CANONICAL LIVING OVERVIEW GUIDE

Kubernetes, Controllers, Operators, and Cluster Engineering

A complete Kubernetes engineering guide covering API machinery, desired state, controllers, CRDs, operators, scheduling, nodes, networking, storage, security, upgrades, multi-cluster operations, batch and AI workloads, and cluster assurance.

PERMANENT ID
MYTHOS-FG-CLD-0002

PUBLICATION CLASS
FIELD GUIDE

VERSION / STATUS
1.0.0-candidate / CANDIDATE

PERMANENT PUBLICATION IDENTITY

Kubernetes, Controllers, Operators, and Cluster Engineering

Document ID
MYTHOS-FG-CLD-0002

Version
1.0.0-candidate

Status
Candidate Focused Field Guide

Review date
2026-10-26

Publication position
This is a living field guide. It is designed for frequent revision while retaining a stable permanent identity. Candidate status means the content is ready for structured use and review but is not represented as an external certification or authority publication.

LEVEL	CLASS	CADENCE	EVIDENCE
L1-L4	FIELD GUIDE	QUARTERLY	SOURCE-BOUND
Foundational through frontier	Practical and educational	Interim updates as needed	Limitations remain visible

Reader orientation

Dimension	Engineering position
Primary domain	MYTHOS-CLD - Cloud, Platform & Distributed Systems
Secondary domain	MYTHOS-SWE; MYTHOS-SEC; MYTHOS-DAT
Audience	Kubernetes, platform, cloud, SRE, software, security, and infrastructure engineers.
Prerequisites	Containers, Linux, networking, distributed systems, APIs, storage, identity, and Go fundamentals.
Publisher / owner	Mythos Systems / Aaron Stovall
Public classification	Public technical guidance

Expected outcomes

Explain Kubernetes from API request and etcd state through reconciliation, scheduling, nodes, and workload outcome.

Build controllers and operators with correct ownership, idempotency, concurrency, status, finalizers, and recovery.

Engineer clusters for networking, storage, security, observability, upgrades, and disaster recovery.

Schedule services, batch, GPU, and AI workloads with quotas, topology, fairness, and failure isolation.

Validate desired state against applied state, workload behavior, SLOs, and recoverability.

How to use this guide

Read the guide as a progressive system rather than a disconnected encyclopedia. Foundational material establishes the mental model; practitioner sections convert it into repeatable implementation and operations; advanced sections expose architecture, scale, security, and failure behavior; frontier sections describe technologies whose evidence or maturity is still evolving.

Suggested reading path
New practitioners should follow the chapters in order and complete the laboratories. Experienced engineers may begin with the architecture, failure, validation, or frontier sections, then use the related-document map to deepen a specific topic.

Learning and assurance model

Level	Reader outcome
L1 - Foundational	Terminology, first principles, component roles, packet or state flow, and simple working examples.
L2 - Practitioner	Implementation, configuration, automation, testing, troubleshooting, and operational evidence.
L3 - Advanced	Architecture, scale, concurrency, resilience, threat models, performance, and complex migrations.
L4 - Frontier	Emerging systems, research directions, technical uncertainty, readiness gates, and adoption signals.

Contents

CHAPTER 01

Kubernetes architecture and API machinery

Trace desired state through the control plane and resource store.

Chapter operating position
Trace desired state through the control plane and resource store.

1.1 API server and resource model

API server and resource model must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Understand kinds, resources, namespaces, metadata, spec, status, subresources, validation, and discovery. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for api server and resource model.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating api server and resource model as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for api server and resource model.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

1.2 etcd and authoritative state

etcd and authoritative state must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Use consistency, revisions, transactions, compaction, snapshots, quorum, and restore. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for etcd and authoritative state.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Implementation sketch

```
apiVersion: platform.mythos.systems/v1
kind: AgentMission
metadata:
  name: mission-042
spec:
  workers: 16
  modelClasses: [local-fast, frontier-reasoner, verifier]
  budget:
    maxTokens: 2000000
    maxGpuHours: 8
  completion:
    requireIndependentVerification: true
status:
  observedGeneration: 4
  conditions:
    - type: Ready
      status: "False"
      reason: VerificationPending
```

