Hermes Agent connector configuration

This reference describes the configuration accepted by aip-host-hermes-agent at the reviewed source revision. A complete process combines two Hermes-specific file settings with the 45 shared connector-host settings.

SECTION	Connectors
TYPE	Connector
PROTOCOL	AIP 1.0
SOURCE	docs/connectors/hermes-agent/reference/configuration.md

Endpoint, operator-policy, identity, topology, limit, and provider-token values are read at startup. Only the optional shared credential-revision policy is re-read before provider side effects.

Configuration identity

Property	Value
Binary	`aip-host-hermes-agent`
AIP workspace version	`1.0.0`
Reviewed AIP source	`97be86e9efedf07ecf1783b03800f683f107fb04`
Connector ID	`hermes-agent`
Connector profile	`aip.connector.hermes_agent.v1`
Pinned Hermes source	`7426c09beee73bdff94d916015bac71384f6bc92`
Capability shape	`cap:hermes_agent:<endpoint_id>:<operation>`
Capability count	35 per configured endpoint

Every setting is available through the listed environment variable. The two Hermes-specific settings also expose the listed command-line flag. Shared settings use the equivalent kebab-case connector-host flag.

Environment variables and flags contain non-secret values or paths. Endpoint tokens, database credentials, and signing material belong in bounded files.

Hermes-specific settings

Environment variable	Command-line flag	Required	Default	Meaning
`AIP_HERMES_ENDPOINTS_FILE`	`--endpoints-file`	Yes	None	JSON array of endpoint descriptors
`AIP_HERMES_OPERATOR_POLICY_FILE`	`--operator-policy-file`	No	Built-in policy	JSON operator and first-class delegation policy

The endpoint file is required even when it contains only one endpoint. The operator-policy file is optional; omission uses the built-in governed operator policy with first-class delegation disabled.

Endpoint descriptor file

The endpoint file is a non-empty JSON array. This example contains no secret bytes:

```
JSON
[
  {
    "id": "primary",
    "base_url": "https://hermes-primary.example.internal:8642",
    "api_key_file": "/run/secrets/hermes-primary-api-key",
    "display_name": "Primary Hermes",
    "tenant_id": "tenant-acme"
  }
]
```

Descriptor fields

Field	Required	Default	Validation and effect
`id`	Yes	None	Trimmed, restricted to ASCII alphanumeric, `-`, or `_`, then lowercased; becomes the capability segment
`base_url`	Yes	None	URL using the `http` or `https` scheme; fixed source of Hermes request URLs
`api_key_file`	No	None	Owner-only file containing the Bearer token for protected operations
`display_name`	No	None	Retained in endpoint configuration; not projected into capability names, routing, or the manifest
`tenant_id`	No	Host tenant	Must exactly equal `AIP_CONNECTOR_HOST_TENANT_ID` when supplied

Endpoint objects reject unknown fields. IDs that collide after trimming and lowercasing make connector construction fail. Action input cannot replace an endpoint ID, URL, tenant, or token.

The URL parser enforces only a valid `http` or `https` URL. It does not prove destination ownership or impose the stricter credential, query, fragment, and origin rules used by shared host endpoints. Deployment policy should require a credential-free HTTPS origin outside a controlled local network.

Endpoint bounds and manifest effects

Limit	Reviewed value
Descriptor file	1 byte through 4 MiB
Descriptor count accepted by parser	1 through 1000
API-key file	1 byte through 16 KiB before UTF-8 and trimming checks
Generated capabilities	35 per endpoint
Common host capability ceiling	512
Effective endpoint ceiling	14, producing 490 capabilities

Fifteen endpoints produce 525 capabilities and fail common host preparation. The parser's 1000-item bound is therefore not a deployable catalog limit.

Changing endpoint IDs, base URLs, tenant values, or membership changes the discovered manifest and requires a corresponding immutable version and admission update. Changing only token bytes or `display_name` does not alter the current manifest, but a new process must load those startup values.

Provider authentication and client constants

Basic `health` is the only operation that does not require an endpoint token. The other 34 operations fail with `connector.hermes_agent.missing_credential` when no usable key is loaded. Protected requests use `Authorization: Bearer` .

