

HADOOP TRAINING:

Hadoop Course will provide the essential concepts of MapReduce applications developed using Hadoop, including an in depth check out framework components, use of Hadoop for a spread of knowledge analysis tasks, and various samples of Hadoop in action. This course will further examine related technologies like Hive, Pig, and Apache Accumulo.

Course Content:

Introduction, the Motivation for Hadoop:

- Problems with traditional large-scale systems
- Requirements for a new approach

Hadoop Basic Concepts:

- An Overview of Hadoop
- The Hadoop Distributed File System
- Hands-on Exercise
- How MapReduce Works
- Hands-on Exercise
- Anatomy of a Hadoop Cluster
- Other Hadoop Ecosystem Components

Writing a MapReduce Program:

- Examining a Sample MapReduce Program
- With several examples
- Basic API Concepts
- The Driver Code
- The Mapper
- The Reducer
- Hadoop's Streaming API

Delving Deeper Into the Hadoop API:

- More About Tool Runner
- Testing with MR Unit
- Reducing Intermediate Data With Combiners
- The configure and close methods for Map/Reduce Setup and Teardown
- Writing Partitioners for Better Load Balancing
- Hands-On Exercise
- Directly Accessing HDFS
- Using the Distributed Cache

- Hands-On Exercise

Performing several Hadoop jobs:

- The configure and close Methods
- Sequence Files
- Record Reader
- Record Writer
- Role of Reporter
- Output Collector
- Processing video files and audio files
- Processing image files
- Processing XML files
- Counters
- Directly Accessing HDFS
- Tool Runner
- Using The Distributed Cache

Common MapReduce Algorithms:

- Sorting and Searching

- Indexing
- Classification/Machine Learning
- Term Frequency-Inverse Document Frequency
- Word Co-Occurrence
- Hands-On Exercise: Creating an Inverted Index
- Identity Mapper
- Identity Reducer
- Exploring well known problems using MapReduce applications

Using HBase:

- What is HBase?
- HBase API
- Managing large data sets with HBase
- Using HBase in Hadoop applications
- Hands-on Exercise

Using Hive and Pig:

- Hive Basics

- Pig Basics
- Hands-on Exercise
- Practical Development Tips and Techniques
- Debugging MapReduce Code
- Using LocalJobRunner Mode for Easier Debugging
- Retrieving Job Information with Countries
- Logging
- Splittable File Formats
- Determining the Optimal Number of Reducers
- Map-Only MapReduce Jobs
- Hands-on Exercise

Debugging MapReduce Programs:

- Testing with MR Unit
- Logging
- Classification/Machine Learning
- Advanced MapReduce Programming
- A Recap of the MapReduce Flow

- The Secondary Sort
- CustomizedInputFormats and Output Formats
- Pipelining Jobs With Oozie
- Map-Side Joins
- Reduce-Side Joins

Joining Data Sets in MapReduce:

- Map-Side Joins
- The Secondary Sort
- Reduce-Side Joins

Monitoring and debugging on a Production Cluster:

- Counters
- Skipping Bad Records
- Rerunning failed tasks with Isolation Runner

Tuning for Performance in MapReduce:

- Reducing network traffic with combiner
- Partitioners



+17207384411
info@procareer.io

- Reducing the amount of input data
- Using Compression
- Reusing the JVM
- Running with speculative execution
- Refactoring code and rewriting algorithms Parameters affecting Performance
- Other Performance Aspects

Our learning methods include:

- Comprehensive course selection of Instructor-Led Training
- Logistical convenience and interactive classroom experience of Online Training
- Flexible pacing and instructor-guided support of Mentored Learning
- Self-paced convenience of Online ANYTIME

In addition:

- Interview preparation with mock interview drills
- Effective resume building
- Process of applying jobs at the right places



+17207384411
info@procareer.io

Reach us:

Call : +1 720 738 4411

Email ID: info@procareer.io

Website: <https://www.procareer.io/>

