Datastat Research
Building Capacity in Training and Research

DATASTAT TRAINING IN

OFFICIAL COURSE PROSPECTUS

Center for Advanced Research & Professional Development

Nairobi Training Center • Global Virtual Academy • Corporate On-Site

DATA SCIENCE

Scalable Big Data Systems and Architecture Training Course

Comprehensive Course Syllabus, Practical Modules & Upcoming Schedules

PROGRAM SPECIFICATIONS & HIGHLIGHTS

- **DURATION**
  5 Days Practical Hands-on
- **DELIVERY MODES**
  In-Person Class & Live Virtual
- **CERTIFICATION**
  Accredited Certificate of Completion
- **INTAKES / DATES**
  Multiple Cohort Schedules Included

PREPARED FOR:

Valued Participant

Email: training@datastatresearch.org | Issued: November 24, 2026

Academic & Professional Training Year: 2026

DATASTAT TRAINING INSTITUTE

P.O. Box 102802-00101, Nairobi, Kenya

Training Centers: Nairobi • Mombasa • Kigali • Dubai • Live Virtual Academy

Tel: +254 724 527 104 | +254 734 969 612

Email: training@datastatresearch.org

www.datastatresearch.org

Scalable Big Data Systems and Architecture Training Course

Data Science • 5 Days Intensive • Accredited Certification

Course Overview

The Scalable Big Data Systems and Architecture Training Course is an advanced professional programme designed to equip participants with the knowledge and practical skills required to design, implement, manage, and optimize scalable big data architectures for modern enterprises

Key Skills & Software Tools Covered

Apache Hadoop • Apache Spark • Apache Kafka • Docker • Kubernetes

Upcoming Training Schedules

Confirmed cohort intakes and venues. Both physical in-person classroom and live virtual online seats are available.

TRAINING DATES	LOCATION / VENUE	FORMAT / DURATION	INVESTMENT (FEE)
30 Nov 2026 – 4 Dec 2026	Nairobi	Classroom / 5d	**$1,500 USD**
7 Dec 2026 – 11 Dec 2026	Nairobi	Classroom / 5d	**$1,500 USD**
14 Dec 2026 – 18 Dec 2026	Nairobi	Classroom / 5d	**$1,500 USD**
21 Dec 2026 – 25 Dec 2026	Nairobi	Classroom / 5d	**$1,500 USD**

* Corporate discounts and customized cohort scheduling are available for teams of 3 or more participants.

Course Curriculum & Modules

Scalable Big Data Systems and Architecture Training Course

Course Introduction

The Scalable Big Data Systems and Architecture Training Course is an advanced professional programme designed to equip participants with the knowledge and practical skills required to design, implement, manage, and optimize scalable big data architectures for modern enterprises. As organizations generate increasingly large volumes of structured, semi-structured, and unstructured data, traditional data-processing infrastructures often struggle with scalability, speed, availability, and cost. This course provides a comprehensive understanding of distributed computing, cloud data platforms, data lakes, data warehouses, data lakehouses, real-time data processing, parallel computing, and enterprise big data architecture.

The training emphasizes practical approaches to building high-performance, fault-tolerant, secure, and cost-effective big data systems capable of scaling as organizational data and processing requirements grow. Participants will explore technologies and architectural patterns such as Apache Hadoop, Apache Spark, Apache Kafka, cloud computing, Lambda architecture, Kappa architecture, microservices, containerization, and distributed databases. Through practical exercises and industry-based case studies, participants will learn how to translate business requirements into scalable data architectures that support advanced analytics, artificial intelligence, machine learning, digital transformation, real-time decision-making, and enterprise data management.

