



Meritshot

Data Engineering with Generative & Agentic AI Specialisation

Become an AI-Powered Data Engineer with Meritshot's 360° Career Assistance & AI focused curriculum

Table of Contents

- 1 The Future Belongs to Data Engineers
- 2 Build the Skills That Top Companies Demand
- 3 Real stories, Real Impact
- 4 Where Can This Take You?
- 5 Industry-Vetted Curriculum
- 6 Hands-on Projects & Case Studies
- 7 Defining your outcomes
- 8 The Meritshot Edge
- 9 Meet Your Instructors
- 10 Sample Certificate
- 11 Your Next Steps



The Future Belongs to Data Engineers

Data is now the world's most valuable resource, but without the right systems, it's just noise. Companies across industries — finance, healthcare, e-commerce, logistics, and even space research — rely on Data Engineers to build the pipelines, platforms, and infrastructure that power Artificial Intelligence and real-time decision-making.

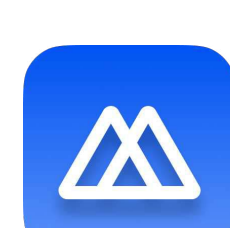
From Netflix recommendations to fraud detection in banks, and from self-driving cars to AI-driven healthcare diagnostics, none of these innovations can exist without a solid data engineering backbone. As organizations move towards AI-first strategies, demand for skilled Data Engineers who can design scalable, secure, and cloud-native data systems is growing faster than ever.

This program ensures that you're not just learning today's tools but preparing for the future of data engineering — with training in real-time streaming systems (Kafka, Kinesis), cloud-native pipelines (AWS, Azure, GCP), and AI-powered automation (Agentic AI for pipelines, GenAI for documentation and monitoring).

Why it matters:

- 97% of organizations are investing in Big Data and AI-driven strategies.
- Data Engineering is ranked among the fastest-growing tech roles globally.
- AI, IoT, and 5G are creating explosive demand for real-time, scalable data systems.

With Meritshot, you'll not only learn how to manage data but also how to engineer the future of AI-driven enterprises.



Build the Skills That Top Companies Demand

Top companies don't just need data scientists — they need data engineers who can make data usable, reliable, and actionable at scale. This program equips you with the end-to-end skill set that top product companies, consulting firms, and fast-growing startups demand.

You'll master:

Programming Foundations

Python, SQL, and data structures tailored for data engineering.

ETL Pipelines & Data Warehousing

Build robust, production-ready workflows with Apache Airflow, Redshift, BigQuery, Synapse, and ADF.

Big Data Frameworks

Work with Spark, Hadoop, and PySpark to process terabytes of structured & unstructured data.

Cloud Ecosystems

Gain hands-on expertise in AWS, GCP, and Azure for storage, compute, streaming, and deployment.

Streaming & Real-time Processing

Master Kafka, Kinesis, and Pub/Sub to build real-time pipelines for IoT and financial systems.

Advanced Data Engineering

Learn DevOps for Data, CI/CD for pipelines, and governance principles like GDPR, HIPAA compliance, and role-based access control.

AI-Integrated Workflows

Upskill with electives in Generative AI, Agentic AI, and Prompt Engineering — future-proofing your profile for AI-driven data roles.

With our industry-vetted curriculum, capstone projects, and AI electives, you'll build a portfolio that signals to employers that you're deal-ready, deployment-ready, and future-ready.



Where can this take you?

Launch your career in cyber security with real-world labs, hands-on projects, and 360° career support, bridging the gap between learning and landing your dream job.



Data Engineer

↗ High

Growth: Senior Data Engineer in 3–4 years

Description:

Build and maintain scalable data pipelines, transform raw data into usable formats, and ensure data accessibility for analytics teams.

Key Skills:

Data Pipeline Design

SQL & NoSQL Databases

Python/Scala for Data Processing

Data Warehousing



Big Data Engineer

↗ High

Growth: Big Data Architect in 4–5 years

Description:

Specialize in processing and managing large-scale datasets using distributed frameworks for analytics and business insights.

Key Skills:

Hadoop & Spark

Distributed Computing

Data Lake Architecture

Real-Time Data Processing



Cloud Engineer (Data Focused)

↗ High

Growth: Cloud Solutions Architect in 3–5 years

Description:

Design and manage cloud-based data infrastructure to support storage, processing, and security of enterprise data.

Key Skills:

AWS/GCP/Azure

Cloud Data Warehousing (BigQuery, Redshift, Snowflake)

Infrastructure as Code (Terraform)

Cloud Security & Monitoring



ETL Developer

↗ High

Growth: Data Integration Lead in 3–4 years

Description:

Design ETL pipelines to extract, transform, and load data efficiently across systems while ensuring data quality.

