INFOSOFT IT SOLUTIONS

Training | Projects | Placements

Revathi Apartments, Ameerpet, 1st Floor, Opposite Annapurna Block,

Infosoft It solutions, Software Training & Development Institute, 9059683947|9182540872

Map Reduce

Introduction to Distributed Computing and Big Data

- Overview of Distributed Computing: Concepts, challenges, and benefits
- Introduction to Big Data: Characteristics, sources, and applications
- Challenges in Big Data Processing: Volume, velocity, variety, and veracity

Introduction to MapReduce

- Evolution of MapReduce: Origins and development
- MapReduce Programming Model: Map and Reduce phases
- Advantages and Limitations of MapReduce

Hadoop Ecosystem Overview

- Introduction to Hadoop: Architecture and components (HDFS, YARN)
- Hadoop Distributed File System (HDFS): Storage and data replication
- Resource Management with YARN: Job scheduling and execution

Setting Up a Hadoop Cluster

- Installing Hadoop: Single-node and multi-node cluster setup
- Configuring Hadoop: XML configuration files and parameters
- Hadoop Cluster Management: Monitoring and administration tools

MapReduce Basics

- Anatomy of a MapReduce Job: Mapper, Reducer, InputFormat, OutputFormat
- MapReduce Execution Flow: Job submission and execution steps
- Writing and Running a MapReduce Program: Example walkthrough

MapReduce Advanced Concepts

- Combiner and Partitioner: Optimization techniques in MapReduce
- Distributed Cache: Sharing files across MapReduce tasks
- Input and Output Formats: Handling different data formats (Text, SequenceFile, Avro)

MapReduce Design Patterns

- MapReduce Design Patterns: Common patterns (Filtering, Summarization, Joins)
- Optimization Techniques: Data locality, speculative execution, task tuning
- Advanced MapReduce Patterns: Secondary sort, count distinct, inverted index

Handling Large-scale Data with MapReduce

- Scaling MapReduce: Handling large datasets with partitioning and shuffling
- Performance Tuning: Optimization strategies for MapReduce jobs

MapReduce Algorithms and Applications

- PageRank Algorithm: Implementing PageRank using MapReduce
- Word Count Example: Variations and optimizations
- Graph Processing: Graph algorithms using MapReduce (BFS, DFS)

Integration with Other Big Data Technologies

- Integration with Hive: Using HiveQL for data analysis
- Integration with Pig: Using Pig Latin for data processing
- Integration with Spark: Comparing MapReduce and Spark

Real-time Big Data Processing

- Overview of Real-time Processing: Stream processing vs. batch processing
- Apache Kafka Integration: Using Kafka with Hadoop ecosystem
- Processing Streaming Data: Using tools like Apache Storm or Apache Flink

MapReduce Project Work and Case Studies

- Real-world MapReduce Projects: Implementation and evaluation
- Case Studies: Industry-specific applications (e.g., retail, healthcare)
- Presentation and Documentation of MapReduce Projects