Learning Objectives

By the end of the Scalable Big Data Systems and Architecture Training Course, participants will be able to:

- **Explain the fundamental principles and characteristics of scalable big data systems.**
- **Design distributed architectures capable of processing and storing massive datasets.**
- **Apply principles of horizontal and vertical scalability to enterprise data systems.**
- **Design highly available, fault-tolerant, and resilient big data infrastructures.**
- **Evaluate and select appropriate big data technologies and architectural frameworks.**
- **Design and implement scalable data lakes, data warehouses, and data lakehouse architectures.**
- **Apply batch processing and real-time stream processing architectures to large-scale data environments.**
- **Integrate cloud computing, distributed databases, and container technologies into scalable data solutions.**
- **Apply data security, governance, monitoring, performance optimization, and disaster recovery principles.**
- **Develop an enterprise big data architecture roadmap aligned with organizational objectives and digital transformation strategies.**

Target Audience

This training course is suitable for:

- Big Data Engineers responsible for designing and implementing distributed data systems.
- Data Architects designing enterprise-scale data infrastructure.
- Data engineers and data scientists working with large and complex datasets.
- IT managers and enterprise technology managers.
- Cloud architects and cloud infrastructure professionals.
- Database administrators and systems administrators.
- Software developers working with distributed and data-intensive applications.
- Business intelligence and analytics professionals.
- Digital transformation, enterprise architecture, and technology consultants.
- Senior executives and technical decision-makers responsible for enterprise data strategy and big data investments.

Course Modules

Module 1: Foundations of Scalable Big Data Systems

This module introduces the principles underlying scalable big data systems and distributed data architectures. Participants explore why conventional systems struggle with massive datasets and how distributed architectures provide greater scalability, availability, processing capacity, and resilience.

Key Topics

- Introduction to scalable big data systems
- Characteristics of big data
- Scalability and elasticity
- Horizontal versus vertical scaling
- Distributed computing principles
- Parallel processing
- Distributed storage
- Fault tolerance and resilience
- High availability
- Performance and capacity planning

Case Study

Global E-Commerce Platform: Participants examine how a rapidly growing e-commerce company can redesign its data infrastructure to handle millions of transactions, customer interactions, product searches, and recommendations without compromising performance.

Module 2: Big Data Architecture Design and Architectural Patterns

This module focuses on designing effective enterprise big data architectures. Participants learn how to translate organizational requirements into scalable technical architectures and evaluate different architectural approaches.

Key Topics

- Principles of big data architecture
- Enterprise data architecture
- Distributed system architecture
- Data ingestion and integration layers
- Storage and processing layers
- Analytics and visualization layers
- Lambda architecture
- Kappa architecture
- Event-driven architecture
- Microservices and data architecture

Case Study

Digital Banking Platform: Participants design a scalable architecture for a digital bank that must process customer transactions, mobile banking events, fraud alerts, and customer analytics in both batch and real time.

Module 3: Distributed Storage, Data Lakes and Data Lakehouse Architecture

This module examines modern approaches to large-scale data storage, focusing on data lakes, data warehouses, and data lakehouses. Participants learn how distributed storage can accommodate diverse data types while supporting analytics and machine learning.

Key Topics

- Distributed file systems
- Data lake architecture
- Data warehouse architecture
- Data lakehouse architecture
- Structured and unstructured data
- Data partitioning
- Data replication
- Data compression
- Storage optimization
- Data lifecycle management

Case Study

Healthcare Data Lake: Participants develop a conceptual data lake architecture capable of integrating electronic health records, medical images, laboratory data, pharmaceutical information, and administrative datasets for advanced healthcare analytics.

Module 4: Distributed Data Processing with Hadoop and Apache Spark

This module explores technologies for high-volume distributed data processing. Participants examine how processing workloads can be distributed across multiple computing nodes to improve speed, scalability, and resource utilization.

Key Topics

- Hadoop architecture
- Hadoop Distributed File System
- MapReduce
- YARN resource management
- Apache Spark architecture
- Spark DataFrames and Spark SQL
- Parallel data processing
- Batch processing
- Distributed machine learning
- Spark performance optimization

Case Study

Telecommunications Analytics: Participants explore how a telecommunications company can use distributed processing to analyse billions of call records, network events, customer interactions, and service-quality records.

