ADVANCED
BIG DATA AND HADOOP
COURSE CURRICULUM

YOUR JOB HUNTING ENDS HERE
Start your career with Big Data and Hadoop course that gets Your dream job!
Become A Big Data & Hadoop Certified Professional

It is only skills and not degree that can help you grow. But if you are one of those individuals who believe in getting certified along with skills then we have got you covered. After completion of the training not only will you become an expert in Big Data but you will also be a Big Data & Hadoop certified professional.

One Training Program
4 Certifications

- CCA ADMINISTRATOR
- CCA SPARK AND HADOOP DEVELOPER
- CCP DATA ENGINEER
- Industry Recognized INVENTATEQ Certificate
GET TRAINING AND GET EMPLOYED

CLASS ROOM TRAINING
ONLINE TRAINING
CORPORATE TRAINING

23,409+ Trainees
500+ Batches
4.9/5 Ratings
Inventateq BTM, it's indeed a great platform to learn Big Data Hadoop for fresher's n working professionals as well. The trainer, shivank. He been the best with excellent subject knowledge and gr8 real time experience. His interaction amongst learners made the session even more effective. He focused more on practical sessions which really helped us a lot. Quality of learning is gr8 with good placement assistance. Highly recommended.

DEEPA NE LIGI

Thank you, Shivank Sir, for all your help & support throughout the journey for Hadoop and Big Data technologies your assignments and real time scenarios help me to crack interviews. Got offer in hand.

RAVINDRA RAGHUWANSHI

A good institute to boost your Hadoop knowledge. Helped me a lot to gain a good knowhow of Hadoop ecosystem. Simply best institute in BTM for big data.

MANOHAR JISHU

I have taken Hadoop Course with INVENTATEQ early this year where the Instructor was very knowledgeable with excellent communication and presentation skills with awesome time management in wrapping up of each session with real time scenarios. Now, I am with full confidence that I can crack any interview with my enhanced skills in this technology and get placed in Infosys.

NANI VINOD
23,409+ Success Stories
Here is what they say about INVENTATEQ.

I joined to learn HADOOP and got good understanding of the subject. Classes are interactive and teaching methodology is very good. Recommended for anyone who’s looking for this course.

LAKSHIMI ROY

Inventateq marathahalli is one of the best software training institute for fresher’s to start their journey in the field of Hadoop and big data. Here taught us with real time experience, hands on experience was really great. I suggest Inventateq is the best platform for fresher’s to start their career.

KOMARA SUMA

After completing graduation I have joined for big data and Hadoop course in Inventateq Marathahalli. It was amazing experience with real time classes, and with one hands on project. I suggest the Inventateq institute if you are a fresher and want to become expert in Hadoop.

VIJAYA LAKSHIMI

Best instructor + Best study material + 100% support is what made my experience with Inventateq very amazing and satisfied too. Can't recommend a better institute. Not only did i get to learn Data Analytics Hadoop but i got 3 great job opportunities within no time and i chose the best amongst them.

LAKSHMI TEJU
Training Process

- Theory
- Practical
- Assignments
- Certification
- Resume Preparation
- Attend Interview

YOU GOT THE JOB!
# DETAILED SYLLABUS

## TABLE OF CONTENT

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Big Data</td>
</tr>
<tr>
<td>2</td>
<td>Hadoop Admin</td>
</tr>
<tr>
<td>3</td>
<td>Hadoop Developer</td>
</tr>
<tr>
<td>4</td>
<td>Python for Hadoop</td>
</tr>
<tr>
<td>5</td>
<td>Java for Hadoop</td>
</tr>
<tr>
<td>6</td>
<td>SQL for Hadoop</td>
</tr>
<tr>
<td>7</td>
<td>Big data with Spark</td>
</tr>
</tbody>
</table>
Module 1

Big Data introduction and Hadoop

- Fundamental
- Data Storage and Analysis
- Comparison with RDBMS

HDFS ARCHITECTURE

- Basic Terminologies
- HDFS Block Concepts
- Replication Concepts
- Basic reading & writing of files in HDFS
- Basic processing concepts in MapReduce
- Data Flow
- Anatomy of file READ and WRITE
Module 2
HADOOP ADMINISTRATOR

