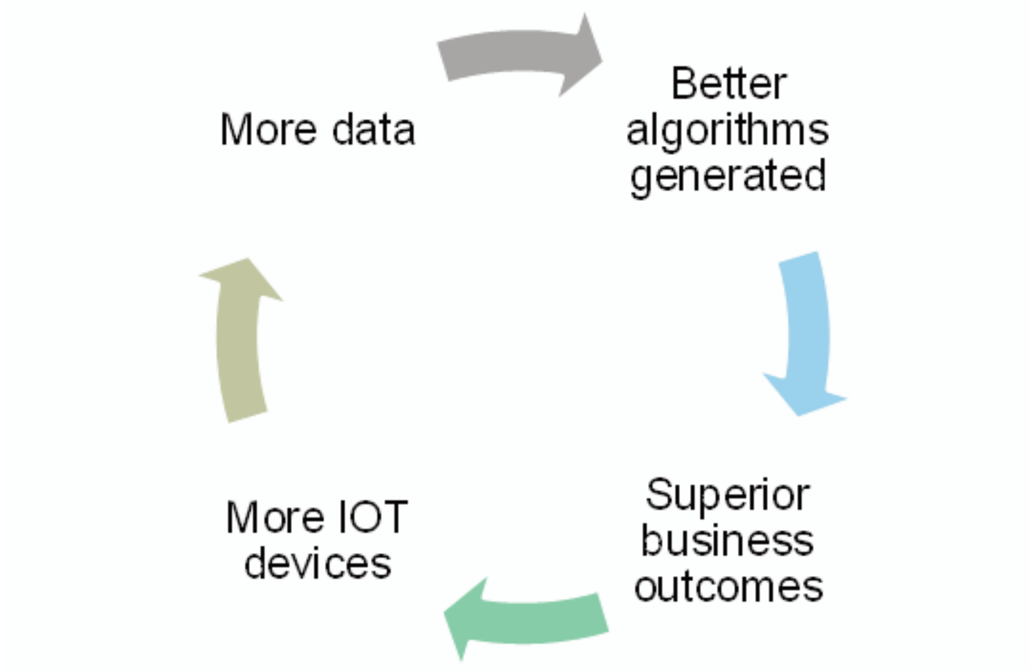
Source: Goldman Sachs Global Investment Research

With every paradigm shift and oscillation, the number of applications, devices, users, and therefore market size, have increased dramatically. For the mainframe era, the cost and physical size of mainframes placed constraints on the technology's potential. IBM, a company that has been a part of mainframe computing from its beginning through today, estimates that there are approximately 10,000 mainframe "footprints" in the world; if we assume a thousand users per mainframe, that would imply a maximum of 10mn users using mainframe resources.

In the PC era, as the "unit of purchase" was shrunk to a manageable level, this led Bill Gates to famously declare Microsoft's mission as "a computer on every desk and in every home." Today, factoring in emerging markets, Forrester estimates that there are approximately two billion PCs in the world – not quite a PC for every person in the world, but nearly so. In the mobile and cloud era, the total addressable market for computing quickly became the number of humans on the planet. In addition to the world's two billion PCs, the GSMA (the trade organization that represents mobile network operators worldwide) estimates that there are currently over five billion mobile phones subscribers globally, meaning that there is essentially one computing device (PC or phone) per human.

In the same vein, with the shift to edge computing, coupled with the rise of autonomous driving, IoT, and AR/VR, as well as the explosion of data sources, we would expect that the number of applications, devices, users, and market size will rise rapidly. The number of computing devices is no longer tethered to human beings: even if every human has a computing device (or multiple), there can be trillions of additional semi-autonomous devices, ranging from connected sensors to smart devices to industrial machinery.

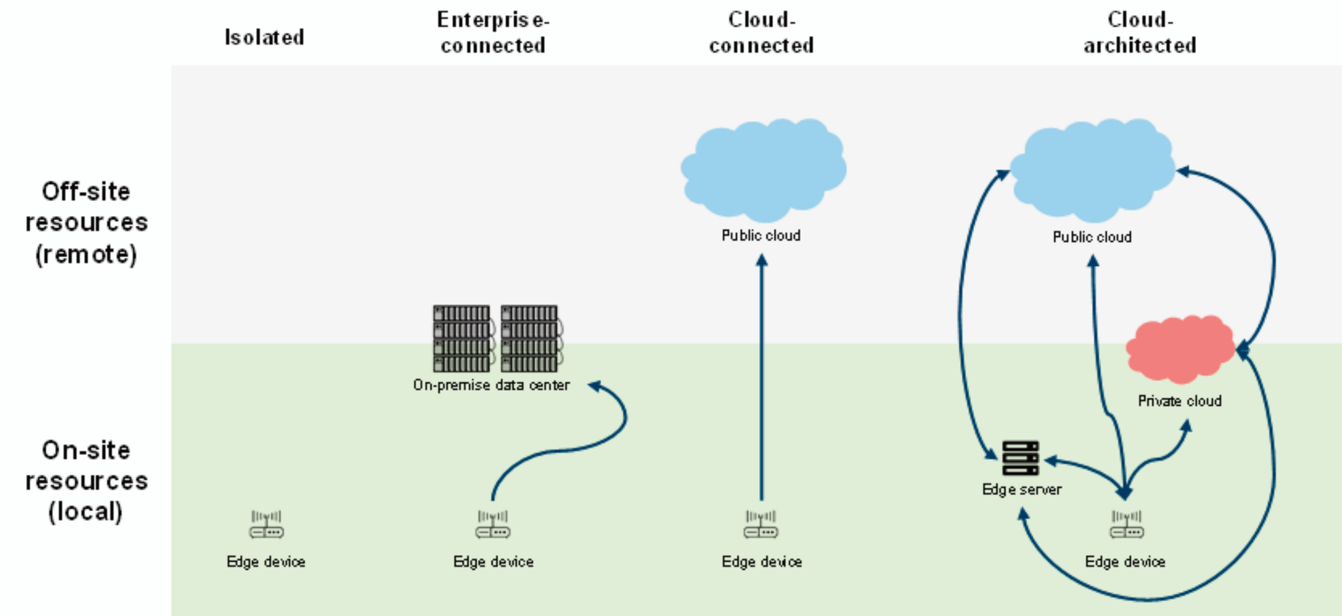
Exhibit 8: Big data drives better outcomes



Source: Goldman Sachs Global Investment Research

We note that some may view cloud computing and edge computing as competing paradigms, with cloud computing aggregating computing into highly centralized and hyperscalable resources and edge computing dispersing computing resources away from data centers. However, we believe that cloud computing and edge computing do not preclude one another: cloud computing is simply an archetype of computing where elastically scalable services are delivered via the internet, while edge computing is an implementation of this model, helping to deliver cloud services and features to the edge. As a result, our view is the cloud and the edge are highly complementary versus competing models of computing. Edge computing is *not* a replacement for cloud computing; rather, we believe it is the natural evolution of the public cloud – a step that allows the public cloud to permeate away from centralized data centers to interact more fluidly with devices at the edge of the network.

Exhibit 9: Edge computing complements cloud computing by bringing cloud services to the edge
Empowering devices at the edge



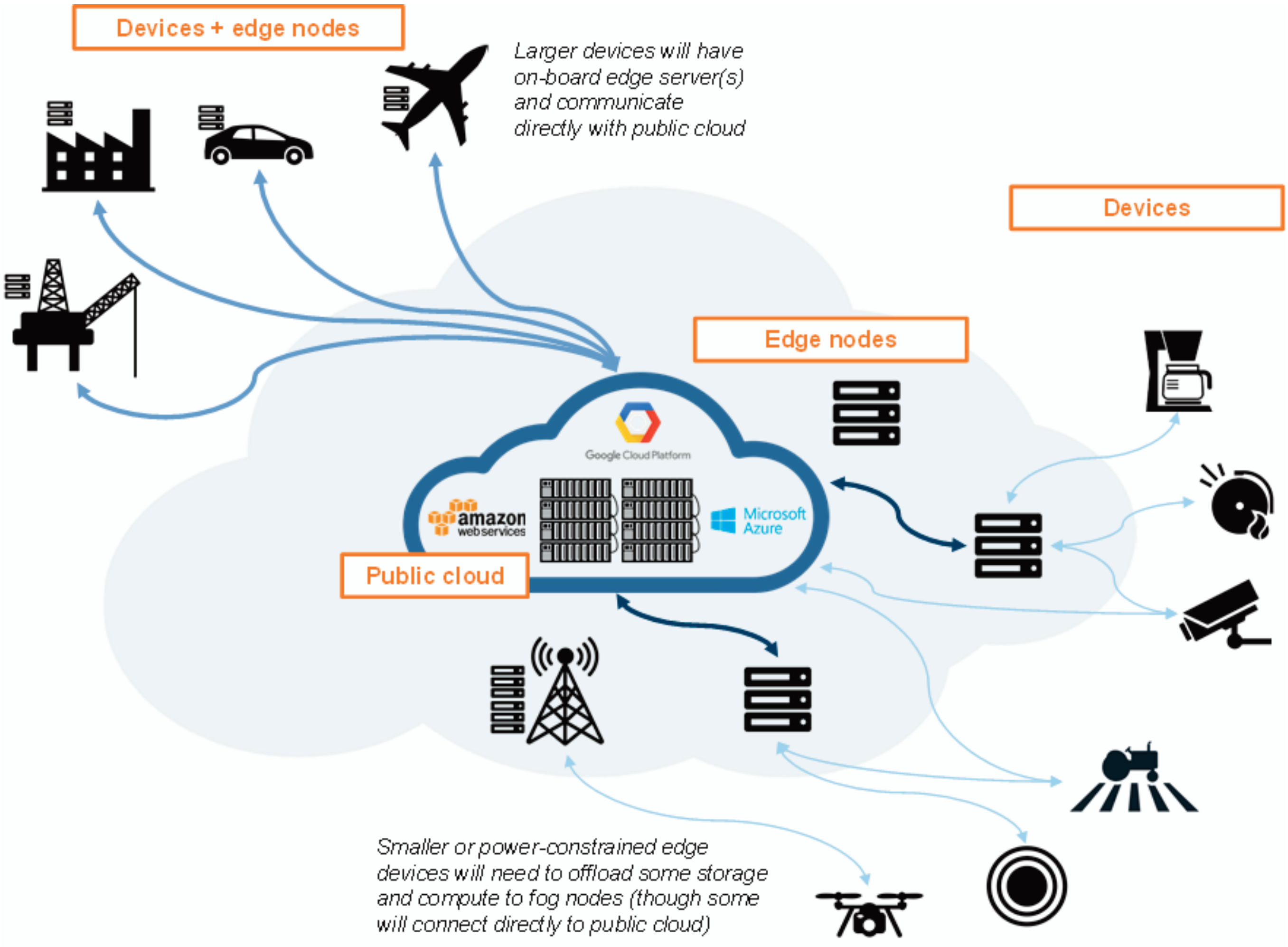
Source: Goldman Sachs Global Investment Research

What is edge computing?

In edge computing, data is processed, analyzed, and acted upon at (or close to) the source of data generation, as opposed to raw data being sent directly to a public or private cloud to be acted upon. To accomplish this, edge computing adds the core building blocks of public cloud – including compute, networking, and storage – closer to the origin of the data, allowing insights to be generated and executed in real-time. In contrast with centrally-located traditional and purpose-built on-premise data centers or private clouds, edge servers can be placed far from centralized computing cores – in (or around) factories, airplanes, cars, oil rigs, or in conjunction with cell phone towers. For this report, we take a fairly broad definition of the edge, defining it as any server *not* in a public cloud data center.

In an edge + cloud world, processing is therefore divided between the edge and the cloud, and fundamentally, our view is that edge computing is complementary to (and not a substitute for) the public cloud – moving all compute to the edge would result in distributed and unmanageable clusters of chaos and forgo the scale benefits of public cloud.

Exhibit 10: We envision the public cloud and edge working together
Public cloud & edge server paradigm



Source: Goldman Sachs Global Investment Research

In this new paradigm, processing responsibilities would be allocated to the computing component best suited for the task. While the pubic cloud will continue to far outclass the edge in terms of raw compute and storage capabilities, which means that they will continue to be the ideal environment for big data analytics or data storage, edge servers have the advantage of being adjacent to the data and the source of data generation. As a result, edge computing minimizes latency by bringing pieces and capabilities of the public cloud closer to where data is generated, making it ideal for use cases that require real-time processing or where networking (i.e. connectivity to the public cloud) is limited. Edge servers can therefore serve as the junction between edge devices that have limited compute, storage, and battery and the public cloud, which has these resources in abundance but is too far away to address real-time needs. The edge server can sit near the device but mimic the capabilities of the public cloud, supporting local ingestion of the data coupled with real-time processing of the results.

For instance, one potential use case would be machine learning, where the algorithms are initially trained and refined in the public cloud using massive data sets and vast compute resources, and once they are sufficiently accurate, the algorithms can be pushed out to the edge devices, which can then leverage the algorithm with real-time data. Subsequently, only the most valuable data (e.g. anomalies that can help to refine the model) is uploaded to the cloud, and as the model is refined, new iterations of the model are pushed to the device. With real-time processing offloaded to the edge, public cloud capacity can be allocated towards heavier tasks (i.e. analysis of large historical data sets).

Exhibit 11: Public cloud and edge servers have different (and complementary) strengths
Public cloud vs. edge servers

	Public cloud	Edge server
Big data analytics	✓	
Data consolidation	✓	
Long-term data storage	✓	
Need for real-time processing		✓
Networking (connectivity & bandwidth) limitations		✓

Source: Goldman Sachs Global Investment Research

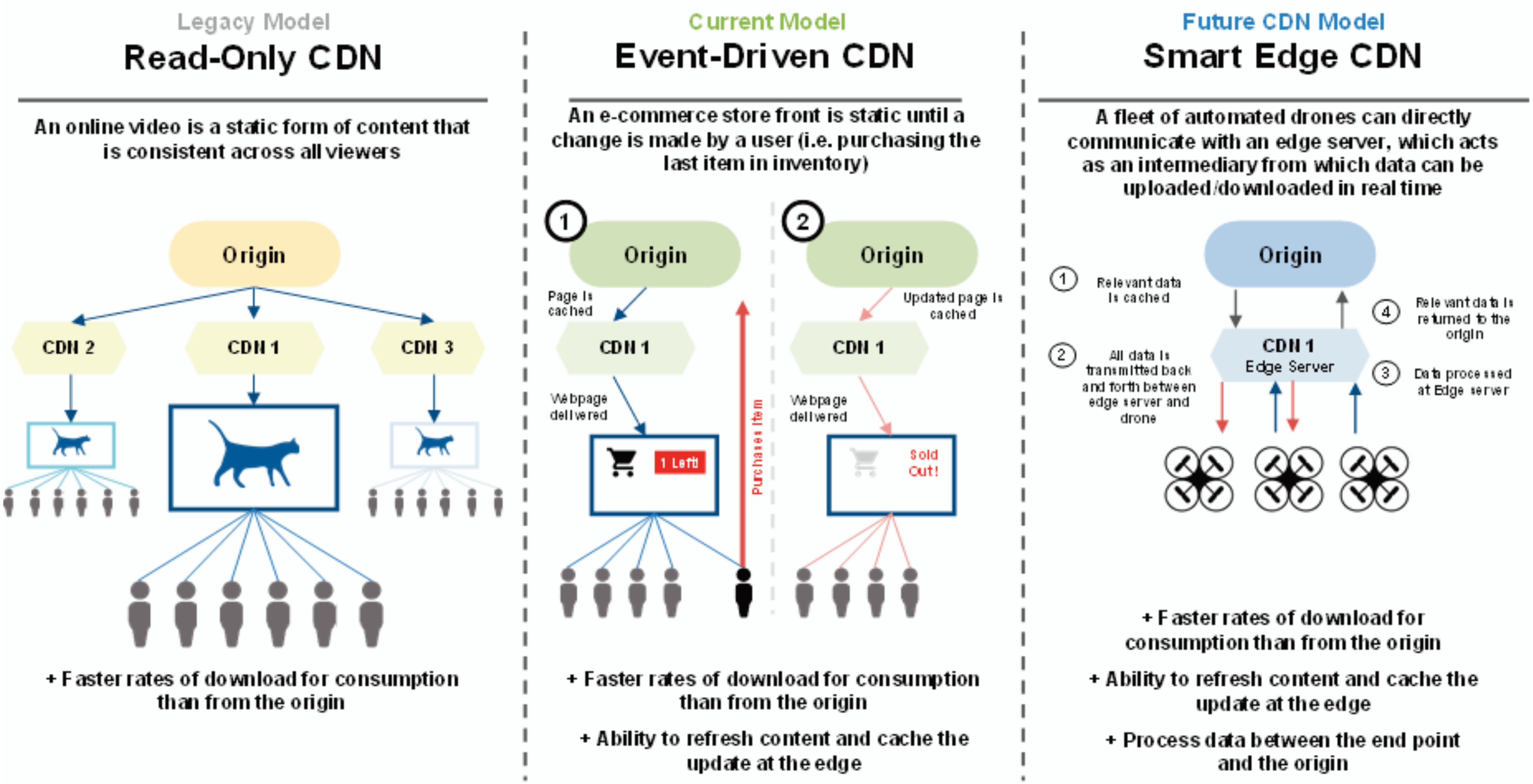
Edge servers are simultaneously an extension of public cloud and an emulation of the services provided by the public cloud running on hardware at the edge, in an edge + cloud computing paradigm; we believe that edge servers will need to be placed near connected devices to supplement public cloud capabilities, given that the inherent limitations of public cloud – requirement for connectivity, latency, bandwidth limitations, and security concerns – preclude a variety of use cases. Edge servers will effectively be micro data centers, including all required IT functionalities in data centers (e.g. uninterruptible power supply, servers, storage, networking, and cooling); in contrast to a traditional data center, however, these edge servers are self-contained and mobile, able to be easily moved and operated with a minimal amount of external inputs outside of power, networking, and airflow. We would expect edge servers to be virtualized devices with built-in compute, storage, and networking capabilities, with the ability to communicate with edge devices via single-hop wireless connections, including WiFi or Bluetooth, as well as with public cloud via a high-speed internet connection.

CDNs

Content-delivery networks (CDNs) are the natural precursors to edge computing. With CDNs, static content is cached and delivered from geographically distributed edge servers. By pre-positioning content at the edge – geographically closer to the end user – CDNs allow for faster and smoother delivery of content. We note, however, that primary purpose of a CDN is localized *storage*, as CDNs are typically not designed for localized *compute*. CDNs are physical networks comprised of geographically distributed servers, which accelerate the delivery of files, media, and webpages with the objective of improving the experience for the end-user. CDNs do this by 'caching' content obtained at the origin, at the edge of the internet (often in data centers operated by last-mile internet service providers), which limits the distance that these packets of information must travel to reach the endpoint. Further, these networks dynamically assign resources based on congestion and operating conditions in order to optimize performance.

The primary objective for CDNs have always been to reduce bandwidth requirements and latency; however, up to this point, this has generally been oriented towards storing static content at the edge, rather than providing localized compute resources. The next generation of content delivery networks however, could integrate processing capabilities into the existing nodes in order to bypass congestion and improve latency further by handling certain requests closer to the users it serves, creating a logical extension of the business model into edge computing. While we have yet to see a fully commercialized offering from the major CDN players such as Akamai and Fastly, we believe these companies could be among the early players in this market.

Exhibit 12: Future CDNs could begin to incorporate compute capabilities
Evolution of CDN models

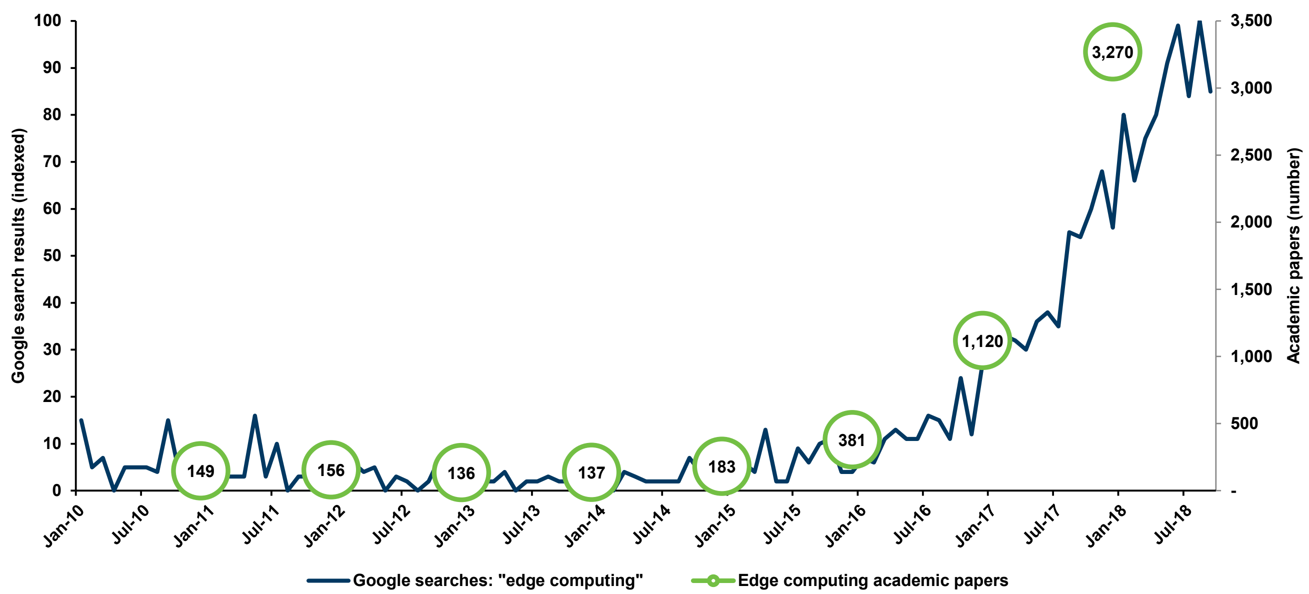


Source: Goldman Sachs Global Investment Research

Edge computing demand drivers

Over the past 18-24 months, we have seen interest and mentions of edge computing increase sharply. In terms of the number of Google searches for the term "edge computing," as well as the number of edge computing academic papers being written on edge computing. In 2016, there were ~3x the number of papers on edge computing as there were in 2015, and in 2017, the number tripled again.

