PREPZEE
Prepare For Future

DATA ENGINEER

JOB ORIENTED PROGRAM

Become a **Data Engineer** with Top Trending Tools and Land High Paying Jobs at **Top Companies**

Azure

Fabric

Databricks

Snowflake

dbt

Applied AI

+1 415-481-4467

+91-77930 68396

Index

About Prepzee

Preparing Learners for

Modern Data Engineering Careers.

We help professionals move beyond tools and theory—gain real engineering skills, build production-ready projects, and become job-ready data engineers.

4.8/5
Google Rating

1000+
Career Transformations

100K+
Learning Hours Delivered

Production-Ready Learning

Mentors from Brands like

We are Featured In

Featured In

WED, MAY 24, 2023 | UPDATED 16:48 IST

ANI
South Asia's Leading Multimedia News Agency

National | Business | World | Sports | Lifestyle | Entertainment | Health | Science

EdTech Company Prepzee to train 1,00,000 IT aspirants globally

ANI | Updated: Dec 01, 2022 18:02 IST

New Delhi [India], December 1 (ANI/PNN): As 2022 nears its end, Prepzee Learning Solutions Pvt Ltd takes the initiative to train 1,00,000 IT aspirants around the world over the next few years. The EdTech company intends to provide the global IT sector with skilled, competent, and passionate IT professionals who help businesses worldwide drive digitization.

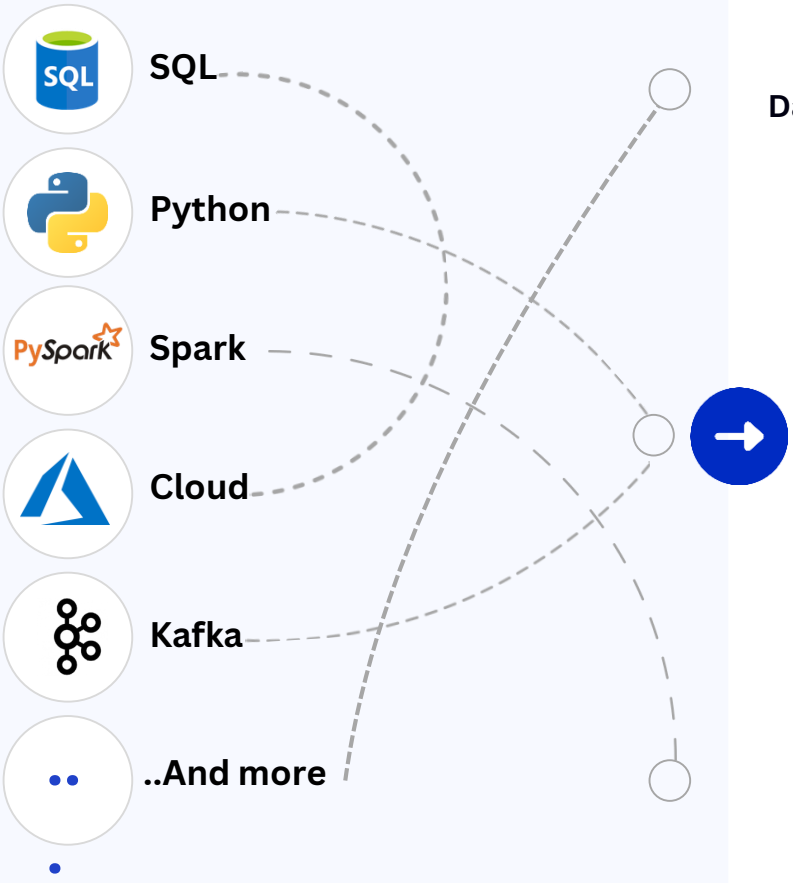
Two years of a global pandemic have made businesses and professionals across the world realize the importance of embracing digitization. The last two years have witnessed alarming spikes in the adoption of online collaborative platforms, virtual workspaces, cloud-based applications, and several other software systems that facilitate efficient, flexible, and scalable operations. Irrespective of their operational scales, businesses belonging to almost every sector are driving digitization to stay in tune with the current environment.

Click to read the whole article

The Gap Between Learning & Getting Hired

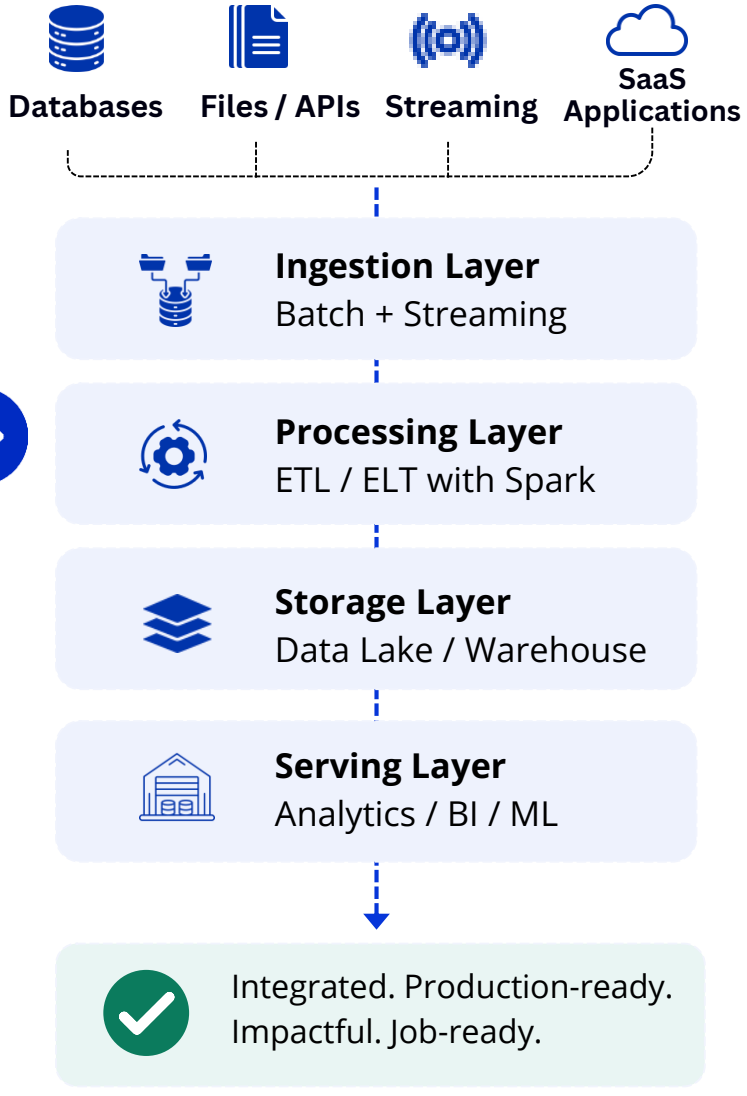
You learn the tools.
The industry expects **production-ready** engineers.