- HADOOP GEN1 VS HADOOP GEN 2(YARN)
- Linux commands
- Single and Multinode cluster installation (HADOOP Gen 2)
- AWS (EC2, RDS, S3, IAM and Cloud formation)
- Cloudera and Hortonworks distribution installation on AWS
- Cloudera Manager and Ambari
- Hadoop Security and Commissioning and Decommissioning of nodes
- Sizing of Hadoop Cluster and Name Node High Availability
Module 3
DATA INGESTION

Sqoop:
- Migration of data from MYSQL/ ORACLE to HDFS.
- Creating SQOOP job.
- Scheduling and Monitoring SQOOP job using OOZIE and Crontab.
- Incremental and Last modified mode in sqoop.

Talend:
- Installation of Talend big data studio on windows server.
- Creating and Scheduling talend Jobs.
- Components: tmap, tmssqlinput, tmssqloutput, tFileInputDelimited,
  tfileoutputdelimited, tmssqloutputbulkexec, tunique,
  tFlowToIterate, tIterateToFlow, tlogcatcher, tflowmetercatcher, tfilelist,
  taggregate, tsort, thdfsinput, thdfsoutput, tFilterRow, thiveload.

Flume:
- Flume Architecture
- Data Ingest in HDFS with Flume
- Flume Sources
- Flume Sinks
- Topology Design Considerations
Module 4
DATA PROCESSING

MapReduce:

- Env Setup
- Tool and ToolRunner
- Mapper
- Reducer
- Driver program
- How to package the job?
- MapReduce WebUI
- How MapReduce Job run?
- Shuffle & Sort
- Speculative Execution
- InputFormats
- Input Splits and Record Reader
- Default Input Formats
- Implement Custom Input Format
- OutputFormats
- Default Output formats
- Output Record Reader
- Compression
- Map Output
- Final Output
- Data types – default
- Writable vs Writable Comparable
- Custom Data types – Custom Writable/Comparable
- File Based Data structures
- Sequence file
- Reading and Writing into Sequence file
- Map File
- Tuning MapReduce Jobs
- Advanced MapReduce
- Sorting
- Partial Sort
- Total Sort
- Secondary Sort
- Joins

**Hive:**
- Comparison with RDBMS
- HQL
- Data types
- Tables
- Importing and Exporting
- Partitioning and Bucketing – Advanced.
- Joins and Join Optimization.
- Functions- Built in & user defined
- Advanced Optimization of HQL
- Storage File Formats – Advanced
- Loading and Storing Data
- SerDes – Advanced

**Pig:**
- Important basics
- Pig Latin
- Data types
- Functions – Built-in, User Defined
- Loading and Storing Data
Module 5

PYTHON FOR HADOOP

1: An Introduction to Python
   1.1 Brief about the course
   1.2 History/timelines of python
   1.3 What is python?
   1.4 What python can do?
   1.5 How the name was put up as python
   1.6 Why python?
   1.7 Who all are using python
   1.8 Features of python
   1.9 Python installation
   1.10 Hello world
      1. using cmd
      2. IDLE
      3. By py script
      4. python command line
2: Beginning Python Basics
   2.1. The print statements
   2.2. Comments
   2.3. Python Data Structures
   2.4. variables & Data Types
     1. rules for variable
     2. declaring variables
     3. Assignment in variables
     4. operations with variables
     5. Reserved keyword
   2.5. Operators in Python
   2.6. Simple Input & Output
   2.7. Examples for variables, Data Types, operators
3: Python Program Flow
   3.1. Indentation
   3.2. The If statement and its related statement
   3.3. An example with if and it's related statement
   3.4. The while loop
   3.5. The for loop
   3.6. The range statement
   3.7. Break
   3.8. Continue
   3.9. pass
   3.9. Examples for looping
4: Functions & Modules
   4.1. system define function(number system and its sdf, string and its sdf)
   4.2. Create your own functions (user define function)
   4.3. Functions Parameters
   4.4. Variable Arguments
5: Exceptions
   5.1. Errors
   5.2. Exception Handling with try
   5.3. Handling Multiple Exceptions
   5.4. raise
   5.5. finally
   5.6. else

