virtana

White Paper

How to Ensure Successful Cloud Migrations by Understanding Performance and Cost Before Moving Your Applications

Ricardo A. Negrete
Sr. Director Services and Technology
Virtana

Table of Contents

1 Introduction

For many organizations, the route to digital transformation is to migrate a set of applications or workloads to an Infrastructure as a Service (IaaS platform, aka "Lift and Shift". Industry analysts estimate "Lift and Shift" will comprise a significant percentage of the initial migrations, with the remainder of the migrations focused on re-engineering applications to take advantage of cloud native or SaaS capabilities. This white paper will focus on how to de-risk the "Lift and Shift" migrations with the Virtana Cloud Migration Readiness (CMR) Service.

1.1 Background

One enabler of Information Technology's digital transformation is the "Cloud", which promises organizations significant value in the form of business agility, faster innovation, on-demand scalability and cost savings.

Once organizations decide to move forward with the cloud, the question often asked is: "Where do I start?"

Even if your deployed on-premise applications are well understood and documented, the first step is to evaluate the current application portfolio from a business function and life cycle perspective. Understanding where applications are deployed, and their interdependencies is an essential first step. Due to business pressures or resource constraints, many organizations find themselves with a limited understanding of their on-premise infrastructure particularly when it comes to application interdependencies and infrastructure components utilized.

There is range of possibilities when evaluating applications for digital transformation, such as:

Decommission. There are new and better ways to provide the business function. The digital transformation driver makes the application obsolete (e.g. implementing mobile apps). Why move it, if it is at the end of its useful life?

Modernize the Application: Should the application be re-engineered using Cloud Service Provider (CSP) native services? Or should the application be **"lifted and shifted"** to the cloud with minimal to no modifications?

Rent the business function: If the business function can be adequately delivered by a Software as a Service (SaaS) provider, then why worry about infrastructure, licenses and other data center concerns when a SaaS provider can adequately provide that business function at a reasonable cost?

Based on these possibilities, the organization must evaluate and decide which cloud service model makes sense for their targeted applications and business transformation strategy, taking into consideration the roles and responsibilities of their organization and CSP inherent in each option.The table below shows the different levels of responsibilities between on-premises IT staff and the CSP:

On-Premises	Infrastructure as a Service IaaS	Platform as a Service PaaS	Software as a Service SaaS
Application	Application	Application	Application
Security	Security	Security	Security
OS	OS	OS	OS
Hypervisor	Hypervisor	Hypervisor	Hypervisor
Infrastructure	Infrastructure	Infrastructure	Infrastructure
Data Center	Data Center	Data Center	Data Center

Table no.1 On-Premise vs CSP (highlighted) responsibilities

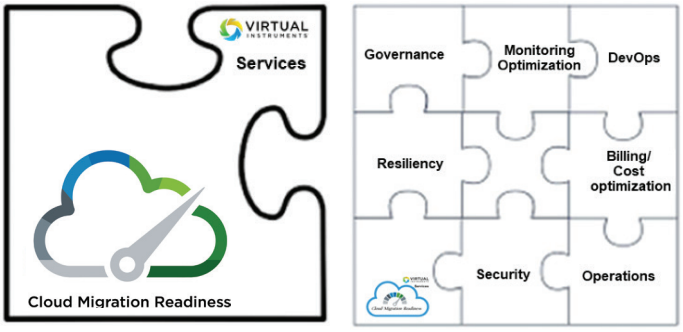
2. The Cloud Migration Puzzle – Start with a corner piece

The "end to end" migration process to an IaaS cloud is a challenging undertaking consisting of many critical phases and spanning IT infrastructure and application organizations. Digital transformation will require changes within the IT and Business organizations and processes in order to achieve the business objectives.

There are several major pieces in the puzzle consisting of different areas such as:

- **Cloud configuration**
- **CSP Cost**
- Security
- Compliance
- Resiliency
- Governance
- Billing and cost optimization architecture

So where do we start? As with any puzzle, we start with the corner piece!



There are three fundamental questions that need to be answered when migrating applications to a cloud delivery platform:

1. Will my applications perform as expected in a public cloud? (Application Fitness)
2. How much will it cost to run my applications in a public cloud? (OpEX)
3. Which Cloud Service Provider is the best choice for my applications? (Cost and Fit)

The CMR service represents a corner piece of the puzzle, organizing and making sense of massive amounts of data to which enables cloud migration initiatives to make informed decisions and move forward.

A Virtana CMR engagement answers the following more detailed questions:

- How do I know which workloads to migrate and which to retain in the data center?
- How do I choose the right cloud service provider for each of my applications?
- How do I simplify the analysis and reduce the time to migrate a large number of diverse workloads?
- How do I "Rightsize" CPU, Memory, Network and Storage configuration for each migrated workload rather than simply replicate my on-premise configurations?
- How do I test cloud workload performance before migrating the workloads?
- How do I prevent migrated workloads from having unforeseen dependencies back to the data center?
- How do I determine if migrated workloads are performing adequately and what can I do if they aren't?

CMR offers a solution to de-risk cloud migration by validating the suitability of the targeted applications based on their on-premises performance SLAs, their dependencies, the preservation of SLA performance in the target CSPs and the estimated costs.

CMR leverages the Virtana Professional Services team, the Infrastructure Performance Management experts. A comprehensive investigation using workload analytics, workload playback and automation capabilities that are based on the Virtana VirtualWisdom and WorkloadWisdom performance and capacity analysis tools will be used throughout the services engagement, providing unparalleled insights into your production application infrastructure estates.

The term "application" or "workload" or "virtual machine" will be used interchangeably throughout this document. It refers to a business application, which is a unit of interacting software that will typically be moved as a single component. A workload has connections of various kinds with other applications and systems.

3. VI Cloud Migration Readiness – CMR Phases

To successfully reduce the cloud migration complexity, right-size cloud configurations, and validate cloud performance, CMR defines four distinct phases:

Discovery - Discover workload characteristics and identify dependencies between compute, networking and storage elements.

Profiling - Distillation of hundreds or thousands of workloads into a small set of representative synthetic workloads that accurately characterize performance.

Playback - Accurate playback of representative synthetic workloads in the cloud to select cost-optimal configurations without compromising workload performance.

Monitor - Monitor actual workloads post migration to the cloud to identify any unforeseen performance or capacity issues

Figure no.2 below shows a brief description of the tasks, deliverables and estimated time-frames (time-frames will vary depending on the scope) for each phase. Also, noted is the phase handled by a 3rd party provider who performs the actual "movement of data" and cloud configurations based on the recommendations provided by CMR's outputs and deliverables.

