

Data Engineer

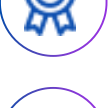
JOB ORIENTED PROGRAM



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



Index

	About Prepzee	01
	The Gap Between Learning & Getting Hired	04
	Real People Real Transitions	05
	Who Is This Program For?	06
	Where Can This Program Take You?	07
	Program Features	08
	Roadmap to Become a Modern Data Engineer	09
	Tools Covered	10
	Clear Industry Recognised Certification	11
	Unlock Bonuses worth ₹32000	12
	Portfolio-Worthy Data Projects	13
	Career Acceleration Support	19
	Celebrating Learner Achievements	20
	Data Engineer Curriculum	22
	What Our Learners Have To Say	35

About Prepzee

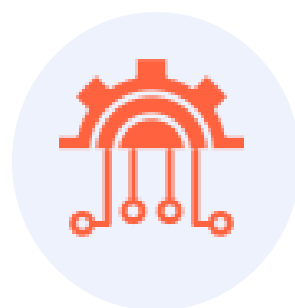
Preparing Learners for Modern Data Engineering Careers.



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

ANI

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

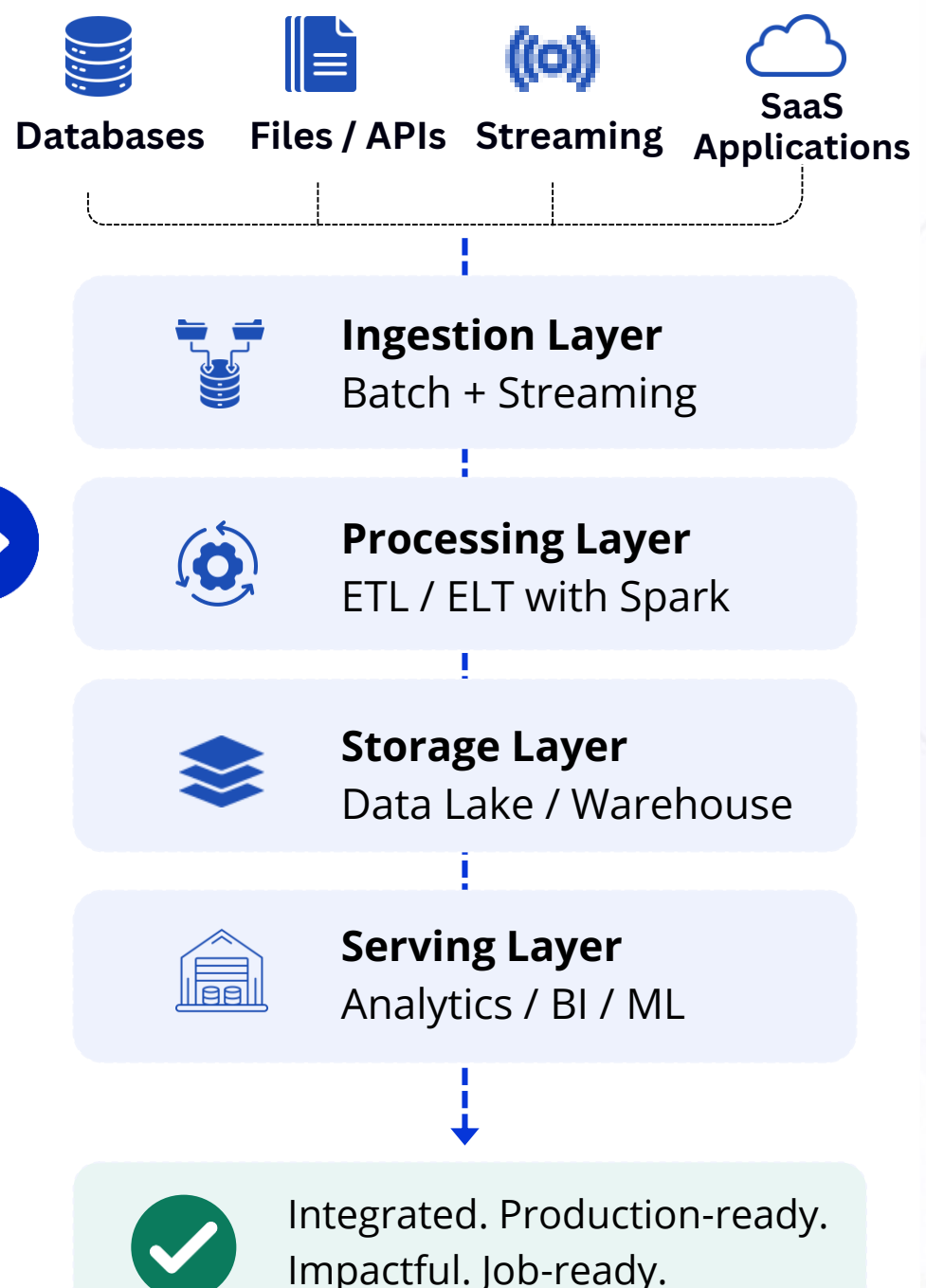
You learn the tools.
The industry expects **production-ready** engineers.

WHAT MOST LEARNERS DO (DISCONNECTED LEARNING)



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

WHAT INDUSTRY EXPECTS (JOB-READY DATA ENGINEER)



Real People. Real Transitions.

