# DIGICROME ACADEMY®

# Master's Program in Big Data Engineering & Cloud Computing
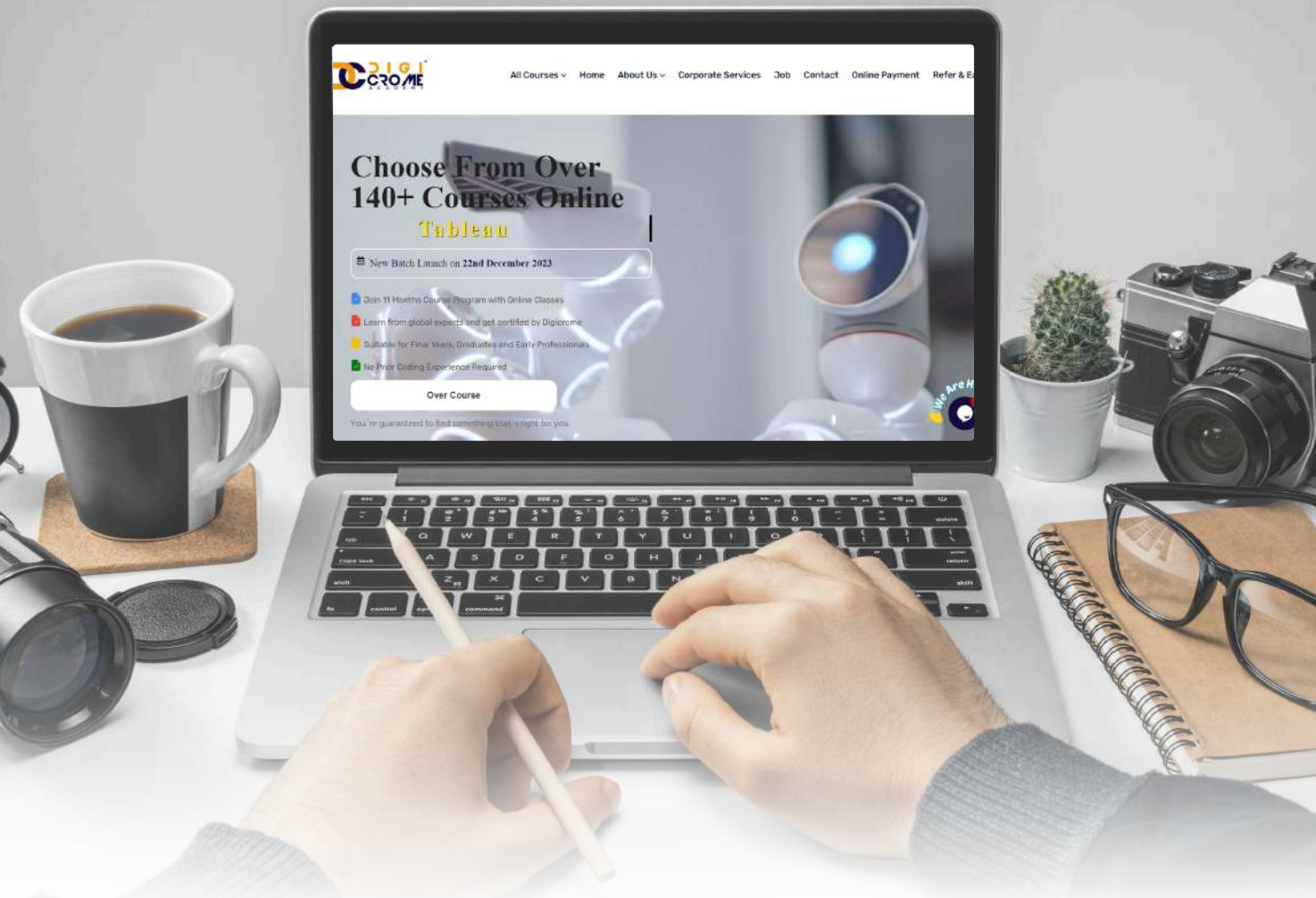
**10K+** Learners

**1:1 Personlized** Mentorship

**40% Average** Salary Hike

# About Digicrome

Digicrome is the world's #1 online bootcamp provider that enables learners through rigorous and highly specialized training. We focus on emerging technologies and processes that are transforming the digital world, at a fraction of the cost and time of traditional approaches. Over one thousand professionals have harnessed our award-winning programs to achieve their career and business goals

## Master's Program in Big Data Engineering & Cloud Computing

**Open For Registration!**
Period 2024 - 2025 ⫸

**Read More**
www.digicrome.com

**10k+**

**5k+**

S

**250+**

Batches

Since our inception, we have been focused on developing cutting-edge learning methodologies by involving learners and experienced faculty, along with providing individuals and corporations with high-quality training materials that aid professionals in accomplishing their career objectives and furthering their careers.

We work with some of the world's finest institutions and certifying authorities and we aspire to provide high-quality training to professionals all across the globe. We have a proven track record of effectively training thousands of professionals in both classroom and online training. Come join us and let us transform your professional lives via digital skills.

❝

Our mission is to offer affordable and industry-relevant education that enables the advancement and development of India's workforce.

G ★★★★★ 

4.6/5     Google Rating

# Hottest Job of 21st Century

### 1.5 Million Job Postings
There is a global estimate of 1.1 million job postings for Data Engineers in 2024

### Skill Development
Data Engineers are equipped with various relevant skills that fetch lucrative job offers

### Growing Data Engineering Industry
50% CAGR in the global data industry

### Future-oriented Career
Data Engineering is a budding field; a head start will prove to be beneficial

### Popular Degree
48% of Data Engineers have a master's degree

### High Demand
In 2024, India and the US will face a demand-supply gap of 500,000 Data Engineers

## Our Credentials

**10K +**
Aspiring Active Students

**100+**
Industry-expert Instructors

**250+**
Global Hiring Partners

**750+**
Corporates Upskilled

**40%**
Average Salary Hike

**65+**
Countries' Learners

# Get Skills To Fulfill Every Role:

Our **Data Engineering** Courses are designed for a wide range of people looking for skills and opportunities across all major **IT sectors**

## Hands-On Projects:

Gain practical experience by working on real-world projects, and building a robust portfolio that will impress potential employers.

