



CeTune -- Benchmarking and tuning your Ceph cluster

CHENDI.XUE@INTEL.COM



CeTune Overview

- What is CeTune?
 - CeTune is a toolkit/framework to deploy, benchmark, analyze and tuning ceph.
- CeTune's objective?
 - For beginners: Shorten landing time of Ceph based storage solution.
 - For performance engineer: Simplify the procedure to deploy, tune ceph, easily finding device bottlenecks, identify unexpected software behavior from processed data.
 - For Developers: Providing an easy way to verify codes with a quick stress test and a clear performance report.

CeTune is open sourced today!!

- Github:

- <https://github.com/01org/CeTune>



- License:

- Apache License v2.0

- Main feature:

- Manage CeTune Configuration and Execute CeTune with Webui
 - Deploy module: install ceph with CeTune Cli, deploy ceph with Webui
 - Benchmark module: support qemurbd, fiorbd, cosbench
 - Analyzer module: support iostat, sar, interrupt, performance counter
 - Report Visualize: Support Config download, csv download, all data present by line chart.

- Maillist:

- maillist: cephperformance@lists.01.org
 - Subscribe maillist: <https://lists.01.org/mailman/listinfo/cephperformance>

CeTune Functionality Details

CeTune WebUI

- Configuration view: In this view, user can add cluster description, ceph configuration, benchmark testcase, etc. Then click 'Execute' to kickoff CeTune.

Test Configuration

CeTune Status

Result Reports

17-4 ...

CeTune – Ceph tuning and profiling

CeTune Status: idle Ceph Status: health HEALTH_OK

CeTune

☐ Deploy

✓ Cluster Configuration

✓ Ceph Configuration

☒ Benchmark

✓ Executvie Configuration

✓ System Configuration

✓ Ceph Tuning

✓ Benchmark Configuration

✓ Analyzer Configuration

Execute

Cluster Configuration

Add

Delete

| <input type="checkbox"/> | Key | Value |
|--------------------------|---------------------------|---|
| <input type="checkbox"/> | clean_build | true |
| <input type="checkbox"/> | head | client01 |
| <input type="checkbox"/> | user | root |
| <input type="checkbox"/> | enable_rgw | true |
| <input type="checkbox"/> | aceph02 | /dev/sdb1:/dev/sdg1:/dev/sdc1:/dev/sdg2:/dev/sdd1:/dev/sdg3:/dev/sde1:/dev/sdg4:/dev/sdh1:/dev/sdg5:/dev/sdi1:/dev/sdf1:/dev/sdj1:/dev/sdf2:/dev/sdk1:/dev/sdf3:/dev/sdl1:/dev/sdf4:/dev/sdm1:/dev/sdf5 |
| <input type="checkbox"/> | aceph03 | /dev/sdb1:/dev/sdg1:/dev/sdc1:/dev/sdg2:/dev/sdd1:/dev/sdg3:/dev/sde1:/dev/sdg4:/dev/sdh1:/dev/sdg5:/dev/sdi1:/dev/sdf1:/dev/sdj1:/dev/sdf2:/dev/sdk1:/dev/sdf3:/dev/sdl1:/dev/sdf4:/dev/sdm1:/dev/sdf5 |
| <input type="checkbox"/> | aceph04 | /dev/sdc1:/dev/sda1:/dev/sdd1:/dev/sda2:/dev/sde1:/dev/sda3:/dev/sdf1:/dev/sda4:/dev/sdg1:/dev/sda5:/dev/sdh1:/dev/sdb1:/dev/sdj1:/dev/sdb2:/dev/sdk1:/dev/sdb3:/dev/sdl1:/dev/sdb4:/dev/sdm1:/dev/sdb5 |
| <input type="checkbox"/> | list_server | aceph02,aceph03,aceph04 |
| <input type="checkbox"/> | list_client | client01,client02,client03,client04 |
| <input type="checkbox"/> | list_mon | aceph02 |
| <input type="checkbox"/> | rgw_num_per_server | 5 |
| <input type="checkbox"/> | cosbench_auth_username | cosbench:operator |
| <input type="checkbox"/> | cosbench_controller_proxy | |
| <input type="checkbox"/> | rgw_start_index | 1 |
| <input type="checkbox"/> | cosbench_auth_password | intel2012 |
| <input type="checkbox"/> | rgw_server | rgw |

True: destroy ceph cluster and build
False: compare current cluster with configuration, incremental build

Deploy radosgw only when this is 'True'

CeTune WebUI

- Configuration view: In this view, user can add cluster description, ceph configuration, benchmark testcase, etc. Then click 'Execute' to kickoff CeTune.

The screenshot displays the CeTune WebUI interface. At the top, there are tabs for 'Test Configuration', 'CeTune Status', and 'Result Reports'. Below the tabs, a blue header bar shows 'CeTune – Ceph tuning and profiling' on the left and 'CeTune Status: idle Ceph Status: health HEALTH_OK' on the right. On the left side, there is a sidebar with a 'CeTune' section containing four items: 'Deploy' (checked), 'Cluster Configuration' (with a green checkmark), 'Ceph Configuration' (with a green checkmark and a red vertical bar to its left), and 'Benchmark' (unchecked). Below the sidebar is a blue 'Execute' button. The main content area is titled 'Ceph Configuration' and features a table with three columns: 'Key', 'Value', and 'Description'. The table contains four rows of configuration data. To the right of the table are 'Add' and 'Delete' buttons. A yellow callout box with a pointer to the 'Ceph Configuration' section contains the text: 'Ceph configurations need to set before deploy'.

| <input type="checkbox"/> | Key | Value | Description |
|--------------------------|-----------------|-------------------------|-------------|
| <input type="checkbox"/> | mon_data | /var/lib/ceph/ceph.\$id | |
| <input type="checkbox"/> | osd_objectstore | filestore | |
| <input type="checkbox"/> | public_network | 10.10.5.0/24 | |
| <input type="checkbox"/> | cluster_network | 10.10.5.0/24 | |

Ceph configurations need to set before deploy

CeTune WebUI

- Configuration view: In this view, user can add cluster description, ceph configuration, benchmark testcase, etc. Then click 'Execute' to kickoff CeTune.

Test Configuration

CeTune Status

Result Reports

CeTune – Ceph tuning and profiling

CeTune Status: idle Ceph Status: health HEALTH_OK

CeTune

☒ Deploy

☐ Benchmark

✓ Executvie Configuration

✓ System Configuration

✓ Ceph Tuning

✓ Benchmark Configuration

✓ Analyzer Configuration

Execute

Executvie Configuration

Add

Delete

| <input type="checkbox"/> | Key | Value | Description |
|--------------------------|------------|--------|-------------|
| <input type="checkbox"/> | workstages | deploy | |

CeTune WebUI

- Configuration view: In this view, user can add cluster description, ceph configuration, benchmark testcase, etc. Then click 'Execute' to kickoff CeTune.

CeTune – Ceph tuning and profiling CeTune Status: idle Ceph Status: health HEALTH_OK

System Configuration Add Delete

| <input type="checkbox"/> | Key | Value | Description |
|--------------------------|--------------------|-------|-------------|
| <input type="checkbox"/> | disk read_ahead_kb | 2048 | |

System configuration, including disk read_ahead, scheduler, sector size

Execute

CeTune WebUI

- Configuration view: In this view, user can add cluster description, ceph configuration, benchmark testcase, etc. Then click 'Execute' to kickoff CeTune.

The screenshot displays the CeTune WebUI interface. At the top, there are tabs for 'Test Configuration', 'CeTune Status', and 'Result Reports'. Below these, a blue header bar shows 'CeTune – Ceph tuning and profiling' on the left and 'CeTune Status: idle Ceph Status: health HEALTH_OK' on the right. On the left side, a sidebar menu lists various configuration options: 'Deploy' (checked), 'Benchmark' (unchecked), 'Executive Configuration' (checked), 'System Configuration' (checked), 'Ceph Tuning' (checked and highlighted with a red border), 'Benchmark Configuration' (checked), and 'Analyzer Configuration' (checked). The main content area is titled 'Ceph Tuning' and contains a table with three columns: 'Key', 'Value', and 'Description'. The table lists three configurations: 'global|osd_op_threads' with a value of 20, 'global|mon_pg_warn_max_per_osd' with a value of 1000, and 'pool|rbd|size' with a value of 2. The row for 'global|mon_pg_warn_max_per_osd' is highlighted in yellow. To the right of the table are 'Add' and 'Delete' buttons. An orange callout box points to the table with the text: 'Ceph tuning, pool configuration and ceph.conf tuning. When CeTune start to run benchmark, will firstly compared ceph tuning with this configuration, apply tuning if needed then start test'. At the bottom left, there is a large blue 'Execute' button.

