



Vendor: IBM

Exam Code: C1000-173

Exam Name: IBM Cloud Pak for Data v4.7 Architect

Version: DEMO

QUESTION 1

An architect is working with a team to configure Dynamic Workload Management for a single DataStage instance on Cloud Pak for Data. Auto-scaling has been disabled and the maximum concurrent jobs has been set to 5. What will happen if a sixth concurrent job is executed?

- A. The sixth job will fail and will need to be restarted.
- B. The sixth job will queue until one of the other concurrent jobs completes.
- C. The sixth job will start if resources are available and will automatic.

Answer: B

Explanation:

Queueing mechanism is in place. Since the maximum concurrent jobs are limited to 5, the sixth job will not start immediately. Instead, it will queue and wait for one of the currently running jobs to complete, making room for the sixth job to execute.

QUESTION 2

How are Knowledge Accelerators deployed?

- A. Deployed as part of the sample assets.
- B. Deploy from the Cloud Pak for Data marketplace.
- C. Deploy by IBM support upon request.
- D. Deploy from the IBM Knowledge Accelerator API.

Answer: D

Explanation:

Knowledge Accelerators are a part of IBM Knowledge Catalog in IBM Cloud Pak for Data and provide predefined industry-specific business terms and relationships. These accelerators are not included by default with sample assets nor automatically deployed through the Cloud Pak for Data marketplace. Instead, they are imported using dedicated API endpoints provided by IBM for Knowledge Accelerators. Deployment involves uploading the accelerator assets using the IBM Knowledge Accelerator API, and once imported, they can be customized and published within the governance framework. This approach provides the flexibility required for various enterprise governance models.

QUESTION 3

Which Watson Pipeline component manages pipeline errors, typically used with DataStage?

- A. Fault Settings
- B. Default Control
- C. Error Handling
- D. Process Termination Window

Answer: C

Explanation:

In Watson Pipelines within IBM Cloud Pak for Data, error management is handled by the Error Handling component. This feature allows developers and pipeline administrators to define how pipeline failures are processed--whether to stop execution, continue, or trigger alternate flows. It ensures controlled behavior in response to job failures, particularly in complex ETL pipelines like those built with DataStage. Error Handling is a configurable element of pipeline orchestration and is typically used to enhance fault tolerance and control error propagation in production workflows.

QUESTION 4

Which Db2 Big SQL component uses system resources efficiently to maximize throughput and minimize response time?

- A. Hive
- B. Scheduler
- C. Analyzer
- D. StreamThrough

Answer: D

Explanation:

StreamThrough is a high-performance component used in Db2 Big SQL within IBM Cloud Pak for Data that is optimized to manage data streams and queries efficiently. It is designed to maximize throughput and minimize query response times by optimizing memory usage, resource allocation, and processing logic. Unlike Hive or Analyzer, which are used for query execution and analysis, StreamThrough enables efficient pipeline execution by streamlining data handling. Scheduler is used for job timing but does not influence runtime efficiency directly. StreamThrough is purpose-built to enhance performance through optimal resource usage.

QUESTION 5

What must be created to enable the Cloud Pak for Data platform to use a company's custom CA certificate to validate certificates from internal servers?

- A. A secret containing a wildcard certificate for all internal servers.
- B. A configmap that contains all internal server certificate chains.
- C. A secret that contains the company's CA certificate.
- D. A configmap that contains the company's CA certificate.

Answer: D

Explanation:

To enable IBM Cloud Pak for Data to trust certificates from internal servers using a custom Certificate Authority (CA), the correct method is to create a Kubernetes ConfigMap that contains the CA certificate. This ConfigMap is referenced by the platform's foundational services to include the CA in the trusted root store. Secrets are typically used for storing sensitive data like private keys and TLS certificates but are not used for adding trusted root CAs at the platform level. A ConfigMap is explicitly required by the platform to inject the CA trust into the certificate validation chain.

QUESTION 6

Which statement is true about governing data lakes in IBM Knowledge Catalog?

- A. It increases the time and effort required for data discovery and cataloging.
- B. It automates data discovery and provides access to enterprise data through virtualization.
- C. All data is masked with no user intervention.
- D. It relies on manual processes for accessing enterprise data.

Answer: B

Explanation:

Within IBM Knowledge Catalog as part of IBM Cloud Pak for Data, governing data lakes is enabled via integration with Data Virtualization. This approach supports automated data discovery, cataloging, tagging, and virtualization, allowing users to access enterprise data virtually--without physical movement. Policies and governance metadata are applied

automatically to virtualized assets, enabling secure and efficient data consumption. Manual processes are not required for discovery, and data is masked selectively based on policies--not completely masked without user intervention. Thus automation and virtualization are central, making statement B correct.

QUESTION 7

Which two of the following can be used with Watson Pipelines?

- A. Postgres
- B. Notebooks
- C. PowerShell
- D. Bash scripts
- E. Db2 Big SQL

Answer: BD

Explanation:

Watson Pipelines in Cloud Pak for Data support orchestration of diverse workload types including notebooks (Python or similar interactive environments) and scripts such as Bash. These pipeline components allow integration of notebook cells or shell scripts as tasks. There is no built-in support for executing PowerShell tasks directly (unless wrapped in Bash-like containers), and Postgres is used as a data source--not a pipeline component type. While Db2 Big SQL can be invoked within a notebook or script, it is not itself a pipeline component. Therefore the supported types in pipelines are notebooks and Bash scripts.

QUESTION 8

Which two Cloud Pak for Data services support the multi-tenancy mechanism of installing the service once and provisioning the workloads in tethered projects?

