

# System Requirements

- [Overview](#)
- [Java](#)
- [Networking](#)
- [Memory](#)
  - [Low-Memory Environments](#)
- [Environments](#)
  - [Development Systems](#)
  - [Production Systems](#)
- [DNS Settings](#)
  - [Hostname](#)
    - [on CentOS](#)
  - [Hosts File Entries](#)
    - [file : /etc/hosts](#)
    - [A few things to note:](#)
  - [Validating DNS Settings](#)
  - [Certificates](#)
- [Operating Systems](#)
  - [Linux Requirements](#)
    - [Linux Memory and Child Processes](#)
    - [Locale](#)
    - [Shell](#)
    - [SELinux](#)
    - [Synchronize Clocks](#)
    - [Required Libraries and Tools](#)
      - [libstdc++](#)
      - [zlib](#)
      - [glibc](#)
      - [Python](#)
      - [Run the Linux Checker Script](#)
  - [Windows Requirements and Recommended Settings](#)
    - [Required Packages](#)
    - [Shell](#)
    - [Language Pack \(for non-English Windows installations\)](#)
    - [Deprecations](#)
    - [Browser Known Issues](#)
- [Hadoop Cluster Requirements](#)
  - [Supported Hadoop Versions](#)
  - [Hadoop Configuration Requirements](#)
  - [Hadoop Linux Settings](#)
  - [Minimal Recommended Hadoop Configuration for a Test System](#)
  - [Minimal Recommended Hadoop Configuration for a Production System](#)
  - [Memory Calculations for a Cluster](#)
- [Best Practices](#)
  - [Virtual Environments](#)
    - [Oversubscription of Resources](#)
    - [VMware with VMotion](#)
  - [Antivirus and Firewall](#)
  - [Cloud Deployments](#)
  - [One Project Per Host](#)

## Overview

The Attivio Cognitive Search and Insight Platform is designed to scale up or down to meet the needs of your situation.

- Attivio can scale up to handle very large solutions, such as email archives containing hundreds of millions of records,
- It can also scale down to run on a typical developer's laptop.

While our minimum and recommended hardware specification reflect these possibilities, it is important to consider the size and complexity of your development or production system when choosing hardware. For instance, a standard laptop is not a good choice for prototyping a large production system, because certain features may require more resources than the laptop can provide.

If you have any questions about hardware, contact [Attivio Support](#) for help in properly sizing your hardware to meet your application requirements.

## Java

Attivio uses its own internal Java Virtual Machine, so project systems do not need to have Java installed separately. Systems on which API development is performed require the [Java SE Development Kit 11](#) (JDK 11).

## Networking

In order to install and configure Attivio, all systems must be able to communicate with each other via their primary IP addresses. There is no requirement that the machines have access to the Internet.

# Memory

## Low-Memory Environments

Attivio allows developers to run in environments with limited memory. Before running in a low-memory environment, read the [Memory Usage Tuning](#) guide

Machines with limited memory should only be used in production environments with small indexes and very light query load. In addition, certain content types and/or modules such as binary file formats and text extraction require more memory than a baseline system.

# Environments

## Development Systems

Follow these configuration guidelines on systems used to develop Attivio applications:

