UNOFFICIAL PRACTICE EXAM

Claude Certified Architect - Foundations

Candidate Edition - 60 advanced scenario-based questions in English

60	6	5	1
Questions	Scenarios	Domains	Best answer

PA

Created and curated by Paolo Amato

Independent practice material designed to train single-best-answer architectural judgment.

Version 1.0.1 | Claude Certified Architect - Foundations Exam Guide, Version 1.0, effective July 2026

How to Use This Practice Exam

This candidate edition contains 60 original, scenario-based questions. Choose one best answer for each item, complete the answer sheet, and consult the separate Solutions Edition only after finishing the full set.

Exam-aligned, not exam-derived. The structure follows the six scenario families and five domain weightings in the official guide baseline. The questions are independently written and do not reproduce live exam content.

Format and scoring baseline. The public Version 1.0 guide, effective July 2026, lists multiple-choice and multiple-response items. This edition intentionally remains single-best-answer and reports only a raw score and domain breakdown; no official raw-to-scaled conversion is available for this practice set.

FORMAT

60 multiple-choice questions, A-D, one best answer

DIFFICULTY

Advanced and intentionally challenging; not psychometrically calibrated

RECOMMENDED USE

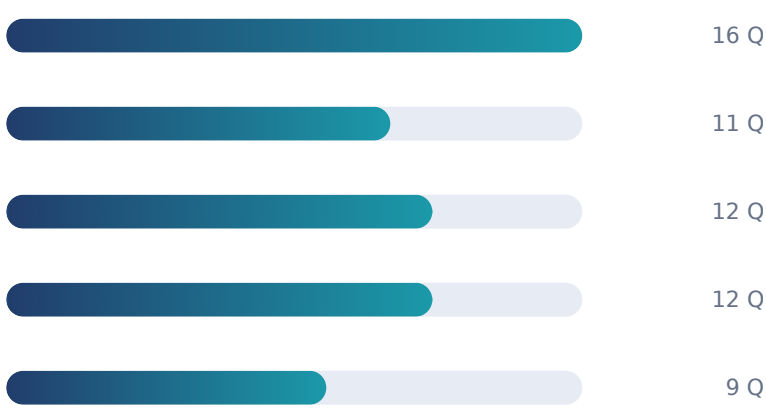
Closed-book first pass, then domain-by-domain review

AUTHORSHIP

Created and curated by Paolo Amato

Domain Distribution

The 60-question bank mirrors the guide's domain weighting as closely as whole-question counts allow.



STUDY WORKFLOW

Practice Protocol and Scenario Map

Treat the bank as an architectural judgment exercise. The goal is not to recognize vocabulary; it is to identify the control, interface, or workflow that best addresses the stated failure mode.

01 Complete a closed-book pass
Do not open the Solutions Edition. Choose one answer even when two options appear plausible.

02 Mark uncertainty
Flag guesses and low-confidence answers, including items that later prove correct.

03 Review the decision boundary
Compare the correct answer with the strongest distractor and identify the scenario phrase that separates them.

04 Re-test by weakness
Convert each miss into a rule, then practice new questions in the weak domain or task focus.

Six Scenario Families

This full-bank edition includes all six guide-aligned scenario families so that every major decision context is represented.

01 Customer Support
Agentic loops, deterministic policy controls, tool selection, ambiguity, and escalation.

02 Claude Code
Configuration scope, rules, skills, plan mode, iterative refinement, and session handling.

03 Multi-Agent Research
Coordinator-subagent orchestration, explicit context transfer, errors, provenance, and conflicts.

04 Developer Productivity
Built-in tool choice, codebase exploration, scoped MCP access, state persistence, and safe edits.

05 CI/CD
Headless CI, structured findings, precision, independent review, multi-pass analysis, and batching.

06 Structured Extraction
Schema design, few-shot extraction, validation, context, confidence calibration, and provenance.

Single-best-answer lens. Prefer the option that fixes the root cause at the correct architectural layer, preserves required context and provenance, and adds no unnecessary machinery.

Answer Sheet

Mark one answer per question. Use the notes area to flag items that felt ambiguous or relied on a guess.

Q	A	B	C	D
1	○	○	○	○
2	○	○	○	○
3	○	○	○	○
4	○	○	○	○
5	○	○	○	○
6	○	○	○	○
7	○	○	○	○
8	○	○	○	○
9	○	○	○	○
10	○	○	○	○
11	○	○	○	○
12	○	○	○	○
13	○	○	○	○
14	○	○	○	○
15	○	○	○	○
16	○	○	○	○
17	○	○	○	○
18	○	○	○	○
19	○	○	○	○
20	○	○	○	○

Q	A	B	C	D
21	○	○	○	○
22	○	○	○	○
23	○	○	○	○
24	○	○	○	○
25	○	○	○	○
26	○	○	○	○
27	○	○	○	○
28	○	○	○	○
29	○	○	○	○
30	○	○	○	○
31	○	○	○	○
32	○	○	○	○
33	○	○	○	○
34	○	○	○	○
35	○	○	○	○
36	○	○	○	○
37	○	○	○	○
38	○	○	○	○
39	○	○	○	○
40	○	○	○	○

Q	A	B	C	D
41	○	○	○	○
42	○	○	○	○
43	○	○	○	○
44	○	○	○	○
45	○	○	○	○
46	○	○	○	○
47	○	○	○	○
48	○	○	○	○
49	○	○	○	○
50	○	○	○	○
51	○	○	○	○
52	○	○	○	○
53	○	○	○	○
54	○	○	○	○
55	○	○	○	○
56	○	○	○	○
57	○	○	○	○
58	○	○	○	○
59	○	○	○	○
60	○	○	○	○

Uncertain or guessed questions

Rules to revisit

SCENARIO 1 OF 6

Customer Support Resolution Agent

Northstar Retail is deploying a customer support agent with the Claude Agent SDK. The agent handles returns, billing disputes, replacements, and account issues through MCP tools: get_customer, lookup_order, get_policy, process_refund, create_replacement, and escalate_to_human. Identity verification is mandatory before account-specific or financial actions. The business target is high first-contact resolution without weakening policy compliance or customer privacy.

