Cognixia®

Big Data Hadoop Developer Training

About Cognixia

Cognixia- A Digital Workforce Solutions Company is dedicated to delivering exceptional trainings and certifications in digital technologies. Founded in 2014, we provide interactive, customized training courses to individuals and organizations alike, and have served more than 100,000 professionals across 37 countries worldwide.

Our team of more than 4,500 industry experts facilitate more than 400 comprehensive digital technologies courses, along with state-of-the-art infrastructure, to deliver the best learning experience for everyone. Our comprehensive series of instructor-led online trainings, classroom trainings and on-demand self-paced online trainings cover a wide array of specialty areas, including all of the following:

- IoT
- Big Data
- Cloud Computing
- Cyber Security
- Machine Learning
- AI & Deep Learning
- Blockchain Technologies
- DevOps

Cognixia is ranked amongst the top five emerging technologies training companies by various prestigious bodies. We're also an MAPR Advantage Partner, Hortonworks Community Partner, RedHat Enterprise Partner, Microsoft Silver Learning Partner and an authorized training partner for Dell EMC, Pivotal, VMware and RSA technologies.

OUR AWARDS & AFFILIATIONS

siliconindia BEST TRAINING PROVIDER
BIG DATA HADOOP DEVELOPER OF THE YEAR 2016

Top 10 IoT Training Institutes – 2016
Analytics India MAGAZINE

Asian Training & Leadership Award, Dubai

TOP 5 Emerging Technologies 2016 TRAINING COMPANIES
siliconindia

TOP 5 CLOUD TRAINING INSTITUTIONS INDIA 2016
HIGHER EDUCATION Review

TOP 5 Big Data TRAINING INSTITUTES 2017
siliconindia

TOP 10 PROMISING IoT TRAINING INSTITUTES IN INDIA 2017
HIGHER EDUCATION Review

Best Workplace Amongst Emerging Enterprises Awarded By
Great Indian Workplace Awards – 2018
Mumbai, India

Top 10 IoT Training Companies for the **Second consecutive year.**
2017 IoT India MAGAZINE

Best Training Provider of the Year
Awarded By
The Golden Globe Tiger Awards – 2018
Kuala Lampur, Malaysia

World HRD Congress has awarded for Excellence in Training (Asia)

WORLD HRD CONGRESS
Cognixia is awarded as Training Company of the Year, 2018

ISO 9001 2015 CERTIFIED
ISO 9001:2015 Certified Quality Management System

ISO
ISO/IEC 27001:2013 Certified Information Security Management System

AFRICA
Best Emerging Technology Training Organization at the Middle East Training & Development Leadership Awards 2018

SOME FORMIDABLE NAMES AS

OUR TRAINING PARTNERS

Red Hat | MAPR | PARTNERWORKS CONSULTANT COMMUNITY | Silver Microsoft Partner

AUTOMATION ANYWHERE | ITIL® | (ISC)²®

WHO SHOULD STUDY BIG DATA?

The market for Big Data analytics is growing across the world and it provides a great opportunity for all IT Professionals.

Being trained in Big Data could be highly beneficial for -

- Software Developers and Architects
- BI/ETL/DW Professionals
- Senior IT Professionals
- Testing & Quality Assurance Professionals
- Mainframe Professionals
- Database Administrators
- Program Managers, Engagement & Relationship Managers
- Business Analysts, Sales Enablement team
- Freshers in the IT domain

