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

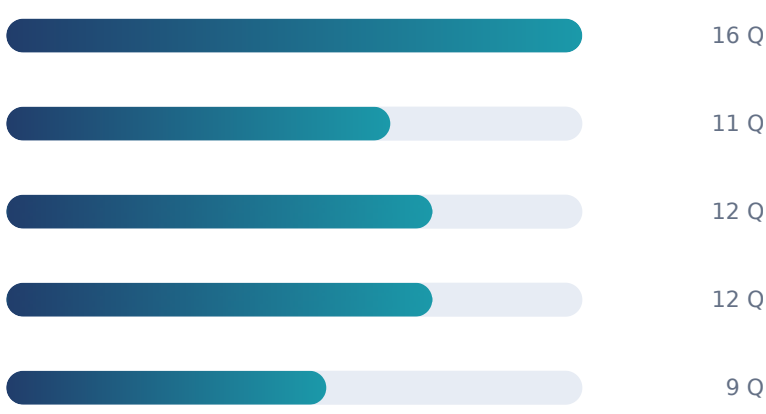
Closed-book first pass, then domain-by-domain review

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.

SOLUTIONS EDITION

Rationales, Domain Mapping, and Exam Traps

Detailed analysis for all 60 questions, including every distractor

60	240	5	v1.0.1
Rationales	Options analyzed	Domains	Release

PA

Created and curated by Paolo Amato

Created and curated by Paolo Amato for disciplined post-exam review.

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

Solutions and Rationales

Use this edition after completing the candidate exam. Each item identifies the guide-aligned domain and a paraphrased task focus, states the governing decision rule, explains the correct option, and diagnoses every distractor.

> **Review method.** For every miss, write the rule in your own words and identify the phrase in the scenario that makes the correct option better than the most plausible distractor.

Answer Key

Q	Answer	Q	Answer	Q	Answer	Q	Answer
1	A	16	A	31	C	46	B
2	C	17	B	32	C	47	C
3	A	18	A	33	C	48	C
4	B	19	D	34	B	49	B
5	B	20	C	35	D	50	A
6	A	21	B	36	B	51	C
7	D	22	C	37	A	52	D
8	D	23	C	38	A	53	B
9	C	24	C	39	D	54	D
10	C	25	B	40	A	55	A
11	B	26	B	41	A	56	B
12	A	27	D	42	D	57	C
13	D	28	A	43	D	58	B
14	D	29	D	44	D	59	A
15	C	30	A	45	B	60	D

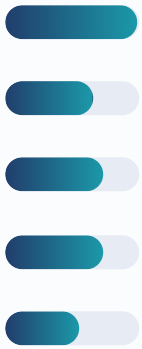
> **Interpretation.** No official raw-to-scaled score conversion is available for this practice set. Use results only for domain-level diagnosis. A high raw score can still conceal a weak task statement, so review the domain and task mapping rather than relying on the total alone.

POST-EXAM ANALYSIS

Review Dashboard

Use three passes: verify correctness, diagnose the decision error, then convert the lesson into a reusable rule.

Domain Counts



Three-Pass Review

1. **Correctness**
   Read the rationale before checking your own explanation.
2. **Decision boundary**
   Identify why the best distractor is useful but still wrong here.
3. **Retention**
   Write one rule and one anti-pattern in your own words.

Useful Error Labels

Wrong layer
A prompt fix was chosen for a code-enforceable invariant, or infrastructure was added before fixing an interface.

Probabilistic vs deterministic
Guidance was mistaken for a guarantee when the requirement was mandatory.

Over-engineering
A classifier, router, or new service was selected before the lower-cost root-cause correction.

Context or provenance loss
The option discarded facts, source mappings, partial results, or recovery detail needed downstream.

Scope confusion
User, project, directory, tool, or subagent boundaries were mixed up.

Version drift
Current product terminology was substituted for the stated exam-guide framing.

Best use of the rationales. Do not merely confirm the correct letter. Explain why the strongest alternative fails under the exact scenario constraints.

Scenario 1 - Customer Support Resolution Agent

Questions 1-10 | Review the decision rule before memorizing the answer letter.

Question 1

Correct answer: A

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?

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.1 - Agentic loop control flow

> **Decision rule:** Control the loop with stop_reason, not assistant prose.

Why A is correct: A tool_use stop means the model is waiting for tool results. The runtime executes the calls, returns tool_result content in history, and lets Claude decide the next action; end_turn is the normal completion signal.

A **Execute all requested tools, append matching tool_result blocks to history, and call Claude again until stop_reason is end_turn.**
Correct option; see the question-specific rationale above.

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.**
The runtime must honor all tool_use blocks returned in that response. Treating them as alternatives silently drops model-requested work.

C **Continue until either the assistant writes an explicit completion phrase or a fixed iteration ceiling is reached, regardless of stop_reason.**
Natural-language phrases and arbitrary caps are not the primary completion signal. The exam rule is to continue on tool_use and terminate normally on end_turn.

D **Send the preamble to the customer immediately, store the tool requests for the next user turn, and resume only if the customer replies.**
Text preceding a tool call does not override stop_reason. Deferring requested tools breaks the agentic loop and withholds information Claude expects on the next turn.

Exam trap: Do not mistake conversational text for completion when the protocol-level stop_reason says tool_use.

Question 2

Correct answer: C

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?

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.4 / 1.5 - Deterministic enforcement and hooks

> **Decision rule:** Use deterministic enforcement for mandatory business prerequisites.

Why C is correct: A hook or programmatic gate can prevent the prohibited call from executing at all. Prompt instructions and examples remain useful guidance, but they are probabilistic and therefore insufficient for a must-not-fail rule.

A **Remove process_refund from the agent and have it draft refund instructions for a human to execute in every case.**
This avoids the risk by eliminating autonomous resolution, but it sacrifices the stated first-contact-resolution objective rather than enforcing the required gate.

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.**
Forcing the first call does not enforce the business condition. A failed or ambiguous lookup could still expose the refund tool prematurely.

C **Enforce a programmatic precondition or hook that blocks process_refund unless verified customer state exists, with a safe recovery path.**
Correct option; see the question-specific rationale above.

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.**
Prompting can reduce failure probability, but it cannot provide the deterministic guarantee required for a financial prerequisite.

Exam trap: The most persuasive prompt is still not a deterministic control.

Question 3

Correct answer: A

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?

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.4 - Multi-concern workflow decomposition

> **Decision rule:** Decompose multi-concern requests, preserve shared facts, and synthesize a unified outcome.

Why A is correct: The agent should track each concern explicitly, investigate them without losing verified context, and ensure the final response covers every item. Escalation is reserved for policy gaps, explicit human requests, or inability to progress.

A **Represent the message as separate concerns, investigate independent concerns using the shared verified case facts, then synthesize one complete resolution or structured handoff.**
Correct option; see the question-specific rationale above.

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.**
Stopping after one concern recreates the failure: the customer's remaining issues are silently abandoned.

C **Open three isolated subagent conversations and return their answers directly to the customer without sharing identity, order facts, or policy context between them.**
Isolated investigations need explicit shared facts and coordinator synthesis; otherwise they can duplicate work, contradict one another, or lose verification state.

D **Escalate the entire case immediately because multiple concerns imply that autonomous resolution is no longer appropriate.**
Multiplicity alone is not an escalation trigger. If the agent has policy coverage and tools, it should resolve what it can and escalate only the blocked or ambiguous portion.

Exam trap: Do not confuse a complex-looking message with an automatic need for human escalation.

Question 4

Correct answer: B

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?

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.4 - Structured handoff protocols

> **Decision rule:** A handoff must be actionable for a reviewer who lacks the transcript.

Why B is correct: Structured fields preserve the facts, provenance, attempted work, and decision boundary while excluding irrelevant context. That lets a human continue rather than restart the case.

A **The full transcript and every raw tool response, including unrelated backend fields, so the human can reconstruct the case without information loss.**
Raw dumps preserve volume rather than relevance. They increase review burden and can obscure the facts the human needs to act.

B **Send a structured handoff with verified identity, issue status, key order facts, evidence, attempted actions, the policy gap, and a recommendation.**
Correct option; see the question-specific rationale above.

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.**
Sentiment and generic confidence do not replace verified facts, attempted actions, or the reason autonomous resolution stopped.

D **Only the unresolved billing request and the last tool error, omitting completed investigations so the handoff remains concise.**
Omitting completed work creates duplication and may hide facts that change the human's decision. Concision should come from structure, not missing context.

Exam trap: More context is not automatically better context; relevance and structure matter.

Question 5

Correct answer: B

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?

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.5 - PostToolUse normalization

> **Decision rule:** Normalize heterogeneous tool results deterministically before model reasoning.

