MYTHOS-CMP-PORT-0001

Complete Comparative Evaluation Library Portfolio

The final benchmark method, current evidence refreshes for CMP-001 through CMP-008, and retained baseline analytic records.

VERSION 1.2.0-candidate / 2026-09-17

PERMANENT PORTFOLIO EDITION / PUBLIC

Complete Comparative Evaluation Library Portfolio

The final benchmark method, current evidence refreshes for CMP-001 through CMP-008, and retained baseline analytic records.

CMP-001	NetBox and Nautobot	CURRENT LAB RERUN PENDING
CMP-002	Enterprise NGFW Platforms	CURRENT LAB RERUN PENDING
CMP-003	Enterprise SIEM Platforms	CURRENT LAB RERUN PENDING
CMP-004	Agentic Frameworks	CURRENT LAB RERUN PENDING
CMP-005	Modern Private Networking	CURRENT LAB RERUN PENDING
CMP-006	Campus Networking Platforms	CURRENT LAB RERUN PENDING
CMP-007	IaC and Control Planes	CURRENT LAB RERUN PENDING
CMP-008	Configuration Management	CURRENT LAB RERUN PENDING

COMPARATIVE EVALUATION / EVIDENCE REFRESH

CMP-001

NetBox and Nautobot

Network source-of-truth platforms

CURRENT EVIDENCE SNAPSHOT

Official product sources refreshed through 2026-09-17

Controlled Mythos laboratory rerun: NOT ASSERTED

BENCHMARK DOCTRINE

Gates before scores. Evidence before certainty. Profile fit before universal ranking.

VERSION 1.2.0-candidate

CANDIDATE COMPARATIVE EVALUATION / PUBLIC

NetBox and Nautobot

Network source-of-truth platforms

DOCUMENT ID
CMP-001
VERSION
1.2.0-candidate
STATUS
Candidate Comparative Evaluation
PUBLICATION DATE
2026-09-17
SCHEDULED REVIEW
2026-12-16
CANONICAL ROUTE
/library/evaluations/cmp-001/

2 CANDIDATES

18 ATOMIC CRITERIA

9 MANDATORY GATES

3 WORKLOAD PROFILES

E2 CURRENT MAX EVIDENCE

CURRENT DECISION BOUNDARY

Do not issue a current winner. The July baseline is a near-tie and both platforms changed materially. Freeze the candidate versions, then rerun upgrade, reconciliation, integration, failure, and export tests before selecting an authority platform.

NO CURRENT UNIVERSAL WINNER IS PUBLISHED FROM THIS REFRESH.

July 24 baseline scores are preserved as lineage and sensitivity evidence, not silently reissued as September laboratory results.

NetBox

v4.7.1 / 2026-09-15

REFRESH TRIGGER

4.7 line adds new infrastructure models and REST/background-processing changes.

Nautobot

v3.2.5 / 2026-09-14

REFRESH TRIGGER

3.2 includes security fixes and job-execution API behavior changes requiring migration attention.

FINAL BENCHMARK SYSTEM / DECISION ARCHITECTURE

01 SCOPE

Freeze decision, candidates, versions and exclusions.

02 GATES

Test non-compensable trust, state, recovery, change and exit.

03 SCORES

Measure atomic capability with approved profile weights.

04 CONFIDENCE

Discount by source grade, freshness, repeatability and relevance.

05 PROOF

Run normal, failure, migration, recovery and operator scenarios.

06 DECISION

Record conditional fit, residual risk, owner and revalidation trigger.

EVIDENCE CONFIDENCE / MODIFIERS

E0	E1	E2	E3	E4	E5
0.00	0.38	0.64	0.78	0.91	1.00
UNSUPPORTED	VENDOR ASSERTION	OFFICIAL SOURCE	MYTHOS LAB	PRODUCTION	INDEPENDENT

THIS EDITION

Current official product and release sources have been refreshed to E2. No E3 controlled Mythos lab rerun is claimed. Therefore current profile rankings remain intentionally unissued until the proof plan is executed.

What changed since the July benchmark snapshot

Changes below are sourced from current official release documentation. They identify revalidation triggers; they are not translated into new numeric scores without controlled proof.

NetBox

v4.7.1 / 2026-09-15

MATERIAL DELTA

4.7 line adds new infrastructure models and REST/background-processing changes.

CRITERIA TOUCHED

C01 / C05 / C08 / C14

MODERATE REVALIDATION

Nautobot

v3.2.5 / 2026-09-14

MATERIAL DELTA

3.2 includes security fixes and job-execution API behavior changes requiring migration attention.

CRITERIA TOUCHED

C02 / C05 / C12 / C14

HIGH REVALIDATION

Mandatory gates remain non-compensable

No current release is marked PASS solely from documentation. E2 means the official source supports the capability path; LAB and RETEST identify proof still required. HOLD blocks a current decision.

MANDATORY GATE	NETBOX	NAUTOBOT
C01 Functional coverage	RETEST	E2
C02 Security / trust	E2	RETEST
C03 Governance / approval	E2	E2
C04 State integrity	E2	E2
C07 Observability / evidence	E2	E2
C09 Resilience / continuity	LAB	LAB
C11 Change safety / rollback	LAB	LAB
C16 Portability / exit	LAB	LAB
C18 Assurance confidence	HOLD	HOLD

LEGEND

E2 official-source support only

LAB controlled proof required

RETEST release delta touches this gate

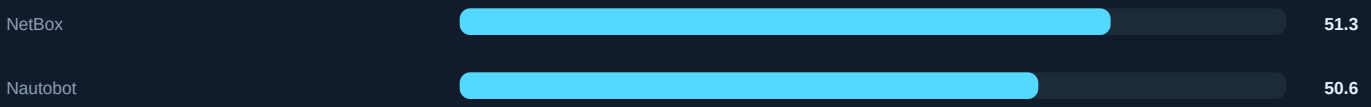
HOLD decision gate remains closed

Historical profile sensitivity - preserved, not reissued

Values below are the 2026-07-24 evidence-adjusted baseline. They show sensitivity to operating-model weights. The September refresh does not recompute rank because E3 lab proof has not been reproduced.

AUTHORITATIVE DATA CORE

BASELINE SPREAD 0.7 PTS

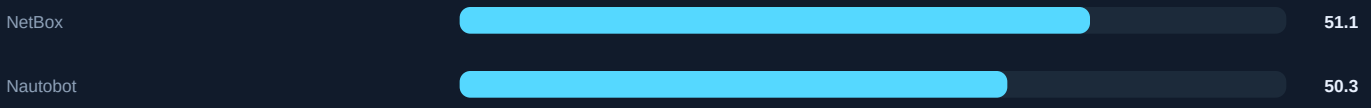


INTEGRATED AUTOMATION PLATFORM

BASELINE SPREAD 0.5 PTS

SOVEREIGN ENGINEERING PLATFORM

BASELINE SPREAD 0.8 PTS



CURRENT RANK STATUS: WITHHELD PENDING CONTROLLED PROOF.

Evidence confidence is separate from capability

E5 — Independent repeatable validation
confidence modifier 1.00

E4 — Production evidence in adopting environment
confidence modifier 0.91

E3 — Controlled Mythos laboratory result
confidence modifier 0.78

E2 — Official documentation / release / source
confidence modifier 0.64

E1 — Vendor assertion or marketing
confidence modifier 0.38

E0 — Unsupported or unverified
confidence modifier 0.00

CURRENT EVIDENCE STATE

NetBox — E2 REFRESHED
Historical adjusted confidence: 59.7%
E3 LAB: NOT REPRODUCED

Nautobot — E2 REFRESHED
Historical adjusted confidence: 59.7%
E3 LAB: NOT REPRODUCED

ASSURANCE RULE

A current release note can change capability context, but it cannot raise assurance beyond E2. Current recommendation strength is therefore bounded until the test plan produces reproducible artifacts.

Lifecycle risk is evaluated independently of features

The benchmark models migration, upgrade, compatibility, support, staffing, data/state export, downtime and exit. Current release changes below are explicit proof triggers.

NetBox

MODERATE REVALIDATION

4.7 line adds new infrastructure models and REST/background-processing changes.

RETEST: C01 / C05 / C08 / C14

Nautobot

HIGH REVALIDATION

3.2 includes security fixes and job-execution API behavior changes requiring migration attention.

RETEST: C02 / C05 / C12 / C14

IRREVERSIBLE COMMITMENT GATE

Before migration or procurement lock-in, prove rollback, state portability, operational reconstruction, and supplier-exit assumptions.

Controlled proof is the next decision-producing step

REPRODUCTION STATE / NOT EXECUTED IN THIS EVIDENCE-REFRESH RELEASE

01

CONTROLLED SCENARIO

Model a representative multi-site network with IPAM, circuits, devices, virtualization, tenancy, and custom data.

02

CONTROLLED SCENARIO

Import authoritative and conflicting data from two systems and prove deterministic reconciliation.

03

CONTROLLED SCENARIO

Execute a guarded change workflow with approval, dry-run, evidence, and rollback.

04

CONTROLLED SCENARIO

Measure API throughput, query latency, job concurrency, and database recovery.

05

CONTROLLED SCENARIO

Upgrade a realistic plugin or application set and record compatibility failures.

06

CONTROLLED SCENARIO

Export the full data model and rebuild the environment from controlled backups.

EVIDENCE PACKAGE

Preserve exact versions, architecture, configuration, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

Current decision posture

Do not issue a current winner. The July baseline is a near-tie and both platforms changed materially. Freeze the candidate versions, then rerun upgrade, reconciliation, integration, failure, and export tests before selecting an authority platform.

CURRENT RANKING: NOT REISSUED / CONTROLLED LAB REQUIRED

CURRENT OFFICIAL-SOURCE REGISTER / CHECKED 2026-09-17

NetBox	E2 / OFFICIAL	
NetBox GitHub releases		github.com
Nautobot	E2 / OFFICIAL	
Nautobot v3.2 release notes		docs.nautobot.com

REVISION LINEAGE

1.2.0-candidate / 2026-09-17 - current-source evidence refresh; no lab rescore issued.

1.1.0-candidate / 2026-07-24 - baseline benchmark using the final comparative evaluation system.

Trigger earlier review after major release, acquisition, licensing change, critical vulnerability, material outage, failed PoC, or workload change.

BASELINE ANALYTIC RECORD CONTINUES ON PAGE 11 FOR TRACEABILITY.

ATOMIC CRITERIA MODEL

WEIGHTED CAPABILITY WITH EXPLICIT MINIMUM EVIDENCE

ID	CRITERION	WEIGHT	GATE	MINIMUM EVIDENCE
C01	Network source-of-truth functional coverage	5	YES	Feature documentation, APIs, configuration exports, and scenario demonstration.
C02	Security and trust architecture	5	YES	Threat model, identity flows, RBAC tests, audit records, and security configuration.
C03	Governance and approval controls	5	YES	Approval workflow, policy controls, separation-of-duties tests, and decision records.
C04	Data-model integrity and reconciliation	5	YES	Backup, restore, rollback, drift, and corruption-recovery evidence.
C05	Automation and API quality	4	NO	API specifications, authentication tests, rate limits, event samples, and automation runs.
C06	Integration ecosystem	4	NO	Supported integrations, connector tests, failure behavior, and lifecycle ownership.
C07	Observability and evidence	4	YES	Logs, traces, metrics, reports, export tests, and evidence-retention controls.
C08	Object, API, and job scale	4	NO	Controlled load tests, topology scale, saturation behavior, and capacity model.
C09	Resilience and continuity	5	YES	Failure injection, HA tests, recovery timing, degraded-mode operation, and runbooks.
C10	Usability and operator cognition	3	NO	Task observation, error-rate measures, navigation tests, and operator interviews.
C11	Change safety and rollback	5	YES	Plan or preview output, canary tests, rollback execution, and post-change verification.
C12	Extensibility and programmability	3	NO	Plugin, module, provider, workflow, or code-extension tests and upgrade impact analysis.
C13	Deployment and control-plane sovereignty	3	NO	Deployment architecture, data-flow mapping, residency evidence, and offline tests.
C14	Lifecycle and upgrade discipline	4	NO	Release notes, upgrade test, compatibility matrix, support policy, and migration plan.
C15	Economics and cost predictability	3	NO	Bill-of-materials, licensing model, staffing estimates, metering, and sensitivity analysis.
C16	Portability and exit	4	YES	Export tests, format analysis, replacement workflow, and decommission evidence.
C17	Vendor and ecosystem risk	3	NO	License analysis, roadmap evidence, bus-factor indicators, and dependency inventory.
C18	Assurance confidence	5	YES	Source provenance, lab artifacts, production evidence, independent replication, and review date.

SCORE SCALE

0 absent; 1 materially deficient; 2 partial; 3 adequate with limits; 4 strong; 5 leading for the defined profile. Scores remain provisional until the required evidence grade is achieved.

RAW CAPABILITY SCORECARD

DOCUMENTED CAPABILITY BEFORE EVIDENCE DISCOUNT

CRITERION	NETBOX	NAUTOBOT
C01 Network source-of-truth functional coverage	4.6	4.5
C02 Security and trust architecture	4.1	4.0
C03 Governance and approval controls	3.8	4.0
C04 Data-model integrity and reconciliation	4.4	4.2
C05 Automation and API quality	4.7	4.6
C06 Integration ecosystem	4.4	4.5
C07 Observability and evidence	4.1	4.2
C08 Object, API, and job scale	4.3	4.0
C09 Resilience and continuity	4.0	3.9
C10 Usability and operator cognition	4.2	4.1
C11 Change safety and rollback	4.0	4.2
C12 Extensibility and programmability	4.4	4.8
C13 Deployment and control-plane sovereignty	4.7	4.5
C14 Lifecycle and upgrade discipline	4.2	4.0
C15 Economics and cost predictability	4.5	4.1
C16 Portability and exit	4.6	4.3
C17 Vendor and ecosystem risk	4.4	4.1
C18 Assurance confidence	3.5	3.5

WEIGHTED BASELINE / 100

NetBox	Nautobot
84.8	83.4

EVIDENCE-ADJUSTED SCORECARD

CAPABILITY DISCOUNTED BY CURRENT EVIDENCE CONFIDENCE

CANDIDATE	RAW	EVIDENCE CONFIDENCE	ADJUSTED	CURRENT STATE
NetBox	84.8	59.7%	50.7	CONTROLLED LAB PENDING
Nautobot	83.4	59.7%	50.0	CONTROLLED LAB PENDING

EVIDENCE SCALE

E0 / 0.00
Unsupported or unverified

E1 / 0.38
Vendor assertion or marketing statement

E2 / 0.64
Official documentation, release notes, or source repository

E3 / 0.78
Controlled Mythos laboratory result

E4 / 0.91
Production evidence from the adopting environment

E5 / 1.00
Independent repeatable validation

CURRENT CONFIDENCE BOUNDARY

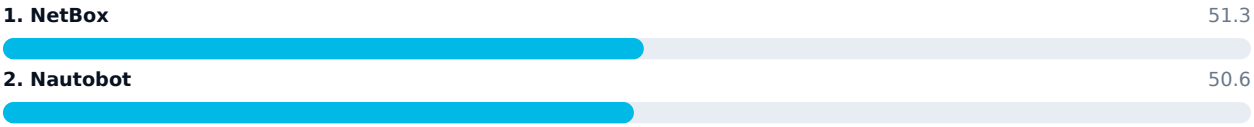
Most candidate criteria are supported at E2: official documentation or source evidence. Environment-specific laboratory, production, and independent evidence remain future gates.

PROFILE-SPECIFIC RANKINGS

EVIDENCE-ADJUSTED SENSITIVITY ANALYSIS

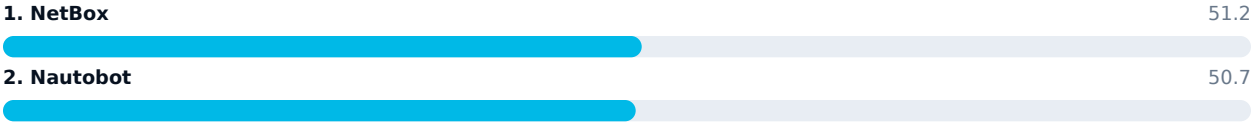
AUTHORITATIVE DATA CORE

Prioritizes clean network modeling, external-tool interoperability, portability, and low coupling.



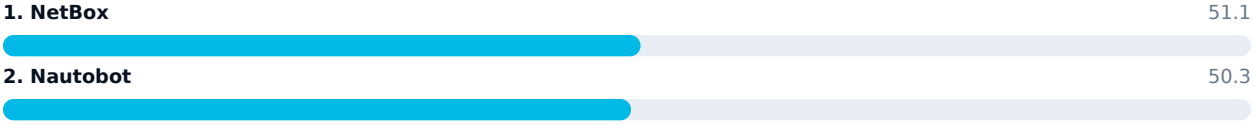
INTEGRATED AUTOMATION PLATFORM

Prioritizes embedded execution, extensibility, synchronization, workflow governance, and operator productivity.



SOVEREIGN ENGINEERING PLATFORM

Prioritizes self-hosting, open interfaces, exportability, and control over data and runtime dependencies.



RANK SENSITIVITY

Small score differences are not decision certainty. Treat near-ties as a requirement for deeper PoC, commercial, migration, and operating-model evidence.

CANDIDATE DEEP DIVE / 01

NETBOX

OPERATING MODEL

Focused network source-of-truth platform centered on DCIM, IPAM, extensible data modeling, APIs, plugins, and integration with external automation systems.

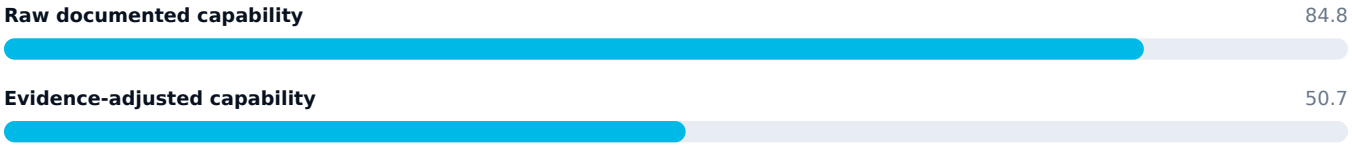
DOCUMENTED STRENGTHS

- Mature network modeling and documentation posture.
- Strong REST API and broad ecosystem adoption.
- Clear separation between source-of-truth data and external execution systems.
- Open-source core with commercial support paths.

CAUTIONS AND PROOF OBLIGATIONS

- Complex orchestration generally remains external.
- Plugin governance and upgrade compatibility require discipline.
- Desired-state workflows must be engineered around the authoritative model.

BASELINE SCORE PROFILE



GATE DISPOSITION

PASS - core source-of-truth profile | CONDITIONAL - embedded workflow profile requires external orchestration

CANDIDATE DEEP DIVE / 02

NAUTOBOT

OPERATING MODEL

Network automation platform combining source-of-truth data, Jobs, applications, design workflows, and synchronization patterns.

DOCUMENTED STRENGTHS

- Integrated Jobs and application framework.
- Strong desired-state and automation-platform orientation.
- SSoT synchronization patterns and DiffSync ecosystem.
- Extensible workflows can reduce handoffs between data and execution.

CAUTIONS AND PROOF OBLIGATIONS

- Integrated execution increases platform responsibility and governance scope.
- Application compatibility and lifecycle must be controlled.
- Operating model can be heavier than a narrowly scoped source of truth.

BASELINE SCORE PROFILE

Raw documented capability 83.4

Evidence-adjusted capability 50.0

GATE DISPOSITION

PASS - integrated automation profile | CONDITIONAL - operating complexity and app lifecycle must be accepted

CROSS-CANDIDATE CONTRASTS

WHERE ARCHITECTURAL DIFFERENCES MATTER MOST

CONTROL-LOOP MODEL

How authority moves from intent to plan, approval, execution, observation, reconciliation, and recovery.

STATE OWNERSHIP

Which system owns desired state, operational state, evidence, history, and conflict resolution.

MANAGED VERSUS SOVEREIGN BOUNDARY

Which runtime, identity, data, policy, relay, or service dependencies remain outside organizational control.

EXTENSION MODEL

Plugins, providers, modules, applications, workflows, SDKs, APIs, and custom code introduce different lifecycle risks.

FAILURE TRANSPARENCY

Candidates differ in how clearly they expose partial execution, degraded mode, retries, inconsistency, and recovery state.

EXIT MECHANICS

Export formats, state semantics, policy portability, knowledge transfer, and replacement effort are materially different.

CANDIDATE	CURRENT ADVANCEMENT POSITION
NetBox	PASS - core source-of-truth profile
Nautobot	PASS - integrated automation profile

CROSS-CANDIDATE CONTRASTS

WHERE ARCHITECTURAL DIFFERENCES MATTER MOST

CONTROL-LOOP MODEL

How authority moves from intent to plan, approval, execution, observation, reconciliation, and recovery.

STATE OWNERSHIP

Which system owns desired state, operational state, evidence, history, and conflict resolution.

MANAGED VERSUS SOVEREIGN BOUNDARY

Which runtime, identity, data, policy, relay, or service dependencies remain outside organizational control.

EXTENSION MODEL

Plugins, providers, modules, applications, workflows, SDKs, APIs, and custom code introduce different lifecycle risks.

FAILURE TRANSPARENCY

Candidates differ in how clearly they expose partial execution, degraded mode, retries, inconsistency, and recovery state.

EXIT MECHANICS

Export formats, state semantics, policy portability, knowledge transfer, and replacement effort are materially different.

CANDIDATE	CURRENT ADVANCEMENT POSITION
NetBox	PASS - core source-of-truth profile
Nautobot	PASS - integrated automation profile

ARCHITECTURE AND INTEGRATION FIT

THE PRODUCT IS ONLY ONE PART OF THE OPERATING SYSTEM

REFERENCE OPERATING LAYERS

01 Authoritative data model and ownership boundaries

02 API and event integration layer

03 Validation, reconciliation, and drift workflow

04 Execution platform or external orchestrators

05 Evidence, audit, and recovery store

INTEGRATION BOUNDARIES

- Identity provider and RBAC
- Git and CI/CD
- Telemetry and event bus
- ITSM and change management
- Device discovery and collectors
- Ansible, Terraform/OpenTofu, and network automation frameworks

ARCHITECTURE GATE

Every external dependency requires an owner, identity boundary, failure mode, evidence path, version policy, recovery procedure, and exit strategy.

SECURITY, OPERATIONS, LIFECYCLE, ECONOMICS, AND EXIT

CROSS-DOMAIN DECISION RECORD

SECURITY AND AUTHORITY

Identity, credentials, secrets, roles, privileged actions, signing, approvals, isolation, audit, and incident response must be tested as an integrated system.

RELIABILITY AND RECOVERY

Control-plane, data-plane, dependency, region, database, queue, network, and operator failures require defined degraded modes and recovery objectives.

CHANGE AND LIFECYCLE

Version, compatibility, migration, canary, rollback, content, plugin, provider, firmware, and decommission procedures are part of product fitness.

ECONOMICS AND CAPACITY

Licenses, metering, data, compute, hardware, storage, support, staffing, training, integration, migration, and opportunity cost require sensitivity analysis.

SOVEREIGNTY AND PRIVACY

Data location, support access, telemetry, subprocessors, encryption, key control, disconnected operation, and legal obligations must align with policy.

PORTABILITY AND EXIT

Configuration, policy, data, state, evidence, workflows, identities, and operational knowledge must be exportable and reconstructable.

DECISION OWNERSHIP

Architecture, security, operations, finance, procurement, legal, data, and service owners jointly accept the final profile, evidence, residual risk, and exit obligations.

RISK AND FAILURE REGISTER

SCENARIOS THAT CAN INVALIDATE A FEATURE-BASED SELECTION

ID	SEVERITY	FAILURE SCENARIO	REQUIRED TREATMENT
R-01	CRITICAL	A source of truth becomes a passive inventory because no authority model is enforced.	PREVENT / DETECT / RECOVER / EVIDENCE
R-02	HIGH	Multiple systems write overlapping fields without reconciliation ownership.	PREVENT / DETECT / RECOVER / EVIDENCE
R-03	HIGH	Plugins or applications create upgrade deadlocks.	PREVENT / DETECT / RECOVER / EVIDENCE
R-04	HIGH	Automation consumes stale or incomplete data and produces unsafe changes.	PREVENT / DETECT / RECOVER / EVIDENCE
R-05	MEDIUM	Backup restores records but not integrations, jobs, secrets, or evidence.	PREVENT / DETECT / RECOVER / EVIDENCE
R-06	MEDIUM	A platform is selected for feature breadth without a sustainable operating model.	PREVENT / DETECT / RECOVER / EVIDENCE

RISK RULE

A candidate with an unresolved critical failure path cannot advance because optional capability, market position, or commercial discount cannot compensate for missing safety or recovery evidence.

CONTROLLED PROOF-OF-CAPABILITY PLAN

REPEATABLE SCENARIOS, INSTRUMENTATION, AND ACCEPTANCE

TEST 01

Model a representative multi-site network with IPAM, circuits, devices, virtualization, tenancy, and custom data.

TEST 02

Import authoritative and conflicting data from two systems and prove deterministic reconciliation.

TEST 03

Execute a guarded change workflow with approval, dry-run, evidence, and rollback.

TEST 04

Measure API throughput, query latency, job concurrency, and database recovery.

TEST 05

Upgrade a realistic plugin or application set and record compatibility failures.

TEST 06

Export the full data model and rebuild the environment from controlled backups.

EVIDENCE PACKAGE

Preserve versions, architecture, configuration, data set, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

CONDITIONAL RECOMMENDATION AND REVALIDATION

DECISION RECORD, ACCEPTANCE, AND NEXT EVIDENCE

RECOMMENDATION POSITION

- Select NetBox for a focused data authority with external deterministic execution.
- Select Nautobot for an integrated network automation platform with governed application lifecycle.
- Require an environment-specific PoC before committing the authoritative data model.
- Revalidate releases, plugins, commercial packaging, and operational scale every 90 days.

CANDIDATE	ADJ. SCORE	GATE POSITION	LAB STATE
NetBox	50.7	PASS - core source-of-truth profile; CONDITIONAL - embedded workflow profile requires external orchestration	PENDING
Nautobot	50.0	PASS - integrated automation profile; CONDITIONAL - operating complexity and app lifecycle must be accepted	PENDING

REVALIDATION SCHEDULE

Review no later than 2026-10-24. Review earlier after a major release, commercial change, critical vulnerability, material outage, architecture transition, failed acceptance test, or change to the target workload profile.

SOURCE AUTHORITY AND REVISION

CURRENT EVIDENCE SNAPSHOT AND PUBLICATION HISTORY

SOURCE ID	PUBLISHER	TITLE	CHECKED
NIST-CSF-20	NIST	Cybersecurity Framework 2.0	2026-07-24
NIST-SP800-53	NIST	Security and Privacy Controls for Information Systems and Organizations	2026-07-24
NETBOX-DOCS	NetBox Labs	NetBox Documentation	2026-07-24
NETBOX-API	NetBox Labs	NetBox REST API Overview	2026-07-24
NETBOX-GH	NetBox Community	NetBox Source Repository	2026-07-24
NAUTOBOT-DESIGN	Network to Code	Nautobot Design Philosophy	2026-07-24
NAUTOBOT-JOBS	Network to Code	Nautobot Jobs Developer Guide	2026-07-24
NAUTOBOT-SSOT	Network to Code	Nautobot Single Source of Truth Application	2026-07-24

SOURCE POSITION

Official documentation and source repositories support the current capability model. Source records preserve publisher, title, URL, purpose, and review date in the machine-readable authority register. Commercial terms, performance, and environment-specific behavior remain unverified until controlled proof.

REVISION HISTORY

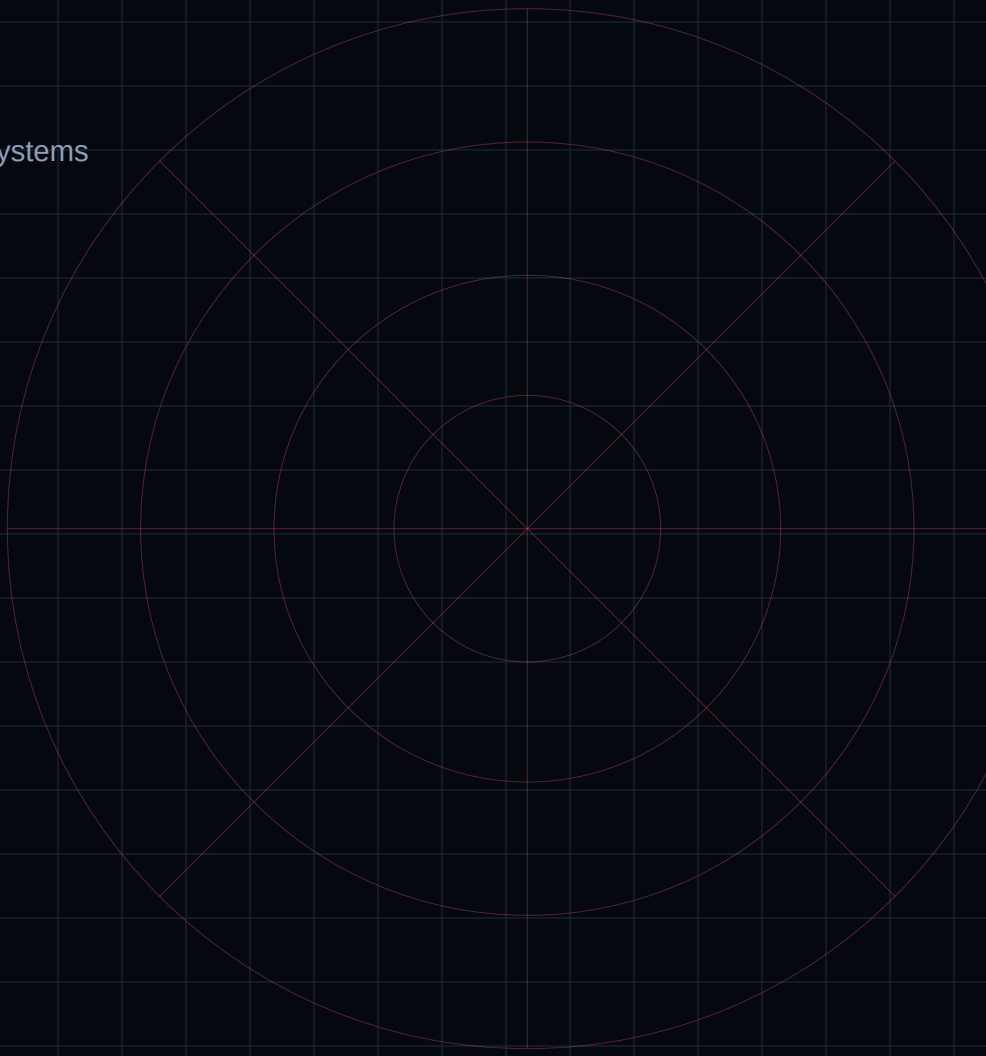
1.1.0-candidate / 2026-07-24 - permanent-publication rebuild using the final benchmark system, profile-specific rankings, evidence adjustment, mandatory gates, and controlled PoC requirements.

COMPARATIVE EVALUATION / EVIDENCE REFRESH

CMP-002

Enterprise NGFW Platforms

Enterprise firewall and centralized policy systems



CURRENT EVIDENCE SNAPSHOT

Official product sources refreshed through 2026-09-17

Controlled Mythos laboratory rerun: NOT ASSERTED

BENCHMARK DOCTRINE

Gates before scores. Evidence before certainty. Profile fit before universal ranking.

VERSION 1.2.0-candidate

CANDIDATE COMPARATIVE EVALUATION / PUBLIC

Enterprise NGFW Platforms

Enterprise firewall and centralized policy systems

DOCUMENT ID
CMP-002
VERSION
1.2.0-candidate
STATUS
Candidate Comparative Evaluation
PUBLICATION DATE
2026-09-17
SCHEDULED REVIEW
2026-12-16
CANONICAL ROUTE
/library/evaluations/cmp-002/

4
CANDIDATES

18
ATOMIC CRITERIA

9
MANDATORY GATES

3
WORKLOAD PROFILES

E2
CURRENT MAX EVIDENCE

CURRENT DECISION BOUNDARY

No universal NGFW winner is defensible. The July ordering is historical only. Current major-version and management-plane changes require profile-specific policy, HA, upgrade, rollback, logging, automation, and migration proof.