01

1 SELECT THE SINGLE BEST ANSWER

The runtime receives an assistant message containing a short customer-facing preamble and two tool_use blocks. The response has stop_reason set to tool_use. What should the runtime do next?

A Execute all requested tools, append matching tool_result blocks to history, and call Claude again until stop_reason is end_turn.

B Execute only the first tool call, because multiple calls in one response should be treated as alternative actions rather than a batch of requests.

C Continue until either the assistant writes an explicit completion phrase or a fixed iteration ceiling is reached, regardless of stop_reason.

D Send the preamble to the customer immediately, store the tool requests for the next user turn, and resume only if the customer replies.

2 SELECT THE SINGLE BEST ANSWER

Identity verification is required before any refund. Logs show that rare failures still occur despite a strong system prompt and several correct examples. Which change most directly closes the remaining safety gap?

A. Remove process_refund from the agent and have it draft refund instructions for a human to execute in every case.

B. Force get_customer as the first tool on every conversation, then expose process_refund after any get_customer response without checking whether verification actually succeeded.

C. Enforce a programmatic precondition or hook that blocks process_refund unless verified customer state exists, with a safe recovery path.

D. Expand the system prompt with stronger wording, add more few-shot examples, and alert on violations so the team can tune the prompt later.

3 SELECT THE SINGLE BEST ANSWER

A verified customer reports a damaged item, a duplicate charge on another order, and a promised replacement that never shipped. The current agent resolves the first issue and ends the turn. Which redesign is best?

A. Represent the message as separate concerns, investigate independent concerns using the shared verified case facts, then synthesize one complete resolution or structured handoff.

B. Handle the concerns strictly in the order mentioned and stop as soon as one concern produces an actionable resolution, to keep each turn short.

C. Open three isolated subagent conversations and return their answers directly to the customer without sharing identity, order facts, or policy context between them.

D. Escalate the entire case immediately because multiple concerns imply that autonomous resolution is no longer appropriate.

4 SELECT THE SINGLE BEST ANSWER

A billing dispute must be escalated because policy is silent on the requested exception. Human agents cannot see the original conversation. What should the escalation payload contain?

A. The full transcript and every raw tool response, including unrelated backend fields, so the human can reconstruct the case without information loss.

B. Send a structured handoff with verified identity, issue status, key order facts, evidence, attempted actions, the policy gap, and a recommendation.

C. A one-line summary, the customer's sentiment score, and the model's overall confidence, because the human can repeat the investigation if necessary.

D. Only the unresolved billing request and the last tool error, omitting completed investigations so the handoff remains concise.

5 SELECT THE SINGLE BEST ANSWER

Three MCP tools return eligibility dates as Unix seconds, ISO 8601 strings, and locale-specific text, while statuses use incompatible numeric codes. The model occasionally compares them incorrectly. What is the strongest first control?

A. Expose a normalize_data tool and allow the model to call it when it notices that two results are not directly comparable.

B. Normalize every result into a canonical schema in a PostToolUse hook before adding it to model context.

C. Document every source format in the system prompt and instruct the model to convert values before making a decision.

D. Return longer human-readable explanations from each backend so Claude can infer the meaning of each timestamp and status code.

6 SELECT THE SINGLE BEST ANSWER

The agent often chooses search_customer_orders when the user supplies an exact order ID, and lookup_order_by_id when the user asks for all recent orders. The tools have terse, overlapping descriptions. What should the team change first?

A. Clarify names and descriptions with distinct purposes, accepted identifiers, output contracts, examples, edge cases, and explicit boundaries.

B. Add a system-prompt rule that routes any message containing digits to lookup_order_by_id and all other messages to search_customer_orders.

C. Force lookup_order_by_id on the first turn whenever an order-related word appears, then allow the model to correct the route after an error.

D. Wrap both operations behind one generic lookup_records tool and let the backend infer whether the user intended an order or customer search.

7 SELECT THE SINGLE BEST ANSWER

process_refund rejects a request because the item is final sale. The current tool returns only 'Operation failed', so the agent retries several times. What response design best supports recovery?

A. Return the raw backend exception, policy table name, and internal rule ID so the model can derive a customer-facing explanation.

B. Return a transient category with isRetryable: true and a recommended backoff, because policy data might change later.

C. Return an empty successful result with refund_created: false, because the backend completed normally even though no refund was issued.

D. Return isError: true with a non-retryable business-policy category, a readable explanation, and safe details for the next allowed action.

8 SELECT THE SINGLE BEST ANSWER

get_customer returns three plausible accounts for the same name, and the user has provided no email, phone number, or account ID. What is the best next action?