CeTune – Ceph tuning and profiling

CeTune Status: idle Ceph Status: health HEALTH_OK

CeTune

- ☒ Deploy
- ☐ Benchmark
- ✓ Executive Configuration
- ✓ System Configuration
- ✓ Ceph Tuning
- ✓ Benchmark Configuration
- ✓ Analyzer Configuration

Execute

Ceph Tuning

| <input type="checkbox"/> | Key | Value | Description |
|--------------------------|--------------------------------|-------|-------------|
| <input type="checkbox"/> | global osd_op_threads | 20 | |
| <input type="checkbox"/> | global mon_pg_warn_max_per_osd | 1000 | |
| <input type="checkbox"/> | pool rbd size | 2 | |

Add Delete

Ceph tuning, pool configuration and ceph.conf tuning
When CeTune start to run benchmark, will firstly compared ceph tuning with this configuration, apply tuning if needed then start test

CeTune WebUI

- Configuration view: In this view, user can add cluster description, ceph configuration, benchmark testcase, etc. Then click 'Execute' to kickoff CeTune.

Test Configuration

CeTune Status

Result Reports

CeTune -- Ceph tuning and profiling

CeTune Status: idle Ceph Status: health HEALTH_OK

CeTune

☒ Deploy

☐ Benchmark

✓ Executvie Configuration

✓ System Configuration

✓ Ceph Tuning

✓ Benchmark Configuration

✓ Analyzer Configuration

Execute

Benchmark Configuration

AddDelete

| <input type="checkbox"/> | Key | Value | Description |
|--------------------------|---------------------|----------------------|-------------|
| <input type="checkbox"/> | tmp_dir | /opt/ | |
| <input type="checkbox"/> | dest_dir | /mnt/data1/ | |
| <input type="checkbox"/> | cache_drop_level | 1 | |
| <input type="checkbox"/> | monitoring_interval | 1 | |
| <input type="checkbox"/> | fio_capping | false | |
| <input type="checkbox"/> | volume_size | 40960 | |
| <input type="checkbox"/> | rbd_volume_count | 1 | |
| <input type="checkbox"/> | disk_num_per_client | 35,35,35,35 | |
| <input type="checkbox"/> | rwmixread | 100 | |
| <input type="checkbox"/> | cosbench_version | v0.4.2.c2 | |
| <input type="checkbox"/> | cosbench_folder | /opt/cosbench | |
| <input type="checkbox"/> | cosbench_config_dir | /opt/cosbench_config | |
| <input type="checkbox"/> | cosbench_cluster_ip | 10.10.5.5 | |
| <input type="checkbox"/> | cosbench_admin_ip | 192.168.5.1 | |
| <input type="checkbox"/> | cosbench_network | 192.168.5.0/24 | |

Benchmark configuration

AddDelete

| <input type="checkbox"/> | benchmark_driver | worker | container_size | iopattern | opsize | object_size/QD | rampup | runtime | device |
|--------------------------|------------------|--------|----------------|-----------|--------|----------------|--------|---------|----------|
| <input type="checkbox"/> | qemurbd | 4 | 40g | seqwrite | 64k | 64 | 100 | 400 | /dev/vdb |
| <input type="checkbox"/> | cosbench | 160 | r(1,100) | write | 128KB | r(1,100) | 0 | 400 | cosbench |

Benchmark testcase configuration

CeTune WebUI

- Configuration view: In this view, user can add cluster description, ceph configuration, benchmark testcase, etc. Then click 'Execute' to kickoff CeTune.

The screenshot shows the CeTune WebUI interface. At the top, there are tabs for 'Test Configuration', 'CeTune Status', and 'Result Reports'. Below the tabs, a blue header bar displays 'CeTune – Ceph tuning and profiling' on the left and 'CeTune Status: idle Ceph Status: health HEALTH_OK' on the right. The main content area is titled 'Analyzer Configuration' and features a table with columns 'Key', 'Value', and 'Description'. The table contains one entry: 'analyzer' with the value 'all'. To the left of the table is a sidebar with a 'CeTune' section containing checkboxes for 'Deploy' (checked), 'Benchmark', and several configuration items with green checkmarks: 'Executvie Config', 'System Configur', 'Ceph Tuning', 'Benchmark Configuration', and 'Analyzer Configuration'. An 'Execute' button is located at the bottom left. Three yellow callout boxes provide instructions: one points to the 'Deploy' and 'Benchmark' checkboxes with the text 'Select deploy, benchmark'; another points to the 'analyzer' row in the table with the text 'System metrics: iostat, sar, interrupt Perf counter process'; and a third points to the 'Execute' button with the text 'Click to execute CeTune'.

CeTune

- ☒ Deploy
- ☐ Benchmark
- ✓ Executvie Config
- ✓ System Configur
- ✓ Ceph Tuning
- ✓ Benchmark Configuration
- ✓ Analyzer Configuration

Analyzer Configuration

| Key | Value | Description |
|----------|-------|-------------|
| analyzer | all | |

Execute

Select deploy, benchmark

System metrics: iostat, sar, interrupt
Perf counter process

Click to execute CeTune

CeTune WebUI

- Status Monitoring View: Reports current CeTune working status, so user can interrupt CeTune if necessary.

The screenshot displays the CeTune WebUI interface. At the top, there is a navigation bar with tabs: "Test Configuration", "CeTune Status" (which is selected), "Result Reports", and "2-20 ...". Below the navigation bar, a header bar shows "CeTune -- Ceph tuning and profiling" on the left and "CeTune Status: idle Ceph Status: health HEALTH_OK" on the right. On the left side of the main content area, there is a sidebar with a "CeTune" section containing a "CeTune Status" link. The main content area displays a log of system events and warnings. The logs include timestamps, log levels (WARNING, ERROR, LOG), and messages regarding Radosgw access, Ceph cluster health, and benchmark execution. The logs show a sequence of events from 2015-09-26T01:30:13 to 2015-09-26T01:31:38, indicating that the benchmark is still running.

CeTune Status: idle Ceph Status: health HEALTH_OK

CeTune

CeTune Status

[2015-09-26T01:30:13.065478][WARNING]: Radosgw is not able to be accessed by swift interface yet, need to wait, will time out in 18
[2015-09-26T01:30:14.089116][ERROR]: Cosbench connect to Radosgw Connection Failed
[2015-09-26T01:30:14.089218][WARNING]: Radosgw is not able to be accessed by swift interface yet, need to wait, will time out in 17
[2015-09-26T01:30:15.109063][LOG]: Radosgw now is working
[2015-09-26T01:30:15.402126][LOG]: Tuning has applied to ceph cluster, ceph is Healthy now
[2015-09-26T01:30:18.838180][LOG]: RUNID: 17, RESULT_DIR: //mnt/data1//17-4-qemurbd-seqwrite-64k-qd64-40g-100-400-vdb
[2015-09-26T01:30:18.838385][LOG]: Prerun_check: check if rbd volume be initialized
[2015-09-26T01:30:19.144237][WARNING]: Ceph cluster used data occupied: 250609664.0 KB, planned_space: 167772160.0 KB
[2015-09-26T01:30:19.144348][LOG]: Prerun_check: check if fio installed in vclient
[2015-09-26T01:30:19.466931][LOG]: Prerun_check: check if rbd volume attached
[2015-09-26T01:30:19.766539][WARNING]: vclients are not attached with rbd volume
[2015-09-26T01:30:19.766657][LOG]: Attach rbd image to vclient106
[2015-09-26T01:30:20.649765][LOG]: Attach rbd image to vclient01
[2015-09-26T01:30:21.190023][LOG]: Attach rbd image to vclient71
[2015-09-26T01:30:22.072345][LOG]: Attach rbd image to vclient36
[2015-09-26T01:30:22.980489][WARNING]: vclients attached rbd volume now
[2015-09-26T01:30:22.980574][LOG]: Prerun_check: check if sysstat installed
[2015-09-26T01:30:31.517288][LOG]: Prepare_run: distribute fio.conf to vclient
[2015-09-26T01:30:34.185856][LOG]: Run Benchmark Status: collect system metrics and run benchmark
[2015-09-26T01:30:34.186019][LOG]: This test will run 500 secs until finish.
[2015-09-26T01:30:54.524887][WARNING]: 8 fio job still running
[2015-09-26T01:30:54.525020][LOG]: FIO Jobs starts on [vclient106', 'vclient01', 'vclient71', 'vclient36]
[2015-09-26T01:30:54.826976][WARNING]: 8 fio job still running
[2015-09-26T01:31:00.133062][WARNING]: 8 fio job still running
[2015-09-26T01:31:05.460242][WARNING]: 8 fio job still running
[2015-09-26T01:31:12.326243][WARNING]: 8 fio job still running
[2015-09-26T01:31:17.660522][WARNING]: 8 fio job still running
[2015-09-26T01:31:23.000393][WARNING]: 8 fio job still running
[2015-09-26T01:31:28.333579][WARNING]: 8 fio job still running
[2015-09-26T01:31:33.661189][WARNING]: 8 fio job still running
[2015-09-26T01:31:38.999589][WARNING]: 8 fio job still running

CeTune WebUI

- Reports view: Result report provides two report view, summary view shows history report list, double click to view the detail report of one specific test run.

