

Study Guide

MarkLogic Professional Certification

Taking a Written Exam

General Preparation

Developer Written Exam Guide

Administrator Written Exam Guide

Example Written Exam Questions

Hands-On Exam Overview

Taking a Written Exam

In this section we will cover important general information about what to expect during your testing appointment, including what will happen before the exam begins, during the exam, and after the exam is completed.

- Before the Exam
 1. The exam proctor will direct you to a site where you will review the instructions for taking the exam.
 2. The exam Proctor will then validate your contact information, identity, and testing environment through the use of your web cam and government issued photo ID.
 3. Once your identity has been confirmed, the exam Proctor will provide you with an exam key which you must enter in order to begin. Exam keys are only valid for your specific testing event and will not be provided in advance. Without a valid exam key, the system will not enable you to take the exam.
 4. The final step before starting the exam is to review and accept the Terms and Conditions of the Certification Program.
- During the Exam
 1. Each exam will consist of 50 questions.
 2. Every exam will be different; however the distribution of questions across categories and the weight of questions within categories will be the same for each exam attempt.
 3. You will have 60 minutes to complete the exam, starting from the time you are presented the first question.
 4. Each question may be skipped once. Skipped questions will then be presented back to you after all other questions have been answered.
 5. Do not close the browser window until your exam is complete. If you accidentally close it, please contact the exam proctor and they can help you resume the exam, as long as there is time remaining.
 6. The back button on your browser will not function during the exam. Instead you may skip questions once and answer them at the end.
 7. The exam will automatically finish when you have either submitted answers for all the questions or time expires.
- After the Exam
 1. Upon completion of the exam, your results will be saved in the system, displayed on your screen, and sent to the email address provided when you entered your contact information.
 2. To pass the exam you must answer at least 37 out of 50 questions correctly.

3. You will be provided a breakdown of your score overall and by category. You will not be shown the exact questions that you answered correctly or incorrectly.
4. If successful in passing the exam, the next step is to enroll in a hands-on testing event.
5. If unsuccessful in passing the exam, you may retry by enrolling in another written exam testing event.

General Preparation:

To prepare for the MarkLogic Professional Certification program it is recommended that you have attended formal MarkLogic University training, performed your own research and self-study, and obtained real world professional experience using MarkLogic.

Resources that are available to you include:

- Free MarkLogic University Instructor led training and On Demand training:
 - <http://www.marklogic.com/services/training/>
- Developer Community resources, blogs and tutorials:
 - <http://developer.marklogic.com>
- Inside MarkLogic Server whitepaper:
 - <http://developer.marklogic.com/inside-marklogic>
- Documentation:
 - <http://docs.marklogic.com>

Developer Written Exam Details

In this section you will learn more about the MarkLogic Certified Developer written exam. You will learn about the category breakdown and some example topics within each category to help focus your studies.

Category	Key Topics
Architecture	<ul style="list-style-type: none"> • General understanding of key MarkLogic architecture components including: <ul style="list-style-type: none"> ○ Databases ○ Forests ○ Application servers ○ Hosts (E node / D node concepts) ○ Groups ○ Caches ○ Clusters • Merges • Rebalancing • Scalability
XPath & FLWOR	<ul style="list-style-type: none"> • XPath expressions, predicates, and axes • FLWOR expression logic • XQuery programming structure and syntax
Indexing	<ul style="list-style-type: none"> • Element and attribute range indexes • String range indexes and collation • Term List indexes • Reverse index • Tokenization and stemming concepts • Working with data across multiple languages and its impact on indexes • Reindexing

Search	<ul style="list-style-type: none"> • Search relevancy, query weighting, document weighting • Filtered versus Unfiltered search • Using CTS query constructors • Query tuning functions such as xdm:plan • Search across multiple languages
Data Modeling	<ul style="list-style-type: none"> • Using mlcp to ingest and transform documents • Loading and managing data using the Client REST API • Managing binary documents • Using fields • Data modeling choices and the impact on indexes and query resolution • Denormalization • Fragmentation
Transactions	<ul style="list-style-type: none"> • MVCC • Single statement transactions • Multi statement transactions • Distributed transactions • Point in time queries versus queries that require locks • Using xdm:invoke(), xdm:eval(), and xdm:spawn()
Security	<ul style="list-style-type: none"> • Roles and users • Application server security • Permissions • Privileges • Managing security on code in a modules database • Amps