Requirement	Minimum Configuration	Recommended Configuration
<b>Operating system</b>	64-bit Windows Server 2016 64-bit Windows Server 2012 64-bit Windows Server 2008 64-bit Windows 10 64-bit Windows 7  64-bit Debian (versions 6.x and 7.x) 64-bit Red Hat Enterprise Linux (versions 6.x and 7.x) (See <a href="#">installation note</a> .) 64-bit CentOS Linux (versions 6.x and 7.x)	64-bit Windows Server 2016 (single node, unclustered only) 64-bit Windows Server 2012 (single node, unclustered only) 64-bit Windows Server 2008 (single node, unclustered only)  64-bit Debian (versions 6.x and 7.x) 64-bit Red Hat Enterprise Linux (versions 6.x and 7.x) 64-bit CentOS Linux (versions 6.x and 7.x)  Starting with Attivio v5.2, clustered systems require at least Debian / Red Hat / CentOS version 7.
<b>CPU</b>	Quad-core	Quad-core+
<b>RAM</b>	16 GB	16+ GB
<b>Disk space</b>	Based on indexed content and enabled features	Based on indexed content and enabled features
<b>Storage</b>	n/a	SAS or SATA direct-attached  <div style="border: 1px solid #f0e68c; padding: 5px; text-align: center;">  N/A is not supported for index storage         </div>
<b>Network connection</b>	10 Gigabit Ethernet using the same physical VLAN and IP subnet for all internal Attivio communication. External client traffic such as queries can be on a separate subnet (if using multiple nodes).	See the required configuration.
<b>Web browser</b> (Admin UI, Search UI)	See the recommended configuration.	Attivio was tested with the following browsers at release time:  <b>Windows Clients:</b> <ul style="list-style-type: none"> <li>• Chrome stable version that is current at time of Attivio release</li> <li>• Microsoft Edge latest stable version at Attivio release date</li> <li>• Internet Explorer 11 (see <a href="#">Browser Known Issues</a> below)</li> </ul> <b>Mac clients:</b> Chrome stable version that is current at time of Attivio release. <b>Linux clients:</b> Chrome stable version that is current at time of Attivio release.

<b>Screen Resolution</b>	1280 x 800 pixels	1600 x 900 pixels or higher
<b>Attivio Deployment</b>	See the recommended configuration.	All files in the Attivio installation folders on each node must match, including patches, JDBC drivers, and configuration settings.

## Production Systems

Follow these configuration guidelines on systems used to deploy Attivio to end-users:

Requirement	Minimum Configuration	Recommended Configuration
<b>Operating system</b>	64-bit Windows Server 2016 64-bit Windows Server 2012 64-bit Windows Server 2008  64-bit Debian (versions 6.x and 7.x) 64-bit Red Hat Enterprise Linux (versions 6.x and 7.x) (See <a href="#">installation note.</a> ) 64-bit CentOS Linux (versions 5.x, 6.x, and 7.x)	64-bit Windows Server 2016 (unclustered only) 64-bit Windows Server 2012 (unclustered only) 64-bit Windows Server 2008 (unclustered only)  64-bit Debian (versions 6.x and 7.x) 64-bit Red Hat Enterprise Linux (versions 6.x and 7.x) 64-bit CentOS Linux (versions 5.x, 6.x, and 7.x)  Starting with Attivio v5.2, clustered systems require at least Debian / Red Hat / CentOS version 7.
<b>CPU</b>	Quad-core	8-core, with at least one core per indexer
<b>RAM</b>	32 GB	32 GB for small indexes/projects. See <a href="#">Memory on the Cluster</a> and <a href="#">Memory Usage Tuning</a> for more information.
<b>Disk space</b>	Based on indexed content and enabled features.	Based on indexed content and enabled features.
<b>Storage</b>	n/a	High performance direct-attached disk or SAN.   <b>NAS is not supported for index storage.</b>
<b>Network connection</b>	In order to meet typical SLA requirements, Attivio requires a 10 Gigabit Ethernet using the same physical VLAN and IP subnet for all internal Attivio communication. External client traffic such as queries can be on a separate subnet (if using multiple nodes).  <b>Additional requirements for optimal performance:</b> <ul style="list-style-type: none"> <li>No more than 10ms latency between all nodes</li> <li>Monitoring across the network with the ability to provide detailed status/logs of the network health related to Attivio events/logs as needed</li> <li>The entire system including Network, CPU, Memory and Disk IO is NOT oversubscribed</li> </ul>	See the required configuration.
<b>Web browser</b> (Admin UI, Search UI)	See the recommended configuration.	Attivio was tested with the following browsers at release time:  <b>Windows Clients:</b> <ul style="list-style-type: none"> <li>Chrome latest stable version at Attivio release date</li> <li>Microsoft Edge latest stable version at Attivio release date</li> <li>Internet Explorer 11 (see <a href="#">Browser Known Issues</a> below)</li> </ul> <b>Mac clients:</b> Chrome latest stable version at Attivio release date <b>Linux clients:</b> Chrome latest stable version at Attivio release date

<b>Attivio Deployment</b>	See the recommended configuration.	All files in the Attivio installation folders on each node must match, including patches, JDBC drivers, and configuration settings.
---------------------------	------------------------------------	---

## DNS Settings

When deploying Attivio, whether an unclustered or multi-node topology, it is critical that each of the Attivio hosts can communicate with themselves and each other properly. The following steps are recommended when configuring Linux servers for a multi-node environment.