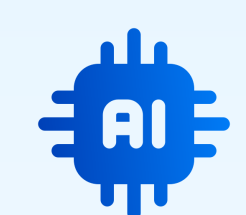
Key Skills:

ETL Tools (Informatica, Talend, Airflow)

SQL Optimization

Data Cleansing & Transformation

Workflow Automation



AI Engineer (Data-Oriented)

↗ High

Growth: Senior AI/ML Engineer in 3–4 years

Description:

Leverage big data and machine learning algorithms to design AI solutions that improve predictions, automation, and decision-making.

Key Skills:

Machine Learning & Deep Learning

Model Deployment on Cloud

Feature Engineering

Python (TensorFlow, PyTorch, Scikit-Learn)



DevOps Engineer (Data Infrastructure)

↗ High

Growth: Platform Engineer in 3–5 years

Description:

Automate, deploy, and monitor data applications with CI/CD pipelines, ensuring scalability and high availability.

Key Skills:

CI/CD (GitHub Actions, Jenkins)

Cloud DevOps Practices

Infrastructure Monitoring (Prometheus, Grafana)

Docker & Kubernetes



Solutions Architect (Data Systems)

↗ High

Growth: Enterprise Architect in 5–7 years

Description:

Design end-to-end enterprise data solutions, integrating storage, analytics, AI, and cloud services into scalable architectures.

Key Skills:

Enterprise Data Architecture

Cloud Solution Design

API & Microservices

Security & Compliance



Real Stories Real Impact

Your Blueprint to Become an Industry-Ready Data Engineer



Sonal Mehta | Excel Analyst → Data Engineer

"From Excel pivots to building HIPAA-compliant pipelines with SQL, PySpark & Airflow. Created real-time patient dashboards and automated schema governance. Reduced reporting latency from 8 hours to 40 minutes, driving a 120% salary hike."



Rohit Khanna | BI Developer → Cloud Data Engineer

"Shifted from BI reports to AWS pipelines (Kinesis, Glue, Redshift, Athena). Built streaming fraud detection pipelines that cut detection time from 2 hours to 8 minutes. Optimized infra to save ₹16 lakhs annually, securing my role as a cloud-first engineer."



Amit Saha | ETL Developer → Modern Data Engineer

"Upgraded from legacy ETL (Informatica) to Spark, Airflow & dbt. Migrated batch jobs into event-driven pipelines with Kafka and Delta Lake. Reduced job failures by 30% while building scalable pipelines for e-commerce data at production scale."