Why B is correct: PostToolUse is the appropriate interception point for transforming inconsistent backend outputs into one stable representation. This removes a recurring reasoning burden from the model.

A **Expose a normalize_data tool and allow the model to call it when it notices that two results are not directly comparable.**
This leaves detection and invocation to the same model that is already missing the inconsistency in some cases.

B **Normalize every result into a canonical schema in a PostToolUse hook before adding it to model context.**
Correct option; see the question-specific rationale above.

C **Document every source format in the system prompt and instruct the model to convert values before making a decision.**
The model may still convert inconsistently. The requirement is better met by deterministic normalization before reasoning begins.

D **Return longer human-readable explanations from each backend so Claude can infer the meaning of each timestamp and status code.**
Longer outputs increase context cost and still require probabilistic interpretation. They do not create a canonical representation.

Exam trap: Do not use prompt text to solve a data-contract problem that can be normalized in code.

Question 6

Correct answer: A

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?

Domain: Domain 2 - Tool Design & MCP Integration

Task: Task 2.1 - Tool descriptions and boundaries

> **Decision rule:** Clear descriptions and non-overlapping boundaries are the primary tool-selection controls.

Why A is correct: Models use tool names and descriptions to choose among similar capabilities. Improving the interface is the lowest-complexity, highest-leverage first fix before adding routing infrastructure.

A **Clarify names and descriptions with distinct purposes, accepted identifiers, output contracts, examples, edge cases, and explicit boundaries.**
Correct option; see the question-specific rationale above.

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.**
A keyword rule is brittle: dates, phone numbers, and mixed requests can contain digits without representing an exact order lookup.

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.**
Forced misrouting adds avoidable calls and does not address the root cause: unclear tool boundaries.

D **Wrap both operations behind one generic lookup_records tool and let the backend infer whether the user intended an order or customer search.**
A generic interface can hide ambiguity rather than resolve it, and it removes useful semantic distinctions from the model's tool-selection surface.

Exam trap: Do not over-engineer a router before fixing ambiguous tool contracts.

Question 7

Correct answer: D

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?

Domain: Domain 2 - Tool Design & MCP Integration

Task: Task 2.2 - Structured MCP error responses

> **Decision rule:** Differentiate business, validation, permission, and transient failures with actionable structure.

Why D is correct: The agent needs to know that the failure is non-retryable and policy-based so it can stop retrying, explain the outcome, and select an allowed next action.

A **Return the raw backend exception, policy table name, and internal rule ID so the model can derive a customer-facing explanation.**
Raw internal exceptions may expose implementation details and still fail to provide a clear recovery contract.

B **Return a transient category with isRetryable: true and a recommended backoff, because policy data might change later.**
The current request violates a stable policy rule. Retrying the same operation wastes calls and may create inconsistent customer messaging.

C **Return an empty successful result with refund_created: false, because the backend completed normally even though no refund was issued.**
A business-rule rejection is a failure the model must understand. Marking it as success hides the reason and can lead to incorrect downstream behavior.

D **Return isError: true with a non-retryable business-policy category, a readable explanation, and safe details for the next allowed action.**
Correct option; see the question-specific rationale above.

Exam trap: A failed action is not a successful empty result, and a policy rejection is not transient.

Question 8

Correct answer: D

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?

Domain: Domain 5 - Context Management & Reliability

Task: Task 5.2 - Ambiguity resolution and escalation

> **Decision rule:** When multiple customer records match, ask for clarification rather than selecting heuristically.

Why D is correct: An additional identifier resolves the ambiguity with minimal risk. The agent should not expose or act on account data until it knows which record belongs to the user.

A **Choose the account with the most recent order, but limit the first action to a read-only lookup before requesting confirmation.**
A read-only action can still disclose another customer's information. Heuristic selection is not acceptable when identity is unresolved.

B **Escalate immediately to a human because any duplicate-name result is a policy exception that the agent cannot handle.**
The ambiguity can usually be resolved by asking for another identifier; escalation is unnecessary unless clarification fails or policy requires it.

C **Query all three accounts, compare order details with the user's description, and use the best match as implicit verification.**
Cross-account investigation increases privacy risk and substitutes inference for explicit disambiguation.

D **Ask for an additional identifier and avoid account-specific lookups or actions until the customer record is unambiguously resolved.**
Correct option; see the question-specific rationale above.

Exam trap: Low-risk-looking access can still be privacy-sensitive when the identity match is uncertain.

Question 9

Correct answer: C

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?

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.6 - Task decomposition strategy

> **Decision rule:** Use dynamic decomposition for variable investigations and deterministic gates for non-negotiable controls.

Why C is correct: Support cases differ substantially, so the agent should select only relevant steps and revise its plan as evidence arrives. Safety-critical prerequisites remain enforced outside the model's discretion.

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.**
Parallelizing irrelevant work hides some latency but increases cost and can still produce conflicting or distracting context.

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.**
A fixed initial route cannot adapt when intermediate findings reveal additional concerns or a different root cause.

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.**
Correct option; see the question-specific rationale above.

D **Keep the fixed pipeline because a predictable sequence is always more reliable than model-directed tool selection, even when most steps are irrelevant.**
Fixed chains are useful for uniform workflows, but here they create unnecessary latency and cost without improving the required safety gates.

Exam trap: Do not equate predictability with efficiency, or adaptability with the absence of guardrails.

Question 10

Correct answer: C

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?

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.3 - Subagent invocation and context passing

> **Decision rule:** Subagents have isolated context and should receive only the information and tools required for their role.

Why C is correct: Explicit context passing preserves correctness and provenance, while scoped tools reduce accidental cross-role behavior. The coordinator remains responsible for orchestration and final decisions.

A **Store the case facts in the first subagent invocation and rely on that hidden state for later invocations of the same specialist type.**
Separate subagent invocations do not share memory automatically. Each call must receive the context it needs.

B **Invoke the subagent with only a ticket ID because subagents automatically inherit the coordinator's conversation and verified state.**
Subagents have isolated context. Without explicit facts, the specialist cannot reliably know what the coordinator already established.

C **Pass verified facts, policy evidence, provenance, constraints, and output criteria explicitly; restrict the specialist to policy-review tools.**
Correct option; see the question-specific rationale above.

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.**
Over-provisioning increases selection risk and violates role scoping. Missing context should be passed explicitly, not rediscovered with sensitive tools.

Exam trap: Do not assume context inheritance or compensate for missing context by giving a specialist every tool.

Scenario 2 - Code Generation with Claude Code

Questions 11-20 | Review the decision rule before memorizing the answer letter.

Question 11

Correct answer: B

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?

Domain: Domain 3 - Claude Code Configuration & Workflows

Task: Task 3.1 - CLAUDE.md hierarchy and scope

> **Decision rule:** Team-wide instructions belong in project-scoped configuration; personal preferences belong at user scope.

Why B is correct: Project-level CLAUDE.md files are version-controlled and available to the whole team. /memory helps diagnose which memory files are actually loaded.

A **Add the conventions to the shell profile that launches Claude Code so the environment variable is present in every terminal.**
Shell configuration is not the documented mechanism for project memory and is still user-specific.

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.**
Correct option; see the question-specific rationale above.

C **Keep the file at user level and add onboarding instructions that require every developer to copy it into the same home-directory path.**
Manual copying creates drift and is not automatically shared through the repository.

D **Create a personal slash command that injects the conventions at the beginning of each session for developers who remember to run it.**
On-demand personal commands do not satisfy an always-loaded, team-wide requirement.

Exam trap: A rule that exists only in one developer's home directory is not a repository standard.

Question 12

Correct answer: A

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?

Domain: Domain 3 - Claude Code Configuration & Workflows

Task: Task 3.2 - Project commands and skills

> **Decision rule:** Project-scoped commands are stored in .claude/commands/ and travel with the repository.

Why A is correct: The project directory makes the command available to teammates through version control without requiring personal setup.

A **In the repository's .claude/commands/ directory, so the definition is project-scoped and version-controlled.**
Correct option; see the question-specific rationale above.

B **In each developer's ~/.claude/commands/ directory, because user-level commands are loaded before project configuration.**
User-level commands are personal and are not distributed with the repository.

C **In .claude/skills/domain-review/SKILL.md with the same slash-command name, because skills and commands are interchangeable configuration objects.**
Skills and commands serve related but distinct workflows in the exam guide. The question asks specifically for a shared custom command.

D **Inside the root CLAUDE.md as a fenced prompt block named /domain-review, because always-loaded memory also defines commands.**
CLAUDE.md supplies instructions and context; it is not the command-definition directory described by the guide.

Exam trap: Do not confuse always-loaded memory, on-demand skills, and custom command definitions.

Question 13

Correct answer: D

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?