NO CURRENT UNIVERSAL WINNER IS PUBLISHED FROM THIS REFRESH.

July 24 baseline scores are preserved as lineage and sensitivity evidence, not silently reissued as September laboratory results.

Palo Alto Networks

PAN-OS 12.2 / SCM 3.0

REFRESH TRIGGER

PAN-OS 12.2 and Strata Cloud Manager 3.0 expand management, routing coexistence, audit, and modern TLS/PQC surfaces.

Check Point

R82.10 SC Build 428 / 2026-09-09

REFRESH TRIGGER

R82.10 SmartConsole remains active with current fixes and change-report updates.

Fortinet

FortiOS 7.6.7 / 2026-06-02

REFRESH TRIGGER

7.6.7 adds operational and network features while retaining upgrade/known-issue considerations; 8.0 documentation is also present.

Cisco Secure Firewall

Threat Defense 10.0.2 / 2026-09-15

REFRESH TRIGGER

Version 10 is on a new release numbering/cadence with active 7.x maintenance streams alongside it.

FINAL BENCHMARK SYSTEM / DECISION ARCHITECTURE

01 SCOPE

Freeze decision, candidates, versions and exclusions.

02 GATES

Test non-compensable trust, state, recovery, change and exit.

03 SCORES

Measure atomic capability with approved profile weights.

04 CONFIDENCE

Discount by source grade, freshness, repeatability and relevance.

05 PROOF

Run normal, failure, migration, recovery and operator scenarios.

06 DECISION

Record conditional fit, residual risk, owner and revalidation trigger.

EVIDENCE CONFIDENCE / MODIFIERS

E0	E1	E2	E3	E4	E5
0.00	0.38	0.64	0.78	0.91	1.00
UNSUPPORTED	VENDOR ASSERTION	OFFICIAL SOURCE	MYTHOS LAB	PRODUCTION	INDEPENDENT

THIS EDITION

Current official product and release sources have been refreshed to E2. No E3 controlled Mythos lab rerun is claimed. Therefore current profile rankings remain intentionally unissued until the proof plan is executed.

What changed since the July benchmark snapshot

Changes below are sourced from current official release documentation. They identify revalidation triggers; they are not translated into new numeric scores without controlled proof.

Palo Alto Networks

PAN-OS 12.2 / SCM 3.0

MATERIAL DELTA

PAN-OS 12.2 and Strata Cloud Manager 3.0 expand management, routing coexistence, audit, and modern TLS/PQC surfaces.

CRITERIA TOUCHED

C02 / C03 / C07 / C11 / C14

MODERATE REVALIDATION

Check Point

R82.10 SC Build 428 / 2026-09-09

MATERIAL DELTA

R82.10 SmartConsole remains active with current fixes and change-report updates.

CRITERIA TOUCHED

C03 / C07 / C11 / C14

MODERATE REVALIDATION

Fortinet

FortiOS 7.6.7 / 2026-06-02

MATERIAL DELTA

7.6.7 adds operational and network features while retaining upgrade/known-issue considerations; 8.0 documentation is also present.

CRITERIA TOUCHED

C01 / C09 / C11 / C14

HIGH REVALIDATION

Cisco Secure Firewall

Threat Defense 10.0.2 / 2026-09-15

MATERIAL DELTA

Version 10 is on a new release numbering/cadence with active 7.x maintenance streams alongside it.

CRITERIA TOUCHED

C01 / C09 / C11 / C14

HIGH REVALIDATION

Mandatory gates remain non-compensable

No current release is marked PASS solely from documentation. E2 means the official source supports the capability path; LAB and RETEST identify proof still required. HOLD blocks a current decision.

MANDATORY GATE	PALO ALTO NETWO	CHECK POINT	FORTINET	CISCO SECURE FI
C01 Functional coverage	E2	E2	RETEST	RETEST
C02 Security / trust	RETEST	E2	E2	E2
C03 Governance / approval	RETEST	RETEST	E2	E2
C04 State integrity	E2	E2	E2	E2
C07 Observability / evidence	RETEST	RETEST	E2	E2
C09 Resilience / continuity	LAB	LAB	RETEST	RETEST
C11 Change safety / rollback	RETEST	RETEST	RETEST	RETEST
C16 Portability / exit	LAB	LAB	LAB	LAB
C18 Assurance confidence	HOLD	HOLD	HOLD	HOLD

LEGEND

E2 official-source support only

LAB controlled proof required

RETEST release delta touches this gate

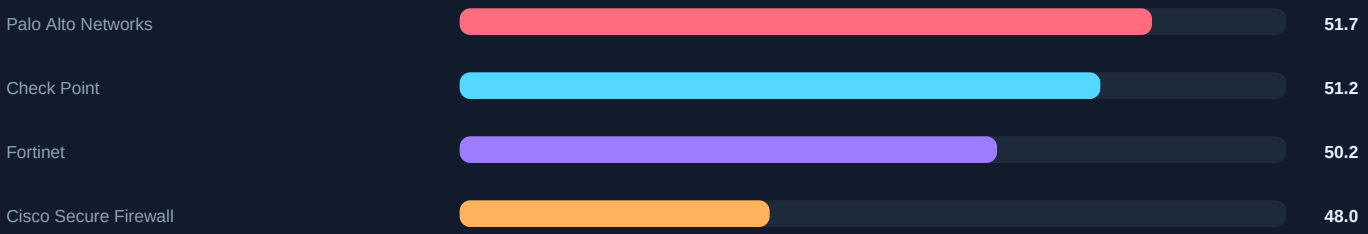
HOLD decision gate remains closed

Historical profile sensitivity - preserved, not reissued

Values below are the 2026-07-24 evidence-adjusted baseline. They show sensitivity to operating-model weights. The September refresh does not recompute rank because E3 lab proof has not been reproduced.

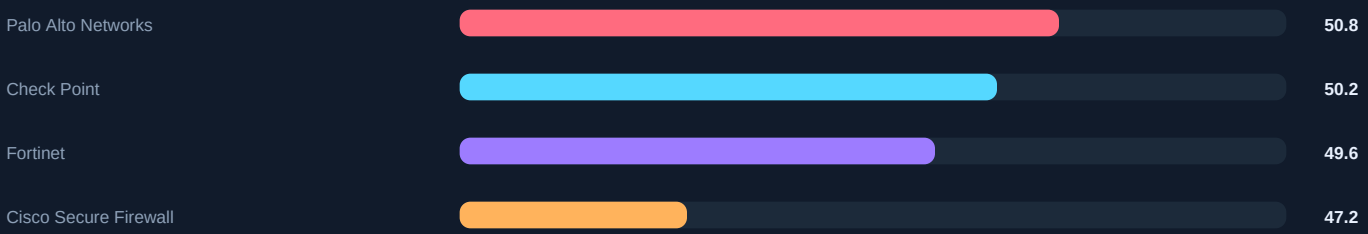
HIGH-ASSURANCE ENTERPRISE SECURITY

BASELINE SPREAD 3.7 PTS



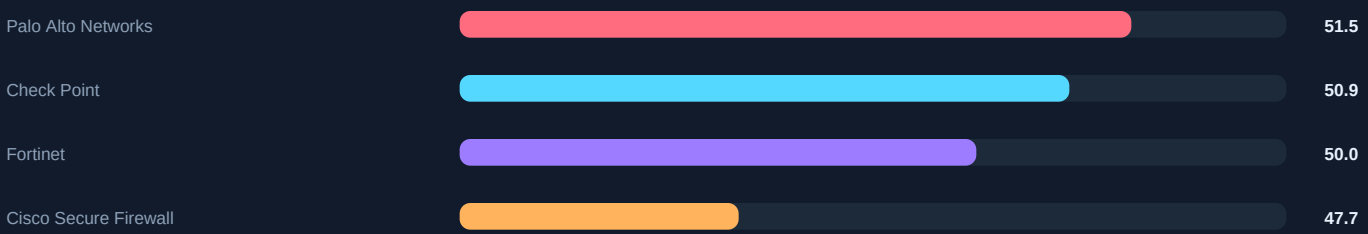
DISTRIBUTED BRANCH AND CAMPUS

BASELINE SPREAD 3.6 PTS



MULTI-DOMAIN SERVICE PROVIDER

BASELINE SPREAD 3.8 PTS



CURRENT RANK STATUS: WITHHELD PENDING CONTROLLED PROOF.

Evidence confidence is separate from capability

E5 — Independent repeatable validation
confidence modifier 1.00

E4 — Production evidence in adopting environment
confidence modifier 0.91

E3 — Controlled Mythos laboratory result
confidence modifier 0.78

E2 — Official documentation / release / source
confidence modifier 0.64

E1 — Vendor assertion or marketing
confidence modifier 0.38

E0 — Unsupported or unverified
confidence modifier 0.00

CURRENT EVIDENCE STATE

Palo Alto Networks — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

Check Point — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

Fortinet — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

Cisco Secure Firewall — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

ASSURANCE RULE

A current release note can change capability context, but it cannot raise assurance beyond E2. Current recommendation strength is therefore bounded until the test plan produces reproducible artifacts.

Lifecycle risk is evaluated independently of features

The benchmark models migration, upgrade, compatibility, support, staffing, data/state export, downtime and exit. Current release changes below are explicit proof triggers.

Palo Alto Networks

MODERATE REVALIDATION

PAN-OS 12.2 and Strata Cloud Manager 3.0 expand management, routing coexistence, audit, and modern TLS/PQC surfaces.

RETEST: C02 / C03 / C07 / C11 / C14

Check Point

MODERATE REVALIDATION

R82.10 SmartConsole remains active with current fixes and change-report updates.

RETEST: C03 / C07 / C11 / C14

Fortinet

HIGH REVALIDATION

7.6.7 adds operational and network features while retaining upgrade/known-issue considerations; 8.0 documentation is also present.

RETEST: C01 / C09 / C11 / C14

Cisco Secure Firewall

HIGH REVALIDATION

Version 10 is on a new release numbering/cadence with active 7.x maintenance streams alongside it.

RETEST: C01 / C09 / C11 / C14

IRREVERSIBLE COMMITMENT GATE

Before migration or procurement lock-in, prove rollback, state portability, operational reconstruction, and supplier-exit assumptions.

Controlled proof is the next decision-producing step

REPRODUCTION STATE / NOT EXECUTED IN THIS EVIDENCE-REFRESH RELEASE

01

CONTROLLED SCENARIO

Import a representative policy estate and compare rule intent, objects, NAT, VPN, routing, and segmentation.

02

CONTROLLED SCENARIO

Exercise high availability, management loss, log loss, backup, restore, and rollback.

03

CONTROLLED SCENARIO

Measure policy installation, session scale, inspection throughput, and logging under representative traffic.

04

CONTROLLED SCENARIO

Validate API-driven object and policy lifecycle with approval and evidence.

05

CONTROLLED SCENARIO

Test decryption, certificate lifecycle, exception handling, and privacy controls.

06

CONTROLLED SCENARIO

Conduct controlled migration with semantic review and post-cutover validation.

EVIDENCE PACKAGE

Preserve exact versions, architecture, configuration, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

Current decision posture

No universal NGFW winner is defensible. The July ordering is historical only. Current major-version and management-plane changes require profile-specific policy, HA, upgrade, rollback, logging, automation, and migration proof.

CURRENT RANKING: NOT REISSUED / CONTROLLED LAB REQUIRED

CURRENT OFFICIAL-SOURCE REGISTER / CHECKED 2026-09-17

Palo Alto Networks	E2 / OFFICIAL	
Palo Alto Networks Aug 2026 new features		docs.paloaltonetworks.com
Check Point	E2 / OFFICIAL	
Check Point R82.10 SmartConsole Build 428		sc1.checkpoint.com
Fortinet	E2 / OFFICIAL	
FortiOS 7.6.7 release notes		docs.fortinet.com
Cisco Secure Firewall	E2 / OFFICIAL	
Cisco Secure Firewall Version 10 release notes		www.cisco.com

REVISION LINEAGE

1.2.0-candidate / 2026-09-17 - current-source evidence refresh; no lab rescore issued.

1.1.0-candidate / 2026-07-24 - baseline benchmark using the final comparative evaluation system.

Trigger earlier review after major release, acquisition, licensing change, critical vulnerability, material outage, failed PoC, or workload change.

BASELINE ANALYTIC RECORD CONTINUES ON PAGE 11 FOR TRACEABILITY.

ATOMIC CRITERIA MODEL

WEIGHTED CAPABILITY WITH EXPLICIT MINIMUM EVIDENCE

ID	CRITERION	WEIGHT	GATE	MINIMUM EVIDENCE
C01	Enterprise NGFW and centralized-management coverage	5	YES	Feature documentation, APIs, configuration exports, and scenario demonstration.
C02	Policy, inspection, segmentation, and trust architecture	5	YES	Threat model, identity flows, RBAC tests, audit records, and security configuration.
C03	Governance and approval controls	5	YES	Approval workflow, policy controls, separation-of-duties tests, and decision records.
C04	Configuration, object, policy, and log integrity	5	YES	Backup, restore, rollback, drift, and corruption-recovery evidence.
C05	Automation and API quality	4	NO	API specifications, authentication tests, rate limits, event samples, and automation runs.
C06	Integration ecosystem	4	NO	Supported integrations, connector tests, failure behavior, and lifecycle ownership.
C07	Observability and evidence	4	YES	Logs, traces, metrics, reports, export tests, and evidence-retention controls.
C08	Gateway, tenant, policy, and log scale	4	NO	Controlled load tests, topology scale, saturation behavior, and capacity model.
C09	Resilience and continuity	5	YES	Failure injection, HA tests, recovery timing, degraded-mode operation, and runbooks.
C10	Usability and operator cognition	3	NO	Task observation, error-rate measures, navigation tests, and operator interviews.
C11	Change safety and rollback	5	YES	Plan or preview output, canary tests, rollback execution, and post-change verification.
C12	Extensibility and programmability	3	NO	Plugin, module, provider, workflow, or code-extension tests and upgrade impact analysis.
C13	Cloud, on-premises, sovereign, and disconnected deployment options	3	NO	Deployment architecture, data-flow mapping, residency evidence, and offline tests.
C14	Lifecycle and upgrade discipline	4	NO	Release notes, upgrade test, compatibility matrix, support policy, and migration plan.
C15	Economics and cost predictability	3	NO	Bill-of-materials, licensing model, staffing estimates, metering, and sensitivity analysis.
C16	Portability and exit	4	YES	Export tests, format analysis, replacement workflow, and decommission evidence.
C17	Vendor and ecosystem risk	3	NO	License analysis, roadmap evidence, bus-factor indicators, and dependency inventory.
C18	Assurance confidence	5	YES	Source provenance, lab artifacts, production evidence, independent replication, and review date.

SCORE SCALE

0 absent; 1 materially deficient; 2 partial; 3 adequate with limits; 4 strong; 5 leading for the defined profile. Scores remain provisional until the required evidence grade is achieved.

RAW CAPABILITY SCORECARD

DOCUMENTED CAPABILITY BEFORE EVIDENCE DISCOUNT

CRITERION	PALO ALTO	CHECK POINT	FORTINET	CISCO
C01 Enterprise NGFW and centralized-management coverage	4.8	4.7	4.6	4.3
C02 Policy, inspection, segmentation, and trust architecture	4.8	4.7	4.5	4.4
C03 Governance and approval controls	4.5	4.7	4.3	4.2
C04 Configuration, object, policy, and log integrity	4.5	4.5	4.3	4.1
C05 Automation and API quality	4.6	4.2	4.5	4.2
C06 Integration ecosystem	4.6	4.3	4.5	4.4
C07 Observability and evidence	4.6	4.5	4.3	4.3
C08 Gateway, tenant, policy, and log scale	4.6	4.5	4.6	4.2
C09 Resilience and continuity	4.5	4.6	4.3	4.2
C10 Usability and operator cognition	4.4	4.2	4.2	4.0
C11 Change safety and rollback	4.5	4.5	4.3	4.1
C12 Extensibility and programmability	4.4	4.1	4.3	3.9
C13 Cloud, on-premises, sovereign, and disconnected deployment options	4.0	4.2	4.4	4.0
C14 Lifecycle and upgrade discipline	4.3	4.2	4.0	3.8
C15 Economics and cost predictability	3.4	3.5	4.2	3.7
C16 Portability and exit	3.7	3.8	3.9	3.6
C17 Vendor and ecosystem risk	3.7	3.8	3.7	3.6
C18 Assurance confidence	3.6	3.6	3.5	3.4

WEIGHTED BASELINE / 100

Palo Alto	Check Point	Fortinet	Cisco
86.8	85.9	85.0	80.8

EVIDENCE-ADJUSTED SCORECARD

CAPABILITY DISCOUNTED BY CURRENT EVIDENCE CONFIDENCE

CANDIDATE	RAW	EVIDENCE CONFIDENCE	ADJUSTED	CURRENT STATE
Palo Alto	86.8	58.2%	51.2	CONTROLLED LAB PENDING
Check Point	85.9	58.2%	50.6	CONTROLLED LAB PENDING
Fortinet	85.0	58.2%	49.9	CONTROLLED LAB PENDING
Cisco	80.8	58.2%	47.5	CONTROLLED LAB PENDING

EVIDENCE SCALE

E0 / 0.00
Unsupported or unverified

E1 / 0.38
Vendor assertion or marketing statement

E2 / 0.64
Official documentation, release notes, or source repository

E3 / 0.78
Controlled Mythos laboratory result

E4 / 0.91
Production evidence from the adopting environment

E5 / 1.00
Independent repeatable validation

CURRENT CONFIDENCE BOUNDARY

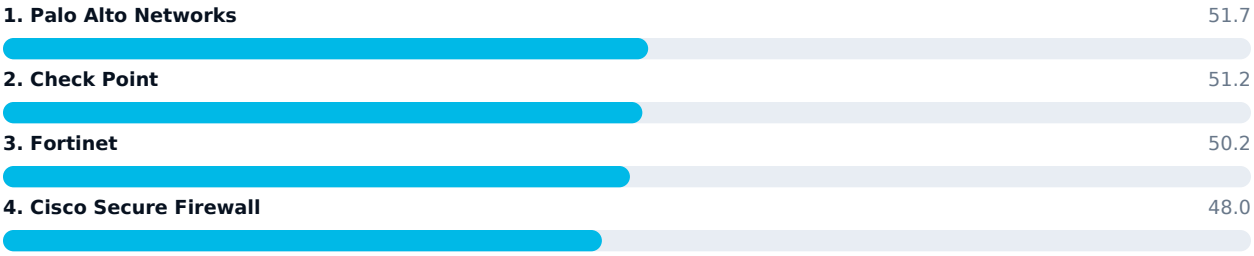
Most candidate criteria are supported at E2: official documentation or source evidence. Environment-specific laboratory, production, and independent evidence remain future gates.

PROFILE-SPECIFIC RANKINGS

EVIDENCE-ADJUSTED SENSITIVITY ANALYSIS

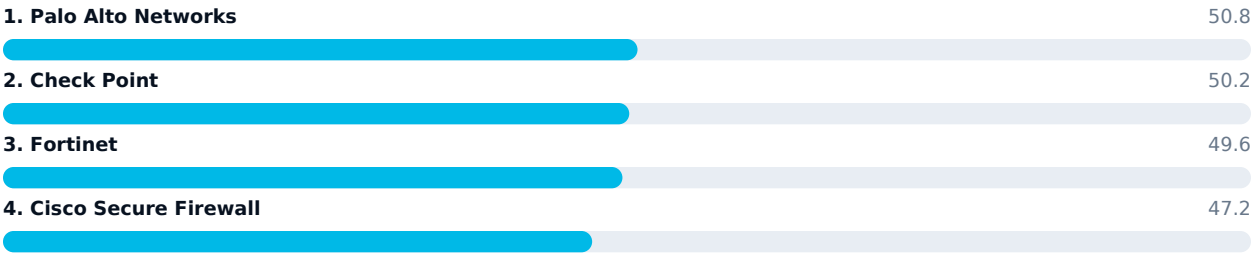
HIGH-ASSURANCE ENTERPRISE SECURITY

Prioritizes inspection depth, policy governance, segmentation, audit, resilience, and controlled change.



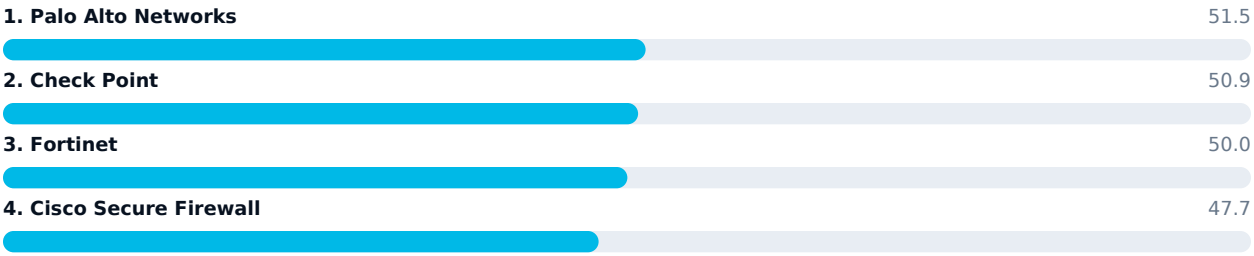
DISTRIBUTED BRANCH AND CAMPUS

Prioritizes appliance breadth, centralized operations, SD-WAN adjacency, economics, and remote lifecycle.



MULTI-DOMAIN SERVICE PROVIDER

Prioritizes delegation, tenancy, scale, policy isolation, automation, evidence, and lifecycle governance.



RANK SENSITIVITY

Small score differences are not decision certainty. Treat near-ties as a requirement for deeper PoC, commercial, migration, and operating-model evidence.

CANDIDATE DEEP DIVE / 01

PALO ALTO NETWORKS

OPERATING MODEL

Application- and identity-aware NGFW ecosystem with Panorama and Strata Cloud Manager management paths.

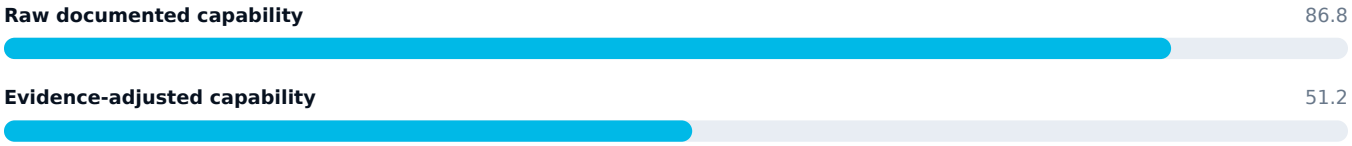
DOCUMENTED STRENGTHS

- Rich application, threat-prevention, decryption, and policy capabilities.
- Strong centralized management and cloud-delivered evolution.
- Broad automation, telemetry, and ecosystem integration.
- Mature enterprise operating patterns.

CAUTIONS AND PROOF OBLIGATIONS

- Commercial complexity and feature licensing require careful modeling.
- Panorama and cloud-management trajectories require explicit architecture decisions.
- Policy quality still depends on disciplined objects, ownership, and cleanup.

BASELINE SCORE PROFILE



GATE DISPOSITION

PASS - high-assurance enterprise profile | CONDITIONAL - commercial and management-path validation

CANDIDATE DEEP DIVE / 02

CHECK POINT

OPERATING MODEL

Management-centric security architecture with SmartConsole, Security Management, MDSM, gateways, and software blades.

DOCUMENTED STRENGTHS

- Deep centralized policy and object management.
- Strong multi-domain and delegated-management patterns.
- Mature clustering, logging, and enterprise control-plane architecture.
- Granular rulebase and security-blade integration.

CAUTIONS AND PROOF OBLIGATIONS

- Operational expertise is specialized and upgrades require careful planning.
- Architecture can become complex across domains, gateways, and management layers.
- Automation and cloud-management fit must be validated for each workflow.

BASELINE SCORE PROFILE

Raw documented capability 85.9

Evidence-adjusted capability 50.6

GATE DISPOSITION

PASS - management-centric and multi-domain profile | CONDITIONAL - specialized operational capability required

CANDIDATE DEEP DIVE / 03

FORTINET

OPERATING MODEL

Security Fabric ecosystem with FortiGate gateways and FortiManager-centered centralized operations.

DOCUMENTED STRENGTHS

- Broad appliance and platform coverage.
- Strong price-performance potential and integrated Security Fabric.
- Central management and automation options across the estate.
- Flexible branch, campus, datacenter, and service-provider deployment.

CAUTIONS AND PROOF OBLIGATIONS

- Feature and licensing boundaries require precise validation.
- Cross-product integration can increase ecosystem coupling.
- Operational consistency depends on firmware and management discipline.

BASELINE SCORE PROFILE

Raw documented capability 85.0

Evidence-adjusted capability 49.9

GATE DISPOSITION

PASS - distributed and value-sensitive profile | CONDITIONAL - firmware and ecosystem coupling controls

CANDIDATE DEEP DIVE / 04

CISCO SECURE FIREWALL

OPERATING MODEL

Secure Firewall Threat Defense managed through Firewall Management Center and integrated with the Cisco security ecosystem.

DOCUMENTED STRENGTHS

- Strong fit for Cisco-heavy enterprises and integrated security telemetry.
- Central management, policy, health, and migration tooling.
- Broad enterprise support and hardware options.
- Potential consolidation with existing Cisco identity and security investments.

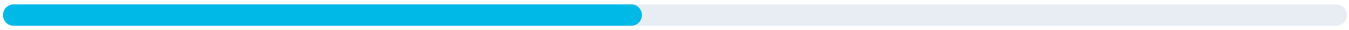
CAUTIONS AND PROOF OBLIGATIONS

- Operational experience and feature behavior must be validated carefully.
- Migration from ASA or other platforms can expose policy translation gaps.
- Management and release complexity require strict lab qualification.

BASELINE SCORE PROFILE

Raw documented capability 80.8

Evidence-adjusted capability 47.5



GATE DISPOSITION

CONDITIONAL - Cisco-aligned estate | CONTROLLED POC REQUIRED - migration and operations

ARCHITECTURE AND INTEGRATION FIT

THE PRODUCT IS ONLY ONE PART OF THE OPERATING SYSTEM

REFERENCE OPERATING LAYERS

01 Central management and policy authority

02 Gateway and inspection enforcement plane

03 Identity, endpoint, cloud, and threat-intelligence integrations

04 Logging, analytics, and evidence plane

05 Change, upgrade, backup, and recovery control system

INTEGRATION BOUNDARIES

- Identity and NAC
- SIEM, SOAR, and data lake
- Endpoint and cloud security
- DNS, DHCP, IPAM, and certificate services
- ITSM, Git, CI/CD, and automation
- SASE and remote-access services

ARCHITECTURE GATE

Every external dependency requires an owner, identity boundary, failure mode, evidence path, version policy, recovery procedure, and exit strategy.

SECURITY, OPERATIONS, LIFECYCLE, ECONOMICS, AND EXIT

CROSS-DOMAIN DECISION RECORD

SECURITY AND AUTHORITY

Identity, credentials, secrets, roles, privileged actions, signing, approvals, isolation, audit, and incident response must be tested as an integrated system.

RELIABILITY AND RECOVERY

Control-plane, data-plane, dependency, region, database, queue, network, and operator failures require defined degraded modes and recovery objectives.

CHANGE AND LIFECYCLE

Version, compatibility, migration, canary, rollback, content, plugin, provider, firmware, and decommission procedures are part of product fitness.

ECONOMICS AND CAPACITY

Licenses, metering, data, compute, hardware, storage, support, staffing, training, integration, migration, and opportunity cost require sensitivity analysis.

SOVEREIGNTY AND PRIVACY

Data location, support access, telemetry, subprocessors, encryption, key control, disconnected operation, and legal obligations must align with policy.

PORTABILITY AND EXIT

Configuration, policy, data, state, evidence, workflows, identities, and operational knowledge must be exportable and reconstructable.

DECISION OWNERSHIP

Architecture, security, operations, finance, procurement, legal, data, and service owners jointly accept the final profile, evidence, residual risk, and exit obligations.

RISK AND FAILURE REGISTER

SCENARIOS THAT CAN INVALIDATE A FEATURE-BASED SELECTION

ID	SEVERITY	FAILURE SCENARIO	REQUIRED TREATMENT
R-01	CRITICAL	Feature scores conceal mandatory policy or recovery failures.	PREVENT / DETECT / RECOVER / EVIDENCE
R-02	HIGH	Licensing assumptions make the preferred architecture economically invalid.	PREVENT / DETECT / RECOVER / EVIDENCE
R-03	HIGH	Management failure or compromise creates estate-wide exposure.	PREVENT / DETECT / RECOVER / EVIDENCE
R-04	HIGH	Decryption deployment causes application outage or privacy violations.	PREVENT / DETECT / RECOVER / EVIDENCE
R-05	MEDIUM	Upgrade paths are accepted without representative HA, VPN, routing, and policy tests.	PREVENT / DETECT / RECOVER / EVIDENCE
R-06	MEDIUM	Migration tools translate syntax but not intent, exceptions, or operational knowledge.	PREVENT / DETECT / RECOVER / EVIDENCE

RISK RULE

A candidate with an unresolved critical failure path cannot advance because optional capability, market position, or commercial discount cannot compensate for missing safety or recovery evidence.

CONTROLLED PROOF-OF-CAPABILITY PLAN

REPEATABLE SCENARIOS, INSTRUMENTATION, AND ACCEPTANCE

TEST 01

Import a representative policy estate and compare rule intent, objects, NAT, VPN, routing, and segmentation.

TEST 02

Exercise high availability, management loss, log loss, backup, restore, and rollback.

TEST 03

Measure policy installation, session scale, inspection throughput, and logging under representative traffic.

TEST 04

Validate API-driven object and policy lifecycle with approval and evidence.

TEST 05

Test decryption, certificate lifecycle, exception handling, and privacy controls.

TEST 06

Conduct controlled migration with semantic review and post-cutover validation.

EVIDENCE PACKAGE

Preserve versions, architecture, configuration, data set, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

CONDITIONAL RECOMMENDATION AND REVALIDATION

DECISION RECORD, ACCEPTANCE, AND NEXT EVIDENCE

RECOMMENDATION POSITION

- Use profile-specific shortlists rather than a universal rank.
- Require a production-representative firewall lab and migration rehearsal.
- Model five-year license, support, staffing, logging, and hardware costs.
- Revalidate quarterly and before any major release or management-plane transition.

CANDIDATE	ADJ. SCORE	GATE POSITION	LAB STATE
Palo Alto	51.2	PASS - high-assurance enterprise profile; CONDITIONAL - commercial and management-path validation	PENDING
Check Point	50.6	PASS - management-centric and multi-domain profile; CONDITIONAL - specialized operational capability required	PENDING
Fortinet	49.9	PASS - distributed and value-sensitive profile; CONDITIONAL - firmware and ecosystem coupling controls	PENDING
Cisco	47.5	CONDITIONAL - Cisco-aligned estate; CONTROLLED POC REQUIRED - migration and operations	PENDING

REVALIDATION SCHEDULE

Review no later than 2026-10-24. Review earlier after a major release, commercial change, critical vulnerability, material outage, architecture transition, failed acceptance test, or change to the target workload profile.