A. Choose the account with the most recent order, but limit the first action to a read-only lookup before requesting confirmation.

B. Escalate immediately to a human because any duplicate-name result is a policy exception that the agent cannot handle.

C. Query all three accounts, compare order details with the user's description, and use the best match as implicit verification.

D. Ask for an additional identifier and avoid account-specific lookups or actions until the customer record is unambiguously resolved.

9 SELECT THE SINGLE BEST ANSWER

Every support request currently runs the same six-step chain, including refund, replacement, and shipping-policy checks. Most tickets need only one branch. Which redesign best balances efficiency and safety?

A. Run every read-only policy check in parallel on every ticket, then choose the first result that appears relevant to the user's wording.

B. Use a keyword classifier to select one of six hard-coded flows at the first turn and never revise the route after new evidence appears.

C. Let the agent choose an adaptive investigation path from the request and intermediate results, while retaining deterministic gates for identity and other mandatory prerequisites.

D. Keep the fixed pipeline because a predictable sequence is always more reliable than model-directed tool selection, even when most steps are irrelevant.

10 SELECT THE SINGLE BEST ANSWER

The coordinator delegates refund-policy review to a specialist subagent. The specialist must use verified case facts but should not access customer-search or payment tools. How should the coordinator invoke it?

A. Store the case facts in the first subagent invocation and rely on that hidden state for later invocations of the same specialist type.

B. Invoke the subagent with only a ticket ID because subagents automatically inherit the coordinator's conversation and verified state.

C. Pass verified facts, policy evidence, provenance, constraints, and output criteria explicitly; restrict the specialist to policy-review tools.

D. Give the subagent all customer and payment tools so it can reconstruct any missing context, and rely on its system prompt not to call unnecessary tools.

SCENARIO 2 OF 6

Code Generation with Claude Code

A platform team uses Claude Code in a large monorepo containing apps, packages, services, infrastructure, and tests. Some standards must be shared through version control, while developers may keep personal preferences locally. The team uses CLAUDE.md, .claude/rules/, custom commands, skills, plan mode, subagents, and session resumption for both small fixes and multi-file architectural work.

02

11 SELECT THE SINGLE BEST ANSWER

A senior engineer put mandatory API conventions in ~/.claude/CLAUDE.md. The rules work for that engineer but not for new teammates after cloning the repository. What is the best correction?

A. Add the conventions to the shell profile that launches Claude Code so the environment variable is present in every terminal.

B. Move the shared conventions into a project-level CLAUDE.md or project .claude/CLAUDE.md under version control, and use /memory to verify the loaded hierarchy.

C. Keep the file at user level and add onboarding instructions that require every developer to copy it into the same home-directory path.

D. Create a personal slash command that injects the conventions at the beginning of each session for developers who remember to run it.

12 SELECT THE SINGLE BEST ANSWER

The team wants a /domain-review workflow to be available immediately to everyone who clones the repository. Where should its command definition live for exam-aligned behavior?

- **A** In the repository's .claude/commands/ directory, so the definition is project-scoped and version-controlled.
- **B** In each developer's ~/.claude/commands/ directory, because user-level commands are loaded before project configuration.
- **C** In .claude/skills/domain-review/SKILL.md with the same slash-command name, because skills and commands are interchangeable configuration objects.
- **D** Inside the root CLAUDE.md as a fenced prompt block named /domain-review, because always-loaded memory also defines commands.

13 SELECT THE SINGLE BEST ANSWER

Test files are colocated with source code throughout apps/, packages/, and services/. All files matching **/*.test.tsx must follow the same conventions. Which configuration is most maintainable?

- **A** Add a separate CLAUDE.md beside every test file so the correct convention is inherited from the nearest directory.
- **B** Create a test-generation skill and require developers to invoke it before editing any test file.
- **C** Put every test convention in the root CLAUDE.md and rely on Claude to infer when the section is relevant.
- **D** Create a focused rule file in .claude/rules/ with YAML frontmatter paths matching **/*.test.tsx, so the conventions load only for matching files.

14 SELECT THE SINGLE BEST ANSWER

A library migration affects 38 files and has two viable approaches with different dependency and rollout implications. What workflow is most appropriate?

- **A** Begin direct execution in the first package and let the resulting compile errors reveal which migration approach should be used elsewhere.
- **B** Ask Claude to change only the most central file, then decide whether plan mode is necessary after the first commit.
- **C** Use direct execution with a long prompt prescribing every file change before Claude has inspected the repository.
- **D** Use plan mode to map dependencies and compare approaches, obtain agreement on the design, then execute the selected plan.

15 SELECT THE SINGLE BEST ANSWER

A data-migration function is almost correct but mishandles null legacy values and inconsistent date strings. Prose instructions have produced different interpretations across attempts. What is the strongest next step?

- **A** Ask Claude to implement a generic fallback that converts any unrecognized value to an empty string so the migration always completes.
- **B** Rewrite the requirement in more emphatic prose and ask Claude to reason carefully about every possible edge case before editing.
- **C** Provide several concrete input/output examples for the ambiguous cases, add tests for those examples, and iterate using the resulting failures.
- **D** Give Claude the entire migration history and all production logs in one prompt so it has enough context to infer the intended behavior.

16 SELECT THE SINGLE BEST ANSWER

Claude follows packaging rules in some directories but ignores them in others. You suspect that a user-level file, project memory, and directory-level files are interacting. What should you do first?

