NoSQL Quiz

Question 1

Which of the following is NOT correct about the Relational Databases (RDB)?

It separate the logical organization of data structures from the physical storage of those structures

Correct Answer

All other options are correct

Large enterprises can support complex applications serving thousands of users

Ensure anyone reading data will have a consistent view of the data

Designed to support hundreds and even thousands of users simultaneously

Question 2

If the layout of records in a flat file data management system changes, what else must change?

Data security measures will have to change

No change will be necessary in a flat-file data management system

Correct Answer

Data access programs will have to change.

Database operating system will have to change

Question 3

NoSQL databases are likely to displace relational databases as relational databases displaced earlier types of data management systems.

True

Correct!

False

Question 4

Three common major components of a database application are:

A user interface, Data Security Module, and Database code

Large-Volume Storage, Business logic, and Database code

A user interface, Business logic, and Data Back-up

Correct Answer

A user interface, Business logic, and Database code

Question 5

Which are motivating factors for database designers and other IT professionals to develop and use NoSQL databases.

Security, Cost, Flexibility, Availability

Correct!

Cost, Flexibility, Availability, Scalability

Scalability, Commonality, Flexibility, Availability

Scalability, Security, Dynamic, Availability

Question 6

Which of the following are required components of a relational database management system (RDMS)(check all that apply)

Correct Answer

Memory management programs

Correct!

Data dictionary

Correct!

Storage management programs

Data Nodes and Links

Data Security Module

Report Management System

Question 7

What kind of relation is supported in a hierarchical data management system?

No relations are allowed.

Correct!

Parent-child

Many-to-many

Many-to-many-to-many

Question 8

If the layout of records in a flat file data management system changes, what else must change?

Database operating system will have to change

Data security measures will have to change

No change will be necessary in a flat-file data management system

Correct Answer

Data access programs will have to change.

Question 9

Scaling up entails upgrading an existing database server to add additional processors, memory, network bandwidth, or other resources that would improve performance.
Correct!
True
False

Question 10

Which of the following SQL is NOT correct grammatically in PostgreSQL?

INSERT INTO test.countries (country_code, country_name)
VALUES ('us', 'United States'), ('un', 'United Nation');
Correct Answer

UPDATE test.countries
SET name = 'United Nations'
WHERE code = 'un'
All other options are correct

INSERT INTO test.countries
VALUES ('kr', 'South Korea');

CREATE TABLE test.countries (
country_code char(2) PRIMARY KEY,
country_name text UNIQUE);

Question 1

In a Key-Value Database Keys must be unique within a Namespace.
Correct!
True
False

Question 2

Eventually consistent means that some replicas might be inconsistent for some period of time but will become consistent at some point.
Correct!

True
False

Question 3

How many values can be stored with a single key in a key-value database?
Correct!
1
2
Unlimited
None

Question 4

Document databases require a fixed, predefined schema.
True
Correct!
False

Question 5

Which of the following statements is NOT true about Document Databases?

Store multiple attributes in a single document instead of storing each attribute of an entity with a separate key
You Answered
Provides application programming interfaces (APIs) or query languages
Correct Answer
All other options are true
Store multiple key value pairs in groups known as documents
Documents can have embedded documents

Question 6

What do the BA and E in the BASE properties in NoSQL stand for? Check all that apply.

E: Easy to Use
BA: Baseline
BA: Best Accessible
Correct!
BA: Basically Available
E: Expandable
Correct!
E: Eventually Consistent

Question 7

You are assigned the task of building a database to model employees and who they work with in your company. The database must be able to answer queries such as how many employees does Employee A work with? And, does Employee A work with anyone who works with Employee B? Which type of NoSQL database would naturally fit with these requirements?

A key-value database

Correct!

A graph database

A column-family database

A document database

Question 8

Which of the following statements is NOT true about Column Family Databases (CFDB)?

Correct!

All other options are true

Share some terms with relational databases: rows and columns

Use a map of maps model to store columns and attributes in column families

CFDB and RDB are superficially simila, but, CFDB does NOT support ‘table join’

The most complex of the NoSQL database types

Question 9

What do the C and A in the CAP theorem stand for?

Compression and Availability

Correct!

Consistency and Availability

Commonality and Availability

Consistency and Anti-entropy

Question 10

What are the two fundamental data structures in a graph database?