WHAT MOST LEARNERS DO
(DISCONNECTED LEARNING)



No direction.
No integration.
No real-world impact.

WHAT INDUSTRY EXPECTS
(JOB-READY DATA ENGINEER)



Integrated. Production-ready.
Impactful. Job-ready.

Real People.
Real Transitions.

Professionals from diverse backgrounds transformed into top data engineering roles.

Kishan Yadav

Data Engineer @ Microsoft

Microsoft

5 MONTHS Journey

2.6x Salary Hike

Mayank Sharma

Data Engineer @ Databricks

databricks

4.5 MONTHS Journey

3x Salary Hike

The right skills. The right guidance.
The Right Transformation.

Who Is This Program For?

Designed for professionals looking to transition into **modern data engineering** roles.

Data Analysts

Move beyond dashboards into scalable data systems.

ETL Professionals

Transition from legacy ETL tools to modern cloud pipelines.

SQL / PL SQL Developers

Build production-ready cloud data engineering skills.

Database Administrators (DBA)

Work with modern data platforms & lakehouse architectures.

Software Engineers

Expand your skills into big data, streaming & cloud engineering.

Career Transitioners

Start your journey from support, testing, or non-DE roles into modern data engineering.

Where Can This Program Take You?

Build in-demand data engineering skills and step into high-impact career roles.

Data Engineer

Build scalable batch & streaming data pipelines and transform data into business value.

Azure Data Engineer

Design and develop cloud-native data solutions using the powerful Azure ecosystem.

Databricks Engineer

Process, analyze and optimize large-scale data using Databricks, Spark and Delta Lake.

Snowflake Data Engineer

Build modern data warehouse and analytics solutions on the Snowflake platform.

Data Platform Engineer

Design enterprise-grade lakehouse architectures and robust data platforms.

Big Data Engineer

Work with distributed systems, orchestration tools and real-time data technologies.

From data pipelines to cloud platforms — build skills that **power the world's data**.

Program Features

Everything you need to learn, build, and advance your **Data Engineering** career.

100+

Hours of Live Training

Instructor-led sessions covering concepts to real-world systems.

80+

Hours Hands-on & Exercises

Practice with industry tools, datasets & engineering tasks.

8+

Projects & Case Studies

End-to-end real-world projects to build your portfolio.

24*7

Technical Support

Get expert help anytime you need it.

Top 1%

Industry Experts

Learn from data engineers working at top companies.

Lifetime

Live Training Access

Access all live sessions, updates & new content for lifetime.

PREPZEE
Prepare For Future

Roadmap to Become a Modern Data Engineer

A step-by-step learning path to build production-ready data engineering skills.

FOUNDATIONS

- Python
- Data Engineering Fundamentals

CLOUD & STORAGE

- Azure
- ADLS Gen2
- Fabric

DATA PROCESSING

- Databricks
- PySpark
- Snowflake

STREAMING & ORCHESTRATION

- Kafka
- Airflow
- Streaming Pipelines

PRODUCTION SYSTEMS

- Lakehouse Architecture
- Medallion Architecture
- Performance Optimization

DATA ENGINEERING FOR AI

- AI Pipelines
- Analytics & BI
- Applied AI

TOOLS COVERED

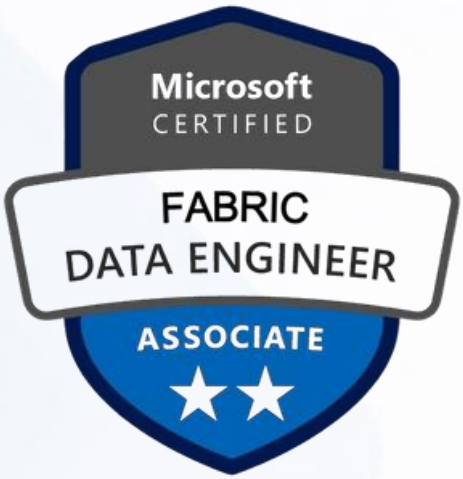
Master the industry-leading tools and platforms to build, orchestrate and scale modern data solutions.

- Azure
- Databricks
- Snowflake
- Fabric
- Airflow
- Kafka
- DBT
- AI
- Cortex AI
- Python
- Medallion Architecture
- Delta Lake
- Azure Data Factory
- ADLS Gen2
- Azure Key Vault

Industry-relevant tools. Real-world impact.
Build scalable, secure and reliable data engineering solutions.

Clear Industry
Recognised Certification

Real-world projects.
End-to-end engineering experience.



Industry-relevant tools. Real-world impact.
Build scalable, secure and reliable data engineering solutions.

UNLOCK BONUSES WORTH $500

Get access to premium courses absolutely **FREE** with our program.

BONUS 1 FREE COURSE

AWS Solutions Architect Associate Certification Course

- Comprehensive coverage of AWS services
- Hands-on labs and real-world scenarios
- Exam preparation and practice tests
- Learn from AWS certified experts

BONUS 2 FREE COURSE

- SQL basics to advanced concepts
- Query writing & optimization
- Database design fundamentals
- Hands-on practice exercises

BONUS 2 FREE COURSE

- CI/CD and automation
- Docker, Kubernetes & more
- Infrastructure as Code
- Real-world DevOps projects

Premium learning resources
to accelerate your career growth.

Portfolio-Worthy Data Projects.

Real-world projects.
End-to-end engineering experience.

