Hadoop

About the Industry (Hadoop)

- Hadoop is a distributed processing technology used for Big Data analysis. Hadoop market is expanding at a significant rate, as Hadoop technology provides cost effective and quick solutions compared to traditional data analysis tools such as RDBMS. The Hadoop Market has great future prospects in trade and transportation, BFSI and retail sector. Global Hadoop market was valued at $1.5 billion in 2012, and is expected to grow at a CAGR of 58.2% during 2013 to 2020 to reach $50.2 billion by 2020.
- The major drivers for the market growth is the growing volume of structured and unstructured data, increasing demand for big data analytics and quick and affordable data processing services offered by Hadoop technology.

IIHT's Approach

- We at IIHT always believe in catering to the latest demands of IT industry. To match and exceed their expectations, we have Hadoop as an offering where we train you on the below technologies

  - Java Fundamentals
  - Hadoop Fundamentals
  - HDFS
  - Map Reduce
  - Spark
  - Hive
  - Pig
  - HBase
  - Sqoop
  - Yarn
  - MongoDB
  - Hadoop Security

IIHT
Connecting Talent to Opportunity

Java Fundamentals

- Java is a high-level programming language originally developed by Sun Microsystems and released in 1995. Java runs on a variety of platforms, such as Windows, Mac OS, and the various versions of UNIX. This tutorial gives a complete understanding of Java.
- This reference will take you through simple and practical approach while learning Java Programming language.

Note:

This consists of the essentials that a candidate should know to begin learning about Hadoop.

Hadoop Fundamentals

- Hadoop is indispensable when it comes to processing big data—as necessary to understanding your information as servers are to storing it. This course is your introduction to Hadoop, its file system (HDFS), its processing engine (MapReduce), and its many libraries and programming tools.

HDFS

- The Hadoop Distributed File System (HDFS) is the primary storage system used by Hadoop applications.
- HDFS is a distributed file system that provides high-performance access to data across Hadoop clusters. Like other Hadoop-related technologies, HDFS has become a key tool for managing pools of big data and supporting big data analytics applications.
- HDFS is built to support applications with large data sets, including individual files that reach into the terabytes. It uses a master/slave architecture, with each cluster consisting of a single NameNode that manages file system operations and supporting DataNodes that manage data storage on individual compute nodes.

Map Reduce

- MapReduce is a core component of the Apache Hadoop software framework.
- Hadoop enables resilient, distributed processing of massive unstructured data sets across commodity computer clusters, in which each node of the cluster includes its own storage. MapReduce serves two essential functions: It parcels out work to various nodes within the cluster or map, and it organizes and reduces the results from each node into a cohesive answer to a query.

Spark

- A new name has entered many of the conversations around big data recently. Some see the popular newcomer Apache Spark as a more accessible and more powerful replacement for Hadoop, big data's original technology of choice. Others recognize Spark as a powerful complement to Hadoop and other more established technologies, with its own set of strengths, quirks and limitations.
- Spark, like other big data tools, is powerful, capable, and well-suited to tackling a range of data challenges. Spark, like other big data technologies, is not necessarily the best choice for every data processing task.

Hive

- Apache Hive is an open-source data warehouse system for querying and analyzing large datasets stored in Hadoop files. Hadoop is a framework for handling large datasets in a distributed computing environment.

Pig

- Apache Pig is a platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs, coupled with infrastructure for evaluating these programs. The salient property of Pig programs is that their structure is amenable to substantial parallelization, which in turns enables them to handle very large data sets.
- At the present time, Pig's infrastructure layer consists of a compiler that produces sequences of Map-Reduce programs, for which large-scale parallel implementations already exist (e.g., the Hadoop subproject). Pig's language layer currently consists of a textual language called Pig Latin.

HBase

- HBase is an open source, non-relational, distributed database modeled after Google's BigTable and written in Java.
- It is developed as part of Apache Software Foundation's Apache Hadoop project and runs on top of HDFS (Hadoop Distributed Filesystem), providing BigTable-like capabilities for Hadoop.
- It provides a fault-tolerant way of storing large quantities of sparse data

Sqoop

- Sqoop is a tool designed to transfer data between Hadoop and relational database servers.
- It is used to import data from relational databases such as MySQL, Oracle to Hadoop HDFS, and export from Hadoop file system to relational databases.

Yarn

- Apache Hadoop YARN (Yet Another Resource Negotiator) is a cluster management technology.
- YARN is one of the key features in the second-generation Hadoop 2 version of the Apache Software Foundation's open source distributed processing framework. Originally described by Apache as a redesigned resource manager, YARN is now characterized as a large-scale, distributed operating system for big data applications.

MongoDB

- MongoDB is an open source database that uses a document-oriented data model.
- MongoDB is one of several database types to arise in the mid-2000s under the NoSQL banner. Instead of using tables and rows as in relational databases, MongoDB is built on an architecture of collections and documents.
- Documents comprise sets of key-value pairs and are the basic unit of data in MongoDB. Collections contain sets of documents and function as the equivalent of relational database tables.

Hadoop Security

- Security is a top agenda item and represents critical requirements for Hadoop projects. Over the years, Hadoop has evolved to address key concerns regarding authentication, authorization, accounting, and data protection natively within a cluster and there are many secure Hadoop clusters in production. Hadoop is being used securely and successfully today in sensitive financial services applications, private healthcare initiatives and in a range of other security-sensitive environments. As enterprise adoption of Hadoop grows, so do the security concerns and a roadmap to embrace and incorporate these enterprise security features has emerged.

Job Profile

- Hadoop Developer on Spark
- Hadoop Consultant
- Technical Lead – Big Data
- Hadoop Engineer
- Senior Hadoop Engineer
- Computer Scientist – Hadoop Developer
- Analytics Tech Lead

FAQs

Who should do this programme?

- This programme is designed to cater the needs of freshers as well as experienced professionals. You get a complete exposure to the Hadoop environment and can do the tasks independently.

Duration of this programme?

- 100 hours

Does IIHT provide placement assistance after finishing this?

- Yes, IIHT has got tie-ups with MNCs and other companies. However, the candidate needs to have good soft skill and interview-facing skills.

IIHT
Connecting Talent to Opportunity

FAQs

Benefits of doing this programme?

- This is a custom tailored programme that opens the doors for you to enter the Hadoop era. Here you learn all the thrilling tools and the once which are gaining popularity in the market rather than learning the tools which will be obsolete in days to come.
- While you learn development, this also gives you an overview on all major tools which makes you the first preference of the recruiters.

IIHT Edge

Why IIHT?

- IIHT is the only pan India company to have specialised and quality programmes in IT-IMS, Social, Mobility, Analytics and Cloud.
- IIHT has a heritage of over 23 years
- IIHT has about 150 centres across the globe
- IIHT trains corporates like IBM, Intel, HP, HCL 150+ Fortune 500 companies. This ensures that our course curriculum is mapped to industry demands much better than other institutes.
- IIHT has trained over 15 Lakh students till date

IIHT
Connecting Talent to Opportunity

Reach Us

- For Big data Training & Hadoop Training

- No: 15, 4th Floor, Sri Lakshmi Complex, Off MG Road, Near SBI LHO, St. Marks Road
  Bangalore - 560 001, India.
- Call us :**1800-123-321-5 (Toll Free)**

- **Visit our Official website For more Information:**
  http://www.iiht.com/big-data-hadoop-sqoop-training-institute/

IIHT
Connecting Talent to Opportunity