## Career Support:

We're dedicated to your success! Benefit from career guidance, resume building, interview prep, and job placement

## Flexibility:

Our flexible schedule options allow you to learn at your own pace, making it perfect for both beginners and experienced developers looking to upskill.

## Community:

Join a vibrant community of like-minded learners, where you can collaborate, share ideas, and network with peers.

## Our students succeed in top-tier tech companies

# Who Should Enroll in this Program?

This program welcomes graduates from any discipline and working professionals across diverse backgrounds, provided they possess basic programming skills. The diversity among our students enhances the depth of class discussions and interactions.

A data engineer constructs and maintains data structures and architectures essential for ingesting, processing, and deploying data in large-scale, data-intensive applications. This career path holds promise for both newcomers and seasoned professionals passionate about data, including:

- IT professionals
- Database administrators
- Beginners in the data engineering domain
- BI Developers
- Data science professionals who want to expand their skills
- Students in UG/ PG programs.

# Eligibility criteria and application process

## Who can apply?

- Freshers and Undergraduates willing to pursue a career in Data Engineering
- Anyone looking for a career transition to Data Engineering
- IT Professionals
- Experienced Professionals willing to learn Data Engineering
- Technical and Non-Technical Professionals with basic-level programming knowledge can also apply
- Project Managers

## Additional Scholarships

- Refer Someone and avail a discount of up to 5%

For more details, look at the program **Terms & Conditions**

## How can apply?

**1**

### Submit Application
Apply for the Program by filling up the 1 min Application form.

**2**

### Join the Prestigious Program
The admissions office will send the acceptance letter. You can secure your seat by depositing the registration fee.

**3**

### Admission
You can secure admission by accepting the offer letter and completing the payment.

# Key Features of the Program

| 11 Months Live Classes | 3 Months Internship | 11 Types of Certification by Digicrome |
|---|---|---|

## Key Highlights

- 11 Months of Live Sessions by Industry Experts
- 200+ Hrs of Self-paced Videos
- 24*7 Career Support
- Personalized project reviews
- Real-Time Data Processing
- Soft Skills Essential Training

- Learn industry-based practices
- 1:1 with Industry Mentors
- 50+ Industry Relevant & Case Studies
- E-learning Videos from LMS Support
- Lifetime Career Assistance
- Placement Assistance Support*

## Skills you'll learn

- Automate Data Pipelines
- Data Modeling

- Cloud Data Warehouses
- Spark and Data Lakes

## Skills Covered

- SQL: ◄┈┈┈┈┈► Master complex queries and database design
- ETL Processes: ◄┈┈┈┈┈► Extract, Transform, Load data with precision
- Data Warehousing: ◄┈┈┈┈┈► Build robust data storage solutions
- Python: ◄┈┈┈┈┈► Leverage powerful programming languages
- Big Data Technologies: ◄┈┈┈┈┈► Harness the power of Hadoop and Spark
- Data Visualization: ◄┈┈┈┈┈► Turn data into compelling stories
- Cloud Services: ◄┈┈┈┈┈► AWS, Azure, GCP for scalable solutions
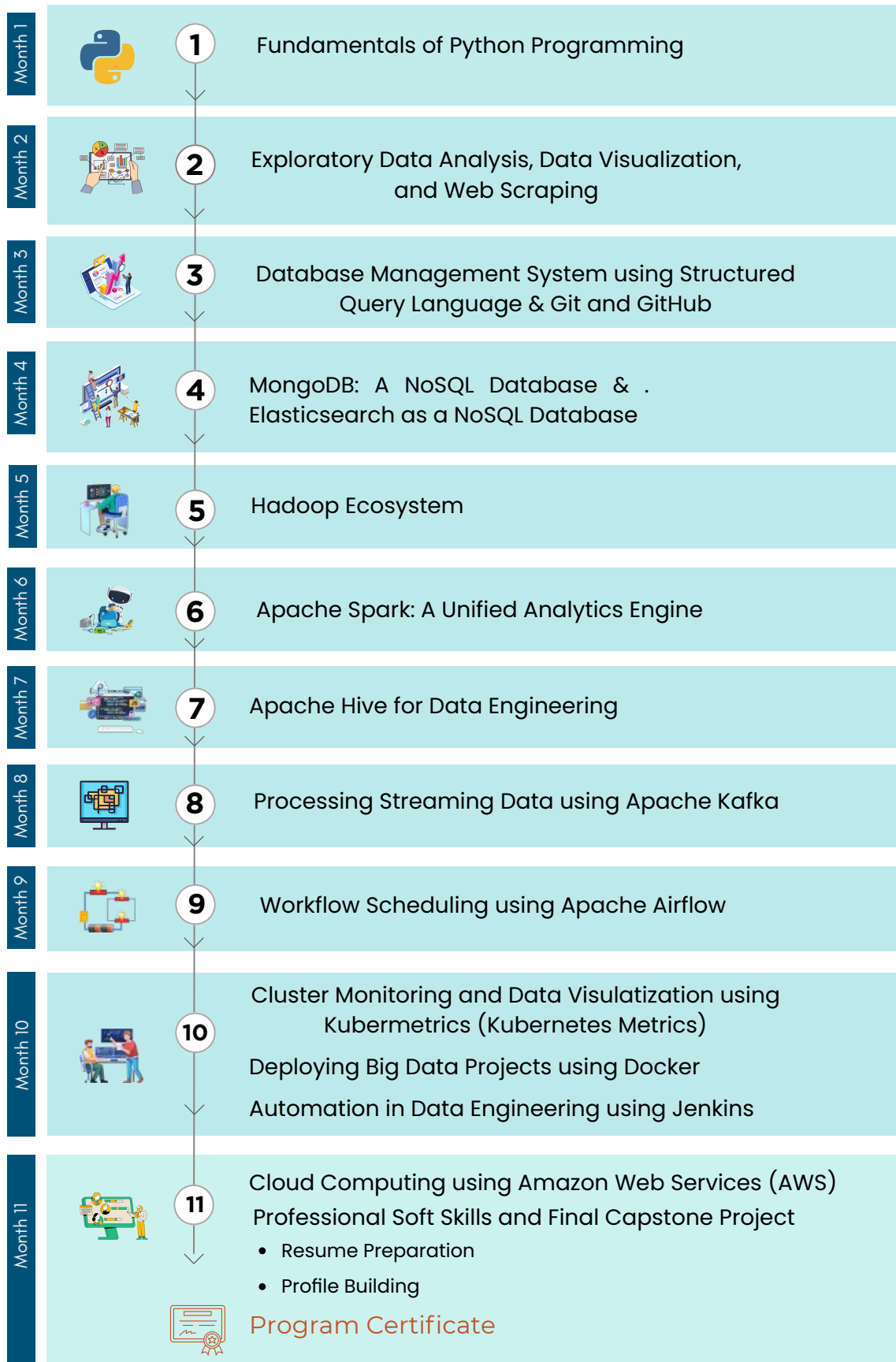
# Course Tools & More