Semantics	<ul style="list-style-type: none">• Loading and managing RDF data with mlcp• Data modeling of triples in MarkLogic• SPARQL query language• Named graphs and collections• Using the sem: XQuery functions• Using the CTS semantics functions (ex: cts:triples)
-----------	--

Administrator Written Exam Details

In this section you will learn more about the MarkLogic Certified Administrator written exam. You will learn about the category breakdown, number of questions per category, and some example topics within each category to help focus your studies.

Category	Key Topics
Architecture	<ul style="list-style-type: none"> • General understanding of key MarkLogic architecture components including: <ul style="list-style-type: none"> ○ Databases ○ Forests ○ Application servers ○ Hosts (E node / D node concepts) ○ Groups ○ Caches ○ Clusters • Merges • Rebalancing • Scalability
Configuration and Deployment	<ul style="list-style-type: none"> • Installing, licensing and initializing a MarkLogic instance • Using the Management and Packaging REST APIs • Installing and configuring CPF • Managing merge policy • Configuring groups • Configuring caches
Indexing	<ul style="list-style-type: none"> • Element and attribute range indexes • String range indexes and collation • Term List indexes • Reverse index • Tokenization and stemming concepts

	<ul style="list-style-type: none"> • Working with data across multiple languages and its impact on indexes • Reindexing
Search	<ul style="list-style-type: none"> • Search relevancy • Impacting relevancy through managing document quality and word query configurations • Filtered versus Unfiltered search
Data Modeling	<ul style="list-style-type: none"> • Using mlep to ingest and transform documents • Loading and managing data using the Client REST API • Managing binary documents • Configuring fields • Data modeling choices and the impact on indexes and query resolution • Denormalization • Fragmentation
Transactions	<ul style="list-style-type: none"> • MVCC • Single statement transactions • Multi statement transactions • Distributed transactions • Point in time queries versus queries that require locks
Security	<ul style="list-style-type: none"> • Roles and users • Application server security • Permissions • Privileges • Managing security on code in a modules database • Amps

High Availability (HA) and Disaster Recovery (DR)	<ul style="list-style-type: none">• Backup and restore• Journal Archiving and point in time recover• Local disk failover• Shared disk failover• Database replication• Distributed transactions
Monitoring	<ul style="list-style-type: none">• Using the Management REST API• Using and managing the Access Log, Error Log and Audit Log• Trace events• The Meters database and monitoring history• Identifying poorly performing queries

Example Questions:

In this section you will see a few sample questions so you have a better feel for how the questions are structured. All questions are multiple choice. Some questions are scenario driven, and require you to analyze a scenario and answer questions based on that scenario. Other questions are designed to validate your knowledge of core principles that relate to the specific category.

Question Text: Which of the following features directly benefit from Range Indexes?

Answers: Snippets
Word query
Element value query
Sorted value list

Question Text: Which of the following search types is both fast and accurate?

Answers: Unfiltered search with the default indexes
Unfiltered search with all of the required indexes
Filtered search with the default indexes
Filtered search with all of the required indexes

Question Text: How do you configure a host as E node?

Answers: Set it as E node at group level
Set up an application server on this host to which user requests are sent
Handle it at the application code level
Either of 1st or 2nd approach

Scenario Information:

Scenario Image:

```
▼<?xml version="1.0" encoding="UTF-8"?>
▼<doc>
  ▼<name>ashutosh</name>
  ▼<date>2015-01-01</date>
  ▼<desc>MarkLogic Trainer</desc>
</doc>


---


▼<?xml version="1.0" encoding="UTF-8"?>
▼<doc>
  ▼<name>tamas</name>
  ▼<date>2015-01-01</date>
  ▼<desc>MarkLogic trainer</desc>
</doc>
```

Scenario Text: Assume that database has only above two documents and fast case sensitive index is set to false.