Failure patterns

- Treating etcd and authoritative state as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for etcd and authoritative state.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

1.3 Watch, cache, and reconciliation

Watch, cache, and reconciliation must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Handle resource versions, informers, eventual caches, missed events, relist, and idempotency. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for watch, cache, and reconciliation.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating watch, cache, and reconciliation as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for watch, cache, and reconciliation.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

1.4 Admission and authorization

Admission and authorization must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Apply authentication, authorization, validation, mutation, policy, and audit before persistence. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

Define authority, ownership, desired state, dependencies, limits, and evidence for admission and authorization.

Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.

Separate control plane, execution plane, data plane, evidence plane, and human decision authority.

Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.

Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.

Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

Treating admission and authorization as a feature rather than an end-to-end stateful system.

Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.

Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for admission and authorization.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

Chapter 01 laboratory and review

Practical laboratory

Trace a Deployment create request through admission, storage, controller reconciliation, scheduling, kubelet, and status.

Laboratory evidence package

- Architecture or topology showing boundaries, identities, dependencies, and authoritative inputs.
- Versioned configuration or code with explanatory comments and reproducible setup instructions.
- Nominal-path validation plus at least two injected failure or boundary conditions.
- Structured logs, traces, metrics, packet captures, test output, or other appropriate evidence.
- A concise decision record covering findings, limitations, residual risks, and next actions.

Frontier extension

Explore strongly typed control-plane APIs generated from domain specifications.

Review gate

Advance only when another engineer can reproduce the result, explain the architecture, identify the failure boundary, and distinguish measured evidence from assumptions or projections.

CHAPTER 02

Workloads, scheduling, and resource management

Place and operate pods according to resources, topology, and policy.

Chapter operating position
Place and operate pods according to resources, topology, and policy.

2.1 Pod and workload controllers

Pod and workload controllers must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Use Deployments, StatefulSets, DaemonSets, Jobs, CronJobs, disruption, rollout, and ownership. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for pod and workload controllers.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating pod and workload controllers as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for pod and workload controllers.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

2.2 Scheduler pipeline

Scheduler pipeline must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Understand queue, filtering, scoring, binding, plugins, preemption, and profiles. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for scheduler pipeline.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Implementation sketch

```
apiVersion: platform.mythos.systems/v1
kind: AgentMission
metadata:
  name: mission-042
spec:
  workers: 16
  modelClasses: [local-fast, frontier-reasoner, verifier]
  budget:
    maxTokens: 2000000
    maxGpuHours: 8
  completion:
    requireIndependentVerification: true
status:
  observedGeneration: 4
  conditions:
    - type: Ready
      status: "False"
      reason: VerificationPending
```

Failure patterns

- Treating scheduler pipeline as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for scheduler pipeline.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

2.3 Resources and QoS

Resources and QoS must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Set requests, limits, CPU, memory, huge pages, ephemeral storage, GPUs, QoS, and eviction. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for resources and qos.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating resources and qos as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for resources and qos.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

2.4 Topology and placement

Topology and placement must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Use affinity, anti-affinity, topology spread, taints, tolerations, zones, nodes, and device locality. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

Define authority, ownership, desired state, dependencies, limits, and evidence for topology and placement.

Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.

Separate control plane, execution plane, data plane, evidence plane, and human decision authority.

Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.

Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.

Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

Treating topology and placement as a feature rather than an end-to-end stateful system.

Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.

Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for topology and placement.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

Chapter 02 laboratory and review

Practical laboratory
Schedule mixed web, stateful, batch, and GPU workloads and test pressure, preemption, and zone failure.

Laboratory evidence package

- Architecture or topology showing boundaries, identities, dependencies, and authoritative inputs.
- Versioned configuration or code with explanatory comments and reproducible setup instructions.
- Nominal-path validation plus at least two injected failure or boundary conditions.
- Structured logs, traces, metrics, packet captures, test output, or other appropriate evidence.
- A concise decision record covering findings, limitations, residual risks, and next actions.