5 INDUSTRY-FOCUSED PROJECTS

01

Azure Platform Architecture

Insurance Domain

02

Real-Time Data Engineering

Event-Driven Architecture

03

Financial Data Platform & Risk Intelligence

FinTech Domain

04

Healthcare Data Platform & AI Analytics

Healthcare Domain

05

End-to-End Retail Data Platform

Production Grade

Project 1

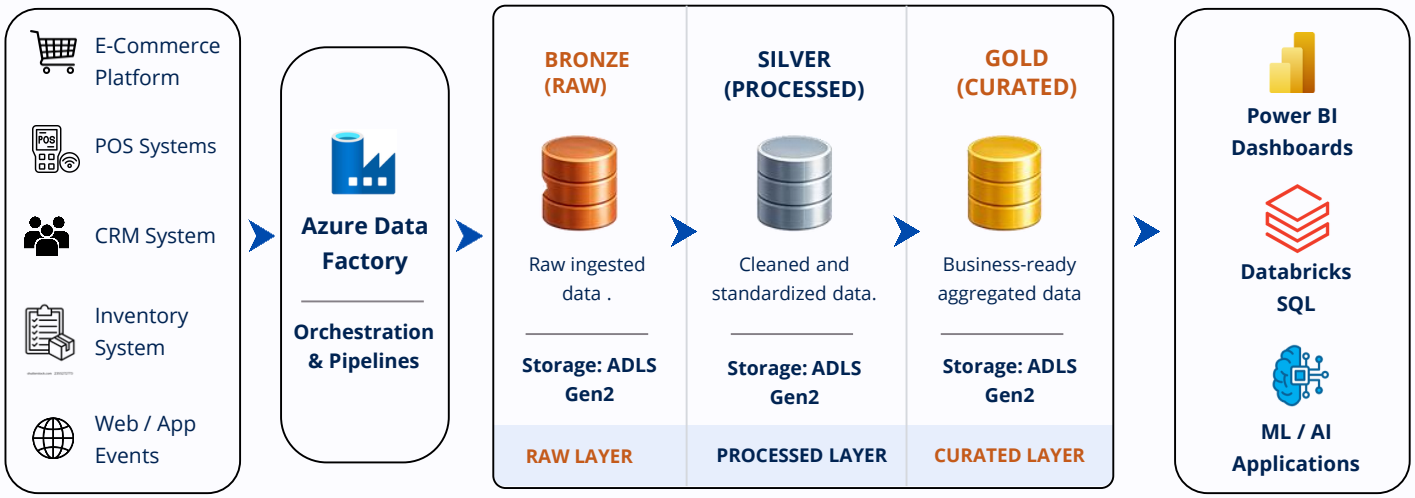
Enterprise Retail Data Platform

(Medallion Architecture)

PROJECT OVERVIEW

Build a cloud-native retail data platform on Microsoft Azure using Medallion Architecture (Bronze → Silver → Gold). Ingest data from multiple sources, process and transform it at scale, and deliver trusted data for analytics, dashboards, and machine learning.

ARCHITECTURE OVERVIEW



TECH STACK

BUSINESS IMPACT

- Single source of truth for all retail data
- Faster & accurate decision making
- 360° view of customers & products

KEY FEATURES

- Medallion architecture for high data quality
- Scalable ingestion from multiple systems
- Unified analytics for sales, customers & products

Project 2

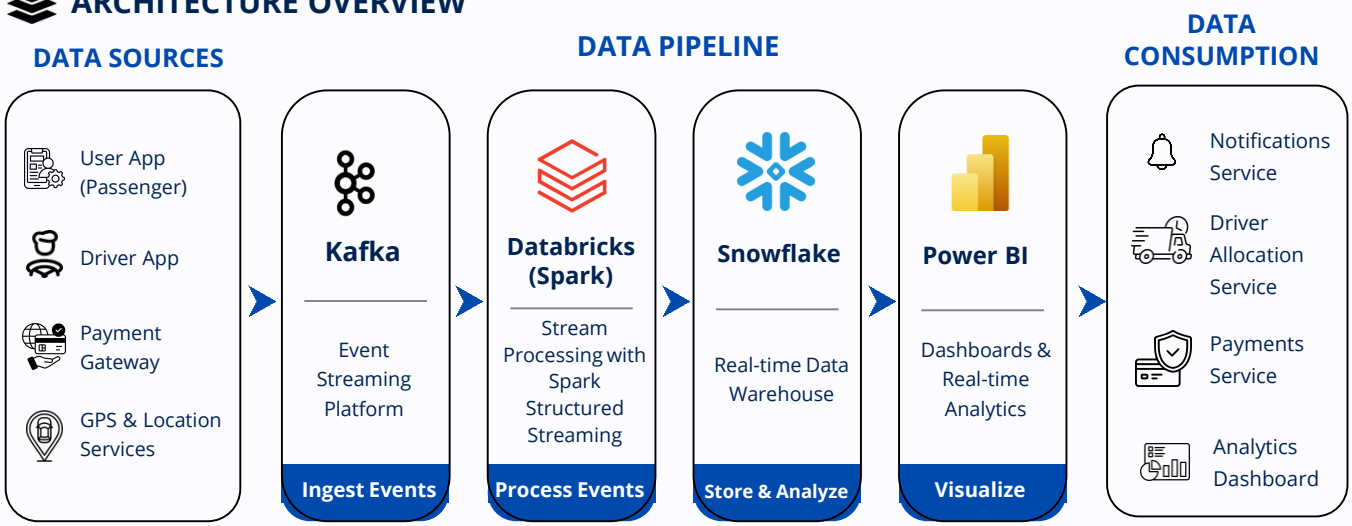
Real-Time Data Engineering System

(Event-Driven Architecture)

PROJECT OVERVIEW

Build a real-time data engineering system inspired by Uber using Apache Kafka, where a single booking triggers driver allocation, payments, notifications, and analytics instantly.