6: File Handling
   6.1. File Handling Modes
   6.2. Reading Files
   6.3. Writing & Appending to Files
   6.4. Handling File Exceptions

7: Data Structures and Data Structures functions
   7.1. List and its sdf
   7.2. tuple and its sdf
   7.3. Dictionary and its sdf
   7.4. set and its sdf
   7.5. use cases and practical examples

8: casting
   8:1 intro to casting
Module 6

NOSQL

Cassandra:

- Cassandra cluster installation
- Cassandra Architecture
- Cqlsh
- Replication strategy
- Tools: Opscenter, Nodetool and CCM
- Cassandra use cases

Labs:

- Real Time use cases and Data sets covered (10+ Real Time datasets)
- Word count, Sensors (Weather Sensors) Dataset, Social Media data sets like YouTube, Twitter data analysis
Module 7
Scala & Spark training
Course Outline

Spark Batch processing API

Introduction
- Why Spark?
- Evolution of Distributed systems
- Challenges with existing distributed systems
- Need of new generation
- Hardware/software evolution in last decade
- Spark History
- Unification in Spark
- Spark ecosystem vs Hadoop
- Spark with Hadoop
- Who are using Spark?

Scala Basics Required for Spark
- Spark Architecture
- RDD
- Immutability
- Laziness
- Type inference
- Cacheable
- Spark on cluster management frameworks
- Spark task distribution

Spark installation
- Local
- Spark on YARN
- Stand alone
- Spark on Mesos
Spark API Hands on
- RDD operations
- Key-value pair RDD
- Map Reduce
- Double RDD

Advanced operations
- Aggregate
- Fold
- mapPartitions
- glom
- Broadcasters

Integration with HDFS
- Introduction to HDFS
- HDFS architecture
- Using HDFS

Caching and Lineage
- RDD caching
- Fault recovery

Spark streaming API
- Introduction
- Spark streaming Architecture
- DStreams
- DStream vs RDD
- Receivers
- Batch vs Streaming

Input Streams
- Socket
- HDFS
- Twitter
- Kafka
Spark API Hands on
- RDD operations
- Key-value pair RDD
- Map Reduce
- Double RDD

Advanced operations
- Aggregate
- Fold
- mapPartitions
- glom
- Broadcasters

Integration with HDFS
- Introduction to HDFS
- HDFS architecture
- Using HDFS

Caching and Lineage
- RDD caching
- Fault recovery

Spark streaming API
- Introduction
- Spark streaming Architecture
- DStreams
- DStream vs RDD
- Receivers
- Batch vs Streaming

Input Streams
- Socket
- HDFS
- Twitter
- Kafka

Streaming API Hands-on
- DStream creation
- Transformations
- Stateful operations
Check pointing
- Recoverable computations
- Error handling

Combining batch and Streaming
- Foreach
- Transform
- Joins

Persist and Caching
- Saving DStream
- Caching DStream

Window Operations
- window
- countByWindow
- reduceByWindow

Deploying Spark Streaming
- Clustering
- Check pointing
- Driver fallback

Spark SQL
- Introduction

Spark SQL DDL
- Case classes
- Inferred schema
- Parquet files
- JSON
- Schema RDD

Spark SQL DML
- Projection
- Condition
- groupBy
- joins
- partitioning
Extending Spark SQL
- User defined functions
- User defined aggregate function

Spark SQL in Streaming
- Querying DStreaming
- DStream joins

Demo Project
- Ecommerce Log Analytics using Kafka, Spark Streaming and Cassandra.

Master Project:
Inventateq

- Real-time Data Warehouse migration:
- Real-time concepts covered are
- Hive - Advanced topics
- Sqoop import/export
- Oozie Scheduling
- How Hadoop MR/Spark used in DW
- RDBMS concepts
- ETL tool concepts
- Integration with Reporting tools
OUR HIRING PARTNERED COMPANIES LIST

Mphasis
Unleash the Next

imshealth
INTELLIGENCE APPLIED.

