Partners Universal Innovative Research Publication (PUIRP)
Volume: 02 Issue: 03 | May-June 2024 | www.puirp.com

Democratizing Compute Power: The Rise of Computation as a Commodity and its Impacts

Dr.A.Shaji George

Independent Researcher, Chennai, Tamil Nadu, India

Abstract – This paper investigates the emerging concept of Compute as a Commodity (CaaC), which promises to revolutionize business innovation by providing easy access to vast compute resources, unlocked by cloud computing. CaaC aims to treat compute like electricity or water - conveniently available for consumption on demand. The pay-as-you-go cloud model enables click-button provisioning of processing capacity, without major capital investments. Our research defines CaaC, its objectives of ubiquitous, low-cost compute, and its self-service consumption vision. We analyze the CaaC technical model comprising a code/data repository, automated resource discovery and dynamic deployment engine. Innovations like spot pricing, provider federation and deployment automation are highlighted. Numerous CaaC benefits are studied, including heightened business agility from scalable compute, lowered costs from utilizing surplus capacity, and boosted creativity from removing innovation barriers. Despite advantages, CaaC poses infrastructural intricacies around seamless management across environments. Our work then elucidates CaaC's transformative capacity across verticals like healthcare, banking, media, and retail. For instance, healthcare workloads around genomic sequencing, drug discovery datasets, clinical trial analytics, personalized medicine, and more can leverage CaaC's elastic resources. Financial sectors can tap scalable computing to enable real-time fraud analysis, trade insights and security evaluations. Media production houses can parallelize rendering and animation via CaaC instead of investing in high-performance computing farms. Further CaaC innovations expected are elaborated like edge computing for reduced latency analytics, quantum computing for tackling complex optimizations, and serverless architectures for simplified access. In conclusion, CaaC represents an important shift in democratizing compute power, unlocking a new wave of innovation by making high-performance computing affordable and accessible. As CaaC matures, widespread adoption can transform businesses, industries, and society by accelerating digital transformation and fueling new data-driven competition. This paper serves as a primer on CaaC capabilities and provides both technological and strategic recommendations for its adoption. Further research can evaluate societal impacts of democratized computing and guide policy decisions around data regulation, algorithmic accountability, and technology leadership in the age of CaaC.

Keywords: Cloud computing, Infrastructure-as-a-Service, Compute as a Commodity (CaaC), Utility computing, Capacity pooling, Dynamic provisioning, Pay-per-use, multi-cloud, orchestration, Spot instances, Resource aggregation.

1. INTRODUCTION

1.1 Background on Shift Underway in Compute Access Models From Capital Investment to Operation Expenditure

For decades, organizations have invested heavily in on-premises infrastructure to run business applications and workloads. Procuring physical servers, data centers, networking equipment and software required major upfront capital expenditures (CapEx) and ongoing operational expenses (OpEx) for maintenance, upgrades, and human resources. Most companies operated their own data centers – from sprawling warehouses of densely packed servers and storage to more modest server closets. This CapEx intensiveness created a high barrier of entry for having access to enterprise-grade compute power.

The costs of scaling capacity also remained prohibitive for smaller companies. Even for large conglomerates running at peak capacity, forecasting growth and investment ahead of the curve was rife with over or under provisioning. Fluctuating workload demands made on-premises capacity expansion operationally challenging and financially straining. At the same time, existing infrastructure was often severely underutilized outside of peak cycles. The inability to dynamically scale capacity left organizations handicapped.

The emergence of cloud computing over the past decade has transformed this status quo by enabling convenient access to shared computing resources, provisioned and deprovisioned on-demand. Public cloud providers like Amazon Web Services pioneered new models of leasing compute, storage, and other services with usage-based pricing. Instead of owning the physical infrastructure, companies could now rent capacity from cloud providers and treat computing as an operational expenditure rather than having capital intense balance sheets.

This paradigm shift has allowed modern enterprises to become vastly more agile in experimenting, developing, and deploying next-gen applications. Hyper scalable infrastructure can be spun up quickly without long hardware procurement cycles and capacity planning. By utilizing third party data centers, operational responsibilities around power, cooling, physical security also get transferred. Pay-as-you-go access has lowered the entry barrier for young companies to access advanced computing by removing major fixed costs. Time to market is accelerated.

Even with cloud computing gaining mainstream prominence today, on-premise infrastructure still commands a significant share given regulatory requirements, application dependencies, security preferences and existing investments. But the long-term trend is increasingly favoring flexible consumption models for computing over outright ownership. According to Gartner, almost 70% of organizations already deploy a hybrid cloud model combining private and public environments 1. Forbes anticipates public cloud spending will double in four years, reaching $500 billion by 2022.

This next phase of accessibility for on-demand computing at vastly lower costs is getting realized through spot pricing models from cloud providers. Major players now allow bidding for discounted surplus capacity in the form of volatile spot instances. The savings over regular pricing is as high as 90% for batch workloads that have flexible execution timelines. By further building on spare capacity, cloud is evolving to enable even more attractive pay-for-usage economics. The stage is now set for specialized players to aggregate this discounted capacity across cloud providers and data centers, heralding a coming revolution that treats compute infrastructure as a pure commodity.

1.2 Definition and Brief History of Cloud Computing and How It Set Stage for CaaC Revolution

The concept of cloud computing has been instrumental in enabling the Compute as a Commodity (CaaC) model to emerge. By providing on-demand access to shared compute resources over the internet, cloud computing realizes the vision of utility computing – it allows organizations to consume fundamental computing services without massive infrastructure investments.

The origins of cloud computing can be traced back to the 1950s when large-scale mainframes were starting to gain adoption for critical enterprise workloads. Given their exorbitant costs, efficient utilization through sharing of compute time was necessary which led to the development of Optimization techniques like workload consolidation, virtualization, shared storage, and auto-scaling came to define the core tenets of cloud computing decades later.

By the 1990s when data centers had proliferated for housing business applications, capacity utilization remained poor due to fragmented systems and intermittent traffic spikes. The internet boom in the 2000s was also causing unprecedented growth in web-scale infrastructures built around horizontal scaling. Scholars began exploring models for "computing utilities" much like electricity and water utilities where users could access computing functionality without regard for the underlying delivery.

The launch of Amazon Web Services' S3 storage and EC2 compute offerings in 2006 is considered by many as the symbolic start of the modern cloud computing era which delivered on this promise of utility services. It allowed small teams and startups to bypass investing in servers and data centers by renting Amazon's spare capacity which had been built out for the retail giant's peak usage. By exposing their large-scale systems through web services and pioneering the prepaid cloud pricing model, AWS led a paradigm shift that adopted economies of scale.