A. Use /memory to inspect the active memory files for the current directory, then correct the scope or hierarchy of conflicting rules.

B. Delete all directory-level CLAUDE.md files and keep only user-level memory so every path receives exactly the same instructions.

C. Reinstall Claude Code and clear every session because inconsistent conventions usually indicate corrupted local state.

D. Add a session prompt that restates the packaging rules and avoid changing configuration until the behavior stabilizes.

17 SELECT THE SINGLE BEST ANSWER

After a long legacy-system investigation, Claude stops citing discovered classes and instead answers from 'typical repository patterns.' Which operating pattern best restores repository-specific reasoning?

A. Keep the entire investigation in one session and ask Claude to reread all previously opened files before answering each new question.

B. Persist key findings in a structured scratchpad, delegate focused exploration, and feed compact evidence-backed summaries into the main session.

C. Start a new session for every question without carrying forward any state, because fresh context is always more reliable than summaries.

D. Increase the number of always-loaded instructions in CLAUDE.md so the model gives more consistent architectural answers.

18 SELECT THE SINGLE BEST ANSWER

A named refactoring session analyzed yesterday's dependency graph. Since then, another developer has changed several central modules and moved files. What is the most reliable way to continue?

A. Start a fresh session with a structured summary of still-valid findings and an explicit list of changed areas that require re-analysis.

B. Discard all prior findings and perform a full repository exploration from scratch without providing a summary of validated work.

C. Resume the named session unchanged because session history is more complete than any manually curated summary.

D. Fork the old session twice and compare the branches, because independent branches automatically refresh any files that changed on disk.

19 SELECT THE SINGLE BEST ANSWER

A project skill performs broad dependency discovery and returns many pages of exploratory output. The result is useful, but it overwhelms the main conversation. How should the skill be configured?

A. Convert it to a user-level command so only advanced developers can invoke it, while leaving the main-context behavior unchanged.

B. Move the workflow into the root CLAUDE.md so its full instructions are always available before any development task.

C. Keep the skill in the main context but ask it to suppress intermediate output and trust it to avoid write operations.

D. Use context: fork so the skill runs in an isolated context, and restrict allowed-tools to the read-only tools required for discovery.

20 SELECT THE SINGLE BEST ANSWER

Before implementing a feature in an unfamiliar service, Claude needs to trace request flow, persistence, and test conventions. The discovery output will be large. What interaction pattern is best?

A Ask one subagent to return its complete transcript and every file listing so the main session can independently verify all details.

B Have the main session read every source and test file up front so no discovery detail is lost before implementation begins.

C Delegate focused discovery to an Explore-style subagent, require concise evidence-backed summaries, and plan implementation in the main session.

D Skip exploration and begin with the most likely implementation pattern, then repair mismatches as tests fail.

SCENARIO 3 OF 6

Multi-Agent Research System

A research platform uses a coordinator plus specialized web-search, document-analysis, synthesis, and report-generation subagents. Reports must be comprehensive, source-attributed, and explicit about conflicts and coverage gaps. Sources may be unavailable, stale, or contradictory, and subagents operate in isolated contexts unless the coordinator passes information explicitly.

03

21 SELECT THE SINGLE BEST ANSWER

Search and analysis subagents currently send tasks directly to one another. Failures are retried inconsistently, and the coordinator cannot reconstruct how a claim was produced. Which topology best addresses the problem?

A. Keep peer-to-peer communication but require every subagent to append a shared log entry after sending a task.

B. Route delegation, result aggregation, error handling, and inter-subagent communication through a hub-and-spoke coordinator.

C. Replace delegation with one fixed sequential pipeline so every query invokes all subagents in the same order.

D. Make the synthesis agent the permanent owner of all other agents because it already sees the final report requirements.

22 SELECT THE SINGLE BEST ANSWER

The synthesis subagent receives only 'write the final report' and produces uncited prose, even though earlier agents collected strong sources. What should the coordinator change?

A. Give the synthesis subagent web-search tools so it can rediscover any evidence that was not automatically inherited.

B. Store the findings in the coordinator's hidden state and reference that state by name, because subagents can resolve parent variables at invocation time.

C. Include the relevant findings and structured source metadata directly in the synthesis prompt, together with the report criteria and required claim-source mapping format.

D. Ask the synthesis subagent to add citations only after drafting, without passing source metadata until the final editing step.

23 SELECT THE SINGLE BEST ANSWER

The coordinator has identified three independent research questions. How should it start the work with the lowest avoidable orchestration latency?

A. Send one broad Task call asking a single subagent to cover all three questions so the coordinator performs fewer tool calls.

B. Invoke one Task call, wait for its result, then decide whether to start the next independent question in a later turn.

C. Emit multiple Task tool calls in the same coordinator response, each with a scoped goal, context, tools, and expected output schema.

D. Open three unrelated user sessions manually and paste their final answers into the coordinator after all sessions finish.

24 SELECT THE SINGLE BEST ANSWER

A report on AI adoption in healthcare is coherent but covers only radiology. Logs show that the coordinator assigned tasks on imaging diagnostics, radiology workflow, and image-model regulation. What is the root cause?

A. The web-search subagent needs a larger result limit because more radiology sources would eventually reveal other healthcare domains.

B. The synthesis agent failed to invent missing domains from general knowledge after receiving the radiology findings.

C. The coordinator decomposed the broad topic too narrowly, so every specialist executed a valid assignment within an incomplete research scope.