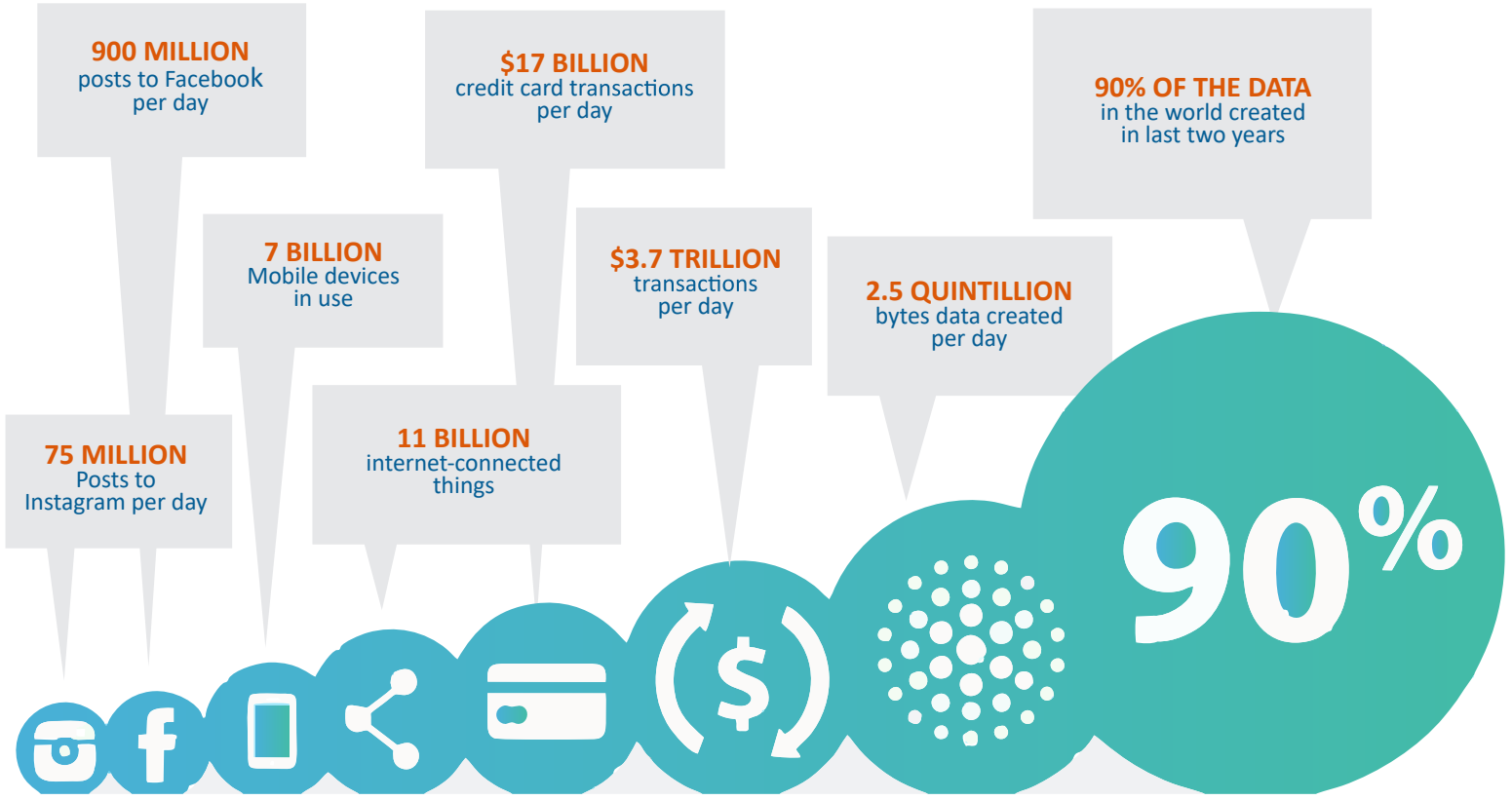
ELIGIBILITY / PREREQUISITES

- Basic Computer Knowledge
- Core Java/SQL will be an added advantage, however, not mandatory

THE WORLD OF BIG DATA

Big data is a term used for data sets that are so large or complex that traditional data processing applications and software are inadequate to deal with them. Challenges in dealing with such huge volumes of data include capturing, storing, analyzing, curating, searching, sharing, transferring, visualizing, querying, updating the data and keeping it private. With each passing year, the amount of data required to store and analyze grows exponentially.

Organizations increasingly want to aggregate all data about their users and analyze it to derive valuable insights from it. To be able to process such humongous amounts of data, special technological tools get employed that distribute multiple computations and make the process more efficient.
Big Data comes with its own set of tools, technologies and frameworks that have the capability to efficiently crunch massive amounts of data and derive valuable insights out of it that would be immensely useful for organizations.



TRENDS IN BIG DATA HADOOP DEVELOPER

"Hadoop Market is expected to reach $99.31 bn by 2022 at a CAGR of 42.1%." Forbes

“Average Salary of Big Data Hadoop Developers is $135k." Indeed.com Salary Data

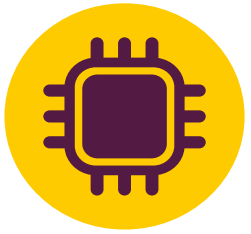
"McKinsey predicts that by 2018 there will be a shortage of 1.5 mn data experts." McKinsey

"The survey revealed that 48 percent of companies have invested in big data in 2016." Gartner

TOOLS AND FRAMEWORKS USED

We provide 42 hours of live online training including live point of contact & assignments.