**Microsoft Azure**   **snowflake**   **python**   **ChatGPT**   **kafka**

**GitHub**   **mongoDB**   **elasticsearch**   **seaborn**

**docker**

**Working Tools**

**Matplotlib**

**Metrics Server**

**Apache Airflow**

**Apache hadoop**   **aws**   **Apache Spark**   **HIVE**

**Jenkins**   **NumPy**   **pandas**   **Apache Pig**   **MySQL**   **Google Cloud**

CAPSTONE PROJECTS

CASE STUDIES

EXPERIENTIAL LEARNING

HACKATHONS

ASSIGNMENTS

# Learning Path **& Career Services**

| | | | |
|---|---|---|---|
| Month 1 | | **1** | Fundamentals of Python Programming |
| Month 2 | | **2** | Exploratory Data Analysis, Data Visualization, and Web Scraping |
| Month 3 | | **3** | Database Management System using Structured Query Language & Git and GitHub |
| Month 4 | | **4** | MongoDB: A NoSQL Database & . Elasticsearch as a NoSQL Database |
| Month 5 | | **5** | Hadoop Ecosystem |
| Month 6 | | **6** | Apache Spark: A Unified Analytics Engine |
| Month 7 | | **7** | Apache Hive for Data Engineering |
| Month 8 | | **8** | Processing Streaming Data using Apache Kafka |
| Month 9 | | **9** | Workflow Scheduling using Apache Airflow |
| Month 10 | | **10** | Cluster Monitoring and Data Visulatization using Kubermetrics (Kubernetes Metrics) |
| | | | Deploying Big Data Projects using Docker |
| | | | Automation in Data Engineering using Jenkins |
| Month 11 | | **11** | Cloud Computing using Amazon Web Services (AWS) |
| | | | Professional Soft Skills and Final Capstone Project |

- Resume Preparation
- Profile Building

Program Certificate

# Syllabus **Overview**

| COURSE MODULE | TIME DURATION: 11 MONTHS |
|---|---|
| | 45+Weeks: 200+ HOURS |

- **11 Months** of Practical & Module Training
- **200+** Hours of Comprehensive Learning
- **100+** Online Expert-led Classes
- **14** Specialized Modules

Note: 03 Months Internship (Internship will start from last three month of the course)

## PROGRAM CURRICULUM

| 14 Modules | | 100+ Classes |
|---|---|---|

**Introduction to Data Engineering – ETL Process and other important theoretical concepts**

### Month 1

## Fundamentals of Python Programming

- 1. Introduction to Python Programming - Basics of Python, Data Types, Print and Input functions
- 2. Data Structures in Python - Different Methods of Each Data Structure
- 3. Control Statements, Loops, and Functions in Python
- 4. Object Oriented Programming - Classes and Objects, Inheritance, etc.

**Month 2**

# Exploratory Data Analysis, Data Visualization, and Web Scraping

- 1. Advanced Numerical Python - Numpy Tutorial
- 2. Data Cleaning and Wrangling with Pandas - Advanced Pandas Tutorial
- 3. Exploratory Data Analysis using Matplotlib and Seaborn
- 4. Basic and Advanced Web Scraping using BeautifulSoup
- 5. Basic and Advanced Web Scraping using Selenium

**Month 3**     **Part 1**

# Practical Statistics for Data Scientists

- 1. Introduction to Databases - (What are Databases), (What is MySQL), (What is RDBMS), (RDBMS v/s NoSQL)
- 2. Data Base Workflows - Understanding Entity Relationship Diagram, Understanding Normalization (1NF, 2NF, 3NF, BCNF)
- 3. Structured Query Language (SQL) - CRUD Operations
- 4. Data Aggregation Functions - (GroupBy), (OrderBy), (HAVING), (COUNT, SUM, MIN, MAX, AVG)
- 5. Joins in SQL - Primary Key and Foreign Key, Constraints, Set Operations, DML - Savepoint, Rollback

**Month 3**   **Part 2**

## Git and GitHub

- ✓ 1. Introduction and Overview - (Understanding Repositories, Branches, and Commits), (Setting up Git and GitHub)
- ✓ 2. Versions Control with Git, Repository Management, and Collaboration
- ✓ 3. Workflow Automation with GitHub - GitHub Actions, Automating CI/CD Pipelines for Data Engineering Projects
- ✓ 4. Security and Compliance in GitHub - Managing Access Permissions, GitHub Pages, Documentation and Wikis
- ✓ 5. Integrating GitHub with other tools and Cloud Platforms
- ✓ 6. Advanced Techniques - Rebasing and Cherry picking Commits, Git for Automated tasks, Stashing changes, and using the reflog

**Month 4**   **Part 1**

## MongoDB: A NoSQL Database

### 1. Installation and Configuration

- ✓ 1. Introduction to MongoDB - Overview, Applications and Use Cases, Relational Databases vs Document Oriented Databases
- ✓ 2. MongoDB Architecture - (Core Concepts - Documents, Collections, Databases), (JSON & BSON Data Formats), (MongoDB Server), (Replica Sets and Sharding)
- ✓ 3. NoSQL Queries - CRUD Operations, Querying Documents
- ✓ 4. Data Modeling - Schema, Design, Embedding vs Referencing, Arrays & Nested Documents, Indexing Strategies

