



# Launch Your Career in Data Engineering

Master the skills required to design, build, and manage powerful data systems with DataStream Institute. Learn from industry experts through a structured, hands-on program built for the data-driven future.



# TABLE OF CONTENTS

<b>01</b>	<b>COVER PAGE</b>	
<b>02</b>	<b>ABOUT DATASTREAM INSTITUTE</b>	
<b>03</b>	<b>VISION &amp; MISSION</b>	
<b>04</b>	<b>WHY DATA ENGINEERING IS IN DEMAND</b>	
<b>05</b>	<b>PROGRAM OVERVIEW</b>	
<b>06</b>	<b>TOOLS &amp; TECHNOLOGIES YOU'LL LEARN</b>	
<b>07</b>	<b>CURRICULUM BREAKDOWN</b>	
<b>08</b>	<b>PROJECTS &amp; HANDS-ON LEARNING</b>	
<b>09</b>	<b>CAREER OPPORTUNITIES</b>	
<b>10</b>	<b>PLACEMENT SUPPORT</b>	
<b>11</b>	<b>TESTIMONIALS</b>	
<b>12</b>	<b>WHO CAN APPLY</b>	
<b>13</b>	<b>CONTACT INFORMATION</b>	

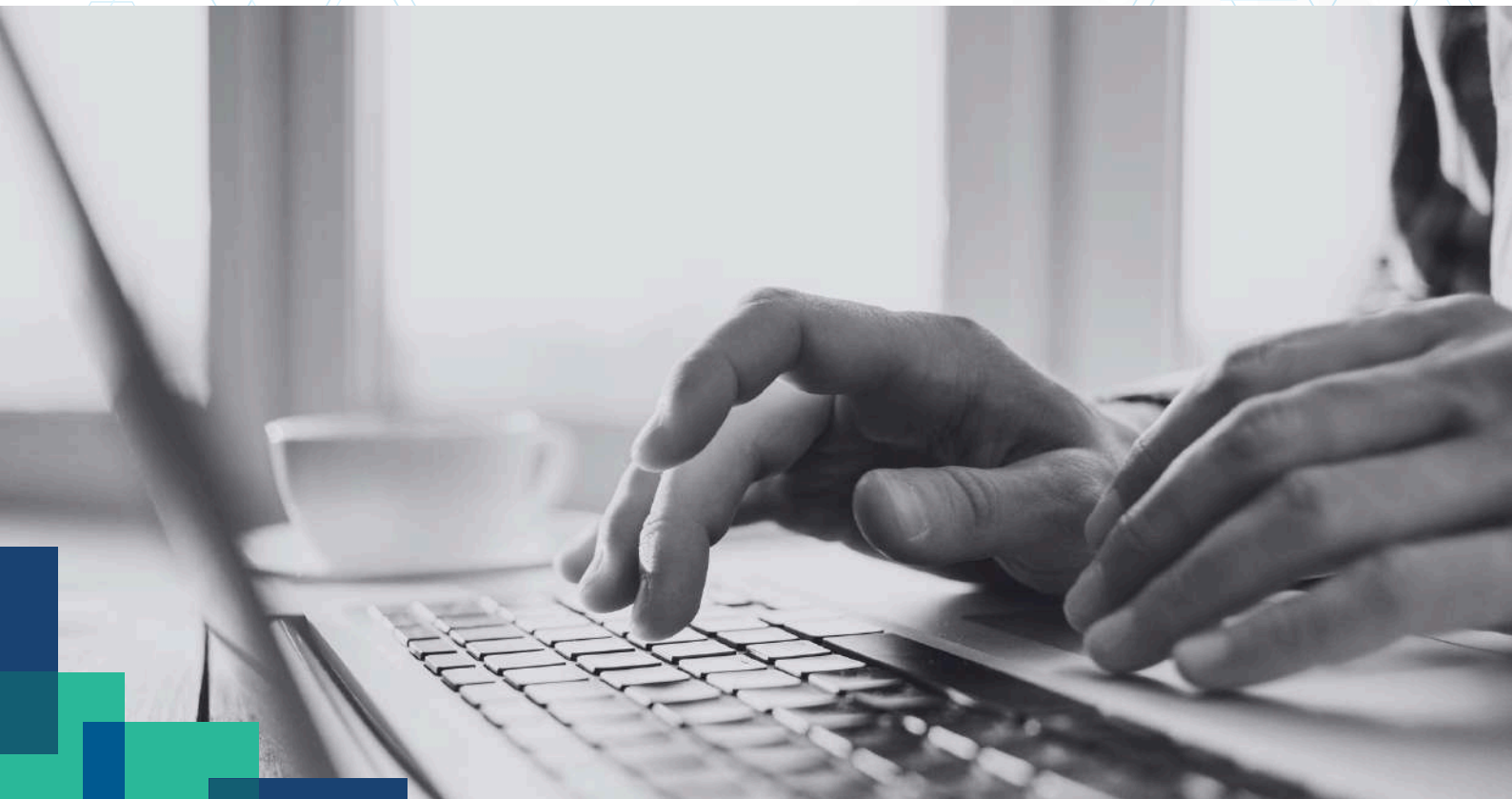
# Introduction

## Building the Next Generation of Data Professionals

In today's digital economy, data has become one of the most valuable assets for businesses. Organizations across industries are investing in Data Engineering technologies to collect, process, and analyze massive volumes of data to drive smarter decisions.