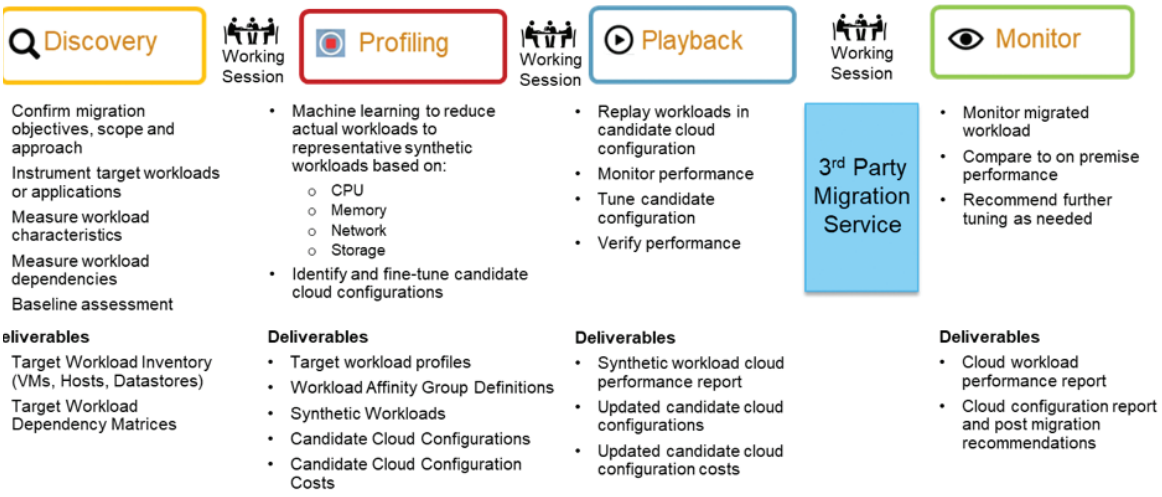


Figure 2: Virtual Instruments Cloud Migration Readiness (CMR) Phases

4. CMR – Discovery Phase

During the CMR Discovery phase the business objectives, the application portfolio and the targeted workloads are identified. Once the targeted workloads are identified, CMR discovers the **inventory** that makes up the workloads and their **inter-dependencies.**

Once the inventory and dependencies have been captured, the workload's **utilization and performance characteristics are measured.** The information collected during the Discovery phase will be an input to the Profiling phase.

A **baseline assessment** is conducted to understand the health, utilization and performance aspects of the application's on-premises infrastructure. It will serve as the reference point for comparing the cloud's utilization and performance and also identify potential issues which should be addressed prior to migration.

4.1 Inventory

To create the inventory, Virtana Services VirtualWisdom Software and CMR Tools are deployed in the target environments.

The inventory documents the workloads' usage components such as number of vCPU, Memory size, and configured storage. The information collected serves as a baseline to understand the resource demands and requirements of the targeted workloads.

	A	B	C	D	E	F	G	H	I	J	K	L	T	U	V	W	X
1	moref	Workload	memoryS	powerState	numCpu	ethernet	VirtualD	guestFullN	maxCpuU	maxMem	capacityInK	ipAddress	cluster	host	hostcpuM	hostcpuM	hostmem
2	vm-166	AWS VW5 2	65532	poweredOn	8	1	2	Other Linu	16800	65532	3221225472	10.36.4.41	SVCS_Pro	vim.Host	2100	Intel(R) X	2.06E+11
3	vm-581	BETA 5.4.0-	32768	poweredOn	4	1	1	SUSE Linux	13300	32768	1073741824		SVCS_Pro	vim.Host	3325	Intel(R) X	1.03E+11
4	vm-718	CMR- Appli	8192	poweredOn	2	1	1	Ubuntu Lin	6650	8192	67108864	10.36.4.10	SVCS_Pro	vim.Host	3325	Intel(R) X	1.03E+11
5	vm-721	CMR- Appli	8192	poweredOn	2	1	1	Ubuntu Lin	6650	8192	67108864	10.36.4.99	SVCS_Pro	vim.Host	3325	Intel(R) X	1.03E+11
6	vm-716	CMR Rapid	8192	poweredOn	2	1	1	Ubuntu Lin	6650	8192	102400000	10.36.4.10	SVCS_Pro	vim.Host	3325	Intel(R) X	1.03E+11
7	vm-710	CMR_2012	8192	poweredOn	2	1	2	Microsoft V	6650	8192	260046848	fe80::dac:	SVCS_Pro	vim.Host	3325	Intel(R) X	1.03E+11
8	vm-714	CMR-001 Ap	8192	poweredOn	2	1	1	Ubuntu Lin	6650	8192	67108864	10.36.4.89	SVCS_Pro	vim.Host	3325	Intel(R) X	1.03E+11
9	vm-638	CMR-002 - F	16384	poweredOn	4	1	1	Ubuntu Lin	9576	16384	209715200	10.36.4.76	SVCS_Pro	vim.Host	2394	Intel(R) X	2.06E+11
10	vm-655	CMR-003-U	8192	poweredOf	4	1	1	Ubuntu Lin	12236	8192	209715200		SVCS_Pro	vim.Host	3059	Intel(R) X	1.03E+11
11	vm-715	CMR-CentC	8192	poweredOn	1	1	1	CentOS 4/5	3325	8192	67108864	10.36.4.73	SVCS_Pro	vim.Host	3325	Intel(R) X	1.03E+11
12	vm-51	DCFM1	8192	poweredOn	2	1	1	Microsoft V	6118	8192	209715200	fe80::244	SVCS_Pro	vim.Host	3059	Intel(R) X	1.03E+11
13	vm-706	demo-base	65536	poweredOf	8	1	2	SUSE Linux	16800	65536	131072000		SVCS_Pro	vim.Host	2100	Intel(R) X	2.06E+11
14	vm-684	demo-pm0	65532	poweredOf	8	1	1	SUSE Linux	19152	65532	1073741824		SVCS_Pro	vim.Host	2394	Intel(R) X	2.06E+11
15	vm-252	demo-vw5	32768	poweredOf	8	1	1	Other Linu	16800	32768	1073741824		SVCS_Pro	vim.Host	2100	Intel(R) X	2.06E+11
16	vm-264	demo-vw5	65536	poweredOn	8	1	1	Other Linu	19152	65536	1073741824	10.36.4.31	SVCS_Pro	vim.Host	2394	Intel(R) X	2.06E+11
17	vm-485	ESTEEVM_2	4096	poweredOn	4	1	1	Other (32-b	9576	4096	83886080	fe80::5d0:	SVCS_Pro	vim.Host	2394	Intel(R) X	2.06E+11
18	vm-411	GetReal Re	36864	poweredOn	6	1	2	Ubuntu Lin	18354	36864	587202560	10.36.4.11	SVCS_Pro	vim.Host	3059	Intel(R) X	1.03E+11
19	vm-593	GigaVUE-FM	4096	poweredOn	1	1	1	CentOS 4/5	3059	4096	41943040	10.36.4.68	SVCS_Pro	vim.Host	3059	Intel(R) X	1.03E+11
20	vm-151	LDX-E-Appl	16384	poweredOn	8	2	2	Debian GN	4000	16384	348127232	192.168.1	SVCS_Pro	vim.Host	1900	Intel(R) X	2.06E+11
21	vm-152	LDX-E-db-2	16384	poweredOn	8	2	2	Debian GN	3000	16384	2181038079	192.168.1	SVCS_Pro	vim.Host	1900	Intel(R) X	2.06E+11
22	vm-153	LDX-V-5.2-L	5120	poweredOf	5	5	1	Debian GN	None	None	4194304		SVCS_Pro	vim.Host	1900	Intel(R) X	2.06E+11

Figure 3: CMR Inventory example (partial view)

4.2 Workload Inter-dependencies

For any migration plan, it is critical to understand the **workload's interdependencies**. During the CMR Discovery phase, various types of workload dependencies are discovered. It is vital to understanding the boundaries and landscape of services supporting the workloads. The workload interdependencies help to determine the sequence the workloads and services are migrated to the cloud and what workloads and services stay on-premise.

The CMR Discovery phase produces detailed analysis of workload communications leveraging NetFlow-based network and hypervisor-based storage-access information such as:

- Counts of workloads communicating to a particular workload or service.
- Workload to workload communications.
- Workload to external communications.
- Workloads using one or multiple datastores.
- Counts of workloads using a particular datastore.
- Fit/Unfit applications based on-premise dependencies.