Exhibit 13: Interest in edge computing is taking off
Google searches for "edge computing" and edge computing academic papers



Source: Google, Fastly, Goldman Sachs Global Investment Research

On conference calls and at analyst days, we have also picked up increasing mentions of the rise of edge computing, as well as the growing realization of the importance of hybrid cloud even in a public cloud world.

"The mobile-first, cloud-first world evolving to this intelligent cloud and intelligent edge world we think is definitely what we're going to be talking for years to come." *Satya Nadella, 2017 financial analyst day*

"I think of our servers as the edge of our cloud, and as I said there's a huge software asset in there which is becoming increasingly competitive." *Satya Nadella, Microsoft F3Q15 conference call*

"Microsoft has bet on a strategy to build a hyper-scale public cloud as well as reinvent servers as the edge of our cloud." *Satya Nadella, Microsoft F1Q16 conference call*

"You need real-time command and control, data aggregation, alerting…having compute at the edge, close to the device." *Frank Leighton, Akamai F1Q18 conference call*

"We'd also emphasize that some of the new IoT and edge use cases tend to bring things back on-premise, where now customers sort of say, oh, I can't round-trip to the cloud if I need this latency or have that amount of bandwidth as well. So we believe all of these indicate a very robust hybrid environment, where it's going to be a combination of on-premise, as well as in the cloud private and public." *Pat Gelsinger, VMware F3Q18 conference call*

"In looking to the future, we see Edge computing as a significant adjacent opportunity." *Pat Gelsinger, VMware F2Q18 conference call*

"And it has an architecture where it runs partly on premise, and that's one of the reasons it's able to do everything that it can do from an integration layer. From Salesforce's core platform, we're still 100% public cloud. I don't see that changing. There's going to be little instances here and there, especially when we acquire a company like MuleSoft or maybe other things in the future…The idea that, look, we're not attached to any kind of religious dogma around the cloud. We're going to do what's best for our customers and what's best for our company. And in the case of MuleSoft, I think it very much reflects that vision, that idea, that we're going to be able to deliver the best Integration Cloud." *Marc Benioff, Salesforce F1Q19 conference call*

"The second trend that we've seen is around moving that inference – taking trained models and deploying them into connected devices to run them at the edge…you still want that intelligence to operate on the device, even if it's disconnected from the cloud." *Matt Wood, GM Deep Learning and AI, Amazon Web Services, re:Invent 2017*

Device resource limitations

Almost by definition, edge devices whose main purpose is collection of video, audio, sensor, or text data have more limited hardware resources relative to full-fledged servers in a data center. As a result, the edge device is typically limited in the amount of processing by the on-board hardware; this includes battery/energy consumption, which may be effectively a limitless resource for data centers but is typically a finite (and one of the most precious) resource for edge devices.

As a result, if complex analytics need to be performed, front-end devices, faced inherent processing and power limitations, may not be able to complete the task; edge servers, located near the edge device, would be perfectly positioned to run the analytics, given the constant availability of power (energy), as well as compute resources orders of magnitude above what the edge device is able to offer. Edge nodes would also be able to act as triage centers, quickly providing not only the results required to the edge device but also analyzing and filtering raw data, and only uploading relevant portions to the public cloud, where truly compute-intensive analytics, such as machine learning or AI, can reason over the data to refine the algorithm.

Latency

For use cases where reaction time is critical to the success of the overall system, the latency inherent with a round trip to the cloud via a hub-and-spoke model may be not be acceptable. Latency can be influenced by a plethora of uncontrollable factors, including the network connectivity of the location, the network provider, other network traffic, as well as the specific region, availability zone, and data center that the user connects to.

According to a white paper by Interxion, a provider of carrier and cloud-neutral colocation data center services, decreased latency has a direct, measurable impact on overall system performance. For instance, every 20ms of network latency results in a 7-15% decrease in page load times, and for e-commerce, page load times are correlated to web traffic and sales. A 500ms delay can cause a 20% drop in Google's traffic, while just a 100ms delay can cause a 1% drop in Amazon's sales. Real-time video applications (e.g. a visual guiding service on a wearable camera) typically demand a latency better than 25-50ms, meaning that a round-trip to the public cloud, plus processing time, is typically too long.

Although network latency continues to improve, physics dictates that further improvements will be asymptotic, tapering off as latency approaches theoretical maximums. In the exhibit below, we note that the maximum speed of light in fiber results in a 56ms round-trip time between New York and London – and this does not take into account real-world fiber performance, time for local network routing, and compute times.

Exhibit 14: Network connectivity speeds have hard limits based on the speed of light and geographical distances

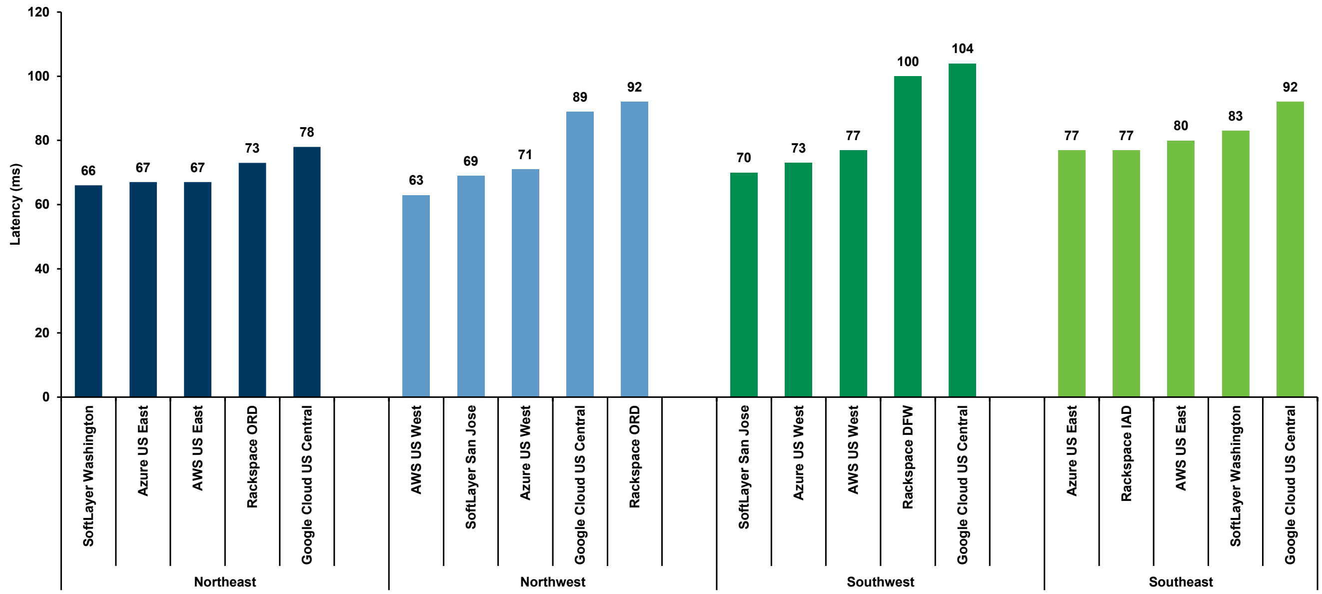
Theoretical "speed limits" of light in fiber

Route	Distance	Time (light in vacuum)	Time (light in fiber with refractive index of 1.5)	Round-trip time (RTT) in fiber
New York to Washington DC	177 mi	1 ms	1 ms	3 ms
New York to San Francisco	2,569 mi	14 ms	21 ms	41 ms
New York to London	3,465 mi	19 ms	28 ms	56 ms
New York to Sydney	9,946 mi	53 ms	80 ms	160 ms
Equatorial circumference	24,901 mi	134 ms	201 ms	401 ms

Source: Goldman Sachs Global Investment Research

To take into account real-world latency times, network-monitoring company Cedexis (since acquired by Citrix) and Network World tested the latency of five major IaaS providers (Amazon Web Services, Microsoft Azure, Google Cloud, IBM SoftLayer, and Rackspace) across four regions of the United States. Within each region, the fastest IaaS providers generally had latencies of 60-70ms, with the lowest latency in the northwest, at AWS US West (63ms).

Exhibit 15: IaaS vendors have a minimum latency of 63ms
Latency of five major IaaS providers across four regions of the United States



Source: Cedexis (Citrix), Network World

We note, however, that the 63ms latency (or 126ms round-trip) does *not* account for any computing or processing time.

Network connectivity & reliability

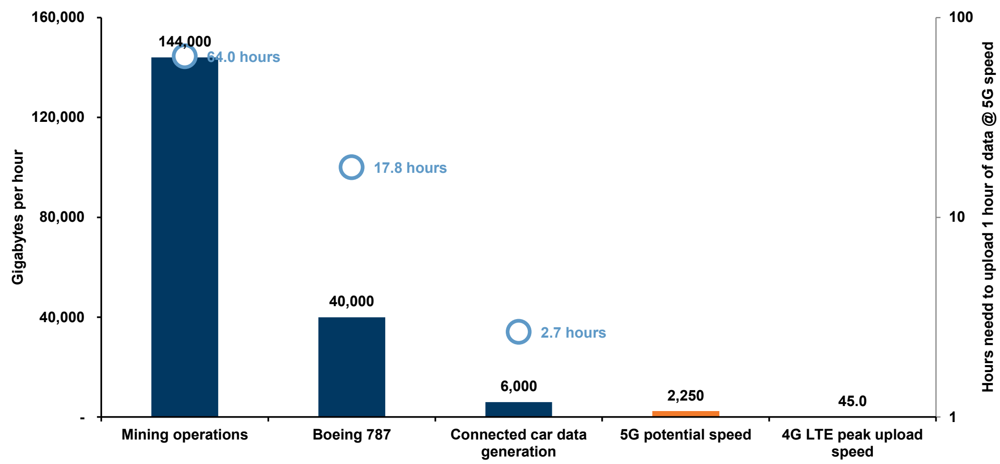
Dependence on public cloud for all data processing and analytics may not be suitable for many use cases, particularly those that feature low or intermittent network connectivity. For instance, physical obstructions (buildings, hills, forests), interference, or atmospheric conditions (bad weather) may result in poor connection, making it critical, for use cases like a connected car, for processing to be local and unaffected by network connectivity.

Bandwidth & storage

With the advent of public cloud, the speed of compute and data processing has far outclassed network bandwidth. With billions of devices generating hundreds-to-thousands of gigabytes of data every second, bandwidth (i.e. the ability to transmit the data to public cloud) and storage (i.e. the ability to retain the data in the public cloud) become impossible, as the sheer quantity of data produced will overwhelm even public cloud capabilities. By the time the data arrives, it will already be stale, and its value will have eroded dramatically.

Former Intel CEO Brian Krzanich estimates that one average connected autonomous car will generate 4,000 gigabytes of data per hour; Microsoft CEO Satya Nadella's estimate is similar – 6,000 gigabytes of data per hour. From a bandwidth perspective, even 5G networks, which are anticipated to become available in the near-future, are expected to have speeds of potentially 10 giga*bits*/second – which would equate to just 2,300 gigabytes per hour at full capacity– less than half of what would be required to continuously upload the autonomous car's data. In these cases, the data clearly needs to be processed at the edge for timely insights and to alleviate network congestion.

Exhibit 16: Even 5G bandwidth is inadequate to upload the vast quantities of data generated by IoT devices
Data generation/capability



Source: Cisco, Microsoft, Goldman Sachs Global Investment Research

Truly big data use cases will also create massive data generation, orders of magnitude above what could be transmitted back to the public cloud; in fact, these big data use cases will generate sufficient data that simply *storing* it, even with the resources of the public cloud (assuming that the data can be transmitted there), will be challenging.

As every electrical device from lightbulbs to jet engines becomes connected, billions of sensors will each be producing tremendous amounts of raw data. Pratt & Whitney's newest Geared Turbo Fan (GTF) jet engines contain 5,000 sensors apiece (50x more sensors than their predecessors), with each engine generating 10 gigabytes of data every second (i.e. 36 terabytes of data an hour) of flight time; the GTF engine leverages AI in conjunction with this data to predict the demands of the engine to adjust thrust levels, and as a result, GTF engines have the potential to reduce fuel consumption by 10-15%, while simultaneously decreasing engine noise and emissions. A 12-hour flight in a twin-engined aircraft could therefore generate 864 terabytes of data, and Pratt & Whitney have an order book of more than 7,000 engines. For context, in 2012, Facebook revealed that its systems processed 500 terabytes of data per day.

Cisco estimates that a Boeing 787 aircraft could generate 40 terabytes of data every hour in flight, and mining operations (including status, performance, and condition data from sensors and devices in mining equipment and transport vehicles) generate 2.4 terabytes of data in a minute. Even *if* networks had the capacity to transfer this amount of data, despite the seemingly endless capacity of the public cloud compared to the compute and storage needs of a single application, every piece of data that is stored in the public cloud still ultimately 1) requires hardware capacity and 2) represents a cost to the enterprise storing the data. By placing an edge server at the source of data collection (e.g. in the airplane), however, the edge server can quickly process the data (e.g. running the analytics and algorithms needed to increase fuel efficiency, decrease engine noise, and lower emissions), discard the vast majority of the data, and stream only the necessary portions of the data to the data center or public cloud (i.e. anomalies or engine maintenance requirements). One of the prime benefits of edge computing, therefore, is the ability to consume and process the data at the edge of the cloud, discard the data that does not need to be kept long-term. As a result, the vast majority of the data produced by edge devices will never be transmitted to public cloud, helping to ensure that the public cloud does not become a data landfill, indefinitely storing the plethora of data generated by IoT devices.

Security & privacy

Processing the data on the device or at edge, versus uploading raw data to the public cloud, yields superior results for security and privacy, as there are inherent risks in transmission. For instance, in use cases where video is captured by the edge device, if the edge device is capable of doing pre-processing (e.g. masking all the faces in the video), privacy concerns may be partially assuaged; if all of the processing happens in the device – the video never physically leaves the device and only the required, distilled data is passed to the public cloud – then privacy concerns could be dramatically alleviated. Regulatory issues, including data residency, could also potentially be addressed by leaving the data at the source of generation.

Furthermore, we would note that edge computing would tend to disaggregate information, preventing the concentration of information relative to a cloud computing paradigm that simultaneously makes it an attractive target and makes breaches disastrous. Cloud security research on proper protection and encryption of fragmented data, coupled with decentralized overlay technologies could help ensure data security for regulated and sensitive data.

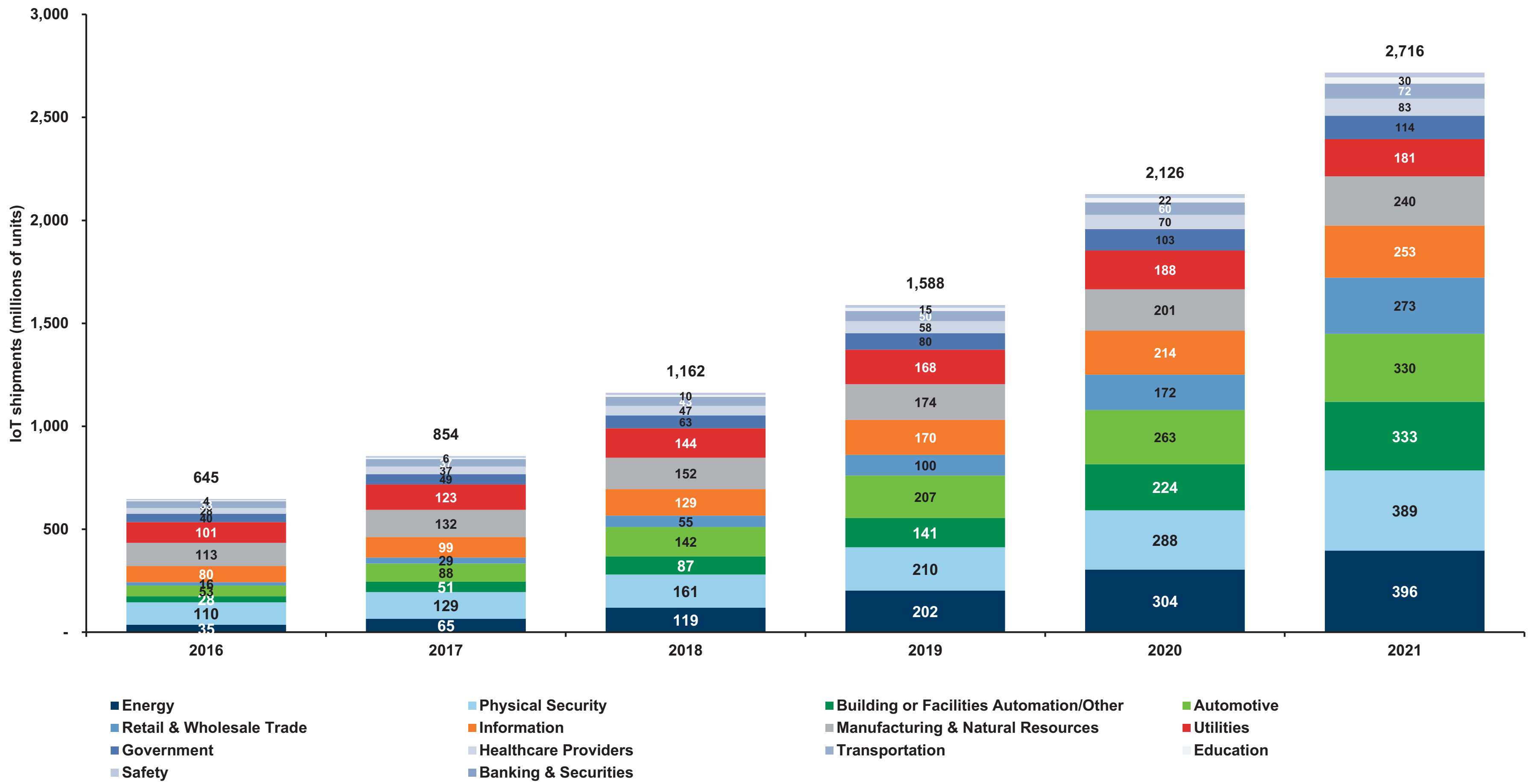
Sizing the potential market opportunity for virtualization and server operating systems

Our view is that edge computing is simply simultaneously an extension of public cloud and an emulation of the services provided by the public cloud running on hardware at the edge; as such, this market is difficult to size as it likely encapsulates both on-premise infrastructure software and public cloud spending.

We therefore evaluate the potential *incremental* infrastructure software spend that could be attributed to an increase in edge servers, driven by the need to perform processing closer to the source of data generation. According to Gartner, IoT shipments (enterprise-only; excluding consumer) will grow at a 33% CAGR from 2016 through 2021, or from 645mn units to 2.72bn units.

Exhibit 17: IoT shipments are growing at a 33% CAGR through 2021 – we believe that edge servers will be required to manage them

Enterprise (ex-consumer) IoT shipments by year vertical/cross-industry use case



Source: Gartner, Goldman Sachs Global Investment Research

With 2.72bn IoT endpoint (i.e. the connected "things" themselves) shipments in 2021, we conservatively assume that only 50% will be connected to an edge server – the remaining are either designed to function completely offline or they are connected directly to the public cloud, without ever connecting to an edge server. Furthermore, we conservatively do not assume any incremental software revenue from consumer products; we note that consumer automotive (consisting of automotive subsystems and connected cars), would likely require on-board compute and would thus be additive to this market size estimate.

We then examine three different possibilities, where there are either 1,000, 500, or 200 IoT endpoints connected to a single edge server. Given that AWS Greengrass Groups (software that allows AWS users to run local compute, including Lambda, as well as messaging, syncing, and machine learning) are designed to represent (for instance) a floor of a building, a single truck, or a home, we believe that 1,000 is likely the most that a single edge server, with a single physical core, could support; this is our most conservative case, as a high number of IoT endpoints per server implies a lower number of incremental edge servers required. On the other end of the spectrum, we assume that each edge server supports just 200 IoT endpoints; we note that AWS Greengrass Groups have a limit of 200 AWS IoT devices and 200 Lambda functions.

For each edge server required, we assume that at a minimum, the edge server infrastructure software consists of 1) virtualization, and 2) a server operating system. For the virtualization portion of the stack, we note that the list price of a single VMware vSphere Enterprise Plus license (not including maintenance and support) is $3,495. The vSphere license entitles the user to a single physical processor, which matches our (conservative) assumption of a single edge server with a single core. For the operating system portion of the stack, we note that the list price of a Red Hat Enterprise Linux Server subscription is $799 per year, and Red Hat's license requires a subscription per socket pair per server node. We note, however, that in many cases, particularly with enterprise agreements with software vendors, companies may receive discounts from the list price.

For VMware, we also account for the maintenance revenue attached to new licenses, which is listed at ~25% of the price of the license (annually). We discount this annual maintenance revenue stream at VMware's WACC (~10%), plus estimated gross churn (~3%), and add it to the incremental new license revenue spend to arrive at the estimated annual value of incremental spending on virtualization.

For Red Hat, we discount the annual subscription revenue stream at Red Hat's WACC (~11%), plus estimated gross churn (~5%), to arrive at the estimated annual value of incremental spending on sever operating systems.

Using these assumptions, we estimate that in the most conservative scenario (1,000 IoT endpoints per edge server), the incremental annual incremental spend would be $14bn for virtualization and $7bn for server operating systems; in the most aggressive scenario (200 IoT endpoints per edge server, or a lower density of IoT endpoints per edge server, equating to more servers required for the same number of IoT endpoints), the incremental annual license spend would be $69bn for virtualization and $34bn for server operating systems. This incremental spend would primarily be driven by use cases like energy, physical security, and building/facilities automation, and industries like retail, manufacturing, and utilities, as Gartner forecasts the highest number of IoT endpoints in these areas.

