OROCK TECHNOLOGIES

A Practical Guide to Mission-Critical Cloud Computing

Table of Contents

Forward

Different Cloud Journeys with One Destination

Whether you have crisscrossed the globe or are just starting your cloud journey, this guide aims to help organizations understand how they can be successful managing their cloud initiatives. Hopefully, you will learn about powering up your cloud environment, the many flavors of cloud computing, the different ways to mitigate risk, how to best feed the cloud and, most importantly, what we believe 'right' looks like.

Quick question. Are you currently researching how to migrate to the cloud, or does your organization already have a cloud environment up and running? Regardless of where you are in your journey, it is important to understand that each organization will take a different route to a cloud future. Much of that will depend on business needs and applications. Before you reach for the clouds, understanding why you are moving to new or different cloud technologies and having confidence in the solution you choose is critical. While there's no one-size-fits-all approach, most organizations do share one common goal: they are aiming for major gains in performance that can provide the competitive edge they need. In short, applying the right

cloud technologies can drive greater capacity, scalability, agility and security. Organizations that run on the cloud quickly discover new avenues and approaches to innovation while realizing cost savings in the process.

Accomplishing your 'Mission' with Cloud

The relevance to organizations with mission-critical applications is particularly timely given the demand for world-class performance and interoperability. We define mission-critical applications as any workload that has absolute requirements on predictability (determinism), performance and security. If your workload requires any or all of these dependencies and the lack of meeting any of them can directly impact your business, then partnering with the right cloud service provider is essential. Finding the right partner with matching cloud technologies to meet your IT requirements can help to enhance every level of your organization.

When researching where and how to leverage a cloud computing environment, careful due diligence is a must. Ask yourself, am I moving my applications and workloads to a proprietary cloud, having to refactor applications and sacrifice open source economies, security and technology lock-in? Will my applications and workloads be deployed on industry-leading next-generation hardware and software? Can my cloud service provider offer a range of cloud computing options for workloads ranging from non-mission critical to mission-critical? Will I be able to work with cloud support personnel that can help me make the right technology choices and help get my cloud environment operational?

As you continue reading, this guide will offer recommendations on what to ask and what to look for in a cloud service provider and how to know you are making the best decisions in selecting a cloud technology and a service provider partner that is right for you.

A Foundation for Cloud Computing

Move to the Cloud with People Who Know How

Before you embark on your journey to the cloud, you should consider a cloud partner willing to understand your unique requirements, challenges and goals. Consider a cloud provider who cares as much about your customers and your success as you do and choose one that understands foundational issues, such as security, OPEX vs CAPEX, bare metal vs virtualization, general workloads vs mission-critical ones, proprietary vs open source, active support vs email support and more. Their expertise in these areas will vastly increase your speed of relevance. Once you decide leveraging cloud CPUs or storage is the direction you would like to pursue, you will want guidance from a vendor with first-hand knowledge on how to build mission-critical cloud infrastructure and who will work closely with you to advance your cloud-first or cloud-complementary journey.

Building a Cloud for Mission-Critical and High-Performance Applications

Whether looking at a "green field" deployment or redesigning your existing systems, infrastructure, methods, and/or processes, we recommend you focus on these foundational principles: security, performance and predictability (determinism). By understanding these key principles and how they impact your business and technology requirements, this guide will assist you with determining which cloud technology is right for your organization and which cloud services provider can best support your requirements.

Let's start with defining two types of cloud resource sharing technologies. One method is designed on "Time Sharing Technology," more commonly known as multi-tenant cloud and the other is based on "Resource Partitioning Technology" multi-tenant environment with next-generation performance and security technologies.

In both cloud resource sharing models, it is important to understand how they work and why there is a difference in how applications and workloads can be impacted. Understanding these differences can help an organization choose on which model to run their mission-critical workloads. While both cloud resource models serve as excellent options, we recommend working with a cloud service provider to help determine which is best suited to meet your requirements. In the following sections, each of these options are further explained.