| Test Configuration CeTune Status Result Reports | | | | | | | | | | | | | | | |
|--|-------------|---------|----------|------|---------|-----------|-----------|--------|--|----------|----------|-------------|---------|-------------|----------------|
| CeTune -- Ceph tuning and profiling | | | | | | | | | CeTune Status: idle Ceph Status: HEALTH_OK | | | | | | |
| runid | Status | Op_Size | Op_Type | QD | Driver | SN_Number | CN_Number | Worker | Runtime(sec) | IOPS | BW(MB/s) | Latency(ms) | SN_IOPS | SN_BW(MB/s) | SN_Latency(ms) |
| 2 | completed | 512k | seqwrite | qd8 | florbd | 2 | 1 | 1 | 100 | 169.000 | 84.888 | 47.020 | 376.830 | 168.075 | 221.487 |
| 7 | Interrupted | 512k | seqwrite | qd8 | florbd | 2 | 1 | 1 | 100 | 0.000 | 0.000 | 0.000 | 312.606 | 142.987 | 325.264 |
| 8 | Completed | 512k | seqwrite | qd8 | florbd | 2 | 1 | 1 | 100 | 178.000 | 89.465 | 44.610 | 392.480 | 175.547 | 88.852 |
| 20 | Interrupted | 512k | seqwrite | qd8 | florbd | 2 | 1 | 1 | 60 | 0.000 | 0.000 | 0.000 | 332.069 | 140.860 | 175.668 |
| 21 | Completed | 512k | seqwrite | qd8 | qemurbd | 2 | 1 | 1 | 60 | 86.000 | 43.443 | 91.960 | 184.300 | 84.086 | 55.742 |
| 22 | Completed | 64k | seqwrite | qd64 | qemurbd | 2 | 1 | 1 | 100 | 1071.000 | 66.986 | 59.680 | 297.290 | 132.146 | 109.619 |
| 23 | Unknown | 64k | seqwrite | qd64 | qemurbd | 2 | 1 | 1 | 100 | 0.000 | 0.000 | 0.000 | 284.547 | 126.603 | 132.696 |
| 24 | Completed | 64k | seqwrite | qd64 | qemurbd | 2 | 1 | 2 | 100 | 1112.000 | 69.513 | 57.510 | 306.850 | 137.110 | 287.760 |
| 26 | Completed | 64k | seqwrite | qd64 | qemurbd | 2 | 1 | 1 | 100 | 1070.000 | 66.876 | 59.790 | 298.393 | 132.360 | 172.425 |
| 27 | Completed | 512k | seqwrite | qd8 | florbd | 2 | 1 | 1 | 100 | 167.000 | 83.585 | 47.750 | 385.080 | 164.420 | 171.115 |
| 29 | Completed | 64k | seqwrite | qd64 | qemurbd | 2 | 1 | 1 | 100 | 1132.000 | 70.779 | 56.490 | 317.650 | 140.303 | 61.132 |
| 30 | Unknown | 64k | seqwrite | qd64 | qemurbd | 2 | 1 | 1 | 100 | 0.000 | 0.000 | 0.000 | 210.434 | 102.087 | 56.139 |
| 31 | Unknown | 512k | seqwrite | qd8 | qemurbd | 2 | 1 | 1 | 100 | 0.000 | 0.000 | 0.000 | 176.449 | 80.089 | 62.904 |
| 34 | Completed | 512k | seqwrite | qd8 | generic | 2 | 1 | 1 | 100 | 353.000 | 176.767 | 22.550 | 0.000 | 0.000 | 0.000 |
| 35 | Unknown | 512k | seqwrite | qd8 | generic | 2 | 1 | 2 | 100 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 36 | Interrupted | 512k | write | qd8 | florbd | 2 | 1 | 1 | 100 | 0.000 | 0.000 | 0.000 | 375.862 | 169.453 | 191.704 |
| 38 | Completed | 64k | seqwrite | qd64 | qemurbd | 2 | 1 | 1 | 100 | 1148.000 | 71.759 | 55.720 | 316.040 | 142.924 | 144.671 |
| 40 | Completed | 64k | seqwrite | qd64 | qemurbd | 2 | 1 | 1 | 100 | 1132.000 | 70.753 | 56.510 | 306.370 | 139.551 | 127.109 |
| 41 | Completed | 64k | seqwrite | qd64 | qemurbd | 2 | 1 | 1 | 100 | 1137.000 | 71.095 | 56.230 | 312.200 | 140.228 | 150.494 |

CeTune WebUI

- Reports view: Result report provides two report view, summary view shows history report list, also user can view the detail report of one specific testrun.

Test Configuration

CeTune Status

Result Reports

17-4 ...

CeTune – Ceph tuning and profiling

CeTune Status: idle Ceph Status: health HEALTH_OK

summary

workload

ceph

client

vclient

| run_id | Status | Op_size | Op_Type | QD | Driver | SN_Number | CN_Number | Worker | Runtime | IOPS | BW(MB/s) | Latency(ms) | SN_IOPS | SN_BW(MB/s) | SN_Latency(ms) |
|--------|-----------|---------|----------|------|---------|-----------|-----------|--------|---------|----------|----------|-------------|----------|-------------|----------------|
| 17 | Completed | 64k | seqwrite | qd64 | qemurbd | 3 | 4 | 4 | 400 | 3835.000 | 239.953 | 31.943 | 1076.684 | 492.900 | 50.167 |

Download

URL

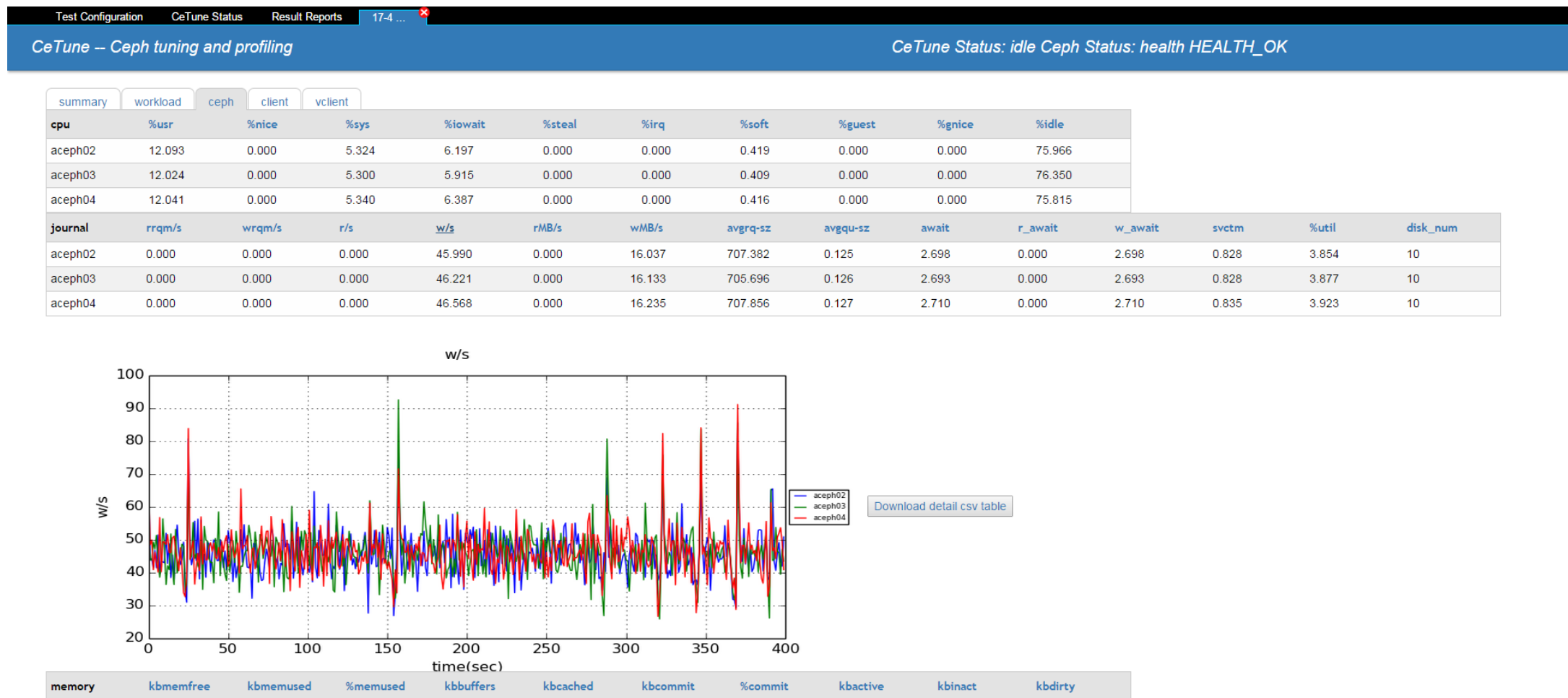
Configuration

Click TO Download

Ceph.conf configuration,
cluster description,
Cetune_process_log,
Fio/cosbench error log

