



In Collaboration with



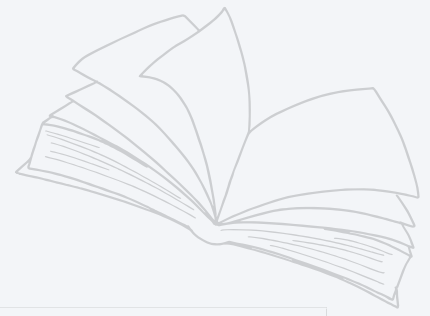
EXECUTIVE PROGRAM IN

DATA ENGINEERING

Designed for Data Engineering
Enthusiasts to redefine the Future
of Data Engineering

DURATION
12 MONTHS

Table of Contents



3	About Meritshot
4	Data Engineering Emergence
5	Why Data Engineering with Meristhot?
6	Positions Offered
7	Learn with Top Industry Experts
8	Program Curriculum
10	Industry Projects & Case-studies
11	Libraries and Tools Covered
13	Program Outcomes
14	Career Support
15	Success Stories
17	Program Certification
18	Hiring Partners
19	Program Details

About

Meritshot

Meritshot is an online education platform to help data engineering enthusiasts and working professionals looking to transition their careers in the Data Analytics, Data Science, Data Engineering and AI Fields. We foster a flourishing community of working professionals who wish to upskill themselves.

At Meritshot, we are working towards transforming the traditional approach of learning with the modern way of learning where individuals learn experientially with ideas and creativity coming from top industry mentors and coaches. Based on our market research and conversation with industry veterans, we have identified Data Science as one of the sectors with a very huge demand and very low supply.

Our vision is to design and deliver a high-octane online Executive program in Data Engineering, Data Science, Data Analytics to feed the supply-demand imbalance surge in India and make India a global hub for data science. If you are reading this, you may wish to accelerate your career in Data Engineering with Meritshot, we promise to equip you with the perfect mix of business understanding and technical skills to help you achieve the same.

Roshan Sharma

Founder & CEO

Meritshot

Data Engineering Emergence

The Modern Era

In the modern era, data engineering has emerged as a critical discipline, evolving far beyond its early roots in basic ETL processes and relational databases. Today, it sits at the heart of data-driven organizations, enabling everything from real-time analytics to machine learning at scale.

This transformation has been fueled by an explosion in data volume, the rise of cloud computing, and the proliferation of powerful open-source tools. Modern data engineers work with scalable technologies like Apache Spark, Kafka, dbt, and Snowflake, orchestrated through platforms such as Airflow or Prefect.



Data engineers today go beyond moving data—they design scalable systems that power real-time analytics and informed decision-making. With the rise of cloud-native stacks, AI tools, and data mesh, their role has become central to innovation and digital transformation.

Why data Engineering with Meritshot?



Certification Program

Get certified on successful completion of the Program in Data Analytics.



Personalized Evaluation

Get personalized evaluation on every assignment and project by domain experts.



Domain Specialization

Get full fledged experience of a specific domain to showcase a focused Data science portfolio.



Personalized Projects

Work only on personalized Portfolio projects that suits your work profile and background.



Dedicated Mentorship

Get dedicated mentorship on every step of learning Data Analytics throughout program.



Amazing Job Offers

Great Job Offers after successful completion of the Data Analytics Program.

Positions Offered



**Data
Engineer**

**Junior
Data
Engineer**

**Senior
Data
Engineer**

**Data
Platform
Engineer**

**Data
Architect**

**ETL
Engineer**

**Big Data
Engineer**

Learn with Top Industry Experts



Saurabh Daund

Data Scientist
Aidetic



Sakshi Pandey

Principle Researcher
Cialfor Research



Dr. Shweta Singh

Associate Professor
Delhi Univesity



S.D. Siddiqui

Associate Professor
Delhi University



Jilu Balakrishnan

AI Specialist
ISDC GLOBAL



Jiohn Martini

Data Evangelist
Meritshot



Heena Arora

Data Scientist
PwC



Bharath Reddy

Data Scientist
Meritshot

Program Curriculum

Preparatory Course

MASTERING ESSENTIAL EXCEL FUNCTIONS

Learn essential Excel functions such as IF, VLOOKUP, INDEX/MATCH, and TEXT, along with logical operators to perform dynamic data analysis.

