Big Data Analytics for Research Training Course

Professional Development Training Course Outline

Category	Research and Data Analysis	**Duration**	5 Days
Base Fee	$1,500 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

In today’s data-driven world, Big Data Analytics plays a vital role in accelerating research productivity, enhancing data insights, and enabling evidence-based decision-making. Big Data Analytics for Research Training Course equips professionals, researchers, and data scientists with advanced skills in Hadoop and Spark ecosystems, focusing on real-world data analysis, machine learning applications, and distributed computing. Participants will gain hands-on experience with structured and unstructured data, unlocking the full potential of big data frameworks for research optimization.

The course bridges the gap between data science theory and practical application using tools like HDFS, MapReduce, Hive, Pig, Spark MLlib, and GraphX. Whether in academic, public, or corporate sectors, learners will master data wrangling, analytics, visualization, and predictive modeling through dynamic case studies from healthcare, finance, environment, and education domains. This course is designed for those seeking to harness big data for strategic research innovation and real-time solutions.

Programme Curriculum

Big Data Analytics for Research Training Course

Introduction

In today’s data-driven world, Big Data Analytics plays a vital role in accelerating research productivity, enhancing data insights, and enabling evidence-based decision-making. Big Data Analytics for Research Training Course equips professionals, researchers, and data scientists with advanced skills in Hadoop and Spark ecosystems, focusing on real-world data analysis, machine learning applications, and distributed computing. Participants will gain hands-on experience with structured and unstructured data, unlocking the full potential of big data frameworks for research optimization.

The course bridges the gap between data science theory and practical application using tools like HDFS, MapReduce, Hive, Pig, Spark MLlib, and GraphX. Whether in academic, public, or corporate sectors, learners will master data

wrangling, analytics, visualization, and predictive modeling through dynamic case studies from healthcare, finance, environment, and education domains. This course is designed for those seeking to harness big data for strategic research innovation and real-time solutions.

Course Objectives

1. Understand the fundamentals of Big Data and Hadoop architecture
2. Explore the core components of the Apache Spark ecosystem
3. Learn data ingestion and data preprocessing for research datasets
4. Apply MapReduce algorithms to real-world problems
5. Use HDFS for efficient data storage and management
6. Execute real-time data analytics using Spark Streaming
7. Perform machine learning on large datasets with Spark MLlib
8. Create data visualizations using Hive and Tableau
9. Implement data querying with Pig and Hive
10. Develop predictive models for academic research and policy planning
11. Analyze graph data using Spark GraphX for social network analysis
12. Build data pipelines for research automation
13. Understand and apply data ethics and privacy in big data research

Target Audiences

1. Academic researchers in data-intensive fields
2. Data scientists and analysts
3. University faculty and PhD scholars
4. Government research officers
5. ICT and analytics professionals
6. Research think-tanks and NGOs
7. Graduate and postgraduate students
8. Corporate R&D teams

Course Duration: 5 days

Course Modules

Module 1: Introduction to Big Data & Hadoop Framework

- Big Data evolution and use cases
- Components of the Hadoop ecosystem
- HDFS architecture and fault tolerance
- Understanding MapReduce paradigm
- Setting up Hadoop environment
- **Case Study**: Analyzing healthcare datasets with Hadoop

Module 2: Working with Apache Spark

- Spark vs. Hadoop: Key differences
- Spark Core and architecture
- Running Spark on YARN
- Spark Shell and RDD fundamentals
- Transformations and actions in Spark
- **Case Study**: Climate change modeling using Spark

Module 3: Data Ingestion and Preprocessing

- Data ingestion with Sqoop and Flume

- Real-time ingestion using Kafka
- Data cleaning and transformation pipelines
- ETL processes with Spark
- Schema evolution in Hive
- **Case Study**: Preparing financial data for risk modeling

Module 4: Data Querying and Hive/Pig

- HiveQL: Structured querying
- Partitioning and bucketing
- UDFs and joins in Hive
- Pig scripting basics
- Use cases for Pig vs Hive
- **Case Study**: Educational performance analysis using Hive

Module 5: Machine Learning with Spark MLlib

- Overview of MLlib algorithms
- Classification and clustering techniques
- Feature extraction and selection
- Model evaluation and tuning
- Building ML pipelines
- **Case Study**: Predicting public health trends

Module 6: Real-Time Analytics with Spark Streaming

- Architecture of Spark Streaming
- DStreams and window operations
- Integrating Kafka with Spark Streaming
- Handling real-time research feeds
- Monitoring and debugging
- **Case Study**: Social media data for crisis management

Module 7: Graph Processing with GraphX

- Introduction to graph theory
- GraphX operators and optimizations
- PageRank and connected components
- Graph data modeling in research
- Visualizing research networks
- **Case Study**: Mapping academic citation networks

Module 8: Data Ethics, Security & Research Compliance

- Ethical principles in data science
- Ensuring data privacy and anonymization
- Compliance frameworks (GDPR, HIPAA)
- Risk assessment in data sharing
- Governance in academic research
- **Case Study**: Ethical handling of sensitive health records

Training Methodology

- Instructor-led live virtual sessions
- Hands-on practical exercises with datasets
- Collaborative research-based projects

- Interactive quizzes and assessments
- Personalized feedback and one-on-one mentorship
- Final capstone project using Spark and Hadoop

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.org or call *+254769199797*

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

f. Payment should be done at least a week before commence of the training, to Fineskill Training Center account, as indicated in the invoice so as to enable us prepare better for you.

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	27 Nov 2026	Online	$1,000
23 Nov 2026	27 Nov 2026	Physical	$1,500
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$6,000
23 Nov 2026	27 Nov 2026	Physical	$4,500
23 Nov 2026	27 Nov 2026	Physical	$6,000

Start Date	End Date	Location	Price
23 Nov 2026	27 Nov 2026	Physical	$6,000

Ready to enroll or request a customized program? Book online or contact us:

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com