Anthropic.CCAR-P.v2026-10-20.q61

Exam Code:	CCAR-P
Exam Name:	Claude Certified Architect - Professional
Certification Provider:	Anthropic
Free Question Number:	61
Version:	v2026-10-20
# of views:	127
# of Questions views:	976
https://www.freecram.net/torrent/Anthropic.CCAR-P.v2026-10-20.q61.html	

NEW QUESTION: 1
A pilot AI assistant for procurement specialists shows 89 percent first-response acceptance, but follow-up surveys reveal that specialists frequently override the assistant's vendor recommendations after considering criteria the assistant did not evaluate. The pilot owner wants to ship the assistant unchanged because of the strong acceptance rate.
Which two Discernment-competency observations should you raise BEFORE approving the launch? (Select two.)
A. The acceptance rate alone proves readiness for general production use.
B. The survey response rate may not be statistically representative of all specialists.
C. The unconsidered criteria represent a scope gap in the assistant's input space.
D. Acceptance does not establish whether recommendations remain correct after specialist review.
E. The pilot duration was probably too short to demonstrate reliability across the full year.
Answer: (SHOW ANSWER)
The assistant is making recommendations without considering information that specialists regard as material.
That is a direct input-scope and requirements gap, not merely a presentation issue. The 89 percent acceptance metric is also incomplete because initial acceptance does not reveal whether recommendations remain correct after users apply the omitted criteria. Production readiness requires outcome-focused evaluation, including final recommendation agreement, override reasons, financial impact, supplier risk, and performance across representative procurement scenarios.
Options B and E raise possible concerns, but the scenario supplies no evidence that response rate or pilot duration is the demonstrated deficiency. Option A incorrectly treats a single proxy metric as sufficient. Anthropic recommends multidimensional, task-specific success criteria aligned with the actual user outcome. Define success criteria and evaluations

NEW QUESTION: 2
A Claude architect is designing a HIPAA-compliant pipeline that processes patient records.
Which two design decisions directly support HIPAA compliance requirements? (Select two.)
A. Setting max_tokens to a low value to minimize the volume of text generated per request.
B. Selecting the highest-capability Claude model to maximize diagnostic accuracy.
C. Enforcing role-based access controls so that PHI is retrievable only by authorized personnel.
D. Ensuring patient data is never included in training feedback loops sent to the model provider without a BAA in place.
E. Using streaming responses to reduce perceived latency for clinical users.
Answer: (SHOW ANSWER)
Role-based access control directly supports HIPAA's access-control and minimum-necessary principles. The retrieval and orchestration layers must verify the authenticated user's role before protected health information enters the model context. Authorization should be applied to individual records, data sources, tools, and actions, with audit events recording the requesting identity, accessed resource, purpose, and result. Model instructions alone are not a security boundary.

Patient information must also remain within services and processing activities covered by the organization's Business Associate Agreement. Anthropic states that PHI processing requires a HIPAA-ready organization and an accepted BAA, and that only specified eligible services and configurations are covered. Patient data must not be submitted through training, feedback, beta, third-party, or integration pathways that fall outside the applicable agreement. A BAA is not blanket authorization; architects must verify feature eligibility and disable inappropriate data-sharing mechanisms.
Reducing max_tokens, selecting a more capable model, and enabling streaming affect output length, quality, or latency. They do not establish authorization, contractual coverage, minimum-necessary access, or compliant PHI handling.
Study Guide references/topics: Anthropic Business Associate Agreement requirements; HIPAA-ready services; PHI access control; minimum necessary; eligible-service boundaries; auditability.

NEW QUESTION: 3
A research summarization assistant has been deployed for six months. A user has flagged that a generated summary contained a fabricated citation. The product team has asked whether the incident requires architectural action or whether it is an isolated case.
Which two Diligence-competency actions should you take? (Select two.)
Each correct answer presents part of the solution.
A. Sample recent summaries to estimate the fabrication frequency across the population.
B. Disable the assistant immediately for all current users without any prior diagnostic analysis.
C. Review whether citation-grounding controls exist anywhere in the current generation pipeline.
D. Communicate to users that the assistant does not fabricate output as a matter of design.
E. Treat the incident as an anecdotal isolated case and take no further investigative action.
Answer: (SHOW ANSWER)
A single confirmed fabrication requires investigation, but it does not establish the population-level failure rate. Sampling recent summaries determines prevalence, affected content types, and whether the incident correlates with particular prompts, documents, or model versions. The architecture must also be examined for grounding controls, including source-constrained generation, citation rendering, claim-to-source verification, and rejection of unsupported statements. Anthropic recommends making factual outputs auditable through citations and requiring supporting quotations for important claims. Immediate universal shutdown may become necessary if the risk is severe, but doing so before establishing scope is not the most disciplined first response. Options D and E dismiss evidence without validation and would prevent corrective action.
Reducing hallucinations

NEW QUESTION: 4
You are comparing two configurations for an internal compliance Q & A assistant. Configuration X reduces top-k retrieval and uses a lighter model, lowering p95 latency by 38 percent while producing a measured 2 percent reduction in the accuracy benchmark. Configuration Y retains the existing top-k retrieval setting and heavier model, with unchanged p95 latency and no measurable change in accuracy. The deployment's stated priorities are accuracy first and latency second.
Which configuration is the better recommendation?
A. Configuration X, because the 2 percent accuracy reduction falls within normal benchmark variance and is not a meaningful regression.
B. Configuration Y, because the deployment's stated priority is accuracy first, and Configuration X produces a measured reduction in accuracy.
C. Configuration X, because lower latency generally matters more than accuracy, regardless of the deployment's stated priorities.
D. Configuration X, because a 38 percent latency improvement outweighs a 2 percent accuracy reduction as a general rule across compliance deployments.
Answer: (SHOW ANSWER)
Configuration Y is the appropriate recommendation because architectural optimization must follow the deployment's explicitly defined priorities and acceptance criteria. Accuracy is the primary requirement for this compliance assistant, and Configuration X introduces a measured two-percent accuracy regression. That reduction cannot be dismissed as ordinary benchmark variation unless repeated evaluation establishes that it is statistically insignificant. In a compliance workload, even a small accuracy decline can increase incorrect guidance, missed policy requirements, human-review effort, and regulatory exposure. Configuration Y preserves the established accuracy level while maintaining the current latency profile. Although its latency does not improve, latency is explicitly the secondary objective. Configuration X could be reconsidered only if further testing eliminates the accuracy regression or stakeholders formally approve a revised trade-off.
Options C and D incorrectly impose a general preference for latency over the documented deployment priorities.

NEW QUESTION: 5

You are integrating human review into a high-volume classification pipeline where reviewing every output is infeasible.
Which sampling strategy best balances throughput with quality oversight?
A. No sampling, relying entirely on user complaints to reveal quality and safety problems after they affect users.
B. Risk-stratified sampling that reviews all low-confidence and high-impact outputs and a smaller random sample of high-confidence routine outputs.
C. Inverse sampling that reviews only high-confidence routine outputs and skips low-confidence and high- impact outputs.
D. Universal review of every output regardless of confidence or throughput impact.
Answer: B (LEAVE A REPLY)
Risk-stratified sampling concentrates limited reviewer capacity where error consequences and uncertainty are greatest. All low-confidence and high-impact cases should receive mandatory review, while random sampling of routine high-confidence outputs provides an unbiased signal for drift, unexpected failure modes, and overconfident errors. Complaint-only monitoring discovers defects after harm occurs. Reviewing only high- confidence cases systematically ignores the most dangerous outputs. Universal review provides maximum coverage but contradicts the stated throughput constraint and may create queues that undermine service objectives. The sampling policy should define confidence calibration, impact categories, escalation thresholds, review SLAs, and periodic adjustment based on observed error rates. Reviewer decisions should feed the evaluation dataset so monitoring quality improves over time.

NEW QUESTION: 6
You are configuring tool permissions for a Claude-based assistant. The assistant's defined responsibilities require read access to a knowledge base and write access to a draft queue, and nothing else.
Which scoping design best applies least privilege?
A. Allow access to all tools with read permissions globally and restrict write permissions to the draft- queue tool, without scoping to only the specific tools required for defined tasks.
B. Per-role allow-list of exactly the read-knowledge-base and write-draft-queue tools, enforced at the orchestration layer.
C. Allow access to every tool in the catalog and rely on the model to decline tools it should not use.
D. Disable all tools entirely to achieve a minimal permission surface, accepting that the assistant can no longer perform the read-knowledge-base or write-draft-queue tasks it was designed for.
Answer: B (LEAVE A REPLY)
Least privilege means granting the exact capabilities required for the defined responsibility-no more and no less. Option B permits the assistant to read the approved knowledge base and write only to the draft queue.
Enforcing the allow-list at the orchestration layer creates a deterministic authorization boundary outside the model's prompt and reasoning.
The rule should also be role-specific so different users or operational roles receive only the tools and data scopes they require. The knowledge-base tool should remain read-only, while the draft-queue tool should be restricted to draft creation rather than external sending, approval, or production publication.
Option A exposes every globally readable tool, creating unnecessary access to unrelated information. Option C relies on voluntary model behavior instead of enforceable authorization. A mistaken tool selection or successful prompt injection could therefore invoke an inappropriate capability. Option D minimizes permissions by eliminating the assistant's required functionality and is not a viable least-privilege implementation.
Anthropic's permission guidance distinguishes allow, ask, and deny rules and confirms that these controls are enforced by Claude Code rather than by model instructions. Configure Permissions Study Guide references/topics: Least privilege; per-role authorization; orchestration-layer enforcement; capability scoping; read/write separation; tool allow-lists.

NEW QUESTION: 7
You are classifying token-management tactics by where each tactic applies in the request lifecycle: "Input Preparation," "Prompt Construction," or "Output Handling."

Answer Area

Tactic:	Select the Correct Request Lifecycle:
Persist the validated output for downstream consumption and audit.	Input Preparation / Prompt Construction / Output Handling
Order the prompt sections so cacheable content sits before per-request content.	Input Preparation / Prompt Construction / Output Handling
Summarize prior conversation history when full history is no longer needed.	Input Preparation / Prompt Construction / Output Handling
Validate the model's structured output against the expected schema.	Input Preparation / Prompt Construction / Output Handling
Move stable repeated content into a cacheable prefix at the head of the prompt.	Input Preparation / Prompt Construction / Output Handling
Trim retrieved passages to spans relevant to the user's question.	Input Preparation / Prompt Construction / Output Handling

freecram.net

Answer:

Answer Area

Tactic:	Select the Correct Request Lifecycle:
Persist the validated output for downstream consumption and audit.	Input Preparation / Prompt Construction / Output Handling
Order the prompt sections so cacheable content sits before per-request content.	Input Preparation / Prompt Construction / Output Handling
Summarize prior conversation history when full history is no longer needed.	Input Preparation / Prompt Construction / Output Handling
Validate the model's structured output against the expected schema.	Input Preparation / Prompt Construction / Output Handling
Move stable repeated content into a cacheable prefix at the head of the prompt.	Input Preparation / Prompt Construction / Output Handling
Trim retrieved passages to spans relevant to the user's question.	Input Preparation / Prompt Construction / Output Handling

Explanation:
Persist the validated output for downstream consumption and audit - Output Handling Order prompt sections so cacheable content appears before per-request content - Prompt Construction Summarize prior conversation history when full history is no longer needed - Input Preparation Validate the model's structured output against the expected schema - Output Handling Move stable repeated content into a cacheable prefix at the beginning of the prompt - Prompt Construction Trim retrieved passages to spans relevant to the user's question - Input Preparation Input preparation determines what evidence and history should enter the request, so summarization and retrieval trimming belong there. Prompt construction determines ordering and cache boundaries; stable repeated instructions must precede dynamic per-request content to maximize cache reuse. Output handling begins after generation and includes schema validation, persistence, auditing, and downstream delivery.
Mixing these responsibilities creates inefficient prompts and weak validation boundaries. Anthropic's context guidance emphasizes curating the smallest useful context because unnecessary tokens can reduce recall and accuracy. Prompt caching similarly depends on a stable reusable prefix, while structured outputs or schema validation protect downstream systems from malformed responses. Context windows; prompt caching

NEW QUESTION: 8
You are preparing a HIPAA-eligible deployment for a healthcare customer.
Which configuration supports HIPAA compliance using Anthropic-offered tools?
A. Claude Free with no contractual addendum, since consumer products meet HIPAA requirements out of the box.
B. Claude Enterprise with a signed Business Associate Agreement, Zero Data Retention enabled, and audit logging configured for compliance tracking.