ARCHITECTURE OVERVIEW



TECH STACK

kafka → Databricks → Azure → Snowflake → Power BI

BUSINESS IMPACT

- **Process millions of events in real time and deliver instant insights.**

KEY FEATURES

- **Event-driven architecture**
- **Real-time stream processing**
- **Low-latency data pipeline**
- **Scalable & fault-tolerant**

Project 3

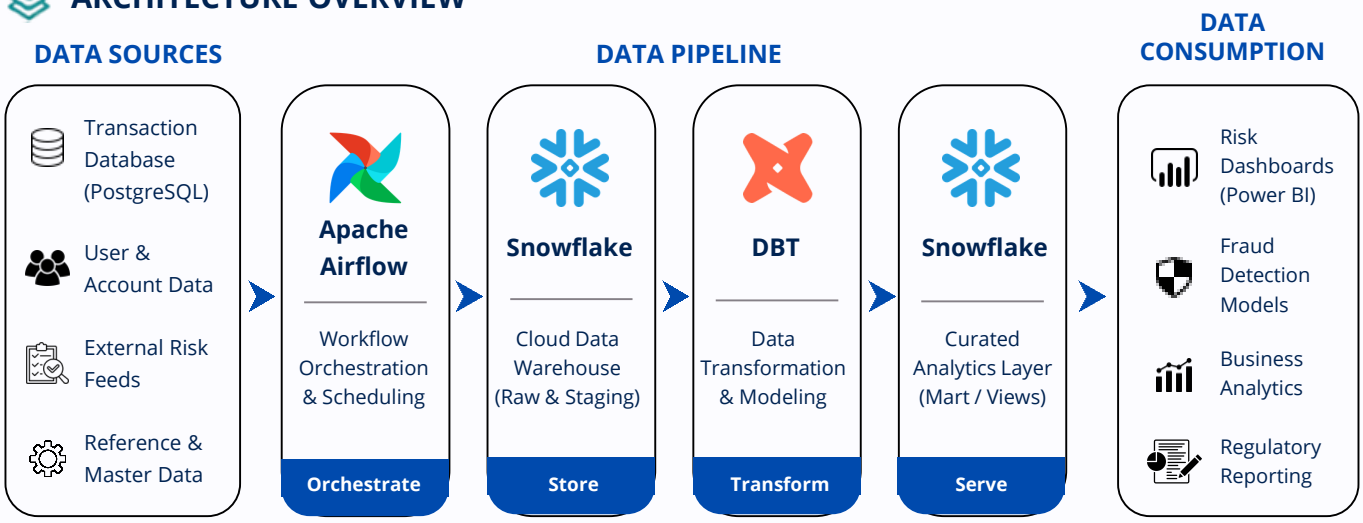
Financial Data Platform & Risk Intelligence

(FinTech Domain)

PROJECT OVERVIEW

Build an enterprise-grade financial data platform using Apache Airflow, Snowflake, and dbt—transforming transaction data from PostgreSQL into actionable insights for risk, fraud, and analytics.

ARCHITECTURE OVERVIEW



TECH STACK

BUSINESS IMPACT

- **Actionable insights for risk & fraud detection, better decision making, and regulatory compliance.**

KEY FEATURES

- **Automated data pipelines with Airflow**
- **Centralized data warehouse on Snowflake**
- **Transformed data models with dbt**
- **Scalable & reliable analytics platform**

Project 4

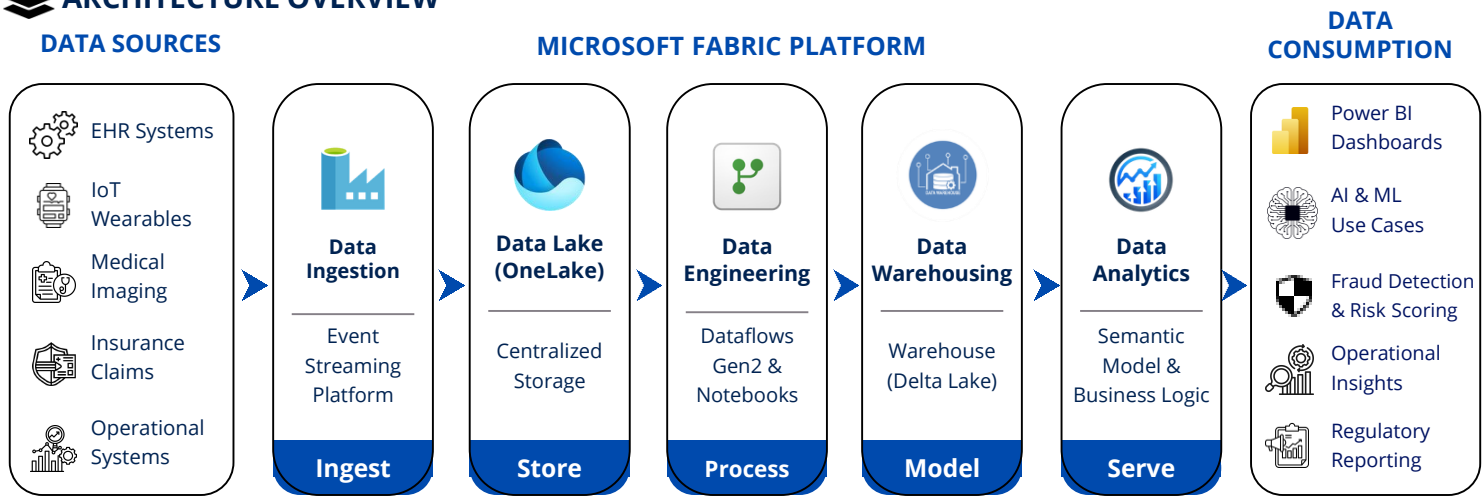
Healthcare Data Platform & AI Analytics

(Healthcare Domain)

PROJECT OVERVIEW

Design an end-to-end healthcare data platform on Microsoft Fabric, processing data from EHR systems, IoT wearables, medical imaging, and insurance claims to power AI-driven use cases like patient risk prediction, fraud detection, and operational optimization.

ARCHITECTURE OVERVIEW



TECH STACK

BUSINESS IMPACT

- AI-driven patient risk prediction and outcomes
- Fraud detection & claim validation
- Operational efficiency & cost optimization
- 300+ patient & population health insights

KEY FEATURES

- Unified data platform on Microsoft Fabric
- Real-time & batch data processing
- Scalable, secure & governed healthcare data
- AI-ready data for advanced analytics & ML