Stacks and queues

Columns and column families.

Heaps and Hash Tables

Correct!
Nodes and relations

Question 1

The three common features of key-value databases are:
Simplicity, Flexibility, Speed
Flexibility, Speed, Scalability
Security, Speed, and Simplicity
Correct!
Simplicity, Speed, and Scalability

Question 2

What is an advantage to having a large number of replicas?
The more replicas you have, the faster your queries can run.
You will have improved consistency with a large number of replicas
Correct!
The more replicas you have, the less likely you will lose data
You will have improved performance with a large number or replicas.

Question 3

What is the relation between speed of compression and the size of compressed data.
Faster compression algorithms can lead to smaller compressed data
Slower compression algorithms can lead to larger compressed data
There is no relation between speed of compression and the size of the compressed data
Correct!
Faster compression algorithms can lead to larger compressed data

Question 4

Which statement is not true in regards to Data Structures

Data structures are implemented using elements of underlying hardware, particularly random access memory and persistent data storage .

Data structures are well-defined data storage structures.

Correct!

Data structures provide a level of organization and abstraction above data models.

Data structures provide a high-level abstraction of data storage.

Question 5

Data models are abstractions that help organize the information conveyed by the data in databases.

Correct!

True

False

Question 6

When working with Hash Functions, what is a collision?

A collision occurs when two identical inputs to a hash function produce the same output.

A collision occurs when two distinct inputs to a hash function produce different outputs.

Correct!

A collision occurs when two distinct inputs to a hash function produce the same output.

A collision occurs when two identical inputs to a hash function produce the same output.

Question 7

Key-Value database clusters tend to be

Correct!

Loosely coupled

Completely Coupled

Tightly coupled

Key-value databases cannot run on clusters

Question 8

Key-value databases do Not support query languages for searching over values.

Correct!

True

False

Question 9

Select all statements that apply to Associative Arrays

Associative arrays are used to improve relational database scalability

Associative arrays are used to improve relational database security.

An associative array is used to implement search in a Key-Value Database

Correct!

Associative arrays generalize the idea of an ordered list indexed by an identifier to include arbitrary values for identifiers and values.

Correct!

An associative array is a data structure, like an array, but is not restricted to using integers as indexes or limiting values to the same type.

Correct Answer
An in-memory cache is an associative array.

Question 10

Which of the following is Not true about a namespace

Keys must be unique within a namespace

A namespace is a logical data structure for organizing key-value pairs

Namespaces are sometimes called buckets

Correct!

A namespace is used to improve relational database performance

---

Question 1

Which of the following is NOT true about key-value DB or its designing?

Key-value databases are relatively simple, but designing is challenging
Keys should have some logical structure to make code readable and extensible
Correct Answer
All other options are true
Key design has an impact on database performance

Use meaningful and unambiguous naming components

Question 2

Which of the following is NOT true about key-value DB design considerations?

Correct!
Well-designed key pattern increase the amount of code a developer must write to create functions that access and set values
In some organizations, custom applications should have code to log specific events
All other options are true
All code running in a production environment must have error checking code.
One criteria for assessing a naming convention is to consider how much code is required to generate keys in a generic set or get function

Question 3

Which of the following is NOT true about the limitations of the Key-Value Database?

Some key-value databases do not support range queries
Correct Answer
All other options are true
There is no standard query language comparable to SQL for RDB

Some key-value databases require that all keys must be strings
The only way to look up values is by key

Question 4

When working with Hash Functions, what is a collision?
You Answered
A collision occurs when two identical inputs to a hash function produce the same output.
A collision occurs when two identical inputs to a hash function produce the same output.
A collision occurs when two distinct inputs to a hash function produce different outputs.

Correct Answer

A collision occurs when two distinct inputs to a hash function produce the same output.

Question 5

What is range partitioning?

Correct!

Range partitioning works by grouping contiguous values and sending them to the same node in a cluster.

Range partitioning divides contiguous values and uses a hash function to distribute them to nodes in the cluster.

Range partitioning divides contiguous values and distributes them evenly across the cluster.

Range partitioning works by grouping contiguous values and sending the group to multiple nodes in a cluster.

Question 6

It is not possible for replicas to have different versions of data. If they did, it would be impossible to ever restore consistency to the database.

