



Data Engineering

# DATA ENGINEER- PROFESSIONAL TRAINING PROGRAM



**500K+**  
Satisfied Students



**100K+**  
Online Students



**300+**  
Trainers



**300K+**  
Placements



**200+**  
Global Certifications



**150+**  
Companies

# Module 1: Programming Languages for Data Engineering

---

## ✓ Python For Data Engineering:

- Data types, loops, and conditionals
- File handling (CSV, JSON, TXT)
- Functions, lambda expressions, error handling
- Working with libraries: Pandas, NumPy
- Data manipulation and cleaning
- Working with APIs and JSON data
- Python for automation and scripting

## ✓ Scala (for Apache Spark):

- Basics of Scala syntax and structure
- Functional programming (map, reduce, filter)
- Case classes and immutability
- Using Scala with Spark (RDDs, DataFrames)

## ✓ Java Basics:

- OOP principles: class, object, inheritance
- Exception handling and file I/O
- Working with Hadoop Java APIs (MapReduce basics)

# Module 2: Distributed Processing Frameworks

---

## ✓ Apache Spark:

- Spark architecture and cluster modes
- RDDs vs DataFrames vs Datasets
- Spark SQL: working with structured data
- Transformations and actions
- Spark Streaming: real-time data processing
- Machine learning pipelines with MLlib (intro level)

## ✓ HDFS (Hadoop Distributed File System):

- Blocks and replication
- NameNode and DataNode
- File storage commands and usage
- Integration with Spark, Hive, Sqoop

## ✓ YARN:

- YARN architecture
- ResourceManager and NodeManager
- Application lifecycle and scheduling

## ✓ Apache Mesos:

- Mesos vs YARN

- Resource allocation and scheduling
- Deploying Spark/Flink on Mesos

## Module 3: Storage Systems

---

### ✓ Relational Databases:

- Database design concepts
- SQL queries (SELECT, JOIN, GROUP BY)
- Indexing, constraints, normalization
- MySQL, PostgreSQL / PostGIS, Oracle

### ✓ NoSQL Databases:

- Concepts: CAP Theorem, BASE vs ACID
- Key-Value, Document, Columnar types

### ✓ Cassandra:

- Data model, keyspaces, replication, CQL
- Write-heavy optimizations

### ✓ MongoDB:

- BSON, collections, indexing, aggregation framework
- CRUD operations and performance tuning

### ✓ HBase:

- Column-family model, integration with Hadoop
- Use cases for wide-column storage

## Module 4: Data Warehousing

---

### ✓ Apache Hive:

- Hive architecture: Metastore, Driver, Executor
- HiveQL syntax: DDL, DML, Joins, UDFs
- Partitions and Bucketing
- Integration with HDFS and Spark

### ✓ Snowflake:

- Cloud-based architecture: compute vs storage
- Warehouses, databases, and schemas
- Time Travel, cloning, and zero-copy restore
- Working with semi-structured data (JSON, XML)
- Performance tuning and caching

## Module 5: Real-Time Data Processing

---

### ✓ Apache Kafka:

- Producer-consumer architecture
- Topics, partitions, brokers
- Kafka Streams basics
- Kafka + Spark/Flink integration

#### ✓ **Apache Flink:**

- Stream vs batch processing
- Windows, time semantics (event vs processing time)
- Stateful operators
- Fault-tolerant streaming with checkpoints

#### ✓ **Apache Storm:**

- Storm architecture: Spouts and Bolts
- Building topologies
- Real-time analytics pipelines

---

## **Module 6: ETL & Workflow Orchestration**

#### ✓ **Apache Airflow:**

- DAGs, Tasks, Operators (PythonOperator, BashOperator, etc.)
- Scheduling and triggering jobs
- XComs, task dependencies, retries
- Airflow UI for monitoring and management
- Integrating with DBs, APIs, Spark, and cloud services

---

## **Module 7: Cloud Platforms (AWS, Azure, GCP)**

#### ✓ **AWS:**

- S3, IAM, EC2 basics
- Redshift for data warehousing
- AWS Glue for ETL
- EMR for big data

#### ✓ **Azure:**

- Azure Data Factory (pipelines)
- Azure Blob Storage
- Azure Synapse for analytics

#### ✓ **GCP:**

- BigQuery for analytics
- Google Cloud Storage
- Dataflow for stream and batch pipelines

---

## **Module 8: Data Modeling**

- ✓ OLTP Vs OLAP Systems
- ✓ Entity-Relationship (ER) Modeling
- ✓ Star Schema And Snowflake Schema
- ✓ Dimension And Fact Tables
- ✓ Slowly Changing Dimensions (SCD Types 1, 2, 3)
- ✓ Data Vault Modeling Basics

## Module 9: System Design for Data Engineering

---

- ✓ Designing Batch And Stream Pipelines
- ✓ Data Lake Vs Data Warehouse Architectures
- ✓ Scalable Ingestion Pipelines
- ✓ Fault Tolerance, Latency, High Availability
- ✓ Case Studies: Uber, Netflix, Spotify Pipelines
- ✓ Choosing Tools For Use Cases

## Module 10: Data Ingestion Tools

---

- ✓ **Apache Sqoop:**
  - Sqoop architecture and setup
  - Import/export from MySQL, Oracle to HDFS/Hive
  - Incremental loads (append, last-modified)
  - Performance tuning using mappers
- ✓ **Capstone Projects**
  - Batch ETL Pipeline using Airflow + Hive + MySQL
  - Real-Time Streaming using Kafka + Spark + MongoDB
  - Cloud Data Pipeline using AWS S3 + Glue + Redshift
  - System Design Case Study (architecture presentation)

# e-Learning through LMS

## Learning **Management System**

Our LMS (LearnPitch) is for the administration, documentation, tracking, reporting, automation, and delivery of educational courses, training programs, or learning and development programs.

