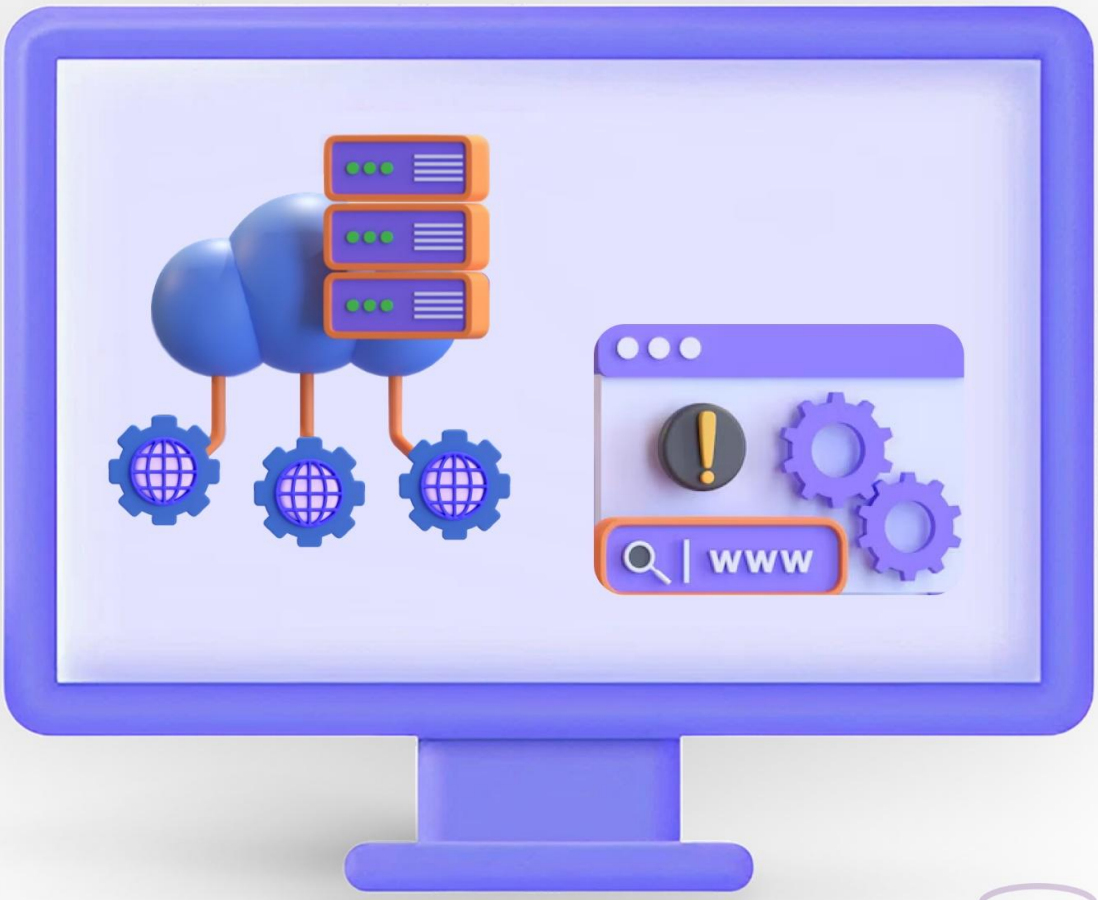
TeqCertify
Elevate Your Data Journey

DATA ENGINEERING



Why Data Engineering?

Have you ever wondered how **millions of transactions,** social media updates, and AI-powered apps run smoothly every second? That's the power of **Data Engineering!** Data Engineers build the pipelines that collect, store, and process massive datasets, ensuring businesses can access real-time insights. Whether it's **handling e-commerce orders, enabling AI, or managing cloud storage,** data engineering is the foundation of the digital world.

Why Data Engineering?

Explosive Growth

The market is set to reach ₹7.5 lakh crore by 2028.