### Hostname

Set the hostname of each host to the FULLY QUALIFIED NAME.

For example, `attivio1.lab.mycompany.com` rather than simply `attivio1`.

### on CENTOS

set this in `/etc/sysconfig/network`

```
HOSTNAME=attivio1.lab.mycompany.com
```

Reboot the host for hostname settings to take effect. Repeat this for every node.

### Hosts File Entries

Edit the `/etc/hosts` file to statically define the host forward and reverse lookup for each node in the cluster.

#### file : `/etc/hosts`

add the following \*AFTER\* the entries already present in `/etc/hosts`.

```
### Attivio cluster
# format
# ip_address fully_qualified_hostname alias
10.1.1.101 attivio1.lab.mycompany.com attivio1
10.1.1.102 attivio2.lab.mycompany.com attivio2
# etc...
```

### A few things to note:

1. The content of this file has to be distributed across the cluster on all machines. DO NOT copy the file onto target machines, the Attivio section needs to be APPENDED to `/etc/hosts`
2. The first entry is IP address (usually an internal IP address)
3. The second entry is the FULLY QUALIFIED HOST NAME. This makes sure reverse lookup picks up the correct hostname
4. The third entry is a shorthand alias
5. One common mistake that happens here is when host alias and fully qualified hostnames are swapped. The following is incorrect:

```
10.1.1.101 attivio1 attivio1.lab.mycompany.com
```

Aliases should follow fully qualified host names.

6. The Attivio cluster section of `/etc/hosts` file has to be distributed to all cluster nodes.

## Validating DNS Settings

In order to validate the DNS settings, run the following commands on each host in the cluster.

First, run `hostname -f`

```
hostname -f
attivio1.lab.mycompany.com
```

The value returned should be the fully qualified host name for the current host.

Next, run `getent hosts` for all other hosts and their respective IPs in the cluster.

```
getent hosts attivio2.lab.mycompany.com
10.1.1.102    attivio2.lab.mycompany.com

getent hosts 10.1.1.101
10.1.1.101    attivio2.lab.mycompany.com
```

The values returned should be the fully qualified host names for the other hosts. Repeat this process on each host to confirm consistent values are returned from each host to each other host.



Once all DNS settings have been validated, be sure to restart all Attivio processes, including the Attivio Agent.

## Certificates

When using SSL certificates, it is critical that the common name (CN) that is used in the certificate match the fully qualified domain name that is resolved by reverse DNS lookup.

# Operating Systems

## Linux Requirements

The settings described here are required for proper execution of Attivio processes on Linux. Modify any settings on your Linux OS to meet the following requirements.

View the current limits using the `ulimit -a` command and then change the settings by adding the following entries to your `/etc/security/limits.conf` file:

```
* soft nofile 65536
* hard nofile 65536
* soft memlock unlimited
* hard memlock unlimited
* soft as unlimited
* hard as unlimited
* soft nproc 10240
* hard nproc 10240
```

- Note that the above settings use the `***` wildcard to specify all regular non-root users on the system. Attivio strongly discourages using the root account to run any Attivio processes.
- On CentOS and RHEL systems, if the `nproc` limit is specified in `/etc/security/limits.d/90-nproc.conf` (RHEL 6.x/CentOS 6.x) or `/etc/security/limits.d/20-nproc.conf` (RHEL 7.x/CentOS 7.x), it will override the value set in `/etc/security/limits.conf`. Please see the section above about Attivio in the cloud.

The following setting is not in the `limits.conf` file. To make it permanent, add the following line to `/etc/sysctl.conf` and reboot the machine:

```
vm.max_map_count = 131072
```

## Linux Memory and Child Processes

Sometimes the Attivio process may spawn one or more child processes to execute certain commands or functionality. For example, on a system configured with replication, or when Advanced Text Extraction is being used, child processes will be forked and executed. On Linux, if the machine does not have sufficient free memory during the fork stage, these child processes will fail. The easiest way to avoid this problem is by setting the Linux kernel `vm.overcommit_memory` parameter to a non-zero value. The recommended value for `vm.overcommit_memory` is:

```
vm.overcommit_memory=1
```

To set this value permanently on Linux, you must modify the `/etc/sysctl.conf` file and reboot as with the `vm.max_map_count` setting mentioned above.