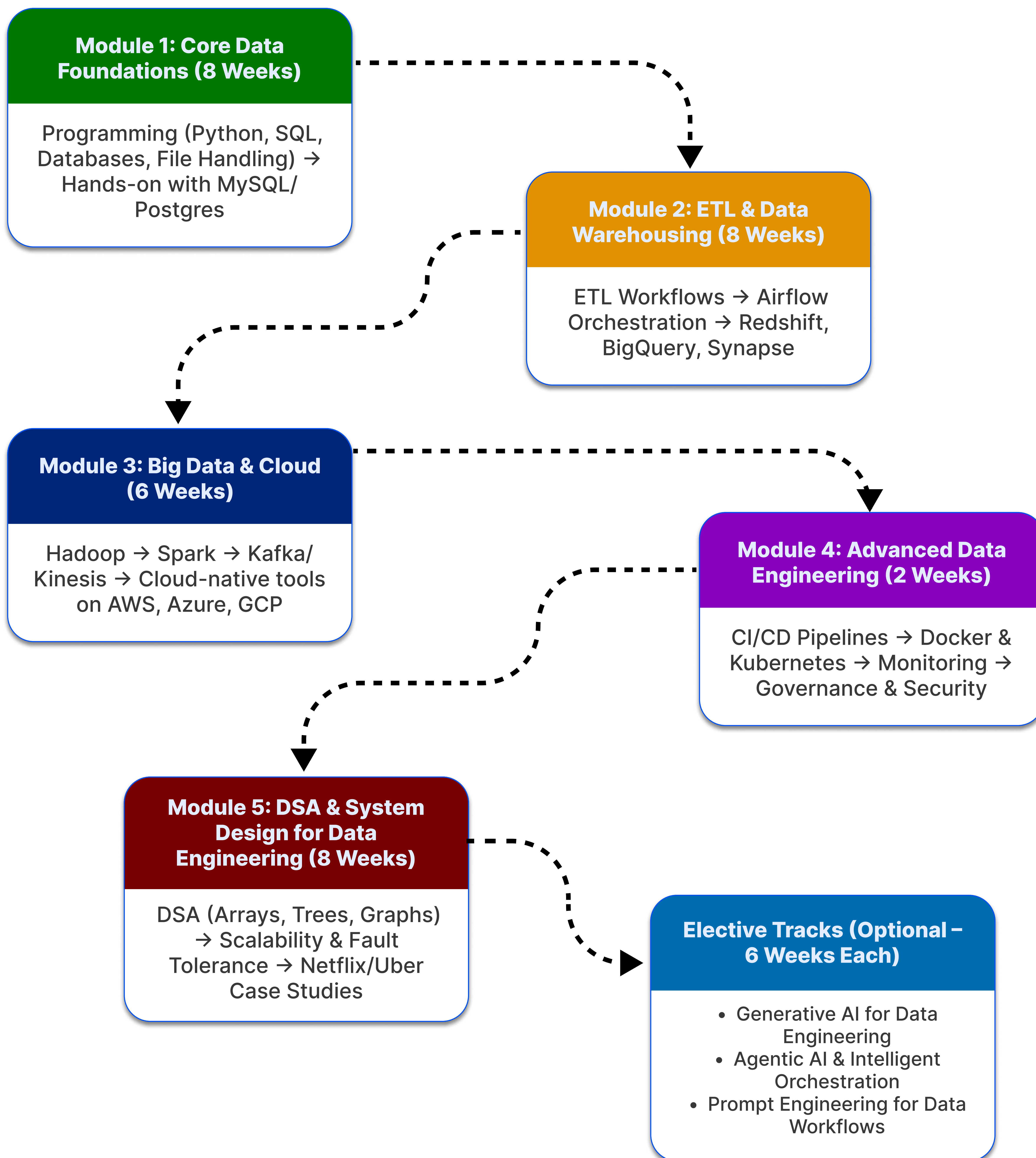


Nisha Rao | Support Engineer → DataOps Engineer

"From firefighting tickets to building resilient GCP data pipelines. Designed monitoring, lineage, and anomaly detection systems for campaign data feeds. Reduced escalation tickets by 62% and shifted to a proactive DataOps role."



Data Engineering Curriculum Roadmap



Outcome:

Land your dream role as Data Engineer, Big Data Engineer, Cloud Data Engineer, or AI-Powered Data Engineer

Module 1

Core Data Foundations (6 Weeks)

Overview:

Crack tough SQL/Python rounds, handle large datasets, write clean & optimized code.

What You'll Learn:

Python

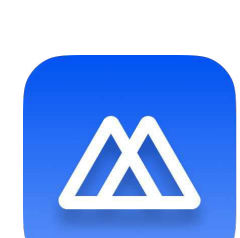
- Flowcharts, Data Types, Operations
- Conditional Statements, Loops, and Strings
- Inbuilt Data structures- List, Tuples, Dictionary, Set, Matrix Algebra, Number system
- Advanced OOPs, Exception Handling, Functional Programming
- Time & Space Complexity in Python (Big-O analysis)
- Libraries: Pandas, NumPy, Matplotlib, Seaborn, and PySpark Basics
- Debugging, Unit Testing, Code Optimization

SQL + NoSQL

- Introduction to Database and BigQuery setup
- Extracting data using SQL
- Functions, Filtering, and Subqueries
- Advanced Joins, Window Functions, and Recursive Queries
- Query Optimization, Indexing, Partitioning
- Transactions & Isolation Levels (ACID vs BASE)
- NoSQL: MongoDB, Cassandra use cases at scale

Learning Outcomes:

- ✓ Confidently manipulate data using Python & SQL.
- ✓ Build structured queries for real-world datasets.
- ✓ Apply core programming and database concepts to solve business problems.



Module 2

ETL & Data Warehousing (8 Weeks)

Overview:

Build enterprise-scale ETL and data warehouse systems like those at Netflix/Amazon.

What You'll Learn

ETL Pipelines

- Batch & Real-Time ETL Concepts (Extract, Transform, Load)
- Apache Airflow (Workflow Orchestration)
- Apache NiFi, AWS Glue, dbt (modern transformations)
- Data Quality Frameworks (Great Expectations, Deequ)

Data Warehousing

- OLTP vs OLAP, Dimensional Modeling (Kimball/Inmon)
- Star & Snowflake Schema Design
- Cloud Warehouses: BigQuery, Redshift, Snowflake
- Query Optimization & Partitioning Strategies

Learning Outcomes:

- ✓ Design and deploy scalable ETL pipelines.
- ✓ Implement data warehouses with industry-standard schemas.
- ✓ Integrate multiple data sources for business-ready reporting.



Module 3

Big Data & Cloud (7 Weeks)

Overview:

FAANG-level distributed systems & cloud mastery.

What You'll Learn

Hadoop

- HDFS(Hadoop Distributed File System)
- YARN (Yet Another Resource Negotiator)
- Map Reduce

Pyspark

- Spark core concepts: RDDs, DataFrames, and Spark SQL
- Parallel processing and distributed computing with Spark
- Spark for data transformation, aggregation, and analytics
- Powerful data processing with PySpark for scalable analytics.