DATA VISUALIZATION AND REPORTING TOOLS

Explore tools like Pivot Tables, Charts, and Conditional Formatting to summarize and visualize data effectively.

SCENARIO MODELING AND WHAT-IF ANALYSIS

Use Scenario Manager, Goal Seek, and What-If Analysis tools to model data-driven decision-making and explore business scenarios.

SQL & Database Fundamentals

SQL BASICS: DDL, DML, AND DQL

Understand SQL basics including Data Definition (DDL), Data Manipulation (DML), and Data Query (DQL) operations.

DATA VISUALIZATION AND REPORTING TOOLS

Gain insight into database structure—learn about tables, keys, and constraints to maintain data integrity

DATA CLEANING, PREPARATION, & PROCESSING

Use Scenario Manager, Goal Seek, and Data Tables to perform data-driven simulations and support decision-making.

Program Curriculum

Programming with Python

CORE PYTHON PROGRAMMING

Learn the basics of Python: data types, variables, loops, functions, and conditionals for writing clean, logical code.

FILE HANDLING & DATA STRUCTURES

Work with files, handle exceptions, and use list/dictionary comprehensions for efficient data processing.

PANDAS & API DATA ACCESS

Use Pandas for reading and transforming data (CSV, JSON, XML), and make basic API requests to access web data.

Data Visualization & BI Tools

BUILDING INTERACTIVE DASHBOARDS

Learn to create visual dashboards using Power BI and Tableau, incorporating features like DAX formulas, slicers, filters, storyboards, and calculated fields for rich, interactive reporting.

CONNECTING AND MODELING DATA

Connect BI tools to various data sources, build accurate data models, and define relationships to enable meaningful and scalable analysis.

APPLYING BI TO REAL-WORLD SCENARIOS

Develop a practical sales and marketing dashboard to apply your BI skills in real business use cases and decision-making.

Program Curriculum

Git & Version Control

INTRODUCTION TO GIT AND VERSION CONTROL

Learn the fundamentals of Git, including initializing repositories, cloning, staging changes, committing, and pushing to remote repositories.

BRANCHING AND GITHUB COLLABORATION

Understand how to create branches, collaborate using GitHub, and manage code changes in a team environment.

WORKING WITH GIT IN VSCODE AND TEAM PROJECTS

Use Git within Visual Studio Code and apply your knowledge in a mini team project to practice real-world collaboration.

AI Business Tools

CHATGPT & BARD

ChatGPT is a free-to-use AI system. Use it for engaging conversations, gain insights, automate tasks, and witness the future of AI, all in one place.

MIDJOURNEY & DALLE2

Midjourney generates images from natural language descriptions, called prompts, similar to OpenAI's DALL-E and Stability AI's Stable Diffusion.

DATA ROBOT

DataRobot is an AI-driven automated machine learning platform. It streamlines the process of building and deploying machine learning models etc.

Program Curriculum

Git & Version Control

INTRODUCTION TO GIT AND VERSION CONTROL

Learn the fundamentals of Git, including initializing repositories, cloning, staging changes, committing, and pushing to remote repositories.

BRANCHING AND GITHUB COLLABORATION

Understand how to create branches, collaborate using GitHub, and manage code changes in a team environment.

WORKING WITH GIT IN VS CODE AND TEAM PROJECTS

Use Git within Visual Studio Code and apply your knowledge in a mini team project to practice real-world collaboration.

