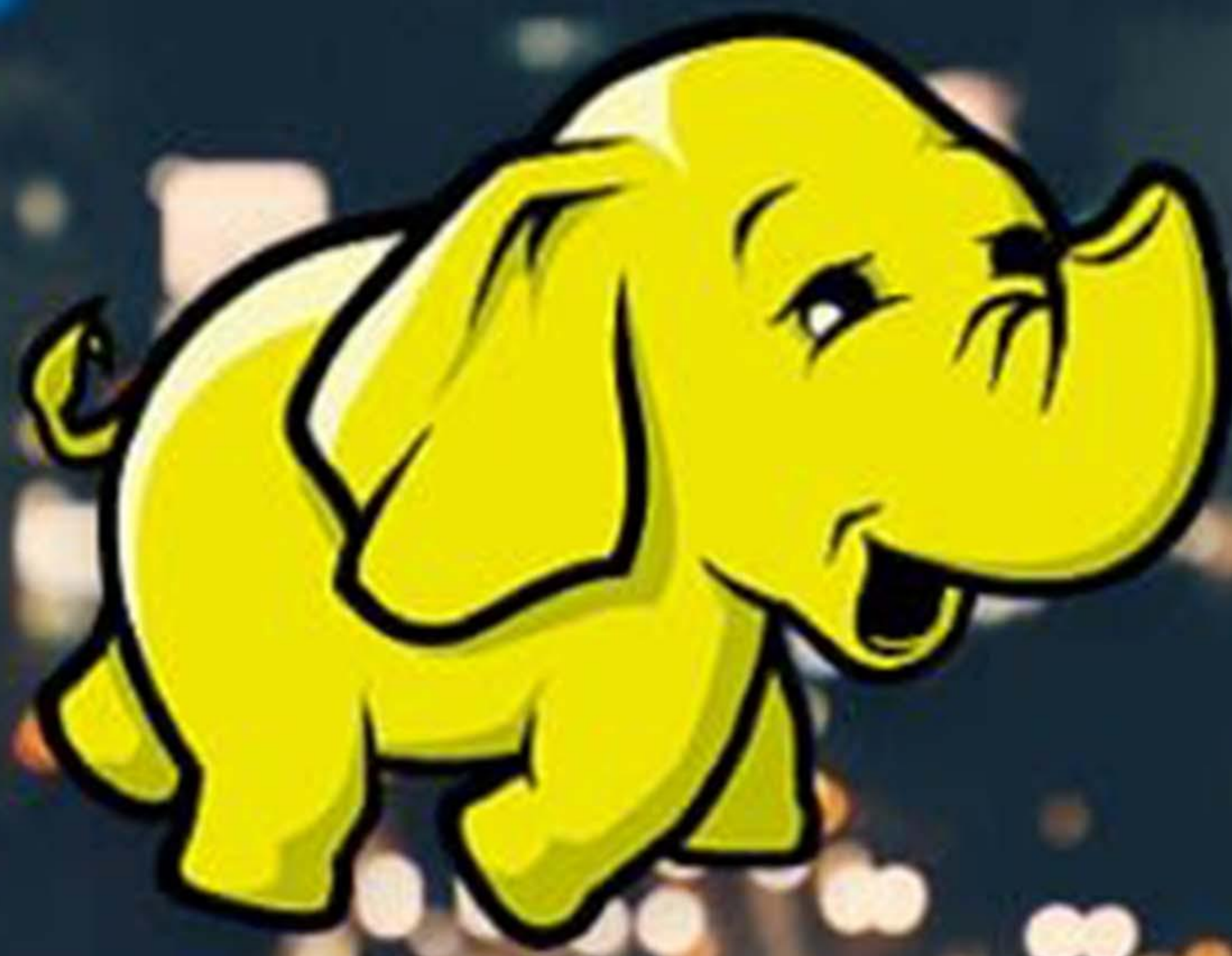


Hadoop Industrial Training



About Course

Apache Hadoop is an open source framework that is used to efficiently store and process large datasets ranging in size from gigabytes to petabytes of data. Instead of using one large computer to store and process the data, Hadoop allows clustering multiple computers to analyze massive datasets in parallel more quickly.

HADOOP INDUSTRIAL TRAINING

CURRICULUM

① Introduction To Hadoop

High Scalability

Scaling

Advantages and Challenges

② Introduction To Big Data

What Is Big Data

Big Data Opportunities

Big Data Challenges

Characteristics Of Big Data

③ Interfacing To Hadoop

Hadoop Distributed File System

Hadoop Vs SQL

Data Locality

Hadoop Architecture

Hadoop Single Node Image

④ Hadoop Distributed File System

- Design and Concepts
- Block Nodes, Data Nodes
- HDFS Availability and Federation
- HDFS Command-Line interface
- Basic File System
- How to add new data node dynamically
- HDFS Availability and Federation
- FSCK Utility (Block Report)
- How to override default configuration at System and Programming level
- Anatomy of File Read/Write

5 Data Integration With Sqoop

- Hadoop Integration
- Introduction Telend
- Loading Data from RDBMS

6 Hadoop Project


- Solving Big Data Issues
- Discussing DataSets
- Live Training On Hadoop
- Assignments



 www.softcrayons.com

 info@softcrayons.com

 (+91) 854 501 2345

 693, Sector 14-A, Vasundhara,
Ghaziabad (U.P.), 201012

   @softcrayons