- A. Db2
- B. Cognos Analytics
- C. Analytics Engine Powered by Apache Spark
- D. IBM Data Virtualization
- E. Watson Pipelines

Answer: CD

Explanation:

Cloud Pak for Data supports service-level multi-tenancy by enabling some services to be installed centrally and then provisioned into user "tethered" namespaces (projects). In version 4.7, Analytics Engine (Apache Spark) and IBM Data Virtualization services support this tethered-project model: they can be installed once and then instantiated per project across multiple tenants. Core services like Db2, Cognos Analytics, and Watson Pipelines in 4.7 require installation per instance and do not yet support tethered-namespace provisioning via multi-tenant model.

QUESTION 9

Which two Cloud Pak for Data services support storage class NFS?

- A. watsonx Assistant
- B. IBM Knowledge Catalog
- C. Watson Knowledge Studio
- D. Watson Discovery

E. Planning Analytics

Answer: BD

Explanation:

IBM Cloud Pak for Data supports NFS-backed persistent volumes (RWX storage class) for services that require access to shared file storage. Among the available services, IBM Knowledge Catalog and Watson Discovery are known to support NFS storage classes in installation configuration (e.g. "managed-nfs-storage") for persistent metadata and document storage. Other services like watsonx Assistant, Watson Knowledge Studio, and Planning Analytics use different storage mechanisms and do not necessarily support NFS shared storage in the standard CP4D 4.7 deployment.

QUESTION 1

How do Cloud Pak for Data administrators obtain access to Match 360?

- A. Each user who is assigned to a service group has access.
- B. Each user is automatically granted read access.
- C. Administrative users automatically have access.
- D. Administrative users must belong to the appropriate service group.

Answer: D

Explanation:

Access to Match 360 within IBM Cloud Pak for Data is role-based and governed by service groups. Even administrative users are not granted automatic access unless they are explicitly assigned to the appropriate Match 360 service group. This allows fine-grained control over who can access master data management capabilities. Service group membership defines the roles and privileges needed for interacting with Match 360 functionalities like entity resolution and golden record management.

QUESTION 11

What endpoint will an application use to interact with Db2 Big SQL?

- A. Representative State Transfer (REST) Endpoint
- B. System Local Efficient Endpoint Pathways (SLEEP)
- C. Simple Normalized Optimum Representative Endpoint (SNORE)
- D. Dynamic Representative Endpoint Activation Mobility (DREAM)

Answer: A

Explanation:

Applications interact with Db2 Big SQL using industry-standard protocols. The most common and supported interface is through a REST (Representative State Transfer) API endpoint. REST endpoints allow for external applications to query, manage, and manipulate data within Big SQL using simple HTTP calls. None of the other options--SLEEP, SNORE, or DREAM--are valid or recognized interfaces in IBM Cloud Pak for Data or Db2 Big SQL documentation.

QUESTION 12

How many service instances can be provisioned for Watson Discovery at one time?

- A. 15
- B. 20
- C. 5

D. 10

Answer: D

Explanation:

"You can create a maximum of 10 instances per deployment. After you reach the maximum number, the New instance button is not displayed in IBM Cloud Pak for Data."

QUESTION 13

Which statement describes MPP (Massively Parallel Processing) Database architecture?

- A. A data warehouse that needs all compute nodes to access a shared data store simultaneously.
- B. Two or more databases kept in sync via two-phase commit.
- C. An analytics environment that improves performance by dividing the data across many nodes.
- D. A transactional system which uses multiple nodes for maximum availability.

Answer: C

Explanation:

MPP, or Massively Parallel Processing, is a database architecture model where data is divided and processed across multiple compute nodes in parallel. Each node works independently on a portion of the data, dramatically improving query performance and throughput for analytics workloads. This model is ideal for big data and analytical queries, not transactional workloads. It differs from shared-disk models or replication strategies like two-phase commit. The correct definition involves distributed data and parallel query execution, as described in option C.

QUESTION 14

What does Watson OpenScale require to generate statistics?

- A. Test data subset (10%)
- B. Complete test data set
- C. Training data
- D. Access to Pipeline

Answer: C

Explanation:

Training Data Statistics: Watson OpenScale needs to understand the characteristics of the data the model was trained on. This includes things like the distribution of features, sensitive attributes (for fairness monitoring), and how the model performed on this initial data. These "training data statistics" are crucial for:

Fairness Configuration: Recommending fairness attributes, reference, and monitored groups.

Bias Detection: Calculating fairness metrics (like disparate impact) by comparing runtime behavior to the learned training data distribution.

Explainability: Generating explanations by understanding the distribution of values in the training data to create meaningful perturbations.

Drift Detection: Building a drift detection model that compares runtime data to the training data to identify shifts.

While Watson OpenScale also consumes payload data (the data sent to the deployed model for predictions) at runtime to calculate various metrics and perform monitoring, the initial setup and the ability to generate meaningful statistics for things like fairness and drift fundamentally rely on understanding the training data.

Thank You for Trying Our Product

Lead2pass Certification Exam Features:

- ★ More than **99,900** Satisfied Customers Worldwide.
- ★ Average **99.9%** Success Rate.
- ★ **Free Update** to match latest and real exam scenarios.
- ★ **Instant Download** Access! No Setup required.
- ★ Questions & Answers are downloadable in **PDF** format and **VCE** test engine format.
- ★ Multi-Platform capabilities - **Windows, Laptop, Mac, Android, iPhone, iPod, iPad**.
- ★ **100%** Guaranteed Success or **100%** Money Back Guarantee.
- ★ **Fast**, helpful support **24x7**.



View list of all certification exams: <http://www.lead2pass.com/all-products.html>



Microsoft



ORACLE



JUNIPER
NETWORKS



EMC²
where information lives

10% Discount Coupon Code: ASTR14