CeTune WebUI

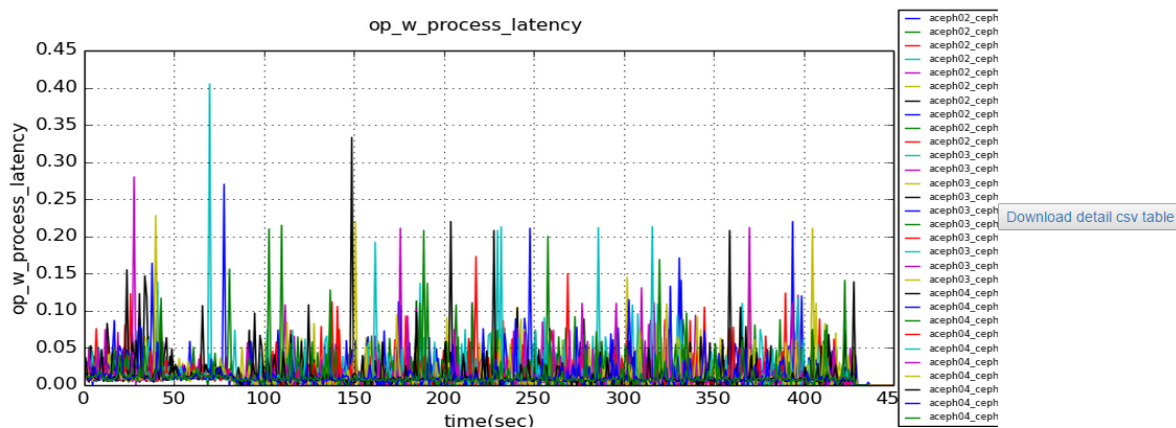
- Reports view: Result report provides two report view, summary view shows history report list, also user can view the detail report of one specific testrun.



CeTune WebUI

- Reports view: Result report provides two report view, summary view shows history report list, also user can view the detail report of one specific testrun.

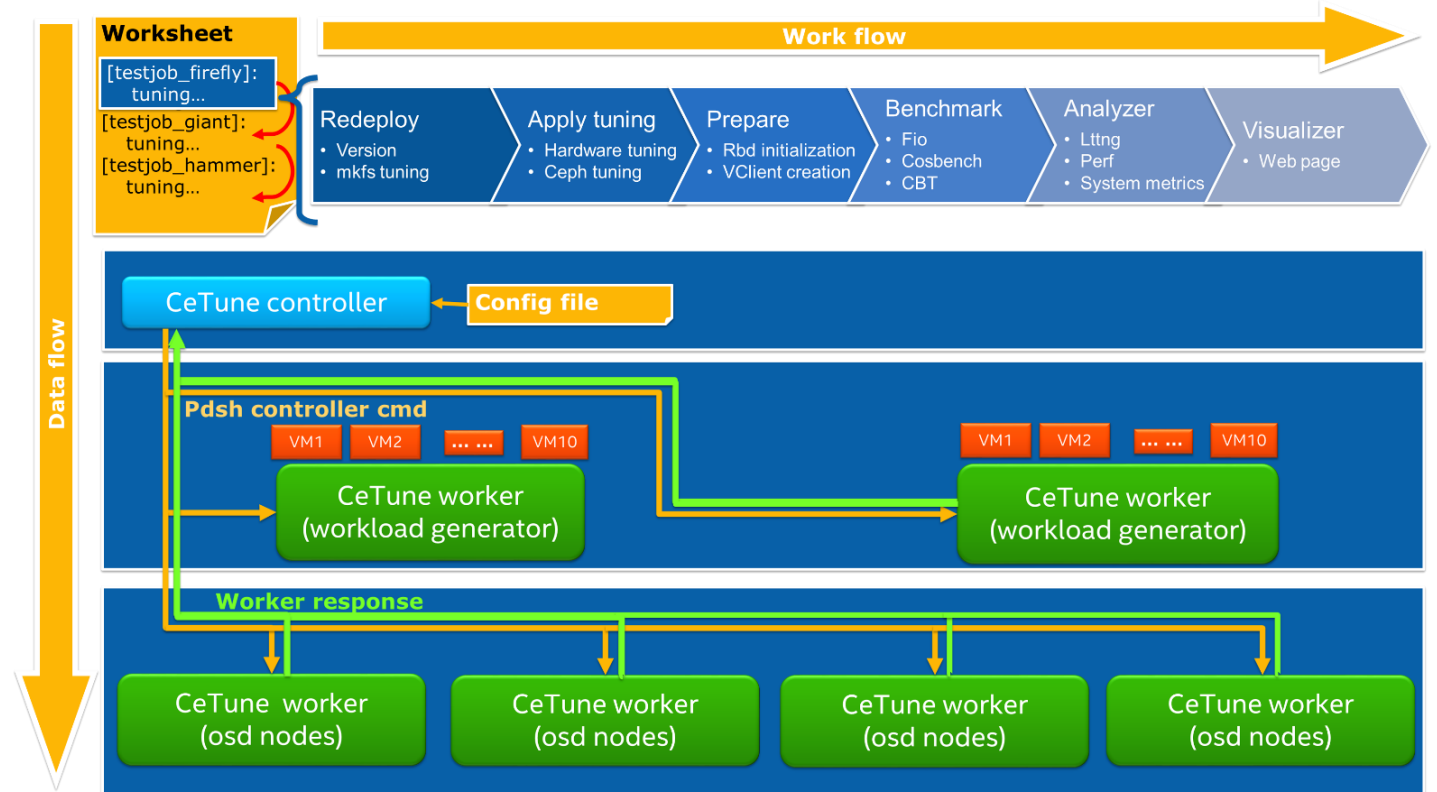
| | | | | | | | | | | | | | | | | | | |
|----------------------------------|-------|------------------------|-------|-----------------|---------------------------|----------|----------------------|-----------------|-----------------|------------------------|------------------------|------------|----------------|-------------------------------|-------|-------|----------|-------|
| aceph03_ceph- osd.19.asok.txt | 0.008 | 3531.069 | 0.014 | 0.006 | 3795651289.865 0.010 | 5059.379 | 2730152369.805 0.010 | 446.163 0.010 | 419.004 | 18294389056.989 | 981215950527.011 0.014 | 0.141 | 0.008 | 2730152369.805 3795654199.865 | | | | |
| aceph04_ceph- osd.20.asok.txt | 0.008 | 4410.479 | 0.016 | 0.012 | 4404588884.418 0.011 | 5917.356 | 3486973172.977 0.012 | 470.327 379.880 | 22254882107.450 | 977255457476.550 0.016 | 0.158 | 0.008 | 3486973172.977 | 4404593045.418 5942.356 | | | | |
| aceph04_ceph- osd.21.asok.txt | 0.008 | 5153.612 | 0.017 | 0.005 | 4596548937.100 0.010 | 6138.454 | 4012276479.039 0.012 | 480.097 0.012 | 382.084 | 22845696210.542 | 976664643373.458 0.017 | 0.169 | 0.008 | 4012276479.039 4596556645.100 | | | | |
| aceph04_ceph- osd.22.asok.txt | 0.008 | 3958.550 | 0.017 | 0.003 | 4201536667.594 0.012 | 5514.349 | 3049298998.795 0.012 | 476.109 0.012 | 381.998 | 19809480176.699 | 979700859407.301 0.017 | 0.144 | 0.008 | 3049298998.795 4201548329.594 | | | | |
| aceph04_ceph- osd.23.asok.txt | 0.016 | 976786960289.185 0.009 | | 22723379294.815 | 4859209670.256 644384.767 | 0.011 | 0.009 | 0.000 | 0.116 | 3630543254.984 | 4859209670.256 | 644384.767 | 6488.557 | 157.274 | 0.012 | 0.016 | 4710.205 | 0.000 |
| aceph04_ceph- osd.24.asok.txt | 0.008 | 4480.386 | 0.019 | 0.014 | 4706469669.701 0.013 | 6345.267 | 3413317172.212 0.014 | 492.048 380.872 | 21340181957.246 | 978170157626.754 0.019 | 0.131 | 0.008 | 3413317172.212 | 4706472882.701 6364.267 | | | | |
| aceph04_ceph- osd.25.asok.txt | 0.017 | 977667332234.443 0.008 | | 21843007349.557 | 4206734080.694 646569.087 | 0.011 | 0.008 | 0.000 | 0.158 | 3699564277.425 | 4206730109.694 | 644930.087 | 5563.893 | 162.388 | 0.013 | 0.017 | 4731.103 | 0.000 |
| aceph04_ceph- osd.26.asok.txt | 0.008 | 4406.525 | 0.017 | 0.010 | 4337477703.936 0.012 | 5832.169 | 3457001486.742 0.013 | 476.337 0.013 | 380.959 | 20886391929.169 | 978623947654.831 0.017 | 0.151 | 0.008 | 3457001486.742 4337496634.936 | | | | |
| aceph04_ceph- osd.27.asok.txt | 0.008 | 4569.429 | 0.017 | 0.006 | 4285844221.959 0.011 | 5715.852 | 3529147986.144 0.012 | 444.870 0.012 | 382.538 | 22017680887.646 | 977492658696.354 0.017 | 0.139 | 0.008 | 3529147986.144 4285844221.959 | | | | |
| aceph04_ceph- osd.28.asok.txt | 0.008 | 4589.651 | 0.016 | 0.009 | 3950713229.516 0.011 | 5274.292 | 3545602719.571 0.011 | 407.850 0.011 | 379.553 | 20827125038.873 | 978683214545.127 0.016 | 0.176 | 0.008 | 3545602719.571 3950716412.516 | | | | |
| aceph04_ceph- osd.29.asok.txt | 0.008 | 3876.790 | 0.018 | 0.013 | 3988665263.824 0.012 | 5171.034 | 3031876963.952 0.013 | 426.211 380.762 | 19200876113.346 | 980309463470.654 0.018 | 0.133 | 0.008 | 3031876963.952 | 3988673399.824 5216.034 | | | | |