Distributed Databases

- CAP Theorem, consistency, availability, partition tolerance
- Cassandra, HBase: Columnar data stores for large-scale datasets

Data Streaming

- Apache Kafka (producers, consumers, partitions, offsets)
- Kafka Streams, Spark Streaming, Flink basics



Module 3**Big Data & Cloud****AWS**

- AWS EMR
- On-Prem vs Cloud
- HDFS vs S3
- What is S3
- EC2
- Elastic IP
- AWS Storage, Networking
- S3 and EBS
- AWS Glue
- AWS Redshift

AZURE

- Azure Data Factory
- Azure Databricks
- Azure Synaps analytics
- Azure Blob Storage

GCP

- Bigquery
- Pub/sub

Linux

- Introduction to Linux
- File system navigation
- Process Management
- Shell Scripting
- System configuration and advanced Linux commands

Learning Outcomes:

- ✓ Process massive datasets with Hadoop & Spark.
- ✓ Build cloud-native ETL and analytics pipelines.
- ✓ Apply real-time streaming for clickstream & IoT data.



Module 4

Advanced Data Ops (4 Weeks)

Overview:

Handle scale, resilience, automation, and security like FAANG engineers.

What You'll Learn

Advanced Data Engineering

- High Availability & Fault-Tolerant Architectures
- Data Lake vs Lakehouse (Delta Lake, Apache Iceberg)
- Scalable pipeline design patterns

DevOps for Data Engineering

- CI/CD for Data Pipelines (GitLab/Jenkins)
- Docker & Kubernetes for data workloads
- Infra as Code (Terraform basics for data infra)

Data Security

- Encryption (AES, KMS)
- Authentication & Authorization (IAM, RBAC)
- GDPR, HIPAA, and compliance basics

Learning Outcomes:

- ✓ Automate and monitor large-scale pipelines.
- ✓ Ensure compliance & governance in data projects.
- ✓ Secure enterprise data systems end-to-end.



Module 5**DSA & System Design (8 Weeks)****Overview:**

Ace FAANG-style coding + design interviews.

What You'll Learn:**DSA**

- Arrays, Strings, HashMaps
- Linked Lists, Stacks, Queues
- Trees, Tries, Graphs
- Dynamic Programming
- Sorting & Searching (Binary Search Variants, Quick/Merge Sort)

System Design

- Designing Scalable Data Platforms
- Event-Driven Architecture, Messaging Systems
- Sharding, Replication, Consistency Models (CAP theorem)

Learning Outcomes:

- ✓ Automate and monitor large-scale pipelines.
- ✓ Ensure compliance & governance in data projects.
- ✓ Secure enterprise data systems end-to-end.



Elective Tracks (Optional, 6 Weeks Each)

Overview:

Specialize in cutting-edge AI-driven data engineering by choosing electives aligned with your career goals.

Tracks Available:

Generative AI for Data Engineering

Automating ETL, AI-driven documentation, LLMs in pipelines

Agentic AI (AI-Driven Workflows)

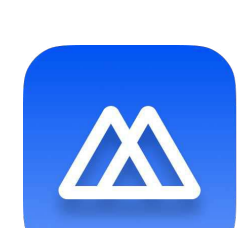
Multi-agent systems, intelligent orchestration, AI agents for monitoring

Prompt Engineering for Data Engineers

Automating SQL/NoSQL queries, integrating LLM APIs

Learning Outcomes:

- ✓ Apply AI to optimize and monitor pipelines.
- ✓ Automate data engineering tasks with AI agents.
- ✓ Gain future-ready skills in AI-powered data workflows.



Industry Projects & Case Studies

Data Engineering Program

Work on projects designed in collaboration with real-world industry use cases. These flagship projects build your portfolio, showcase your expertise, and prepare you for high-impact data engineering roles.



Supply Chain Optimization for Walmart

Business Challenge:

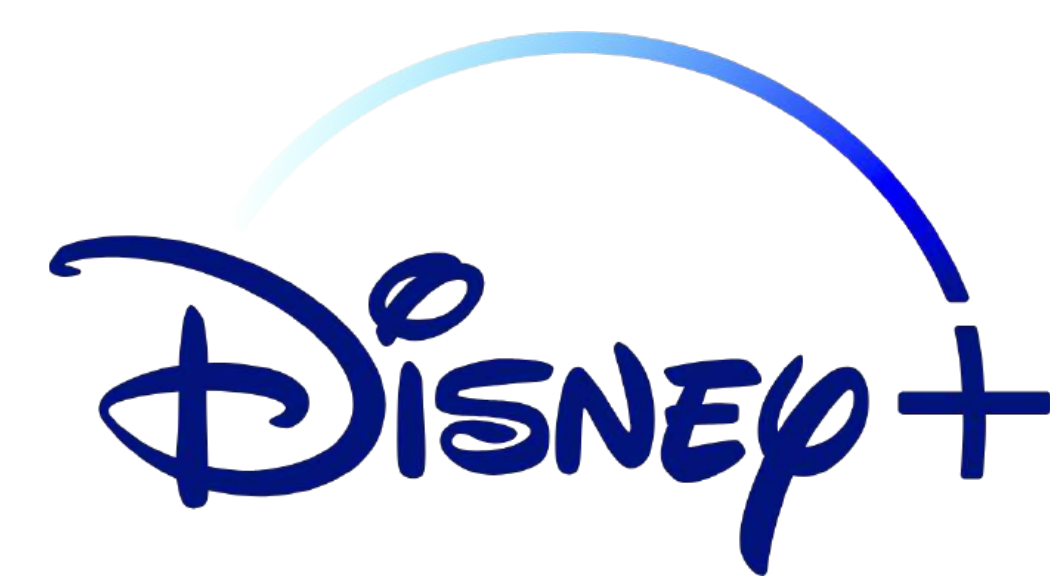
Walmart faces challenges in aligning supplier deliveries, warehouse inventories, and retail demand, leading to frequent stock-outs and inefficiencies.

Your Task:

Design a scalable data pipeline that integrates supplier, warehouse, and retail data into a unified warehouse. Use this to forecast demand, optimize inventory, and reduce stock-outs.

Tech Stack:

Apache NiFi · Apache Spark · AWS Redshift · Tableau



Streaming Content Insights for Disney+

Business Challenge:

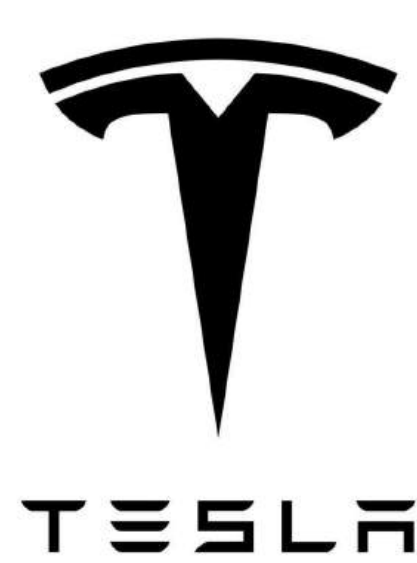
Disney+ needs to analyze billions of real-time user events to identify trending shows and improve recommendation engines.

Your Task:

Build a real-time data pipeline to capture and process user activity logs, analyze peak watch hours, and provide content insights for personalization.

Tech Stack:

Apache Kafka · Spark Structured Streaming · Google BigQuery · Looker Studio



Energy Consumption Analytics for Tesla

Business Challenge:

Tesla charging stations produce massive IoT sensor data that must be monitored in real-time to optimize energy grid usage.

Your Task:

Develop a data pipeline that collects and analyzes IoT sensor streams to forecast demand spikes and improve grid efficiency for sustainable energy management.

Tech Stack:

Azure Event Hubs · Azure Databricks · Power BI



Flight Delay Prediction for Delta Airlines

Business Challenge:

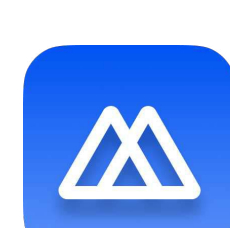
Flight delays impact both passenger experience and airline profitability. Delta needs real-time insights into delays caused by weather and operational issues.

Your Task:

Create a real-time pipeline that processes flight logs and weather feeds to predict delays and send proactive alerts to airline operations teams.

Tech Stack:

Apache Kafka · Apache Flink · Azure Synapse Analytics · Python ML Libraries



Industry Projects & Case Studies

Data Engineering Program

Work on projects designed in collaboration with real-world industry use cases. These flagship projects build your portfolio, showcase your expertise, and prepare you for high-impact data engineering roles.



Personalized Marketing Analytics for Starbucks

Business Challenge:

Starbucks generates millions of daily transactions across outlets but struggles to leverage this data for personalized promotions.

Your Task:

Build a big data solution that processes large-scale transactions, segments customers, and generates personalized offers and recommendations.

Tech Stack:

Hadoop · Spark MLlib · AWS S3 · AWS Redshift · Tableau



Fraud Detection for Mastercard

Business Challenge:

Mastercard needs to detect and block fraudulent transactions instantly to minimize losses.

Your Task:

Develop a fraud detection pipeline that processes streaming credit card transactions in real time, flags anomalies, and updates fraud detection dashboards.

Tech Stack:

Apache Kafka · Apache Flink · GCP BigQuery · Python



Smart City Traffic Management for Singapore Govt.

Business Challenge:

Singapore struggles with urban traffic congestion and needs a smart solution for live route optimization.

