RackWare

Delivering on the Promise of Cloud

RackWare Solutions | White Paper

www.rackwareinc.com

The Continuum of Virtualization to the Cloud

The Continuum of Virtualization to the Cloud
Enterprises of all sizes engage in virtualization to consolidate and reduce costs, and many organizations are expressing an interest in expanding virtualization to improve automation and operational expenses. Indeed, basic virtualization is just the beginning of abstracting underlying platforms and infrastructure to improve efficiency of operations and services. Organizations are experimenting with building private clouds for speed, extending into a hybrid cloud and perhaps even migration to the public cloud for maximum flexibility, and to optimize capital expense.

Most enterprises look to virtualization efforts as a way to save money. But basic virtualization is just the beginning of abstracting underlying platforms and infrastructure to improve efficiency of operations and services. Virtualization is really an entry point to a datacenter architecture evolution that will lead to both private and public cloud computing. Thus, virtualization should not be seen as just a consolidation or cost-savings project, but rather as the beginning of an overall long term plan that provides an on-ramp to cloud computing.

There are several stages along this evolutionary path. Most organizations have elements of their operations at one stage or another, working to move to the next stage. It's tempting to pursue this continuum focusing only on the immediate next step, and investing in changes that may be tactically fine, but are strategically dead ends. Each transformation from one stage to the next must be planned strategically, rather than reactionary, and more importantly with a view of the entire evolutionary path from basic virtualization to maximum usage of the cloud.

The challenge is to prepare for the cloud while at the same time continuing to optimize ongoing basic virtualization.

Necessary Elements for Success for Virtualization and Cloud Enablement
Workload mobility is a critical element to address optimization of virtualization as well as migration to the cloud. Beyond basic mobility, a set of elasticity capabilities, with intelligence and automation, are required to unleash the true potential of the cloud. While mobility is fundamental to achieve virtualized and cloud operational state, it's really the elasticity components that bring the largest ROI.

In the datacenter virtualization continuum, most IT shops are already virtualizing workloads. And initially, virtualizing workloads brings substantial improvements in

efficiency to the datacenter. But there are still great improvements that can be realized by optimizing the virtualization stage.

- Some environments are still struggling to migrate physical workloads to virtualized infrastructure, improving the overall percentage of virtualized systems.
- Often existing virtualized workloads are not optimally deployed (e.g. - the need to "right size" a VM), and would benefit from moving to a different VM, environment, or hypervisor architecture.
- Some workloads already deployed or migrated to a VM would function far better having been moved back to bare metal.
- Avoiding vendor lock in, Hypervisor agnostic tools, are a great benefit. This includes the ability to move workloads between disparate hypervisor architectures. With the advent of private clouds, often utilizing different hypervisor technologies, this is an increasingly important element of success in this evolution. Large-scale enterprise data centers today are necessarily a mix of heterogeneous hardware, operating systems, network and storage devices; hypervisors will be no different.

While faced with optimizing existing virtualized environments Enterprises are simultaneously planning for utilization of cloud computing. Organizations envision building on-premises private clouds for speed, and extending to Hybrid models by migrating workloads to the public cloud to increase flexibility and reduce capital expense. The cloud has its own set of challenges

The move to cloud has its own set of challenges.

- Managing cost and quickly realizing the benefits of those technologies requires that legacy systems be leveraged to make this feasible, practical, and economical.
- Cloud resources are fundamentally different from datacenter resources. If you deploy a cloud resource and it's not used you are wasting money! Policy and automation around activating and deactivating cloud resources, per need, is fundamental to an economical cloud solution.
- With the milieu and changing landscape of Cloud Service Providers, it's vital to avoid vendor lock in.
- It's not just about moving workloads into the Cloud, workloads should be able to be moved back to the datacenter. Bi-directional Cloud migration is required.
- The Cloud should be position for integrated disaster recovery solutions.

