# getaip-server configuration

This reference lists the complete command-line configuration of the product-neutral getaip-server gateway. Use it to look up process identity, native authentication, durable state, callbacks, connector-fleet routing, NATS, delegation, and M

| | |
|---|---|
| SECTION | Deploy and Operate |
| TYPE | Reference |
| PROTOCOL | AIP 1.0 |
| SOURCE | docs/reference/configuration.md |

The option surface is implemented by `CoreArgs` at source revision `d7cce13d1d555644d04a4d73c66c95b113737635`. It does not include Cal.diy, Hermes Agent, Chatwoot, Dify, CrewAI, or Twenty provider settings. Those belong to standalone connector hosts. Existing bundled Cal.diy or Hermes Agent deployments use the separate legacy bundled deployment.

All defaults below are implementation defaults for the reviewed revision, not capacity recommendations or evidence that a deployment has qualified for its intended load.

## Value resolution

Most single-value settings resolve in this order:

1. command-line option;
2. corresponding environment variable;
3. implementation default or unset.

Boolean command-line and environment switches are combined: either one can enable the behavior. Repeatable command-line lists are extended with values from their environment list. The complete environment inventory and delimiter rules live in Environment variables.

There are two precedence exceptions. `GETAIP_SERVER_NATIVE_PRINCIPAL` overrides the value parsed from `--native-principal`, and `GETAIP_SERVER_MCP_PRINCIPAL` overrides the value parsed from `--mcp-principal`. `GETAIP_SERVER_NATS_REQUEST_TIMEOUT_MS` also overrides the CLI timeout value. Configure only one source for these values.

Inline and file forms of the same secret are mutually exclusive even when one form comes from the environment. Prefer the file form when it exists.

## Process identity and modes

| Option | Default | Meaning and constraint |
| --- | --- | --- |
| `--bind ADDRESS` | `127.0.0.1:8080` | HTTP listener. Insecure development mode is restricted to a loopback address. |
| `--public-base-url URL` | Unset | External origin advertised for A2A and capability discovery. It must be an origin without credentials, path, query, or fragment. |
| `--service-id PRINCIPAL_ID` | `agent:getaip:server:local` | Stable principal reported in the manifest and used by configured signing identities. |
| `--trust-domain DOMAIN` | Unset | Trust domain reported by the deployment. Fleet and delegation composition use `local` when it is unset. |
| `--print-manifest` | Off | Prepare the complete deployment, print its manifest as JSON, and exit without serving. |
| `--mcp-stdio` | Off | Prepare the deployment and serve newline-delimited MCP through stdin and stdout instead of HTTP. |

A non-loopback listener requires an explicit HTTPS public base URL. Insecure development is restricted to a loopback listener. HTTP public origins are accepted only when both the listener and origin are loopback-scoped and insecure development is explicit.

`--print-manifest` takes precedence over `--mcp-stdio` when both are present. Startup failures are emitted as structured JSON on stderr and exit with a non-zero status.

## Native authentication and trust

| Option | Default | Meaning and constraint |
| --- | --- | --- |
| `--require-signed-envelopes` | Signed by default | Explicitly require Ed25519 signatures on incoming native AIP envelopes. |
| `--trusted-signer DID=PRINCIPAL_ID` | None | Add one Ed25519 `did:key` to principal binding; repeatable. |
| `--trusted-signer-file PATH` | Unset | Load additional signer bindings from a bounded JSON file. |
| `--native-bearer-token TOKEN` | Unset | Inline bearer token for native HTTP routes. Prefer the file form. |
| `--native-bearer-token-file FILE` | Unset | Load the native bearer token from a mode-0600 file. |
| `--native-principal PRINCIPAL_ID` | `service:getaip:server:http-edge` | Principal established after native bearer authentication. |
| `--native-tenant-id TENANT_ID` | Unset | Verified tenant bound to the native bearer identity for fleet discovery and routing. |
| `--native-principal-scope SCOPE` | None | Add one scope to the authenticated native principal; repeatable. |
| `--trusted-identity-file PATH` | Deny all | Load revisioned trusted principal, tenant, credential, and identity bindings. |
| `--approval-authority-file PATH` | Deny all | Load revisioned approval-authority memberships. |
| `--allow-insecure-development` | Off | Permit unauthenticated native traffic only on a loopback listener. |