We note, however, that these estimates likely skew conservative, as it does not account for other infrastructure software like NoSQL databases, which could potentially be a lightweight option for edge computing; nor does it account for analytics and application software, which will depend heavily on the types of use cases leveraged for edge computing resources. Finally, we believe that container adoption could serve as a multiplier for spending, as Red Hat has commented that OpenShift is "almost 20x the price of RHEL on the same two-socket server."

We walk through our calculations below in Exhibit 18.

Exhibit 18: We estimate that the incremental infrastructure software spend needed to support edge servers could reach $100bn

Virtualization and server operating system spend needed for IoT endpoints and edge servers

IoT market size (2021)				IoT endpoints per server: 1,000			IoT endpoints per server: 500			IoT endpoints per server: 200		
	Number of IoT endpoint shipments (mn)	% connected to an edge server	Total hardware spending ($ mn)	Number of servers (mn)	VMware vSphere Enterprise Plus ($3,495 list) spend ($ mn)	RHEL Standard ($799 list) spend ($ mn)	Number of servers (mn)	VMware vSphere Enterprise Plus ($3,495 list) spend ($ mn)	RHEL Standard ($799 list) spend ($ mn)	Number of servers (mn)	VMware vSphere Enterprise Plus ($3,495 list) spend ($ mn)	RHEL Standard ($799 list) spend ($ mn)
Consumer	**4,921.1**		**$1,766,991**	**-**	**$0**	**$0**	**-**	**$0**	**$0**	**-**	**$0**	**$0**
Automotive	1,183.9	0%	$1,464,058	-	$0	$0	-	$0	$0	-	$0	$0
Health and Fitness	212.9	0%	$31,766	-	$0	$0	-	$0	$0	-	$0	$0
Home Automation/Other	1,070.7	0%	$92,324	-	$0	$0	-	$0	$0	-	$0	$0
Home Energy Management	493.0	0%	$21,223	-	$0	$0	-	$0	$0	-	$0	$0
Home Security and Safety	659.6	0%	$12,997	-	$0	$0	-	$0	$0	-	$0	$0
Information and Entertainment	1,300.9	0%	$144,623	-	$0	$0	-	$0	$0	-	$0	$0
Cross-Industry	**1,723.0**	**50%**	**$658,704**	**0.9**	**$3,011**	**$688**	**1.7**	**$6,022**	**$1,377**	**4.3**	**$15,054**	**$3,442**
Automotive	329.8	50%	$371,131	0.16	$576	$132	0.33	$1,153	$264	0.82	$2,882	$659
Building or Facilities Automation/Other	333.0	50%	$155,185	0.17	$582	$133	0.33	$1,164	$266	0.83	$2,910	$665
Energy	396.5	50%	$19,955	0.20	$693	$158	0.40	$1,386	$317	0.99	$3,464	$792
Information	253.0	50%	$71,228	0.13	$442	$101	0.25	$884	$202	0.63	$2,210	$505
Physical Security	389.2	50%	$39,989	0.19	$680	$156	0.39	$1,360	$311	0.97	$3,401	$778
Safety	21.5	50%	$1,215	0.01	$37	$9	0.02	$75	$17	0.05	$187	$43
							0					
Vertical-Specific	**992.9**	**50%**	**$964,084**	**0.5**	**$1,735**	**$397**	**1.0**	**$3,470**	**$793**	**2.5**	**$8,675**	**$1,983**
Banking & Securities	0.3	50%	$1,855	0.00	$1	$0	0.00	$1	$0	0.00	$3	$1
Education	30.4	50%	$6,949	0.02	$53	$12	0.03	$106	$24	0.08	$266	$61
Government	113.6	50%	$195,904	0.06	$198	$45	0.11	$397	$91	0.28	$992	$227
Healthcare Providers	83.3	50%	$32,403	0.04	$146	$33	0.08	$291	$67	0.21	$728	$166
Manufacturing & Natural Resources	239.5	50%	$160,016	0.12	$419	$96	0.24	$837	$191	0.60	$2,093	$478
Retail & Wholesale Trade	273.2	50%	$14,391	0.14	$477	$109	0.27	$955	$218	0.68	$2,387	$546
Transportation	72.0	50%	$528,755	0.04	$126	$29	0.07	$252	$58	0.18	$629	$144
Utilities	180.5	50%	$23,810	0.09	$315	$72	0.18	$631	$144	0.45	$1,577	$361
Total	**7,636.9**		**$3,389,779**	**1.4**	**$4,746**	**$1,085**	**2.7**	**$9,492**	**$2,170**	**6.8**	**$23,730**	**$5,425**
License list price					$3,495			$3,495			$3,495	
1-year maintenance list					$874			$874			$874	
Maintenance % of license					25.0%			25.0%			25.0%	
Annual new maintenance/subscription revenue					**$1,187**	**$1,085**		**$2,374**	**$2,170**		**$5,934**	**$5,425**
Discount rate					10%	11%		10%	11%		10%	11%
Gross churn assumption					3%	5%		3%	5%		3%	5%
PV of maintenance/subscription revenue					**$9,129**	**$6,781**		**$18,259**	**$13,562**		**$45,647**	**$33,906**
License total					$4,746	$0		$9,492	$0		$23,730	$0
Maintenance/subscription total					$9,129	$6,781		$18,259	$13,562		$45,647	$33,906
Total annual value					**$13,875**	**$6,781**		**$27,751**	**$13,562**		**$69,377**	**$33,906**

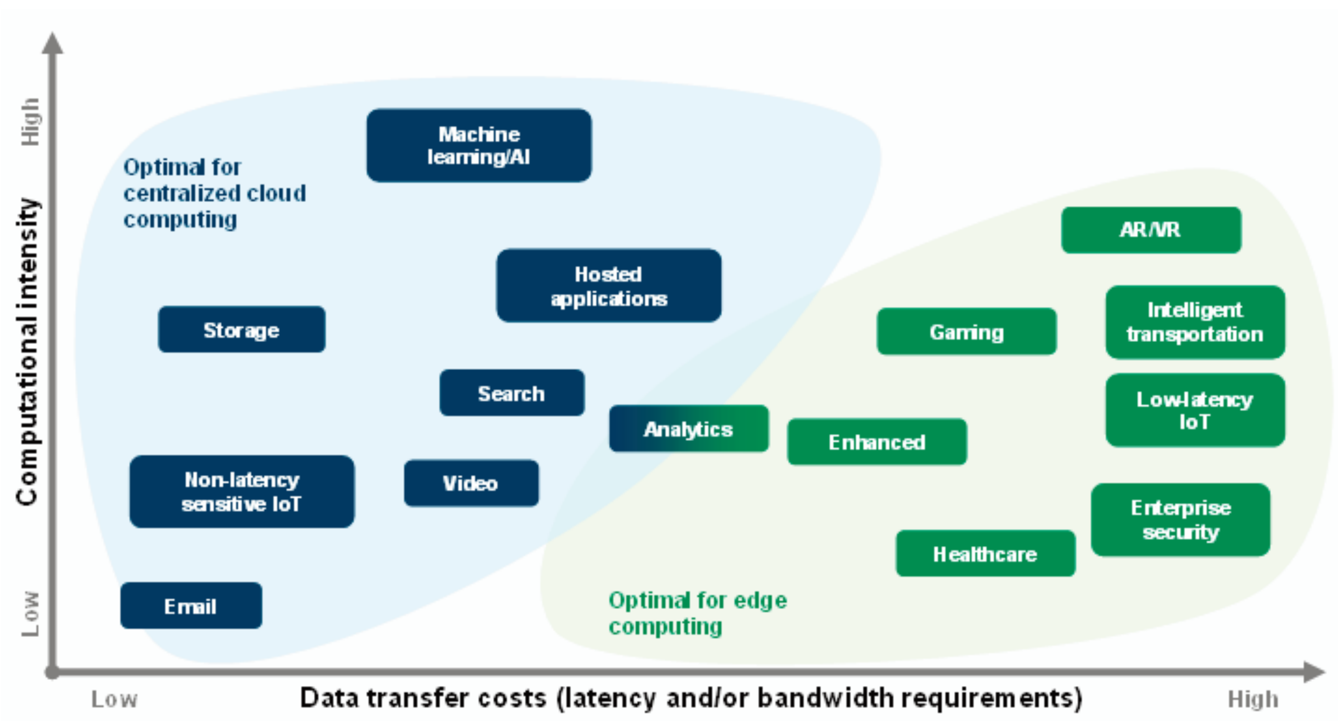
Source: Gartner, Goldman Sachs Global Investment Research

Edge computing vs. cloud computing performance

As we have previously noted, we believe that cloud computing and edge computing are complementary, as opposed to competing, architectures. While cloud computing aggregates compute resources into highly centralized and scalable resources and edge computing disperses these resources, our view is that there is a need for both these modes, as computing will become increasingly pervasive. Edge computing helps to deliver the public cloud's elastically scalable services where the public cloud is either inaccessible or too distant.

The public cloud, whether delivered by Amazon Web Services, Azure, or Google Cloud Platform, will continue to be densely packed with cutting edge servers, storage devices, and networking equipment. With elastic scaling, or the ability to horizontally add additional compute, storage, or networking resources as the need arises, the processing power of the public cloud will be essentially immeasurably more vast than a single edge server. As a result, the public cloud will continue to be uniquely suited to computationally intensive tasks, including storage, reasoning over large data sets (e.g. machine learning), and hosted applications. However, given the physical distance of the public cloud, it is suitable only for tasks that do not require latency of under 100-200 milliseconds or excessive bandwidth (i.e. requires large datasets to be sent to the public cloud). For these types of use cases, including AR, transportation, and low-latency IoT, an edge server, located near the source of data, is more suitable.

Exhibit 19: Some workloads will continue to be most effectively run in the public cloud; some are more suitable for edge computing
Workloads for public cloud vs. edge computing



Source: Goldman Sachs Global Investment Research

For a given task, the time to completion is a function of both 1) the processing power available (favoring the public cloud), and 2) the latency/bandwidth of the connection to the processing source (favoring edge computing); there is, however, a fundamental tradeoff between processing power and latency/bandwidth. We would expect that for highly computationally-intensive use cases, the efficiencies gained by processing in the public cloud would overwhelm latency/bandwidth concerns; conversely, for highly data-intensive use cases, the time needed to upload to the public cloud would overwhelm the benefits gained by more powerful public cloud computing resources.

In a 2013 Carnegie Mellon University paper *(The Impact of Mobile Multimedia Applications on Data Center Consolidation)*[1], the researchers experimented, using real-world use cases, with the balance between consolidated public cloud compute resources against latency-sensitive and resource-intensive applications. While on campus in Pittsburgh, Pennsylvania, the researchers tested six distinct use cases (facial recognition, speech recognition, object & pose identification, augmented reality, and a physics simulation) that would potentially be suitable for edge computing on six different types of infrastructure, ranging from mobile to edge to public cloud, to test the total performance, including processing and transmission.

[1] https://www.cs.cmu.edu/~satya/docdir/ha-ic2e2013.pdf
Satyanarayanan, Mahadev, Carnegie Mellon University, et al. "The Impact of Mobile Multimedia Applications on Data Center Consolidation." 2013 IEEE International Conference on Cloud Engineering (IC2E), 2013, doi:10.1109/ic2e.2013.17.

Exhibit 20: The Carnegie Mellon paper evaluated six different types of use cases…
Carnegie Mellon paper use cases

Use case	Details	Average request size	Average response size
1) Facial recognition Detects and identifies faces	▪ Detects faces in an image (using a Haar Cascade of classifiers) ▪ Attempts to identify the face from a pre-populated database (using the Eigenfaces method based on principal component analysis) ▪ Implementation based on an OpenCV image processing and computer vision routines ▪ Runs in a Windows environment ▪ Experiment considers a pre-trained system (training the classifiers and populating the database are done offline	62 KB	< 60 bytes
2) Speech recognition Extracts text from digitized audio	▪ Input digitized audio of a spoken English sentence ▪ Attempts to extract all of the words in plain text format ▪ Uses an open source speech-to-text framework based on Hidden Markov Model recognition systems ▪ Single-threaded application ▪ Application is written in Java (can run on Linux and Windows); for the purposes of the experiment, runs on Linux	243 KB	< 50 bytes
3) Object and pose identification Identifies known objects in an environment & recognizes the position and orientation relative to the user	▪ Identifies and locates known objects in a scene (extracts key visual elements [SIFT features] from an image and matches against a database of features from a known set of objects) ▪ Performs geometric computations to determine the pose of the identified object ▪ Database is populated with thousands of features extracted from more than 500 images of 13 different objects ▪ Application based on a computer vision algorithm originally developed for robotics; modified for use by handicapped users	73 KB	< 50 bytes
4) Augmented reality Identifies buildings and landmarks & label them	▪ Displays timely and relevant information as an overlay on top of a live view of a scene (e.g. street names, restaurant ratings, directional arrows overlaid on a scene captured via a smartphone camera) ▪ Extracts a set of features from the scene image ▪ Uses the feature descriptors to find similar-looking entries in a database constructed using features from labeled images of known landmarks and buildings ▪ Database search is kept tractable by spatially indexing the data by geographic locations, with searches limited to a slice of the data relevant to current GPS coordinates ▪ Application uses a dataset of 1,005 labeled images of 200 buildings as the relevant database slide ▪ Multi-threaded; runs on Windows, using OpenCV libraries and Intel Performance Primitives	26 KB	< 20 bytes
5) Physics simulation Models the motion of fluids with which the user can interact	▪ Physically models the motion of imaginary fluids, allowing users to interact (e.g. liquid in a container on a smartphone screen, with the liquid reacting as the smartphone is moved) ▪ Application backend runs a physics simulation based on predictive-corrective incompressible smoothed particles hydrodynamics method ▪ Simulates a 2,218 particle system with 20ms timesteps, generating up to 50 frames/second ▪ Implemented as multi-threaded Linux application ▪ Requires latency of ~100ms or less	16 bytes	25 KB

1KB = 1,024 bytes

Source: Carnegie Mellon University

Exhibit 21: …across six separate infrastructure types (four AWS locations)
Carnegie Mellon paper infrastructure types

Hardware	Details
Mobile Replicates a state-of-the-art mobile device	• Uses a netbook (Dell Latitude 2102) to replicate a cutting-edge mobile phone • **CPU:** Intel Atom N550, 1.5 GHz per core, 2 cores (4 threads) • **RAM:** 2 GB • **Storage:** 320 GB • **OS:** Linux, Windows • **VMM:** none
Edge server Replicates an obsolete server	• Replicates a minimal data center using a six-year old (i.e. near-obsolete) WiFi-connected server to deliberately stack the deck against the edge server • One network hop away from the mobile device (WiFi) • **CPU:** Xeon N550, 1.86 GHz, 4 cores • **RAM:** 4 GB • **VMM:** KVM
AWS – US East (N. Virginia) X-Large instance	• AWS EC2 instance, physically in N. Virginia • **CPU:** 20 Compute Unites, 8 virtual cores • **RAM:** 7 GB • **VMM:** Xen, VMware
AWS – US West (Oregon) X-Large instance	• AWS EC2 instance, physically in Oregon • **CPU:** 20 Compute Unites, 8 virtual cores • **RAM:** 7 GB • **VMM:** Xen, VMware
AWS – EU (Ireland) X-Large instance	• AWS EC2 instance, physically in Ireland • **CPU:** 20 Compute Unites, 8 virtual cores • **RAM:** 7 GB • **VMM:** Xen, VMware
AWS – Asia (Singapore) X-Large instance	• AWS EC2 instance, physically in Singapore • **CPU:** 20 Compute Unites, 8 virtual cores • **RAM:** 7 GB • **VMM:** Xen, VMware

VMM = virtual machine monitor

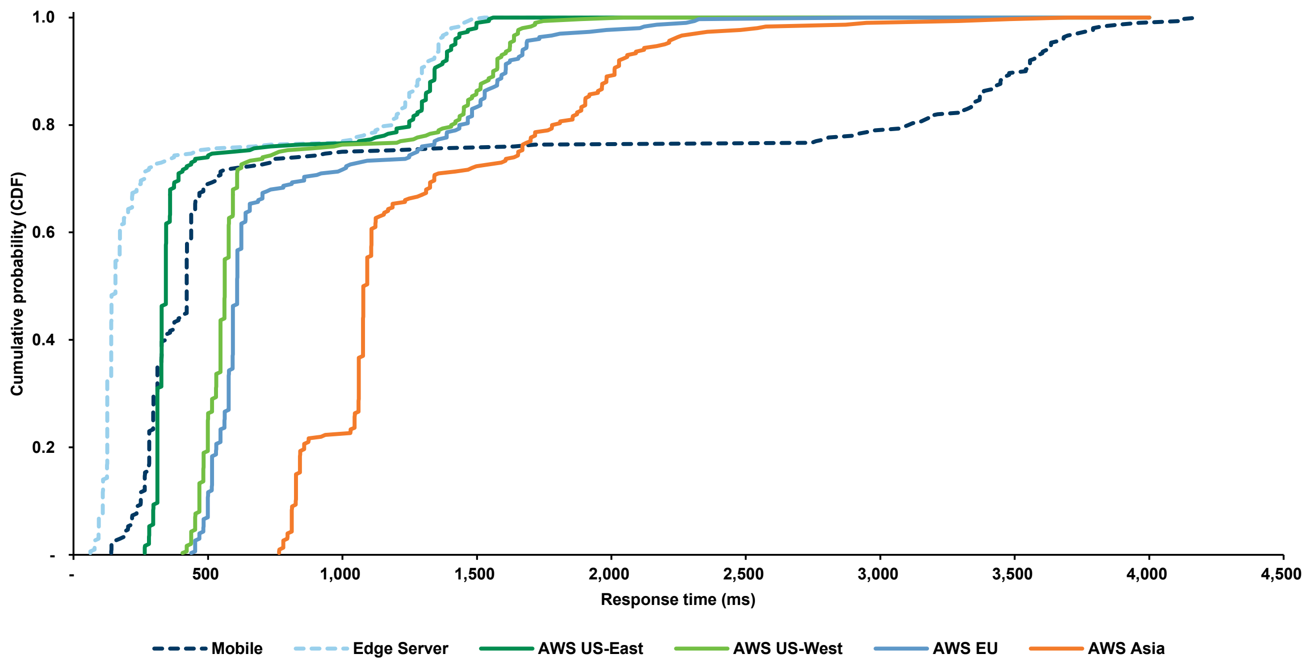
Source: Carnegie Mellon University

For the first use case, facial recognition, the researchers tested the ability of the system to process images that may have known faces, unknown faces, or no faces at all; for the images with faces, the system attempts to identify the face based on a database of faces. We note that training of the models were completed ahead of time, with the test measuring only the length of time needed to perform the recognition task on a pre-trained system.

Overall, the mobile device fared poorly: while it performed well, with tolerable response times, on single large recognizable faces, in cases where the image contained only small faces, the mobile device took upwards of 4 seconds to return the result. These types of images, which require higher levels of processing, lead to a heavy tail for the mobile device. By contrast, humans generally take just 370 milliseconds for the fastest responses to familiar faces to 620 milliseconds for the slowest response to an unfamiliar face; humans take under 700 milliseconds to determine that a scene contains no faces.

The edge computing device performed the best, with a response time of under 200 milliseconds for 66% of the images, and a worst-case response time of 1,500 milliseconds. This outperformed the cloud, with AWS US-East's *best* response times in the 250-300 millisecond range; 66% of the images were processed under 360 milliseconds. We note that for images, the data transfer costs (in terms of time) are likely high, leading to the relatively poor performance of the public cloud relative to the edge server. For this use case, as well as the others, the other AWS regions followed generally similar distributions of results, plus an additional fixed latency for the further geographic distance.

Exhibit 22: Facial recognition: edge server is faster than public cloud, given the high data transfer costs
Cumulative probability of response time (ms)



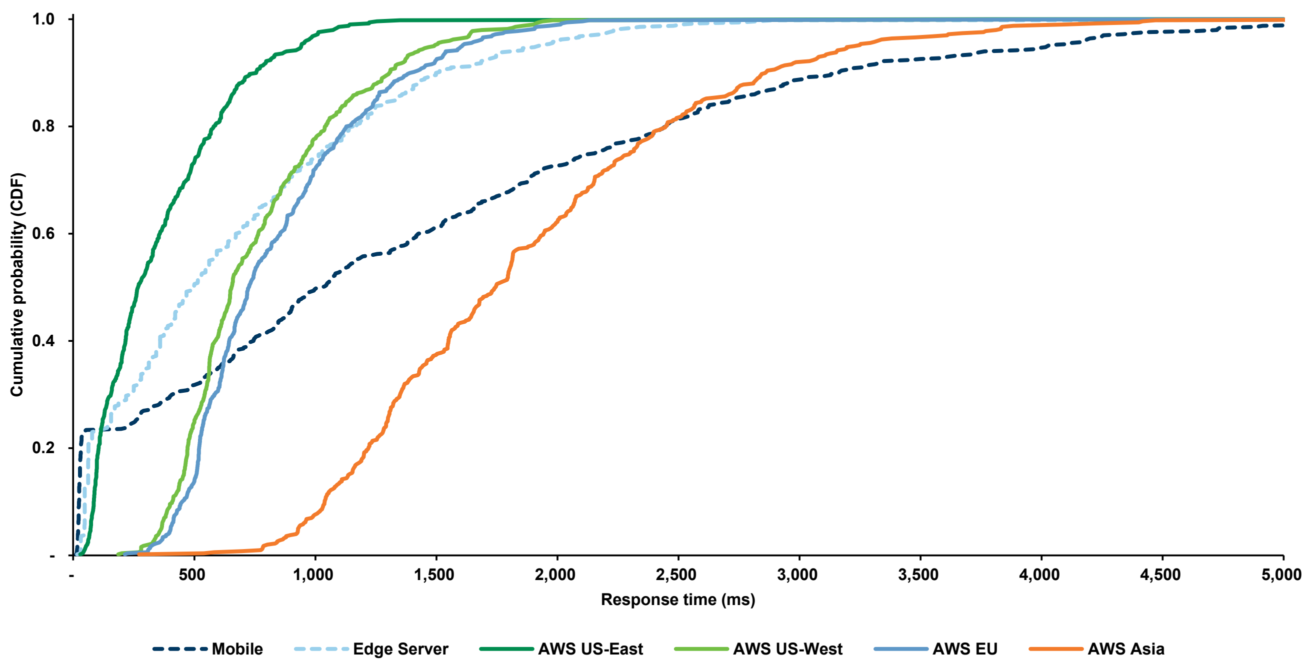
Source: Carnegie Mellon University

For the second use case, speech recognition, the researchers tested the ability of the system to extract text from a digital audio recording of a single English sentence. Similar to image recognition, speech recognition requires significant processing; however, in contrast with images, the data transfer costs of audio tend to be dramatically lower. Effectively, speech recognition incurs a lower "cost" for offloading the processing to the cloud. As a result, the response time is dominated by processing time versus data transfer time – this dynamic favors leveraging the computational prowess of the public cloud (please see our note *TAC today and "talk" tomorrow* for our views on voice search potentially upending over $150bn in search spending over the next 10 years).