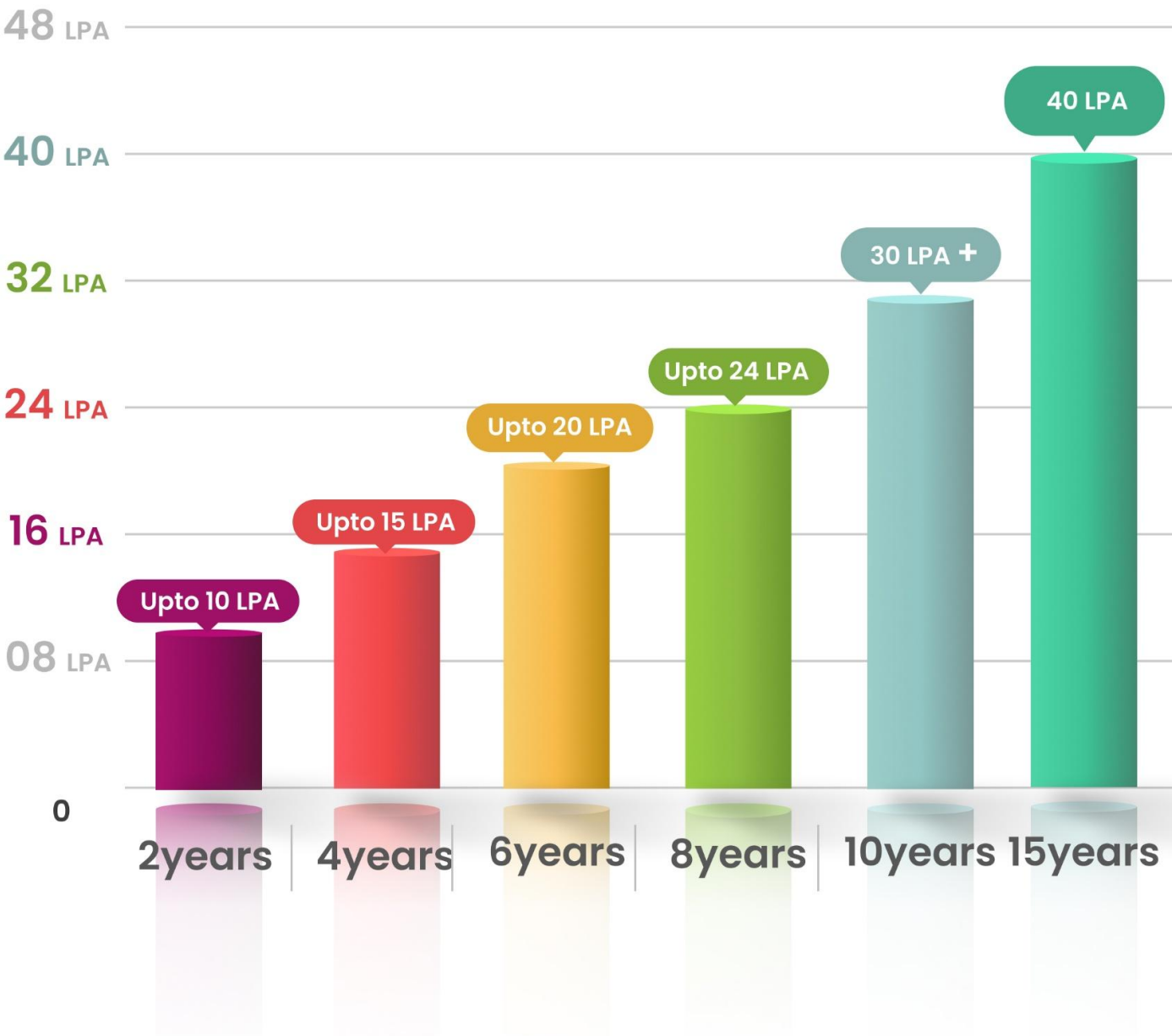
High-Paying Jobs

Become a Data Engineer, Cloud Data Architect, or ETL Developer.

Industry Applications

Essential for AI, e-commerce, banking, and big tech companies.

Career Opportunities



Years of Experience

Journey With TeqCertify

Capstone Project
Designing Scalable Data Solutions

DSA & System Design
Architecting Efficient & Scalable Systems

Advanced Data Ops
Enhancing Efficiency, Automating Workflows

Big Data & Cloud
Powering Scalable & Agile Solutions

ETL & Data Warehousing
Engineering Reliable Data Pipelines

Python
Empowering Scalable Data Workflows

Path to your Dream Job

Build a Standout Resume & Portfolio

01

Craft an ATS-friendly resume.
Highlight key skills & achievements.
Showcase your best projects & experience.

