

Module 1 : Introduction to BigData, Hadoop (HDFS and MapReduce) : Available (Length 35 Minutes)

1. BigData Introduction
2. Hadoop Introduction
3. HDFS Introduction
4. MapReduce Introduction

Video URL : [Watch](#)

Module 2 : Deep Dive in HDFS : Available (Length 48 Minutes)

1. HDFS Design
2. Fundamental of HDFS (Blocks, NameNode, DataNode, Secondary Name Node)
3. Rack Awareness
4. Read/Write from HDFS
5. HDFS Federation and High Availability
6. Parallel Copying using DistCp
7. HDFS Command Line Interface

Video URL : [Watch](#)

Module 3: Understanding MapReduce: Available (Length 60 Minutes)

1. JobTracker and TaskTracker
2. Topology Hadoop cluster
3. Example of MapReduce
 - Map Function
 - Reduce Function
4. Java Implementation of MapReduce
5. DataFlow of MapReduce
6. Use of Combiner

Video URL : [Watch](#)

Module 4 : MapReduce Internals -1 (In Detail) : Available (Length 57 Minutes)

1. How MapReduce Works
2. Anatomy of MapReduce Job (MR-1)
3. Submission & Initialization of MapReduce Job (What Happen?)
4. Assigning & Execution of Tasks
5. Monitoring & Progress of MapReduce Job
6. Completion of Job
7. Handling of MapReduce Job
 - Task Failure
 - TaskTracker Failure
 - JobTracker Failure

Video URL : [Watch](#)

Module 5 : MapReduce-2 (YARN : Yet Another Resource Negotiator)
: Available (Length 52 Minutes)

1. Limitation of Current Architecture (Classic)
2. What are the Requirement?
3. YARN Architecture
4. Job Submission and Job Initialization
5. Task Assignment and Task Execution
6. Progress and Monitoring of the Job
7. Failure Handling in YARN
 - Task Failure
 - Application Master Failure
 - Node Manager Failure
 - Resource Manager Failure

Video URL : [Watch](#)

Module 6: Advanced Topic for MapReduce (Performance and Optimization): Available (Length 58 Minutes)

1. Job Scheduling
2. In Depth Shuffle and Sorting
3. Speculative Execution
4. Output Committers
5. JVM Reuse in MR1
6. Configuration and Performance Tuning

Video URL : [Watch](#)

Module 7: Advanced MapReduce Algorithm: Available (Length 87 Minutes)

File Based Data Structure

- Sequence File
- MapFile

Default Sorting In MapReduce

- Data Filtering (Map-only jobs)
- Partial Sorting

Data Lookup Strategies

- In MapFiles

Sorting Algorithm

- Total Sort (Globally Sorted Data)
 - InputSampler
- Secondary Sort

Video URL : [Watch](#)

Module 8: Advanced MapReduce Algorithm-2: Available (Length 67 Minutes)

1. MapReduce Joining

- Reduce Side Join
- MapSide Join
- Semi Join

2. MapReduce Job Chaining

- MapReduce Sequence Chaining
- MapReduce Complex Chaining

Video URL : [Watch](#)

Module 9 : Features of MapReduce : Available : Private (Length 61 Minutes)

Introduction to MapReduce Counters

Types of Counters

Task Counters

Job Counters

User Defined Counters

Propagation of Counters

Side Data Distribution

Using JobConfiguration

Distributed Cache

Steps to Read and Delete Cache File

Video URL : [Watch](#)

Module 10: MapReduce DataTypes and Formats : Available : Private (Length 77 Minutes)

1. Serialization In Hadoop

2. Hadoop Writable and Comparable

3. Hadoop RawComparator and Custom Writable

4. MapReduce Types and Formats

5. Understand Difference Between Block and InputSplit

6. Role of RecordReader

7. FileInputFormat

8. CombineFileInputFormat and Processing whole file Single Mapper

9. Each input File as a record

10. Text/KeyValue/NLine InputFormat

11. BinaryInput processing

12. MultipleInputs Format

13. DatabaseInput and Output

14. Text/Binary/Multiple/Lazy OutputFormat MapReduce Type

Video URL : [Watch](#)

Module 11: Apache Pig : Available Length 52 Minutes

1. What is Pig ?

2. Introduction to Pig Data Flow Engine

3. Pig and MapReduce in Detail

4. When should Pig Used?

5. Pig and Hadoop Cluster

6. Pig Interpreter and MapReduce

7. Pig Relations and Data Types

8. PigLatin Example in Detail

9. Debugging and Generating Example in Apache Pig

Video URL : [Watch](#)

Module 12 : Fundamental of Apache Hive Part-1 : Available Length 60 Minutes

1. What is Hive?

2. Architecture of Hive

3. Hive Services

4. Hive Clients

5. how Hive Differs from Traditional RDBMS

6. Introduction to HiveQL

7. Data Types and File Formats in Hive

8. File Encoding

9. Common problems while working with Hive

Video URL : [Watch](#)

Module 13 : Deep Dive in HDFS: Available (73 Minutes)

1. Hive Tables
 - managed
 - external
2. Partitions and Buckets
 - Code Examples
 - Sorted Buckets
3. TABLESAMPLES
4. Hive Storage Formats
 - File Format
 - Row Format
5. Hive SerDe
 - Understanding of SerDe
 - Object Inspector
 - LifeCycle of SerDe
 - Writing Custom SerDe and Code Example

Video URL : [Watch](#)

Module 14 : Single Node Hadoop Cluster Set Up In Amazon Cloud : Available (Length 60 Minutes Hands On Practice Session)

1. How to create instance on Amazon EC2
2. How to connect that Instance Using putty
3. Installing Hadoop framework on this instance
4. Run sample wordcount example which come with Hadoop framework.

In 30 minutes you can create Hadoop Single Node Cluster in Amazon cloud, does it interest you?

Video URL : [Watch](#)

Module 15 Hands On : Implementation on NGram algorithm: Available (Length 48 Minutes Hands On Practice Session)

1. Understand the NGram concept using (Google Books NGram)
2. Step by Step Process creating and Configuring eclipse for writing MapReduce Code
3. Deploying the NGram application in Hadoop Installed in Amazon EC2
4. Analyzing the Result by Running NGram application (UniGram, BiGram, TriGram etc.)

Video URL : [Watch](#)

To buy the training [Click Here](#) or Visit www.HadoopExam.com

[For Corporate Bulk License Please Contact Us](#)

Email: admin@hadoopexam.com , hadoopexam@gmail.com , pappupass2004@gmail.com

Phone: 022-42669636

Mobile: +91-8879712614

HadoopExam Learning Resource

B-902 Shah Trade Centre

Malad-E Mumbai 400097