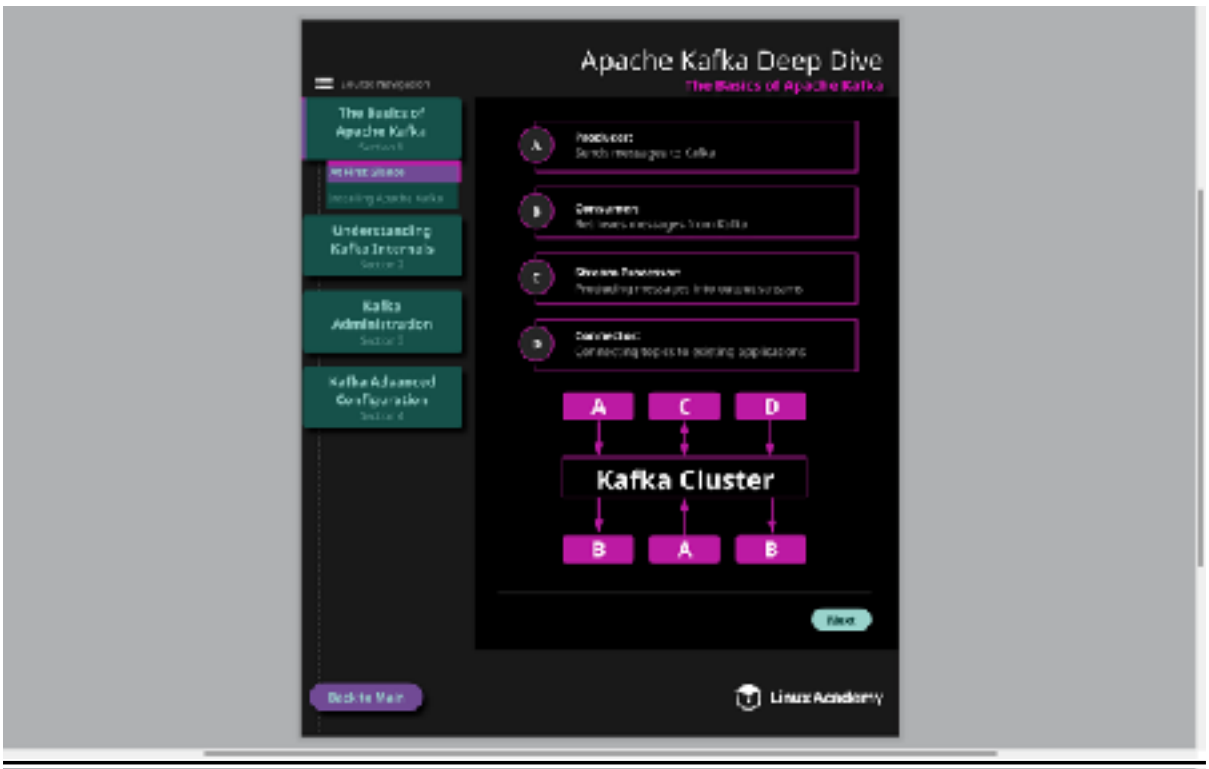
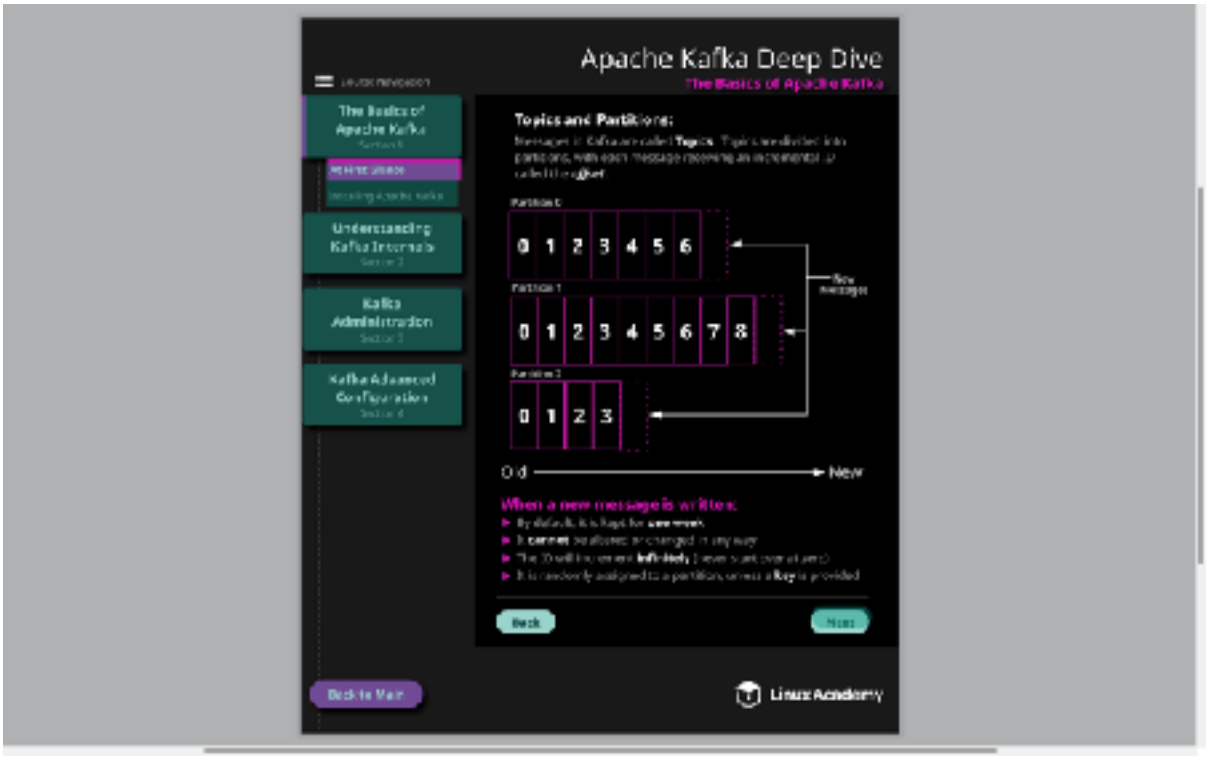
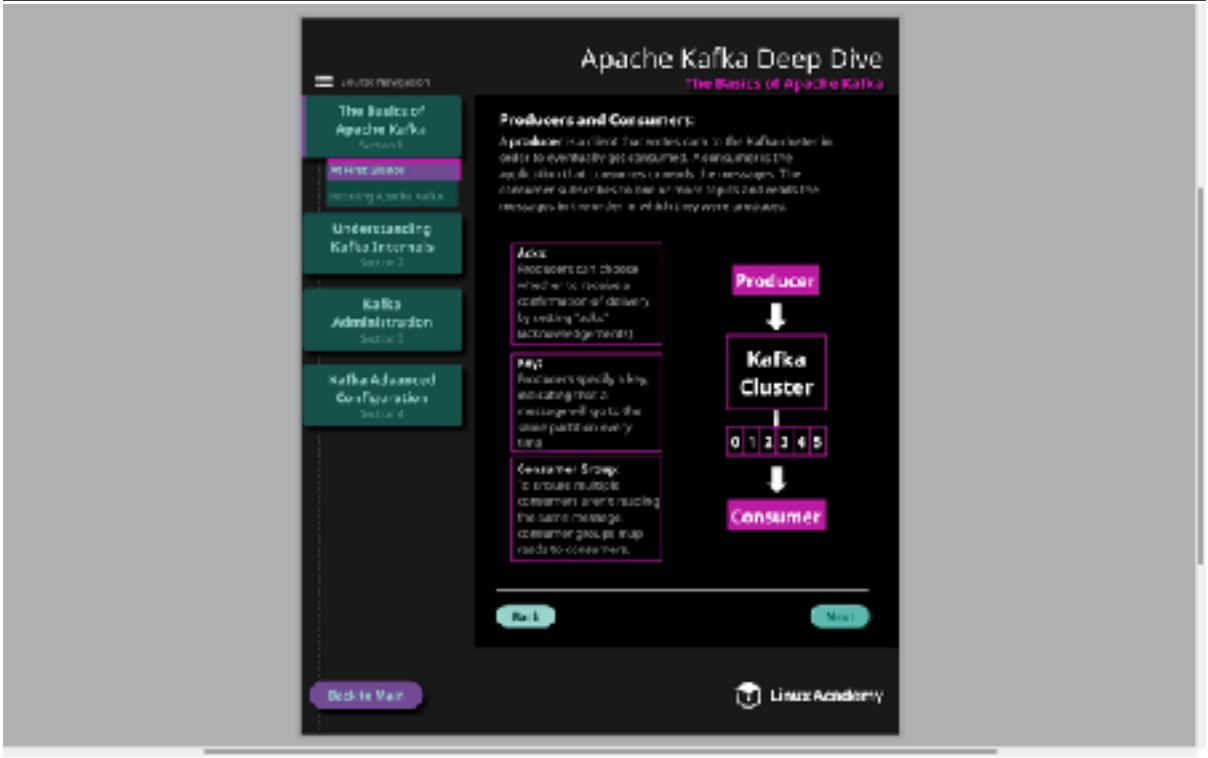
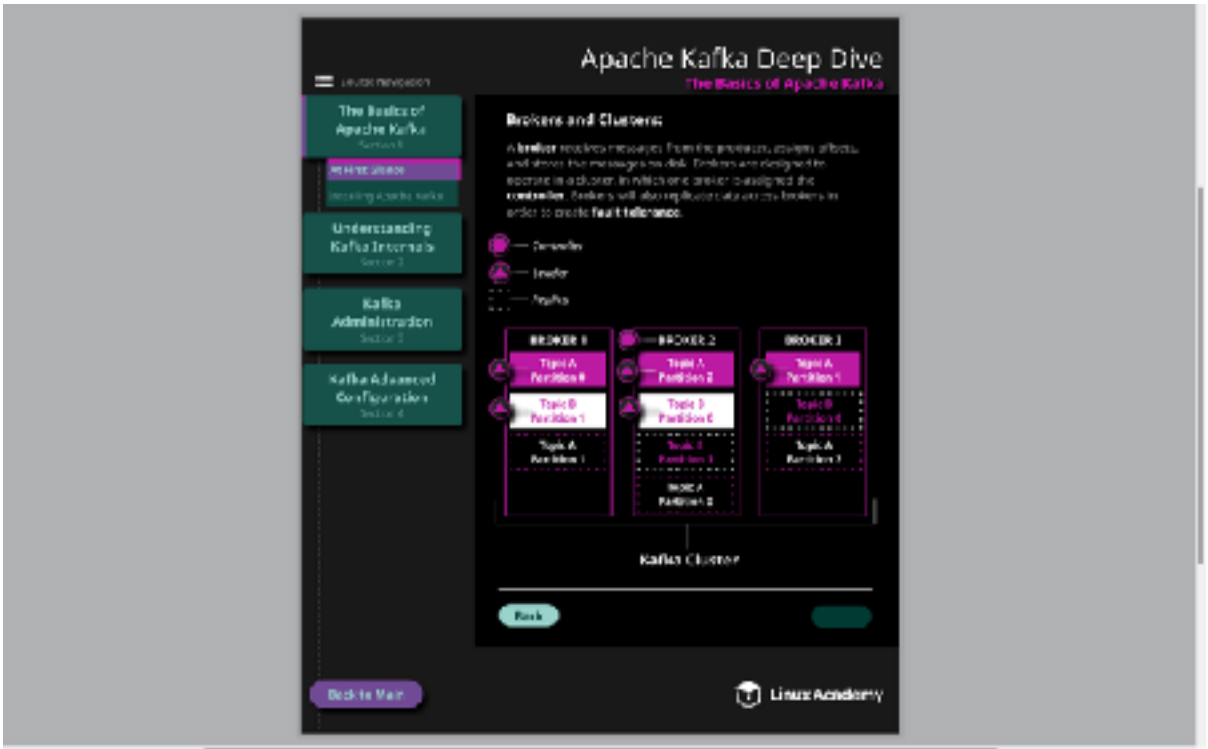
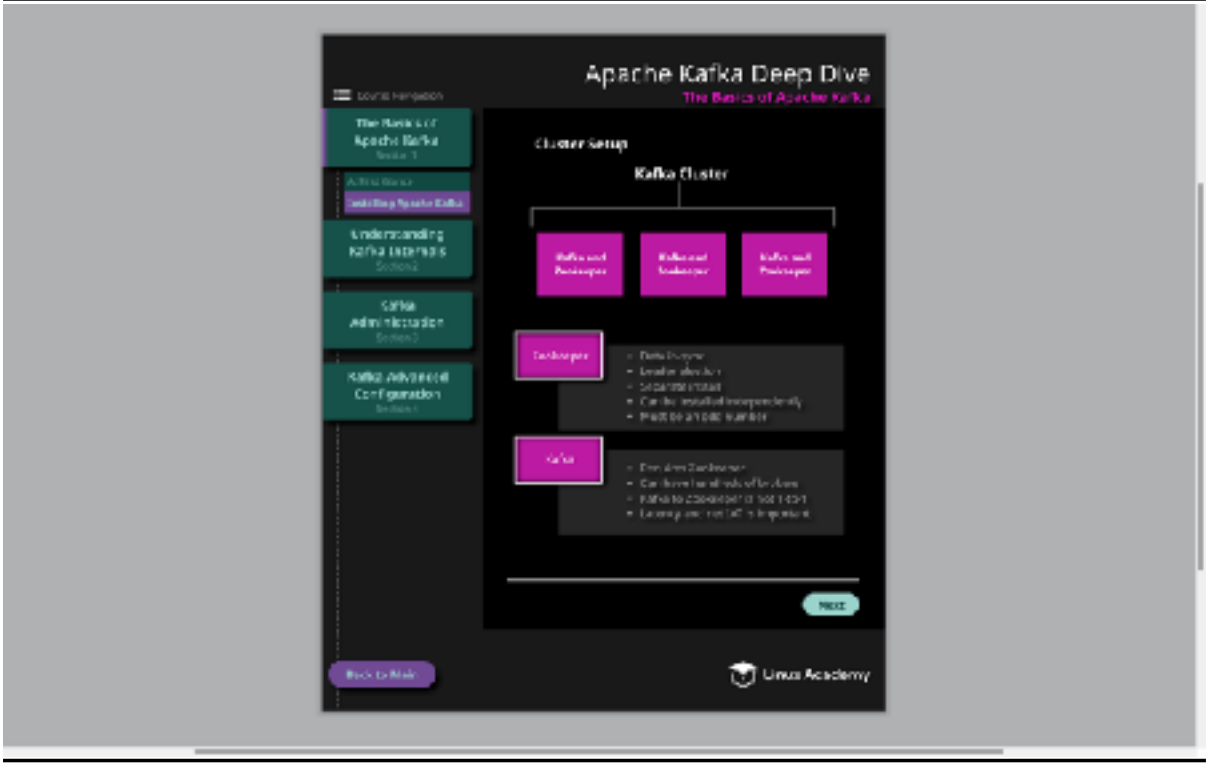
Kafka Linux Academy Basics