C. Disabling all audit logging so that no PHI is recorded in any log store, on the assumption that the absence of logs satisfies HIPAA requirements without a signed BAA.
D. An ad-hoc personal Claude account used by individual clinicians for PHI-related tasks, with no Business Associate Agreement, no Zero Data Retention, and no audit logging configured.
Answer: B (LEAVE A REPLY)
Option B is the only configuration establishing the contractual, data-handling, and audit foundations needed for a HIPAA-eligible deployment. Anthropic requires an eligible commercial organization to activate the HIPAA-ready configuration and accept a Business Associate Agreement. A standard Enterprise subscription is not automatically covered merely because it is an enterprise product. Anthropic BAA Guidance Zero Data Retention may be required for particular covered services, including eligible Claude Code configurations, and is enabled per organization for qualified accounts. Audit logging supports access review, configuration monitoring, incident investigation, and compliance evidence. Anthropic provides audit logs for Enterprise organizations. Claude Code Legal and Compliance, Access Audit Logs Options A and D use consumer or personal accounts without the required contractual framework. Option C incorrectly treats the elimination of audit evidence as a substitute for a BAA and compliant configuration.
The signed BAA remains the authoritative source for covered products and features. The customer must also configure access control, minimum-necessary data use, workforce policies, retention, integrations, incident response, and vendor-risk controls. The configuration supports compliance; it does not independently guarantee it.
Study Guide references/topics: HIPAA readiness; BAA; ZDR; PHI handling; audit logging; eligible services; shared compliance responsibility.

NEW QUESTION: 9
An architect is reviewing a set of architecture documentation packages before handing off a Claude-based pipeline to an implementation team.
Which two characteristics indicate that a documentation package is sufficient to support implementation without ongoing architect involvement? (Select two.)
A. The document specifies integration contracts, configuration schemas, and expected input/output shapes for each component.
B. The document includes a decision log that records the rationale for key architectural choices and the alternatives rejected.
C. The document provides a high-level narrative description of the business problem without component- level detail.
D. The document includes the architect's contact information for questions arising during implementation.
E. The document lists all Claude models that were evaluated but does not specify which was selected or why.
Answer: (SHOW ANSWER)
Implementation-ready documentation must provide both executable specifications and preserved architectural reasoning. Option A supplies the contracts an implementation team needs to build components independently: configuration schemas, accepted inputs, produced outputs, validation requirements, and integration boundaries. These details prevent teams from inventing incompatible assumptions about data structures, tool behavior, error handling, or component ownership.
Option B complements the technical contracts with a decision log. Recording the selected design, underlying constraints, alternatives considered, and reasons for rejection enables implementers to distinguish deliberate architectural decisions from accidental omissions. It also prevents previously rejected approaches from being reintroduced without new evidence.
A high-level business narrative in Option C may explain purpose but cannot direct component-level implementation. Contact information in Option D provides an escalation route, but documentation is not sufficient if routine progress still depends on the architect's availability. Option E records activity rather than a decision; a model list without the selected model and selection rationale leaves critical cost, latency, capability, and evaluation assumptions unresolved.
A strong handoff package should permit implementation, testing, and support teams to proceed while escalating only genuinely new architectural questions.
Study Guide references/topics: Architecture documentation; interface contracts; configuration schemas; Architecture Decision Records; implementation handoff; decision traceability.

NEW QUESTION: 10
You are integrating Claude Code into a team workflow and must complete the design and configuration steps before piloting the integrated workflow with a small group.
Which two steps must be completed BEFORE piloting the integrated workflow? (Select two.) Each correct answer presents part of the solution.
Freecram.net
A. Document the integrated workflow with handoff criteria, permission boundaries, and on-call runbooks.
B. Identify the workflow steps where AI-assisted tooling adds value and where human authority must remain.
C. Negotiate the workflow change with affected teams and obtain formal sign-off from each manager.
D. Iterate the integrated workflow based on pilot findings before broader rollout to the team.
E. Configure the project-scope MCP servers, permissions, subagents, and persistent project context.

Answer: (SHOW ANSWER)
The pilot must test a deliberately designed workflow, not technology added without role boundaries. Option B identifies where Claude provides measurable value and where approval, accountability, or consequential decisions must remain human responsibilities. Option E establishes the technical environment required for consistent pilot execution: repository-shared MCP servers, scoped permissions, specialized subagents, and persistent project instructions. Iteration in Option D necessarily follows observed pilot findings. Formal approval from every manager is unnecessarily broad unless required by organizational governance. Complete runbooks and final handoff documentation mature after the workflow has been validated, although essential pilot procedures must exist. Anthropic documents project-level instructions, permissions, MCP servers, and subagents as complementary mechanisms for team workflows. Claude Code extension overview

NEW QUESTION: 11
You are evaluating retrieval-strategy claims used by a peer team.
For each claim, select yes if the statement is generally accurate. Otherwise, select no.

Answer Area

Statements:	Select Yes or No:
Dense vector retrieval is well suited to recovering paraphrased intent against unstructured text.	Yes / No
Sparse term-matching retrieval recovers exact terms (such as identifiers and rare tokens) more reliably than dense alone.	Yes / No
Structured query retrieval is the right fit for filter-and-aggregate workloads against a relational store with schema and indexes.	Yes / No
Hybrid retrieval is identical to dense-only retrieval and shares the same recall characteristics.	Yes / No
Random sampling of a data store is a sound retrieval strategy for production Q&A.	Yes / No

Answer:

Answer Area

Statements:	Select Yes or No:
Dense vector retrieval is well suited to recovering paraphrased intent against unstructured text.	Yes / No
Sparse term-matching retrieval recovers exact terms (such as identifiers and rare tokens) more reliably than dense alone.	Yes / No
Structured query retrieval is the right fit for filter-and-aggregate workloads against a relational store with schema and indexes.	Yes / No
Hybrid retrieval is identical to dense-only retrieval and shares the same recall characteristics.	Yes / No
Random sampling of a data store is a sound retrieval strategy for production Q&A.	Yes / No

Explanation:
Yes, Yes, Yes, No, No
Dense vector retrieval represents semantic similarity, making it suitable for matching paraphrases and conceptually related language even when the query and source do not share identical words. Sparse lexical retrieval, including BM25-style matching, retains strong sensitivity to exact tokens and is therefore valuable for identifiers, product codes, technical names, and uncommon terminology.
Structured query retrieval is appropriate when the required operation depends on explicit fields, predicates, joins, counts, grouping, or aggregation over a relational schema. In that situation, generating or invoking a constrained database query is more precise than approximating the operation through semantic similarity.
Hybrid retrieval is not identical to dense retrieval. It combines semantic and lexical candidate sets, typically followed by rank fusion or reranking. Anthropic's contextual-retrieval guidance explains that semantic search captures meaning and paraphrases, while BM25 captures exact terminology; combining them improves coverage. Contextual Retrieval, Contextual Retrieval Cookbook Random sampling is not a relevance strategy. It provides no systematic relationship between the query and selected evidence, producing unstable coverage and preventable hallucination risk. Production Q & A requires deterministic or evaluated relevance mechanisms, access filters, suitable indexes, and measurable retrieval metrics such as recall at k.
Study Guide references/topics: Dense retrieval; sparse retrieval; structured queries; hybrid search; rank fusion; retrieval evaluation; production RAG design.

NEW QUESTION: 12

You are defining where human review must remain in a planned automated pipeline. Which placement reflects sound human-in-the-loop design?
A. Place a human reviewer only after the irreversible action has already executed, making the review a post-hoc audit rather than a meaningful pre-action check or approval gate.
B. Place a human reviewer between the model's output and any high-impact, irreversible action, with explicit criteria for what the reviewer must check before approval.
C. Place a human reviewer in the loop for a randomly selected sample of requests, without defining criteria for what the reviewer should check or which output categories require mandatory review.
D. Remove all human review steps from the pipeline to maximize throughput, accepting that high-impact and irreversible actions will be taken without any human approval or oversight.
Answer: (SHOW ANSWER)
Sound human-in-the-loop design places review before a high-impact or irreversible action occurs. The reviewer must receive sufficient context and explicit approval criteria so that the checkpoint functions as a meaningful control rather than a ceremonial step. This is especially important when Claude outputs could initiate financial transactions, publish sensitive material, alter access rights, affect employment decisions, or trigger other consequential actions. Option A provides only retrospective auditing and cannot prevent harm.
Option C may support monitoring or quality measurement, but random sampling does not protect every case that requires mandatory approval, particularly when review criteria are undefined. Option D removes the accountability and intervention mechanism entirely. Therefore, B provides the strongest design by combining risk-based placement, human authority, and clearly defined verification requirements before execution.

NEW QUESTION: 13
You are identifying the highest-impact optimization for a deployment whose token cost is dominated by a long, repeated system prompt and a large retrieved context per request.
Which optimization most directly targets the dominant cost driver?
A. Increase retrieval depth on every request to maximize recall, worsening the dominant cost driver by adding more retrieved tokens per request rather than reducing them.
B. Add additional repeated content to the system prompt to give the model more guidance.
C. Move the long, repeated system prompt into a cacheable prefix and trim retrieved context to the spans relevant to each query.
D. Switch every request to the heaviest available model to maximize output quality, accepting that higher per-request inference cost compounds rather than addresses the dominant cost driver.
Answer: (SHOW ANSWER)
Option C addresses both components responsible for the excessive cost. Anthropic prompt caching allows stable, repeatedly submitted prompt material-such as system instructions, tool definitions, and reusable background information-to be placed in a consistent prefix. After that prefix is written to the cache, qualifying subsequent requests can reuse it at the lower cache-read cost instead of repeatedly processing the same content at the standard input-token rate.
Retrieval must be optimized separately. Supplying an entire document collection or excessively deep search results increases cost, consumes context capacity, and may reduce answer quality by surrounding the relevant evidence with distracting material. Retrieval should select the smallest set of authoritative passages that provides sufficient evidence for the current query. This typically requires relevance scoring, deduplication, metadata filtering, reranking, and explicit token-budget limits.
Increasing retrieval depth or expanding the repeated system prompt directly worsens the identified cost driver.
Moving every request to a heavier model changes the unit economics but does not correct inefficient context construction. The recommended optimization therefore combines prefix caching with query-specific context pruning, followed by evaluation to confirm that reduced context does not lower task accuracy.
Study Guide references/topics: [Prompt caching](https://docs.anthropic.com/en/docs/build-with-claude
/prompt-caching); [effective context engineering] (https://www.anthropic.com/engineering/effective-context- engineering-for-ai-agents); retrieval precision; token-cost analysis; cache-prefix stability.

NEW QUESTION: 14
You are responding to a stakeholder request that adds significant scope to a deployment already in flight.
Which response best aligns expectations while respecting the stakeholder's underlying need?
A. Hide the scope-addition request from the broader delivery team to avoid a difficult conversation, allowing the change to proceed without team awareness of its timeline implications.
B. Decline the request outright without acknowledging the stakeholder's underlying need or offering alternatives such as deferral, descoping, or timeline extension.
C. Acknowledge the request, describe the scope and timeline implications transparently, and propose options-defer it, descope something else, or extend the timeline-for the stakeholder to choose.
D. Accept the additional scope into the current release without describing the timeline or descoping implications, allowing the delivery plan to absorb unacknowledged changes until they surface as delays.
Answer: (SHOW ANSWER)
Option C preserves both transparency and stakeholder agency. The architect acknowledges the underlying business need while making the delivery consequences explicit. Presenting controlled alternatives converts an unbounded request into a governed decision: defer the capability, exchange it for existing scope, or accept a later release date. The selected option should be documented with revised success criteria, dependencies, risk, cost, ownership, and approval.

Concealing the request prevents accurate planning and creates unmanaged delivery risk. Rejecting it without exploring alternatives ignores legitimate stakeholder value. Quietly absorbing additional work produces scope creep and makes later delays appear unexpected. Effective lifecycle communication does not promise every requested capability; it presents the trade-offs necessary for an informed decision.
Study Guide references/topics: Scope governance; expectation alignment; transparent trade-offs; change control; stakeholder decision rights.