Your Task:

Design a pipeline that processes live GPS feeds and traffic data to build real-time congestion heatmaps and recommend optimized routes for public transport.

Tech Stack:

Kafka · Azure IoT Hub · Spark Streaming · Azure Data Lake · Tableau



J.P.Morgan

Real-Time Stock Market Analytics for JP Morgan

Business Challenge:

JP Morgan traders need real-time insights on volatility, liquidity, and trading opportunities in high-frequency markets.

Your Task:

Build a high-performance analytics system to process market tick data and deliver real-time dashboards with risk indicators and trade recommendations.

Tech Stack:

Apache Kafka · Apache Flink · AWS Redshift · Python · Power BI



Defining Your Outcomes

At Meritshot, our Data Engineering Program is not just about learning tools — it's about transforming you into an industry-ready professional who can design, build, and manage enterprise-grade data systems. By the end of this program, you'll have the skills, confidence, and portfolio to thrive in top product-based companies and global enterprises.

1. Data Engineering Specialist

Design and maintain scalable pipelines that move data seamlessly across systems. Build end-to-end ETL workflows and ensure data quality for enterprise reporting.

2. Cloud & Big Data Engineer

Deploy and manage big data frameworks (Hadoop, Spark) and cloud-native tools (AWS, Azure, GCP) to process petabyte-scale datasets in real time.

3. Real-Time Streaming Expert

Leverage Apache Kafka, Flink, and Kinesis to process live data streams for use cases like fraud detection, IoT monitoring, and stock market analytics.

4. Data Platform Architect

Apply system design and DevOps practices to architect high-performance, fault-tolerant data platforms inspired by Netflix, Uber, and Airbnb.

5. AI-Driven Data Engineer

Integrate Generative AI and Agentic AI into data workflows, automate ETL, query generation, and intelligent monitoring with AI-powered agents.

6. Industry-Certified Professional

Graduate with the official Meritshot Program Completion Certificate, validated by recruiters and recognized globally, showcasing your data engineering expertise.



The Meritshot Edge

Our Data Engineering Program goes beyond theory. It's designed to give you a career-defining advantage through practical learning, AI-powered electives, and unmatched placement support.

Beginner to Pro Tracks

Whether you're new to data or already in tech, our structured roadmap takes you from Python and SQL foundations → to advanced data engineering → to AI-powered pipelines.

Hands-On Projects with Real Data

Work on industry-scale projects across finance, e-commerce, IoT, and cloud. From building fraud detection systems to real-time stock market pipelines, every project is designed to mirror real-world challenges.

Cloud & Big Data Mastery

Gain multi-cloud expertise across AWS, Azure, and GCP, and hands-on experience with Hadoop, Spark, Kafka, and Flink. Skills directly aligned with what top employers demand.

AI-Driven Specialisations

Stay ahead of the curve with electives in Generative AI, Agentic AI, and Prompt Engineering for Data Engineers — making you future-ready for AI-driven data workflows.

Placement-First Approach

400+ recruiter network, alumni referrals, AI-powered resume builder, LinkedIn branding, and advanced mock interviews — all designed to help you land high-paying roles in top product-based companies.

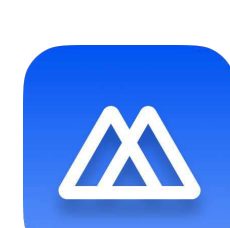
Small Batches, 1:1 Mentorship

Learn in small cohorts with dedicated mentors who give personalised feedback, career guidance, and doubt resolution.

Globally Recognized Certification

Earn the official Meritshot Program Completion Certificate, proof of your expertise in Data Engineering & Cloud Systems, trusted by employers worldwide.

With The Meritshot Edge, you don't just learn Data Engineering — you graduate as a pipeline-ready engineer prepared for roles like Big Data Engineer, Cloud Engineer, Real-Time Streaming Specialist, and Data Platform Architect.



Learn from the Industry's Finest Engineers

Learn practical skills and real-world systems from senior data engineers and cloud architects at top product companies