The default Hermes client has these fixed properties:

Property	Fixed value
Redirect policy	No redirects
Connect timeout	10 seconds
Overall client timeout	900 seconds
Basic health timeout	30 seconds
Default direct-stream event limit	4096 events

Operation-specific request and stream paths can apply narrower bounds. These constants are not environment settings.

Operator-policy file

The optional operator-policy file is a non-secret JSON object of at most 512 KiB. It is parsed and validated once at startup.

Omitting the file constructs the complete built-in defaults. Supplying a file does not merge it with those defaults: every field marked required below must be present.

JSON field	Required in file	Built-in default	Validation or meaning
`delegation_enabled`	Yes	`false`	Enables first-class delegation routing; does not enable the direct operator capability
`delegation_requires_approval`	Yes	`true`	Requires runtime-verified durable approval for delegated work
`delegation_approval_exempt_capability_prefixes`	No	`[]`	Narrow capability prefixes that bypass only the connector-level delegation approval gate
`model`	No	`null`	Optional Hermes model route selected by deployment policy
`system_instructions`	Yes	Built-in governed operator instructions	Non-empty, at most 131072 bytes
`timeout_ms`	Yes	`900000`	1000 through 86400000
`poll_interval_ms`	Yes	`250`	10 through 60000
`max_events`	Yes	`4096`	1 through 100000
`max_input_bytes`	Yes	`1048576`	1024 through 16777216
`max_delegation_depth`	Yes	`16`	1 through 128
`start_claim_ttl_ms`	Yes	`60000`	1000 through 300000
`cancel_grace_ms`	Yes	`15000`	1000 through 300000
`allowed_capability_prefixes`	No	`[]`	Delegated child-capability prefixes
`allowed_scope_prefixes`	No	`[]`	Delegated scope prefixes

Every prefix must be non-empty after trimming and at most 512 bytes. When `delegation_enabled` is `true`, both allowed-prefix lists must be non-empty. The literal `*` explicitly opts into an unrestricted dimension; use a complete capability or narrow namespace whenever possible.

Approval-exempt prefixes waive only this connector's delegation approval gate. They do not bypass native AIP authorization, tenant, scope, capability, approval, or lifecycle controls.

Unlike endpoint descriptors, `HermesOperatorPolicy` does not deny unknown JSON fields at this revision. Serde ignores an unrecognized key. Treat such a key as an invalid deployment input during review; its presence is not evidence that a control is active.

Disabled-delegation example

This complete file preserves first-class delegation as disabled while making the operator limits explicit:

```
JSON
{
  "delegation_enabled": false,
  "delegation_requires_approval": true,
  "model": null,
  "system_instructions": "Operate only within the admitted AIP contract.",
  "timeout_ms": 900000,
  "poll_interval_ms": 250,
  "max_events": 4096,
  "max_input_bytes": 1048576,
  "max_delegation_depth": 16,
  "start_claim_ttl_ms": 60000,
  "cancel_grace_ms": 15000,
  "allowed_capability_prefixes": [],
  "allowed_scope_prefixes": []
}
```

The built-in system instructions contain additional invariants. Replacing them with this short example changes behavior; use a reviewed deployment policy, not the example text, for production.

Shared process and peer settings

Environment variable	Required	Default	Meaning and constraint
`AIP_CONNECTOR_HOST_BIND`	No	`0.0.0.0:8081`	Local listener behind deployment ingress
`AIP_CONNECTOR_HOST_PUBLIC_ENDPOINT`	Yes	None	Advertised URL ending exactly in `/aip/v1/messages`
`AIP_CONNECTOR_HOST_CONTROL_ENDPOINT`	Yes	None	Lifecycle URL using the exact connector-control path
`AIP_CONNECTOR_HOST_CONTROL_DID`	Alternative	None	Inline lifecycle control-plane `did:key`
`AIP_CONNECTOR_HOST_CONTROL_DID_FILE`	Alternative	None	Owner-only file containing the control-plane `did:key`
`AIP_CONNECTOR_HOST_GATEWAY_PRINCIPAL_ID`	Yes	None	Only central gateway principal accepted by the host
`AIP_CONNECTOR_HOST_GATEWAY_DID`	Alternative	None	Inline central gateway `did:key`
`AIP_CONNECTOR_HOST_GATEWAY_DID_FILE`	Alternative	None	Owner-only file containing the gateway `did:key`
`AIP_CONNECTOR_HOST_TLS_CA_FILE`	No	None	Optional PEM root for control-plane, callback, and event clients
`AIP_CONNECTOR_HOST_ALLOW_LOOPBACK_HTTP`	No	`false`	Allows public and control HTTP only with loopback bind and loopback URLs

