White Paper

Assuring and Accelerating Cloud Application Migrations

The Challenge

Enterprises may operate hundreds or perhaps even thousands of line-of-business applications, hosted in private datacenters (DCs) on VMware infrastructure. These applications generate widely-varying workloads. Some are relatively stable, but others show periodic variation (e.g. end-of-month billing), seasonality (e.g. changes in demand around holidays), or anomalous events (e.g. spikes due to high-profile product launches).

When workloads are stable, enterprises are able to employ traditional capacity planning and make efficient use of their private capital equipment in the data centers. However, for varying workloads, this virtual infrastructure is sized for the worst case. This is a safe approach, as it ensures that application performance doesn't degrade during periods of peak workload, potentially impacting business outcomes. However, it does not make efficient use of capital investment. During non-peak periods, these workloads are effectively running on over-provisioned virtual infrastructure.

Enterprises should see the benefits of migrating selected applications to public cloud providers, taking advantage of elastic computing and other resources. With this approach, private infrastructure can be resized for the "normal" case, with cloud computing resources handling extra load up to peaks. Applications may also be completely migrated to the cloud and then scaled up and down as needed. Either way, efficiency can be increased by shifting CapEx to OpEx and reducing overall spend.

Problem Statement

Operationally, an enterprise's primary challenge is that application migration to the public cloud is both time-consuming and risky. When considering migrating candidate applications, decision makers need to be confident that public cloud infrastructure can handle the intended workloads, and that latencies on critical inter-application interfaces will not increase. There are many cloud providers and myriad service offerings. Given a candidate application and prospective cloud configuration, the relevant question is, *"can this cloud configuration effectively support the application?"* Since every application migration requires time and effort, this question must be answered even before a trial is attempted.

Over the longer-term, enterprises are faced with an additional challenge: *"Which applications should be migrated in order to maximize cost savings?"* Collectively, hundreds or thousands of applications form a huge set, far too many for each application to be manually considered by analysts. *How to select the right combinations of applications, application configurations, and cloud providers to maximize CapEx savings / minimize new OpEx spend?*

Assuring Cloud Migrations

In order to assure successful cloud migrations, it is necessary to have a solution capable of actively measuring cloud infrastructure performance in multiple dimensions: compute, memory, storage, and network.

Active measurement means simultaneously generating workloads and measuring the infrastructure's ability to support the same:

- **Compute workload**—requires running computationally-intensive code, forcing hypervisors to schedule the necessary amount of vCPU.
- **Memory load**—requires generating read and write memory accesses, forcing underlying hardware to exercise caches and physical memory.
- **Storage load**—requires generating read and write I/O accesses, forcing directly- or network-attached storage subsystems to read and write to physical disks.
- **Network load**—requires generating bidirectional packet flows between the local VM and remote endpoints, exercising virtual and physical networks.

In addition to measuring the network's ability to deliver application traffic, it is also important to measure the quality of the network, including *one-way latency, packet loss, duplication, reordering, and integrity*.

When an existing application is being considered for migration, enterprises need to extract historical time-series data from their monitoring system (see Figure 1 showing VMware vRealize Operations, or vROps, example) to form the basis of a test workload. The workload may be scaled up or down in any dimension. For example, the whole workload may be doubled or tripled to simulate the impact of anticipated growth. Or, a specific dimension like network traffic may be scaled up or down to explore application-specific scenarios. The time scale of the workload may also be compressed. For example, to represent a year's worth of application activity may be compressed down to a 15-minute test workload.