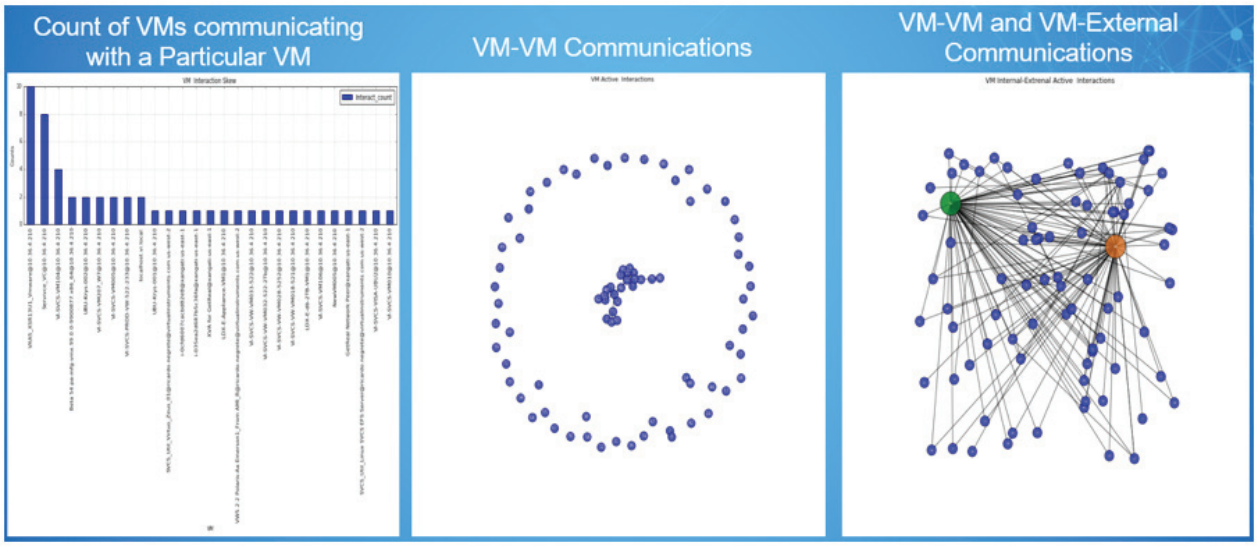


Figure 4: CMR Workloads Inter-Dependencies example (partial view)

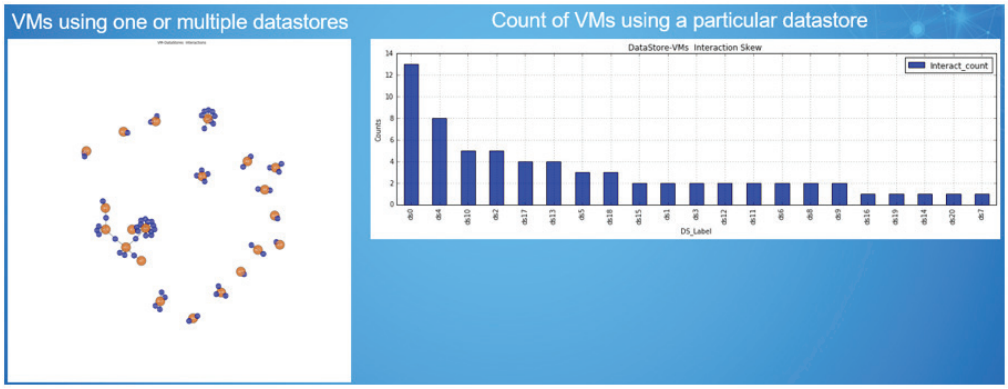


Figure 5: CMR Workload Datastores Inter-Dependencies example (partial view)

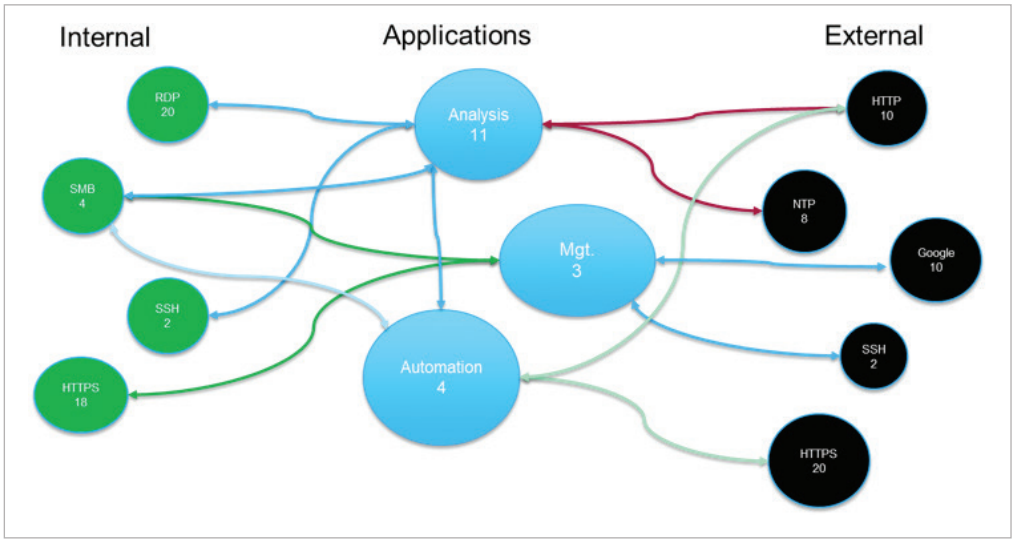


Figure 6: Example of on-premises application external and internal dependencies before migration

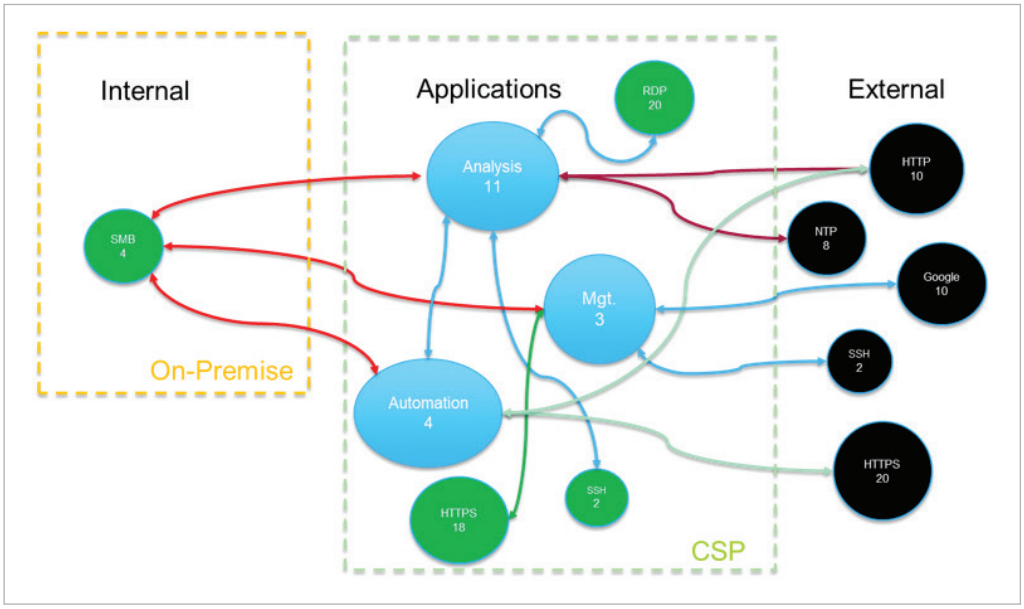


Figure 7: Example of application external and internal dependencies after migration. A dependency to an on-premise service remains which will require network connectivity back to the data center.