Professionals from diverse backgrounds transformed into top data engineering roles.



Kishan Yadav

Data Engineer @ Microsoft



5
MONTHS
Journey



2.6x
Salary Hike



Mayank Sharma

Data Engineer @ 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.



AWS Data Engineer

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



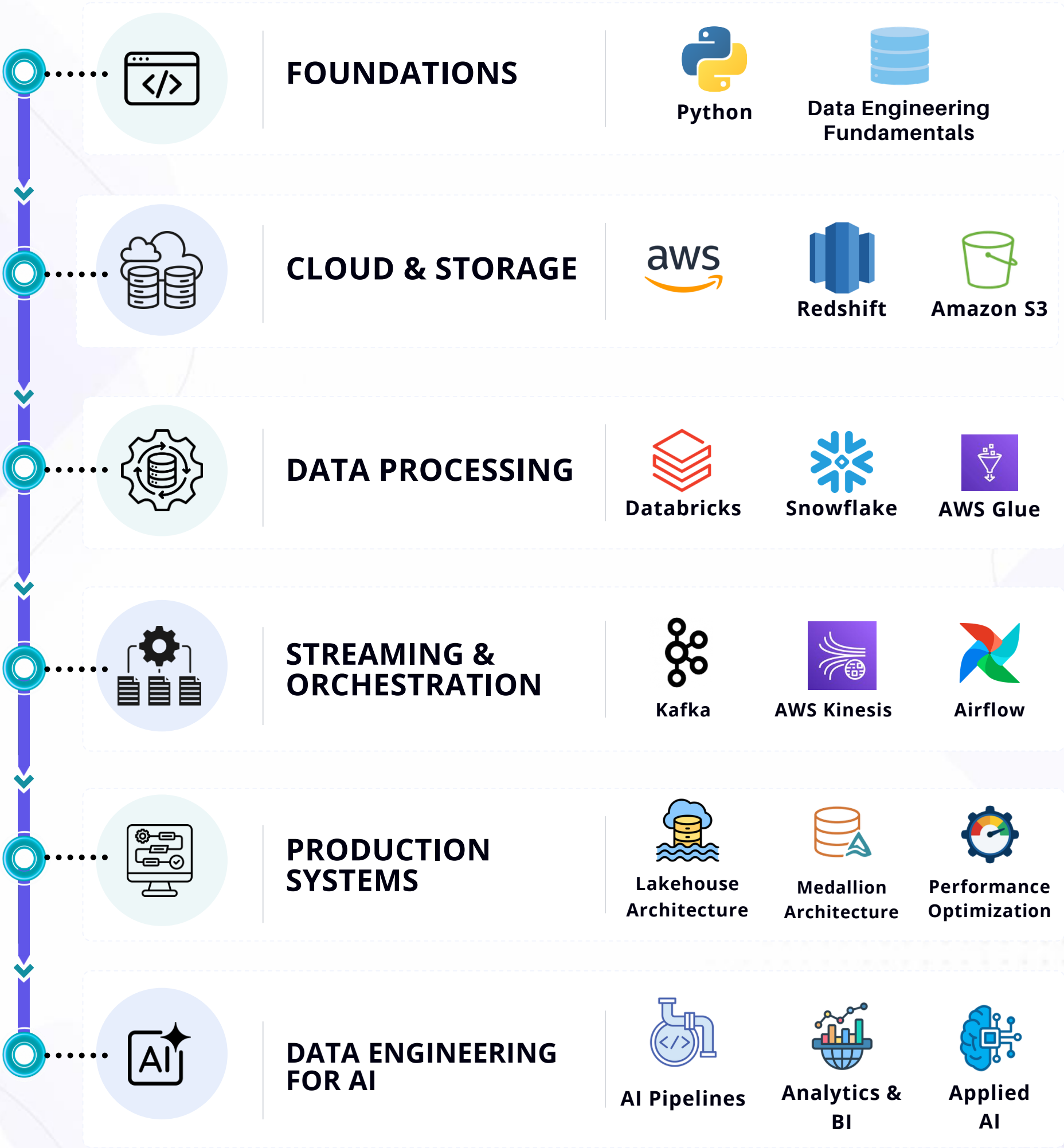
Lifetime

Live Training Access

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

Roadmap to Become a Modern Data Engineer

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



TOOLS COVERED

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



Redshift



Airflow



Kafka



DBT



Cortex AI



Python



Medallion
Architecture



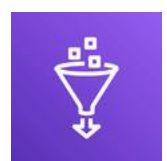
Delta Lake



AWS Kinesis



AWS Lambda



AWS Glue



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 ₹32000

Get access to premium courses absolutely
FREE with our program.

NEW BONUS



DBT Live Training



Modern Data Transformation for Data Engineers

- ✓ 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 [REC]

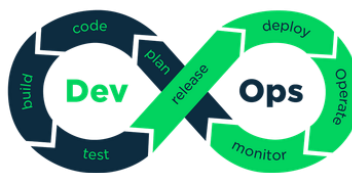


SQL Fundamentals Course

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

BONUS 3

FREE COURSE [REC]



DevOps Certification 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.



