.[REGEX]*
Software Services- Learn To Create

DATA ENGINEERING (BIG-DATA)

Data Engineering (BIG-DATA) Specialization Program

contact@regexsoftware.com

www.regexsoftware.com

ABOUT THE PROGRAM:

With the belief to build a healthy ecosystem as per the Industry Standards REGex Software brings a Training/Internship Program on "Power BI & BigData". We organize Industrial Training/Internship Program for improving the knowledge and skills of the Students/Professionals, so that they can become expert in the field of BigData and get their Dream Job in Software Development Field in Big MNCs.

REGex Software Services's Big-Data program is a valuable resource for both beginners and experts. This specialization program will introduce you to the domain of Data Engineering include Hadoop, Map-Reduce, Apache Spark, Snowflake, ETL pipelines, Kafka Streaming ,SQL, Power BI, Amazon EMR and much more starting from Basics to Advance. If you want to become Data Engineer / Business Analyst, REGex introduce this program for you.

Mode:	**Duration:**	**Weekly Duration:**	**Participants**
Physical (Jaipur) or Online (Google Meet)	8 Months + 6 Months Additional Support	20 Hours Per week	18-20 per Batch

What you will Learn ?

- Python
- SQL
- Hadoop
- Map Reduce
- Delta Lake
- Data Warehousing
- Databricks
- ETL
- Spark
- Airflow
- Snowflake
- Apache Kafka
- No SQL
- Power BI
- AWS

Study Material

- E-Notes
- Assignments per day
- Poll test every week
- Live Video Lectures
- Access of Lecture Videos & Notes
- 24*7 Mentorship Support
- Working on Live Projects

Output

- Help you in Data Analytics Domain
- Able to think out of the box
- Expertise in different Big Data Tools like HDFS, Apache Spark, AWS
- Become expert in multiple technology domains like: Python, SQL, Power BI etc.
- Able to solve many Interview Questions of Top MNCs
- Understand creating Data Insights by connecting data sets, transform & clean the data into data models and then create chars/graphs to provide visuals of the data
- Build Charts/Graphs/Insights of data sets
- Work on multiple data sets
- Become Power BI Analyst / Power BI Developer after completion of this program
- Able to get package in Big MNCs upto 40 LPA

Package Offered So Far

IT Candidates	Non-IT Candidates
Minimum Package: 4 LPA	Minimum Package: 3 LPA
Average Package: 4 – 6 LPA	Average Package: 3.5 – 5 LPA
Overall Highest Package: 39 LPA	Overall Highest Package: 14.5 LPA

EXTRA SESSIONS:

Additional Session on GIT, Linux, Docker, AWS Basics, Jenkins and many more for all students

OUR STUDENTS PLACED // PARTNERSHIP

amazon Microsoft Goldman Sachs intel Google fractal paytm Myntra Adobe IBM
ORACLE vmware HCL hp Honeywell cognizant Capgemini wipro Infosys DELL

PLACEMENT PROCESS:

At **REGex Software**, we are committed to providing a structured and results-driven training approach to ensure your career success.

- 🔹 **Training & Performance Analysis:**
  - Your training will begin from day 1 of your joining, focusing on hands-on learning and practical implementation.
  - Our team will analyze your performance based on assignments, projects and weekly assessments from the **second week onwards** and we will provide **weekly feedback** to help you improve.
- 🔹 **Mandatory Criteria for Placement Opportunities:**

To be eligible for placement opportunities, you must meet the following criteria:

- ✅ **80% attendance** in live training sessions.
- ✅ **80% completion and timely submission** of assignments & projects.
- ✅ **80% attendance** in assessments, including **tests, mock interviews, HR interviews and group discussions**.

- 🔹 **Resume Preparation & Placement Process:**
  - Between **5.5 to 6 months**, our team will provide guidance on **resume building** and evaluate your resumes accordingly.
  - After completing **75-80% of the program**, you will receive details about the placement opportunities based on your **performance and company requirements**.
  - Placement opportunities will be provided **continuously** via **email, calls and WhatsApp groups**, depending on your performance.

- 🔹 **Placement Assurance & Refund Policy (Applicable only for Indian Students Only):**
  - This is a **Placement Assured Program**, with an additional **6-month post-program assistance**.
  - IT Graduates who passed out in 2025 or later (Regular B.tech, BCA, M.tech, MCA programs) are assured a minimum salary package of 4LPA upon placement. However, for IT Graduates who passed out in 2024 or earlier, having gaps in their academics, as well as for Non-IT Graduates (graduates other than regular B.tech, BCA, M.tech, MCA programs), the minimum guaranteed package will be 3LPA.
  - In the event that you have attended & completed at least 80% of the program, submitted and finished at least 80% of the assignments, Tests, Mock Interviews & HR Interview and still do not secure a placement then REGEX will refund your fees with a 9% Annual interest rate. Furthermore, Refunds are applicable only within the first 3 days of the demo period and solely in cases where a specific concern is raised regarding the quality of the learning experience provided. You will receive an official notification email from our team on the third day at 7:30 PM, confirming the completion of your demo period. Requests for a refund of the registration amount must be submitted prior to the issuance of this email. No refund requests will be entertained after this time and Even if you discontinue the program prematurely, you are still obligated to pay the full fee to REGex.
- 🔹 **Our Commitment to Your Success:**