Domain: Domain 3 - Claude Code Configuration & Workflows

Task: Task 3.3 - Path-specific rules

> **Decision rule:** Use path-scoped rules when conventions apply by file pattern across many directories.

Why D is correct: A glob-scoped rule activates only for matching files, reducing irrelevant context while centralizing the convention.

A **Add a separate CLAUDE.md beside every test file so the correct convention is inherited from the nearest directory.**
Test files are spread across many directories, so this creates duplication and maintenance drift.

B **Create a test-generation skill and require developers to invoke it before editing any test file.**
A manually invoked skill does not automatically enforce conventions whenever a matching file is edited.

C **Put every test convention in the root CLAUDE.md and rely on Claude to infer when the section is relevant.**
This loads irrelevant guidance into every task and depends on inference rather than an explicit path condition.

D **Create a focused rule file in .claude/rules/ with YAML frontmatter paths matching **/*.test.tsx, so the conventions load only for matching files.**
Correct option; see the question-specific rationale above.

Exam trap: Do not use directory hierarchy to model a rule that is fundamentally based on file type.

Question 14

Correct answer: D

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

Domain: Domain 3 - Claude Code Configuration & Workflows

Task: Task 3.4 - Plan mode versus direct execution

> **Decision rule:** Use plan mode for multi-file, architectural, or multi-approach work; use direct execution for small, clear changes.

Why D is correct: Plan mode supports safe exploration and tradeoff analysis before irreversible edits. Implementation can then proceed with a coherent approach.

A **Begin direct execution in the first package and let the resulting compile errors reveal which migration approach should be used elsewhere.**
The architecture and scope are already known to be complex. Starting implementation first risks rework and inconsistent changes.

B **Ask Claude to change only the most central file, then decide whether plan mode is necessary after the first commit.**
The need for planning follows from the stated multi-file and architectural uncertainty, not from whether the first edit succeeds.

C **Use direct execution with a long prompt prescribing every file change before Claude has inspected the repository.**
An unverified upfront prescription cannot account for actual dependency structure or hidden constraints.

D **Use plan mode to map dependencies and compare approaches, obtain agreement on the design, then execute the selected plan.**
Correct option; see the question-specific rationale above.

Exam trap: Do not treat plan mode as a fallback after predictable complexity has already caused rework.

Question 15

Correct answer: C

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?

Domain: Domain 3 - Claude Code Configuration & Workflows

Task: Task 3.5 - Iterative refinement

> **Decision rule:** Use concrete examples and test-driven feedback when prose is interpreted inconsistently.

Why C is correct: Examples make the transformation contract explicit, while tests turn each failure into targeted evidence for the next iteration.

A **Ask Claude to implement a generic fallback that converts any unrecognized value to an empty string so the migration always completes.**
A blanket fallback can silently corrupt data and does not encode the intended semantics for nulls or invalid dates.

B **Rewrite the requirement in more emphatic prose and ask Claude to reason carefully about every possible edge case before editing.**
More prose may remain ambiguous. Concrete transformations and executable tests provide clearer, verifiable feedback.

C **Provide several concrete input/output examples for the ambiguous cases, add tests for those examples, and iterate using the resulting failures.**
Correct option; see the question-specific rationale above.

D **Give Claude the entire migration history and all production logs in one prompt so it has enough context to infer the intended behavior.**
More context does not guarantee a consistent interpretation and may dilute the specific edge cases that need correction.

Exam trap: Do not respond to ambiguity by adding more vague text or more unrelated context.

Question 16

Correct answer: A

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?

Domain: Domain 3 - Claude Code Configuration & Workflows

Task: Task 3.1 - Diagnosing loaded memory

> **Decision rule:** Diagnose the loaded memory hierarchy before changing files.

Why A is correct: /memory reveals the active configuration sources, allowing the team to move, remove, or narrow only the rules responsible for the inconsistency.

A **Use /memory to inspect the active memory files for the current directory, then correct the scope or hierarchy of conflicting rules.**
Correct option; see the question-specific rationale above.

B **Delete all directory-level CLAUDE.md files and keep only user-level memory so every path receives exactly the same instructions.**
This removes potentially valid scoped conventions before diagnosing which files are actually active.

C **Reinstall Claude Code and clear every session because inconsistent conventions usually indicate corrupted local state.**
The described symptom is a scope and hierarchy issue, not evidence of installation corruption.

D **Add a session prompt that restates the packaging rules and avoid changing configuration until the behavior stabilizes.**
A one-off prompt masks the configuration problem rather than identifying the source of inconsistent loading.

Exam trap: Do not delete or duplicate configuration before confirming what Claude actually loaded.

Question 17

Correct answer: B

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?

Domain: Domain 5 - Context Management & Reliability

Task: Task 5.4 - Large codebase context management

> **Decision rule:** Persist repository-specific facts outside the transient conversation and isolate verbose exploration.

Why B is correct: Scratchpads and summaries preserve concrete discoveries, while subagents keep noisy exploration from crowding out the main coordination context.

A **Keep the entire investigation in one session and ask Claude to reread all previously opened files before answering each new question.**
Repeatedly rereading everything consumes context and worsens the degradation the design is trying to solve.

B **Persist key findings in a structured scratchpad, delegate focused exploration, and feed compact evidence-backed summaries into the main session.**
Correct option; see the question-specific rationale above.

C **Start a new session for every question without carrying forward any state, because fresh context is always more reliable than summaries.**
Fresh sessions help only if they receive a curated state summary; otherwise the system repeatedly rediscovers the same facts.

D **Increase the number of always-loaded instructions in CLAUDE.md so the model gives more consistent architectural answers.**
The issue is loss of discovered repository facts, not a shortage of generic instructions.

Exam trap: Do not try to solve context degradation by putting still more raw exploration into the same context.

Question 18

Correct answer: A

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?

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.7 - Session resumption versus fresh context

> **Decision rule:** Resume only when prior context is mostly valid; otherwise start fresh with a curated state transfer.

Why A is correct: Substantial code changes make old tool results unreliable. A new session with explicit valid facts and change boundaries preserves value without carrying stale assumptions.

A **Start a fresh session with a structured summary of still-valid findings and an explicit list of changed areas that require re-analysis.**
Correct option; see the question-specific rationale above.

B **Discard all prior findings and perform a full repository exploration from scratch without providing a summary of validated work.**
A fresh session is appropriate, but throwing away still-valid findings creates unnecessary cost and loses useful context.

C **Resume the named session unchanged because session history is more complete than any manually curated summary.**
The history contains stale tool results and assumptions that can mislead the resumed agent.

D **Fork the old session twice and compare the branches, because independent branches automatically refresh any files that changed on disk.**
Forking preserves the same stale baseline unless the new branches are explicitly told what changed and re-run the relevant analysis.

Exam trap: A longer history is not more reliable when its evidence no longer matches the codebase.

Question 19

Correct answer: D

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?

Domain: Domain 3 - Claude Code Configuration & Workflows

Task: Task 3.2 - Skill isolation and tool restrictions

> **Decision rule:** Use skills for on-demand workflows, context: fork for verbose isolation, and allowed-tools for role scoping.

Why D is correct: The forked context contains the exploration noise and returns a concise result to the main session. Tool restrictions reduce the risk of unintended edits during discovery.

A **Convert it to a user-level command so only advanced developers can invoke it, while leaving the main-context behavior unchanged.**
Changing scope does not solve context pollution or the need to restrict tools.

B **Move the workflow into the root CLAUDE.md so its full instructions are always available before any development task.**
Always loading a verbose, task-specific workflow increases context cost even when the workflow is not needed.

C **Keep the skill in the main context but ask it to suppress intermediate output and trust it to avoid write operations.**
Prompt-based restraint does not provide the same isolation or tool-level restriction as skill configuration.

D **Use context: fork so the skill runs in an isolated context, and restrict allowed-tools to the read-only tools required for discovery.**
Correct option; see the question-specific rationale above.

Exam trap: Do not solve context pollution merely by hiding the workflow from some users.

Question 20

Correct answer: C

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?

Domain: Domain 5 - Context Management & Reliability

Task: Task 5.4 - Subagent delegation for exploration

> **Decision rule:** Use subagents to isolate verbose discovery and return compact, decision-ready findings.

Why C is correct: Focused exploration preserves the main context for coordination and design while still grounding the plan in repository evidence.

A **Ask one subagent to return its complete transcript and every file listing so the main session can independently verify all details.**
Returning raw exploration recreates the same context problem. The subagent should return structured findings and evidence, not its entire working history.

B **Have the main session read every source and test file up front so no discovery detail is lost before implementation begins.**
Reading everything exhausts context and dilutes attention before the relevant architecture is known.

C **Delegate focused discovery to an Explore-style subagent, require concise evidence-backed summaries, and plan implementation in the main session.**
Correct option; see the question-specific rationale above.