It would be live & interactive online session with Industry Expert Instructor.



Apache Hadoop Ecosystem.

APACHE HADOOP

APACHE MAP REDUCE

APACHE HIVE

APACHE PIG

APACHE SPARK

APACHE SQOOP

APACHE FLUME

COURSE OBJECTIVES

Massive amount of data are being generated every day and everywhere. As a result, a number of organisations are focusing on big data processing.

The chief objectives of this course are -

• Help the learner understand how Hadoop, as an ecosystem, helps store, process, and analyze data

• Familiarize the learner with the context of big data within a data warehouse - structured, unstructured data, and raw data

• Help the learner comprehend the process of developing large-scale distributed data processing applications using Hadoop, Map Reduce, Apache Pig, Apache Hive and Apache Spark

• Familiarize the learner with the entire process of designing and building data applications that can visualize, navigate, and interpret humongous amounts of data

DETAILED CURRICULUM : MODULES

Introduction:

Become an expert in Hadoop by getting hands-on knowledge of MapReduce, Hadoop Architecture, Pig & Hive, Oozie, Flume and Apache workflow scheduler. Also, get familiarized with HBase, Zookeeper, and Sqoop concepts while working on industry-based, use-cases and projects.

Understanding Big Data

- 4V (Volume, Velocity, Variety and Veracity) characteristics
- Structured and Unstructured Data
- Application and use cases of Big Data
- Limitations of traditional large Scale systems
- How a distributed way of computing is superior (cost and scale)
- Opportunities and challenges with Big Data

Understanding Linux

- Introduction to Linux and Big Data Virtual Machine (VM)
- Introduction to Linux - Why Linux?
- Windows and the Linux equivalents
- Different favors of Linux
- Unity Shell (Ubuntu UI)
- Basic Linux
- Commands (enough to get started with Hadoop)

HDFS (The Hadoop Distributed File System)

- HDFS Overview and Architecture
- Deployment Architecture
- Name Node, Data Node and Checkpoint Node (aka Secondary Name Node)
- Safe mode
- Configuration fles
- HDFS Data Flows (Read v/s Write)

DETAILED CURRICULUM : MODULES

Advanced HDFS Features
- Load Balancer
- Dist Cp
- HDFS Federation
- HDFS High Availability
- Hadoop Archives

How HDFS Addresses Fault Tolerance?
- CRC Checksum
- Data replication
- Rack awareness and Block placement policy
- Small fles problem

HDFS Interfaces
- Command Line Interface
- File System
- Web Interface

MapReduce Architecture
- Legacy MR v/s Next Generation MapReduce (aka YARN/ MRv2)
- Slots v/s Containers
- Schedulers
- Shuffing, Sorting
- Hadoop Data Types
- Input and Output Formats
- Input Splits - Partitioning (Hash Partitioner v/s Customer Partitioner)

Optimization techniques
- Speculative execution
- Combiners
- JVM Reuse
- Compression

DETAILED CURRICULUM : MODULES

MR Algorithms (Non-graph)

- Word Count
- Term Frequency
- Inverse Document Frequency
- Log Data Analysis
- Different ways of joining data
- Purchases Data Analysis
- Max Temperature

MR Algorithms (Graph)

- PageRank
- Inverted Index

Higher Level Abstractions for MR (Pig)

- Introduction and Architecture
- Different Modes of executing Pig constructs
- Data Types
- Dynamic invokers Pig streaming Macros
- Pig Latin language Constructs (LOAD, STORE, DUMP, SPLIT, etc)
- User Defined Functions
- Use Cases

Higher Level Abstractions for MR (Hive)

- Introduction and Architecture
- Different Modes of executing Hive queries
- Metastore Implementations
- HiveQL (DDL & DML Operations)
- External v/s Managed Tables
- Views Partitions & Buckets
- Joins, Group by, Order by
- User Defined Functions

NoSQL Databases (Theoretical Concepts)

- Review of RDBMS

DETAILED CURRICULUM : MODULES

- Need for NoSQL
- Brewers CAP Theorem
- ACID v/s BASE
- Schema on Read vs. Schema on Write
- Different levels of consistency

Different Types of NoSQL Databases
- Key Value
- Columnar
- Document
- Graph

Apache HBase
- HBase Architecture
- Master and Region Server
- Catalog Tables (Root and Meta)
- HBase Data Modeling
- Loading data in HBase