Master Interview Skills

02

Learn how to answer tricky questions.
Improve communication & confidence.
Practice with mock interviews.

Get Expert Feedback & Identify Gaps

03

Receive constructive feedback on interviews.
Understand areas that need improvement.
Make necessary revisions.

Improve & Upskill

04

Gain new skills based on feedback.
Strengthen technical & problem-solving abilities.
Enhance communication & presentation skills.

Land Your Dream Job

05

Negotiate salary & job offers smartly.
Secure the right job that fits your goals.
Get mentorship for long-term career growth.

Why should you invest in the track?

Build the Invisible Systems That Power the World

Love tech? Data engineering powers Netflix recommendations, self-driving cars, and real-time trends! High demand, great pay, and exciting challenges—no genius needed, just curiosity. Master Python, SQL, and cloud tools to build the invisible systems that run the world.

Syllabus Breakdown

Build the backbone of AI and analytics! Learn how to manage, store, and process large-scale data.

01 **Data Engineering 101**
Understand how data moves from collection to insights.

02 **SQL & Databases**
Work with relational & NoSQL databases.

03 **Big Data Tools**
Process massive datasets with Hadoop, Spark & Kafka.

04 **ETL & Data Pipelines**
Automate data cleaning and transformation.

05 **Cloud Data Warehousing**
Store & manage data with AWS, GCP, & Snowflake.

06 **Real-time Data Processing**
Work with live data streams for instant insights.

07 **Scalable Systems**
Build high-performance data infrastructure.

Track Curriculum

Phase 1: Programming & Database Foundations (18 Modules)

Programming & Database Foundations

Lay a solid foundation in Python, SQL, and databases for data engineering workflows.

Unit 1: Python for Data Engineering

Learn Python syntax, data types, loops, functions, file handling, OOP, and exception handling.

Unit 2: SQL for Data Analysis & Engineering

Master SELECT, JOINs, GROUP BY, CTEs, subqueries, window functions, indexing, and optimization.

Unit 3: Relational Databases & ER Modeling

Understand database design, normalization, schema creation, and relational database principles.

Track Curriculum

Phase 2: Data Warehousing & ETL Pipelines (8 Modules)

Data Warehousing & ETL Pipelines

Build scalable systems to move and transform large volumes of data efficiently.

Unit 4: Introduction to ETL & Data Pipelines

Understand the ETL process, data pipeline architecture, and orchestration strategies.

Unit 5: Data Warehousing Concepts

Learn about OLAP vs OLTP, star/snowflake schema, and warehouse architecture.

Unit 6: Dimensional Modeling

Design fact and dimension tables for analytical queries.

Unit 7: Working with Apache Airflow

Build, schedule, and monitor data pipelines using DAGs and Airflow operators.

Unit 8: Using Apache NiFi for ETL

Design data flows with NiFi processors and monitor ETL in real-time.

Unit 9: Building ETL with Python & Pandas

Extract, clean, transform, and load data using custom Python scripts.

Unit 10: Data Ingestion with Apache Kafka

Ingest real-time data using Kafka producers and consumers.

Unit 11: Connecting APIs & External Sources

Fetch and load data from APIs, flat files, cloud buckets, and more.

Track Curriculum

Phase 3: Big Data & Distributed Systems (7 Modules)

Big Data & Distributed Systems

Harness the power of distributed systems to store and process massive datasets.

Unit 12: Hadoop Ecosystem Overview

Understand HDFS, MapReduce, and core Hadoop architecture.

Unit 13: Working with HDFS & Hive

Store and query large datasets using Hadoop Distributed File System and Hive.

Unit 14: Apache Spark for Data Processing

Use Spark for in-memory distributed computing with PySpark.

Unit 15: Spark DataFrames & RDDs

Transform and manipulate large datasets efficiently using Spark APIs.

Unit 16: Spark SQL & Spark Streaming

Query structured data and process streaming data in real-time.

Unit 17: Introduction to Databricks

Collaborate and process big data in the cloud with Databricks notebooks.