D **Skip exploration and begin with the most likely implementation pattern, then repair mismatches as tests fail.**
The service is explicitly unfamiliar; implementation-first increases rework and can miss architectural constraints.

Exam trap: The value of a subagent is lost if its entire verbose transcript is copied back into the parent context.

Scenario 3 - Multi-Agent Research System

Questions 21-30 | Review the decision rule before memorizing the answer letter.

Question 21

Correct answer: B

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?

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.2 - Coordinator-subagent orchestration

> **Decision rule:** Use a coordinator as the control plane for delegation, routing, aggregation, recovery, and observability.

Why B is correct: Hub-and-spoke orchestration centralizes decisions and preserves a coherent audit trail while keeping specialists focused on their roles.

A **Keep peer-to-peer communication but require every subagent to append a shared log entry after sending a task.**
A log improves visibility but leaves routing and recovery distributed across agents, so behavior can remain inconsistent.

B **Route delegation, result aggregation, error handling, and inter-subagent communication through a hub-and-spoke coordinator.**
Correct option; see the question-specific rationale above.

C **Replace delegation with one fixed sequential pipeline so every query invokes all subagents in the same order.**
A fixed pipeline improves predictability but wastes work and cannot adapt to query complexity or intermediate findings.

D **Make the synthesis agent the permanent owner of all other agents because it already sees the final report requirements.**
Synthesis is a specialized role. Combining it with orchestration increases coupling and weakens separation of concerns.

Exam trap: A shared log is not a substitute for a single orchestration authority.

Question 22

Correct answer: C

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?

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.3 - Explicit subagent context passing

> **Decision rule:** Subagents receive isolated context; pass content, metadata, constraints, and quality criteria explicitly.

Why C is correct: The synthesis agent can preserve attribution only when the coordinator supplies both the findings and their provenance in a structured form.

A **Give the synthesis subagent web-search tools so it can rediscover any evidence that was not automatically inherited.**
Rediscovery adds cost and can select different evidence. The coordinator should pass the already gathered findings explicitly.

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.**
Subagents do not automatically inherit the parent's hidden state or conversation history.

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.**
Correct option; see the question-specific rationale above.

D **Ask the synthesis subagent to add citations only after drafting, without passing source metadata until the final editing step.**
Citations cannot be reliably reconstructed after claims have already been composed without source mappings.

Exam trap: Do not grant extra tools to compensate for context that should have been passed during handoff.

Question 23

Correct answer: C

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

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.3 - Parallel subagent spawning

> **Decision rule:** Spawn independent subagents in parallel by issuing multiple Task calls in one coordinator response.

Why C is correct: Parallel calls reduce orchestration latency while preserving distinct scopes and structured outputs for later aggregation.

A **Send one broad Task call asking a single subagent to cover all three questions so the coordinator performs fewer tool calls.**
Combining independent scopes increases attention dilution and reduces the benefits of specialization and parallelism.

B **Invoke one Task call, wait for its result, then decide whether to start the next independent question in a later turn.**
Sequential invocation adds unnecessary round trips when the tasks are already known to be independent.

C **Emit multiple Task tool calls in the same coordinator response, each with a scoped goal, context, tools, and expected output schema.**
Correct option; see the question-specific rationale above.

D **Open three unrelated user sessions manually and paste their final answers into the coordinator after all sessions finish.**
Manual sessions weaken provenance, reproducibility, and coordinator control over context and failure handling.

Exam trap: Do not serialize work merely because the coordinator receives results one conversation turn at a time.

Question 24

Correct answer: C

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?

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.2 - Coverage-aware decomposition

> **Decision rule:** The coordinator owns breadth of decomposition; successful subagents cannot repair an omitted scope they were never assigned.

Why C is correct: The assignments reveal the failure directly: all subtasks are radiology variants. The coordinator must reason about coverage before delegation.

A **The web-search subagent needs a larger result limit because more radiology sources would eventually reveal other healthcare domains.**
More results within the same narrow queries do not reliably cover omitted sectors such as administration, drug discovery, or patient operations.

B **The synthesis agent failed to invent missing domains from general knowledge after receiving the radiology findings.**
The synthesis agent should not fabricate unresearched coverage. The upstream decomposition must allocate the missing scope.

C **The coordinator decomposed the broad topic too narrowly, so every specialist executed a valid assignment within an incomplete research scope.**
Correct option; see the question-specific rationale above.

D **The document-analysis agent filtered out non-radiology sources because its relevance threshold was too strict.**
The logs already show that no non-radiology assignments were issued; the failure occurred before document filtering.

Exam trap: Do not blame downstream execution when the coordinator defined an incomplete search space.

Question 25

Correct answer: B

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

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.2 - Iterative refinement loops

> **Decision rule:** Use structured gap detection to drive targeted re-delegation and iterative re-synthesis.

Why B is correct: The coordinator should decide whether the gap is material, gather the missing evidence efficiently, and preserve the relationship between new claims and sources.

A **Restart the entire research workflow with broader prompts so all agents can produce a more comprehensive first pass.**
A full restart discards useful work. Iterative refinement should target the identified deficiency.

B **Evaluate the gap against the objective, delegate targeted follow-up research, then re-run synthesis with the new evidence and preserved provenance.**
Correct option; see the question-specific rationale above.

C **Delete the coverage_gaps field before publication because exposing uncertainty reduces the perceived quality of the report.**
Suppressing the gap hides a material limitation and damages trustworthiness.

D **Publish the current report and add a generic caveat that more research may be needed in future versions.**
The system has enough information to pursue the specific missing area now; a generic caveat is weaker than targeted recovery.

Exam trap: A coverage gap is a control signal for refinement, not a formatting defect to hide.

Question 26

Correct answer: B

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

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.6 - Fixed versus adaptive decomposition

> **Decision rule:** Match decomposition to task structure: prompt chaining for predictable stages, adaptive planning for open-ended discovery.

Why B is correct: The standardized workflow benefits from repeatability; the investigation benefits from revising subtasks when intermediate findings reveal new directions.

A **Choose the pattern only by document count: fixed below a threshold and dynamic above it, regardless of task uncertainty.**
Task structure and uncertainty, not raw volume alone, determine whether a fixed or adaptive plan is appropriate.

B **Use a predictable prompt chain for the standardized review and adaptive decomposition for the open-ended investigation.**
Correct option; see the question-specific rationale above.

C **Use the same fixed pipeline for both so orchestration, monitoring, and provenance remain identical across workloads.**
A fixed pipeline can miss newly discovered branches in open-ended research and perform unnecessary steps when the evidence changes direction.

D **Use dynamic decomposition for both because any model-driven workflow becomes less reliable when stages are predetermined.**
Predictable, repeatable reviews can benefit from a fixed chain that ensures every required stage is covered.

Exam trap: Do not apply one orchestration pattern universally merely for implementation convenience.

Question 27

Correct answer: D

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?

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.7 - Fork-based session management

> **Decision rule:** Use fork_session to explore divergent approaches from a common validated baseline.

Why D is correct: Forking preserves the shared evidence while isolating subsequent assumptions, prompts, and revisions in each branch.

A **Start two blank sessions and have each rediscover the 120-source baseline so their reasoning is completely independent.**
This preserves independence but wastes the validated shared analysis and can produce incomparable evidence sets.

B **Ask one subagent to generate both strategies in a single response, then split the resulting text into two reports.**
One reasoning context encourages convergence and does not create independent branches that can evolve over time.

C **Continue both strategies sequentially in one session and ask Claude to forget the assumptions from the first strategy before starting the second.**
The second analysis remains exposed to the first branch's reasoning and decisions, weakening independence.

D **Create separate fork_session branches from the shared baseline and give each branch its own strategy-specific instructions and evaluation criteria.**
Correct option; see the question-specific rationale above.

Exam trap: Independence requires isolated branches, not a request to mentally ignore prior reasoning.

Question 28

Correct answer: A

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?

Domain: Domain 2 - Tool Design & MCP Integration

Task: Task 2.4 - MCP resources for catalogs

> **Decision rule:** Use MCP resources for catalogs and reference content; use tools for actions.

Why A is correct: Resources give agents visibility into available content without expanding the action set or requiring repeated discovery calls.

A **Expose the collection catalog and descriptive metadata as MCP resources, while reserving tools for retrieval, filtering, and other actions.**
Correct option; see the question-specific rationale above.

B **Paste the full catalog into the coordinator system prompt on every turn so no discovery request is required.**
Always-loaded catalog data consumes context even when irrelevant and becomes harder to update consistently.

C **Create one tool per collection so the model sees every possible source as a separate action in its tool list.**
Hundreds of collection-specific tools overload selection and turn static catalog content into an action surface.