Project 5

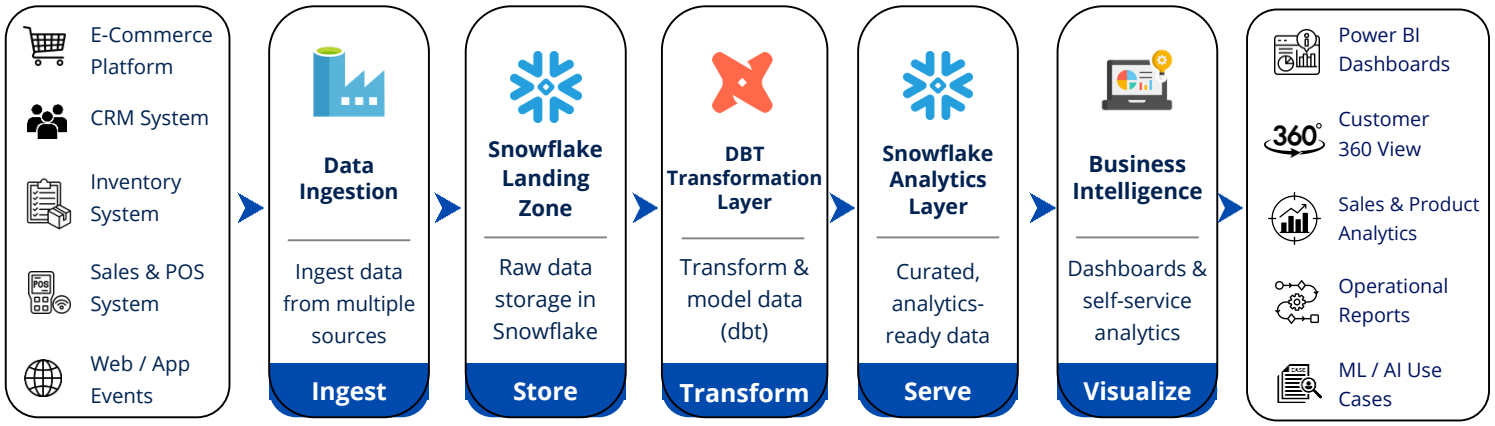
End-to-End Retail Data Platform

(Production Grade)

PROJECT OVERVIEW

Build a production-grade retail analytics platform using Snowflake, transforming sales, customer, and product data into scalable, analytics-ready datasets for business insights and reporting.

ARCHITECTURE OVERVIEW



TECH STACK

Snowflake → DBT → Power BI

STACK USAGE

- Snowflake for scalable data warehouse
- DBT for transformation & data modeling
- Power BI for visualization & reporting

BUSINESS IMPACT

- Unified view of sales, customers & products
- Data-driven decisions for growth & efficiency
- Scalable platform for analytics & reporting

KEY FEATURES

- End-to-end production grade data platform
- Reliable data transformation with dbt
- Scalable & cost-efficient architecture
- Rich dashboards & self-service analytics

CAREER ACCELERATION SUPPORT

Career support built for real data engineering transitions.

Resume Engineering

Industry-focused resume reviews that highlight your skills, projects and impact.

LinkedIn Optimization

Optimize your profile to attract recruiters and create new opportunities.

Mock Interviews

Personalized mock interviews with expert feedback to help you crack real interviews.

Referral Support

Referral opportunities through our hiring network and mentor ecosystem.

Transition Backgrounds

- Analysts
- DBA
- ETL
- Support
- Testing
- Developers

Celebrating Learner Achievements.

Our learners actively share their achievements on LinkedIn, celebrating their new skills and career journey.

Kishor M.

Data Engineer

Azure & Big Data Specialist

10
Years of Experience

India

Kishor M. 3rd+
Senior Data Engineer | Azure & Big Data Specialist
Dec 2025

Thrilled to have completed the Data Engineer certification with Prepzee Learning! Gained hands-on skills in building scalable data pipelines and managing cloud data infrastructure on Azure.

6 3 Comments

Jason King

AI - Data Architect

Azure | Fabric | Snowflake

22
Years of Experience

United States

Jason King 6th+
AI - Data Platform Architect | Azure | Fabric | Snowflake
2025 .

Thrilled to have completed the Data Engineer certification with Prepzee Learning! Gained hands-on skills in building scalable data pipelines and managing cloud data infrastructure on Azure.

6 3 Comments

Waqar Shareef

Data Engineer

Azure | Fabric | Snowflake

22+
Years of Experience

Canada

Waqar Shareef 3rd+
Senior Data Analyst \ Data Engineer | Azure | Fabric | Snowflake
Aug 2025

Thrilled to have completed the Data Engineer certification with Prepzee Learning! Gained hands-on skills in building scalable data pipelines and managing cloud data infrastructure on Azure.

20 1 Comments

Dileep Kumar

Senior Data Engineer

AWS | PySpark | Airflow

5+
Years of Experience

United States

Dileep Kumar 3rd+
Senior Data Engineer | AWS | PySpark | Airflow
1mo .

Excited to share that I've successfully completed the Data Engineer Professional Program from Prepzee Learning.

15 3 Comments

DATA ENGINEER CURRICULUM

Structured learning path to build industry-ready data engineering skills.

Module 1 — Python for Data Engineering

1. Python Fundamentals

- Python setup & environment
- Variables, data types
- Functions & modular code
- Loops, conditions & logic building

2. Data Structures

- Lists, tuples, sets & dictionaries
- Iteration techniques
- Working with different data structures

3. Pandas & NumPy

- Pandas Series & DataFrames
- Data reading & writing
- Data cleaning & manipulation
- Filtering, sorting & aggregation
- NumPy for numerical operations

4. Scripting & Automation

- Running Python scripts
- Script structure & best practices
- Functions & modular programming
- Building reusable components

Module 2 — Data Handling with Python

1. Working with Tabular Data

- Reading & writing data (CSV, Excel, JSON)
- Handling missing data
- Data types & conversions

2. Data Cleaning & Transformation

- Filtering & sorting data
- Grouping & aggregation
- Merging & joining datasets
- Reshaping data

3. Data Analysis Essentials

- Exploratory data analysis
- Summary statistics
- Data visualization basics (Matplotlib / Seaborn)

4. Mini Projects

- Hands-on exercises
- Data cleaning tasks
- Small end-to-end analysis projects

Module 3 Cloud & Microsoft Azure Foundations for Data Engineers

1. Cloud Fundamentals

- Fundamentals of Cloud Computing for Data Engineering
- Understanding Public, Private & Hybrid Cloud Models
- Real-World Cloud Use Cases

2. Azure Fundamentals

- Setting Up and Navigating Microsoft Azure Portal
- Azure Services Used in Modern Data Engineering
- Key Azure Resources Overview