5. Aggregation Framework - Introduction, Pipeline Stages - $map $group $Project etc., Aggregation Transformation and Analysis, Performance Tuning

6. Transactions and Concurrency - ACID transactions, Section based transactions, Optimistic Concurrency Control

7. MongoDB Indexing - Creation and Management of Indexes, Types of Indexes

8. Replication and High Availability - Setting up and managing Replica Sets, Automatic Failover and Recovery, Read and Write Concerns in Replica Sets

9. Sharding for Scalability - Introduction, Shard Keys and Chunk Distribution, Balancing and Mitigating Chunks, Managing Sharded Clusters

10. Advanced MongoDB Concepts

10.1 Authentication and Authorization for Data Security, Role-based access control, Data Encryption, Auditing and Logging

10.2 Performance Tuning - MongoDB Atlas, Ops Manager, Profiting and Optimizing Queries, Resource Management and Capacity Planning

10.3 MongoDB with different Programming Languages, Integration with Big Data Tools - Spark & Hadoop, Using MongoDB and BI tools for Data Visualization

**Month 4**     **Part 2**

# Elasticsearch as a NoSQL Database

### 1. Installation and Configuration

1. Introduction to Elasticsearch - Use Cases, Core Features, Elasticsearch vs Other NoSQL Databases

2. Elasticsearch Architecture - Cluster, Nodes, Indices, Shards, Replicas, how data is stored in JSON documents, Role of Mappings

✔ 3. Data Modelling and Ingestion

3.1 Schema Design, Nested and Parent-Child Relationships

3.2 Using Rest APIs for CRUD Operations, Bulk API for effectively indexing large volumes of data

3.3 Integration with tools like Logstah, Beats, Apache Kafka, etc. for ingesting data

✔ 4. Search and Querying - Query DSL (Domain Specific Language), Search APIs for basic and complex queries, Filters and Aggregations

✔ 5. Miscellaneous

5.1 Data Management - Index Management, Aliases, Reindexing

5.2 Performance Optimization- Indexing Performance, Search Performance, Shard Allocation

5.3 Monitoring and Maintenance - (Monitoring Tools like Kabana, Elastic Stack Monitoring, etc.), Cluster Health, Backup and Restore

5.4 Security - Authentication and Authorization, Role-based access control, SSL/TLS for secure data communication

✔ 6. Advanced Techniques using Elasticsearch

6.1 Real-Time Data Analytics and Monitoring, Data Pipelines, Log Management and Event Data Analysis

6.2 Machine Learning for Anomaly Detection and Predictive Analytics

6.3 Geo Spatial Queries, Scripting and Extensions

# Hadoop Ecosystem

## 1.Installation and Configuration

- ✓ 1. Introduction to Apache Hadoop - Overview, History, Use Cases for Big Data

- ✓ 2. Hadoop Architecture - HDFS, YARN, MapReduce, Data Replication, and Block Management, How does Hadoop achieve fault tolerance?

- ✓ 3. Hadoop Distributed File System (HDFS) - Architecture, File operations, Data Replication, Control and Security in HDFS

- ✓ 4. Yet Another Resource Negotiator (YARN) - Architecture, Resource Allocation and Management, Job Scheduling and Monitoring

- ✓ 5. MapReduce Model - (Basics - Mapper, Reducer, Combiner, Partitioner), Running and Writing MapReduce Jobs, Optimization Techniques

- ✓ 6. Data Ingestion and ETL Processes
    6.1 Apache Sqoop for Importing and Exporting Data
    6.2 Apache Flume for Real-time Data
    6.3 Apache Pig and Apache Hive for Data Transformation

- ✓ 7. Data Processing and Analytics - Apache Pig and Apache Hive, Integrating Apache Spark with Hadoop for Advanced Analytics

- ✓ 8. Hadoop Security - Authentication, Authorization, Encryption, Data Security at transit and rest, Role-based access control

- ✓ 9. Deployment, Monitoring, and Management in the Hadoop Ecosystem
    9.1 Monitoring using Ambari and Cloudera
    9.2 Collection and Analysis of Metrics
    9.3 Scaling, Backup, and Recovery Practices
    9.4 Troubleshooting and Debugging

✓ 10. Hadoop Ecosystem Advanced Concepts

10.1 HDFS Federation and High Availability, YARN Capacity and Fair Schedulers

10.2 Customizing and Fine Tuning MapReduce

10.3 Integration with Apache Kafka for Real-Time Data

10.4 Using Hadoop with Cloud Platforms - AWS, GoogleCloud, Microsoft Azure

## Month 6

# Apache Spark: A Unified Analytics Engine

### Installation and Configuration

✓ 1. Introduction to Apache Spark - Overview, Key Theoretical concepts: RDD, DAG, SparkContext, SparkSession

✓ 2. Spark Architecture

2.1 Cluster Managers - Standalone, YARN, Mesos, Kubernetes

2.2 Executors, Workers, Job, Task, and Stage Concepts

2.3 How does Spark achieve fault tolerance?

✓ 3. Spark Core Concepts - RDDs Working, Pair RDDs operations and usage, Broadcast variables and accumulators

✓ 4. Spark SQL - SQL Operations, DataFrames and Datasets, Interoperability b/w DataFrames and RDDs, Performance Optimization

✓ 5. Data Processing with Spark - ETL Processes, (Data Sources - HDFS, S3, Cassandra, HBase, JDBC), File formats

✓ 6. Spark Streaming - DStreams, Structured Streaming, Window Operations, Stateful Transformations, Checkpointing

- 7. Machine Learning with Spark
  - 7.1 Pipelines, Transformers, Estimators
  - 7.2 Feature Engineering with Spark
  - 7.3 Classification, Clustering, Regression, Collaborative Filtering
- 8. Graph Processing - Creating and Transforming Graphs using GraphX, PageNet, Triangle Counting, Connected Components
- 9. Advanced Spark Concepts - Shuffles, Data Serialization, Performance Tuning and Configuration
- 10. Deployment using Spark - Debugging and Troubleshooting

