PassQueen
a.k.a dumpsqueen.com

ACA Big Data Certification Exam

Alibaba Cloud ACA-BigData1

Version Demo

Total Demo Questions: 10

Total Premium Questions: 107

Buy Premium PDF

https://passqueen.com

support@passqueen.com

Topic Break Down

Topic	No. of Questions
Topic 1, Big Data Basics	13
Topic 2, Alibaba Cloud Big Data Products	74
Topic 3, Big Data Application Development	20
Total	107

QUESTION NO: 1

DataWorks Data Integration is used to configure a batch synchronization task from a MySQL source to a MaxCompute destination. Which of the following configurations can be used in the synchronization task? (Number of correct answers: 3)

A. Map source fields to destination fields.

B. Use constants or variables in field mappings where supported.

C. Configure a source-side filter condition for incremental extraction.

D. Run the task without configuring source and destination data sources.

ANSWER: A B C

Explanation:

A synchronization task can configure source-to-destination field mappings, allowing source columns to be mapped to destination fields. Constants and variables can also be used where supported to populate destination fields or parameterize task behavior. A source query or filtering condition can be configured to select the rows that need to be extracted, which is commonly used for incremental synchronization.

Data Integration does not eliminate the need to configure data-source connection information. A source and destination must be configured and authorized so that the task can connect to the relevant systems. For details, see Configure a data source and Configure a batch synchronization task.

QUESTION NO: 2

DataWorks can be used to create all types of tasks and configure scheduling cycles as

needed. The supported granularity levels of scheduling cycles include days, weeks,

months, hours, minutes and seconds.

A. True

B. False

ANSWER: B

Explanation:

False is correct because DataWorks scheduling does not use seconds as a supported scheduling-cycle granularity. In DataWorks, periodic scheduling is configured using defined cycle types, such as minute, hour, day, week, month, and year, together with the relevant scheduling parameters for that cycle. Minute-level scheduling is the finest standard periodic granularity; it is not a per-second scheduler. Therefore, a statement that lists seconds among the supported granularity levels is inaccurate, even though it correctly identifies several supported higher-level cycles.

This distinction matters when designing DataWorks workflows: tasks that require frequent execution must be planned according to the platform's minute-level scheduling capabilities and applicable scheduling constraints, rather than assuming a cron-like per-second trigger is available. DataWorks also provides scheduling properties and dependency controls so that recurring tasks can run in an ordered, managed production workflow. Alibaba Cloud's documentation describes the available scheduling-cycle configuration and parameters in its DataWorks scheduling guidance: Configure scheduling properties. For general DataWorks workflow and scheduling concepts, see DataWorks overview.

QUESTION NO: 3

Which of the following statements about MaxCompute user-defined functions are correct? (Number of correct answers: 3)

A. A Java UDF can be packaged as a JAR resource and uploaded to MaxCompute.

B. A function must be created before its name can be used in MaxCompute SQL.

C. Users require appropriate permissions to use a UDF.

D. A UDF automatically creates a target table when it is invoked.

E. UDF code must be stored only on a local client computer.

ANSWER: A B C

Explanation:

MaxCompute supports user-defined functions to extend SQL processing logic. A Java UDF implementation is commonly compiled into a JAR file and uploaded as a MaxCompute resource before the function is created. After the function is created, users need the required permissions to use it. UDFs are invoked in SQL expressions; they do not replace the need to define the target table schema. See User-defined functions and Resources.

QUESTION NO: 4

DataWorks provides powerful scheduling capabilities including time-based or dependencybased

task trigger mechanisms to perform tens of millions of tasks accurately and punctually each day based

on DAG relationships. It supports multiple scheduling frequency configurations like:

(Number of correct

answers: 4)

Score 2

A. By Minute

B. By Hour

C. By Day

D. By Week

E. By Second

ANSWER: A B C D

Explanation:

DataWorks scheduling supports several recurring cycle types, allowing workflow nodes to run at intervals appropriate to operational and analytical workloads. The supported frequencies represented here are **By Minute**, **By Hour**, **By Day**, and **By Week**. Minute scheduling is suitable for near-real-time periodic processing, while hourly scheduling is commonly used for regular incremental loads and aggregation. Daily scheduling supports standard batch data processing, and weekly scheduling is appropriate for reporting or maintenance workloads that occur on designated days.