Frontier extension

Explore topology-aware AI workload placement and dynamic resource allocation.

Review gate
Advance only when another engineer can reproduce the result, explain the architecture, identify the failure boundary, and distinguish measured evidence from assumptions or projections.

CHAPTER 03

Controllers and reconciliation engineering

Build reliable feedback loops around desired and observed state.

Chapter operating position
Build reliable feedback loops around desired and observed state.

3.1 Reconcile contract

Reconcile contract must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Read current state, compute intent, apply minimal change, update status, and return retry or watch dependencies. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for reconcile contract.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating reconcile contract as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for reconcile contract.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

3.2 Idempotency and concurrency

Idempotency and concurrency must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Use owner references, resource versions, patch, compare-and-swap, deduplication, and deterministic naming. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for idempotency and concurrency.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Implementation sketch

```
apiVersion: platform.mythos.systems/v1
kind: AgentMission
metadata:
  name: mission-042
spec:
  workers: 16
  modelClasses: [local-fast, frontier-reasoner, verifier]
  budget:
    maxTokens: 2000000
    maxGpuHours: 8
  completion:
    requireIndependentVerification: true
status:
  observedGeneration: 4
  conditions:
    - type: Ready
      status: "False"
      reason: VerificationPending
```

Failure patterns

- Treating idempotency and concurrency as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for idempotency and concurrency.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

3.3 Finalizers and deletion

Finalizers and deletion must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Perform external cleanup, handle retries, timeouts, stuck resources, force removal, and evidence. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for finalizers and deletion.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating finalizers and deletion as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for finalizers and deletion.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

3.4 Conditions, status, and events

Conditions, status, and events must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Expose observed generation, readiness, progress, degradation, reason, message, and actionable diagnostics. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

Define authority, ownership, desired state, dependencies, limits, and evidence for conditions, status, and events.

Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.

Separate control plane, execution plane, data plane, evidence plane, and human decision authority.

Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.

Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.

Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

Treating conditions, status, and events as a feature rather than an end-to-end stateful system.

Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.

Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for conditions, status, and events.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

Chapter 03 laboratory and review

Practical laboratory
Implement a small controller simulator and test duplicate events, conflicts, partial external effects, deletion, and recovery.

Laboratory evidence package

- Architecture or topology showing boundaries, identities, dependencies, and authoritative inputs.
- Versioned configuration or code with explanatory comments and reproducible setup instructions.
- Nominal-path validation plus at least two injected failure or boundary conditions.
- Structured logs, traces, metrics, packet captures, test output, or other appropriate evidence.
- A concise decision record covering findings, limitations, residual risks, and next actions.

Frontier extension

Explore formally verified reconciliation and controller synthesis.

Review gate
Advance only when another engineer can reproduce the result, explain the architecture, identify the failure boundary, and distinguish measured evidence from assumptions or projections.

CHAPTER 04

CRDs, operators, and platform APIs

Turn operational knowledge into safe declarative APIs.

Chapter operating position
Turn operational knowledge into safe declarative APIs.

4.1 API and schema design

API and schema design must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Define versioned spec, status, defaults, validation, immutability, list semantics, and compatibility. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for api and schema design.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating api and schema design as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for api and schema design.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

4.2 Operator capability levels

Operator capability levels must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Encode installation, configuration, upgrades, backup, failure recovery, and application-specific automation. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for operator capability levels.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Implementation sketch

```
apiVersion: platform.mythos.systems/v1
kind: AgentMission
metadata:
  name: mission-042
spec:
  workers: 16
  modelClasses: [local-fast, frontier-reasoner, verifier]
  budget:
    maxTokens: 2000000
    maxGpuHours: 8
  completion:
    requireIndependentVerification: true
status:
  observedGeneration: 4
  conditions:
    - type: Ready
      status: "False"
      reason: VerificationPending
```

Failure patterns

- Treating operator capability levels as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for operator capability levels.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

4.3 Version conversion and migration

Version conversion and migration must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Use storage versions, conversion, deprecation, data migration, rollback, and client compatibility. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for version conversion and migration.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating version conversion and migration as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for version conversion and migration.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

4.4 Packaging and lifecycle

Packaging and lifecycle must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Distribute controllers, CRDs, RBAC, webhooks, images, charts, operators, and release evidence. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

Define authority, ownership, desired state, dependencies, limits, and evidence for packaging and lifecycle.

Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.

Separate control plane, execution plane, data plane, evidence plane, and human decision authority.

Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.

Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.

Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

Treating packaging and lifecycle as a feature rather than an end-to-end stateful system.

Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.

Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for packaging and lifecycle.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

Chapter 04 laboratory and review

Practical laboratory
Design an application CRD and operator that performs safe upgrades and backup with version migration.

Laboratory evidence package

- Architecture or topology showing boundaries, identities, dependencies, and authoritative inputs.
- Versioned configuration or code with explanatory comments and reproducible setup instructions.
- Nominal-path validation plus at least two injected failure or boundary conditions.
- Structured logs, traces, metrics, packet captures, test output, or other appropriate evidence.
- A concise decision record covering findings, limitations, residual risks, and next actions.

Frontier extension

Explore Crossplane-style compositions that combine applications, infrastructure, policy, and operations.

Review gate
Advance only when another engineer can reproduce the result, explain the architecture, identify the failure boundary, and distinguish measured evidence from assumptions or projections.

CHAPTER 05

Networking, service, ingress, and policy

Engineer communication from pod interfaces through service and gateway paths.

Chapter operating position
Engineer communication from pod interfaces through service and gateway paths.

5.1 CNI and pod networking

CNI and pod networking must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Understand IP allocation, routes, overlays, underlays, MTU, policy, eBPF, and network failure. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for cni and pod networking.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating cni and pod networking as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for cni and pod networking.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

5.2 Services and discovery

Services and discovery must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Use ClusterIP, headless, load balancing, endpoint slices, DNS, session behavior, and topology. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for services and discovery.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Implementation sketch

```
apiVersion: platform.mythos.systems/v1
kind: AgentMission
metadata:
  name: mission-042
spec:
  workers: 16
  modelClasses: [local-fast, frontier-reasoner, verifier]
  budget:
    maxTokens: 2000000
    maxGpuHours: 8
  completion:
    requireIndependentVerification: true
status:
  observedGeneration: 4
  conditions:
    - type: Ready
      status: "False"
      reason: VerificationPending
```

Failure patterns

- Treating services and discovery as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for services and discovery.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

5.3 Gateway API and ingress

Gateway API and ingress must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Separate infrastructure, application routing, policies, certificates, and implementation capabilities. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for gateway api and ingress.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating gateway api and ingress as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for gateway api and ingress.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

5.4 Network policy and service identity

Network policy and service identity must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Control ingress, egress, DNS, host paths, service mesh, mTLS, and workload authorization. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

Define authority, ownership, desired state, dependencies, limits, and evidence for network policy and service identity.

Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.

Separate control plane, execution plane, data plane, evidence plane, and human decision authority.

Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.

Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.

Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

Treating network policy and service identity as a feature rather than an end-to-end stateful system.

Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.

Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for network policy and service identity.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

Chapter 05 laboratory and review

Practical laboratory
Build a service path using Gateway API and default-deny network policy, then troubleshoot DNS, MTU, and return-path faults.

Laboratory evidence package

- Architecture or topology showing boundaries, identities, dependencies, and authoritative inputs.
- Versioned configuration or code with explanatory comments and reproducible setup instructions.
- Nominal-path validation plus at least two injected failure or boundary conditions.
- Structured logs, traces, metrics, packet captures, test output, or other appropriate evidence.
- A concise decision record covering findings, limitations, residual risks, and next actions.

Frontier extension

Explore unified north-south and east-west policy through Gateway API GAMMA and ambient data planes.

Review gate
Advance only when another engineer can reproduce the result, explain the architecture, identify the failure boundary, and distinguish measured evidence from assumptions or projections.