SOURCE AUTHORITY AND REVISION

CURRENT EVIDENCE SNAPSHOT AND PUBLICATION HISTORY

SOURCE ID	PUBLISHER	TITLE	CHECKED
NIST-CSF-20	NIST	Cybersecurity Framework 2.0	2026-07-24
NIST-SP800-53	NIST	Security and Privacy Controls for Information Systems and Organizations	2026-07-24
CIS-CONTROLS	Center for Internet Security	CIS Critical Security Controls v8.1	2026-07-24
PAN-SCM	Palo Alto Networks	Strata Cloud Manager Documentation	2026-07-24
PAN-PANORAMA	Palo Alto Networks	Panorama Administrator Documentation	2026-07-24
CP-R82-MGMT	Check Point	R82 Security Management Administration Guide	2026-07-24
CP-R82-MDSM	Check Point	R82 Multi-Domain Security Management Guide	2026-07-24
FORTIMANAGER	Fortinet	FortiManager 8.0 Documentation	2026-07-24
CISCO-FMC	Cisco	Cisco Secure Firewall Management Center Administration Guide	2026-07-24

SOURCE POSITION

Official documentation and source repositories support the current capability model. Source records preserve publisher, title, URL, purpose, and review date in the machine-readable authority register. Commercial terms, performance, and environment-specific behavior remain unverified until controlled proof.

REVISION HISTORY

1.1.0-candidate / 2026-07-24 - permanent-publication rebuild using the final benchmark system, profile-specific rankings, evidence adjustment, mandatory gates, and controlled PoC requirements.

COMPARATIVE EVALUATION / EVIDENCE REFRESH

CMP-003

Enterprise SIEM Platforms

Security analytics, detection, investigation and response

CURRENT EVIDENCE SNAPSHOT

Official product sources refreshed through 2026-09-17

Controlled Mythos laboratory rerun: NOT ASSERTED

BENCHMARK DOCTRINE

Gates before scores. Evidence before certainty. Profile fit before universal ranking.

VERSION 1.2.0-candidate

CANDIDATE COMPARATIVE EVALUATION / PUBLIC

Enterprise SIEM Platforms

Security analytics, detection, investigation and response

DOCUMENT ID
CMP-003
VERSION
1.2.0-candidate
STATUS
Candidate Comparative Evaluation
PUBLICATION DATE
2026-09-17
SCHEDULED REVIEW
2026-12-16
CANONICAL ROUTE
/library/evaluations/cmp-003/

4 CANDIDATES

18 ATOMIC CRITERIA

9 MANDATORY GATES

3 WORKLOAD PROFILES

E2 CURRENT MAX EVIDENCE

CURRENT DECISION BOUNDARY

The July 0.1-point Sentinel/Splunk difference is not decision certainty. Current data-lake, graph, AI, SOAR, and endpoint changes are material. Reproduce ingestion, detection, investigation, response, retention, migration, and cost behavior before profile selection.

NO CURRENT UNIVERSAL WINNER IS PUBLISHED FROM THIS REFRESH.

July 24 baseline scores are preserved as lineage and sensitivity evidence, not silently reissued as September laboratory results.

Microsoft Sentinel

Cloud service / updates through 2026-08-24

REFRESH TRIGGER

Data lake, graph, MCP-assisted investigation, detections-as-code, and expanded UEBA materially change the platform boundary.

Splunk Enterprise Security

8.7.0 / 2026-09-02

REFRESH TRIGGER

8.7.0 adds AI SOC Analyst and synchronized CIM updates; known issues and upgrade compatibility remain decision inputs.

Google Security Operations

SIEM 2026-09-15 / SOAR 6.3.100

REFRESH TRIGGER

Parser updates plus preview reaction triggers and case playbooks expand response automation and lifecycle considerations.

Elastic Security

9.5.4 current docs

REFRESH TRIGGER

Current 9.x line continues rapid detection, endpoint, AI-assistant, and migration changes; exact fit depends on deployment model.

FINAL BENCHMARK SYSTEM / DECISION ARCHITECTURE

01 SCOPE

Freeze decision, candidates, versions and exclusions.

02 GATES

Test non-compensable trust, state, recovery, change and exit.

03 SCORES

Measure atomic capability with approved profile weights.

04 CONFIDENCE

Discount by source grade, freshness, repeatability and relevance.

05 PROOF

Run normal, failure, migration, recovery and operator scenarios.

06 DECISION

Record conditional fit, residual risk, owner and revalidation trigger.

EVIDENCE CONFIDENCE / MODIFIERS

E0	E1	E2	E3	E4	E5
0.00	0.38	0.64	0.78	0.91	1.00
UNSUPPORTED	VENDOR ASSERTION	OFFICIAL SOURCE	MYTHOS LAB	PRODUCTION	INDEPENDENT

THIS EDITION

Current official product and release sources have been refreshed to E2. No E3 controlled Mythos lab rerun is claimed. Therefore current profile rankings remain intentionally unissued until the proof plan is executed.

What changed since the July benchmark snapshot

Changes below are sourced from current official release documentation. They identify revalidation triggers; they are not translated into new numeric scores without controlled proof.

Microsoft Sentinel

Cloud service / updates through 2026-08-24

MATERIAL DELTA

Data lake, graph, MCP-assisted investigation, detections-as-code, and expanded UEBA materially change the platform boundary.

CRITERIA TOUCHED

C01 / C05 / C06 / C07 / C08 / C12 / C14

MODERATE REVALIDATION

Splunk Enterprise Security

8.7.0 / 2026-09-02

MATERIAL DELTA

8.7.0 adds AI SOC Analyst and synchronized CIM updates; known issues and upgrade compatibility remain decision inputs.

CRITERIA TOUCHED

C01 / C07 / C10 / C11 / C14

MODERATE REVALIDATION

Google Security Operations

SIEM 2026-09-15 / SOAR 6.3.100

MATERIAL DELTA

Parser updates plus preview reaction triggers and case playbooks expand response automation and lifecycle considerations.

CRITERIA TOUCHED

C01 / C05 / C06 / C07 / C11 / C14

MODERATE REVALIDATION

Elastic Security

9.5.4 current docs

MATERIAL DELTA

Current 9.x line continues rapid detection, endpoint, AI-assistant, and migration changes; exact fit depends on deployment model.

CRITERIA TOUCHED

C01 / C05 / C07 / C12 / C13 / C14

MODERATE REVALIDATION

Mandatory gates remain non-compensable

No current release is marked PASS solely from documentation. E2 means the official source supports the capability path; LAB and RETEST identify proof still required. HOLD blocks a current decision.

MANDATORY GATE	MICROSOFT SENTI	SPLUNK ENTERPRI	GOOGLE SECURITY	ELASTIC SECURIT
C01 Functional coverage	RETEST	RETEST	RETEST	RETEST
C02 Security / trust	E2	E2	E2	E2
C03 Governance / approval	E2	E2	E2	E2
C04 State integrity	E2	E2	E2	E2
C07 Observability / evidence	RETEST	RETEST	RETEST	RETEST
C09 Resilience / continuity	LAB	LAB	LAB	LAB
C11 Change safety / rollback	LAB	RETEST	RETEST	LAB
C16 Portability / exit	LAB	LAB	LAB	LAB
C18 Assurance confidence	HOLD	HOLD	HOLD	HOLD

LEGEND

E2 official-source support only

LAB controlled proof required

RETEST release delta touches this gate

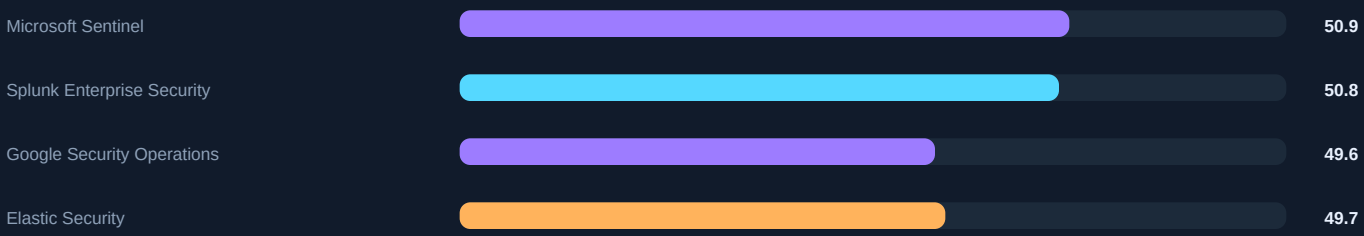
HOLD decision gate remains closed

Historical profile sensitivity - preserved, not reissued

Values below are the 2026-07-24 evidence-adjusted baseline. They show sensitivity to operating-model weights. The September refresh does not recompute rank because E3 lab proof has not been reproduced.

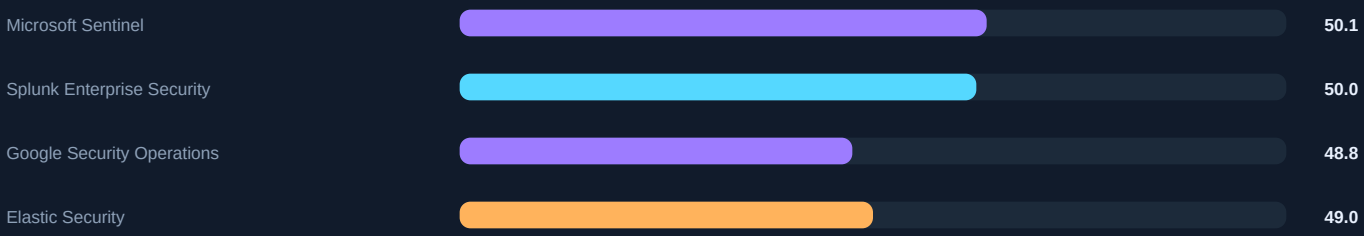
MICROSOFT-CENTRIC SOC

BASELINE SPREAD 1.3 PTS



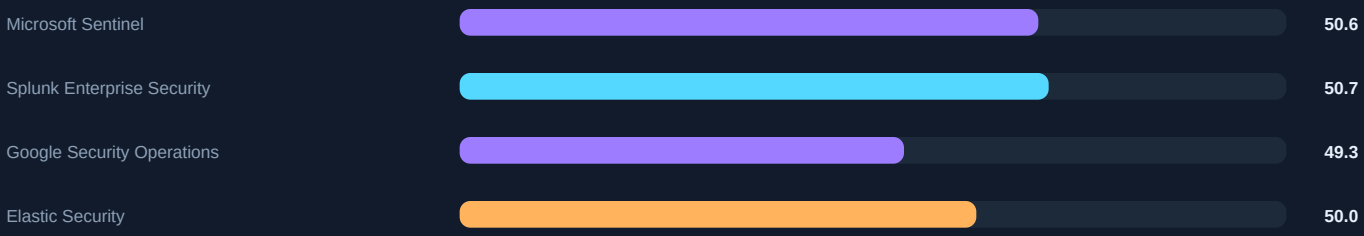
HIGH-SCALE SECURITY DATA PLATFORM

BASELINE SPREAD 1.3 PTS



SOVEREIGN AND EXTENSIBLE SOC

BASELINE SPREAD 1.4 PTS



CURRENT RANK STATUS: WITHHELD PENDING CONTROLLED PROOF.

Evidence confidence is separate from capability

E5 — Independent repeatable validation
confidence modifier 1.00

E4 — Production evidence in adopting environment
confidence modifier 0.91

E3 — Controlled Mythos laboratory result
confidence modifier 0.78

E2 — Official documentation / release / source
confidence modifier 0.64

E1 — Vendor assertion or marketing
confidence modifier 0.38

E0 — Unsupported or unverified
confidence modifier 0.00

CURRENT EVIDENCE STATE

Microsoft Sentinel	E2 REFRESHED
Historical adjusted confidence: 58.2%	
E3 LAB: NOT REPRODUCED	
Splunk Enterprise Security	E2 REFRESHED
Historical adjusted confidence: 58.2%	
E3 LAB: NOT REPRODUCED	
Google Security Operations	E2 REFRESHED
Historical adjusted confidence: 58.2%	
E3 LAB: NOT REPRODUCED	
Elastic Security	E2 REFRESHED
Historical adjusted confidence: 58.2%	
E3 LAB: NOT REPRODUCED	

ASSURANCE RULE

A current release note can change capability context, but it cannot raise assurance beyond E2. Current recommendation strength is therefore bounded until the test plan produces reproducible artifacts.

Lifecycle risk is evaluated independently of features

The benchmark models migration, upgrade, compatibility, support, staffing, data/state export, downtime and exit. Current release changes below are explicit proof triggers.

Microsoft Sentinel

MODERATE REVALIDATION

Data lake, graph, MCP-assisted investigation, detections-as-code, and expanded UEBA materially change the platform boundary.

RETEST: C01 / C05 / C06 / C07 / C08 / C12 / C14

Splunk Enterprise Security

MODERATE REVALIDATION

8.7.0 adds AI SOC Analyst and synchronized CIM updates; known issues and upgrade compatibility remain decision inputs.

RETEST: C01 / C07 / C10 / C11 / C14

Google Security Operations

MODERATE REVALIDATION

Parser updates plus preview reaction triggers and case playbooks expand response automation and lifecycle considerations.

RETEST: C01 / C05 / C06 / C07 / C11 / C14

Elastic Security

MODERATE REVALIDATION

Current 9.x line continues rapid detection, endpoint, AI-assistant, and migration changes; exact fit depends on deployment model.

RETEST: C01 / C05 / C07 / C12 / C13 / C14

IRREVERSIBLE COMMITMENT GATE

Before migration or procurement lock-in, prove rollback, state portability, operational reconstruction, and supplier-exit assumptions.

Controlled proof is the next decision-producing step

REPRODUCTION STATE / NOT EXECUTED IN THIS EVIDENCE-REFRESH RELEASE

01

CONTROLLED SCENARIO

Replay a representative 30-day telemetry corpus with known attacks and operational noise.

02

CONTROLLED SCENARIO

Measure ingest latency, normalization quality, query latency, detection delay, and analyst workload.

03

CONTROLLED SCENARIO

Implement equivalent detections, risk scoring, cases, and response playbooks.

04

CONTROLLED SCENARIO

Model hot, warm, archive, replay, and legal-hold economics.

05

CONTROLLED SCENARIO

Exercise regional loss, collector outage, schema change, and source backpressure.

06

CONTROLLED SCENARIO

Export detections, cases, evidence, and normalized data to prove portability.

EVIDENCE PACKAGE

Preserve exact versions, architecture, configuration, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

Current decision posture

The July 0.1-point Sentinel/Splunk difference is not decision certainty. Current data-lake, graph, AI, SOAR, and endpoint changes are material. Reproduce ingestion, detection, investigation, response, retention, migration, and cost behavior before profile selection.

CURRENT RANKING: NOT REISSUED / CONTROLLED LAB REQUIRED

CURRENT OFFICIAL-SOURCE REGISTER / CHECKED 2026-09-17

Microsoft Sentinel	E2 / OFFICIAL	
Microsoft Sentinel - What is new		learn.microsoft.com
Splunk Enterprise Security	E2 / OFFICIAL	
Splunk Enterprise Security 8.7 release notes		help.splunk.com
Google Security Operations	E2 / OFFICIAL	
Google SecOps SIEM/SOAR release notes		docs.cloud.google.com
Elastic Security	E2 / OFFICIAL	
Elastic Security release notes		www.elastic.co

REVISION LINEAGE

1.2.0-candidate / 2026-09-17 - current-source evidence refresh; no lab rescore issued.

1.1.0-candidate / 2026-07-24 - baseline benchmark using the final comparative evaluation system.

Trigger earlier review after major release, acquisition, licensing change, critical vulnerability, material outage, failed PoC, or workload change.

BASELINE ANALYTIC RECORD CONTINUES ON PAGE 11 FOR TRACEABILITY.

ATOMIC CRITERIA MODEL

WEIGHTED CAPABILITY WITH EXPLICIT MINIMUM EVIDENCE

ID	CRITERION	WEIGHT	GATE	MINIMUM EVIDENCE
C01	SIEM, detection, investigation, hunting, and response coverage	5	YES	Feature documentation, APIs, configuration exports, and scenario demonstration.
C02	Security and trust architecture	5	YES	Threat model, identity flows, RBAC tests, audit records, and security configuration.
C03	Governance and approval controls	5	YES	Approval workflow, policy controls, separation-of-duties tests, and decision records.
C04	Telemetry, schema, detection, and case-state integrity	5	YES	Backup, restore, rollback, drift, and corruption-recovery evidence.
C05	Automation and API quality	4	NO	API specifications, authentication tests, rate limits, event samples, and automation runs.
C06	Integration ecosystem	4	NO	Supported integrations, connector tests, failure behavior, and lifecycle ownership.
C07	Observability and evidence	4	YES	Logs, traces, metrics, reports, export tests, and evidence-retention controls.
C08	Ingestion, retention, query, detection, and case scale	4	NO	Controlled load tests, topology scale, saturation behavior, and capacity model.
C09	Resilience and continuity	5	YES	Failure injection, HA tests, recovery timing, degraded-mode operation, and runbooks.
C10	Usability and operator cognition	3	NO	Task observation, error-rate measures, navigation tests, and operator interviews.
C11	Change safety and rollback	5	YES	Plan or preview output, canary tests, rollback execution, and post-change verification.
C12	Extensibility and programmability	3	NO	Plugin, module, provider, workflow, or code-extension tests and upgrade impact analysis.
C13	Deployment and sovereignty options	3	NO	Deployment architecture, data-flow mapping, residency evidence, and offline tests.
C14	Lifecycle and upgrade discipline	4	NO	Release notes, upgrade test, compatibility matrix, support policy, and migration plan.
C15	Ingestion, retention, compute, staffing, and response economics	3	NO	Bill-of-materials, licensing model, staffing estimates, metering, and sensitivity analysis.
C16	Portability and exit	4	YES	Export tests, format analysis, replacement workflow, and decommission evidence.
C17	Vendor and ecosystem risk	3	NO	License analysis, roadmap evidence, bus-factor indicators, and dependency inventory.
C18	Assurance confidence	5	YES	Source provenance, lab artifacts, production evidence, independent replication, and review date.

SCORE SCALE

0 absent; 1 materially deficient; 2 partial; 3 adequate with limits; 4 strong; 5 leading for the defined profile. Scores remain provisional until the required evidence grade is achieved.

RAW CAPABILITY SCORECARD

DOCUMENTED CAPABILITY BEFORE EVIDENCE DISCOUNT

CRITERION	SENTINEL	SPLUNK ES	GOOGLE SECOPS	ELASTIC
C01 SIEM, detection, investigation, hunting, and response coverage	4.6	4.7	4.5	4.4
C02 Security and trust architecture	4.5	4.4	4.4	4.3
C03 Governance and approval controls	4.3	4.2	4.2	4.0
C04 Telemetry, schema, detection, and case-state integrity	4.4	4.4	4.3	4.2
C05 Automation and API quality	4.5	4.6	4.4	4.6
C06 Integration ecosystem	4.7	4.7	4.4	4.3
C07 Observability and evidence	4.6	4.7	4.6	4.4
C08 Ingestion, retention, query, detection, and case scale	4.7	4.6	4.8	4.5
C09 Resilience and continuity	4.5	4.4	4.5	4.2
C10 Usability and operator cognition	4.4	4.5	4.3	4.2
C11 Change safety and rollback	4.4	4.2	4.3	4.1
C12 Extensibility and programmability	4.2	4.6	4.1	4.6
C13 Deployment and sovereignty options	3.8	4.2	3.7	4.6
C14 Lifecycle and upgrade discipline	4.3	4.0	4.0	3.9
C15 Ingestion, retention, compute, staffing, and response economics	3.6	3.0	3.5	4.0
C16 Portability and exit	3.9	4.0	3.7	4.4
C17 Vendor and ecosystem risk	4.0	3.9	3.8	4.1
C18 Assurance confidence	3.6	3.6	3.5	3.5

WEIGHTED BASELINE / 100

Sentinel	Splunk ES	Google SecOp	Elastic
86.0	85.5	83.9	84.4

EVIDENCE-ADJUSTED SCORECARD

CAPABILITY DISCOUNTED BY CURRENT EVIDENCE CONFIDENCE

CANDIDATE	RAW	EVIDENCE CONFIDENCE	ADJUSTED	CURRENT STATE
Sentinel	86.0	58.2%	50.6	CONTROLLED LAB PENDING
Splunk ES	85.5	58.2%	50.5	CONTROLLED LAB PENDING
Google SecOps	83.9	58.2%	49.2	CONTROLLED LAB PENDING
Elastic	84.4	58.2%	49.6	CONTROLLED LAB PENDING

EVIDENCE SCALE

E0 / 0.00
Unsupported or unverified

E1 / 0.38
Vendor assertion or marketing statement

E2 / 0.64
Official documentation, release notes, or source repository

E3 / 0.78
Controlled Mythos laboratory result

E4 / 0.91
Production evidence from the adopting environment

E5 / 1.00
Independent repeatable validation

CURRENT CONFIDENCE BOUNDARY

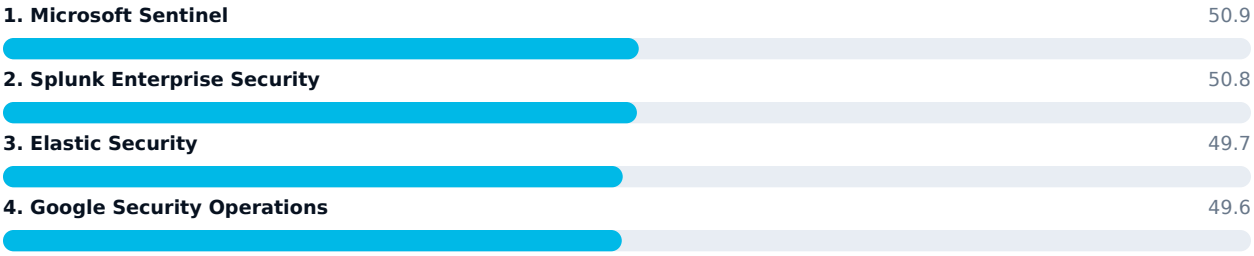
Most candidate criteria are supported at E2: official documentation or source evidence. Environment-specific laboratory, production, and independent evidence remain future gates.

PROFILE-SPECIFIC RANKINGS

EVIDENCE-ADJUSTED SENSITIVITY ANALYSIS

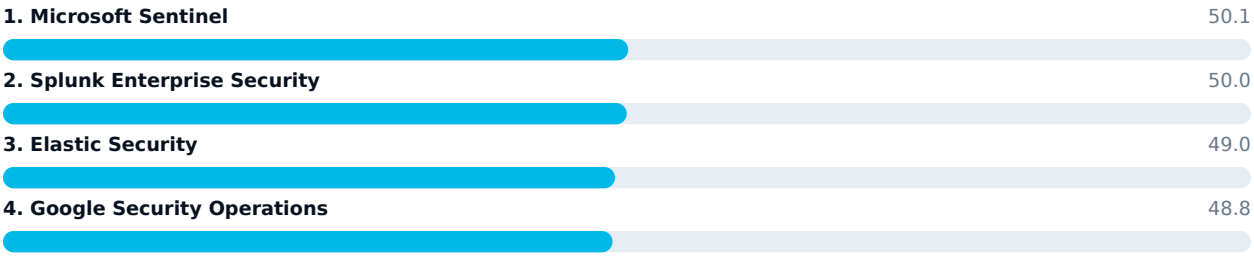
MICROSOFT-CENTRIC SOC

Prioritizes Microsoft identity, endpoint, cloud, collaboration, data-lake, and automation integration.



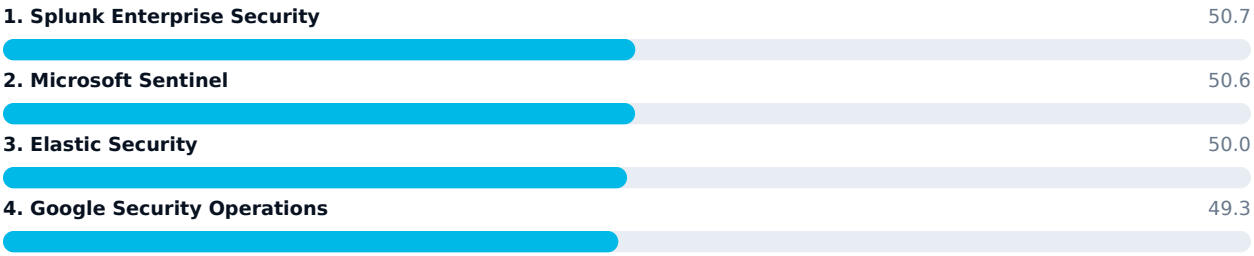
HIGH-SCALE SECURITY DATA PLATFORM

Prioritizes ingestion, retention, search, detection throughput, data architecture, and reliability.



SOVEREIGN AND EXTENSIBLE SOC

Prioritizes deployment control, data portability, open integration, customization, and exit.



RANK SENSITIVITY

Small score differences are not decision certainty. Treat near-ties as a requirement for deeper PoC, commercial, migration, and operating-model evidence.

CANDIDATE DEEP DIVE / 01

MICROSOFT SENTINEL

OPERATING MODEL

Cloud-native Microsoft security operations platform spanning SIEM, data lake, analytics, investigation, automation, and Defender integration.

DOCUMENTED STRENGTHS

- Strong Microsoft ecosystem integration.
- Cloud-native analytics and automation.
- Broad connector and content ecosystem.
- Unified trajectory through the Defender portal and security data lake.

CAUTIONS AND PROOF OBLIGATIONS

- Cost depends heavily on ingestion, retention, query, and architecture.
- Portal and platform transitions require operational planning.
- Non-Microsoft data and response depth must be validated per source.

BASELINE SCORE PROFILE

Raw documented capability 86.0

Evidence-adjusted capability 50.6

GATE DISPOSITION

PASS - Microsoft-centric and cloud-native profiles | CONDITIONAL - ingestion economics and portal transition

CANDIDATE DEEP DIVE / 02

SPLUNK ENTERPRISE SECURITY

OPERATING MODEL

Search- and analytics-centric enterprise SIEM operating on Splunk platform capabilities with risk-based alerting and broad ecosystem support.

DOCUMENTED STRENGTHS

- Powerful search, data exploration, and extensibility.
- Mature enterprise SIEM content and operations.
- Risk-based alerting supports entity-centered detection.
- Flexible deployment and rich ecosystem.

CAUTIONS AND PROOF OBLIGATIONS

- Cost and platform engineering can be substantial.
- Data-model and content discipline are prerequisites.
- Scale and search performance require experienced architecture and operations.

BASELINE SCORE PROFILE

Raw documented capability 85.5

Evidence-adjusted capability 50.5

GATE DISPOSITION

PASS - search-intensive mature SOC profile | CONDITIONAL - cost and platform engineering

CANDIDATE DEEP DIVE / 03

GOOGLE SECURITY OPERATIONS

OPERATING MODEL

Cloud-native SIEM and SOAR platform with high-scale normalization, threat intelligence, detection, investigation, and playbooks.

DOCUMENTED STRENGTHS

- Strong cloud-scale data processing and search.
- Integrated threat intelligence and security analytics.
- SIEM and SOAR convergence.
- High-volume security-data orientation.

CAUTIONS AND PROOF OBLIGATIONS

- Commercial, regional, integration, and migration details require validation.
- Analyst and content migration can be significant.
- Operating model may depend on Google Cloud adjacency and partner ecosystem.

BASELINE SCORE PROFILE

Raw documented capability 83.9

Evidence-adjusted capability 49.2

GATE DISPOSITION

PASS - high-scale cloud security-data profile | CONDITIONAL - integration and operating-model fit

CANDIDATE DEEP DIVE / 04

ELASTIC SECURITY

OPERATING MODEL

Unified SIEM, XDR, endpoint, and cloud-security solution built on Elasticsearch and Kibana with flexible deployment options.

DOCUMENTED STRENGTHS

- Strong search and analytics foundation.
- Flexible self-managed and cloud deployment.
- Integrated endpoint and cloud-security capabilities.
- Open ecosystem and extensibility.

CAUTIONS AND PROOF OBLIGATIONS

- Enterprise operations require Elastic engineering capability.
- Feature and subscription boundaries need validation.
- Content, tuning, and response maturity vary by operating model.

BASELINE SCORE PROFILE

Raw documented capability 84.4

Evidence-adjusted capability 49.6

GATE DISPOSITION

PASS - sovereign and extensible profile | CONDITIONAL - engineering and content maturity

ARCHITECTURE AND INTEGRATION FIT

THE PRODUCT IS ONLY ONE PART OF THE OPERATING SYSTEM

REFERENCE OPERATING LAYERS

01 Telemetry acquisition and routing

02 Normalization and security data platform

03 Detection, analytics, hunting, and intelligence

04 Case, investigation, automation, and response

05 Evidence retention, reporting, and assurance

INTEGRATION BOUNDARIES

- Identity, endpoint, cloud, network, application, and threat-intelligence sources
- SOAR and response tools
- ITSM and collaboration
- Data lake, archive, and eDiscovery
- Detection-as-code and CI/CD
- Asset, vulnerability, and exposure systems

ARCHITECTURE GATE

Every external dependency requires an owner, identity boundary, failure mode, evidence path, version policy, recovery procedure, and exit strategy.

SECURITY, OPERATIONS, LIFECYCLE, ECONOMICS, AND EXIT

CROSS-DOMAIN DECISION RECORD

SECURITY AND AUTHORITY

Identity, credentials, secrets, roles, privileged actions, signing, approvals, isolation, audit, and incident response must be tested as an integrated system.

RELIABILITY AND RECOVERY

Control-plane, data-plane, dependency, region, database, queue, network, and operator failures require defined degraded modes and recovery objectives.

CHANGE AND LIFECYCLE

Version, compatibility, migration, canary, rollback, content, plugin, provider, firmware, and decommission procedures are part of product fitness.

ECONOMICS AND CAPACITY

Licenses, metering, data, compute, hardware, storage, support, staffing, training, integration, migration, and opportunity cost require sensitivity analysis.

SOVEREIGNTY AND PRIVACY

Data location, support access, telemetry, subprocessors, encryption, key control, disconnected operation, and legal obligations must align with policy.

PORTABILITY AND EXIT

Configuration, policy, data, state, evidence, workflows, identities, and operational knowledge must be exportable and reconstructable.

DECISION OWNERSHIP

Architecture, security, operations, finance, procurement, legal, data, and service owners jointly accept the final profile, evidence, residual risk, and exit obligations.

RISK AND FAILURE REGISTER

SCENARIOS THAT CAN INVALIDATE A FEATURE-BASED SELECTION

ID	SEVERITY	FAILURE SCENARIO	REQUIRED TREATMENT
R-01	CRITICAL	The cheapest ingest design removes telemetry required for detection or investigation.	PREVENT / DETECT / RECOVER / EVIDENCE
R-02	HIGH	A high query benchmark conceals weak schema, detection, and case workflows.	PREVENT / DETECT / RECOVER / EVIDENCE
R-03	HIGH	Automation executes destructive response without adequate approval and rollback.	PREVENT / DETECT / RECOVER / EVIDENCE
R-04	HIGH	Data residency or retention assumptions violate legal or contractual obligations.	PREVENT / DETECT / RECOVER / EVIDENCE
R-05	MEDIUM	Detection content is imported without validation against local log semantics.	PREVENT / DETECT / RECOVER / EVIDENCE
R-06	MEDIUM	Migration loses searches, cases, data models, risk history, and analyst knowledge.	PREVENT / DETECT / RECOVER / EVIDENCE

RISK RULE

A candidate with an unresolved critical failure path cannot advance because optional capability, market position, or commercial discount cannot compensate for missing safety or recovery evidence.

CONTROLLED PROOF-OF-CAPABILITY PLAN

REPEATABLE SCENARIOS, INSTRUMENTATION, AND ACCEPTANCE

TEST 01

Replay a representative 30-day telemetry corpus with known attacks and operational noise.

TEST 02

Measure ingest latency, normalization quality, query latency, detection delay, and analyst workload.

TEST 03

Implement equivalent detections, risk scoring, cases, and response playbooks.

TEST 04

Model hot, warm, archive, replay, and legal-hold economics.

TEST 05

Exercise regional loss, collector outage, schema change, and source backpressure.

TEST 06

Export detections, cases, evidence, and normalized data to prove portability.