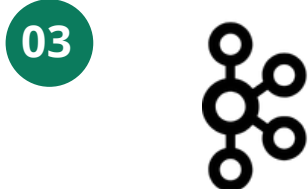
5 INDUSTRY-FOCUSED PROJECTS



Enterprise CDC Platform on AWS
Insurance Domain



Real-Time Customer Analytics Platform
Event-Driven Architecture



Real-Time Event Streaming Platform
Mobility & Transportation Domain



Payment Fraud Detection Analytics Platform
Fintech Domain



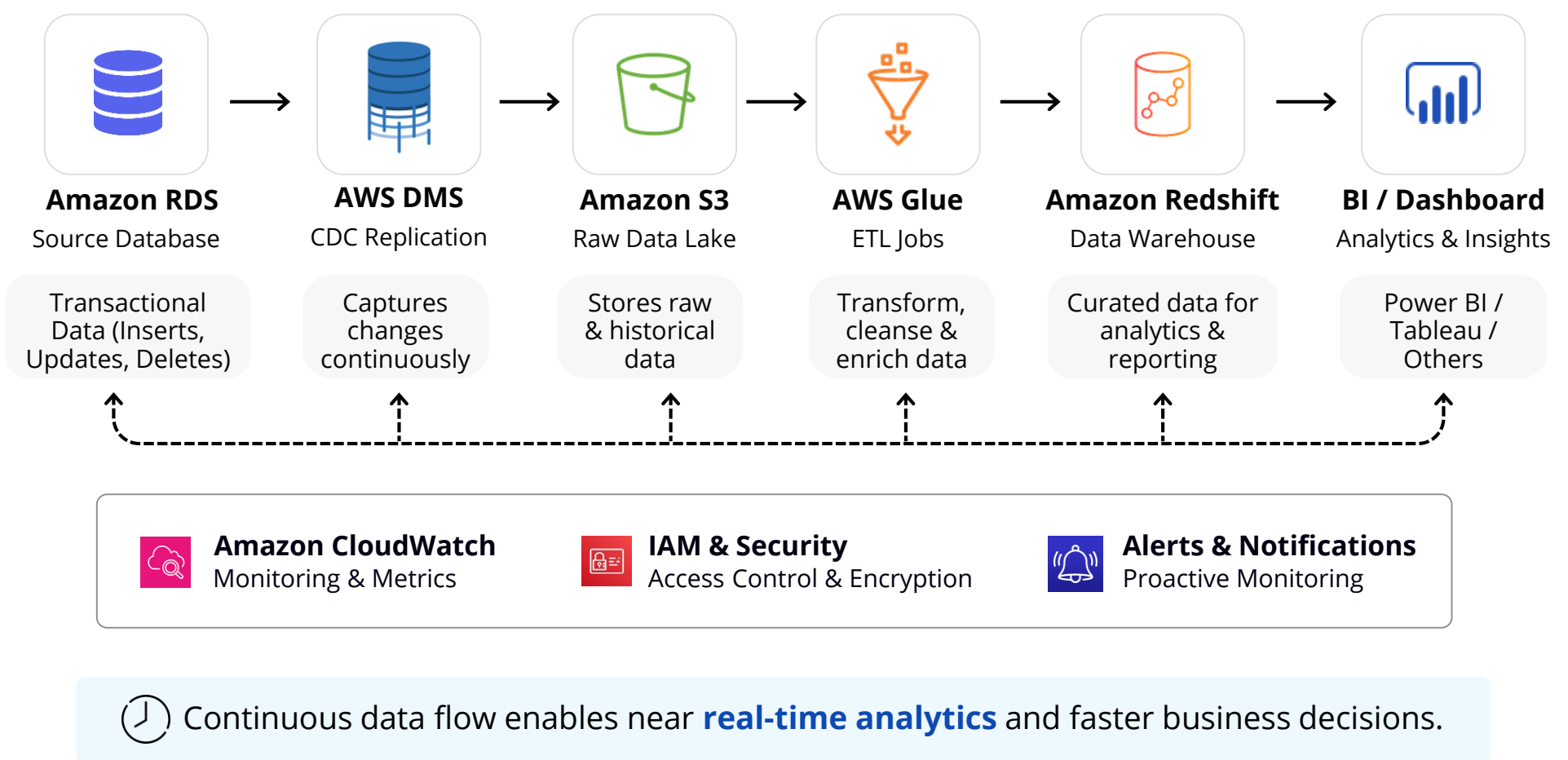
Real-Time Data Streaming Pipeline
Production Grade

Enterprise Change Data Capture (CDC) Platform on AWS

Designed and implemented a production-ready CDC pipeline to replicate transactional data from Amazon RDS into Amazon Redshift for **near real-time analytics** and reporting.