D. The document-analysis agent filtered out non-radiology sources because its relevance threshold was too strict.

25 SELECT THE SINGLE BEST ANSWER

The draft report is strong but its structured coverage_gaps field identifies no evidence for small-business impacts. What should the coordinator do next?

A. Restart the entire research workflow with broader prompts so all agents can produce a more comprehensive first pass.

B. Evaluate the gap against the objective, delegate targeted follow-up research, then re-run synthesis with the new evidence and preserved provenance.

C. Delete the coverage_gaps field before publication because exposing uncertainty reduces the perceived quality of the report.

D. Publish the current report and add a generic caveat that more research may be needed in future versions.

26 SELECT THE SINGLE BEST ANSWER

The platform supports a standardized compliance review with known stages and an open-ended investigation into an emerging technology. Which decomposition policy is best?

A. Choose the pattern only by document count: fixed below a threshold and dynamic above it, regardless of task uncertainty.

B. Use a predictable prompt chain for the standardized review and adaptive decomposition for the open-ended investigation.

C. Use the same fixed pipeline for both so orchestration, monitoring, and provenance remain identical across workloads.

D. Use dynamic decomposition for both because any model-driven workflow becomes less reliable when stages are predetermined.

27 SELECT THE SINGLE BEST ANSWER

After a shared baseline analysis of 120 sources, the team wants two independent report strategies to evolve from the same evidence without influencing one another. What should it use?

A. Start two blank sessions and have each rediscover the 120-source baseline so their reasoning is completely independent.

B. Ask one subagent to generate both strategies in a single response, then split the resulting text into two reports.

C. Continue both strategies sequentially in one session and ask Claude to forget the assumptions from the first strategy before starting the second.

D. Create separate fork_session branches from the shared baseline and give each branch its own strategy-specific instructions and evaluation criteria.

28 SELECT THE SINGLE BEST ANSWER

Before retrieval, agents need to discover which internal collections exist, including policy briefs, survey data, and regulatory archives. The catalog changes infrequently. Which MCP interface is most suitable?

- **A** Expose the collection catalog and descriptive metadata as MCP resources, while reserving tools for retrieval, filtering, and other actions.
- **B** Paste the full catalog into the coordinator system prompt on every turn so no discovery request is required.
- **C** Create one tool per collection so the model sees every possible source as a separate action in its tool list.
- **D** Force a list_collections tool call before every research request, even when the coordinator already has the current catalog.

29 SELECT THE SINGLE BEST ANSWER

Two credible sources report different market-size figures because one measures 2022 revenue and the other forecasts 2024. How should the report represent the evidence?

- **A** Average the two figures and report the mean as a balanced estimate with a lower confidence score.
- **B** Use the newer publication automatically, because recency is the most reliable way to resolve conflicting statistics.
- **C** Omit both values from the final report to avoid presenting a contradiction that readers may misinterpret.
- **D** Preserve both values with source attribution, dates, and methodology, explaining that the apparent conflict comes from different time bases.

30 SELECT THE SINGLE BEST ANSWER

The document-analysis subagent cannot access a paywalled source after two local recovery attempts, but it extracted several useful facts from an abstract. What should it return to the coordinator?

A Structured failure context describing the access problem, attempted recovery, partial findings with provenance, and suggested alternatives or coverage implications.

B Return only 'source unavailable' after retries, because detailed failure context belongs in infrastructure logs rather than agent messages.

C Propagate the raw exception to a top-level handler and terminate the entire research workflow.

D Return an empty successful result so the synthesis agent is not distracted by an error it cannot resolve.

SCENARIO 4 OF 6

Developer Productivity with Claude

An internal developer-productivity agent helps engineers explore unfamiliar repositories, trace dependencies, generate boilerplate, and perform targeted edits. It can use Read, Write, Edit, Bash, Grep, Glob, selected MCP integrations, and specialized subagents. The architecture must keep tool selection reliable, preserve context during long investigations, and separate team configuration from personal customization.

04

31 SELECT THE SINGLE BEST ANSWER

An engineer asks the agent to find every caller of calculatePremium and then locate all test files that exercise those call paths. Which tool sequence is most appropriate?

A. Read every source file first to build a complete mental model, then search the loaded context for callers and tests.

B. Use Bash for both searches because shell commands are always more precise than the built-in repository tools.

C. Use Grep for content references, Read for relevant flows, and Glob for matching test-file paths.

D. Use Glob for the function name and Grep for **/*.test.* because Glob searches repository contents while Grep matches file paths.

32 SELECT THE SINGLE BEST ANSWER

Edit cannot replace a logging block because the anchor text appears four times in the file. What is the safest fallback?

A. Use a global search-and-replace command in Bash so all four occurrences stay consistent.

B. Modify the first matching occurrence and rely on tests to reveal whether the wrong block was changed.

C. Read the full file, construct the intended complete content with the correct occurrence changed, and use Write to replace the file.

D. Run Edit again with the same anchor and instruct Claude to choose the occurrence closest to the reported line number.

33 SELECT THE SINGLE BEST ANSWER

A productivity agent exposes analyze_content and analyze_document. Their descriptions, inputs, and outputs overlap, and logs show near-random selection. What is the best interface redesign?

A. Force tool_choice to alternate between the two tools across turns so evaluation data remains balanced.

B. Merge both into analyze_anything and let one backend dispatch to the appropriate implementation from the raw user message.

