High-Performance Computing (HPC) for Data Analysis Training Course

Professional Development Training Course Outline

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

Course Introduction

In today’s era of massive data generation, the demand for real-time insights, scalable storage, and high-throughput processing has driven the adoption of High-Performance Computing (HPC) in data analysis. High-Performance Computing (HPC) for Data Analysis Training Course is designed to equip participants with the essential tools and techniques to leverage HPC for complex data processing, enabling faster computations, enhanced simulations, and actionable analytics across diverse sectors including finance, healthcare, climate modeling, AI, and bioinformatics.

With a strong focus on parallel computing, cluster computing, cloud-based HPC systems, and data-intensive applications, this training blends theoretical knowledge with hands-on labs and case studies. Participants will explore how to optimize performance, manage big data workflows, utilize cutting-edge tools such as Spark and MPI, and apply HPC to solve real-world data analysis problems with maximum efficiency.

Programme Curriculum

High-Performance Computing (HPC) for Data Analysis Training Course

Introduction

In today’s era of massive data generation, the demand for real-time insights, scalable storage, and high-throughput processing has driven the adoption of High-Performance Computing (HPC) in data analysis. High-Performance Computing (HPC) for Data Analysis Training Course is designed to equip participants with the essential tools and techniques to leverage HPC for complex data processing, enabling faster computations, enhanced simulations, and actionable analytics across diverse sectors including finance, healthcare, climate modeling, AI, and bioinformatics.

With a strong focus on parallel computing, cluster computing, cloud-based HPC systems, and data-intensive applications, this training blends theoretical knowledge with hands-on labs and case studies. Participants will explore how to optimize

performance, manage big data workflows, utilize cutting-edge tools such as Spark and MPI, and apply HPC to solve real-world data analysis problems with maximum efficiency.

Course Objectives

1. Understand the fundamentals of High-Performance Computing and its architecture.
2. Apply HPC techniques for large-scale data analysis and scientific computing.
3. Deploy parallel programming models (MPI, OpenMP, CUDA) in data analytics workflows.
4. Implement HPC clusters using cloud platforms (AWS, Azure, GCP).
5. Analyze big data using distributed computing frameworks like Apache Spark.
6. Optimize computation and storage performance in HPC systems.
7. Explore fault tolerance and load balancing in HPC environments.
8. Use containerization tools (Docker, Singularity) in HPC workflows.
9. Visualize HPC-generated data for actionable insights.
10. Manage HPC workloads using SLURM and other job schedulers.
11. Apply machine learning and AI models using GPU-accelerated computing.
12. Address security, compliance, and ethical issues in HPC data analysis.
13. Solve domain-specific challenges using HPC: bioinformatics, finance, and climate models.

Target Audience

1. Data Scientists and Analysts
2. Research Scientists in Bioinformatics, Climate, and Physics
3. AI and Machine Learning Engineers
4. IT Infrastructure Architects
5. University Students in STEM
6. Software Developers and Engineers
7. Government and Defense Data Analysts
8. Professionals in Financial and Healthcare Sectors

Course Duration: 10 days

Course Modules

Module 1: Introduction to HPC

- Definition and scope of HPC
- HPC architecture (CPU, GPU, interconnects)
- Use cases in real-world data analysis
- Basics of cluster computing
- HPC system setup overview
- **Case Study**: HPC in COVID-19 genome analysis

Module 2: Parallel Computing Fundamentals

- Principles of parallelism
- Types of parallelism: data vs task
- Shared vs distributed memory
- Amdahl’s Law and scalability
- Debugging and performance tools
- **Case Study**: Simulating seismic activity with parallel code

Module 3: Programming with MPI and OpenMP

- Message Passing Interface (MPI) basics

- OpenMP syntax and directives
- Comparison of MPI vs OpenMP
- Hybrid programming strategies
- Sample applications and benchmarking
- **Case Study**: Weather prediction modeling using MPI

Module 4: HPC Job Scheduling with SLURM

- Overview of job schedulers
- Writing SLURM scripts
- Job queues and dependencies
- Resource allocation and monitoring
- Cluster job optimization techniques
- **Case Study**: Resource scheduling for genomic pipelines

Module 5: Apache Spark for Distributed Data Processing

- Spark architecture and components
- RDDs and DataFrames
- Spark MLlib for machine learning
- Integrating Spark with Hadoop and HPC
- Spark performance tuning
- **Case Study**: Financial fraud detection using Spark

Module 6: GPU Computing and CUDA

- GPU vs CPU processing
- CUDA programming model
- GPU-accelerated libraries
- Matrix multiplication using CUDA
- Performance analysis tools
- **Case Study**: Deep learning on GPUs for image recognition

Module 7: HPC in the Cloud

- Comparing cloud platforms for HPC
- Cluster deployment on AWS/GCP/Azure
- Cost optimization techniques
- Cloud security and compliance
- Cloud-native tools for HPC
- **Case Study**: Earthquake simulation in AWS Cloud HPC

Module 8: Containerization in HPC

- Introduction to Docker and Singularity
- Creating HPC-ready containers
- Running containers on clusters
- Container orchestration (Kubernetes)
- Security and portability advantages
- **Case Study**: Reproducible research with Singularity

Module 9: High-Speed Storage and File Systems

- Overview of parallel file systems
- Lustre, GPFS, HDFS
- I/O optimization techniques

- Data locality and caching
- Storage benchmarking tools
- **Case Study**: Storage strategy for climate simulation

Module 10: Fault Tolerance and Load Balancing

- Understanding node failures
- Checkpointing and recovery techniques
- Load balancing algorithms
- Redundancy and replication
- Performance impact analysis
- **Case Study**: HPC in disaster modeling simulations

Module 11: Data Visualization in HPC

- Tools for large-scale data viz (ParaView, VisIt)
- Plotting HPC outputs in Python/R
- Real-time visualization techniques
- Integration with dashboards
- 3D rendering for scientific data
- **Case Study**: Visualizing ocean currents with HPC

Module 12: Security and Compliance in HPC

- Data privacy and encryption in HPC
- Secure access to HPC systems
- Compliance frameworks (HIPAA, GDPR)
- Role-based access and auditing
- Secure job submissions
- **Case Study**: Secure healthcare analytics with HPC

Module 13: AI and HPC Integration

- GPU acceleration in ML/DL models
- Distributed training with TensorFlow
- HPC for NLP and image processing
- Scaling models with Horovod
- Model deployment and testing
- **Case Study**: NLP model training on HPC clusters

Module 14: Domain-Specific Applications

- HPC in bioinformatics
- HPC in finance and stock prediction
- HPC in weather and climate modeling
- Engineering simulations with HPC
- Energy and transportation case uses
- **Case Study**: Predicting flood risks using HPC

Module 15: Final Project and Capstone

- Problem definition and proposal
- Dataset selection and pre-processing
- HPC pipeline setup
- Execution, benchmarking, and optimization
- Final presentation and peer review

- **Case Study**: Capstone – Solve a real-world data analysis problem with HPC

Training Methodology

- Instructor-led live sessions with real-time interaction
- Hands-on labs with cloud and cluster access
- Step-by-step coding walkthroughs
- Group projects and peer collaboration
- Case study presentations for practical insights
- End-of-module quizzes and feedback assessments

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	04 Dec 2026	Online	$2,000
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

Start Date	End Date	Location	Price
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