Training Course on Big Geospatial Data Analytics with Spark and Hadoop

Professional Development Training Course Outline

Category	GIS	**Duration**	10 Days
Base Fee	$3,000 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

The convergence of **Big Data** and **Geospatial Analytics** is revolutionizing how organizations understand and interact with the world. This intensive training course delves into the powerful synergy of **Apache Spark** and **Apache Hadoop** for processing, analyzing, and visualizing massive volumes of **location-based data**. Participants will gain expertise in handling diverse geospatial datasets, from **satellite imagery** and **IoT sensor data** to **GIS layers** and **GPS tracks**, enabling them to extract actionable insights and make data-driven decisions across various domains like **urban planning**, **environmental monitoring**, **logistics optimization**, and **disaster management**.

Training Course on Big Geospatial Data Analytics with Spark and Hadoop provides a comprehensive understanding of **distributed computing frameworks** for geospatial applications. Leveraging Spark's in-memory processing capabilities and Hadoop's robust storage and batch processing, attendees will master techniques for **spatial data warehousing**, **real-time geospatial streaming**, **machine learning for spatial patterns**, and advanced **geospatial visualization**. Through practical exercises and real-world **case studies**, participants will develop the skills to design, implement, and optimize scalable solutions for complex geospatial challenges, positioning them at the forefront of **geospatial AI** and **location intelligence**.

Programme Curriculum

Training Course on Big Geospatial Data Analytics with Spark and Hadoop

Introduction

The convergence of **Big Data** and **Geospatial Analytics** is revolutionizing how organizations understand and interact with the world. This intensive training course delves into the powerful synergy of **Apache Spark** and **Apache Hadoop** for processing, analyzing, and visualizing massive volumes of **location-based data**. Participants will gain expertise in handling diverse geospatial

datasets, from **satellite imagery** and **IoT sensor data** to **GIS layers** and **GPS tracks**, enabling them to extract actionable insights and make data-driven decisions across various domains like **urban planning**, **environmental monitoring**, **logistics optimization**, and **disaster management**.

Training Course on Big Geospatial Data Analytics with Spark and Hadoop provides a comprehensive understanding of **distributed computing frameworks** for geospatial applications. Leveraging Spark's in-memory processing capabilities and Hadoop's robust storage and batch processing, attendees will master techniques for **spatial data warehousing**, **real-time geospatial streaming**, **machine learning for spatial patterns**, and advanced **geospatial visualization**. Through practical exercises and real-world **case studies**, participants will develop the skills to design, implement, and optimize scalable solutions for complex geospatial challenges, positioning them at the forefront of **geospatial AI** and **location intelligence**.

Course Duration

10 days

Course Objectives

1. Understand the fundamental concepts of Big Data technologies (Hadoop, Spark) and their application to geospatial data.
2. Differentiate between various geospatial data formats (vector, raster) and their implications for big data processing.
3. Set up and configure a Spark and Hadoop cluster for geospatial analytics workloads.
4. Ingest and manage diverse big geospatial datasets using HDFS, Hive, and other relevant tools.
5. Perform scalable spatial data transformations and geoprocessing operations using Spark.
6. Implement advanced geospatial queries and spatial joins on large datasets with Spark SQL.
7. Develop real-time geospatial streaming applications using Spark Streaming for live location intelligence.
8. Apply machine learning algorithms (e.g., clustering, classification, regression) to identify spatial patterns and make predictions.
9. Utilize geospatial libraries and extensions within Spark (e.g., GeoSpark, Magellan) for enhanced analytical capabilities.
10. Optimize Spark and Hadoop performance for efficient processing of large-scale geospatial data.
11. Visualize complex geospatial insights using various mapping and data visualization tools.
12. Troubleshoot common issues and best practices in big geospatial data analytics.
13. Design and architect end-to-end big geospatial data pipelines for various industry use cases.

Organizational Benefits

- Leverage **location intelligence** for more informed strategic and operational decisions.
- Optimize processes in logistics, resource management, and urban planning through **geospatial insights**.
- Utilize **open-source big data frameworks** to reduce infrastructure and processing costs associated with large datasets.
- Identify market trends and customer needs through **spatial data analysis**, fostering innovation.
- Better assess and respond to risks in areas like **disaster preparedness** and **environmental monitoring**.

- Build in-house expertise in cutting-edge **geospatial big data technologies**, staying ahead of the curve.
- Implement solutions capable of handling ever-increasing volumes of **geospatial information**.
- Transition to a more **data-centric approach** for addressing geographical challenges.

Target Audience

- Data Scientists and Data Analysts.
- GIS Professionals and Geospatial Analysts
- Big Data Engineers and Architects.
- Software Developers
- Researchers and Academics.
- Business Intelligence Professionals.
- Urban Planners and Environmental Scientists.
- Anyone involved in projects requiring scalable geospatial data processing and analysis.

Course Outline

Module 1: Introduction to Big Geospatial Data and Ecosystems

- Understanding Big Data Concepts.
- Evolution of Geospatial Data.
- Introduction to Hadoop Ecosystem
- Introduction to Apache Spark
- Challenges and Opportunities in Big Geospatial Data.
- **Case Study**: Analyzing **global climate data** (e.g., temperature anomalies, precipitation patterns) stored in HDFS for long-term trends, comparing traditional GIS approaches vs. a distributed approach.