True

Correct!

False

Question 7

Key-Value database clusters tend to be

Key-value databases cannot run on clusters

Correct!

Loosely coupled

Completely Coupled

Tightly coupled

Question 8

Which design pattern provides some of the features of relational transactions?

Emulating tables

Correct Answer

Atomic Aggregate Pattern

Indexes

A TTL key pattern

Question 9

Which design pattern specifies a time that a key-value record is allowed to exist ?
Aggregate Pattern
Indexes
Emulating tables
Correct!
A TTL key pattern

Question 10

Which of the following is NOT one of the four characteristics of a well-designed key-naming convention.
Correct!
Avoid appending components to make a key
Keep keys as short as possible
Use meaningful and unambiguous naming components
Use range-based components when you would like to retrieve ranges of values

---

Question 1

Two types of formats for storing data in a document database are
JSON and Java
XML and Java
Java and SQL
Correct!
XML and JSON

Question 2

When is it reasonable to use highly abstract entities?
Abstract entities should be used when many of the queries used against a collection are unique and the collection contains few subtypes
Abstract entities should never be used in modeling document collections
Correct!
Abstract entities should be used when many of the queries used against a collection apply to all or many subtypes.
Abstract entities should be used if there is a potential for the number of subtypes to grow into the tens or hundreds.

Question 3

Which query operator is used to search for values in multiple keys?
$let

$gte
Correct Answer
$or

$in

Question 4

Why would you want to denormalize collections in a document database?
To reduce the amount of redundant data in the database
To evenly distribute documents over shards
TO reduce the potential for data anomalies
Correct!
To improve performance

Question 5

Because Document databases do not require data modelers to formally specify the structure of documents they are considered _____.
Correct!
schemaless
Correct Answers
Schemaless
schemaless

Question 6

Because the documents that exist in Document Database collections can have many different forms, Document Databases are considered _____.

schemaless
Correct Answers
polymorphic
Polymorphic

Question 7

Which of the following MongoDB commands is NOT correct grammatically?

db.towns.find({population : population_range}, {name: 1})
show collections
db.towns.insert({ name: "New York", population: 8406000})
Correct!
All are correct

use book
db.towns.find({"_id": ObjectId("5da65ce1403c741a5baa7d64") })

Question 8

This question has been regraded.
One of the weaknesses of a Document Database is that Developers can not make up their own set of attribute tags; they are constrained to a predefined set of tags for specifying structure.

True
Correct!
False

Question 9

Documents in document databases are composed of a set of key and value pairs.

Correct!
True
False

Question 10

Which of the following are tips on designing collections for document databases (check all that apply) ?

Correct!
Avoid overly abstract document types
Correct Answer
Avoid overusing denormalization
Correct!
Watch for separate functions for manipulating different document types
Correct!
Use document subtypes when entities are frequently aggregated or share substantial code
Correct Answer
Avoid explicit schema definitions

How do document database modelers avoid costly joins?
Correct!
Use denormalized data models.

Use few Indexes
Use normalized data models.
Use many Indexes

Question 2

Three ways to model hierarchies in a document database are:

Parent references, child references, listing all ancestors.

Question 3

Too much denormalization will lead to large documents that will likely lead to unnecessary data read from persistent storage.
Correct!
True

Question 4

When would it make sense to have many indexes on your document collections?
When you do not need to filter queries on large documents
Correct!
When you have read-heavy applications
When you have write-heavy applications
When you do not have any ad hoc query requirements

Question 5

Which of the following statements is FALSE in regards to Vertical Partitioning
Correct!
These parts of the database,known as shards, are stored on separate servers.
is a technique for improving database performance by separating columns of a relational table into multiple separate tables.
is particularly useful when you have some columns that are frequently accessed and others that are not.
Vertical partitioning is typically used with relational tables because they have a fixed structure.

Question 6

Which statement is FALSE in regards to JOIN operations
As the size of tables grows, JOIN operations take longer.
JOINS retrieve data from multiple tables.

JOINS use loops, hashes, and merge operations.

Correct Answer

JOINS can improve performance, but they also require disk seeks to retrieve data blocks

Question 7

Match the term to its meaning

Correct!

Horizontal partitioning -the process of dividing a database by documents in a document database or by rows in a relational database

Shards -The parts of a horizontally partitioned database that are stored on separate servers

Correct!