Time Sharing Technology - Time Division CPU Access (TDCA)

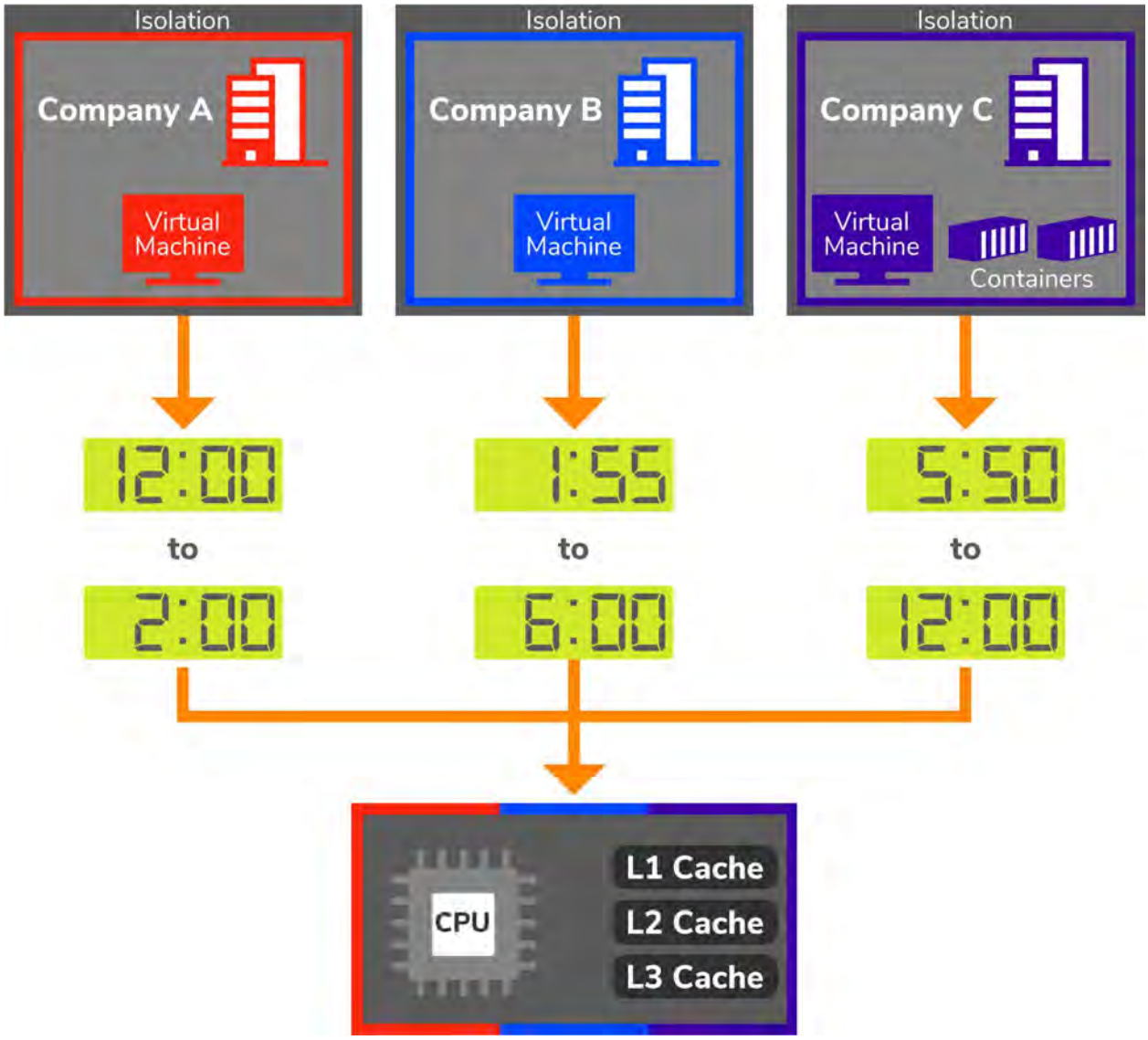
The public cloud was built on this model and it has greatly changed the way organizations view their technology landscapes. By providing organizations with an alternative to deploying equipment in their own datacenters, they can now use virtual machine resources that can be consumed as needed, providing vastly greater flexibility and agility. It is important to understand how this technology is delivered.

Time Division CPU Access (TDCA) is a method of scheduling a CPU where every instance—virtual machine (VM) or container—gets an amount, or slice, of time of the underlying CPU. (see Diagram 1) This is the existing model for how resources are shared in a multi-tenant cloud and legacy VM infrastructure. At the most basic of layer, the CPU uses a time splicing methodology to share CPU resources across applications. On the surface, this appears to provide dedicated CPU time to your workload. The reality is you are paying for seconds on a clock. This works fine when you are not using your CPU. However, for workloads that are using CPU, performance, security and determinism suffer greatly.

