DataStax

Certified-Cassandra-Developer

DataStax Apache Cassandra Developer Associate Certification

Latest Version: 6.0

Question: 1

An enterprise development team is designing a highly available, high-throughput system. They select Apache Cassandra specifically because they must deploy the cluster across multiple commodity servers and guarantee that system performance scales directly in proportion to the number of nodes added to manage increasing data volume. Which fundamental feature of Cassandra directly addresses this requirement?

A. Continuous availability across multiple datacenters.
B. Built-in integration of Materialized Views for flexible query patterns.
C. Dynamic data model enabling maximum flexibility.
D. Linear scalability across many commodity servers.
E. The automatic removal of single points of failure via the Dynamo architecture.

Answer: D

Explanation:
Apache Cassandra delivers linear scalability across many commodity servers, meaning that as nodes are added to increase capacity, the performance characteristics (throughput and latency) per node remain constant and overall cluster performance scales linearly. While continuous availability and no single point of failure are also key features, linear scalability is the specific feature guaranteeing predictable performance growth commensurate with adding hardware to manage increasing data loads.

Question: 2

A developer is modeling data for a customer profile system using CQL. The goal is to store customer-specific data (last_login, total_purchases) while ensuring the field applies universally to all records for a given customer, irrespective of how many secondary data records exist for that customer's partition. Which specific CQL feature enables this behavior?

A. The use of a Collection type (like a SET or MAP) to group customer data within a single cell.
B. Declaring the ‘customer_segment’ column as part of the Clustering Columns.
C. Defining ‘customer_segment using a User-Defined Type (UDT).
D. Declaring the ‘customer_segment’ column as a STATIC column.
E. Ensuring the data model uses the Composite Partition Key feature.

Answer: D

Explanation:

A column defined with the STATIC keyword holds a single value that is shared by all rows within the same partition. This feature ensures that even if multiple clustering rows exist for a single partition key (customer), the `customer_segment` will remain constant across that partition. The row definition syntax supports the STATIC keyword in the column definition: `column_definition::= column_name cql_type [ STATIC ] [ column_mask ] [ PRIMARY KEY]`. The example in the practice material demonstrates that a STATIC column, like `nickname`, holds a single value for the entire partition, even if multiple rows exist.

Question: 3

A software architect reviewing Cassandra's features identifies several advanced capabilities aimed at enhancing query performance, handling complex data types, and supporting modern analytical workloads. Which two of the following technologies, prominently listed as features/components in Cassandra documentation, relate to these specialized indexing and data handling capabilities? (Choose two.)

A. Improved Streaming for better internode data movement.
B. Storage-Attached Indexing (SAI).
C. Change Data Capture (CDC).
D. Vector Data Type (used in Vector Search).
E. Bloom Filters for optimizing read paths.

Answer: B,D

Explanation:
Storage-Attached Indexing (SAI) is a key feature listed in the sources under Indexing concepts and is documented separately for its concepts and quickstart, indicating its importance for querying capabilities. The Vector Data Type and associated Vector Search functionality are also listed as features under Data Types and Indexing, providing specialized capabilities for use cases like Approximate Nearest Neighbor (ANN) queries. Both SAI and Vector Search extend querying capabilities significantly. Improved Streaming and Bloom Filters are core architectural/operational features, while Change Data Capture (CDC) is a mechanism for capturing data changes, but SAI and Vector Data Type represent specific indexing/data handling features aimed at complex modern workloads.

Question: 4

A data architect is modeling a high-read performance application in Apache Cassandra, moving away from a traditional RDBMS design that relied heavily on normalized tables and complex JOINs. The architect consciously opts to denormalize data, duplicating customer details across and tables. What core architectural limitation in Cassandra primarily dictates this denormalized, query-centric modeling approach?

A. RDBMS uses row-level locking during transactions, but Cassandra uses lightweight transactions (LWT) which mandate denormalization.
B. Cassandra's internal architecture, unlike RDBMS, cannot perform disk I/O concurrently across multiple SSTables on the same physical node.
C. The core design of CQL strictly prohibits the use of JOINs and sub-queries, requiring all necessary data for a query to reside within a single table to ensure low latency.
D. Cassandra utilizes client-side joins in application code to manage complex relationships, whereas RDBMS performs server-side joins.