C. Rename or split the tools around distinct purposes, and give each a clear contract and boundary against the alternative.

D. Keep both tools unchanged but add a system-prompt rule that selects analyze_document whenever the word 'file' appears.

34 SELECT THE SINGLE BEST ANSWER

Every subagent receives 18 tools, including deployment, ticketing, search, code editing, and report generation. A summarizer sometimes edits files, while an editor opens tickets. What is the best correction?

- **A** Force one tool call on every turn so the model spends less time choosing among the 18 available options.
- **B** Give each role only its required tools, adding narrowly constrained cross-role tools only for frequent, well-defined needs.
- **C** Create more specialized subagents but preserve the same 18-tool set so routing, rather than permissions, determines behavior.
- **D** Keep all tools available and strengthen each system prompt with a list of tools the role should avoid.

35 SELECT THE SINGLE BEST ANSWER

The repository needs a shared Jira MCP integration, but each developer must supply a personal token without committing secrets. Which configuration is best?

- **A** Store the token directly in .mcp.json so every clone has a complete, reproducible configuration.
- **B** Describe the server endpoint and token variable in CLAUDE.md, then ask Claude to construct the MCP connection dynamically when needed.
- **C** Configure the Jira server only in each developer's ~/.claude.json because authentication is personal even when the integration is team-standard.
- **D** Commit the shared server in project .mcp.json, reference an environment token, and keep experimental servers at user scope.

36 SELECT THE SINGLE BEST ANSWER

Before any enrichment tool runs, the agent must call extract_repo_metadata exactly once to establish language, package manager, and framework. After that, Claude may choose tools freely. Which configuration fits?

A. Use tool_choice: auto and emphasize the metadata requirement in the prompt so Claude can skip it when the repository appears obvious.

B. Force extract_repo_metadata first, return its tool_result, then continue with tool_choice: auto.

C. Force extract_repo_metadata on every request so the prerequisite remains visible throughout the workflow.

D. Use tool_choice: any for the first request because any guarantees that the model will select the required metadata tool.

37 SELECT THE SINGLE BEST ANSWER

A developer wants a more verbose personal variant of the team's code-review skill without changing team behavior. What should the developer do?

A. Create a user-scoped skill with a distinct name in ~/.claude/skills/, preserving the project skill unchanged.

B. Create a user-level skill with the same name so it silently overrides the project skill in every context.

C. Put the personal preference in the root CLAUDE.md under a heading that mentions the developer's name.

D. Edit the project SKILL.md and rely on teammates to ignore the additional verbosity when they do not need it.

38 SELECT THE SINGLE BEST ANSWER

An engineer asks Claude to design a cache for an unfamiliar legacy service. Requirements do not specify consistency, invalidation, failure handling, or traffic patterns. What is the best first interaction?

A. Ask targeted questions that surface hidden constraints and tradeoffs before proposing an implementation.

B. Ask whether the engineer prefers Redis, and treat the selected product as the main architectural requirement.

C. Provide a long generic explanation of caching strategies and ask the engineer to choose one without investigating the service.

D. Implement a standard cache-aside pattern immediately, then use test failures to discover any missing requirements.

39 SELECT THE SINGLE BEST ANSWER

The request is 'add comprehensive tests to this legacy billing module.' The module has unclear boundaries, hidden dependencies, and no test plan. What decomposition should Claude use?

A. Create snapshot tests for all public outputs before reading the implementation, because snapshots maximize coverage quickly.

B. Generate one test for every file in alphabetical order so coverage grows predictably and no file is skipped.

C. Use a fixed prompt chain of unit tests, integration tests, and performance tests without changing the plan after repository discovery.

D. Map dependencies and risks first, then build and revise a prioritized test plan as evidence emerges.

40 SELECT THE SINGLE BEST ANSWER

The monorepo root CLAUDE.md has grown into a large file containing package-specific API, testing, and deployment rules. Only some packages need each standards document. What is the most maintainable organization?

A. Keep universal rules at the project root, then use package-level CLAUDE.md files with @import references to the specific shared standards each package needs.

B. Keep every rule in the root CLAUDE.md and add a table of contents so Claude can identify which sections apply to the current package.

C. Move package-specific standards into each maintainer's user-level CLAUDE.md so only the developers working on that package load them.

D. Convert every standards document into a skill and require developers to invoke the relevant skills before editing a package.

SCENARIO 5 OF 6

Claude Code for Continuous Integration

A CI pipeline invokes Claude Code for pull-request review, security analysis, and test generation. Blocking checks must finish during the merge workflow, while some audits can run asynchronously. Findings are posted automatically, so output must be machine-parseable, actionable, and calibrated to minimize false positives that would erode developer trust.

05

41 SELECT THE SINGLE BEST ANSWER

A pipeline pipes a pull-request diff into Claude Code, but the job waits indefinitely for interactive input. Which invocation pattern is required?

A Use claude -p or --print so the command runs non-interactively, emits output, and exits.

B Use a --batch flag for the CLI so the job is submitted asynchronously and the pipeline can poll for completion.

C Set a CLAUDE_HEADLESS environment variable before the command so the same prompt syntax works in CI.

D Redirect stdin from /dev/null while keeping the normal interactive command, because the absence of input forces Claude Code to finish.