The data therefore show that for speech recognition, offloading to the closest AWS region (in this case, AWS US-East, from Pittsburg) is the most efficient infrastructure, as the faster processing in the cloud outstrips the (relatively minor) latency penalty needed to upload the audio to the cloud. The edge server lagged AWS US-East in all but the easiest audio clips, although it generally compared favorably relative to the next closes AWS region (US-West) in all but the toughest audio clips. The researchers noted, however, that when they replaced the edge server with a more powerful version (i.e. an Intel i-3770 desktop), the edge server was superior to AWS US-East.

Processing purely on a mobile device, without the support of an edge server or the public cloud) is untenable for speech recognition: although 23% of the audio samples could be processed nearly instantly (<50 milliseconds), processing times for audio on a mobile device has an enormous right tail, with the worst-case scenario taking more than 5,000 milliseconds.

Exhibit 23: Speech recognition: public cloud is faster than the edge server, given the relatively low data transfer costs
Cumulative probability of response time (ms)



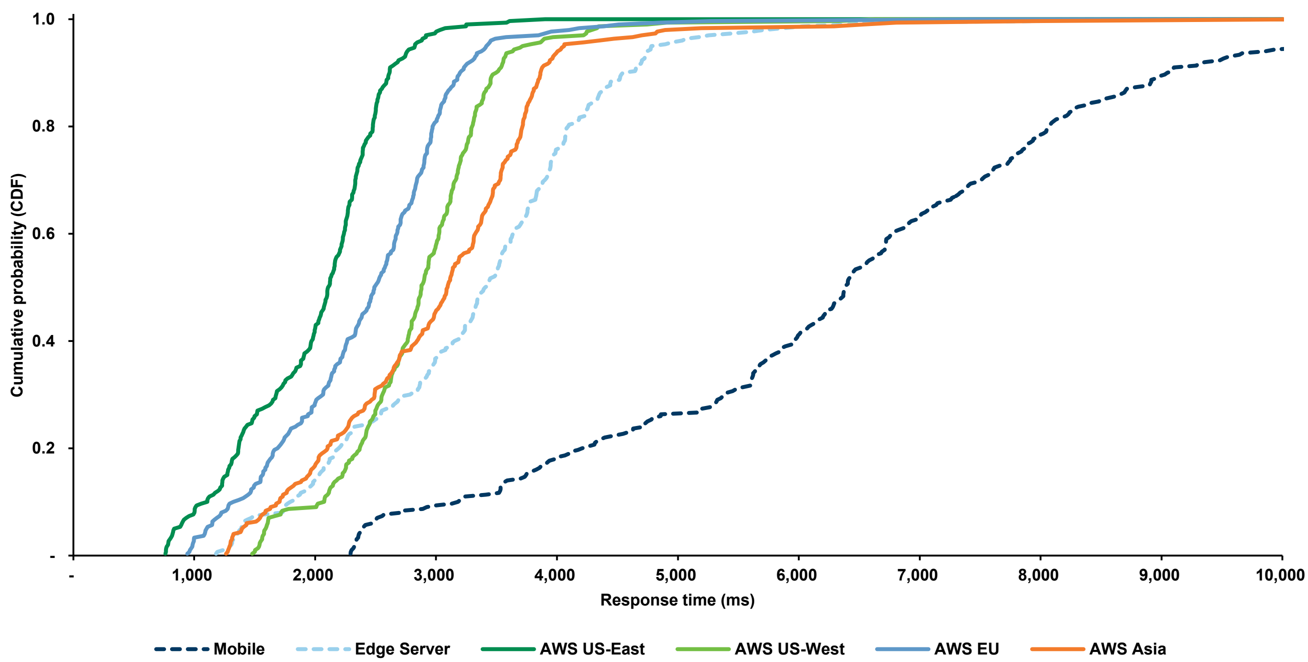
Source: Carnegie Mellon University

Object and pose identification was the most computationally intensive use case tested, and as would be expected, this tilts the scales more in favor of the public cloud. In fact, the processing load is so high that it overwhelms even the relatively robust AWS X-Large Instance, with 20 Compute Units (8 virtual cores). The best-case for the AWS instance was ~1,000 milliseconds (i.e. 1 second), with the 50th percentile taking roughly ~2,000 milliseconds (2 seconds). The researchers noted that to decrease response times to real-world acceptable levels, more than a single VM was likely required, potentially in conjunction with specialized hardware (e.g. GPUs) to expedite critical routines.

The inferior processing capabilities of the edge server led to it performing worse than all of the AWS regions, including the Asia region, demonstrating the high relative importance of computational power versus latency and bandwidth for this object and pose identification use case. Similar to speech recognition, however, when the researchers changed the edge server to the more powerful version (the Intel i-3770 desktop), the edge server was superior to AWS US-East, with 50% of the trials completed in 200 milliseconds or less.

Processing on a mobile device for object and pose identification was completely ineffective, with the best-case taking over 2,000 milliseconds; 5% of the trials took over 10,000 milliseconds (i.e. 10+ seconds).

Exhibit 24: Object and pose identification: extremely computationally-intensive, so public cloud performs the best
Cumulative probability of response time (ms)

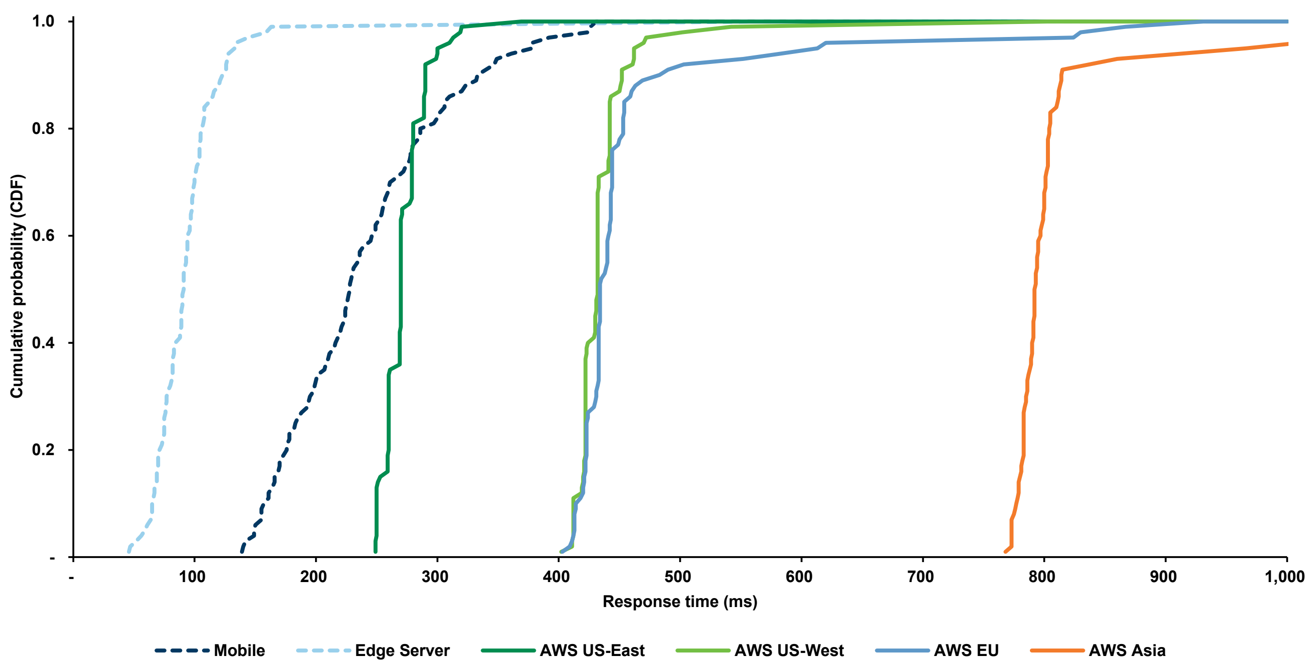


Source: Carnegie Mellon University

In the researchers' augmented reality use case, computer vision is leveraged to overlay timely and relevant information on top of a live view of a scene – for instance, street names, restaurant ratings, or directional arrows overlaid on top of a scene capture by a smartphone camera. In terms of the type of resources required, augmented reality is effectively the inverse of object and pose identification: processing costs are modest, with a low-cost feature extraction algorithm coupled with an efficient nearest-neighbor algorithm to match features in a database (constrained by GPS coordinates). While data transfer costs are high, as the image stream from the camera needs to be continually uploaded – this combination of requirements favors the edge server versus the public cloud.

As expected, local processing resources performed better, with the edge server generally completing the task in fewer than 100 milliseconds – demonstrating its suitability to provide crisp augmented reality interactions. The mobile device also generally performed well, besting the AWS EC2 instance in most cases, which took 250-300 milliseconds to complete the task – too slow for this augmented reality use case, given the need for data transfer to the public cloud. For additional details on AR and VR, please see our recent Profiles in Innovation report on Extended Reality.

Exhibit 25: Augmented reality: local devices are superior, given low processing costs and high data transfer costs
Cumulative probability of response time (ms)

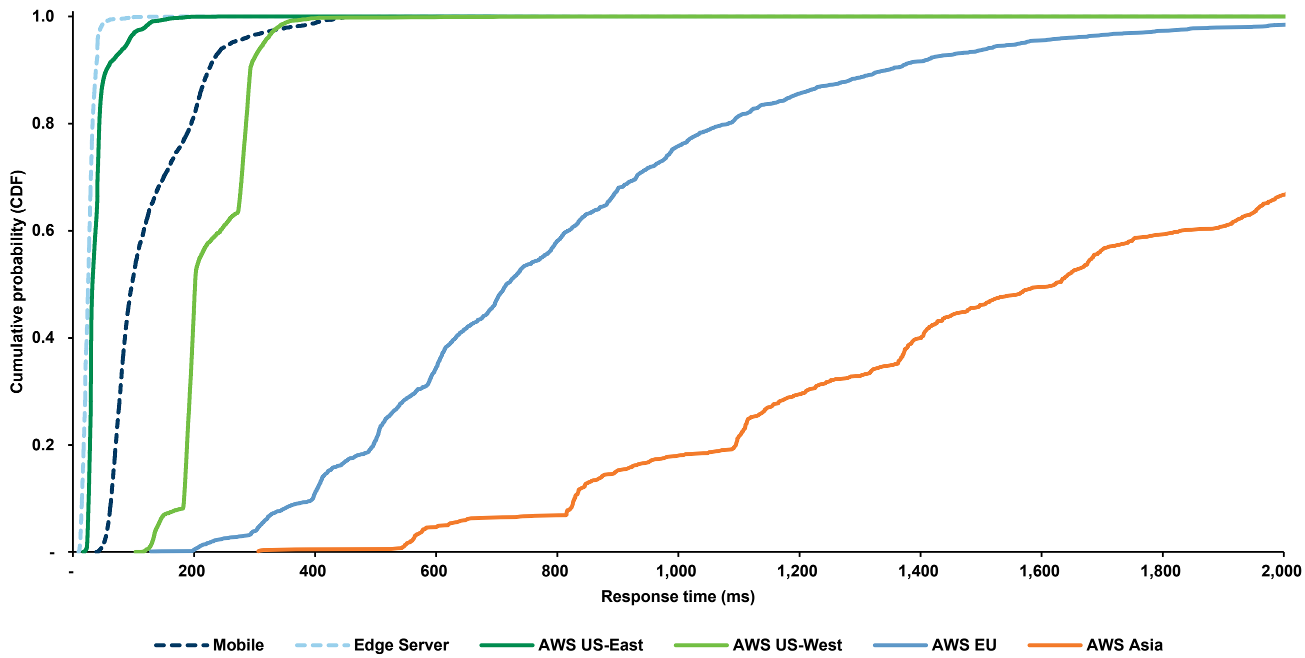


Source: Carnegie Mellon University

The final use case tested was a physics demonstration – simulating a fluid with which the user can react (e.g. a glass of water that can be moved by tilting the smartphone screen), with the response time defined as the time between the sensing of the user action (accelerometer reading) to the time that the output is reflected (water movement on the smartphone screen). The researchers noted that this process reflected three distinct steps: the network latency, the simulation and computation step, as well as the data transfer time needed to receive a frame from the simulation thread.

Although the mobile device has effectively zero network latency and data transfer time (as computation is local), its limited computational capacity results in the inability to execute the simulation quickly enough to produce a real-time simulation, with an appropriate frame rate (the researchers note that fluid motions on the mobile device were just one-fifth of realistic speeds). At the other extreme, public cloud infrastructure in distant geographies, though more than capable of producing real-time simulations, cannot deliver the results quickly enough due to network latency and data transfer time. As a result of the balance of capabilities required for this specific use case, only the edge server and AWS US-East were able to perform the simulation in real-time, with the appropriate frame rate.

Exhibit 26: Physics simulation: moderate computational and data transfer costs; both edge servers and public cloud perform well
Cumulative probability of response time (ms)



Source: Carnegie Mellon University

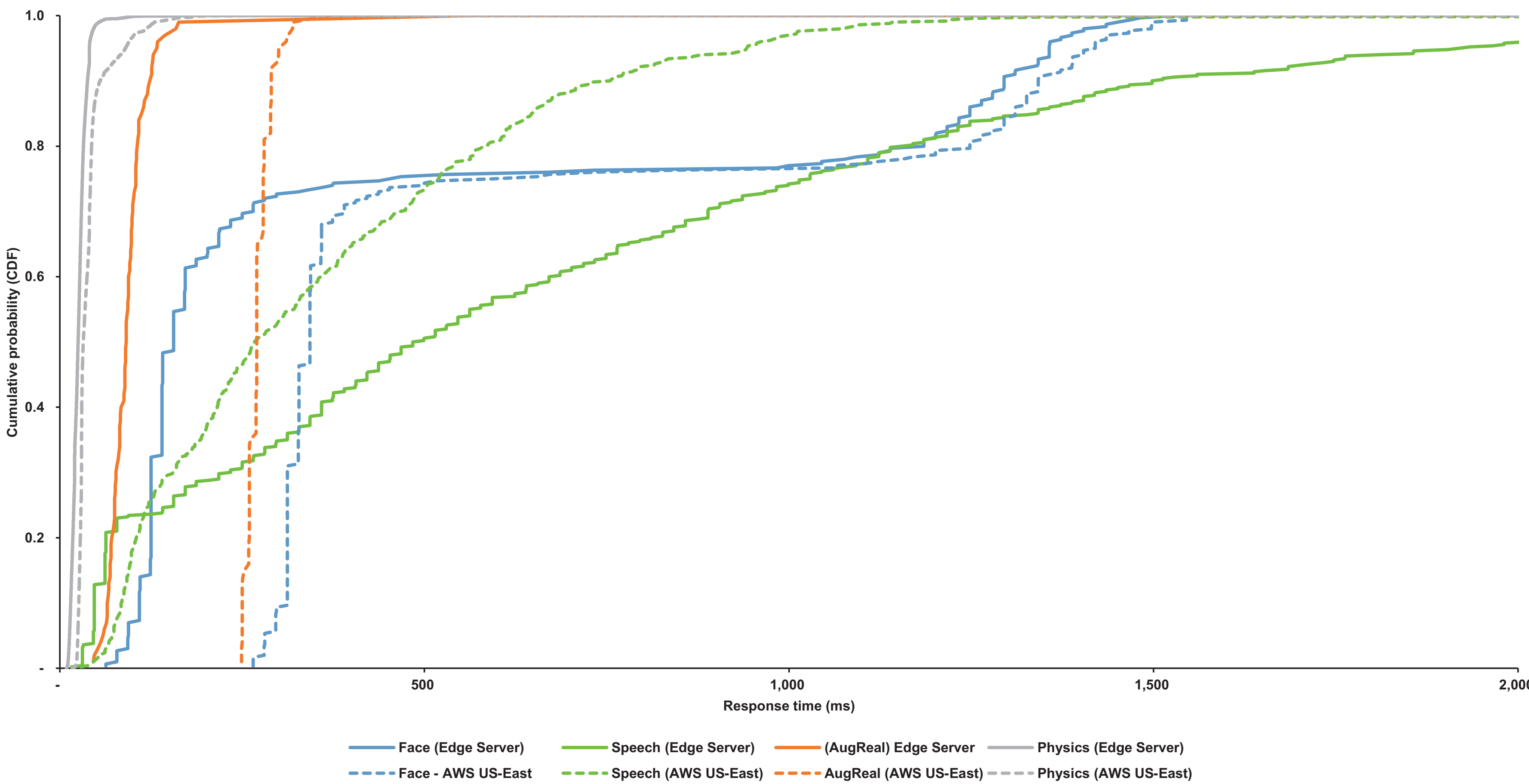
In general across the various use cases, the mobile device itself performed poorly, emphasizing the need to offload compute to either an edge server or the public cloud. Of the five use cases tested, three performed best on the edge server, while two were most suitable for the public cloud.

Exhibit 27: Summary of use cases and performance

Use case	Computational intensity	Data transfer costs	Fastest infrastructure
1) Facial recognition Detects and identifies faces	Medium	High	Edge
2) Speech recognition Extracts text from digitized audio	Medium	Low	Cloud
3) Object and pose identification Identifies known objects in an environment & recognizes the position and orientation relative to the user	High	High	Cloud
4) Augmented reality Identifies buildings and landmarks & label them	Low	High	Edge
5) Physics simulation Models the motion of fluids with which the user can interact	Medium	Medium	Edge

Source: Carnegie Mellon University, Goldman Sachs Global Investment Research

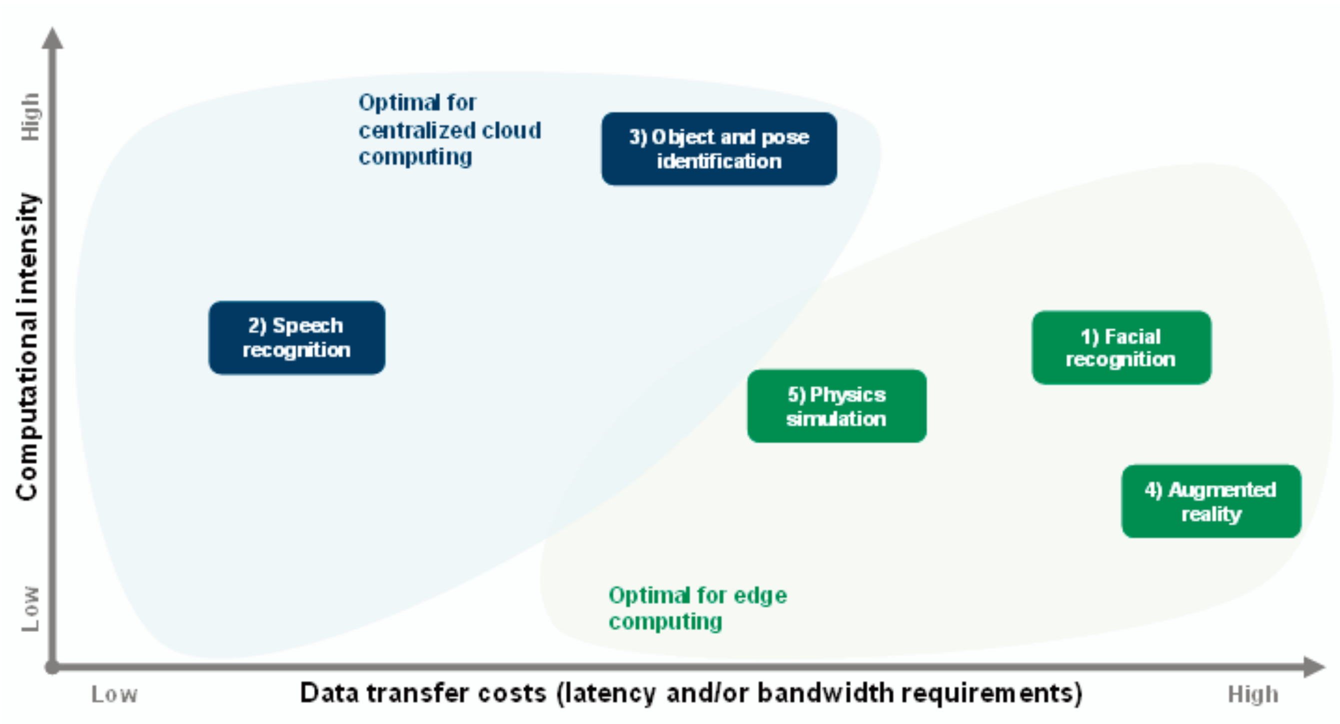
Exhibit 28: Summary of use cases and performance: edge vs. AWS (object recognition use case excluded)



Source: Carnegie Mellon University

We also plot the results on the framework introduced earlier. As we noted earlier, although the edge server does not feature the near-limitless compute capacity of the public cloud, it does vastly outperform the mobile device, and importantly, it is physically near the source of data generation and able to deliver near-instantaneous results. Edge servers will therefore be the optimal vehicle for compute for use cases where the computational intensity is not excessive and the data transfer costs (in terms of latency needs or bandwidth requirements) are high; conversely (and complementarily), the public cloud will be leveraged for use cases where sheer computational capacity is required *and* where there are low data transfer costs (e.g. latency is not important and the use case is not bandwidth-intensive).