CeTune Internal

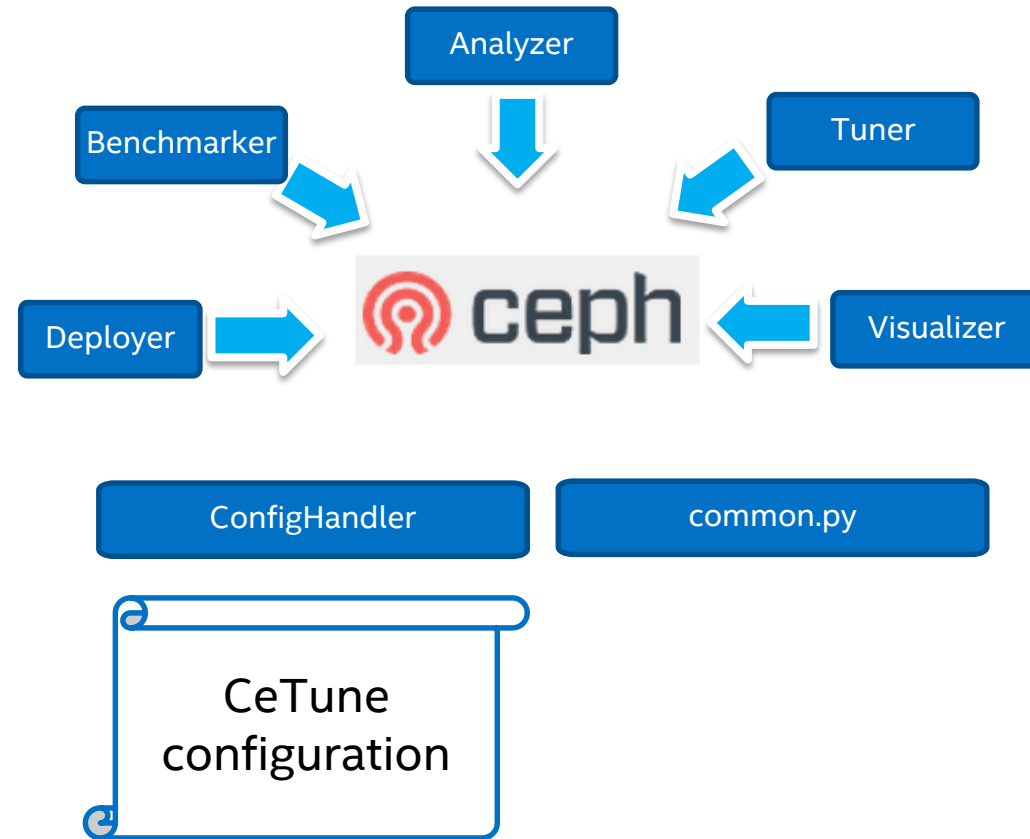
Terms definition

- Cetune controller
 - reads config files and controls the process to deploy, benchmark and analyze the collected data;
- Cetune workers
 - controlled by CeTune controller working as workload generator, system metrics collector.
- CeTune Configuration files
 - all.conf
 - tuner.conf
 - testcase.conf



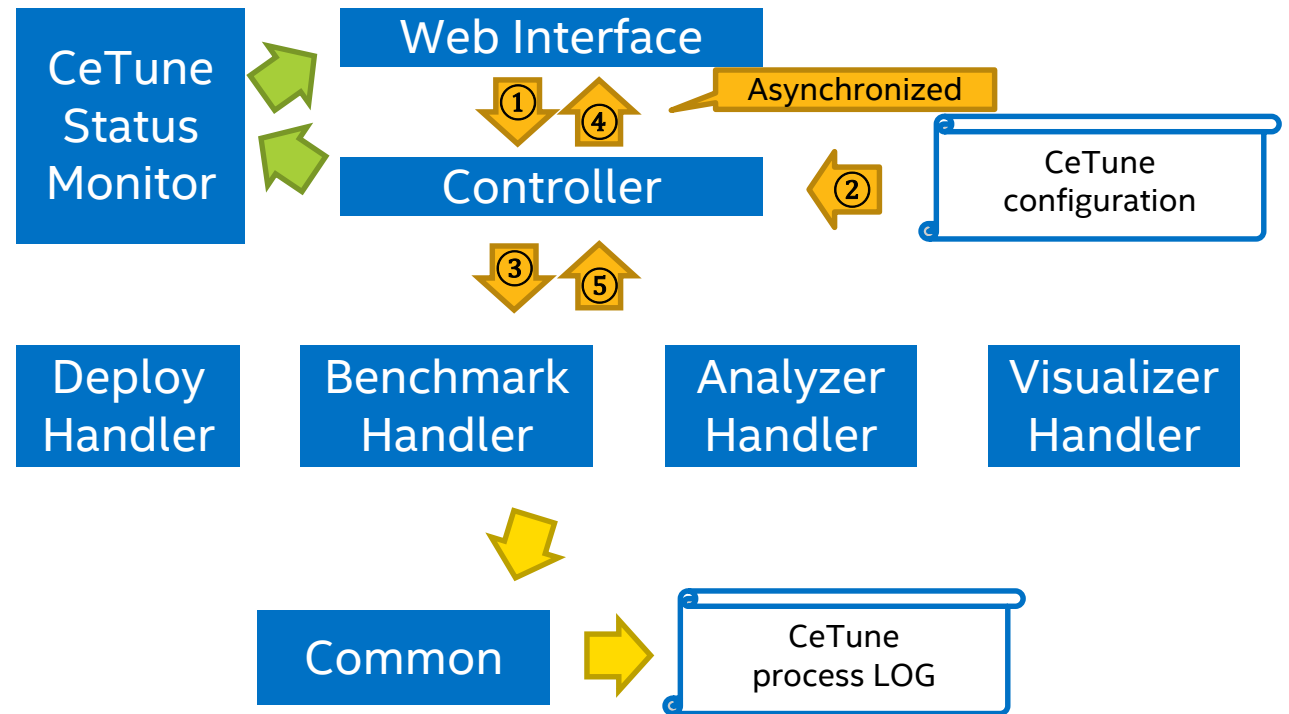
Modules

- Deployment:
 - clean build, incremental build
 - ceph cluster, radosgw
- Benchmark:
 - qemurbd, fiordb, cosbench
 - seqwrite, seqread, randwrite, randread, mixreadwrite
- Analyze:
 - System_metrics: iostat, sar, interrupt
 - Performance_counter
 - Latency_breakdown (WIP)
 - Output as a json file
- Tuner:
 - ceph tuning after comparing
 - pool tuning
 - disk tuning
 - Other System Tuning?
- Visualizer:
 - Read from json file
 - Output as html



Modules

- Deployment:
 - clean build, incremental build
 - ceph cluster, radosgw
- Benchmark:
 - qemurbd, fiordb, cosbench
 - seqwrite, seqread, randwrite, randread, mixreadwrite
- Analyze:
 - System_metrics: iostat, sar, interrupt
 - Performance_counter
 - Latency_breakdown (WIP)
 - Output as a json file
- Tuner:
 - ceph tuning after comparing
 - pool tuning
 - disk tuning
 - Other System Tuning?
- Visualizer:
 - Read from json file
 - Output as html



Deployment

- Ceph Package installation:
 - Use ceph-deploy to do installation.
 - Check and maybe reinstall nodes defined in CeTune description file to make sure all nodes under same Ceph Major release.
 - Ex: if some nodes are 0.94.3, some are 0.94.2 won't reinstall 0.94.2 nodes
- Deployment:
 - Clean_build: Remove current osd, mon, and radosgw then deploy ceph cluster from zero.
 - Incremental_build(Non clean build): compare current ceph cluster with desired configuration, then add osd devices, new node, radosgw daemon if necessary.