E. Data duplication in Cassandra (denormalization) is the only method to achieve write consistency across data centers via anti-entropy mechanisms.

Answer: C

Explanation:
Cassandra employs a query-centric data model where data is designed around expected application queries, unlike the table-centric RDBMS approach. CQL explicitly does not support joins or sub-queries, meaning a SELECT statement must apply to a single table. To fulfill query requirements efficiently, all necessary fields must be grouped into a single table, necessitating denormalization (data duplication) to achieve high read performance.

Question: 5

A developer transitioning from PostgreSQL wants to enforce data integrity, specifically ensuring that a column referencing another entity's key holds a valid value. In an RDBMS, this is handled by foreign keys. Regarding data structure definition, what essential RDBMS capability is explicitly stated as absent in Apache Cassandra's core data model?

A. The concept of a Primary Key, which combines partition and clustering columns.
B. Support for Data Definition Language (DDL) statements like CREATE KEYSPACE or CREATE TABLE.
C. The mechanism of logical data modeling based on eventual consistency principles.
D. The concept of foreign keys or relational integrity constraints.
E. The use of CQL (Cassandra Query Language) statements for data manipulation (DML).

Answer: D

Explanation:
Relational databases use foreign keys to maintain relations and integrity. By contrast, Apache Cassandra does not have the concept of foreign keys or relational integrity. Integrity enforcement must typically be handled at the application level.

Question: 6

The fundamental difference in data organization between Apache Cassandra and an RDBMS lies in Cassandra prioritizing a query-centric model that employs denormalization, versus the RDBMS reliance on normalization to manage relations and reduce duplication. Based on the principles described in the data modeling process, which two statements correctly describe the consequences or implications of using Cassandra's denormalized data model instead of a highly normalized RDBMS structure? (Choose Two)

A. Data duplication across multiple tables is generally required to support different application query patterns efficiently.
B. Client-side joins in application code are completely eliminated, ensuring that all relational complexities are handled solely within the database layer.

C. The query latency is minimized because all required columns for a specific query are located within a single, wide row (or minimal partitions).
D. Referential integrity and foreign key constraints are enforced automatically at the cluster level by the keyspace definition.
E. The storage efficiency and total dataset size are substantially improved due to the elimination of join overhead.

Answer: A,C

Explanation:
In Cassandra, denormalization means data is duplicated across various tables (Option A) to serve specific application query patterns (query-driven design). This design ensures that a query reads only the necessary partitions/rows to fulfill the request, minimizing overhead and reducing latency (Option C). Option D is incorrect because Cassandra lacks foreign keys. Option E is incorrect because duplicating data *increases* the total data storage size. Option B is incorrect because while Cassandra aims to minimize complexity, client-side joins may still be needed if the data model cannot fully integrate complex relationships.

Question: 7

A developer is working on an Approximate Nearest Neighbor (ANN) search application using a Cassandra 5.0 cluster. They have a table cycling.comments_vs containing a comment_vector VECTOR <FLOAT, column, indexed using SAI for ANN search. The developer needs to retrieve the 3 comments closest to a specific query vector [0.1, 0.2, 0.3, 0.4, 0.5] based on vector similarity. Which CQL statement correctly executes this vector query, leveraging the new features available in Cassandra 5.0?

- ○ SELECT * FROM cycling.comments_vs WHERE vector_distance(comment_vector, [0.1, 0.2, 0.3, 0.4, 0.5]) LIMIT 3;
- ○ SELECT * FROM cycling.comments_vs WHERE comment_vector == [0.1, 0.2, 0.3, 0.4, 0.5] ORDER BY ANN LIMIT 3;
- ○ SELECT * FROM cycling.comments_vs ANNSearch OF [0.1, 0.2, 0.3, 0.4, 0.5] LIMIT 3;
- ○ SELECT * FROM cycling.comments_vs ORDER BY comment_vector ANN OF [0.1, 0.2, 0.3, 0.4, 0.5] LIMIT 3;
- ○ SELECT * FROM cycling.comments_vs MATCH VECTOR [0.1, 0.2, 0.3, 0.4, 0.5] LIMIT 3;

A. Option A
B. Option B
C. Option C
D. Option D
E. Option E

Answer: D

Explanation:
Cassandra 5.0 introduces Vector Search functionality along with the `VECTOR` data type. The correct syntax for performing an Approximate Nearest Neighbor (ANN) search utilizes the `ORDER BY <vector_column> ANN OF <query_vector>` clause, followed by `LIMIT`. Option D accurately reflects this required syntax for querying vector data in Cassandra 5.0.