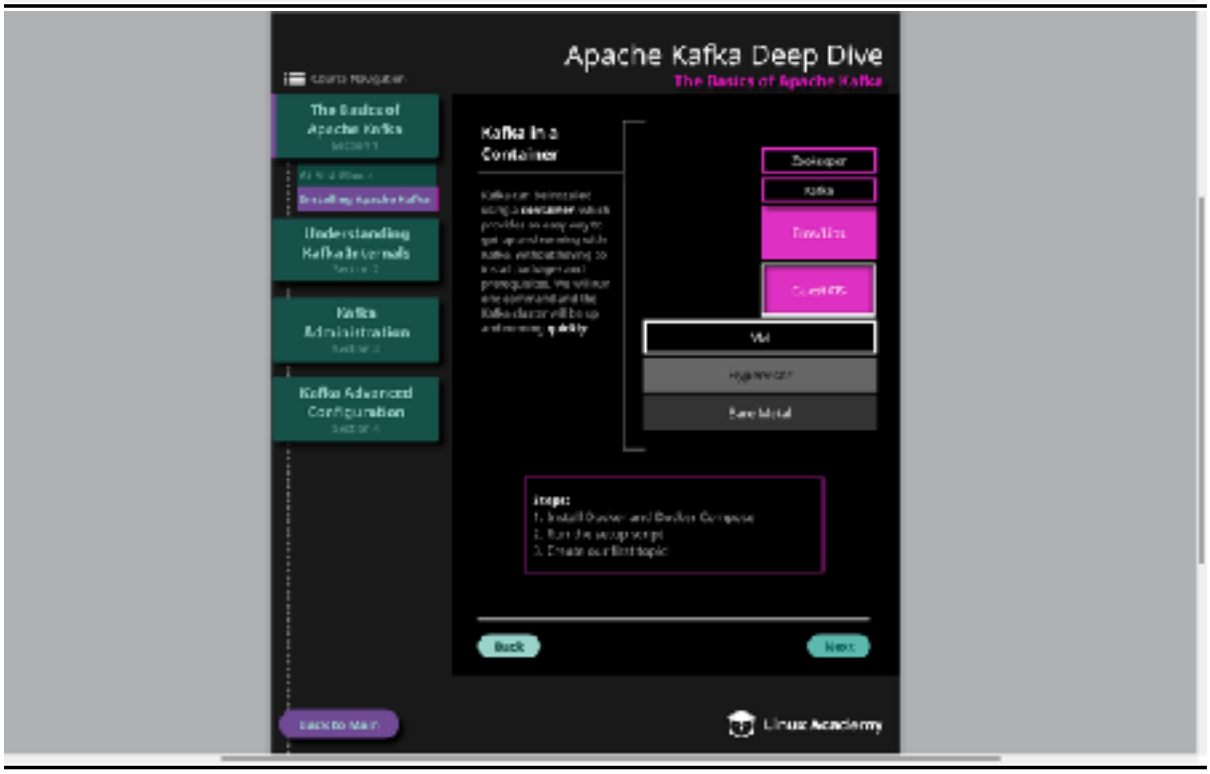

Apache Kafka Deep Dive
The Basics of Apache Kafka
Level navigation
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configurations
Topics and Partitions:
Partition 0
0 1 2 3 4 5 6
Partition 1
0 1 2 3 4 5 6 7 8
New messages
Partition 2
0 1 2 3
Old
New
When a new message is written:
Back
Next
Back to Main
Linux Academy

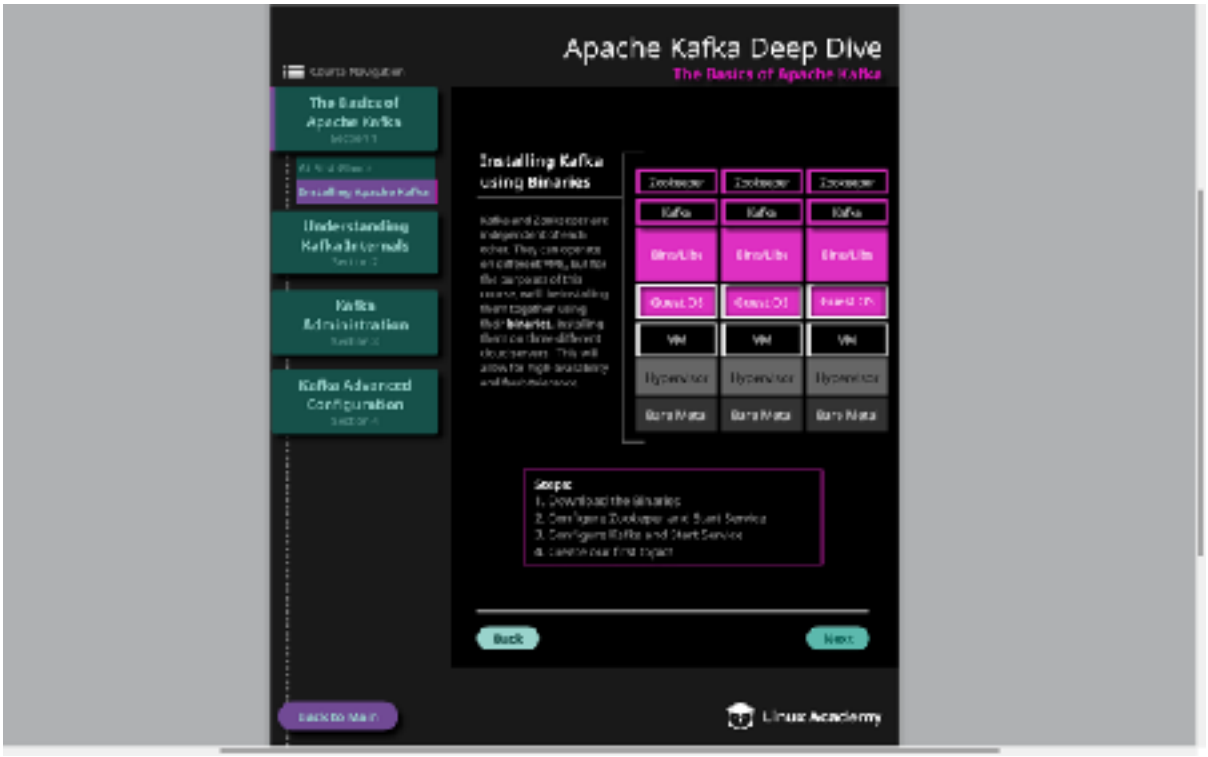

Apache Kafka Deep Dive
The Basics of Apache Kafka
Level navigation
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configurations
Producers and Consumers:
Acks:
Key:
Consumer Group:
Producer
Kafka Cluster
0 1 2 3 4 5
Consumer
Back
Next
Back to Main
Linux Academy


Apache Kafka Deep Dive
The Basics of Apache Kafka
Brokers and Clusters
Kafka Cluster
Linux Academy

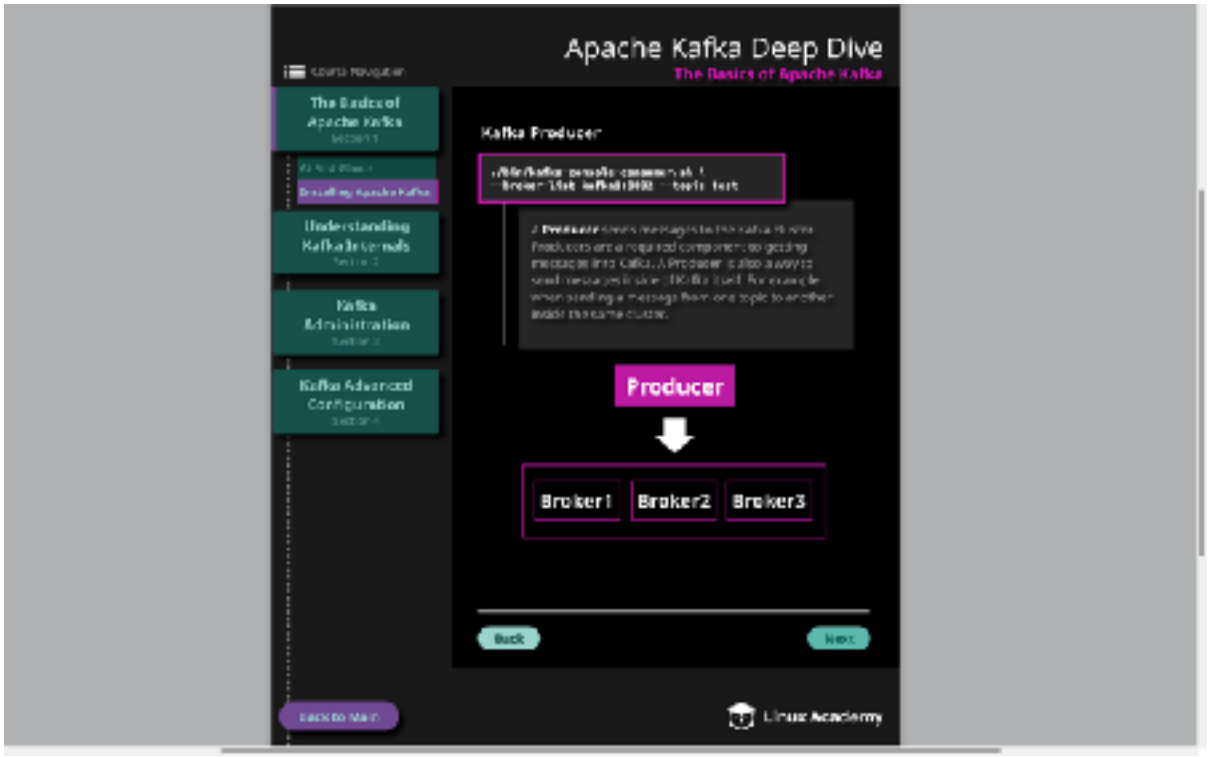

Apache Kafka Deep Dive
The Basics of Apache Kafka
Cluster Setup
Kafka Cluster
Linux Academy

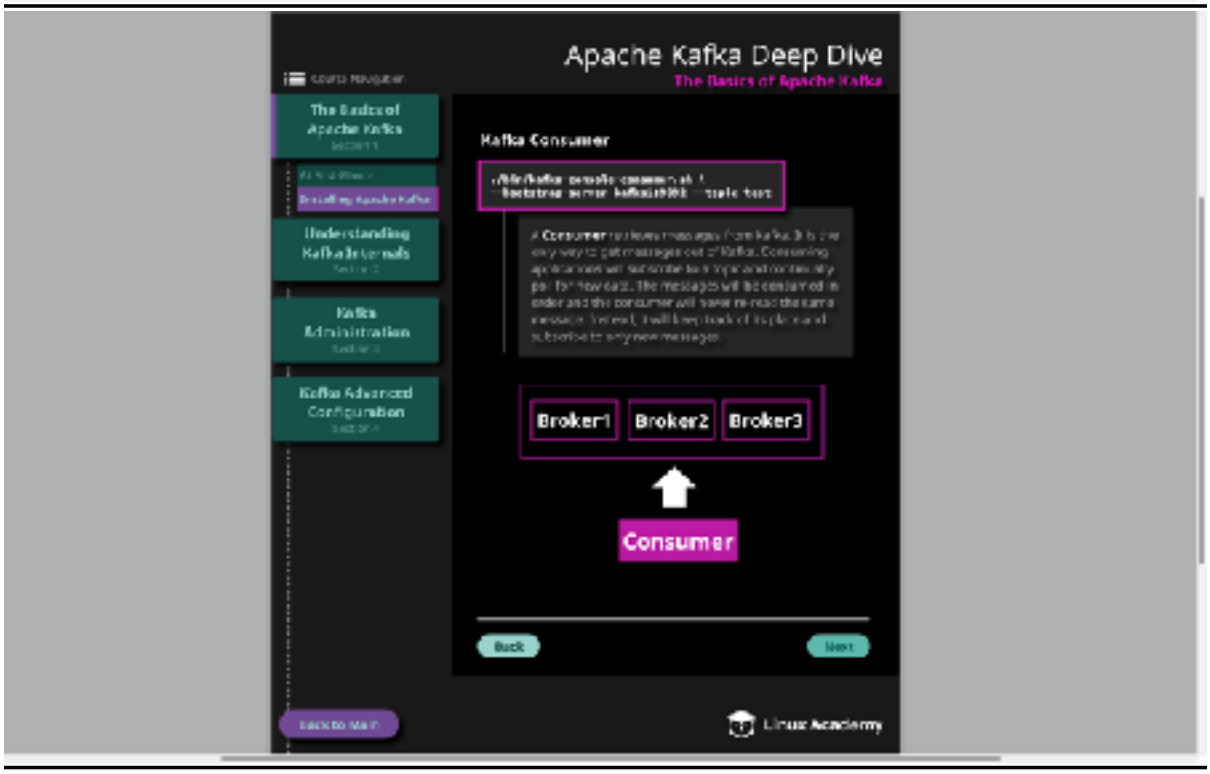

Apache Kafka Deep Dive
The Basics of Apache Kafka
Kafka in a Container
Linux Academy