Incoming envelopes are signed by default unless `GETAIP_SERVER_ALLOW_UNSIGNED_ENVELOPES` explicitly disables the default and neither the CLI nor `GETAIP_SERVER_REQUIRE_SIGNED_ENVELOPES` restores it. Outside insecure development, unsigned native envelopes still require native bearer authentication. Caller-supplied `from` data is not a replacement for an authenticated edge identity.

Inline and file signer bindings are merged. Every signer DID must be a valid Ed25519 `did:key`, and DIDs must be unique across both sources. The signer file uses this exact shape and rejects unknown fields:

```
JSON
[
  {
    "signer_did": "did:key:<ed25519-multibase-value>",
    "principal_id": "service:trusted-edge"
  }
]
```

The trusted identity file rejects an empty directory, zero revisions, revoked or expired entries, duplicate principals, and invalid tenant or credential bindings. The approval file rejects an empty directory, zero revisions, duplicate principals, and invalid memberships. When either file is absent, the corresponding resolver denies the operation rather than trusting an implicit identity.

## Callback policy and runtime work budgets

| Option | Default | Meaning and constraint |
| --- | --- | --- |
| `--callback-allowed-host HOST` | None | Add one exact outbound callback host; repeatable. |
| `--callback-signing-seed-hex HEX` | Unset | Inline 32-byte Ed25519 seed as 64 hexadecimal characters. Prefer the file form. |
| `--callback-signing-seed-file FILE` | Unset | Load the response and callback signing seed from a mode-0600 file. |
| `--a2a-push-encryption-key-hex HEX` | Unset | Inline 32-byte key used to encrypt stored A2A push credentials. |
| `--callback-allow-http` | Off | Permit plaintext HTTP callback targets. |
| `--callback-allow-private-networks` | Off | Permit private, loopback, link-local, or otherwise non-public callback targets. |
| `--callback-max-in-flight COUNT` | `64` | Maximum callback deliveries executing concurrently in one runtime. |
| `--callback-recovery-batch COUNT` | `64` | Maximum callback records leased by one recovery cycle. |
| `--reconciliation-max-in-flight COUNT` | `32` | Maximum transaction reconciliations executing concurrently. |

An allowed callback host requires a response and callback signing seed. An A2A push encryption key also requires that signing seed. Each work budget must be greater than zero.

The callback request timeout and maximum response size are environment-only settings in this revision. Their defaults are 5,000 milliseconds and 4 MiB. See the environment reference for their exact names.

The HTTP and private-network switches weaken the default destination policy. They affect callbacks, HTTP delegation destinations, and other routes that use the shared callback policy. Keep them disabled unless the deployment owns and reviews the complete isolated network path.

## Runtime storage

| Option | Default | Meaning and constraint |
| --- | --- | --- |
| `--storage-dir DIR` | Unset | Durable local runtime and auxiliary file state for a single-process topology. |
| `--postgres-url URL` | Unset | Inline clustered runtime PostgreSQL URL. Prefer the file form. |
| `--postgres-url-file FILE` | Unset | Load the clustered runtime PostgreSQL URL from a mode-0600 file. |

`--postgres-url` and `--postgres-url-file` are mutually exclusive. PostgreSQL is the core runtime store when either form resolves. Otherwise, a configured storage directory supplies durable local runtime state; without either, core runtime state is in memory.

The storage directory can still own MCP correlation and replay files when PostgreSQL owns core runtime state. Treat both stores as one recovery boundary when both are configured. Use PostgreSQL for replicas that share one logical runtime; do not point several processes at one local storage directory.