Question: 8

A Data Architect is designing a new table for product embeddings in Cassandra 5.0. They need to create a SAI index on the product_embedding column (which is a VECTOR) and must use the Cosine similarity metric for calculating vector distances, as this aligns with the embedding model used. The base table is inventory. products. Which set of options is required in the CREATE INDEX statement to specify the necessary similarity function for the vector index?

- ○ WITH OPTIONS = { 'index_type': 'ANN', 'distance_metric': 'COSINE' }
- ○ WITH OPTIONS = { 'similarity_function': 'COSINE' }
- ○ WITH OPTIONS = { 'similarity_metric': 'COSINE_DISTANCE' }
- ○ WITH OPTIONS = { 'function': 'SAI_COSINE' }
- ○ WITH OPTIONS = { 'vector_metric': 'COSINE' }

A. Option A
B. Option B
C. Option C
D. Option D
E. Option E

Answer: B

Explanation:
When creating a SAI index for Vector Search in Cassandra 5.0, the indexing options must define the similarity function using the key similarity _ function. Valid values for this option include DOT_PRODUCT, COSINE, or EUCLIDEAN. Therefore, option B correctly specifies the required configuration option and value for Cosine similarity.

Question: 9

A development team is preparing to migrate their application code and infrastructure to utilize the new capabilities introduced in Apache Cassandra 5.0. They must ensure that the updated application environment meets the core dependencies and leverages the enhanced architectural features. Which two capabilities are identified as major new features or core requirements/references associated with Apache Cassandra 5.0 in the documentation structure? (Choose two.)

A. Removal of Storage Attached Indexing (SAI) in favor of legacy Secondary Indexes (2i).
B. The required minimum platform support/reference environment for Java 17.
C. The official inclusion of Dynamic Data Masking (DDM) within CQL definitions.
D. Removal of all bulk loading tools, requiring ETL exclusively through client drivers.
E. Deprecation of the VECTOR data type in favor of User-Defined Types (UDTs) for embeddings.

Answer: B,C

Explanation:
Apache Cassandra 5.0 documentation highlights several key updates and requirements: Option B: The documentation structure explicitly references Java 17 under the Reference section, indicating its status as the supported platform. Option C: Dynamic Data Masking (DOM) is explicitly listed as a component within the Cassandra Query Language (CQL) sections defining Data Manipulation (DML), signifying its inclusion in this major version. Option A is incorrect; SAI is clearly highlighted in the Cassandra 5.0 documentation. Option D is incorrect; Bulk loading is listed under operating commands. Option E is incorrect; the VECTOR data type is a significant feature introduced in Cassandra 5.0 for Vector Search.

Question: 10

A Cassandra developer needs to dynamically enable audit logging across the cluster but must explicitly omit logging activity related to internal Cassandra operations within the system_traces keyspace. Which nodetool option is correctly used to apply this exclusion at runtime?

- ○ `--excluded-categories 'DDL, DML'`
- ○ `--ignored-keyspaces 'system_traces'`
- ○ `--excluded-users 'system'`
- ○ `--excluded-keyspaces 'system_traces'`
- ○ `--filter-tables 'system_traces.*'`

A. Option A
B. Option B
C. Option C
D. Option D
E. Option E

Answer: D

Explanation:
The `nodetool enableauditlog` command allows runtime configuration of audit logging. To specify a comma-separated list of keyspaces whose activity should be excluded from the audit log, the `--excluded-keyspaces` option is used. Other options exist for excluding categories (`--excluded-categories`) or users (`--excluded-users`).

Thank You for Trying Our Product

For More Information – Visit link below:

https://www.examsboost.com/

15 USD Discount Coupon Code:

G74JA8UF

FEATURES

- ✓ 90 Days Free Updates
- ✓ Money Back Pass Guarantee
- ✓ Instant Download or Email Attachment
- ✓ 24/7 Live Chat Support
- ✓ PDF file could be used at any Platform
- ✓ 50,000 Happy Customer