Management

Objectives

We will be covering:

- Hardware Considerations
  - CPUs
  - RAM
  - Network
  - Flash/SSD
  - RAID
- Capacity Planning
  - RAM
  - Flash/SSD

Hardware Considerations

Hardware - CPU

- Choose at least quad-core CPUs. Most current Aerospike users use 6 or 8 core CPUs.
- Aerospike is multi-core and multi-threaded and can take advantage of more cores if the storage throughput allows.
- When using Flash/SSD, CPU is usually not a bottleneck.
- When checking if CPU will be an issue, you may notice that top shows software interrupts ("si" column at >30%). This is normally an indication you are actually bound by a single network queue.

There are 2 reasons to choose multiple high-powered CPUs:

1) You need to have the extra CPU to be able to access additional RAM
2) Your throughput requirements are high (> 80 ktps per node). If so, you should also be aware of your networking equipment

Using NICs that have 8+ network queues is essential to getting the best performance. If you only need < 80ktps per node, this will not likely pose an issue.

If you do have high throughput requirements, please contact Aerospike in assistance configuring your hardware.

Hardware RAM

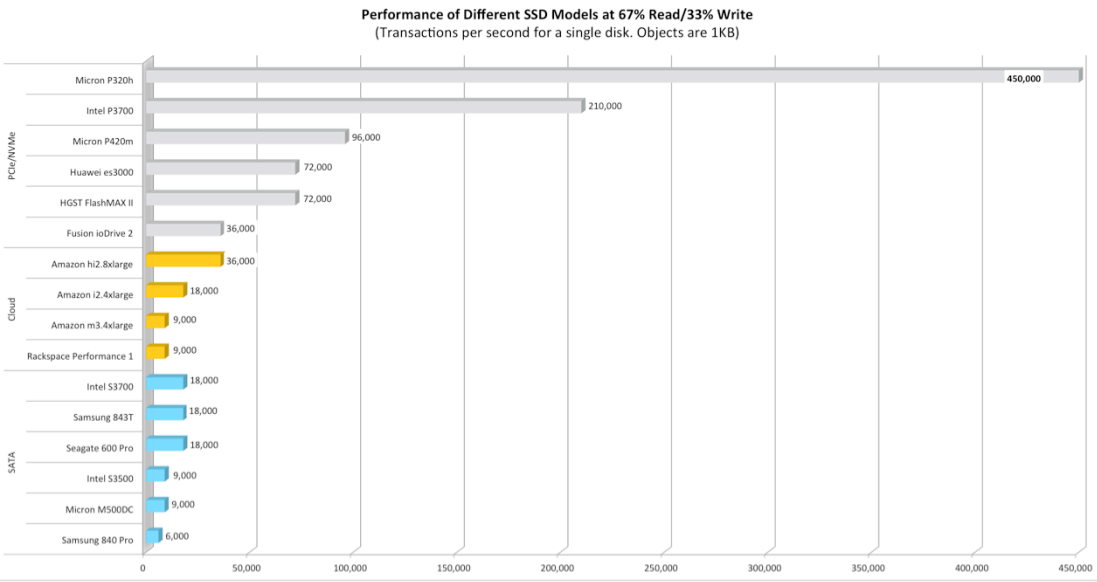
- RAM is used for:
  - Storage of primary index
  - Storage of secondary index (if using)
  - Storage of data (if using RAM for back end storage)
- Except in very high transaction loads (>~ 1 Mtps per node), RAM speed does not significantly affect performance.

Hardware - Network

- Many Aerospike customers use 2 NICs: 1 that is forward facing and is used to communicate with the clients. Also a second one that is used for intra-cluster communication. The settings for these can be configured.
- Aerospike can use either unicast (a.k.a mesh) or multicast methods for cluster heartbeats. Most people will use multicast, but you must make sure your network switches can handle this.
- People often underestimate the need for bandwidth. If you have 4 KB objects and need 30 ktps of transactions from a single node, this is 120 MB per second, or 0.96 Gb per second. This will obviously strain a 1 Gbps ethernet connection.
- If your bandwidth requirements are very high > 6Gbps, you can link aggregate multiple connections into a single one. Make sure the bonding method you use actually improves throughput and is not a simple failover.
- Even if your bandwidth requirements are met, it is possible that you will be bottlenecked by pure transactions per second due to software interrupts on a single core. Modern Linux kernels (version > 3.10) have a version of irqbalance that can handle this. Otherwise there are manual methods for doing the same thing. Contact Aerospike, if you wish to do this. There are hardware dependencies.

It is very common for Aerospike hosts to have more than 1 NIC. The details in selecting the best method for how this works depends on the details of the environment and how you wish to connect to them. This topic is covered in the configuration of an Aerospike cluster.

SSD Certification Tests



Aerospike maintains a list of SSD performance numbers at:

http://www.aerospike.com/docs/operations/plan/ssd/ssd_certification.html

Hardware - RAID

- Direct connect of flash/SSD to the motherboard often works very well.
- For SATA/SAS based flash devices, the choice and configuration of the RAID controller is very important. RAID controllers based on the LSI 2208 (used by many manufacturers) are especially good if you use the LSI Fastpath. The method for turning this on varies depending on the RAID controller manufacturer. In many cases this requires a low-cost software license.
- PCIe based flash devices are still new. It is not known how linearly these scale with multiple cards.

Performance of even the best drives can be hurt by using improper connections to the server.

Normally the best performance is to use direct attach to the motherboard.

If this is not possible, you will want to use very good RAID controllers. The lower end ones get very poor performance with multiple drives.

Aerospike has found that a single disk on a RAID controller may have very good performance, but as you add more drives, this degrades very rapidly.

In some cases your RAID controller may have settings that are optimal for control of individual disks rather than as a single RAIDed drive. This is common among controllers optimized for hadoop.

Capacity Planning

Capacity Planning

For a exact, detailed planning guide, go to:

http://www.aerospike.com/docs/operations/plan/capacity/

For a quick estimate (no secondary indexes) you will need the following:

- number of records
- replication factor
- size of records
- number of bins
- average set size
- type of storage (RAM or flash/SSD)

The exact method for estimating is often too stringent. To create a simple estimate collect the following.

Capacity Planning – Quick Estimate

Area	How stored	Formula	Note
Primary Index	RAM	n * r * 64	The amount of RAM needed for the primary index is fixed at 64 bytes.
Data storage	RAM	n * r * (2 + (17 * b) + v)	Every objects needs 2 bytes for overhead, 17 bytes per bin, and the actual data
Data storage	Flash/SSD	n * r * p Where p is ((64 + (9 + s) + (28 * b) +5 + v) -> round up to nearest 128 bytes	Every object needs to store the index (64 bytes), set overhead (9 +s bytes) , general overhead (28 bytes), type info (5 bytes), and the actual data. Because Aerospike stores data in 128 byte blocks, you must round up to the nearest 128 byte amount.

- n = number of records
- r = replication factor
- v = average size of records
- b = number of bins
- s = average set size

Note that this does not take into account secondary indexes. Aerospike always stores secondary indexes in RAM, so you should plan accordingly.

The size of the index varies on the type of value (string or numeric), but as a quick estimate, you can assume 40 bytes of RAM per string and 64 bytes per integer indexed.

Summary

What we have covered:

- Hardware Considerations
  - CPUs
  - RAM
  - Network
  - Flash/SSD
  - RAID
- Capacity Planning
  - RAM
  - Flash/SSD