## Connector-fleet activation and transport

Fleet mode is disabled when no connector-registry URL resolves. Supplying any other fleet, topology, admission, or connector-event CLI option without that URL is a configuration error. The environment reference records the exact activation behavior for environment-only settings.

| Option | Default | Meaning and constraint |
|---|---|---|
| `--connector-registry-url URL` | Unset | Inline PostgreSQL data-plane URL for catalog and routing. Prefer the file form. |
| `--connector-registry-url-file FILE` | Unset | Load the registry data-plane URL from a mode-0600 file. |
| `--connector-registry-data-max-connections COUNT` | `32` | Maximum registry data-plane connections; greater than zero. |
| `--connector-registry-acquire-timeout-ms MILLISECONDS` | `5000` | Maximum registry connection wait; greater than zero and no more than 60 seconds. |
| `--connector-fleet-signing-seed-file FILE` | Required in fleet mode | Mode-0600 file containing the gateway's 32-byte Ed25519 seed as hexadecimal text. |
| `--connector-fleet-callback-url URL` | Unset | Central streamed-action callback URL ending exactly in `/aip/v1/connector-callbacks`. |
| `--connector-fleet-allowed-host HOST` | None | Add one exact connector-host DNS name or IP literal; repeatable. |
| `--connector-fleet-trust-registry-endpoints` | Off | Derive the destination host allowlist from admitted registry endpoints per route. |
| `--connector-fleet-allow-http` | Off | Permit plaintext connector-host or fleet callback URLs. |
| `--connector-fleet-allow-private-networks` | Off | Permit private, loopback, link-local, or otherwise non-public connector-host addresses. |
| `--connector-fleet-timeout-ms MILLISECONDS` | `5000` | Timeout for one connector-host request; greater than zero. |
| `--connector-fleet-max-response-bytes BYTES` | `4194304` | Maximum accepted connector-host response; greater than zero. |
| `--connector-fleet-retry-budget COUNT` | `2` | Transport retries after the first connector-host attempt. |
| `--connector-fleet-max-cached-clients COUNT` | `256` | Maximum cached connector-host HTTP connection pools; greater than zero. |
| `--connector-fleet-tls-ca-file FILE` | Unset | Additional PEM root used to verify private-PKI connector-host TLS. |

Fleet mode requires a registry URL, signing-seed file, and one of these destination policies:

- at least one explicit `--connector-fleet-allowed-host`; or
- `--connector-fleet-trust-registry-endpoints`.

The gateway opens the registry through the data-plane constructor. The configured database role must not have registry-administrator authority. The additional CA augments the public trust roots; it does not disable certificate or hostname verification.

The fleet callback URL must be an absolute HTTP(S) URL with no credentials, query, or fragment and the exact callback path. Plain HTTP requires the explicit fleet HTTP exception. Configuring the URL makes remote streaming available; without it, the remote implementation support does not advertise streaming.

## Fleet topology and local admission

| Option | Default | Meaning and constraint |
| --- | --- | --- |
| `--connector-fleet-region REGION` | No preference | Prefer this region for a new route assignment. |
| `--connector-fleet-zone ZONE` | No preference | Prefer this zone for a new route assignment. |
| `--connector-fleet-capacity-class CLASS` | No preference | Require this capacity class for a new route assignment. |
| `--connector-fleet-disable-cross-region-failover` | Off | Reject fallback to another region for a new assignment. |
| `--connector-fleet-max-in-flight COUNT` | `1024` | Maximum remote connector actions executing through this daemon. |
| `--connector-fleet-max-in-flight-per-tenant COUNT` | `128` | Maximum executing remote actions for one verified tenant. |
| `--connector-fleet-max-queued COUNT` | `4096` | Maximum remote actions waiting for a local dispatch permit. |
| `--connector-fleet-max-queued-per-tenant COUNT` | `512` | Maximum waiting actions for one verified tenant. |
| `--connector-fleet-max-queue-bytes BYTES` | `67108864` | Maximum canonical bytes retained by all waiting actions. |
| `--connector-fleet-max-request-bytes BYTES` | `4194304` | Maximum canonical size of one remote action. |
| `--connector-fleet-max-queue-age-ms MILLISECONDS` | `30000` | Maximum time an action may wait for local admission. |
| `--connector-fleet-tenant-weight TENANT_ID=WEIGHT` | None | Add one positive scheduling weight for a verified tenant; repeatable. |

Every numeric admission bound must be greater than zero. The per-tenant in-flight bound cannot exceed the global in-flight bound, and the per-tenant queue bound cannot exceed the global queue bound. Tenant IDs in weight entries contain 1–256 bytes, weights are positive, and a tenant can appear only once across the merged CLI and environment list.