D **Force a list_collections tool call before every research request, even when the coordinator already has the current catalog.**
Repeated exploratory calls add latency and tokens without adding new information.

Exam trap: Do not model every piece of discoverable content as a separate tool.

Question 29

Correct answer: D

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?

Domain: Domain 5 - Context Management & Reliability

Task: Task 5.6 - Provenance and conflicting sources

> **Decision rule:** Preserve claim-source mappings, dates, and methodology; annotate conflicts instead of selecting arbitrarily.

Why D is correct: The two claims can coexist when their temporal and methodological contexts are explicit. Provenance lets readers understand rather than erase the difference.

A **Average the two figures and report the mean as a balanced estimate with a lower confidence score.**
Averaging incompatible measurements invents a number that neither source supports and destroys methodological meaning.

B **Use the newer publication automatically, because recency is the most reliable way to resolve conflicting statistics.**
The figures measure different periods and possibly different constructs; recency alone does not make one a replacement for the other.

C **Omit both values from the final report to avoid presenting a contradiction that readers may misinterpret.**
The evidence is useful when properly qualified. Suppression creates an unnecessary coverage gap.

D **Preserve both values with source attribution, dates, and methodology, explaining that the apparent conflict comes from different time bases.**
Correct option; see the question-specific rationale above.

Exam trap: A conflict is not always an error; it may be a missing time or methodology label.

Question 30

Correct answer: A

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?

Domain: Domain 5 - Context Management & Reliability

Task: Task 5.3 - Error propagation across agents

> **Decision rule:** Recover locally when possible, then propagate unresolved errors with partial results and actionable context.

Why A is correct: Structured error context lets the coordinator choose among retry, alternative sourcing, partial completion, or explicit coverage-gap reporting.

A **Structured failure context describing the access problem, attempted recovery, partial findings with provenance, and suggested alternatives or coverage implications.**
Correct option; see the question-specific rationale above.

B **Return only 'source unavailable' after retries, because detailed failure context belongs in infrastructure logs rather than agent messages.**
The coordinator needs the attempted query, partial results, and alternatives to decide whether to retry, replace the source, or annotate a gap.

C **Propagate the raw exception to a top-level handler and terminate the entire research workflow.**
A single inaccessible source need not invalidate the whole report when partial results and alternatives exist.

D **Return an empty successful result so the synthesis agent is not distracted by an error it cannot resolve.**
Marking failure as success hides both the access problem and useful partial evidence, preventing intelligent recovery and honest coverage reporting.

Exam trap: Do not collapse a recoverable partial failure into either silent success or total workflow termination.

Scenario 4 - Developer Productivity with Claude

Questions 31-40 | Review the decision rule before memorizing the answer letter.

Question 31

Correct answer: C

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?

Domain: Domain 2 - Tool Design & MCP Integration

Task: Task 2.5 - Built-in tool selection

> **Decision rule:** Use Grep for content, Glob for paths, and Read to inspect the relevant flow incrementally.

Why C is correct: The sequence narrows the search before consuming file contents and handles wrappers or aliases by following actual references.

A **Read every source file first to build a complete mental model, then search the loaded context for callers and tests.**
Reading all files up front wastes context and is less reliable than targeted search followed by focused reading.

B **Use Bash for both searches because shell commands are always more precise than the built-in repository tools.**
Bash can work, but the built-ins express the intent more clearly and reduce unnecessary shell complexity.

C **Use Grep for content references, Read for relevant flows, and Glob for matching test-file paths.**
Correct option; see the question-specific rationale above.

D **Use Glob for the function name and Grep for **/*.test.* because Glob searches repository contents while Grep matches file paths.**
The tools are reversed: Grep searches content, while Glob matches file paths.

Exam trap: Do not reverse Grep and Glob or read the entire repository before searching.

Question 32

Correct answer: C

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

Domain: Domain 2 - Tool Design & MCP Integration

Task: Task 2.5 - Edit fallback behavior

> **Decision rule:** When Edit lacks a unique anchor, fall back to Read plus a deliberate full-file Write.

Why C is correct: Reading the file provides enough context to identify the intended block and produce one deterministic replacement.

A **Use a global search-and-replace command in Bash so all four occurrences stay consistent.**
The request targets one occurrence. Global replacement can alter unrelated behavior.

B **Modify the first matching occurrence and rely on tests to reveal whether the wrong block was changed.**
Deliberately making an ambiguous edit creates avoidable risk and rework.

C **Read the full file, construct the intended complete content with the correct occurrence changed, and use Write to replace the file.**
Correct option; see the question-specific rationale above.

D **Run Edit again with the same anchor and instruct Claude to choose the occurrence closest to the reported line number.**
The non-unique anchor remains ambiguous; retrying the same operation does not make the edit deterministic.

Exam trap: Do not turn an ambiguous targeted edit into an uncontrolled global change.

Question 33

Correct answer: C

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?

Domain: Domain 2 - Tool Design & MCP Integration

Task: Task 2.1 - Tool boundaries and naming

> **Decision rule:** Tool interfaces should have distinct names, purposes, boundaries, and contracts.

Why C is correct: Clear specialization reduces selection ambiguity and makes evaluation failures attributable to a defined interface rather than overlapping semantics.

A **Force tool_choice to alternate between the two tools across turns so evaluation data remains balanced.**
Balancing call counts does not improve task-tool fit and can intentionally misroute requests.

B **Merge both into analyze_anything and let one backend dispatch to the appropriate implementation from the raw user message.**
A generic tool reduces the model's control and hides ambiguity inside the backend rather than defining clear capabilities.

C **Rename or split the tools around distinct purposes, and give each a clear contract and boundary against the alternative.**
Correct option; see the question-specific rationale above.

D **Keep both tools unchanged but add a system-prompt rule that selects analyze_document whenever the word 'file' appears.**
Keyword routing does not establish a semantic boundary and fails on mixed or indirect requests.

Exam trap: Do not patch an interface-design problem with keywords or forced call rotation.

Question 34

Correct answer: B

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?

Domain: Domain 2 - Tool Design & MCP Integration

Task: Task 2.3 - Tool scoping across agents

> **Decision rule:** Scope tools to the agent's role; expose cross-role capabilities only when justified and constrained.

Why B is correct: A smaller, relevant tool set improves selection reliability and enforces separation of concerns at the capability boundary.

A **Force one tool call on every turn so the model spends less time choosing among the 18 available options.**
tool_choice does not solve over-broad access and can force unnecessary actions.

B **Give each role only its required tools, adding narrowly constrained cross-role tools only for frequent, well-defined needs.**
Correct option; see the question-specific rationale above.

C **Create more specialized subagents but preserve the same 18-tool set so routing, rather than permissions, determines behavior.**
Specialization without tool scoping leaves the original misuse pathways intact.

D **Keep all tools available and strengthen each system prompt with a list of tools the role should avoid.**
Prompt restraint is probabilistic, while removing irrelevant tools directly reduces decision complexity and misuse.

Exam trap: A specialist with every tool is still a generalist at execution time.

Question 35

Correct answer: D

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

Domain: Domain 2 - Tool Design & MCP Integration

Task: Task 2.4 - MCP server scope and credentials

> **Decision rule:** Use project scope for shared MCP definitions and environment-variable expansion for secrets; use user scope for personal servers.

Why D is correct: The repository can standardize the integration without distributing credentials, while individual developers supply their own environment values.

A **Store the token directly in .mcp.json so every clone has a complete, reproducible configuration.**
Committing credentials exposes secrets and prevents per-user authentication.

B **Describe the server endpoint and token variable in CLAUDE.md, then ask Claude to construct the MCP connection dynamically when needed.**
CLAUDE.md is not the MCP server configuration mechanism and should not contain connection secrets.

C **Configure the Jira server only in each developer's ~/.claude.json because authentication is personal even when the integration is team-standard.**
This avoids committed secrets but loses the shared, version-controlled server definition and creates setup drift.

D **Commit the shared server in project .mcp.json, reference an environment token, and keep experimental servers at user scope.**
Correct option; see the question-specific rationale above.

Exam trap: Shared configuration and shared secrets are different concerns.

Question 36

Correct answer: B

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?

Domain: Domain 2 - Tool Design & MCP Integration

Task: Task 2.3 - tool_choice configuration

> **Decision rule:** Use forced tool selection for a required specific call, then restore auto for model-driven continuation.

Why B is correct: The first turn deterministically establishes the prerequisite; later turns preserve flexibility without repeating the same tool.

A **Use tool_choice: auto and emphasize the metadata requirement in the prompt so Claude can skip it when the repository appears obvious.**
The requirement says the call is mandatory. auto leaves the decision probabilistic.

B **Force extract_repo_metadata first, return its tool_result, then continue with tool_choice: auto.**
Correct option; see the question-specific rationale above.