Shard Key -one or more keys or fields that exist in all documents in a collection that is used to separate documents into different partitions.

Partitioning Algorithm -determines how to distribute documents overshards.

Question 8

What process reduces the amount of redundant data in a database

Vertical Partitioning

Correct!

Normalization

De-Duplication

Horizontal Partitioning

Question 9

Indexes support the efficient execution of queries in MongoDB. The explain(“executionStats”) method outputs details, e.g., information if an index is used, of a given operation. Which of the following stage names is found under the queryPlanner.winningPlan, we considered having used an index?

COLL

INDEX

Correct Answer

IXSCAN

USEIDX

FETCH

Question 10

Which type of Relation is the following: Students can be enrolled in many courses and courses can have many students enrolled.

One-to-One
One-to-Many
Correct!
Many-to-Many
Many-to-One

Question 1

Column family databases are appropriate choices for large-scale database deployments that require high levels of write performance or when a large number of servers are required to meet expected workloads
Correct!
True
False

Question 2

HBase uses what types of Hadoop nodes?
Name Nodes and Job Tracker
Data Nodes and Task Tracker
Correct!
Data Nodes and Name Nodes
All of Hadoop's architectural nodes are used by HBase

Question 3

Which of the following is NOT true about peer-to-peer networks or approach that are used in Cassandra?
All Cassandra nodes run the same software
No node can be a single point of failure
Correct!
Have a single master coordinating server
Servers in a peer-to-peer network communicate with each other and, eventually, new nodes are assigned a set of data to manage

Scaling up and down is fairly straightforward

Question 4

Cassandra uses an anti-entropy algorithm, that is, one that increases order, to correct inconsistencies between replicas. When a server initiates an anti-entropy session with another server, it sends a hash data structure, known as a Merkle or hash tree, derived from the data in a column family. The receiving server calculates a hash data structure from its copy of the column family. If they do not match, the servers determine which of the two has the latest information and updates the server with the outdated data

Correct!

True

False

Question 5

Which of the following is/are NOT true about the core features of Google's BigTable? (Check all that apply)

Correct!

Reads of a row are atomic, but writes are not atomic

Correct Answer

Columns are maintained in a sorted order

Developers have dynamic control over columns

Data modelers and developers have control over location of data

Data values are indexed by row identifier, column name, and a time stamp

Question 6

Which of the following is/are NOT true about the gossip protocol used in Cassandra?

First, a node in the cluster initiates gossip sessions with all nodes

A more-efficient method of sharing information is to have each server update another server about itself as well as all the servers it knows about

The problem of sharing information about the state of servers in a cluster is that all-servers-to-all-other-servers protocol can quickly increase the volume of traffic on the network

Servers in a peer-to-peer network communicate with each other and, eventually, new nodes are assigned a set of data to manage

In the course of message exchange, each server is updated about the state of servers as known by the other server

Question 7

A hinted handoff entails storing information about the write operation on a proxy node and periodically checking the status of the unavailable node. When that node becomes available again, the node with the write information sends, or “hands off,” the write request to the recently recovered node- **True**

Question 8

Which of the following is NOT true about similarities to and differences from key-value/document databases?

Key-value databases do not support querying based on contents of the JSON or XML string
Document databases enable you to query and filter based on elements in the document
Both DBs do not require all columns in all rows

In key-value database, you could query and extract a subset of the data
In both column family and document databases, columns or fields can be added as needed by developers

Question 9

Peer-to-peer architectures use one node type so that any node can assume responsibility for any service or task.- True

Question 10

Which of the following HBase command(s) is NOT correct? (Check that all apply)

You Answered
hbase> create 'myns:safe_table', 'will_NOT_be_deleted_cf'
hbase> create "myns:mytable2_tbl", "mycolfam_cf"
hbase> scan 'myns:mytable_tbl'
Correct Answer
hbase> put 'myns:mytable_tbl', 'myrowkey_key', 'Welcome to My HBase!'
All other choices are correct
You Answered
hbase> create_namespace 'myns'
You Answered
hbase> create 'del_table', 'will_NOT_be_deleted_cf'
—----------