PRODUCTION ARCHITECTURE



WHAT YOU'LL ENGINEER

- Build CDC pipelines using AWS DMS
- Design scalable Data Lakes on Amazon S3
- Transform enterprise data using AWS Glue
- Build analytical models in Amazon Redshift

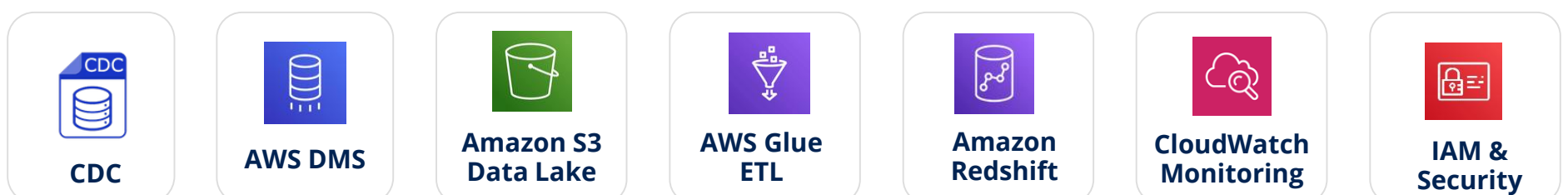


RECRUITER PERSPECTIVE

After completing this project, you'll be able to confidently discuss the following in interviews:

- End-to-end CDC Architecture
- AWS DMS & Glue Workflows
- Data Lake & Warehouse Design
- Real-world Production Challenges

ENTERPRISE TECHNOLOGIES USED



Real-Time Customer Analytics Platform using Databricks

Build a production-ready CDC Lakehouse that processes millions of customer transactions into analytics-ready datasets using **Databricks & Delta Lake**.

PRODUCTION ARCHITECTURE



WHAT YOU'LL ENGINEER

- Production-ready CDC pipeline from Amazon RDS to Databricks
- Medallion Lakehouse with Bronze, Silver & Gold layers
- Incremental PySpark pipelines with Delta Lake
- Analytics-ready data models for enterprise reporting



RECRUITER PERSPECTIVE

After completing this project, you'll be able to confidently discuss the following in interviews:

- End-to-end Lakehouse architecture design
- Incremental ETL & CDC implementation
- Delta Lake optimization & data quality
- Databricks SQL for enterprise analytics

ENTERPRISE TECHNOLOGIES USED



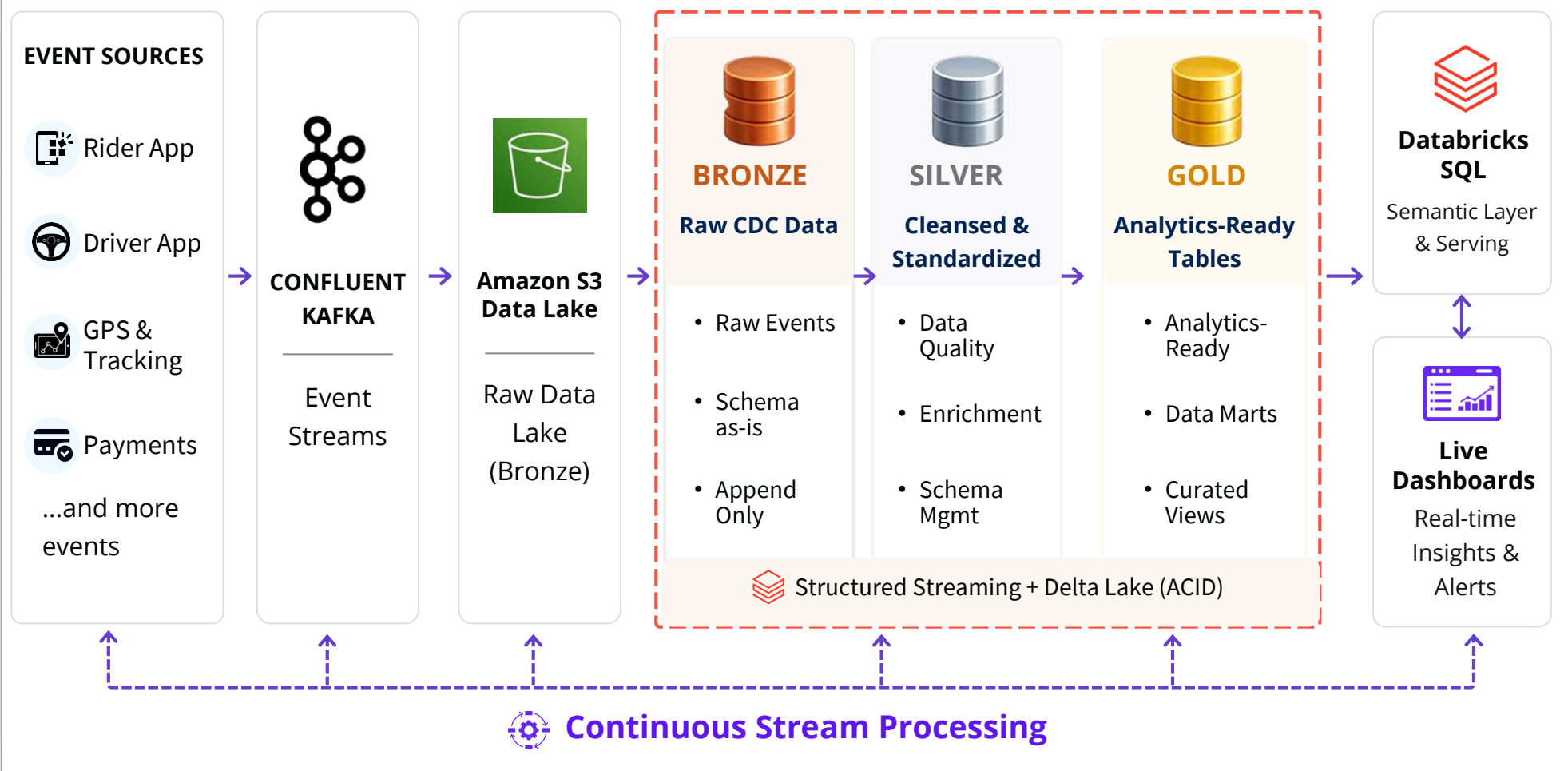
Real-Time Event Streaming Platform



Ride Sharing Live Analytics Platform

Build a real-time event streaming platform to capture, process and **analyze ride events** from millions of users and power live dashboards with operational & business insights.