The RackWare Software Solution

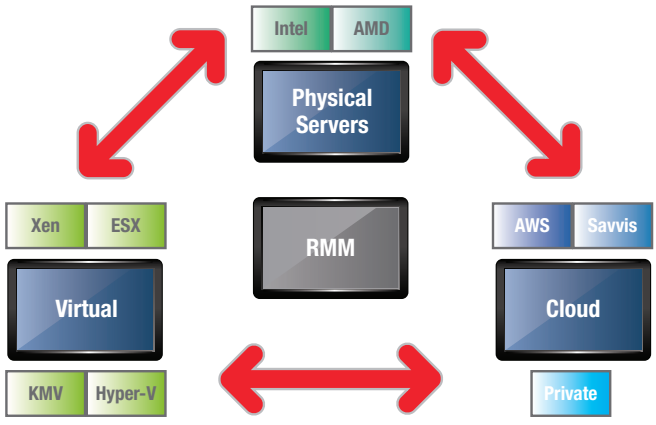
RackWare has a unique blend of capabilities that address the evolutionary flow of workloads, from optimizing virtualization to Cloud computing. The RackWare Management Module (RMM) layers the necessary automation and intelligence on top of workload mobility needed to meet these changing needs.

Key Features:

- Automated Intelligence for the Cloud
- Broadest platform support for Enterprise IT to optimize choice of performance, cost, and reliability
- On ramp to Cloud for both physical and virtual workloads
- Cost optimization without sacrificing availability and reliability
- Workload viability and sizing analysis for Cloud
- Robust Policy Framework
- Offline Network and Storage configuration of Images and Systems
- Data migration and Cloning independent of image migration
- Bi-directional Cloud migration
- Ability to sync to remote data center/cloud

RackWare's software vision is a holistic solution where an enterprise can evolve seamlessly towards a Cloud compute model. This solution can then be used to provision work objects to bare metal systems, hypervisors and CSPs – creating a seamless, end-to-end, utility computing model.

The RMM implements a push-button installation process for existing and Green Field data centers, essentially creating a physical-virtual-cloud infrastructure. The RMM provides a powerful a solution that encompasses multiple, heterogeneous platforms (bare metal systems, hypervisors and CSPs).



The RMM has been architected from the ground up to include the appropriate hooks to execute policies that are vital to automating and economizing datacenter work flow. Central to the RMM is a comprehensive policy framework that focuses on ease of use and adaptability.

The RMM is embodied by multiple solution sets.

- Onboarding
- AutoParking
- AutoScaling
- Cloud Bursting
- Disaster Recovery

Onboarding

RackWare Management Module's (RMM) On-Boarding solution enables mobility of legacy and Greenfield applications, realizing the full benefits of cloud and virtualized environments. On-Boarding is a RackWare solution that provides a mechanism to "capture" and "push" workloads, possibly from a remote location over the WAN, to a different datacenter or environment. It is an easy and convenient process that has the advantage of creating an exact duplicate of a running Image without the burden of rebuilding or recreating template images and applications. All Operating System and application settings are maintained in the process. Images can be moved with network configuration either retained or removed, whatever is most convenient.

Once the Image is pulled into the destination datacenter or Cloud, if desired, a new network and storage identity can be defined. The Image is then simply "assigned" to a System (Physical, Virtual Machine, or cloud entity). Image transfer can be via the WAN, or with an appliance that collects a large number of Images and then physically moved to the destination.

Scope of Platform Support

RackWare Onboarding supports "capture" of virtually any x86 Host in the datacenter. The capture process supports bare metal Hosts, and is Hypervisor agnostic, handling Virtual Machines across a wide variety of technologies.

The RMM does not simply manipulate VMDKs but utilizes an architecture that analyzes and catalogues Image metadata. This approach makes mobility to disparate environments and dissimilar hardware and VMs inconsequential.

Easy Deployment and Operation

The RMM first discovers Hosts and logically disassociates the Image information from the underlying System (hardware or Hypervisor). Once captured the Images are prepared and can be moved to the appropriate location.

Image Transfer Options

In cases where Images are smaller or a high speed connection exists between sites, Images are easily transferred via simple export/import mechanisms in the RMM. However, in many environments, connecting over the WAN may not be practical due to network topology or image size. The results of capture operations can be placed on a portable disk or simply stored in an appliance. The contents are then shipped securely to the destination site.

Two Way Mobility Operations

Image mobility is not a one way street. Simply moving Images in one direction is not enough to realize the maximum benefits of the cloud. The RMM can move Images between clouds as well as back to the original datacenter, even back to a bare metal server should the need arise. This two way mobility is supported for Images as well as data only transfers, where the data can be synchronized from a primary source to a destination source with only delta updates.