CHAPTER 06

Storage, stateful systems, and data protection

Provide durable state through dynamic infrastructure and failure.

Chapter operating position
Provide durable state through dynamic infrastructure and failure.

6.1 Volumes and CSI

Volumes and CSI must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Use classes, claims, binding, topology, snapshots, expansion, attachment, access modes, and reclaim. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for volumes and csi.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating volumes and csi as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for volumes and csi.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

6.2 Stateful applications

Stateful applications must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Design identity, ordering, replication, quorum, anti-affinity, backups, upgrades, and failover. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for stateful applications.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Implementation sketch

```
apiVersion: platform.mythos.systems/v1
kind: AgentMission
metadata:
  name: mission-042
spec:
  workers: 16
  modelClasses: [local-fast, frontier-reasoner, verifier]
  budget:
    maxTokens: 2000000
    maxGpuHours: 8
  completion:
    requireIndependentVerification: true
status:
  observedGeneration: 4
  conditions:
    - type: Ready
      status: "False"
      reason: VerificationPending
```

Failure patterns

- Treating stateful applications as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for stateful applications.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

6.3 Data protection

Data protection must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Coordinate application consistency, snapshots, backups, object copies, encryption, restore, and validation. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for data protection.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating data protection as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for data protection.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

6.4 Storage incidents and recovery

Storage incidents and recovery must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Handle lost nodes, attach conflicts, corrupt volumes, full disks, zone failure, and control-plane restore. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

Define authority, ownership, desired state, dependencies, limits, and evidence for storage incidents and recovery.

Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.

Separate control plane, execution plane, data plane, evidence plane, and human decision authority.

Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.

Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.

Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

Treating storage incidents and recovery as a feature rather than an end-to-end stateful system.

Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.

Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for storage incidents and recovery.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

Chapter 06 laboratory and review

Practical laboratory
Deploy a stateful service, take an application-consistent backup, fail a node and zone, and restore to a new cluster.

Laboratory evidence package

- Architecture or topology showing boundaries, identities, dependencies, and authoritative inputs.
- Versioned configuration or code with explanatory comments and reproducible setup instructions.
- Nominal-path validation plus at least two injected failure or boundary conditions.
- Structured logs, traces, metrics, packet captures, test output, or other appropriate evidence.
- A concise decision record covering findings, limitations, residual risks, and next actions.

Frontier extension

Explore disaggregated storage and verifiable state migration.

Review gate
Advance only when another engineer can reproduce the result, explain the architecture, identify the failure boundary, and distinguish measured evidence from assumptions or projections.

CHAPTER 07

Cluster security, observability, and operations

Protect and operate the cluster as critical infrastructure.

Chapter operating position
Protect and operate the cluster as critical infrastructure.

7.1 Identity, RBAC, and workload security

Identity, RBAC, and workload security must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Use least privilege, service accounts, pod security, secrets, admission, image verification, and node controls. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for identity, rbac, and workload security.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating identity, rbac, and workload security as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for identity, rbac, and workload security.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

7.2 Telemetry and diagnostics

Telemetry and diagnostics must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Observe API, etcd, scheduler, controllers, nodes, runtimes, network, storage, workloads, and user outcomes. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for telemetry and diagnostics.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Implementation sketch

```
apiVersion: platform.mythos.systems/v1
kind: AgentMission
metadata:
  name: mission-042
spec:
  workers: 16
  modelClasses: [local-fast, frontier-reasoner, verifier]
  budget:
    maxTokens: 2000000
    maxGpuHours: 8
  completion:
    requireIndependentVerification: true
status:
  observedGeneration: 4
  conditions:
    - type: Ready
      status: "False"
      reason: VerificationPending
```

Failure patterns

- Treating telemetry and diagnostics as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for telemetry and diagnostics.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

7.3 Upgrades and version skew

Upgrades and version skew must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Plan control-plane, node, add-on, CRD, API, runtime, network, storage, and workload compatibility. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for upgrades and version skew.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating upgrades and version skew as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for upgrades and version skew.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

7.4 Backup and disaster recovery

Backup and disaster recovery must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Protect etcd, manifests, certificates, external resources, images, data, policy, and runbooks. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

Define authority, ownership, desired state, dependencies, limits, and evidence for backup and disaster recovery.

Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.

Separate control plane, execution plane, data plane, evidence plane, and human decision authority.

Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.

Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.

Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

Treating backup and disaster recovery as a feature rather than an end-to-end stateful system.

Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.

Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for backup and disaster recovery.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

Chapter 07 laboratory and review

Practical laboratory
Execute a cluster upgrade simulation and recover from an incompatible CRD, node failure, and etcd restore.

Laboratory evidence package

- Architecture or topology showing boundaries, identities, dependencies, and authoritative inputs.
- Versioned configuration or code with explanatory comments and reproducible setup instructions.
- Nominal-path validation plus at least two injected failure or boundary conditions.
- Structured logs, traces, metrics, packet captures, test output, or other appropriate evidence.
- A concise decision record covering findings, limitations, residual risks, and next actions.

Frontier extension

Explore immutable clusters and continuously verified upgrade twins.

Review gate
Advance only when another engineer can reproduce the result, explain the architecture, identify the failure boundary, and distinguish measured evidence from assumptions or projections.

CHAPTER 08

Multi-cluster, fleet, batch, and AI engineering

Scale beyond one cluster and one workload class.

Chapter operating position
Scale beyond one cluster and one workload class.

8.1 Fleet architecture

Fleet architecture must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Manage clusters, regions, versions, policy, identity, networking, configuration, and lifecycle. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for fleet architecture.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating fleet architecture as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for fleet architecture.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

8.2 Multi-cluster service and data

Multi-cluster service and data must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Handle discovery, routing, failover, replication, consistency, sovereignty, and observability. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for multi-cluster service and data.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Implementation sketch

```
apiVersion: platform.mythos.systems/v1
kind: AgentMission
metadata:
  name: mission-042
spec:
  workers: 16
  modelClasses: [local-fast, frontier-reasoner, verifier]
  budget:
    maxTokens: 2000000
    maxGpuHours: 8
  completion:
    requireIndependentVerification: true
status:
  observedGeneration: 4
  conditions:
    - type: Ready
      status: "False"
      reason: VerificationPending
```

Failure patterns

- Treating multi-cluster service and data as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.
- Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.
- Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.
- Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for multi-cluster service and data.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

8.3 Batch and queueing

Batch and queueing must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Use Jobs, priorities, Kueue admission, quotas, cohorts, preemption, and fair sharing. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

- Define authority, ownership, desired state, dependencies, limits, and evidence for batch and queueing.
- Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.
- Separate control plane, execution plane, data plane, evidence plane, and human decision authority.
- Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.
- Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.
- Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

- Treating batch and queueing as a feature rather than an end-to-end stateful system.
- Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.
- Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for batch and queueing.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

8.4 AI and accelerator workloads

AI and accelerator workloads must be engineered as a governed distributed-system mechanism, not a product label, agent persona, YAML object, or diagram. Schedule distributed training, inference, GPUs, topology, checkpoints, model data, and failure recovery. The design should identify authoritative state, workload identities, typed contracts, resource and token budgets, trust boundaries, failure domains, telemetry, rollback, and acceptance evidence.

This guide traces the subject through the API request, admission, persistence, watch, reconciliation, scheduling, node execution, service, status, and recovery chain. A running agent, reconciled object, successful API call, healthy proxy, or completed function is not proof of the intended outcome. Desired state must be compared with applied state, external effects, negative tests, degraded behavior, causal evidence, cost, and user or mission results.

Production engineering begins with a minimal known-good control plane and explicit invariants. Add one worker class, model route, controller, policy, event source, gateway, or platform feature at a time; preserve before-state evidence; test authorized and prohibited behavior; inject realistic failures; and retain a tested compensation or rollback path.