Exhibit 29: The best compute vehicle depends on both the use case's computation intensity and its data transfer costs



Source: Carnegie Mellon University

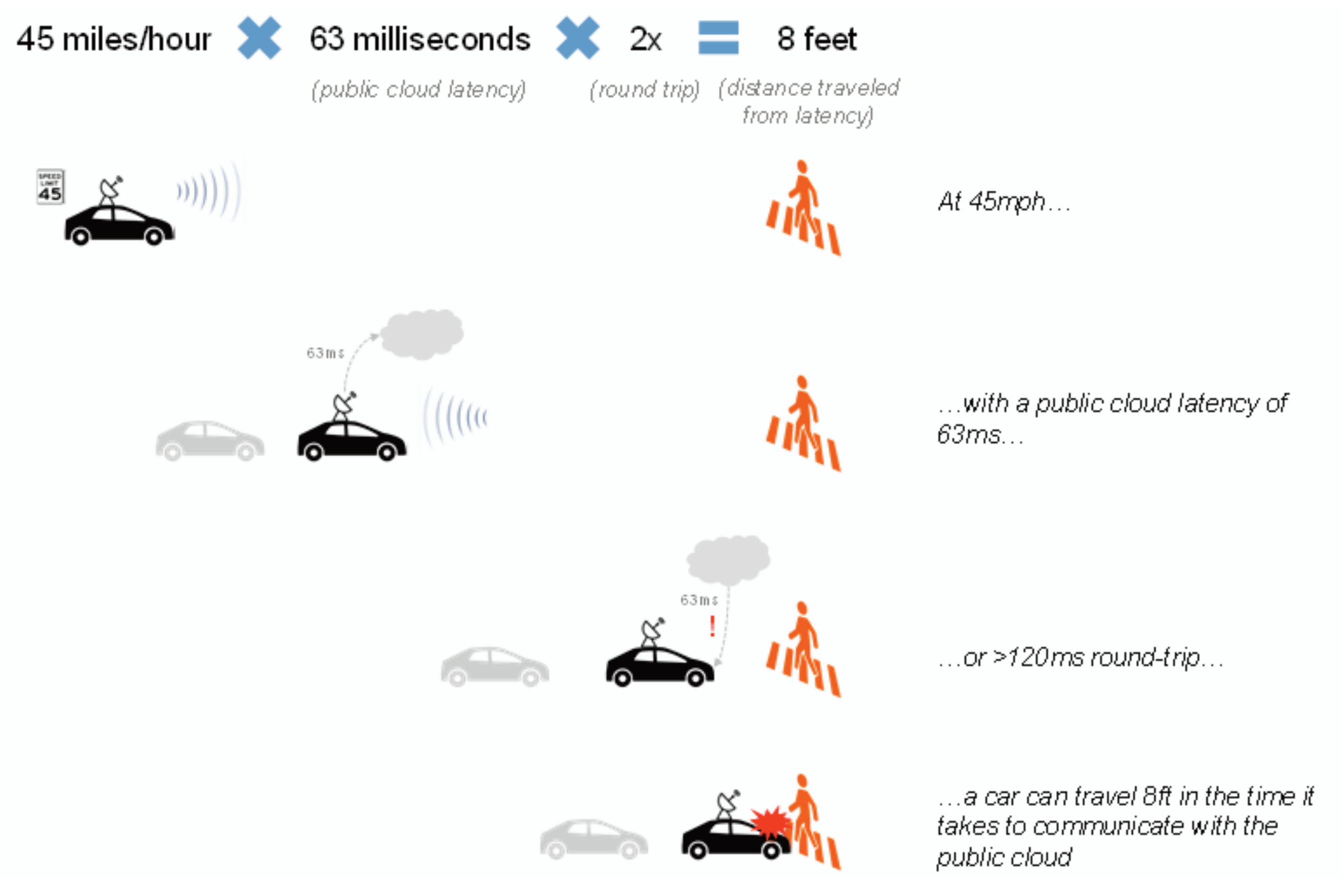
Killer apps for edge computing

Autonomous cars & trucks

We believe real-time processing via an onboard edge server is critical to the safe operation of an autonomous vehicle, for both the passengers as well as the general public; an autonomous vehicle cannot afford the latency required to access the public cloud, as any delays in reaction speed could be potentially catastrophic. For this use case, analyzing the data in real-time – a task that can only be accomplished by an edge server – is critical to maintaining the vehicle's safety, efficient, and performance. We estimate that the market opportunity for autonomous vehicles will reach $100bn by 2025 (Cars 2025: Vol. 3 – Monetizing the rise of Autonomous Vehicles), and we believe that edge computing will be a key capability required by autonomous vehicles.

We noted previously that IaaS vendors have, at a minimum, 63ms of latency, or 126ms round-trip (and this does not include any compute or processing time). However, with just 63ms of latency, an autonomous car traveling 45mph would travel 8ft in the time that it takes to communicate with the public cloud – not counting image recognition/analysis time, time to process the algorithm, and braking distance, all of which would add incremental distance.

Exhibit 30: 63ms of latency is unacceptable for many use cases, including autonomous cars
With 63ms of latency, a 45mph car would travel 8ft



Source: Goldman Sachs Global Investment Research

We note that this case would represent a theoretical best-case distance, as it does not factor in image recognition/analysis time, time to process the algorithm, and braking distance. As a result, we believe that autonomous cars will require processing and computing at the edge in order to maximize safety by minimizing latency; the public cloud will simply be too distant to achieve the performance required to control an autonomous car.

In terms of operating the vehicle, former Intel CEO Brian Krzanich estimates that one average connected autonomous car will generate 4,000 gigabytes of data per hour, given the plethora of onboard sensors (GPS, cameras/video, radar, LIDAR, ultrasonic) recording telematics, resulting in "each car driving on the road [generating] about as much data as about 3,000 people," Krzanich notes. In addition to data generation, autonomous vehicles will also be vociferous data *consumers*, as maps used by the vehicle will need to be accurate down to the inch and be continuously updated to account for construction and road hazards.

In addition to operating the vehicle, the onboard edge server can provide maintenance and analytics to monitor the operational health of key components without the need to stream the data to public cloud. For instance, log data from consumable components (e.g. brakes, fluids, tires, and batteries) would be ingested and analyzed by the onboard edge server. Key data could then be filtered out and uploaded to the public cloud for recommended actions, aggregation, and analysis across the entire fleet of vehicles, helping the operator track key performance metrics that impact business value.

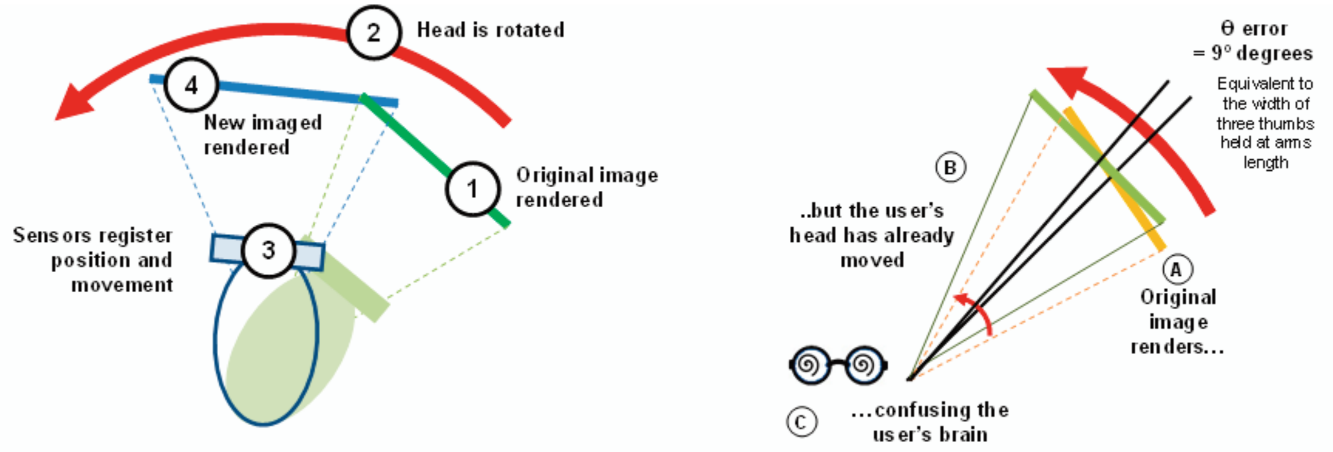
Extended reality (AR/VR): Is 'edge' the sweet spot between latency and form factor?

Augmented and virtual reality use cases require large amounts of processing power; however, users are heavily sensitive to latency, precluding AR/VR from leveraging public cloud given the networking capabilities available today. We estimate that the market opportunity for AR/VR will reach $107bn by 2025 (Profiles in Innovation: Extended Reality).

The case against public cloud: A common roadblock cited in adoption of VR technology is "simulator sickness" – the nauseating effects that stem from the prolonged use of a VR headset – and technologists have come to the conclusion that this is in part due to the lag between a user's movement, and what is rendered on the screen. If a head rotates left, the VR headset must render a new image based on the orientation of the user's field of view to reflect what exists to the user's left in the virtual world, which changes with every movement. Latency is one of the determining factors driving the frequency at which the image can be refreshed and delivered to the user, thus determining the responsiveness of the device.

For example, assuming a user rotates their head at a rate of ~90 degrees per second (i.e. one full revolution every 4 seconds), a latency of 100ms would mean that by the time the headset registered the movement and produced the image, the user's gaze would have changed by 9 degrees, resulting in an image that is slightly "off" compared to what the brain would naturally expect, thus inducing a feeling of dizziness.

Exhibit 31: XR is heavily sensitive to latency
Images are rendered too slowly for true real-time movement



Source: Goldman Sachs Global Investment Research

Typical PC games on average generally have a latency of ~50ms from mouse movement to screen update, however technical papers published by academics indicate that due to the unique requirements of VR, 15ms may be the threshold for truly immersive experiences. Considering only 50% of AR responses fell under ~275ms in the Carnegie Mellon study for AWS-US East (a comparable, but not identical use case), we generally do not believe that streamed VR experiences from public cloud are likely to be the solution in the near term. In contrast, Oculus claims to have achieved a 60-80ms average latency for its Rift headset, where the compute resources are located on a tethered PC.

The case for edge: The same Carnegie Mellon study cited earlier demonstrated that edge-servers could deliver end-to-end response for Augmented Reality of <100ms, 75% of the time, and while we would expect PCs remain the primary mode of compute for the time being, we could see use cases develop for the use of edge servers if this latency can be improved over time (i.e. through 5G), particularly where device-level compute is too difficult to achieve in a form factor that meets the needs of the user. For instance, by eliminating the need for a powerful on-board processing unit, lighter, more compact form factors could be achieved for products such as AR glasses, or wireless VR headsets.

Digital oilfields

Edge computing is slated to play an increasingly vital role in oil and gas exploration, given the remote locations in which the industry operates.

- **Increased productivity:** For exploration wells, using real-time processing can help to maximize drills' output while minimizing energy consumption. Drills operating in remote locations, oftentimes several miles underground, can generate gigabytes of geological data in real-time (Cisco estimates that a typical offshore oil platform generates 1-2 TB of data per day, or ~1 GB every second). While much of this valuable captured data can be leveraged to update models of the Earth's internal structure and layers, the difficulty lies in processing and analyzing the data in real-time, as the data becomes stale quickly. Teams operating in the field need to make instant decisions about the next best course of action – should the drill continue, change direction, drill horizontally, or stop? Although manual analysis and manual adjustments are potentially feasible, given the need to drive real-time decisions from large data sets, for maximum efficiency, data from sensors would ideally be automatically processed and deployed to fine-tune equipment rather than incorporate additional latency from manual processes. Edge computing at the point of data collection (i.e. on the oil platform) would be critical to driving real-time insights and recommendations from data generated by oil platform equipment.

- **Systems uptime:** Apache Corporation, a petroleum and natural gas exploration and production company, estimates that downtime can cost up to $1mn per hour, or $16,000 per minute. Equipment difficulties can be spotted (or predicted) much faster, minimizing the expensive downtime.
- **Lower costs:** Drilling frequently occurs in remote locations, with limited (or very expensive) satellite connectivity – typically at 64 Kbps to 2 Mbps, implying ~12 days to upload a single day's worth of data from an oil rig. Processing raw data at the edge would preclude the need to send data back to a data center or the public cloud, which dramatically lowers network and communication expenses.

IoT enterprises

We expect edge computing to play a pivotal role in the development of new IoT software platforms. As increasing amounts of compute, storage, and analytics capabilities are integrated into ever-smaller devices, we expect IoT devices to continue to proliferate, and as noted previously, Gartner expects IoT endpoints to grow at a 33% CAGR through 2021. In cases where reaction time is the raison d'être of the IoT system, the latency associated with sending data to the cloud for processing would eliminate the value of the system, necessitating processing at the edge; public cloud could still be leveraged where processing is less time sensitive or in instances where the scale and sophistication of public cloud need to be brought to bear. Gartner projects $3.4 trillion of annual spending on IoT hardware alone by 2021.

For instance, C3 IoT provides an application platform for enterprises to deploy IoT solutions. The company began by targeting energy companies, but has since expanded to other industries. Customers include Enel SpA, conEdison, Exelon, PG&E and the U.S. Department of State. C3 IoT's solution monitors real-time and aggregates data from connected sensors (e.g. smart meters, thermostats, transformers) to provide predictive analytics and performance insights. The company targets data-intensive industries where analyzing the data can drive meaningful operational improvements for the business. C3 IoT's software leverages artificial intelligence (AI), so that its algorithms become more accurate the more information it is provided. The platform currently leverages the public cloud (AWS) and has an open architecture that leverages 3rd party libraries/plug-ins. Edge computing, in our view, could serve to accelerate the AI and provide more timely recommendations by bringing processing power to the source of data generation.

The company has also highlighted predictive maintenance as a potential "killer app" for IoT due to the cost savings it facilitates. For instance, C3 IoT is deploying its technology with Enel (utility company) across smart meters in Europe to drive €261mn in recurring cost savings through automation. We believe that edge computing could play a vital role by expediting the realization that predictive maintenance is required, rather than uploading to and/or batch processing data in the public cloud.

Public safety (Amber Alerts)

Video analytics is an example where bandwidth limitations, long latency, and privacy concerns converge to favor edge computing over leveraging public cloud. For instance, locating a lost child in a city is one potential real-world application of video analytics where public cloud limitations would prevent successful deployment. In today's world, urban areas typically have a wide variety of cameras covering large proportions of areas, including security, traffic, and vehicle-borne cameras. When a child needs to be located, these cameras can be leveraged, as it is likely that the child will be captured on a camera at some point. However, the data from these cameras typically is *not* uploaded to the public cloud, in light of both bandwidth and privacy considerations. Even excluding these considerations, the ability of even public cloud computing resources to analyze the amount of raw data being generated would be overwhelmed, with real-time analysis – which would be critical in searching for a missing child – essentially impossible. However, with an edge computing paradigm, the request to locate the missing child can instead be pushed out to all of the relevant devices: each camera would perform the search independently using nearby compute resources. If, and only if, the camera registers a positive match would it then upload data to the cloud: by distributing the analytics to the small-but-numerous devices in the edge (where the data resides), tasks can be quickly and efficiently processed.

Winners & losers: edge computing could sustain a renaissance in on-premise software

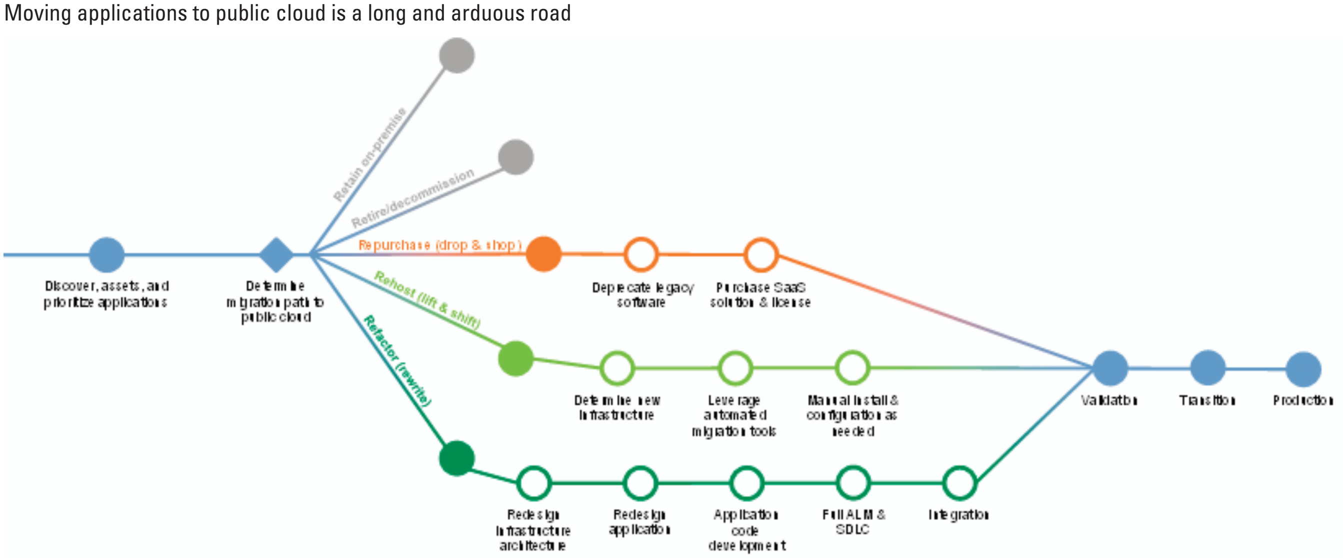
In our conversations with partners and resellers over the past several months, many have noted a generally robust IT spending environment, not just for public cloud but also for on-premise and hybrid offerings. Although the move to public cloud continues in earnest, enterprises are increasingly confronting the challenges of migrating workloads to public cloud and digesting their public cloud spending. As enterprises come to the conclusion that their IT paradigm will likely be hybrid for longer than anticipated, with servers at the edge to augment public cloud resources, this dynamic is helping drive a renaissance in on-premise spending.

With the initial positive sentiment, elevated expectations, and initial curve of the hype cycle of public cloud now past, CIOs are starting to work through the challenging task of migrating legacy workloads to public cloud. While one path of moving to public cloud is lift & shift, to take full advantage of the scaling and elastic capabilities of public cloud, legacy workloads must be refactored – redesigning, rearchitecting, and rebuilding the application on a public cloud PaaS in order to use innovative, cloud-native features. Unfortunately, refactoring applications can be a difficult and time-consuming process.

Even in one of the best of cases, Expedia, which is listed by AWS as a case study (and was on stage at 2017's AWS re:Invent conference), has taken 9+ years thus far on their journey to move 100% of workloads to AWS from 100% on-premise at their data center in Chandler, Arizona. Starting in 2009, Expedia began a massive replatforming effort to rewrite every line of their 10 million+ lines of code. Even with this concentrated, top-down effort to refactor its base of applications, Expedia estimates that it is still 2-3 years away from achieving 80% of its applications on AWS, with presumably the most challenging 20% of its on-premise applications remaining to be refactored.

Exhibit 32: The path to public cloud is more challenging than many originally anticipated

Moving applications to public cloud is a long and arduous road

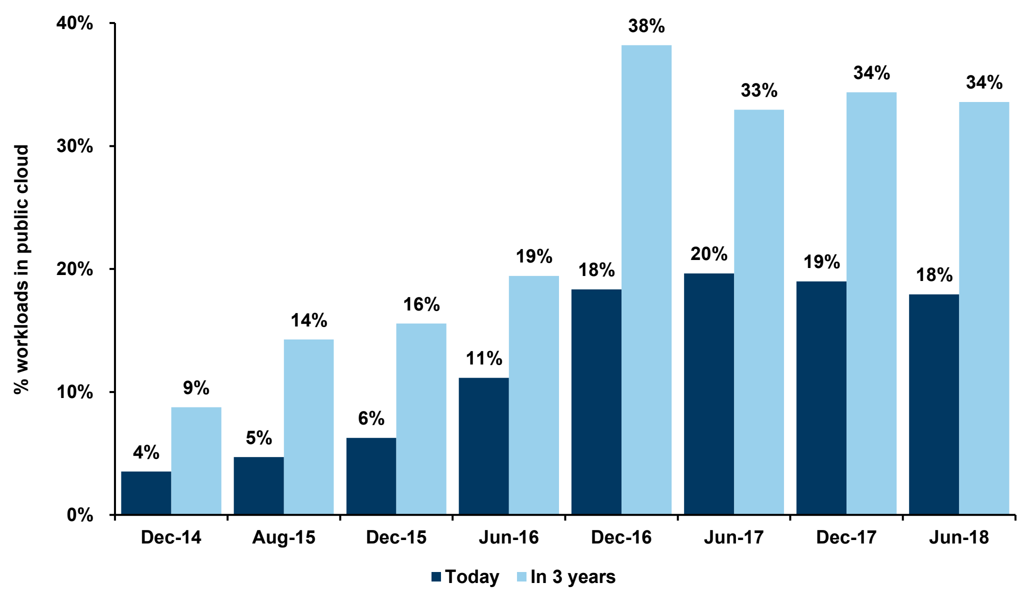


Source: Goldman Sachs Global Investment Research

As CIOs begin to operationalize workloads in the public cloud, this has led to "sticker shock" for many CIOs: our conversations with public cloud partners also reveal that almost every customer that leverages the public cloud has ended up over-consuming relative to their original budget and planned spend. Particularly if the application is simply lifted & shifted and *not* refactored (resulting in the application having low deployment density, not ensuring maximum utilization of system resources, or deallocation/reallocation when idle), public cloud workloads can, in fact, be *more* expensive to run than on-premise. We would note that VMware's announced acquisition of CloudHealth Technologies, which helps customers analyze and manage cloud cost, usage, security, and performance centrally for public cloud, might help to address this pain point.

