



HP Vertica Hardware Planning Guide

This document provides information that helps you to appropriately size and recommend Hardware for HP Vertica.

Optimal Hardware for HP Vertica

Processor

For optimal performance, HP recommends running two socket servers with six or eight core CPUs, clocked at or above 2.2 GHz (Note: in some smaller applications it may be acceptable to run four core CPUs).

Memory

HP requires a minimum of 4 GB of memory per physical CPU core in the server, however you should run 8 GB of memory per physical core. (Note: Some heavy analytics workloads may require up to 16 GB of memory per physical core.) The memory should be at least DDR3-1333 (preferably DDR3-1600), and should be appropriately distributed across all memory channels in the server.

Storage

HP requires a minimum read/write speed of 20 MB/s per physical core of the CPU. However, you should have 40–60 MB/s per physical core. Each node should have 1–9 TB of storage post RAID. In a production setting, RAID 10 is recommended but, in some cases, RAID 50 is acceptable. (Note: SSDs are not required due to the heavy compression/encoding done by HP Vertica. In most cases, a RAID array of more, less expensive HDDs will work just as well as a RAID array of fewer SSDs to satisfy HP Vertica's requirements.)

Network

On a cluster with more than three nodes, and more than 10 TB, 10 GB networking is highly recommended. HP Vertica works well in smaller instances on 1 GB networking, but for large cluster sizes and large data sets, 10 GB networking is highly recommended. In some cases, it may also be acceptable to run quad-bonded 1G networking (4G throughput).

Sizing Your Cluster

You should consider several factors to properly size a cluster, but for the sake of simplicity we will focus on three:

1. **Data Volume (Compression):** First look at the total raw data volume for the cluster, and then apply a reasonable compression number (in most cases 2:1 compression with high availability is a good start). You should use a previously attained compression number or run some quick tests to figure out a reasonable compression number for your data. To do this, install HP Vertica on an existing system and run Database Designer (DBD), or use a ddl you already have and load in several hundred gigabytes to 1 TB of data.
2. **Data Growth:** Once you have a good idea of the starting compressed data volume, look at the amount of data ingest (the amount of data you will be saving each day) and the retention policy (how long you will be saving the data). Consider your starting compressed data volume and your ingest rate multiplied by the retention period; this gives you a good idea of the total data volume for your cluster.
3. **Workload:** The next important factor to look into is the workload itself. Is it a high-concurrency workload? What type of workload management will you be doing? How much data will the average query be running against (that is, will partition pruning be happening)? These can all be major factors, but for the most part a high concurrency workload will require more disk speed (high spindle count) and more memory on each node.

Server Configuration

HP recommends a minimum of three nodes for high availability. (Note: All configurations are based on compressed data volumes.) Always run your hardware BOM through the HP Vertica team before purchasing hardware for a production cluster.

| Data Volume Size | Recommended Server Configuration and Comments |
|------------------|---|
| Up to 2 TB | Three nodes with at least eight 10 k rpm spinning disks for the data partition (total of 1–2 TB per node), dual 4–6 core processors with 4–8 GB of memory per core, and 1 GB networking. (Note: in cases with heavy workload, run the suggested configuration for a 2–5 TB cluster). |
| 2–5 TB | Three nodes with at least 10 10 k rpm spinning disks for the data partition (total of 1–4 TB per node), dual 6 core processors with 8 GB of memory per core, and 1 GB networking. (Note: in cases with heavy workload, run the suggested configuration for a 5–10 TB cluster). |
| 5–10 TB | Three nodes with at least 20 10 k rpm spinning disks (total of 5–8 TB per node), dual 8 core processors with 8 GB of memory per core, and either quad-bonded 1G networking (4G throughput) or 10 GB networking. (Note: in cases with heavy workload, or high concurrency, it may be beneficial to run 22 drives or 15 k rpm drives, with 16 GB of memory per core). |
| 10 TB – 1 PB | Given the node configuration for the 5–10 TB cluster, and 5 TB of usable storage per node, calculate the number of nodes you will need. (Note: the same changes apply for heavy workload environments). |
| Larger than 1 PB | In most cases the same configuration as listed above will work, but contact an HP Vertica Technical Representative for recommendations on sizing clusters over 1 PB (compressed). (Note: if you have 10:1 compression this becomes a 10 PB cluster.) |

Get connected

hp.com/go/getconnected

Current HP driver, support, and security alerts delivered directly to your desktop.