Apache Kafka Deep Dive
The Basics of Apache Kafka
Kafka in a Container
Linux Academy

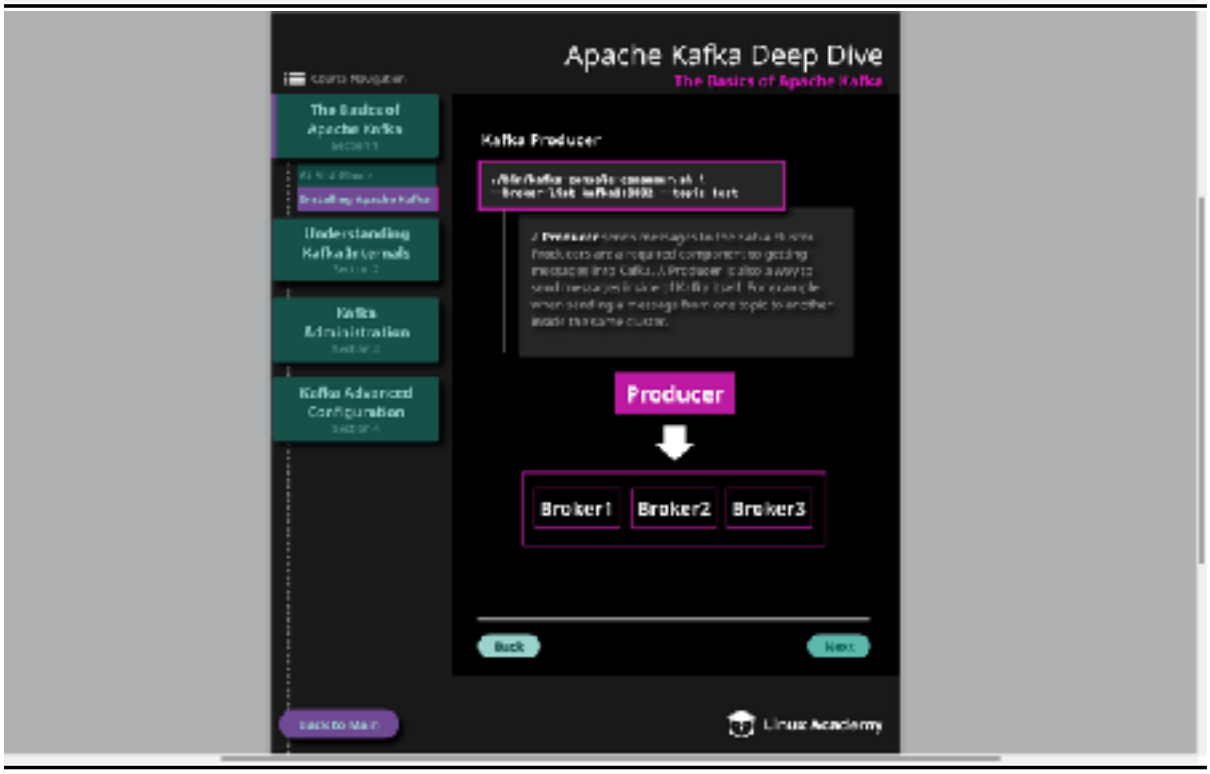

Apache Kafka Deep Dive
The Basics of Apache Kafka
Course Navigator
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Back to Main
Installing Kafka using Binaries
Zookeeper
Kafka
Bins/Libs
Guest OS
VM
Hypervisor
Bare Metal
Scope
1. Download the binaries
Buck
Next
Linux Academy

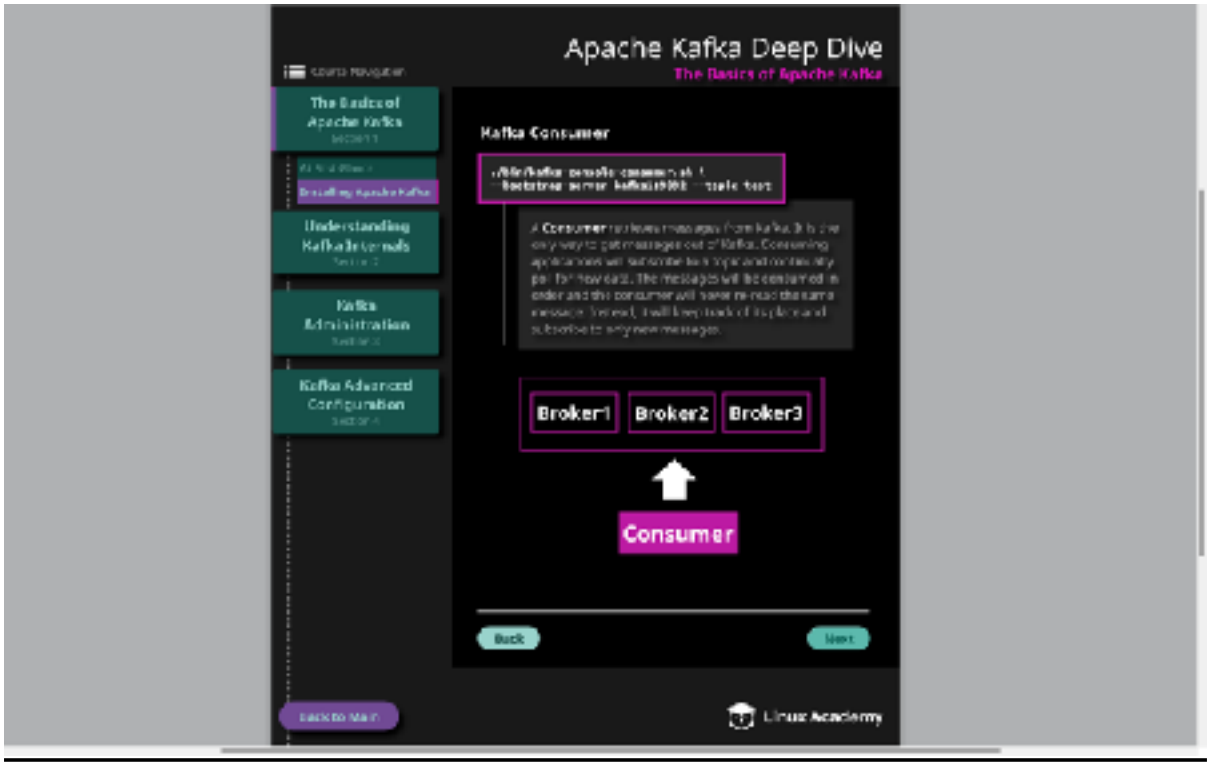

Apache Kafka Deep Dive
The Basics of Apache Kafka
Course Navigator
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Back to Main
Breaking Down the Commands
--zookeeper
--create
--topic
--partitions
--group
Buck
Next
Linux Academy


Apache Kafka Deep Dive
The Basics of Apache Kafka
Course Navigation
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Back to Main
Kafka Producer
Producer
Broker1
Broker2
Broker3
Back
Next
Linux Academy


Apache Kafka Deep Dive
The Basics of Apache Kafka
Course Navigation
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Back to Main
Kafka Consumer
Broker1
Broker2
Broker3
Consumer
Back
Next
Linux Academy

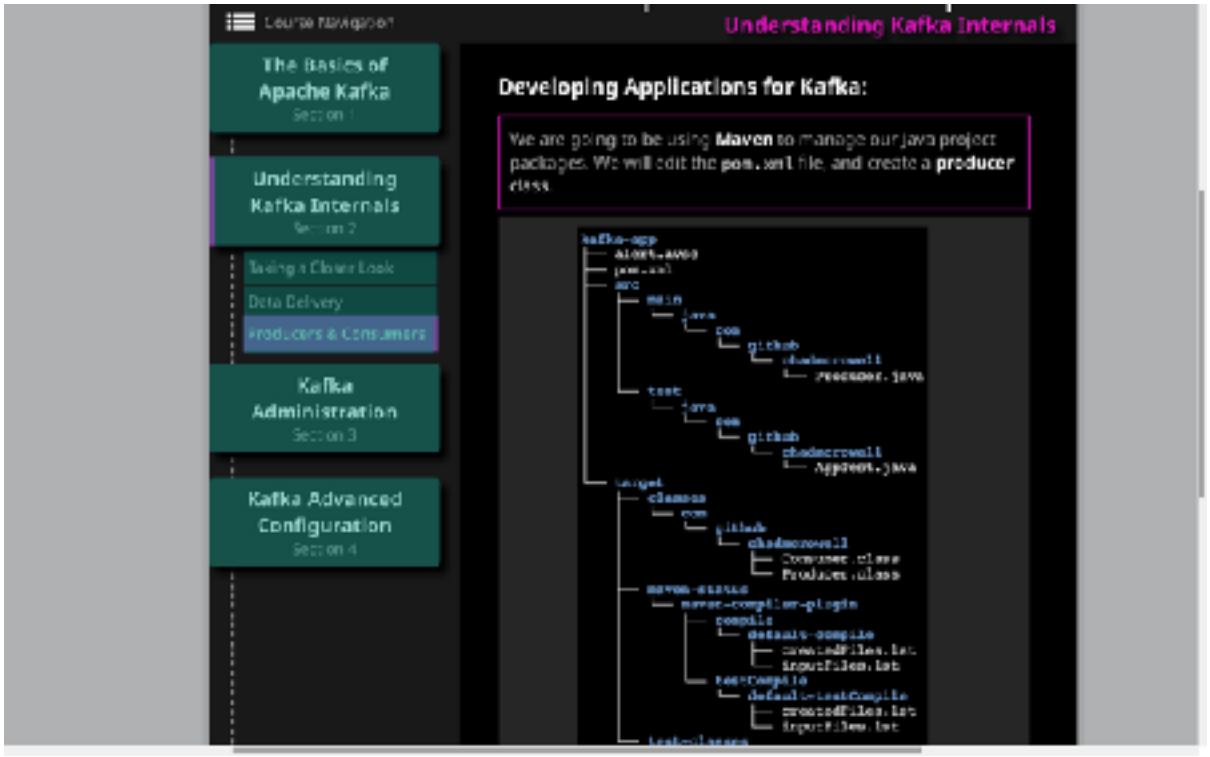

Apache Kafka Deep Dive
The Basics of Apache Kafka
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Kafka Consumer Group
Broker1
Broker2
Broker3
Consumer
Consumer
Consumer
Back
Next
Back to Main
Linux Academy

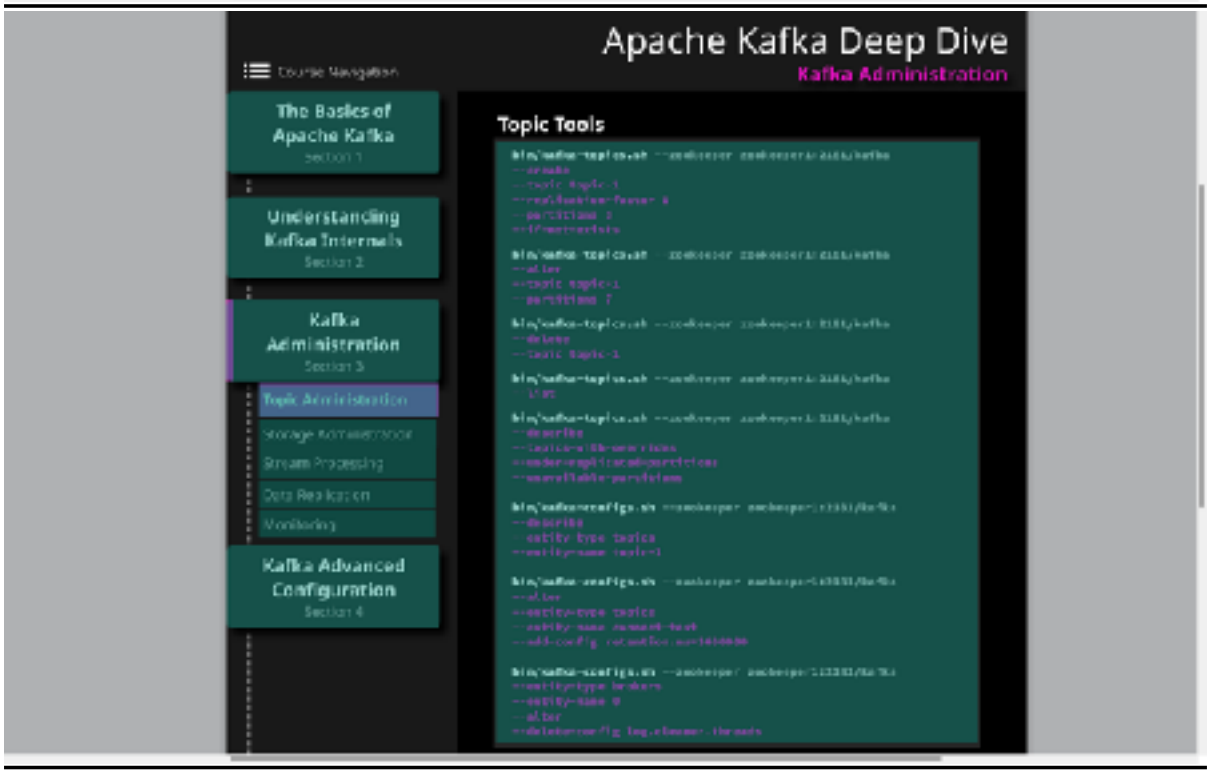

Apache Kafka Deep Dive
The Basics of Apache Kafka
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Kafka Producer
Producer
Broker1
Broker2
Broker3
Back
Next
Back to Main
Linux Academy


Apache Kafka Deep Dive
The Basics of Apache Kafka
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Kafka Consumer
A Consumer
Broker1
Broker2
Broker3
Consumer
Back
Next
Back to Main
Linux Academy


Apache Kafka Deep Dive
Understanding Kafka Internals
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Handling Requests
Broker
Acceptor Thread
Processor Thread
IO Thread
Processor Thread
Producer or Consumer
Acceptor Thread
Processor Thread
IO Thread
Request Queue
Response Queue
Back
Next