Introduction to Data Modeling and Warehousing

DIMENSIONAL MODELING CONCEPTS

Learn the fundamentals of dimensional modeling, including the Star and Snowflake schemas, and the distinction between fact and dimension tables.

DATA LAKE VS DATA WAREHOUSE

Understand the differences between Data Lakes and Data Warehouses, and when to use each for storing and processing large datasets.

DATA MODELING WITH DBT

Explore data modeling with dbt (data build tool) to efficiently build and manage your data pipelines and transformations.

Program Curriculum

Data Warehousing and ETL Automation

DATA WAREHOUSING SOLUTIONS

Explore popular cloud data warehouses such as Amazon Redshift, Snowflake, and BigQuery, and understand how they support large-scale data storage and analysis.

DATA PARTITIONING, CLUSTERING, AND OPTIMIZATION

Learn how to optimize data performance with techniques like partitioning, clustering, and various performance tuning strategies.

AUTOMATING ETL WITH DBT AND APACHE AIRFLOW

Build and automate an end-to-end ETL pipeline for a Retail Sales Analysis project, using dbt for data transformation and Apache Airflow for scheduling and orchestrating the pipeline.

Introduction to Big Data and Distributed Computing

HADOOP ECOSYSTEM OVERVIEW

Get introduced to key components of the Hadoop ecosystem, including HDFS, YARN, Hive, and Pig for distributed storage and processing.

SPARK BASICS: RDDS AND DATAFRAMES

Learn the fundamentals of Apache Spark, including SparkContext, RDDs (Resilient Distributed Datasets), and DataFrames for distributed data processing.

DATA TRANSFORMATIONS IN PYSPARK

Explore how to perform data transformations and manipulations using PySpark, leveraging its powerful capabilities for big data processing.

Program Curriculum

Advanced Spark for Data Processing

SPARK SQL AND ANALYTICAL FUNCTIONS

Use Spark SQL for powerful data analysis with filtering, joins, and window functions to handle large-scale structured data.

PERFORMANCE OPTIMIZATION IN SPARK

Learn key techniques like data caching, partitioning, and broadcasting to optimize Spark jobs for speed and efficiency.

HANDS-ON PROJECT: SERVER LOG ANALYSIS

Apply Spark skills in a practical project by analyzing server logs to extract insights and demonstrate end-to-end data processing.

Real-time Data Engineering

KAFKA ARCHITECTURE AND CORE COMPONENTS

Understand the building blocks of Kafka, including producers, consumers, brokers, and topics, for reliable message streaming.

KAFKA STREAMS AND CONNECTORS

Learn how to process real-time data using Kafka Streams and integrate systems using Kafka Connectors.

STREAM VS BATCH PROCESSING

Compare stream and batch processing models, and understand when to use each for different data engineering use cases.

Program Curriculum

Real-Time Data Processing with Spark and Kafka

WINDOWING AND STATEFUL STREAM PROCESSING

Understand sliding windows and manage stateful streams to handle complex event processing in real time.

INTEGRATING SPARK STREAMING WITH KAFKA

Learn how to connect Spark Streaming with Kafka to build scalable, real-time data pipelines.

REAL-TIME USE CASE AND MONITORING

Implement a real-time fraud detection system using transaction data, and monitor stream performance using Kafka Manager and Grafana.

Cloud Data Engineering with AWS and GCP

CORE CLOUD SERVICES: AWS & GCP

Get hands-on with key cloud services including AWS IAM, S3, EC2, Lambda and GCP BigQuery, Cloud Storage, and Cloud Functions.

SERVERLESS ETL AND DATA LAKE ANALYTICS

Build ETL pipelines using AWS Glue or GCP Dataflow, and run SQL queries on data lakes with Athena or BigQuery.

CLOUD-BASED ETL PROJECT

Develop an end-to-end ETL pipeline, moving data from AWS S3 to Redshift, and apply your skills in a real-world cloud scenario.