These schedules work with the workflow's directed acyclic graph dependencies: DataWorks can trigger a node only after its upstream dependencies and its configured scheduling conditions are satisfied. This enables reliable orchestration of large numbers of data-development tasks without manually coordinating execution order. DataWorks also documents additional cycle configuration capabilities, including monthly cycles, illustrating that its scheduler is designed for a range of periodic batch workloads. The relevant available choices therefore correctly identify four supported scheduling frequencies. See the DataWorks scheduling-cycle documentation and the DataWorks scheduling basics documentation.

QUESTION NO: 5

You are working on a project where you need to chain together MapReduce, Hive jobs.

You also need the ability to use forks, decision points, and path joins. Which ecosystem

project should you use to perform these actions?

A. Apache HUE

B. Apache Zookeeper

C. Apache Oozie

D. Apache Spark

ANSWER: C

Explanation:

Apache Oozie is the appropriate ecosystem project because it is a workflow scheduler designed specifically to coordinate and manage Hadoop jobs. An Oozie workflow is expressed as a directed acyclic graph of control-flow nodes and action nodes, allowing related data-processing steps to be run in a defined sequence. Its action nodes support Hadoop workload types including MapReduce and Hive, so one workflow can launch a MapReduce operation and then trigger one or more Hive actions with the required dependencies.

Oozie also directly provides the required control-flow features. A *fork* control node starts parallel execution paths, a *join* node synchronizes those paths before downstream processing continues, and a *decision* node evaluates an expression to select a path dynamically. This lets a team model conditional processing and parallel branches without having to build custom orchestration code around each individual job. Oozie can additionally use coordinators to schedule workflow execution based on time or data availability, which makes it suitable for repeatable production pipelines. The Apache Oozie workflow specification documents its MapReduce and Hive actions as well as fork, join, and decision control nodes: Apache Oozie Workflow Functional Specification. General project documentation is available at Apache Oozie.

QUESTION NO: 6

DataWorks Data Quality can be used to monitor data in a MaxCompute table. Which of the following quality rules are appropriate for validating a daily sales table? (Number of correct answers: 3)

A. Check whether the current daily partition contains records.

B. Check whether the order_id column contains null values.

C. Check whether the row count is within an expected threshold.

D. Automatically change the data type of an invalid source column.

E. Automatically grant SELECT permission to all project users.

ANSWER: A B C

Explanation:

Data quality rules can validate whether expected data has arrived and whether it meets defined business requirements. For a daily sales table, checking that the current partition contains records, verifying that an order identifier does not contain null values, and comparing the row count with an expected threshold are valid checks. A quality rule does not itself change the table schema or automatically repair invalid source data. See the Data Quality overview.

QUESTION NO: 7

Resource is a particular concept of MaxCompute. If you want to use user-defined

function UDF or MapReduce, resource is needed. For example: After you have prepared

UDF, you must

upload the compiled jar package to MaxCompute as resource. Which of the following objects are

MaxCompute resources? (Number of correct answers: 4)

Score 2

A. Files

B. Tables: Tables in MaxCompute

C. Jar: Compiled Java jar package

D. Archive: Recognize the compression type according to the postfix in the resource name

E. ACL Policy

ANSWER: A B C D

Explanation:

MaxCompute resources are managed objects that provide code or data required when executing SQL jobs, user-defined functions, or MapReduce tasks. **Files** are a supported resource type and can hold auxiliary files that a job needs at runtime. **Tables: Tables in MaxCompute** is also a resource type, enabling a MaxCompute table to be referenced as resource data by supported processing tasks. **Jar: Compiled Java jar package** is specifically intended for Java code dependencies, including Java UDF, UDAF, and UDTF implementations as well as MapReduce programs. The JAR is uploaded first, then associated with the relevant function or task so MaxCompute can load it during execution. **Archive: Recognize the compression type according to the postfix in the resource name** is another supported resource type. Archive resources package multiple files and directories; MaxCompute identifies supported archive formats from the resource filename extension and decompresses them for task use. These four selections match the documented MaxCompute resource categories. Alibaba Cloud documents file, archive, table, Python, and JAR resources, and explains that resources are used to make external code and supporting data available to computation jobs. See MaxCompute Resource documentation and Create a resource.

QUESTION NO: 8

Function Studio is a web project coding and development tool independently developed by the

Alibaba Group for function development scenarios. It is an important component of

DataWorks.

Function Studio supports several programming languages and platform-based function development

scenarios except for _____ .

Score 2

A. Real-time computing

B. Python

C. Java

D. Scala

ANSWER: D

Explanation:

Scala is the correct answer. Function Studio is DataWorks' browser-based environment for developing, managing, and publishing user-defined functions. Its documented development capabilities focus on Java and Python function development, providing an integrated workflow for writing code, configuring dependencies, testing functions, and deploying them for supported compute-engine scenarios. The supported platform scenarios include real-time computing, so this is not solely a

language-specific feature: Function Studio is designed to streamline function development across DataWorks-supported processing environments. Scala is therefore the exception in the stated set because it is not listed as a directly supported Function Studio development language in the relevant Function Studio capability description. This distinction matters in practice: a developer selecting Function Studio should choose a supported language and use the platform's dependency and deployment workflow rather than assume that every JVM-based language can be authored as a native Function Studio function project. Alibaba Cloud's DataWorks documentation describes Function Studio and its supported function-development workflow at Function Studio. Additional DataWorks product and development documentation is available through the DataWorks documentation center.

QUESTION NO: 9

If a task node of DataWorks is deleted from the recycle bin, it can still be restored.

A. True

B. False

ANSWER: B

Explanation:

False is correct. In DataWorks, the recycle bin is the temporary recovery location for deleted workflow objects such as task nodes. A node can be restored only while its deleted record remains in the recycle bin. If the node is deleted from the recycle bin, that action permanently removes the recycle-bin record and its associated recoverable metadata; it is not a second reversible deletion stage. Consequently, DataWorks no longer provides a restore operation for that node through the workspace interface.

This behavior is important when maintaining production workflows. Before permanently deleting a node from the recycle bin, confirm that its code, configuration, dependencies, scheduling settings, and any required historical context have been retained elsewhere or are no longer needed. Teams should use the recycle bin as a short-term safeguard and restore objects before performing permanent removal. Alibaba Cloud's DataWorks documentation describes the recycle bin as the mechanism for recovering deleted objects and documents its management behavior. See DataWorks Recycle Bin documentation and the Alibaba Cloud DataWorks documentation for current workspace and object-management guidance.

QUESTION NO: 10

A DataWorks workflow contains an upstream node that produces a daily partition and a downstream MaxCompute SQL node that consumes the same business date. Which configuration is most appropriate to ensure that the downstream node runs only after the corresponding upstream instance succeeds?

A. Configure the upstream node as a resource in MaxCompute.

B. Configure a scheduling dependency from the downstream node to the upstream node.

C. Place both SQL statements in the same SELECT statement.

D. Configure the downstream node as a zero-load node.

ANSWER: B

Explanation:

Configure a scheduling dependency between the downstream node and its upstream node. DataWorks schedules periodic task instances according to both their configured cycle and dependency relationships. A downstream instance waits until the matching upstream instance is successfully completed before it is dispatched. This avoids reading incomplete daily data. See Configure scheduling dependencies.