Unit 18: Data Lake Architecture

Design scalable data lakes using storage formats like Parquet and Delta Lake.

Track Curriculum

Phase 4: Cloud Data Engineering (6 Modules)

Cloud Data Engineering

Move your pipelines to the cloud and work with cloud-native tools.

Unit 19: Introduction to Cloud Platforms (AWS/GCP/Azure)

Understand cloud architecture, billing, and resource management.

Unit 20: Cloud Storage (S3, GCS, Azure Blob)

Store and manage structured/unstructured data in the cloud.

Unit 21: Cloud-Based Databases (BigQuery, Redshift, Snowflake)

Query massive datasets with serverless and MPP database engines.

Unit 22: Serverless Data Pipelines with Cloud Functions

Automate pipelines using event-driven cloud functions.

Unit 23: Using Cloud Composer & Dataflow

Build and monitor workflows with managed Apache Airflow and Dataflow.

Unit 24: Infrastructure as Code with Terraform

Deploy data infrastructure using code (IaC) with Terraform basics.

Track Curriculum

Phase 5: Real-Time Processing & Streaming Analytics (5 Modules)

Real-Time Processing & Streaming Analytics

Design systems that handle real-time events and stream processing at scale.

Unit 25: Streaming vs Batch Processing

Compare architectures and use-cases for batch and real-time data.

Unit 26: Kafka for Real-Time Ingestion

Produce and consume high-throughput streaming data with Kafka.

Unit 27: Apache Flink Basics

Stream process events with Flink's powerful APIs.

Unit 28: Spark Structured Streaming

Process and analyze structured streams in real time.

Unit 29: Monitoring & Observability for Pipelines

Set up logging, metrics, and alerts for production pipelines.

Track Curriculum

Phase 6: Data Governance, Security & CI/CD (5 Modules)

Data Governance, Security & CI/CD

Ensure data reliability, privacy, and maintainability in enterprise-grade pipelines.

Unit 30: Data Quality & Validation Tools

Automate data validation using Great Expectations or custom scripts.

Unit 31: Data Lineage & Cataloging

Track data sources and transformations using tools like Apache Atlas or DataHub.

Unit 32: Security & Access Control

Implement encryption, IAM, and role-based access to data.

Unit 32: Security & Access Control

Implement encryption, IAM, and role-based access to data.

Unit 33: CI/CD for Data Pipelines

Version control, automate testing, and deploy pipelines with GitHub Actions or Jenkins.

Unit 34: Documentation & Best Practices

Maintain project documentation and follow engineering best practices.

Sample Projects

Real-time Data Pipeline for Ride-sharing

Created a pipeline with Kafka, Spark Streaming, and Redshift to process real-time trip data for pricing and availability.

Data Lake Architecture for BFSI

Built a cloud-based data lake using AWS S3, Glue, and Athena for centralized, scalable data access and analysis.

ETL Automation for Logistics

Automated daily ETL workflows with Apache Airflow to ingest, clean, and store package data in a data warehouse.

Batch Job Scheduling for Healthcare Claims

Implemented PySpark-based batch jobs to clean, transform, and load healthcare claims, improving analytics readiness.

Few of our hiring partners

TATA ELECTRONICS

Henkel

amazon

TOYOTA

J.P.Morgan

accenture

AVEVA

Sangeetha mobiles.com

ALTIMETRIK

BANGALORE BIOINNOVATION CENTRE
Sculpting the Future of Bioinnovation

Pfizer

COLAN INFOTECH

GSK

Kmart

lululemon

virtusa
Accelerating Business Outcomes

Initiative

Tech Mahindra

TATA CONSULTANCY SERVICES

T2

Parents At Work

STYLE UNION
NEVER BLEND IN

LEVELAI

GWC DATA.AI
SOLUTION MATTERS

Student Testimonials

Vinoth Kumar

Data Engineer

J.P.Morgan

Valli Raja Sekar

Sr. Data Scientist

TATA CONSULTANCY SERVICES

Rajashekaran

Sr. Data Analyst

accenture

Your Name

Your Role

You can be here

Contact us

TeqCertify

Elevate Your Data Journey

For further details write to us at

info@teqcertify.ai 9944414491