A partition is a subset of servers configured to function together.-- False

Question 2

What are the three parts of a column in a column family database
  Column name, Row Key, Time/Version stamp
  Row Key, Value, Time/Version stamp
  Row Key, Column name, Value
Correct!
  Column name, Time/Version stamp, Value

Question 3

Which process or protocol can improve the availability of write operations
  Gossip Protocol
Correct!
  Hinted Handoff
  Replication
  Anti-Entropy Process

Question 4

Which process or protocol detects and resolves inconsistencies in distributed replicas
  Hinted Handoff
  Gossip Protocol
  Replication
Correct!
  Anti-Entropy Process

Question 5

Which of the following is NOT a characteristic of column family databases
  Columns can vary between rows
Correct!
  Entities are modeled as columns in column family database
  Columns can be added dynamically
  Joins are not used

Question 6

Which of the following scenarios would be a good use case for descriptive statistics
Correct Answer
  Compare the average salaries of IT professionals by job title

Examine the monthly change in the average price of a gallon of gasoline and predict what the price will be next month

survey 100 students to predict the outcome of the SGA elections

All of the scenarios

Question 7

An output of an HBase shell command is shown below. Which command produces the output below?

```
hbase(main):001:0> [Command?] 'truck_event'
 ROW                                COLUMN + CELL
 10|23|9223370572126391280          column=events::eventTime,
timestamp=1600813667091, value=59:44.5
 10|23|9223370572126391280          column=events:CorrelationId,
timestamp=1600813667091, value=1000
 10|23|9223370572126391280          column=events:driverId, timestamp=1600813667091,
value=10
 10|23|9223370572126391280          column=events:driverName,
timestamp=1600813667091, value=George Vetticaden
 10|23|9223370572126391280          column=events:eventDate,
timestamp=1600813667091, value=2016-06-02-20

 <Skipped>

 10|39|9223370571956618141          column=events:driverId, timestamp=1600813667091,
value=10
 10|39|9223370571956618141          column=events:driverName,
timestamp=1600813667091, value=George Vetticaden
 10|39|9223370571956618141          column=events:eventDate,
timestamp=1600813667091, value=2016-06-02-20
17077 row(s)
Took 32.4493 seconds
```

Correct!

Scan

Show

Describe

Select

Put

Get

List

Question 8

HBase commands are shown below. Which of the following commands is NOT correct? (Check all that apply)

Correct Answer
enable 'mytable', 'my_cf'
disable 'mytable'; drop 'mytable'
alter 'mytable', {NAME => 'my_cf', VERSIONS => 5}
create 'mytable', 'my_cf'
Correct!
remove 'mytable', 'my_cf'

Question 9

Which of the following is NOT an advantage of using a Bloom Filter

A negative response guarantees an element is not in the set
Correct Answer
False positive matches are never possible

False negative matches are never possible
Bloom filters are used to reduce the number blocks read.

Question 10

Hotspotting occurs when many operations are performed on a small number of servers which leads to underutilization of cluster resources

Correct Answer
True

False

Question 1

Which of the following is NOT true about Apache HBase?
Provides Fast record lookup
Support for record-level insertion
Correct!
Good for batch processing (scans over big files)
Support for updating records
None of the above (All true)

Question 2

Which of the following is NOT true about column families/qualifiers of HBase?
Compression is configurable by column family
The version retention policy is configurable by column family
None of the above (All true)
Column families are stored separately on disk
Correct!
Recommend all column family members have different access patterns

Question 3

The roles of HBase components or instances are described below. Which of the following is correct?“The <A> opens the region and creates a corresponding <B> object. When the <B> is opened it sets up an <C> instance for each HColumnFamily for every table as defined by the user beforehand”
A: HRegionServer, B: HRegion, C: HLog
Correct!
A: HRegionServer, B: HRegion, C: Store
A: HRegion, B: HLog, C: Memstore
None of the above
A: HBase, B: HRegionServer, C: HRegion

Question 4

To see region information, startkey/endkey, and region server information of a table, which command of the following is used in HBase shell?
Correct!
scan ‘hbase:meta’, {FILTER=>“PrefixFilter(‘table_name’)”}
create ‘table_name’, ‘columnfamily_name’
describe ‘table_name’, {name=>‘regionserver’}
show region ‘table_name’
None of the above

Question 5

Which of the following is NOT true about B+ Trees and Log-Structured Merge-Trees (LSM-Trees)?
B+ trees achieve a much higher fanout compared to binary trees, resulting in a much lower number of I/O operations to find a specific key
None of the above (All true)
Correct!
LSM-trees work at disk seek time and scale much better to handle large amounts of data

B+ trees have some specific features that allow for efficient insertion, lookup, and deletion of records that are identified by keys

LSM-trees guarantee a very consistent insert rate, as they transform random writes into sequential writes using the logfile plus in-memory store

Question 6

Which of the following is NOT true about split and compaction in HBase?