The next step most organizations take is to pay for a larger slice of time. You are very likely overbuying from hyperscale cloud providers to get the performance you need. Current technology deployed at all hyperscale cloud companies are based on the concept of time-sharing. It does not matter whether you are using a VM or a container; they are both based on time-sharing. Even when you buy dedicated instances, you still end up sharing the very valuable CPU cache with other customers. Time-sharing results in poor performance and poor security. TDCA affects virtual machines, containers and Kubernetes because all these technologies are up the stack from the underlying time-sharing system.

By running applications and workloads that require greater performance and predictability, there can be drawbacks to this method. However, it does provide for a more economical approach for workloads that can tolerate this kind of resource sharing.

Diagram 1: **Time Division CPU Access (TDCA)**



Resource Partitioning Technology - Core Division CPU Access (CDCA)

As more organizations seek to move additional advanced and mission-critical applications and workloads to the cloud, new technologies are emerging that provide multi-tenant environments and resources to be logically and physically separated from one another. This evolution promises to drastically change the virtual environments' performance, security, and predictability, allowing organizations to now consider cloud environments as an option for running critical workloads.

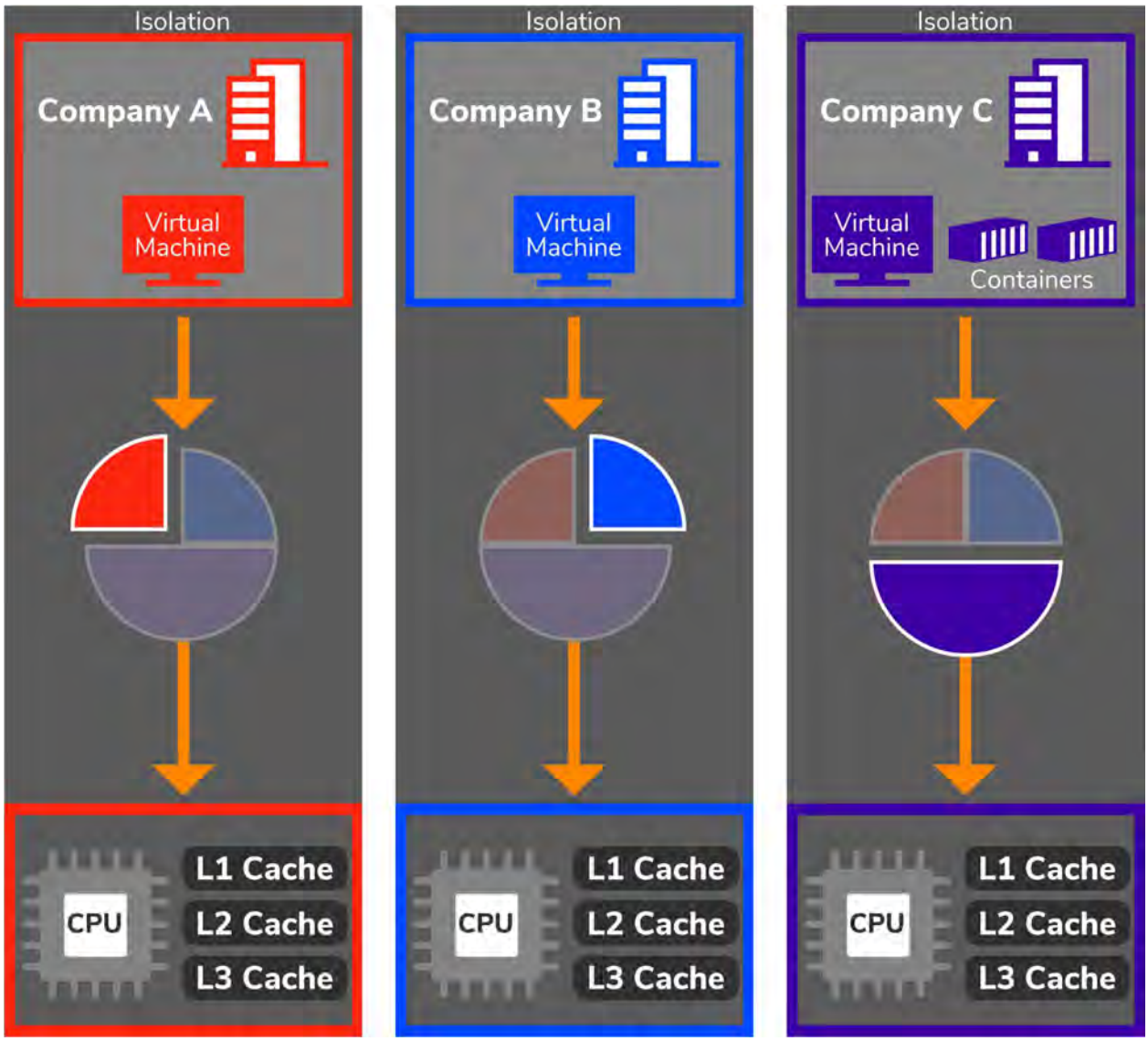
These new cloud technology breakthroughs will provide new opportunities for organizations to derive cloud benefits but at security and performance levels typically found in dedicated on-premises server environments. What are the key elements for organizations to know?

