Tuning Objectives

The hardware you use determines the performance limits of your system and your Ceph cluster. The objective of tuning performance is to use your hardware as efficiently as possible.

It is a common observation that tuning a specific subsystem can adversely affect the performance of another. For example, you can tune your system for low latency at the expense of high throughput. Therefore, before starting to tune, establish your goals to align with the expected workload of your Ceph cluster:

IOPS optimized
Workloads on block devices are often IOPS intensive, for example, databases running on virtual machines in OpenStack. Typical deployments require high-performance SAS drives for storage and journals placed on SSDs or NVMe devices.

Throughput optimized
Workloads on a RADOS Gateway are often throughput intensive. Objects can store significant amounts of data, such as audio and video content.

Capacity optimized
Workloads that require the ability to store a large quantity of data as inexpensively as possible usually trade performance for price. Selecting less-expensive and slower SATA drives is the solution for this kind of workload.

Depending on your workload, tuning objectives should include:

- Reduce latency
- Increase IOPS at the device
- Increase block size

Optimizing Ceph Performance

The following section describes recommended practices for tuning Ceph.

Ceph Deployment

It is important to plan a Ceph cluster deployment correctly. The MONs performance is critical for overall cluster performance. MONs should be on dedicated nodes for large deployments. To ensure a correct quorum, an odd number of MONs is required.

Designed to handle large quantities of data, Ceph can achieve improved performance if the correct hardware is used and the cluster is tuned correctly.

After the cluster installation, begin continuous monitoring of the cluster to troubleshoot failures and schedule maintenance activities. Although Ceph has significant self-healing abilities, many types of failure events require rapid notification and human intervention. Should performance issues occur, begin troubleshooting at the disk, network, and hardware level. Then, continue with diagnosing RADOS block devices and the Ceph RADOS Gateways.

Recommendations for OSDs
Use SSDs or NVMes to maximize efficiency when writing to BlueStore block database and write-ahead log (WAL). An OSD might have its data, block database, and WAL *collocated* on the same storage device, or *non-collocated* by using separate devices for each of these components.

In a typical deployment, OSDs use traditional spinning disks with high latency because they provide satisfactory metrics that meet defined goals at a lower cost per megabyte. By default, BlueStore OSDs place the data, block database, and WAL on the same block device. However,