AutoParking

In many environments, workloads running in the cloud are not required 24/7, but rather at particular times throughout the day or through the week. Substantial cost savings can be realized to customers by employing strategically timed parking and unparking of cloud Hosts. Beyond cloud environments, in the datacenter, simply powering down Systems can yield cost savings, either in terms of reducing CPU running on Hypervisor systems or minimizing power and cooling on bare metal servers. Automated parking and unparking of Hosts, in any environment, at critical pre-determined times ensures adequate capacity during peak periods and cost savings during non-peak periods.

The RackWare Management Module (RMM) Policy Engine supports schedule-based AutoParking.

Policy objects can be easily created using RMM templates or customer provided media. These “Policy-schedule” objects define a Power Schedule Map that can be applied to a Host or group of Hosts. The RMM then executes automated triggers and downstream operations to actualize the Power Schedule Map for Hosts to which the policy has been applied. The benefits of AutoParking can be realized by Hosts running on bare metal, Virtual Machines, or CloudHosts, each with their own set of benefits depending on the environment.

Policy-schedules can be defined with great flexibility and detail, including daily, weekday (Monday through Friday), weekend (Saturday and Sunday), and specific day (e.g., Friday) attributes. Any number of triggers can be defined for any given day. The types of power events are also flexible, including power-on, power-off, and “park”.

An example of a Power Schedule Map might look like this:

```
Monday - Saturday          1200am - 0700am          park
Monday - Saturday          0700am - 1000pm          poweron
Monday - Saturday          1000pm - 1200pm          park
Sunday                     1200am - 1200pm          park
```

In this Power Schedule Map the Hosts to which the policy is applied are powered on every day at 7:00am and powered off at 10pm. The exception is Sunday where the System is not powered on at all.

It's possible to easily create more complex Power Schedule Maps such as:

```
Monday  - Thursday        1200am - 0700am        Park
Monday  - Thursday        0700am - 1000pm        Poweron
Monday  - Thursday        1000pm - 1200pm        Park

Friday                    1200am - 0700am        Park
Friday                    0700am - 0600pm        Poweron
Friday                    0800pm - 1200pm        Park

Saturday - Sunday         1200am - 0900am        Park
Saturday - Sunday         0900am - 0300pm        Poweron
Saturday - Sunday         0300pm - 1200pm        Park
```

In this example, Friday is an exception day where the System gets powered off several hours earlier than the rest of the weekday, achieving maximum cost savings.

As Hosts are powered on and off and parked, it may be important to coordinate with other infrastructure such as a load balancing engine or related tiered applications. This can be accomplished by pre and post action API calls. An optional script file or software module can be executed for every trigger action. The API is called prior to the event, as well as after the event.

An administrator can manually control a Host even when a Host falls under policy control of the RMM. For example, if needed, the administrator can poweron a Host prior to its scheduled poweron event. Then when the poweron event is triggered, the RMM detects that the Host is already powered on, and skips that operation. For longer term manual situations the Policy execution can be "paused" and "resumed" for specific Hosts as needed. This may be useful during a maintenance cycle or to address some issue with the Host.

Case Study: Customer saves substantial operational costs with RMM AutoParking
Recently a RackWare customer used a Cloud service to support its customer training courses. Because the customer's server usage started at 8:00 a.m. weekdays and ended at 6:00 p.m., the customer was able to set up a policy schedule that powered up resources when needed and parked them when not needed.

On weekdays, the customer required a minimum of 20 training applications to be running 98 percent of the time, at least 60 applications running 58 percent of the time, and a peak of 110 applications running 5 percent of the time. Each application required Microsoft Windows on a 2-core VM with a 50 GB boot drive and 25 GB data drive.

Sample Power Schedule Map:

```
Monday - Friday          0600pm -1200pm          park
Monday - Friday          1200am -0745am          park
Monday - Friday          0745am -0600pm          power-on
Saturday                 1200am -1200pm          park
Sunday                   1200am -1200pm          park
Monday                   1200am -0745am          park
```

The total monthly compute cost was lowered by 44% using the RMM AutoParking.