Core Division CPU Access (CDCA) is a new method of scheduling work on the CPU. CDCA technology can break the processor and associated cache down into physical slices. This means that instead of getting seconds on a clock, users buy the number of CPUs they need and get access to them all the time. (see Diagram 2) So instead of the cloud service provider (CSP) over allocating processor and cache resources by "time splicing" these resources, your organization no longer has to share these resources with everyone in that overallocated resource pool. Without CDCA, CSPs are hedging their bets that these companies will never overexert that shared resource all at the same time.

For many applications and workloads that can tolerate performance variations and fluctuations due to time sharing (or TCDA), it makes perfect economic sense to leverage these options from a CSP. However, in cases where security, performance and predictability are paramount, the introduction of CDCA is a significant advancement in cloud computing. For CSPs that can deliver this advance cloud computing option, organizations can now benefit from greater options and more flexibility.

Today, the days of being limited to running mission-critical applications only on bare metal environments has changed. The benefits of using CDCA technology center on security, performance and predictability (determinism), and with it, you get truly dedicated performance at the edge, on-premises or in the cloud whenever this technology is deployed.

Diagram 2: **Core Division CPU Access**



So why is security, performance and predictability (or determinism) so important? Here's why.

Security – Information Security (INFOSEC) continues to be an increasingly hot topic in the minds of everyone from engineering to the C-suite. In multi-tenant environments, security is even more mission-critical as basic layers of your perimeter are shared with other organizations. Depending on your application, a security breach could be the difference between your business thriving or shutting down. For organizations responsible for defense applications, it could be the difference between people living or dying. Security must be a primary driver in selecting a suite of cloud technologies. Security is packed in layers and if the most basic layers are not secure, then the mitigations up the stack become more costly in terms of performance and actual monies spent. So knowing what security measures are provided from the lowest levels including the CPU and up the stack would be important factors when selecting a cloud platform.

Performance – Cost and performance are directly correlated properties in cloud computing. If a system

performs poorly in the cloud, the course of action is typically to spend more money on it. When trying to build out a budget to move an application to the cloud, the performance of the infrastructure is of the utmost importance. To keep budgets from being maxed out by cloud infrastructure costs, performance must be a measurement. For example, the length of time it takes for your app to process a web request or a database transaction can be how performance is measured—otherwise known as latency. Measuring latency across your environment is important because it all adds up and can impact your cloud costs.

Predictability (or Determinism) – General purpose cloud architecture is built on the simple premise of time slicing as defined above. When a customer requests an instance with 8 CPU cores, that means 8 CPU cores worth of time is on the processor. The time-sharing mechanism is why predictability (or determinism) is an important factor to consider when building or buying cloud technology. This is a key foundational property to understand when moving applications to the cloud as predictability and will determine how long an application takes to process workloads. Variations of performance over time is referred to as "jitter." Performance measured over time is important to understand as it can have a significant impact on processing application request which for critical applications should not vary over time. Predictability (or determinism) can now be delivered in multi-tenant cloud environments thanks to new CDCA technologies as defined above. This technology can help ensure your organization is getting the performance it is paying for down to the microsecond.