For each peer DID, configure the inline value or file, not both. Both peers are required and must decode to usable verification keys.

Public and control endpoints require HTTPS, contain no embedded credentials, query, or fragment, and use their protocol paths. The HTTP exception is valid only when explicitly enabled for a loopback listener and loopback URLs.

The shared TLS CA is not passed into the default Hermes HTTP client by this binary. Configure Hermes provider trust through the process or container trust store, and review it independently from the control, callback, and event root.

Durable storage and replica identity

Environment variable	Required	Default	Meaning and constraint
`AIP_CONNECTOR_HOST_DATABASE_URL_FILE`	Yes	None	Owner-only PostgreSQL URL file for all host and Hermes operator state
`AIP_CONNECTOR_HOST_SIGNING_SEED_FILE`	Yes	None	Owner-only file containing a 32-byte Ed25519 seed as 64 hexadecimal characters
`AIP_CONNECTOR_HOST_TYPE_ID`	Yes	None	Admitted Hermes connector type ID
`AIP_CONNECTOR_HOST_VERSION_ID`	Yes	None	Immutable admitted connector version ID
`AIP_CONNECTOR_HOST_INSTANCE_ID`	Yes	None	Logical tenant-owned Hermes instance
`AIP_CONNECTOR_HOST_REPLICA_ID`	Yes	None	Pre-provisioned concrete process identity
`AIP_CONNECTOR_HOST_ARTIFACT_DIGEST`	Yes	None	Exact immutable OCI SHA-256 digest for the running host
`AIP_CONNECTOR_HOST_CONFIG_REVISION`	No	`1`	Positive monotonic non-secret configuration revision

Hermes operator run bindings, approvals, native actions, streams, and other runtime projections share this PostgreSQL boundary. Do not move one replica to a different database while claiming the same instance and durable state.

Tenant and external-account settings

Environment variable	Required	Default	Meaning and constraint
`AIP_CONNECTOR_HOST_TENANT_ID`	Yes	None	Verified tenant owned by this instance; 1 through 512 bytes
`AIP_CONNECTOR_HOST_TENANT_SYSTEM`	No	None	Optional source system for the tenant reference
`AIP_CONNECTOR_HOST_MEMBERSHIP_ID`	Yes	None	Stable verified membership assertion; 1 through 512 bytes
`AIP_CONNECTOR_HOST_EXTERNAL_ACCOUNT_ID`	Pair	None	Stable Hermes deployment or endpoint-set reference; 1 through 512 bytes
`AIP_CONNECTOR_HOST_EXTERNAL_ACCOUNT_SYSTEM`	Pair	None	Product name for the account reference; 1 through 128 bytes

External-account ID and system must be configured together or omitted together. Every endpoint tenant must equal the shared host tenant. A normal Hermes host uses `hermes_agent` as its external-account system by deployment convention; the common parser validates the pair but does not require that literal.

Trust, topology, and secret-provider identity

Environment variable	Required	Default	Meaning and constraint
`AIP_CONNECTOR_HOST_TRUST_DOMAIN`	Yes	None	Shared host, gateway, and lifecycle trust domain; at most 253 bytes
`AIP_CONNECTOR_HOST_REGION`	No	`global`	Topology-aware routing region
`AIP_CONNECTOR_HOST_ZONE`	No	`default`	Independent failure zone
`AIP_CONNECTOR_HOST_CAPACITY_CLASS`	No	`standard`	Operator-defined capacity class
`AIP_CONNECTOR_HOST_SECRET_PROVIDER_REF`	Yes	None	Non-secret secret-provider reference; at most 2048 bytes

Topology values describe the replica to the registry. They do not move a process, configure the network, or raise local concurrency.