## Month 7

# Apache Hive for Data Engineering

### Installation and Configuration

- 1. Introduction to Apache Hive - Overview, Use Cases in Real World, Comparison with SQL and RDBMS
- 2. Hive Architecture - (HiveQL, Metastore, Driver, Compiler, Execution Engine, and HiveServer2), Interaction of Hive with Hadoop
- 3. Hive Query Language (HiveQL)
  - 3.1 Basics of HiveQL, Data Definition Language, Data Manipulation Language
  - 3.2 Create, Read, Update, Delete, Insert operations
  - 3.3 Querying Data in HiveQL - SELECT, JOIN Subqueries
  - 3.4 Functions in HiveQL - User-Defined Functions, User-Defined Aggregate Functions
  - 3.5 Windowing and Analytics Functions
  - 3.6 Using Views and Indexes
  - 3.7 Writing and Optimizing Complex Queries

✓ 4. Hive Data Modeling, Data Types, and File Formats

     4.1 Partitioning and Bucketing

     4.2 Table Properties and Constraints, External vs Managed Tables

     4.3 Complex Data Types - Arrays, Maps, Structs

     4.4 File Formats - TextFile, SequenceFile, Optimized Row Columnar (ORC), Parquet, Avro

✓ 5. Hive Security and Performance Tuning - Authentication, Authorization, Role-Based Access Control

✓ 6. Query Optimization Techniques, Indexing and Partition Pruning, Configuring and Tuning Hive Parameters

✓ 7. Data Ingestion and ETL Processes, Monitoring Hive Performance

✓ 8. Hive on Cloud Platforms - AWS EMR, GoogleCloud Dataproc, Azure HDInsight

## Month 8

# Processing Streaming Data using Apache Kafka

### Installation and Configuration

✓ 1. Introduction to Apache Kafka - Overview, History, Use Cases for Real-Time Data Processing

✓ 2. Kafka Architecture - (Brokers, Topics, Partitions, Replicas, Producers, Consumers, and ZooKeeper), How does Kafka achieve fault tolerance?

✓ 3. Creating and Managing Topics in Kafka

✓ 4. Understanding Partitions and Replicas in Kafka

✓ 5. Producers and Consumers

5.1 Writing producers to send data to Kafka

5.2 Writing consumers to read data from Kafka

5.3 (Producer and consumer configurations), (Synchronous vs. asynchronous producer sends), (Consumer groups and message consumption)

✓ 6. Kafka Connect

6.1 Introduction to Kafka Connect - Setting up and configuring connectors

6.2 Source and Sink connectors, Custom connectors development

6.3 Common connectors for databases, file systems, and other data sources

✓ 7. Kafka Streams

7.1 Introduction to Kafka Streams API

7.2 Stream processing concepts: KStreams, KTables

7.3 Writing and deploying stream processing applications

7.4 Stateful processing and windowing operations

7.5 Interactive queries and materialized views

✓ 8. Schema Management with Confluent Schema Registry - Introduction, Registering, and managing schemas, (Avra, Protobuf, and JSON Schemas)

✓ 9. Kafka Security, Monitoring, Management, Performance Tuning

✅ 10. Advanced Concepts in Apache Kafka

10.1 Kafka Transactions and exactly one semantics, Joins and Aggregation in Kafka Streams, Log Compaction, Handling Rebalances in Kafka Streams

10.2 Designing Pipelines with Kafka, Using Kafka for ETL Processes, Real-Time Data Transformation, Data Integration Patterns

10.3 Kafka Integration with Hadoop and Spark

10.4 Using Kafka with Data Warehouses - Amazon Redshift, Snowflake

10.5 Real-time Data Analytics with Kafka

10.6 Kafka Integration with Cloud (AWS, GoogleCloud, Microsoft Azure)

## Month 9



# Workflow Scheduling using Apache Airflow

### Installation and Configuration

✅ 1. Introduction to Apache Airflow - Overview, Use Cases, Comparison with other workflow orchestration tools

✅ 2. Airflow Architecture - Schedulers, Web Server, Worker, Metadata Database, Directed Acyclic Graphs, Task Instances and their states, Executors

✅ 3. Directed Acyclic Graphs (DAGs) - (Writing DAGs using Python), (Understanding Operators, tasks, and dependencies), (Built-in Operators), (Custom Operators and sensors)

- 4. Scheduling Workflows and Data Pipelines

  4.1 Designing Efficient Workflows

  4.2 Scheduling Complex Pipelines

  4.3 Implementing ETL pipelines

  4.4 Orchestrating data pipelines across multiple systems

  4.5 Handling dependencies and data consistency

- 5. Scheduling and Triggering

  5.1 Scheduling DAGs with cron expressions and time delta

  5.2 Trigger rules and conditional task execution

  5.3 Manual triggering of DAGs

  5.4 Handling time zones and daylight saving time

- 6. Task Management and Execution

  6.1 Task retries, SLA, and timeouts

  6.2 Task dependencies and triggering rules

  6.3 Parallel task execution and task queues

  6.4 Handling task failures and retries

- 7. Airflow Plugins, Airflow UI, Security, Monitoring and Integrations

  7.1 Creating and using Airflow Plugins

  7.2 Navigating Airflow web interface

  7.3 Authentication, Authorization, Role-based access control, Securing Airflow web server

  7.4 Monitoring DAGs, Managing DAGs using Airflow CLI, Viewing Logs and Task Details

  7.5 Integration of Airflow with Databases and Cloud Services, Using Hooks for External System Interactions

✅ 8. Advanced Level Apache Airflow Tutorial

8.1 Dynamic DAG generation, SubDAGs, Branching and conditional workflows

8.2 Optimizing Airflow Performance, Scaling Airflow with Celery and Kubernetes, Resource Management

8.3 Version Controlling DAGs and Configurations, Implementing CI/CD for Airflow DAGs, Testing and Validating before deployment

8.4 Integration with Cloud Services - AWS, GoogleCloud, Microsoft Azure

**Month 10**   **Part 1**

# Deploying Big Data Projects using Docker

**Installation and Configuration**

✅ 1. Introduction and Overview to Docker, Benefits of using Docker for Data Engineering, (Key Concepts - Containers, Images, Docker Engine)

✅ 2. Working with Docker Images - Pulling Images from Docker Hub, Custom Docker Images using Dockerfiles