EVIDENCE PACKAGE

Preserve versions, architecture, configuration, data set, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

CONDITIONAL RECOMMENDATION AND REVALIDATION

DECISION RECORD, ACCEPTANCE, AND NEXT EVIDENCE

RECOMMENDATION POSITION

- Select by workload profile, data architecture, and analyst operating model.
- Use a controlled attack-and-noise replay rather than vendor demos.
- Model full data, compute, content, staffing, and migration economics.
- Revalidate detection quality and cost monthly during pilot and quarterly after adoption.

CANDIDATE	ADJ. SCORE	GATE POSITION	LAB STATE
Sentinel	50.6	PASS - Microsoft-centric and cloud-native profiles; CONDITIONAL - ingestion economics and portal transition	PENDING
Splunk ES	50.5	PASS - search-intensive mature SOC profile; CONDITIONAL - cost and platform engineering	PENDING
Google SecOps	49.2	PASS - high-scale cloud security-data profile; CONDITIONAL - integration and operating-model fit	PENDING
Elastic	49.6	PASS - sovereign and extensible profile; CONDITIONAL - engineering and content maturity	PENDING

REVALIDATION SCHEDULE

Review no later than 2026-10-24. Review earlier after a major release, commercial change, critical vulnerability, material outage, architecture transition, failed acceptance test, or change to the target workload profile.

SOURCE AUTHORITY AND REVISION

CURRENT EVIDENCE SNAPSHOT AND PUBLICATION HISTORY

SOURCE ID	PUBLISHER	TITLE	CHECKED
NIST-CSF-20	NIST	Cybersecurity Framework 2.0	2026-07-24
NIST-SP800-53	NIST	Security and Privacy Controls for Information Systems and Organizations	2026-07-24
CIS-CONTROLS	Center for Internet Security	CIS Critical Security Controls v8.1	2026-07-24
SENTINEL	Microsoft	Microsoft Sentinel Overview	2026-07-24
SENTINEL-AUTO	Microsoft	Microsoft Sentinel Automation Rules	2026-07-24
SPLUNK-ES	Splunk	Splunk Enterprise Security Documentation	2026-07-24
SPLUNK-RBA	Splunk	Splunk Risk-Based Alerting	2026-07-24
GOOGLE-SECOPS	Google Cloud	Google Security Operations Overview	2026-07-24
GOOGLE-SOAR	Google Cloud	Google Security Operations SOAR Overview	2026-07-24
ELASTIC-SEC	Elastic	Elastic Security Documentation	2026-07-24

SOURCE POSITION

Official documentation and source repositories support the current capability model. Source records preserve publisher, title, URL, purpose, and review date in the machine-readable authority register. Commercial terms, performance, and environment-specific behavior remain unverified until controlled proof.

REVISION HISTORY

1.1.0-candidate / 2026-07-24 - permanent-publication rebuild using the final benchmark system, profile-specific rankings, evidence adjustment, mandatory gates, and controlled PoC requirements.

COMPARATIVE EVALUATION / EVIDENCE REFRESH

CMP-004

Agentic Frameworks

Agent runtimes, orchestration and governed infrastructure automation

CURRENT EVIDENCE SNAPSHOT

Official product sources refreshed through 2026-09-17

Controlled Mythos laboratory rerun: NOT ASSERTED

VERSION 1.2.0-candidate

CANDIDATE COMPARATIVE EVALUATION / PUBLIC

BENCHMARK DOCTRINE

Gates before scores. Evidence before certainty. Profile fit before universal ranking.

Agentic Frameworks

Agent runtimes, orchestration and governed infrastructure automation

DOCUMENT ID
CMP-004
VERSION
1.2.0-candidate
STATUS
Candidate Comparative Evaluation
PUBLICATION DATE
2026-09-17
SCHEDULED REVIEW
2026-12-16
CANONICAL ROUTE
/library/evaluations/cmp-004/

3
CANDIDATES

18
ATOMIC CRITERIA

9
MANDATORY GATES

3
WORKLOAD PROFILES

E2
CURRENT MAX EVIDENCE

CURRENT DECISION BOUNDARY

Reopen the comparison. AgentCore and Itential changed materially after the July baseline while LangGraph continued to evolve. No inherited ranking should be promoted to a current recommendation until durable execution, authority, observability, replay, migration, and failure tests are rerun.

NO CURRENT UNIVERSAL WINNER IS PUBLISHED FROM THIS REFRESH.

July 24 baseline scores are preserved as lineage and sensitivity evidence, not silently reissued as September laboratory results.

LangGraph

Python 1.2.11 / JS 1.4.15

REFRESH TRIGGER

Durable graph runtime and checkpoint ecosystem continue to evolve across Python and JavaScript lines.

Amazon Bedrock AgentCore

Managed service / Sept 2026 updates

REFRESH TRIGGER

Batch Evaluations and A/B Testing are now GA, materially strengthening the observe-evaluate-improve proof loop.

Itential Automation Platform

Platform 6.5 / Cloud Jul-Aug 2026

REFRESH TRIGGER

FlowAI, Work Center, asynchronous workflow support, and auditable agent sessions expand governed infrastructure-agent operation.

FINAL BENCHMARK SYSTEM / DECISION ARCHITECTURE

01 SCOPE

Freeze decision, candidates, versions and exclusions.

02 GATES

Test non-compensable trust, state, recovery, change and exit.

03 SCORES

Measure atomic capability with approved profile weights.

04 CONFIDENCE

Discount by source grade, freshness, repeatability and relevance.

05 PROOF

Run normal, failure, migration, recovery and operator scenarios.

06 DECISION

Record conditional fit, residual risk, owner and revalidation trigger.

EVIDENCE CONFIDENCE / MODIFIERS

E0	E1	E2	E3	E4	E5
0.00	0.38	0.64	0.78	0.91	1.00
UNSUPPORTED	VENDOR ASSERTION	OFFICIAL SOURCE	MYTHOS LAB	PRODUCTION	INDEPENDENT

THIS EDITION

Current official product and release sources have been refreshed to E2. No E3 controlled Mythos lab rerun is claimed. Therefore current profile rankings remain intentionally unissued until the proof plan is executed.

What changed since the July benchmark snapshot

Changes below are sourced from current official release documentation. They identify revalidation triggers; they are not translated into new numeric scores without controlled proof.

LangGraph

Python 1.2.11 / JS 1.4.15

MATERIAL DELTA

Durable graph runtime and checkpoint ecosystem continue to evolve across Python and JavaScript lines.

CRITERIA TOUCHED

C01 / C04 / C07 / C11 / C12 / C14

MODERATE REVALIDATION

Amazon Bedrock AgentCore

Managed service / Sept 2026 updates

MATERIAL DELTA

Batch Evaluations and A/B Testing are now GA, materially strengthening the observe-evaluate-improve proof loop.

CRITERIA TOUCHED

C01 / C03 / C07 / C08 / C11 / C18

MATERIAL REVALIDATION

Itential Automation Platform

Platform 6.5 / Cloud Jul-Aug 2026

MATERIAL DELTA

FlowAI, Work Center, asynchronous workflow support, and auditable agent sessions expand governed infrastructure-agent operation.

CRITERIA TOUCHED

C01 / C03 / C05 / C06 / C07 / C11

MATERIAL REVALIDATION

Mandatory gates remain non-compensable

No current release is marked PASS solely from documentation. E2 means the official source supports the capability path; LAB and RETEST identify proof still required. HOLD blocks a current decision.

MANDATORY GATE	LANGGRAPH	AMAZON BEDROCK	ITENTIAL AUTOMA
C01 Functional coverage	RETEST	RETEST	RETEST
C02 Security / trust	E2	E2	E2
C03 Governance / approval	E2	RETEST	RETEST
C04 State integrity	RETEST	E2	E2
C07 Observability / evidence	RETEST	RETEST	RETEST
C09 Resilience / continuity	LAB	LAB	LAB
C11 Change safety / rollback	RETEST	RETEST	RETEST
C16 Portability / exit	LAB	LAB	LAB
C18 Assurance confidence	HOLD	HOLD	HOLD

LEGEND

E2 official-source support only

LAB controlled proof required

RETEST release delta touches this gate

HOLD decision gate remains closed

Historical profile sensitivity - preserved, not reissued

Values below are the 2026-07-24 evidence-adjusted baseline. They show sensitivity to operating-model weights. The September refresh does not recompute rank because E3 lab proof has not been reproduced.

CUSTOM GOVERNED AGENT RUNTIME

BASELINE SPREAD 2.4 PTS

MANAGED CLOUD AGENT OPERATIONS

BASELINE SPREAD 1.5 PTS

INFRASTRUCTURE AUTOMATION CONTROL PLANE

BASELINE SPREAD 1.6 PTS

CURRENT RANK STATUS: WITHHELD PENDING CONTROLLED PROOF.

Evidence confidence is separate from capability

E5 — Independent repeatable validation
confidence modifier 1.00

E4 — Production evidence in adopting environment
confidence modifier 0.91

E3 — Controlled Mythos laboratory result
confidence modifier 0.78

E2 — Official documentation / release / source
confidence modifier 0.64

E1 — Vendor assertion or marketing
confidence modifier 0.38

E0 — Unsupported or unverified
confidence modifier 0.00

CURRENT EVIDENCE STATE

LangGraph — E2 REFRESHED
Historical adjusted confidence: 59.7%
E3 LAB: NOT REPRODUCED

Amazon Bedrock AgentCore — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

Itential Automation Platform — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

ASSURANCE RULE

A current release note can change capability context, but it cannot raise assurance beyond E2. Current recommendation strength is therefore bounded until the test plan produces reproducible artifacts.

Lifecycle risk is evaluated independently of features

The benchmark models migration, upgrade, compatibility, support, staffing, data/state export, downtime and exit. Current release changes below are explicit proof triggers.

LangGraph

MODERATE REVALIDATION

Durable graph runtime and checkpoint ecosystem continue to evolve across Python and JavaScript lines.

RETEST: C01 / C04 / C07 / C11 / C12 / C14

Amazon Bedrock AgentCore

MATERIAL REVALIDATION

Batch Evaluations and A/B Testing are now GA, materially strengthening the observe-evaluate-improve proof loop.

RETEST: C01 / C03 / C07 / C08 / C11 / C18

Itential Automation Platform

MATERIAL REVALIDATION

FlowAI, Work Center, asynchronous workflow support, and auditable agent sessions expand governed infrastructure-agent operation.

RETEST: C01 / C03 / C05 / C06 / C07 / C11

IRREVERSIBLE COMMITMENT GATE

Before migration or procurement lock-in, prove rollback, state portability, operational reconstruction, and supplier-exit assumptions.

Controlled proof is the next decision-producing step

REPRODUCTION STATE / NOT EXECUTED IN THIS EVIDENCE-REFRESH RELEASE

01

CONTROLLED SCENARIO

Run a long-lived mission with branching, retry, interruption, approval, restart, replay, and compensation.

02

CONTROLLED SCENARIO

Integrate local and frontier models, MCP tools, secrets, identity, and policy.

03

CONTROLLED SCENARIO

Inject model timeout, malformed tool output, duplicate event, lost checkpoint, and operator cancellation.

04

CONTROLLED SCENARIO

Measure task throughput, state growth, token use, tool latency, and evidence completeness.

05

CONTROLLED SCENARIO

Export mission state, traces, memory, policies, and tool definitions.

06

CONTROLLED SCENARIO

Red-team prompt injection, confused-deputy behavior, excess authority, and unsafe recovery.

EVIDENCE PACKAGE

Preserve exact versions, architecture, configuration, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

Current decision posture

Reopen the comparison. AgentCore and Itential changed materially after the July baseline while LangGraph continued to evolve. No inherited ranking should be promoted to a current recommendation until durable execution, authority, observability, replay, migration, and failure tests are rerun.

CURRENT RANKING: NOT REISSUED / CONTROLLED LAB REQUIRED

CURRENT OFFICIAL-SOURCE REGISTER / CHECKED 2026-09-17

LangGraph	E2 / OFFICIAL	
LangGraph GitHub releases		github.com
Amazon Bedrock AgentCore	E2 / OFFICIAL	
Amazon Bedrock AgentCore release notes		docs.aws.amazon.com
Itential Automation Platform	E2 / OFFICIAL	
Itential Platform 6.5 release notes		docs.itential.com

REVISION LINEAGE

1.2.0-candidate / 2026-09-17 - current-source evidence refresh; no lab rescore issued.

1.1.0-candidate / 2026-07-24 - baseline benchmark using the final comparative evaluation system.

Trigger earlier review after major release, acquisition, licensing change, critical vulnerability, material outage, failed PoC, or workload change.

BASELINE ANALYTIC RECORD CONTINUES ON PAGE 11 FOR TRACEABILITY.

ATOMIC CRITERIA MODEL

WEIGHTED CAPABILITY WITH EXPLICIT MINIMUM EVIDENCE

ID	CRITERION	WEIGHT	GATE	MINIMUM EVIDENCE
C01	Agent orchestration, execution, memory, tool, and human-control coverage	5	YES	Feature documentation, APIs, configuration exports, and scenario demonstration.
C02	Agent identity, tool authority, isolation, and policy architecture	5	YES	Threat model, identity flows, RBAC tests, audit records, and security configuration.
C03	Governance and approval controls	5	YES	Approval workflow, policy controls, separation-of-duties tests, and decision records.
C04	Durable mission, state, checkpoint, and memory integrity	5	YES	Backup, restore, rollback, drift, and corruption-recovery evidence.
C05	Automation and API quality	4	NO	API specifications, authentication tests, rate limits, event samples, and automation runs.
C06	Integration ecosystem	4	NO	Supported integrations, connector tests, failure behavior, and lifecycle ownership.
C07	Observability and evidence	4	YES	Logs, traces, metrics, reports, export tests, and evidence-retention controls.
C08	Agent, task, model, tool, and concurrency scale	4	NO	Controlled load tests, topology scale, saturation behavior, and capacity model.
C09	Resilience and continuity	5	YES	Failure injection, HA tests, recovery timing, degraded-mode operation, and runbooks.
C10	Usability and operator cognition	3	NO	Task observation, error-rate measures, navigation tests, and operator interviews.
C11	Interrupt, approval, replay, compensation, and rollback safety	5	YES	Plan or preview output, canary tests, rollback execution, and post-change verification.
C12	Extensibility and programmability	3	NO	Plugin, module, provider, workflow, or code-extension tests and upgrade impact analysis.
C13	Deployment and sovereignty options	3	NO	Deployment architecture, data-flow mapping, residency evidence, and offline tests.
C14	Lifecycle and upgrade discipline	4	NO	Release notes, upgrade test, compatibility matrix, support policy, and migration plan.
C15	Economics and cost predictability	3	NO	Bill-of-materials, licensing model, staffing estimates, metering, and sensitivity analysis.
C16	Portability and exit	4	YES	Export tests, format analysis, replacement workflow, and decommission evidence.
C17	Vendor and ecosystem risk	3	NO	License analysis, roadmap evidence, bus-factor indicators, and dependency inventory.
C18	Assurance confidence	5	YES	Source provenance, lab artifacts, production evidence, independent replication, and review date.

SCORE SCALE

0 absent; 1 materially deficient; 2 partial; 3 adequate with limits; 4 strong; 5 leading for the defined profile. Scores remain provisional until the required evidence grade is achieved.

RAW CAPABILITY SCORECARD

DOCUMENTED CAPABILITY BEFORE EVIDENCE DISCOUNT

CRITERION	LANGGRAPH	AGENTCORE	ITENTIAL
C01 Agent orchestration, execution, memory, tool, and human-control coverage	4.5	4.7	4.0
C02 Agent identity, tool authority, isolation, and policy architecture	4.0	4.6	4.4
C03 Governance and approval controls	4.1	4.5	4.7
C04 Durable mission, state, checkpoint, and memory integrity	4.6	4.4	4.2
C05 Automation and API quality	4.5	4.6	4.5
C06 Integration ecosystem	4.3	4.6	4.8
C07 Observability and evidence	4.4	4.7	4.4
C08 Agent, task, model, tool, and concurrency scale	4.2	4.7	4.2
C09 Resilience and continuity	4.1	4.5	4.3
C10 Usability and operator cognition	4.0	4.4	4.3
C11 Interrupt, approval, replay, compensation, and rollback safety	4.8	4.4	4.7
C12 Extensibility and programmability	4.8	4.2	4.2
C13 Deployment and sovereignty options	4.5	3.2	3.8
C14 Lifecycle and upgrade discipline	4.1	4.2	4.1
C15 Economics and cost predictability	4.4	3.5	3.4
C16 Portability and exit	4.7	3.3	3.5
C17 Vendor and ecosystem risk	4.2	3.4	3.6
C18 Assurance confidence	3.5	3.4	3.4

WEIGHTED BASELINE / 100

LangGraph	AgentCore	Itential
86.1	84.6	83.4

EVIDENCE-ADJUSTED SCORECARD

CAPABILITY DISCOUNTED BY CURRENT EVIDENCE CONFIDENCE

CANDIDATE	RAW	EVIDENCE CONFIDENCE	ADJUSTED	CURRENT STATE
LangGraph	86.1	59.7%	51.5	CONTROLLED LAB PENDING
AgentCore	84.6	58.2%	49.8	CONTROLLED LAB PENDING
Itential	83.4	58.2%	49.2	CONTROLLED LAB PENDING

EVIDENCE SCALE

E0 / 0.00
Unsupported or unverified

E1 / 0.38
Vendor assertion or marketing statement

E2 / 0.64
Official documentation, release notes, or source repository

E3 / 0.78
Controlled Mythos laboratory result

E4 / 0.91
Production evidence from the adopting environment

E5 / 1.00
Independent repeatable validation

CURRENT CONFIDENCE BOUNDARY

Most candidate criteria are supported at E2: official documentation or source evidence. Environment-specific laboratory, production, and independent evidence remain future gates.

PROFILE-SPECIFIC RANKINGS

EVIDENCE-ADJUSTED SENSITIVITY ANALYSIS

CUSTOM GOVERNED AGENT RUNTIME

Prioritizes deterministic graphs, durable execution, interrupts, model independence, portability, and deep engineering control.

MANAGED CLOUD AGENT OPERATIONS

Prioritizes managed runtime, identity, gateway, memory, observability, scale, and reduced platform burden.

INFRASTRUCTURE AUTOMATION CONTROL PLANE

Prioritizes approvals, integration, deterministic execution, infrastructure workflows, audit, and enterprise operations.

RANK SENSITIVITY

Small score differences are not decision certainty. Treat near-ties as a requirement for deeper PoC, commercial, migration, and operating-model evidence.

CANDIDATE DEEP DIVE / 01

LANGGRAPH

OPERATING MODEL

Low-level open-source orchestration framework and runtime for long-running, stateful agent graphs.

DOCUMENTED STRENGTHS

- Fine-grained control over graph, state, branching, interrupts, and persistence.
- Open-source and model-provider flexibility.
- Strong fit for custom deterministic agent runtimes.
- Developer-centric composability and testability.

CAUTIONS AND PROOF OBLIGATIONS

- Production platform services must be assembled or adopted separately.
- Security, tenancy, identity, policy, and operations remain architectural responsibilities.
- Low-level flexibility can become complexity without strong engineering standards.

BASELINE SCORE PROFILE

GATE DISPOSITION

PASS - custom governed agent runtime | CONDITIONAL - production platform services must be engineered

CANDIDATE DEEP DIVE / 02

AMAZON BEDROCK AGENTCORE

OPERATING MODEL

Managed framework-agnostic agent platform providing runtime, gateway, memory, identity, policy, observability, and operational services.

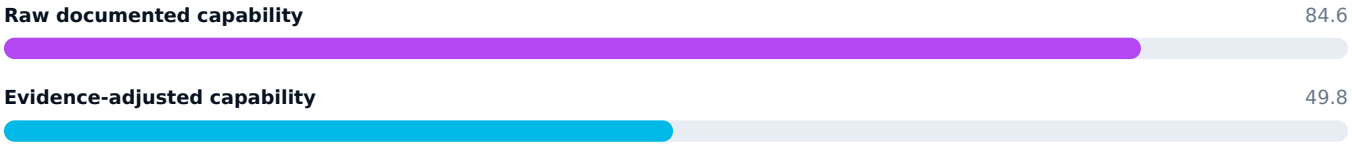
DOCUMENTED STRENGTHS

- Managed runtime and scaling.
- Integrated gateway, memory, identity, policy, and observability services.
- Framework and model flexibility within the AWS operating boundary.
- Reduced infrastructure burden for production agent operations.

CAUTIONS AND PROOF OBLIGATIONS

- AWS dependency and service maturity must be accepted.
- Portability, data boundaries, and cost require controlled validation.
- Managed abstractions can constrain low-level control and debugging.

BASELINE SCORE PROFILE



GATE DISPOSITION

PASS - managed cloud agent operations | CONDITIONAL - AWS boundary, portability, and cost

CANDIDATE DEEP DIVE / 03

ITENTIAL AUTOMATION PLATFORM

OPERATING MODEL

Enterprise network and infrastructure orchestration platform with integration, workflow, governance, and operational automation capabilities.

DOCUMENTED STRENGTHS

- Strong network and infrastructure orchestration orientation.
- Enterprise workflow, integration, approval, and operational controls.
- Useful bridge between intent, existing automation, controllers, and ITSM.
- Fits environments where deterministic infrastructure execution is primary.

CAUTIONS AND PROOF OBLIGATIONS

- Not a general-purpose low-level agent runtime.
- Model, memory, reasoning, and multi-agent features require integration or custom architecture.
- Commercial platform coupling and connector lifecycle require governance.

BASELINE SCORE PROFILE

Raw documented capability 83.4

Evidence-adjusted capability 49.2

GATE DISPOSITION

PASS - infrastructure automation control-plane profile | NOT EQUIVALENT - general-purpose agent-runtime comparison

CROSS-CANDIDATE CONTRASTS

WHERE ARCHITECTURAL DIFFERENCES MATTER MOST

CONTROL-LOOP MODEL

How authority moves from intent to plan, approval, execution, observation, reconciliation, and recovery.

STATE OWNERSHIP

Which system owns desired state, operational state, evidence, history, and conflict resolution.

MANAGED VERSUS SOVEREIGN BOUNDARY

Which runtime, identity, data, policy, relay, or service dependencies remain outside organizational control.

EXTENSION MODEL

Plugins, providers, modules, applications, workflows, SDKs, APIs, and custom code introduce different lifecycle risks.

FAILURE TRANSPARENCY

Candidates differ in how clearly they expose partial execution, degraded mode, retries, inconsistency, and recovery state.

EXIT MECHANICS

Export formats, state semantics, policy portability, knowledge transfer, and replacement effort are materially different.

CANDIDATE	CURRENT ADVANCEMENT POSITION
LangGraph	PASS - custom governed agent runtime
AgentCore	PASS - managed cloud agent operations
Itential	PASS - infrastructure automation control-plane profile

ARCHITECTURE AND INTEGRATION FIT

THE PRODUCT IS ONLY ONE PART OF THE OPERATING SYSTEM

REFERENCE OPERATING LAYERS

01 Intent, task, and mission model

02 Supervisor, graph, workflow, and durable state

03 Model, memory, context, and knowledge services

04 Tool gateway, identity, policy, approvals, and execution

05 Observation, evaluation, evidence, replay, and recovery

INTEGRATION BOUNDARIES

- Model providers and local runtimes
- MCP, APIs, CLIs, infrastructure controllers, and automation tools
- Identity, secrets, policy, and approval systems
- Vector, graph, object, event, and relational data stores
- Observability, evaluation, incident, and evidence platforms

ARCHITECTURE GATE

Every external dependency requires an owner, identity boundary, failure mode, evidence path, version policy, recovery procedure, and exit strategy.

SECURITY, OPERATIONS, LIFECYCLE, ECONOMICS, AND EXIT

CROSS-DOMAIN DECISION RECORD

SECURITY AND AUTHORITY

Identity, credentials, secrets, roles, privileged actions, signing, approvals, isolation, audit, and incident response must be tested as an integrated system.

RELIABILITY AND RECOVERY

Control-plane, data-plane, dependency, region, database, queue, network, and operator failures require defined degraded modes and recovery objectives.

CHANGE AND LIFECYCLE

Version, compatibility, migration, canary, rollback, content, plugin, provider, firmware, and decommission procedures are part of product fitness.

ECONOMICS AND CAPACITY

Licenses, metering, data, compute, hardware, storage, support, staffing, training, integration, migration, and opportunity cost require sensitivity analysis.

SOVEREIGNTY AND PRIVACY

Data location, support access, telemetry, subprocessors, encryption, key control, disconnected operation, and legal obligations must align with policy.

PORTABILITY AND EXIT

Configuration, policy, data, state, evidence, workflows, identities, and operational knowledge must be exportable and reconstructable.

DECISION OWNERSHIP

Architecture, security, operations, finance, procurement, legal, data, and service owners jointly accept the final profile, evidence, residual risk, and exit obligations.

RISK AND FAILURE REGISTER

SCENARIOS THAT CAN INVALIDATE A FEATURE-BASED SELECTION

ID	SEVERITY	FAILURE SCENARIO	REQUIRED TREATMENT
R-01	CRITICAL	A flexible framework is mistaken for a complete production platform.	PREVENT / DETECT / RECOVER / EVIDENCE
R-02	HIGH	A managed platform is accepted without testing portability, cost, data boundaries, and failure transparency.	PREVENT / DETECT / RECOVER / EVIDENCE
R-03	HIGH	Agent authority exceeds tool, data, or environment permissions.	PREVENT / DETECT / RECOVER / EVIDENCE
R-04	HIGH	State replay duplicates irreversible actions.	PREVENT / DETECT / RECOVER / EVIDENCE
R-05	MEDIUM	Human approval is cosmetic because plans, diffs, and consequences are not understandable.	PREVENT / DETECT / RECOVER / EVIDENCE
R-06	MEDIUM	Model quality is used to compensate for weak deterministic execution and verification.	PREVENT / DETECT / RECOVER / EVIDENCE

RISK RULE

A candidate with an unresolved critical failure path cannot advance because optional capability, market position, or commercial discount cannot compensate for missing safety or recovery evidence.

CONTROLLED PROOF-OF-CAPABILITY PLAN

REPEATABLE SCENARIOS, INSTRUMENTATION, AND ACCEPTANCE

TEST 01

Run a long-lived mission with branching, retry, interruption, approval, restart, replay, and compensation.

TEST 02

Integrate local and frontier models, MCP tools, secrets, identity, and policy.

TEST 03

Inject model timeout, malformed tool output, duplicate event, lost checkpoint, and operator cancellation.

TEST 04

Measure task throughput, state growth, token use, tool latency, and evidence completeness.

TEST 05

Export mission state, traces, memory, policies, and tool definitions.

TEST 06

Red-team prompt injection, confused-deputy behavior, excess authority, and unsafe recovery.

EVIDENCE PACKAGE

Preserve versions, architecture, configuration, data set, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

CONDITIONAL RECOMMENDATION AND REVALIDATION

DECISION RECORD, ACCEPTANCE, AND NEXT EVIDENCE

RECOMMENDATION POSITION

- Select the architecture layer before selecting the product.
- Require durable-state, authority, replay, approval, and failure-injection tests.
- Preserve model and tool portability through typed contracts and evidence.
- Revalidate monthly because managed agent platforms and framework capabilities are changing rapidly.

CANDIDATE	ADJ. SCORE	GATE POSITION	LAB STATE
LangGraph	51.5	PASS - custom governed agent runtime; CONDITIONAL - production platform services must be engineered	PENDING
AgentCore	49.8	PASS - managed cloud agent operations; CONDITIONAL - AWS boundary, portability, and cost	PENDING
Itential	49.2	PASS - infrastructure automation control-plane profile; NOT EQUIVALENT - general-purpose agent-runtime comparison	PENDING

REVALIDATION SCHEDULE

Review no later than 2026-10-24. Review earlier after a major release, commercial change, critical vulnerability, material outage, architecture transition, failed acceptance test, or change to the target workload profile.

SOURCE AUTHORITY AND REVISION

CURRENT EVIDENCE SNAPSHOT AND PUBLICATION HISTORY

SOURCE ID	PUBLISHER	TITLE	CHECKED
NIST-AI-RMF	NIST	AI Risk Management Framework	2026-07-24
NIST-SP800-53	NIST	Security and Privacy Controls for Information Systems and Organizations	2026-07-24
LANGGRAPH	LangChain	LangGraph Overview	2026-07-24
LANGGRAPH-GH	LangChain	LangGraph Source Repository	2026-07-24
AGENTCORE	Amazon Web Services	Amazon Bedrock AgentCore Overview	2026-07-24
AGENTCORE-GATEWAY	Amazon Web Services	Amazon Bedrock AgentCore Gateway	2026-07-24
ITENTIAL	Itential	Itential Automation Platform Documentation	2026-07-24

SOURCE POSITION

Official documentation and source repositories support the current capability model. Source records preserve publisher, title, URL, purpose, and review date in the machine-readable authority register. Commercial terms, performance, and environment-specific behavior remain unverified until controlled proof.

REVISION HISTORY

1.1.0-candidate / 2026-07-24 - permanent-publication rebuild using the final benchmark system, profile-specific rankings, evidence adjustment, mandatory gates, and controlled PoC requirements.

COMPARATIVE EVALUATION / EVIDENCE REFRESH

CMP-005

Modern Private Networking

Overlay networks, identity-aware access and private application fabrics

CURRENT EVIDENCE SNAPSHOT

Official product sources refreshed through 2026-09-17

Controlled Mythos laboratory rerun: NOT ASSERTED

BENCHMARK DOCTRINE

Gates before scores. Evidence before certainty. Profile fit before universal ranking.

VERSION 1.2.0-candidate

CANDIDATE COMPARATIVE EVALUATION / PUBLIC

Modern Private Networking

Overlay networks, identity-aware access and private application fabrics

DOCUMENT ID
CMP-005
VERSION
1.2.0-candidate
STATUS
Candidate Comparative Evaluation
PUBLICATION DATE
2026-09-17
SCHEDULED REVIEW
2026-12-16
CANONICAL ROUTE
/library/evaluations/cmp-005/

4 CANDIDATES

18 ATOMIC CRITERIA

9 MANDATORY GATES

3 WORKLOAD PROFILES

E2 CURRENT MAX EVIDENCE

CURRENT DECISION BOUNDARY

No universal private-networking winner. The products continue to optimize different authority, hosting, access, and routing models. Reproduce identity, relay, route convergence, self-host, outage, export, and migration behavior for the target profile.

NO CURRENT UNIVERSAL WINNER IS PUBLISHED FROM THIS REFRESH.

July 24 baseline scores are preserved as lineage and sensitivity evidence, not silently reissued as September laboratory results.

Tailscale

v1.102.4 / 2026-09-10

REFRESH TRIGGER

Client and Kubernetes-operator work continues to improve connectivity, scale, IPv6, and multi-cluster operations.

NetBird

v0.78.2 / 2026-09-14

REFRESH TRIGGER

Rapid self-hosted and policy-management evolution keeps sovereignty attractive but increases upgrade/revalidation importance.

Cloudflare One

Cloud service updates through 2026-09-15

REFRESH TRIGGER

Tagged infrastructure targets, fresh-auth controls, DLP discovery, and route-management improvements broaden the access fabric.

ZeroTier

ZeroTier One 1.16.2 / New Central Jun 2026

REFRESH TRIGGER

New Central API v2, webhooks and full custom Flow Rules materially strengthen automation and policy management.

FINAL BENCHMARK SYSTEM / DECISION ARCHITECTURE

01 SCOPE

Freeze decision, candidates, versions and exclusions.

02 GATES

Test non-compensable trust, state, recovery, change and exit.

03 SCORES

Measure atomic capability with approved profile weights.

04 CONFIDENCE

Discount by source grade, freshness, repeatability and relevance.

05 PROOF

Run normal, failure, migration, recovery and operator scenarios.

06 DECISION

Record conditional fit, residual risk, owner and revalidation trigger.

EVIDENCE CONFIDENCE / MODIFIERS

E0	E1	E2	E3	E4	E5
0.00	0.38	0.64	0.78	0.91	1.00
UNSUPPORTED	VENDOR ASSERTION	OFFICIAL SOURCE	MYTHOS LAB	PRODUCTION	INDEPENDENT

THIS EDITION

Current official product and release sources have been refreshed to E2. No E3 controlled Mythos lab rerun is claimed. Therefore current profile rankings remain intentionally unissued until the proof plan is executed.

