

Big Data in Life Sciences — HPC & Cloud — Hands-on

Engineer production-grade big data platforms for life sciences. This module blends HPC scheduling, distributed compute frameworks, and cloud-native storage/processing so you can handle terabyte-scale omics, imaging, and EHR data. You will build secure, cost-aware pipelines using Spark/Dask/Ray on top of Parquet/Delta/Iceberg with reproducible IaC and monitoring.

Big Data in Life Sciences — HPC & Cloud

[Help Desk](#) · [WhatsApp](#)

Session Index

[Session 1 — Storage, Formats & Data Lakes](#) [Session 2 — Distributed Compute: Spark/Dask/Ray](#)

[Session 3 — HPC Schedulers & Hybrid Architectures](#) [Session 4 — Mini Capstone: Cloud-Scale Pipeline](#)

Session 1

Fee: Rs 15,800 [Apply Now](#)

Storage, Formats & Data Lakes

Object storage & POSIX/HDFS basics

[S3/GS/Azure Blob](#) [HDFS/Lustre/IB](#) [throughput vs IOPS](#)

Columnar formats & table layers

[Parquet/ORC](#) [Delta Lake/Iceberg/Hudi](#) [partitioning/z-order](#)

Governance & cost controls

[lifecycle policies](#) [catalogs/Glue/Hive](#)

compression/compaction

Session 2

Fee: Rs 21,200 Apply Now

Distributed Compute: Spark/Dask/Ray

Cluster setup & execution models

YARN/Kubernetes **autoscaling** **shuffle/IO tuning**

Spark/Dask dataframes & ML

Spark SQL/MLlib **Dask-ML** **Arrow interchange**

Ray for scalable Python & AI

Ray Data/Train/Tune **distributed inference** **GPU scheduling**

Session 3

Fee: Rs 27,200 Apply Now

HPC Schedulers & Hybrid Architectures

Schedulers & job orchestration

SLURM/PBS **Singularity/Apptainer** **array jobs/checkpointing**

Hybrid (HPC↔Cloud) patterns

bursting to cloud **VPN/VPC peering** **data locality**

Security & compliance basics*

IAM/KMS/secrets **network segmentation** **audit/monitoring**

*Educational guidance only; not legal compliance advice.

Session 4

Fee: Rs 33,800 Apply Now

Mini Capstone: Cloud-Scale Pipeline

Implement an end-to-end omics/EHR big data pipeline

Theory + Practical

CI/CD & IaC for data platforms

Terraform/CloudFormation **GitHub Actions**

observability (logs/metrics)

Deliverables: repo, infra plan & cost report

code/notebooks **IaC templates** **FinOps summary**