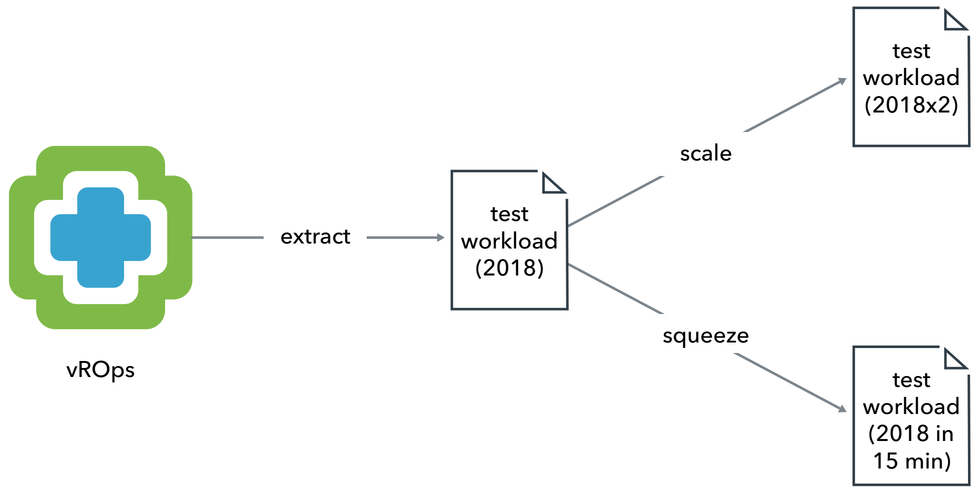

Figure 1. Developing test workloads using historical data from vROps

Using their knowledge of specific applications, enterprise engineers or system integrators may also add application flows to the workload. A flow supplements monitoring-derived data with additional information about a critical inter-application interface:

- The application/service (in a private DC or elsewhere in public cloud) is this application talking to
- The protocol being used
- Typical packet/request/transaction rates and direction
- The required network service level objectives (SLOs)

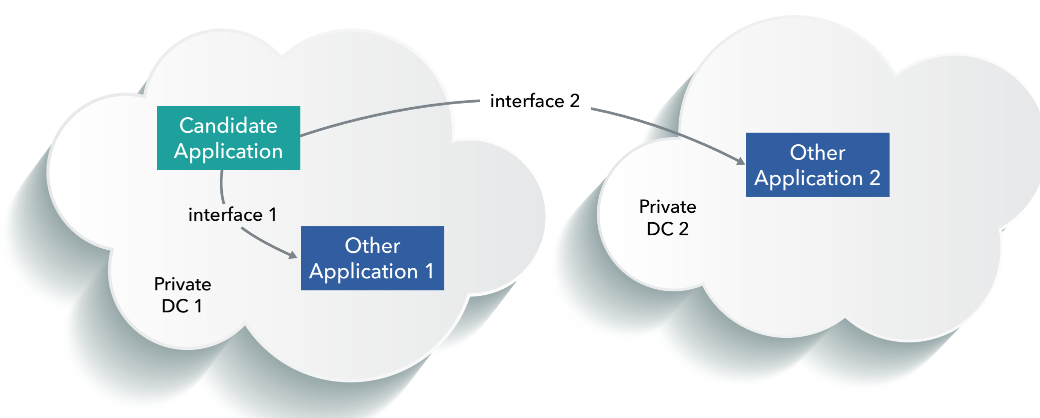

Figure 2. A candidate application with two critical interfaces

Given a proposed cloud configuration, the test workload can help answer, *"Can this cloud configuration effectively support the application?"*

Test agents can be deployed on the public cloud to generate a test workload that is representative of the application. Each of the CPU, memory, storage, and network dimensions are loaded up to levels indicated by the test. This process may be run as a short-duration test (e.g. compressing a long time-series into a much shorter one for a quick assessment) or may be left to run in actual time (e.g. for soak testing).

If it is not possible to position test agents in a enterprise's private datacenter (DC), application flows that are part of the workload can still be run between the test agents deployed solely within the public cloud.

The primary impacts of not positioning test agents in the DC pertain to latency measurements:

- Latency measurements are then determined using ICMP ping packets instead of test traffic. Test agents can still generate a workload of test traffic that emulates the migrated application and its interfaces, however this traffic will stay within the public cloud and use ping packets instead (assuming that hosts in the DC will respond to pings).
- Latency measurements are measured as round-trip times. Using ICMP ping means that the latency of inter-application interfaces is measured as a round-trip time of both a request and a response, instead of measuring requests and responses individually.