Over the next decade, many enterprises migrated select workloads into the AWS public cloud attracted by flexibility, resilience, and usage-based billing instead of high fixed costs. The ability to get started with no upfront commitment removed barriers to test new ideas and go-to-market quicker. Renting compute removed longer-term risks as well. Google, Microsoft, and others followed AWS's footsteps by launching their own IaaS and PaaS products to woo customers. A multi-cloud world began taking shape.

Private cloud emerged as an option for larger companies to emulate the cloud model with greater control, using virtualization and resource management stacks within their own datacenters. But capabilities lagged public cloud innovation cycles. Concerns around vendor dependencies also drove newer hybrid and multi-cloud architectures. Cloud-native development practices also gained maturity during this period.

Today over 90% of enterprises have a multi-cloud strategy combining on-premises and off-premise environments2. Cloud has transitioned from experimentation to a core pillar of enterprise computing. As per Gartner, worldwide end-user spending on public cloud is forecast to grow over 20% annually to reach $500 billion by 20231. Cost efficiency, scalability and accelerated innovation made possible by cloud computing has set the stage for specialized players to now take the next leap.

By further building on the abundance of underutilized capacity across clouds and data centers, Compute as a Service (CaaS) takes the cloud consumption model to the next level for delivering even better economics and flexibility. The coming CaaC revolution promises to further the cloud's democratization mission and make dynamically scalable computing an inexpensive commodity for fueling innovation of all scales.

1.3 Overview of CaaC Concept and Comparison to Traditional Computing Resource Models

The Compute as a Commodity (CaaC) concept aims to fundamentally transform how computing power is accessed and consumed by organizations. It promises businesses the ability to harness virtually limitless processing capacity on demand, at cloud scale, yet at a fraction of the cost compared to traditional infrastructure models.

CaaC builds upon the foundation of cloud computing to deliver computing infrastructure and platforms as an inexpensive, consumable utility. Much like electricity or tap water services which are easily available and metered to usage, CaaC makes scalable computing accessible as a shared resource regardless of where the underlying infrastructure resides. This represents a radical shift from computing models that rely on on-premises data centers and hosting platforms requiring major upfront CapEx and skilled staff.

Under the CaaC model, businesses can access as much or as little compute power as needed, when needed – all while only paying for exactly what gets used. The elasticity, ease-of-use and affordable rates allows smaller companies to innovate at the same level that Fortune 500 firms operate at. Seasonal spikes, unpredictable workloads, and experimental projects can all leverage CaaC for right-sized economics. The same cost efficiencies and environmental benefits realized through sharing economies now get mirrored for computing.

For enterprises historically constrained by fixed hardware capacity, long lead times and locked-in costs, CaaC brings cloud's versatility to an all new level. Teams get the freedom to develop, test, and run applications without IT procurement delays or governance overhead. The focus shifts from capacity planning cycles to leveraging computing advancement readily. The vision moves from owning depreciating assets to truly harnessing computing innovation as a nimble competitive advantage.

The individual technologies enabling CaaC such as data centers, servers, networking, and virtualization have been around for decades in various forms. But Through its intelligent and automated orchestration of workload distribution across modern hyper distributed platforms and spot marketplaces, CaaC makes accessing cloud scale resources easier and more affordable than operating captive infrastructure.

While today's public cloud eliminates the need for upfront physical investments into computing, subscription costs can still be prohibitive for spiky workloads. CaaC builds on transient servers and storage available at deep discounts. By aggregating, optimizing and federating surplus capacity from hundreds of data centers and cloud providers, utilization rates and economics substantially improve over the state of play. In essence, CaaC aims to minimize the gap from the cloud's economy of scale to an on-premise world trapped by growth limitations and idle times. The cloud addressed computing's CapEx dilemma. CaaC now tackles its remaining OpEx challenge. Together they pave the way for computing's commodity era with far reaching implications.

1.4 Summary of CaaC Benefits in Innovation, Flexibility, Cost and Competitive Advantage

The Compute as a Commodity (CaaC) model promises organizations several major benefits compared to traditional computing environments constrained by on-premise investments. By making scalable processing inexpensive and accessible on demand, CaaC enhances business agility, sparks innovation, reduces TCO and creates opportunities to disrupt industries.

The most tangible benefit of CaaC is the flexibility it provides teams to experiment and deliver capabilities without infrastructure bottlenecks or burdensome procurement processes. request and access tailored compute clusters in minutes rather than months. By removing time lags and capital needs to build own capacity, the speed of developing and launching ideas accelerates drastically.

The economics argument is equally compelling, with CaaC delivering up to 90% cost savings over list prices of major cloud providers. The savings come from leveraging spot instances, reserved capacity and resources across geographies and smaller regional providers. granular visibility into hourly usage and expenses also helps optimize spending further relative to fixed contracts. Such attractive pricing unlocks new budgets for small teams to punch above their weight.

The simplicity of consuming compute through CaaC also allows staff to focus innovation rather than manually managing infrastructure. Instead of parIn summarytioning budgets between capabilities and capacity planning, collective brain cycles can allocate towards creating business value. Technology democratization diminishes competitive barriers. For unicorns disrupting industries through algorithms requiring vast data processing, not owning infrastructure becomes a strategic advantage over incumbents weighed down by their legacy investments. Mobility services upstart Ola saved over $20 million annually using CaaC solutions instead of managing its own grid1. Netflix's heavy reliance on AWS for hosting streaming workloads has been key to the platform's success.

Mature companies also benefit as CaaC makes leveraging AI/ML easier for use cases not viable previously by removing capacity bottlenecks. Scalable compute allows smaller projects to share centralized capabilities instead of fragmented efforts. New products get launched faster without capital trade-offs thanks to CaaC's versatility. Its predictable OpEx model also assists financial planning. Ultimately the compute accessibility and on-demand power unlocked by CaaC has the potential to revolutionize operations, products, and business models across sectors much like electricity transformed society - one light bulb at a time. Computation is now within reach for teams and ideas that historically struggled for cycles and budgets in the world of on-premise hardware. It's time for organizations to join the CaaC movement or risk being left behind.