If `vm.overcommit_memory=0`, Attivio will likely run fine without errors as long as at least 1x the amount of memory allocated to Attivio is free. For example, on a machine with 16Gig of memory (including swap space), if Attivio is configured to run with 4Gig of memory, then as long as there is at least 4Gig of free space, Attivio will be able to fork these child processes.

## Locale

For Linux, the `en_US.UTF-8` encoding must be installed.

## Shell

On Linux systems, Attivio only supports the bash shell.

## SELinux

Ensure that [SELinux](#) has been disabled or put into permissive mode.

## Synchronize Clocks

Attivio recommends synchronizing clocks for all nodes in the topology using a tool such as NTP ( <http://www.ntp.org/> ). This will help with reconciling logs across nodes when troubleshooting. Synchronizing clocks is required for the Hadoop master and slave nodes when configuring a [multi-node topology](#) .

## Required Libraries and Tools

### libstdc++

Ensure that the `libstdc++` related libraries are installed and up to date. One way to check which libraries you may need to update would be to run the following command on RHEL/CentOS Linux:

```
yum list available | grep libstdc++
```

### zlib

If you are running Red Hat Enterprise Linux or CentOS 7.1, you may also need to install the 32-bit version of the `zlib` library, as it is not included in these distributions by default.

```
sudo yum install zlib.i686
```

### glibc

You must have `glibc` version 2.3 or later installed.

Attivio takes advantage of features introduced in `glibc` 2.13 when monitoring resource limits for Advanced Text Extraction processes. For best results, please install `glibc` 2.13 or later if offered for your Linux distribution and version.

## Python

Ensure that Python version 2.6.6 or higher is available.

## Run the Linux Checker Script

The Attivio Platform installation's `<INSTALL_DIR>/bin` directory includes a `linux_checker.sh` Bash script. Run this script from the `<INSTALL_DIR>/bin` directory and review its output to ensure that your current system configuration meets Attivio's requirements and recommendations.

```
> cd /opt/attivio/platform/bin
> ./linux_checker.sh
```

The script writes its output to `stdout` and to a log file named `linux_checker.<HOSTNAME>.log`. The output includes entries for a number of system tests. Each test returns one of three results:

Test Result	Meaning
[PASS]	The system meets Attivio's requirement and recommendation for the tested component.
[WARN]	The system meets Attivio's minimum requirement for the tested component, but does not meet the recommendation for this component.
[FAIL]	The system does not meet Attivio's minimum requirement for the tested component.

For [WARN] and [FAIL] results, a message displays below the test result with additional details.

Make appropriate changes to address any system settings marked with [WARN] or [FAIL] messages in the Linux checker script's output.

## Windows Requirements and Recommended Settings

### Required Packages

Attivio Platform 5.6.1 on Windows requires libraries from the [Microsoft Visual C++ 2010 Redistributable Package \(x64\)](#). Download the `vc_redist_x64.exe` installer for this package and execute it on each Attivio Platform host to install the required libraries.

As of Attivio Platform 5.6.2 these libraries are included with the Attivio installer, and no separate download or installation step is required for Windows hosts.

### Shell

On Windows systems, the only shell supported is the Command Prompt window. Other shells (in particular Cygwin) are not supported.

### Language Pack (for non-English Windows installations)

Attivio may fail to start on **non-English only** installations of Windows. To work around this issue, the English language pack for Windows should be installed.

### Deprecations

Deprecations are called out in the tables below as needed.

### Browser Known Issues

Business Center UI

Browser Version	Description
IE11	The Profile Import button is unresponsive.
IE11	Clicking on ranking chevrons makes the top few results disappear temporarily.

