Performance Testing Minutes

Agenda

1. Welcome / Intro - Lincoln / Rabi / Trevor (5 minutes) - set expected outcomes
2. Update on where CNTT is todate - Rabi (10 minutes)
3. Definition of performance - working session to document performance terms that will be used in RC outputs and OVP - Group (15 minutes)
4. Performance Testing: CNTT RC and OVP2.0 Performance Testing Requirements - Trevor (30 minutes)

Attendees

1. Lincoln Lavoie
2. Rabi Abdel
3. Trevor Cooper(Intel)
4. Al Morton
5. eth Cohen
6. Walter Kozlowski
7. Petar Torre (Intel)
8. Gergely Csatari (Nokia)
9. Karine Sevilla (Orange)
10. Toshiyasu Wakayama (KDDI)
11. Bob Monkman (Intel)
12. Padmakumar Subramani (Nokia)
13. Carlo Cavazzoni (TIM / Telecom Italia)
14. Georg Kunz (Ericsson Software Technology)
15. Ahmed El Sawaf (STC)
16. Sridhar Rao
17. Parth Yadav
18. Mark Beierl
19. Michael Pedersen
20. Heather Kirksey
21. Pankaj Goyal (AT&T)

Minutes

Minutes of June 23 7am EDT session "Joint Topic: Performance"

- CNTT_vF2F_June_Performance-edits TCAMv02.pptx
- Work plan - use the next CVC calls (July 6) to begin documenting the agreed terms for performance. Then map how these terms apply within the context of the CNTT RC definitions.

Definitions of Performance

Calibration:

The process of checking and/or adjusting a stimulus generation or measurement device with a known reference value, to improve the overall quality of the measured results. Calibration may be very simple, such as a comparison of the configured traffic generator sending rate and measured rate using a simple SUT (loop-back cable between interfaces), such that that the known reference value is the published nominal interface rate.

Reference Value:

A measured or established result or outcome for comparison with new measurements. For example, the reference value or expected outcome of a Functional Test is "PASS". The reference value or expected outcome of a Performance Measurement or Benchmarking test may be the value measured for the previous SUT release, or the published value or theoretical limit of a simple SUT.

API Testing:

Testing against a protocol specification for conformance,

(Discussion: 1000's of tests possible for a complex API. Are these tests meaningful? Where do we draw the line on meaningful API tests for a Conformance test program? ETSI NFV SOL/MANO testing includes causation == behavior).

Functional Testing:

The main object of functional testing is the verification of a specific stimulus / response within the SUT, **including the expected behavior.** These tests generally result in a binary outcome, i.e. pass / fail. For example, verification of an "API call" and its associated response, such as the instantiation of a VM (or container) and verification of the VM's (or container's) existence (expected behavior), or the ability to activate a specific feature of the SUT (i.e. SR-IOV).

Performance Measurement: <lead to benchmarking more explicitly>

The procedure or set of operations having the objective of determining a Measured Value or Measurement Result of an infrastructure in operation according to a defined metric.

In the context of telemetry, Performance Measurements reflect data generated and collected within the cloud infrastructure, that reflects a performance aspect of the cloud infrastructure. For example, a count of frames or packets traversing an interface per unit of time, memory usage information, other resource usage and availability, etc. This data may be instantaneous or accumulated, and made available (i.e. exposed) based on permissions and contexts (e.g., workload vs. infra).

Other Performance Measurements are designed to assess the efficiency of SUT Functions, such as the time to successfully instantiate one or more VMs or containers, or the percentage of instantiation Function failures in a set of many attempts.

Still other Performance Measurements are conducted under controlled conditions using Calibrated test systems, such that the measured results are more likely to comparable with other such measurements.

Performance Testing:

The main objective of performance testing is to understand if the SUT is able to achieve the expected performance, through conducting a series of performance measurements and comparing those results against a reference value.

Operators need to do performance testing to help dimension a solution or to assess that a platform (particular hardware + software combination) is configured correctly and performing as expected (as compared with capacity / performance claims made by the infrastructure and VNF/CNF vendors).

Performance Testing may be useful to compare infrastructure capabilities between a particular SUT and a CNTT reference implementation with well understood known good configurations and previously established performance ranges. Performance testing for the purpose of comparing between different commercial implementations is not a goal of CNTT and hence out of scope for the purposes of this definition. Performance testing relies on well established benchmark specifications to help establish appropriate methodologies and accuracy tolerances.

Benchmarking:

Benchmarking is a type of performance test that assesses a key aspect of the computing environment in its role as the infrastructure for network functions, using calibrated test systems and controlled conditions. In general the benchmark testing attempts to isolate the feature or parameter under test, to reduce the impact of other system components or operations on the test result. The benchmarks (and related metrics) have been agreed by the Industry and documented in publications of an accredited standards body. As a result, benchmarks are a sub-set of all possible performance tests and metrics i.e. they are selected measurements which are more important than others. Example benchmarks include Zero-loss Throughput, Latency, and Loss ratio (ref to ETSI NFV TST009, RFC 2544) of various components of the environment, expressed in quantitative units to allow direct comparison between different systems treated as a black box (vendor-independence).

Because the demands on a particular system may vary from deployment to deployment, benchmarking assessments do not define acceptance criteria or numerical performance requirements.

Benchmark testing and conformance testing intersect when a specific requirement in the reference architecture specification is very important to the performance of the system. Correct execution of the performance test with the valid result constitutes conformance. The completion time for a single conforming execution, or the number of conforming executions per second are potential benchmark metrics, and sources of known reference values.

<picture of the hierarchy>

Bin list of related terms / topics

- Threshold -
- Measurement Performance of functions
- Exception Threshold
- Computation
- - CPU
  - GPU
- Life cycle
  - Time to create VM / container
  - Time to delete VM / container
  - Maximum number of VMs / containers
- SUT = System Under Test

Related CNTT issues / pull requests

[RC2 Ch01] Terminology definition for testing per issue #1495 #1649 https://github.com/cntt-n/CNTT/pull/1649

- Once terminology agreed will update this pull request to align