NEW QUESTION: 15
You are compiling continuity practices that span the deployment lifecycle.
Which two practices belong on the list? (Select two.)
Each correct answer presents a complete solution.
A. Lock decisions made in early phases to prevent revisiting them as later phases begin.
B. Archive every phase deliverable in long-term storage to preserve a record of what was produced.
C. Capture lessons learned at the end of each phase and surface them as inputs to the next phase.
D. Carry the evaluation framework and reference set forward across iterations rather than rebuilding each time.
E. Maintain a stakeholder register and notify the listed parties at every phase transition event.
Answer: (SHOW ANSWER)

NEW QUESTION: 16
You are operating an interactive assistant whose dominant performance constraint is per-turn latency. Quality on routine turns is already acceptable.
Which configuration adjustment most directly improves latency without disproportionately damaging quality?
A. Disable prompt caching entirely to ensure fresh context processing on every request, preventing stale prefix content from affecting latency-sensitive interactions.
B. Reduce retrieval depth to the top-k passages that historically cover the answer, and cache stable system- prompt content.
C. Increase retrieval depth to the corpus maximum to improve recall regardless of latency.
D. Switch every turn to the heaviest available model to maximize output quality, accepting that the increased model latency will worsen the per-turn SLO rather than improve it.
Answer: (SHOW ANSWER)
Option B targets two major contributors to per-turn latency: retrieved-context size and repeated prompt processing. Reducing retrieval to an empirically validated top-k limits the tokens Claude must process while preserving the passages that historically provide sufficient answer coverage. Caching the stable system- prompt prefix avoids repeatedly processing identical instructions.
Anthropic's latency guidance recommends choosing an appropriate model and reducing unnecessary input and output tokens. Its prompt-caching documentation explains that caching reusable prompt prefixes can reduce both repeated processing cost and latency. Reducing Latency, Prompt Caching The retrieval depth must be selected through evaluation rather than arbitrary truncation. The team should compare answer quality, retrieval recall, latency, and failure rates across candidate values and retain escalation logic for complex queries requiring deeper retrieval.
Option A removes an applicable latency optimization. Option C maximizes context consumption regardless of demonstrated benefit. Option D deliberately selects the slowest model class even though routine-turn quality already meets requirements.
Study Guide references/topics: Latency optimization; top-k retrieval; prompt caching; token reduction; quality-latency evaluation; adaptive retrieval.

Valid CCAR-P Dumps shared by EduDump.com for Helping Passing CCAR-P Exam! EduDump.com now offer the **newest CCAR-P exam dumps**, the EduDump.com CCAR-P exam **questions have been updated** and **answers have been corrected** get the **newest** EduDump.com CCAR-P dumps with Test Engine here: https://www.edudump.com/exams/Anthropic/CCAR-P/premium/ (**130** Q&As Dumps, **35%OFF** **Special Discount Code:** **freecram**)

NEW QUESTION: 17
You are selecting the documentation set that should accompany a Claude-based deployment at handoff.
Which set is most complete?
A. Architecture overview, ADRs, component-level diagrams, runbooks for common operations, an on-call playbook, and a list of known limitations.

B. Code comments scattered across individual source files with no integrating architecture overview, no ADRs, no runbooks, and no on-call playbook to orient operators or new team members.
C. A single one-page architecture diagram with no ADRs, no runbooks, no on-call playbook, and no list of known limitations to support operators or future maintainers.
D. A verbal handover walkthrough conducted on the day of transition, with no written architecture overview, ADRs, runbooks, playbook, or known limitations captured for future reference.
Answer: (SHOW ANSWER)
A complete handoff package must help different stakeholders understand, operate, troubleshoot, and evolve the Claude-based solution. An architecture overview explains the system's structure and responsibilities, while Architecture Decision Records document important decisions, alternatives, and trade-offs. Component- level diagrams provide implementation context without forcing maintainers to reconstruct dependencies from source code. Operational runbooks and an on-call playbook enable reliable incident detection, diagnosis, escalation, and recovery. Documenting known limitations is particularly important for Claude applications because model behavior can vary with prompts, context, input distribution, and external dependencies.
Options B, C, and D each provide only a partial or temporary understanding of the system. They would create knowledge gaps, increase operational risk, and make future maintenance dependent on individuals rather than durable organizational documentation.

NEW QUESTION: 18
You are reviewing instrumentation in a multi-agent system.
Which two findings constitute valid observability gaps in the instrumentation? (Select two.) Each correct answer presents a complete solution.
A. Trace spans for each agent step are exported to the shared distributed-tracing backend.
B. Latency and token usage on every span are emitted to the central metrics pipeline.
C. Tool-call payloads and outcomes are recorded with redaction applied to known sensitive fields.
D. Model identity and version on each turn are not recorded with the turn artifacts.
E. Request-scoped correlation identifiers do not propagate across agent and tool calls.
Answer: (SHOW ANSWER)
Model identity and version are essential diagnostic dimensions. Without them, operators cannot correlate a behavioral change with a model migration, reproduce the conditions of a failed turn, compare performance among model versions, or distinguish model drift from prompt, retrieval, and application defects. Therefore, the omission described in Option D is a genuine observability gap.
Option E is equally significant. A multi-agent execution normally crosses orchestration, model, agent, MCP, and tool boundaries. If the request-scoped correlation identifier or distributed trace context is lost at any boundary, the resulting spans cannot reliably be reconstructed into one end-to-end transaction. Anthropic's monitoring guidance explains that distributed tracing links a user prompt to the API requests and tool executions it initiates; it also documents propagation of W3C trace context to supported subprocesses and outbound MCP requests. Claude Code Monitoring Options A and B describe healthy telemetry coverage rather than gaps. Option C is also appropriate because recording tool interactions supports investigation, while redaction reduces the risk of exposing sensitive data in logs.
Study Guide references/topics: Integration-observability challenges at scale; distributed tracing; correlation identifiers; model-version attribution; privacy-aware tool telemetry.

NEW QUESTION: 19
You are integrating Claude Code into the team's pull-request workflow. The team wants AI-assisted review without removing human approval.
Which integration design best fits this requirement?
A. Claude Code reviews the pull request and posts a structured analysis as a comment, while a human reviewer retains the approval decision under the existing branch-protection rules.
B. Claude Code merges every pull request automatically after completing its analysis, bypassing human approval and the existing branch-protection rules.
C. Claude Code disables all existing branch-protection rules to streamline the merge process, removing human approval as a required gate.
D. Claude Code silently deletes pull requests it assesses as low quality without posting a comment or notifying the author.
Answer: (SHOW ANSWER)
Option A augments the existing review process without transferring final accountability to the model. Claude Code can inspect the change in repository context, identify logic defects, security weaknesses, regressions, and maintainability concerns, and publish structured findings for human evaluation. Existing branch protection then preserves required approvals and separation of duties. Automatic merging or disabling protection rules would convert an advisory capability into an uncontrolled authorization mechanism. Silent deletion would eliminate transparency and prevent authors from challenging or correcting the assessment.
Anthropic's Code Review documentation states that findings are posted as comments and do not approve or block the pull request, leaving the established review workflow intact. Claude Code Review

NEW QUESTION: 20

A security audit uncovers two issues: (1) all end users share a single API key, and (2) tool calls are executed without logging the initiating user.
Which two mitigations directly address these specific findings? (Select two.)

freecram.net

A. Validate structured outputs against a schema before downstream actions are executed.
B. Enforce RBAC at the retrieval layer before content enters the model context.
C. Move credentials out of the prompt context and resolve them from a server-side secret store.
D. Add actor attribution to tool-call logs so each call records the initiating user identity.
E. Replace the shared API key with per-user OAuth tokens carrying scope-restricted permissions.
Answer: D,E (LEAVE A REPLY)
Per-user, scope-restricted OAuth tokens directly eliminate the shared-credential weakness. Each token represents a distinct user or delegated identity, carries only the permissions required for that user's role, and can be revoked independently. This supports least privilege, limits the blast radius of credential compromise, and enables authorization decisions to be tied to a specific principal instead of an undifferentiated shared API key.
Actor attribution in tool-call logs directly resolves the second finding. Each audit event should record the authenticated user, delegated service identity, session or request identifier, tool name, authorization decision, relevant resource, timestamp, and outcome. This establishes accountability and permits incident investigators to reconstruct who initiated a consequential action.
Moving credentials into a server-side secret store is sound secret-management practice, but it does not by itself correct the shared-identity problem: a centrally stored key can still represent every user as the same principal. Schema validation protects downstream processing, while retrieval-layer RBAC protects information access; neither addresses the two specific audit findings.
These controls should be reinforced by short-lived tokens, server-side authorization enforcement, sensitive- field redaction, and tamper-resistant audit-log retention.
Study Guide references/topics: [Claude Code security](https://docs.anthropic.com/en/docs/claude-code
/security); [MCP authentication and OAuth] (https://docs.anthropic.com/en/docs/claude-code/mcp); least privilege; identity propagation; auditable tool execution.

NEW QUESTION: 21

You are reviewing a peer's draft system prompt that contains contradictory instructions: one section says never to speculate beyond the supplied source, while another says to confidently fill in any gaps.
Which response is most appropriate?
A. Add a priority instruction directing the model to evaluate all instructions and apply whichever appears most contextually appropriate on each request.
B. Remove or rewrite the gap-filling instruction so the prompt consistently constrains the model to source- supported content.
C. Increase temperature so output randomness masks the contradiction.
D. Keep both instructions and rely on the model to decide which one to follow on each request.
Answer: B (LEAVE A REPLY)
The prompt contains mutually incompatible behavioral requirements. The architect should resolve the contradiction explicitly so the model receives one coherent grounding policy. If the application requires source-supported answers, the prompt should prohibit unsupported completion, permit an insufficient- evidence response, and require citations or supporting passages. Adding another priority instruction leaves the underlying conflict in place and creates ambiguity about when speculation is permitted. Increasing temperature makes behavior more variable rather than more correct. Retaining both instructions produces inconsistent outputs across requests and model versions. Anthropic recommends clear, direct, specific prompting and notes that conflicting project instructions can cause Claude to select among them unpredictably. Prompting best practices ; project-instruction troubleshooting

NEW QUESTION: 22

A healthcare organization is evaluating two Claude-powered AI architectures for a clinical documentation assistant. Architecture X produces higher output quality scores but costs $0.18 per documentation session and averages 4.2 seconds per response. Architecture Y scores slightly lower on quality metrics but costs $0.09 per session and averages 2.1 seconds per response. The stated SLA requires responses under 3 seconds, and the annual volume is projected at 2 million documentation sessions.
Which evaluation approach correctly applies business value pillar analysis to this decision?
A. Select Architecture Y based solely on the 50% cost reduction, since solution cost is the most important value pillar in healthcare budget-constrained environments.
B. Select Architecture X because the higher quality scores justify the cost premium, and any SLA gap can be addressed through infrastructure optimization after deployment.

C. Eliminate Architecture X on SLA grounds, then evaluate Architecture Y against the efficiency and solution cost pillars by calculating annual cost difference and assessing whether the quality delta materially impacts clinical workflow productivity.
D. Recommend a hybrid approach using Architecture X for complex cases and Architecture Y for routine cases, without additional analysis, since this preserves quality where it matters most.
Answer: (SHOW ANSWER)
Architecture X fails a stated production gate: its 4.2-second average exceeds the required three-second response SLA. It should therefore not remain a qualifying architecture unless it is redesigned and retested before selection. A promise to optimize after deployment is not evidence of compliance.
Architecture Y meets the latency requirement at 2.1 seconds. Its annual inference cost is $180,000 at two million sessions, compared with $360,000 for Architecture X-a difference of $180,000 annually. However, cost alone does not establish suitability. The organization must determine whether Y's lower quality score causes material omissions, correction effort, clinician dissatisfaction, documentation risk, or reduced workflow productivity.
Option C correctly treats the SLA as a threshold and then analyzes the remaining architecture across cost, efficiency, quality, and productivity. Option A ignores clinical quality. Option B disregards an explicit SLA failure. Option D may eventually be viable, but it introduces routing complexity and leaves unanswered how complex cases are detected, whether X's latency becomes acceptable for them, and how both paths are evaluated.
The final choice should be gated by clinical-quality evaluation, human review requirements, safety metrics, operational latency, and total annual cost.
Study Guide references/topics: Business-value pillars; performance gates; solution cost; workflow efficiency; quality trade-offs; healthcare architecture evaluation.