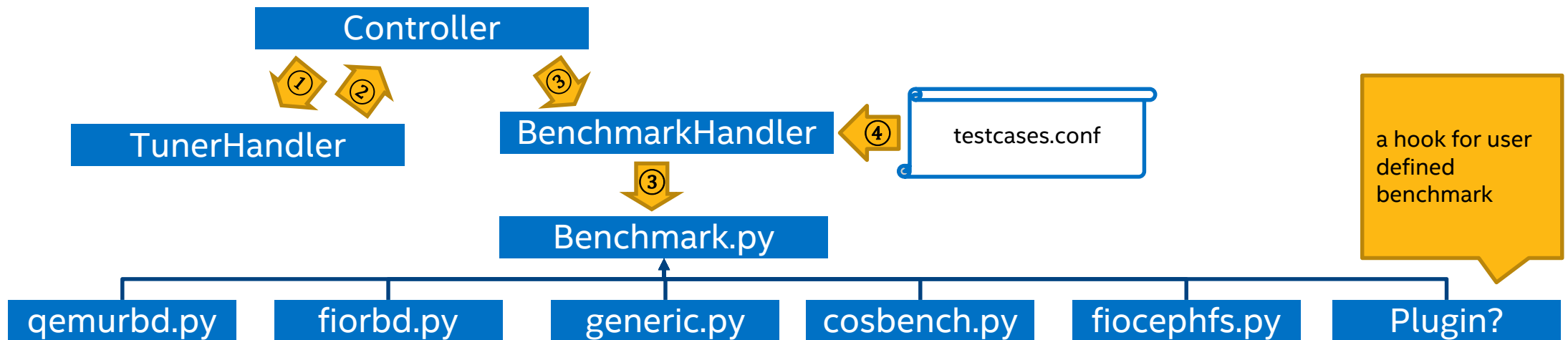
DeployHandler

Deploy.py

Deploy_rgw.py

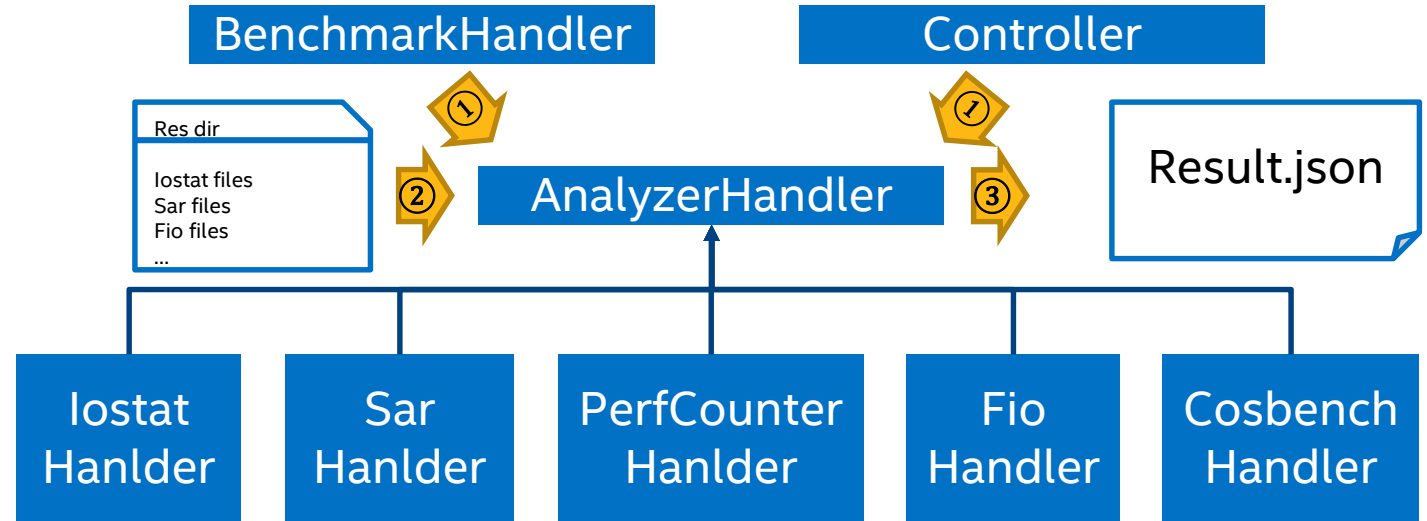
Benchmark

- RBD
 - Fio running inside VM, fio-rbd engine
- Cosbench
 - Adopts Radosgw to test object
- Cephfs
 - Fio cephfs engine(not recommend, will working on a more generic benchmark at CeTune v2)
- Generic devices
 - Distribute Fio test job to multi nodes multi disks.



Analyzer

- System metrics:
 - iostat: partition
 - Sar: cpu, mem, nic
 - Interrupt
 - Top: raw data
- Performance counter:
 - Indicates software behavior
 - Stable and well format in ceph codes
 - Well supported in cetune
- Lttnng(blkin):
 - Better reveal of code path
 - Current lttnng codes lack of an uniform identifier to mark one op, but doable by chaining thread id and object id.
 - Blkin branch is not merged to ceph master by now.
 - Not fully supported in CeTune, but have experience on visualize/process blkin lttnng data, and able to help.



Visualizer

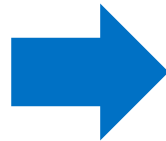
- Read from result.json
- Output as html

Result.json

```
1 {
2   "summary": {
3     "run_id": {
4       "2": {
5         "Status": "Completed\n",-
6         "Op_size": "64k",-
7         "Op_Type": "seqwrite",-
8         "QD": "qd64",-
9         "Driver": "qemurbd",-
10        "SN_Number": 3,-
11        "CN_Number": 4,-
12        "Worker": "20",-
13        "Runtime": "100",-
14        "IOPS": "12518.000",-
15        "BW(MB/s)": "782.922",-
16        "Latency(ms)": "102.758",-
17        "SN_IOPS": "3687.939",-
18        "SN_BW(MB/s)": "1556.812",-
19        "SN_Latency(ms)": "50.789"
20      }
21    },-
22    "Download": {}
23  },-
24  "workload": {},-
25  "ceph": {},-
26  "client": {},-
27  "vclient": {},-
28  "runtime": 100,-
29  "status": "Completed\n",-
30  "session_name": "2-20-qemurbd-seqwrite-64k-qd64-40g-0-100-vdb"
31}
```

Annotations:

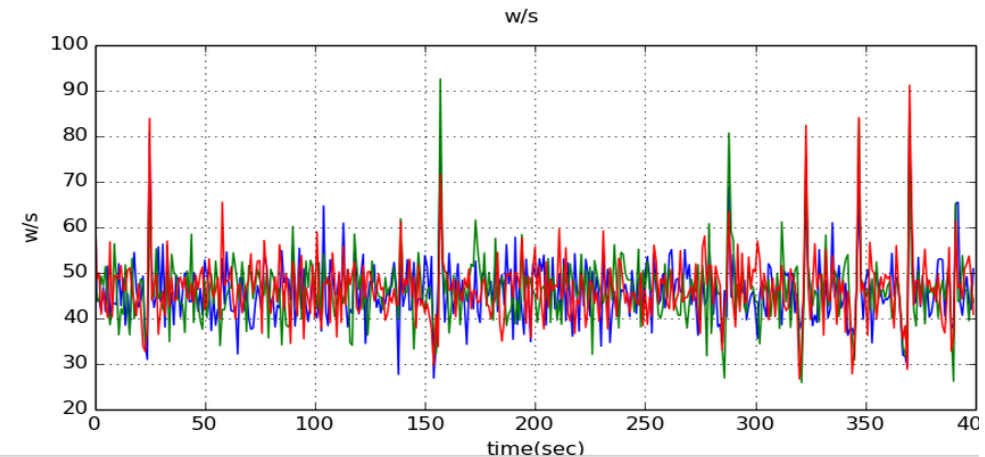
- tab (pointing to line 1)
- table (pointing to line 2)
- row (pointing to line 4)
- column, data (pointing to line 5)



Test Configuration CeTune Status Result Reports 17-4 ...

CeTune – Ceph tuning and profiling

| summary | workload | ceph | client | vclient | | | |
|---------|----------|--------|--------|---------|--------|--------|----------|
| cpu | %usr | %nice | %sys | %iowait | %steal | %irq | %soft |
| aceph02 | 12.093 | 0.000 | 5.324 | 6.197 | 0.000 | 0.000 | 0.419 |
| aceph03 | 12.024 | 0.000 | 5.300 | 5.915 | 0.000 | 0.000 | 0.409 |
| aceph04 | 12.041 | 0.000 | 5.340 | 6.387 | 0.000 | 0.000 | 0.416 |
| journal | rrqm/s | wrqm/s | r/s | w/s | rMB/s | wMB/s | avgrq-sz |
| aceph02 | 0.000 | 0.000 | 0.000 | 45.990 | 0.000 | 16.037 | 707.382 |
| aceph03 | 0.000 | 0.000 | 0.000 | 46.221 | 0.000 | 16.133 | 705.696 |
| aceph04 | 0.000 | 0.000 | 0.000 | 46.568 | 0.000 | 16.235 | 707.856 |