4.3 Workload Characterization – Performance Measurement and Baseline Assessment

Workload characterization is the next step in the process. VirtualWisdom Virtual Edition is deployed including the compute (ProbeVM) and network (NetFlow) monitoring capabilities to measure utilization and performance for the targeted workloads. The duration of the collection of data is dependent on the organization's objectives (e.g. capturing peak activity) and utilization seasonality.

Typically, the monitoring time-period varies between two weeks to one month. The collected data is used to measure resource utilization and to identify ongoing performance issues such as CPU contention or Memory pressure due to oversubscribed hosts or virtual machines. It is often assumed that such performance conditions will go away once the workload is running in the scalable cloud, however, this could come at significant expense and verification after migration is essential.

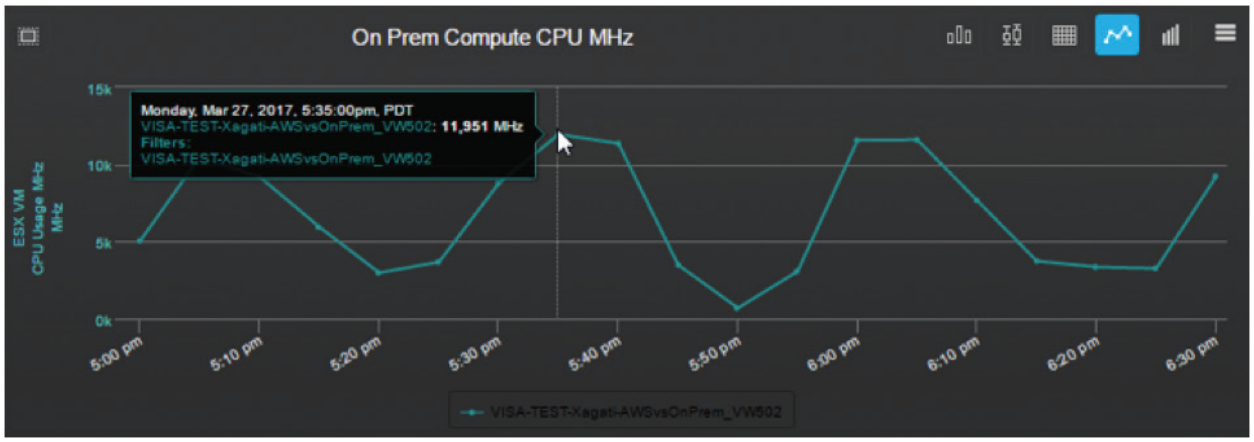


Figure 8: CMR Workload Characterization Example – CPU MHz

Utilization – "Bully" Virtual Machines

Assessment: The list below shows the "Bully VMs" in the environment. They could be run away processes or misconfigured applications.
VI Advice: Further investigation is required to determine if the workloads are fit to migrate to the cloud.

	VM CPU Utilization - %							
ESX VM	Summ...	Min	5th	25th	Median	75th	95th	Max
SNOW-dockerba1	100	100	100	100	100	100	100	100
App4-ubuntu1704-...	100	100	100	100	100	100	100	100
Auto3-docker1	75.624	75.28	75.41	75.52	75.61	75.73	75.95	75.96
Netflow_Sim_Maruti	59.101	58.57	58.62	58.75	58.89	59.33	59.8	59.99
CMR-001 Appliance	53.973	53.85	53.86	53.94	53.98	54.02	54.03	54.04

Figure: 9 Example of Baseline Assessment (partial). "Run Away" processes or idle virtual machines must be identified and investigated before considering them for cloud migration. The health, utilization and performance of target environment should be well understood before cloud migration.

4.4 Discovery Deliverables

During the CMR Discovery phase the following deliverables are produced:

- Target workload inventory
- Dependency matrices and map
- Workload Utilization and Performance assessment
- Workload characterization –capturing eight distinct dimensions:
  - CPU MHz
  - Memory Utilization - GiB
  - Read and Writes - IOPs
  - Read and Writes Throughput - Bps
  - Network Received and Transmit - Bps

5. CMR Profiling Phase

During the CMR Profiling Phase, Virtana's advanced analytics are applied to the workloads that were characterized during the Discovery phase. Depending on the organization's goals, the number of workloads under consideration may range from hundreds to thousands. In either case, the first step in the Profiling phase is to consume the workload characterization data to produce "Affinity Groups."

5.1 "Affinity Groups"

The "Affinity Groups" consist of workloads with similar resource utilization levels and temporal (time-based) characteristics. The Virtana CMR analytics clusters workloads by analyzing and comparing resource utilization dimensions from both the temporal and resource level dimensions. CPU, Memory, IOPs, and Network utilization levels are aligned with temporal data to create groups with statistically similar traits such as workload with CPU intensive demands or Memory or IO or a combination.

5.2 "Representative Workloads"

Once the "Affinity Groups" have been created, CMR's machine learning algorithms are applied to create a "Representative Workload". The Representative Workload captures the relevant utilization attributes of the workloads within an Affinity Group. It is constructed by concatenating temporal segments of various workloads to preserve relevant utilization levels for the group. The result is a high fidelity synthetic representation of the hundreds or thousands of workloads. Armed with Representative workloads, the number of Workloads that require verification in the cloud is significantly reduced. The complexity and scale are condensed to a manageable set without losing the original and critical characteristics of the on-premise workloads.

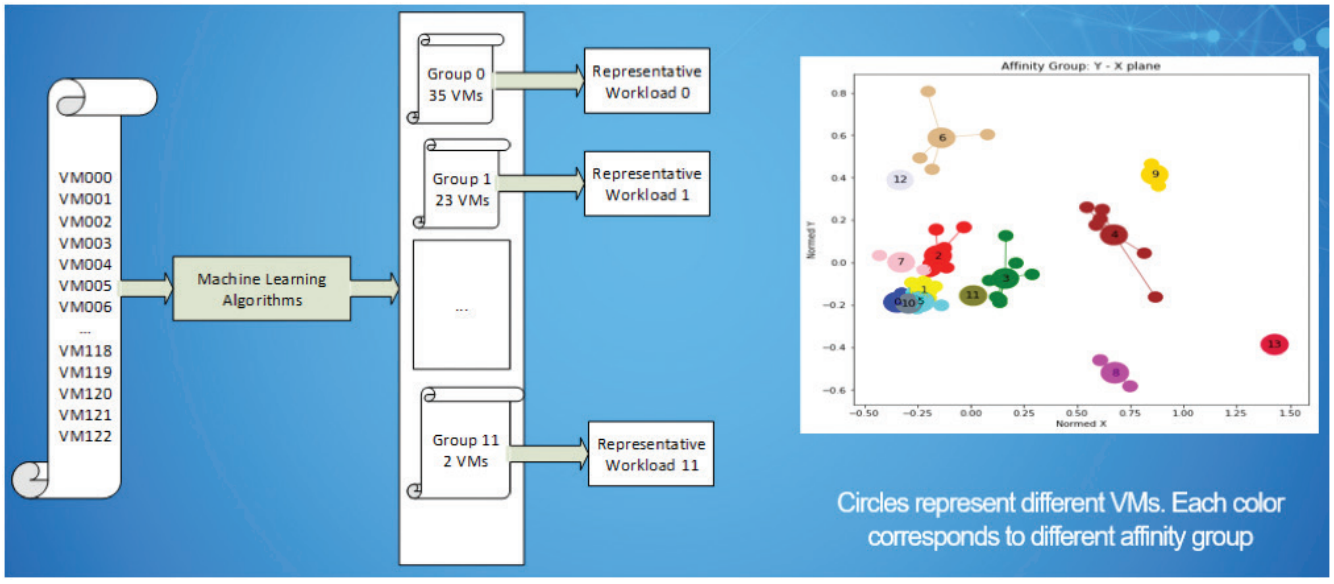


Figure 10: CMR Profiling Phase – "Affinity Groups" and "Representative" Workloads