AutoScaling
Computing demands ebb and flow, which can test the limits of some resources while other resources sit idle. RackWare AutoScaling delivers automatic scaling for both private and public clouds. RMM AutoScaling provides server resources on demand and is based on real-time usage, load and performance. Groups of hosts can be defined with operational parameters that influence when hosts are activated and deactivated to meet these fluctuating demands

RMM's automation and intelligent layers deliver the right resources to applications at the right time, providing improved availability and significant savings. With AutoScaling, groups can span multiple environments and selectively utilize multiple pools of resources, even in a specified priority. AutoScaling groups can include both bare metal and virtual data center resources, private clouds, and public clouds.

Flexible Host Scaling for Any Environment
While most AutoScaling groups are composed of identical host images deployed in various locations, this is not a requirement — host images do not have to be identical or even have the same applications installed for AutoScaling to be effective. The solution can be applied even where a group has a subset of images within the same priority but with a different configura¬tion, or where the environment location of a set of host images implies a non-identical image. RMM can be used to set up an AutoScaling environment including the creation of cloned im¬ages from a master or primary image.

AutoScaling Setup in 1, 2, 3..
RMM AutoScaling is easy to set up and provides for ultimate flexibility: Simply define an AutoScaling group (CLI/API/GUI), define your minimum and maximum resource levels and define Triggers such as RW Monitoring, third-party monitoring and load balancing, etc. Capture the host used as the basis for AutoScaling, and then any number of cloned images can be created in terms of application installation and configuration. There is no need to create each custom image – each clone can be configured with its unique network configuration and assigned to a system.

Each host can be configured with the same monitoring parameters, with frequency of snap¬shots defined by the interval parameter. Statistics may be gathered at any frequency desired and averaged over time for decision-making metrics; this avoids a powering on or off action being taken on a peak or valley of a metric.

Powerful AutoScaling Features

- Flexible Triggering is set based on metrics gathered by RMM Monitoring, or generated externally via administrator-defined algorithms. ScaleOut and ScaleIn operations are then 'triggered' as load increases above or below set trigger points.
- Minimum and Maximum Resource Levels enables scaling to the number of hosts specified by a maximum parameter and also ensures a specific level of application resources regardless of the trigger status. Setting a minimum value also provides built-in failover.
- Prioritization allows for orchestration of resource allocation via the ability for hosts in an AutoScaling group to be defined as a Priority field. This then defines the order that hosts are activated or deactivated.
- As hosts are powered on and off, AutoScaling coordinates with other infrastructure such as a load-balancing engine.

Cloud Bursting

The Cloud is leveraged for all kinds of business processes such as retail transactions, website traffic, and video applications to name a few. At the same time, Cloud resources rarely work optimally in isolation but are deployed most effectively working in conjunction with other resources in the datacenter. In fact, an ideal use for Cloud Computing, is to address peak loads of applications that have great swings of demand.

It is typical for a core set of resources to be deployed to meet normal level of demand. But building out datacenter infrastructure in anticipation of these peaks is not cost-effective as they may be idle a majority of the time. The investment in capital and heavy maintenance costs such as power, cooling and administration are prohibitive.

To achieve cost objectives, and meet peak demands without degraded or interrupted service, requires a cohesive system of workload mobility and policy framework that spans multiple environments. The necessary elements of a Cloud Bursting solution include efficient On- Boarding of existing applications to the Cloud, along with dynamic scaling of resources based on policy (AutoScaling).
A Unique Blend of Technologies

A Unique Blend of Technologies

RackWare Management Module (RMM) Cloud Bursting is the culmination of AutoScaling and Onboarding working synergistically together. This unique combination provides scaling across heterogeneous environments. The environments can be single or multiple clouds, or simply extending the datacenter. As with the RMM AutoScaling solution, the RMM

policy framework can be used to set up a variety of Triggers to launch, stop, park, delete, or provision CloudHosts. The RackWare policy engine automatically launches Hosts and CloudHosts during peak periods based on applicable metrics. This information is gathered and analyzed by the RMM Monitoring Engine, or via external data from a Load Balancer.

In the case of extreme high load, the RMM can provision machines on the fly in a private or public cloud environment. When traffic or load drops back to more modest levels, RMM automatically decommissions Cloud resources and optionally copies application data back to the local Datacenter.

Cloudify and Uncloudify

Cloudify and Uncloudify provide the ultimate in Image mobility. The RMM "Cloudify" operation provides a single command to automate the process of deploying Images to a Cloud Service Provider (CSP). The RMM first performs a "capture" of the Image, and then makes the appropriate configuration modifications to the Image based on the target CSP. The "cloudified" images are then uploaded by RMM to the CSP of choice. Once uploaded to a Cloud, the Image can be synced with primary image in the Data Center to maintain data consistency across environments.

