SYMPOSIUM'16
AMERICA – EUROPE – ASIA
THE ASIAN SPEC SYMPOSIUM ON SERVER EFFICIENCY

Measuring Cloud Performance

SPEC Cloud IaaS 2016

- **SPEC's first benchmark suite to measure infrastructure-as-a-service (IaaS) performance on either public or private cloud environments**
- **Targets cloud providers, cloud consumers, hardware vendors, virtualization software vendors, application software vendors, and academic researchers**

Main Features

- **No restriction on how a cloud is configured, no hypervisor or virtualization layer is required**
- **Supports optional multi-tenancy and strict Quality of Service (QoS) metrics**
- **Stresses the provisioning and run-time of a cloud with multiple multi-instance workloads**
- **Uses I/O and CPU intensive workloads that run in multi-instance configuration. Each multi-instance application cluster is known as an application instance.**
- **Allows any instance types, that is a baremetal, container, or a virtual machine, with a particular configuration. Instance types cannot be changed during a compliant run.**
- **NoSQL database transaction workload (YCSB (one instance) / Cassandra cluster (six instances))**
- **K-Means clustering using map/reduce (Hadoop master (one instance) / Hadoop data nodes (five instances)**

Metrics

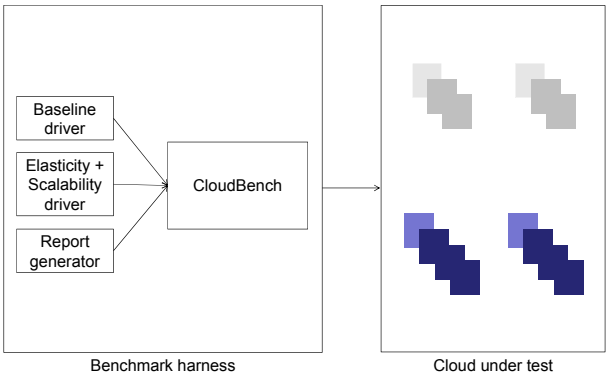
- **Scalability** - Total amount of work performed by application instances running in a cloud
- **Elasticity** - Measures whether the work performed by application instances scales linearly in a cloud when compared to the performance of application instances during baseline phase.
- **Mean Instance Provisioning Time** - Time interval between the instance provisioning request and connectivity to the instance. This metric is an average across all instances in valid application instances.

Baseline phase

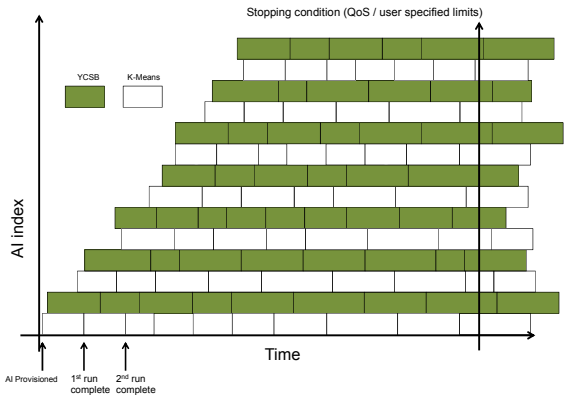
- Determine the performance of a single application instance for determining QoS criteria.
- Only a single application instance is run in the cloud (white-box) or in the user account (black-box)

Elasticity + Scalability phase

- Instantiates multiple application instances for two workloads over time. The workloads use data that is statistically similar
- Stops when QoS criteria as determined from baseline is not met, or maximum AIs as set by a tester are reached
- Scalability results are normalized against a reference platform



SPEC Cloud IaaS 2016 Benchmark Setup



Elasticity and Scalability Phase

Contact: info@spec.org