memory kbmemfree kbmemused %memused kbbuffers kbcached kbcommit %commit

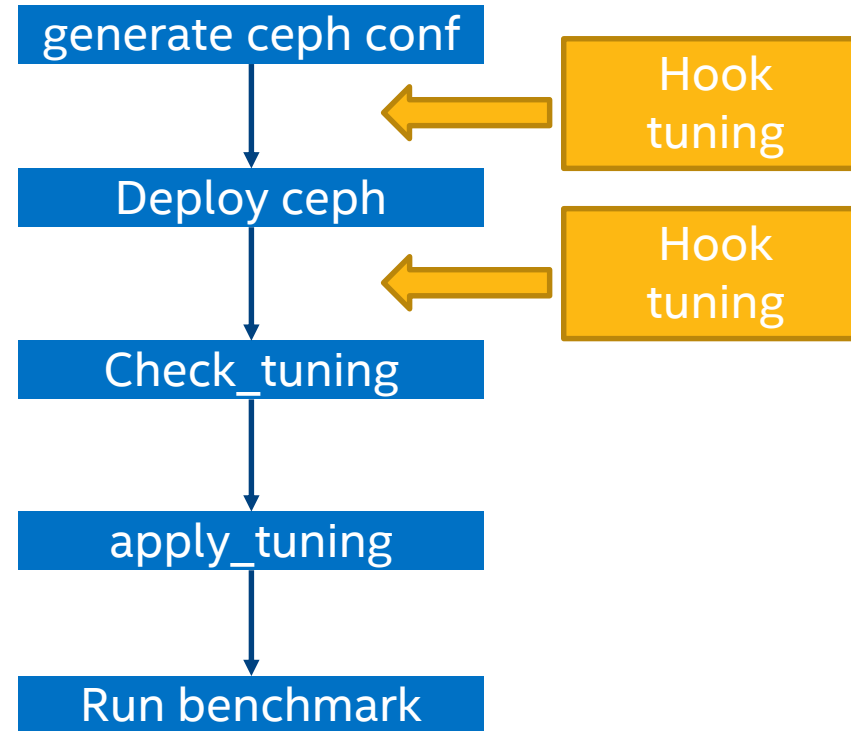
Tuning

- Current CeTune Supports:

- POOL tuning
- Ceph.conf tuning
 - Debug to 0
 - Op_thread_num
 - Xattr size
 - Etc.
- Disk tuning

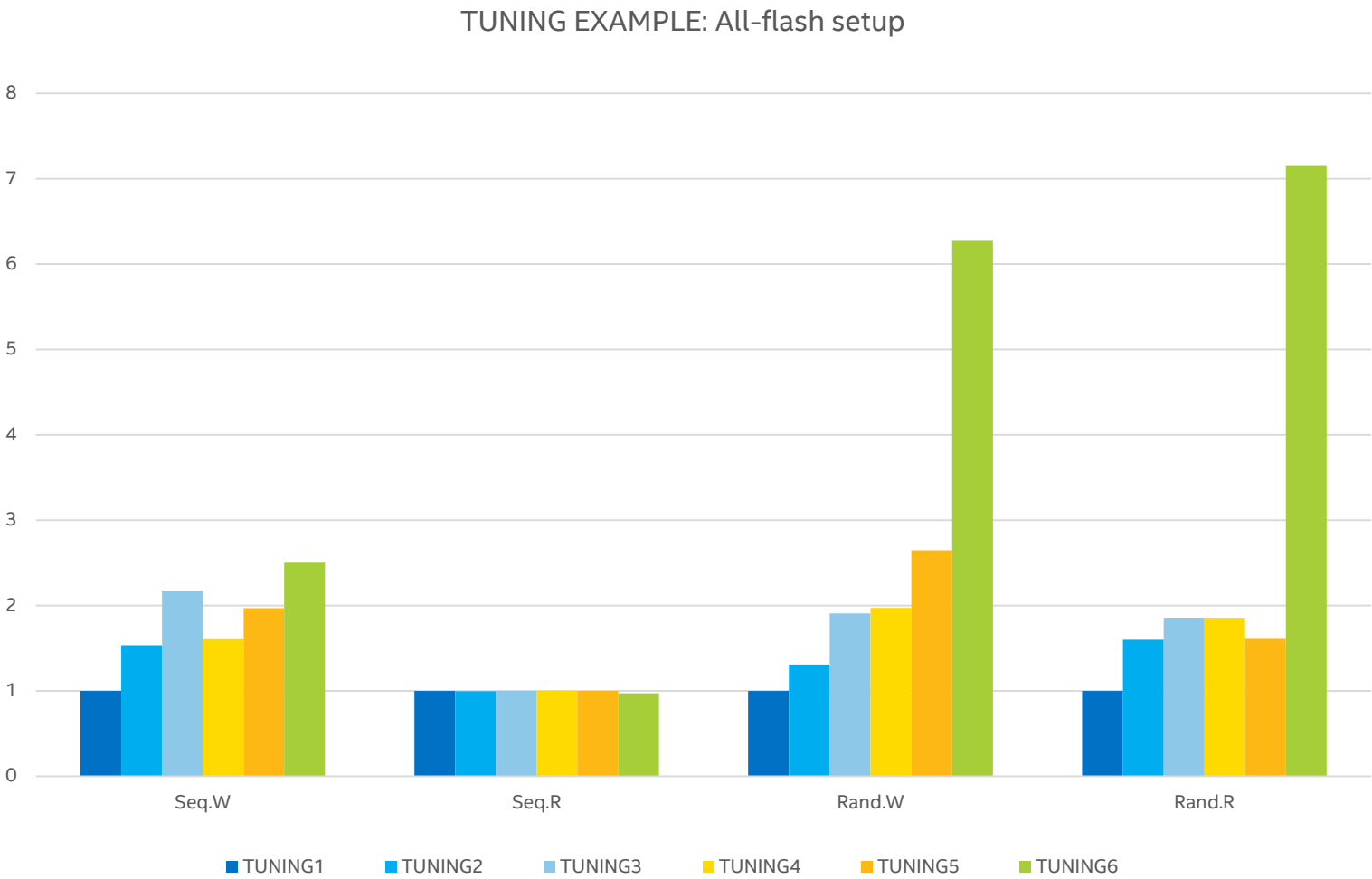
- To be extend:

- Mount omap to a separate device
- Multi osd daemons on one physical device
- Adopting flashcache device as osd
- Multi pools ?

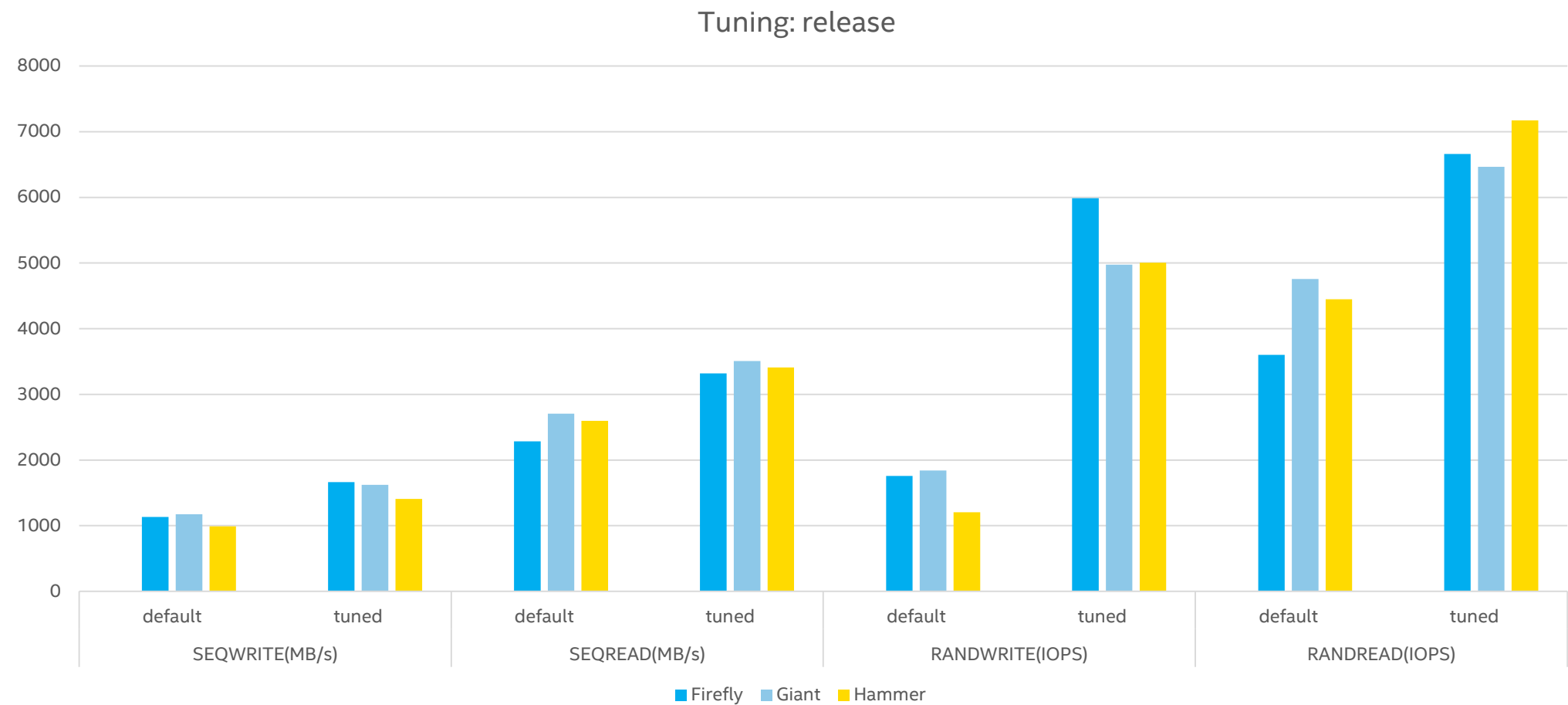


Tuning(All flash Setup)

| Case | Tuning Items |
|----------|---|
| TUNIN G1 | Default |
| TUNIN G2 | 2 OSD instances running on 2 partitions of same SSD |
| TUNIN G3 | Case-2 + Set debug log to 0 |
| TUNIN G4 | Case-3 + Set throttle values to 10x of default value |
| TUNIN G5 | Case-4 + disable RBD cache, OPtracker, tuning FD cache size to 64 |
| TUNIN G6 | Case-5 + Replace Tcmalloc with Jemalloc |



