Apache Kafka Contents 3 days
1. Getting Started
● Introduction
● Use Cases
● Architecture
● Components of Kafka - Broker, Producer, Consumer, Topic, Partition
● Ecosystem
● Kafka vs Flume
Installing Kafka
First Things First
Installing a Kafka Broker
Broker Configuration
General Broker
Topic Defaults
num.partitions
log.retention.ms
log.retention.bytes
log.segment.bytes
log.segment.ms
message.max.bytes
Hardware Selection
Kafka in the Cloud
Kafka Clusters
How Many Brokers
Broker Configuration
Operating System Tuning
Virtual Memory
Disk
Networking
Production Concerns
Garbage Collector Options
Datacenter Layout
Colocating Applications on Zookeeper
Getting Started With Clients
Best Practices of Kafka on Production
Best Practices of Zookpeer on Production

2. Cluster Setup
● Zookeeper
● Single node kafka
● **Hands-On - Setting Up**
● Multi node kafka
● **Hands-On - Multi Node Setup**
● Console Producer & Console Consumer
● **Hands-On - Producer & Consumer**
● High Availability & Performance
● **Considerations for adding only zookeeper node in exisiting cluster**
● **Considerations for adding a kafka node in existing cluster**
● **Considerations for removing only zookeeper node in exisiting cluster**
● **Considerations for removing kafka node in existing cluster**
Kafka Producers - Writing Messages to Kafka
Producer overview
Constructing a Kafka Producer
Sending a Message to Kafka
Serializers
Custom Serializers
Serializing using Apache Avro
Using Avro records with Kafka
Partitions
Configuring Producers
acks
buffer.memory
compression.type
retries
batch.size
linger.ms
client.id
max.in.flight.requests.per.connection
timeout.ms and metadata.fetch.timeout.ms
Old Producer APIs
4. Detailed Design

● Performance tuning
● Serialization, Compression
● Message Delivery Semantics
● Replication
● Log Compaction
● Quotas
● **Hands-On**
Kafka Consumers - Reading Data from Kafka
KafkaConsumer Concepts
Consumers and Consumer Groups
Consumer Groups - Partition Rebalance
Creating a Kafka Consumer
Subscribing to Topics
The Poll Loop
Commits and Offsets
Automatic Commit
Commit Current Offset
Asynchronous Commit
Combining Synchronous and Asynchronous commits
Commit Specified Offset
Rebalance Listeners
Seek and Exactly Once Processing
But How Do We Exit?
Deserializers
Configuring Consumers
fetch.min.bytes
fetch.max.wait.ms
max.partition.fetch.bytes
session.timeout.ms
auto.offset.reset
enable.auto.commit
partition.assignment.strategy
client.id
Stand Alone Consumer - Why and How to Use a Consumer without a Group
Older consumer APIs
Kafka Internals

Cluster Membership
Replication
Request Processing
Produce Requests
Fetch Requests
Other Requests
Physical Storage
Partition Allocation
File Management
File Format
Indexes
Compaction
How Compaction Works
Deleted Events
When Are Topics Compacted
Advanced Configuration
● Broker Configs
● **Hands-On**
● Producer Configs
● Consumer Configs
● Consumer groups
● **Hands-On**
5. Implementation
● API Design
● Producer and Consumer APIs (Java)
● **Hands-On Producer & Consumer API**
● Message format
● Log
● **Hands-On**
6. Operations
● Managing Topics

● Decommissioning nodes
● Data mirroring
● Data centers and Racks
● Monitoring
● Security
● Authorization and ACL
● REST API
● **Hands-On**
7. Confluent Platform
● Overview
● Confluent Platform vs Apache Kafka
● Kafka Streams
● Kafka Connectors
Confluent Platform Hands On Usecases
8. Kafka UseCases HandsOn
● Millions of Messages per second
How to Handle with Kafka?
IoT HandsOn Usecase
● Kafka with Spark
● **Hands-On**
● Kafka with Flume (for Hadoop/Hbase/Hive)
● **Hands-On**
Google Scale Data Handling 100 million messages per sec