Data Engineering combines raw data, advanced algorithms, and scalable software systems to create powerful data infrastructures that support modern analytics, machine learning, and artificial intelligence applications.

The DataStream Institute Data Engineering Program is designed to equip learners with the practical knowledge and hands-on experience required to build and manage modern data systems. Guided by experienced professionals, this program prepares you to meet the growing demand for skilled Data Engineers and thrive in the data-driven future.



# Our Vision

DATASTREAM INSTITUTE

“ To bridge the gap between education and industry by providing practical, career-focused training in emerging data technologies. ”

# Our Mission

1

### Industry-Focused Training

Specialized training in modern data and technology domains shaping the digital future.

2

### Practical Learning Approach

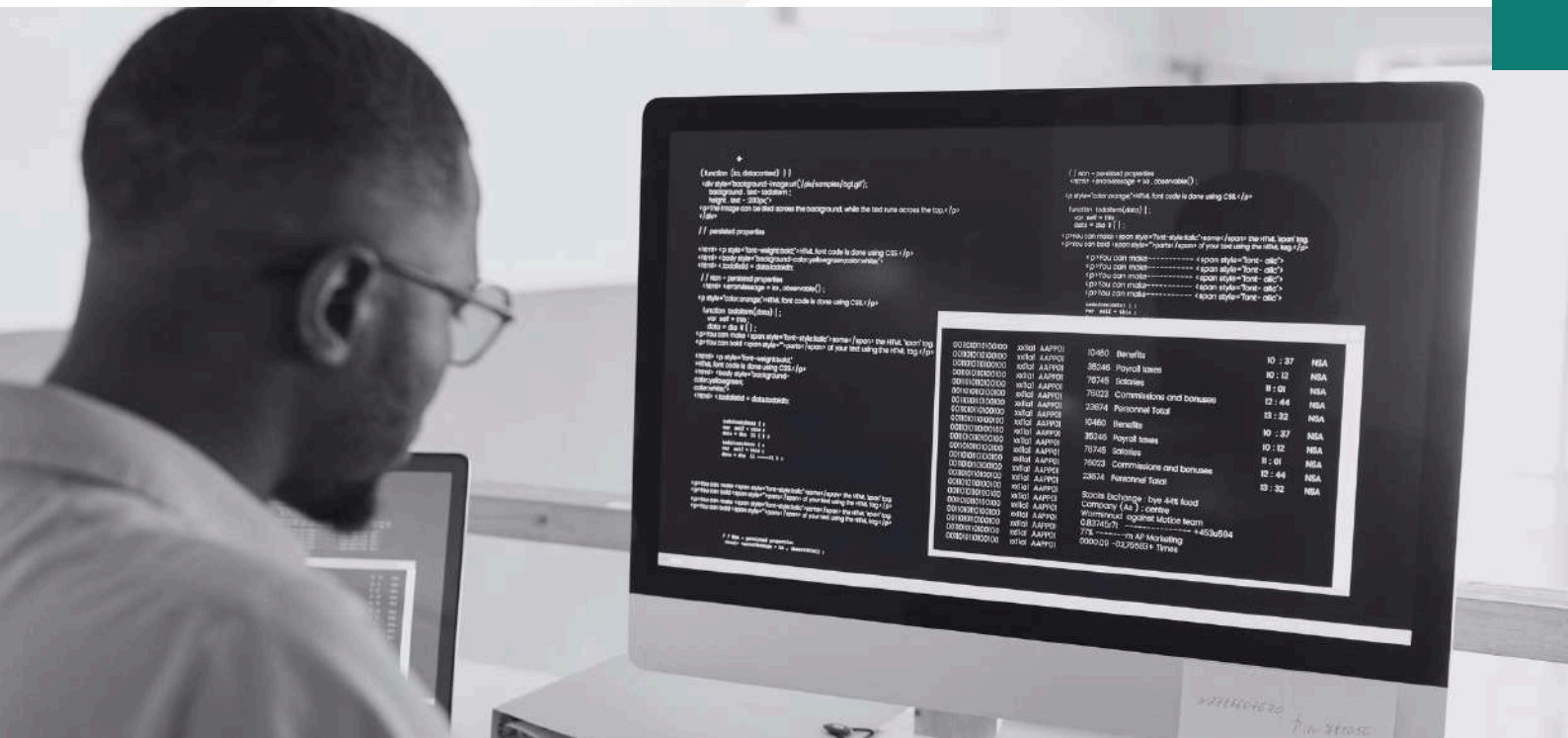
Fast-paced, hands-on education designed to build job-ready skills quickly.

