Learn end-to-end chemical data sourcing and curation for drug discovery and QSAR workflows. You will fetch structures and bioactivity metadata from major public repositories, reconcile identifiers, remove duplicates and salts/solvents, and ship clean QSAR-ready datasets with full provenance and licensing compliance.

Chemical Databases & Data Curation (PubChem/ChEMBL/ChEBI)

Help Desk · WhatsApp

Session Index

Session 1

Fee: Rs 10800 Apply Now

Public Chemical Repositories & Identifiers

Landscape & data models

PubChem CID/SID **ChEMBL compound/activity** **ChEBI ontology**

Identifiers & cross-refs

SMILES / InChI / InChIKey **synonyms & registry IDs** **provenance & versioning**

Licensing & attribution

reuse guidelines | citations & metadata | compliance checklist

Session 2

Fee: Rs 13800 Apply Now

Programmatic Access & Bulk Download

APIs & clients

PubChem PUG-REST | ChEMBL web services | FTP/SDF dumps

Query design & filters

substructure/similarity | assay/endpoint selection | species & units

ETL basics

SDF/CSV to pandas | rate limiting & retries | caching & checkpoints

Session 3

Fee: Rs 16800 Apply Now

Standardization, Dedup & QA/QC

Structure normalization

salt/solvent stripping | tautomer handling | charge neutralization

Deduplication & merges

InChIKey-based | synonym collapse | conflict resolution

QA/QC & audit trails

data dictionaries | unit harmonization | validation reports

Session 4

Fee: Rs 20800 Apply Now

Mini Capstone: QSAR-Ready Dataset

Assemble endpoint-specific data (e.g., pIC50)

Theory + Practical

Curation pipeline & export

RDKit + pandas SMI/SDF/CSV outputs provenance log

Deliverables

clean dataset + report license & citation file reproducible notebook/script