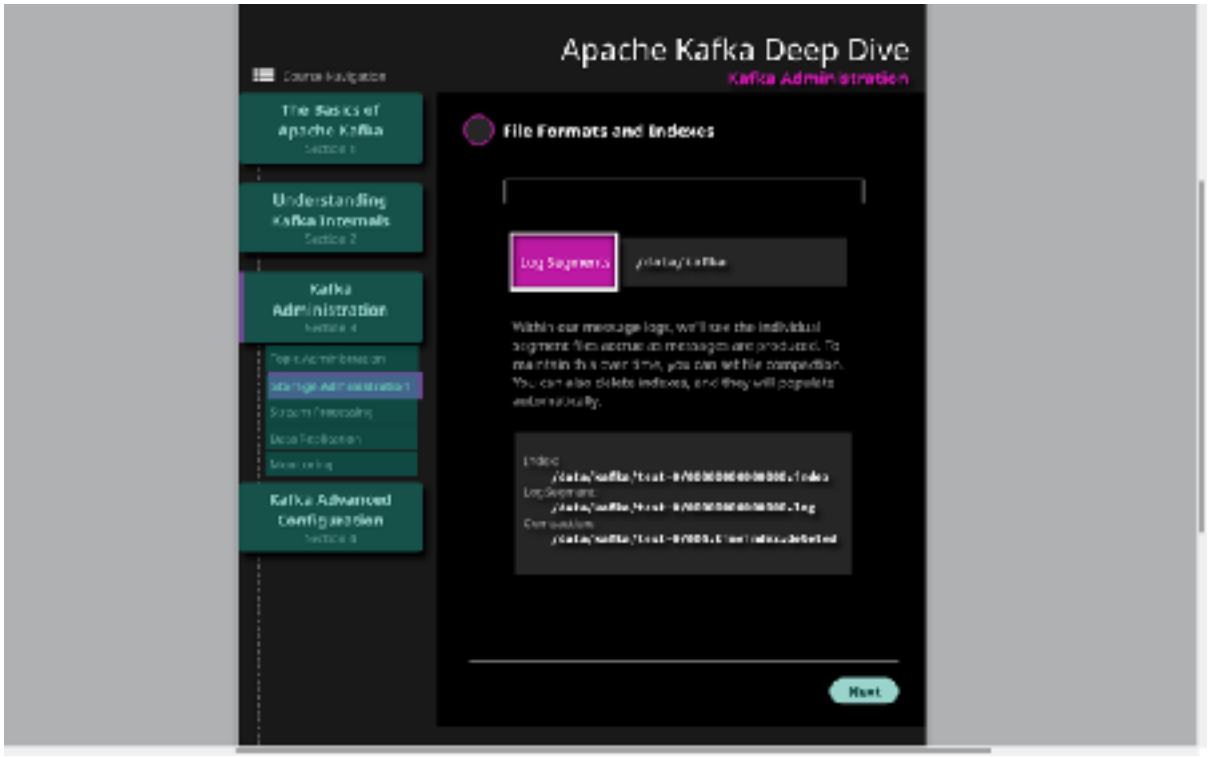

Apache Kafka Deep Dive
Understanding Kafka Internals
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Partitions
Broker
TOPIC 1
Partition 0
Partition 1, Replica 1
Partition 2, Replica 2
Back

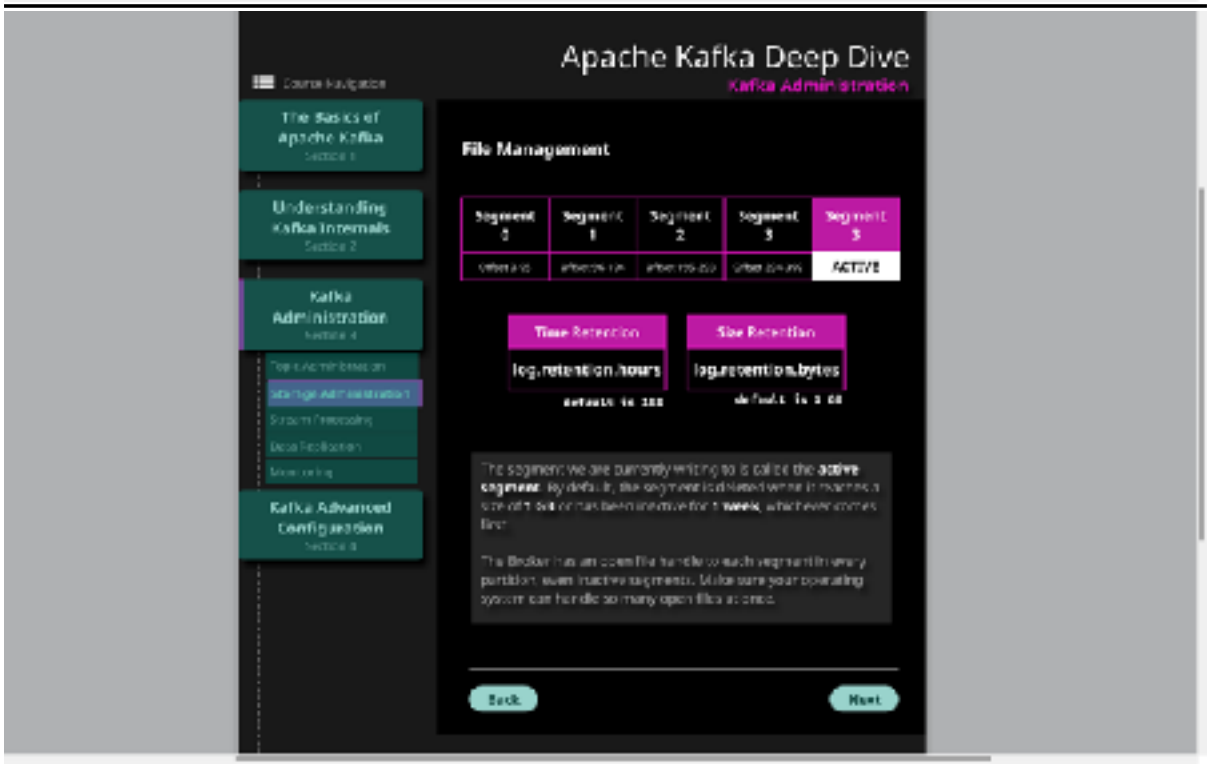

Course Navigation
The Basics of Apache Kafka
Section 1
Understanding Kafka Internals
Section 2
Taking a Closer Look
Data Delivery
Producers & Consumers
Kafka Administration
Section 3
Kafka Advanced Configuration
Section 4
Understanding Kafka Internals
Developing Applications for Kafka:
We are going to be using Maven to manage our java project packages. We will edit the pom.xml file, and create a producer class

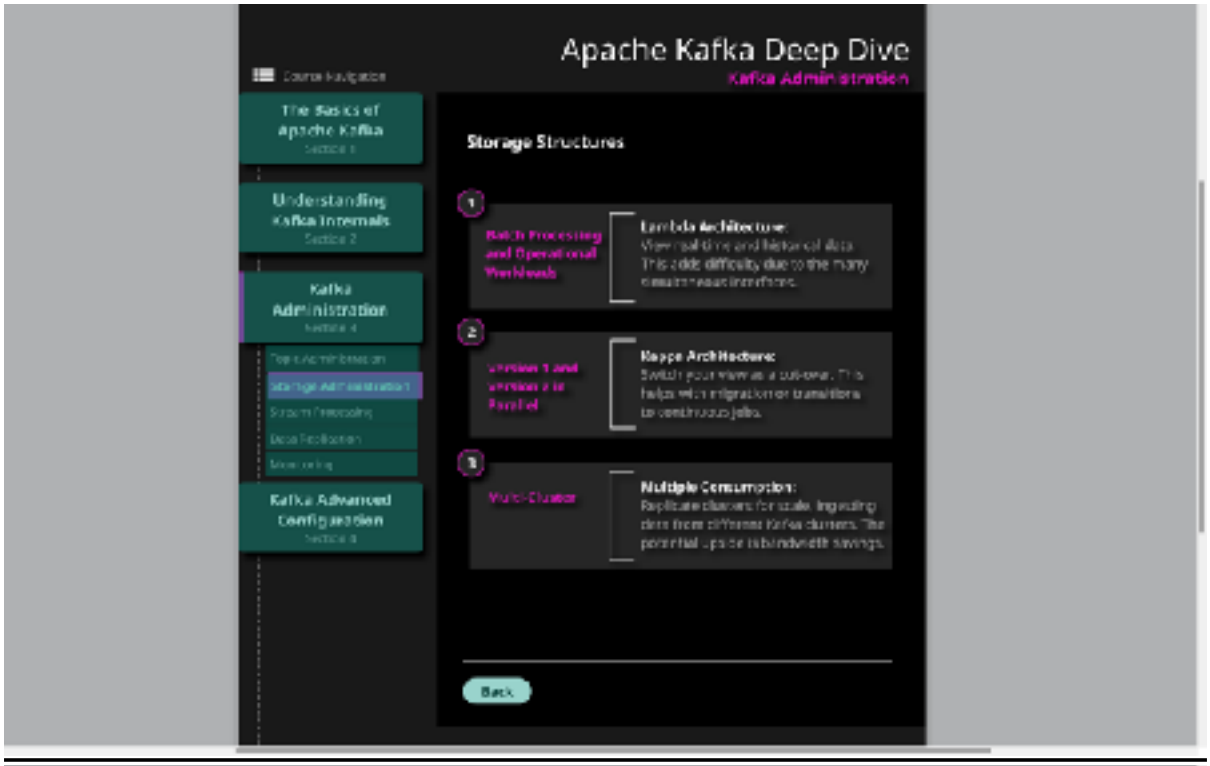

Course Navigation
Apache Kafka Deep Dive
Kafka Administration
The Basics of Apache Kafka
Section 1
Understanding Kafka Internals
Section 2
Kafka Administration
Section 3
Topic Administration
Storage Administration
Stream Processing
Data Replication
Monitoring
Kafka Advanced Configuration
Section 4
Topic Tools

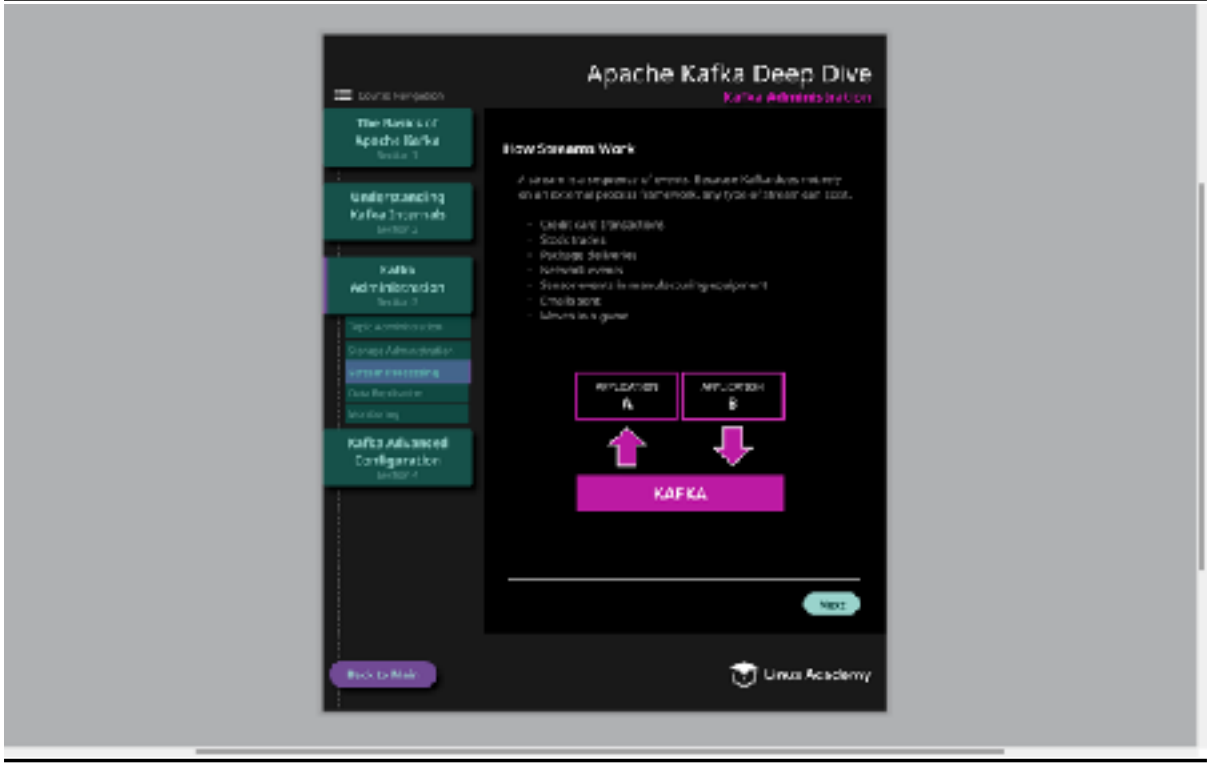

Apache Kafka Deep Dive
Kafka Administration
Course Navigator
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Topic Configurations
Back
Next

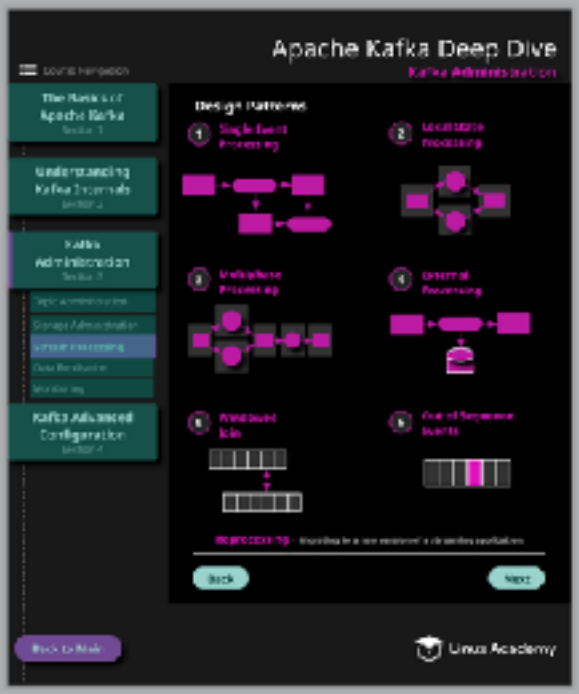

Apache Kafka Deep Dive
Kafka Administration
Course Navigator
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Message Behavior
A
B
Line Racing:
C
Back