42 SELECT THE SINGLE BEST ANSWER

The review job must post inline comments with file, line, severity, issue, and suggested fix. Free-form Markdown intermittently breaks the parser. What should the job configure?

A. Return a raw transcript of the review session and have a second model extract comment fields after the job finishes.

B. Ask Claude to emit a comma-separated line per finding and split the result in the pipeline.

C. Keep Markdown but add page numbers and stronger headings so the parser can recover fields from a more consistent visual layout.

D. Use -p with --output-format json and --json-schema so the CI consumer receives a defined machine-readable structure.

43 SELECT THE SINGLE BEST ANSWER

Developers ignore the review bot because it reports harmless style differences alongside real defects. Which prompt change is most likely to restore precision?

A. Require a confidence score above 0.9 for every finding but leave the existing issue criteria unchanged.

B. Increase the prompt length with more general software-engineering best practices so the bot has a broader basis for critique.

C. Ask the model to be conservative and report only issues it considers high confidence.

D. Define reportable and excluded categories, concrete severity boundaries, and examples of local patterns that must not be flagged.

44 SELECT THE SINGLE BEST ANSWER

The bot inconsistently decides whether a missing branch test is a meaningful coverage gap or an acceptable omission. Detailed prose has not stabilized the judgment. What is the best next step?

A. Add one positive example of a missing test and ask the model to generalize the same pattern to all branches.

B. Lower the sampling temperature so repeated runs are more deterministic even though the decision criteria remain ambiguous.

C. List every possible branching construct in the codebase and instruct the model to flag each one unless a matching test name is found.

D. Add a small set of targeted examples showing genuine branch-level gaps, acceptable omissions, and the exact finding format expected.

45 SELECT THE SINGLE BEST ANSWER

The same Claude session generates a patch and then reviews it. Subtle assumptions from generation survive the role switch, and defects are missed. What should the pipeline change?

A. Expose the generator's confidence and route only low-confidence files to a second reviewer.

B. Use a separate review instance that receives the patch and relevant repository context but not the generator's reasoning history.

C. Keep the same session and add a system message instructing Claude to become a skeptical reviewer before reading the patch.

D. Ask the generator to perform extended self-critique twice and accept only defects found in both passes.

46 SELECT THE SINGLE BEST ANSWER

A 22-file pull request receives deep comments on early files, shallow comments on later files, and contradictory treatment of the same pattern. How should the review be restructured?

A. Use one larger-context request containing all files and ask the model to allocate equal attention to every file.

B. Run focused per-file passes for local issues, followed by a separate integration pass for cross-file data flow and consistency.

C. Require developers to split every large pull request into four-file chunks before the bot runs.

D. Run three complete reviews of the full pull request and report only findings that appear in at least two runs.

47 SELECT THE SINGLE BEST ANSWER

The organization has a blocking pre-merge security check and an overnight technical-debt audit. It wants lower cost without violating workflow latency. What is the best API split?

A. Submit the pre-merge check to a batch and fall back to a synchronous request after a short timeout if the result has not arrived.

B. Move both workflows to Message Batches and poll aggressively so the pre-merge result is likely to arrive before developers become impatient.

C. Keep the pre-merge check synchronous; use the Message Batches API for the overnight audit, correlate results with custom_id, and resubmit only failed items.

D. Keep both workflows synchronous because batch result ordering makes it unsafe to associate responses with the original documents.

48 SELECT THE SINGLE BEST ANSWER

The team wants to discover which code constructs repeatedly trigger dismissed findings. Which schema change most directly supports that analysis?

A. Record the model temperature and mode for each review so analysts can correlate dismissals with generation randomness.

B. Add a reviewer_sentiment field that classifies whether the dismissal comment sounded frustrated, neutral, or positive.

C. Add a detected_pattern field that records the concrete heuristic or construct associated with each finding.

D. Store the model's private reasoning chain alongside every finding so reviewers can inspect how the conclusion was reached.

49 SELECT THE SINGLE BEST ANSWER

Security findings are usually accurate, but comment-accuracy findings are dismissed 70% of the time and are causing developers to ignore the bot. What is the best short-term reliability move?

A. Keep every category enabled so the evaluation dataset remains representative, but add a disclaimer that some findings may be noisy.

B. Temporarily disable or suppress the noisy category while refining its criteria and evaluation, leaving the trustworthy categories active.

C. Merge all findings into one severity level so developers cannot distinguish the category with the high dismissal rate.

D. Raise one global confidence threshold for all finding types so the overall number of comments decreases.

50 SELECT THE SINGLE BEST ANSWER

The review bot always returns schema-valid JSON, but it sometimes labels low-impact issues as critical. What conclusion is correct?

A. The schema guarantees syntax and shape, not semantic judgment; severity still needs explicit criteria, examples, calibration, and possibly downstream validation.

B. The JSON schema is defective, because a schema-compliant output should also guarantee that every field value is factually correct.

C. The team should abandon structured output and use prose, because classification errors prove that schemas distort reasoning.

D. tool_choice should be changed from forced to auto so Claude can return explanatory text when it is uncertain about severity.

SCENARIO 6 OF 6

Structured Data Extraction

A document-processing service extracts invoices, contracts, and narrative reports into structured records. Documents vary widely, fields may be absent, source sections may conflict, and downstream systems require reliable schemas. The workflow uses validation, targeted retries, batch processing, human review, and provenance-aware outputs.