C **Force extract_repo_metadata on every request so the prerequisite remains visible throughout the workflow.**
Repeated forced calls waste tokens and prevent the model from selecting the tools needed after metadata is established.

D **Use tool_choice: any for the first request because any guarantees that the model will select the required metadata tool.**
any guarantees a tool call, not a particular tool call. The model could choose an enrichment tool first.

Exam trap: Do not confuse 'must call some tool' with 'must call this exact tool.'

Question 37

Correct answer: A

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

Domain: Domain 3 - Claude Code Configuration & Workflows

Task: Task 3.2 - Personal skill customization

> **Decision rule:** Personal variants belong at user scope and should use distinct names to avoid affecting shared workflows.

Why A is correct: A separately named skill makes the customization explicit and leaves the version-controlled team skill stable.

A **Create a user-scoped skill with a distinct name in ~/.claude/skills/, preserving the project skill unchanged.**
Correct option; see the question-specific rationale above.

B **Create a user-level skill with the same name so it silently overrides the project skill in every context.**
Name collisions create ambiguity and can make behavior depend on loading precedence rather than explicit intent.

C **Put the personal preference in the root CLAUDE.md under a heading that mentions the developer's name.**
The root file is shared and always loaded, so it is not appropriate for one person's optional variant.

D **Edit the project SKILL.md and rely on teammates to ignore the additional verbosity when they do not need it.**
A project edit changes the shared workflow for everyone.

Exam trap: Do not use shared files or name collisions for personal behavior.

Question 38

Correct answer: A

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?

Domain: Domain 3 - Claude Code Configuration & Workflows

Task: Task 3.5 - Interview pattern

> **Decision rule:** Use the interview pattern to expose requirements the requester may not have considered.

Why A is correct: Targeted questions reveal the decision criteria needed to choose an appropriate cache design and avoid premature implementation.

A **Ask targeted questions that surface hidden constraints and tradeoffs before proposing an implementation.**
Correct option; see the question-specific rationale above.

B **Ask whether the engineer prefers Redis, and treat the selected product as the main architectural requirement.**
A product choice does not answer the more fundamental questions about semantics and operational constraints.

C **Provide a long generic explanation of caching strategies and ask the engineer to choose one without investigating the service.**
A catalog of patterns does not connect the decision to the actual workload, consistency needs, or failure modes.

D **Implement a standard cache-aside pattern immediately, then use test failures to discover any missing requirements.**
Caching choices can affect correctness and operations in ways tests may not reveal until after costly design work.

Exam trap: Do not let a familiar pattern or product substitute for missing requirements.

Question 39

Correct answer: D

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?

Domain: Domain 1 - Agentic Architecture & Orchestration

Task: Task 1.6 - Adaptive decomposition for open-ended work

> **Decision rule:** Open-ended codebase work should begin with discovery and use an adaptive, prioritized plan.

Why D is correct: Mapping the system first directs effort toward the most consequential behaviors and allows the plan to evolve as dependencies become visible.

A **Create snapshot tests for all public outputs before reading the implementation, because snapshots maximize coverage quickly.**
Snapshots may capture incidental behavior and miss the critical logic or failure modes that require targeted assertions.

B **Generate one test for every file in alphabetical order so coverage grows predictably and no file is skipped.**
File order is unrelated to behavioral risk, dependency boundaries, or business impact.

C **Use a fixed prompt chain of unit tests, integration tests, and performance tests without changing the plan after repository discovery.**
A fixed sequence ignores the stated uncertainty and cannot adapt to hidden dependencies or missing seams.

D **Map dependencies and risks first, then build and revise a prioritized test plan as evidence emerges.**
Correct option; see the question-specific rationale above.

Exam trap: Comprehensive does not mean uniform effort across every file.

Question 40

Correct answer: A

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?

Domain: Domain 3 - Claude Code Configuration & Workflows

Task: Task 3.1 - Modular CLAUDE.md organization with @import

> **Decision rule:** Keep universal guidance central, and import only the standards relevant to each package.

Why A is correct: Package-level CLAUDE.md files can use @import to load the standards that apply in that area, avoiding one always-loaded monolith while keeping shared source files maintainable.

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.**
Correct option; see the question-specific rationale above.

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.**
The entire file remains always loaded, consuming context and relying on inference to ignore irrelevant package rules.

C **Move package-specific standards into each maintainer's user-level CLAUDE.md so only the developers working on that package load them.**
User-level files are personal, not version-controlled team configuration, and would produce inconsistent behavior across maintainers.

D **Convert every standards document into a skill and require developers to invoke the relevant skills before editing a package.**
These conventions should load automatically in the relevant area; manual skill invocation is less reliable for always-applicable package standards.

Exam trap: Do not load every specialized rule everywhere merely to keep configuration in one file.

Scenario 5 - Claude Code for Continuous Integration

Questions 41-50 | Review the decision rule before memorizing the answer letter.

Question 41

Correct answer: A

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

Domain: Domain 3 - Claude Code Configuration & Workflows

Task: Task 3.6 - Non-interactive CI execution

> **Decision rule:** Use -p/--print for Claude Code in non-interactive pipelines.

Why A is correct: Print mode accepts the prompt, produces output, and exits rather than waiting for a conversational session.

A **Use claude -p or --print so the command runs non-interactively, emits output, and exits.**
Correct option; see the question-specific rationale above.

B **Use a --batch flag for the CLI so the job is submitted asynchronously and the pipeline can poll for completion.**
The Message Batches API is separate from the Claude Code CLI, and the stated flag is not the documented CI mode.

C **Set a CLAUDE_HEADLESS environment variable before the command so the same prompt syntax works in CI.**
The guide specifies -p/--print; the proposed environment variable is not the exam-aligned mechanism.

D **Redirect stdin from /dev/null while keeping the normal interactive command, because the absence of input forces Claude Code to finish.**
Shell redirection does not switch Claude Code into its documented non-interactive execution mode.

Exam trap: Do not replace a documented CLI mode with shell workarounds or invented flags.

Question 42

Correct answer: D

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?

Domain: Domain 3 - Claude Code Configuration & Workflows

Task: Task 3.6 - Structured CI output

> **Decision rule:** Use the CLI's structured-output flags for machine-consumed CI results.

Why D is correct: A JSON schema makes the output predictable enough for automated posting and validation, avoiding fragile text parsing.

A **Return a raw transcript of the review session and have a second model extract comment fields after the job finishes.**
A second parsing pass adds latency and another failure point when the CLI can produce structured output directly.

B **Ask Claude to emit a comma-separated line per finding and split the result in the pipeline.**
Ad hoc delimiters are fragile when fields contain commas, newlines, or optional values.

C **Keep Markdown but add page numbers and stronger headings so the parser can recover fields from a more consistent visual layout.**
Markdown remains a text format with no schema guarantee for field presence or type.

D **Use -p with --output-format json and --json-schema so the CI consumer receives a defined machine-readable structure.**
Correct option; see the question-specific rationale above.

Exam trap: Visual consistency is not the same as a machine-enforced contract.

Question 43

Correct answer: D

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

Domain: Domain 4 - Prompt Engineering & Structured Output

Task: Task 4.1 - Explicit review criteria

> **Decision rule:** Precision improves through explicit categorical criteria, not generic caution or self-reported confidence.

Why D is correct: The prompt should say exactly what to report, what to skip, and how severity is determined in this repository.

A **Require a confidence score above 0.9 for every finding but leave the existing issue criteria unchanged.**
A numerical threshold does not repair ambiguous categories or false-positive definitions.

B **Increase the prompt length with more general software-engineering best practices so the bot has a broader basis for critique.**
More generic guidance can expand the set of stylistic opinions rather than narrow it to actionable defects.

C **Ask the model to be conservative and report only issues it considers high confidence.**
Vague confidence language does not define what counts as a defect and is often poorly calibrated.

D **Define reportable and excluded categories, concrete severity boundaries, and examples of local patterns that must not be flagged.**
Correct option; see the question-specific rationale above.

Exam trap: 'Be conservative' sounds precise but does not create a decision boundary.

Question 44

Correct answer: D

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?

Domain: Domain 4 - Prompt Engineering & Structured Output

Task: Task 4.2 - Few-shot examples

> **Decision rule:** Use targeted few-shot examples to demonstrate ambiguous decision boundaries and output format.

Why D is correct: Examples show both sides of the distinction and let the model generalize the intended judgment to new code patterns.

A **Add one positive example of a missing test and ask the model to generalize the same pattern to all branches.**
A single positive example does not teach the boundary between reportable and acceptable cases.

B **Lower the sampling temperature so repeated runs are more deterministic even though the decision criteria remain ambiguous.**
Lower variability does not make the underlying judgment boundary more correct.