Apache Kafka Deep Dive
Kafka Administration
Course Navigation
The Basics of Apache Kafka
Section 1
Understanding Kafka Internals
Section 2
Kafka Administration
Section 4
Kafka Advanced Configuration
Section 5
File Formats and Indexes
Log Segments
/data/kafka
Next

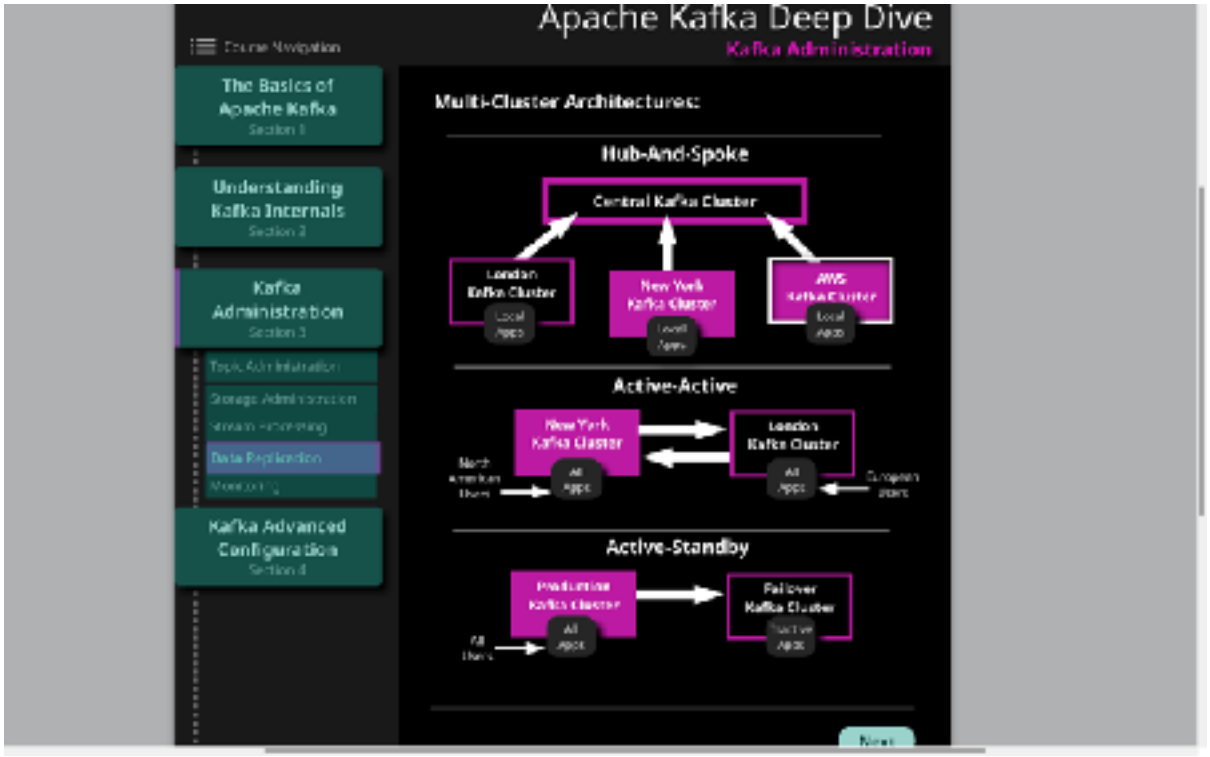

Apache Kafka Deep Dive
Kafka Administration
Course Navigation
The Basics of Apache Kafka
Section 1
Understanding Kafka Internals
Section 2
Kafka Administration
Section 4
Kafka Advanced Configuration
Section 5
File Management
Segment 0
Segment 1
Segment 2
Segment 3
Segment 4
ACTIVE
Time Retention
log.retention.hours
Size Retention
log.retention.bytes
Back
Next

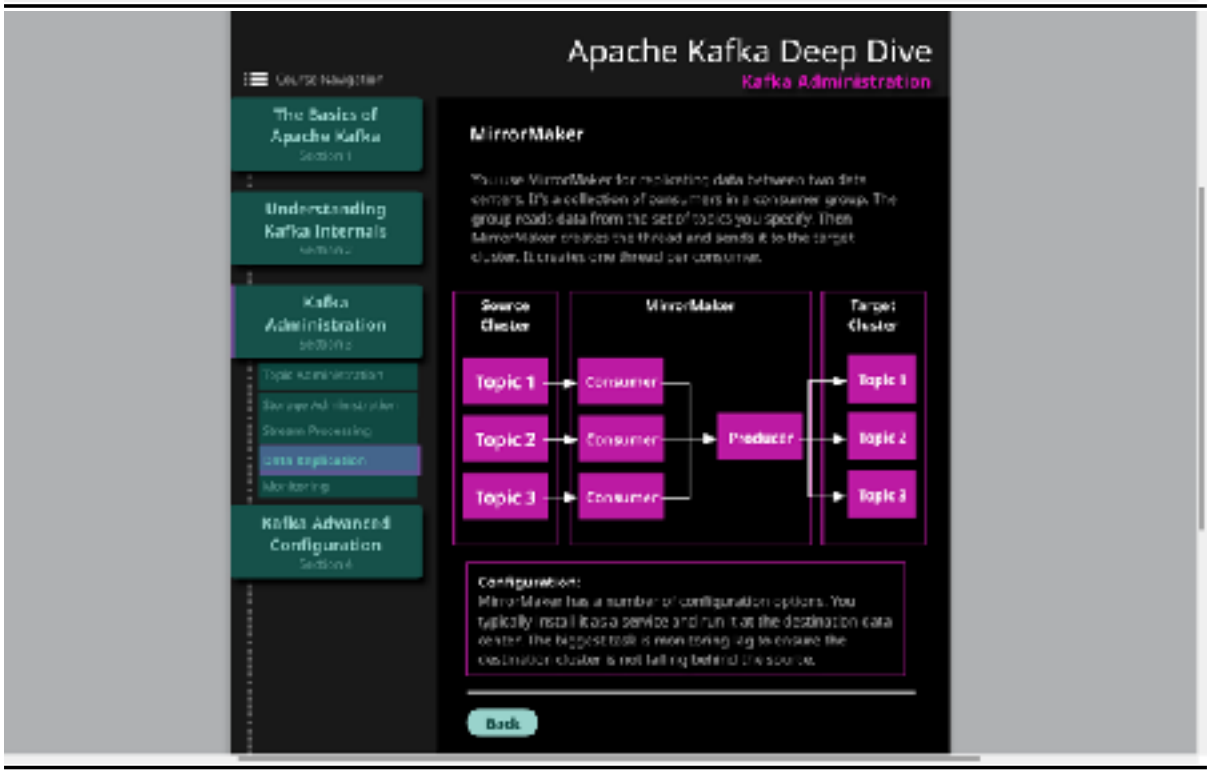

Apache Kafka Deep Dive
Kafka Administration
Course Navigation
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Storage Structures
Batch Processing and Operational Workloads
Lambda Architecture:
Version 1 and version 2 in Parallel
Kappa Architecture:
Multi-Cluster
Multiple Consumption:
Back

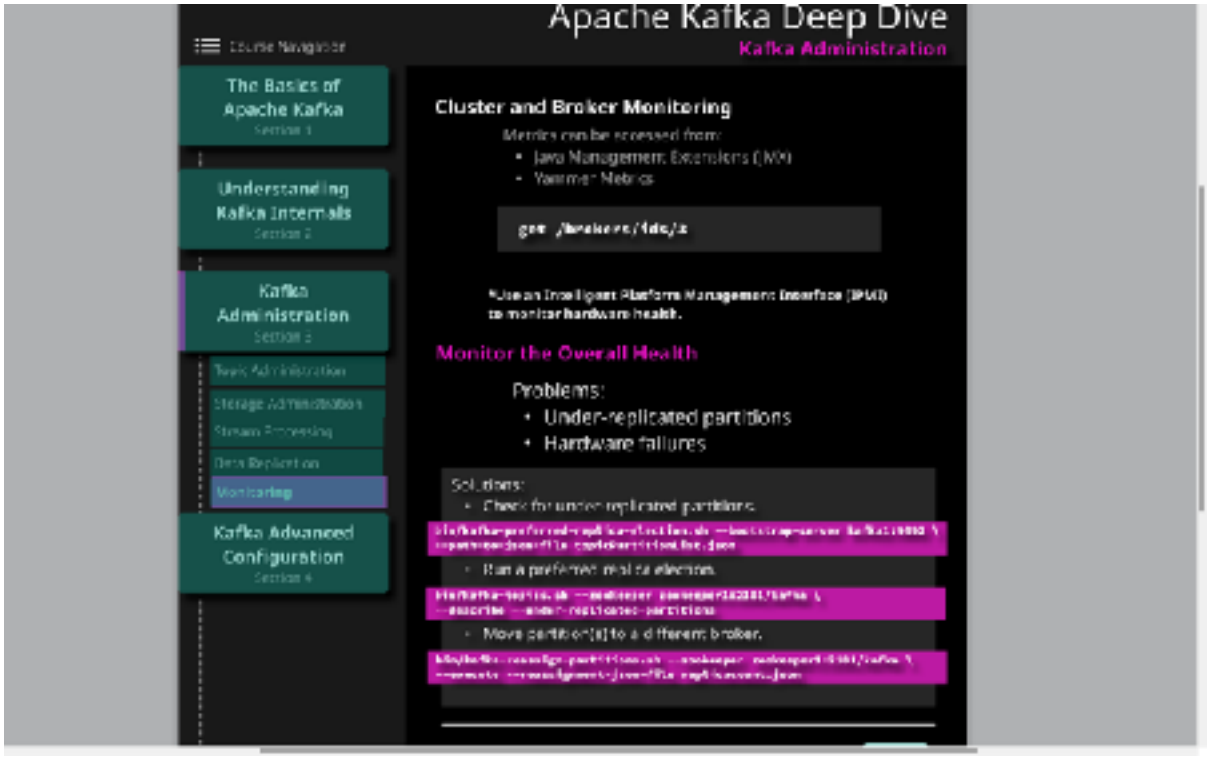

Apache Kafka Deep Dive
Kafka Administration
Course Navigation
How Streams Work
KAFKA
Next
Back to Main
Linux Academy

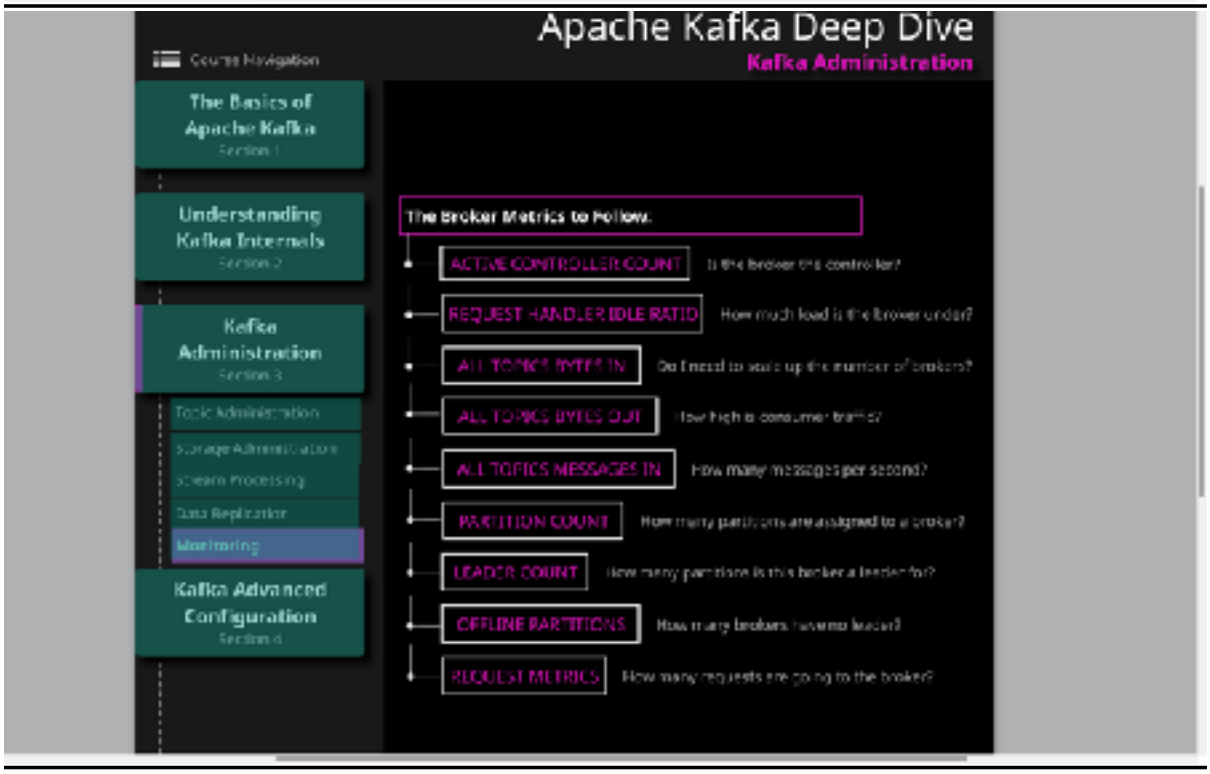

Apache Kafka Deep Dive
Kafka Administration
Design Patterns
Back
Next
Linux Academy