STREAMING DATA FLOW ARCHITECTURE



WHAT YOU'LL ENGINEER

- Enterprise event streaming platform
- Fault-tolerant streaming ETL pipelines
- Delta Lake Medallion architecture
- Real-time business intelligence layer



RECRUITER PERSPECTIVE

After completing this project, you'll be able to confidently discuss the following in interviews:

- Event-driven architecture design
- Streaming data engineering concepts
- Lakehouse optimization strategies
- Real-time analytics solutions

ENTERPRISE TECHNOLOGIES USED

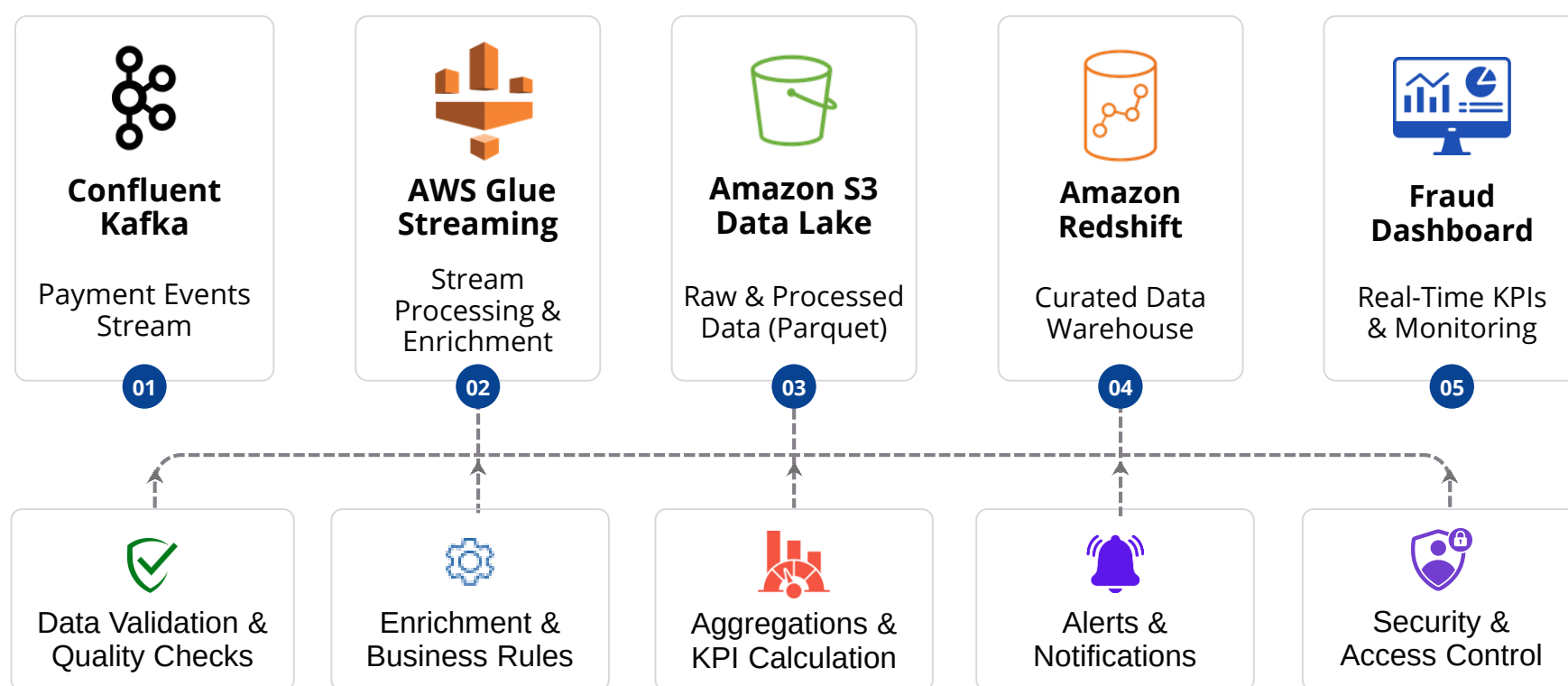


Payment Fraud Detection Analytics Platform

(Confluent Kafka → AWS Glue → Amazon S3 → Amazon Redshift)

Build a real-time **fraud detection platform** that processes live payment events into actionable business insights in real time.

