



Certified Data Governance Specialist (CDGS)

Ten Sample Questions

These ten questions are written at the difficulty of the live exam. Six carry a figure, a data table, a configuration or a code listing; four are plain scenario questions, as on the live paper. Every one is a scenario a working data governance specialist would meet, and none of them appears in the live question bank.

The live exam is multistage adaptive: how you score on one stage selects the difficulty of the next, so no two candidates sit the same paper. It is a performance-based exam: each form can carry up to 5 to 12 performance-based questions alongside the multiple-choice questions, which means you build, sequence or complete something rather than only choosing from a list. Answers and full explanations follow each question here; the live exam does not disclose item-level answers.

Question 1

Domain 1: Data Governance Foundations

DATA

Emma Wilson matches each governance role to its responsibility.

Role	Responsibility
Data owner	(to choose)
Data steward	(to choose)
Data custodian	(to choose)
Governance council	(to choose)

Table 1.1 Roles and their responsibilities.

Which set of responsibilities is correct?

- ☐ A Handles storage and controls, owns the domain and sets policy, minds daily quality, and steers direction, so the steward approves all access requests.
- ☐ B Owns the domain and sets policy, minds daily quality and definitions, handles storage and controls, and steers direction, matching each role to its job.
- ☐ C Steers direction, handles the storage, owns the domain and minds daily quality, so the custodian sets policy and approves access across each data domain.
- ☐ D Minds daily quality, steers direction, handles storage and owns the domain, so the council performs the steward daily definition work across the estate.

Correct answer: B

A data owner is accountable for a domain and sets policy, a data steward minds daily quality and definitions, a data custodian handles technical storage and controls, and the governance council steers direction. Only option B pairs each role with its actual responsibility.

Question 2

Domain 1: Data Governance Foundations

CONFIG

James Carter reviews a new governance program described below.

```
policy owner      : unassigned
decision rights   : undefined
escalation path   : none
data domains      : unmapped
```

Listing 2.1 The current state.

What is the soundest first step?

- ☐ A Buy a new catalog tool right away, since purchasing technology first is the only reliable way to make any governance program succeed at an organization.
- ☐ B Delete a large share of the existing data now, on the assumption that a smaller estate removes the need to define any decision rights or ownership at all.
- ☐ C Assign owners and stewards, define the decision rights and escalation, and map the domains, so clear accountability exists well before any tool is chosen.
- ☐ D Raise the storage budget for every team at once, on the assumption that more capacity alone resolves unassigned ownership and the undefined decision rights.

Correct answer: C

A governance program needs accountability before tooling, so the first step is to assign owners and stewards, define decision rights and escalation, and map the data domains. Buying a tool, deleting data, or raising budgets does none of that and leaves the program without ownership.

Question 3

Domain 2: Data Quality Management

DATA

Olivia Davis maps each data problem to the quality dimension it violates.

Problem	Dimension
Same customer stored twice	(to choose)
Postal code field left blank	(to choose)
Birth date falls in the future	(to choose)
Address differs across systems	(to choose)

Table 3.1 Problems and dimensions.

Which set of dimensions is correct?

- ☐ A Uniqueness, completeness, validity and consistency, naming the duplicate, the blank field, the impossible value and the cross-system mismatch in that order.

- B** Completeness, uniqueness, consistency and validity, so completeness names the duplicated customer record rather than the field that was left blank by a user.
- C** Validity, consistency, uniqueness and completeness, so validity names the duplicated customer record rather than the birth date value that falls in the future.
- D** Consistency, validity, completeness and uniqueness, so consistency names the field left blank rather than the address that differs from system to system in use.

Correct answer: A

A duplicate customer violates uniqueness, a blank required field violates completeness, an impossible birth date violates validity, and an address that differs across systems violates consistency. Only option A pairs each problem with the dimension it actually breaks.

Question 4

Domain 3: Metadata, Lineage and Cataloging

CONFIG

Liam Anderson lists the stages a record passes through, shown out of order.

- A. Curated data published to the catalog
- B. Source system captures the raw record
- C. Transformation applies business rules
- D. Ingestion loads raw data into the store
- E. Quality checks validate the loaded data

Listing 4.1 Stages to order.

What is the correct order of these stages?

- A** B, then D, then C, then A, then E, validating the loaded data only after it has already been published to the catalog for consumers downstream to use.
- B** D, then B, then E, then A, then C, capturing the source after ingestion loaded it and applying the business rules after publication to the catalog itself.
- C** C, then A, then B, then D, then E, transforming and publishing the data before the source system has even captured or loaded the raw record at all yet.
- D** B, then D, then E, then C, then A, capturing the source, ingesting it, validating it, applying the business rules, then publishing to the catalog in turn.

Correct answer: D

Lineage runs from capture at the source (B), ingestion into the store (D), quality validation of the loaded data (E), transformation that applies business rules (C), and publication to the catalog (A). The other orders validate, transform, or publish before the record exists or is loaded.

Question 5

Domain 4: Master and Reference Data Management

DATA

Ava Thompson matches each master data term to its meaning.

Term	Meaning
Golden record	(to choose)
Survivorship rule	(to choose)
Match and merge	(to choose)
Reference data	(to choose)

Table 5.1 Terms and their meanings.