3

### Career & Placement Support

Dedicated guidance to help students confidently enter the technology workforce.





# Why Data Engineering?

In today's digital economy, the volume of data generated every day is growing rapidly, and businesses rely heavily on this data to make strategic decisions and drive innovation. However, raw data needs to be collected, processed, and structured before it becomes useful. This has made Data Engineering a critical function in modern organizations, especially with the rise of Big Data, cloud computing, and AI technologies. As companies continue to adopt data-driven approaches, the demand for skilled Data Engineers is increasing across industries.

## 90%

Companies Are Investing in Data & AI. Businesses across industries are adopting data-driven strategies.

## 40%+

Growth in Data Engineering Roles Data-related jobs are among the fastest-growing careers in the tech industry.

## ₹10-30 LPA

Average Salary in India Data Engineers are highly paid due to their critical role in building data systems.

# Program Overview

Elevate your career with the DataStream Institute Data Engineering Program, thoughtfully designed to equip learners with the skills required to thrive in today's data-driven world. This comprehensive program combines strong foundational concepts, real-world case studies, and hands-on projects, ensuring a deep and practical understanding of Data Engineering.

Designed for both working professionals and aspiring learners from diverse backgrounds, the program offers a flexible and immersive learning experience through interactive online sessions and expert-led training. With continuous mentorship and high-engagement sessions, learners gain valuable industry insights and practical exposure to real-world data challenges.

The curriculum covers a wide range of modern tools and technologies essential for today's data ecosystem, including SQL, Python, Big Data frameworks, Cloud platforms (AWS, GCP, Azure), Spark, Kafka, and data visualization tools. This ensures that students develop the technical expertise required to build and manage scalable data systems.

Join DataStream Institute and take the next step toward becoming a skilled Data Engineer, equipped with the knowledge and confidence to succeed in today's competitive job market.