NEW QUESTION: 23
You are compiling guardrail tactics for a customer-facing assistant.
Which two tactics belong on the guardrail list?
Each correct answer presents a complete solution.
A. Embed approved override phrases that let trusted users relax guardrails on demand.
B. Lower sampling temperature globally to reduce the chance of off-policy completions.
C. Validate the model output against a structured schema before downstream actions are taken.
D. Layer prompt-level guardrails with runtime content checks rather than relying on either alone.
E. Rely on a single hardened system prompt that enumerates every disallowed behavior.
Answer: (SHOW ANSWER)
Option C introduces a deterministic validation boundary between probabilistic model output and downstream execution. Structured validation can reject missing fields, invalid types, unauthorized actions, malformed arguments, and values outside approved enumerations before any tool or business process acts on the response. Anthropic recommends Structured Outputs when guaranteed schema conformance is required.
Increase Output Consistency
Option D implements defense in depth. Prompt-level instructions establish intended behavior, while runtime classifiers, content filters, policy engines, output validators, tool permissions, and human confirmation gates provide controls outside the model. Anthropic's guardrail guidance recommends combining safeguards rather than depending on one mechanism. Mitigate Jailbreaks and Prompt Injections Option A creates a bypass mechanism that adversaries may discover or imitate. Option B may alter output variability but is not a security boundary and cannot guarantee policy compliance. Option E places complete reliance on a system prompt, which remains vulnerable to ambiguity, conflicting inputs, model errors, and indirect prompt injection.
Guardrails should be risk-based, independently enforceable where possible, observable, tested against adversarial cases, and proportional to the consequences of downstream actions.
Study Guide references/topics: Layered guardrails; structured validation; runtime content checks; prompt injection; deterministic enforcement; defense in depth.

NEW QUESTION: 24
A Claude architect is configuring a shared Claude Code environment for a twelve-person development team.
Which two configuration decisions most directly establish a consistent, secure team environment? (Select two.)
A. Commit shared MCP server definitions and tool permissions in project-scope configuration files alongside the repository.
B. Apply managed configuration to enforce non-overridable security and compliance policies across all team members.
C. Set editor theme and display preferences at the project scope so they are uniform across workstations.
D. Instruct each engineer to configure their preferred Claude model in personal user-scope settings.
E. Store the team's shared API key in the project-scope settings.json so all engineers use the same credential.

Answer: (SHOW ANSWER)
Project-scope configuration is appropriate for shared MCP definitions, permissions, hooks, and other repository-specific behavior that every team member should receive consistently. Managed configuration complements this by enforcing organization-wide security and compliance requirements that individual developers cannot override. Together, these mechanisms separate collaborative project configuration from centrally controlled policy. Editor themes are personal preferences and do not materially establish a secure environment. Allowing each engineer to select an arbitrary model can undermine consistency when model choice affects behavior, cost, or approved-use requirements. Storing a shared API key in a committed project file is a serious credential-management failure. Claude Code documentation identifies .claude/settings.json as the shared project file and managed settings as the highest-precedence organizational policy layer. Claude Code: Settings files and precedence

NEW QUESTION: 25
An engineer inadvertently commits an API key to the repository by placing it in .claude/settings.json.
Which configuration design principle was violated?
A. Project-scope configuration should always override user-scope configuration.
B. Managed configuration must be applied before project-scope configuration is read.
C. Credentials must be resolved at runtime from a secret store and never stored in version-controlled files.
D. MCP server definitions must reside in environment variables to prevent scope conflicts.
Answer: (SHOW ANSWER)
A shared project settings file is intended to be committed so that team-level configuration can travel with the repository. Consequently, it must never contain API keys, access tokens, passwords, or comparable secrets.
Credentials should be injected at runtime through an approved secret-management mechanism, environment- specific credential provider, or delegated authentication flow. This separates distributable configuration from sensitive authentication material and supports rotation, revocation, auditing, and least privilege. Options A and B concern configuration precedence rather than secret handling. Option D is incorrect because MCP server definitions may legitimately be stored in project configuration; only their sensitive credential values must remain outside version control. Because the key was exposed, the operational response should also include immediate revocation or rotation and investigation of repository history. Claude Code: Settings files and precedence

NEW QUESTION: 26
Engineering leadership wants to roll out Claude Skills to 280 developers across 14 teams. Skills will encode internal coding standards, code-review checklists, and incident-postmortem templates. Leadership has asked how to govern Skill authorship so that Skills remain trustworthy without bottlenecking on a single central team.
Which governance model should you recommend?
A. Per-developer authorship across the 280 engineers with no team-level coordination required.
B. Centralized authorship by a single platform team responsible for every Skill produced.
C. Fully decentralized authorship across the 14 teams with no review before publication.
D. Federated authorship across the 14 teams with a central review and publication gate.
Answer: (SHOW ANSWER)
Federated authorship allows each domain team to maintain the procedural knowledge it understands while preserving organization-wide trust through a central publication gate. Team authors can update coding standards, review practices, and postmortem procedures without waiting for a single platform group to write every change. Central reviewers can validate ownership, security, tool permissions, duplication, versioning, testing evidence, and compliance before publication. Per-developer or fully decentralized publication creates inconsistent and potentially unsafe Skills. Fully centralized authorship becomes a scalability bottleneck and separates content maintenance from domain expertise. Anthropic describes Skills as packages of instructions, scripts, and resources and warns that repository Skills may grant broad tool access, making formal review essential. Claude Code Skills

NEW QUESTION: 27
You are diagnosing a Claude Code session whose subagent uses 50,000 tokens of context before the engineer types a single message.
Which root cause is most likely?
A. The developer ' s keyboard layout or input-method configuration is the cause of the elevated context consumption, introducing extra tokens before the engineer types any message.
B. Many MCP servers are configured, each contributing tool definitions to the context budget; Tool Search is not enabled, so all definitions load upfront.
C. The model has internal personal preferences or default behaviors that silently consume large portions of context budget before any user message is processed, independent of tool configuration.

D. The font rendering or display-scaling settings of the IDE are converting visual output into additional context tokens, causing the high pre-session context consumption.
Answer: (SHOW ANSWER)
Claude Code constructs the model context before the first user message. That initial context can include the system prompt, project instructions, subagent configuration, MCP server instructions, and tool definitions expressed as JSON schemas. When numerous MCP servers expose many tools and Tool Search is disabled, those schemas must be loaded upfront. A large collection of verbose tool descriptions and parameter schemas can therefore consume tens of thousands of tokens before the engineer enters any text.
Tool Search is specifically designed to prevent this form of context inflation. With Tool Search enabled, Claude initially receives only server instructions and compact tool-identification information; full tool schemas are deferred until a relevant tool is discovered and required. Anthropic consequently describes Tool Search as the mechanism for keeping MCP context consumption low as integrations scale.
Keyboard layouts, font rendering, and display scaling operate outside the language-model context and cannot generate API input tokens. Likewise, model behavior does not silently reserve 50,000 tokens for personal preferences. The diagnostic action is to inspect configured MCP servers and tool counts, determine whether Tool Search has been disabled, and remove or defer unnecessary definitions.
Study Guide references/topics: [MCP Tool Search and context scaling](https://docs.anthropic.com/en/docs
/claude-code/mcp); MCP server configuration; subagent context budgeting; tool-schema optimization.

NEW QUESTION: 28
You are rolling out monitoring for a Claude-based deployment and must complete the specification steps before instrumenting the deployment.
Which two steps must be completed BEFORE instrumenting the deployment to emit metrics and traces?
(Select two.)
Each correct answer presents part of the solution.
A. Define the metrics and the slices the deployment will monitor across normal and adversarial traffic.
B. Tune the alert thresholds based on observed normal-state distributions to reduce false positives.
C. Define the service-level objectives (SLOs) and the error budgets the deployment will be held to.
D. Build the dashboards that surface metrics across slices at the cadence the team operates on.
E. Document the dashboards, alerts, and runbooks for the on-call rotation that will respond.
Answer: (SHOW ANSWER)
Instrumentation must be driven by an explicit measurement specification. The team must first define the metrics and evaluation slices required to detect performance, cost, quality, safety, or reliability failures. Slices should include representative traffic, adversarial inputs, user segments, tools, and relevant content categories.
SLOs and error budgets must also be established before instrumentation because they determine required measurements, aggregation windows, and acceptable failure levels. Alert thresholds cannot be tuned until normal-state data has been collected. Dashboards and on-call documentation depend on the final metric definitions and observed operational behavior. Anthropic supports metrics, event, and trace export through OpenTelemetry, but telemetry is useful only when tied to measurable success and operating objectives.
Claude Code monitoring

NEW QUESTION: 29
An architect is reviewing a Claude-based candidate-screening tool prior to deployment. A stakeholder asserts that because the model was not trained on company data, no bias evaluation is necessary.
Which two responses most accurately challenge this assertion? (Select two.)
A. Transparency obligations are satisfied by disclosing that an AI system is in use, without further evaluation.
B. Fairness testing is required only when the training dataset is known to contain protected-class labels.
C. The model may carry demographic biases from pretraining that manifest in screening outcomes regardless of fine-tuning.
D. The system should be evaluated on outcome-disparity metrics across protected groups before deployment.
E. Bias evaluation is unnecessary when the model provider has published a responsible-use policy.
Answer: (SHOW ANSWER)
The absence of company-specific training does not establish unbiased behavior. A foundation model may reflect demographic associations or unequal treatment patterns originating from pretraining data, model behavior, prompt framing, retrieval content, or the surrounding decision workflow. Because candidate screening can materially affect employment opportunities, the complete system must be evaluated across relevant protected groups before deployment. Appropriate

measures include selection-rate differences, false- positive and false-negative rates, calibration, intersectional slices, and human override patterns. Disclosure that AI is used does not replace empirical fairness testing. Protected-class labels need not have been explicit in training data for proxy variables or inferred attributes to produce disparities. Provider policies similarly do not validate a customer's specific implementation. Anthropic system cards document discriminatory-bias evaluations, reinforcing the need for application-level testing. Claude model system card

NEW QUESTION: 30

A solutions architect is analyzing stakeholder feedback collected after the first quarter of a Claude-powered procurement automation deployment. The feedback includes four statements: (1) "Our procurement team is processing 3x more purchase orders per analyst per day." (2) "We have eliminated the manual data entry role entirely and redeployed those staff to vendor relationship management." (3) "The API integration costs are running 40% over the projected per-transaction budget." (4) "Response latency during end-of-month batch runs is averaging 11 seconds, against our committed 5-second SLA." Which of the following correctly identifies the primary business value pillar each stakeholder statement represents?

Freecram.net

A. Statement 1: Efficiency; Statement 2: Transformation; Statement 3: Solution Cost; Statement 4: Performance SLA

B. Statement 1: Productivity; Statement 2: Efficiency; Statement 3: Solution Cost; Statement 4: Performance SLA

C. Statement 1: Transformation; Statement 2: Productivity; Statement 3: Solution Cost; Statement 4: Efficiency

D. Statement 1: Efficiency; Statement 2: Productivity; Statement 3: Performance SLA; Statement 4: Solution Cost

Answer: A (LEAVE A REPLY)

Statement 1 measures efficiency because the same operating unit-one procurement analyst-now processes three times as many purchase orders per day. It reflects increased throughput and reduced effort per transaction within the existing procurement process.

Statement 2 represents transformation. The organization has not merely accelerated manual data entry; it has eliminated that role from the workflow and redirected employees toward higher-value vendor relationship management. This is a structural redesign of responsibilities and operating practices.

Statement 3 is a solution-cost issue because the measured API integration expense exceeds the approved per- transaction financial model by 40%. The architect must investigate token consumption, model routing, caching, retrieval overhead, tool calls, and integration charges while protecting the required quality level.

Statement 4 directly concerns the performance SLA. The observed 11-second latency is compared with an explicit contractual or operational target of five seconds. This is not simply an efficiency complaint; it is a measurable breach of a defined service-level commitment.

The architect should map each stakeholder statement to its dominant value pillar before recommending remediation because throughput, organizational transformation, cost control, and SLA compliance require different architectural responses.

Study Guide references/topics: Business-value pillars; efficiency; transformation; solution cost; performance SLAs; stakeholder-outcome classification.

NEW QUESTION: 31

A Claude architect is implementing safety controls for a customer-facing advice assistant that must never provide regulated investment recommendations.

Which two guardrail implementations most directly enforce this constraint? (Select two.)

A. Increase response temperature to introduce variability that reduces the likelihood of specific recommendations.

B. Add an output classifier that detects and blocks responses containing regulated investment- recommendation language.

C. Limit session length to reduce the volume of queries processed per user per day.

D. Log all user queries to a SIEM for post-hoc compliance review.

E. Define explicit out-of-scope categories in the system prompt with fixed refusal phrasing for investment advice requests.