What changed since the July benchmark snapshot

Changes below are sourced from current official release documentation. They identify revalidation triggers; they are not translated into new numeric scores without controlled proof.

Tailscale

v1.102.4 / 2026-09-10

MATERIAL DELTA

Client and Kubernetes-operator work continues to improve connectivity, scale, IPv6, and multi-cluster operations.

CRITERIA TOUCHED

C01 / C07 / C08 / C09 / C14

LOW REVALIDATION

NetBird

v0.78.2 / 2026-09-14

MATERIAL DELTA

Rapid self-hosted and policy-management evolution keeps sovereignty attractive but increases upgrade/revalidation importance.

CRITERIA TOUCHED

C01 / C03 / C05 / C13 / C14 / C16

MODERATE REVALIDATION

Cloudflare One

Cloud service updates through 2026-09-15

MATERIAL DELTA

Tagged infrastructure targets, fresh-auth controls, DLP discovery, and route-management improvements broaden the access fabric.

CRITERIA TOUCHED

C01 / C02 / C03 / C05 / C07 / C14

MODERATE REVALIDATION

ZeroTier

ZeroTier One 1.16.2 / New Central Jun 2026

MATERIAL DELTA

New Central API v2, webhooks and full custom Flow Rules materially strengthen automation and policy management.

CRITERIA TOUCHED

C01 / C03 / C05 / C07 / C12 / C14

MODERATE REVALIDATION

Mandatory gates remain non-compensable

No current release is marked PASS solely from documentation. E2 means the official source supports the capability path; LAB and RETEST identify proof still required. HOLD blocks a current decision.

MANDATORY GATE	TAILSCALE	NETBIRD	CLOUDFLARE ONE	ZEROTIER
C01 Functional coverage	RETEST	RETEST	RETEST	RETEST
C02 Security / trust	E2	E2	RETEST	E2
C03 Governance / approval	E2	RETEST	RETEST	RETEST
C04 State integrity	E2	E2	E2	E2
C07 Observability / evidence	RETEST	E2	RETEST	RETEST
C09 Resilience / continuity	RETEST	LAB	LAB	LAB
C11 Change safety / rollback	LAB	LAB	LAB	LAB
C16 Portability / exit	LAB	RETEST	LAB	LAB
C18 Assurance confidence	HOLD	HOLD	HOLD	HOLD

LEGEND

E2 official-source support only

LAB controlled proof required

RETEST release delta touches this gate

HOLD decision gate remains closed

Historical profile sensitivity - preserved, not reissued

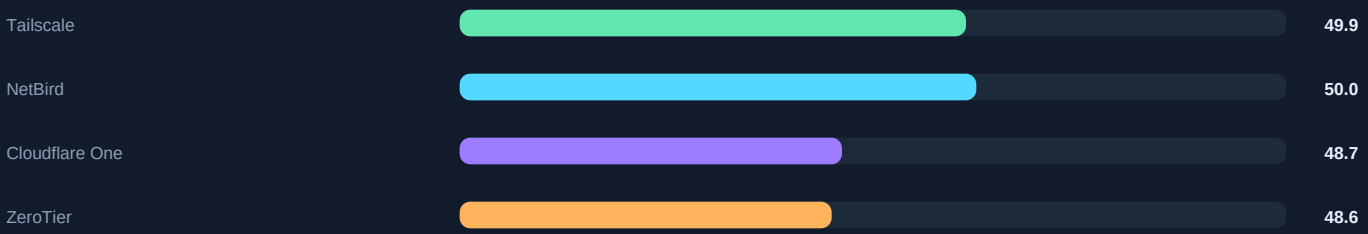
Values below are the 2026-07-24 evidence-adjusted baseline. They show sensitivity to operating-model weights. The September refresh does not recompute rank because E3 lab proof has not been reproduced.

IDENTITY-AWARE WORKFORCE ACCESS

BASELINE SPREAD 2.9 PTS

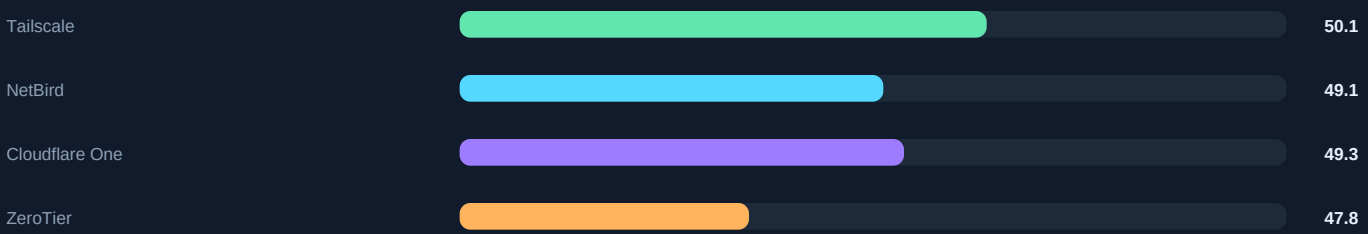
SOVEREIGN PRIVATE NETWORK

BASELINE SPREAD 1.4 PTS



MULTI-SITE APPLICATION AND NETWORK FABRIC

BASELINE SPREAD 2.3 PTS



CURRENT RANK STATUS: WITHHELD PENDING CONTROLLED PROOF.

Evidence confidence is separate from capability

E5 — Independent repeatable validation
confidence modifier 1.00

E4 — Production evidence in adopting environment
confidence modifier 0.91

E3 — Controlled Mythos laboratory result
confidence modifier 0.78

E2 — Official documentation / release / source
confidence modifier 0.64

E1 — Vendor assertion or marketing
confidence modifier 0.38

E0 — Unsupported or unverified
confidence modifier 0.00

CURRENT EVIDENCE STATE

Tailscale — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

NetBird — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

Cloudflare One — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

ZeroTier — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

ASSURANCE RULE

A current release note can change capability context, but it cannot raise assurance beyond E2. Current recommendation strength is therefore bounded until the test plan produces reproducible artifacts.

Lifecycle risk is evaluated independently of features

The benchmark models migration, upgrade, compatibility, support, staffing, data/state export, downtime and exit. Current release changes below are explicit proof triggers.

Tailscale

LOW REVALIDATION

Client and Kubernetes-operator work continues to improve connectivity, scale, IPv6, and multi-cluster operations.

RETEST: C01 / C07 / C08 / C09 / C14

NetBird

MODERATE REVALIDATION

Rapid self-hosted and policy-management evolution keeps sovereignty attractive but increases upgrade/revalidation importance.

RETEST: C01 / C03 / C05 / C13 / C14 / C16

Cloudflare One

MODERATE REVALIDATION

Tagged infrastructure targets, fresh-auth controls, DLP discovery, and route-management improvements broaden the access fabric.

RETEST: C01 / C02 / C03 / C05 / C07 / C14

ZeroTier

MODERATE REVALIDATION

New Central API v2, webhooks and full custom Flow Rules materially strengthen automation and policy management.

RETEST: C01 / C03 / C05 / C07 / C12 / C14

IRREVERSIBLE COMMITMENT GATE

Before migration or procurement lock-in, prove rollback, state portability, operational reconstruction, and supplier-exit assumptions.

Controlled proof is the next decision-producing step

REPRODUCTION STATE / NOT EXECUTED IN THIS EVIDENCE-REFRESH RELEASE

01

CONTROLLED SCENARIO

Connect workforce endpoints, servers, containers, mobile devices, sites, and overlapping networks.

02

CONTROLLED SCENARIO

Test direct, relayed, site-to-site, subnet-router, exit-node, application, SSH, and DNS flows.

03

CONTROLLED SCENARIO

Exercise identity outage, relay loss, controller loss, route conflict, key rotation, and revoked-device behavior.

04

CONTROLLED SCENARIO

Measure connection establishment, throughput, latency, roaming, and recovery.

05

CONTROLLED SCENARIO

Validate policy-as-code, posture, logging, API, and incident response.

06

CONTROLLED SCENARIO

Export or reconstruct all routes, policies, devices, identities, and operational evidence.

EVIDENCE PACKAGE

Preserve exact versions, architecture, configuration, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

Current decision posture

No universal private-networking winner. The products continue to optimize different authority, hosting, access, and routing models. Reproduce identity, relay, route convergence, self-host, outage, export, and migration behavior for the target profile.

CURRENT RANKING: NOT REISSUED / CONTROLLED LAB REQUIRED

CURRENT OFFICIAL-SOURCE REGISTER / CHECKED 2026-09-17

Tailscale	E2 / OFFICIAL	
Tailscale changelog		tailscale.com
NetBird	E2 / OFFICIAL	
NetBird documentation and releases		docs.netbird.io
Cloudflare One	E2 / OFFICIAL	
Cloudflare One changelog		developers.cloudflare.com
ZeroTier	E2 / OFFICIAL	
ZeroTier changelog		docs.zerotier.com

REVISION LINEAGE

1.2.0-candidate / 2026-09-17 - current-source evidence refresh; no lab rescore issued.

1.1.0-candidate / 2026-07-24 - baseline benchmark using the final comparative evaluation system.

Trigger earlier review after major release, acquisition, licensing change, critical vulnerability, material outage, failed PoC, or workload change.

BASELINE ANALYTIC RECORD CONTINUES ON PAGE 11 FOR TRACEABILITY.

ATOMIC CRITERIA MODEL

WEIGHTED CAPABILITY WITH EXPLICIT MINIMUM EVIDENCE

ID	CRITERION	WEIGHT	GATE	MINIMUM EVIDENCE
C01	Identity-aware private networking and application-access coverage	5	YES	Feature documentation, APIs, configuration exports, and scenario demonstration.
C02	Peer identity, key management, policy, relay, and trust architecture	5	YES	Threat model, identity flows, RBAC tests, audit records, and security configuration.
C03	Governance and approval controls	5	YES	Approval workflow, policy controls, separation-of-duties tests, and decision records.
C04	Identity, device, route, policy, and key-state integrity	5	YES	Backup, restore, rollback, drift, and corruption-recovery evidence.
C05	Automation and API quality	4	NO	API specifications, authentication tests, rate limits, event samples, and automation runs.
C06	Integration ecosystem	4	NO	Supported integrations, connector tests, failure behavior, and lifecycle ownership.
C07	Observability and evidence	4	YES	Logs, traces, metrics, reports, export tests, and evidence-retention controls.
C08	Peer, route, site, policy, and relay scale	4	NO	Controlled load tests, topology scale, saturation behavior, and capacity model.
C09	Resilience and continuity	5	YES	Failure injection, HA tests, recovery timing, degraded-mode operation, and runbooks.
C10	Usability and operator cognition	3	NO	Task observation, error-rate measures, navigation tests, and operator interviews.
C11	Change safety and rollback	5	YES	Plan or preview output, canary tests, rollback execution, and post-change verification.
C12	Extensibility and programmability	3	NO	Plugin, module, provider, workflow, or code-extension tests and upgrade impact analysis.
C13	SaaS, self-hosted, hybrid, and sovereign control-plane options	3	NO	Deployment architecture, data-flow mapping, residency evidence, and offline tests.
C14	Lifecycle and upgrade discipline	4	NO	Release notes, upgrade test, compatibility matrix, support policy, and migration plan.
C15	Economics and cost predictability	3	NO	Bill-of-materials, licensing model, staffing estimates, metering, and sensitivity analysis.
C16	Portability and exit	4	YES	Export tests, format analysis, replacement workflow, and decommission evidence.
C17	Vendor and ecosystem risk	3	NO	License analysis, roadmap evidence, bus-factor indicators, and dependency inventory.
C18	Assurance confidence	5	YES	Source provenance, lab artifacts, production evidence, independent replication, and review date.

SCORE SCALE

0 absent; 1 materially deficient; 2 partial; 3 adequate with limits; 4 strong; 5 leading for the defined profile. Scores remain provisional until the required evidence grade is achieved.

RAW CAPABILITY SCORECARD

DOCUMENTED CAPABILITY BEFORE EVIDENCE DISCOUNT

CRITERION	TAILSCALE	NETBIRD	CLOUDFLARE	ZEROTIER
C01 Identity-aware private networking and application-access coverage	4.7	4.4	4.5	4.3
C02 Peer identity, key management, policy, relay, and trust architecture	4.6	4.4	4.6	4.1
C03 Governance and approval controls	4.4	4.1	4.4	3.9
C04 Identity, device, route, policy, and key-state integrity	4.3	4.2	4.2	4.2
C05 Automation and API quality	4.7	4.4	4.6	4.2
C06 Integration ecosystem	4.5	4.2	4.8	4.0
C07 Observability and evidence	4.4	4.1	4.5	4.0
C08 Peer, route, site, policy, and relay scale	4.5	4.0	4.7	4.3
C09 Resilience and continuity	4.3	4.0	4.5	4.1
C10 Usability and operator cognition	4.8	4.4	4.5	3.8
C11 Change safety and rollback	4.3	4.2	4.2	4.1
C12 Extensibility and programmability	4.2	4.3	4.0	4.2
C13 SaaS, self-hosted, hybrid, and sovereign control-plane options	3.0	4.9	2.8	4.7
C14 Lifecycle and upgrade discipline	4.4	4.0	4.3	3.9
C15 Economics and cost predictability	4.0	4.4	3.6	4.4
C16 Portability and exit	3.7	4.6	3.2	4.5
C17 Vendor and ecosystem risk	3.8	4.0	3.4	3.8
C18 Assurance confidence	3.6	3.4	3.5	3.4

WEIGHTED BASELINE / 100

Tailscale	NetBird	Cloudflare	ZeroTier
85.2	84.0	83.5	81.8

EVIDENCE-ADJUSTED SCORECARD

CAPABILITY DISCOUNTED BY CURRENT EVIDENCE CONFIDENCE

CANDIDATE	RAW	EVIDENCE CONFIDENCE	ADJUSTED	CURRENT STATE
Tailscale	85.2	58.2%	50.1	CONTROLLED LAB PENDING
NetBird	84.0	58.2%	49.4	CONTROLLED LAB PENDING
Cloudflare	83.5	58.2%	49.0	CONTROLLED LAB PENDING
ZeroTier	81.8	58.2%	48.0	CONTROLLED LAB PENDING

EVIDENCE SCALE

E0 / 0.00
Unsupported or unverified

E1 / 0.38
Vendor assertion or marketing statement

E2 / 0.64
Official documentation, release notes, or source repository

E3 / 0.78
Controlled Mythos laboratory result

E4 / 0.91
Production evidence from the adopting environment

E5 / 1.00
Independent repeatable validation

CURRENT CONFIDENCE BOUNDARY

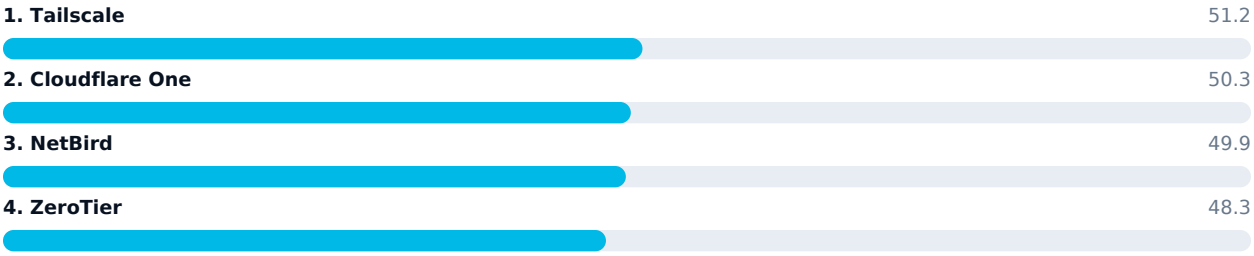
Most candidate criteria are supported at E2: official documentation or source evidence. Environment-specific laboratory, production, and independent evidence remain future gates.

PROFILE-SPECIFIC RANKINGS

EVIDENCE-ADJUSTED SENSITIVITY ANALYSIS

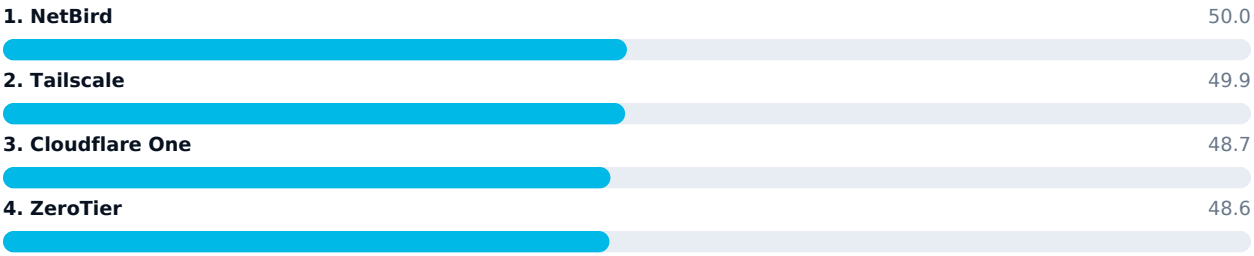
IDENTITY-AWARE WORKFORCE ACCESS

Prioritizes user experience, identity policy, posture, application access, audit, and rapid deployment.



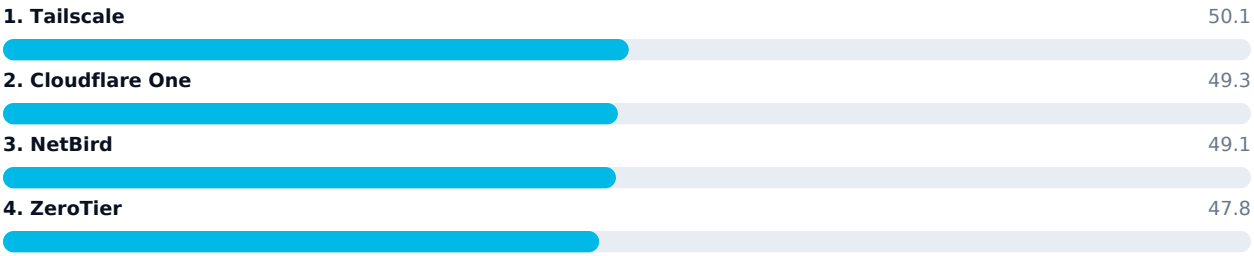
SOVEREIGN PRIVATE NETWORK

Prioritizes self-hosting, key and route control, exportability, offline tolerance, and reduced external dependency.



MULTI-SITE APPLICATION AND NETWORK FABRIC

Prioritizes sites, routes, application publishing, relay performance, global operations, and policy.



RANK SENSITIVITY

Small score differences are not decision certainty. Treat near-ties as a requirement for deeper PoC, commercial, migration, and operating-model evidence.

CANDIDATE DEEP DIVE / 01

TAILSCALE

OPERATING MODEL

Managed WireGuard-based tailnet with identity integration, grants, device posture, routing, and application-layer access extensions.

DOCUMENTED STRENGTHS

- Excellent operator and end-user experience.
- Strong identity-aware policy and modern grants model.
- Broad endpoint support and rapid deployment.
- Mature managed coordination and operational simplicity.

CAUTIONS AND PROOF OBLIGATIONS

- Managed control-plane dependency is material.
- Advanced routing and application patterns require careful architecture.
- Sovereignty and full exit requirements must be validated explicitly.

BASELINE SCORE PROFILE

Raw documented capability 85.2

Evidence-adjusted capability 50.1

GATE DISPOSITION

PASS - workforce and rapid-deployment profile | CONDITIONAL - control-plane and exit boundaries

CANDIDATE DEEP DIVE / 02

NETBIRD

OPERATING MODEL

WireGuard-based private-network platform available as managed SaaS or self-hosted control plane with groups, policies, posture, routes, and relays.

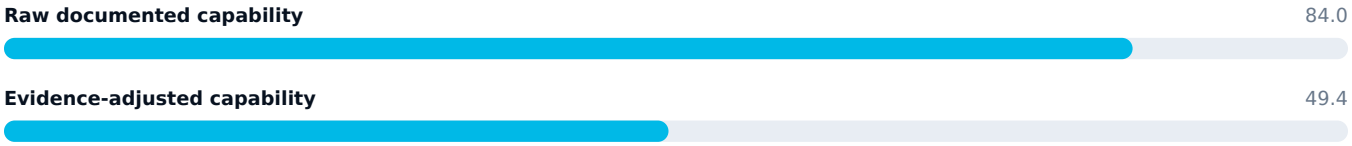
DOCUMENTED STRENGTHS

- Self-hosted and managed deployment options.
- Identity-aware access policies and posture checks.
- Open-source transparency and growing integration surface.
- Strong fit for sovereign or hybrid operating models.

CAUTIONS AND PROOF OBLIGATIONS

- Self-hosting transfers reliability and security responsibility to the operator.
- Ecosystem maturity and scale must be proven for large estates.
- Upgrade, relay, identity, and database operations require disciplined ownership.

BASELINE SCORE PROFILE



GATE DISPOSITION

PASS - sovereign and hybrid profile | CONDITIONAL - self-hosted operations and scale

CANDIDATE DEEP DIVE / 03

CLOUDFLARE ONE

OPERATING MODEL

Cloud-delivered connectivity and access platform using Tunnel, Mesh, Access, global network services, and policy.

DOCUMENTED STRENGTHS

- Strong application publishing and identity-aware access.
- Global network and managed connector operations.
- Outbound-only tunnel model reduces inbound exposure.
- Converges application, private-network, and broader security services.

CAUTIONS AND PROOF OBLIGATIONS

- Cloudflare dependency and service coupling are significant.
- Peer-mesh and site-routing requirements must be tested carefully.
- Exit, routing behavior, logging, and pricing require profile-specific validation.

BASELINE SCORE PROFILE

Raw documented capability 83.5

Evidence-adjusted capability 49.0

GATE DISPOSITION

PASS - application access and global managed fabric | CONDITIONAL - platform coupling and route behavior

CANDIDATE DEEP DIVE / 04

ZEROTIER

OPERATING MODEL

Virtual Ethernet networking platform with controllers, roots, distributed rules engine, routing, bridging, and enterprise deployment options.

DOCUMENTED STRENGTHS

- Layer-2 and Layer-3 virtual networking flexibility.
- Distributed rules engine and network virtualization model.
- Self-hosting and private-root options support sovereignty.
- Useful for complex device, site, lab, and embedded networking.

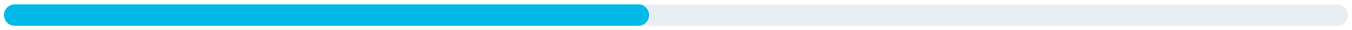
CAUTIONS AND PROOF OBLIGATIONS

- Identity-aware application access is less turnkey.
- Operational model and troubleshooting can be more network-centric.
- Policy, controller, and route design require specialized engineering.

BASELINE SCORE PROFILE

Raw documented capability 81.8

Evidence-adjusted capability 48.0



GATE DISPOSITION

PASS - virtual network and sovereign network profile | CONDITIONAL - identity and operational model

ARCHITECTURE AND INTEGRATION FIT

THE PRODUCT IS ONLY ONE PART OF THE OPERATING SYSTEM

REFERENCE OPERATING LAYERS

01 Identity, device enrollment, and key authority

02 Coordination, discovery, relay, and routing control plane

03 Peer or connector data plane

04 Policy, posture, application access, and segmentation

05 Telemetry, evidence, lifecycle, recovery, and exit

INTEGRATION BOUNDARIES

- Identity providers and device management
- DNS and service discovery
- Endpoint security and posture
- Cloud, datacenter, branch, home, lab, and embedded networks
- SIEM, ITSM, and automation
- Application proxies, SSH, RDP, database, and web services

ARCHITECTURE GATE

Every external dependency requires an owner, identity boundary, failure mode, evidence path, version policy, recovery procedure, and exit strategy.

SECURITY, OPERATIONS, LIFECYCLE, ECONOMICS, AND EXIT

CROSS-DOMAIN DECISION RECORD

SECURITY AND AUTHORITY

Identity, credentials, secrets, roles, privileged actions, signing, approvals, isolation, audit, and incident response must be tested as an integrated system.

RELIABILITY AND RECOVERY

Control-plane, data-plane, dependency, region, database, queue, network, and operator failures require defined degraded modes and recovery objectives.

CHANGE AND LIFECYCLE

Version, compatibility, migration, canary, rollback, content, plugin, provider, firmware, and decommission procedures are part of product fitness.

ECONOMICS AND CAPACITY

Licenses, metering, data, compute, hardware, storage, support, staffing, training, integration, migration, and opportunity cost require sensitivity analysis.

SOVEREIGNTY AND PRIVACY

Data location, support access, telemetry, subprocessors, encryption, key control, disconnected operation, and legal obligations must align with policy.

PORTABILITY AND EXIT

Configuration, policy, data, state, evidence, workflows, identities, and operational knowledge must be exportable and reconstructable.

DECISION OWNERSHIP

Architecture, security, operations, finance, procurement, legal, data, and service owners jointly accept the final profile, evidence, residual risk, and exit obligations.

RISK AND FAILURE REGISTER

SCENARIOS THAT CAN INVALIDATE A FEATURE-BASED SELECTION

ID	SEVERITY	FAILURE SCENARIO	REQUIRED TREATMENT
R-01	CRITICAL	A proof of connectivity is mistaken for proof of policy, resilience, and operations.	PREVENT / DETECT / RECOVER / EVIDENCE
R-02	HIGH	Relay dependence or route asymmetry causes hidden latency and failure.	PREVENT / DETECT / RECOVER / EVIDENCE
R-03	HIGH	Identity or posture outages lock out critical operations.	PREVENT / DETECT / RECOVER / EVIDENCE
R-04	HIGH	Self-hosted control planes are deployed without HA, patching, backup, and monitoring.	PREVENT / DETECT / RECOVER / EVIDENCE
R-05	MEDIUM	Application publishing bypasses segmentation or origin hardening.	PREVENT / DETECT / RECOVER / EVIDENCE
R-06	MEDIUM	Exit testing reveals that device, route, policy, and key state cannot be migrated safely.	PREVENT / DETECT / RECOVER / EVIDENCE

RISK RULE

A candidate with an unresolved critical failure path cannot advance because optional capability, market position, or commercial discount cannot compensate for missing safety or recovery evidence.

CONTROLLED PROOF-OF-CAPABILITY PLAN

REPEATABLE SCENARIOS, INSTRUMENTATION, AND ACCEPTANCE

TEST 01

Connect workforce endpoints, servers, containers, mobile devices, sites, and overlapping networks.

TEST 02

Test direct, relayed, site-to-site, subnet-router, exit-node, application, SSH, and DNS flows.

TEST 03

Exercise identity outage, relay loss, controller loss, route conflict, key rotation, and revoked-device behavior.

TEST 04

Measure connection establishment, throughput, latency, roaming, and recovery.

TEST 05

Validate policy-as-code, posture, logging, API, and incident response.

TEST 06

Export or reconstruct all routes, policies, devices, identities, and operational evidence.

EVIDENCE PACKAGE

Preserve versions, architecture, configuration, data set, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

CONDITIONAL RECOMMENDATION AND REVALIDATION

DECISION RECORD, ACCEPTANCE, AND NEXT EVIDENCE

RECOMMENDATION POSITION

- Choose the dominant access pattern before ranking candidates.
- Test direct and relayed behavior from representative geographies and restricted networks.
- Treat identity, DNS, relay, and control-plane failure as mandatory scenarios.
- Revalidate quarterly and after material routing, policy, or pricing changes.

CANDIDATE	ADJ. SCORE	GATE POSITION	LAB STATE
Tailscale	50.1	PASS - workforce and rapid-deployment profile; CONDITIONAL - control-plane and exit boundaries	PENDING
NetBird	49.4	PASS - sovereign and hybrid profile; CONDITIONAL - self-hosted operations and scale	PENDING
Cloudflare	49.0	PASS - application access and global managed fabric; CONDITIONAL - platform coupling and route behavior	PENDING
ZeroTier	48.0	PASS - virtual network and sovereign network profile; CONDITIONAL - identity and operational model	PENDING

REVALIDATION SCHEDULE

Review no later than 2026-10-24. Review earlier after a major release, commercial change, critical vulnerability, material outage, architecture transition, failed acceptance test, or change to the target workload profile.

SOURCE AUTHORITY AND REVISION

CURRENT EVIDENCE SNAPSHOT AND PUBLICATION HISTORY

SOURCE ID	PUBLISHER	TITLE	CHECKED
NIST-CSF-20	NIST	Cybersecurity Framework 2.0	2026-07-24
NIST-SP800-53	NIST	Security and Privacy Controls for Information Systems and Organizations	2026-07-24
TAILSCALE-GRANTS	Tailscale	Tailscale Grants	2026-07-24
TAILSCALE-POLICY	Tailscale	Tailnet Policy File Syntax	2026-07-24
NETBIRD-ACCESS	NetBird	NetBird Access Control	2026-07-24
NETBIRD-SELFHOST	NetBird	NetBird Self-Hosting Guide	2026-07-24
CLOUDFLARE-TUNNEL	Cloudflare	Cloudflare Tunnel Documentation	2026-07-24
CLOUDFLARE-MESH	Cloudflare	Cloudflare Mesh Routes	2026-07-24
ZEROTIER	ZeroTier	ZeroTier Enterprise Deployment Guide	2026-07-24
ZEROTIER-RULES	ZeroTier	ZeroTier Rules Engine	2026-07-24

SOURCE POSITION

Official documentation and source repositories support the current capability model. Source records preserve publisher, title, URL, purpose, and review date in the machine-readable authority register. Commercial terms, performance, and environment-specific behavior remain unverified until controlled proof.

REVISION HISTORY

1.1.0-candidate / 2026-07-24 - permanent-publication rebuild using the final benchmark system, profile-specific rankings, evidence adjustment, mandatory gates, and controlled PoC requirements.

COMPARATIVE EVALUATION / EVIDENCE REFRESH

CMP-006

Campus Networking Platforms

Wired, wireless, assurance and campus operations

CURRENT EVIDENCE SNAPSHOT

Official product sources refreshed through 2026-09-17

Controlled Mythos laboratory rerun: NOT ASSERTED

BENCHMARK DOCTRINE

Gates before scores. Evidence before certainty. Profile fit before universal ranking.

VERSION 1.2.0-candidate

CANDIDATE COMPARATIVE EVALUATION / PUBLIC

Campus Networking Platforms

Wired, wireless, assurance and campus operations

DOCUMENT ID
CMP-006
VERSION
1.2.0-candidate
STATUS
Candidate Comparative Evaluation
PUBLICATION DATE
2026-09-17
SCHEDULED REVIEW
2026-12-16
CANONICAL ROUTE
/library/evaluations/cmp-006/

4 CANDIDATES

18 ATOMIC CRITERIA

9 MANDATORY GATES

3 WORKLOAD PROFILES

E2 CURRENT MAX EVIDENCE

CURRENT DECISION BOUNDARY

Reopen the cloud-native assurance and mixed-vendor profile. Mist support for selected AOS-CX switches changes the candidate boundary. Current versions require lab proof on provisioning, assurance, NAC, upgrade, failure, and rollback before any portfolio preference is asserted.

NO CURRENT UNIVERSAL WINNER IS PUBLISHED FROM THIS REFRESH.

July 24 baseline scores are preserved as lineage and sensitivity evidence, not silently reissued as September laboratory results.

Cisco Catalyst Center

3.3.1 / 2026-09-09

REFRESH TRIGGER

3.3.1 refreshes integrations and operations across the Catalyst management surface.

HPE Aruba Networking Central

Cloud updates through 2026-08-18

REFRESH TRIGGER

Central continues rapid cloud delivery across configuration, NAC, assurance, and Copilot-oriented operations.

Juniper Mist

2026-09-09 update

REFRESH TRIGGER

Mist now onboards, configures, monitors, upgrades, and troubleshoots selected HPE AOS-CX switches, materially changing mixed-vendor fit.