Chintada Abhilash

7+ years experience

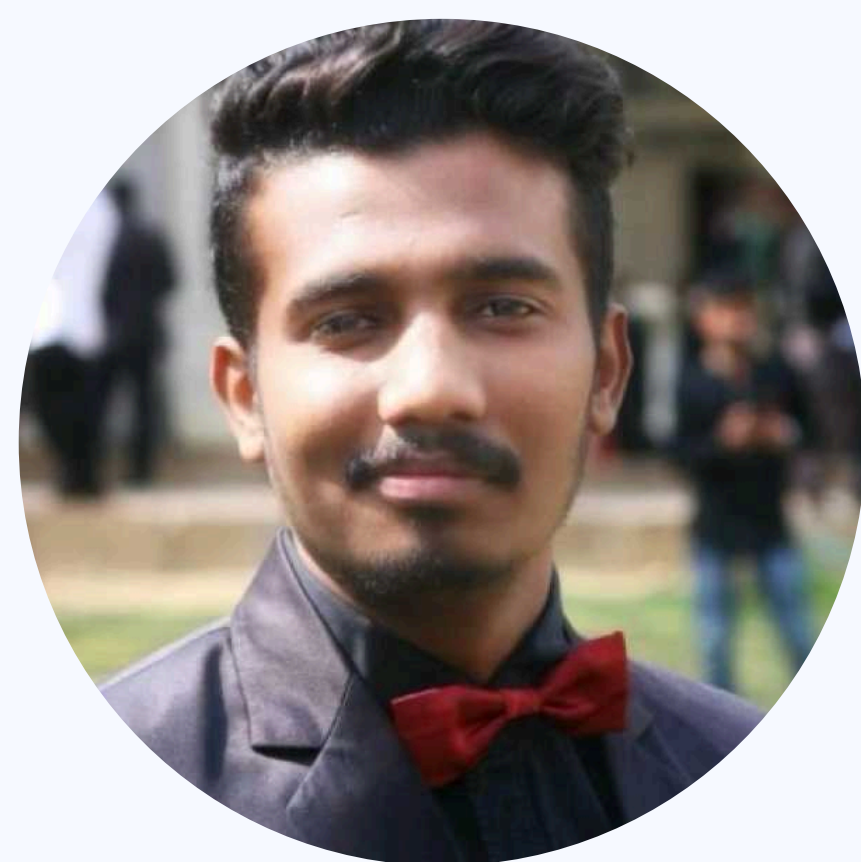
Data Science leader with 7+ years of experience in AI & ML, driving impactful solutions from predictive analytics to automation. Skilled in building scalable models and advanced algorithms, he has led teams to deliver data-driven business outcomes. Abhilash is passionate about turning data into actionable insights and mentoring the next generation of innovators.



Heena Arora

3+ years experience

Data Scientist at PwC with 3+ years of experience in predictive modeling, advanced analytics, and large-scale data solutions. Previously at Amazon, she gained expertise in machine learning and process optimization. Heena specializes in transforming raw data into business insights that drive strategic impact.



Saurabh Daund

5+ years experience

AI & Data Science professional with 5+ years of expertise in NLP, Generative AI, LLMs, and intelligent system design. He has built scalable, AI-driven products that enhance business performance and user experience. Saurabh blends technical depth with practical problem-solving to deliver innovative AI solutions.



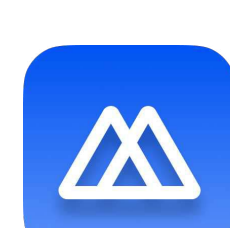
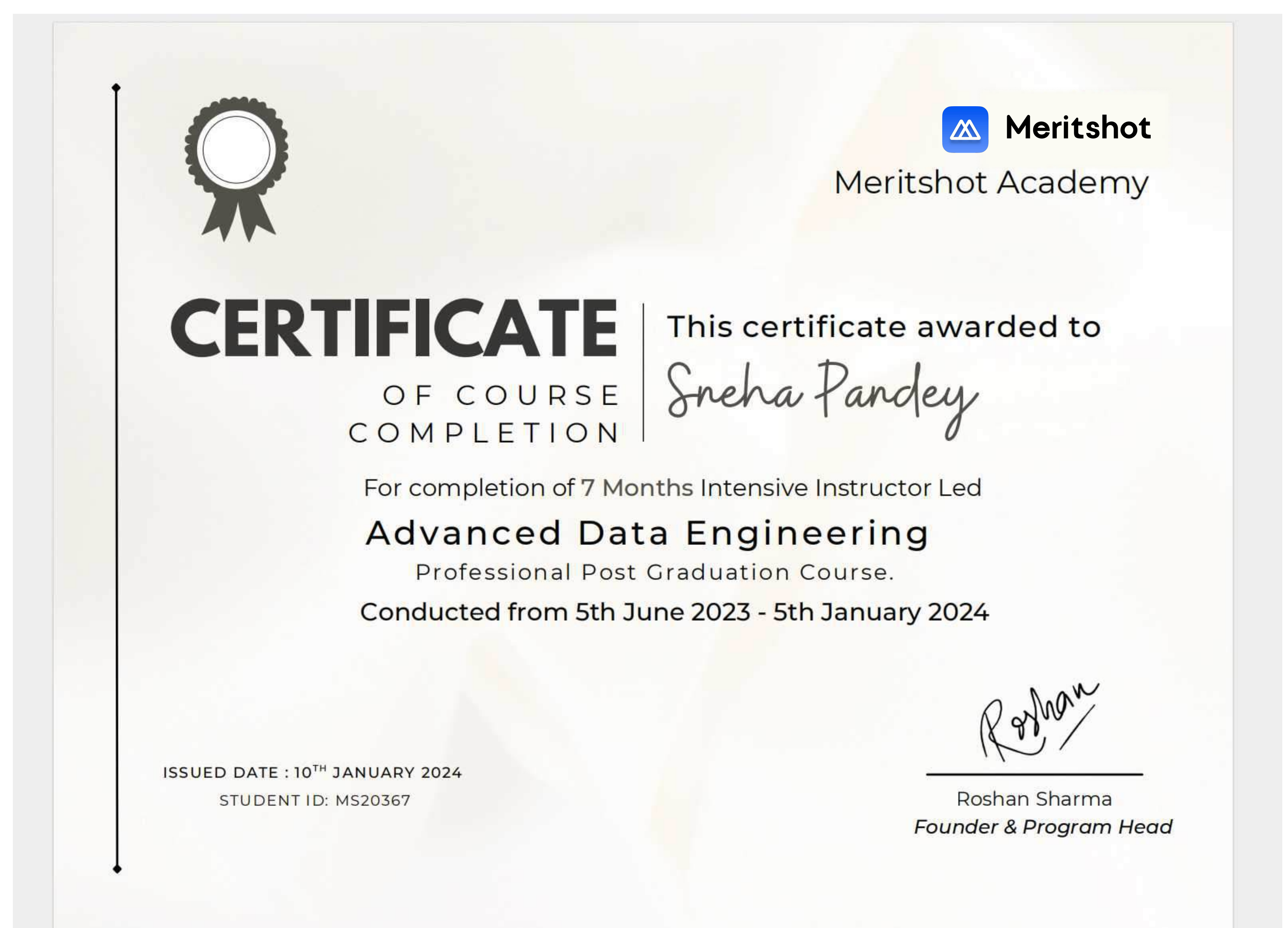
Saadh Khan