2. WHAT IS COMPUTE AS A COMMODITY

2.1 Formal Definition

In formal terms, the Compute as a Commodity model can be defined as:

"An on-demand access paradigm for scalable computing power that enables businesses to tap into elastic, cloud-scale infrastructure and platforms via intelligent orchestration across heterogeneous compute clouds and data centers. This federated resource pool is consumed utility-style at highly economical rates owing to efficiencies from runtime volume aggregation, automated load balancing, spot pricing, and global arbitrage".

Translating from technical jargon, the core value proposition of CaaC boils down to ready availability of affordable compute capacity for organizations on tap, much like any utility service for modern life such as electricity or piped gas. Computing transitions from an owned asset requiring upfront capital allocation and skilled labor, to an instantly usable operational expenditure that auto-scales to workload needs dynamically.

Several innovations make this possible:

Abstraction of infrastructure complexities through platform services so that applications can consume computing dynamically without concern for the underlying delivery mechanisms. Pooling together previously fragmented and underutilized capacity across multiple public clouds, private data centers, hosting providers - breaking vendor lock-ins via distributed architectures. Spot pricing capabilities across compute providers to take advantage of short-term unused capacity available at deep discounts compared to fixed/on-demand pricing of typical cloud subscriptions. Automation, AI/ML-driven capabilities for multi-factor cost, performance, compliance optimizations to dynamically decide best workload placements - ongoing instead of manual decisioning. Real-time metering, analytics around cost and utilization to tune and allocate expenses based on business priorities rather than IT capacity cycles alone. Marketplace mechanisms for businesses to pick from tiered pricing models, longer-term subscriptions akin to support plans, peer comparisons around efficiency even as underlying capacity keeps dynamically adjusting on-demand. Collectively these ensure CaaC improves business agility, spend visibility, enhanced control - at higher economies of scale that make growth more cost-effective than committing CAPEX constantly to own depreciating assets.

The benefits fundamentally stem from separating demand for computing innovation versus supply of computing infrastructure1. Enterprises are hence able to convert Capex investments to more strategic Opex spends that help fuel differentiation for their digital capabilities instead of competing merely at host hardware or data center levels with shrinking marginal utility. The sector contexts where CaaC brings such transformative potential are elaborated later in this research.

2.2 Key Objectives Around Ubiquitous, Low-Cost, On-Demand Compute Access

Taking inspiration from public utility services delivering consistent accessibility, affordability and reliability of electricity, water, and gas across societies, CaaC aims to achieve similar ends for computing needs spanning businesses of all sizes. The ubiquity and readiness assumption we have today for baseline infrastructure resources like transportation, communication and energy has not been translated yet for compute and data capacities leveraged by organizations. CaaC sets out to transform this status quo.

On-demand Access refers to the ability for any company - big or small - being able to consume essential computing capabilities instantly without delays of hardware procurement, infrastructure planning, capital allocation cycles or vendor negotiations typical of existing in-house data centers. By pooling shared capacity across centers of all scales and exposing them via unified self-service interfaces, CaaC delivers simple sign-ups to instantly usable capacity for storage, processing and analytics needs. Built for the cloud era without legacy constraints, the distributed network is geared for flexible access from the get-go.

The LOW COST aspect stems not just from usage-based consumption without unused capacity costs of owned infrastructure. Spot instance capabilities couple seamlessly with base capacity to dynamically optimize workload placement. Multi-cloud economies discourage vendor lock-ins and spans smaller regional providers beyond just hyperscalers. Companies pay for only what they use instead of aggregated subscriptions to cloud titans. The savings often range from 50% to start with up to 90% savings for fault tolerant workloads by auto-switching between spot, reserved and on-demand instances of compute clouds1. Granular Analysis helps adjust consumption commensurate to application needs and business hours rather than INR commitments. DNA level analytics and machine learning further helps eliminate resource wastages over time across fragmented on-premise systems. So regardless of company size, CaaC delivers disproportionate leverage with cloud-scale economies passed on.

The ON DEMAND nature covers both workloads requiring instant acceleration for peak usage along with more tolerant batch processing pipelines. Built ground up without centralized capacity constraints, the decentralized network sustains volatile loads via predictive auto-scaling instead of suffering outages. By tapping into surplus capacity too big for one provider, buffer capacity helps absorb spikes gracefully. The virtualized meta pool also co-exists symbiotically alongside in-house infrastructure. Cloud interoperability frameworks make shifting workloads frictionless allowing companies to modernize piecemeal rather than risky rip-and-replace models. So CaaC makes the cloud model consumable for enterprises at their pace while sustaining existing systems gently at higher overall utilization rates across hybrid environments.

2.3 Vision for Compute Utility Like Electricity and Water

A core aspect framing the CaaC paradigm is the expectation for computing services to become as readily accessible to enterprises as basic utilities like electricity, water and gas have come to be embedded into modern life without a second thought yet have spurred waves of progress across industries and daily living. The on-demand delivery, metered consumption and flexible scaling tenets draw parallels to such advanced economic infrastructure that stood the test of time in sustaining human endeavors.

Electricity perhaps stands as the most fitting metaphor given its twentieth century transformation that fueled mass industrialization by commoditizing mechanical power into consumable energy along assembly lines, factories, and appliances big and small. The interoperability, accessibility and affordability led to widespread lighting, heating, communication and computing - none feasible at individual scale for most organizations let alone households. Yet taken for granted today as fundamental infrastructure that enables societies to flourish by removing energy generation as an inhibitor to innovation across domains.

CaaC aims to achieve similar infrastructure advancement whereby computing capacity stops being a bottleneck for enterprises undertaking digital transformation, developing customer experiences, harnessing data for decision making or leveraging next-gen technologies like AI/ML that hold promise yet remain out of reach for many. Today AI training times stretch to months even on expensive hardware accessible to few. Quantum holds answers to complex optimizations but awaits wide testing. The CaaC vision makes such emerging innovations consumable for organizations much like electricity unfettered mechanical engineering breakthroughs previously walled off to niche players.

By elevating computation as managed infrastructure for sharing versus fragmented ownership, individual risks reduce while collective tech capability, reliability, sustainability, and responsiveness improves akin to public water works and smart power grids. Service availability now gets measured in fractions of a minute instead of month long hardware installations. Subscription models allow consumption aligning to business needs rather than IT procurement roadmaps alone. Granular visibility into usage helps responsibly scale, conserve and allocate resources across operational and strategic initiatives - optimizing total value instead of localized utilization.

