

Cloudera CDP-6001

CDP Machine Learning Engineer Certification Exam

For More Information – Visit link below:

<https://www.examsempire.com/>

Product Version

1. Up to Date products, reliable and verified.
2. Questions and Answers in PDF Format.



<https://examsempire.com/>

Visit us at: <https://www.examsempire.com/cdp-6001>

Latest Version: 6.1

Question: 1

When working with window functions that involve aggregation, how do you handle potential null values that could arise within the window frame?

- A. Utilize functions like coalesce or when/otherwise to provide default values.
- B. Employ filtering before calculating the window aggregation.
- C. Use aggregation functions that inherently ignore nulls (e.g., sum, avg)
- D. All of the above

Answer: D

Explanation:

Nulls need careful attention: Coalesce/Imputation: Replace nulls with meaningful defaults.

Filtering: Exclude problematic rows if appropriate.

Null-aware aggregations: Some functions handle nulls gracefully.

Question: 2

You have a DataFrame with user website visits (userID, sessionID, eventTimestamp, page). You need to calculate the time difference between the first and last event within each session for every user. How would you achieve this?

- A.
Use `min(eventTimestamp)` and `max(eventTimestamp)` with a window partitioned by 'userID' and 'sessionID'.
- B.
Create a window partitioned by 'userID' and ordered by 'eventTimestamp', then use lead and lag functions.
- C.
Employ `first_value()` and `last_value()` over a window partitioned by 'userID' and 'sessionID' ordered by 'eventTimestamp'.
- D.
Write a UDF to calculate the difference because window functions cannot directly handle this.

Answer: C

Explanation:

Partitioning: We need to calculate this per user and per session.

`first_value()` and `last_value()`: These functions let us grab the necessary timestamps from the beginning and end of each window.

Question: 3

You are deploying a complex machine learning pipeline in CML. Which of the following actions would be the MOST suitable way to break down the pipeline for improved modularity and collaboration?

- A. Create multiple interconnected workspaces, each handling a pipeline stage.
- B. Define the entire pipeline within a single session.
- C. Utilize separate virtual machines for each pipeline step.
- D. Write the entire pipeline as a monolithic Python script.

Answer: A

Explanation:

Workspaces promote modular design. Creating interconnected workspaces for each stage allows developers to focus on specific components, facilitates versioning, encourages collaboration, and enables easier troubleshooting.

Question: 4

Your data science team needs to collaborate on a project involving sensitive financial data

a. What is the primary way to enforce access controls and manage permissions within a CML workspace?

- A. Employing network-level firewalls at the instance level.
- B. Configuring user roles and project-level permissions.
- C. Relying on operating system-level user groups.
- D. Implementing encryption solely within the machine learning models.

Answer: B

Explanation:

CML provides granular access controls. User roles (e.g., Admin, Editor, Viewer) can be assigned, and these roles directly map to project-level permissions, ensuring control over who can access and modify work within a workspace.

Question: 5

A workspace has been running for an extended period, consuming significant resources. You want to optimize costs without losing work. Which actions would be appropriate?

- A. Delete the workspace and all its contents.
- B. Stop the workspace, then restart it when needed.
- C. Migrate the workspace to a less powerful instance type.
- D. Manually reduce the number of running sessions within the workspace.
- E. Convert the workspace instance to a spot instance type.

Answer: B,C,D

Explanation:

Stopping a workspace releases compute resources while preserving the work.

Resizing to a smaller instance type, if feasible, reduces ongoing costs.

Reducing active sessions frees up memory and CPU within the workspace.

Question: 6

You need to share a reproducible machine learning model, including its dependencies and environment configuration, with external stakeholders. Which of the following CML features would be the BEST fit for this task?

- A. Creating a detailed README file within the project.
- B. Exporting the workspace as a Docker image.
- C. Publishing the model as an API endpoint.
- D. Sharing the project's Git repository.

Answer: B

Explanation:

Exporting the workspace as a Docker image encapsulates the entire project environment, ensuring reproducibility regardless of the target machine where the stakeholders intend to run it.

Question: 7

While deploying a new model to a CML workspace, you encounter conflicts with existing library versions. Which approach would be effective in resolving this?

- A. Manually upgrade all project libraries to their latest versions.
- B. Create a new workspace with the required libraries already installed.
- C. Utilize a virtual environment or Conda environment within the workspace.
- D. Rely on the default CML environment without any modifications.

Answer: C

Explanation:

Virtual or Conda environments isolate project dependencies. This prevents conflicts with other projects in the workspace and maintains compatibility with the specific model requirements.

Question: 8

You have a computationally intensive model training workload in a CML workspace. Which of the following would likely lead to the most significant performance improvement?

- A. Increasing the workspace's storage capacity.
- B. Switching to an instance type with a GPU.

- C Upgrading to a newer version of CML.
- D. Refactoring the training code to use a more efficient algorithm.

Answer: B

Explanation:

GPU acceleration is often the most impactful factor for computationally heavy machine learning workloads, especially deep learning. Other options might help marginally but are unlikely to match the performance gains provided by a GPU.

Question: 9

Your CML workspace suddenly becomes unresponsive, and you suspect a memory leak. What tools or techniques would help you diagnose the issue?

- A. CML's built-in resource monitoring dashboard.
- B. System-level tools like top or htop.
- C. Python profiling libraries (e.g., cProfile, line_profiler).
- D. Examining CML application logs.
- E. Restarting the workspace instance.

Answer: A,B,C,D

Explanation:

Multiple tools can pinpoint memory leaks: CML Dashboard: Provides workspace-level CPU/memory usage

System Tools: Show process-level resource consumption.

Profilers: Identify memory-intensive code sections.

Logs: May reveal errors or warnings related to memory-

Question: 10

You're automating the creation of CML workspaces based on templates. Which technologies or approaches would be suitable?

- A. CML REST API
- B. Infrastructure as Code tools (Terraform, CloudFormation)
- C. Configuration management tools (Ansible, Chef)
- D Apache NiFi dataflows
- E. Manually configuring each workspace through the CML CLI

Answer: A,B,C

Explanation:

CML REST API: Allows programmatic workspace management.

Iac Tools: Provision infrastructure, including workspace instances, with desired configurations.
Config Management Tools: Can configure workspaces after they're created (software, libraries, etc.).

Question: 11

You want to track the evolution of experiments within a long-running CML project. Which combination of features would be the most effective?

- A. CML Jobs for scheduling regular training runs.
- B. Git version control for code and configuration.
- C. Experiment tracking tools like MLflow.
- D. CML's built-in file versioning.

Answer: B,C

Explanation:

Git: Tracks code history and is essential for collaboration.

Experiment Tracking: Logs parameters, metrics, and artifacts for each experiment, enabling comparison and reproducibility.

Question: 12

Due to a change in requirements, you need to modify the underlying operating system of your CML workspaces. How would you best achieve this?

- A. Install the required packages directly within the running workspace.
- B. Create a new CML project and migrate your work.
- C. Utilize a custom base image when creating new workspaces.
- D. Operating system modifications are not directly supported in CML.

Answer: C

Explanation:

Custom base images provide the most control over the workspace's OS environment. Modifying running workspaces is discouraged as it can lead to instability.

Thank You for Trying Our Product
Special 16 USD Discount Coupon: NSZUBG3X

Email: support@examsempire.com

**Check our Customer Testimonials and ratings
available on every product page.**

Visit our website.

<https://examsempire.com/>