Answer: (SHOW ANSWER)
The required control is best implemented through layered prevention. The system prompt should explicitly classify personalized investment recommendations as out of scope, instruct Claude not to recommend particular securities or transactions, and provide standardized refusal language. Fixed refusal behavior reduces ambiguity and gives the model a clear alternative response when users request prohibited advice or attempt to reframe it indirectly.
An output classifier supplies an independent enforcement layer after generation but before delivery. It should detect prohibited recommendation language, individualized buy-or-sell instructions, portfolio allocations, and equivalent regulated content. A detected violation must block or replace the response rather than merely record it. This protects against prompt circumvention, unexpected phrasing, and occasional model noncompliance.
Increasing temperature makes responses less predictable and weakens control reliability. Session-length limits regulate volume, not content. SIEM logging is valuable for monitoring, investigation, and compliance evidence, but it is detective rather than preventative: the prohibited recommendation may already have reached the customer.
The completed design should evaluate both layers using adversarial prompts, paraphrases, multi-turn escalation, and classifier false-negative testing.
Study Guide references/topics: [Mitigating jailbreaks and prompt injections](https://docs.anthropic.com/en
/docs/test-and-evaluate/strengthen-guardrails/mitigate-jailbreaks); layered guardrails; explicit scope boundaries; standardized refusals; output classification; pre-delivery enforcement.

Valid CCAR-P Dumps shared by EduDump.com for Helping Passing CCAR-P Exam! EduDump.com now offer the **newest CCAR-P exam dumps**, the EduDump.com CCAR-P exam **questions have been updated** and **answers have been corrected** get the **newest** EduDump.com CCAR-P dumps with Test Engine here: https://www.edudump.com/exams/Anthropic/CCAR-P/premium/ (**130** Q&As Dumps, **35%OFF** **Special Discount Code:** **freecram**)

NEW QUESTION: 32
The engineering lead at Trenova Systems, Inc. is evaluating two proposals for improving developer workflows using Claude-assisted tooling. Proposal A adds Claude Code to the IDE for inline code generation and review. Proposal B routes all code-generation requests through a shared Slack bot without IDE integration.
Which two observations most accurately evaluate these proposals against workflow-improvement objectives?
(Select two.)
A. Proposal B is superior because centralizing requests in Slack creates an auditable log of all code- generation activity.
B. Proposal B introduces workflow friction by requiring developers to leave the IDE, undermining the productivity objective.
C. Proposal A reduces context switching by providing AI assistance at the point of development without requiring a separate tool.
D. Proposal B improves workflow velocity because Slack notifications create an asynchronous review queue.
E. Both proposals are equivalent because the model capabilities are identical regardless of the integration point.
Answer: B,C (LEAVE A REPLY)
Workflow improvement depends on where assistance is delivered, not solely on the underlying model. IDE integration makes code, selections, diagnostics, diffs, and repository context available at the point where developers work. This reduces context switching and avoids repeatedly transferring code between systems.
Routing every request through Slack introduces an additional interaction surface, separates assistance from development context, and increases friction. A Slack record may support auditability, but it does not outweigh the stated productivity objective, and equivalent audit data can be captured through approved telemetry and repository history. Notifications do not inherently improve engineering velocity. Anthropic's Claude Code integrations are designed to place agentic coding capabilities directly within terminal and IDE workflows.

NEW QUESTION: 33
You are supporting an EU-based deployment with GDPR obligations.
Which combination of measures best supports the deployment's GDPR posture?
A. enterprise-tier deployment with a signed Data Processing Addendum, defined data-retention configuration, redaction of personal data not needed for the task, and documented data-subject-rights handling
B. disabling all data-retention configuration, redaction controls, and data-subject-rights handling to simplify day-to-day operations, accepting the resulting GDPR compliance exposure
C. using a personal Claude account tier for processing EU personal data at scale, with no signed Data Processing Addendum and no documented data-subject-rights handling procedure
D. pasting full EU personal data into every prompt to "give Claude complete context" without considering purpose limitation, data minimization, or the organization's Data Processing Addendum obligations
Answer: (SHOW ANSWER)

Option A combines the contractual, technical, and operational measures relevant to a defensible GDPR posture. For commercial services, the Data Processing Addendum defines processor obligations and incorporates safeguards such as Standard Contractual Clauses where applicable. Anthropic states that its DPA is incorporated into its Commercial Terms, while the customer remains responsible for determining lawful purposes, issuing processing instructions, and fulfilling controller obligations.
A defined retention configuration ensures personal data is not retained indefinitely without a documented purpose. Redacting personal information that is unnecessary for the task implements data minimization and reduces exposure in prompts, logs, retrieval stores, and model outputs. Documented data-subject-rights procedures support access, correction, deletion, restriction, and other applicable requests by ensuring the organization can locate and act on relevant records.
Enterprise deployment alone does not establish compliance; the organization must configure and govern the service consistently with its processing activities, risk assessment, lawful basis, and contractual commitments.
Options B, C, and D intentionally remove those controls or promote excessive collection.
Study Guide references/topics: [Anthropic Data Processing Addendum guidance](https://privacy.anthropic.
com/en/articles/7996862-how-do-i-view-and-sign-your-data-processing-addendum-dpa); [commercial data- retention practices] (https://privacy.anthropic.com/en/articles/7996866-how-long-do-you-store-my- organization-s-data); GDPR data minimization; retention governance; data-subject-rights procedures.

NEW QUESTION: 34
You are selecting the documentation set that should accompany a Claude-based deployment at handoff.
Which set is most complete?
A. A single one-page architecture diagram with no ADRs, no runbooks, no on-call playbook, and no list of known limitations to support operators or future maintainers.
B. Code comments scattered across individual source files with no integrating architecture overview, no ADRs, no runbooks, and no on-call playbook to orient operators or new team members.
C. A verbal handover walkthrough conducted on the day of transition, with no written architecture overview, ADRs, runbooks, playbook, or known limitations captured for future reference.
D. Architecture overview, ADRs, component-level diagrams, runbooks for common operations, an on-call playbook, and a list of known limitations.
Answer: (SHOW ANSWER)

NEW QUESTION: 35
You are designing a feedback session for a deployment in flight.
Which structure best supports productive stakeholder feedback?
A. Hold an open-ended meeting with no pre-distributed agenda or artifacts, and rely on participants' memory to carry decisions and follow-up owners forward.
B. Define a focused agenda, share the artifacts in advance, capture decisions and follow-ups in writing, and confirm action owners and dates.
C. Distribute artifacts at the session start rather than in advance, so stakeholders review materials in real time without preparation before the discussion begins.
D. End the session without recording decisions, follow-up items, action owners, or dates, relying on participant memory to carry the session's outcomes forward.
Answer: B (LEAVE A REPLY)
A structured feedback session must enable informed review, traceable decisions, and accountable follow- through. Option B allows stakeholders to examine architecture diagrams, evaluation results, risk registers, or prototypes before the meeting. The focused agenda keeps discussion aligned with decisions that must be made rather than allowing the session to become an unstructured status conversation. Written decision records prevent different participants from leaving with incompatible interpretations. Assigning an owner and due date to every action converts feedback into executable delivery work. Options A, C, and D depend heavily on memory, provide inadequate preparation time, and create weak decision traceability. In-flight deployments particularly require documented feedback because accepted changes may affect scope, cost, risk, testing, and release timing.
Study Guide references/topics: Stakeholder feedback; decision records; action ownership; lifecycle alignment; communication planning.

NEW QUESTION: 36
A revenue projection assistant has missed its monthly cost target by 38 percent. Profiling shows three contributors: a 6,000-token policy preamble repeated on every call (45 percent of cost), retrieval of historical sales chunks averaging 3,000 tokens per call (30 percent), and inference on a flagship-tier model (25 percent).
Stakeholders require that projection accuracy remain unchanged.
Which two optimizations should you sequence first to reduce cost without affecting accuracy? (Select two.) Each correct answer presents part of the solution.
A. Reduce the number of historical sales chunks retrieved across each query run.

B. Truncate the policy preamble to remove non-essential clauses from the prompt.
C. Enable prompt caching on the static policy preamble across the recurring calls.
D. Switch the workload to a smaller, faster Claude model tier across all queries.
E. Cache common retrieved sales chunks accessed across many of the daily queries.
Answer: (SHOW ANSWER)
The required sequence must reduce repeated computation without changing the information or model capability used to generate projections. Prompt caching the static 6,000-token policy preamble directly addresses the largest cost contributor while preserving the complete instruction set. Anthropic states that cache reads cost substantially less than uncached input tokens, making repeated, stable prompt prefixes an ideal caching target. Prompt Caching Caching frequently reused historical-sales chunks applies the same principle to the retrieval layer. When identical, version-controlled chunks are repeatedly fetched and supplied to the model, caching their retrieval or reusable representation eliminates redundant work while maintaining the same evidence available to the projection process. The cache must use appropriate invalidation or source-version keys so updated sales data cannot be replaced by stale content.
Options A and B modify the information supplied to Claude. Fewer sales chunks could remove relevant historical evidence, while truncating policy instructions could alter constraints or projection behavior. Option D introduces a model-capability change and therefore cannot guarantee unchanged accuracy without a comparative evaluation. Those interventions may be considered later, but only after representative regression testing establishes equivalence.
Study Guide references/topics: Cost profiling; prompt caching; retrieval caching; cache invalidation; accuracy- preserving optimization; model and context trade-offs.

NEW QUESTION: 37
You are producing an architecture guide for a new deployment and must complete the planning steps before drafting each section.
Which two steps must be completed BEFORE drafting each section of the guide? (Select two.) Each correct answer presents part of the solution.
A. Identify the audience and the questions the guide must answer for that audience.
B. Translate the guide into the supported regional languages for the candidate population.
C. Validate the guide with the implementation team and incorporate corrections.
D. Establish the document under version control with a defined review cadence and approver list.
E. Outline the guide sections covering the overview, components, contracts, flows, runbooks, and limitations.
Answer: A,E (LEAVE A REPLY)
Effective documentation begins with audience analysis and structural planning. Identifying the audience determines the required technical depth, terminology, decision context, and operational emphasis. Defining the questions the guide must answer prevents sections from becoming generic descriptions. The outline then establishes complete coverage across architecture overview, component responsibilities, interface contracts, dominant flows, operational runbooks, and known limitations. Translation occurs only after stable source content exists. Implementation-team validation also follows an initial draft because there must be material to review. Version control, reviewers, and cadence are important governance controls, but they do not replace the two content-planning activities required before section drafting. The guide should be planned around the decisions and tasks its intended readers must perform.

NEW QUESTION: 38
You are a solution architect designing a Claude-based assistant with access to 60 internal tools across multiple business domains. Loading every tool definition on every request increases token usage and time to first response.
Which design pattern best addresses this issue without sacrificing capability breadth?
A. Apply progressive tool discovery so a curated initial subset is exposed and additional tools are loaded on demand based on the task.
B. Use a separate model call to summarize all 60 tool definitions before each user turn.
C. Increase the maximum context length and load all 60 tool definitions on every request, accepting the higher token cost and latency as necessary for full capability.
D. Hard-code a fixed set of five tools per request to reduce token usage, regardless of whether those tools are relevant to the current task.
Answer: (SHOW ANSWER)
Progressive tool discovery preserves the complete capability catalog while limiting initial context to a curated set of common tools and a discovery mechanism. When the task requires an additional capability, Claude searches the catalog and loads only the relevant tool definitions.
Anthropic's Tool Search mechanism implements this pattern through deferred loading. The context initially contains the search tool and any deliberately non-deferred tools; relevant definitions are expanded only when discovered. This reduces token consumption and helps maintain tool-selection accuracy in large catalogs.
Tool Search Tool

Option B adds a model call and still repeatedly processes all definitions, increasing latency and cost while potentially losing important schema details. Option C treats a larger context window as permission to waste context and does not address time-to-first-response overhead. Option D reduces capability breadth because the fixed five tools may be irrelevant to a given task.
The architect should provide clear tool and server descriptions, define which high-frequency tools remain immediately available, and evaluate discovery recall, selection accuracy, latency, and context savings across representative workflows.
Study Guide references/topics: Progressive discovery; deferred tool loading; MCP scaling; context optimization; tool-selection accuracy; on-demand capability loading.

NEW QUESTION: 39
You are a solution architect evaluating candidate use cases for a Claude-based program.
For each scenario, select Yes if Claude is appropriate as the primary solution at the architectural level.
Otherwise, select No.