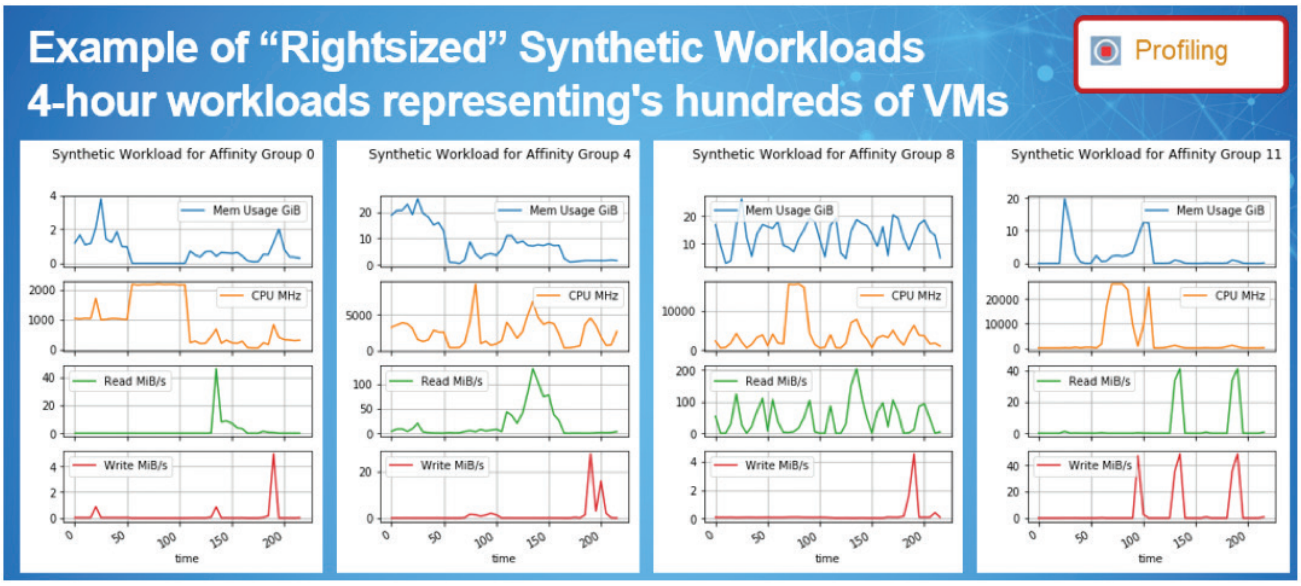
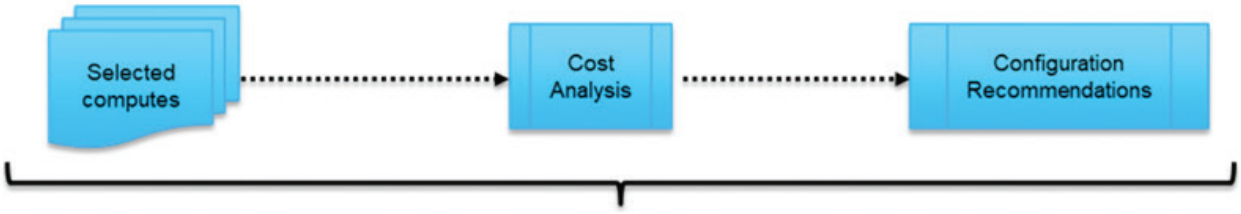


Figure 11: CMR Profiling Phase – Example of synthetic 4-hour workloads for 4 Affinity Groups

5.3 Initial candidate cloud configuration and cost analysis

Initial cloud configuration choices are selected based on the utilization attributes of the “Representative Workloads” and considering the original on-premise inventory configuration documented during the Discovery phase.

The candidate cloud configuration choices are the targets for “Playback” of the “Representative Workloads”.



Application	Virtual Machines	Azure Instance	Azure Disk Type	Instance Monthly Cost	App Monthly Cost
Analysis	VM014-523-2TB VM016-523-1TB	L8s D4sv3	P30 P15	$449. 28 $276.48	$725.76
Management	VM001-523-64gig	F16v2	P10	$937.44	$937.44
Automation	VM027-522-HF1	L4s	P40	$224.64	$224.64
WinTel1	Del-030318	D2sv3	P20	$69.12	$69.12
⋮					
WinTel16	Del-030318	D2sv3	P20	$69.12	$69.12
Grand Total					**$25,407**

Figure 12: CMR Profiling Phase – Candidate CSP Computes and On-demand Cost.

5.4 CMR Profiling Deliverables

During the CMR Profiling phase the following deliverables are produced:

- **Affinity Groups**
- **Representative Workload** – Temporal workload metrics to be consumed by the CMR Playback engine.
- **Candidate Cloud Configuration**
- **Initial Cost Estimate**

6. CMR Playback Phase

During the "Playback" phase, cloud candidate configurations and cost estimates are finalized. The complexity of assessing the migration of a large spectrum of on-premise workloads is dramatically reduced.

6.1 Playback Feature

To determine if a workload is fit to run in the cloud, the profile (or representative characteristics) of the workload are played back at a high fidelity. Performance is compared to the on-premises levels and cost estimates are re-calculated. The process is iterative until a cloud configuration that meets the performance criteria while minimizing cost is found. That is, the goal is to determine a cloud configuration that meets the original on-premises levels at the lowest cost. Optionally, organizations may be interested in testing higher levels of utilization or performance beyond those of the on-premises baseline. Higher levels of utilization can be executed as multiple of 2xCPU or 4x Memory Utilization etc.

A key benefit of the CMR Playback feature is the ability to evaluate migration before the actual application is migrated. A cloud configuration may meet all the performance criteria. However, in some cases, the cost to achieve such high-performance demands may result in a very expensive option. Worst, there are cases where the CSP compute and storage options may not be able to meet the resource demands. This is particularly true for workloads with very high IO intensive characteristics. Throughput, IOPS, and latency depend on each other. Typically, MB/s are impacted by the number of IOPs. The larger the number of IOPS the smaller the throughput gets. Assuming all other aspects in the infrastructure are equal, latency is impacted. How a CSP will handle the actual demand can only be revealed by playing back representative workloads or by the applications themselves.

Playing back a representative workload is accomplished by feeding the temporal workload metrics into the CMR playback engine.