✅ 3. Docker Containers - Creating, Running, Stopping, Starting, Restarting, Inspecting, and Managing Docker Containers

✅ 4. Data Persistence in Docker - Managing data within containers, Docker Volumes for Persistent storage, Bind mounts vs volumes, Backup and Restore for Data Volumes

✅ 5. Docker Networking - (Bridge, Host, Overlay), Connecting Containers using Networks, Exposing and Publishing Container Ports, Network Security

✅ 6. Docker compose - (Multi Container applications with docker-compose.yml), (Docker compose for development, testing, and Production)

✅ 7. Applications of Docker in Data Engineering

     7.1 Data Ingestion and ETL Processes

     7.2 Data Processing and Analytics - Running Big Data Frameworks in Docker, Containerizing Jupyter Notebooks for Data Analysis

     7.3 Database Management - Running Relational and NoSQL Databases like MySQL and MongoDB in Docker, Data Backup and Recovery for Databases

     7.4 Machine Learning and Artificial Intelligence - Deploying ML & AI models in Docker, Docker for Reproducible ML Experiments

     7.5 Distributed Systems - Distributed Data Pipelines using Docker, Orchestrating Distributed Systems with Kubernetes

## Month 10     Part 2

# Cluster Monitoring and Data Visualization using Kubermetrics (Kubernetes Metrics)

**Installation and Configuration**

✅ 1. Introduction to Kubermetrics - Clusters, Nodes, Pods, Namespaces, Deployments, Services, ConfigMaps and Secrets, Benefits of Kubermetrics in Data Engineering

✅ 2. Storage Management - Persistent Volumes (PVs) and Persistent Volume Claims (PVCs), Stateful Sets and Storage Classes

✅ 3. Networking - Networking Fundamentals for Kubernetes, Network Policies, Controllers for Load Balancing and Routing

✅ **4. Advanced Kubermetrics Techniques**

4.1 Integrating Kubermetrics with CI/CD Pipelines, Deploying Updates and Managing Rollbacks

4.2 Deploying Data Processing Frameworks on Kubernetes, Managing Databases in Kubernetes, Containerizing ETL Pipelines

4.3 Multicluster Management, Service Mesh for Microservices Management, Kubernetes Operators for Managing Complex Applications

✅ **5. Miscellaneous**

5.1 Resource Management, Monitoring and Logging, Using Helm for Package Management

5.2 Role-based access control, Secrets Management, Pod Security

`Month 10`   `Part 3`

# Automation in Data Engineering using Jenkins

**Installation and Configuration**

✅ 1. Understanding Jenkins as a continuous integration and continuous delivery (CI/CD) tool.

✅ 2. Jobs and Pipelines in Jenkins

✅ 3. Version Control using Git and GitHub in Jenkins

✅ 4. Pipelines in Jenkins - Managing Dependencies and Sequential Tasks, Stages and Steps for Pipeline Workflows, Handling Errors in Pipelines

✅ 5. Building and Testing - Automating processes for Data Engineering Applications, Implementing Automated Testing, Jenkins agents and executors

- 6. Deployment and Release - Automating Deployment with Jenkins Pipelines, Continuous Deployment, Blue-Green Deployments and Canary Releases

- 7. Miscellaneous

  7.1 Monitoring Jenkins Pipelines executions and build statuses, Configuring Email Notifications for Failures and successes in Pipelines, Pipeline Health Check

  7.2 Role-based access control, Securing Credentials and Sensitive Information in Jenkins

  7.3 Optimizing Performance for Data Engineering Pipelines, Scaling Jenkins Infrastructure with Distributed Builds and Agents

  7.4 Jenkins Plugins, Building Custom Jenkins Plugins, Automated Data Pipelines with Jenkins

## Month 11

# Cloud Computing using Amazon Web Services (AWS)

- 1. Introduction to Amazon Web Services (AWS) - Overview, Key Services for Data Engineering, AWS Global Infrastructure

- 2. Core AWS Services
  2.1 Compute - Amazon EC2, AWS Lambda, AWS Fargate

  2.2 Storage - Amazon S3, Amazon EBS, Amazon EFS

  2.3 Networking - Amazon VPC, AWS Direct Connect, Amazon Route 53

- 3. Data Warehousing using Amazon Redshift - Architecture, Redshift Spectrum, Performance Tuning

- 4. Processing Data Lakes using AWS Lake Formation and Integrating Data Lakes with other AWS Services

- **5. Database Services with Amazon Web Services**
    5.1 Relational Databases: Amazon RDS, Amazon Aurora

    5.2 NoSQL Databases: Amazon DynamoDB, Amazon DocumentDB (with MongoDB compatibility)

    5.3 Database Migration Service (DMS) for Data Migration

- **6. Big Data Processing**
    6.1 Amazon EMR (Elastic MapReduce) - (Setting up clusters), (Using Apache Hadoop, Spark, Hive, Presto)

    6.2 AWS Glue - ETL (Extract, Transform, Load), Data Cataloging

    6.3 Amazon Kinesis - Kinesis Data Streams, Kinesis Data Firehose, Kinesis Data Analytics

- **7. Analytics** - Amazon Athena (Querying data in Amazon S3), (Inegrating with AWS Glue), Amazon QuickSight (Visualizations)

- **8. Data Integration and Orchestration** - AWS Step Functions, AWS Glue Workflows, Amazon Managed Workflows for Apache Airflow

- **9. Machine Learning and Artificial Intelligence** - Amazon SageMaker, AWS ML Services (Comprehend, Rekognition, Polly, etc.), Building Data Pipelines for Machine Learning

- **10. Miscellaneous**
    10.1 Security and Compliance - Amazon Virtual Private Cloud, Amazon Identity and Access Management, Amazon Key Management Services, AWS Shield, AWS WAF

    10.2 Monitoring and Logging - Amazon CloudWatch, AWS CloudTrail, AWS Config

    10.3 AWS Cost Management - AWS Cost Explorer, AWS Budgets, AWS Pricing Calculator

    10.4 DevOps and Automation - AWS CodePipeline, AWS CodeBuild, AWS CodeDeploy, Infrastructure as Code (IaC) with AWS CloudFormation and AWS CDK

# Course + Internship

## Course Overview & Time Duration:

- ✔ **11 Months** of Practical & Module Training
- ✔ **100+** Expert-led Classes
- ✔ **200+ Hours** of Comprehensive Learning
- ✔ **14 Specialized** Modules
- ✔ **3-Month** Internship

## After the Preparation Placement opportunity

Tech Mahindra · wipro · HCL Technologies · TATA

Google · Microsoft · facebook · amazon

Capgemini · accenture · Walmart · Deloitte.