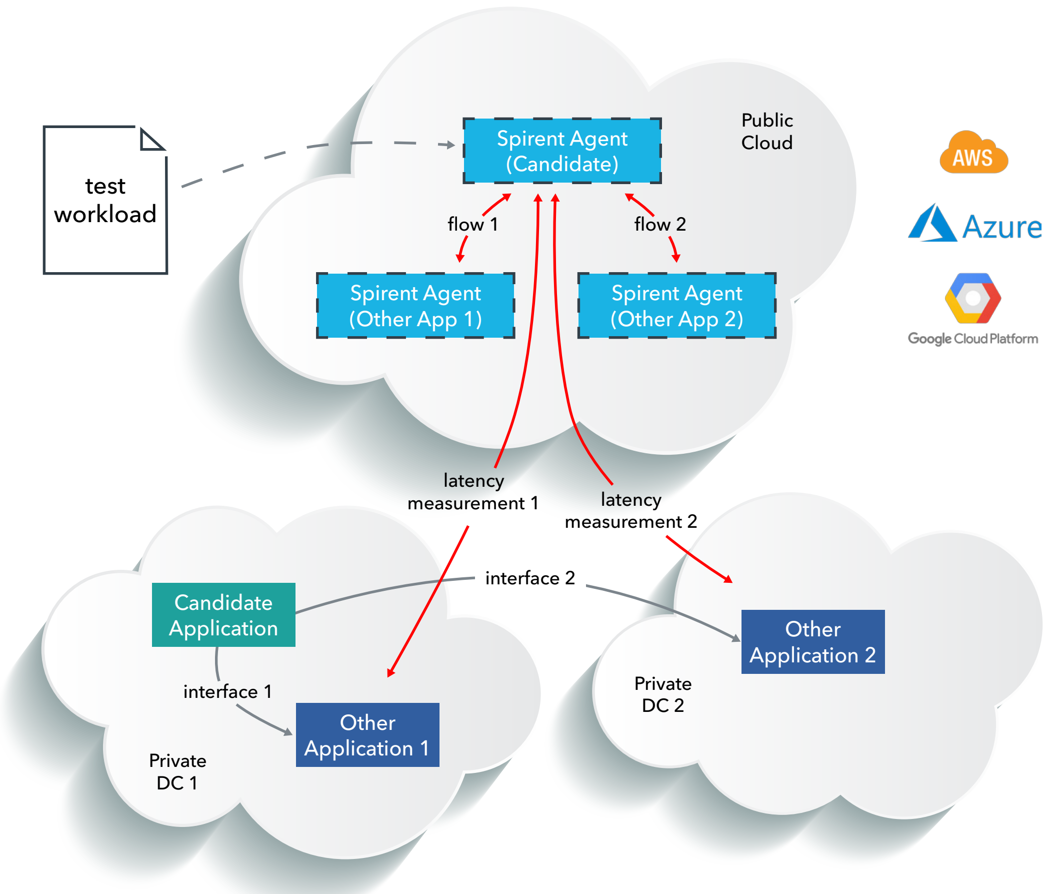

Figure 3. Using a test workload to assure public cloud infrastructure's ability to support the application

From the perspective of the virtual infrastructure, the workload generated by the test solution will be as close to the migrated application as possible, without actually migrating the application.

If network SLOs are unknown, then the test solution deployed on the DC side may be used to establish a baseline. For example, given an existing application that has interfaces across a DC boundary (e.g. as shown in Figure 2 above), measure the existing latencies between these DCs. These metrics may be used as the basis for subsequent latency measurements, i.e. *"What is the change in network latency that this application will experience, post-migration?"*

Significant variations between the intended workload and the actual/delivered workload may indicate that cloud infrastructure cannot support application migration. This could be an indication that reconfiguration is required (e.g. choosing a different compute instance type, switching to single-tenancy to avoid noisy neighbors, increasing provisioned IOPS, or selecting a different network interface). It may also be an indication that the application should not be migrated, no matter what the configuration (e.g. network latencies on critical application interfaces have increased beyond acceptable limits).