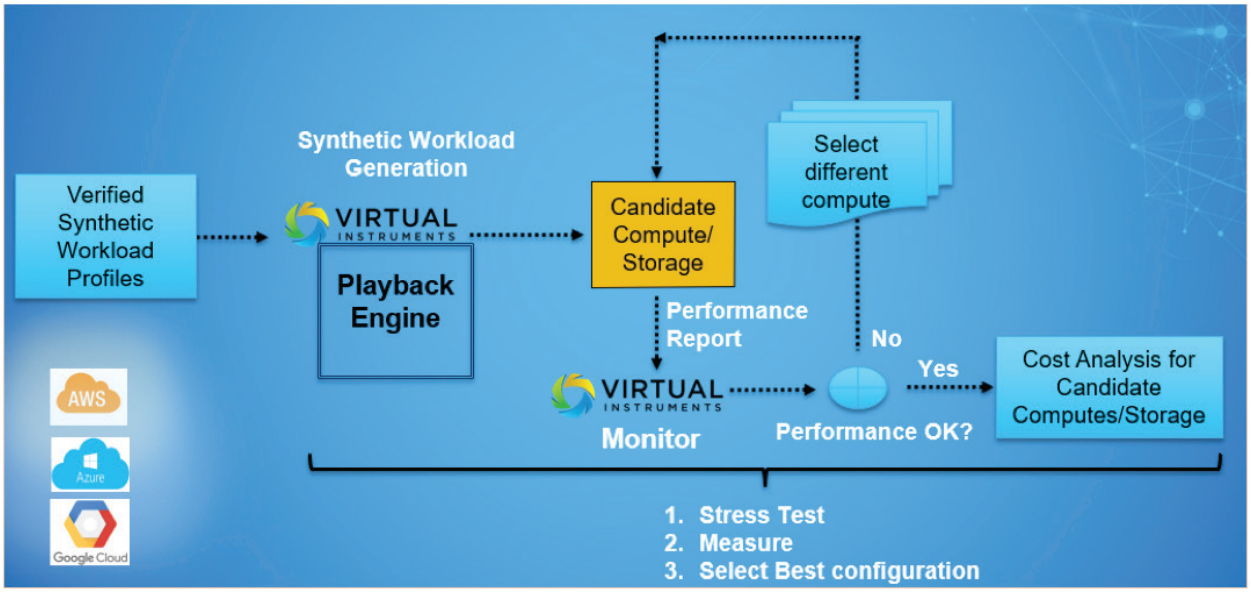


Figure 13: Playback Phase Process Flow

6.2 CMR's Playback Engine

The CMR playback engine was designed to emulate the key dimensions of a workload with high fidelity. The dimensions are:

- CPU MHz
- Memory Utilization GiB
- Read and Writes - IOPs
- Read and Writes Throughput - Bps
- Network Received and Transmit – Bps

The CMR playback engine consumes the representative workloads created during CMR Profiling phase. The CMR Playback engine runs on top of the operating system in the target cloud environment and configuration.

6.3 CMR's Playback Engine Output

The CMR Playback engine generates utilization levels matching, as much as possible, the on-premises profile levels. The on-premise profile of synthetic workloads is the reference that is compared to the synthetic workload profile in the cloud. CMR applies adapted linear correlation and relative error to compare cloud values against on-premises values (the reference).

The figures below show visualization comparing two different CSPs against the on-premises values. In one case, the Workload is fit for migration to both CSPs. In the second case, the Workload is unfit to migrate to the cloud due to the network demand not being met by the CSP.

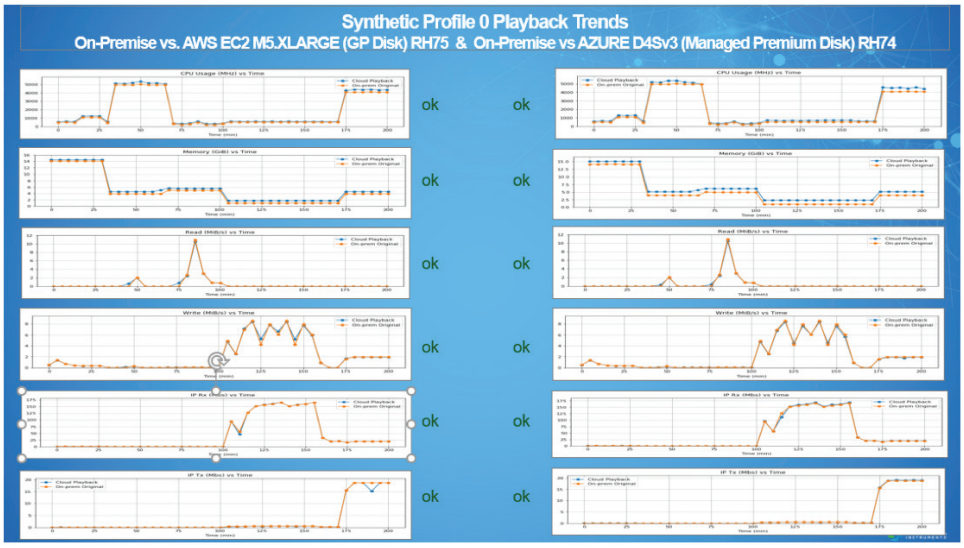


Figure 14: Example comparing two CSPs - On-premise vs Cloud playback – Fit to move to the Cloud

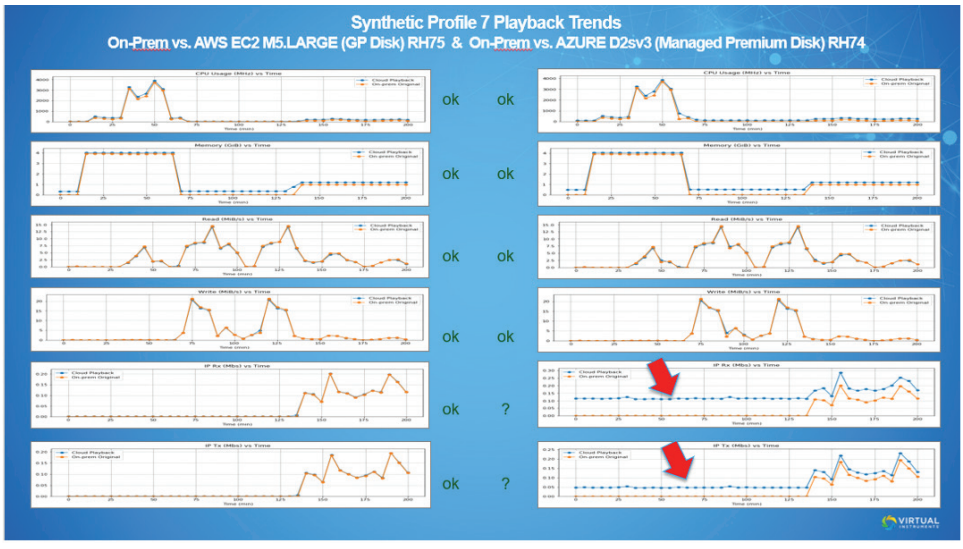


Figure 15: Two CSPs comparison - On-premise vs Cloud playback – Suboptimal fit.