Program Curriculum

Electives for Specialized Data Engineering Skills

DEVOPS FOR DATA ENGINEERING

Learn Docker and containerization with Docker Compose, set up CI/CD pipelines using GitHub Actions and Jenkins, and implement logging and monitoring with the ELK Stack.

DATA GOVERNANCE AND SECURITY

Understand encryption methods (symmetric & asymmetric), enforce data access controls like role-based and row-level security, and track data lineage and auditing for compliance.

ML PIPELINES FOR DATA ENGINEERS

Get started with MLflow for experiment tracking, explore feature stores and model registries, and deploy scalable ML pipelines as part of the data engineering workflow.

Capstone Projects: Real-World Data Engineering

ECOMMERCE ANALYTICS

Build a clickstream ingestion pipeline using Kafka and develop a recommendation engine with Spark.

FINANCE DATA ENGINEERING

Ingest and clean daily trade and transaction data, then architect a Delta Lake-based data lakehouse.

LOGISTICS AND PREDICTIVE ANALYTICS

Build a GPS-based vehicle tracking system and use streaming analytics for predictive maintenance.

Domain Specialization



BANKING FINANCE & INSURANCE

You will work on Projects such as **Churn Analysis, Risk-Reward Analysis, Stock Market Analysis and Fraud Analysis** to understand core BFSI Concepts.



ECOMMERCE AND MARKETING

You will learn **Customer lifetime Analysis, Ad Campaign Analysis, Market Basket Analysis & Dynamic Pricing** from this segment.



HEALTH CARE AND PHARMACY

You will learn **Payer and Provider Analytics, Analytics used in Pharmaceutical Industry** to understand core Healthcare Concepts.



HUMAN RESOURCES & OPERATIONS

You will learn **Attrition Analysis, Promotion Analysis, Productivity Analysis & Resource Optimization** to understand core HR & Operations Concepts.

Capstone Projects with Data Engineering



ECOMMERCE

- * Clickstream Ingestion with Kafka
- * Building a Recommendation **Pipeline** with **Spark**



FINANCE

- * Ingesting and cleaning daily trades and transactions
- * Building a **Data Lakehouse** with Delta Lake



LOGISTICS

- * GPS-based Vehicle Tracking Data Pipeline
- * Predictive Maintenance with **streaming analytics**

Industry Projects & Case studies



Deliver Real-time Product Recommendations at Scale

Utilize scalable data pipelines and streaming technologies to process user activity logs in real time, enabling product recommendation systems to deliver timely and relevant suggestions.

Design and implement ETL workflows and real-time data ingestion frameworks to support ML-driven personalization, improving latency and data freshness for ecommerce platforms.

Skills: Apache Kafka, Apache Spark, Data Pipelines, ETL, AWS/GCP, SQL, NoSQL, Real-time Processing

Streamline Patient Health Records for Real-time Insights

Enable healthcare providers to access up-to-date patient data by building a unified, secure, and scalable data platform.

Integrate disparate health data sources (EHRs, labs, wearable devices) into a centralized data lake using batch and stream processing, enabling real-time monitoring and analytics for patient care and operational efficiency.

Skills: Apache NiFi, Spark Streaming, Hadoop, Data Lakes, HIPAA-compliant data handling, SQL, Parquet/Avro

Industry Projects & Case studies



Deliver Real-time Product Recommendations at Scale

Utilize scalable data pipelines and streaming technologies to process user activity logs in real time, enabling product recommendation systems to deliver timely and relevant suggestions.

Design and implement ETL workflows and real-time data ingestion frameworks to support ML-driven personalization, improving latency and data freshness for ecommerce platforms.

Skills: Apache Kafka, Apache Spark, Data Pipelines, ETL, AWS/GCP, SQL, NoSQL, Real-time Processing

Streamline Patient Health Records for Real-time Insights

Enable healthcare providers to access up-to-date patient data by building a unified, secure, and scalable data platform.