As our July 2018 CIO survey (*IT Spending Survey: Spending intentions tick down, but remain near record highs*, 7/9/18) helped to highlight, although the overall trend of a shift to public cloud continues, the expectations around the pace of the shift over the last year is now expected to be somewhat more gradual then originally estimated by many.

Exhibit 33: The shift to cloud continues, but expectations have been tempered in the past 6-12 months
GS CIO survey: percentage of workloads in public cloud today (navy) vs. percentage of workloads in public cloud in three years (light blue)



Source: Goldman Sachs Global Investment Research

In the near-term, we would expect that edge servers leverage very similar architectures as on-premise data centers today, to ensure maximum compatibility between the edge server and data center.

- **Virtualization:** In the near-term, we would expect that virtualization will play a critical role with edge servers, much as it has for data centers over the past two decades. Virtualization would likely be mandatory for edge servers, allowing multiple applications to share a single physical edge server by running inside a virtual machine. **VMware**'s vSphere, which has ~90%+ market share in what has emerged as a winner-takes-all virtualization market, would clearly be the vendor to benefit from the need for virtualization in edge servers, in our view.

- **Operating system:** Linux continues to be the fastest-growing server operating system, with Gartner projections indicating that Linux's share of the overall market will grow from 15% in 2014 to 26% in 2020. In our view, **Red Hat**, as the preeminent enterprise Linux provider (Red Hat Enterprise Linux, or RHEL), would be the primary beneficiary of edge server computing, as we envision that enterprises would strive to maintain a consistent environment between the edge server, their data centers, and the public cloud, which is dominated by Linux distributions. Our view is that this potential mobilization of edge computing could drive a resurgence in RHEL, thereby helping to offset RHEL growth deceleration as workloads moving to the public cloud replatform off of RHEL. We believe that although Windows continues to retain its market share (~55%), few net new workloads, especially those designed with the public cloud in mind, are being architected on Windows Server. We note, however, that Azure has the ability to span the public cloud, on-premise, and the edge, which could potentially help to support Windows Server market share.

The role of containers in the edge

Given that edge nodes will certainly not have the same caliber of compute, memory, and storage resources as the public cloud (or an on-premise data center), edge node infrastructure software will likely need to be much more efficient and consume fewer resources, in addition to being optimized for quick boot-up and resource isolation. As a result, we would expect that containers play an increasing role in edge computing, given the necessity of wringing out every possible bit of performance from a finite and constrained resource like an edge server.

Traditionally, software virtualization leveraged virtual machines (VMs), which use a hypervisor to abstract away the system hardware – via the hypervisor, this allowed multiple VMs to run atop a single physical server. Each VM contains its own guest operating system (OS), with the applications installed within the guest OS

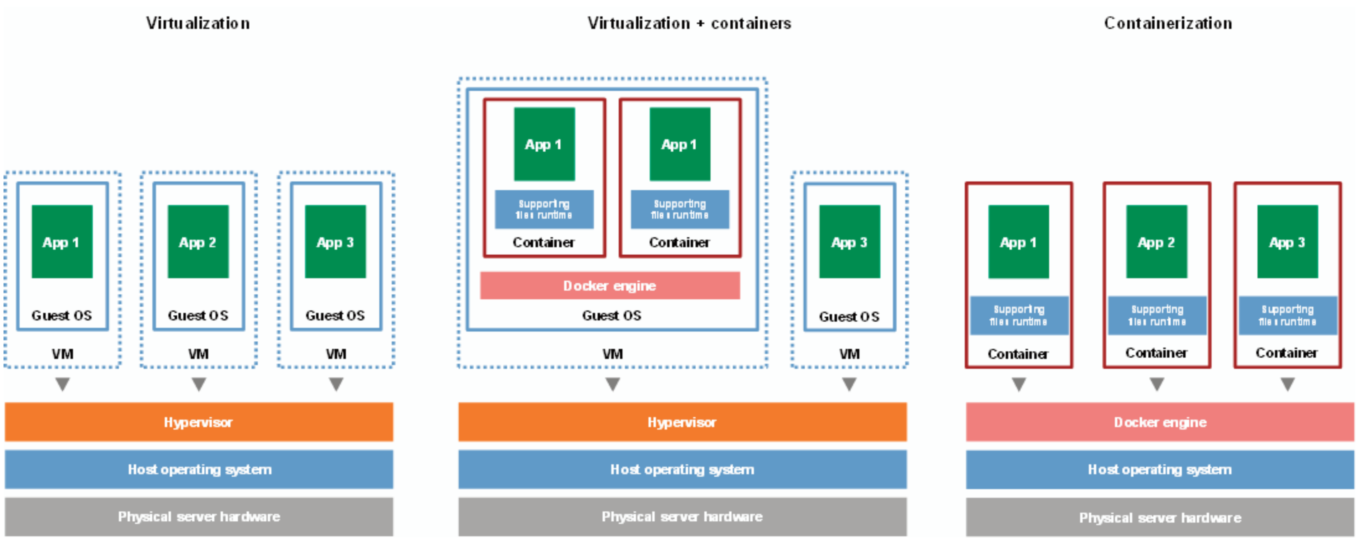
Because the VMs are completely isolated and independent from each other, a single physical server can be shared among many VMs and many applications, with the VMs providing high levels of isolation and security, given that each application runs inside its own dedicated environment. However, this architecture necessitates virtualizing a set of hardware and running a separate guest OS within each VM: this results in a performance penalty, in the form of overhead, which lowers the number of VMs and applications that can be run within a single physical server. Additionally, the process of spinning up a new VM and starting a new guest OS is not instantaneous, resulting in increased latency.

Containers help to solve the performance overhead issue by implementing a lighter-weight type of virtualization. Rather than abstracting the system hardware, containers essentially virtualize one level up – the operating system. Containers package up the application with the supporting files and runtime (i.e. everything that the application needs to run). As a result, multiple containers could theoretically be run atop the same host operating system, without the need to virtualize a set of hardware and run a guest operating system for each container. Instead, containers are designed to isolate (from other containers and from the host OS), a set of processes and resources, including compute, memory, and storage resources.

However, given that VMs are thought to provide superior isolation and security (as they virtualize at the hardware level vs. at the operating system level), organizations that leverage containers today typically run the containers inside VMs – this provides portability and flexibility of containerized applications, reduces overhead, and provides the security benefits of VMs. Additionally, management and tooling are much more mature for VMs, allowing for a wide range of out-of-the-box capabilities, including moving workloads among hosts and live upgrading of software.

We note that container adoption remains in very early stages, with Gartner survey data indicating that just ~40% of survey respondents have deployed any containers in production; of these adopters, the median company had just 20 container instances (the typical enterprise has thousands of application, each potentially with large numbers of instances, depending on the application capacity required).

Exhibit 34: Edge servers could use lightweight containers to run applications
Virtualization vs. virtualization + containers vs. containers on bare metal



Source: Goldman Sachs Global Investment Research

With the rise of containers in edge computing, we would expect that container platforms like Red Hat OpenShift and Pivotal Cloud Foundry would benefit.

Public cloud winners

We believe that longer-term, the winners of a shift towards edge computing will be 1) the scaled public cloud vendors and 2) infrastructure software vendors who can seamlessly bridge the gap between on-premise and public cloud. Like public cloud computing, edge computing requires an efficient software stack that can be deployed in a cohesive and scalable fashion, with automation key to ensuring that the multitude of edge servers and edge devices are properly maintained, updated, and secured. Without a cohesive public cloud and edge cloud solution, there could conceivably be three distinct software stacks: one at the edge device, a different one at the edge server and data center, and yet another one in the public cloud. With three disjointed software stacks comes three different application stacks and three different development teams, in addition to the need to integrate among the three. In order to leverage true interoperability and elastic scalability, a single software stack that can span the public cloud, edge cloud, and edge device is required.

To support the emerging intelligent cloud, intelligent edge application pattern, a user needs a consistent stack across the public cloud and the edge. Merely providing colocation services or connectivity between on-premise data centers and the public cloud is not sufficient to meet customer needs. Users need consistency across the development environment, operating models and technology stacks. Azure provides this consistency across the entire stack, inclusive of identity, data, app platform, security and management at the edge and in the cloud.

We highlight Microsoft, Amazon (Web Services), Pivotal Software, and Red Hat as potential long-term winners of a shift towards edge + cloud computing.

Microsoft

Azure Stack

Azure Stack is an on-premise extension of Microsoft's Azure cloud, enabling customers to run a consistent environment whether in public cloud or on-premise. As a result of this consistency, developers can build and deploy applications with exactly the same approach – including the same APIs and the same DevOps tools – regardless of where the application is run.

Physically, Azure Stack is delivered via an integrated, hyperconverged hardware system offered by Dell EMC, HPE, Cisco, and Huawei partners and certified by Microsoft. Microsoft estimates that a full-sized, 12-rack server unit of Azure Stack hardware can run ~400 virtual machines (each with 2 CPUs and 7 GB of RAM).

Exhibit 35: Azure Stack is a hyperconverged system fully consistent with Azure public cloud

Source: Microsoft, Goldman Sachs Global Investment Research

After purchasing the hardware, customers can pay for Azure Stack's software with "pay-as-you-go" pricing – mirroring the public cloud pricing paradigm where customers are billed for consumption.

Exhibit 36: Azure Stack pay-as-you-go pricing

Service	Pricing
Base virtual machine	$0.008/vCPU/hour ($6/vCPU/month)
Windows Server virtual machine	$0.046/vCPU/hour ($34/vCPU/month)
Azure Blob Storage	$0.006/GB/month (no transaction fee)
Azure Table and Queue Storage	$0.018/GB/month (no transaction fee)
Azure App Service (Web Apps, Mobile Apps, API Apps, Functions)	$0.056/vCPU/hour ($42/vCPU/month)

Source: Microsoft, Goldman Sachs Global Investment Research

Customers can also choose to use existing licenses, including Windows Server and SQL Server, with Azure Stack; in that case, customers would pay only the base virtual machine consumption fee.

However, given the hybrid nature of Azure Stack, customers can also choose a capacity model pricing package, more reminiscent of on-premise license pricing. In this case, pricing is a fixed fee annual subscription based on the number of physical cores in the deployment. Two capacity packages are offered by Microsoft: the App Service package ($400/core/year), which includes App Service, base virtual machines, and Azure Storage, as well as the IaaS package ($144/core/year), which is only base virtual machines and Azure Storage.

For the hardware, we note that the HPE ProLiant for Microsoft Azure Stack version starts at $300,000 to $400,000 for the hardware and support, depending on the exact configuration.

The Azure Stack software includes many of the cloud features found in Azure, including basic IaaS functions (VMs, storage, and virtualized networking), as well as some of Azure's basic PaaS services (Azure Container Service, Azure Functions, and Azure Active Directory). Natalia Mackevicius, director of program management for Azure Stack, expects that Azure Stack will appeal to verticals like oil and gas, manufacturing, retail, healthcare, and government, where connectivity/bandwidth, regulatory, and/or security concerns prevent full utilization of the public cloud. Microsoft notes that a key use case will likely be addressing "latency and connectivity requirements by processing data locally in Azure Stack and then aggregating in Azure for further analytics, with common application logic across both. We're seeing lots of interest in this Edge scenario across different contexts, including factory floor, cruise ships, and mine shafts."

Azure Stack is effectively an edge server that is a miniaturized data center. By running a completely consistent environment (updates to Azure Stack software are pushed to customers as they are completed) between Azure Stack and the Azure public cloud, Azure Stack effectively allows Azure to be placed next to the source of data generation and operate with intermittent (or a complete lack of) connectivity, while leveraging Azure cloud services. We note that neither AWS nor GCP offers anything similar, as Microsoft has a unique position as both a leading on-premises vendor as well as a top-tier public cloud provider.

Early adopters of Azure Stack

- **Carnival Cruise Lines:** Carnival's cruise ships have limited network connectivity while underway, in the middle of the ocean. To satisfy the cruise ship's compute and storage needs, Carnival leverages Azure Stack as a private cloud to collect and process data while in transit; after the ship returns to port, the data can then be uploaded to Azure for additional processing and analysis. Additionally, Carnival can write applications once and then run it either in Azure or in Azure Stack.
- **Avid Technology:** Avid provides content editing, content management, newsroom graphics, and news production solutions for customers, including nearly all of the major film studios, 9 of the top 10 global news networks, all of the major music companies, and the majority of leading subscription streaming services. For its MediaCentral solution, Avid's technology platform for providing the tools and services required for news production, Avid chose a hybrid architecture, deploying some pieces in Azure public cloud datacenters and some in Azure Stack on-premises at the customer's data center. In addition to public cloud's in-built advantages relative to on-premises infrastructure (scale, redundancy, cost, elasticity, etc.), the public cloud aspect allows news producers to leverage Azure's breadth of

media services, including encoding, transcoding, media analytics, and streaming, as well as Azure AI technologies like speech-to-text, keyword extraction, and face, object, and action identification. Azure Stack allows media to be colocated where it is needed (e.g. for editing) – with large raw media files, latency and bandwidth become major issues even with a fast internet connection. Additionally, during critical and time-sensitive moments, Azure Stack enables journalists, editors, and producers to continue to work even if network connectivity is reduced.

- **Mitsui Knowledge Industry (MKI):** MKI provides IT consulting and digital transformation services to large Japanese enterprise customers, helping them modernize business-critical apps. However, with many of its customers forced to remain partially on-premise due to regulatory and security requirements, MKI leverages Azure Stack to architect a consistent hybrid cloud model for its customers, allowing them to deploy applications both on-premise and in Azure, as the need arises. For instance, financial institutions may not be able to store sensitive data in public cloud; in some cases, manufacturers' factories may not even be connected to the internet. With Azure Stack on-premises at the customer, MKI can develop and deploy applications consistently across both Azure and Azure Stack, accelerating application deployment times and providing flexibility for the application infrastructure.

Azure IoT Edge

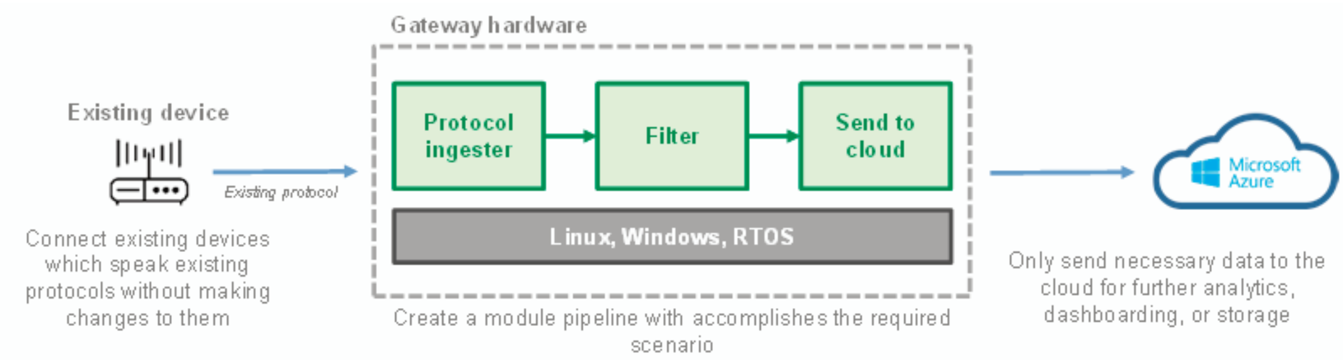
Originally announced at Microsoft Build in 2017, Azure IoT Edge is a feature of Azure that allows customers to do edge computing by bringing cloud intelligence down to the device. For instance, users can run custom logic or container-based modules using C, C#, Node.js, Python, and Java, or leverage Microsoft APIs or marketplace services with pre-built code. Additionally, Azure services like Azure Cognitive Services, Machine Learning, Stream Analytics, and Functions can run on the device.

Azure IoT Edge is designed to help businesses generate actionable insights from data collected by sensors and devices residing at the edge of networks, without the need to send that data back to the public cloud for processing. According to Microsoft corporate VP of communications Frank Shaw, "In our customers' worlds, devices and data are often locked in remote places, like oil wells and farms, or in mission-critical places like hospitals and factories. Where connectivity can be expensive or unreliable, having IoT devices that can do local processing outside of the cloud is a big advantage under these conditions."

As previously mentioned, we believe that containers will play an increasingly important role in the edge, where resources are at a premium. As such, Azure IoT includes support for the Moby container management system (which is the open-source platform upon which Docker is built), which helps to ensure that containers can be extended from the cloud to edge devices with full isolation and management. Moby containers and Docker containers are compatible, allowing Moby containers to work on Docker-based systems (and vice-versa), and as a result, no changes are required to existing Docker-based modules to run on Azure IoT Edge.

Azure IoT Edge was recently announced as generally available for enterprise-grade and scaled deployments (June 27, 2018). Along with IoT Device Provisioning Service, devices can be provisioned in the field with no operator intervention, and customers can provision tens of thousands of devices. Device options are diverse and range from Raspberry Pis to industrial equipment. Three components are required for Azure IoT Edge deployment: Azure IoT Edge Runtime, Azure IoT Hub, and edge modules. The Azure IoT Edge runtime is free and available as open source code; however, customers will require an Azure IoT Hub instance to manage and deploy edge devices. Pricing is dependent on the Microsoft Azure services used.

Exhibit 37: Azure IoT Edge



Source: Microsoft, Goldman Sachs Global Investment Research

Amazon Web Services (AMZN covered by Heath Terrry)

Snowball Edge

AWS Snowball Edge is a 100TB data transfer device that includes both storage and compute capabilities. Unlike the standard Snowball, which is essentially a pure storage appliance for transferring large quantities of data into and out of AWS, Snowball Edge incorporates compute capabilities as well: in addition to transferring data to/from AWS, on-board compute allows Snowball Edge to process data locally using Lambda functions in Python. This edge compute capability allows the appliance to analyze data streams or process data locally, without the need to transfer it back to the AWS public cloud. As an edge appliance, Snowball Edge does not require an internet connection to function. Snowball Edge appliances can also be clustered to 1) scale and shrink storage capacity and 2) achieve 99.999% data durability on-premises.

Exhibit 38: AWS Snowball Edge device

Source: Amazon

In terms of pricing, users are charged a fixed service fee per job ($300 in the US), which includes 10 days of onsite appliance usage; extra days beyond the first 10 initial days are billed at $30/day. Data transferred into AWS is free of charge, and data transferred out of AWS ranges from $0.03 to $0.05/GB, depending on the region. Standard AWS S3 pricing applies after the data is transferred into AWS.

AWS cites use cases like remote locations with embedded applications, remote locations, IoT, and manufacturing as prime examples of where Snowball Edge's compute functionality can add value:

- **Embedded applications:** Snowball Edge can support medical imaging or optical scanning MRI machines by storing the images as they are captured to give doctors and administrators immediate access as they are generated. With a cluster of Snowball Edges, devices can be removed from the cluster to upload the images into the AWS public cloud with no on-premise downtime.
- **Remote locations:** Airplane engine manufacturers can use a Snowball Edge to gather data while the airplane is in flight, providing immediate analysis of performance and maintenance needs. After the plane lands, the Snowball Edge appliance can then be sent back to AWS, where the data is uploaded into the public cloud, and more complex data analysis can be performed (e.g. across all engines and all flights).
- **Manufacturing:** Manufacturers can use a Snowball Edge to help manage automated machinery (equipped with sensors). The Snowball Edge uses Lambda functions to run analysis, take actions, and respond to sensor inputs in real-time (e.g. log alerts, order spare parts, or tune machine performance). The data in the appliance can

either by uploaded back to AWS via the internet or by shipping the Snowball Edge device for deeper analysis.

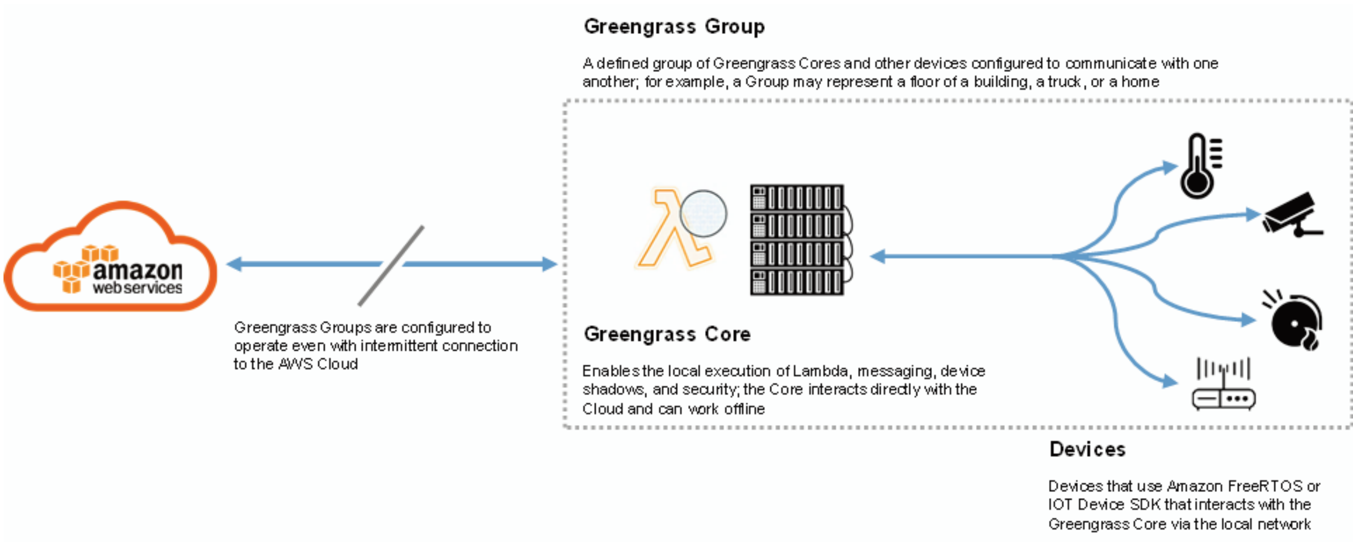
Greengrass

AWS Greengrass is software that brings IoT cloud programming and functionality to sets of devices, enabling devices to run programming even when disconnected from the cloud. Using AWS Greengrass, Lambda functions are developed and refined in the cloud and then deployed to the Greengrass Core device (effectively an edge server), allowing it to use AWS Lambda functions to respond immediately to data or local events, while continuing to leverage the cloud for management, analytics, and durable storage.