Troubleshooting locates the first divergence from the expected state machine. Account for stale watches, leases, retries, eventual consistency, duplicated work, cache state, identity rotation, partial external effects, version skew, prompt injection, overload, network partitions, and missing telemetry. Closure requires causal explanation, reproducibility, validated recovery, and residual limitations.

Engineering practices

Define authority, ownership, desired state, dependencies, limits, and evidence for ai and accelerator workloads.

Use typed APIs, stable identities, explicit state machines, idempotency, and bounded effects.

Separate control plane, execution plane, data plane, evidence plane, and human decision authority.

Test nominal, prohibited, overloaded, partitioned, partially failed, recovery, upgrade, and rollback cases.

Preserve configuration, policy, model, source, trace, event, artifact, approval, and user-outcome evidence.

Scale through quotas, queues, backpressure, fairness, isolation, canaries, and deterministic fan-in.

Failure patterns

Treating ai and accelerator workloads as a feature rather than an end-to-end stateful system.

Scaling worker count or deployment scope before bounding queues, budgets, retries, and fan-in.

Trusting controller, agent, or proxy status without validating external state and user outcome.

Sharing credentials, workspaces, caches, or context across tenants and roles without explicit policy.

Allowing AI-generated plans, policies, or changes to bypass deterministic validation and human authority.

Declaring recovery after process restart without reconciling orphan effects, leases, artifacts, and evidence.

Validation and evidence

Method	Expected evidence
Examine	Requirements, APIs, identities, state machines, policies, versions, resources, and dependencies for ai and accelerator workloads.
Observe	Queue, task, controller, model, tool, route, proxy, event, resource, trace, artifact, and user-visible state.
Test	Expected, prohibited, malformed, duplicate, overloaded, partitioned, failed, recovery, upgrade, and rollback conditions.
Measure	Task success, reliability, latency, throughput, fairness, cost, token and compute use, blast radius, and operator burden.
Reconcile	Desired and planned state against actual cloud, cluster, agent, network, data, external-effect, and evidence state.

Chapter 08 laboratory and review

Practical laboratory
Design a three-cluster fleet with service failover, policy, GPU cohorts, and sovereign workload placement.

Laboratory evidence package

- Architecture or topology showing boundaries, identities, dependencies, and authoritative inputs.
- Versioned configuration or code with explanatory comments and reproducible setup instructions.
- Nominal-path validation plus at least two injected failure or boundary conditions.
- Structured logs, traces, metrics, packet captures, test output, or other appropriate evidence.
- A concise decision record covering findings, limitations, residual risks, and next actions.

Frontier extension

Explore global workload schedulers combining Kubernetes, Kueue, agent fabrics, and heterogeneous accelerators.

Review gate
Advance only when another engineer can reproduce the result, explain the architecture, identify the failure boundary, and distinguish measured evidence from assumptions or projections.

Integrated learning roadmap

The guide is intended to be revisited. First establish fluency, then build a small implementation, operate it under failure, automate repeatable work, and finally evaluate emerging techniques against a stable baseline. Progress is demonstrated through evidence rather than reading completion alone.

Stage	Demonstrated capability
Foundation	Explain the system from first principles and trace its critical path without relying on product terminology.
Build	Implement the smallest representative system with explicit contracts, identities, telemetry, and tests.
Operate	Diagnose injected failures, recover safely, and preserve evidence of the causal chain.
Automate	Convert a proven manual workflow into bounded, idempotent, observable automation.
Architect	Design for scale, resilience, security, interoperability, and lifecycle change.
Explore	Evaluate frontier techniques using stated hypotheses, limitations, and adoption gates.

Source authority register

Source policy
Official standards, specifications, and project documentation remain with their issuing organizations. Mythos Systems provides original engineering interpretation, synthesis, implementation guidance, and relationship mapping. Source verification dates do not guarantee that an authority has not changed since review.

