

QuickStart

获取指定版本的etcd二进制文件

```
export ETCD_VER=v3.5.0
export ETCD_BIN=/tmp/test-etcd

export GOOGLE_URL=https://storage.googleapis.com/etcd
export GITHUB_URL=https://github.com/etcd-io/etcd/releases/download

export DOWNLOAD_URL=${GITHUB_URL}

rm -f /tmp/etcd-${ETCD_VER}-linux-amd64.tar.gz
rm -rf $ETCD_BIN
mkdir -p $ETCD_BIN

curl -L ${DOWNLOAD_URL}/${ETCD_VER}/etcd-${ETCD_VER}-linux-amd64.tar.gz -o /tmp/etcd-${ETCD_VER}-linux-amd64.tar.gz

tar xzvf /tmp/etcd-${ETCD_VER}-linux-amd64.tar.gz -C $ETCD_BIN --strip-components=1
```

启动etcd

```
[root@etcd01 test-etcd]# ./etcd
{"level":"info","ts":"2021-08-05T09:53:36.043+0800","caller":"etcdmain/etcd.go:72","msg":"Running:", "args":["./etcd"]}
{"level":"warn","ts":"2021-08-05T09:53:36.048+0800","caller":"etcdmain/etcd.go:104","msg":"'data-dir' was empty; using default","data-dir":"default.etcd"}
{"level":"info","ts":"2021-08-05T09:53:36.048+0800","caller":"embed/etcd.go:131","msg":"configuring peer listeners","listen-peer-urls":["http://localhost:2380"]}
{"level":"info","ts":"2021-08-05T09:53:36.049+0800","caller":"embed/etcd.go:139","msg":"configuring client listeners","listen-client-urls":["http://localhost:2379"]}
{"level":"info","ts":"2021-08-05T09:53:36.049+0800","caller":"embed/etcd.go:307","msg":"starting an etcd server","etcd-version":"3.5.0","git-sha":"946a5a6f2","go-version":"go1.16.3","go-os":"linux","go-arch":"amd64","max-cpu-set":2,"max-cpu-available":2,"member-initialized":false,"name":"default","data-dir":"default.etcd","wal-dir":"","wal-dir-dedicated":"","member-dir":"default.etcd/member","force-new-cluster":false,"heartbeat-interval":"100ms","election-timeout":"1s","initial-election-tick-advance":true,"snapshot-count":100000,"snapshot-catchup-entries":5000,"initial-advertise-peer-urls":["http://localhost:2380"],"listen-peer-urls":["http://localhost:2380"],"advertise-client-urls":["http://localhost:2379"],"listen-client-urls":["http://localhost:2379"],"listen-metrics-urls":[],"cors":["*"],"host-whitelist":["*"],"initial-cluster":"default=http://localhost:2380","initial-cluster-state":"new","initial-cluster-token":"etcd-cluster","quota-size-bytes":2147483648,"pre-vote":true,"initial-corrupt-check":false,"corrupt-check-time-interval":"0s","auto-compaction-mode":"periodic","auto-compaction-retention":"0s","auto-compaction-interval":"0s","discovery-url":"","discovery-proxy":"","downgrade-check-interval":"5s"}
{"level":"info","ts":"2021-08-05T09:53:36.050+0800","caller":"etcdserver/backend.go:81","msg":"opened backend
```

```
db", "path": "default/etcd/member/snap/db", "took": "1.144864ms"}
{"level": "info", "ts": "2021-08-05T09:53:36.053+0800", "caller": "etcdserver/raft.go:448", "msg": "starting local member", "local-member-id": "8e9e05c52164694d", "cluster-id": "cdf818194e3a8c32"}
{"level": "info", "ts": "2021-08-05T09:53:36.053+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "8e9e05c52164694d switched to configuration voters=()"}
{"level": "info", "ts": "2021-08-05T09:53:36.053+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "8e9e05c52164694d became follower at term 0"}
{"level": "info", "ts": "2021-08-05T09:53:36.053+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "newRaft 8e9e05c52164694d [peers: [], term: 0, commit: 0, applied: 0, lastindex: 0, lastterm: 0]"}
{"level": "info", "ts": "2021-08-05T09:53:36.053+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "8e9e05c52164694d became follower at term 1"}
{"level": "info", "ts": "2021-08-05T09:53:36.053+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "8e9e05c52164694d switched to configuration voters=(10276657743932975437)"}
{"level": "warn", "ts": "2021-08-05T09:53:36.055+0800", "caller": "auth/store.go:1220", "msg": "simple token is not cryptographically signed"}
{"level": "info", "ts": "2021-08-05T09:53:36.057+0800", "caller": "mvcc/kvstore.go:415", "msg": "kvstore restored", "current-rev": 1}
{"level": "info", "ts": "2021-08-05T09:53:36.058+0800", "caller": "etcdserver/quota.go:94", "msg": "enabled backend quota with default value", "quota-name": "v3-applier", "quota-size-bytes": 2147483648, "quota-size": "2.1 GB"}
{"level": "info", "ts": "2021-08-05T09:53:36.060+0800", "caller": "etcdserver/server.go:843", "msg": "starting etcd server", "local-member-id": "8e9e05c52164694d", "local-server-version": "3.5.0", "cluster-version": "to_be_decided"}