Question Information:

Question Text: Which of the following query would fetch only one document?

Answers:

- cts:search(doc(),"Trainer","unfiltered")
- cts:search(doc(),"Trainer")**
- cts:search(doc(),"trainer")
- cts:search(doc(),"trainer","unfiltered")

Question Text: How many content databases does a typical MarkLogic application need?

Answers:

- 2
- Depends upon the application scenario
- As many as you want
- 1**

Question Text:	Which of the following is NOT configured at group level?
Answers:	in-memory sizes cache sizes failover enable http timeout

Scenario Information:	
Scenario Image:	
<pre> ▼ <doc> ▼ <desc>Lovers paradise, Paris!!!</desc> </doc> </pre>	
Scenario Text:	Assume that database indexes are set to the default settings and your database has the above document.
Question Information:	
Question Text:	Which of the following would be found in the universal term index
Answers:	paradise, Paris Paris!!! <desc>Paris!!!</desc>

Question Text:	Which of the following is NOT a transaction related feature of the database?
Answers:	MVCC Failover Journal xdmp:update()

Question Text:	Which of the following can a MarkLogic user be directly assigned to?
Answers:	Execute privileges for application code URI read privileges Roles All of the above

Question Text: Which of the following operators are by default understood by Search API grammar?

Answers:

- and
-
- OR

Answer 2 and 3

Question Text: What is the maximum number of forests that should be configured on a host machine which has 1 TB of storage and 2 quad core CPUs assuming you cannot add more storage to it?

Answers:

- Two
- Three**
- Four
- Five

Question Text: Which is the valid state of forest if local disk failover is configured?

Answers:

- SYNC Replicating
- ERROR
- OPEN

All of the above

Question Text: Which is NOT the data modeling best practice?

Answers:

- Denormalize
- Use collections
- Avoid fragmentation
- Document Joins**

Question Text: Which is NOT a valid mlcp parameter?

Answers: host
input_file_type
output_file_type
directory_filter

Question Text: Which transaction mode never acquires any locks?

Answers: Locks can be acquired in any transaction mode depending upon the application code
Update
Query
Auto

Scenario Information:

Scenario Image:

```
1 declare option xdmp:transaction-mode 'query';  
2  
3 let $a := 1  
4 let $b := 2  
5 return  
6 if ( $a = $b )  
7 then xdmp:document-insert("/sample-song.xml", <top-song/>)  
8 else "Not Inserted"
```

Scenario Text: Suppose you run the above code from query console

Question Information:

Question Text: What would be the output?

Answers: Error is thrown—XDMP:UPDATEFUNCTIONFROMQUERY
Document is inserted in DB
"Not Inserted"
Throws a syntaxerror

Question Text: Which file log level is most appropriate in production environment?

Answers: Error
Fine
Info
Debug

Question Text: What is NOT true about meters database?

Answers: **It is used to store performance data at database level**
It is used to store performance data at group level
Performance data retention policy can be applied for meters database
History data is available on port 8002

Question Text: Which of the following would you use for writing alerting applications?

Answers: Co-occurrence query
Matchmaking query
Reverse query
Registered query

Question Text: What are the valid transaction modes in the list below?

Answers: Update
Auto
Default
Answer 1 and 2

Scenario Information:

Scenario Image:

```

1 declare option xdm:transaction-mode 'update';
2
3 let $a := 1
4 let $b := 1
5 return
6 if ( $a = $b )
7 then xdm:document-insert("/sample-song.xml", <top-song/>)
8 else "Not Inserted"

```

Scenario Text: Suppose you run the above code from query console

Question Information:

Question Text: What would be the output?

Answers:

- Server Error is thrown
- Empty sequence and document is inserted in DB
- Empty sequence and document is not inserted in DB**
- Not Inserted

Question Text: What is correct about triple indexes and data?