C **List every possible branching construct in the codebase and instruct the model to flag each one unless a matching test name is found.**
This creates a brittle checklist and encourages false positives when behavior is covered indirectly.

D **Add a small set of targeted examples showing genuine branch-level gaps, acceptable omissions, and the exact finding format expected.**
Correct option; see the question-specific rationale above.

Exam trap: Consistency without a correct boundary merely makes the same mistake repeatable.

Question 45

Correct answer: B

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?

Domain: Domain 4 - Prompt Engineering & Structured Output

Task: Task 4.6 - Independent review instances

> **Decision rule:** Independent review is more effective than self-review within the generation context.

Why B is correct: A fresh instance approaches the code without the generator's justifications and is therefore more likely to challenge hidden assumptions.

A **Expose the generator's confidence and route only low-confidence files to a second reviewer.**
The generator's confidence is not a reliable indicator of where its own assumptions caused blind spots.

B **Use a separate review instance that receives the patch and relevant repository context but not the generator's reasoning history.**
Correct option; see the question-specific rationale above.

C **Keep the same session and add a system message instructing Claude to become a skeptical reviewer before reading the patch.**
A role instruction does not remove the retained reasoning context and assumptions from generation.

D **Ask the generator to perform extended self-critique twice and accept only defects found in both passes.**
Repeated self-review remains anchored to the same reasoning context and can suppress real issues through consensus.

Exam trap: Changing roles inside one conversation does not create independent reasoning.

Question 46

Correct answer: B

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?

Domain: Domain 4 - Prompt Engineering & Structured Output

Task: Task 4.6 - Multi-pass review architecture

> **Decision rule:** Separate local analysis from cross-file integration to avoid attention dilution.

Why B is correct: Per-file passes provide consistent depth, while the integration pass examines relationships that cannot be evaluated in isolation.

A **Use one larger-context request containing all files and ask the model to allocate equal attention to every file.**
A larger window does not guarantee even attention or solve the cross-file versus local-analysis tradeoff.

B **Run focused per-file passes for local issues, followed by a separate integration pass for cross-file data flow and consistency.**
Correct option; see the question-specific rationale above.

C **Require developers to split every large pull request into four-file chunks before the bot runs.**
Smaller pull requests can help, but shifting the burden to developers does not provide a robust review architecture for legitimate multi-file changes.

D **Run three complete reviews of the full pull request and report only findings that appear in at least two runs.**
Full-pass repetition preserves attention dilution and consensus can discard legitimate findings that are detected inconsistently.

Exam trap: A bigger context window is not a substitute for a better review decomposition.

Question 47

Correct answer: C

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?

Domain: Domain 4 - Prompt Engineering & Structured Output

Task: Task 4.5 - Batch processing strategy

> **Decision rule:** Use batch processing for non-blocking, latency-tolerant workloads; use synchronous requests for blocking workflows.

Why C is correct: The split matches each workload's latency requirement and uses custom_id for reliable correlation and selective retry.

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.**
This adds duplicate work and complexity to a workflow that is inherently latency-sensitive and should remain synchronous.

B **Move both workflows to Message Batches and poll aggressively so the pre-merge result is likely to arrive before developers become impatient.**
Batch processing has no blocking-latency guarantee suitable for a merge gate.

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.**
Correct option; see the question-specific rationale above.

D **Keep both workflows synchronous because batch result ordering makes it unsafe to associate responses with the original documents.**
custom_id is designed for request-response correlation, so ordering is not the limiting factor.

Exam trap: Cost savings do not make an asynchronous service appropriate for a blocking gate.

Question 48

Correct answer: C

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

Domain: Domain 4 - Prompt Engineering & Structured Output

Task: Task 4.4 - Feedback loops for false positives

> **Decision rule:** Capture the observable pattern that triggered a finding so dismissal data can drive systematic prompt and rule improvement.

Why C is correct: A detected_pattern field enables aggregation by construct and reveals which heuristics need new criteria, examples, or suppression.

A **Record the model temperature and mode for each review so analysts can correlate dismissals with generation randomness.**
Runtime settings do not identify which code pattern caused the finding category to fire.

B **Add a reviewer_sentiment field that classifies whether the dismissal comment sounded frustrated, neutral, or positive.**
Sentiment does not explain the technical trigger responsible for the false positive.

C **Add a detected_pattern field that records the concrete heuristic or construct associated with each finding.**
Correct option; see the question-specific rationale above.

D **Store the model's private reasoning chain alongside every finding so reviewers can inspect how the conclusion was reached.**
Hidden reasoning is not an appropriate or necessary telemetry field; structured observable features are more useful and safer.

Exam trap: Measure the failure mechanism, not unrelated model settings or reviewer emotion.

Question 49

Correct answer: B

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?

Domain: Domain 4 - Prompt Engineering & Structured Output

Task: Task 4.1 - Restoring trust after noisy categories

> **Decision rule:** Protect trust by isolating a known high-false-positive category while it is being improved.

Why B is correct: The accurate categories can continue delivering value, and the weak category can be reintroduced after targeted prompt and evaluation work.

A **Keep every category enabled so the evaluation dataset remains representative, but add a disclaimer that some findings may be noisy.**
Continuing to flood developers with a known weak category further erodes trust in accurate categories.

B **Temporarily disable or suppress the noisy category while refining its criteria and evaluation, leaving the trustworthy categories active.**
Correct option; see the question-specific rationale above.

C **Merge all findings into one severity level so developers cannot distinguish the category with the high dismissal rate.**
Hiding the category prevents diagnosis and does not improve its accuracy.

D **Raise one global confidence threshold for all finding types so the overall number of comments decreases.**
A global threshold can suppress good security findings without specifically repairing the failing category.

Exam trap: Do not preserve a noisy feature at the expense of confidence in the entire system.

Question 50

Correct answer: A

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

Domain: Domain 4 - Prompt Engineering & Structured Output

Task: Task 4.3 - Schema limits and semantic validation

> **Decision rule:** Structured output eliminates shape errors but does not guarantee semantic correctness.

Why A is correct: Severity is a domain judgment that must be specified, demonstrated, evaluated, and validated independently of JSON conformance.

A **The schema guarantees syntax and shape, not semantic judgment; severity still needs explicit criteria, examples, calibration, and possibly downstream validation.**
Correct option; see the question-specific rationale above.

B **The JSON schema is defective, because a schema-compliant output should also guarantee that every field value is factually correct.**
Schemas validate structure and allowed forms, not whether the model's interpretation of impact is accurate.

C **The team should abandon structured output and use prose, because classification errors prove that schemas distort reasoning.**
Removing the schema reintroduces parsing failures without solving severity calibration.

D **tool_choice should be changed from forced to auto so Claude can return explanatory text when it is uncertain about severity.**
The selection mode does not define severity criteria or validate the semantic correctness of the chosen enum.

Exam trap: Do not confuse 'valid data structure' with 'correct decision.'

Scenario 6 - Structured Data Extraction

Questions 51-60 | Review the decision rule before memorizing the answer letter.

Question 51

Correct answer: C

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?

Domain: Domain 4 - Prompt Engineering & Structured Output

Task: Task 4.3 - tool_use and JSON schemas

> **Decision rule:** For the exam, use tool_use with a JSON schema to obtain schema-shaped structured data.

Why C is correct: The tool contract constrains the returned arguments and avoids the common syntax and trailing-commentary failures of prompt-only JSON.

A **Wrap the response in fenced code blocks and strip all text before the first brace and after the last brace.**
String cleanup cannot guarantee that the remaining JSON matches the required schema or semantics.

B **Prompt Claude to return only valid JSON, set temperature to zero, and reject any response containing markdown.**
Prompt-only controls reduce but do not eliminate syntax, extra-text, or field-shape failures.

C **Define a schema-backed extraction tool and consume its tool_use arguments, using tool_choice when the call must be guaranteed.**
Correct option; see the question-specific rationale above.

D **Accept natural-language output and use a downstream parser to infer fields because a second pass can repair malformed structures.**
This adds another probabilistic step and compounds extraction and parsing errors.

Exam trap: Do not treat 'return JSON' as equivalent to an enforced schema contract.

Question 52

Correct answer: D

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?

Domain: Domain 4 - Prompt Engineering & Structured Output

Task: Task 4.3 - Optional and nullable fields

Decision rule: Represent legitimate absence in the schema instead of forcing a value.

Why D is correct: Optional or nullable fields align the output contract with the source reality and reduce fabrication pressure.

A **Require the model to calculate renewal_date from signature date and contract length whenever the source omits it.**
An inferred date is not the same as a stated source fact and can create false data.

B **Keep the field required and add a stronger instruction that fabricated dates are prohibited.**
The schema still pressures the model to supply a string even when the source has no value.