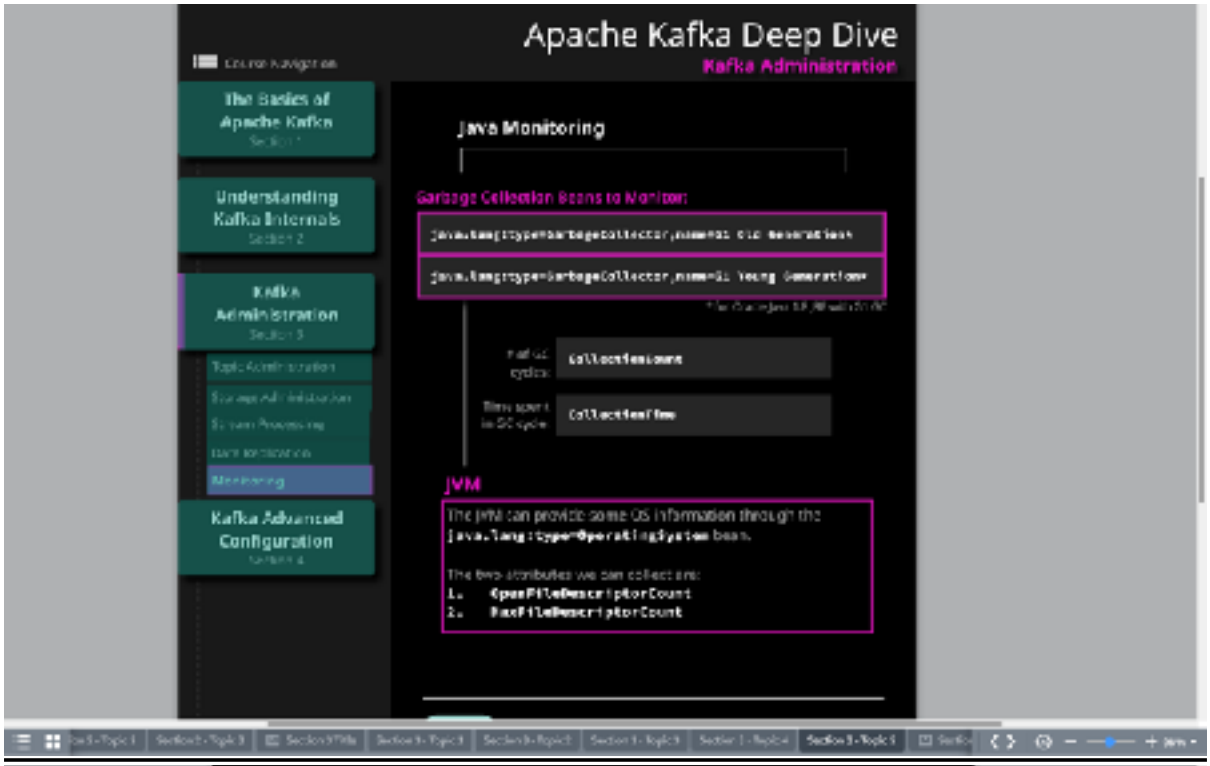

Apache Kafka Deep Dive
Kafka Administration
Frameworks
Back
Linux Academy

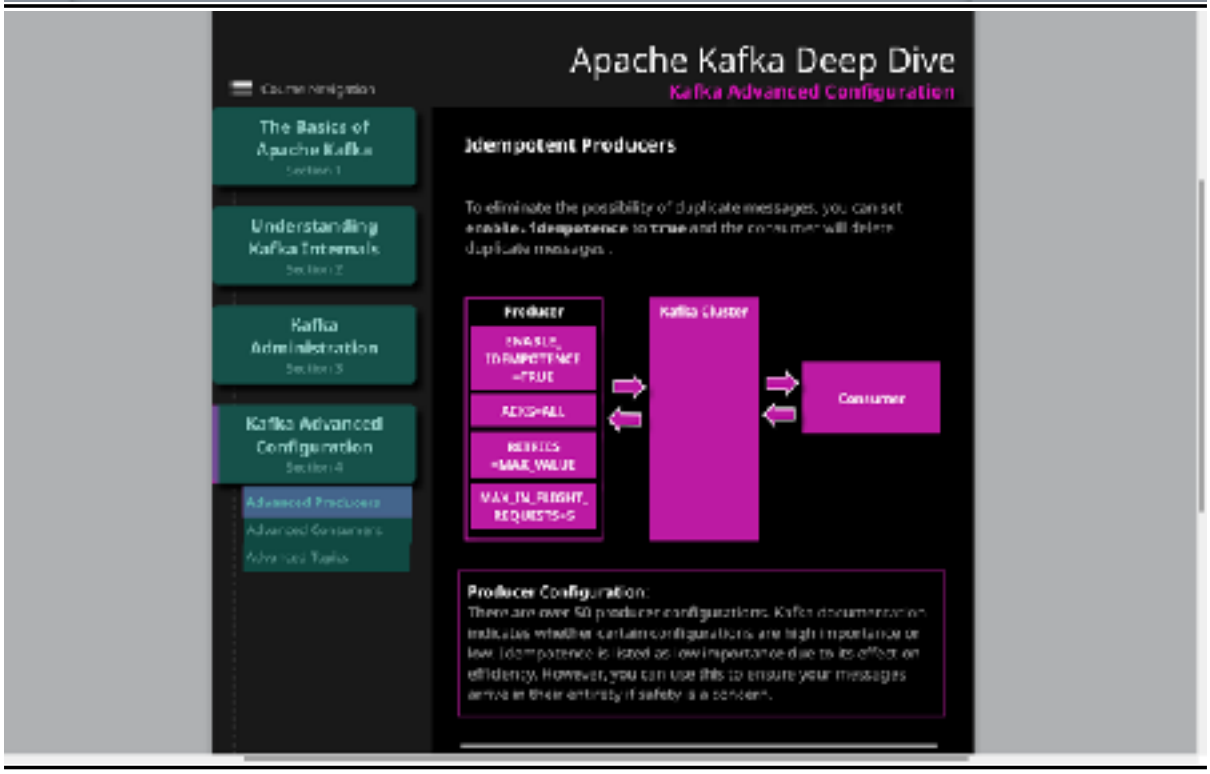

Apache Kafka Deep Dive
Kafka Administration
Course Navigation
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Multi-Cluster Architectures:
Hub-And-Spoke
Central Kafka Cluster
Active-Active
Active-Standby

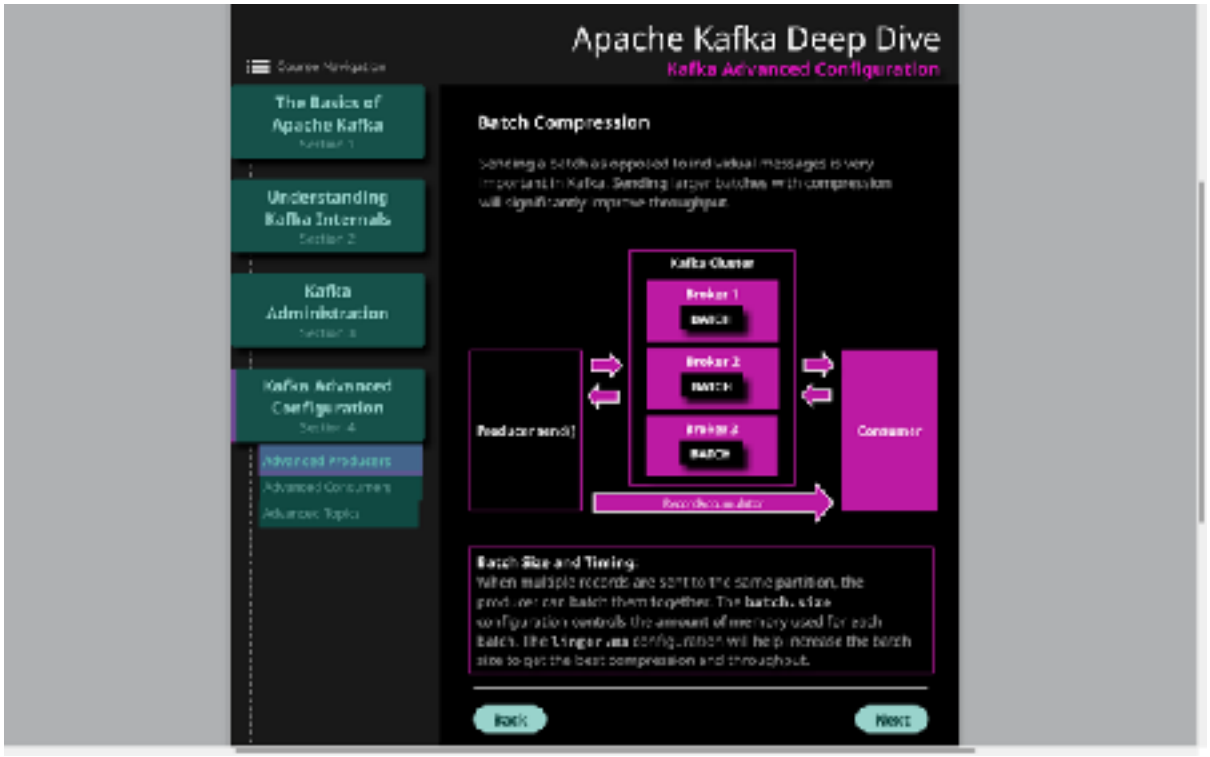

Apache Kafka Deep Dive
Kafka Administration
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
MirrorMaker
Source Cluster
MirrorMaker
Target Cluster
Topic 1
Topic 2
Topic 3
Consumer
Producer
Back

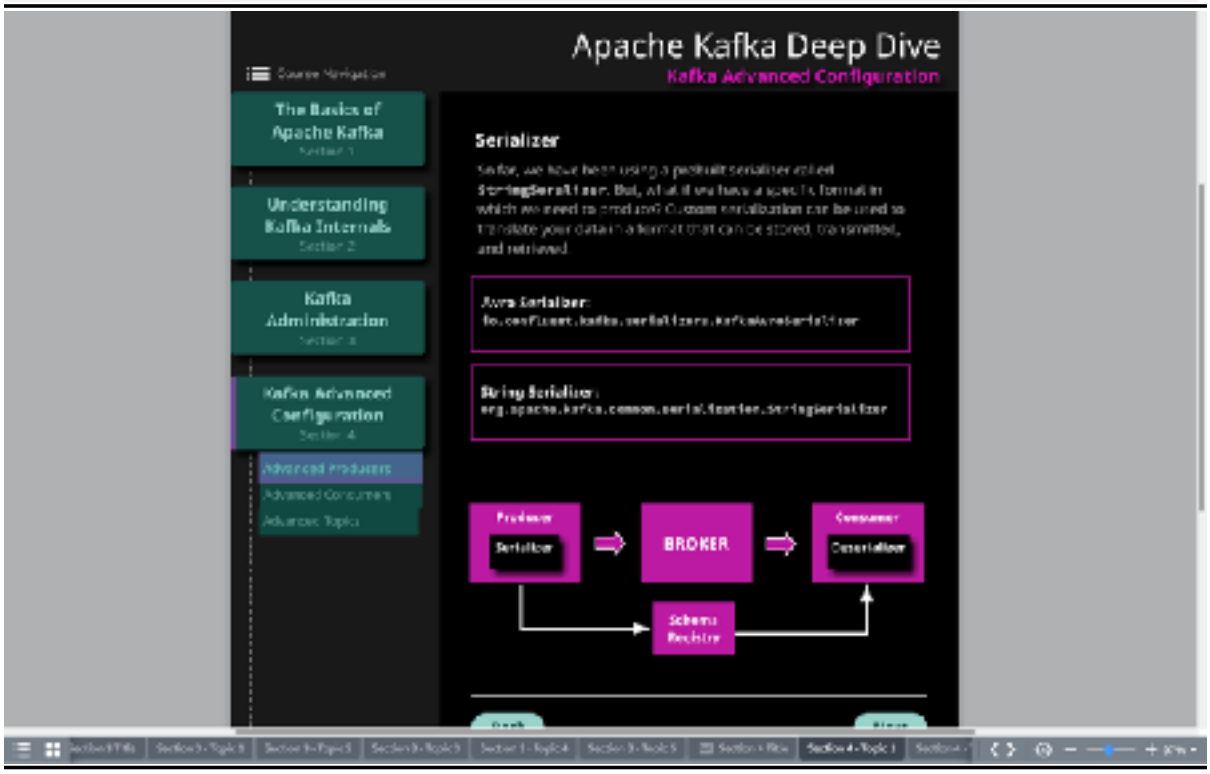

Apache Kafka Deep Dive
Kafka Administration
Course Navigation
The Basics of Apache Kafka
Section 1
Understanding Kafka Internals
Section 2
Kafka Administration
Section 3
Topic Administration
Storage Administration
Stream Processing
Data Replication
Monitoring
Kafka Advanced Configuration
Section 4
Cluster and Broker Monitoring
Metrics can be accessed from:
• Java Management Extensions (JMX)
• Yammer Metrics
Monitor the Overall Health
Problems:
• Under-replicated partitions
• Hardware failures
Solutions:
• Check for under-replicated partitions
• Run a preferred replica election
• Move partition(s) to a different broker.