Major compaction Compact all files into a single file (in a store)

HBase handles the splitting of regions automatically (auto-sharding)

Correct Answer

None of the above (All true)

Once the regions reach the configured maximum size, split into two halves and can start taking on more data

Minor compaction rewrites the last few files into a larger one (in a store)

Question 7

Which of the following is NOT true about the physical storage layout of HBase?

The cells are sorted in descending order by that timestamp

Correct Answer

Store both set and unset cells

None of the above (All true)

Store the row key and column key with every cell

Multiple versions of the same cell are stored as separate, consecutive cells

Question 8

10 / 10 pts

Which of the followings are true about Auto Sharding of HBase?

If you exceed limit when you add data, the region is split into multiple pieces

Correct!

Auto sharding start with one region, then start adding data

Region splitting is very slow because the split regions read from the original storage files

All other answers are true

Correct!

The region is split at the row key in the middle of the region

Question 9

Which of the following is NOT true about write path in HBase?

The first step is to write the data to the write- ahead log (the WAL), represented by the HLog class

None of the above (All true)

Once the data is written to the WAL, it is placed in the MemStore

C**orrect!**

Memstore is checked to see if the MemStore is full and, if so, a flush is requested. The data is flushed into the memory (RAM)

The WAL is a standard Hadoop SequenceFile and it stores HLogKey instances

Question 10

The figure shows files and directories of HBase Storage in HDFS. Which level of the directory is shown in the figure?

/hbase/

/hbase/data/<namespace>/table-name/

Correct!

/hbase/data/<namespace>/table-name/region-name/

/hbase/data/<namespace>/

/hbase/data/<namespace>/table-name/region-name/column-family-name/

Question 1

Which of the following is NOT true about Apache Cassandra?

Due to the decentralized model, it has no single point of failure.

Correct Answer

Support stored procedures through coprocessors

Cassandra trades some consistency to achieve total availability

Elastic scalability in Cassandra means seamless scale out and scale back

It has Dynamo-style replication model

Question 2

Which of the following is NOT true about differences between Cassandra vs. RDBMS?

HBase don't provide SQL-like Query Language for developers' transition from RDBMS

HBase provides RDBMS-like triggers.

Cassandra allows you to tune for performance or consistency or a balance of both

HBase provides strong consistency at the record level

Correct Answer

Cassandra's symmetric architecture makes it difficult to create and scale large clusters

Question 3

Which of the following is NOT true about data modeling differences in Cassandra (vs. RDBMS)?

Query-first design

Do not enforce referential integrity

Correct!

Cassandra performs best when the data model is normalized

None of the above (all true)

No Joins, so do the work on the client side

Question 4

Cassandra's data model can be described as a partitioned row store and data is stored in sparse multidimensional hash tables.

Correct!

True

False

Question 5

Which of the following is NOT true about tuneable consistency of Cassandra?

Cassandra trades some consistency to achieve total availability

Read consistency level "one" means immediately return the record held by the first node that respond to the query

Correct!

Means it allows you to easily decide the level of consistency you require, in balance with the level of partition tolerance

None of the above (all true)

Read consistency level "quorum" means once a majority of replicas ((replication factor / 2) + 1) respond, return to the client the value with the most recent timestamp

Question 6

Cassandra's replication strategy, named 'NetworkTopologyStrategy', allows you to specify a different replication factor for each data center. Within a data center, it allocates replicas to different racks in order to maximize availability.

Correct!

True

False

Question 9

Let’s assume that the table, user, has a massive number of rows in it, and the column is heavily used in many queries. So, we should design the table to allow users to use the column ‘last_name’ in the ‘where’ clause. In this case, ‘Allow filtering’ is the best solution.

True
Correct!
False

Question 10

We create a table, use, using below CQL.