Topology preferences apply when the registry creates a new assignment. An existing action remains pinned to its assigned replica; changing a preference does not move active work.

## Connector-event and stream ingress

These options configure the gateway's shared authenticated connector-event ingress. They are valid only in fleet mode.

| Option | Default | Meaning and constraint |
| --- | --- | --- |
| `--connector-event-max-envelope-bytes BYTES` | `4194304` | Maximum serialized signed envelope size. |
| `--connector-event-max-events-per-envelope COUNT` | `100` | Maximum events in one existing AIP event stream; range 1–1000. |
| `--connector-event-max-event-bytes BYTES` | `262144` | Maximum serialized size of one enriched event; no more than the envelope bound. |
| `--connector-stream-max-chunk-bytes BYTES` | `262144` | Maximum serialized connector stream chunk; no more than the envelope bound. |
| `--connector-event-max-in-flight COUNT` | `128` | Maximum concurrent connector-event storage operations. |
| `--connector-event-max-age-seconds SECONDS` | `86400` | Maximum accepted age of a connector-originated event. |
| `--connector-event-max-future-skew-seconds SECONDS` | `300` | Maximum accepted future clock skew for a connector event. |
| `--connector-stream-callback-max-age-seconds SECONDS` | `300` | Maximum accepted age of a signed connector stream callback. |

Every limit must be greater than zero. Event and stream envelopes still require the admitted route identity, signature, timestamp, sequence, replay, and assignment checks; increasing a size or time limit does not bypass those checks.

## Native NATS

| Option | Default | Meaning and constraint |
| --- | --- | --- |
| `--nats-url URL` | Unset | Enable the native NATS request/reply listener. |
| `--nats-username USERNAME` | Unset | NATS username; requires a password file and server URL. |
| `--nats-password-file FILE` | Unset | Mode-0600 NATS password file; requires a username and server URL. |
| `--nats-trust-domain DOMAIN` | Daemon trust domain | Trust-domain subject segment. |
| `--nats-service SERVICE` | `getaip-server` | Service subject segment. |
| `--nats-version VERSION` | `v1` | Version subject segment. |
| `--nats-queue-group GROUP` | Unset | Optional queue group for scaled consumers. |
| `--nats-request-timeout-ms MILLISECONDS` | `30000` | Outbound NATS request timeout. |

NATS is disabled when the server URL is absent. Supplying authentication without a server URL is an error. Username and password file must appear together; configuring neither selects an unauthenticated connection and does not create TLS, account, or subject authorization outside the process.

## Remote delegation routes

| Option | Default | Meaning and constraint |
| --- | --- | --- |
| `--delegation-http-route SPEC` | None | Route one delegate principal to an authenticated native HTTP peer; repeatable. |
| `--delegation-nats-route SPEC` | None | Route one child capability to an authenticated native NATS peer; repeatable. |

HTTP route syntax is:

```
TEXT
DELEGATE_ID=URL,PEER_ID,PEER_DID[,TRUST_DOMAIN]
```

NATS route syntax is:

```
TEXT
CAPABILITY_ID=SERVER_URL,SUBJECT,PEER_ID,PEER_DID[,TIMEOUT_MS[,TRUST_DOMAIN]]
```

The NATS route timeout defaults to 30,000 milliseconds. Both route forms use the daemon trust domain when the optional route trust domain is absent. Routes require a response and callback signing seed, a valid peer principal and `did:key`, and a destination accepted by the shared callback policy. The daemon validates each route before serving.