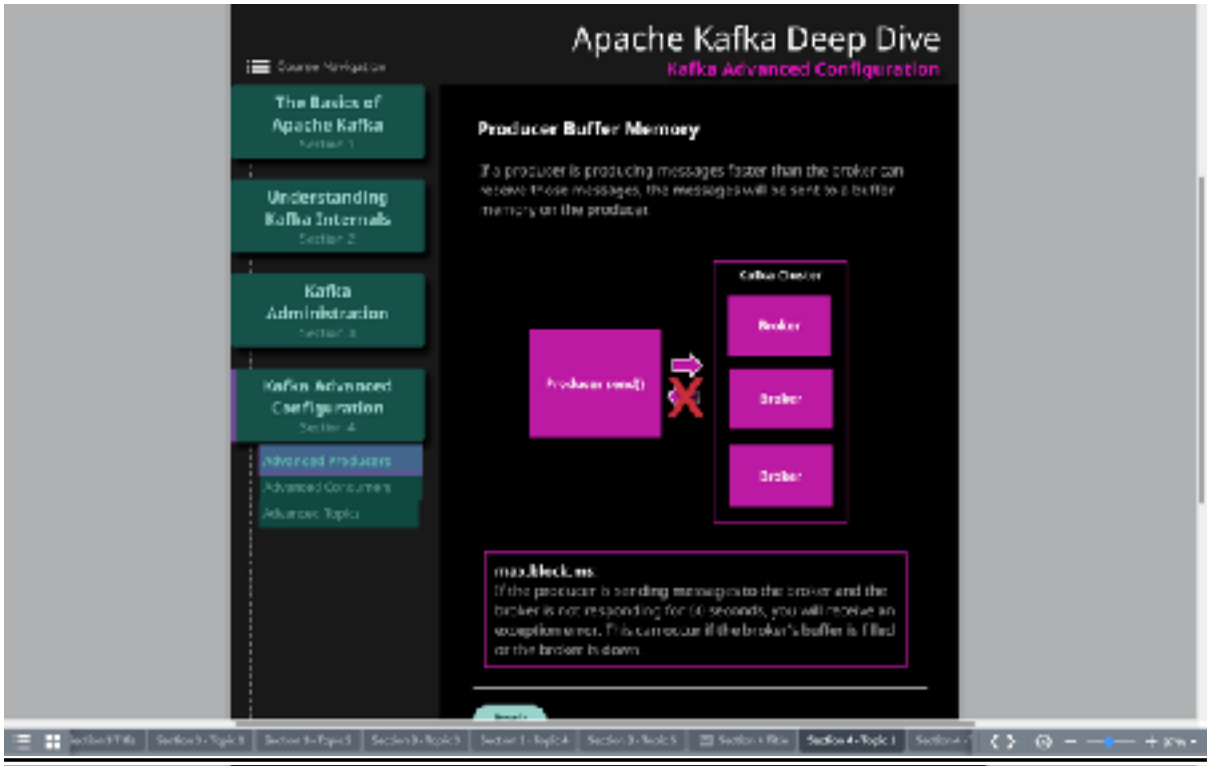

Apache Kafka Deep Dive
Kafka Administration
Course Navigation
The Basics of Apache Kafka
Section 1
Understanding Kafka Internals
Section 2
Kafka Administration
Section 3
Topic Administration
Storage Administration
Stream Processing
Data Replication
Monitoring
Kafka Advanced Configuration
Section 4
The Broker Metrics to Follow:
ACTIVE CONTROLLER COUNT
Is the broker the controller?
REQUEST HANDLER IDLE RATIO
How much load is the broker under?
ALL TOPICS BYTES IN
Do I need to scale up the number of brokers?
ALL TOPICS BYTES OUT
How high is consumer traffic?
ALL TOPICS MESSAGES IN
How many messages per second?
PARTITION COUNT
How many partitions are assigned to a broker?
LEADER COUNT
How many partitions is this broker a leader for?
OFFLINE PARTITIONS
How many brokers have no leader?
REQUEST METRICS
How many requests are going to the broker?

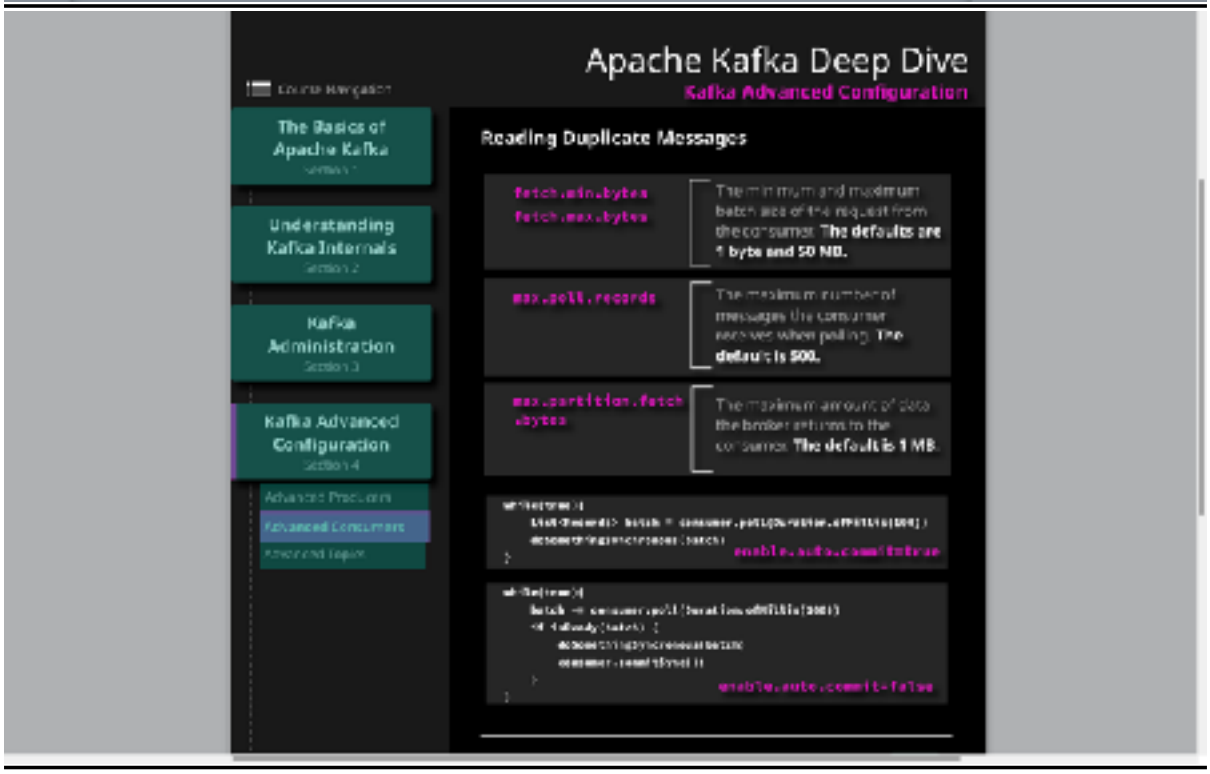

Apache Kafka Deep Dive
Kafka Administration
Course Navigation
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Topic Administration
Monitoring
Kafka Advanced Configuration
Java Monitoring
Garbage Collection Beans to Monitor:
java.lang:type=GarbageCollector,name=G1 Old Generation
java.lang:type=GarbageCollector,name=G1 Young Generation
collectionCount
CollectionTime
JVM
The JVM can provide some OS information through the java.lang:type=OperatingSystem bean.
The two attributes we can collect are:
1. OpenFileDescriptorCount
2. MaxFileDescriptorCount

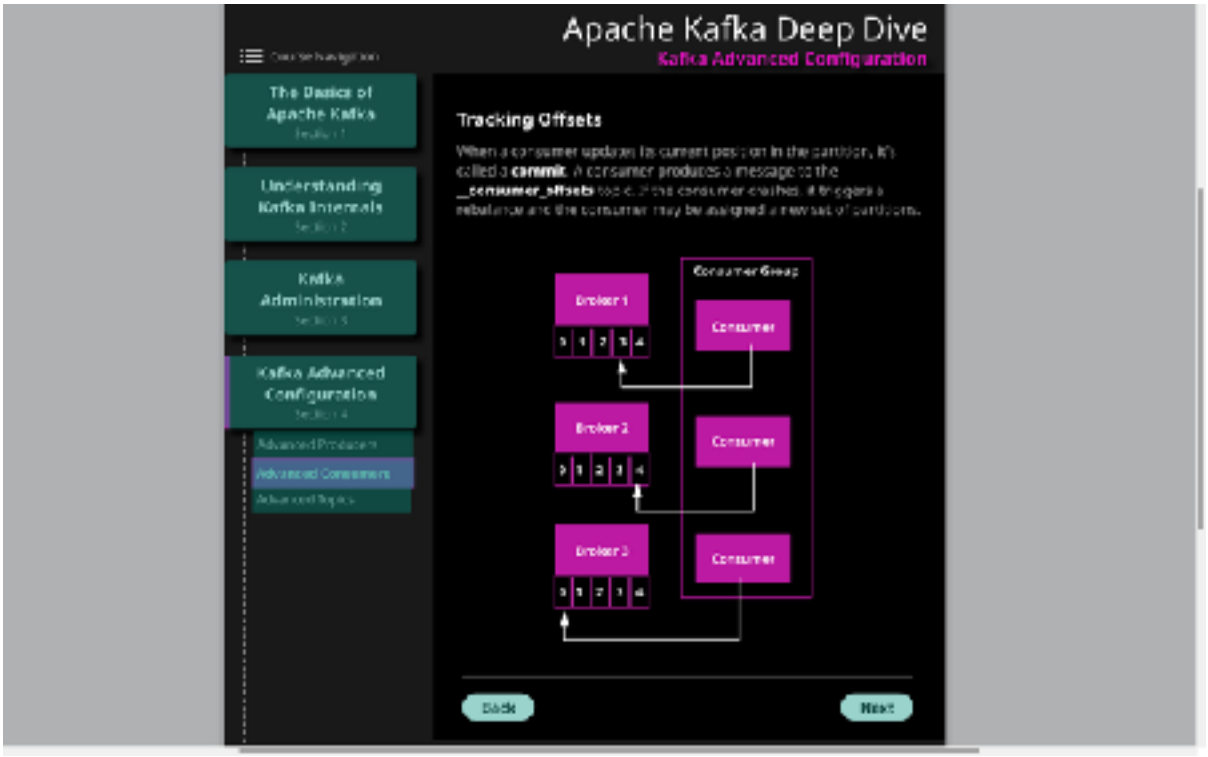

Apache Kafka Deep Dive
Kafka Advanced Configuration
Course Navigation
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Advanced Producers
Idempotent Producers
To eliminate the possibility of duplicate messages, you can set enable.idempotence to true and the consumer will drop duplicate messages.
PRODUCER
ENABLE_IDEMPOTENCY =TRUE
ACKS=ALL
RETRIES =MAX_VALUE
MAX_IN_FLIGHT_REQUESTS=5
Kafka Cluster
Consumer
Producer Configuration:
There are over 50 producer configurations. Kafka documentation indicates whether certain configurations are high importance or low. Idempotence is listed as low importance due to its effect on efficiency. However, you can use this to ensure your messages arrive in their entirety if safety is a concern.

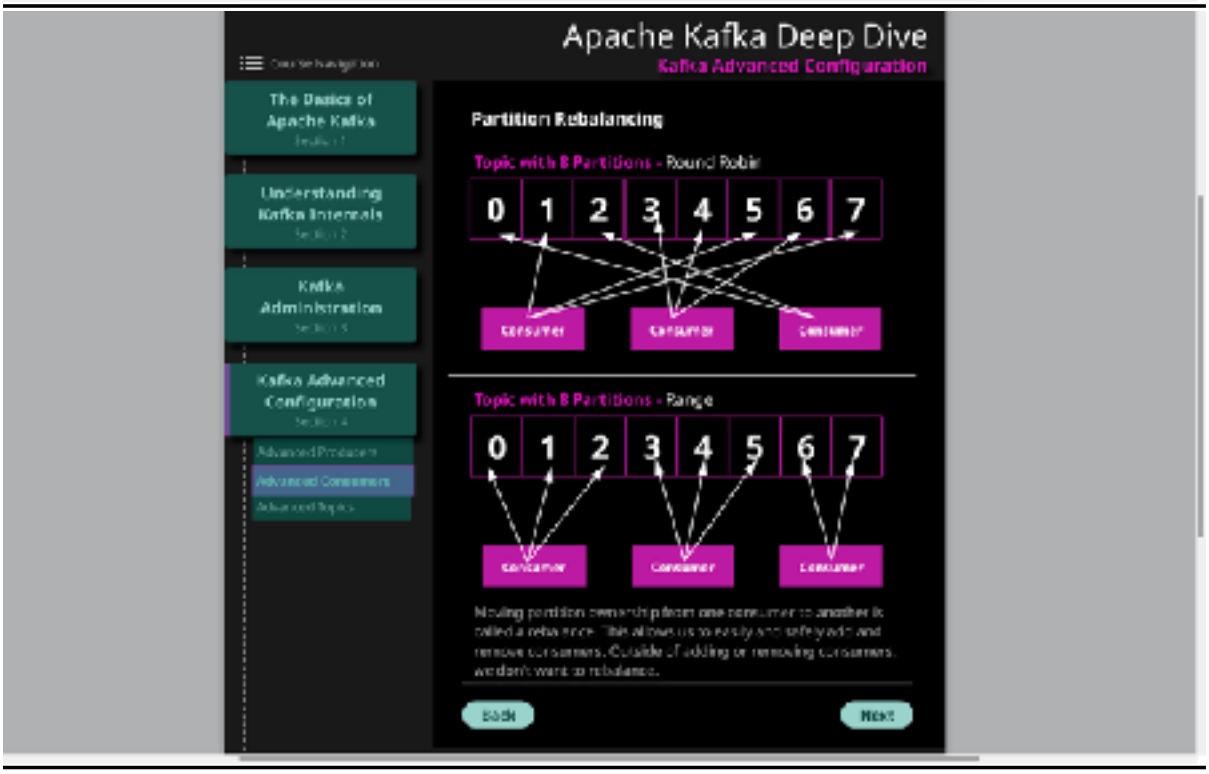

Apache Kafka Deep Dive
Kafka Advanced Configuration
Course Navigation
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Advanced Producers
Advanced Consumers
Advanced Topics
Batch Compression
Kafka Cluster
Broker 1
Broker 2
Broker 3
Producer(s)
Consumer
Batch Size and Timing:
Back
Next