3. Cloud Service Models

- IaaS, PaaS & SaaS Explained with Real-World Use Cases
- Comparing Service Models with Examples
- Choosing the Right Model for Data Workloads

4. Azure Environment Setup

- Creating Your First Azure Account
- Building Your First Azure Resource Group
- Hands-on: Launch Your First Azure Environment

Module 4 Secure Identity & Access Management in Azure

1. IAM Fundamentals

- Introduction to Identity & Access Management in Azure
- Importance of Security in Data Engineering
- Authentication vs Authorization

2. Microsoft Entra ID (Azure AD)

- Microsoft Entra ID Fundamentals
- Understanding Tenants, Users, Groups & Roles
- Role-Based Access Control (RBAC)

3. SPN & Managed Identity

- Service Principals (SPN) for Secure Workflows
- Managed Identity vs Service Principal
- When to Use SPN vs Managed Identity

4. Security Best Practices

- Authentication & Authorization for Azure Resources
- Secure Access to Storage, Databases & Services
- Real-World Security Best Practices for Data Engineers

Module 5 Azure Storage & Data Lake Fundamentals

1. Storage Fundamentals

- Azure Storage Account Fundamentals
- Storage Architecture for Modern Data Platforms

2. Blob Storage Essentials

- Blob Storage Concepts & Real-World Use Cases
- Containers, Folder Structures & File Organization
- Block, Append & Page Blobs

3. Advanced Storage Concepts

- Metadata Storage
- Audit Logs & Configuration Management

4. Queue Storage & Workflows

- Azure Table Storage Concepts
- Azure Queue Storage for Data Engineering Workflows

Module 6 Azure Data Lake Storage Gen2 (ADLS Gen2)

1. ADLS Gen2 Overview

- Introduction to Azure Data Lake Storage Gen2
- ADLS Gen2 Architecture & Hierarchical Namespace

2. Working with Data Lake

- Working with File Systems, Directories & Files
- Creating & Configuring ADLS Gen2

3. Data Ingestion & Management

- Ingesting Data into Data Lake Storage
- Accessing & Managing Data in ADLS Gen2

4. Data Lake Best Practices

- Real-World Data Lake Design Concepts
- Performance & Cost Optimization Tips

Module 7 REST APIs & Data Integration Fundamentals

1. APIs Fundamentals

- Understanding REST APIs in Data Engineering
- REST vs SOAP APIs

2. API Integration Concepts

- API-Based Data Integration Concepts
- Real-World API Integration Scenarios

3. Working with APIs

- Working with API Testing Tools
- CRUD Operations Explained

4. API Data Handling

- Extracting & Working with API Data
- Handling API Responses & Error Management

Module 8 Azure Data Factory (ADF) for Building Data Pipelines

1. ADF Fundamentals

- Introduction to Azure Data Factory (ADF)
- ADF Architecture & Pipeline Execution Flow

2. Pipelines, Datasets & Integration Runtime

- Understanding Pipelines, Datasets & Linked Services
- Creating & Managing Different Dataset Types
- Working with Integration Runtime

3. Data Integration & Orchestration

- Connecting ADF with Azure Storage, SQL & REST APIs
- Data Movement & Transformation Pipelines
- Scheduling & Triggering Data Pipelines

4. Monitoring & Best Practices

- Monitoring, Debugging & Pipeline Optimization
- Real-World Best Practices for Production Data Pipelines

Module 9 PySpark Foundations with Databricks

1. Introduction to Databricks & Spark

- Introduction to Databricks Platform
- Understanding SparkSession & Spark Execution Flow
- Transformations vs Actions in Spark
- Lazy Evaluation in Spark

2. Core Concepts & Data Structures

- Basics of RDDs (Conceptual Overview)
- DataFrames & Distributed Data Processing
- Spark SQL for Analytics & Transformations

3. Data Ingestion in PySpark

- Reading Data from CSV, JSON & Parquet
- Schema Inference & Explicit Schema
- Working with Different File Formats

4. Data Output in PySpark

- Writing Data using DataFrame APIs
- Save Modes & Partitioning Options
- Best Practices for Data Output

Module 10 Data Processing with PySpark in Databricks

1. Data Transformation & Analysis

- Data Ingestion & Transformation using PySpark
- Working with Joins & Aggregations
- Data Cleaning & Type Conversions

2. Advanced Analytics with Spark SQL

- Window Functions in Spark SQL
- Complex Queries & Subqueries
- Using SQL & DataFrame APIs Together

3. Performance Optimization

- Partitioning & Shuffle Operations
- Caching & Persistence in Spark
- Performance Optimization Basics in Spark

4. Monitoring & Tuning

- Monitoring Jobs using Spark UI
- Understanding Stages, Tasks & Executors
- Performance Tuning Best Practices

Module 11 Databricks Platform, Delta Lake & Governance

1. Databricks Platform Essentials

- Azure Databricks Architecture Overview
- Databricks Workspace & Notebook Management
- Managing Clusters & Cluster Pools

2. Data Connectivity & Storage

- Connecting Databricks with ADLS Gen2
- Mounting & Accessing Storage
- Secure Data Access Best Practices

3. Delta Lake Essentials

- Understanding Delta Lake Architecture
- Working with Delta Tables in Databricks
- ACID Transactions in Delta Lake

4. Governance & Unity Catalog

- Understanding Unity Catalog
- Managing Permissions & Access Control
- Data Governance using Unity Catalog

Module 12 Databricks Lakehouse Architecture & Pipelines

1. Lakehouse & Medallion Architecture

- Understanding Lakehouse Architecture
- Medallion Architecture: Bronze, Silver, Gold
- Designing Layered Data Pipelines

2. Building Data Pipelines

- Building Batch Data Pipelines
- Building Real-time (Streaming) Pipelines
- End-to-End Data Processing using Databricks

3. Advanced Pipeline Concepts

- Incremental Processing Techniques
- CDC (Change Data Capture) Concepts
- Pipeline Orchestration Workflow Concepts

4. Use Cases & Best Practices

- Real-world Data Engineering Use Cases
- Performance & Cost Optimization
- Best Practices for Scalable Data Pipelines

Module 13 Introduction to Microsoft Fabric

1. What is Microsoft Fabric

- What is Microsoft Fabric
- How Microsoft Fabric differs from Azure Data Engineering