Our LMS has been designed to identify training and learning gaps, using analytical data and reporting to keep you up with the class activities.

## Key Features **Learning Management System**



Live Sessions with Class recordings



Get study material with Assignments.



Track your curriculum covered.



Track your class wise attendance



Share your feedback for Trainer & Training



Get your Training Certificate from LMS



# | Training Certification

## Earn Your **Certificate**

Your certificate and skills are vital to the extent of jump-starting your career and giving you a chance to compete in a global space.



|Croma Campus is Nasscom Certified

Croma Campus is now

**NASSCOM**<sup>®</sup>

Certified Member



[www.cromacampus.com](http://www.cromacampus.com)

00 | [www.cromacampus.com](http://www.cromacampus.com)

# |Croma Campus! Reviews



"The most rewarding part of my experience has been achieving a prestigious certification in the subject that I love. Moreover, the training offered out by the specialists are of world-class and prepares out the students for corporate world. For me Croma Campus means a lot."

*"By The Students For The Students,"*

## Your Success is **Our Story**



**Bharat**

"I am fully satisfied with the excellent training services received by the expert staff at Croma Campus. I want to thank Croma Campus for providing me with the most innovative and affordable training services for learning all the software testing procedures and guidelines. "



**Ankit**

"It was a lifetime experience for me to get trained by IT Experts of Croma Campus. What I liked most about the training was the consistent high-quality education, which was friendly and co-active. The placement department was also proactive, they keep me updated regarding new job opportunities and provide the grooming session to crack the interview. At last, I would like to thank all faculty members of Croma Campus for their immense help and support. "



**Umesh**

"Without any second thought, I will give Croma Campus 10/10. Their placement department is highly proactive. I remember they started scheduling interviews for me from the very next day when I told them my course has been completed. These people are doing a phenomenal job and I highly recommend Croma Campus to everyone. "



**Shams Khan**

"Croma Campus is doing a phenomenal job in the IT training industry. The reason why I decided to join their training program was that they provide quality training at very a nominal price. Plus, the online training mode was also a factor due to which I decided to join the training program of Croma Campus as I didn't want to attend physical classes. "



# | Meet Our Team



## Sales Team

Our Sales team is highly passionate, emphatic, positive attitude, great listening skills, ability to deliver quick solutions, and they are multitasker too. Our team always remains up-to-date about all the latest technologies and market trends. With effective communication skills, they always work to deliver the right information to customers when it is needed.

## Product Team

Our product team is highly functional and collaborative working together to achieve the common outcome of designing exceptional digital experiences. Each of our members is a contributor to help us achieve success in long-run. Sitting at the high-end of technology and innovation, team helps to deliver high-end customer experiences and always comes out with a big idea as a game-changing plan.



## Marketing Team

Our Marketing team works as gladiators and helps us to achieve business success in all possible ways. They are included in almost everything either it is building a brand, creating brand awareness, promoting products or services, delivering trailblazing customer experiences or increasing engagement at public forums. They are the true backbone of the Company.

## Content Team

Our content team is responsible for ideation, creation, optimization, and distribution of content throughout the company. The team always starts its work with a strategy, how to create high-quality contents, and how to promote or share the content. Our in-house content team help us to produce all types of contents either they are educational content pieces, marketing content, SEO content, or any other forms too.



## Customer Access Team

This is the team that has actually been taken up us from reactive state to a pro-active state. The team utilizes high-valued solutions to satisfy customers in all possible ways. It is truly said that no company can succeed if your customers are not satisfied. And our customer success team is dedicatedly working to keep all the customers satisfied and we always consider our customer feedback on priority.

## HR Team

Our HR team is committed to provide high-end solutions to employees as they require. Our HR team has the right skills and knowledge to make sure that the HR department can always be legally and strategically successful. They know how to keep employees motivated all the time with the best HR policies and fun activities too from time to time.



# | Glimpse Of Our Office

## Look **Who We are**

Our office's infrastructure comprises all the necessary software and network resources that are required to deliver IT & Design, Human Resources, Digital Marketing, and training services.

We are well-equipped with bright designed work bays for employees and managers having separate cabins with spacious cafeteria and training classrooms.



# | About Croma Campus

“ Our Mission is to Build Nation through Education & Beyond Limitation. ”

**CROMA** CAMPUS | #1 Asia's Leading Edtech Company

Croma Campus Training & Development Private Limited is an education platform providing rigorous industry-relevant programs designed and delivered in collaboration with world-class faculty, industry & Infrastructure. In the past 14 years we have trained 18000+ candidates and out of which we are able to place 12000+ professionals in various industries successfully.



We're Here to Help –  
**Reach Out to Our Global Offices**

 **India**

Croma Campus Training & Development Private Limited

G-21, Block-G, Sector 3, Noida, Uttar Pradesh-201301

+91-971 152 6942 [Chat on WhatsApp](#)

[View on Google Maps](#)

 **Canada**

Croma Campus Training & Development Private Limited

279-B Baltzan BLVD, Saskatoon, Saskatchewan, Canada-S7W0S1

+1 (647) 370-6470 [Chat on WhatsApp](#)

[View on Google Maps](#)

 **USA**

Croma Campus Training & Development Private Limited

454 church street Cairo, GA-39828, United States of America

+1 (647) 370-6470 [Chat on WhatsApp](#)

[View on Google Maps](#)

follow us on:



CromaCampusNoidaOfficial/facebook



cromacampus/pinterest



cromacampus/instagram



cromacampusofficial/linkedin



cromacampus/twitter



CromaCampusOfficial/youtube