06

51 SELECT THE SINGLE BEST ANSWER

Free-form invoice extraction occasionally includes commentary or malformed JSON, breaking ingestion. According to the exam guide's structured-output framing, which design is most reliable?

A. Wrap the response in fenced code blocks and strip all text before the first brace and after the last brace.

B. Prompt Claude to return only valid JSON, set temperature to zero, and reject any response containing markdown.

C. Define a schema-backed extraction tool and consume its tool_use arguments, using tool_choice when the call must be guaranteed.

D. Accept natural-language output and use a downstream parser to infer fields because a second pass can repair malformed structures.

52 SELECT THE SINGLE BEST ANSWER

Some contracts have no renewal term. The current schema requires renewal_date as a string, and the model invents plausible dates. What is the best schema change?

A. Require the model to calculate renewal_date from signature date and contract length whenever the source omits it.

B. Keep the field required and add a stronger instruction that fabricated dates are prohibited.

C. Remove renewal_date from the schema entirely so the extractor cannot hallucinate it.

D. Make renewal_date optional or nullable and return null when the source does not contain it.

53 SELECT THE SINGLE BEST ANSWER

Validation reports that supplier_tax_id is missing. A human check confirms the document contains no tax identifier. How should the retry controller respond?

A. Copy another identifier, such as the invoice number, into supplier_tax_id so the record passes schema validation and can be corrected later.

B. Do not retry to manufacture the value; return null or route the record according to the workflow's missing-data and human-review policy.

C. Retry with progressively stronger wording until the model fills the required field, because validation failures should always trigger self-correction.

D. Ask the model to search external sources for the supplier's tax ID and insert it into the extraction result.

54 SELECT THE SINGLE BEST ANSWER

The extractor performs well on standardized forms but misses measurements embedded informally in narrative reports. Which prompt improvement is most likely to generalize?

A. Make every field required so the model searches more aggressively for values in unstructured text.

B. Tell Claude to be more creative and infer likely measurements whenever the document seems incomplete.

C. Add a long dictionary of every known field-name synonym and require exact phrase matching before a value is extracted.

D. Add targeted few-shot examples that demonstrate extraction from varied layouts and informal measurement phrasing while preserving the same output schema.

55 SELECT THE SINGLE BEST ANSWER

A 40-page source is followed by extraction instructions. Important exception clauses in the middle are missed more often than information near the beginning and end. How should the input be reorganized?

A. Surface a key-facts and exceptions block early, preserve exact numbers and dates, and organize detailed evidence under explicit section headings.

B. Move all instructions to the very end, because the final part of a long context is always processed accurately.

C. Randomize document section order on each request so no clause is consistently placed in the middle.

D. Compress amounts, dates, and percentages into approximate prose so the summary occupies fewer tokens.

56 SELECT THE SINGLE BEST ANSWER

The system reports one confidence score per document. Overall accuracy is high, but handwritten addenda have poor termination-date accuracy. What review-routing redesign is strongest?

A. Remove confidence scores and route only schema-invalid records to humans, because valid structured output is sufficient for automation.

B. Produce field-level confidence, calibrate thresholds on labeled validation data, and evaluate accuracy by document type and field before reducing review.

C. Raise the single document-level threshold for every record until the handwritten-addendum error rate becomes acceptable.

D. Keep the aggregate confidence score but add a rule that all handwritten documents receive the same fixed penalty.

57 SELECT THE SINGLE BEST ANSWER

High-confidence extractions will be auto-approved, but the team still needs to measure rare errors across document types and fields. Which sampling strategy is best?

A. Review the first fixed number of records processed each week, regardless of source type or field distribution.

B. Stop sampling once overall accuracy exceeds the target for one month, because continued review no longer changes the deployment decision.

C. Stratify random samples of high-confidence outputs by document type and field, tracking errors and new patterns over time.

D. Review only the lowest-confidence records because high-confidence outputs have already passed the model's reliability threshold.

58 SELECT THE SINGLE BEST ANSWER

A contract summary states a value of $1.2M, while the signed appendix states $1.4M. Both sections are credible and downstream users must reconcile the discrepancy. What should the output contain?

A. Use the appendix value only because signed sections are generally more authoritative, and omit the conflicting summary to simplify downstream processing.

B. Both values with source locations, conflict_detected metadata, relevant dates or section context, and enough provenance for downstream reconciliation or human review.

C. Use the first value encountered and add a generic warning that the document may contain inconsistencies.

D. Average the values and report $1.3M with low confidence so the record contains one normalized number.

59 SELECT THE SINGLE BEST ANSWER

The agent repeatedly calls list_templates with different filters just to discover which of hundreds of templates and policy references exist. What MCP redesign reduces exploratory calls?

A. Expose the template and policy catalog as MCP resources, while keeping tools for retrieval, validation, and actions.

B. Turn every template into a separate tool so the model can call the exact document type it recognizes.

C. Hide the catalog and require the model to infer the template from document text without consulting external metadata.

D. Force list_templates before every extraction so the model always starts with a fresh view of the catalog.

60 SELECT THE SINGLE BEST ANSWER

A retrieval tool returns [] both when no matching document exists and when its repository token has expired. What should change?

A Throw a top-level exception for both cases so any uncertainty is routed directly to human review.

B Return a generic not_found error for both situations so the agent retries with a broader query.

C Keep returning [] and let the agent ask the user whether the repository might be unavailable.

D Return empty success only for a valid no-match; use structured error metadata for access or permission failures.