2. Microsoft Fabric Components

- Explore core components of Microsoft Fabric
- Unified platform for data engineering, science & analytics

3. Real-time Analytics in Fabric

- Introduction to real-time analytics
- Benefits for modern data engineering

4. Microsoft Fabric Eventhouse

- Introduction to Eventhouse
- Real-time data ingestion & processing

Module 14 Data Integration & Orchestration with Microsoft Fabric

1. Dataflows Gen2 in Fabric

- Understand Dataflows Gen2 in Microsoft Fabric
- Explore & integrate Dataflows Gen2

2. Pipelines for Data Engineering

- Integrate pipelines in Microsoft Fabric
- Understand pipeline concepts
- Use pipeline templates

3. Run & Monitor Pipelines

- Run & monitor pipelines
- Track pipeline performance & execution

4. Real-time Data Processing

- Ingest, transform, store & query real-time data
- Visualize real-time data in Microsoft Fabric

Module 15 Lakehouse, Delta Tables & Warehousing in Microsoft Fabric

1. Lakehouse & Medallion Architecture

- Master Fabric Lakehouse with Medallion Architecture (Bronze, Silver, Gold)
- Use Microsoft Fabric for data ingestion, transformation & analysis
- Manage & utilize Lakehouses for Data Engineering certification

2. Delta Tables in Fabric

- Understand Delta Lake & Delta Tables within Fabric
- Create & handle Delta Tables using Spark
- Enhance Delta Table performance
- Work with Delta Tables using structured streaming

3. Data Warehousing in Fabric

- Define Data Warehouses within Fabric
- Differentiate between Data Warehouse & Data Lakehouse
- Work on Data Warehouses in Fabric
- Create & manage fact tables and dimensions
- Load data into a warehouse using T-SQL

4. Dataflows & Data Management

- Load & transform data with Dataflows Gen2
- Manage and protect data in Fabric Data Warehouse

Module 16 CI/CD, Deployment & Governance in Microsoft Fabric

1. Data Loading & Pipeline Strategies

- Explore strategies for loading data into Fabric Data Warehouse
- Construct data pipelines to populate warehouses in Fabric

2. CI/CD in Microsoft Fabric

- Understand CI/CD concepts in Microsoft Fabric
- Configure version control using Git repositories
- Leverage deployment pipelines for streamlined workflows

3. Automation with APIs & KQL

- Automate CI/CD tasks using Fabric APIs
- Work with KQL effectively

4. Advanced SQL & Analytics

- Explore materialized views in Microsoft Fabric
- Explore stored functions in Microsoft Fabric

Module 17 Snowflake Fundamentals & Architecture

1. Snowflake Basics

- Introduction to Snowflake
- Snowflake's use cases in Data Engineering
- Why Snowflake for modern data platforms

2. Setup & Environment

- Setting up Snowflake
- Creating a Snowflake account
- Setting up the Snowflake environment
- Navigating the Snowflake Web UI

3. Architecture & Storage Layers

- Snowflake Architecture Deep Dive
- Cloud Services Layer, Compute Layer, Storage Layer
- Micro-partitioning and its benefits
- How data is stored and accessed in Snowflake

4. Data Types & Tables

- Supported data types (BOOLEAN, INTEGER, STRING, etc.)
- VARIANT data type for semi-structured data (JSON, XML, Parquet)
- Tables (Permanent, Temporary, Transient)

Module 18 Data Loading & Transformation in Snowflake

1. Data Loading Fundamentals

- Time Travel and Fail-safe
- Zero Copy Cloning
- Snowflake's automatic scaling and partitioning
- Loading Data into Snowflake (Data Engineering)

2. Stages & COPY Command

- Creating and managing stages
- File formats supported by Snowflake (CSV, JSON, Parquet, Avro)
- Using Snowflake's COPY command

3. Streams, Tasks & Automation

- Snowflake SQL capabilities for ETL
- Data transformation using Streams and Tasks
- What are Streams and Tasks?
- Automation and scheduling tasks in Snowflake

4. Real-time ETL & Integrations

- Implementing real-time ETL pipelines using Snowflake
- Snowflake's integration with Data Lake and Data Science tools
- Connecting Snowflake to BI tools like Tableau, Looker, Power BI

Module 19 Performance Optimization & Warehousing in Snowflake

1. Virtual Warehouses

- Understanding virtual warehouses in Snowflake
- Optimizing virtual warehouse size and performance
- Auto-suspend and auto-resume configurations

2. Performance Tuning

- Clustering Keys and their benefits
- Query profiling and performance tuning
- Best practices for improving query performance

3. Caching & Optimization Techniques

- Caching in Snowflake
- Managing compute resources efficiently
- Monitoring and analyzing performance

4. Schema Design & Best Practices

- Star schema vs Snowflake schema
- Data modeling best practices
- Designing schemas for high performance

Module 20 Security, Governance & Data Sharing in Snowflake

1. Authentication & Access Control

- Authentication and Authorization
- Role-Based Access Control (RBAC)
- Access controls for sensitive data

2. Security & Monitoring

- Data encryption at rest and in transit
- Auditing and monitoring usage
- Ensuring compliance and security standards

3. Data Sharing & Masking

- Setting up data sharing and data masking
- Sharing data securely with other Snowflake accounts
- Using Snowflake's secure data sharing feature

4. Governance Best Practices

- Data governance in Snowflake
- Data sharing best practices
- Governance, compliance and operational best practices

Module 21 Airflow

1. Airflow Fundamentals

- Introduction of Airflow
- Different Components of Airflow

2. Installation & Setup

- Installing Airflow
- Understanding Airflow Web UI

3. DAGs, Operators & Tasks

- DAG Operators & Tasks in Airflow Job
- Understanding workflow and dependencies

4. Scheduling & Jobs

- Create & Schedule Airflow Jobs
- Trigger, Monitor & Manage Airflow Jobs for Data Processing

Module 22 Kafka

1. Kafka Fundamentals

- Need for Kafka
- What is Kafka
- Core Concepts of Kafka

2. Kafka Architecture

- Kafka Architecture Overview
- Where is Kafka Used
- Understanding the Components of Kafka Cluster

3. Kafka Cluster Management

- Configuring Kafka Cluster
- Brokers, Topics, Partitions & Replication
- Producers, Consumers & Consumer Groups