Greengrass also includes ML Inference, which allows the device to perform machine learning inference locally using models that have been built and pre-trained in the cloud, with its near-limitless computing resources. However, because inference using a pre-trained model is typically less taxing, it can be performed real-time on the device itself, and as a result, most data generated by the device never needs to be uploaded to the cloud. Given its nature, minimum hardware requirements to run Lambda are fairly trivial: a single 1 GHz core (x86 or ARM), 128 MB of RAM, and Linux (either Ubuntu or Amazon Linux).

With AWS Lambda, the Greengrass Cores (i.e. edge servers) are defined and configured in the cloud; each device must have the Greengrass software installed. Greengrass Cores are then assigned to Greengrass Groups, and according to AWS, Greengrass Groups could represent a floor of a building, a single truck, or a home. In terms of the IoT endpoint devices, any device running Amazon FreeRTOS or the IoT Device SDK can then be configured to interact with the Greengrass Core (via the local network). Lambda functions can then be built, edited, and pushed to the Cores, allowing them to communicate and react (run local compute, message, cache data) to data from the devices, even without cloud connectivity. Once cloud connectivity is reachieved, data can then be synched.

Exhibit 39: AWS Greengrass



Source: AWS, Goldman Sachs Global Investment Research

Pivotal Software

Pivotal Container Service (PKS)

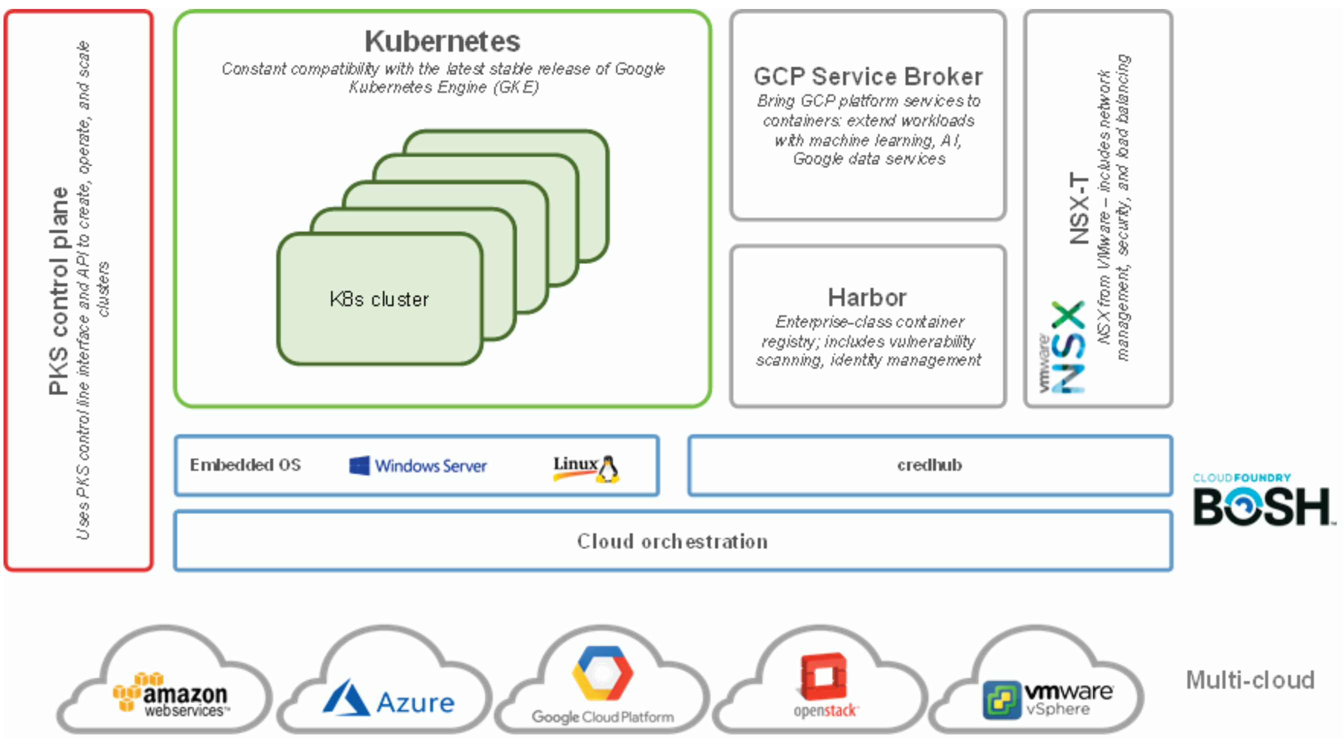
Unlike full-fledged servers in data centers, edge servers working with more constrained resources may need lighter versions of software given the hardware constraints. We believe that there is an opportunity to leverage containers and Kubernetes in edge computing, as it has the ability to support diverse workloads, enable quicker deployment of applications on a broad variety of platforms, in addition to improved performance.

Running containers inside virtual machines incurs a "double virtualization" penalty – the hypervisor layer upon which the VM runs plus the container engine. CenturyLink performed a study comparing the performance of Kubernetes clusters on both VMs and bare metal servers by measuring the network latency of a testing utility running inside a Docker container. When running the workload on a bare metal Kubernetes cluster, CenturyLink saw a 3x improvement in latency. In addition to higher network latency, CPU consumption when the cluster was run on VMs was noticeably higher than when the cluster was run on bare metal.

We believe that offerings like Pivotal Container Service (PKS), which abstract away the underlying infrastructure, build in resiliency, and utilize portable, lightweight compute platforms like containers, are a potential fit for edge computing architectures. PKS is the company's container management platform, delivering an enterprise-grade Kubernetes solution for running, updating, and maintaining container-based applications. The Kubernetes framework provides a number of key application layer capabilities such as application orchestration, scaling, health monitoring and healing; PKS operationalizes Kubernetes by including BOSH to orchestrate deployment, lifecycle management, and operations, VMware's NSX offering to manage software-defined virtual networks, and production-grade features, including high-availability, security, identity and access management, logging, and monitoring. PKS allows enterprise developers to build containers and/or run containerized applications developed by third parties. Popular containerized applications (as measured by Docker Hub pulls by software review site G2 Crowd) include operating systems (Ubuntu, CentOS, Fedora), databases (MySQL, MongoDB, PostgreSQL), web frameworks (Ruby On Rails, Django), web content management (WordPress, Joomla), and business content management & file storage (Owncloud), among others.

Launched together with Google and VMware in August 2017 (GA in February 2018), PKS leverages common services and core technologies from Pivotal Cloud Foundry (PCF) to allow enterprises to leverage Kubernetes, the market-leading open-source offering for container operations, across various public, private, hybrid, and multi-cloud IT environments. PKS also includes VMware's NSX offering to manage software-defined virtual networks; this results in a 50/50 revenue share with VMware, with the VMware sales force incentivized to leverage its enterprise relationships to help drive PKS adoption.

Exhibit 40: PKS architecture



Source: Pivotal Software, Goldman Sachs Global Investment Research

Red Hat

OpenShift

Similar to Pivotal Container Service, as enterprises begin to develop applications to take advantage of the wealth of data generated by their edge devices, we would expect that Kubernetes -based platform-as-a-service offerings like OpenShift, could benefit. We note that Red Hat has commented that OpenShift is "almost 20x the price of RHEL on the same two-socket server."

We believe that the next generation of applications will be built on modern software development platforms, leverage newer development methodologies (e.g. DevOps), and architected via microservices.

Exhibit 41: Pivotal Container Service vs. Red Hat OpenShift

	Pivotal Container Service (PKS)	Red Hat OpenShift
Platform	▪ vSphere ▪ Google Container Engine (GKE)	▪ Red Hat Enterprise Linux ▪ Red Hat CoreOS (previously CoreOS Container Linux)
Infrastructure	Multi-cloud	Multi-cloud ▪ OpenShift Online (multi-tenant public cloud offering) ▪ OpenShift Dedicated (single-tenant public cloud offering) ▪ OpenShift Container Platform (private cloud offering)
Container Runtime	Cloud Foundry Container Runtime (previously Kubo)	Docker
Orchestration	Kubernetes	Kubernetes (starting in 2015), previously proprietary
Registry	Harbor (VMware)	▪ Proprietary ▪ CoreOS Quay
Security features	NSX-T (VMware)	Proprietary
Storage	vSAN (VMware)	Proprietary
Application States	▪ Stateful ▪ Stateless	▪ Stateful ▪ Stateless
Pricing	▪ Higher price point (approximately 2x OpenShift price) ▪ Consumption-based pricing based on number of pods (collection of containers that share the same computing resources)	▪ Lower price point ▪ Capacity-based pricing based on core and socket count

Source: Gartner, Goldman Sachs Global Investment Research

OpenStack

We believe that many enterprises could choose OpenStack as the infrastructure-as-a-service foundation of their IoT cloud, leveraging familiar on-premise components and concepts like server virtualization, shared storage, data lakes, and containers for building a new IT stack cohesive across the edge, private cloud, and public cloud.

OpenStack is a set of open source tools that help provide fundamental infrastructure building blocks – including compute, networking, and storage – that can be deployed in private clouds, including the edge of the network. In virtualization, resources abstracted from a vendor-specific programs (e.g. vSphere for compute, NSX for networking, and vSAN for storage), allowing the hypervisor to distribute the resources as required. By contrast, OpenStack uses a consistent set of APIs to further abstract resources like compute, networking, and storage into a commoditized "pool," to create a cloud environment that lets administrators, operators, and users interact with them directly.

Thus far, our channel checks indicate that OpenStack remains highly complex and difficult to deploy. As a result, we believe that Red Hat's traction with OpenStack remains limited to the very highest end of the company's customer base (namely, telcos and financial services), owing to its complexity and the substantial expertise necessary for deployment and management. However, given the flexible and modular nature of OpenStack, we believe that OpenStack has the potential to allow users to efficiently run the minimal services required at the edge, yet simultaneously provide robust support for bare metal, container technologies and virtual machines.

VMware

In the near-term, our view is that virtualization would likely be mandatory for edge servers, allowing multiple applications to share a single physical edge server by running inside a virtual machine, given the resource limitations inherent in an edge server. VMware's vSphere, which has ~90%+ market share in what has emerged as a winner-takes-all virtualization market, would clearly be the vendor to benefit from the need for virtualization in edge servers, in our view. VMware believes that eventually, edge computing could potentially become a $1bn+ business. We note that at VMworld 2018, VMware announced support for its ESXi hypervisors on Arm architectures, which are typically used for consumer electronic devices (e.g. smartphones, tablets, wearables) and are lighter-weight than Intel x86 servers. VMware CEO Pat Gelsinger believes that edge computing could be the next $1bn market for VMware.

VMware expects **Project Dimension,** announced at VMworld 2018, to extend VMware Cloud to deliver VMware's software-defined data center (SDDC) infrastructure and hardware as-a-service to on-premises locations – essentially extending the cloud infrastructure experience to the edge. Essentially, VMware takes a hyperconverged infrastructure appliance (i.e. a self-contained, optimized server with compute, storage, and networking capabilities) and delivers it to a location that needs edge computing services (i.e. a customer's datacenter, a factory, or an oil rig), providing the same cloud environment at the edge as in a customer's data center or in VMware cloud. As a fully-managed service, VMware manages the infrastructure, troubleshoots issues, and performs patching and maintenance as required, monitoring all Project Dimension infrastructure locations for any problems to proactively fix them. In addition, all the operational tooling that customers use for their existing datacenter and cloud locations will be applied to Project Dimension locations. Project Dimension is expected to be a key portion of VMware's edge computing strategy, delivering edge computing-as-a-service; VMware expects that with Project Dimension, customers will be able to produce real-time business decisions (avoiding the latency associated with sending data to the public cloud), cost reduction (avoiding bandwidth usage), and improving business continuity (ensuring that the edge remains operational even with network failures or downtime).

VMware Cloud on AWS places VMware's software-defined data center in the AWS public cloud as a service, integrating vSphere, NSX, and vSAN with vCenter management, optimized to run on bare-metal AWS infrastructure. This enables customers to run applications across vSphere environments both on-premise as well as on AWS, allowing bi-directional migrations between environments. VMware Cloud on AWS is now available on several AWS availability zones (US East – N. Virginia, US West – Oregon, as well as Europe – London, Europe – Frankfurt, and Asia-Pacific – Sydney), with the company announcing at VMworld 2018 that they expect VMware Cloud on AWS to soon be available for Tokyo and AWS GovCloud – US.

Despite strong early adoption trends, however, management does not expect material uplift to FY19 (ending January) revenue. In our recent conversations with customers and distributors, we have also picked up an increased sensitivity and interest for cross-cloud management, which we expect VMware to benefit from with the platform, which is expected to offer greater efficiency in their customers' hybrid environments.

Amazon RDS on vSphere: Amazon's relational database service (RDS) is a fully managed database offering that allows customers to choose from one of six database engines (Amazon Aurora, MySQL, MariaDB, Oracle, Microsoft SQL Server, and PostgreSQL), with AWS managing the entire infrastructure stack, from the VM through routine database tasks like provisioning, patching, backup, recovery, failure detection, and repair. Since inception, RDS has been limited to AWS infrastructure; however, at VMworld 2018, VMware announced Amazon RDS on vSphere, allowing RDS to operate for on-premises deployments yet remain compatible with AWS, meaning that RDS databases can be moved into and out of AWS as needed. Our view is that Amazon RDS on vSphere could be a key capability for edge computing, as a cloud-native database can now straddle both the public cloud and the edge.

vSphere Platinum Edition, announced at VMworld 2018, is a premium SKU at a higher price point and includes VMware AppDefense, originally announced a year ago at VMworld 2017. VMware AppDefense is a data center endpoint security product that protects applications running in virtualized environments. AppDefense leverages VMware's hypervisor competencies and uses machine learning to create a baseline of a VM's "good state" (intended behavior), flagging changes that could indicate a threat. This is in contrast to continuously monitoring for threats: with attack vectors becoming increasingly diverse, hunting for a small section of malicious code among the legitimate code ("chasing bad") is much more difficult than performing the inverse ("ensuring good"). With the need to secure a variety of new endpoints and edge servers, we would expect to see solid uptake as the edge becomes increasingly important, given the criticality of securing the edge and the connection to public cloud.

VMware Kubernetes Engine (VKE) is a fully managed (i.e. a software-as-a-service offering) that allows customers to create Kubernetes clusters on public clouds; VKE, in public beta, was initially launched with support for AWS, and the company anticipates adding Azure in the future. By hosting Kubernetes in the cloud for its customers, VMware plans to eliminate the complexity associated with standing up Kubernetes clusters for its customers, who focus solely on the applications that run atop Kubernetes (rather than the infrastructure itself).

We note, however, that VMware's 10-Q states as a risk:

"We may also fail to adequately anticipate and prepare for the commercialization of emerging technologies such as blockchain and the development of new markets and applications for our technology such as the Internet of Things and "edge" computing and thereby fail to take advantage of new market opportunities or fall behind early movers in those markets."

Exhibit 42: Microsoft utilities comps

$ in mn, except per share items

Microsoft utilities comps (FactSet consensus estimates)																									
					EPS				P/E			EV/OCF		Sales				Sales growth				Operating Margin		OCF Margin	
Ticker	Name	Share price	Market cap	Enterprise value	CY17	CY18	CY19	CY20	CY18	CY19	CY20	CY18	CY19	CY17	CY18	CY19	CY20	CY17	CY18	CY19	CY20	CY18	CY19	CY18	CY19
NEE	NextEra Energy, Inc.	$169.09	79,744	114,887	$6.70	$7.76	$8.37	$8.99	21.8x	20.2x	18.8x	18.0x	13.2x	17,195	17,624	18,868	19,796	6%	2%	7%	5%	29%	31%	36%	46%
DUK	Duke Energy Corporation	$79.98	56,974	111,939	$4.57	$4.73	$4.97	$5.22	16.9x	16.1x	15.3x	15.4x	13.4x	23,565	23,953	24,682	24,995	4%	2%	3%	1%	24%	24%	30%	34%
SO	Southern Company	$43.54	44,156	95,251	$3.02	$2.99	$3.01	$3.11	14.6x	14.5x	14.0x	14.7x	14.9x	23,031	22,782	22,640	22,669	16%	(1%)	(1%)	0%	23%	25%	29%	28%
D	Dominion Energy Inc	$71.78	46,927	86,472	$3.60	$4.12	$4.25	$4.43	17.4x	16.9x	16.2x	15.9x	15.5x	12,586	13,374	13,832	14,239	7%	6%	3%	3%	36%	37%	41%	40%
EXC	Exelon Corporation	$42.73	41,273	78,008	$2.60	$3.11	$3.09	$3.09	13.8x	13.9x	13.8x	10.0x	10.7x	33,701	33,426	33,526	33,788	7%	(1%)	0%	1%	15%	15%	23%	22%
AEP	American Electric Power Company, Inc.	$71.22	35,107	59,123	$3.68	$3.93	$4.13	$4.38	18.1x	17.2x	16.3x	13.7x	13.2x	15,400	15,431	15,924	16,594	(6%)	0%	3%	4%	22%	22%	28%	28%
PEG	Public Service Enterprise Group Inc	$53.59	27,080	41,381	$2.93	$3.10	$3.30	$3.55	17.3x	16.3x	15.1x	13.5x	13.1x	9,420	9,650	10,093	10,274	4%	2%	5%	2%	26%	26%	32%	31%
ED	Consolidated Edison, Inc.	$76.05	23,659	40,715	$4.12	$4.26	$4.39	$4.58	17.8x	17.3x	16.6x	14.6x	11.7x	12,033	12,201	12,628	12,864	(0%)	1%	3%	2%	20%	21%	23%	27%
PCG	PG&E Corporation	$46.62	24,110	43,046	$3.68	$3.82	$4.03	$4.18	12.2x	11.6x	11.2x	9.8x	9.8x	17,135	17,365	17,948	18,375	(3%)	1%	3%	2%	17%	20%	25%	24%
PPL	PPL Corporation	$29.97	20,966	42,359	$2.25	$2.34	$2.43	$2.55	12.8x	12.3x	11.8x	15.5x	14.4x	7,447	7,666	8,016	8,345	(1%)	3%	5%	4%	39%	40%	36%	37%
EIX	Edison International	$68.41	22,289	39,002	$4.50	$4.11	$4.51	$4.85	16.7x	15.2x	14.1x	11.2x	10.0x	12,320	12,588	13,031	13,470	4%	2%	4%	3%	18%	20%	28%	30%
AWK	American Water Works Company, Inc.	$87.45	15,784	23,924	$3.03	$3.29	$3.56	$3.87	26.6x	24.5x	22.6x	18.6x	17.2x	3,357	3,410	3,610	3,788	2%	2%	6%	5%	34%	35%	38%	38%
				Mean					**17.2x**	**16.3x**	**15.5x**	**14.2x**	**13.1x**					**3%**	**2%**	**3%**	**3%**	**25%**	**26%**	**31%**	**32%**
				Median					**17.1x**	**16.2x**	**15.2x**	**14.7x**	**13.2x**					**4%**	**2%**	**3%**	**3%**	**23%**	**25%**	**29%**	**31%**
MSFT	**Microsoft**	$105.91	812,141	772,236	$3.60	$4.08	$4.59	$5.36	26.0x	23.1x	19.7x	16.4x	14.4x	103,586	116,657	129,222	144,311	10%	13%	11%	12%	32%	33%	40%	42%

Source: FactSet, Goldman Sachs Global Investment Research

Valuation & risks

Microsoft (MSFT): Reiterate Buy rating and $123 12-month PT, based on equally weighted DCF (WACC 9%), EV/FCF (22x CY19), and P/E (25x CY19). Key risks include adoption of hybrid cloud, Windows and Office performance, IT spending, and macro trends.

Google (GOOGL): Maintain Buy rating (on CL) and $1,500 12-month PT, based on an equal-weighted blend of DCF (11x terminal), P/E (29x CY19), and EV/EBITDA (17x CY19). Risks to our investment thesis include worsening macro, user fatigue, and impact from privacy concerns/GDPR.

Amazon (AMZN, covered by Heath Terry): Buy-rated (on CL) and $2,250 12-month PT, based on our SOTP. Key risks include competition, margin pressures from investment, and valuation.

Pivotal (PVTL): Maintain Buy rating and $25 12-month PT, based on equally weighted DCF (~5% perpetuity growth rate) and EV/sales (8.5x CY19). Key risks include enterprise, IT spending, competition, and ramp to target operating model.

Red Hat (RHT): Maintain Sell rating and $140 12-month PT, based 85% on a fundamental component (equally weighted DCF (~4% perpetuity growth), EV/billings (6x CY19), and EV/FCF (20x CY19), as well as 15% weighted on an M&A component (7x EV NTM sales). Key risks include the trajectory of IT spend and new product uptake.

VMware (VMW): We are Not Rated on VMware.

Rating and pricing information: Alphabet Inc. (B/A, $1,090.74), Amazon.com Inc. (B/A, $1,719.36), Microsoft Corp. (B/A, $105.91), Pivotal Software Inc. (B/A, $17.09), Red Hat Inc. (S/A, $117.38) and VMware Inc. (NR, $141.49).

Financial Advisory Disclosure

Goldman Sachs and/or one of its affiliates is acting as a financial advisor in connection with an announced strategic matter involving the following company or one of its affiliates: VMware, Inc.

Disclosure Appendix

Reg AC

I, Heather Bellini, CFA, hereby certify that all of the views expressed in this report accurately reflect my personal views about the subject company or companies and its or their securities. I also certify that no part of my compensation was, is or will be, directly or indirectly, related to the specific recommendations or views expressed in this report.