There may be instances where it's desirable to move Images back to the datacenter (or to another Cloud). In this case, Images can "uncloudified" by the RMM and returned to the datacenter for deployment there or moved to another Cloud.

Disaster Recovery

Business Continuance via Disaster Recovery is an essential element of IT and takes on many forms. At the high end, for business critical applications, cost is no object and DR solutions are justifiably large and complex. Along the rest of the continuum, however, all too frequently, Disaster Recovery is little more than tape backup. Where Disaster Recovery solutions are more feature rich than tape backup they are often implemented with a disjointed architecture separate from daily operations. Major improvements can be made with an economical disaster recovery capability that provides business continuance in a model that simply extends the existing IT architecture.

With the introduction of virtualization and, in particular, the cloud, many architectural elements that make those technologies successful seem well suited to building an economical DR solution. For this class of DR, datacenters should not have to carry the burden of deploying identical hardware or hypervisors at a DR site. Image mobility must support heterogeneity, being able to move Images among disparate hardware as well as hypervisors. Platform agnostic mobility is important not only when the computing trigger is shifted to the remote site, but it's also important to support easy mobility back to the origin site to resume normal operations. Beyond mobility a policy framework needs to be in place to automate processes and make intelligent decisions regarding workload operations.

RackWare Management Module's (RMM) Disaster Recovery solution builds on its Image mobility and elasticity features to bring economical Disaster Recovery to Enterprises. The building blocks of RackWare's DR solution include Onboarding, Cloud Bursting, and the Policy Framework to automate necessary functions.

Disaster Recovery as an Extension of Existing Infrastructures

RackWare's Disaster Recovery solution does not require a complicated, independent product to be installed, configured, and maintained. Rather the RMM, perhaps already installed for optimizing virtualization and cloud operations can be extended to support DR, even using existing cloud environments. RMM Captured Images can be cloned and pushed out to a DR site (a private or public cloud environment or remote data center). At that site, periodic updates are synchronized with the remote Image, keeping the original Host Image and the DR image in sync. In the event of an outage at the origin site, the up to date Image at the DR site can assume operations. This provides significantly improved DR support over traditional tape backup at a cost that is a small fraction of keeping a replicated datacenter.

Setting up a DR Image utilizes the basic fundamentals of the RMM, beginning with capturing the origin Host. Then a clone of the Image is made, cloudified, and deployed at the desired cloud location. The resulting CloudImage is left in its most economical state. The CloudImage is activated only when assuming production operations following a serious outage at the primary data center.

It's important that Image updates include user data, Operating System updates, and application installations and configuration changes. The cloud DR feature supports all of these. While OS updates are more infrequent it is still important to ensure that kernel patches are kept in sync with the DR Image. When updating the OS, an image refresh operation can performed from the RMM first before the sync to the CloudImage. In the event of a DR event, the CloudImage is launched and is running with the latest synchronized changes.

After a DR event, the CMDB can be resumed at the origin site, and if necessary a rediscovery of the environment to address any changes that occurred. In terms of returning the workload execution to the origin site, it's important to ensure that any changes to the CloudImage are reflected as part of that process. When the origin site is restored to its operational state, the administrator can utilize the capture from cloud feature to refresh the original Image. Once they are captured, the RMM can deploy those Images and operations are resumed as normal.

Conclusion

The RMM provides a key building block for Enterprise virtualized and cloud environments, bringing intelligence and automation. The RackWare solution enables enterprises to optimize operations at every step in the evolutionary path from virtualization to Cloud computing. Computing resources—physical, virtual, and cloud machines—can be easily and automatically scaled up or down as computing needs fluctuate.

About RackWare

RackWare brings intelligence and automation to the cloud, providing greater availability for enterprises, greater flexibility for enterprise IT users, and reduced costs for enterprise IT providers. Computing resources—physical, virtual, and cloud machines—can be easily and automatically scaled up or down as demand fluctuates. On average, RackWare customers realize a cost savings of 40 to 50 percent, while getting the highest performance and availability out of their cloud. RackWare was founded in 2009 and is based in Santa Clara, California.

Phone: 408-430-5821
Email: info@rackwareinc.com
Web: www.rackwareinc.com