## MCP compatibility profile

| Option | Default | Meaning and constraint |
| --- | --- | --- |
| `--mcp-bearer-token TOKEN` | Unset | Static bearer token for MCP HTTP. Do not combine with introspection. |
| `--mcp-resource RESOURCE` | Derived or unset | Protected-resource identifier advertised for MCP HTTP. |
| `--mcp-authorization-server ISSUER` | None | Add one advertised authorization server; repeatable. |
| `--mcp-scope SCOPE` | None | Add one advertised supported scope; repeatable. |
| `--mcp-required-scope SCOPE` | None | Add one scope required to establish an MCP HTTP session; repeatable. |
| `--mcp-introspection-url HTTPS_URL` | Unset | RFC 7662 token introspection endpoint. |
| `--mcp-introspection-issuer ISSUER` | Unset | Trusted issuer represented by the introspection endpoint. |
| `--mcp-introspection-client-id CLIENT_ID` | Unset | OAuth client ID used for introspection. |
| `--mcp-introspection-client-secret-file FILE` | Unset | Mode-0600 introspection client-secret file. |
| `--mcp-introspection-allow-loopback-http` | Off | Permit a loopback plaintext introspection endpoint. |
| `--mcp-resource-documentation URL` | Unset | Human-readable protected-resource documentation URL. |
| `--mcp-allowed-origin ORIGIN` | None | Add one browser origin allowed to call MCP HTTP; repeatable. |
| `--mcp-principal PRINCIPAL_ID` | `service:getaip:server:mcp-edge` | Principal established after MCP transport authentication. |
| `--mcp-principal-scope SCOPE` | None | Add one scope to that authenticated principal; repeatable. |

If a static token, authorization server, supported scope, or documentation URL is set without an explicit resource, the implementation derives `http:///mcp`. Set `--mcp-resource` explicitly for a public deployment so the protected-resource identifier matches the external origin.

Introspection is an all-or-none group: URL, issuer, client ID, and secret file must all resolve. It also requires a protected-resource policy, rejects a simultaneous static token, and requires the introspection issuer to appear in the advertised authorization-server list. The URL must use HTTPS unless the explicit exception permits a loopback HTTP fixture.

The protected-resource configuration permits loopback browser origins in addition to the explicit origin list. Treat MCP stdio as a separate local transport whose authentication boundary is the trusted spawning process.

## File contracts

Secret files must be regular non-symlink files. On Unix the loader rejects any group or other permission bit and reports the expected contract as mode `0600`; an owner-readable `0400` mount also satisfies the implemented bitmask.

Secret values are bounded by their owning option and decoded as UTF-8 where required. The common inline-or-file text resolver trims values and rejects an empty result; other secret consumers apply their own parser.

Trusted signer, trusted identity, approval authority, and fleet TLS CA files use the secure-configuration loader. They must be regular non-symlink files, cannot exceed 1 MiB, and cannot be group- or world-writable on Unix. Read-only access for a deployment-owned group is therefore different from secret-file permission and should be granted only when the owning policy requires it.

Mount configuration and secrets read-only. The process reads them during startup; this revision does not implement an in-process configuration reload. A changed value requires a controlled restart and the recovery procedure owned by that setting.

## Startup coupling summary

| Setting present | Required companion configuration |
|---|---|
| Non-loopback listener | Public HTTPS base URL and an authenticated deployment edge |
| Unsigned native envelopes | Native bearer authentication or loopback-only insecure development |
| Callback allowed hosts | Response and callback signing seed |
| A2A push encryption key | Response and callback signing seed |
| Delegation route | Response and callback signing seed plus valid peer identity and destination |
| PostgreSQL URL | Exactly one inline or file source |
| Any fleet option | Connector-registry URL |
| Fleet registry URL | Signing-seed file and explicit or registry-derived host allowlist |
| Fleet stream callback | Exact central callback URL; HTTP exception if plaintext |
| NATS username or password | NATS URL and the other credential half |
| Any introspection field | Complete four-field introspection group and protected resource |
| MCP introspection | Advertised matching issuer and no static MCP token |

Use Deploy AIP in production for the ordered deployment procedure, Connector host configuration for the separate host process, and Observe and recover for runtime diagnosis. This page owns option lookup; those guides own task flow and recovery decisions.