Module 2: Hadoop for Geospatial Data Storage and Management

- HDFS Architecture for Spatial Data
- Data Ingestion Techniques.
- Hive for Geospatial Data Warehousing
- NoSQL Databases for Geospatial.
- Data Partitioning and Indexing Strategies
- **Case Study**: Storing and managing **historic urban development plans** and **property parcel data** in a distributed file system to enable large-scale historical analysis.

Module 3: Spark Core for Geospatial Data Processing

- Spark RDDs for Spatial Data
- Spark DataFrames and Datasets for Geospatial.
- Common Spatial Transformations scale.
- User-Defined Functions (UDFs) for Spatial Logic.
- Performance Tuning for Spark Geospatial Jobs.
- **Case Study**: Processing **large-scale GPS trace data** from millions of vehicles to identify common routes and congestion points using Spark RDDs and DataFrames.

Module 4: Spark SQL for Advanced Geospatial Queries

- Spark SQL for Structured Geospatial Data
- Geospatial Data Types in Spark SQL

- Spatial Joins and Relationships
- Aggregating Spatial Data.
- Integrating with External GIS Tools.
- **Case Study**: Analyzing **customer demographics** linked to **store locations** and **competitor proximity** to identify optimal expansion sites, using Spark SQL for complex spatial joins.

Module 5: Geospatial Libraries in Spark

- Introduction to GeoSpark/Apache Sedona
- Key Features of GeoSpark
- Working with ESRI ArcGIS GeoAnalytics Engine
- Other Relevant Libraries.
- Building Custom Geospatial Functionality.
- **Case Study**: Using GeoSpark to perform **real-time spatial clustering** of **anomalous sensor readings** from an environmental monitoring network to detect pollution hotspots.

Module 6: Real-time Geospatial Streaming with Spark

- Introduction to Spark Streaming.
- Integrating with Message Queues.
- Windowing Operations for Spatiotemporal Data.
- Stateful Stream Processing for Geospatial Events.
- Real-time Geospatial Alerts and Dashboards.
- **Case Study**: Monitoring **fleet logistics** in real-time by ingesting GPS data streams to optimize delivery routes and respond to traffic incidents instantly.

Module 7: Machine Learning for Geospatial Data with Spark MLlib

- Introduction to Spark MLlib.
- Geospatial Feature Engineering.
- Spatial Clustering Algorithms
- Spatial Classification and Regression.
- Model Evaluation and Deployment.
- **Case Study**: Predicting **deforestation rates** in a region based on **satellite imagery time series data** and various environmental factors using Spark MLlib.

Module 8: Graph Processing with Spark GraphX for Geospatial Networks

- Introduction to GraphX.
- Representing Geospatial Networks
- Graph Algorithms for Spatial Analysis.
- Network Centrality and Influence.
- Visualizing Geospatial Graphs.
- **Case Study**: Analyzing **urban transportation networks** to identify bottlenecks and optimize public transport routes using GraphX algorithms

Module 9: Geospatial Big Data Visualization

- Introduction to Big Geospatial Data Visualization.
- Using Spark with Visualization Tools:
- Interactive Mapping Libraries.
- Heatmaps and Density Maps.
- Time-series Geospatial Visualization.

- **Case Study**: Creating an interactive dashboard to visualize **crime hotspots** and their evolution over time in a metropolitan area based on analyzed big geospatial data.

Module 10: Performance Optimization and Best Practices

- Spark Performance Tuning.
- Data Skew and Remediation workloads.
- Efficient Spatial Joins.
- Monitoring and Debugging Spark Jobs.
- Best Practices for Geospatial Big Data Projects
- **Case Study**: Optimizing a Spark job designed to analyze **global shipping routes** and **maritime traffic patterns** for significant performance improvements.

Module 11: Cloud-based Geospatial Big Data Platforms

- Overview of Cloud Big Data Services
- Leveraging Managed Spark and Hadoop Services.
- Cloud Storage for Geospatial Data.
- Serverless Geospatial Analytics.
- Cost Optimization in Cloud Environments.
- **Case Study**: Migrating an on-premise **environmental impact assessment** pipeline for a large-scale infrastructure project to a cloud-based Spark environment, demonstrating scalability and cost benefits.

Module 12: Geospatial Data Quality and Governance

- Ensuring Geospatial Data Accuracy.
- Data Lineage and Metadata for Spatial Data
- Data Security and Privacy in Geospatial Big Data.
- Compliance and Regulatory Considerations
- Building a Geospatial Data Governance Framework
- **Case Study**: Developing a robust data quality pipeline for **crowdsourced mapping data** to identify and rectify inaccuracies, ensuring reliable outputs for navigation and urban planning.

Module 13: Advanced Geospatial Analytics Techniques

- Spatial Interpolation and Extrapolation.
- Geographically Weighted Regression (GWR)
- Network Analysis with Advanced Metrics.
- Spatio-temporal Analysis.
- Emerging Trends: Geospatial AI and Deep Learning
- **Case Study**: Performing **predictive analyt**

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Online	$2,000

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Physical	$3,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$9,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$12,000

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

Register Online Now

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