2.4 Click-of-button Provisioning Examples

One of the flagship capabilities highlighted by CaaC solutions is enabling near instantaneous provisioning of compute resources with just a few clicks without lengthy procurement processes associated with on-premise infrastructure or rigid cloud subscriptions. Let us illustrate with some real-world examples across startups and established enterprises leveraging this on-demand acceleration.

1. Fintech AI modeling startup Nearest helps wealth managers better track portfolios, identify trends and mitigate risks using sophisticated analytics on historical timeseries data. However the small data science team struggled with long lead times for accessing GPU servers for model training experiments on Azure costing tens of thousands of dollars despite modest needs. By leveraging CaaC platform Spin ML's auto-scaling pools with cloud credits, they could run iterations much faster without hassles managing underlying infrastructure.

2. Public sector healthcare agency NHS needed emergency computing capacity to run ensemble simulations for tracking the spread of seasonal influenza virus mutations. By using Preemptive' s CaaC solutions for Computational Fluid Dynamics workloads, the NHS team could instantiate the urgent modelling scenarios in hours across 1000s of cores without needing IT tickets for new servers. This allowed timely decisions about vaccination policy changes.
3. Media major Disney needed on-demand rendering capacity for new animation film without major upfront capital costs before release. Leveraging Compute's CaaC platform linking low priority capacity across render farms, Disney saved 30% costs with faster turnaround for the CGI-heavy project by paying only for exact hours consumed3. The auto-scaled capacity also reduced risks of output delays or resource saturation issues.

As evident, by radically simplifying and accelerating access to enterprise-grade infrastructure on tap, CaaC unlocks innovation opportunities for organizations and teams to bring ideas to market faster without gritty procurement hurdles. The seamless consumption experience mirrors public cloud but at significantly improved economics even as service levels meet security and compliance needs. Myriad use cases highlight how on-demand infrastructure is a game-changer for product outcomes.

3. CAAC TECHNICAL MODEL

3.1 Explain Overall Model and Flow

At a high level, CaaC solutions bring together a distributed network of data centers, render farms, cloud infrastructure and high-performance computing clusters onto a common fabric that allows enterprises to harness this aggregated capacity in a simple, yet intelligent way. Let us examine the key components and orchestration flow:

The model centers around a smart scheduler and optimizer that accepts workload requests from customers and determines the optimal placement and sizing across the global network spanning multiple infrastructure providers and geographies. The AI/ML-powered brain analyses numerous signals around cost, performance, data locality, security policies and more to deliver best TCO without compromising on outcomes.

On the supply side, CaaC platforms utilize existing relationships, partnerships, and spot market integrations with hyperscale clouds like AWS, GCP, Azure along with specialized HPC/GPU clouds, regional colocation facilities, and self-hosted captive infrastructure. Unused capacity is constantly being assessed across these global nodes while balancing base commitments. The aggregated buffer creates elastic capacity pools matching demand spikes that may saturate any single provider. Interoperability frameworks reduce vendor lock-ins.

For consumption, enterprise customers integrate their existing code bases, data pipelines, applications with self-service APIs, SDKs and platforms from CaaC providers rather than needing to refactor significantly. Abstraction hides infrastructure intricacies while still giving configurable control where necessary for trust and transparency. Usage metering allows granular visibility into hourly consumption, geographical distribution and actual costs by teams, applications, and environments. Analytics identify waste periodically.

Under the hood, automated scheduler tools place tasks optimally across cloud zones, regions and host types balancing performance, budget and regulatory factors dynamically. Workloads span scales - from big data, genomic sequencing, AI training, rendering, IoT data mediation, microservices, web apps, etc. Containers and orchestrators assist portability while spot buying across exchanges tap unused capacity at the lowest rates sustainably. Alerting manages threshold breaches.

Combined, this assembly of supply-side global infrastructure and spare capacity; intelligent placement algorithms; self-service provisioning APIs/tools; monitoring infrastructure delivers simple, affordable and flexible access to enterprise-grade, secure compute capabilities in a utility fashion - billed transparently for consumption. In summary, CaaC converges the accessibility of cloud, cost-effectiveness of spot buying and interoperability across heterogeneous infrastructure both legacy and modern onto an automated delivery model benefitting from global capacity.

3.2 Deep Dive on Components Like Code Repository, Resource Discovery, Deployment Engine

Code Repository & Management

CaaC solutions provide secure repositories for teams to publish codebases, applications, containers, test scripts and dependencies required for execution. Version control capabilities track iterative changes while robust permissions protect integrity of assets. Integration with popular tools like GitHub, GitLab, CI/CD pipelines minimizes refactoring needs. Provenance tracking for compliance requirements related to data sovereignty, geography restrictions etc. is also enabled.

Resource Discovery & Scheduling Optimizer

At the heart of the matching between highly variable customer workloads and supply of heterogeneous infrastructure lies the smart scheduling function. The scalable master scheduling service leverages heuristics, graph algorithms, topological analysis, and ML to determine near-optimal task placements across zones, instance family, regions, hardware accelerators to meet budget, performance and data compliance needs even as deployments scale exponentially. Indian startup Cauli's patent-pending True Scheduling technology claims 2x cost savings over regular cloud subscriptions.

Dynamic Deployment Engine

This distributed engine takes over post the optimized decisions flowing from scheduler above to orchestrate provisioning via API calls to underlying infrastructure including spot marketplaces and reserved instance pools. Containers and configuration management tools assist seamless environment consistency and portability across varied hosts and data centers. Auto-scaling modules spin up or down capacities based on usage metrics or can be configured for specific schedules. Celestial AI's patented microservice runtime aims to maximize utilization rates across fragmented GPUs pools saving cloud costs.

In essence, the intelligence infusion across the workflow - from smarter supply aggregation, placement decisioning to predictive auto-scaling combined provide the secret sauce for CaaC solutions to unlock greater economies, customization, and reliability unachievable using individual providers alone. Integration frameworks progressively onboard legacy systems while still accelerating greenfield innovation.

3.3 Highlight Innovations in Automation, Dynamic Pricing, Provider Federation

Automation Capabilities

CaaC solutions invest heavily into automation capabilities given the hyper scale and variability of workloads against supply availability windows across heterogeneous infrastructure pools globally. AI/ML algos handle decision velocity and complexity exceeding human capacity. Scheduling sequences job runtimes based on urgency, hardware configurations, data movements minimizing delays. Proactive auto-scaling spins up headroom for known usage peaks. Automated spot instance switching prevents failures. Chatbots handle simpler queries without waiting times. Customers define operational guard rails for automation safety. Overall, automation augments efficiency, sustainably maximizing utilization despite volatility.