Answer Area

Scenarios:	Select Yes or No:
drafting first-pass investigative reports from semi-structured incident logs for analyst review	Yes / No
computing real-time fraud scores at sub-50-millisecond latency on a streaming pipeline	Yes / No
long-context contract review with structured clause extraction and deviation flagging	Yes / No
replacing a vector index for semantic retrieval over a multi-million-document corpus	Yes / No
routing inbound support tickets into 30 categories with reasoning for the chosen routing	Yes / No

Answer:

Answer Area

Scenarios:	Select Yes or No:
drafting first-pass investigative reports from semi-structured incident logs for analyst review	Yes / No
computing real-time fraud scores at sub-50-millisecond latency on a streaming pipeline	Yes / No
long-context contract review with structured clause extraction and deviation flagging	Yes / No
replacing a vector index for semantic retrieval over a multi-million-document corpus	Yes / No
routing inbound support tickets into 30 categories with reasoning for the chosen routing	Yes / No

Explanation:
Drafting first-pass investigative reports from semi-structured incident logs for analyst review - Yes Computing real-time fraud scores at sub-50-millisecond latency on a streaming pipeline - No Long-context contract review with structured clause extraction and deviation flagging - Yes Replacing a vector index for semantic retrieval over a multi-million-document corpus - No Routing inbound support tickets into 30 categories with reasoning for the selected route - Yes Claude is well suited to language-intensive analysis, summarization, extraction, classification, and drafting.
Investigative-report drafting, contract analysis, and support-ticket routing all require interpretation of unstructured or semi-structured language and can incorporate human review. A sub-50-millisecond fraud- scoring path requires deterministic, specialized real-time processing and should not place LLM inference on the critical path. Claude also does not replace the retrieval index used to search millions of documents; it consumes the results produced by that infrastructure. The correct architectural decision separates tasks requiring semantic language reasoning from deterministic computation and information-retrieval infrastructure. Model selection and evaluation must then confirm that the chosen Claude model satisfies quality, latency, and cost requirements.

NEW QUESTION: 40
You are explaining the precedence of Claude Code configuration scopes to the team.
Which precedence ordering, from highest to lowest, is correct?
A. Local # managed # user # command-line arguments # project

B. Project # user # managed # local # command-line arguments
C. Managed # command-line arguments # local # project # user
D. User # project # local # command-line arguments # managed
Answer: (SHOW ANSWER)
The correct general precedence is managed configuration, command-line arguments, local configuration, project configuration, and then user configuration.
Managed configuration has the highest authority because it contains centrally administered organizational and compliance controls that ordinary engineers must not be able to override. Command-line arguments provide temporary session-level overrides where the relevant setting permits them. Local settings apply to one user within a specific repository and take precedence over the shared project configuration. Project settings apply to every contributor using that repository and override general user preferences. User settings form the lowest tier because they provide personal defaults across projects and apply only when no higher-precedence source supplies a value.
Anthropic documents this order explicitly: managed, command line, local, project, and user. It also notes that permission rules may merge rather than behave as simple scalar overrides, and certain security-sensitive keys have specialized precedence behavior. Claude Code Configuration Scopes The ordering allows personal customization without permitting it to defeat repository conventions or centrally mandated security requirements. Architects must still examine the merge behavior of the individual configuration property rather than assuming every array or permission rule is replaced wholesale.
Study Guide references/topics: Claude Code configuration; scope precedence; managed policies; local overrides; project settings; user defaults.

NEW QUESTION: 41
You are running a controlled experiment to compare two prompts and must complete the design steps before executing the experiment.
Which two steps must be completed BEFORE running the experiment with random assignment? (Select two.) Each correct answer presents part of the solution.
A. Determine the minimum detectable effect size and the sample size needed for power.
B. Decide whether to promote, reject, or iterate the candidate based on the analysis.
C. Define the hypothesis and the primary success metric for the comparison.
D. Analyze the results against the predefined success metric and significance threshold.
E. Document the recommendation, the trade-offs accepted, and the alternatives considered.
Answer: (SHOW ANSWER)
A controlled prompt experiment must begin with a falsifiable hypothesis and a predefined primary metric.
Option C prevents the team from examining results first and then selecting whichever metric makes the candidate look successful. The metric might measure task accuracy, rubric score, citation validity, escalation rate, latency, cost, or another criterion directly connected to the hypothesis.
Option A determines whether the experiment can detect a practically meaningful improvement. The minimum detectable effect expresses the smallest difference worth acting upon, while the power calculation determines the required sample size. Without this step, the experiment may be too small to detect a real improvement or unnecessarily large and expensive.
Random assignment should then distribute representative traffic between the control and candidate prompts while controlling model version, retrieval configuration, tool availability, and other confounding variables.
Options B and D occur after data collection. Option E follows the completed analysis and decision. The team should also define significance thresholds, stopping rules, guardrail metrics, exclusion criteria, and treatment of repeated observations before launch.
Study Guide references/topics: Prompt A/B testing; hypothesis definition; primary metrics; minimum detectable effect; statistical power; random assignment; decision sequencing.

NEW QUESTION: 42
You are integrating AI-assisted tooling into the team's documentation workflow. The team wants generated documentation that stays grounded in the actual code.
Which integration approach best fits this requirement?
A. Generate documentation from the model's training-data recall without reading any of the actual repository code, accepting that the output will not reflect the current implementation.
B. Have the subagents publish generated documentation directly to the public-facing site without passing through the team's normal review workflow or any human approval step.
C. Configure subagents that read the relevant code files via filesystem and code-search tools, generate the documentation, and emit changes through the team's normal review workflow.
D. Disable all filesystem and code-search tools so the subagents cannot read any repository code, accepting that documentation generation will be entirely disconnected from the actual implementation.
Answer: (SHOW ANSWER)

Documentation remains grounded only when the generating agent can inspect the authoritative implementation. A purpose-built subagent should receive scoped filesystem and code-search capabilities, identify the relevant modules, interfaces, configuration files, tests, and call sites, and generate documentation from that evidence. This approach minimizes reliance on training-data recall, which cannot represent the repository's current private code or recent changes.
Subagents are useful because repository exploration occurs in their own context, allowing them to return focused findings or proposed edits without filling the main conversation with every file read. Their tool access should follow least privilege: read access to required paths, narrowly controlled write access, and explicit exclusions for secrets or unrelated repositories.
Generated documentation should enter the team's established change-control process as a patch, branch, or pull request. Normal review, automated checks, link validation, example execution, and human approval provide accountability and prevent unverified material from being published. Direct publication bypasses these safeguards, while disabling repository tools makes factual grounding impossible.
Study Guide references/topics: [Claude Code custom subagents](https://docs.anthropic.com/en/docs/claude- code/sub-agents); [codebase exploration workflows] (https://docs.anthropic.com/en/docs/claude-code /common-workflows); source-grounded generation; least-privilege tools; human review and lifecycle controls.

NEW QUESTION: 43
The compliance team at a firm has approved a Claude Skill that generates client-facing investment summaries. The Skill includes the firm's required disclaimers and prohibited-language list. A product manager has asked whether additional guardrails are needed at the application layer or whether the Skill alone is sufficient.
Which two guardrail responsibilities should remain at the application layer rather than the Skill? (Select two.) Each correct answer presents part of the solution.
A. Log every generated summary to the firm's compliance audit trail for retention.
B. Verify the requesting user is authorized to generate investment summaries at all.
C. Format output sections according to the firm's standardized house style guidelines.
D. Apply the disclaimer template that the compliance team has standardized firm-wide.
E. Apply the prohibited-language list that the compliance team maintains and updates.
Answer: (SHOW ANSWER)
Authorization and compliance logging must remain enforceable application-layer controls. Option B requires the application to authenticate the requester, resolve the user's role and entitlements, and deny the operation when the user lacks authority. A Skill can instruct Claude how to behave, but it is not a reliable security boundary and cannot replace identity-based access control enforced by the surrounding system.
Option A similarly belongs at the application layer. The application must create an authoritative audit event containing the requester, timestamp, Skill and model version, relevant input or content identifiers, generated output, review status, and retention classification. Audit evidence cannot depend on whether the model remembers or chooses to invoke a logging instruction.
Options C, D, and E describe reusable generation behavior that appropriately belongs within the approved Skill: house style, the standardized disclaimer, and the prohibited-language policy. The application may still validate their presence after generation, but the Skill remains the primary reusable content package.
The correct architecture separates behavioral guidance from deterministic enforcement. Skills supply domain instructions and templates; the application enforces identity, authorization, audit retention, data access, output validation, and consequential-action controls.
Study Guide references/topics: Skills versus application controls; authentication; authorization; compliance logging; deterministic guardrails; policy enforcement boundaries.

NEW QUESTION: 44
You are running a discovery engagement for a new Claude-based capability and must complete the requirements-gathering steps before validating with stakeholders.
Which two steps must be completed BEFORE validating the captured requirements with stakeholders? (Select two.) Each correct answer presents part of the solution.
A. Produce architecture decision records that capture the rationale for major design choices.
B. Capture the business goals, success criteria, and the in-scope user population for the engagement.
C. Document the consolidated requirements with traceability to the source stakeholder for each item.
D. Schedule the rollout milestones and dependencies with engineering and product partners.
E. Identify the non-functional constraints covering latency, cost, audit, data sensitivity, and regulatory needs.
Answer: (SHOW ANSWER)
Requirements must first establish what the capability is expected to accomplish, who will use it, and which constraints bound the solution. Option B captures the business purpose, success criteria, and intended user population. Option E identifies critical non-functional requirements that will influence model selection, integration patterns, deployment boundaries, evaluation design, and operational controls. These inputs must exist before stakeholders can validate whether the captured requirements accurately represent their needs.

Architecture decision records are produced after meaningful alternatives and design decisions exist. Detailed traceability is important but follows the initial elicitation and consolidation process. Rollout scheduling occurs later, after requirements and architecture have been validated. Anthropic likewise recommends defining measurable success criteria and empirical evaluation methods before optimizing prompts or finalizing the implementation. Anthropic: Prompt engineering overview

NEW QUESTION: 45
You are identifying signals that a deployment should re-enter design rather than continue iterating in place.
Which signal most directly indicates the need for a new design cycle?
A. A runbook step requires a clarification edit to improve on-call guidance accuracy, which can be handled as a documentation update without changes to component responsibilities or core contracts.
B. A dashboard alert threshold needs a small numerical adjustment to reduce false-positive noise, which can be handled as an operational configuration change without a new design cycle.
C. A minor copy edit is requested in a customer-facing string within the existing UI, which can be handled as a localized content change without altering component responsibilities or contracts.
D. The system's current architecture cannot meet the new requirements without changes to component responsibilities or core contracts.
Answer: (SHOW ANSWER)
A new design cycle is warranted when the required change crosses an architectural boundary. Option D explicitly states that the existing architecture cannot satisfy the new requirements without redistributing component responsibilities or altering core contracts. Such changes may affect service ownership, orchestration, data flows, integration interfaces, security boundaries, failure handling, evaluation strategy, and operational accountability. They therefore require renewed discovery, impact analysis, design review, stakeholder approval, and regression planning.
Options A, B, and C are localized maintenance activities. Clarifying a runbook improves operational documentation but does not change the deployed system's structure. Adjusting an alert threshold is a controlled operational tuning activity, provided the change remains within established monitoring policy.
Editing customer-facing copy is similarly confined to the presentation or content layer. None of these changes inherently invalidates component contracts or architectural assumptions.
An architect should distinguish between iteration within an approved design and evidence that the design itself no longer supports the required outcome. Re-entering design for every minor adjustment creates unnecessary governance overhead; continuing local iteration after architectural assumptions have failed creates unmanaged technical and operational risk. Option D is the only signal that establishes a structural incompatibility.
Study Guide references/topics: Lifecycle feedback loops; design re-entry criteria; architectural significance; component responsibilities; interface contracts; controlled operational iteration.