Which set of meanings is correct?

- A** Decides which value wins, the single trusted version, links the duplicates, and shared code sets, so the golden record is what picks the winning value here.
- B** The single trusted version, decides which value wins, links duplicate records, and shared code sets, matching each term to what it actually means in practice.
- C** Links the duplicate records, shared code sets, the single trusted version and a winner, so match and merge names the shared standard code sets used widely.
- D** Shared code sets, links the duplicates, picks the winning value and trusted version, so reference data names the single trusted version of the entity itself.

Correct answer: B

A golden record is the single trusted version of an entity, a survivorship rule decides which value wins when sources conflict, match and merge links duplicate records into one, and reference data is shared standard code sets. Only option B pairs each term with its meaning.

Question 6

Domain 5: Data Privacy, Security and Compliance

DATA

Data handling for a sensitive dataset is measured as shown below.

Observation	Result
Sensitive fields classified	No
Access limited by need	No
Data encrypted at rest	No
Retention schedule applied	No

Table 6.1 How the data is protected today.

What is the core weakness, and what should be fixed?

- A** Sensitive data is ungoverned; classify the fields by sensitivity, limit access by need, encrypt the data at rest, and apply a retention schedule per class.
- B** Nothing is wrong, because data that is useful to the business does not need classification, access limits, encryption at rest, or any retention schedule set.
- C** The storage is too small, so buying more capacity alone would classify the sensitive fields and limit the access without any protection policy written first.

- D** The team is too small, so hiring more staff alone would encrypt the data and apply retention without ever classifying a single field by its sensitivity first.

Correct answer: A

Sensitive data with no classification, access limits, encryption, or retention is effectively ungoverned. The fix is to classify fields by sensitivity, limit access by need, encrypt at rest, and apply a retention schedule to each class. More storage or staff without policy fixes none of that.

Question 7 Domain 2: Data Quality Management

A team fixes the same bad values by hand in each report every month, yet the errors keep returning. What is the soundest practice?

- A** Keep correcting the values by hand in every report each month, because manual cleanup at the end is the fastest way to keep the published numbers clean.
- B** Delete every record that contains a bad value outright, on the assumption that removing the data entirely is simpler than correcting it at its true source.
- C** Fix the root cause at the source, add validation at data entry, and profile the data regularly, so the errors are prevented rather than patched downstream.
- D** Raise the report refresh rate sharply, on the assumption that generating the same numbers more often will make the underlying bad values correct themselves.

Correct answer: C

Recurring errors point to an upstream cause, so the fix is to address the root cause at the source, validate data at entry, and profile regularly to catch drift. Manual cleanup, deleting records, or refreshing more often treats the symptom and leaves the defect in place.

Question 8 Domain 6: Governance Operations and Metrics

A governance program reports only how many policies it has written. What best measures whether the program is actually working?

- A** Count the policies written, because the sheer number of documents produced is the clearest proof that the governance program is working exactly as intended.
- B** Count the meetings held each quarter, since a busy governance calendar is the surest sign that the data is being governed well across the whole organization.
- C** Count the tools purchased so far, on the assumption that owning more governance software is what shows the program is delivering real value to the business.
- D** Track outcomes such as steward coverage, policy adherence, issue resolution time and catalog adoption, rather than merely how many policies have been written.

Correct answer: D

Policy counts measure activity, not results. A program is working when outcome metrics improve, such as steward coverage, policy adherence, issue resolution time, and catalog adoption. Counting policies, meetings, or tools mistakes effort and spend for governed data.

Question 9

Domain 4: Master and Reference Data Management

Two systems each keep their own customer master, and the two often disagree about the same customer. What is the soundest approach?

- (A) Let each system keep its own customer master and reconcile the two by hand whenever a report happens to need a single consistent view of a given customer.
- (B) Establish a golden record with match, merge and survivorship rules, so one trusted version of each customer is mastered and shared consistently across systems.
- (C) Pick one of the systems at random to be the correct one, on the assumption that any single source is just as good as another for mastering the customer data.
- (D) Copy both records into a third system unchanged, on the assumption that simply storing every version together in one place resolves the conflict on its own.

Correct answer: B

Conflicting masters are resolved by mastering one golden record, using match and merge to link duplicates and survivorship rules to pick winning values, then sharing it across systems. Manual reconciliation, a random pick, or a third copy leaves the conflict unresolved.

Question 10

Domain 6: Governance Operations and Metrics

New reports flag data issues each week, but no one is assigned to fix them and the backlog keeps growing. What best fixes this?

- (A) Assign each issue to a steward with a due date and track it through to closure, so the reported problems are actually resolved rather than logged and forgotten.
- (B) Keep logging the issues without assigning any of them, because a complete record of the problems is far more valuable to the program than resolving any of them.
- (C) Close every open issue automatically after one week has passed, on the assumption that clearing the list is much the same as fixing the underlying problems it holds.
- (D) Route all of the issues to a single senior leader, on the assumption that one very busy executive can personally resolve every data problem that gets reported.

Correct answer: A

Issues resolve only when someone owns them, so each should be assigned to a steward with a due date and tracked to closure. Logging without assigning, auto-closing, or funneling everything to one leader lets the backlog grow while the underlying problems remain unfixed.