Integrate disparate health data sources (EHRs, labs, wearable devices) into a centralized data lake using batch and stream processing, enabling real-time monitoring and analytics for patient care and operational efficiency.

Skills: Apache NiFi, Spark Streaming, Hadoop, Data Lakes, HIPAA-compliant data handling, SQL, Parquet/Avro

Industry Projects & Case studies



Build a Scalable Financial Transactions Pipeline

Create robust ETL pipelines to handle high-volume financial data for fraud detection, reporting, and compliance.

Design ingestion and transformation workflows that validate, enrich, and store transaction data efficiently for downstream analytics and alerting systems in banking and fintech environments.

Skills: Apache Airflow, Kafka, Snowflake, SQL, Python, Data Quality Checks, Data Warehousing

Optimize Logistics Operations with Real-time Shipment Tracking

Develop a data infrastructure to collect, process, and store live GPS and sensor data from delivery fleets.

Use stream processing and cloud storage to feed real-time dashboards, route optimizations, and predictive maintenance models that improve efficiency in supply chain management.

Skills: Apache Flink, AWS Kinesis or GCP Pub/Sub, Redshift/BigQuery, Docker, Data APIs, Time-series Databases

Industry Projects & Case studies



Real-time Inventory Monitoring for Retail Chains

Track inventory across multiple warehouses and stores by building a centralized realtime data platform.

Ingest and process transactional and sensor data to prevent stockouts, forecast demand, and automate replenishment decisions.

Skills: Kafka, Delta Lake, Spark Structured Streaming, Azure Data Factory, SQL, BI Tools

Automated Data Warehouse for Marketing Analytics

Build a modern data warehouse to unify marketing campaign data from multiple platforms

Design ETL pipelines to collect and normalize data from Google Ads, Facebook, and CRMs for campaign performance tracking and ROI analysis.

Skills: Airbyte, dbt, BigQuery, SQL, Looker, Data Modeling, APIs

IoT Data Pipeline for Smart Agriculture

Process high-frequency sensor data from farms to monitor soil, weather, and irrigation patterns.

Develop edge-to-cloud streaming architecture that enables real-time dashboards and alerts for crop health and yield optimization.

Skills: MQTT, Apache Flink, InfluxDB, AWS IoT Core, Grafana, Time-series Analytics

Industry Projects & Case studies



Education Platform Usage Analytics

Build pipelines to track student activity and engagement across LMS platforms.

Capture and analyze learning patterns, content effectiveness, and course completion rates for data-driven curriculum improvements

Skills: Python, Apache Beam, GCP Dataflow, PostgreSQL, Cloud Storage, Tableau

Data Lakehouse for Customer 360 View

Create a lakehouse architecture to unify customer interactions across apps, support tickets, and purchases

Implement batch and streaming ETL to feed data into a single, queryable layer that supports segmentation, personalization, and churn analysis.

Skills: Databricks, Delta Lake, Spark, Azure Synapse, SQL, Event Hubs

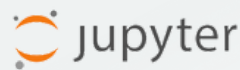
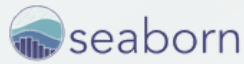
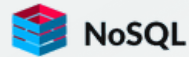
Compliance Data Archival System for Regulated Industries

Design an automated archival pipeline for long-term, secure, and searchable storage of sensitive data.

Build workflows to enforce data retention policies, encryption, and compliance logging for audits.

Skills: AWS S3 Glacier, Glue, Lambda, CloudTrail, Python, IAM, Encryption Standards

Libraries & Tools Covered



Program Outcomes



- **In-depth knowledge** of Data Analytics Concepts, Data Cleaning, Processing, ML, DL, and NLP algorithms.
- **Hands-on Experience** with advanced Data Science and Business Intelligence Tools and Techniques.
- **Ability to solve Real-world Data Science problems** and making data-driven decision making for optimized results.
- **Networking Opportunities** with fellow students, faculties, Domain Experts, Seasoned data Science Professionals.
- **Enhanced Career Opportunities**, as demand for Data Science professionals is very high across all the Industries.



