



1535/2, Sharanya Towers, 14th Main Road, 1st Stage, Kumaraswamy Layout, Bengaluru, Karnataka 560078

- www.techninzaz.com
- contact@techninzaz.com
- [/Techninzaz](https://www.facebook.com/Techninzaz)
- [/company/techninzaz](https://www.linkedin.com/company/techninzaz)
- [/techninzaz](https://www.instagram.com/techninzaz)
- +91 - 9739789657
- 080 - 42149657

DURATION	48 HRS
WeekDays	7:00 - 9:30 AM (Mon-Thurs) 6:30 - 9:00 PM
WeekEnd	9:00 - 1:00 PM (Sat-Sun) 1:30 - 5:30 PM 6:00 - 9:00 PM

* Batch Slots As Per Availability

ECOSYSTEM

ARCHITECTURE

ENVIRONMENT

SCHEDULER

HORTONWORKS 2.6

POWERING THE FUTURE OF DATA™

MEMORY PARALLEL BATCH

ETL

cloudera 5.13

Ask Bigger Questions

I/O PARALLEL BATCH

DISTRIBUTED NoSQL

DATA MINING

DATA WAREHOUSING

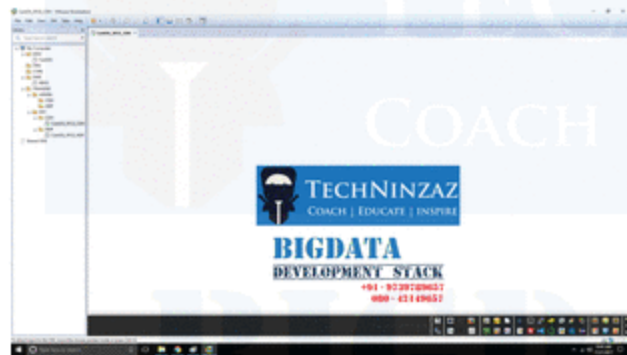
❖ Module I: Introduction to Bigdata and Ecosystem

- Data Types, Distributed / Parallel Processing Concepts.
- Big Data Characteristics, Challenges with Traditional Systems.
- Hadoop TimeLine & History.
- Fundamentals, Core Components, Rack Awareness, Node & Cluster Concept.
- Solution Types, Distributions & Specialities, Challenges & Complexity & Use Cases.
- Stack Insight {On Premise Vs Cloud} (Cloudera, Hortonworks, MapR, AWS).
- Understanding Cluster Development Environment vis-à-vis Virtualization, Linux, FileSystems, Java, Scala, Sql, & Other Terminology.
- Role Expectation, Job Description, Responsibilities & Growth Plan.
- Multi Node "Cloudera" Cluster "First Look".

❖ Module II: HDFS, Hadoop Architecture & YARN

- HDFS Components, Fault Tolerance, Horizontal Scaling, Block Size, Replication Factor, Daemons, HA, Federation, Quotas.
- Anatomy of Read / Write & Failure / Recovery on HDFS.
- Shell Commands & WebUI.
- YARN "The Hadoop OS" In Depth (Architecture, HA, RM, Scheduler, Queues, Node Labels).
- File Formats Avro, JSON, ORC, Parquet & Others.
- Understanding MapReduce Architecture, Logical Data Flow, Driver, Mapper, Partitioner, Reducer, Combiner, Input Splits, Shuffle & Sort, Counters, Distributed Cache.
- Anatomy of MapReduce Jobs Execution MRv2 - YARN(RM, AM, NM).
- Run Simple M/R Job on YARN through Java programming on Eclipse IDE.

❖ Module III: Environment



- State Of The Art Customized (CentOS 7.3 + XFCE) VM Designed with plethora of Latest Items for BigData / ML / Analytics / ETL / DB Development including Scala IDE, Eclipse IDE, VS Code, Anaconda, R Studio & Talend Open Studio.
- Featuring In Built One Node Cloudera 5.13 & HortonWorks 2.6 Full Stack Implementation for Hands-on.
- Numerous Data Patterns for Modeling & Real Life Understanding of Dev Problems & Situations.
- Participants will also be a part of "Batch" Multi Node Cloudera Cluster for Classroom assignments and learning.

We Dont Use / Teach On RAW Apache Hadoop OR Default Cloudera / HortonWorks VM.