NEW QUESTION: 46
You are assessing a Claude-based system whose dominant risk is silent quality drift on safety-relevant outputs after a model-version upgrade.
Which assessment activity most directly addresses this risk?
A. Disable adversarial evaluation entirely during model-version upgrade cycles to reduce evaluation cost, accepting that safety drift will go undetected until it appears in production.
B. Rotate adversarial inputs randomly so no two upgrades are scored on the same set.
C. Skip evaluation on each model-version upgrade and rely on user-submitted complaints to surface safety drift after the upgraded model has already served production traffic.
D. Maintain an adversarial evaluation set with version-attributed scoring so each upgrade is measured against the same set before promotion.
Answer: D (LEAVE A REPLY)
A stable adversarial regression set provides a controlled basis for detecting behavioral change between model versions. Version-attributed results show whether the proposed upgrade improves, preserves, or degrades performance on the same safety-relevant cases. The evaluation must run before promotion so a regression can block deployment rather than becoming a production incident. Randomly replacing the test inputs removes comparability and makes score changes difficult to attribute. Disabling evaluation or relying on complaints converts a preventable release-control failure into reactive incident handling. Anthropic characterizes regression evaluations as tests of whether an agent still handles tasks it previously completed and recommends maintaining them continuously to detect backsliding. Demystifying evals for AI agents

Valid CCAR-P Dumps shared by EduDump.com for Helping Passing CCAR-P Exam! EduDump.com now offer the **newest CCAR-P exam dumps**, the EduDump.com CCAR-P exam **questions have been updated** and **answers have been corrected** get the **newest** EduDump.com CCAR-P dumps with Test Engine here: https://www.edudump.com/exams/Anthropic/CCAR-P/premium/ (**130** Q&As Dumps, **35%OFF** **Special Discount Code:** **freecram**)

NEW QUESTION: 47

You are listing characteristics of strong architectural-decision communication.
Which two characteristics belong on the list? (Select two.)
Each correct answer presents a complete solution.
A. Ownership and review cadence for revisiting the decision when conditions or assumptions change.
B. Distribution to a standing review forum with mandatory attendance from senior engineering leaders.
C. Omission of rejected alternatives to keep the decision document focused and concise for readers.
D. Alternatives considered along with the criteria that were used to evaluate each alternative.
E. Use of a consistent template across the team to standardize the visual presentation of decisions.
Answer: (SHOW ANSWER)
Strong decision communication must explain both how the decision was reached and how it will remain governed. Option D provides the reasoning structure: the alternatives considered and the criteria used to compare them. This allows reviewers to verify that capability, cost, latency, security, maintainability, integration complexity, and business outcomes were evaluated consistently rather than selectively.
Option A establishes lifecycle accountability. Architectural decisions depend on assumptions about requirements, models, data, regulations, usage volume, and operational constraints. Naming an owner and review cadence ensures the decision is reconsidered when those assumptions materially change.
Option B imposes a meeting mechanism rather than describing decision quality. Mandatory attendance by senior leaders may add delay without improving the clarity or correctness of the record. Option C removes important evidence and increases the chance that rejected approaches will be proposed again without addressing the original objections. Option E can improve consistency and discoverability, but visual standardization alone does not establish substantive architectural communication.
A strong decision record should therefore be evidence-based, audience-accessible, traceable to requirements, explicit about alternatives and consequences, assigned to an accountable owner, and governed by defined review triggers.
Study Guide references/topics: Architectural-decision communication; alternatives analysis; evaluation criteria; decision ownership; review cadence; changing assumptions; lifecycle governance.

NEW QUESTION: 48
You are building an ethics review checklist for deployments supported by artificial intelligence (AI). Which two checks belong on the list? (Select two.) Each correct answer presents a complete solution.
A. Confirm that vendor licensing terms permit the planned production use of the model.
B. Confirm that high-impact decisions retain human accountability rather than being attributed to the model.
C. Restrict the ethics review to outputs that exceed a defined model-confidence threshold.
D. Confirm that latency and throughput targets are met across the supported user populations.
E. Verify that outputs do not rely on generalizations about people that the underlying data does not support.
Answer: (SHOW ANSWER)
High-impact AI-supported decisions must retain clearly assigned human accountability. Claude may provide recommendations, classifications, or generated content, but responsibility for consequential outcomes cannot be transferred to the model. An ethics review should also investigate whether outputs make unsupported generalizations about individuals or groups, because such behavior can produce discriminatory, misleading, or harmful outcomes. Therefore, B and E directly address ethical accountability and fairness. Licensing in option A is an important legal and procurement concern, but it is not the strongest ethics-focused control presented.
Latency and throughput in option D are operational performance requirements. Option C is inappropriate because ethics risks are not limited to outputs above a confidence threshold; confidence values may also be poorly calibrated and do not prove that an output is fair, accurate, or ethically acceptable.

NEW QUESTION: 49
A platform team operates a self-hosted multi-agent system on Kubernetes that orchestrates seven specialized agents for invoice processing. The team spends approximately 40 percent of engineering capacity on infrastructure maintenance, message bus reliability, and agent state recovery. The CFO has asked you to evaluate moving to managed agent infrastructure to reclaim engineering capacity. The security officer requires that all customer financial data remain within an approved network boundary.
Which factor should most heavily influence your recommendation?
A. Whether the current seven agents map cleanly to the patterns supported by managed agents.
B. Whether managed agents reduce per-invoice token costs across the existing processing volume.
C. Whether managed agents support the current bus topology used by the platform team.

D. Whether managed agent data handling satisfies the network boundary required by security.
Answer: (SHOW ANSWER)
The network-boundary requirement is a mandatory architectural constraint; therefore, Option D must be resolved before evaluating operational savings. Managed infrastructure can reduce responsibility for agent loops, sandboxes, session state, recovery, and orchestration, but those advantages are immaterial if financial data leaves the approved boundary. Anthropic describes Claude Managed Agents as a hosted agent harness and provides configurable data-residency controls. The architect must still validate inference location, session persistence, encryption, subprocessors, credential handling, auditability, and deletion behavior against the organization's exact security policy. Options A, B, and C affect migration effort or economics, but they cannot override a non-negotiable data-governance requirement.
Study Guide references/topics: Architectural constraints; managed versus self-hosted agents; data boundaries; residency; security-gated decision-making.

NEW QUESTION: 50
You are reviewing a customer-support agent's configuration. Each candidate tool falls into one of four categories: (1) required to complete defined tasks, (2) frequently used and reduces hand-offs, (3) occasionally useful for unrelated work, (4) speculative future utility.
Which categories should typically remain in the agent configuration?
Freecram.net
A. Only category 3, because occasionally useful tools for unrelated work provide broader coverage and should take priority over tools required for the agent's defined tasks.
B. Only category 4, because speculative future-utility tools provide the most flexibility and should be configured even when no defined task currently requires them.
C. Categories 1 and 2 only, because they map to defined tasks and the agent's regular hand-offs.
D. All four categories, because broader tool access is categorically better for agent performance regardless of whether the tools map to defined tasks or regular hand-offs.
Answer: (SHOW ANSWER)
Category 1 tools are necessary for the agent to complete its approved responsibilities and must remain.
Category 2 tools are also justified when they are regularly used and eliminate predictable hand-offs without expanding the agent beyond its defined operating role.
Categories 3 and 4 create capability bloat. A tool that is occasionally useful only for unrelated work does not belong to this agent's responsibility boundary. A speculative tool has no validated requirement and adds attack surface, context consumption, authorization complexity, evaluation burden, and operational dependencies without demonstrated value.
Tool inclusion should be based on task traceability: every configured capability should map to a documented user journey, responsibility, permission scope, and evaluation case. Anthropic recommends least privilege so that a successful injection or model error can cause minimal damage. Mitigate Jailbreaks and Prompt Injections The architect should periodically review tool-call telemetry and remove unused capabilities. New tools can be added when a validated workflow requires them, after security review and evaluation. The objective is not the smallest possible catalog regardless of usefulness; it is the smallest catalog that reliably completes the agent's approved tasks.
Study Guide references/topics: Capability-bloat analysis; tool necessity; hand-off reduction; least privilege; attack-surface management; tool lifecycle review.

NEW QUESTION: 51
A document analysis service processes legal filings averaging 80,000 tokens each. Each filing is queried by attorneys an average of 14 times during a case. The current architecture sends the full filing on every query.
The CFO has asked you to reduce per-query costs while preserving response quality. The security officer requires that filing contents not be stored outside Fabrikam ' s tenancy.
Which optimization approach should you recommend?
A. Summarize each filing once at intake and run all subsequent queries against the summary.
B. Cache the filing as the prompt prefix for reuse across the 14 queries per case.
C. Index filings in a vector store and retrieve only the relevant passages per query.
D. Move the workload to a smaller Claude model to reduce the per-token cost paid.
Answer: (SHOW ANSWER)
The 80,000-token filing is a large, stable prefix reused across fourteen requests, making prompt caching the strongest optimization. The first request creates the cache; subsequent identical-prefix requests use lower- priced cache reads while retaining the complete filing, preserving the evidence available to Claude. Anthropic documents cache reads at a fraction of standard input cost and confirms that caches are isolated by organization and, on supported platforms, by workspace. Cached representations remain in memory rather than persistent storage. Options A and C may omit context through summarization or retrieval and therefore require additional quality evaluation. Option D lowers unit cost but may reduce reasoning quality. The design must preserve prefix identity, choose an appropriate TTL, and verify the tenancy configuration.
Study Guide references/topics: Prompt caching; cache isolation; stable prefixes; long-document analysis; cost optimization.

NEW QUESTION: 52
You are designing a human-in-the-loop validation workflow for a new Claude-based deployment and must complete the design steps before piloting the workflow.
Which two steps must be completed BEFORE piloting the workflow with a representative subset of traffic?
(Select two.)
Each correct answer presents part of the solution.
A. Define the sampling strategy and the escalation criteria at each oversight point in the pipeline.
B. Iterate the workflow design based on observed pilot findings before broader rollout to production.
C. Onboard the reviewer pool with role-specific training on the check criteria and escalation procedures.
D. Document the workflow with check criteria, escalation paths, and service-level agreements (SLAs) for each step.
E. Identify the decision points in the pipeline that require human oversight by impact and reversibility.
Answer: (SHOW ANSWER)
The workflow must first identify where human judgment is necessary. Option E applies risk-based oversight by examining each decision's impact and reversibility. High-impact, irreversible, externally binding, privacy- sensitive, or legally consequential actions require stronger review than low-risk draft generation or easily reversible recommendations.
After locating those control points, Option A defines how work reaches reviewers. The sampling strategy should specify whether review is universal, random, risk-triggered, confidence-based, or stratified. Escalation criteria should cover policy violations, uncertain outputs, conflicting evidence, sensitive data, failed validation, and cases exceeding the reviewer's authority.
Anthropic recommends human confirmation for decisions with meaningful real-world consequences and actions requiring affirmative consent. Computer Use Security Considerations Option B necessarily follows the pilot because findings must exist before design iteration. Reviewer onboarding and operational documentation in Options C and D are important readiness activities, but they follow the core design decisions being tested by this item: where oversight is required and how cases are sampled and escalated.
The completed pilot plan should subsequently include trained reviewers, documented procedures, queue capacity, SLAs, audit evidence, and measured inter-reviewer agreement.
Study Guide references/topics: Human-in-the-loop design; impact and reversibility; oversight points; sampling; escalation criteria; consequential decisions.

NEW QUESTION: 53
You are reviewing an integration specification for security gaps.
Which two findings constitute valid security gaps in the specification? (Select two.) Each correct answer presents a complete solution.
A. Tool calls execute server-side under a least-privilege service principal scoped to the requested action.
B. Service credentials are placed in the prompt context, where they can leak into logs and traces.
C. Role-based access control is enforced only at the response-rendering layer after the model accesses restricted data.
D. Per-user OAuth tokens are exchanged with scope-restricted permissions and refreshed within the active session.
E. Tool inputs and outputs are encrypted in transit using transport-layer security between services.
Answer: (SHOW ANSWER)
Credentials must never be supplied as prompt content. Prompts can be retained in traces, debugging output, caches, or logs, and model output may inadvertently expose included secrets. Credentials should instead be resolved server-side through an approved secret store or delegated authentication mechanism. Option C is also a serious gap because post-response RBAC occurs after unauthorized data has already entered model context.
Authorization must be enforced before retrieval and prompt construction. Options A, D, and E describe appropriate controls: least-privilege service identities, per-user scoped OAuth, and encrypted transport.
Anthropic's prompt-leak guidance recommends excluding unnecessary sensitive information, while Claude Code security relies on permission-based enforcement rather than natural-language instructions alone. Reduce prompt leak

NEW QUESTION: 54
You are designing the prompt for a ticket-triage classifier with thirty well-defined categories and clear category descriptions in the prompt.
Which technique is most appropriate as the starting point?
A. A chain-of-thought prompt for a routing decision that does not require multi-step reasoning.
B. A zero-shot prompt that specifies the categories with their descriptions and the required output format.