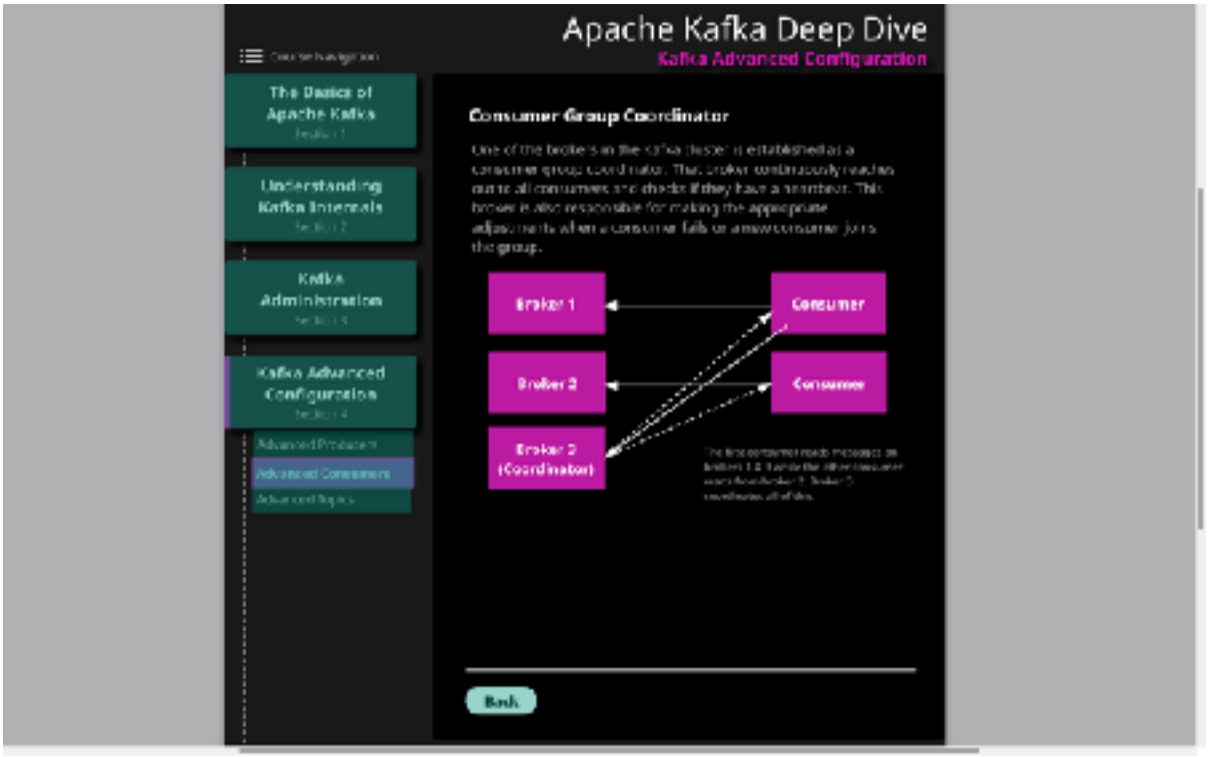

Apache Kafka Deep Dive
Kafka Advanced Configuration
Course Navigation
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Advanced Producers
Advanced Consumers
Advanced Topics
Serializer
Avro Serializer:
String Serializer:
org.apache.kafka.common.serialization.StringSerializer
Producer
Serializer
BROKER
Consumer
Deserializer
Schema Registry

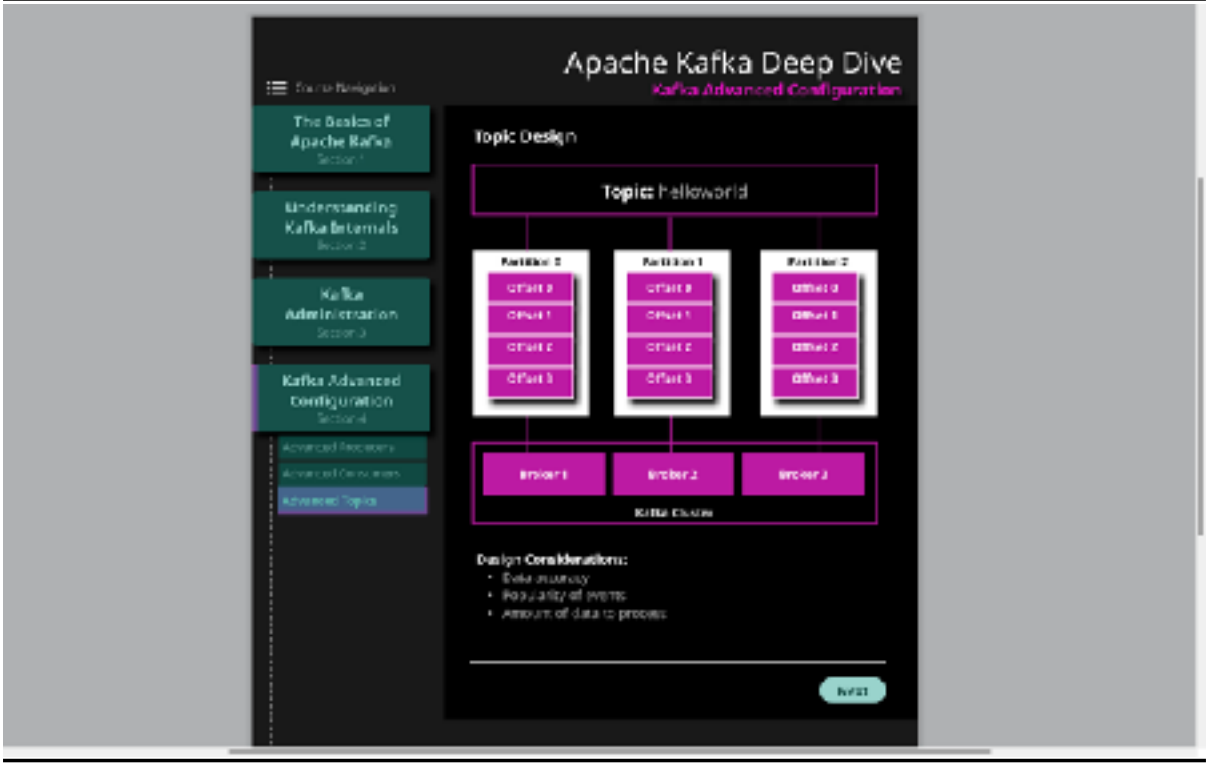

Apache Kafka Deep Dive
Kafka Advanced Configuration
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Advanced Producers
Advanced Consumers
Advanced Topics
Producer Buffer Memory
If a producer is producing messages faster than the broker can receive those messages, the messages will be sent to a buffer memory on the producer
Kafka Cluster
Broker
Broker
Broker
max.block.ms
If the producer is sending messages to the server and the broker is not responding for 60 seconds, you will receive an exception error. This can occur if the broker's buffer is filled or the broker is down

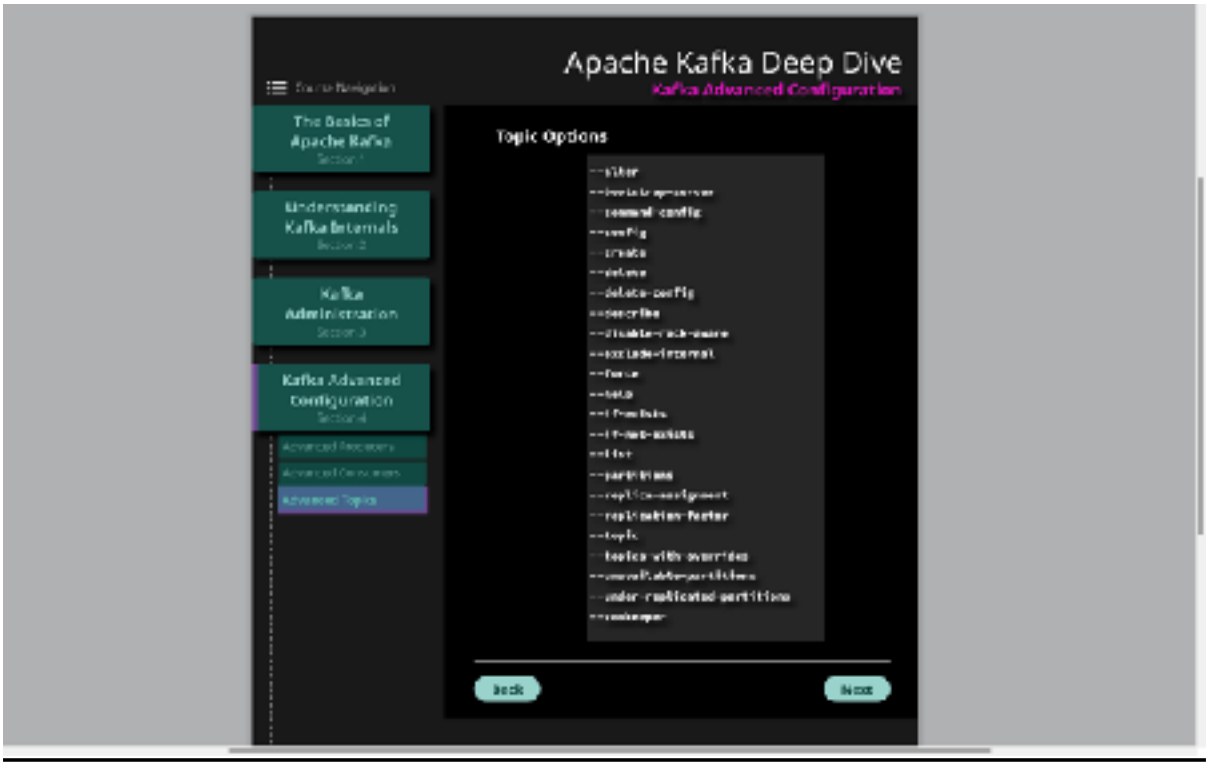

Apache Kafka Deep Dive
Kafka Advanced Configuration
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Advanced Consumers
Reading Duplicate Messages
fetch.min.bytes
fetch.max.bytes
The minimum and maximum batch size of the request from the consumer. The defaults are 1 byte and 50 MB.
max.poll.records
The maximum number of messages the consumer receives when polling. The default is 500.
max.partition.fetch.bytes
The maximum amount of data the broker returns to the consumer. The default is 1 MB.
enable.auto.commit=true
enable.auto.commit=false

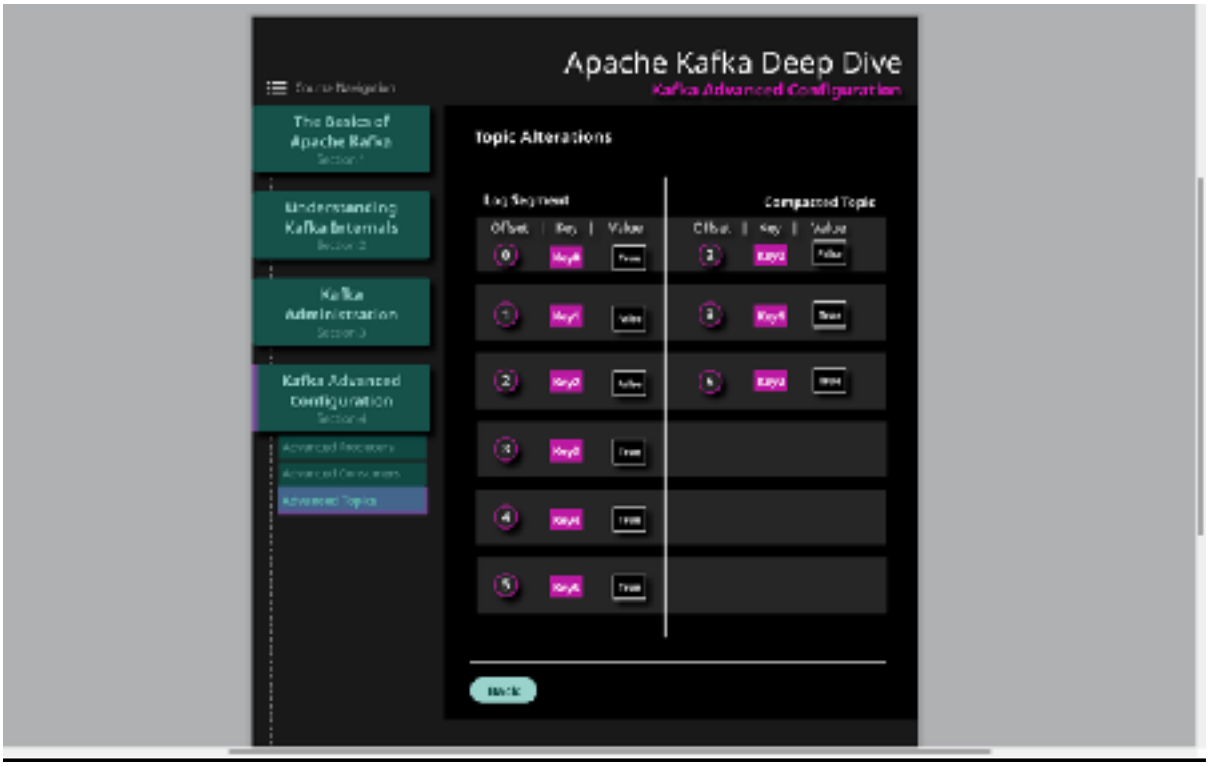

Apache Kafka Deep Dive
Kafka Advanced Configuration
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Tracking Offsets
Consumer Group
Broker 1
Broker 2
Broker 3
Consumer
Consumer
Consumer
Back
Next


Apache Kafka Deep Dive
Kafka Advanced Configuration
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Partition Rebalancing
Topic with 8 Partitions - Round Robin
0 1 2 3 4 5 6 7
Topic with 8 Partitions - Range
0 1 2 3 4 5 6 7
Back
Next


Apache Kafka Deep Dive
Kafka Advanced Configuration
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Advanced Producers
Advanced Consumers
Advanced Topics
Consumer Group Coordinator
Broker 1
Broker 2
Broker 3 (Coordinator)
Consumer
Consumer
Back


Apache Kafka Deep Dive
Kafka Advanced Configuration
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Topic Design
Topic: Helloworld
Design Considerations:


Apache Kafka Deep Dive
Kafka Advanced Configuration
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Topic Options
Back
Next


Apache Kafka Deep Dive
Kafka Advanced Configuration
The Basics of Apache Kafka
Understanding Kafka Internals
Kafka Administration
Kafka Advanced Configuration
Topic Alterations
Compacted Topic
Offset
Key
Value