Comprehensive MapReduce Training helps the delegates in developing their knowledge about MapReduce framework, and the received MapReduce certification will prove their ability to the employers. BY undergoing this training program, the delegates can enhance their career prospects greatly.

Comprehensive MapReduce Training is designed to provide delegates with the knowledge and skills required to master the fundamental and advanced concepts of MapReduce. During the training program, the delegates will gain a comprehensive understanding of Hadoop and its usage in parallel data processing. The delegates will also learn how to write MapReduce programs for analysing Big data and bringing business benefits to an organisation. The training will be delivered by certified and experienced instructors who make sure that the delegates are able to implement the technology in their future projects.

Prerequisites

The professionals who wish to attend this course must have good knowledge of core java and analytics so that they can grasp the concepts and terminologies used in MapReduce easily.

Course Objectives

During the training program, the delegates will learn:

- How to use Hadoop in parallel processing
- Advanced features of MapReduce and YARN
- How to write MapReduce programs
- How to bring more benefits to the organisation

More about MapReduce…

Developed by Google, MapReduce is one of the core building blocks of processing in the framework of Hadoop. It allows the users to perform distributed and parallel processing on large datasets. It consists of two different jobs: Map and Reduces. In the map (first job), the data is read and processed for producing key-value pairs. These key-value pairs are taken as input for a reducer (second job). The reducer then aggregates the intermediate data into a smaller set of tuples which is the final output.

The two major advantages of MapReduce are Parallel Processing and Data Locality.

Parallel Processing :

In MapReduce, the job is divided into various nodes. Each node works with a part of job concurrently. Therefore, MapReduce is based on Divide and Conquer model. This model help users to process data using different machines. The processing time of data is reduced as the multiple machines are used in parallel instead of one.

Data Locality:

Tel +44 20 4571 2395
info@siliconbeachtraining.co.uk

Despite moving data to process unit, the processing unit is moved to data in the framework of MapReduce.

Getting Started

- What is Big Data?
- Huge Data Sets Scalable Systems
- Complexity and its Solution

HDFC and MapReduce

- HDFC: Behind the Concept
- MapReduce Data Flow
- MapReduce Design Patterns

MapReduce Phases

- Map Phase
- Sort Phase
- Reduce Phase

MapReduce Design Patterns

- Input-Map-Reduce-Output
- Input-Map-Output
- Input-Multiple Maps-Reduce-Output
- Input-Map-Combiner-Reduce-Output

Advanced Topics

- Clustering Ensemble
- Clustering in MapReduce
- Semi-Supervised Clustering

Tel +44 20 4571 2395
info@siliconbeachtraining.co.uk

Comprehensive MapReduce Training helps the delegates in developing their knowledge about MapReduce framework, and the received MapReduce certification will prove their ability to the employers.