

# Kafka



Kafka requires a fairly small amount of resources, especially with some configuration tuning. By default, Kafka, can run on as little as 1 core and 1GB memory with storage scaled based on requirements for data retention.

CPU is rarely a bottleneck because Kafka is I/O heavy, but a moderately-sized CPU with enough threads is still important to handle concurrent connections and background tasks.

Kafka brokers tend to have a similar hardware profile to HDFS data nodes. How you build them depends on what is important for your Kafka use cases.

Use the following guidelines:

To affect performance of these features:	Adjust these parameters:
Message Retention	Disk size
Client Throughput (Producer & Consumer)	Network capacity
Producer throughput	Disk I/O
Consumer throughput	Memory

A common choice for a Kafka node is as follows:

Component	Memory/Java Heap	CPU	Disk
Broker	<ul style="list-style-type: none"> <li>RAM: 64 GB</li> <li>Recommended Java heap: 4 GB</li> </ul> <p>Set this value using the Java Heap Size of Broker Kafka configuration property.</p>	12- 24 cores	<ul style="list-style-type: none"> <li>1 HDD For operating system</li> <li>1 HDD for Zookeeper dataLogDir</li> <li>10- HDDs, using Raid 10, for Kafka data</li> </ul>
Cruise Control	1 GB	1 core  <b>Note</b>  A moderately-sized CPU with enough threads is important to handle metric fetching from Kafka and background tasks.	Because Cruise Control stores its data in Kafka the storage requirements will depend on the retention settings of the related Kafka topics.
Kafka Connect	0.5 - 4 GB heap size depending on the Connectors in use.	4 cores  <b>Note</b>  Depends on the Connectors in use.	
MirrorMaker	1 GB heap  Set this value using the Java Heap Size of MirrorMaker Kafka configuration property.	1 core per 3-4 streams	No disk space needed on MirrorMaker instance. Destination brokers should have sufficient disk space to store the topics being copied over.
Schema Registry	1 GB heap	2 cores	1 MB  Serialization JAR files may be uploaded and may be of any size. The disk usage depends on the JAR files uploaded. The files may be stored locally on the same host where SchemaRegistry is running or in HDFS if available.
Streams Messaging Manager	8 GB heap	8 cores	5 GB
<b>Note</b>			

Component	Memory/Java Heap	CPU	Disk
The hardware requirements for SMM depends on the number of Kafka partitions.			
Streams Replication Manager	<ul style="list-style-type: none"> <li>• 1 GB heap for SRM driver</li> <li>• 1 GB heap for SRM Service</li> </ul>	The performance of the SRM driver is mostly impacted by network throughput and latency.	No resources required

**Networking requirements:** Gigabit Ethernet or 10 Gigabit Ethernet. Avoid clusters that span multiple data centers.

**Kafka and Zookeeper:** It is common to run ZooKeeper on 3 broker nodes that are dedicated for Kafka. However, for optimal performance Cloudera recommends the usage of dedicated Zookeeper hosts. This is especially true for larger, production environments.