CREATE TABLE user (first_name text, last_name text, PRIMARY KEY (first_name));
After we populate the table, we try to retrieve data using below CQL.

SELECT * FROM user WHERE last_name='Lee';
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot execute this query as it might involve data filtering and thus may have unpredictable performance. If you want to execute this query despite the performance unpredictability, use ALLOW FILTERING"
We got an error. Explain why this happens, in other words, what causes this.

Allow filtering helps in executing the query but the performance is unpredictable and so it is not recommended. Here the query works fine if it is executed with first_name as it is the primary key whereas the last_name is not the primary key , it gives the above error. It might need a clustering column as the partition key is not mandatory.

—----------

Question 1

A table schema is shown below. The following query works well. Is it true or not?

```
/* Table schema */
Create table Products (
  type text,
  price int,
  desc text,
  primary key (type, price)
```

```
);

/* Query */
SELECT * FROM Products WHERE type = 'computer' AND price < 1000;
```

Correct!

- **True**
- False

Question 2

A table schema is shown below. The following query works well. True or not?

```
/* Table schema */
Create table SocialMedia (
  src  text,
  name text,
  sent text,
  body text,
  date int,
  primary key (src, name, date)
);

/* Query */
SELECT * FROM SocialMedia WHERE src = 'twitter' AND date >= 2020101 AND date <=
20221231;
```

- True

Correct!

- **False**

Question 3

NetworkTopologyStrategy of Cassandra is a replication strategy and places replicas at consecutive nodes around the ring.

- True

Correct!

- **False**

Question 4

Paxos is a consensus algorithm that allows distributed peer nodes to agree on a proposal, without requiring a master to coordinate a transaction.

Correct!

- **True**

False

Question 5

Which of the following is NOT true about Cassandra's snitches?
All true
Snitches gather information about your network topology so that Cassandra can efficiently route requests
Dynamic snitch monitors the performance of requests to the other nodes
Correct!
Cassandra's default snitch is a SimpleSnitch and knows about the racks and data centers in a cluster
The job of a snitch is to determine relative host proximity for each node in a cluster, which is used to determine which nodes to read and write from

Question 6

Which of the following is NOT true about data modeling differences in Cassandra (vs. RDBMS)?
Do not enforce referential integrity
No Joins, so do the work on the client-side
C**orrect!**
Cassandra performs best when the data model is normalized
Query-first design

Question 7

Which of the following is NOT true about the composite key or compound key of Cassandra?
Correct!
A Clustering column (or key) is mandatory
A table can have more than one clustering columns
Partition key determines the node on which rows are stored
Partition key can be multiple columns

Question 8

Which of the following is NOT true about the coordinator node?
For a read, the coordinator contacts enough replicas to ensure the required consistency level is met and returns the data to the client
All true
For a write, the coordinator node contacts all replicas, as determined by the consistency level and replication factor
Coordinator identifies which nodes are replicas for the data that is being written or read and forwards the queries to them
Correct!

A client can connect the master coordinator node only in the cluster to initiate a read or write a query

Question 9

Which of the following is NOT true about the gossip protocol and failure detection?
Accrual failure detector finds a place, suspicion level, in between the extremes of dead and alive
Cassandra uses a gossip protocol that allows each node to keep track of state information about the other nodes in the cluster
Cassandra gossip is used for failure detection, the Gossiper class maintains a list of nodes that are alive and dead
Cassandra does not decide alive/dead based on "heartbeats" only
Correct!
All true

Question 10

Which of the following is NOT true about the tunable consistency of Cassandra?
Read consistency level “quorum” means once a majority of replicas ((replication factor / 2) + 1) respond, return to the client the value with the most recent timestamp
Read consistency level “one” means immediately return the record held by the first node that respond to the query
Cassandra trades some consistency to achieve total availability
Correct!
Means it allows you to easily decide the level of consistency you require, in balance with the level of partition tolerance
All true

Question 1

The data dictionary is the part of the RDBMS that keeps track of information about the structure of data stored in the database. This includes

Correct!
Schemas
Rows
Databases
Correct!
Constraints
Correct Answer
Columns

Question 2

If the layout of records in a flat file data management system changes, what else must change?

Correct!

Data access programs will have to change.