4. Reliability & Best Practices

- High Availability in Kafka
- Data Durability & Fault Tolerance
- Kafka Best Practices for Data Engineering

Module 23 Snowflake with Cortex AI

1. Cortex AI Introduction

- Introduction to Snowflake Cortex AI
- AI Capabilities Inside the Snowflake Platform

2. Cortex AI Search Service

- Cortex AI Search Service Overview
- Powered by Semantic Search & Embeddings
- Use Cases in Data Engineering

3. Cortex Analyst

- DAG Operators & Tasks in Airflow Job
- Understanding workflow and dependencies

4. Document AI for NLP & Analytics

- NLP on Unstructured Documents
- Information Extraction & Classification
- Predictive Analytics Using Document AI

Module 24 Data Engineering for AI Systems

1. RAG Architecture & Data Pipeline

- RAG Architecture from a Data Pipeline View
- End-to-End Flow: Source → Embeddings → Vector DB → LLM
- Role of Data Engineers in RAG Pipelines

2. Vector Databases & Embeddings

- Vector Databases for Storage & Retrieval
- Storing & Managing Embeddings
- Similarity Search & Retrieval Mechanisms

3. Data Processing & Embedding Pipelines

- Configuring Kafka Cluster
- Brokers, Topics, Partitions & Replication
- Producers, Consumers & Consumer Groups

4. Data Ingestion Strategies

- Data Ingestion Strategies for AI Applications
- Batch & Streaming Ingestion for AI Systems

5,000+
Careers Advanced

FEATURED SUCCESS

Eresh Tayanna

Data Engineer
at **Maveric (USA)**

TCS TATA CONSULTANCY SERVICES >>> MAVERIC ACCELERATE NEXT

Transitioned from **On-Premise to Cloud** Data Engineering & moved to **USA**

FEATURED SUCCESS

Abhishek Pareek

Technical Manager
at **Capgemini**

RenoSys >>> Capgemini

200% Salary Growth
Got 2 Job Offers

FEATURED SUCCESS

Vishal Purohit

Product Manager
at **Icertis**

ZYCUS COGNITIVE PROCUREMENT >>> Icertis

Moved to a **Product Based Company**
Immense Job Opportunities

FEATURED SUCCESS

Aman Yadav

Senior Associate
Product Compliance
at **Amazon**

ICICI Bank >>> amazon

Journey From **Non-IT to IT**
Now Building a Successful Tech Career

What Our Learners Have To Say

Real stories. Real transformations. **Real careers.**

Kishan Yadav
Data Engineer
Working at | Microsoft

★★★★★

Placed at Microsoft as a Data Engineer! The program gave me strong practical exposure to real-world data engineering workflows, hands-on projects, and guidance from industry professionals. It really helped me build the skills and confidence needed for this transition. Highly recommended for anyone serious about a career in Data Engineering.

Vasantha Manvitha Bonda
Data Engineer
Working at | Cargill

★★★★★

After spending several years in software development, I wanted to transition into modern Data Engineering. The program provided a structured learning path covering Azure, Databricks, Snowflake, Airflow, Kafka, and real-world project implementation. What helped me the most was the hands-on approach and practical exposure to industry-relevant tools. The knowledge gained during the program played an important role in helping me successfully transition into a Data Engineer role and eventually grow into a Senior Data Engineer position.

Eresh Tayanna
Senior Data Engineer
Working at | MAVERIC ACCELERATE NEXT

★★★★★

Transitioned from traditional Big Data technologies into modern Data Engineering workflows with the help of this program. The practical learning approach, real-world projects, and exposure to new-age tools helped me strengthen my skills in modern data engineering. Highly recommended for professionals looking to upgrade from Big Data to cloud and modern Data Engineering technologies.

Patrick Terkpertey
Product Manager
Working at | Honeywell

★★★★★

The Prepzee Data Engineering course serves as a premier career launchpad that masterfully bridges the gap between certification and industry readiness through expert-led instruction and high-impact hands-on projects. By focusing on end-to-end ETL pipelines and industry-standard big data tools, the program ensures students develop the architectural intuition needed to solve complex data challenges. With a curriculum designed for the modern job market, it is an essential investment for anyone looking to build a professional portfolio and secure a high-paying role in the data engineering field.

Jatin Garg

DBA

Working at | TCS | TATA CONSULTANCY SERVICES

★★★★★

Coming from a PostgreSQL DBA background, I wanted to broaden my expertise beyond database administration and gain hands-on experience in modern Data Engineering technologies. The program provided practical exposure to Azure, Databricks, Snowflake, Airflow, and real-world data engineering workflows. The structured curriculum and project-based learning helped me understand how large-scale data platforms are built and managed in the cloud. It has significantly strengthened my ability to contribute to modern data engineering initiatives alongside my database expertise.

Sheetal Katiyar

Senior Module Lead

Working at | **CONFIDENCIAL**

★★★★★

Prepzee has been a great partner for us and is committed towards upskilling our employee.Their catalog of training content covers a wide range of digital domains and functions, which we appreciate.The best part was there LMS on which videos were posted online for you to review if you missed anything during the class.I would recommend Prepzee to all to boost his/her learning.The trainer was also very knowledgeable and ready to answer individual questions.

Kishor

Azure Data Engineer

Working at | Synechron

★★★★★

The program helped me strengthen my expertise in modern Azure Data Engineering with a strong focus on practical implementation and real-world workflows. The hands-on projects, industry-oriented approach, and guidance from working professionals made the learning highly valuable. A great program for professionals looking to upskill in modern cloud and data engineering technologies.

Waqar Shareef

Senior Data Analyst

Working at | **SHAREEFS**

★★★★★

I have completed the Data Engineering program with Prepzee, where I learned Azure, Snowflake, DBT, and Airflow. It was an amazing experience with a strong focus on practical learning. Special thanks to our trainer, Aneel, who provided great support during the lab sessions. Prepzee is the best platform for beginners who want to become experts in Data Engineering.

Our Alumni Work At

Microsoft databricks MAVERIC ACCELERATE NEXT Synechron Honeywell Sundyne

WELLS FARGO tcs TATA CONSULTANCY SERVICES IBM Deloitte. accenture

PREPZEE

prepzee.com

For Further Details, **Contact Us**

hi@prepzee.com

+1 415-481-4467

+91-7793068396

Chat with Us

We are available 24*7