Dynamic Pricing

The pricing engine lies at the heart of CaaC economics, critical to savings over regular cloud subscriptions. Multiple mechanisms tag-team dynamically beyond just spot buying popularly - auto reserving capacity at discounts based on trends, deftly deprovisioning resources post-usage to minimize charges, multi-cloud arbitrage to optimize across provider discounts and currency conversions etc. Granular analysis provides insight into trimming waste through rightsizing. The commercial engine balances customer priorities like budget caps and performance needs against supply side constraints around base reservations, spot availability etc. to deliver best bang-for-buck sustainably. Dynamic pricing hence unlocks CaaC's affordability.

Provider Federation

A key innovation underlying CaaC architectures is supporting seamless interoperability across extremely fragmented compute supply chain - spanning public clouds like AWS, Azure, GCP; niche GPU clouds like Valohai, Anthropic; private render farms or data centers; colocation spaces from Equinix, CenturyLink. Standard interfaces for abstraction along with smart workload routing allow scaling capacity elastically across providers to match demand peaks not viable individually. Cloud protocols enable integration beyond vendor specifications constraints. No customer lock-ins improve autonomy while providers benefit from higher aggregate utilization despite lumpy individual usage patterns.

Together these innovations crucially differentiate CaaC solutions from adopting any single vendor approach unable to match scale, cost, flexibility, or control advantages collectively offered in this multi-provider federation model glued by intelligent optimization for automation and peak value delivery.

4. CAAC BENEFITS AND CHALLENGES

4.1 Detail Benefits in Scalability, Cost Savings, Spurring Creativity and Experimentation

Scalability

By pooling together excess capacity across multitudes of data centers, clouds and render farms, CaaC creates abundance versus scarcity traditionally faced by owned infrastructure always hitting ceilings. The collective buffer allows seamless scaling to match volatile workloads instead of saturation-led performance drops or delays typifying in-house hardware limitations. Small teams can thus punch above their weight running enterprise-grade workloads by sharing economies of massive centralized systems accessible selectively on demand. Streamlined acquisition reduces procurement vortex issues freeing up delivery timelines and innovations possibilities.

Cost Savings

Cost savings represent the flagship benefit proposition from shifting compute to CaaC models, with reported optimal savings ranging from 60% up to 90% over list prices from major cloud providers. By spreading workloads dynamically across spot instances, reserved capacity, on-demand infrastructure, multi-cloud migration etc., CaaC unlocks best value through availability abundancy. Granular analysis helps trim waste by rightsizing usage and spending to business hours. Aggregate discounts surpass piecemeal commits to individual platforms. Over time DNA-level telemetry guides optimizations automatically identifying savings from eliminating even tiny resource fragmentation adding up across scale.

Spurring Creativity

The on-tap accessibility to leading edge infrastructure like GPUs clusters, FGPAs, quantum etc combined with more budget flexibility from cost savings allows organizations to get creative accelerating ideas not possible previously. Data teams can run experiments faster with more iterations across hardware types, geographies, and cloud interoperability instead of just local limitations. Engineers can easily prototype leveraging global capacity on demand prioritizing innovation velocity over perfection. Startups stand to benefit tremendously by leveling the playing field through sharing excess resources inexpensively rather than making capital commitments early on still validating product-market fit.

In essence by tackling the twin constraints of capacities and costs hampering cloud consumption at scale, CaaC introduces new degrees of freedom spurring creativity, productivity, and outcomes - achieving more collectively by aggregating fragmented infrastructure dotted across isolated domains. Democratization beyond consolidation liberates innovation bottlenecks.

4.2 Discuss Main Limitation Around in-house Infrastructure Management

While public cloud eliminated large upfront Capex for owning depreciating hardware assets over time, significant challenges remain around managing in-house infrastructure at scale leading to sub-optimal utilization, lack of visibility into capacity planning and expenses bloat due to overprovisioning - directly hampering innovation agility and budgets. CaaC looks to overcome these major pitfalls.

Firstly, capacity ceiling constraints get frequently triggered given workload variability against fixed on-premise resources unable to serve demand spikes gracefully or require failover buffers adding redundancy overhead. Procurement delays also worsen matters during surges leading to denied requests and delays even for mission critical needs until additional infrastructure gets commissioned manually - severely slowing down product and innovations. Such rigid ceilings on owned hardware significantly drag business velocity and responsiveness.

Secondly, poor visibility into utilization rates across fragmented on-premise systems leads to significant overprovisioning and resulting expense inefficiencies as capacity planning relies more on guesswork. Actual runtime consumption patterns remain opaque. The inability to automatically optimize and rescale capacities based on fine-grained usage analytics hence compounds inefficiencies over time. Resource sharing across internal teams also remains cumbersome to administer manually.

Thirdly, the lack of usage transparency also limits granular accountability into costs incurred by business units towards infrastructure consumed, making cost optimization harder. Cross-charging complexities disincentivize teams optimizing expenses through disciplined infrastructure sharing models internally. Tight coupling between applications and non-fungible capacity blocks encourage resource silos and hoarding preventing economies of scale.

By federating heterogeneous infrastructure pools onto unified self-service access mediums governed by real-time usage analytics, CaaC platforms overcome these critical limitations hampering innovation velocity and budget efficiencies using captive infrastructure alone. The aggregated capacity buffer managed intelligently unlocks speed, scalability, and cost-effectiveness unmatchable otherwise.

5. INDUSTRY IMPACT

5.1 Vertical Examples in Healthcare, Finance, Media, and Retail

Healthcare

CaaC allows healthcare entities to leverage scalable computing for advancing research insights and patient outcomes through data-intensive applications. For instance pharmaceutical firm AstraZeneca leverages cloud container technologies tapping into bio compute capacity from global CROs for running drug simulations at scale. Mayo Clinic accelerated pediatric tumor analysis 30x faster using cloud credits from CaaC aggregators by leveraging federated GPU pools. Similarly pay-as-you-go infrastructure assists cash-strapped hospitals minimize unused servers. Startup Rapid FT gained flexibility prototyping genomic sequencing workflows at 1/10th cost before raising venture capital. CaaC hence fuels healthcare modernization.

Financial Services

Banks and insurers benefit significantly from CaaC models to enable real-time fraud analysis, conduct micro segmented customer analytics or building trading insights through alternative datasets - all requiring access to versatile computing capacity on-tap without long procurement cycles. Spanish bank BBVA could test algorithms across global stock exchanges through cloud credits saving $12M while also lowering model risk. Financial risk modeling and regulatory compliance workloads also utilize transient servers for secure data sharing. Thus, CaaC drives financial services transformation.