No change will be necessary in a flat-file data management system

Database operating system will have to change

Data security measures will have to change

Question 3

What do the BA and E in the BASE properties in NoSQL stand for? Check all that apply.

Correct!

BA: Basically Available

E: Easy to Use

BA: Baseline

E: Expandable

BA: Best Accessible

Correct!

E: Eventually Consistent

Question 4

What is the relation between speed of compression and the size of compressed data.

There is no relation between speed of compression and the size of the compressed data

Slower compression algorithms can lead to larger compressed data

Faster compression algorithms can lead to smaller compressed data

Correct!

Faster compression algorithms can lead to larger compressed data

Question 5

Which statement is not true in regards to Data Structures

Correct Answer

Data structures provide a level of organization and abstraction above data models.

Data structures are well-defined data storage structures.

Data structures provide a high-level abstraction of data storage.

Data structures are implemented using elements of underlying hardware, particularly random access memory and persistent data storage .

Question 6

It is not possible for replicas to have different versions of data. If they did, it would be impossible to ever restore consistency to the database.

True
Correct!
False

Question 7

Which design pattern provides some of the features of relational transactions?

Indexes
A TTL key pattern

Emulating tables
Correct Answer
Atomic Aggregate Pattern

Question 8

_____ is one or more keys or fields that exist in all documents in a collection that is used to separate documents into different partitions.

Correct!
shard key
Correct Answers
Shard Key

Question 9

Which of the following are TRUE about associative arrays (check all that apply) ?

Only keep data in memory
Correct!
Data structure like an array
Values are all the same type
Correct!
Generalize the idea of an ordered list indexed by an identifier to include arbitrary values for identifiers and values
Correct!
Not restricted to using integers as indexes or limiting values to the same type

Question 10

Which of the following is Not true about a namespace

A namespace is a logical data structure for organizing key-value pairs

Correct Answer

A namespace is used to improve relational database performance

You Answered

Namespaces are sometimes called buckets

Keys must be unique within a namespace

Question 11

What is the purpose of defining a naming convention for keys in key-value databases? (Select all that apply)

To introduce an additional level of abstraction in applications
To provide secondary indexes
To make it difficult for developers to create keys
Correct!
To help developers to document the types of values associated with the keys
Correct!
To allow developers to easily create keys

Question 12

Design pattern creates keys that use counters or sequences to generate new keys is _____Enumerable keys_____.

Question 13

Why would you want to denormalize collections in a document database?

TO reduce the potential for data anomalies

To reduce the amount of redundant data in the database
To evenly distribute documents over shards
C**orrect!**
To improve performance

Question 14

Which query operator is used to search for values in a single key?

$gte
$or
Correct!
$in
$let

Question 15

Which of the following statements is FALSE in regards to Vertical Partitioning

is particularly useful when you have some columns that are frequently accessed and others that are not.
Correct!
These parts of the database,known as shards, are stored on separate servers.
is a technique for improving database performance by separating columns of a relational table into multiple separate tables.
Vertical partitioning is typically used with relational tables because they have a fixed structure.

Question 16

What is range partitioning?

Correct!

Range partitioning works by grouping contiguous values and sending them to the same node in a cluster.

Range partitioning divides contiguous values and distributes them evenly across the cluster.

Range partitioning works by grouping contiguous values and sending the group to multiple nodes in a cluster.

Range partitioning divides contiguous values and uses a hash function to distribute them to nodes in the cluster.

Question 17

When would it make sense to have many indexes on your document collections?

When you do not have any ad hoc query requirements
Correct!
When you have Read-Heavy Applications
When you do not need to filter queries on large documents
When you have Write-Heavy Applications

Question 18

What factors typically drive the use of document databases by modelers and application developers? (Select all that apply)

Data normalization

Query optimization
Correct!
Flexibility
Correct!
Scalability

Question 19

Why might data modelers try to minimize the number of indexes in write-heavy applications? (Select all that apply)

Correct!
Indexes consume CPU, storage, and memory resources
Indexes improve query response time
Balancing competing interests is necessary for optimizing database performance
C**orrect Answer**
Fewer indexes correlate with faster updates

Question 20

What does Scaling out entail?
Increasing the network bandwidth
Correct!

Adding servers to a cluster.

Adding additional memory to upgrade the database server

Adding additional processors to upgrade the database server

—----------THE END—----------