**200 Hours  
Interactive  
Live  
Sessions**

**80 Hours  
Projects &  
Assignments**

**Learn  
from  
Experts**

**Career &  
Placement  
Support**

**Industry  
Ready  
Curriculum**



# Curriculum Breakdown

---

## Module 1: Linux & Shell Scripting

- Linux Fundamentals & Commands
- Shell Scripting Basics
- Stream Editor (SED)
- AWK Programming
- Linux Administration Essentials

## Module 2: Databases & SQL (Oracle)

- Database Fundamentals
- SQL & Client Tools
- Data Integrity Constraints
- Operators, Joins & Subqueries
- SQL Functions & Window Functions
- Indexing, Views & Partitions
- Performance Optimization

## Module 3: Python for Data Engineering

- Python Basics & Data Types
- Operators & Flow Control
- Functions & Advanced Concepts (Lambda, Decorators)
- OOP Concepts (Constructors, Inheritance)
- Regular Expressions
- Libraries: NumPy, Pandas

## Module 4: Big Data & Hadoop Fundamentals

- Introduction to Big Data & Hadoop
- Hadoop Ecosystem Overview
- Big Data Challenges & Solutions
- Hadoop vs Traditional Systems (RDBMS, Data Warehousing)
- NoSQL Introduction & Use Cases
- On-Premise vs Cloud Data Projects

# Curriculum Breakdown

---

## **Module 5: HDFS (Hadoop Distributed File System)**

- Cluster Environment & Hadoop Clusters
- HDFS Architecture & Components
- Storage & Data Replication
- Failover Mechanism
- Hadoop Versions & Evolution

## **Module 6: MapReduce Programming**

- MapReduce Architecture & Lifecycle
- Mapper, Reducer & Driver Code
- Input/Output Formats
- Combiner & Partitioner
- Job Chaining & Joins
- Debugging & Optimization
- Data Locality & Performance Tuning

## **Module 7: Data Ingestion with Sqoop**

- Introduction to Sqoop
- Connecting with Relational Databases
- Import & Export Data
- Incremental Data Load
- Performance Optimization

## **Module 8: Hive & Data Warehousing**

- Hive Architecture & HiveQL
- Partitioning & Bucketing
- Data Modeling & Optimization
- UDFs & SerDe
- Semi-Structured Data Processing
- Real-Time Use Cases & Query Optimization

# Curriculum Breakdown

---

## Module 9: YARN & Resource Management

- Introduction to YARN
- YARN Architecture (Resource Manager, Node Manager)
- Job Scheduling & Process Flow
- Monitoring with YARN UI
- MapReduce vs YARN

## Module 10: Apache Spark & PySpark

- Spark Fundamentals (RDDs, DataFrames)
- Spark SQL & Data Processing
- Working with Multiple Data Sources
- Joins, Transformations & Aggregations
- Window Functions & Analytics
- Performance Optimization & Adaptive Execution

## Module 11: Data Visualization (Power BI)

- Power BI Desktop & Data Modeling
- Data Transformation & Query Editor
- DAX Expressions
- Dashboard Creation & Reporting
- Power BI Service & Deployment

## Module 12: Capstone Projects

- End-to-End Data Pipeline Projects
- Cloud-Based Data Engineering Projects
- Real-World Industry Case Studies

## Module 13: Career Preparation

- Resume Building & LinkedIn Optimization
- Mock Technical Interviews
- Interview Preparation
- Placement Assistance & Referrals

# Curriculum Breakdown

---

## Module 15: Cloud Data Engineering

### AWS

- EC2, S3, RDS
- IAM, Lambda, CloudWatch
- Redshift, EMR, Glue, Athena

### GCP

- Compute Engine & Cloud Storage
- BigQuery & DataProc
- Cloud Functions & Scheduler
- IAM & Monitoring

### Azure

- Virtual Machines & Blob Storage
- Azure SQL Database
- Azure Data Factory & Synapse
- Azure Databricks & Active Directory

## Module 16: Data Modeling & Warehousing Concepts

- Star & Snowflake Schema
- Fact & Dimension Tables
- Data Modeling Best Practices

## Module 17: Data Governance & Security

- Data Security Concepts
- Access Control (IAM)
- Data Privacy & Compliance Basics

## Module 18: Career Preparation

- Resume & LinkedIn Optimization
- Mock Interviews
- Placement Support