6.4 CMR Playback Deliverables

During the CMR Playback phase the following deliverables are produced:

- **Final candidate cloud configuration**
- **Final cloud cost estimate**

Application	Virtual Machines	Azure Instance	Azure Disk Type	Instance Monthly Cost	App Monthly Cost
Analysis	VM014-523-2TB VM016-523-1TB	L8s D4sv3	P30 P15	$449. 28 $276.48	$725.76
Management	VM001-523-64gig	F16v2	P10	$937.44	$937.44
Automation	VM027-522-HF1	L4s	P40	$224.64	$224.64
WinTel1	Del-030318	D2sv3	P20	$69.12	$69.12
⋮					
WinTel16	Del-030318	D2sv3	P20	$69.12	$69.12
Grand Total					**$25,407**

Figure 16: CMR Playback Phase – Example of final monthly cloud cost estimate for one CSP

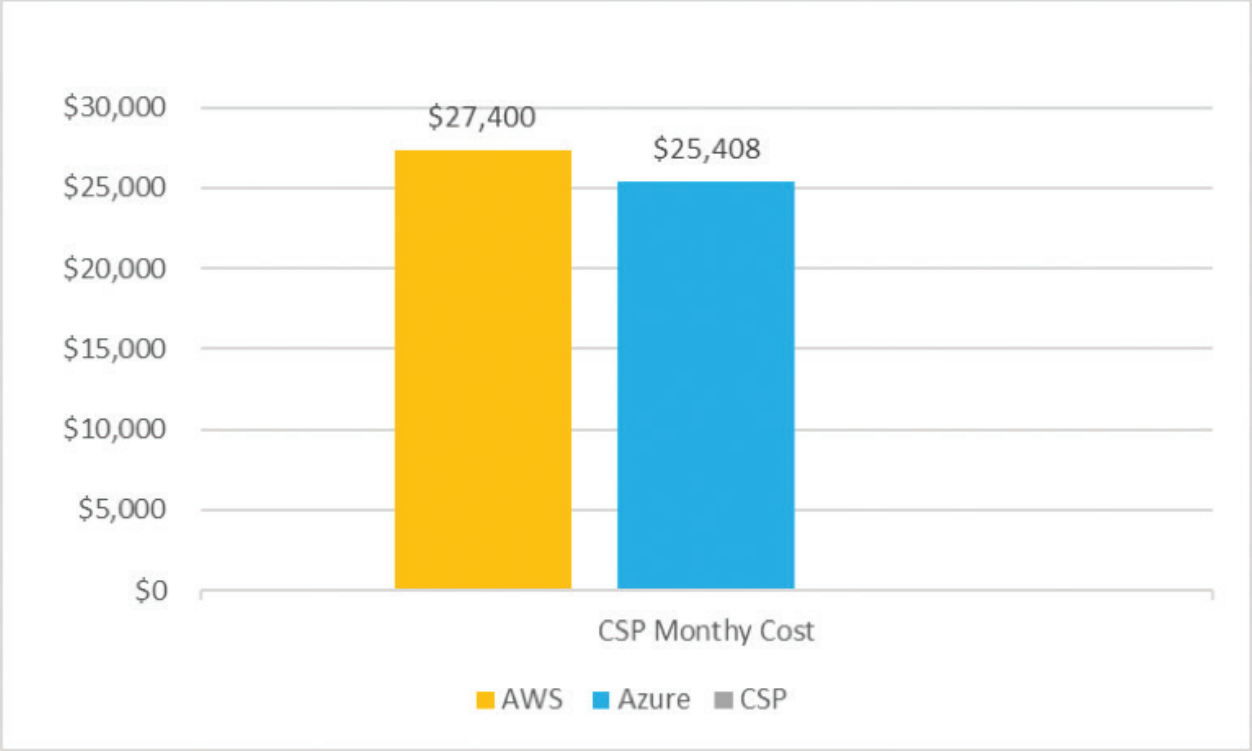


Figure 17: Example of final monthly cloud cost estimate comparison for two CSPs

7. CMR Monitoring Phase

The monitoring phase starts after the actual "Lift and Shift" has taken place. The applications and their corresponding data have been successfully migrated to the cloud by a 3rd party data migration service provider.

During the CMR monitoring phase, Virtana's VirtualWisdom and/or native CSP monitoring tools are used to verify that the applications' performance remains within acceptable levels.

As leader in Application- Centric performance monitoring, VirtualWisdom offers the advantage of monitoring on-premise and cloud applications simultaneously within the same platform – ideal for Hybrid cloud environments.

Monitoring should run for at least one month. However, options are available for longer monitoring periods or a managed service arrangement.

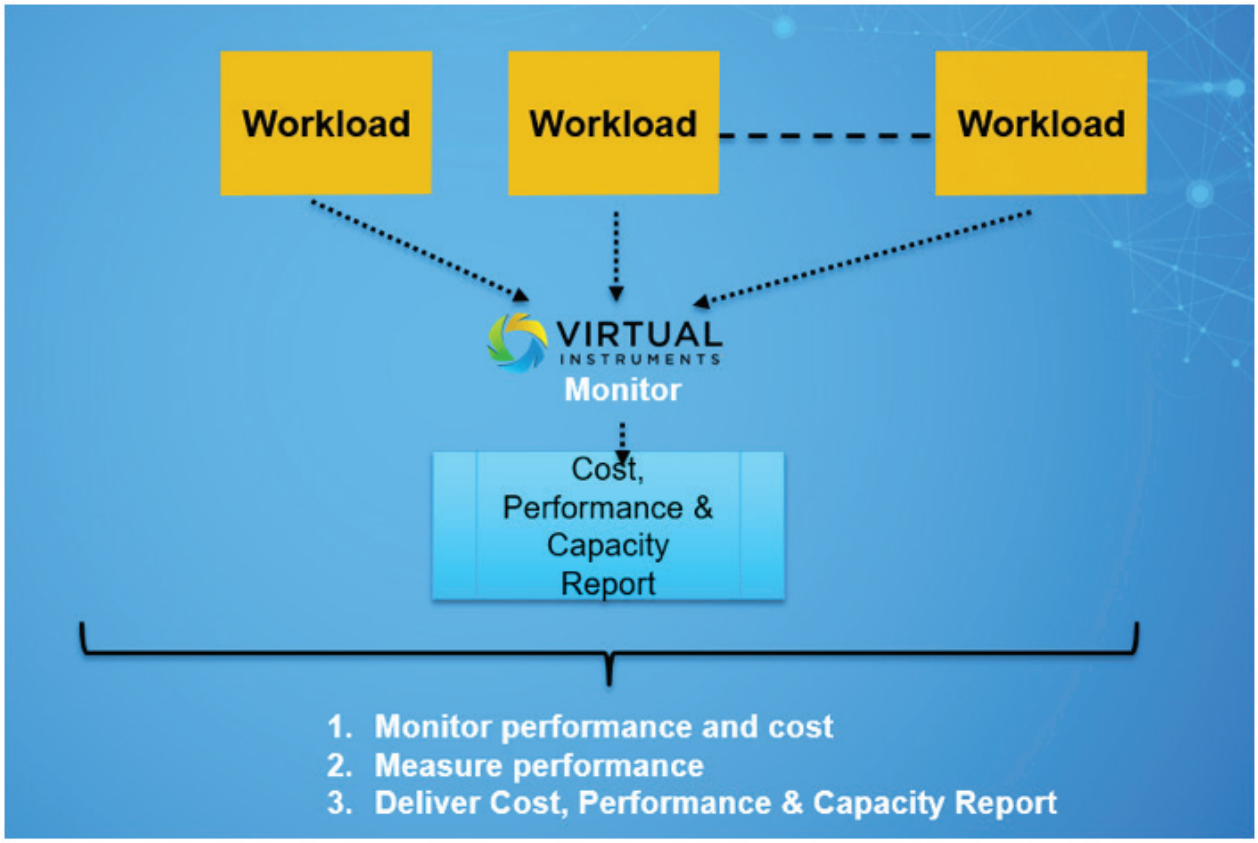


Figure 18: CMR Monitoring Phase

7.1 Utilization and Performance Report

Once the monitoring data has been collected for the agreed period of time, CMR analysts perform an assessment and produce a Utilization and Performance report. The report findings include performance trend deviations and recommendations related to remediation of issues and optimization opportunities.