C. A prompt that demands the model invent new categories when the supplied ones do not fit.
D. A prompt with no category descriptions, expecting the model to infer the category set.
Answer: (SHOW ANSWER)
A clearly defined closed-set classification task should begin with a concise zero-shot prompt containing the permitted categories, unambiguous category descriptions, and the required output structure. This provides a simple, inexpensive baseline that can be measured against a representative evaluation set. Chain-of-thought prompting adds unnecessary latency and token usage when the routing decision does not require substantive multi-step reasoning. Permitting Claude to invent categories would violate the closed-set contract and complicate downstream automation. Omitting descriptions would increase ambiguity among potentially overlapping labels. If evaluation later reveals persistent confusion between particular categories, the architect can introduce targeted examples or refine the definitions. Prompt complexity should be added in response to measured failure patterns, not assumed to be necessary before establishing baseline performance. Anthropic:
Prompt engineering overview

NEW QUESTION: 55
You are selecting a protocol for a single low-latency stateless tool call from a Claude-based assistant to an internal pricing service that already exposes a stable HTTP API.
Which integration mechanism is the most appropriate?
A. A direct API call to the existing endpoint with the appropriate scoped credentials.
B. A long-lived stateful session protocol for a stateless single-call interaction.
C. A bespoke streaming protocol layered over an unrelated asynchronous message bus.
D. An agent-to-agent handoff that introduces another Claude-based agent in front of the pricing service.
Answer: (SHOW ANSWER)
A direct call to the existing stable HTTP endpoint is the simplest mechanism that satisfies the stated requirements. The interaction is stateless, requires only one tool invocation, and has a strict latency objective.
Adding session management, message-bus translation, or another model-mediated agent would introduce unnecessary network hops, operational dependencies, failure modes, and processing latency.
The assistant's tool implementation should validate input parameters, authenticate through a server-side credential mechanism, apply narrowly scoped authorization, set explicit timeouts, and validate the pricing response before returning it to Claude. Credentials must never be supplied by or exposed to the model.
Appropriate logging should capture request attribution, endpoint outcome, and timing without unnecessarily recording sensitive data.
Anthropic describes the Claude API as a RESTful interface and supports custom tools in which the application executes the requested function or API operation. Claude API Overview, Tool Use Overview MCP could be justified when standardized discovery or reuse across clients is required, but the scenario already supplies a stable API and does not establish that additional requirement.
Study Guide references/topics: Protocol selection; direct API integration; stateless calls; latency minimization; scoped credentials; avoiding unnecessary agentic complexity.

NEW QUESTION: 56
You are producing an architecture guide for a new deployment and must complete the planning steps before drafting each section.
Which two steps must be completed BEFORE drafting each section of the guide? (Select two.) Each correct answer presents part of the solution.
A. Validate the guide with the implementation team and incorporate corrections.
B. Identify the audience and the questions the guide must answer for that audience.
C. Establish the document under version control with a defined review cadence and approver list.
D. Outline the guide sections covering the overview, components, contracts, flows, runbooks, and limitations.
E. Translate the guide into the supported regional languages for the candidate population.
Answer: B,D (LEAVE A REPLY)

NEW QUESTION: 57
You are reviewing a peer's end-to-end design for a Claude-based platform expected to scale to thousands of concurrent users.
For each statement, indicate Yes if it reflects sound architectural practice. Otherwise, select No.

Answer Area

Statements:	Select Yes or No:
Authentication and authorization run before retrieval so retrieval can filter by identity.	Yes / No
An asynchronous queue absorbs bursty traffic between intake and the model invocation layer.	Yes / No
Tax computation is encoded directly in the system prompt rather than in code.	Yes / No
Conversation logs include unredacted government identifiers to maximize signal for tuning.	Yes / No

freecram.net

Answer:

Answer Area

Statements:	Select Yes or No:
Authentication and authorization run before retrieval so retrieval can filter by identity.	Yes / No
An asynchronous queue absorbs bursty traffic between intake and the model invocation layer.	Yes / No
Tax computation is encoded directly in the system prompt rather than in code.	Yes / No
Conversation logs include unredacted government identifiers to maximize signal for tuning.	Yes / No

Explanation:
Authentication and authorization run before retrieval so retrieval can filter by identity - Yes An asynchronous queue absorbs bursty traffic between intake and the model-invocation layer - Yes Tax computation is encoded directly in the system prompt rather than in code - No Conversation logs include unredacted government identifiers to maximize tuning signal - No Authentication and authorization must precede retrieval so unauthorized information never enters model context. An asynchronous queue is appropriate for burst absorption, backpressure, controlled concurrency, and retry management at scale. Deterministic tax computation should be implemented in validated code or a controlled calculation tool, not delegated to probabilistic prompt interpretation. Including unredacted government identifiers in conversation logs violates data-minimization principles and creates unnecessary privacy, security, and regulatory exposure. Sensitive fields should be removed, tokenized, or redacted before logging or model use unless specifically required and authorized. The resulting design separates deterministic computation, access control, scalable orchestration, and language reasoning into appropriate architectural layers.

NEW QUESTION: 58
A Claude architect needs to ensure that a security-hardening flag cannot be disabled by any individual engineer after it is set.
Which configuration scope correctly enforces this requirement?
A. User scope (~/.claude/settings.json) on each engineer's machine
B. Environment variable defined in the CI/CD pipeline only
C. Managed configuration applied centrally and marked as non-overridable
D. Project scope (.claude/settings.json) committed to the repository
Answer: (SHOW ANSWER)

Managed configuration is designed for organization-wide security and compliance controls that individual users and repositories must not override. Applying the hardening flag centrally establishes an administrative policy boundary and prevents engineers from disabling the control through user, project, or local settings.
Anthropic identifies managed scope as the appropriate location for security policies, non-overridable compliance requirements, and standardized configurations deployed by IT or DevOps. Managed values ordinarily take precedence over command-line arguments and every user-controlled settings scope. Claude Code Settings Option A depends on every engineer maintaining the setting and allows the user to edit or delete it. Option B affects only pipeline execution and does not protect local, interactive, IDE, or other execution paths. Option D standardizes the setting within the repository but does not make it tamper-resistant; contributors with repository write access could alter the file, use higher-precedence local settings where permitted, or operate outside the repository configuration.
The implementation should also verify active policy delivery, monitor configuration-change events, and test that startup or execution fails safely if the managed setting is absent or invalid. Central definition without enforcement and verification would not fully satisfy the requirement.
Study Guide references/topics: Managed settings; non-overridable controls; enterprise hardening; policy enforcement; configuration governance; defense against local override.

NEW QUESTION: 59
An operations engineer reports that a Claude-based pipeline began returning malformed JSON responses after a scheduled maintenance window, causing downstream processing failures.
Which two investigative steps most directly isolate the root cause? (Select two.)
A. Clear the prompt cache and resubmit all pending requests to eliminate stale cached prefixes.
B. Compare the current system prompt and output schema configuration against the last known good version from before the maintenance window.
C. Switch to a different Claude model tier to rule out provider-side changes as a contributing factor.
D. Increase the max_tokens limit to determine whether output truncation is causing incomplete JSON structures.
E. Replay a set of premaintenance requests against the current configuration and inspect the raw model output before downstream parsing.
Answer: (SHOW ANSWER)
The strongest investigation combines configuration comparison with controlled replay. Comparing the current system prompt and output-schema settings against the last known good version identifies maintenance-related configuration drift. Relevant differences include removed formatting instructions, altered schema requirements, incompatible parser expectations, changed stop sequences, or a disabled structured-output mechanism.
Replaying known premaintenance requests against the current configuration creates a controlled test set.
Inspecting the raw response before parsing then establishes whether malformed JSON originates in model generation or is introduced by streaming assembly, transport handling, post-processing, or the downstream parser. If the raw response is valid but the pipeline rejects it, the fault lies after generation. If the raw response is malformed, the prompt, schema enforcement, stop reason, and request configuration become the primary investigation targets.
Clearing the prompt cache assumes cache corruption without evidence. Changing model tiers introduces another variable and weakens causal attribution. Raising `max_tokens` is appropriate only when raw responses show truncation and the returned stop reason indicates the token limit was reached; it is not the most direct initial test.
Study Guide references/topics: [Increasing output consistency](https://docs.anthropic.com/en/docs/test-and- evaluate/strengthen-guardrails/increase-consistency); [agent evaluation methodology] (https://www.anthropic.com/engineering/demystifying-evals-for-ai-agents); regression replay; last-known-good comparison; raw- output inspection.

NEW QUESTION: 60
A Claude architect is auditing configuration scope assignments.
Which two statements correctly identify an appropriate use of user-scope configuration versus other scopes?
(Select two.)
Freecram.net
A. Persisting personal editor theme preferences that follow an engineer across projects.
B. Saving a preferred Claude response language that applies to all repositories the engineer uses.
C. Enforcing a company-wide policy that disables a feature for all engineers.
D. Defining MCP server endpoints shared by all contributors to a specific repository.
E. Storing API authentication keys so they are not committed to version control.
Answer: (SHOW ANSWER)

User scope applies to one engineer across every project and is therefore appropriate for portable personal preferences. Option A fits this definition because an interface theme belongs to the individual rather than to a repository or organization. Option B also fits because Claude Code provides a language setting for the preferred response language, and placing that preference in user scope makes it apply across repositories.
Option C requires managed configuration because an organization-wide security control must be centrally deployed and resistant to individual override. Option D belongs in project scope because repository-specific MCP endpoints must be shared consistently with that repository's contributors.
Option E describes credential handling rather than merely a preference assignment. Credentials must not be placed in a repository, but avoiding version control alone is insufficient. They must be stored through a supported secure authentication mechanism, protected environment configuration, operating-system credential facility, or approved secrets manager-not inserted as plaintext into an ordinary settings file.
Anthropic defines user scope as affecting one user across all projects, project scope as shared repository configuration, and managed scope as the location for enforceable organizational controls. It also explicitly lists themes and other personal settings as user-scope use cases. Claude Code Configuration Scopes Study Guide references/topics: User scope; project scope; managed configuration; personal preferences; repository-shared MCP configuration; credential protection.

NEW QUESTION: 61
You are choosing the level of detail for an implementation guide. The audience is a delivery team that will build the deployment.
Which guidance composition best serves them?
A. Component responsibilities, contracts between components, sequence diagrams of the dominant flows, configuration parameters with defaults, and operational runbooks.
B. Component responsibilities and interface contracts only, without sequence diagrams of the dominant flows, configuration parameters with defaults, or runbooks to guide operational tasks.
C. An architecture overview and sequence diagrams for the dominant flows, without interface contracts, configuration-parameter tables, or operational runbooks for the delivery team to follow.
D. An architecture overview and a list of known limitations, without component-level diagrams, interface contracts, configuration parameters, or operational runbooks to support implementation.
Answer: (SHOW ANSWER)
A delivery team needs implementation-level direction rather than an executive architecture summary.
Component responsibilities establish ownership and boundaries, while interface contracts define schemas, accepted inputs, outputs, errors, and behavioral expectations. Sequence diagrams show how components interact across dominant and failure flows. Configuration tables with defaults prevent inconsistent assumptions between environments. Operational runbooks ensure that deployment, recovery, monitoring, and routine maintenance are addressed during implementation rather than improvised after release. Options B, C, and D omit one or more elements required to translate architectural intent into a buildable and supportable system. The guide should enable normal delivery work without continuous dependence on undocumented knowledge held by the architect.
Study Guide references/topics: Implementation documentation; component contracts; sequence diagrams; configuration management; operational readiness.

Valid CCAR-P Dumps shared by EduDump.com for Helping Passing CCAR-P Exam! EduDump.com now offer the **newest CCAR-P exam dumps**, the EduDump.com CCAR-P exam **questions have been updated** and **answers have been corrected** get the **newest** EduDump.com CCAR-P dumps with Test Engine here: https://www.edudump.com/exams/Anthropic/CCAR-P/premium/ (**130** Q&As Dumps, **35%OFF** **Special Discount Code: freecram**)

Valid CCAR-P Dumps shared by EduDump.com for Helping Passing CCAR-P Exam! EduDump.com now offer the **newest CCAR-P exam dumps**, the EduDump.com CCAR-P exam **questions have been updated** and **answers have been corrected** get the **newest** EduDump.com CCAR-P dumps with Test Engine here: https://www.edudump.com/exams/Anthropic/CCAR-P/premium/ (**130** Q&As Dumps, **35%OFF** **Special Discount Code: freecram**)