SQBE
INFOTECH

Tech Mahindra

3i Infotech
Innovation • Insight • Integrity

Collabera
Value. Accelerated.

Inventateq
Transformation Happens Here

genpact

hp

Mphasis
The Next Applied

HCL

ITC INFOTECH
Business-friendly Solutions

R SYSTEMS
Caring - Innovating - Transforming

100% JOB ORIENTED TRAINING COURSES WE PROVIDE
<table>
<thead>
<tr>
<th>Digital Marketing (SEO/Social Media/PPC Google Adwords)</th>
<th>BigData Hadoop Course</th>
<th>Machine Learning Certification Training</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Best SEO Training</strong></td>
<td>Spark and Scala Course</td>
<td><strong>Block Chain Training Courses</strong></td>
</tr>
<tr>
<td><strong>AWS Training</strong></td>
<td>DevOps Training</td>
<td><strong>Artificial Intelligence Courses</strong></td>
</tr>
<tr>
<td><strong>Cloud Computing Training</strong></td>
<td>Angularjs and Node JS Training</td>
<td><strong>Tally ERP &amp; GST Accounting classes</strong></td>
</tr>
<tr>
<td><strong>Data Science Courses</strong></td>
<td>Weblogic Training</td>
<td><strong>Java course</strong></td>
</tr>
<tr>
<td><strong>RPA Training</strong></td>
<td>.NET Technologies</td>
<td><strong>Software Testing Course, Manual Testing, QTP, UFT, Loadrunner</strong></td>
</tr>
<tr>
<td><strong>Internet of Things IoT Training</strong></td>
<td>SOA Suite 11g</td>
<td><strong>C C++ Course</strong></td>
</tr>
<tr>
<td><strong>Microsoft Azure Training</strong></td>
<td>Oracle DBA Training</td>
<td><strong>Tableau</strong></td>
</tr>
<tr>
<td><strong>Data Warehousing - Informatica</strong></td>
<td>Oracle SQL, PLSQL, DBA, D2k, Apps</td>
<td><strong>PHP MYSQL, Python</strong></td>
</tr>
<tr>
<td><strong>Selenium Training</strong></td>
<td>ETL Testing Course</td>
<td><strong>Human Resources Classes</strong></td>
</tr>
<tr>
<td><strong>IBM Cognos 10 BI &amp; Cognos TM1</strong></td>
<td>PPC Training Institute</td>
<td><strong>Microstartegy Course</strong></td>
</tr>
<tr>
<td><strong>Qlikview (Designer, Developer, Publisher, Server)</strong></td>
<td>Autodesk Revit Training</td>
<td><strong>Cisco CCNA Networking</strong></td>
</tr>
<tr>
<td><strong>IBM Websphere</strong></td>
<td>Autodesk CAD 2d and 3d Course</td>
<td><strong>SAS Training</strong></td>
</tr>
<tr>
<td><strong>Learn ODI 11g</strong></td>
<td>Catia Training</td>
<td><strong>Softskill Courses</strong></td>
</tr>
<tr>
<td><strong>Python Training</strong></td>
<td>Wiring Harness Training</td>
<td><strong>ITIL Certificate Training</strong></td>
</tr>
<tr>
<td><strong>Abinitio Classes</strong></td>
<td>Content Writing Training</td>
<td><strong>SAP Courses</strong></td>
</tr>
<tr>
<td><strong>Web Designing Course</strong></td>
<td>Siebel CRM Courses</td>
<td><strong>OBIEE</strong></td>
</tr>
<tr>
<td><strong>Data Stage Classes</strong></td>
<td>Clear Case Training</td>
<td></td>
</tr>
</tbody>
</table>

**4 CENTERS AND COUNTING**

No. 1578, 2nd Floor, 11th Main Road, Jayanagar 4th 'T' Block, Next to Jayanagar Head Post-office
IT IS TIME 
YOU GET STARTED 
WITH FREE DEMO CLASSES

www.inventateq.com
info@inventateq.com
76-76-76-5421
Week Days Batches
Weekend Batches