## Hadoop Cluster Requirements

Attivio can be configured to run in multi-node mode with a Linux Hadoop cluster. The following versions of Hadoop and related packages have been tested and are supported:

### Supported Hadoop Versions

Vendor	Version	HDFS	YARN	HBase	ZooKeeper	Hive
Hortonworks	2.3	2.7.1	2.7.1	1.1.1	3.4.6	1.2.1

Hortonworks	2.4	2.7.1	2.7.1	1.1.2	3.4.6	1.2.1
Cloudera	5.7	2.6.0	2.6.0	1.2.0	3.4.5	1.1.0
Cloudera	5.8	2.6.0	2.6.0	1.2.0	3.4.5	1.1.0
Cloudera	5.10	2.6.0	2.6.0	1.2.0	3.4.5	1.1.0
Cloudera	5.12	2.6.0	2.6.0	1.2.0	3.4.5	1.1.0
Cloudera	5.13	2.6.0	2.6.0	1.2.0	3.4.5	1.1.0
Cloudera	5.14	2.6.0	2.6.0	1.2.0	3.4.5	1.1.0
Cloudera	5.16.1	2.6.0	2.6.0	1.2.0	3.4.5	1.1.0

When running with a Hadoop cluster, the Attivio nodes and index are only supported on Linux.

## Hadoop Configuration Requirements

The Hadoop cluster must have the following installed and configured for proper Attivio execution:

- **HDFS**, **ZooKeeper**, and **HBase** installed and running
- The **Java Home Directory**, specified under Host Configuration, pointing to a Java JDK version 1.8.0\_60 or later (ex: `/usr/java/jdk1.8.0_60`)

### Software system specific parameters:

System	Parameter	Setting
ZooKeeper	maxSessionTimeout	600000
ZooKeeper	tickTime	5000
HBase	zookeeper.session.timeout	80000 or higher
YARN	yarn.scheduler.maximum-allocation-mb	MB of memory as calculated in <a href="#">Create an Attivio Project on Hadoop page's "Memory Considerations" section</a>
YARN	yarn.nodemanager.resource.memory-mb	MB of memory as calculated in <a href="#">Create an Attivio Project on Hadoop page's "Memory Considerations" section</a>
YARN	yarn.resourcemanager.recovery.enabled	<code>true</code> to allow the Attivio index YARN application(s) to persist across YARN restarts; if <code>false</code> , indexes must be re-deployed from the AIE-CLI whenever the YARN service is restarted
YARN	yarn.nodemanager.vmem-check-enabled	<code>false</code> , otherwise YARN might kill Attivio projects if virtual memory exceeds a certain limit
YARN	yarn.nodemanager.sleep-delay-before-sigkill.ms	60000 or higher; with a lower setting, index processes may not be able to clean up properly before shutdown, causing index partitions to become <b>LOCKED</b> for an hour or more
YARN HWX only	mapreduce.application.framework.path	Specify the exact HDP version as this will be used directly for directory paths. A. Obtain exact HDP version: Ambari > Admin > Stack and Versions > Versions, for example 2.3.2.0-2950. B. Go to Ambari > MapReduce2 > Configs > Filter search box > type mapreduce.application.framework.path > replace <code>\${hdp.version}</code> with the real version C. When downloading site files from the cluster to be used by the uploadcluster info tool, please include MapReduce2 configs by using Ambari > MapReduce2 > Service Actions drop-down > Download Client Configs.

## Hadoop Linux Settings

Attivio cannot offer generic advice about Linux requirements for Hadoop and HDFS, however, there are a few configuration changes that are needed on every Hadoop node and every Attivio node for proper Attivio execution. These are noted below.

Set `swappiness` to 1 to reduce the amount of swapping to minimum. To do this, edit the `/etc/sysctl.conf` file as root:

```
sudo vi /etc/sysctl.conf
```

And then add this setting to the bottom of the file:

```
vm.swappiness = 1
```

Disable **Transparent Huge Pages** for better performance by editing `/etc/rc.local` as root:

```
sudo vi /etc/rc.local
```

And add this to the bottom of the file:

```
#disable THP at boot time
if test -f /sys/kernel/mm/redhat_transparent_hugepage/enabled; then
  echo never > /sys/kernel/mm/redhat_transparent_hugepage/enabled
fi
if test -f /sys/kernel/mm/redhat_transparent_hugepage/defrag; then
  echo never > /sys/kernel/mm/redhat_transparent_hugepage/defrag
fi
```