REAL-TIME EVENT PROCESSING PIPELINE



Amazon CloudWatch
Monitoring & Metrics



Alerts & Notifications
Proactive Monitoring



Auto Scaling
High Availability



WHAT YOU'LL ENGINEER

- Real-time streaming ETL pipelines using Confluent Kafka & AWS Glue
- Production-ready data lake architecture on Amazon S3
- Curated Redshift data warehouse for fraud analytics & reporting
- Automated monitoring, alerting & data quality validation



RECRUITER PERSPECTIVE

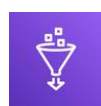
After completing this project, you'll be able to confidently discuss the following in interviews:

- Event-driven data pipeline design
- Streaming ETL implementation concepts
- Redshift analytics optimization
- Production-ready architecture practices

ENTERPRISE TECHNOLOGIES USED



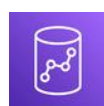
Confluent Kafka



AWS Glue Streaming



Amazon S3 Data Lake



Amazon Redshift



Streaming ETL



Data Quality & Validation



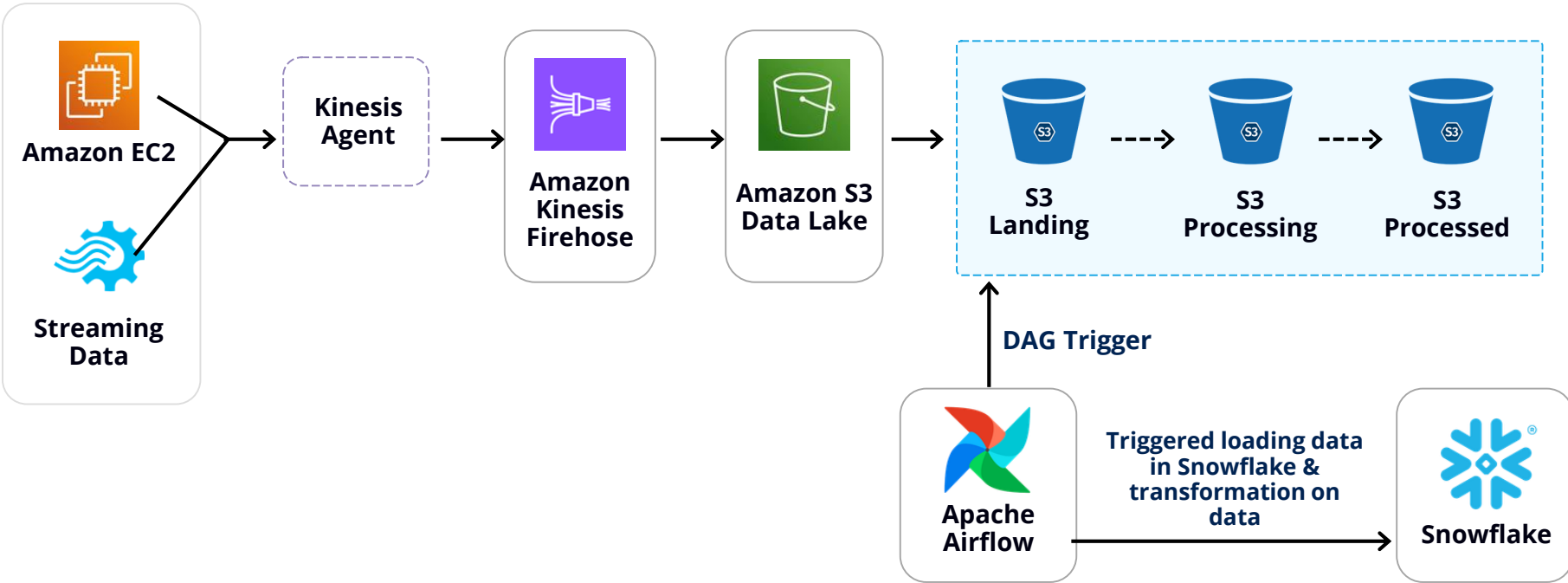
BI / Dashboard

Real-Time Data Streaming Pipeline with AWS Kinesis & Snowflake

(EC2 / Streaming Data → Kinesis Agent → Firehose → S3 → Snowflake)

Implement a live **data streaming pipeline** using AWS Kinesis for ingestion, Amazon S3 for storage, and Snowflake for analytics and reporting.

REAL-TIME ARCHITECTURE



Amazon CloudWatch
Monitoring & Metrics



Alerts & Notifications
Proactive Monitoring



Auto Scaling
High Availability



WHAT YOU'LL ENGINEER

- Real-time streaming pipeline using Amazon Kinesis
- Scalable data lake architecture on Amazon S3
- Automated orchestration with Apache Airflow
- Analytics-ready datasets in Snowflake



RECRUITER PERSPECTIVE

After completing this project, you'll be able to confidently discuss the following in interviews:

- Streaming data pipeline architecture
- Kinesis & Firehose implementation
- Airflow orchestration concepts
- Snowflake data loading & optimization

ENTERPRISE TECHNOLOGIES USED



Kinesis Data
Firehose



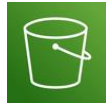
Apache
Airflow



Snowflake



Streaming
Data



Amazon S3



Monitoring &
Scheduling



Amazon EC2

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

 10
Years of Experience

 India



Kishor M.  3rd+
Senior Data Engineer | 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.



  6

3 Comments



Jason King
AI - Data Architect

 22
Years of Experience

 United States



Jason King  6th+
AI - Data Platform Architect | 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.