The secret-provider reference is not a Hermes token. The binary reads actual endpoint tokens only from descriptor-selected files.

Credential identity and revision policy

Environment variable	Required	Default	Meaning and constraint
`AIP_CONNECTOR_HOST_CREDENTIAL_ID`	Group	None	Opaque non-secret credential handle ID
`AIP_CONNECTOR_HOST_CREDENTIAL_ISSUER`	Group	None	Trusted handle issuer
`AIP_CONNECTOR_HOST_CREDENTIAL_SCOPES`	Group	Empty	Comma-delimited or repeatable non-empty verified scopes
`AIP_CONNECTOR_HOST_CREDENTIAL_REVISION`	Group	None	Opaque assignment-pinned revision, 1 through 256 bytes
`AIP_CONNECTOR_HOST_CREDENTIAL_POLICY_FILE`	No	Static current revision	Reloadable non-secret JSON authorization policy

ID, issuer, at least one scope, and revision must be configured together or all omitted. A configured credential handle is tenant-bound. Its revision must be authorized at startup and immediately before provider side effects.

The policy object has `current_revision_ref`, `accepted_previous_revisions`, and `revoked_revisions`. Entries cannot overlap, and each opaque revision is 1 through 256 bytes. A read, parse, validation, or authorization failure denies provider work.

This revision policy does not load, choose, or reload an endpoint token.

Callback and event-delivery settings

Environment variable	Required	Default	Meaning and constraint
`AIP_CONNECTOR_HOST_CALLBACK_ENDPOINT`	No	None	Fixed central result and stream callback endpoint
`AIP_CONNECTOR_HOST_CALLBACK_ALLOW_HTTP`	No	`false`	Controlled-network HTTP exception; requires a callback endpoint
`AIP_CONNECTOR_HOST_CALLBACK_ALLOW_PRIVATE`	No	`false`	Private-address exception; requires a callback endpoint
`AIP_CONNECTOR_HOST_EVENT_ENDPOINT`	No	None	Fixed central connector-event ingress endpoint
`AIP_CONNECTOR_HOST_EVENT_ALLOW_HTTP`	No	`false`	Controlled-network HTTP exception; requires an event endpoint
`AIP_CONNECTOR_HOST_EVENT_ALLOW_PRIVATE`	No	`false`	Private-address exception; requires an event endpoint

Both dispatchers use the host signing identity and optional private TLS root. Setting an exception without its corresponding endpoint fails startup. These settings do not configure Hermes MCP or model-provider callbacks.

Lifecycle and resource limits

Environment variable	Default	Valid range or relation	Effect
`AIP_CONNECTOR_HOST_MAX_IN_FLIGHT`	`32`	Greater than zero	Concurrent action permits for one replica
`AIP_CONNECTOR_HOST_LEASE_TTL_MS`	`30000`	1000 through 300000	Registry lease lifetime
`AIP_CONNECTOR_HOST_HEARTBEAT_MS`	`10000`	Greater than zero and at most half the lease TTL	Lease renewal interval
`AIP_CONNECTOR_HOST_DRAIN_TIMEOUT_MS`	`30000`	1 through 300000	Graceful shutdown deadline
`AIP_CONNECTOR_HOST_HEALTH_TIMEOUT_MS`	`2000`	1 through 30000	Product and storage readiness-probe deadline
`AIP_CONNECTOR_HOST_CONTROL_TIMEOUT_MS`	`5000`	1 through 30000	Lifecycle control request timeout

The common host also fixes a 4 MiB canonical manifest limit, 512 capability definitions, and a 4 MiB native request-body limit. This binary exposes no environment setting for those three bounds.