Answers:

- Triple indexes need to be memory mapped
- There cannot be more than 100 embedded triples in a single document
- Triple indexes are physically not stored in the same forest as other indexes
- Triple index utilize dynamic caches set at group level**

Question Text: What type of constraint do you configure for bucketed facets in search API?

Answers:

- Word
- Date
- Range**
- None of the above

Question Text: How can you search within the text of attribute nodes?

Answers: cts:word-query
cts:element-word-query
cts:word-match
None of the above

Question Text: Which of the following statement is NOT true?

Answers: Search API has simple search grammar which can be extended
Search API internally utilizes CTS functions
Search API by default does unfiltered search
REST API search endpoint returns search:response element

Scenario Information:

Scenario Image:

```
▼<?xml version="1.0" encoding="UTF-8"?>
▼<article>
  ▼<p>this article mentions about mlu</p>
  ▼<p>mlu has multiple trainers located worldwide</p>
  </article>

```

```
▼<?xml version="1.0" encoding="UTF-8"?>
▼<article>
  ▼<p>this article mentions about mlu</p>
  ▼<p>marklogic offers free online instructor led training worldwide</p>
  </article>

```

Scenario Text: Suppose you have above two documents in your database

Question Information:

Question Text: What would be the output of the query
cts:search(/, cts:near-query(("mlu", "worldwide"), 8))

Answers: First article is returned
Second article is returned
Both the articles are returned
It is not a valid query

Question Text: What does . (dot) in an XPATH expression represent?

Answers: Descendent node in any namespace
Child node in any namespace
. (dot) is not permitted in XPATH
Self node

Question Text: How many valid item() are there in sequence (1, 2, <a/>, 1, "hello world")

Answers: 3
4
5
6

Question Text: Which of the following is NOT a valid collection name for storing triples?

Answers: <http://marklogic.com/semantics#default-graph>
<http://marklogic.com/semantics#music-graph>
mytriplecollection
All three names are valid collection names for storing triples

Question Text: Which of the following could be used in marklogic for inference using triples data?

Answers: Chain Based
Path Based
Rule Based
Answer 2 and 3

Question Text: Which of the following is not implemented at the Forest level?

Answers: Journal
Range Index
List Cache
Expanded Tree Cache

Question Text: A deleted document is no longer searchable as soon as:

Answers: **the update transaction commits**
the journal is written to the disk
a merge removes the content with a deleted timestamp
the replica forest is updated

Question Text: Which of the following may benefit from a range index but does not require one?

Answers: Faceting
Order by
Date greater than
Element Value lookup

Scenario Information:

Scenario Image:

```
Documents:
<doc1>
  <name>Joe User</name>
</doc1>

<doc2 role="user">
  <id>djones</id>
  <last-login>2011-03-14T09:27:35</last-login>
</doc2>

<doc3>
  <title>Account Policy</title>
  <text>Users must keep credentials private...</text>
</doc3>

Query:
cts:search(/, cts:word-query("user"))
```

Scenario Text:

Assume these documents have been loaded into a database that has no additional indexes other than the universal index, and you run this query against that database in Query Console.

Question Information:

Question Text: Which of the following statements is true?

Answers:

- doc2 will not be in the results because word-queries ignore attribute values**
- doc2 would appear in the results if there were an element-attribute-range-index on /doc2/@role
- doc2 will appear in the results set
- the query will fail because the document roots do not match

Hands-On Exam Overview

After successful completion of the written exam you will be eligible to take the MarkLogic Professional Certification hands on exam.

The hands-on exam is deliverable-based. You will access a hosted virtual machine which will have all the necessary software, data, and instructions required to complete the hands on exam. You will have up to 4 hours to complete the deliverables according to the requirements defined in the instructions.

The exam proctor will provide you details on how to access the virtual machine and the instructions during your hands on testing session. All your deliverables will be saved to the virtual machine. MarkLogic representatives will evaluate your submissions and provide your results within three (3) weeks of completing the hands on exam.