C **Remove renewal_date from the schema entirely so the extractor cannot hallucinate it.**
The field is still needed for documents that do contain a renewal date; the schema should represent legitimate absence.

D **Make renewal_date optional or nullable and return null when the source does not contain it.**
Correct option; see the question-specific rationale above.

Exam trap: A required field can cause hallucination when the source is allowed to omit the information.

Question 53

Correct answer: B

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

Domain: Domain 4 - Prompt Engineering & Structured Output

Task: Task 4.4 - Retry-with-feedback limits

> **Decision rule:** Retry when the source contains correctable information, not when the fact is absent.

Why B is correct: The workflow should represent missingness honestly and use review or enrichment only when explicitly designed for that purpose.

A **Copy another identifier, such as the invoice number, into supplier_tax_id so the record passes schema validation and can be corrected later.**
Substituting a different field creates known false data and undermines downstream integrity.

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.**
Correct option; see the question-specific rationale above.

C **Retry with progressively stronger wording until the model fills the required field, because validation failures should always trigger self-correction.**
Retries cannot recover information that is absent and may increase fabrication.

D **Ask the model to search external sources for the supplier's tax ID and insert it into the extraction result.**
The task is source-grounded extraction; external enrichment changes provenance and may identify the wrong entity.

Exam trap: Validation is not a command to invent a value.

Question 54

Correct answer: D

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

Domain: Domain 4 - Prompt Engineering & Structured Output

Task: Task 4.2 - Few-shot extraction examples

> **Decision rule:** Use few-shot examples to demonstrate how the same schema maps across ambiguous source structures.

Why D is correct: Examples teach the model what counts as evidence in narrative text and how to normalize it without changing the output contract.

A **Make every field required so the model searches more aggressively for values in unstructured text.**
Required fields can increase hallucination when the value is truly absent.

B **Tell Claude to be more creative and infer likely measurements whenever the document seems incomplete.**
Vague creativity encourages unsupported inference and reduces extraction precision.

C **Add a long dictionary of every known field-name synonym and require exact phrase matching before a value is extracted.**
Narrative reports can express values without explicit field labels, so synonym matching remains brittle.

D **Add targeted few-shot examples that demonstrate extraction from varied layouts and informal measurement phrasing while preserving the same output schema.**
Correct option; see the question-specific rationale above.

Exam trap: Do not trade missing fields for invented fields by increasing pressure or creativity.

Question 55

Correct answer: A

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?

Domain: Domain 5 - Context Management & Reliability

Task: Task 5.1 - Lost-in-the-middle mitigation

> **Decision rule:** Mitigate position effects by surfacing critical facts early and structuring the remaining evidence clearly.

Why A is correct: A persistent fact block protects exact values, while headings and an early summary make key evidence easier to retrieve from a long input.

A **Surface a key-facts and exceptions block early, preserve exact numbers and dates, and organize detailed evidence under explicit section headings.**
Correct option; see the question-specific rationale above.

B **Move all instructions to the very end, because the final part of a long context is always processed accurately.**
End placement alone does not protect important evidence located in the middle of the source.

C **Randomize document section order on each request so no clause is consistently placed in the middle.**
Randomization damages document coherence and does not create a stable mechanism for preserving critical facts.

D **Compress amounts, dates, and percentages into approximate prose so the summary occupies fewer tokens.**
Vague summarization can destroy the exact transactional facts the extraction must preserve.

Exam trap: Do not solve context pressure by blurring the numbers and dates the task depends on.

Question 56

Correct answer: B

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?

Domain: Domain 5 - Context Management & Reliability

Task: Task 5.5 - Field-level confidence and calibration

Decision rule: Calibrate confidence at the field and segment level with labeled data.

Why B is correct: Field-level thresholds expose weak combinations that aggregate metrics hide and let reviewers focus on the values most likely to be wrong.

A **Remove confidence scores and route only schema-invalid records to humans, because valid structured output is sufficient for automation.**
Schema validity does not measure extraction correctness, especially for a known weak document-field combination.

B **Produce field-level confidence, calibrate thresholds on labeled validation data, and evaluate accuracy by document type and field before reducing review.**
Correct option; see the question-specific rationale above.

C **Raise the single document-level threshold for every record until the handwritten-addendum error rate becomes acceptable.**
A global score can over-review strong fields and still hide the weak termination-date segment.

D **Keep the aggregate confidence score but add a rule that all handwritten documents receive the same fixed penalty.**
A blanket penalty is not calibrated to specific fields or actual observed error distributions.

Exam trap: High overall accuracy can conceal a dangerous failure mode in one document type or field.

Question 57

Correct answer: C

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?

Domain: Domain 5 - Context Management & Reliability

Task: Task 5.5 - Stratified sampling

> **Decision rule:** Sample the auto-approved population by meaningful strata, not only the records already flagged as uncertain.

Why C is correct: Stratification makes segment-specific error rates visible and supports continued detection of rare or emerging failures.

A **Review the first fixed number of records processed each week, regardless of source type or field distribution.**
Convenience sampling can overrepresent common documents and miss rare but important segments.

B **Stop sampling once overall accuracy exceeds the target for one month, because continued review no longer changes the deployment decision.**
Failure modes and document mix can change; ongoing sampling is needed to detect drift and novel patterns.

C **Stratify random samples of high-confidence outputs by document type and field, tracking errors and new patterns over time.**
Correct option; see the question-specific rationale above.

D **Review only the lowest-confidence records because high-confidence outputs have already passed the model's reliability threshold.**
That approach cannot measure errors in the population being auto-approved.

Exam trap: You cannot validate high-confidence automation by reviewing only low-confidence cases.

Question 58

Correct answer: B

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?

Domain: Domain 5 - Context Management & Reliability

Task: Task 5.6 - Conflict preservation and provenance

> **Decision rule:** Preserve conflicting claims with source mappings rather than resolving them arbitrarily during extraction.

Why B is correct: The extractor's job is to expose the evidence and uncertainty so a downstream policy or human can reconcile it transparently.

A **Use the appendix value only because signed sections are generally more authoritative, and omit the conflicting summary to simplify downstream processing.**
Even if one source may be more authoritative, suppressing the conflict removes evidence the downstream decision process needs.

B **Both values with source locations, conflict_detected metadata, relevant dates or section context, and enough provenance for downstream reconciliation or human review.**
Correct option; see the question-specific rationale above.

C **Use the first value encountered and add a generic warning that the document may contain inconsistencies.**
A generic warning does not preserve the competing claim or its exact source location.

D **Average the values and report $1.3M with low confidence so the record contains one normalized number.**
Averaging invents a value not stated in the document and obscures the disagreement.

Exam trap: Normalization must not erase a material contradiction.

Question 59

Correct answer: A

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?

Domain: Domain 2 - Tool Design & MCP Integration

Task: Task 2.4 - MCP resources for document catalogs

> **Decision rule:** Use resources for catalogs and tools for actions.

Why A is correct: A resource gives the model visibility into available templates without flooding the tool-selection surface or requiring repeated exploratory calls.

A **Expose the template and policy catalog as MCP resources, while keeping tools for retrieval, validation, and actions.**
Correct option; see the question-specific rationale above.

B **Turn every template into a separate tool so the model can call the exact document type it recognizes.**
A very large tool set degrades selection reliability and misuses tools for static catalog content.

C **Hide the catalog and require the model to infer the template from document text without consulting external metadata.**
The known catalog is valuable context; withholding it increases ambiguity and duplicate discovery work.

D **Force list_templates before every extraction so the model always starts with a fresh view of the catalog.**
Repeated listing wastes latency and context when the catalog can be exposed directly as reference content.

Exam trap: Discoverable content is not automatically an action.

Question 60

Correct answer: D

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

Domain: Domain 2 - Tool Design & MCP Integration

Task: Task 2.2 - Valid empty result versus access failure

> **Decision rule:** Distinguish successful emptiness from tool failure with structured error semantics.

Why D is correct: The coordinator can only choose the right recovery path when it knows whether the query succeeded with no matches or failed because access was unavailable.

A **Throw a top-level exception for both cases so any uncertainty is routed directly to human review.**
A normal no-match result is not exceptional, and many access failures have specific recovery paths short of human escalation.

B **Return a generic not_found error for both situations so the agent retries with a broader query.**
A broader query cannot repair expired credentials, and a valid empty result should not be treated as a failure.

C **Keep returning [] and let the agent ask the user whether the repository might be unavailable.**
The tool knows which condition occurred and should communicate it explicitly instead of forcing the model to guess.

D **Return empty success only for a valid no-match; use structured error metadata for access or permission failures.**
Correct option; see the question-specific rationale above.

Exam trap: An empty result is data; a permission failure is an error.