File requirements

File	Maximum	Content and permission rule
Endpoint descriptors	4 MiB	Non-empty JSON; regular non-symlink; no Unix group or world write bits
Endpoint API key	16 KiB	Non-empty UTF-8 after trimming; regular non-symlink; no Unix group or world permission bits
Operator policy	512 KiB	Valid JSON object; regular non-symlink; no Unix group or world write bits
PostgreSQL URL	16 KiB	Same owner-only secret-file checks as endpoint keys
Host signing seed	256 bytes read bound	Exactly 64 hexadecimal characters encoding 32 bytes; owner-only secret-file checks
Gateway or control-plane DID	512 bytes	Non-empty DID text; owner-only secret-file checks
TLS CA	1 MiB	Regular non-symlink; no Unix group or world write bits
Credential revision policy	64 KiB	Valid bounded JSON; regular non-symlink; no Unix group or world write bits

The reader validates metadata before and after opening a file. On Unix it also rejects an inode or device change during the read.

Configuration supplements

These snippets contain only Hermes additions. They are not complete host or production configurations and contain no token bytes.

One endpoint with built-in operator policy

```
TEXT
AIP_HERMES_ENDPOINTS_FILE=/etc/aip/hermes-endpoints.json
```

Explicit operator policy

```
TEXT
AIP_HERMES_ENDPOINTS_FILE=/etc/aip/hermes-endpoints.json
AIP_HERMES_OPERATOR_POLICY_FILE=/etc/aip/hermes-operator-policy.json
```

Tenant and external-account alignment

```
TEXT
AIP_CONNECTOR_HOST_TENANT_ID=tenant-acme
AIP_CONNECTOR_HOST_EXTERNAL_ACCOUNT_ID=primary-hermes
AIP_CONNECTOR_HOST_EXTERNAL_ACCOUNT_SYSTEM=hermes_agent
```

The endpoint file should either omit `tenant_id` or use the exact `tenant-acme` value in this supplement.

Startup and reload behavior

Startup proceeds in this order:

1. Parse Hermes-specific and common host arguments.
2. Read and validate the endpoint descriptor array.
3. Normalize endpoints, bind every endpoint to the host tenant, and read API keys.
4. Build the 35-operation-per-endpoint connector and default HTTP client.
5. Read and validate the optional operator policy.
6. Validate host identity, trust, endpoints, files, limits, and credential policy.
7. Discover and validate the source-derived manifest without provider I/O.
8. Open PostgreSQL runtime stores and attach operator, approval, action, and lifecycle state.
9. Register, acquire a lease, become ready, and serve the common host surface.

Setting class	Reload behavior at the reviewed source
Credential-revision policy JSON	Re-read before provider side effects
Endpoint API keys	Read once at startup
Endpoint descriptor fields	Parsed once at startup
Hermes operator policy	Parsed once at startup
Identity, topology, delivery endpoints, TLS CA, and limits	Startup-only

Readiness includes durable storage and connector-local health across the configured endpoint set. A syntactically valid configuration can remain unready because admission, registration, storage, TLS, DNS, or provider health fails.

Configuration failures

Failure	Corrective decision
Endpoint array is empty, oversized, or malformed	Correct the bounded owner-controlled JSON file
Descriptor has an unknown field	Remove the unsupported field; do not bypass strict parsing
IDs collide after normalization	Assign distinct stable endpoint IDs and update admission
Endpoint tenant differs	Place it in the matching tenant-owned instance
Endpoint key file is unsafe	Correct the mount and permissions without printing the key
Protected operation lacks a key	Add the correct endpoint file and replace the process
Operator policy omits a required field	Supply a complete object or omit the file to use built-in defaults
Operator policy contains an unknown field	Treat it as invalid even though this revision ignores the key
Delegation lacks allowlists	Set both explicit lists or keep first-class delegation disabled
Public or control endpoint is invalid	Restore its exact HTTPS path; use HTTP only for explicit loopback development
Peer DID alternatives conflict	Keep the inline value or file, not both
External-account or credential group is partial	Configure the complete group or omit it completely
Credential policy denies the route revision	Correct the reviewed policy or roll forward; do not bypass authorization
Heartbeat relation is invalid	Keep a nonzero heartbeat at or below half the lease TTL
Host remains unready	Separate config, admission, storage, lease, TLS, endpoint, and provider evidence

Related documentation

- Authenticate and bind Hermes endpoints
- Hermes Agent connector
- Shared connector-host configuration
- Rotate connector credentials
- Security model