Accelerating Cloud Migrations

Since a rich source of historical time-series data already typically exists (e.g. in the form of vROps, databases as in the example above), it can be extracted and analyzed to provide enterprises additional opportunities to maximize their CapEx / OpEx savings:

- Using analytics to select candidate applications from their huge set of applications
- Estimating the cost of running all or a portion of those applications on various cloud providers
- Recommending specific configurations for the applications to be migrated

Applications that exhibit variable workloads represent potential cost-reduction opportunities. There are several factors to be considered. What dimensions (i.e. compute, memory, storage, and network) contribute to the variability? How significant is this variability? Is there a stable baseline with isolated peaks? Or, is the workload especially bursty? These types of metrics form the basis of initial candidate application selection and ranking.

Once extracted from vROps, these dimensions can be used to develop a candidate migration strategy and cloud configuration for an application. For example, if an application has a stable baseline with seasonal or periodic bursts then it may be a candidate for partial migration to the cloud, i.e. with the cloud used only for excess workload. As a counterexample, if an application is highly variable but has no discernable baseline, then it may make sense to migrate the entire application to the cloud and then to subsequently scale allocated resources up or down. In both cases, the raw data can be used to derive the specific cloud configuration for the application.

Given a set of candidate applications, application configurations, and cloud providers, the test solution for migration assurance can be used to de-risk the migrations as described above. Decision makers can focus their time and energy on those applications that have the greatest cost savings potential and the least risk.

Benefits Realized with Cloud Migration Assurance

- **Time Savings:** existing data sources can be mined to identify the most promising candidates for migration to the cloud. During migration assurance testing large time-series data sets may be compressed so that long-term application performance can be assessed in a much shorter amount of time, e.g. validate that the cloud configuration can handle the last years' worth of application load in under an hour.
- **Cost Savings:** maximize CapEx savings and minimize new OpEx spend.
- **Confidence in Migrations** by actively measuring (not simulating) application workloads: Test workloads are generated directly on public cloud infrastructure, and variations in the actual performance of underlying compute, memory, storage, and network can be are measured. Users can perform as many tests as desired, without the hassle of a trial migration.
- **Derisked Public Cloud Migrations:** Cloud migration assurance solutions help accurately measure the network quality on critical application interfaces, including latency. Any mis-specification or mis-configuration of public cloud instances, storage, and network – which may impact application performance post-migration – can be detected, thereby avoiding failed migrations and the cost of mitigating actions.
- **Data-Driven Decision Making:** selection of candidate applications, application configurations, and test workloads is based on time-series data extracted from enterprise's historical data, allowing for consideration of important periodic, seasonal, or anomalous events in the entire process.
- **"What If?" Analysis:** in addition to cloud provider shopping, test workloads may be scaled up or down, e.g. simulating the effect of doubling or tripling application loads. Compute, memory, storage, and network load may be scaled independently to simulate application-specific scenarios. Noise may be added to each dimension to simulate the impacts of unexpectedly bursty traffic on infrastructure.

About Spirent Communications

Spirent Communications (LSE: SPT) is a global leader with deep expertise and decades of experience in testing, assurance, analytics and security, serving developers, service providers, and enterprise networks.

We help bring clarity to increasingly complex technological and business challenges.

Spirent's customers have made a promise to their customers to deliver superior performance. Spirent assures that those promises are fulfilled.

For more information, visit: www.spirent.com

Spirent is a trusted provider of test, measurement, assurance and analytics solutions for next-generation devices and networks. We help bring clarity to increasingly complex technological and business challenges. Contact us today to learn how to accelerate your cloud application migrations and realize all the benefits by building on our technology for assuring cloud migrations.

spirent™ Promise. Assured.

Contact Us

For more information, call your Spirent sales representative or visit us on the web at www.spirent.com/ContactSpirent.

www.spirent.com

Americas 1-800-SPIRENT
+1-800-774-7368 | sales@spirent.com

Europe and the Middle East
+44 (0) 1293 767979 | emeainfo@spirent.com

Asia and the Pacific
+86-10-8518-2539 | salesasia@spirent.com