Career Support



Get Ready with Domain Experts:

Learn from top Data Analytics Educators with domain expertise through a Live Online classroom. 24X7 Support from Teaching Assistants for Help in Assignments, Discussions, and Doubts Resolutions.

Industry Leaders for Professional Grooming:

Know what's Happening inside the industry, and What's really required to make it through by engaging with Data Scientists and accompany them with sessions on Hiring, Day-to-Day Life of a Data Scientist/Data Analyst etc.

Benefit of Networking with Professionals:

Grow and Develop your Career with the Help of similarly motivated and Highly energetic Peers. Get to know How our Alumni Started their Career in the Emerging Field of Data Science and Artificial Intelligence.

100 % Job Guarantee or Money Back:

Get Guaranteed Job Opportunities from Our Partners from Top MNCs and Startups such as American Express, Tokopedia, Infosys, TCS, PayTm, SpringWorks, MIQ, Tech Mahindra, Genpact, Amazon, Wipro etc.

Success Stories



“ The assignments and case studies are the ones that will really give you a mileage during the job interviews. Through Meritshot career support, I have secured an offer from an MNC. I have learnt quite a lot from peer discussions and project sessions. All-in-all the Meritshot program can really help you transition into data science field easily.

Sahil Khan,
Data Scientist at Kearney



“ The Domain Specialization and Personalized Projects really elevated my Portfolio on Medium, Kaggle, Github. I gained a lot from the special sessions provided for Placements. I have successfully transitioned my career as a Data Science Engineer with a good package. The Live Sessions from Industry Experts are great.

Mansi Pant,
Data Analyst at NTT Data



“ The Program is really structured so that you can learn each and every skill one by one in the easiest way. The mentors are really helpful for any kind of feedbacks related to assignments, projects, case studies etc. The personalization they provided me placements helped me to bag an amazing Offer.

Sobhamayee Sahu,
Data Analyst at Amazon

Success Stories



“ The case studies and tasks are what will actually help you stand out in job interviews. I was able to receive a decent offer from a good company because of Meritshot career support. Alumni Meets and project sessions have taught me a lot. Overall, the Meritshot programme can truly aid with your smooth transition into the profession of data science.

Prabhat Sinha

Data Science Mentor at Top Edtech



“ Highly recommended for individuals planning to transition their Non-IT careers into technical space. The curriculum, and especially the Notes are designed for absolute non-technical individuals or beginners, which can help you a lot to gain confidence to learn such advanced concepts and technologies.

Pratyush Yadav

Data Analyst at MUFG



“ This program is nicely organised so that you can pick up each ability the simplest conceivable way, one at a time. For any kind of guidance or help in assignments, tests, etc., the coaches are incredibly helpful. The Interviews opportunities they provided me enabled me to accept a fantastic offer.

Md. Mannan

Data Specialist at Johnsons Control

Program Certification



Hiring Partners



Program Details

Program Flow

- **600+ Live Training Hours** - Get Trained by seasoned Data Science and AI Professionals and Experts throughout the course for over 600+ Hours.
- **300+ Hiring Partners** - Get Access to the Huge spectrum of Hiring Partners Including both Startups and MNCs across the Country from Various Industries.
- **40+ Tools Covered** - Get Professionally Trained on more than 40+ Popular and Highly Demanded Tools and Libraries.

Program Fee

Flexible Financing Options(EMI) available starting @5500/-
0% Interest rates on EMIs.

Contact Support

For further details, contact us at: **+91 9354902983** or
Mail us at **support@meritshot.com**



M E R I T S H O T

COMPANY INFORMATION

Meritshot Zetta Edutech Pvt. Ltd.
3rd Floor, G-184, G Block, Sector 63,
Noida 201301, UP.