Cloud Basics

Approximately 40 years ago, the manufacturing industry embraced automation and incorporated robots to their assembly lines to become smarter, faster and more competitive. Think of the cloud as a giant, interconnected global robot. When you purchase a compute instance in the cloud, you are buying into the notion of this automation that creates the instance, wires it up to the network and connects the hard drives. Instead of paying a human to create each of your instances, you can instead pay a robot to carry out these tasks and repetitive steps. In plain English, the cloud is powered by various Application Programming Interfaces (APIs) which are messengers that relay commands back and forth between different applications and cloud services. The adoption and expansion of APIs has led to exciting advancements leading to application modernization and the use of cloud services.

Cloud Flavors

While many people think the cloud is a place, the cloud is in fact a model for computing characterized by different cloud flavors, with distinctions based on size and location. The principal elements to consider include whether you own the infrastructure or not, where the cloud is located relative to your network and what is the approximation of its size. Cloud computing follows a basic model that uses APIs to create, update, configure and manage the actual application. The API replaces what you previously paid a physical resource to do.

Technical Debt

Technical debt can be quantified by long-term maintenance costs of an application or service, and like financial debt, it accumulates interest over time. These debts can become so burdensome that innovation is impossible. Debts can grow to occupy 100% (or more) of the budget to maintain and operate applications or services. At some point, technical debt must be

paid. The cloud has helped organizations address technical debt by allowing them to adopt and build new modern applications or refactor existing legacy applications to leverage new technologies that cloud computing has introduced. Simply moving an application to the cloud does not relieve its technical debt. Therefore, it is important to develop a plan and then choose a cloud provider that cares about you reaching your goals and can help you along your journey.

Risk vs Reward

Moving to the cloud can introduce risks in the areas of security and compliance. Companies considering upgrading their infrastructure and moving workloads to the cloud should work with a provider knowledgeable with regulatory compliance and data security and privacy matters. When moving sensitive data such as payment card information (PCI data), protected health information (PHI data) or personally identifiable information (PII data), knowing where and how your data is stored and accessible is critical. Security threats abound and unexpected hidden costs can challenge any organization in ways not previously considered.

Leveraging cloud technology can help organizations lower infrastructure and management costs associated with personnel, data storage and real estate, and enable businesses to ramp up capacity on-demand during peak periods and dial capacity down when it is not needed. Many organizations cannot realize these rewards alone. With the right cloud service provider providing security and compliance standards, many organizations can benefit by partnering with a cloud service provider that prioritizes security, compliance, affordability and customer service.

Cloud Lock

Cloud lock is when an application or service is hard coded or irrevocably tied to a cloud vendor's proprietary technology. Some vendors go to great lengths to make it enormously challenging or cost prohibitive to move from one cloud to another. To prevent cloud lock-in, organizations should seek a cloud solution that is fully transparent and based on open source technologies. Application owners within an organization should be empowered with options as a mitigation method to avoid cloud lock. Applications and services that are developed with open source cloud technology can be moved seamlessly between public and private cloud providers without costly refactoring.

Cost Control

While public cloud infrastructure is designed to be easy, not all cases save you money. Operating an entire data center for a small company typically does not make financial sense, so an all-in public cloud strategy may be applicable. For large organizations, the benefits of a public cloud include scale, agility, flexibility and ease of consumption. Yet not all applications require scale, agility or ease of consumption. Moving the applications of a large organization to the public cloud may indeed be cost prohibitive. Most IT leaders across every industry agree that not all applications and data should reside in a public cloud environment and that a balanced hybrid IT strategy yields greater results. Without a balanced approach and cost control strategy, many organization's public cloud usage could overrun budgets and leave no money to continue the curve of innovation or the reduction of technical debt. To maintain cost controls and avoid surprise charges, work with a cloud service provider that is transparent in their pricing and fees and who won't surprise you with hidden fees, egress charges and unplanned expenses. For mission-critical applications or workloads, it is essential to select a provider that can fully support these applications and workloads without breaking the bank.

Cloud Modes

Public Cloud

A public cloud is a pool of virtual resources—developed from hardware owned and managed by a third-party company—that is automatically provisioned and allocated among multiple clients through a self-service interface. As such, public clouds are useful for smaller organizations that do not want to manage any infrastructure and typically buy their business applications from leading software vendors. A public cloud can be very effective for organizations who have fluctuating demands on compute and storage resources. For instance, if your application is only at full load from 5:00 p.m. to 8:00 p.m. daily, then a public cloud can help save money by allowing more resources to scale during those peak moments and to spin down for other times of the day, thus saving money for non-peak times. For this to work properly, organizations must leverage cloud native workloads and develop or purchase modern application that can benefit from this scaling capability. Simply put, using a public cloud is like renting a car: you rent a car when you need it and return it when you are done.