8+ years experience

With 8+ years in AI & ML, Saadh has led end-to-end projects across industries, from predictive modeling to scalable AI deployment. He brings expertise in advanced analytics and data-informed decision-making. Passionate about innovation, he creates AI solutions that deliver measurable business growth.

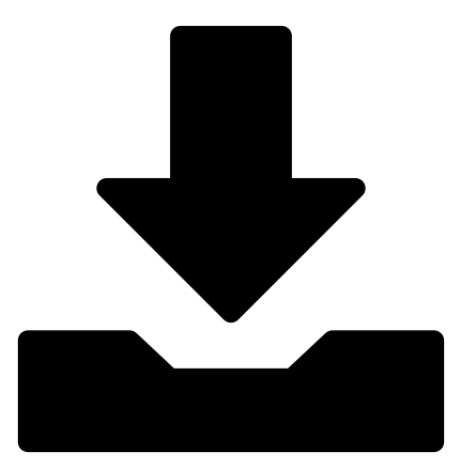


Certificate



How to Enroll?

Your Next Steps



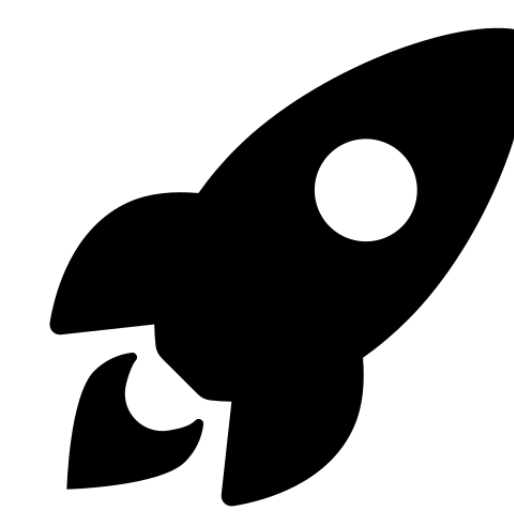
Download Brochure

Get complete curriculum details and program information



Get in Touch with us

Speak with our experts to explore specialisations, career paths, and program fit



Start Your Journey

Join the next batch and make a promise to transition your career.

Start Your Journey with Meritshot

1. Hassle-Free Registration

Sign up on meritshot.com or connect with our counsellor to begin your journey.

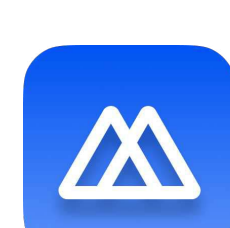
2. Choose Your Track

Pick the batch that suits you: Beginner or Advanced, based on your current skill level.

3. Batch Allotment & Kickstart

Get onboarded into your chosen track with flexible payment options and start learning.

Enroll Now →





Meritshot

THANK YOU

Need more insights?

Our counsellors will guide you on enrollment, specialisations, and career opportunities.

GET IN TOUCH →