Fortinet Campus and Branch

FortiManager 7.6.7 / FortiOS 7.6.7

REFRESH TRIGGER

Current management and FortiOS lines continue security-converged branch/campus development with explicit upgrade constraints.

FINAL BENCHMARK SYSTEM / DECISION ARCHITECTURE

01 SCOPE

Freeze decision, candidates, versions and exclusions.

02 GATES

Test non-compensable trust, state, recovery, change and exit.

03 SCORES

Measure atomic capability with approved profile weights.

04 CONFIDENCE

Discount by source grade, freshness, repeatability and relevance.

05 PROOF

Run normal, failure, migration, recovery and operator scenarios.

06 DECISION

Record conditional fit, residual risk, owner and revalidation trigger.

EVIDENCE CONFIDENCE / MODIFIERS

E0	E1	E2	E3	E4	E5
0.00	0.38	0.64	0.78	0.91	1.00
UNSUPPORTED	VENDOR ASSERTION	OFFICIAL SOURCE	MYTHOS LAB	PRODUCTION	INDEPENDENT

THIS EDITION

Current official product and release sources have been refreshed to E2. No E3 controlled Mythos lab rerun is claimed. Therefore current profile rankings remain intentionally unissued until the proof plan is executed.

What changed since the July benchmark snapshot

Changes below are sourced from current official release documentation. They identify revalidation triggers; they are not translated into new numeric scores without controlled proof.

Cisco Catalyst Center

3.3.1 / 2026-09-09

MATERIAL DELTA

3.3.1 refreshes integrations and operations across the Catalyst management surface.

CRITERIA TOUCHED

C01 / C05 / C06 / C07 / C14

MODERATE REVALIDATION

HPE Aruba Networking Central

Cloud updates through 2026-08-18

MATERIAL DELTA

Central continues rapid cloud delivery across configuration, NAC, assurance, and Copilot-oriented operations.

CRITERIA TOUCHED

C01 / C03 / C07 / C10 / C14

MODERATE REVALIDATION

Juniper Mist

2026-09-09 update

MATERIAL DELTA

Mist now onboards, configures, monitors, upgrades, and troubleshoots selected HPE AOS-CX switches, materially changing mixed-vendor fit.

CRITERIA TOUCHED

C01 / C05 / C06 / C08 / C10 / C14

MATERIAL REVALIDATION

Fortinet Campus and Branch

FortiManager 7.6.7 / FortiOS 7.6.7

MATERIAL DELTA

Current management and FortiOS lines continue security-converged branch/campus development with explicit upgrade constraints.

CRITERIA TOUCHED

C01 / C02 / C05 / C09 / C14

MODERATE REVALIDATION

Mandatory gates remain non-compensable

No current release is marked PASS solely from documentation. E2 means the official source supports the capability path; LAB and RETEST identify proof still required. HOLD blocks a current decision.

MANDATORY GATE	CISCO CATALYST	HPE ARUBA NETWO	JUNIPER MIST	FORTINET CAMPUS
C01 Functional coverage	RETEST	RETEST	RETEST	RETEST
C02 Security / trust	E2	E2	E2	RETEST
C03 Governance / approval	E2	RETEST	E2	E2
C04 State integrity	E2	E2	E2	E2
C07 Observability / evidence	RETEST	RETEST	E2	E2
C09 Resilience / continuity	LAB	LAB	LAB	RETEST
C11 Change safety / rollback	LAB	LAB	LAB	LAB
C16 Portability / exit	LAB	LAB	LAB	LAB
C18 Assurance confidence	HOLD	HOLD	HOLD	HOLD

LEGEND

E2 official-source support only

LAB controlled proof required

RETEST release delta touches this gate

HOLD decision gate remains closed

Historical profile sensitivity - preserved, not reissued

Values below are the 2026-07-24 evidence-adjusted baseline. They show sensitivity to operating-model weights. The September refresh does not recompute rank because E3 lab proof has not been reproduced.

LARGE CISCO CAMPUS

BASELINE SPREAD 2.2 PTS

CLOUD-NATIVE EXPERIENCE ASSURANCE

BASELINE SPREAD 2.2 PTS

SECURITY-CONVERGED BRANCH AND CAMPUS

BASELINE SPREAD 1.8 PTS

CURRENT RANK STATUS: WITHHELD PENDING CONTROLLED PROOF.

Evidence confidence is separate from capability

E5 — Independent repeatable validation
confidence modifier 1.00

E4 — Production evidence in adopting environment
confidence modifier 0.91

E3 — Controlled Mythos laboratory result
confidence modifier 0.78

E2 — Official documentation / release / source
confidence modifier 0.64

E1 — Vendor assertion or marketing
confidence modifier 0.38

E0 — Unsupported or unverified
confidence modifier 0.00

CURRENT EVIDENCE STATE

Cisco Catalyst Center	**E2 REFRESHED**
Historical adjusted confidence: 58.2%	
E3 LAB: NOT REPRODUCED	
HPE Aruba Networking Central	**E2 REFRESHED**
Historical adjusted confidence: 58.2%	
E3 LAB: NOT REPRODUCED	
Juniper Mist	**E2 REFRESHED**
Historical adjusted confidence: 58.2%	
E3 LAB: NOT REPRODUCED	
Fortinet Campus and Branch	**E2 REFRESHED**
Historical adjusted confidence: 58.2%	
E3 LAB: NOT REPRODUCED	

ASSURANCE RULE

A current release note can change capability context, but it cannot raise assurance beyond E2. Current recommendation strength is therefore bounded until the test plan produces reproducible artifacts.

Lifecycle risk is evaluated independently of features

The benchmark models migration, upgrade, compatibility, support, staffing, data/state export, downtime and exit. Current release changes below are explicit proof triggers.

Cisco Catalyst Center

MODERATE REVALIDATION

3.3.1 refreshes integrations and operations across the Catalyst management surface.

RETEST: C01 / C05 / C06 / C07 / C14

HPE Aruba Networking Central

MODERATE REVALIDATION

Central continues rapid cloud delivery across configuration, NAC, assurance, and Copilot-oriented operations.

RETEST: C01 / C03 / C07 / C10 / C14

Juniper Mist

MATERIAL REVALIDATION

Mist now onboards, configures, monitors, upgrades, and troubleshoots selected HPE AOS-CX switches, materially changing mixed-vendor fit.

RETEST: C01 / C05 / C06 / C08 / C10 / C14

Fortinet Campus and Branch

MODERATE REVALIDATION

Current management and FortiOS lines continue security-converged branch/campus development with explicit upgrade constraints.

RETEST: C01 / C02 / C05 / C09 / C14

IRREVERSIBLE COMMITMENT GATE

Before migration or procurement lock-in, prove rollback, state portability, operational reconstruction, and supplier-exit assumptions.

Controlled proof is the next decision-producing step

REPRODUCTION STATE / NOT EXECUTED IN THIS EVIDENCE-REFRESH RELEASE

01

CONTROLLED SCENARIO

Deploy a representative wired and wireless site hierarchy with users, guests, IoT, voice, and critical applications.

02

CONTROLLED SCENARIO

Measure discovery, provisioning, configuration compliance, telemetry completeness, and client-assurance accuracy.

03

CONTROLLED SCENARIO

Inject DHCP, DNS, RADIUS, RF, roaming, PoE, VLAN, routing, WAN, and certificate failures.

04

CONTROLLED SCENARIO

Exercise software upgrade, rollback, device replacement, site outage, and management-plane loss.

05

CONTROLLED SCENARIO

Validate API, webhooks, ITSM, source-of-truth, SIEM, and automated-remediation controls.

06

CONTROLLED SCENARIO

Compare operator time-to-diagnose and time-to-recover using blinded scenarios.

EVIDENCE PACKAGE

Preserve exact versions, architecture, configuration, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

Current decision posture

Reopen the cloud-native assurance and mixed-vendor profile. Mist support for selected AOS-CX switches changes the candidate boundary. Current versions require lab proof on provisioning, assurance, NAC, upgrade, failure, and rollback before any portfolio preference is asserted.

CURRENT RANKING: NOT REISSUED / CONTROLLED LAB REQUIRED

CURRENT OFFICIAL-SOURCE REGISTER / CHECKED 2026-09-17

Cisco Catalyst Center	E2 / OFFICIAL	
Cisco Catalyst Center 3.3.1 release notes		www.cisco.com
HPE Aruba Networking Central	E2 / OFFICIAL	
HPE Aruba Networking Central release notes		arubanetworking.hpe.com
Juniper Mist	E2 / OFFICIAL	
Juniper Mist Sep 9 2026 update		www.juniper.net
Fortinet Campus and Branch	E2 / OFFICIAL	
FortiManager 7.6.7 release notes		docs.fortinet.com

REVISION LINEAGE

1.2.0-candidate / 2026-09-17 - current-source evidence refresh; no lab rescore issued.

1.1.0-candidate / 2026-07-24 - baseline benchmark using the final comparative evaluation system.

Trigger earlier review after major release, acquisition, licensing change, critical vulnerability, material outage, failed PoC, or workload change.

BASELINE ANALYTIC RECORD CONTINUES ON PAGE 11 FOR TRACEABILITY.

ATOMIC CRITERIA MODEL

WEIGHTED CAPABILITY WITH EXPLICIT MINIMUM EVIDENCE

ID	CRITERION	WEIGHT	GATE	MINIMUM EVIDENCE
C01	Campus wired, wireless, policy, assurance, and lifecycle coverage	5	YES	Feature documentation, APIs, configuration exports, and scenario demonstration.
C02	Security and trust architecture	5	YES	Threat model, identity flows, RBAC tests, audit records, and security configuration.
C03	Governance and approval controls	5	YES	Approval workflow, policy controls, separation-of-duties tests, and decision records.
C04	Inventory, topology, intent, configuration, and telemetry integrity	5	YES	Backup, restore, rollback, drift, and corruption-recovery evidence.
C05	Automation and API quality	4	NO	API specifications, authentication tests, rate limits, event samples, and automation runs.
C06	Integration ecosystem	4	NO	Supported integrations, connector tests, failure behavior, and lifecycle ownership.
C07	Observability and evidence	4	YES	Logs, traces, metrics, reports, export tests, and evidence-retention controls.
C08	Site, device, client, event, and telemetry scale	4	NO	Controlled load tests, topology scale, saturation behavior, and capacity model.
C09	Resilience and continuity	5	YES	Failure injection, HA tests, recovery timing, degraded-mode operation, and runbooks.
C10	Network-operator and service-desk cognition	3	NO	Task observation, error-rate measures, navigation tests, and operator interviews.
C11	Change safety and rollback	5	YES	Plan or preview output, canary tests, rollback execution, and post-change verification.
C12	Extensibility and programmability	3	NO	Plugin, module, provider, workflow, or code-extension tests and upgrade impact analysis.
C13	Cloud, on-premises, hybrid, and disconnected management options	3	NO	Deployment architecture, data-flow mapping, residency evidence, and offline tests.
C14	Lifecycle and upgrade discipline	4	NO	Release notes, upgrade test, compatibility matrix, support policy, and migration plan.
C15	Economics and cost predictability	3	NO	Bill-of-materials, licensing model, staffing estimates, metering, and sensitivity analysis.
C16	Portability and exit	4	YES	Export tests, format analysis, replacement workflow, and decommission evidence.
C17	Vendor and ecosystem risk	3	NO	License analysis, roadmap evidence, bus-factor indicators, and dependency inventory.
C18	Assurance confidence	5	YES	Source provenance, lab artifacts, production evidence, independent replication, and review date.

SCORE SCALE

0 absent; 1 materially deficient; 2 partial; 3 adequate with limits; 4 strong; 5 leading for the defined profile. Scores remain provisional until the required evidence grade is achieved.

RAW CAPABILITY SCORECARD

DOCUMENTED CAPABILITY BEFORE EVIDENCE DISCOUNT

CRITERION	CATALYST CENTER	ARUBA CENTRAL	MIST	FORTINET
C01 Campus wired, wireless, policy, assurance, and lifecycle coverage	4.8	4.7	4.6	4.3
C02 Security and trust architecture	4.4	4.4	4.4	4.5
C03 Governance and approval controls	4.3	4.2	4.1	4.1
C04 Inventory, topology, intent, configuration, and telemetry integrity	4.4	4.3	4.2	4.1
C05 Automation and API quality	4.6	4.5	4.5	4.4
C06 Integration ecosystem	4.5	4.4	4.4	4.4
C07 Observability and evidence	4.7	4.6	4.8	4.2
C08 Site, device, client, event, and telemetry scale	4.7	4.5	4.6	4.4
C09 Resilience and continuity	4.4	4.3	4.4	4.2
C10 Network-operator and service-desk cognition	4.4	4.4	4.7	4.0
C11 Change safety and rollback	4.4	4.3	4.2	4.2
C12 Extensibility and programmability	4.3	4.2	4.2	4.0
C13 Cloud, on-premises, hybrid, and disconnected management options	3.8	4.3	3.0	4.1
C14 Lifecycle and upgrade discipline	4.1	4.0	4.3	3.9
C15 Economics and cost predictability	3.3	3.6	3.5	4.1
C16 Portability and exit	3.5	3.7	3.4	3.6
C17 Vendor and ecosystem risk	3.6	3.7	3.5	3.6
C18 Assurance confidence	3.5	3.5	3.4	3.4

WEIGHTED BASELINE / 100

Catalyst Cen	Aruba Centra	Mist	Fortinet
84.8	84.3	83.0	81.9

EVIDENCE-ADJUSTED SCORECARD

CAPABILITY DISCOUNTED BY CURRENT EVIDENCE CONFIDENCE

CANDIDATE	RAW	EVIDENCE CONFIDENCE	ADJUSTED	CURRENT STATE
Catalyst Center	84.8	58.2%	49.9	CONTROLLED LAB PENDING
Aruba Central	84.3	58.2%	49.6	CONTROLLED LAB PENDING
Mist	83.0	58.2%	48.8	CONTROLLED LAB PENDING
Fortinet	81.9	58.2%	48.1	CONTROLLED LAB PENDING

EVIDENCE SCALE

E0 / 0.00
Unsupported or unverified

E1 / 0.38
Vendor assertion or marketing statement

E2 / 0.64
Official documentation, release notes, or source repository

E3 / 0.78
Controlled Mythos laboratory result

E4 / 0.91
Production evidence from the adopting environment

E5 / 1.00
Independent repeatable validation

CURRENT CONFIDENCE BOUNDARY

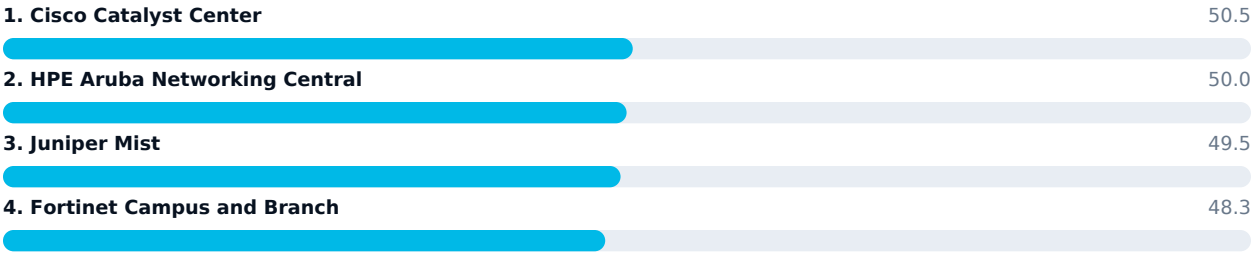
Most candidate criteria are supported at E2: official documentation or source evidence. Environment-specific laboratory, production, and independent evidence remain future gates.

PROFILE-SPECIFIC RANKINGS

EVIDENCE-ADJUSTED SENSITIVITY ANALYSIS

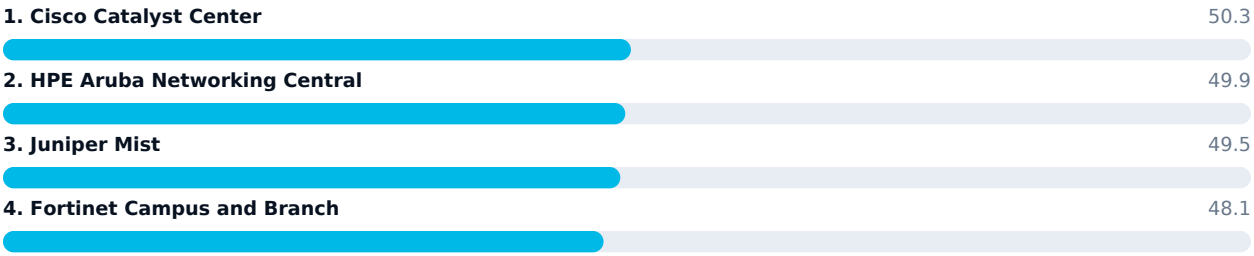
LARGE CISCO CAMPUS

Prioritizes Cisco design, provisioning, SD-Access, assurance, inventory, and enterprise integration.



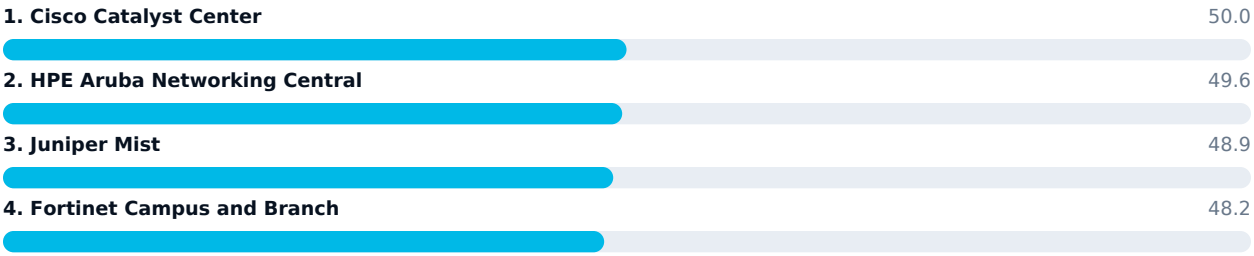
CLOUD-NATIVE EXPERIENCE ASSURANCE

Prioritizes client experience, wired/wireless assurance, API operations, rapid feature delivery, and service cognition.



SECURITY-CONVERGED BRANCH AND CAMPUS

Prioritizes security integration, branch, SD-WAN adjacency, central lifecycle, economics, and appliance breadth.



RANK SENSITIVITY

Small score differences are not decision certainty. Treat near-ties as a requirement for deeper PoC, commercial, migration, and operating-model evidence.

CANDIDATE DEEP DIVE / 01

CISCO CATALYST CENTER

OPERATING MODEL

Cisco campus management, design, provisioning, policy, automation, inventory, topology, and assurance platform.

DOCUMENTED STRENGTHS

- Deep Cisco campus integration.
- Rich inventory, design, provisioning, assurance, and SD-Access workflows.
- Broad API and platform integration.
- Strong fit for large Cisco wired and wireless estates.

CAUTIONS AND PROOF OBLIGATIONS

- Platform and appliance lifecycle can be substantial.
- Best value depends on Cisco-centric infrastructure and licensing.
- Assurance quality depends on telemetry, design, and operational discipline.

BASELINE SCORE PROFILE

Raw documented capability 84.8

Evidence-adjusted capability 49.9

GATE DISPOSITION

PASS - large Cisco campus profile | CONDITIONAL - platform lifecycle and licensing

CANDIDATE DEEP DIVE / 02

HPE ARUBA NETWORKING CENTRAL

OPERATING MODEL

Cloud or on-premises network-management and operations platform for Aruba wired, wireless, gateway, branch, and policy ecosystems.

DOCUMENTED STRENGTHS

- Strong Aruba wired and wireless integration.
- Cloud and on-premises options.
- Broad campus, branch, policy, and operations coverage.
- Validated solution guidance and mature enterprise networking heritage.

CAUTIONS AND PROOF OBLIGATIONS

- Classic and new Central transitions require architecture and feature validation.
- Licensing and feature parity must be verified per deployment model.
- Cross-version and device-family operations require disciplined testing.

BASELINE SCORE PROFILE

Raw documented capability 84.3

Evidence-adjusted capability 49.6

GATE DISPOSITION

PASS - Aruba campus and deployment-flexibility profile | CONDITIONAL - platform-generation parity

CANDIDATE DEEP DIVE / 03

JUNIPER MIST

OPERATING MODEL

Cloud-native AI-driven wired, wireless, access-assurance, location, and operations platform.

DOCUMENTED STRENGTHS

- Strong experience and assurance orientation.
- Cloud-native operational model and Marvis-assisted troubleshooting.
- Wired, wireless, and access assurance convergence.
- Modern APIs and microservices delivery.

CAUTIONS AND PROOF OBLIGATIONS

- Cloud-service dependency and subscription model are material.
- Best fit depends on Juniper and Mist ecosystem alignment.
- AI-driven recommendations require transparent operational validation.

BASELINE SCORE PROFILE

Raw documented capability 83.0

Evidence-adjusted capability 48.8

GATE DISPOSITION

PASS - cloud-native experience-assurance profile | CONDITIONAL - cloud dependency and ecosystem fit

CANDIDATE DEEP DIVE / 04

FORTINET CAMPUS AND BRANCH

OPERATING MODEL

FortiGate, FortiSwitch, FortiAP, FortiManager, and Security Fabric operating model for converged campus, branch, and security.

DOCUMENTED STRENGTHS

- Converged security and network operations.
- Broad branch and campus appliance ecosystem.
- Potential commercial and operational consolidation.
- Central management and automation within the Security Fabric.

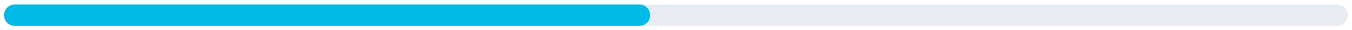
CAUTIONS AND PROOF OBLIGATIONS

- Campus assurance depth must be tested against requirements.
- Ecosystem coupling and firmware discipline are central risks.
- Network and security ownership boundaries must be designed deliberately.

BASELINE SCORE PROFILE

Raw documented capability 81.9

Evidence-adjusted capability 48.1



GATE DISPOSITION

PASS - security-converged branch profile | CONDITIONAL - campus assurance and firmware governance

ARCHITECTURE AND INTEGRATION FIT

THE PRODUCT IS ONLY ONE PART OF THE OPERATING SYSTEM

REFERENCE OPERATING LAYERS

01 Intent, site, policy, and inventory authority

02 Wired, wireless, gateway, access, and identity infrastructure

03 Telemetry, assurance, analytics, and incident workflows

04 Provisioning, automation, software, and configuration lifecycle

05 Evidence, recovery, integration, and service management

INTEGRATION BOUNDARIES

- Identity, NAC, PKI, DHCP, DNS, and IPAM
- ITSM, collaboration, and incident command
- SIEM, NDR, endpoint, and security platforms
- Cloud, WAN, SD-WAN, branch, and datacenter
- APIs, event streaming, automation, and source of truth

ARCHITECTURE GATE

Every external dependency requires an owner, identity boundary, failure mode, evidence path, version policy, recovery procedure, and exit strategy.

SECURITY, OPERATIONS, LIFECYCLE, ECONOMICS, AND EXIT

CROSS-DOMAIN DECISION RECORD

SECURITY AND AUTHORITY

Identity, credentials, secrets, roles, privileged actions, signing, approvals, isolation, audit, and incident response must be tested as an integrated system.

RELIABILITY AND RECOVERY

Control-plane, data-plane, dependency, region, database, queue, network, and operator failures require defined degraded modes and recovery objectives.

CHANGE AND LIFECYCLE

Version, compatibility, migration, canary, rollback, content, plugin, provider, firmware, and decommission procedures are part of product fitness.

ECONOMICS AND CAPACITY

Licenses, metering, data, compute, hardware, storage, support, staffing, training, integration, migration, and opportunity cost require sensitivity analysis.

SOVEREIGNTY AND PRIVACY

Data location, support access, telemetry, subprocessors, encryption, key control, disconnected operation, and legal obligations must align with policy.

PORTABILITY AND EXIT

Configuration, policy, data, state, evidence, workflows, identities, and operational knowledge must be exportable and reconstructable.

DECISION OWNERSHIP

Architecture, security, operations, finance, procurement, legal, data, and service owners jointly accept the final profile, evidence, residual risk, and exit obligations.

RISK AND FAILURE REGISTER

SCENARIOS THAT CAN INVALIDATE A FEATURE-BASED SELECTION

ID	SEVERITY	FAILURE SCENARIO	REQUIRED TREATMENT
R-01	CRITICAL	A polished dashboard conceals incomplete telemetry or weak root-cause accuracy.	PREVENT / DETECT / RECOVER / EVIDENCE
R-02	HIGH	Cloud or management loss removes essential operations without a degraded mode.	PREVENT / DETECT / RECOVER / EVIDENCE
R-03	HIGH	Templates and intent drift from deployed device configuration.	PREVENT / DETECT / RECOVER / EVIDENCE
R-04	HIGH	Automated remediation changes service without sufficient approval or rollback.	PREVENT / DETECT / RECOVER / EVIDENCE
R-05	MEDIUM	Licensing or hardware constraints invalidate the target operating model.	PREVENT / DETECT / RECOVER / EVIDENCE
R-06	MEDIUM	Platform migration breaks inventory, policy, maps, telemetry, historical baselines, and runbooks.	PREVENT / DETECT / RECOVER / EVIDENCE

RISK RULE

A candidate with an unresolved critical failure path cannot advance because optional capability, market position, or commercial discount cannot compensate for missing safety or recovery evidence.

CONTROLLED PROOF-OF-CAPABILITY PLAN

REPEATABLE SCENARIOS, INSTRUMENTATION, AND ACCEPTANCE

TEST 01

Deploy a representative wired and wireless site hierarchy with users, guests, IoT, voice, and critical applications.

TEST 02

Measure discovery, provisioning, configuration compliance, telemetry completeness, and client-assurance accuracy.

TEST 03

Inject DHCP, DNS, RADIUS, RF, roaming, PoE, VLAN, routing, WAN, and certificate failures.

TEST 04

Exercise software upgrade, rollback, device replacement, site outage, and management-plane loss.

TEST 05

Validate API, webhooks, ITSM, source-of-truth, SIEM, and automated-remediation controls.

TEST 06

Compare operator time-to-diagnose and time-to-recover using blinded scenarios.

EVIDENCE PACKAGE

Preserve versions, architecture, configuration, data set, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

CONDITIONAL RECOMMENDATION AND REVALIDATION

DECISION RECORD, ACCEPTANCE, AND NEXT EVIDENCE

RECOMMENDATION POSITION

- Shortlist by installed estate and target operating model.
- Use identical client and service-failure scenarios in a controlled lab.
- Measure actual operator cognition and recovery time.
- Revalidate quarterly and before major cloud, licensing, appliance, or controller transitions.

CANDIDATE	ADJ. SCORE	GATE POSITION	LAB STATE
Catalyst Center	49.9	PASS - large Cisco campus profile; CONDITIONAL - platform lifecycle and licensing	PENDING
Aruba Central	49.6	PASS - Aruba campus and deployment-flexibility profile; CONDITIONAL - platform-generation parity	PENDING
Mist	48.8	PASS - cloud-native experience-assurance profile; CONDITIONAL - cloud dependency and ecosystem fit	PENDING
Fortinet	48.1	PASS - security-converged branch profile; CONDITIONAL - campus assurance and firmware governance	PENDING

REVALIDATION SCHEDULE

Review no later than 2026-10-24. Review earlier after a major release, commercial change, critical vulnerability, material outage, architecture transition, failed acceptance test, or change to the target workload profile.

SOURCE AUTHORITY AND REVISION

CURRENT EVIDENCE SNAPSHOT AND PUBLICATION HISTORY

SOURCE ID	PUBLISHER	TITLE	CHECKED
NIST-CSF-20	NIST	Cybersecurity Framework 2.0	2026-07-24
NIST-SP800-53	NIST	Security and Privacy Controls for Information Systems and Organizations	2026-07-24
CATALYST-CENTER	Cisco	Cisco Catalyst Center User Guide	2026-07-24
CATALYST-ASSURANCE	Cisco	Cisco Catalyst Assurance User Guide	2026-07-24
ARUBA-CENTRAL	HPE Aruba Networking	HPE Aruba Networking Central Documentation	2026-07-24
ARUBA-VSG	HPE Aruba Networking	Aruba Validated Solution Guides	2026-07-24
MIST-WIRED	HPE Juniper Networking	Juniper Mist Wired Assurance Guide	2026-07-24
FORTI-CAMPUS	Fortinet	FortiManager 8.0 Operations	2026-07-24

SOURCE POSITION

Official documentation and source repositories support the current capability model. Source records preserve publisher, title, URL, purpose, and review date in the machine-readable authority register. Commercial terms, performance, and environment-specific behavior remain unverified until controlled proof.

REVISION HISTORY

1.1.0-candidate / 2026-07-24 - permanent-publication rebuild using the final benchmark system, profile-specific rankings, evidence adjustment, mandatory gates, and controlled PoC requirements.

COMPARATIVE EVALUATION / EVIDENCE REFRESH

CMP-007

IaC and Control Planes

Infrastructure as code, declarative delivery and platform control planes

CURRENT EVIDENCE SNAPSHOT

Official product sources refreshed through 2026-09-17

Controlled Mythos laboratory rerun: NOT ASSERTED

BENCHMARK DOCTRINE

Gates before scores. Evidence before certainty. Profile fit before universal ranking.

VERSION 1.2.0-candidate

CANDIDATE COMPARATIVE EVALUATION / PUBLIC

IaC and Control Planes

Infrastructure as code, declarative delivery and platform control planes

DOCUMENT ID
CMP-007
VERSION
1.2.0-candidate
STATUS
Candidate Comparative Evaluation
PUBLICATION DATE
2026-09-17
SCHEDULED REVIEW
2026-12-16
CANONICAL ROUTE
/library/evaluations/cmp-007/

4
CANDIDATES

18
ATOMIC CRITERIA

9
MANDATORY GATES

3
WORKLOAD PROFILES

E2
CURRENT MAX EVIDENCE

CURRENT DECISION BOUNDARY

The July benchmark is a near-tie, not a universal result. Current Terraform, Pulumi, Crossplane, and OpenTofu releases change lifecycle and control-plane assumptions. Reproduce state, policy, drift, rollback, provider failure, migration, and exit under the intended operating model.

NO CURRENT UNIVERSAL WINNER IS PUBLISHED FROM THIS REFRESH.

July 24 baseline scores are preserved as lineage and sensitivity evidence, not silently reissued as September laboratory results.

Terraform / HCP Terraform

Terraform 1.16.3 / 2026-09-16

REFRESH TRIGGER

Stable 1.16.3 is current; 1.17 is beta. HCP Terraform continues policy/Stacks evolution.

OpenTofu

v1.12.0 / 2026-05-14

REFRESH TRIGGER

1.12 adds dynamic prevent_destroy, checksum improvements, output changes, and lifecycle controls; WinRM deprecation affects some estates.

Pulumi

v3.263.0 / 2026-09-15

REFRESH TRIGGER

Current line adds continued policy, Terraform-state, cross-language, and HCL integration improvements.

Crossplane

v2.4 / 2026-08-20

REFRESH TRIGGER

v2.4 is a maintained release in the broader v2 control-plane model; upgrade and composition behavior remain material.

FINAL BENCHMARK SYSTEM / DECISION ARCHITECTURE

01 SCOPE

Freeze decision, candidates, versions and exclusions.

02 GATES

Test non-compensable trust, state, recovery, change and exit.

03 SCORES

Measure atomic capability with approved profile weights.

04 CONFIDENCE

Discount by source grade, freshness, repeatability and relevance.

05 PROOF

Run normal, failure, migration, recovery and operator scenarios.

06 DECISION

Record conditional fit, residual risk, owner and revalidation trigger.