Set `nofile` and `nproc` high to allow for large numbers of files and process needed by Attivio. To do this edit `/etc/security/limits.conf` as root:

```
sudo vi /etc/security/limits.conf
```

And change or add these settings:

```
* - nofile 32768
* - nproc 65536
```

After making these changes, reboot your machines for them to take effect:

```
sudo reboot
```

## Minimal Recommended Hadoop Configuration for a Test System

The minimum recommended Hadoop nodes and memory for a **test system** are as follows:

1 Instance, 32 GB RAM Hadoop Cluster

- 8 GB RAM allocated for the HBase Region Server Maximum Memory
- 4 GB RAM allocated for the HBase Master Maximum Memory

## Minimal Recommended Hadoop Configuration for a Production System

The minimum recommended Hadoop nodes and memory for a **production system** are as follows:

3 Instance, 64 GB RAM per node Hadoop Cluster

- 16 GB RAM allocated for the HBase Region Server Maximum Memory

- 16 GB RAM allocated for the HBase Master Maximum Memory

## Memory Calculations for a Cluster

The actual memory required by Attivio on the cluster is determined by the Hadoop components, Attivio index processes, and Attivio modules. If the cluster does not have enough memory to launch a process, it will silently wait until enough resources become available.

To calculate the amount of memory required for the Hadoop cluster and determine what `yarn.nodemanager.resource.memory-mb` should be set to, follow the memory calculations described [here](#).

For Attivio systems consisting of more than 50 nodes, it is recommended that you set the `maxClientCnxns` property of your Hadoop's ZooKeeper instance to a value larger than the default of 60. The exact recommended setting for this property will depend on the size of your system. Please consult Attivio support for additional assistance.

## Best Practices

### Virtual Environments

#### Oversubscription of Resources

⚠ Attivio cannot meet production SLAs in virtual environments where physical resources (CPU, memory and I/O) are over-subscribed and/or not reserved due to the unpredictable nature of the overall system performance. All virtual resources for Attivio VMs should be backed with physical resources on a 1:1 basis. Over-subscribed / non-reserved virtual environments can be used for development and QA, however performance may vary significantly based on available resources.

#### VMware with VMotion

⚠ Applications using large memory footprints such as Attivio are particularly sensitive to VMotion pauses as it can take significant time to copy in-memory data. If you are hosting Attivio in a cluster with DRS/VMotion enabled, we recommend using VM/Host Affinity rules on the Attivio VMs to avoid VMotion on these systems during normal operation.

### Antivirus and Firewall

- Attivio recommends that no antivirus or firewall software be running on the Attivio server nodes. Security software adversely affects Attivio's execution and performance.
- If antivirus software is running on Attivio server nodes, it should be configured to ignore the running Attivio server processes and avoid scanning the Attivio application and data directories.
- If firewall software is running on Attivio server nodes, it must open all HTTP ports configured for the running Attivio server. See the [Security Guide](#) for more information about what ports Attivio uses for network communication.

Some (not all) systems prompt you to update your firewall the first time the agents start. If no such prompt appears, check the firewall settings to ensure it allows the ports configured for running Attivio. The firewall must allow these ports for the Attivio nodes to start.

#### Digital Guardian

The Verdasys [Digital Guardian](#) data-protection platform has been known to chronically shut down Attivio ingestion nodes. It should be disabled or configured to ignore Attivio applications, port ranges, and file directories.

### Cloud Deployments

Amazon's default Linux settings set the max open file (`nofile`) and the max user process (`nproc`) limits to very low levels which are incompatible with Attivio. It is a good policy to set these properties to high values whether on a cloud server or not. See the [Recommended Linux Settings](#) above and Linux limit instructions [here](#). Also, Amazon's Linux servers do not have any swap configured by default. Configuring a minimum of 16GB of swap space is a reasonable starting point for Attivio hosts.

Although we have encountered this issue with Amazon's cloud, it may also be true of other cloud providers' Linux images and other Linux distributions.

### One Project Per Host

It is a Best Practice to run only one Attivio project at a time on a given host. If a multi-project host crashes, it can be very difficult to determine the cause.