Media & Entertainment

Media production houses and gaming studios stand to benefit tremendously from CaaC solutions by accessing cloud-scale rendering capacity from aggregators for graphics, CGI and animation workloads in a flexible consumption model. Disney's recent shift to leveraging spot compute capacity helped produce blockbuster Shang-Chi film 30% cheaper using pay-as-you-go infrastructure. Gojek's gaming arm ACTOR taps spot capacity via Platform One across Southeast Asia saving 60% over regular cloud rates through transparent consumption model. CaaC is revolutionizing media creation economics thereby.

Retail & CPG

Retailers and consumer brands utilize elastic compute for personalizing shopping experiences via recommendation engines; micro segmenting customers for targeted promotions; optimizing supply chains via sensor data analytics; speeding up loyalty programs; enabling faster feature development and experimentation via shadow IT capabilities unlocked by CaaC access. Walmart observed 4X faster clickstream analytics response times after migrating to Azure spot instances saving over $4 million annually in cloud costs. Chip maker AMD uses Right Scale to architect multi-cloud efficiencies. Thus, CaaC empowers retail innovations significantly.

5.2 How CaaC Unlocks AI/ML Application and Data Intensive Workloads

AI/ML Advancements

A core benefit expected from CaaC adoption is dramatically accelerating AI and machine learning capabilities for organizations by removing infrastructure barriers slowing down development and deployment currently. Model building, training and inference workloads demand extensive compute resources to iterate experiments dealing with vast datasets. But long provisioning delays for such expensive GPU/TPU farms coupled with high cloud costs discourage experimentation velocity - severely hampering mainstream AI adoption.

By making heterogenous hardware ecosystems consisting of specialized silicon and accelerators accessible instantly via unified APIs instead of fragmented capital deployment, CaaC introduces new degrees of flexibility facilitating AI innovation. Teams can run iterative experiments across geographically distributed infrastructure not earlier feasible to identify optimal configurations before committing long-term. Sharing economies allow smaller players to tap into cutting-edge capacity by paying only for consumption instead of full-time cluster subscriptions. Automated pipelines scale seamlessly to manage explosive growth in data and model complexity without delays. Overall CaaC vastly accelerates organizations leveraging predictive insights and automation by removing constraints around capital, capacity ceilings and skills bottlenecking AI progress currently. Democratized access will fuel creativity in applying intelligence across vertical domains much like electricity sparked wider dispersion of mechanical engineering innovations previously gated to select players.

Data-driven Innovation

Similarly CaaC also empowers enterprises break data silos by making storage and distributed processing universally accessible to teams without upfront procurement signoffs. Batch, stream and graph data workloads can tap into abundant capacity to derive cross-functional intelligence - sharing data openly as utility versus fragmented warehouses currently. Product developers can thus conduct fine-grained segment analysis, engineers monitor IoT sensor streams for predictive maintenance, data scientists mine real-world evidence datasets at bigger scales while executives analyze customer journeys holistically by converged data pipelines across the computing fabric stitched by CaaC - all catalyzing data-driven innovation.

5.3 Competitive Advantage and Innovation Acceleration

Competitive Advantage

By scaling affordable excess to versatile computing on-tap, CaaC introduces new competitive dynamics beyond what exists in the capital intensive models of procuring physical hardware traditionally. The versatility and flexibility attributes allow fast experimentation, rapid prototyping, and deployment not earlier feasible - enabling companies to seize first-mover advantages with emerging technologies before rivals burdened by legacy infra commitment cycles.

Decoupling priorities between business innovation needs versus underlying infrastructure supply strategies also creates strategic optionality for enterprises. Software teams get more independence on tools adoption without hardware gatekeeping barriers that previously throttled release velocities if procurement guidelines got breached. Younger firms face less entry barriers leveraging the shared capacity from CaaC that incumbent players accumulated with patient capital over decades. Such democratization forces the pace of innovation faster. Studies have shown technology early adopters consistently outperforming competition across sectors as innovations reaching maturity get commoditized for laggards to imitate. By accelerating tech innovation leverage through shared infrastructure, CaaC promises to reshape competitive leadership dynamics with access to bleeding edge capacity on tap determining firm trajectories. Incumbents weighed by legacy setups face stiffer innovation imperatives to prevent disruption by trusting CaaC strategies alone.

Innovation Velocity

From an agility perspective, CaaC solutions promise 10x faster value realization over traditional models building on-premise infrastructure. Click-of-button access removes delays associated with capital signoffs, vendor selection, hardware procurement, capacity planning, skills development etc. before deriving business value. By pooling heterogeneous infrastructure fragmented currently, capacity utilization also improves holistically beyond localized pockets constrained individually.

CaaC makes compute power as reliable and ubiquitous as electricity access, fueling innovations multiplier effect by removing basic impediments stifling progress repeatedly. Platformization promises to trigger an explosion in computing use cases from AI to IoT much like Android sparked the app economy by standardizing mobile app infrastructure for masses instead of custom hardware variations fragmenting reach earlier. Democratization beyond consolidation thus paces innovation velocity manifold.

6. EMERGING CAAC INNOVATIONS

6.1 Edge Computing, Quantum Computing, Serverless Computing

Edge Computing

Edge computing refers to processing workload locally on distributed hardware physically closer to source instead of routing back haul traffic to centralized clouds far away. Use cases like IoT, AR/VR, autonomous vehicles etc demand real-time response times by reduced latency. Small edge micro data centers embedded across 5G towers, retail spaces, manufacturing floors etc. assist responsiveness. While public cloud providers now offer managed stacks, CaaC introduces new opportunities by harnessing fragmented edge capacity ubiquitously. Retailers can analytics POS data onsite for personalized promotions using shared edge infrastructure from CaaC providers on OpEx model instead of upfront CDN build outs awaiting scale. Thus, CaaC helps proliferate edge ecosystems faster.

Quantum Computing

Quantum promises exponential leap in compute power by leveraging quantum bits and superposition over classical binary constraints to tackle complex optimization constraints efficiently - with early success in areas like chemical simulations, financial risk modelling etc. But high costs of bleeding edge hardware coupled with scarcity of skilled developers has gated progress. However by pooling access across early commercial systems from IBM, Rigetti, IonQ etc., CaaC marketplaces lower entry barriers for enterprises curious to prototype quantum workflows. Teams get to experiment on real quantum systems at modest hourly rates instead of full-stack capital commitments early on. CaaC can hence assist faster quantum democratization.