EVIDENCE CONFIDENCE / MODIFIERS

E0	E1	E2	E3	E4	E5
0.00	0.38	0.64	0.78	0.91	1.00
UNSUPPORTED	VENDOR ASSERTION	OFFICIAL SOURCE	MYTHOS LAB	PRODUCTION	INDEPENDENT

THIS EDITION

Current official product and release sources have been refreshed to E2. No E3 controlled Mythos lab rerun is claimed. Therefore current profile rankings remain intentionally unissued until the proof plan is executed.

What changed since the July benchmark snapshot

Changes below are sourced from current official release documentation. They identify revalidation triggers; they are not translated into new numeric scores without controlled proof.

Terraform / HCP Terraform

Terraform 1.16.3 / 2026-09-16

MATERIAL DELTA

Stable 1.16.3 is current; 1.17 is beta. HCP Terraform continues policy/Stacks evolution.

CRITERIA TOUCHED

C01 / C03 / C04 / C05 / C11 / C14

MODERATE REVALIDATION

OpenTofu

v1.12.0 / 2026-05-14

MATERIAL DELTA

1.12 adds dynamic prevent_destroy, checksum improvements, output changes, and lifecycle controls; WinRM deprecation affects some estates.

CRITERIA TOUCHED

C04 / C11 / C14 / C16

MODERATE REVALIDATION

Pulumi

v3.263.0 / 2026-09-15

MATERIAL DELTA

Current line adds continued policy, Terraform-state, cross-language, and HCL integration improvements.

CRITERIA TOUCHED

C03 / C04 / C05 / C06 / C12 / C14

MODERATE REVALIDATION

Crossplane

v2.4 / 2026-08-20

MATERIAL DELTA

v2.4 is a maintained release in the broader v2 control-plane model; upgrade and composition behavior remain material.

CRITERIA TOUCHED

C01 / C04 / C08 / C09 / C11 / C14

HIGH REVALIDATION

Mandatory gates remain non-compensable

No current release is marked PASS solely from documentation. E2 means the official source supports the capability path; LAB and RETEST identify proof still required. HOLD blocks a current decision.

MANDATORY GATE	TERRAFORM / HCP	OPENTOFU	PULUMI	CROSSPLANE
C01 Functional coverage	RETEST	E2	E2	RETEST
C02 Security / trust	E2	E2	E2	E2
C03 Governance / approval	RETEST	E2	RETEST	E2
C04 State integrity	RETEST	RETEST	RETEST	RETEST
C07 Observability / evidence	E2	E2	E2	E2
C09 Resilience / continuity	LAB	LAB	LAB	RETEST
C11 Change safety / rollback	RETEST	RETEST	LAB	RETEST
C16 Portability / exit	LAB	RETEST	LAB	LAB
C18 Assurance confidence	HOLD	HOLD	HOLD	HOLD

LEGEND

E2 official-source support only

LAB controlled proof required

RETEST release delta touches this gate

HOLD decision gate remains closed

Historical profile sensitivity - preserved, not reissued

Values below are the 2026-07-24 evidence-adjusted baseline. They show sensitivity to operating-model weights. The September refresh does not recompute rank because E3 lab proof has not been reproduced.

ENTERPRISE DECLARATIVE DELIVERY

BASELINE SPREAD 1.7 PTS

OPEN SOVEREIGN IAC

BASELINE SPREAD 1.4 PTS

INTERNAL DEVELOPER PLATFORM CONTROL PLANE

BASELINE SPREAD 1.6 PTS

CURRENT RANK STATUS: WITHHELD PENDING CONTROLLED PROOF.

Evidence confidence is separate from capability

E5 — Independent repeatable validation
confidence modifier 1.00

E4 — Production evidence in adopting environment
confidence modifier 0.91

E3 — Controlled Mythos laboratory result
confidence modifier 0.78

E2 — Official documentation / release / source
confidence modifier 0.64

E1 — Vendor assertion or marketing
confidence modifier 0.38

E0 — Unsupported or unverified
confidence modifier 0.00

CURRENT EVIDENCE STATE

Terraform / HCP Terraform — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

OpenTofu — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

Pulumi — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

Crossplane — E2 REFRESHED
Historical adjusted confidence: 58.2%
E3 LAB: NOT REPRODUCED

ASSURANCE RULE

A current release note can change capability context, but it cannot raise assurance beyond E2. Current recommendation strength is therefore bounded until the test plan produces reproducible artifacts.

Lifecycle risk is evaluated independently of features

The benchmark models migration, upgrade, compatibility, support, staffing, data/state export, downtime and exit. Current release changes below are explicit proof triggers.

Terraform / HCP Terraform

MODERATE REVALIDATION

Stable 1.16.3 is current; 1.17 is beta. HCP Terraform continues policy/Stacks evolution.

RETEST: C01 / C03 / C04 / C05 / C11 / C14

OpenTofu

MODERATE REVALIDATION

1.12 adds dynamic prevent_destroy, checksum improvements, output changes, and lifecycle controls; WinRM deprecation affects some estates.

RETEST: C04 / C11 / C14 / C16

Pulumi

MODERATE REVALIDATION

Current line adds continued policy, Terraform-state, cross-language, and HCL integration improvements.

RETEST: C03 / C04 / C05 / C06 / C12 / C14

Crossplane

HIGH REVALIDATION

v2.4 is a maintained release in the broader v2 control-plane model; upgrade and composition behavior remain material.

RETEST: C01 / C04 / C08 / C09 / C11 / C14

IRREVERSIBLE COMMITMENT GATE

Before migration or procurement lock-in, prove rollback, state portability, operational reconstruction, and supplier-exit assumptions.

Controlled proof is the next decision-producing step

REPRODUCTION STATE / NOT EXECUTED IN THIS EVIDENCE-REFRESH RELEASE

01

CONTROLLED SCENARIO

Provision representative cloud, network, identity, data, and SaaS resources with dependencies.

02

CONTROLLED SCENARIO

Exercise import, drift, replacement, partial failure, retry, refresh, taint or repair, and state recovery.

03

CONTROLLED SCENARIO

Validate policy, approval, secret isolation, provider signing, and execution-environment controls.

04

CONTROLLED SCENARIO

Measure plan, apply, reconcile, provider, and state performance at representative scale.

05

CONTROLLED SCENARIO

Upgrade providers, modules, languages, compositions, and the control plane.

06

CONTROLLED SCENARIO

Migrate a controlled stack between approaches and verify data, state, policy, and rollback.

EVIDENCE PACKAGE

Preserve exact versions, architecture, configuration, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

Current decision posture

The July benchmark is a near-tie, not a universal result. Current Terraform, Pulumi, Crossplane, and OpenTofu releases change lifecycle and control-plane assumptions. Reproduce state, policy, drift, rollback, provider failure, migration, and exit under the intended operating model.

CURRENT RANKING: NOT REISSUED / CONTROLLED LAB REQUIRED

CURRENT OFFICIAL-SOURCE REGISTER / CHECKED 2026-09-17

Terraform / HCP Terraform	E2 / OFFICIAL	
Terraform releases		github.com
OpenTofu	E2 / OFFICIAL	
OpenTofu 1.12.0 release		opentofu.org
Pulumi	E2 / OFFICIAL	
Pulumi versions and changelog		www.pulumi.com
Crossplane	E2 / OFFICIAL	
Crossplane releases		github.com

REVISION LINEAGE

1.2.0-candidate / 2026-09-17 - current-source evidence refresh; no lab rescore issued.

1.1.0-candidate / 2026-07-24 - baseline benchmark using the final comparative evaluation system.

Trigger earlier review after major release, acquisition, licensing change, critical vulnerability, material outage, failed PoC, or workload change.

BASELINE ANALYTIC RECORD CONTINUES ON PAGE 11 FOR TRACEABILITY.

ATOMIC CRITERIA MODEL

WEIGHTED CAPABILITY WITH EXPLICIT MINIMUM EVIDENCE

ID	CRITERION	WEIGHT	GATE	MINIMUM EVIDENCE
C01	Infrastructure lifecycle and control-plane coverage	5	YES	Feature documentation, APIs, configuration exports, and scenario demonstration.
C02	Identity, policy, secret, provider, and execution trust architecture	5	YES	Threat model, identity flows, RBAC tests, audit records, and security configuration.
C03	Governance and approval controls	5	YES	Approval workflow, policy controls, separation-of-duties tests, and decision records.
C04	State, plan, dependency, and reconciliation integrity	5	YES	Backup, restore, rollback, drift, and corruption-recovery evidence.
C05	Automation and API quality	4	NO	API specifications, authentication tests, rate limits, event samples, and automation runs.
C06	Integration ecosystem	4	NO	Supported integrations, connector tests, failure behavior, and lifecycle ownership.
C07	Observability and evidence	4	YES	Logs, traces, metrics, reports, export tests, and evidence-retention controls.
C08	Resource, stack, workspace, provider, and control-loop scale	4	NO	Controlled load tests, topology scale, saturation behavior, and capacity model.
C09	Resilience and continuity	5	YES	Failure injection, HA tests, recovery timing, degraded-mode operation, and runbooks.
C10	Usability and operator cognition	3	NO	Task observation, error-rate measures, navigation tests, and operator interviews.
C11	Plan, approval, policy, canary, rollback, and reconciliation safety	5	YES	Plan or preview output, canary tests, rollback execution, and post-change verification.
C12	Extensibility and programmability	3	NO	Plugin, module, provider, workflow, or code-extension tests and upgrade impact analysis.
C13	Deployment and sovereignty options	3	NO	Deployment architecture, data-flow mapping, residency evidence, and offline tests.
C14	Lifecycle and upgrade discipline	4	NO	Release notes, upgrade test, compatibility matrix, support policy, and migration plan.
C15	Economics and cost predictability	3	NO	Bill-of-materials, licensing model, staffing estimates, metering, and sensitivity analysis.
C16	Portability and exit	4	YES	Export tests, format analysis, replacement workflow, and decommission evidence.
C17	Vendor and ecosystem risk	3	NO	License analysis, roadmap evidence, bus-factor indicators, and dependency inventory.
C18	Assurance confidence	5	YES	Source provenance, lab artifacts, production evidence, independent replication, and review date.

SCORE SCALE

0 absent; 1 materially deficient; 2 partial; 3 adequate with limits; 4 strong; 5 leading for the defined profile. Scores remain provisional until the required evidence grade is achieved.

RAW CAPABILITY SCORECARD

DOCUMENTED CAPABILITY BEFORE EVIDENCE DISCOUNT

CRITERION	TERRAFORM	OPENTOFU	PULUMI	CROSSPLANE
C01 Infrastructure lifecycle and control-plane coverage	4.8	4.6	4.7	4.5
C02 Identity, policy, secret, provider, and execution trust architecture	4.3	4.3	4.3	4.4
C03 Governance and approval controls	4.4	4.1	4.2	4.3
C04 State, plan, dependency, and reconciliation integrity	4.5	4.6	4.4	4.3
C05 Automation and API quality	4.7	4.6	4.7	4.5
C06 Integration ecosystem	4.7	4.5	4.6	4.5
C07 Observability and evidence	4.4	4.2	4.4	4.4
C08 Resource, stack, workspace, provider, and control-loop scale	4.7	4.5	4.5	4.5
C09 Resilience and continuity	4.3	4.1	4.2	4.3
C10 Usability and operator cognition	4.4	4.2	4.3	4.0
C11 Plan, approval, policy, canary, rollback, and reconciliation safety	4.6	4.5	4.5	4.2
C12 Extensibility and programmability	4.6	4.4	4.8	4.8
C13 Deployment and sovereignty options	4.0	4.7	4.5	4.4
C14 Lifecycle and upgrade discipline	4.4	4.1	4.2	4.0
C15 Economics and cost predictability	3.7	4.3	3.8	3.7
C16 Portability and exit	4.1	4.6	4.2	4.0
C17 Vendor and ecosystem risk	3.7	4.0	3.7	3.7
C18 Assurance confidence	3.6	3.5	3.5	3.4

WEIGHTED BASELINE / 100

Terraform	OpenTofu	Pulumi	Crossplane
86.9	86.3	86.1	84.4

EVIDENCE-ADJUSTED SCORECARD

CAPABILITY DISCOUNTED BY CURRENT EVIDENCE CONFIDENCE

CANDIDATE	RAW	EVIDENCE CONFIDENCE	ADJUSTED	CURRENT STATE
Terraform	86.9	58.2%	51.3	CONTROLLED LAB PENDING
OpenTofu	86.3	58.2%	50.8	CONTROLLED LAB PENDING
Pulumi	86.1	58.2%	50.8	CONTROLLED LAB PENDING
Crossplane	84.4	58.2%	49.7	CONTROLLED LAB PENDING

EVIDENCE SCALE

E0 / 0.00
Unsupported or unverified

E1 / 0.38
Vendor assertion or marketing statement

E2 / 0.64
Official documentation, release notes, or source repository

E3 / 0.78
Controlled Mythos laboratory result

E4 / 0.91
Production evidence from the adopting environment

E5 / 1.00
Independent repeatable validation

CURRENT CONFIDENCE BOUNDARY

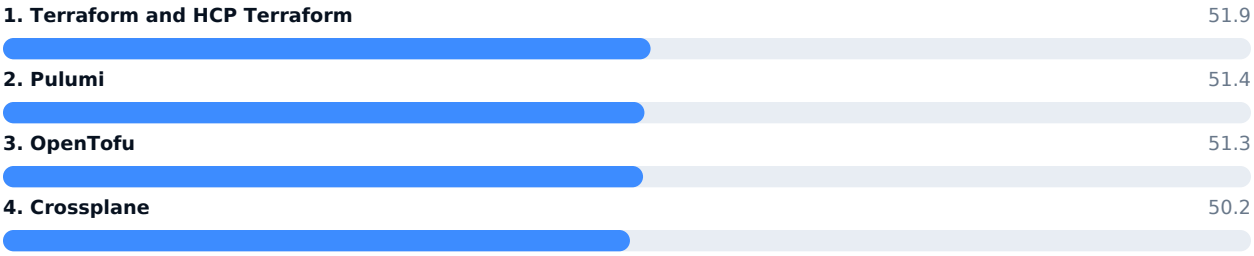
Most candidate criteria are supported at E2: official documentation or source evidence. Environment-specific laboratory, production, and independent evidence remain future gates.

PROFILE-SPECIFIC RANKINGS

EVIDENCE-ADJUSTED SENSITIVITY ANALYSIS

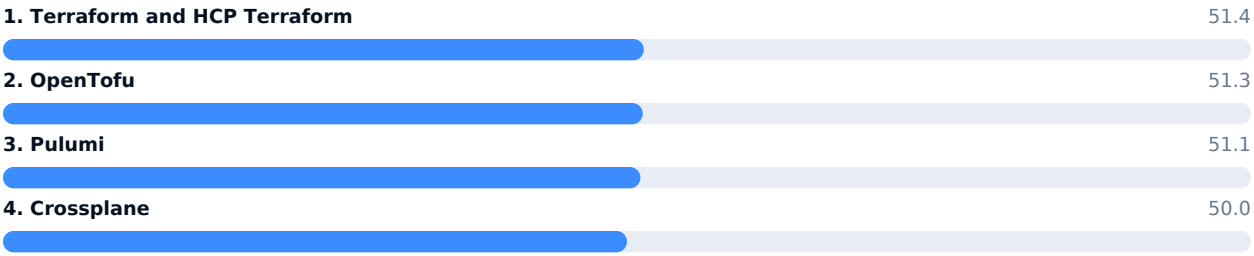
ENTERPRISE DECLARATIVE DELIVERY

Prioritizes broad provider coverage, plans, approvals, policy, collaboration, audit, and established operating patterns.



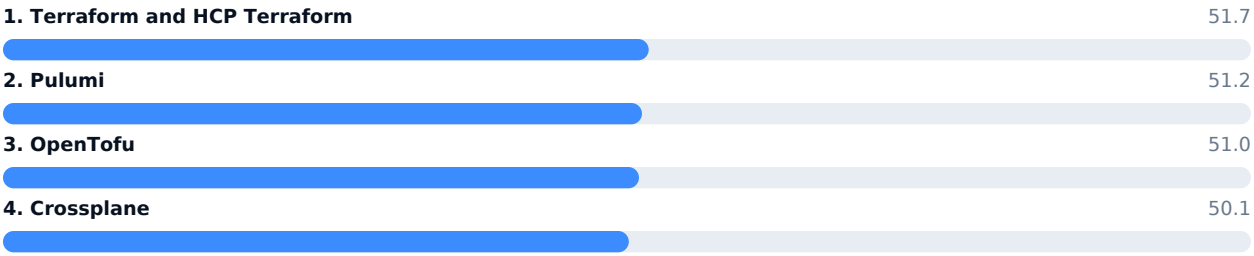
OPEN SOVEREIGN IAC

Prioritizes open governance, compatibility, state control, encryption, exportability, and self-managed operation.



INTERNAL DEVELOPER PLATFORM CONTROL PLANE

Prioritizes reusable platform APIs, continuous reconciliation, self-service, policy, Kubernetes integration, and lifecycle.



RANK SENSITIVITY

Small score differences are not decision certainty. Treat near-ties as a requirement for deeper PoC, commercial, migration, and operating-model evidence.

CANDIDATE DEEP DIVE / 01

TERRAFORM AND HCP TERRAFORM

OPERATING MODEL

Declarative infrastructure provisioning using HCL with local, remote, and managed workflow options.

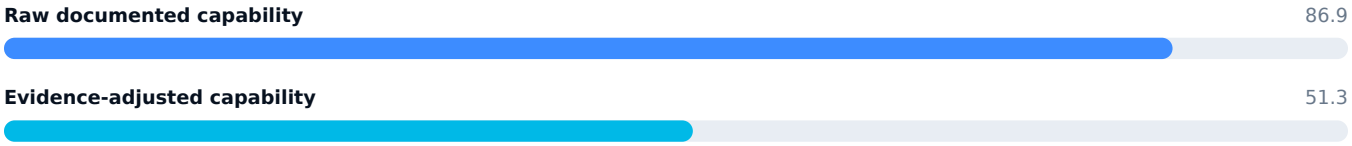
DOCUMENTED STRENGTHS

- Large provider and module ecosystem.
- Widely understood declarative workflow and mature operational patterns.
- Strong plan-and-apply model with managed collaboration options.
- Broad cloud, network, SaaS, and on-premises coverage.

CAUTIONS AND PROOF OBLIGATIONS

- State and provider trust require strict protection.
- Licensing and commercial-platform boundaries must be accepted.
- Rollback is not universally transactional and must be engineered by resource type.

BASELINE SCORE PROFILE



GATE DISPOSITION

PASS - enterprise declarative delivery | CONDITIONAL - licensing, state, and provider trust

CANDIDATE DEEP DIVE / 02

OPENTOFU

OPERATING MODEL

Community-governed open-source Terraform-compatible infrastructure-as-code engine with state and plan encryption capabilities.

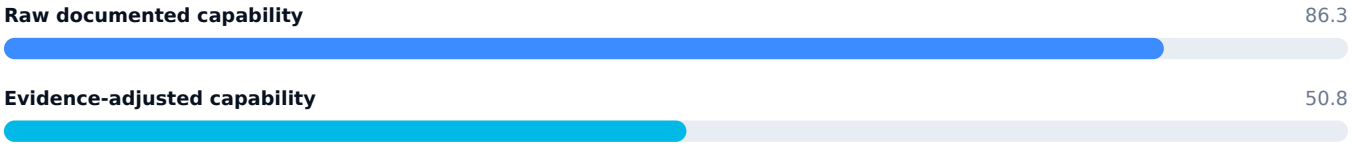
DOCUMENTED STRENGTHS

- Strong compatibility with Terraform workflows and providers.
- Open governance and permissive open-source posture.
- Native state and plan encryption options.
- Lower migration friction for many existing HCL estates.

CAUTIONS AND PROOF OBLIGATIONS

- Commercial ecosystem and enterprise platform depth differ from HCP Terraform.
- Compatibility and provider behavior require release-specific validation.
- Operational tooling, policy, registry, and support model must be selected explicitly.

BASELINE SCORE PROFILE



GATE DISPOSITION

PASS - open sovereign IaC | CONDITIONAL - enterprise workflow and support model

CANDIDATE DEEP DIVE / 03

PULUMI

OPERATING MODEL

Infrastructure as code using general-purpose languages, component resources, managed or DIY state, and policy capabilities.

DOCUMENTED STRENGTHS

- General-purpose languages improve abstraction and reuse for software-oriented teams.
- Strong multi-language SDK and component model.
- Managed and self-managed backend choices.
- Rich testing and software-engineering integration potential.

CAUTIONS AND PROOF OBLIGATIONS

- Language flexibility can increase inconsistency and dependency risk.
- State and provider behavior still require infrastructure discipline.
- Team skill, package lifecycle, and code review standards are critical.

BASELINE SCORE PROFILE

Raw documented capability 86.1

Evidence-adjusted capability 50.8

GATE DISPOSITION

PASS - software-engineering-centric IaC | CONDITIONAL - language and dependency governance

CANDIDATE DEEP DIVE / 04

CROSSPLANE

OPERATING MODEL

Kubernetes-native control-plane framework using composite resources, providers, compositions, and continuous reconciliation.

DOCUMENTED STRENGTHS

- Creates organization-specific cloud and infrastructure APIs.
- Continuous reconciliation supports platform control-plane patterns.
- Strong Kubernetes-native integration and composition model.
- Useful for self-service platform engineering and policy-bound abstractions.

CAUTIONS AND PROOF OBLIGATIONS

- Kubernetes becomes a critical control-plane dependency.
- State and failure semantics differ materially from plan-and-apply IaC.
- Provider and composition lifecycle require specialized platform engineering.

BASELINE SCORE PROFILE

Raw documented capability 84.4

Evidence-adjusted capability 49.7

GATE DISPOSITION

PASS - internal developer platform control plane | NOT EQUIVALENT - continuous reconciliation changes the operating model

ARCHITECTURE AND INTEGRATION FIT

THE PRODUCT IS ONLY ONE PART OF THE OPERATING SYSTEM

REFERENCE OPERATING LAYERS

01 Intent and platform-product interfaces

02 Configuration or API definitions and policy

03 Plan-and-apply engine or reconciliation control loop

04 Providers, credentials, state, dependency, and execution environment

05 Evidence, drift, recovery, migration, and retirement

INTEGRATION BOUNDARIES

- Cloud and infrastructure provider APIs
- Git, CI/CD, artifact, and module registries
- Identity, secrets, policy, approval, and signing
- Kubernetes, service catalog, developer portal, and platform APIs
- Cost, observability, ITSM, security, and evidence systems

ARCHITECTURE GATE

Every external dependency requires an owner, identity boundary, failure mode, evidence path, version policy, recovery procedure, and exit strategy.

SECURITY, OPERATIONS, LIFECYCLE, ECONOMICS, AND EXIT

CROSS-DOMAIN DECISION RECORD

SECURITY AND AUTHORITY

Identity, credentials, secrets, roles, privileged actions, signing, approvals, isolation, audit, and incident response must be tested as an integrated system.

RELIABILITY AND RECOVERY

Control-plane, data-plane, dependency, region, database, queue, network, and operator failures require defined degraded modes and recovery objectives.

CHANGE AND LIFECYCLE

Version, compatibility, migration, canary, rollback, content, plugin, provider, firmware, and decommission procedures are part of product fitness.

ECONOMICS AND CAPACITY

Licenses, metering, data, compute, hardware, storage, support, staffing, training, integration, migration, and opportunity cost require sensitivity analysis.

SOVEREIGNTY AND PRIVACY

Data location, support access, telemetry, subprocessors, encryption, key control, disconnected operation, and legal obligations must align with policy.

PORTABILITY AND EXIT

Configuration, policy, data, state, evidence, workflows, identities, and operational knowledge must be exportable and reconstructable.

DECISION OWNERSHIP

Architecture, security, operations, finance, procurement, legal, data, and service owners jointly accept the final profile, evidence, residual risk, and exit obligations.

RISK AND FAILURE REGISTER

SCENARIOS THAT CAN INVALIDATE A FEATURE-BASED SELECTION

ID	SEVERITY	FAILURE SCENARIO	REQUIRED TREATMENT
R-01	CRITICAL	A green plan conceals provider-side behavior or destructive replacement.	PREVENT / DETECT / RECOVER / EVIDENCE
R-02	HIGH	State compromise exposes secrets or permits unauthorized control.	PREVENT / DETECT / RECOVER / EVIDENCE
R-03	HIGH	Provider or module supply-chain failure changes behavior across many environments.	PREVENT / DETECT / RECOVER / EVIDENCE
R-04	HIGH	Continuous reconciliation fights emergency or manual recovery actions.	PREVENT / DETECT / RECOVER / EVIDENCE
R-05	MEDIUM	Abstractions hide cost, topology, failure, or compliance consequences from users.	PREVENT / DETECT / RECOVER / EVIDENCE
R-06	MEDIUM	Migration proves syntax compatibility but loses state lineage, policy, workflows, and evidence.	PREVENT / DETECT / RECOVER / EVIDENCE

RISK RULE

A candidate with an unresolved critical failure path cannot advance because optional capability, market position, or commercial discount cannot compensate for missing safety or recovery evidence.

CONTROLLED PROOF-OF-CAPABILITY PLAN

REPEATABLE SCENARIOS, INSTRUMENTATION, AND ACCEPTANCE

TEST 01

Provision representative cloud, network, identity, data, and SaaS resources with dependencies.

TEST 02

Exercise import, drift, replacement, partial failure, retry, refresh, taint or repair, and state recovery.

TEST 03

Validate policy, approval, secret isolation, provider signing, and execution-environment controls.

TEST 04

Measure plan, apply, reconcile, provider, and state performance at representative scale.

TEST 05

Upgrade providers, modules, languages, compositions, and the control plane.

TEST 06

Migrate a controlled stack between approaches and verify data, state, policy, and rollback.

EVIDENCE PACKAGE

Preserve versions, architecture, configuration, data set, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

CONDITIONAL RECOMMENDATION AND REVALIDATION

DECISION RECORD, ACCEPTANCE, AND NEXT EVIDENCE

RECOMMENDATION POSITION

- Select the control-loop model before comparing feature scores.
- Treat state, provider, secret, and execution trust as mandatory gates.
- Require resource-specific failure, rollback, and migration testing.
- Revalidate providers, licensing, state formats, and control-plane releases quarterly.

CANDIDATE	ADJ. SCORE	GATE POSITION	LAB STATE
Terraform	51.3	PASS - enterprise declarative delivery; CONDITIONAL - licensing, state, and provider trust	PENDING
OpenTofu	50.8	PASS - open sovereign IaC; CONDITIONAL - enterprise workflow and support model	PENDING
Pulumi	50.8	PASS - software-engineering-centric IaC; CONDITIONAL - language and dependency governance	PENDING
Crossplane	49.7	PASS - internal developer platform control plane; NOT EQUIVALENT - continuous reconciliation changes the operating model	PENDING

REVALIDATION SCHEDULE

Review no later than 2026-10-24. Review earlier after a major release, commercial change, critical vulnerability, material outage, architecture transition, failed acceptance test, or change to the target workload profile.

SOURCE AUTHORITY AND REVISION

CURRENT EVIDENCE SNAPSHOT AND PUBLICATION HISTORY

SOURCE ID	PUBLISHER	TITLE	CHECKED
NIST-CSF-20	NIST	Cybersecurity Framework 2.0	2026-07-24
NIST-SP800-53	NIST	Security and Privacy Controls for Information Systems and Organizations	2026-07-24
TERRAFORM	HashiCorp	Terraform Documentation	2026-07-24
TERRAFORM-STATE	HashiCorp	Terraform State Documentation	2026-07-24
OPENTOFU	OpenTofu	OpenTofu Documentation	2026-07-24
OPENTOFU-ENC	OpenTofu	OpenTofu State and Plan Encryption	2026-07-24
PULUMI	Pulumi	Pulumi Infrastructure as Code Documentation	2026-07-24
PULUMI-STATE	Pulumi	Pulumi State and Backends	2026-07-24
CROSSPLANE	Crossplane	Crossplane Compositions	2026-07-24

SOURCE POSITION

Official documentation and source repositories support the current capability model. Source records preserve publisher, title, URL, purpose, and review date in the machine-readable authority register. Commercial terms, performance, and environment-specific behavior remain unverified until controlled proof.

REVISION HISTORY

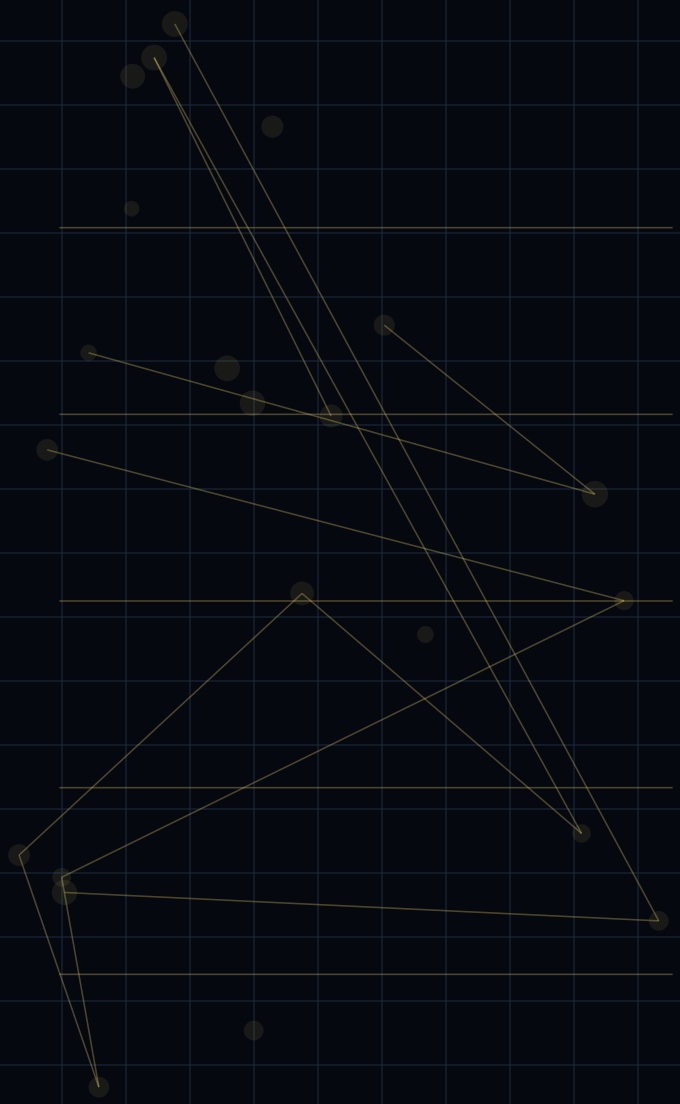
1.1.0-candidate / 2026-07-24 - permanent-publication rebuild using the final benchmark system, profile-specific rankings, evidence adjustment, mandatory gates, and controlled PoC requirements.

COMPARATIVE EVALUATION / EVIDENCE REFRESH

CMP-008

Configuration Management

Fleet configuration, convergence and orchestration platforms



CURRENT EVIDENCE SNAPSHOT

Official product sources refreshed through 2026-09-17

Controlled Mythos laboratory rerun: NOT ASSERTED

BENCHMARK DOCTRINE

Gates before scores. Evidence before certainty. Profile fit before universal ranking.

VERSION 1.2.0-candidate

CANDIDATE COMPARATIVE EVALUATION / PUBLIC

Configuration Management

Fleet configuration, convergence and orchestration platforms

DOCUMENT ID
CMP-008
VERSION
1.2.0-candidate
STATUS
Candidate Comparative Evaluation
PUBLICATION DATE
2026-09-17
SCHEDULED REVIEW
2026-12-16
CANONICAL ROUTE
/library/evaluations/cmp-008/

4
CANDIDATES

18
ATOMIC CRITERIA

9
MANDATORY GATES

3
WORKLOAD PROFILES

E2
CURRENT MAX EVIDENCE

CURRENT DECISION BOUNDARY

Ansible remains a strong July documentation baseline, but that is not a universal selection. Puppet lifecycle changes, Chef 19 migration, and Salt runtime evolution make current lab evidence mandatory. Prove drift, content trust, secrets, fleet scale, rollback, recovery, and upgrade behavior.