Data Ingestion Tools
- Apache Sqoop
- Data movement from Relational databases to Hadoop
- Sqoop Commands
- Sqoop Advanced features
- Apache Flume
- Components of Flume
- Log Data ingestion to Hadoop

Apache Spark
- Introduction to RDD
- Installation and Configuration of Spark
- Spark Architecture
- Different interfaces to Spark

DETAILED CURRICULUM : MODULES

- Data frames and Datasets
- Querying massive data using SparkSql
- Sample Python programs in Spark
- Data Visualization using Apache Zeppelin

Big Data on Cloud

- Cloudera Hadoop cluster on the Amazon Using EMR (Elastic Map Reduce)
- Using EC2 (Elastic Compute Cloud)

Hadoop Industry Solutions

- Importing/exporting data across RDBMS and HDFS using Sqoop
- Getting real-time events into HDFS using Flume
- Creating workflows in Oozie
- Introduction to Graph processing
- Graph processing with Neo4J
- Processing data in real time using Storm
- Interactive Adhoc querying with Impala

COURSE PROJECT WORK

CASE STUDY #1– "Retail Data Analysis"

More and more organizations have developed the capacity to gain greater insight into customer behavior. It is now essential to use innovative analytics practices to succeed.

However, having all the necessary data and processing it into useful information and insights are two different things. As big data analytics tools become more affordable and usable, the technology's influence is expanding in the Retail sector.

Technologies such as HDFS, Map-Reduce, Hive, Pig, Spark, Oozie, Sqoop are used to extract and analyze the necessary data. This case study covers use cases that determine femand forecasting of products, peak time forecasting, payment mode analysis, customer classification and top products.

CASE STUDY #2– "Click Stream Analysis"

Many e-commerce websites have been making quite an impact on the overall economy in many countries for quite some time.

All the e- commerce portals store the user activities on their site as clickstream activity. This data is later analyzed to identify what the user browsed and show the appropriate recommendations to the user when they visit the site again or to send a personalized email.

This case study shows how to analyse the clickstream and the user data together using Pig and Hive. The user data would be from RDBMS and the user behavior (clickstream) data would be got using Flume into HDFS. The case goes on to perform some interesting analysis using both Hive and Pig. The above mentioned Click Stream Analysis will also be automated using the workflow engine Oozie in this case.

COGNIXIA's KEY DIFFERENTIATORS

- LIFETIME LMS ACCESS
- 24 x 7 SUPPORT
- REAL-LIFE PROJECTS & CASE STUDIES
- INDUSTRY EXPERTS AS TRAINERS
- INDUSTRY STANDARD CERTIFICATE

EXIT PROFILE

BIG DATA DEVELOPER

HADOOP DEVELOPER

TESTIMONIALS

VIJAY KUMAR GUPTA, BANGALORE, INDIA

I would really like to thank the Cognixia trainer for providing this training on Hadoop ecosystem focusing on the concepts and their implementation practically. Every time I go through the session recordings , I am able to understand it even better. The good part is the way he explained the concepts with such simplicity and not using any difficult to understand vocabulary, which made it very easy to understand, especially for beginners like me.

RENUKA PATEL, HYDERABAD, INDIA

It was a great initiative from Cognixia to offer training in Big Data. It was indeed a great experience for me. The training was interactive & had lot of practical aspects covered in it. I really enjoyed learning a new technology. Thanks Cognixia team for all your support & help."

KRITI KRISHNA, MUMBAI, INDIA

I was always fascinated with new upcoming technologies. Through my research on the Internet, i came to know about the Big Data Hadoop Technology. I discovered Cognixia then, inquired about the details and gosh!!! Guess what? I got trained by an industry expert on being a Big Data Hadoop Developer.

RICHARD ROMO, NY, USA

I would like to say thanks to Cognixia for conducting the training on Big Data Hadoop Developer, and the quality of training is awesome. I got good knowledge on the hadoop development and the other components of hadoop like Hive, Pig, Hbase, Oozie, Sqoop and many more. It was awesome experience for me.

KRITI KRISHNA, MUMBAI

The training module that Cognixia possesses is excellent! Praveed Sripati's (the trainer) name was enough for me to get enrolled for the training on Big Data Hadoop technology. I had been following his blogs and to interact with him as part of this training was a superb experience.

OLIVER THOMSON, EDINBURGH, UNITED KINGDOM

The state-of-the-art infrastructure and access to the same as part of my training with Cognixia gave me a better understanding and hands-on experience on the practical part of the Big Data technology.

Big Data Hadoop Developer Training

To learn more visit
https://www.cognixia.com/