7.2 CMR Monitoring Deliverables

During the CMR Monitoring phase the following deliverable is produced:

- Utilization and Performance Report

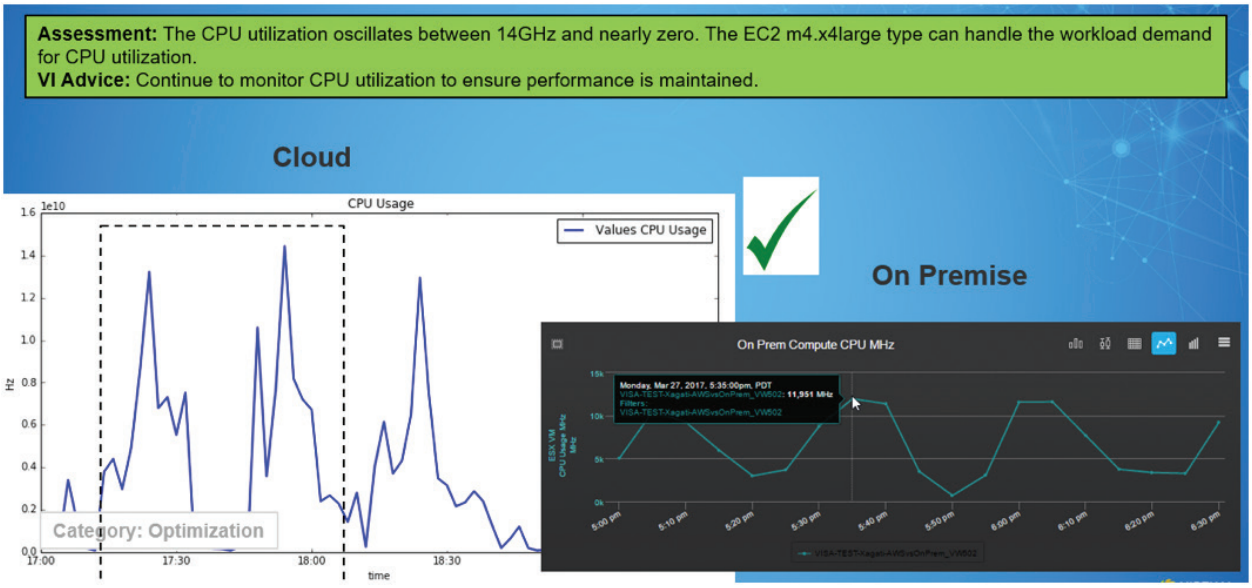


Figure 19: CMR Monitoring – CPU utilization assessment example

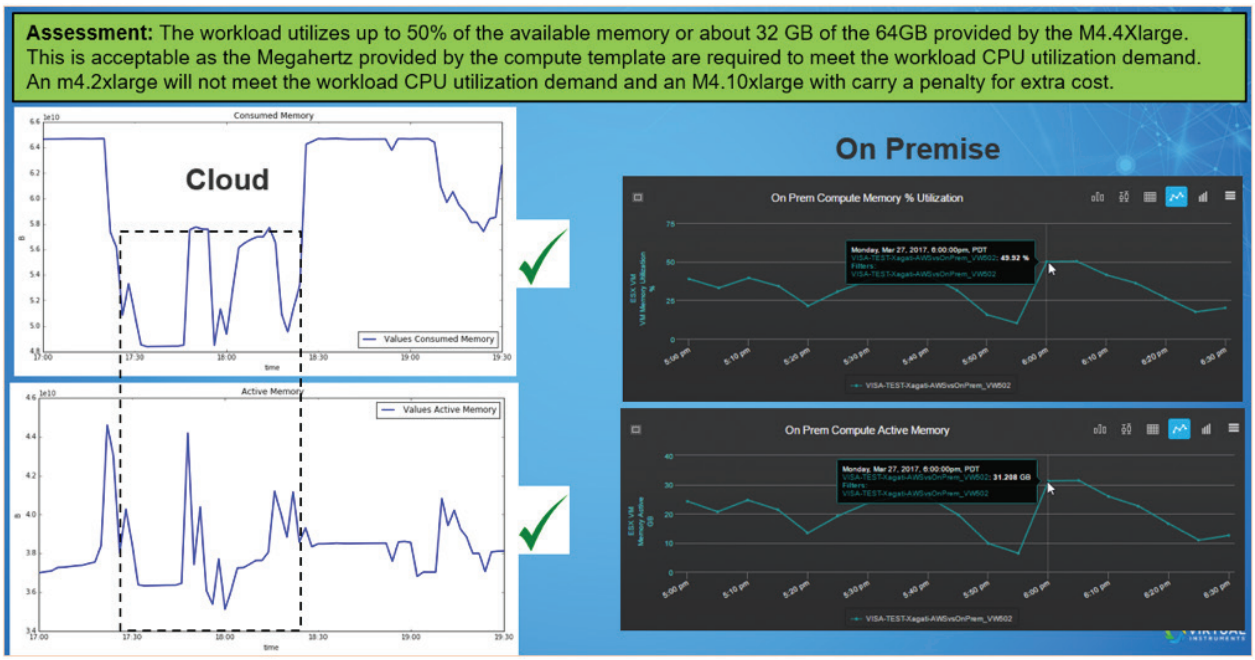


Figure 19: CMR Monitoring – CPU utilization assessment example

8. Conclusion

Once the business transformation project has been initiated and the candidate applications or workloads have been selected, the first step in a "lift and shift" project can take place – a Cloud Migration Readiness assessment.

The initial step is to answer the three fundamental questions (Fitness, Cost and CSP choice) before migrating the actual applications to the cloud or embarking onto the next critical and costly phases.

Understanding the inventory, dependencies and the performance baseline of the targeted workloads is imperative. Some applications may not be fit for migration to the public cloud based on their dependencies to on-premises services alone. Data is iteratively analyzed particularly for dependency mapping.

When dealing with hundreds or thousands of workloads, "Affinity Groups" offer the best way to significantly reduce the analysis allowing the creation of a small, manageable number of "Representative" synthetic workloads.

The playback of the synthetic workloads in the cloud, reveals if the applications are fit for performance in the cloud or rather, whether the CSP can handle the applications resource demands. The playback also confirms the optimal cloud configuration required and the estimated cost.

It is important to understand the value of a service over buying tools for the purposes of "lift and shift" cloud migrations. The Virtana CMR service is delivered by industry experts in workload analytics and infrastructure performance assessments. They are uniquely qualified experts to perform pre and post cloud migration assessments.

The CMR Service provides critical insight into business transformation decisions. Cloud projects are fluid and require an interactive approach - they require expert workload profiling advice - not advice on how to use a tool's features.

The deliverables are reused by data movers or cloud architects.

Finally, it is highly recommended to engage partners in a cloud migration project – don't do it alone. A reputable partner will understand your business goals and offer expert advice on how to achieve them in the cloud.

Whether you are exploring the cloud or have a current cloud initiative, the CMR service will work with your partners or directly for you. Either way, CMR will put you in control of your "lift and shift" cloud migration.


Learn more about VirtualWisdom


Contact a Virtana Expert


View more Resources

Virtana
2331 Zanker Road
San Jose, CA 95131
Phone: +1.408.579.4000

©10/2019 Virtana. All rights reserved. WorkloadWisdom and VirtualWisdom are trademarks or registered trademarks in the United States and/or in other countries. All other trademarks and trade names are the property of their respective holders.