{"level": "info", "ts": "2021-08-05T09:53:36.061+0800", "caller": "etcdserver/server.go:728", "msg": "started as single-node; fast-forwarding election ticks", "local-member-id": "8e9e05c52164694d", "forward-ticks": 9, "forward-duration": "900ms", "election-ticks": 10, "election-timeout": "1s"}
{"level": "info", "ts": "2021-08-05T09:53:36.061+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "8e9e05c52164694d switched to configuration voters=(10276657743932975437)"}
{"level": "info", "ts": "2021-08-05T09:53:36.061+0800", "caller": "membership/cluster.go:393", "msg": "added member", "cluster-id": "cdf818194e3a8c32", "local-member-id": "8e9e05c52164694d", "added-peer-id": "8e9e05c52164694d", "added-peer-peer-urls": ["http://localhost:2380"]}
{"level": "info", "ts": "2021-08-05T09:53:36.063+0800", "caller": "embed/etcd.go:276", "msg": "now serving peer/client/metrics", "local-member-id": "8e9e05c52164694d", "initial-advertise-peer-urls": ["http://localhost:2380"], "listen-peer-urls": ["http://localhost:2380"], "advertise-client-urls": ["http://localhost:2379"], "listen-client-urls": ["http://localhost:2379"], "listen-metrics-urls": []}
{"level": "info", "ts": "2021-08-05T09:53:36.063+0800", "caller": "embed/etcd.go:580", "msg": "serving peer traffic", "address": "127.0.0.1:2380"}
{"level": "info", "ts": "2021-08-05T09:53:36.063+0800", "caller": "embed/etcd.go:552", "msg": "cmux::serve", "address": "127.0.0.1:2380"}
{"level": "info", "ts": "2021-08-05T09:53:36.367+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "8e9e05c52164694d is starting a new election at term 1"}
{"level": "info", "ts": "2021-08-05T09:53:36.367+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "8e9e05c52164694d became pre-candidate at term 1"}
{"level": "info", "ts": "2021-08-05T09:53:36.367+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "8e9e05c52164694d received MsgPreVoteResp from 8e9e05c52164694d at term 1"}
{"level": "info", "ts": "2021-08-05T09:53:36.368+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "8e9e05c52164694d became candidate at term 2"}
{"level": "info", "ts": "2021-08-05T09:53:36.368+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "8e9e05c52164694d became leader at term 2"}
```

```
05T09:53:36.368+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "8e9e05c52164694d
received MsgVoteResp from 8e9e05c52164694d at term 2"}
{"level": "info", "ts": "2021-08-
05T09:53:36.368+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "8e9e05c52164694d
became leader at term 2"}
{"level": "info", "ts": "2021-08-
05T09:53:36.368+0800", "logger": "raft", "caller": "etcdserver/zap_raft.go:77", "msg": "raft.node:
8e9e05c52164694d elected leader 8e9e05c52164694d at term 2"}
{"level": "info", "ts": "2021-08-
05T09:53:36.369+0800", "caller": "etcdserver/server.go:2027", "msg": "published local member to
cluster through raft", "local-member-id": "8e9e05c52164694d", "local-member-attributes": "
{Name: default ClientURLs: [http://localhost:2379]}"}, "request-
path": "/0/members/8e9e05c52164694d/attributes", "cluster-id": "cdf818194e3a8c32", "publish-
timeout": "7s"}
{"level": "info", "ts": "2021-08-
05T09:53:36.370+0800", "caller": "etcdserver/server.go:2476", "msg": "setting up initial cluster
version using v2 API", "cluster-version": "3.5"}
{"level": "info", "ts": "2021-08-05T09:53:36.370+0800", "caller": "embed/serve.go:98", "msg": "ready to
serve client requests"}
{"level": "info", "ts": "2021-08-
05T09:53:36.370+0800", "caller": "membership/cluster.go:531", "msg": "set initial cluster
version", "cluster-id": "cdf818194e3a8c32", "local-member-id": "8e9e05c52164694d", "cluster-
version": "3.5"}
{"level": "info", "ts": "2021-08-05T09:53:36.371+0800", "caller": "api/capability.go:75", "msg": "enabled
capabilities for version", "cluster-version": "3.5"}
{"level": "info", "ts": "2021-08-
05T09:53:36.371+0800", "caller": "etcdserver/server.go:2500", "msg": "cluster version is
updated", "cluster-version": "3.5"}
{"level": "info", "ts": "2021-08-
05T09:53:36.371+0800", "caller": "etcdmain/main.go:47", "msg": "notifying init daemon"}
{"level": "info", "ts": "2021-08-
05T09:53:36.371+0800", "caller": "etcdmain/main.go:53", "msg": "successfully notified init daemon"}
{"level": "info", "ts": "2021-08-05T09:53:36.372+0800", "caller": "embed/serve.go:140", "msg": "serving
client traffic insecurely; this is strongly discouraged!", "address": "127.0.0.1:2379"}
```