  6

3 Comments



Waqar Shareef

Data Engineer



22+
Years of Experience



Canada



Waqar Shareef 3rd+
Senior Data Analyst \ Data Engineer
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.



20

1 Comments



Dileep Kumar

Senior Data Engineer



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

AWS Cloud Computing Fundamentals



1. Cloud Computing Introduction

- Cloud Computing Introduction
- Understand IAAS, PAAS, SAAS
- Real-World Cloud Use Cases



2. Understand IAAS, PAAS, SAAS

- Understand IAAS, PAAS, SAAS
- Deployment Models
- Service Models Comparison



3. AWS Account Setup & Configuration

- AWS Account Setup & Configuration
- Root User vs IAM User
- Billing & Cost Management Basics



4. AWS Regions & Availability Zones

- Understanding AWS Regions
- Availability Zones (AZ) Concept
- High Availability & Fault Tolerance



Module 4

Compute & Storage (AWS EC2 & EBS)



1. Amazon EC2 Introduction

- Introduction to Amazon Elastic Compute Cloud (EC2)
- Benefits of EC2
- EC2 Instance Types



2. Network in EC2

- Public IP vs. Elastic IP
- Security Groups
- Key Pairs
- Best Practices



3. Amazon Machine Image (AMI)

- Introduction to AMI
- Creating Custom AMIs
- Sharing & Using AMIs



4. Amazon EBS (Elastic Block Store)

- Introduction to EBS
- EBS Volume Types
- Snapshots & Backup
- EBS Volume Types and Snapshots



aws

Module 5

Networking (AWS VPC)



1. VPC Fundamentals

- Introduction to Amazon VPC
- Benefits of Virtual Private Cloud
- Public vs Private Networking



2. Core Networking Components

- Route Tables & Internet Gateway
- NAT Gateway & NAT Instances
- Network Interfaces (ENI)



3. IP Addressing & Connectivity

- CIDR Blocks & IP Addressing
- Public IP vs Elastic IP
- VPC Peering & Connectivity



4. Enterprise Network Design

- VPC Endpoints
- Design Patterns & Best Practices
- Real-world Architecture Scenarios

aws

Module 6

Identity & Security Management (IAM)



1. IAM Fundamentals

- Introduction to IAM
- Users, Groups & Roles
- Authentication vs Authorization



2. Access Control

- IAM Policies
- Permissions & Least Privilege
- Identity Federation



3. Resource Security

- Resource Access Manager (RAM)
- Cross-Account Resource Sharing
- Service-Level Permissions



4. Enterprise Security Practices

- Security Best Practices
- Access Governance
- Real-world IAM Use Cases



aws

Module 7

Data Lake & Lakehouse with Amazon S3

1. Amazon S3 Fundamentals



- Introduction to Amazon S3
- Buckets, Objects & Folder Structure
- Uploading, Downloading & Managing Files

2. Enterprise Data Lake Design



- Raw, Processed & Curated Data Zones
- Naming Conventions & Best Practices
- Data Lake Folder Architecture

3. Storage Management & Optimization



- Storage Classes & Lifecycle Policies
- Multipart Upload & Large File Handling
- Cost Optimization Strategies

4. S3 for Data Engineering



- Integration with AWS Data Services
- Lakehouse Architecture using S3
- Real-World Enterprise Design Patterns



Module 8

PySpark Foundations with Databricks

1. Introduction to Databricks & Spark



- Introduction to the Databricks Lakehouse Platform
- Understanding Spark Execution Flow & SparkSession
- DataFrames vs Distributed Processing in Spark

2. PySpark Core Concepts



- Transformations vs Actions in PySpark
- Spark SQL for Data Processing & Analytics
- Lazy Evaluation & Distributed Execution Concepts

3. Data Ingestion with PySpark



- Reading CSV, JSON & Parquet Files
- Schema Inference & Explicit Schema Definition
- Working with Different Data Sources using PySpark

4. Writing & Managing Data



- Writing Data into Delta Lake Tables
- Save Modes, Partitioning & File Formats
- Best Practices for Efficient Data Output



Module 9

Data Processing with PySpark in Databricks



1. Data Transformation & Analytics

- Building ETL Pipelines using PySpark
- Joins, Aggregations & Window Functions
- Data Cleaning & Type Conversion Techniques



2. Delta Lake Processing

- Working with MERGE, UPDATE & DELETE Operations
- Incremental Processing using Delta Lake
- Time Travel & ACID Transactions



3. Performance Optimization

- Partitioning, Caching & Persistence Strategies
- Shuffle Optimization & Query Performance
- Spark Performance Tuning Best Practices



4. Structured Streaming

- Building Real-Time Streaming Pipelines
- Checkpointing & Fault-Tolerant Processing
- Stream Processing using Structured Streaming



Module 10

Databricks Platform, Delta Lake & Governance



1. Databricks Platform Essentials

- Azure/AWS Databricks Workspace Architecture
- Cluster Management & Notebook Development
- Workspace Administration & Collaboration



2. Data Connectivity & Storage

- Connecting Databricks with Amazon S3
- Mounting & Accessing External Storage
- Secure Data Access Best Practices



3. Delta Lake Essentials

- Delta Lake Architecture & Storage Layers
- Working with Delta Tables & Versioning
- ACID Transactions & Schema Evolution