Serverless Computing

By auto-scaling capacity precisely to workload needs and deprovisioning dynamically, serverless models aim for optimum utilization eliminating overprovisioning waste typical for cloud servers with pre-allocated CPUs/memory. Event triggers invoke temporary containers only for peak periods unlike always-on resources. While AWS Lambda, Google Cloud Functions spearhead FaaS adoption, CaaC introduces new dimensions by federating serverless capacity multi-cloud encompassing niche providers like Spotinst while also easing integration with legacy systems. Teams hence gain ability to adopt serverless grants gradually without vendor lock-ins. CaaC thus unlocks next frontier for serverless innovation beyond early adopters.

6.2 Reduced Latency, Solving Impossibly Complex Problems, Simplified Access

Reduced Latency

Ultra-low latency represents an area of active innovation across domains from industrial IoT, virtual spaces in metaverse to financial trading platforms where microseconds matter. However, achieving consistent response times below 50ms thresholds for remote processing remains challenging and expensive. By harnessing decentralized capacity closer to usage source through edge computing advancements, CaaC unlocks new possibilities. Pools can span colocation spaces, telco stacks, retail store compute and smaller cloudlets — all sharing excess capacity via CaaC platforms. Instead of round trips to distant clouds exacerbating delays despite broadband, hybrid models allow selective workload placement minimising lags for priority use cases. Smart routing algorithms also load balance dynamically for predictable outcomes. CaaC hence drives the next threshold in latency reductions materially.

Solving Complex Problems

Certain classes of optimization constraints become solvable efficiently only at cloud scale leveraging massively parallel computing. Pharmacogenomics workflows analyzing gene mutations influencing medicine efficacy demand processing billions of genome datapoints concurrently. Similarly predicting climate risks require modelling years of granular environmental sensor data globally to guide policy interventions. Operating fixed capacity datacenters cannot tackle such problems economically. By simplifying access to unlimited compute, CaaC introduces more degrees of freedom for enterprises addressing such complex challenges using scale while paying only for flexible consumption. Breakthroughs hitherto constrained by resources availability open - be it personalized healthcare, clean energy or smart mobility solutions.

Simplified Access

While hyperscale cloud eliminated large capital investments for upgrading infrastructure over time, complexity persists around managing hybrid landscapes spanning legacy systems and niche workloads not yet cloud ready. CaaC introduces unified self-service access layers above existing infrastructure much like electricity adapters provide ubiquitous power connectivity without appliances needing to know underlying generator mechanisms. Everything connects directly through common APIs and SDKs abstracting underlying delivery procedures. Teams can hence migrate gradually without risky big bang projects. Such simplified access can accelerate emerging technology adoption from AI to Blockchain that promise enterprises value yet suffer integration bottlenecks still. CaaC empowers through flexibility.

7. CONCLUSIONS

7.1 Summary of CaaC Transformative Potential

Concluding section summarizing the transformative potential of Compute as a Commodity (CaaC) evolution promising to reshape infrastructure innovation and digital capabilities over the coming decade much like electrification and mass electromechanical automation transformed entire economies in the previous century:

In closing, Compute delivered on-demand as managed capacity pools accessible via unified self-service constructs represents the next leap forward in infrastructure advancement from electricity, telephony and Internet eras that fueled socioeconomic progress. Much like aggregated energy generation pools continued to expand output capacities over time, growing exponentially versus individual microgrids constraining growth, CaaC opens up new trajectories harnessing the infinitely expanding compute capacities innovated collectively - all made consumable on flexible, pay-per-use basis dynamically for varied needs and budgets.

By providing versatile compute ready for consumption at click-of-a-button akin to utility services today, CaaC solutions promise enterprises capabilities to experiment, iterate, analyze more ambitiously; leverage emerging innovations in ways not conceivable amidst previous infrastructure barriers stifling progress. The coming proliferation of AI, IoT, metaverse and quantum applications can stand on the shoulders of cloud-scale flexible capacity pools instead of reinventing finite wheels locally. Shared abundance sustains innovation better than ownership scarcity models still widespread.

Just as village electrification programs transformed rural economies by removing energy access as an inhibiting factor for grassroots development, CaaC aims to provide a level playing field for digital innovation across entities - minimizing divides allowing smaller teams also to punch above their weight without climate commitments restraining experimentation. Democratization of access holds the key to accelerate mainstream technology assimilation.

With exponential data generation, use cases crossing domains and geographies seeking intelligence from connected insights in real-time and machine learning promising new possibilities, compute holds the promise to advance humanity provided the fundamental capacity barriers get addressed with reliable infrastructure reach. CaaC represents the building blocks for this inclusive, resilient and creative digital future.

7.2 Future Outlook on Growth and Mainstream Adoption

Future Outlook

According to recent research from Markets and Markets, the global CaaC solutions market is estimated to grow multi-fold from $10.3 billion valuations in 2022 to over $106 billion by 2030 as pace of digital transformation and cloud adoption continues accelerating across sectors. Let us examine the key trends fueling massive mainstream adoption.

Firstly, as more enterprise workloads involving data analytics, IoT mediation, AI/ML model building go cloud native, demand for instantly scalable capacity keeps rising beyond what fixed on-premise infrastructure can sustain without procurement delays or cost overruns from overprovisioning. CaaC provides the flexible buffer augmenting capacity available on-tap elastically. Growth mirrors public cloud explosion in recent years - but with far more cost savings. Secondly, spot pricing mechanisms couple seamlessly with base capacity to optimize workload placement across regions, availability zones and instance family types based on urgency, budget caps etc. Expanding ecosystem of regional providers like Bunny.net, NXTCLOUD offer alternatives to hyperscalers improving redundancy and value. Snowflake-style separated storage from compute too minimizes unused capacity. Granular metering assists optimization.

Thirdly, cloud interoperability frameworks like CAST, decimal allow easy portability across heterogeneous infrastructure without vendor lock-ins. Open standards continually improve interchangeability spanning legacy on-premise systems and modern microservices, functions. Abstraction hides low-level complexities while still giving control over configurations, security policies etc. Together these ensures CaaC solutions offer the best of both worlds flexibility - spanning bleeding edge capacity for innovation needs as well as maximizing utilization of existing infrastructure assets using intelligent multi-cloud orchestration. Mainstream adoption uptick seems inevitable as enterprises seek to accelerate digital initiatives balancing costs and agility via flexible capacity.