BT Data and Surveying Services India Pvt. Ltd · Axis Bank · Celebal Technologies · Netflix

Datavant · DataRobot · MyMoneyMantra · teradata.

splunk · airtel · databricks · Oracle

OLA · Nokia Connecting People · Infosys Navigate your next · IBM
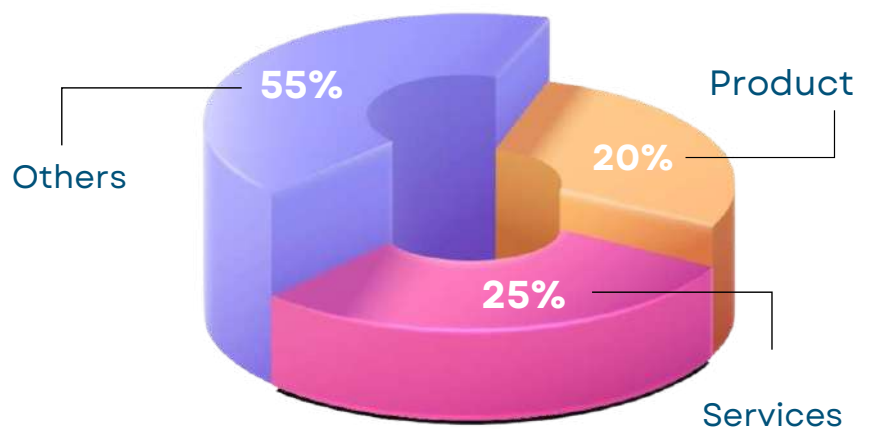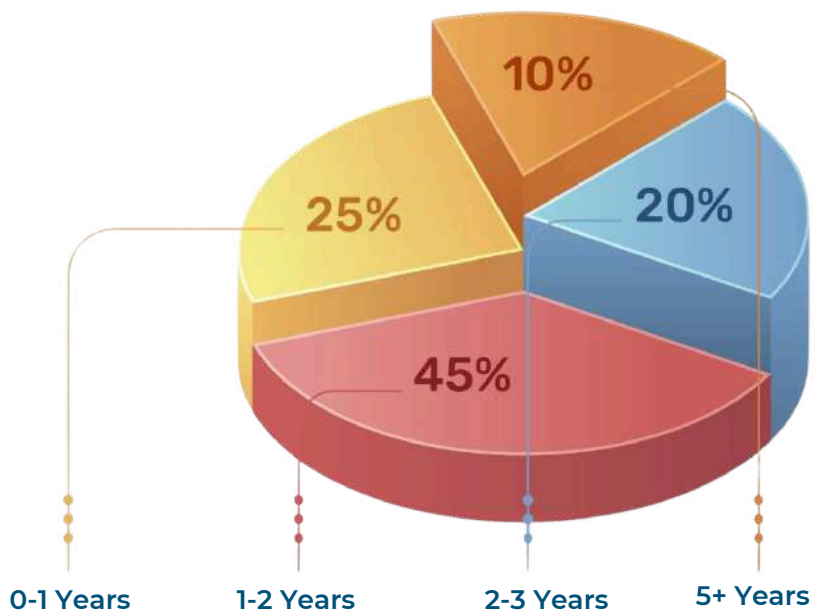
# Data Engineering

Data Engineering professionals are in high demand, ensuring significant job security in this field. This offers great motivation for those seeking a stable and reliable career path.
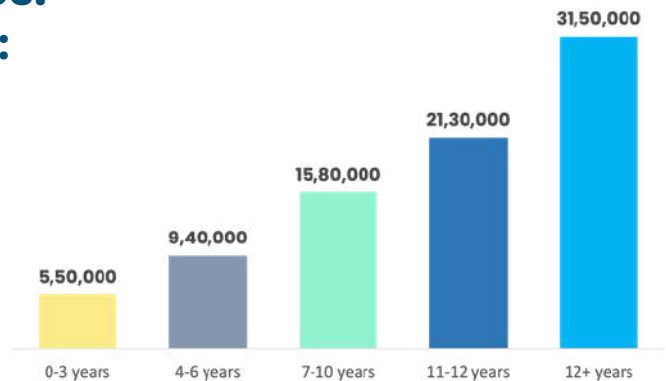
## Learners' Industry Background

55% Others

20% Product

25% Services

## Learners' Work Experience

10%

20%

25%

45%

0-1 Years    1-2 Years    2-3 Years    5+ Years

# Take Your **Career** to the **Next Level**

The global **Data Engineering** job market is projected to grow by **44% from 2021 to 2030**, resulting in the creation of 3.5 million new jobs worldwide. In India, it is anticipated that 309,000 new Data Engineering positions will emerge by 2030, constituting 9% of the global demand. The average salary for a Data Engineering professional in India ranges from approximately **Rs. 5 to Rs. 14 lakhs per annum.**
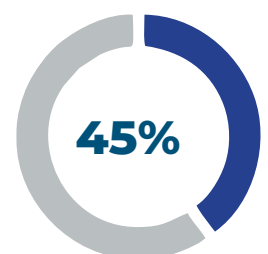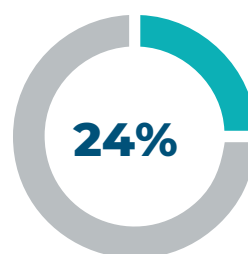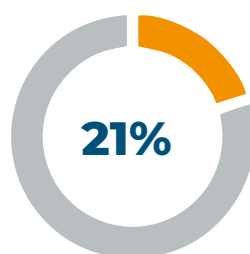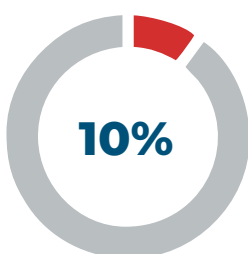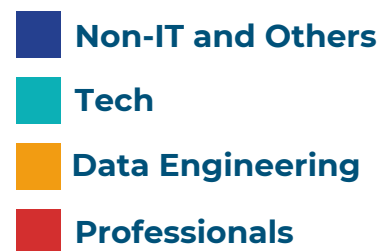
## Data Engineering Job Roles: Salary Trends 2023-2024:

## Salary By Experience

The **average median salary** of a data scientist in India is around **INR 21,00,000**

| Experience | Salary |
| --- | --- |
| 0-3 years | 5,50,000 |
| 4-6 years | 9,40,000 |
| 7-10 years | 15,80,000 |
| 11-12 years | 21,30,000 |
| 12+ years | 31,50,000 |

## Learners' Expertise

- Non-IT and Others
- Tech
- Data Engineering
- Professionals

**10%** Professionals
**21%** Data Engineering
**24%** Tech
**45%** Non-IT and Others

# Program **Details**

## Qualification:

Technical and Nontechnical Professionals with basic-level programming knowledge can also apply

## Course Duration: 11 Months

## Weekend Classes
## Saturday & Sunday: 2 hours/day

| W | W |
|---|---|

**b**     **l**

Experienced Professional Data Engineer imparts valuable real-world expertise and efficient strategies, equipping students for achievement in the field.

### TOTAL FEES:

### ₹ 2,99,000/-

**Included+18% GST**

### EASY EMI

**Registration Fee: ₹ 5000/-**

**Financing Partners**

Razorpay   Credit Fair   BAJAJ FINSERV   UPI   IDFC FIRST Bank

# Certificates

Upon completing this Professional Certificate Program in Data Engineering, you will receive a program completion certificate from Digicrome, validating your expertise as a data engineering professional. In addition to this, you will earn eight more certificates as detailed below, and a three-month internship certificate. All these certificates are globally recognized and approved.

**After the Completion of the Course, You'll get 9 Professional Certificates:**

1 Masters Program in Big Data Engineering and Cloud Computing

2 Fundamentals of Python Programming

3 Machine Learning and Artificial Intelligence

4 Big Data Professional

5 Cloud Computing using Amazon Web Services

6 Database Management using SQL

7 NoSQL Database Professional

8 Data Processing and Analytics

9 Automation and Deployment of Data Engineering Models

### Internship Certificates

1 Internship Acknowledgement
2 Letter of Recommendation

Digicrome's certification programs are crafted to meet the demands of today's dynamic job market. These certifications are recognized for their rigorous standards and practical relevance, ensuring that you gain not only theoretical knowledge but also hands-on experience.

We are proud to have partnerships with multiple multinational corporations, which significantly enrich our certification programs. These collaborations ensure that our exams are industry-relevant and up-to-date with the latest trends and technologies.

Whether you are looking to advance in your current role or seeking new opportunities, Digicrome's certifications provide a robust foundation for career growth. Join us and take the next step towards achieving your professional goals with qualifications that are both respected and demanded by leading companies worldwide.

Digicrome is excited to announce the availability of a range of certifications designed to enhance your career prospects.

**A learner can choose any one option out of the following:**

## Option - 1

## Microsoft Certified: Power BI Data Analyst Associate

**+**

# Choose any one from below:

### Microsoft

| Program |
| --- |
| ✓ Microsoft Azure AI Fundamentals |
| ✓ Microsoft Office Specialist - Excel Expert |
| ✓ Microsoft Certified Educator - Technology Literacy for Educators |
| ✓ Microsoft Azure Data Fundamentals |
| ✓ Microsoft Azure Fundamentals |

### IBM

| Program |
| --- |
| ✓ IBM Certified Data Scientist - Machine Learning Specialist |

### PMI Project Management Ready™

| Program |
| --- |
| ✓ PMI Project Management Ready ™ |

### INFORMATION TECHNOLOGY

| Program |
| --- |
| ✓ Information Technology Specialist - Data Analytics |
| ✓ Information Technology Specialist - Artificial Intelligence |
| ✓ Information Technology Specialist - Python |

### COMMUNICATION SKILLS FOR BUSINESS

| Program |
| --- |
| ✓ Communication Skills for Business - Professional Communication |
| ✓ Communication Skills for Business - English for IT |

### Choose any three from the given:

**Microsoft**

| Program |
| --- |
| ✓ Microsoft Azure AI Fundamentals |
| ✓ Microsoft Office Specialist - Excel Expert |
| ✓ Microsoft Certified Educator - Technology Literacy for Educators |
| ✓ Microsoft Azure Data Fundamentals |
| ✓ Microsoft Azure Fundamentals |
| ✓ Microsoft Office Specialist - Excel Associate |

**INFORMATION TECHNOLOGY**

| Program |
| --- |
| ✓ Information Technology Specialist - Data Analytics |
| ✓ Information Technology Specialist - Artificial Intelligence |
| ✓ Information Technology Specialist - Python |

**COMMUNICATION SKILLS FOR BUSINESS**

| Program |
| --- |
| ✓ Communication Skills for Business - Professional Communication |
| ✓ Communication Skills for Business - English for IT |

**IBM**

| Program |
| --- |
| ✓ IBM Certified Data Scientist - Machine Learning Specialist |

**PMI Project Management Ready™**

| Program |
| --- |
| ✓ PMI Project Management Ready ™ |

**Important Note:** Students must pass the relevant exams in order to obtain these worthwhile certificates. The exam fees are fully covered by Digicrome's tuition fee, so there are no additional costs involved for the first attempt. Please note that this just includes one try. There will be additional charges for each attempt if a student fails the exam the first time and wants to retake it. In order to ensure that financial limitations do not impede your professional development, we want to offer a simple and reasonably priced certification pathway that upholds the quality and worth of our certification process.

# Digicrome

## INDIA
Office (IND)
C-108 Second Floor, Sector 2
Noida, Uttar Pradesh 201301
Contact:- 01203131297

## USA
Office (USA)
30 N Gould St Ste R Sheridan,
Wyoming 82801
Contact:- 013015292014

## More Information

**Phone**
01203113765

www.digicrome.com

**Mail**
info@digicrome.com