4. Governance & Unity Catalog

- Understanding Unity Catalog Architecture
- Managing Permissions & Data Access Control
- Enterprise Data Governance Best Practices



aws

Module 11

Enterprise Data Engineering with Databricks



1. Lakehouse & Medallion Architecture

- Designing Bronze, Silver & Gold Data Layers
- Building Enterprise Lakehouse Architectures
- Layered Data Processing Best Practices



2. Batch & Streaming Pipelines

- Building Batch Data Processing Pipelines
- Developing Structured Streaming Pipelines
- End-to-End Pipeline Design using Databricks



3. CDC & Incremental Processing

- Change Data Capture (CDC) Concepts
- Incremental Data Loading using Delta MERGE
- Designing Production-Ready Data Pipelines



4. Enterprise Best Practices

- Pipeline Monitoring & Workflow Management
- Performance & Cost Optimization Techniques
- Real-World Enterprise Data Engineering Patterns

aws

Module 12

Build Production ETL Pipelines with AWS Glue



1. AWS Glue Fundamentals

- Introduction to AWS Glue & Serverless ETL
- Understanding Glue Components & Architecture
- Glue Data Catalog, Databases & Tables



2. Metadata & Data Discovery

- Crawlers for Automatic Schema Discovery
- Managing Metadata using Glue Data Catalog
- Working with Partitions & Schema Evolution



3. Building ETL Pipelines

- Developing ETL Jobs using AWS Glue
- Data Transformation & Cleansing Techniques
- Connecting Glue with Amazon S3, RDS & Redshift



4. Production Automation & Monitoring

- Glue Triggers & Workflow Orchestration
- Job Scheduling, Monitoring & Logging
- Enterprise ETL Best Practices & Use Cases



Module 13

Enterprise Data Warehousing with Amazon Redshift

1. Amazon Redshift Fundamentals



- Introduction to Amazon Redshift Architecture
- Understanding Clusters, Nodes & Data Warehouse Concepts
- Creating Databases, Schemas & User Management

2. Data Modeling & Storage



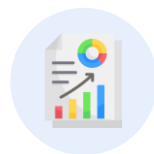
- Designing Fact & Dimension Tables
- Data Types, Keys & Table Design
- Sort Keys, Distribution Keys & Compression Techniques

3. Data Loading & Query Optimization



- Loading Data from Amazon S3 & AWS Glue
- Query Performance & Workload Management (WLM)
- Optimizing Data Warehouse Performance

4. Analytics & Enterprise Reporting



- Connecting Amazon Redshift with BI & Analytics Tools
- Building Analytics-Ready Data Models
- Enterprise Best Practices for Scalable Data Warehouses



Module 14

Real-Time Streaming with Apache Kafka

1. Kafka Fundamentals



- Introduction to Apache Kafka & Event Streaming
- Kafka Architecture: Topics, Brokers & Partitions
- Understanding Producers & Consumers

2. Kafka Core Concepts



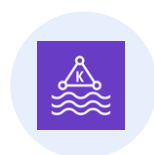
- Consumer Groups & Offset Management
- Replication, Fault Tolerance & High Availability
- Message Delivery Guarantees (At-Least, Exactly-Once)

3. Real-Time Streaming with Kafka



- Building Real-Time Data Pipelines
- Publishing & Consuming Streaming Events
- Enterprise Streaming Use Cases

4. AWS MSK & Production Practices



- Amazon MSK (Managed Kafka Service)
- AWS MSK vs Self-Managed Kafka
- Production Best Practices & Monitoring



aws

Module 15

AWS Kinesis for Streaming Data Ingestion

1. Amazon Kinesis Fundamentals



- Introduction to Amazon Kinesis Services
- Kinesis Data Streams Architecture
- Core Streaming Concepts & Use Cases

2. Building Streaming Pipelines



- Producers & Consumers in Kinesis
- Streaming Data Ingestion Workflows
- Processing Real-Time Data Streams

3. Amazon Kinesis Firehose



- Delivering Data to Amazon S3 & Redshift
- Data Buffering & Automatic Delivery
- Integrating Firehose with Analytics Platforms

4. Kinesis Integration & Best Practices



- Kinesis vs Kafka – Choosing the Right Service
- Integrating Kinesis with AWS Data Services
- Monitoring, Scaling & Production Best Practices

aws

Module 16

Additional AWS Services & Certification Concepts

1. AWS Lambda



- Introduction to Serverless Computing with AWS Lambda
- Creating & Deploying Lambda Functions
- Event Sources & Trigger-Based Execution

2. Amazon Athena



- Querying Data Directly from Amazon S3
- Serverless SQL Analytics
- Athena Integration for Data Exploration

3. Amazon DynamoDB



- Introduction to NoSQL Database Concepts
- Reading & Writing Data using DynamoDB
- DynamoDB vs Amazon RDS for Modern Applications

4. Enterprise Use Cases & Certification Concepts



- Integrating Lambda with Data Pipelines
- Selecting the Right AWS Service for Different Workloads
- Common Data Engineering & Certification Scenarios



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

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 23

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

[Need More Info?](#) 🤔

[Chat with Us](#)



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 | 



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 |



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 AWS, 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 their 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

Data Engineer

Working at | **Synechron**



The program helped me strengthen my expertise in modern 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 AWS, 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



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