

Tuning

The default settings in etcd should work well for installations on a local network where the average network latency is low. However, when using etcd across multiple data centers or over networks with high latency you may need to tweak the heartbeat interval and election timeout settings.

The network isn't the only source of latency. Each request and response may be impacted by slow disks on both the leader and follower. Each of these timeouts represents the total time from request to successful response from the other machine.

Time Parameters

The underlying distributed consensus protocol relies on two separate time parameters to ensure that nodes can handoff leadership if one stalls or goes offline. The first parameter is called the *Heartbeat Interval*. This is the frequency with which the leader will notify followers that it is still the leader. For best practices, the parameter should be set around round-trip time between members. By default, etcd uses a 100ms heartbeat interval.

The second parameter is the *Election Timeout*. This timeout is how long a follower node will go without hearing a heartbeat before attempting to become leader itself. By default, etcd uses a 1000ms election timeout.

Adjusting these values is a trade off. The value of heartbeat interval is recommended to be around the maximum of average round-trip time (RTT) between members, normally around 0.5-1.5x the round-trip time. If heartbeat interval is too low, etcd will send unnecessary messages that increase the usage of CPU and network resources. On the other side, a too high heartbeat interval leads to high election timeout. Higher election timeout takes longer time to detect a leader failure. The easiest way to measure round-trip time (RTT) is to use [PING utility](#).

The election timeout should be set based on the heartbeat interval and average round-trip time between members. Election timeouts must be at least 10 times the round-trip time so it can account for variance in your network. For example, if the round-trip time between your members is 10ms then you should have at least a 100ms election timeout.

You should also set your election timeout to at least 5 to 10 times your heartbeat interval to account for variance in leader replication. For a heartbeat interval of 50ms you should set your election timeout to at least 250ms - 500ms.

The upper limit of election timeout is 50000ms (50s), which should only be used when deploying a globally-distributed etcd cluster. A reasonable round-trip time for the continental United States is 130ms, and the time between US and Japan is around 350-400ms. If your network has uneven performance or regular packet delays/loss then it is possible that a couple of retries may be necessary to successfully send a packet. So 5s is a safe upper limit of global round-trip time. As the election timeout should be an order of magnitude bigger than broadcast time, in the case of ~5s for a globally distributed cluster, then 50 seconds becomes a reasonable maximum.

The heartbeat interval and election timeout value should be the same for all members in one cluster. Setting different values for etcd members may disrupt cluster stability.

You can override the default values on the command line:

```
# Command line arguments:
```

```
$ etcd -heartbeat-interval=100 -election-timeout=500
```

```
# Environment variables:
```

```
$ ETCD_HEARTBEAT_INTERVAL=100 ETCD_ELECTION_TIMEOUT=500 etcd
```

The values are specified in milliseconds.

Snapshots

etcd appends all key changes to a log file. This log grows forever and is a complete linear history of every change made to the keys. A complete history works well for lightly used clusters but clusters that are heavily used would carry around a large log.

To avoid having a huge log etcd makes periodic snapshots. These snapshots provide a way for etcd to compact the log by saving the current state of the system and removing old logs.

Snapshot Tuning

Creating snapshots can be expensive so they're only created after a given number of changes to etcd. By default, snapshots will be made after every 10,000 changes. If etcd's memory usage and disk usage are too high, you can lower the snapshot threshold by setting the following on the command line:

```
# Command line arguments:
```

```
$ etcd -snapshot-count=5000
```

```
# Environment variables:
```

```
$ ETCD_SNAPSHOT_COUNT=5000 etcd
```