❖ Module IV: Scala Programming Language

- Quintessential Hello World On Scala IDE
- Literals, Values, Variables & Types
- Expressions, Conditionals & Iterators
- Functions, Methods, Operators & Closures
- Strings
- Tuples, Sets, Arrays, Lists, Sets, Maps
- Objects, Classes, and Traits
- Advanced Typing
- Exception Handling
- I/O Operations

+91 - 9739789657
080 - 42149657

❖ **Module V: Spark Cluster-Computing Framework**

- Introduction, Architecture, Components, Execution & Related Concepts
- The Resilient Distributed Dataset
- RDD Data Types, Creation & Operations
- Spark Shell In Action & WordCount Job on YARN through Functional Programming on Scala IDE
- RDD Map, Filter, Sort Transformations
- Data Partitioning & Joins
- Accumulators, Broadcast Variables
- Caching and Persistence
- Spark SQL DataFrames and DataSets
- Joins, Strongly Typed Dataset
- DataSets Vs RDD's Choice / Conversion
- Hive queries through Spark

❖ **Module VII: Hive & Impala**

- Hive Components, Architecture, Metastore & HCatalog.
- Driver, Query Compiler, Optimizer and Execution Engine.
- Server and Client components, Data Types, Data Mode, File Formats.
- Primitive Datatypes, Collections_Arrays_Maps, Structs and Unions.
- Hive Shell, Beeline, DDL, DML & QL+CRUD Hands On Practical Scenarios.
- Managed / External / Dynamic Table, Parallel Jobs, Partitioning, Bucketing, Transactions, Indexing.
- Map-Side / Skewed / Cartesian Joins & Optimization, Views, Compression Codecs, Windowing Functions.
- Implementation & Usage of Hive UDF, UDTF and SerDe.
- Impala Shell, Sql in Action, Interaction with Hive & HBase.

❖ **Module VI: MapReduce & Scalding (Scala DSL)**

- Implementing Custom Input / Output Formats, Saving Binary Data Using SequenceFile and Avro Files, Map-Side / Reduce-Side / Skewed Join, Cartesian Product.
- Building Inverted Index, Custom Partitioner, Converting Unstructured to Structured Data, Compression Techniques.
- MapReduce abstractions, Cascading Pipes, Scalding File Read / Write.
- Understanding the core capabilities of Scalding.
- Map-Like, Join, Pipe, Composite Operations, Grouping / Reducing functions.
- Scalding Design Patterns, External Operations, Dependency Injection, Late Bound Dependency.
- Scheduling & Coordinating Execution on YARN.
- Monitoring Jobs, Execution Throttling.
- Scalding HBase Interaction.

❖ **Module VIII: Pig Latin**

- Architecture, Program Structure, Execution Process, Grunt Shell, Interactive / Batch Scripting.
- Data Model, Expression, Data Types, Nulls, Operation.
- Core Relational Operators – Load, Store, Filter, Transform.
- Skewed / Merge / Replicated Join, Cross Operator.
- Group, CoGroup, Union, Foreach, Sort/Order, Combine / Split, Distinct, Filter, Limit.
- Describe, Explain, Illustrate.
- Multiple Loaders, UDF, Macros.
- Interaction With Hive & HBase.

+91 - 9739789657

080 - 42149657

❖ Module IX: The Hadoop DataBase - HBase

- NoSQL Introduction, CAP Theorem & Eventual Consistency, Row Oriented v/s Column Oriented Storage.
- HBase Architecture, Working Components, Data Model, CRUD Operations, Read and Write Operations, Compactions, Filters, Bulk Loading, Scans.
- Column Families, Column Value & Key Pair.
- HMaster, HRegionServer, Zookeeper.
- Memstore / Hfile / WAL.
- HBase Shell In Action.
- Data Storage For Hive & Pig.
- Phoenix Architecture & Capabilities.
- CRUD, Joins, SubQueries, Views, Transactions & UDFs.

❖ Module X: ETL & Orchestration

- Sqoop - Importing Data to HDFS, Hive, HBase & Exporting to MySQL.
- Incremental Imports using Sqoop Jobs.
- Flume Introduction & DataFlow Model.
- Flume Complex Flow - Multiplexing.
- Configuration File, Interceptors, Importing Data to HDFS.
- Oozie Architectural Components, Workflow Lifecycle.
- job.properties & workflow.xml file.
- Control nodes, Action nodes and Global configurations within Workflows.
- Coordinators, Bundles.

- *Industry Compliant Practical Curriculum On Latest Stacks.*
- *Multiple Domains Simulated Data Patterns for Real Life Project Scenarios.*
- *Guidance for Resume Preparation & Interview Questions.*
- *POCs Suggestions for further practice and pursual.*
- *Mock Interviews for Interested Candidates.*
- *Discussing Current Market Standards & New / Incubating Tech.*
- *Other Q & A, Doubt Clearnace.*
- *Discussing Cloudera & Hortonworks Certification Programs.*