Module 5: Real-Time Big Data Processing and Streaming Architecture

This module focuses on real-time big data systems and streaming architectures. Participants learn how organizations can capture, process, analyse, and respond to continuously generated data.

Key Topics

- Batch versus real-time processing
- Data streaming fundamentals
- Event-driven systems
- Apache Kafka architecture
- Stream processing
- Real-time analytics
- Event sourcing
- Real-time dashboards
- Complex event processing
- Stream-processing performance optimization

Case Study

Real-Time Fraud Detection: Participants design a conceptual streaming architecture that analyses financial transactions as they occur, identifies suspicious behaviour, generates alerts, and supports automated fraud prevention.

Module 6: Cloud-Native Big Data Systems and Elastic Scalability

This module explores how cloud computing technologies enable organizations to build flexible and highly scalable big data environments. Participants examine cloud storage, compute resources, managed data platforms, containers, and serverless technologies.

Key Topics

- Cloud computing and big data
- Infrastructure as a Service
- Platform as a Service
- Cloud data storage
- Cloud data warehouses
- Cloud data lakes
- Elastic computing
- Containerization and Kubernetes
- Serverless data processing
- Multi-cloud and hybrid-cloud architectures

Case Study

Global Media Streaming Company: Participants analyse how a media organization can use cloud-based scalable infrastructure to accommodate fluctuating demand, store large volumes of video metadata, and support personalized content recommendations.

Module 7: Scalable Data Security, Governance, Reliability and Performance

This module addresses the operational challenges associated with managing enterprise-scale big data systems. Participants learn how to maintain security, data quality, system reliability, and optimal performance as data environments expand.

Key Topics

- Big data security architecture
- Identity and access management
- Data encryption
- Data governance
- Data quality management
- Metadata management
- System monitoring and observability
- Performance tuning
- Backup and disaster recovery
- High availability and business continuity

Case Study

Government Big Data Platform: Participants develop a conceptual security and governance framework for a government data platform integrating information from multiple agencies while maintaining access controls, data quality, privacy, availability, and auditability.

Module 8: Enterprise Implementation, Optimization and Future Big Data Architecture

The final module integrates the technical and strategic concepts covered throughout the course. Participants learn how to develop an enterprise scalable big data implementation roadmap, evaluate technology choices, manage infrastructure growth, and prepare organizations for emerging data-intensive technologies.

Key Topics

- Enterprise big data architecture planning
- Requirements and workload assessment
- Technology selection
- Capacity planning
- Scalability testing
- Cost optimization
- Architecture performance evaluation
- Migration and modernization strategies
- Big data implementation roadmap
- Future trends in scalable data architecture

Case Study

Smart City Big Data Transformation: Participants develop a scalable architecture integrating IoT sensors, traffic systems, public transport data, environmental monitoring, citizen services, cloud computing, and real-time analytics to support intelligent urban management.

Training Methodology

- Interactive instructor-led sessions
- Hands-on AI tool demonstrations
- Group-based leadership simulations
- Real-life case study discussions
- Personalized leadership development plans
- Post-training mentoring and AI coaching sessions

Register as a group from 3 participants for a Discount

Send us an email: info@datastatresearch.org or call +254724527104

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 DATASTAT CONSULTANCY LTD account, as indicated in the invoice so as to enable us prepare better for you.

Certification & Learning Methodology

- Hands-on practical delivery with real-world datasets, case studies, and lab exercises.
- Post-training mentorship, access to class resources, and customized follow-up support.
- Participants who successfully complete the course receive an accredited Datastat Certificate of Completion.

How to Enroll & Reserve Your Seat

1. Online Registration: Visit https://datastatresearch.org/courses/scalable-big-data-systems-and-architecture-training-course to register online.
2. Email Booking: Send participant details to training@datastatresearch.org for official invoice & reservation.
3. Support & Assistance: Inquire via phone +254 724 527 104 | +254 734 969 612.