Private Cloud — On-Premises or Colocated

Private cloud is cloud infrastructure operated solely for a single organization, whether managed internally or by a third party, and hosted either internally in an organization's own physical data center or externally at a colocation data center. As the name suggests, the private cloud is dedicated to the organization itself. While an on-premises cloud infrastructure seems to contradict a central principle of cloud services, organizations find great IT value in creating virtualized compute and storage environments and benefits by providing internal IT resources the ability to use cloud platform tools to manage resources with the same flexibility and ease found in public cloud deployments. Private clouds are designed to give an organization the same API-driven design as a public cloud provider, but with complete control over the entire system.

Hybrid Cloud

Hybrid cloud is a combination of private infrastructure and public cloud, with software enabling communication between these environments. A hybrid cloud strategy provides organizations with greater flexibility by running workloads between private infrastructure and the public cloud as needs, strategy and costs dictate. Many organizations are moving to a hybrid cloud strategy, but not without careful planning and considerations. Organizations often encounter the requirement of having to mix proprietary technologies into a hybrid model. Issues such as technical incompatibilities, legal limitations and refactoring costs all need to be anticipated and addressed. This is why many organizations are embracing open source cloud technologies like OpenStack and finding this pathway can help address proprietary technology issues, thus saving time, money and effort in standing up a hybrid infrastructure quickly. Additionally, finding a cloud service provider to assist in properly architecting and deploying hybrid cloud infrastructure and applications can make the transition from legacy environments to the cloud much faster. The goal of hybrid cloud is to create a unified, automated, scalable environment that takes advantage of the cloud while still maintaining control over mission-critical applications.

Multicloud

Multicloud is a strategy where an organization leverages two or more cloud computing platforms to perform various tasks. Organizations that do not want to depend on a single cloud provider may choose to use resources from several providers to get the best benefits from each unique service. Multicloud can be challenging if highly skilled talent isn't involved to effectively manage the implementation and application development required for different services and technologies. While

this cloud mode can be effective if managed by the right people and applied to the right use cases, most organizations do not have a need for a multicloud approach.

Edge Computing

Edge computing is a distributed computing model in which computing takes place near the physical location where data is being collected and analyzed rather than on a centralized server or in the cloud. This latest infrastructure often involves sensors being used to collect data and edge servers to securely process that data in real-time at the edge. This configuration typically allows other devices, like laptops and smartphones, to connect to a network to share data and send results to an organization's headquarters or to a public cloud. In a modern IT environment, edge computing is applicable to the Internet of Things (IoT) and an organization's use case plays a big factor in which deployment mode makes the most sense. Trying to fully manage an IoT environment through a full cloud platform is not an optimal way to manage workloads because all that data must traverse the network to where the cloud capability resides, leading to latency in the data and potential data transport (network) costs. That is why edge computing puts compute, network and storage resources in locations where they are being used—away from the center—as opposed to private clouds that are centrally deployed and managed. For organizations with latency sensitive workloads, edge computing makes sense. The growing trend for edge computing capabilities is being driven by many organizations that need to process ever increasing amounts data at or near where that data is being collected.

Beyond the Cloud APIs

No App Left Behind

An organization can have a massive catalog of services and applications to power up. More than simply workstation services, these applications run everything the organization does today. The goal is to transform the methodologies and applications, ensuring that no application or service is left behind. Without a proper strategy, technical debts of the past will continue to grow. Most organizations' program budgets and time are spent on technical debts which stifles innovation. Adopting modern DevSecOps workflows to produce and update these applications will allow an organization to pay down accrued technical debts and enable an organization to grow its knowledge in modern application development, deployment and management. A hybrid cloud solution gives organizations the freedom to innovate and adopt the most modern methods.

Cloud Application Types

Cloud-ready applications are ones that know they are in the cloud yet give you read-only data. Cloud-ready applications are typically a first step on the road to fully autonomous systems. While it is tempting to jump directly to cloud-native applications, it is dependent on where your organization is in terms of cloud skills and knowledge. Building cloud-ready apps can be entry point to the world of cloud computing.