7.3 Broader Societal Impacts on Democratizing Compute Power

Broader Impact

The ubiquitous access to on-demand computing unlocked by CaaC has the potential to reshape societies by tackling one of the pressing digital age divides - the asymmetric access to versatile computing capacity slowing down equitable progress. Much like early electrification brought modern amenities by increasing energy availability multi-fold, CaaC aims to proliferate breakthrough innovations to grassroots instead of concentration with elites hitherto protected by privileged incubation environments locally. Democratization beyond consolidation can pace wider assimilation of emerging technologies. For instance, in healthcare, smaller clinics and regional hospitals unable to invest heavily into precision medicine capacities like gene sequencing infrastructure individually can benefit tremendously via shared CaaC pools. Enriching molecular datasets globally can accelerate discovery of pattern correlations between genetic markers and disease progression. Expanding drug trials diversity also improves efficacy confidence levels currently biased by selective population sampling. All such incremental boosts summate to improve community health outcomes collectively over time by increasing both reach and quality.

Similarly demystifying access to bleeding edge hardware like quantum/neuromorphic computing now gated via high investment barriers temporary can encourage more solution architects to engage with novel architectures beyond Silicon Valley early adopters alone. Student researchers, citizen scientists and SMBs also stand to gain experimenting hands-on at modest hourly access rates instead of full subscriptions unviable individually. Shared prosperity has compounded technology progress historically evident from open source movements lowering barriers across software ecosystem. CaaC promises this for infrastructure abundance. Thus, Compute delivered as managed utility promises to boost digital inclusion tremendously. Mainstream adoption beyond pilots can transform solution outcomes holistically much like off-grid solar and mobile telephony penetration did more to uplift underprivileged communities structurally by adapting innovations for access constraints upfront. Shared abundance sustains human progress equitably by collective upliftment.

REFERENCES

[1] A new era of computing is coming. How can we make sure it is sustainable? (2022, May 20). World Economic Forum. https://www.weforum.org/agenda/2018/09/end-of-an-era-what-computing-will-look-like-after-moores-law/
[2] AI for All: The Power of Democratization and Collaboration. (2024, May 9).
[3] Democratize the Power of Computing: IBM Symphony Is Now Available on IBM Cloud - IBM Blog. (2021, September 30). IBM Blog. https://www.ibm.com/blog/announcement/democratize-the-power-of-computing/
[4] George, A. S. (2024a). Humanoid Robots as Poultry Partners: Enhancing Welfare Through Collaboration on the Farm. puirj.com. https://doi.org/10.5281/zenodo.10850069
[5] Exploring the Benefits of Cloud Computing for Businesses - ValueLabs. (2024a, March 15). ValueLabs - Digital Technology Solutions | IT Consulting Services.
[6] Exploring the Benefits of Cloud Computing for Businesses - ValueLabs. (2024b, March 15). ValueLabs - Digital Technology Solutions | IT Consulting Services.
[7] George, A. S. (2023b). The Potential of Generative AI to Reform Graduate Education. puirj.com. https://doi.org/10.5281/zenodo.10421475
[8] Flexera Releases 2021 State of the Cloud Report. (n.d.). https://www.flexera.com/about-us/press-center/flexera-releases-2021-state-of-the-cloud-report
[9] George, A. S. (2023a). Securing the Future of Finance: How AI, Blockchain, and Machine Learning Safeguard Emerging Neobank Technology Against Evolving Cyber Threats. puirp.com. https://doi.org/10.5281/zenodo.10001735
[10] Goddard, W. (2023, February 3). The Evolution of Cloud Computing – Where's It Going Next? ITChronicles.

[11] Gupta, S. (2017, May 11). The democratization of the supercomputer. Mint. https://www.livemint.com/Industry/GgQpL6eYfTXBmZlkSNzbdN/The-democratization-of-the-supercomputers.html
[12] George, A. S., George, A. S. H., & Baskar, T. (2023). Exploring the Potential of Prompt Engineering in India: A Study on the Future of AI-Driven Job Market and the Role of Higher Education. puirp.com. https://doi.org/10.5281/zenodo.10121998
[13] High-performance computing benefits signal- and data processing in aerospace and defense applications. (2011, July 1). Military Aerospace. https://www.militaryaerospace.com/computers/article/16717088/high-performance-computing-benefits-signal-and-data-processing-in-aerospace-and-defense-applications
[14] How has Bare Metal Cloud democratized High Performance Computing. (n.d.). Cherry Servers. https://www.cherryservers.com/blog/bare-metal-cloud-for-high-performance-computing
[15] George, A. S. (2024b). Weaponizing WhatsApp: Organized Propaganda and the Erosion of Democratic Discourse in India. pumrj.com. https://doi.org/10.5281/zenodo.11178130
[16] Nadeem, R., & Nadeem, R. (2024, April 14). 3. Concerns about democracy in the digital age. Pew Research Center. https://www.pewresearch.org/internet/2020/02/21/concerns-about-democracy-in-the-digital-age/
[17] George, A. S., George, A. S. H., & Baskar, T. (2024). Artificial Intelligence and the Future of Healthcare: Emerging Jobs and Skills in 2035. pumrj.com. https://doi.org/10.5281/zenodo.11176554
[18] Technology is helping SMBs to become a global force — Executive Digest — Presented by FedEx. (n.d.). https://www.wsj.com/ad/article/execdigest-technology
[19] The Barbell Effect of Machine Learning | Nick Beim. (2016, June 3). https://www.nickbeim.com/2016/06/03/the-barbell-effect-of-machine-learning/
[20] The Evolution of Cloud Computing: A Decade in Review. (n.d.). https://www.cloudpwr.com/news/insights/the-evolution-of-cloud-computing-a-decade-in-review/
[21] The Evolution of Cloud Computing: A Foundational Shift. (2024, May 29). AscentCore. https://ascentcore.com/2024/05/29/the-evolution-of-cloud-computing-a-foundational-shift/
[22] what-is-cloud-computing. (n.d.). [Video]. Amazon Web Services, Inc. https://aws.amazon.com/what-is-cloud-computing/
[23] Workshop Session. (n.d.). https://sc23.supercomputing.org/proceedings/workshops/workshop_sessions.html
[24] Yen, L. (2023, November 10). Cloud vs. On-Premises: Pros, Cons, and Use Cases. Datamation. https://www.datamation.com/cloud/cloud-vs-on-premises-pros-cons-and-use-cases/