Unless otherwise stated, the individuals listed on the cover page of this report are analysts in Goldman Sachs' Global Investment Research division.

GS Factor Profile

The Goldman Sachs Factor Profile provides investment context for a stock by comparing key attributes to the market (i.e. our coverage universe) and its sector peers. The four key attributes depicted are: Growth, Financial Returns, Multiple (e.g. valuation) and Integrated (a composite of Growth, Financial Returns and Multiple). Growth, Financial Returns and Multiple are calculated by using normalized ranks for specific metrics for each stock. The normalized ranks for the metrics are then averaged and converted into percentiles for the relevant attribute. The precise calculation of each metric may vary depending on the fiscal year, industry and region, but the standard approach is as follows:

Growth is based on a stock's forward-looking sales growth, EBITDA growth and EPS growth (for financial stocks, only EPS and sales growth), with a higher percentile indicating a higher growth company. **Financial Returns** is based on a stock's forward-looking ROE, ROCE and CROCI (for financial stocks, only ROE), with a higher percentile indicating a company with higher financial returns. **Multiple** is based on a stock's forward-looking P/E, P/B, price/dividend (P/D), EV/EBITDA, EV/FCF and EV/Debt Adjusted Cash Flow (DACF) (for financial stocks, only P/E, P/B and P/D), with a higher percentile indicating a stock trading at a higher multiple. The **Integrated** percentile is calculated as the average of the Growth percentile, Financial Returns percentile and (100% - Multiple percentile).

Financial Returns and Multiple use the Goldman Sachs analyst forecasts at the fiscal year-end at least three quarters in the future. Growth uses inputs for the fiscal year at least seven quarters in the future compared with the year at least three quarters in the future (on a per-share basis for all metrics).

For a more detailed description of how we calculate the GS Factor Profile, please contact your GS representative.

M&A Rank

Across our global coverage, we examine stocks using an M&A framework, considering both qualitative factors and quantitative factors (which may vary across sectors and regions) to incorporate the potential that certain companies could be acquired. We then assign a M&A rank as a means of scoring companies under our rated coverage from 1 to 3, with 1 representing high (30%-50%) probability of the company becoming an acquisition target, 2 representing medium (15%-30%) probability and 3 representing low (0%-15%) probability. For companies ranked 1 or 2, in line with our standard departmental guidelines we incorporate an M&A component into our target price. M&A rank of 3 is considered immaterial and therefore does not factor into our price target, and may or may not be discussed in research.

Quantum

Quantum is Goldman Sachs' proprietary database providing access to detailed financial statement histories, forecasts and ratios. It can be used for in-depth analysis of a single company, or to make comparisons between companies in different sectors and markets.

GS SUSTAIN

GS SUSTAIN is a global investment strategy focused on the generation of long-term alpha through identifying high quality industry leaders. The GS SUSTAIN 50 list includes leaders we believe to be well positioned to deliver long-term outperformance through superior returns on capital, sustainable competitive advantage and effective management of ESG risks vs. global industry peers. Candidates are selected largely on a combination of quantifiable analysis of these three aspects of corporate performance.

Disclosures

Coverage group(s) of stocks by primary analyst(s)

Heather Bellini, CFA: America-Software. Mark Grant: America-Software.

America-Software: Adobe Systems Inc., Akamai Technologies Inc., Alphabet Inc., Atlassian Corp., Autodesk Inc., Citrix Systems Inc., Dropbox Inc., Endurance International Group, Facebook Inc., GoDaddy.com Inc., Microsoft Corp., MongoDB Inc., Okta Inc., Oracle Corp., Pivotal Software Inc., Red Hat Inc., RingCentral, Salesforce.com Inc., Twilio, VMware Inc., Wix.com, Workday Inc..

Company-specific regulatory disclosures

Compendium report: please see disclosures at http://www.gs.com/research/hedge.html. Disclosures applicable to the companies included in this compendium can be found in the latest relevant published research

Distribution of ratings/investment banking relationships

Goldman Sachs Investment Research global Equity coverage universe

	Rating Distribution			Investment Banking Relationships		
	Buy	Hold	Sell	Buy	Hold	Sell
Global	35%	54%	11%	64%	57%	55%

As of October 1, 2018, Goldman Sachs Global Investment Research had investment ratings on 2,814 equity securities. Goldman Sachs assigns stocks as Buys and Sells on various regional Investment Lists; stocks not so assigned are deemed Neutral. Such assignments equate to Buy, Hold and Sell for the purposes of the above disclosure required by the FINRA Rules. See 'Ratings, Coverage groups and views and related definitions' below. The Investment Banking Relationships chart reflects the percentage of subject companies within each rating category for whom Goldman Sachs has provided investment banking services within the previous twelve months.

Price target and rating history chart(s)

Compendium report: please see disclosures at http://www.gs.com/research/hedge.html. Disclosures applicable to the companies included in this compendium can be found in the latest relevant published research

Regulatory disclosures

Disclosures required by United States laws and regulations

See company-specific regulatory disclosures above for any of the following disclosures required as to companies referred to in this report: manager or co-manager in a pending transaction; 1% or other ownership; compensation for certain services; types of client relationships; managed/co-managed public offerings in prior periods; directorships; for equity securities, market making and/or specialist role. Goldman Sachs trades or may trade as a principal in debt securities (or in related derivatives) of issuers discussed in this report.

The following are additional required disclosures: **Ownership and material conflicts of interest:** Goldman Sachs policy prohibits its analysts, professionals reporting to analysts and members of their households from owning securities of any company in the analyst's area of coverage. **Analyst compensation:** Analysts are paid in part based on the profitability of Goldman Sachs, which includes investment banking revenues. **Analyst as officer or director:** Goldman Sachs policy generally prohibits its analysts, persons reporting to analysts or members of their households from serving as an officer, director or advisor of any company in the analyst's area of coverage. **Non-U.S. Analysts:** Non-U.S. analysts may not be associated persons of Goldman Sachs & Co. LLC and therefore may not be subject to FINRA Rule 2241 or FINRA Rule 2242 restrictions on communications with subject company, public appearances and trading securities held by the analysts.

Distribution of ratings: See the distribution of ratings disclosure above. **Price chart:** See the price chart, with changes of ratings and price targets in prior periods, above, or, if electronic format or if with respect to multiple companies which are the subject of this report, on the Goldman Sachs website at http://www.gs.com/research/hedge.html.

Additional disclosures required under the laws and regulations of jurisdictions other than the United States

The following disclosures are those required by the jurisdiction indicated, except to the extent already made above pursuant to United States laws and regulations. **Australia:** Goldman Sachs Australia Pty Ltd and its affiliates are not authorised deposit-taking institutions (as that term is defined in the Banking Act 1959 (Cth)) in Australia and do not provide banking services, nor carry on a banking business, in Australia. This research, and any access to it, is intended only for "wholesale clients" within the meaning of the Australian Corporations Act, unless otherwise agreed by Goldman Sachs. In producing research reports, members of the Global Investment Research Division of Goldman Sachs Australia may attend site visits and other meetings hosted by the companies and other entities which are the subject of its research reports. In some instances the costs of such site visits or meetings may be met in part or in whole by the issuers concerned if Goldman Sachs Australia considers it is appropriate and reasonable in the specific circumstances relating to the site visit or meeting. To the extent that the contents of this document contains any financial product advice, it is general advice only and has been prepared by Goldman Sachs without taking into account a client's objectives, financial situation or needs. A client should, before acting on any such advice, consider the appropriateness of the advice having regard to the client's own objectives, financial situation and needs. **Brazil:** Disclosure information in relation to CVM Instruction 483 is available at http://www.gs.com/worldwide/brazil/area/gir/index.html. Where applicable, the Brazil-registered analyst primarily responsible for the content of this research report, as defined in Article 16 of CVM Instruction 483, is the first author named at the beginning of this report, unless indicated otherwise at the end of the text. **Canada:** Goldman Sachs Canada Inc. is an affiliate of The Goldman Sachs Group Inc. and therefore is included in the company specific disclosures relating to Goldman Sachs (as defined above). Goldman Sachs Canada Inc. has approved of, and agreed to take responsibility for, this research report in Canada if and to the extent that Goldman Sachs Canada Inc. disseminates this research report to its clients. **Hong Kong:** Further information on the securities of covered companies referred to in this research may be obtained on request from Goldman Sachs (Asia) L.L.C. **India:** Further information on the subject company or companies referred to in this research may be obtained from Goldman Sachs (India) Securities Private Limited, Research Analyst - SEBI Registration Number INH000001493, 951-A, Rational House, Appasaheb Marathe Marg, Prabhadevi, Mumbai 400 025, India, Corporate Identity Number U74140MH2006FTC160634, Phone +91 22 6616 9000, Fax +91 22 6616 9001. Goldman Sachs may beneficially own 1% or more of the securities (as such term is defined in clause 2 (h) the Indian Securities Contracts (Regulation) Act, 1956) of the subject company or companies referred to in this research report. **Japan:** See below. **Korea:** Further information on the subject company or companies referred to in this research may be obtained from Goldman Sachs (Asia) L.L.C., Seoul Branch. **New Zealand:** Goldman Sachs New Zealand Limited and its affiliates are neither "registered banks" nor "deposit takers" (as defined in the Reserve Bank of New Zealand Act 1989) in New Zealand. This research, and any access to it, is intended for "wholesale clients" (as defined in the Financial Advisers Act 2008) unless otherwise agreed by Goldman Sachs. **Russia:** Research reports distributed in the Russian Federation are not advertising as defined in the Russian legislation, but are information and analysis not having product promotion as their main purpose and do not provide appraisal within the meaning of the Russian legislation on appraisal activity. **Singapore:** Further information on the covered companies referred to in this research may be obtained from Goldman Sachs (Singapore) Pte. (Company Number: 198602165W). **Taiwan:** This material is for reference only and must not be reprinted without permission. Investors should carefully consider their own investment risk. Investment results are the responsibility of the individual investor. **United Kingdom:** Persons who would be categorized as retail clients in the United Kingdom, as such term is defined in the rules of the Financial Conduct Authority, should read this research in conjunction with prior Goldman Sachs research on the covered companies referred to herein and should refer to the risk warnings that have been sent to them by Goldman Sachs International. A copy of these risks warnings, and a glossary of certain financial terms used in this report, are available from Goldman Sachs International on request.

European Union: Disclosure information in relation to Article 4 (1) (d) and Article 6 (2) of the European Commission Directive 2003/125/EC is available at http://www.gs.com/disclosures/europeanpolicy.html which states the European Policy for Managing Conflicts of Interest in Connection with Investment Research.

Japan: Goldman Sachs Japan Co., Ltd. is a Financial Instrument Dealer registered with the Kanto Financial Bureau under registration number Kinsho 69, and a member of Japan Securities Dealers Association, Financial Futures Association of Japan and Type II Financial Instruments Firms Association. Sales and purchase of equities are subject to commission pre-determined with clients plus consumption tax. See company-specific disclosures as to any applicable disclosures required by Japanese stock exchanges, the Japanese Securities Dealers Association or the Japanese Securities Finance Company.

Ratings, coverage groups and views and related definitions

Buy (B), Neutral (N), Sell (S) -Analysts recommend stocks as Buys or Sells for inclusion on various regional Investment Lists. Being assigned a Buy or Sell on an Investment List is determined by a stock's total return potential relative to its coverage. Any stock not assigned as a Buy or a Sell on an Investment List with an active rating (i.e., a stock that is not Rating Suspended, Not Rated, Coverage Suspended or Not Covered), is deemed Neutral. Each regional Investment Review Committee manages various regional Investment Lists to a global guideline of 25%-35% of stocks as Buy and 10%-15% of stocks as Sell; however, the distribution of Buys and Sells in any particular analyst's coverage group may vary as determined by the regional Investment Review Committee. Additionally, each Investment Review Committee manages Regional Conviction lists, which represent investment recommendations focused on the size of the total return potential and/or the likelihood of the realization of the return across their respective areas of coverage. The addition or removal of stocks from such Conviction lists do not represent a change in the analysts' investment rating for such stocks.

Total return potential represents the upside or downside differential between the current share price and the price target, including all paid or anticipated dividends, expected during the time horizon associated with the price target. Price targets are required for all covered stocks. The total return potential, price target and associated time horizon are stated in each report adding or reiterating an Investment List membership.

Coverage groups and views: A list of all stocks in each coverage group is available by primary analyst, stock and coverage group at http://www.gs.com/research/hedge.html. The analyst assigns one of the following coverage views which represents the analyst's investment outlook on the coverage group relative to the group's historical fundamentals and/or valuation. **Attractive (A).** The investment outlook over the following 12 months is favorable relative to the coverage group's historical fundamentals and/or valuation. **Neutral (N).** The investment outlook over the following 12 months is neutral relative to the coverage group's historical fundamentals and/or valuation. **Cautious (C).** The investment outlook over the following 12 months is unfavorable relative to the coverage group's historical fundamentals and/or valuation.

Not Rated (NR). The investment rating and target price have been removed pursuant to Goldman Sachs policy when Goldman Sachs is acting in an advisory capacity in a merger or strategic transaction involving this company and in certain other circumstances. **Rating Suspended (RS).** Goldman Sachs Research has suspended the investment rating and price target for this stock, because there is not a sufficient fundamental basis for determining, or there are legal, regulatory or policy constraints around publishing, an investment rating or target. The previous investment rating and price target, if any, are no longer in effect for this stock and should not be relied upon. **Coverage Suspended (CS).** Goldman Sachs has suspended coverage of this company. **Not Covered (NC).** Goldman Sachs does not cover this company. **Not Available or Not Applicable (NA).** The information is not available for display or is not applicable. **Not Meaningful (NM).** The information is not meaningful and is therefore excluded.

Global product; distributing entities

The Global Investment Research Division of Goldman Sachs produces and distributes research products for clients of Goldman Sachs on a global basis. Analysts based in Goldman Sachs offices around the world produce equity research on industries and companies, and research on macroeconomics, currencies, commodities and portfolio strategy. This research is disseminated in Australia by Goldman Sachs Australia Pty Ltd (ABN 21 006 797 897); in Brazil by Goldman Sachs do Brasil Corretora de Títulos e Valores Mobiliários S.A.; Ombudsman Goldman Sachs Brazil: 0800 727 5764 and / or ouvidoriagoldmansachs@gs.com. Available Weekdays (except holidays), from 9am to 6pm. Ouvidoria Goldman Sachs Brasil: 0800 727 5764 e/ou ouvidoriagoldmansachs@gs.com. Horário de funcionamento: segunda-feira à sexta-feira (exceto feriados), das 9h às 18h; in Canada by either Goldman Sachs Canada Inc. or Goldman Sachs & Co. LLC; in Hong Kong by Goldman Sachs (Asia) L.L.C.; in India by Goldman Sachs (India) Securities Private Ltd.; in Japan by Goldman Sachs Japan Co., Ltd.; in the Republic of Korea by Goldman Sachs (Asia) L.L.C., Seoul Branch; in New Zealand by Goldman Sachs New Zealand Limited; in Russia by OOO Goldman Sachs; in Singapore by Goldman Sachs (Singapore) Pte. (Company Number: 198602165W); and in the United States of America by Goldman Sachs & Co. LLC. Goldman Sachs International has approved this research in connection with its distribution in the United Kingdom and European Union.

European Union: Goldman Sachs International authorised by the Prudential Regulation Authority and regulated by the Financial Conduct Authority and the Prudential Regulation Authority, has approved this research in connection with its distribution in the European Union and United Kingdom; Goldman Sachs AG and Goldman Sachs International Zweigniederlassung Frankfurt, regulated by the Bundesanstalt für Finanzdienstleistungsaufsicht, may also distribute research in Germany.

General disclosures

This research is for our clients only. Other than disclosures relating to Goldman Sachs, this research is based on current public information that we consider reliable, but we do not represent it is accurate or complete, and it should not be relied on as such. The information, opinions, estimates and forecasts contained herein are as of the date hereof and are subject to change without prior notification. We seek to update our research as appropriate, but various regulations may prevent us from doing so. Other than certain industry reports published on a periodic basis, the large majority of reports are published at irregular intervals as appropriate in the analyst's judgment.

Goldman Sachs conducts a global full-service, integrated investment banking, investment management, and brokerage business. We have investment banking and other business relationships with a substantial percentage of the companies covered by our Global Investment Research Division. Goldman Sachs & Co. LLC, the United States broker dealer, is a member of SIPC (http://www.sipc.org).

Our salespeople, traders, and other professionals may provide oral or written market commentary or trading strategies to our clients and principal trading desks that reflect opinions that are contrary to the opinions expressed in this research. Our asset management area, principal trading desks and investing businesses may make investment decisions that are inconsistent with the recommendations or views expressed in this research.

The analysts named in this report may have from time to time discussed with our clients, including Goldman Sachs salespersons and traders, or may discuss in this report, trading strategies that reference catalysts or events that may have a near-term impact on the market price of the equity securities discussed in this report, which impact may be directionally counter to the analyst's published price target expectations for such stocks. Any such trading strategies are distinct from and do not affect the analyst's fundamental equity rating for such stocks, which rating reflects a stock's return potential relative to its coverage group as described herein.

We and our affiliates, officers, directors, and employees, excluding equity and credit analysts, will from time to time have long or short positions in, act as principal in, and buy or sell, the securities or derivatives, if any, referred to in this research.

The views attributed to third party presenters at Goldman Sachs arranged conferences, including individuals from other parts of Goldman Sachs, do not necessarily reflect those of Global Investment Research and are not an official view of Goldman Sachs.

Any third party referenced herein, including any salespeople, traders and other professionals or members of their household, may have positions in the products mentioned that are inconsistent with the views expressed by analysts named in this report.

This research is not an offer to sell or the solicitation of an offer to buy any security in any jurisdiction where such an offer or solicitation would be illegal. It does not constitute a personal recommendation or take into account the particular investment objectives, financial situations, or needs of individual clients. Clients should consider whether any advice or recommendation in this research is suitable for their particular circumstances and, if appropriate, seek professional advice, including tax advice. The price and value of investments referred to in this research and the income from them may fluctuate. Past performance is not a guide to future performance, future returns are not guaranteed, and a loss of original capital may occur. Fluctuations in exchange rates could have adverse effects on the value or price of, or income derived from, certain investments.

Certain transactions, including those involving futures, options, and other derivatives, give rise to substantial risk and are not suitable for all investors. Investors should review current options disclosure documents which are available from Goldman Sachs sales representatives or at http://www.theocc.com/about/publications/character-risks.jsp. Transaction costs may be significant in option strategies calling for multiple purchase and sales of options such as spreads. Supporting documentation will be supplied upon request.

Differing Levels of Service provided by Global Investment Research: The level and types of services provided to you by the Global Investment Research division of GS may vary as compared to that provided to internal and other external clients of GS, depending on various factors including your individual preferences as to the frequency and manner of receiving communication, your risk profile and investment focus and perspective (e.g., marketwide, sector specific, long term, short term), the size and scope of your overall client relationship with GS, and legal and regulatory constraints.

As an example, certain clients may request to receive notifications when research on specific securities is published, and certain clients may request that specific data underlying analysts' fundamental analysis available on our internal client websites be delivered to them electronically through data feeds or otherwise. No change to an analyst's fundamental research views (e.g., ratings, price targets, or material changes to earnings estimates for equity securities), will be communicated to any client prior to inclusion of such information in a research report broadly disseminated through electronic publication to our internal client websites or through other means, as necessary, to all clients who are entitled to receive such reports.

All research reports are disseminated and available to all clients simultaneously through electronic publication to our internal client websites. Not all research content is redistributed to our clients or available to third-party aggregators, nor is Goldman Sachs responsible for the redistribution of our research by third party aggregators. For research, models or other data related to one or more securities, markets or asset classes (including related services) that may be available to you, please contact your GS representative or go to http://360.gs.com.

Disclosure information is also available at http://www.gs.com/research/hedge.html or from Research Compliance, 200 West Street, New York, NY 10282.

Mindcraft: Our Thematic Deep Dives

Innovation & Disruption

Extended Reality

Drones

Factory of the Future

Blockchain

Precision Farming

eSports

Space

AI Hardware

The Store of the Future

B2B Payments

Banking on Technology

The Genome Revolution

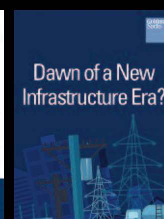
Why Technology is not a bubble

Insights & Policy

Top of Mind

Making Cents

US Infrastructure

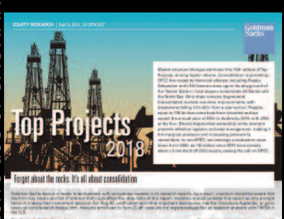


Healthcare's Holy Grail

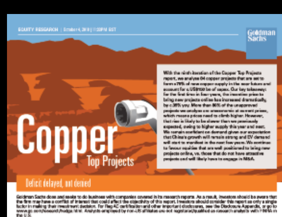
Trade Policy

Commodity Corner

Top Projects 2018

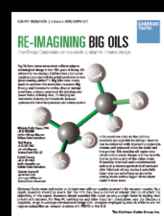


Copper Top Projects



Oil: Age of Restraint

Reimagining Big Oil



Factor Investing in Credit

Music's Return to Growth

The Low Carbon Economy

NextGen Power

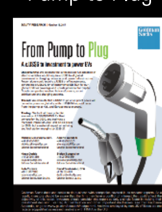
IMO 2020: Global Shipping

Power Shift 2.0

The Great Battery Race

More Lean, More Green

Pump to Plug



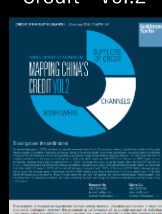
Americas Lodging: 2018 Hospitality Handbook

Rising Asia

Apple Suppliers' Dilemma

E+commerce

Mapping China's Credit - Vol.2



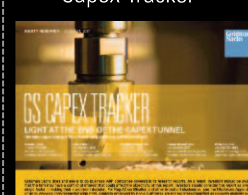
China AI

GS SUSTAIN

The PM's Guide To The ESG Revolution

Capex Upturn

Capex Tracker



... and more in **The Mindcraft Library**