

## PRACTICE QUESTIONS

Anthropic

# CCAR-P

Claude Certified Architect - Professional

SHARED BY [u/Swift-Kettle-4577](#) on aiexamcerts

---

EDITION Free Set 1

---

QUESTIONS 10

---

SHARED October 18, 2026

---

EXAM Anthropic CCAR-P, Claude Certified Architect - Professional

---

Shared by [u/Swift-Kettle-4577](#) on aiexamcerts

10 practice questions for the Anthropic CCAR-P exam, Claude Certified Architect - Professional, each followed by its answer and a short explanation. More free sets and the full practice exam:

[aiexamcerts.com/anthropic/ccar-p](https://aiexamcerts.com/anthropic/ccar-p). Spotted a mistake? Email [support@aiexamcerts.com](mailto:support@aiexamcerts.com) with the exam code CCAR-P and the question number.

1. Which step belongs in the solution design phase rather than in the handoff phase of a Claude project?
  - A Choosing between a workflow and an agent for the planned feature, based on the steps it requires
  - B Writing the runbook that on-call engineers use during an incident, and listing each on-call contact with their phone number
  - C Training the support team on how to read the new dashboard
  - D Signing off the final acceptance test with the business owner

**Answer: A**

Choosing the pattern is a design decision made before implementation, so it belongs to solution design. Runbooks, team training and acceptance sign-off all come later, in the handoff and monitoring phases, though runbooks sound like a technical design artifact and are the tempting confusion.

2. A design that routes simple requests to a small model and complex ones to a larger model is an example of model \_\_\_\_\_.

**Answer: routing** (also accepted: model routing)

Model routing sends each request to the cheapest model that can handle it, trading cost and latency against quality. Caching is tempting here, but it reuses earlier responses or prompt prefixes rather than choosing a model for each request.

3. Which metric best reflects the cost pillar of a production Claude deployment?
  - A Average number of sentences in each model response, measured across every product line and reporting period
  - B Number of prompt templates stored in the repository
  - C Count of distinct system prompts across all environments
  - D Cost per resolved ticket, computed from token usage and infrastructure spend

**Answer: D**

Cost per resolved ticket ties token consumption and infrastructure spend to a business outcome, which is what the cost pillar measures. Response length is tempting because token count drives cost, but on its own it is not a cost-per-outcome measure.

4. An architect must document the decision to use an agent for a planning feature. What should the record include to be useful later?
- A A screenshot of the prototype and the name of the developer who built it, along with the list of files changed in the repository
  - B The model's full reasoning text pasted from one successful run
  - C The options considered, the chosen pattern, the constraints that drove it, and the conditions that would trigger a revisit
  - D A list of every library installed in the development environment

**Answer: C**

A decision record that captures alternatives, constraints and revisit triggers lets later teams judge whether the choice still holds. A prototype screenshot is tempting as evidence, but it records what was built, not why it was chosen.

5. Should a reference architecture show the data flow from input to output and mark where evaluation signals are captured, so the feedback loop can be designed and tested?
- A True
  - B False

**Answer: A (True)**

This is true. Showing data flow and instrumentation points makes the feedback loop concrete and testable. A diagram of component boxes without data flow is the tempting shortcut, but it hides where the system should be measured.

6. Which reason best justifies adding a human approval step before an agent issues a refund?
- A Models cannot reliably call tools, so a person must press a button before each tool call runs in the workflow
  - B The action is irreversible and costly, so a human checks the decision before it takes effect
  - C Human approval makes the model faster, because the agent no longer waits for the tool
  - D Every AI output must be read by a person before it is stored, as a general rule

**Answer: B**

Irreversible, high-impact actions justify a human checkpoint before execution, which limits the cost of an error. Option D is tempting because it sounds like a compliance rule, but human oversight is required for specific high-risk uses, not as a blanket rule for every output.

7. Which two statements describe sound criteria for choosing between a single agent and a multi-agent design? (Choose two.)
- A Split agents when subtasks need different tools or context that would overload one agent's window
  - B Add agents whenever a task has more than three steps, regardless of how the steps depend on each other
  - C Use one agent when the subtasks share the same context and must stay consistent with each other
  - D Give every specialist agent full write access so they can coordinate freely
  - E Prefer multi-agent designs whenever the team wants to show an advanced architecture

**Answer: A, C**

Options A and C state the real trade-offs: separate agents help when subtasks need separate tools or context, and one agent keeps shared context consistent. Option B is tempting because step count feels like complexity, but step count alone does not justify agents, D grants unnecessary permissions, and E is not a technical reason.

8. Model Context Protocol servers expose capabilities to clients as tools, resources and \_\_\_\_\_.

**Answer: prompts** (also accepted: prompt templates)

The three MCP server primitives are tools (model-invoked functions), resources (data the client can read by URI) and prompts (reusable templates a user can select). Sampling is tempting here, but it is a client capability that lets servers request model completions, not a server primitive.

9. A coding assistant has read access to a repository, write access to a branch, and a tool that deploys to production. It is only used for code review. Which change best applies least privilege?
- A Log every deploy call so the team can review misuse after the fact
  - B Keep the deploy tool but ask the model in the system prompt to avoid calling it unless asked
  - C Remove the deploy tool and the write access, leaving only read and comment capabilities
  - D Use a more capable model so it refuses unsafe deploy requests reliably

**Answer: C**

Least privilege removes capabilities the role does not need, which eliminates the risk instead of watching for it. Logging is a detective control, and an instruction to avoid the tool is a request the model can still break, which makes it the tempting choice.

10. Which connection mechanism is most appropriate for exposing an internal inventory system to several different AI clients that each need the same tools?
- A An MCP server that exposes the inventory tools and resources once, for any compliant client
  - B A custom integration written separately inside each client application
  - C Pasting inventory exports into each client's system prompt on a nightly schedule, refreshed by a nightly cron job on every application server
  - D Giving every client direct database credentials so each can write its own queries

**Answer: A**

An MCP server implements the tools and resources once, and any MCP-compatible client can discover and call them. Separate per-client integrations are tempting for a quick start, but they duplicate effort and drift apart over time.