设置kv, 读取kv

```
[root@etcd01 test-etcd]# ./etcdctl put greeting "Hello, etcd"
OK
[root@etcd01 test-etcd]# ./etcdctl get greeting
greeting
Hello, etcd
```

ETCD概要

什么是etcd

ETCD是一个分布式 (distributed) **强一致性** (consistent) kv存储

一致性通过ETCD的Leader保证, 但是client无需知道谁是leader, 所有节点会将请求转发给leader

配置文件

容易混淆的配置参数

- `listen-<client,peer>-urls` , 本地etcd server的地址绑定
- `advertise-client-urls` , 节点间通信地址, 不要使用 `localhost` 或者 `0.0.0.0`
- `initial-advertise-peer-urls`

为防止脑裂, 不要尝试修改 `--initial-advertise-peer-urls` 修改成员, 要用 `etcdctl member update`

部署

系统要求

- 推荐高性能的ssd存储
- 默认2G存储配额, 防止性能降级和未预期的kv重载
- 8G RAM 是推荐的maximum大小

奇数节点cluster members

集群的状态变更需要 `quorum (n/2)+1` 个节点同意, 当n为偶数是, 节点出现异常时, `quorum` 可能没有变化, 从而造成集群的容错性降低

最大集群规模

理论上无限制, 但实践经验建议最大规模7节点, google的实践为5节点, 节点数量的增多对写入性能存在影响。

故障容忍能力

Cluster Size	Majority	Failure Tolerance
1	1	0
2	2	0
3	2	1
4	3	1
5	3	2
6	4	2
7	4	3
8	5	3

Cluster Size	Majority	Failure Tolerance
9	5	4

先删除异常节点，再添加新节点

都与quorum有关

数据模型

适用于不经常变动的KV数据

etcd是多版本的数据结构，kv的更新，并不变更原有结构，而是生成新的结构。原有结构仍可访问可监听。为防止数据的无限增长，可压缩删除最旧的版本数据。

逻辑视图

扁平的二进制键空间，按序存储的keys索引。键空间维护多版本。

物理视图

Etcd将物理数据以键值对的形式存储在持久的b+树中

key 数据结构3-tuple(major,sub,type)