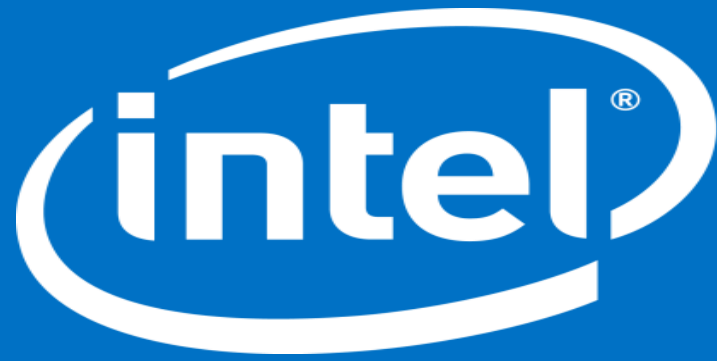
Tuning(releases)



Next Step

Next step

- Benchmark:
 - Cephfs benchmark - vdbench
 - Third party workload hook
- Failover test
- Analyzer:
 - BLKIN(lttng support) support
- Tuning



Legal Notices and Disclaimers

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.

No computer system can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit <http://www.intel.com/performance>.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

Statements in this document that refer to Intel's plans and expectations for the quarter, the year, and the future, are forward-looking statements that involve a number of risks and uncertainties. A detailed discussion of the factors that could affect Intel's results and plans is included in Intel's SEC filings, including the annual report on Form 10-K.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Intel, Xeon and the Intel logo are trademarks of Intel Corporation in the United States and other countries.

*Other names and brands may be claimed as the property of others.

© 2015 Intel Corporation.

Legal Information: Benchmark and Performance Claims Disclaimers

Software and workloads used in performance tests may have been optimized for performance only on Intel® microprocessors. Performance tests, such as SYSmark* and MobileMark*, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase.

Test and System Configurations: See Back up for details.

For more complete information about performance and benchmark results, visit <http://www.intel.com/performance>.

Risk Factors

The above statements and any others in this document that refer to plans and expectations for the first quarter, the year and the future are forward-looking statements that involve a number of risks and uncertainties. Words such as "anticipates," "expects," "intends," "plans," "believes," "seeks," "estimates," "may," "will," "should" and their variations identify forward-looking statements. Statements that refer to or are based on projections, uncertain events or assumptions also identify forward-looking statements. Many factors could affect Intel's actual results, and variances from Intel's current expectations regarding such factors could cause actual results to differ materially from those expressed in these forward-looking statements. Intel presently considers the following to be important factors that could cause actual results to differ materially from the company's expectations. Demand for Intel's products is highly variable and could differ from expectations due to factors including changes in the business and economic conditions; consumer confidence or income levels; customer acceptance of Intel's and competitors' products; competitive and pricing pressures, including actions taken by competitors; supply constraints and other disruptions affecting customers; changes in customer order patterns including order cancellations; and changes in the level of inventory at customers. Intel's gross margin percentage could vary significantly from expectations based on capacity utilization; variations in inventory valuation, including variations related to the timing of qualifying products for sale; changes in revenue levels; segment product mix; the timing and execution of the manufacturing ramp and associated costs; excess or obsolete inventory; changes in unit costs; defects or disruptions in the supply of materials or resources; and product manufacturing quality/yields. Variations in gross margin may also be caused by the timing of Intel product introductions and related expenses, including marketing expenses, and Intel's ability to respond quickly to technological developments and to introduce new features into existing products, which may result in restructuring and asset impairment charges. Intel's results could be affected by adverse economic, social, political and physical/infrastructure conditions in countries where Intel, its customers or its suppliers operate, including military conflict and other security risks, natural disasters, infrastructure disruptions, health concerns and fluctuations in currency exchange rates. Results may also be affected by the formal or informal imposition by countries of new or revised export and/or import and doing-business regulations, which could be changed without prior notice. Intel operates in highly competitive industries and its operations have high costs that are either fixed or difficult to reduce in the short term. The amount, timing and execution of Intel's stock repurchase program and dividend program could be affected by changes in Intel's priorities for the use of cash, such as operational spending, capital spending, acquisitions, and as a result of changes to Intel's cash flows and changes in tax laws. Product defects or errata (deviations from published specifications) may adversely impact our expenses, revenues and reputation. Intel's results could be affected by litigation or regulatory matters involving intellectual property, stockholder, consumer, antitrust, disclosure and other issues. An unfavorable ruling could include monetary damages or an injunction prohibiting Intel from manufacturing or selling one or more products, precluding particular business practices, impacting Intel's ability to design its products, or requiring other remedies such as compulsory licensing of intellectual property. Intel's results may be affected by the timing of closing of acquisitions, divestitures and other significant transactions. A detailed discussion of these and other factors that could affect Intel's results is included in Intel's SEC filings, including the company's most recent reports on Form 10-Q, Form 10-K and earnings release.

Backup

All Flash Setup Configuration Details

| Client Cluster | |
|----------------|---|
| CPU | Intel(R) Xeon(R) CPU E5-2699 v3 @ 2.30GHz 36C/72T |
| Memory | 64GB |
| NIC | 10Gb |
| Disks | 1 HDD for OS |

| Ceph Cluster | |
|--------------|--|
| CPU | Intel(R) Xeon(R) CPU E5-2699 v3 @ 2.30GHz 36C/72T |
| Memory | 64 GB |
| NIC | 10GbE |
| Disks | 4 x 400 GB DC3700 SSD (INTEL SSDSC2BB120G4) each cluster |

| Ceph cluster | |
|--------------|----------------|
| OS | Ubuntu 14.04.2 |
| Kernel | 3.16.0 |
| Ceph | 0.94.2 |

| Client host | |
|-------------|----------------|
| OS | Ubuntu 14.04.2 |
| Kernel | 3.16.0 |

- ***Ceph version is 0.94.2***
- **XFS** as file system for Data Disk
- 4 partitions of each SSD and two of them for OSD daemon
- replication setting (2 replicas), **2048 pgs/OSD**

Tuning release Configuration Details

| Ceph Nodes | |
|------------|--|
| CPU | 1 x Intel Xeon E3-1275 V2 @ 3.5 GHz (4-core, 8 threads) |
| Memory | 32 GB (4 x 8GB DDR3 @ 1600 MHz) |
| NIC | 1 X 82599ES 10GbE SFP+, 4x 82574L 1GbE RJ45 |
| HBA/C204 | {SAS2008 PCI-Express Fusion-MPT SAS-2} / {6 Series/C200 Series Chipset Family SATA AHCI Controller} |
| Disks | 1 x INTEL SSDSC2BW48 2.5" 480GB for OS 1 x Intel P3600 2TB PCI-E SSD (Journal) 2 x Intel S3500 400GB 2.5" SSD as journal 10 x Seagate ST3000NM0033-9ZM 3.5" 3TB 7200rpm SATA HDD (Data) |

| Client Nodes | |
|--------------|---|
| CPU | 2 x Intel Xeon E5-2680 @ 2.8Hz (20-core, 40 threads) (Qty: 3) |
| Memory | 128 GB (8GB * 16 DDR3 1333 MHZ) |
| NIC | 2x 10Gb 82599EB, ECMP (20Gb), 64 GB (8 x 8GB DDR3 @ 1600 MHz) |
| Disks | 1 HDD for OS |

| Client VM | |
|-----------|------------------|
| CPU | 1 X VCPU VCPUPIN |
| Memory | 512 MB |

| Ceph cluster | |
|--------------|--------------|
| OS | Ubuntu 14.04 |
| Kernel | 3.16.0 |

| Client host | |
|-------------|--------------|
| OS | Ubuntu 14.04 |
| Kernel | 3.13.0 |

| Client VM | |
|-----------|--------------|
| OS | Ubuntu 12.10 |
| Kernel | 3.5.0-17 |