NO CURRENT UNIVERSAL WINNER IS PUBLISHED FROM THIS REFRESH.

July 24 baseline scores are preserved as lineage and sensitivity evidence, not silently reissued as September laboratory results.

Red Hat Ansible Automation P

AAP 2.6 patch / 2026-08-04

REFRESH TRIGGER

Current 2.6 patch line includes security fixes and updates across controller, hub, Event-Driven Ansible, and Receptor.

Puppet Enterprise

PE 2025.11.3 / 2026-09-10

REFRESH TRIGGER

Puppet is transitioning support lifecycle models while updating certificate, patching, PostgreSQL, security, and platform support.

Chef Infra

Client 19.3.15 LTS / 2026-05-22

REFRESH TRIGGER

Client 19 LTS moves to Habitat-based packaging; migration and platform compatibility are first-class selection risks.

Salt

3008.2 line / current docs

REFRESH TRIGGER

Current 3008 line includes packaging/runtime changes while 3006.x remains an important long-lived branch.

FINAL BENCHMARK SYSTEM / DECISION ARCHITECTURE

01 SCOPE

Freeze decision, candidates, versions and exclusions.

02 GATES

Test non-compensable trust, state, recovery, change and exit.

03 SCORES

Measure atomic capability with approved profile weights.

04 CONFIDENCE

Discount by source grade, freshness, repeatability and relevance.

05 PROOF

Run normal, failure, migration, recovery and operator scenarios.

06 DECISION

Record conditional fit, residual risk, owner and revalidation trigger.

EVIDENCE CONFIDENCE / MODIFIERS

E0	E1	E2	E3	E4	E5
0.00	0.38	0.64	0.78	0.91	1.00
UNSUPPORTED	VENDOR ASSERTION	OFFICIAL SOURCE	MYTHOS LAB	PRODUCTION	INDEPENDENT

THIS EDITION

Current official product and release sources have been refreshed to E2. No E3 controlled Mythos lab rerun is claimed. Therefore current profile rankings remain intentionally unissued until the proof plan is executed.

What changed since the July benchmark snapshot

Changes below are sourced from current official release documentation. They identify revalidation triggers; they are not translated into new numeric scores without controlled proof.

Red Hat Ansible Automation Platfor

AAP 2.6 patch / 2026-08-04

MATERIAL DELTA

Current 2.6 patch line includes security fixes and updates across controller, hub, Event-Driven Ansible, and Receptor.

CRITERIA TOUCHED

C02 / C05 / C07 / C09 / C14

MODERATE REVALIDATION

Puppet Enterprise

PE 2025.11.3 / 2026-09-10

MATERIAL DELTA

Puppet is transitioning support lifecycle models while updating certificate, patching, PostgreSQL, security, and platform support.

CRITERIA TOUCHED

C02 / C04 / C09 / C14 / C17

HIGH REVALIDATION

Chef Infra

Client 19.3.15 LTS / 2026-05-22

MATERIAL DELTA

Client 19 LTS moves to Habitat-based packaging; migration and platform compatibility are first-class selection risks.

CRITERIA TOUCHED

C04 / C09 / C11 / C14 / C16 / C17

HIGH REVALIDATION

Salt

3008.2 line / current docs

MATERIAL DELTA

Current 3008 line includes packaging/runtime changes while 3006.x remains an important long-lived branch.

CRITERIA TOUCHED

C04 / C08 / C09 / C14 / C17

MODERATE REVALIDATION

Mandatory gates remain non-compensable

No current release is marked PASS solely from documentation. E2 means the official source supports the capability path; LAB and RETEST identify proof still required. HOLD blocks a current decision.

MANDATORY GATE	RED HAT ANSIBLE	PUPPET ENTERPRI	CHEF INFRA	SALT
C01 Functional coverage	E2	E2	E2	E2
C02 Security / trust	RETEST	RETEST	E2	E2
C03 Governance / approval	E2	E2	E2	E2
C04 State integrity	E2	RETEST	RETEST	RETEST
C07 Observability / evidence	RETEST	E2	E2	E2
C09 Resilience / continuity	RETEST	RETEST	RETEST	RETEST
C11 Change safety / rollback	LAB	LAB	RETEST	LAB
C16 Portability / exit	LAB	LAB	RETEST	LAB
C18 Assurance confidence	HOLD	HOLD	HOLD	HOLD

LEGEND

E2 official-source support only

LAB controlled proof required

RETEST release delta touches this gate

HOLD decision gate remains closed

Historical profile sensitivity - preserved, not reissued

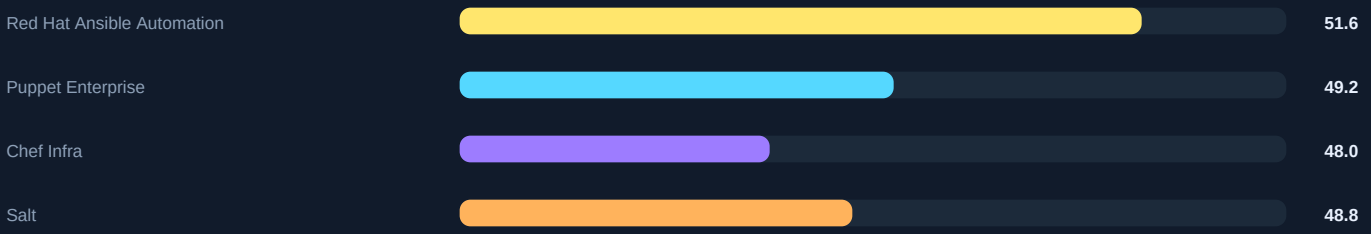
Values below are the 2026-07-24 evidence-adjusted baseline. They show sensitivity to operating-model weights. The September refresh does not recompute rank because E3 lab proof has not been reproduced.

AGENTLESS ENTERPRISE AUTOMATION

BASELINE SPREAD 4.0 PTS

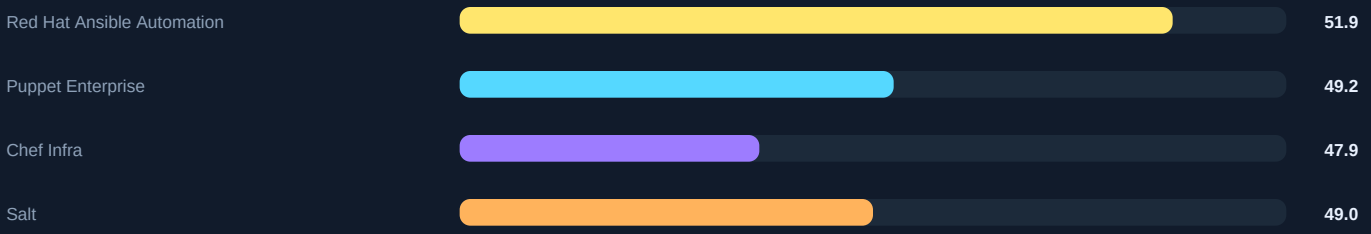
CONTINUOUS DESIRED-STATE ENFORCEMENT

BASELINE SPREAD 3.6 PTS



EVENT-DRIVEN OPERATIONS FABRIC

BASELINE SPREAD 4.0 PTS



CURRENT RANK STATUS: WITHHELD PENDING CONTROLLED PROOF.

Evidence confidence is separate from capability

E5 — Independent repeatable validation
confidence modifier 1.00

E4 — Production evidence in adopting environment
confidence modifier 0.91

E3 — Controlled Mythos laboratory result
confidence modifier 0.78

E2 — Official documentation / release / source
confidence modifier 0.64

E1 — Vendor assertion or marketing
confidence modifier 0.38

E0 — Unsupported or unverified
confidence modifier 0.00

CURRENT EVIDENCE STATE

Red Hat Ansible Automation P	E2 REFRESHED
Historical adjusted confidence: 58.2%	
E3 LAB: NOT REPRODUCED	
Puppet Enterprise	E2 REFRESHED
Historical adjusted confidence: 58.2%	
E3 LAB: NOT REPRODUCED	
Chef Infra	E2 REFRESHED
Historical adjusted confidence: 58.2%	
E3 LAB: NOT REPRODUCED	
Salt	E2 REFRESHED
Historical adjusted confidence: 58.2%	
E3 LAB: NOT REPRODUCED	

ASSURANCE RULE

A current release note can change capability context, but it cannot raise assurance beyond E2. Current recommendation strength is therefore bounded until the test plan produces reproducible artifacts.

Lifecycle risk is evaluated independently of features

The benchmark models migration, upgrade, compatibility, support, staffing, data/state export, downtime and exit. Current release changes below are explicit proof triggers.

Red Hat Ansible Automation Platform

MODERATE REVALIDATION

Current 2.6 patch line includes security fixes and updates across controller, hub, Event-Driven Ansible, and Receptor.

RETEST: C02 / C05 / C07 / C09 / C14

Puppet Enterprise

HIGH REVALIDATION

Puppet is transitioning support lifecycle models while updating certificate, patching, PostgreSQL, security, and platform support.

RETEST: C02 / C04 / C09 / C14 / C17

Chef Infra

HIGH REVALIDATION

Client 19 LTS moves to Habitat-based packaging; migration and platform compatibility are first-class selection risks.

RETEST: C04 / C09 / C11 / C14 / C16 / C17

Salt

MODERATE REVALIDATION

Current 3008 line includes packaging/runtime changes while 3006.x remains an important long-lived branch.

RETEST: C04 / C08 / C09 / C14 / C17

IRREVERSIBLE COMMITMENT GATE

Before migration or procurement lock-in, prove rollback, state portability, operational reconstruction, and supplier-exit assumptions.

Controlled proof is the next decision-producing step

REPRODUCTION STATE / NOT EXECUTED IN THIS EVIDENCE-REFRESH RELEASE

01

CONTROLLED SCENARIO

Manage representative Linux, Windows, network, cloud, middleware, and application targets.

02

CONTROLLED SCENARIO

Execute configuration, patch, compliance, orchestration, secret, and remediation workflows.

03

CONTROLLED SCENARIO

Test dry-run, check or noop mode, canary, concurrency, failure, retry, rollback, and recovery.

04

CONTROLLED SCENARIO

Inject controller loss, certificate expiry, repository outage, partial fleet reachability, and conflicting state.

05

CONTROLLED SCENARIO

Measure convergence time, drift detection, job throughput, operator effort, and evidence completeness.

06

CONTROLLED SCENARIO

Rebuild the control plane and representative nodes from versioned content and backups.

EVIDENCE PACKAGE

Preserve exact versions, architecture, configuration, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

Current decision posture

Ansible remains a strong July documentation baseline, but that is not a universal selection. Puppet lifecycle changes, Chef 19 migration, and Salt runtime evolution make current lab evidence mandatory. Prove drift, content trust, secrets, fleet scale, rollback, recovery, and upgrade behavior.

CURRENT RANKING: NOT REISSUED / CONTROLLED LAB REQUIRED

CURRENT OFFICIAL-SOURCE REGISTER / CHECKED 2026-09-17

Red Hat Ansible Automation P	E2 / OFFICIAL	
Red Hat AAP 2.6 Aug 4 release		docs.redhat.com
Puppet Enterprise	E2 / OFFICIAL	
Puppet Enterprise release notes		help.puppet.com
Chef Infra	E2 / OFFICIAL	
Chef Infra Client release notes		docs.chef.io
Salt	E2 / OFFICIAL	
Salt 3008.2 release notes		docs.saltproject.io

REVISION LINEAGE

1.2.0-candidate / 2026-09-17 - current-source evidence refresh; no lab rescore issued.

1.1.0-candidate / 2026-07-24 - baseline benchmark using the final comparative evaluation system.

Trigger earlier review after major release, acquisition, licensing change, critical vulnerability, material outage, failed PoC, or workload change.

BASELINE ANALYTIC RECORD CONTINUES ON PAGE 11 FOR TRACEABILITY.

ATOMIC CRITERIA MODEL

WEIGHTED CAPABILITY WITH EXPLICIT MINIMUM EVIDENCE

ID	CRITERION	WEIGHT	GATE	MINIMUM EVIDENCE
C01	Configuration, patch, compliance, orchestration, and remediation coverage	5	YES	Feature documentation, APIs, configuration exports, and scenario demonstration.
C02	Controller, agent, credential, content, and execution trust architecture	5	YES	Threat model, identity flows, RBAC tests, audit records, and security configuration.
C03	Governance and approval controls	5	YES	Approval workflow, policy controls, separation-of-duties tests, and decision records.
C04	Desired-state, inventory, content, report, and run-state integrity	5	YES	Backup, restore, rollback, drift, and corruption-recovery evidence.
C05	Automation and API quality	4	NO	API specifications, authentication tests, rate limits, event samples, and automation runs.
C06	Integration ecosystem	4	NO	Supported integrations, connector tests, failure behavior, and lifecycle ownership.
C07	Observability and evidence	4	YES	Logs, traces, metrics, reports, export tests, and evidence-retention controls.
C08	Node, job, event, content, and controller scale	4	NO	Controlled load tests, topology scale, saturation behavior, and capacity model.
C09	Resilience and continuity	5	YES	Failure injection, HA tests, recovery timing, degraded-mode operation, and runbooks.
C10	Usability and operator cognition	3	NO	Task observation, error-rate measures, navigation tests, and operator interviews.
C11	Dry-run, change window, canary, rollback, and recovery safety	5	YES	Plan or preview output, canary tests, rollback execution, and post-change verification.
C12	Extensibility and programmability	3	NO	Plugin, module, provider, workflow, or code-extension tests and upgrade impact analysis.
C13	Deployment and sovereignty options	3	NO	Deployment architecture, data-flow mapping, residency evidence, and offline tests.
C14	Lifecycle and upgrade discipline	4	NO	Release notes, upgrade test, compatibility matrix, support policy, and migration plan.
C15	Economics and cost predictability	3	NO	Bill-of-materials, licensing model, staffing estimates, metering, and sensitivity analysis.
C16	Portability and exit	4	YES	Export tests, format analysis, replacement workflow, and decommission evidence.
C17	Vendor and ecosystem risk	3	NO	License analysis, roadmap evidence, bus-factor indicators, and dependency inventory.
C18	Assurance confidence	5	YES	Source provenance, lab artifacts, production evidence, independent replication, and review date.

SCORE SCALE

0 absent; 1 materially deficient; 2 partial; 3 adequate with limits; 4 strong; 5 leading for the defined profile. Scores remain provisional until the required evidence grade is achieved.

RAW CAPABILITY SCORECARD

DOCUMENTED CAPABILITY BEFORE EVIDENCE DISCOUNT

CRITERION	ANSIBLE AAP	PUPPET	CHEF	SALT
C01 Configuration, patch, compliance, orchestration, and remediation coverage	4.8	4.4	4.3	4.5
C02 Controller, agent, credential, content, and execution trust architecture	4.4	4.4	4.3	4.2
C03 Governance and approval controls	4.4	4.2	4.1	4.0
C04 Desired-state, inventory, content, report, and run-state integrity	4.2	4.5	4.4	4.3
C05 Automation and API quality	4.8	4.1	4.0	4.6
C06 Integration ecosystem	4.7	4.2	4.1	4.2
C07 Observability and evidence	4.5	4.4	4.3	4.2
C08 Node, job, event, content, and controller scale	4.6	4.6	4.4	4.6
C09 Resilience and continuity	4.3	4.4	4.2	4.1
C10 Usability and operator cognition	4.5	4.0	3.9	4.0
C11 Dry-run, change window, canary, rollback, and recovery safety	4.5	4.2	4.1	4.2
C12 Extensibility and programmability	4.8	4.3	4.4	4.6
C13 Deployment and sovereignty options	4.4	4.3	4.2	4.4
C14 Lifecycle and upgrade discipline	4.3	4.1	3.9	3.8
C15 Economics and cost predictability	3.8	3.7	3.6	4.0
C16 Portability and exit	4.4	4.0	3.9	4.2
C17 Vendor and ecosystem risk	4.1	3.7	3.5	3.6
C18 Assurance confidence	3.7	3.5	3.4	3.4

WEIGHTED BASELINE / 100

Ansible AAP	Puppet	Chef	Salt
88.0	**83.7**	**81.4**	**83.2**

EVIDENCE-ADJUSTED SCORECARD

CAPABILITY DISCOUNTED BY CURRENT EVIDENCE CONFIDENCE

CANDIDATE	RAW	EVIDENCE CONFIDENCE	ADJUSTED	CURRENT STATE
Ansible AAP	88.0	58.2%	51.8	CONTROLLED LAB PENDING
Puppet	83.7	58.2%	49.2	CONTROLLED LAB PENDING
Chef	81.4	58.2%	47.9	CONTROLLED LAB PENDING
Salt	83.2	58.2%	48.9	CONTROLLED LAB PENDING

EVIDENCE SCALE

E0 / 0.00
Unsupported or unverified

E1 / 0.38
Vendor assertion or marketing statement

E2 / 0.64
Official documentation, release notes, or source repository

E3 / 0.78
Controlled Mythos laboratory result

E4 / 0.91
Production evidence from the adopting environment

E5 / 1.00
Independent repeatable validation

CURRENT CONFIDENCE BOUNDARY

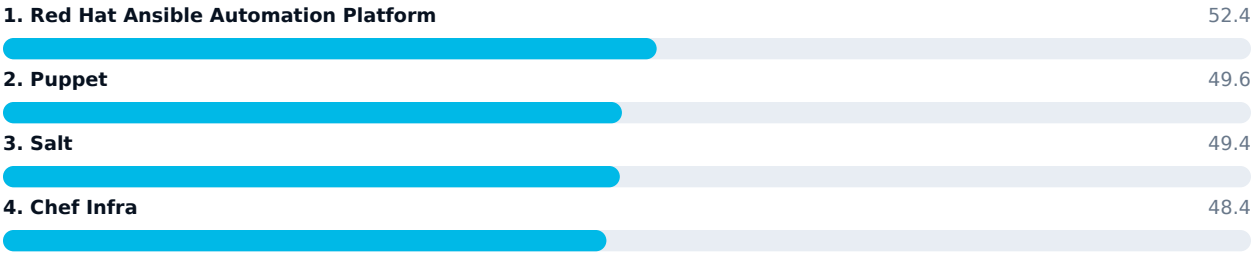
Most candidate criteria are supported at E2: official documentation or source evidence. Environment-specific laboratory, production, and independent evidence remain future gates.

PROFILE-SPECIFIC RANKINGS

EVIDENCE-ADJUSTED SENSITIVITY ANALYSIS

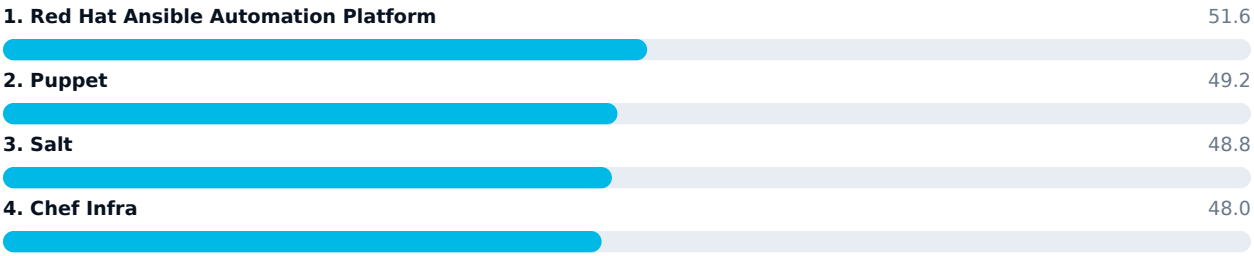
AGENTLESS ENTERPRISE AUTOMATION

Prioritizes broad automation, workflows, execution portability, network and cloud coverage, and low endpoint footprint.



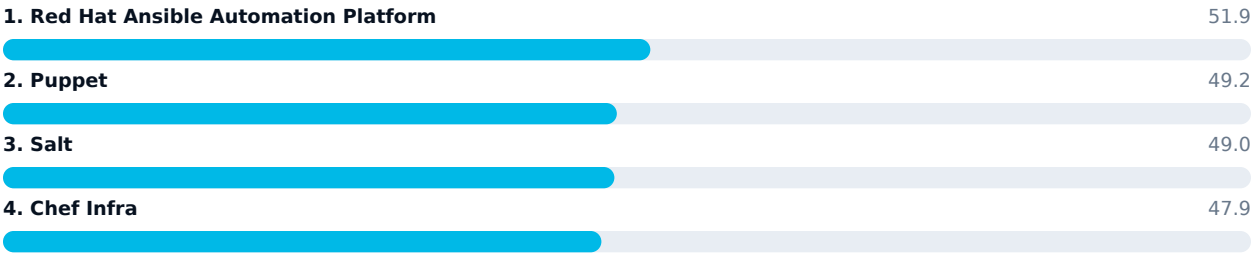
CONTINUOUS DESIRED-STATE ENFORCEMENT

Prioritizes periodic convergence, drift correction, reporting, fleet scale, and stable server configuration.



EVENT-DRIVEN OPERATIONS FABRIC

Prioritizes high-speed remote execution, event reactions, targeting, scale, extensibility, and operational control.



RANK SENSITIVITY

Small score differences are not decision certainty. Treat near-ties as a requirement for deeper PoC, commercial, migration, and operating-model evidence.

CANDIDATE DEEP DIVE / 01

RED HAT ANSIBLE AUTOMATION PLATFORM

OPERATING MODEL

Agentless automation platform using playbooks, inventories, collections, execution environments, controller, API, and event-driven capabilities.

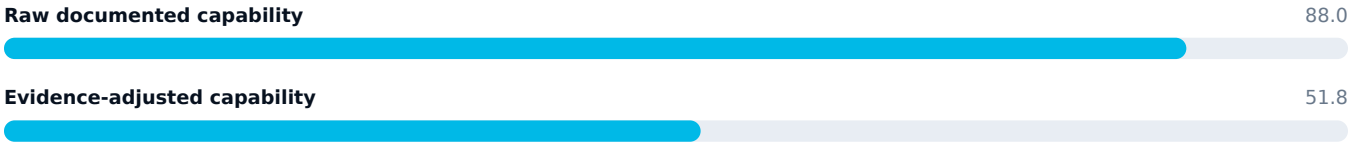
DOCUMENTED STRENGTHS

- Broad automation ecosystem and strong network, cloud, server, and application coverage.
- Agentless SSH and API execution reduces endpoint footprint.
- Execution environments improve reproducibility.
- Controller, RBAC, workflow, and API support enterprise operations.

CAUTIONS AND PROOF OBLIGATIONS

- Idempotence and rollback depend on module and playbook engineering.
- Push execution can miss drift between runs without additional controls.
- Content sprawl and inventory quality require governance.

BASELINE SCORE PROFILE



GATE DISPOSITION

PASS - agentless enterprise automation | CONDITIONAL - content quality and drift architecture

CANDIDATE DEEP DIVE / 02

PUPPET

OPERATING MODEL

Agent-server desired-state configuration management using catalogs, facts, modules, reports, and periodic enforcement.

DOCUMENTED STRENGTHS

- Strong desired-state convergence and drift correction.
- Mature agent-server architecture and reporting.
- Rich module and policy ecosystem.
- Suitable for large persistent server fleets.

CAUTIONS AND PROOF OBLIGATIONS

- Agent and server lifecycle add operational responsibility.
- General orchestration and network/cloud breadth may require complementary tools.
- Language, module, certificate, and upgrade governance are specialized.

BASELINE SCORE PROFILE

Raw documented capability 83.7

Evidence-adjusted capability 49.2

GATE DISPOSITION

PASS - continuous desired-state server fleet | CONDITIONAL - agent and platform lifecycle

CANDIDATE DEEP DIVE / 03

CHEF INFRA

OPERATING MODEL

Agent-based configuration management using recipes, cookbooks, Policyfiles, Chef Infra Server, and client convergence.

DOCUMENTED STRENGTHS

- Powerful code-driven configuration model.
- Policyfiles support versioned dependency and run-list control.
- Mature convergence architecture and ecosystem.
- Strong fit for organizations with existing Chef engineering capability.

CAUTIONS AND PROOF OBLIGATIONS

- Ruby and cookbook engineering skill are required.
- Platform and product transitions must be evaluated.
- Agent, server, policy, and content lifecycle add complexity.

BASELINE SCORE PROFILE

Raw documented capability 81.4

Evidence-adjusted capability 47.9

GATE DISPOSITION

PASS - code-driven convergence for established Chef estates | CONDITIONAL - skills and product trajectory

CANDIDATE DEEP DIVE / 04

SALT

OPERATING MODEL

Event-driven remote execution and configuration-management platform using master/minion, states, pillars, reactors, and execution modules.

DOCUMENTED STRENGTHS

- Fast remote execution and event-driven operations.
- Flexible states, modules, targeting, and orchestration.
- Strong fit for large-scale operational control and reactive automation.
- Extensible Python-based architecture.

CAUTIONS AND PROOF OBLIGATIONS

- Master, key, event, and minion architecture require strong security and operations.
- Content and state complexity can become difficult to govern.
- Enterprise support, packaging, and lifecycle expectations need validation.

BASELINE SCORE PROFILE

Raw documented capability 83.2

Evidence-adjusted capability 48.9

GATE DISPOSITION

PASS - event-driven operations fabric | CONDITIONAL - security, content, and support model

ARCHITECTURE AND INTEGRATION FIT

THE PRODUCT IS ONLY ONE PART OF THE OPERATING SYSTEM

REFERENCE OPERATING LAYERS

01 Inventory, facts, targeting, and desired-state authority

02 Controller, server, master, agent, or execution environment

03 Content, modules, cookbooks, states, playbooks, policy, and secrets

04 Execution, convergence, orchestration, events, and remediation

05 Reports, evidence, recovery, content supply chain, and lifecycle

INTEGRATION BOUNDARIES

- Identity, secrets, PKI, SSH, and privilege management
- Git, artifact, collection, module, cookbook, and package registries
- CMDB, source of truth, ITSM, CI/CD, and policy
- Cloud, network, operating system, application, and endpoint platforms
- Observability, compliance, vulnerability, and security operations

ARCHITECTURE GATE

Every external dependency requires an owner, identity boundary, failure mode, evidence path, version policy, recovery procedure, and exit strategy.

SECURITY, OPERATIONS, LIFECYCLE, ECONOMICS, AND EXIT

CROSS-DOMAIN DECISION RECORD

SECURITY AND AUTHORITY

Identity, credentials, secrets, roles, privileged actions, signing, approvals, isolation, audit, and incident response must be tested as an integrated system.

RELIABILITY AND RECOVERY

Control-plane, data-plane, dependency, region, database, queue, network, and operator failures require defined degraded modes and recovery objectives.

CHANGE AND LIFECYCLE

Version, compatibility, migration, canary, rollback, content, plugin, provider, firmware, and decommission procedures are part of product fitness.

ECONOMICS AND CAPACITY

Licenses, metering, data, compute, hardware, storage, support, staffing, training, integration, migration, and opportunity cost require sensitivity analysis.

SOVEREIGNTY AND PRIVACY

Data location, support access, telemetry, subprocessors, encryption, key control, disconnected operation, and legal obligations must align with policy.

PORTABILITY AND EXIT

Configuration, policy, data, state, evidence, workflows, identities, and operational knowledge must be exportable and reconstructable.

DECISION OWNERSHIP

Architecture, security, operations, finance, procurement, legal, data, and service owners jointly accept the final profile, evidence, residual risk, and exit obligations.

RISK AND FAILURE REGISTER

SCENARIOS THAT CAN INVALIDATE A FEATURE-BASED SELECTION

ID	SEVERITY	FAILURE SCENARIO	REQUIRED TREATMENT
R-01	CRITICAL	Automation changes systems successfully but cannot prove intended service outcome.	PREVENT / DETECT / RECOVER / EVIDENCE
R-02	HIGH	Agent or controller compromise enables fleet-wide execution.	PREVENT / DETECT / RECOVER / EVIDENCE
R-03	HIGH	Content dependencies change without controlled locking or signing.	PREVENT / DETECT / RECOVER / EVIDENCE
R-04	HIGH	Dry-run behavior diverges from actual enforcement.	PREVENT / DETECT / RECOVER / EVIDENCE
R-05	MEDIUM	Desired-state engines fight manual incident recovery or application controllers.	PREVENT / DETECT / RECOVER / EVIDENCE
R-06	MEDIUM	Migration loses inventory semantics, encrypted data, schedules, reports, exceptions, and operator knowledge.	PREVENT / DETECT / RECOVER / EVIDENCE

RISK RULE

A candidate with an unresolved critical failure path cannot advance because optional capability, market position, or commercial discount cannot compensate for missing safety or recovery evidence.

CONTROLLED PROOF-OF-CAPABILITY PLAN

REPEATABLE SCENARIOS, INSTRUMENTATION, AND ACCEPTANCE

TEST 01

Manage representative Linux, Windows, network, cloud, middleware, and application targets.

TEST 02

Execute configuration, patch, compliance, orchestration, secret, and remediation workflows.

TEST 03

Test dry-run, check or noop mode, canary, concurrency, failure, retry, rollback, and recovery.

TEST 04

Inject controller loss, certificate expiry, repository outage, partial fleet reachability, and conflicting state.

TEST 05

Measure convergence time, drift detection, job throughput, operator effort, and evidence completeness.

TEST 06

Rebuild the control plane and representative nodes from versioned content and backups.

EVIDENCE PACKAGE

Preserve versions, architecture, configuration, data set, scenario, expected result, instrumentation, raw output, timing, failures, operator actions, deviations, acceptance state, owner, and artifact hashes.

CONDITIONAL RECOMMENDATION AND REVALIDATION

DECISION RECORD, ACCEPTANCE, AND NEXT EVIDENCE

RECOMMENDATION POSITION

- Select by enforcement model and target estate, not language preference alone.
- Require content-locking, secret, privilege, dry-run, and recovery evidence.
- Define which system owns each configuration domain.
- Revalidate quarterly and before major controller, agent, content, or packaging changes.

CANDIDATE	ADJ. SCORE	GATE POSITION	LAB STATE
Ansible AAP	51.8	PASS - agentless enterprise automation; CONDITIONAL - content quality and drift architecture	PENDING
Puppet	49.2	PASS - continuous desired-state server fleet; CONDITIONAL - agent and platform lifecycle	PENDING
Chef	47.9	PASS - code-driven convergence for established Chef estates; CONDITIONAL - skills and product trajectory	PENDING
Salt	48.9	PASS - event-driven operations fabric; CONDITIONAL - security, content, and support model	PENDING

REVALIDATION SCHEDULE

Review no later than 2026-10-24. Review earlier after a major release, commercial change, critical vulnerability, material outage, architecture transition, failed acceptance test, or change to the target workload profile.

SOURCE AUTHORITY AND REVISION

CURRENT EVIDENCE SNAPSHOT AND PUBLICATION HISTORY

SOURCE ID	PUBLISHER	TITLE	CHECKED
NIST-CSF-20	NIST	Cybersecurity Framework 2.0	2026-07-24
NIST-SP800-53	NIST	Security and Privacy Controls for Information Systems and Organizations	2026-07-24
CIS-CONTROLS	Center for Internet Security	CIS Critical Security Controls v8.1	2026-07-24
ANSIBLE-AAP	Ansible	Red Hat Ansible Automation Platform	2026-07-24
ANSIBLE-EE	Ansible	Ansible Execution Environments	2026-07-24
PUPPET	Perforce	Puppet Platform Components	2026-07-24
CHEF	Progress Chef	Chef Policyfiles	2026-07-24
SALT	Salt Project	Salt States	2026-07-24
SALT-TEST	Salt Project	Salt State Testing	2026-07-24

SOURCE POSITION

Official documentation and source repositories support the current capability model. Source records preserve publisher, title, URL, purpose, and review date in the machine-readable authority register. Commercial terms, performance, and environment-specific behavior remain unverified until controlled proof.

REVISION HISTORY

1.1.0-candidate / 2026-07-24 - permanent-publication rebuild using the final benchmark system, profile-specific rankings, evidence adjustment, mandatory gates, and controlled PoC requirements.