Cloud native applications are like fully autonomous trucks; they are self-aware and drive themselves around with a predetermined path. Cloud native apps will do impressive things like auto-scale their own compute or automatically connect storage resources when they are needed. In general, most advanced applications require a significant amount of skill to write complex applications and properly leverage the cloud APIs that drive them.

Cloud Image Types

There are two common flavors of image types that work in the cloud: virtual machines (VM) or containers such as Docker and Kubernetes. A VM runs an operating system that functions like a separate computer that is constantly running and executing tasks. Containers are a different approach to breaking up your applications and allow you to create applications using smaller microservices. Examples of application microservices within an e-commerce application include a shopping cart service, a catalog service and an ordering service. Using containers is a more modern way to test and deploy applications in the cloud with greater efficiency and at a lower cost than using traditional compute resources.

How to Go Fast

Automated Manufacturing

As with automated manufacturing, automation in cloud computing improves performance, quality and productivity. Today's automotive industry uses an army of high-tech robots to rapidly optimize supply chains and drive greater production levels. While there is some capital investment with robotics, they are less expensive for over the long run. By applying a robotic fleet across their manufacturing process, larger automakers are able to move at a blinding pace, as highly skilled staff are dedicated to the design, coding and continuous improvement of their robot force. Like the automotive industry example, cloud technology can be viewed as a cluster of robots, where robots take care of provisioning compute, network and storage resources instead of having employees manage these cloud tasks. The human is now focused on talking to the robots via an API. If you are moving to cloud technology, you will need to know how to automate the environment, including the role of talking to robots. The aim is to strive for 100% automation where humans insert variables and the APIs perform useful tasks.

Feeding the Cloud

To properly feed a hybrid cloud, your organization should initially decide how to migrate applications to a cloud environment. For an application to be moved to a cloud, an image of the application needs to be created continuously using modern processes. Developing a method for how an organization will deliver and lifecycle manage applications in any type of cloud is a critical step. There are many approaches to feeding a cloud and a litany of buzzwords that surround the topic, with the basic method involving building, scanning, configuring and deploying one or more applications to an organization's chosen cloud mode.

DevSecOps

DevSecOps is an organizational framework and set of processes that provides common tooling for three distinct departments in most IT shops. DevSecOps places the deliverables from the development team, security team, and operations team on a single assembly line. The objective is to implement a workflow that can automatically build, test, scan, deploy and monitor one or more applications. This set of processes works in a flexible, interoperable environment. For example, the development team, security team and operations team would all use git to house their source code. When an application or infrastructure moves from the hands of one team to the next, the handoff is 100% automated and seamless. DevSecOps processes can be leveraged by any organization to produce high-quality code and infrastructure. Leveraging the DevSecOps framework does require organizations to embrace a continuous integration and continuous deployment (CI/CD) process as well as agile development methodologies for success.

Pre-purchase Testing

Would you buy a car before you drove it? Would you buy a house without a walk-through? Probably not. Your software inventory decisions should be no different. Establishing a method for pre-purchase testing will save you money and time in the long run. Pre-purchase testing is a step that can be added into the organizations DevSecOps processes. To enable this method, the organization must create a sandbox where vendors and the IT shop can collaborate in code. The sandbox should mimic the cloud mode in which the organization is operating. When done properly, pre-purchase testing will allow the organization to sort out which vendors provide the cheapest, fastest and easiest-to-learn technologies.

Speed of Relevance

The speed of relevance relates to how fast technology can be churned out and how relevant those products are in the market. To stay at the cutting-edge, you must move fast. Speed of relevance is about how fast an organization can innovate. To stay relevant, one must be able to move at machine speed for everything possible.

Machine Speed

Machine speed is where applications talk to each other without human intervention. Since computers are much faster than humans at doing certain tasks, we can assume a properly designed cloud application will talk and function at machine speed on every facet. While a human can take care of telling these applications how to talk to each other, the actual approval, update, lifecycle, provisioning, deletion, executions and more are entirely automated. A modern DevSecOps process is set up in a manner where everything moves at machine speed.

Looking Ahead

Embarking on your cloud journey can be daunting, especially as applications grow more complex, data becomes more distributed, and compliance standards become stricter. ORock Technologies can help every organization fully understand the cloud landscape and support their successful shift to a cloud-driven future. To learn more about how we can help you, contact Eric Van Arsdall at ORock Technologies, sales@orocktech.com or 571.386.0201.