01. Kubernetes - Kubernetes Releases
Role: Current supported-release and lifecycle information.
Verified: 2026-07-26
Official source: https://kubernetes.io/releases/

02. Kubernetes - Kubernetes Documentation
Role: Official architecture, API, workload, scheduling, security, storage, networking, and operations documentation.
Verified: 2026-07-26
Official source: https://kubernetes.io/docs/home/

03. Kubernetes - Kubernetes Controllers
Role: Official reconciliation and controller-loop model.
Verified: 2026-07-26
Official source: https://kubernetes.io/docs/concepts/architecture/controller/

04. Kubernetes - Kubernetes API Concepts
Role: API object, resource version, watch, patch, consistency, and concurrency semantics.
Verified: 2026-07-26
Official source: https://kubernetes.io/docs/reference/using-api/api-concepts/

05. Kubernetes SIGs - controller-runtime
Role: Libraries and patterns for building Kubernetes controllers.
Verified: 2026-07-26
Official source: https://pkg.go.dev/sigs.k8s.io/controller-runtime

06. Operator Framework - Operator SDK Documentation
Role: Operator construction, scaffolding, testing, and lifecycle guidance.
Verified: 2026-07-26
Official source: https://sdk.operatorframework.io/

07. etcd - etcd Documentation
Role: Distributed key-value state, consistency, watch, transactions, operations, and recovery.
Verified: 2026-07-26
Official source: https://etcd.io/docs/

08. Kubernetes - Kueue Documentation
Role: Kubernetes-native batch and AI workload queueing, admission, cohorts, and resource sharing.
Verified: 2026-07-26
Official source: https://kueue.sigs.k8s.io/

09. Kubernetes SIG Network - Gateway API
Role: Kubernetes-native role-oriented L4/L7 routing and service-mesh APIs.
Verified: 2026-07-26
Official source: https://gateway-api.sigs.k8s.io/docs/introduction/

Authority application and refresh triggers

Authority class	Required library action
Protocols and specifications	Recheck interoperability, security considerations, defaults, and migration guidance when a referenced specification changes.
Security and assurance frameworks	Update control mappings, evidence expectations, risk language, and assessment procedures after material framework revision.
Languages, runtimes, and projects	Re-run examples, tests, deprecation checks, and operational recommendations against supported releases.
Benchmarks and evaluations	Re-execute claims when workloads, methods, models, compilers, drivers, hardware, or scoring rules change.
Operational evidence	Incorporate validated incidents, failure tests, reader corrections, and measured implementation outcomes into the next revision.

Limitations, freshness, and revision control

Boundary	Statement
Publication limitation	This guide synthesizes broad engineering practice. It does not replace product documentation, legal advice, safety certification, or environment-specific design review.
Evidence limitation	Not every pattern is benchmarked in every environment. Examples illustrate engineering methods and must be validated against actual versions, workloads, hardware, policy, and failure conditions.
Freshness limitation	Vendor behavior, software libraries, AI models, security threats, and emerging standards can change quickly. Volatile claims require source revalidation before operational adoption.
Adoption limitation	Frontier content is not an implicit adoption recommendation. Adoption requires defined value, qualification gates, controlled proof, rollback, ownership, and residual-risk acceptance.
Status boundary	Candidate Focused Field Guide describes the publication, not certification of a product, implementation, engineer, or organization.

Revision record

Version	Revision statement
1.0.0-candidate	Initial candidate living-guide edition. Published 2026-07-26; scheduled review 2026-10-26.

Search and relationship metadata

Dimension	Engineering position
Primary domain	MYTHOS-CLD - Cloud, Platform & Distributed Systems
Secondary domain	MYTHOS-SWE; MYTHOS-SEC; MYTHOS-DAT
Publication class	Field Guide / Canonical Living Overview Guide
Skill levels	L1 Foundational; L2 Practitioner; L3 Advanced; L4 Frontier
Tags	Kubernetes, controllers, operators, CRD, scheduler, etcd, cluster engineering, Kueue, GPU workloads, multi-cluster
Canonical route	/library/field-guides/mythos-fg-cld-0002/

Living publication

Corrections, expanded laboratories, improved diagrams, authority changes, and new engineering evidence should revise this permanent publication rather than create disconnected replacements.