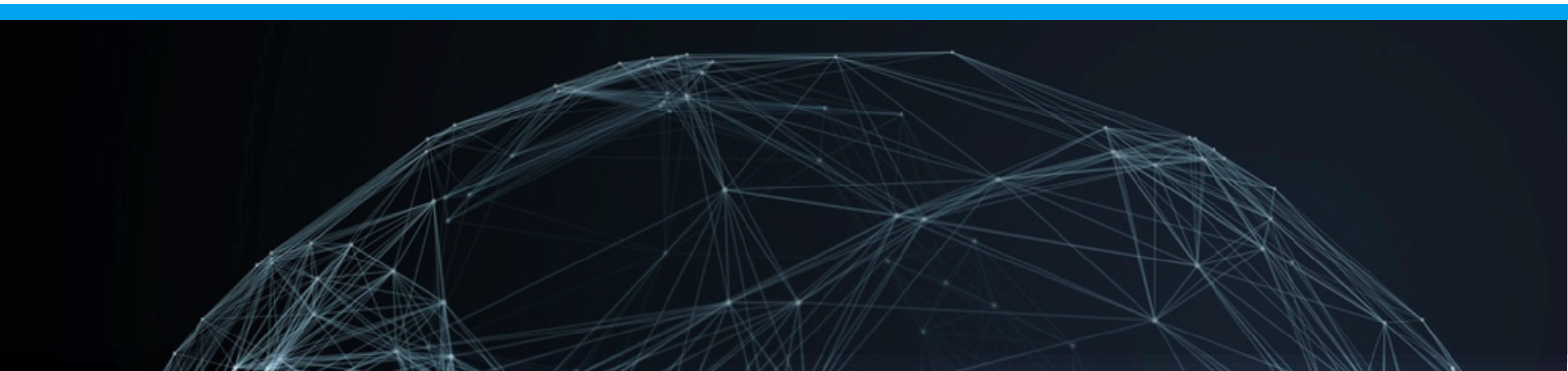
At **REGex Software**, **Placement Assurance = Skills + Opportunities**

.

We equip you with **industry-relevant skills** and provide continuous **job opportunities** based on your performance. However, it is the **student's responsibility to crack interviews** and enhance their skills based on feedback.

For additional support, we offer the flexibility to **rejoin previous batches** to reinforce concepts and improve understanding.

We are dedicated to your career success! 🚀



COURSE CONTENT:

Week 1 - 4

C Logical Programming

- Data Types
- Variables
- Variable Scope – Local, Global
- Constants
- Operators
- Decision Making Statements
  - if Statement
  - if...else
  - switch
- Loops
  - while Loop
  - do...while Loop
  - for Loop
- Basic I/O Functions
  - scanf() and printf() usage

Week 5

Linux

- Basic Commands of linux os
- Vi editor
- Tar Archieve
- User Management and permission
- Grep and advance commands

GitHub

- Git vs GitHub
- Git commands
- Git Branches
- Branching and merging
- Git push vs pull commands

Week 6

Deep Dive into Python Programming

Topics Covered

- **Introduction to Python**
- **Data types**
- **Immutable vs mutable data types**
- **String formatting and slicing**
- **Conditional statements**

Setup and Tools

- **Install Python:** Watch Tutorial
- **Jupyter Notebook:** Setup Guide

Week 7

Deep Dive into Python Programming

Topics Covered

- **Introduction to looping statements**
- **For loop vs while loop**
- **String data type**
- **String manipulation**
- **Sequential data types**
- **List vs tuple**
- **Problem solving with list data type**

Week 8

Deep Dive into Python Programming

Topics Covered

Dictionary and Sets

- In-built functions related to sequential data types
- List comprehension, set comprehension, and dictionary comprehension
- Problem solving with data types

Practice and Problem Solving

- **LeetCode: Two Sum**
- **LeetCode: Contains Duplicate**
- **LeetCode: Two Sum II (Input Array is Sorted)**

Functions and Beyond

- Introduction to functions
- In-built vs user-defined functions
- Types of parameter passing in functions
- Functions vs lambda functions
- Exception handling
- File handling
- Generators and iterators

Week 9

Deep Dive into Python Programming

Topics Covered

- Introduction to OOP concepts
- Creating your first class and objects
- Inheritance, polymorphism, abstraction, and encapsulation
- Use cases of __init__ and __main__
- Advanced topics in Python

Milestone

- Achieve Hackerrank 5 Star

Practice and Problem Solving

- LeetCode: Valid Parentheses
- Extract elements with frequency greater than K – GeeksforGeeks
- Remove multiple empty spaces from string list – GeeksforGeeks

Week 10

Microsoft Excel Walkthrough

Topics:

- Intro to Excel
- HLookup vs VLookup
- Basic & Advanced Functions
- Pivot Tables & Data Visualization

Week 11

SQL Mastery - Basics

Topics:

- **E-R Diagrams & Relationships**
- **Types of Databases (DBMS vs RDBMS)**
- **Installing MySQL**
- **SQL Statement Types**
- **Basic SELECT Queries**

Resources:

- **Install MySQL**
- **Import Sakila DB**
- **Select Statement**
- **Wildcard Characters**

Week 12

SQL Mastery - Intermediate

Topics:

- **SQL Operators**
- **Functions: String, Date, Number**
- **Aggregation & Grouping**
- **WHERE vs HAVING, Sorting**

Resources:

- **MySQL String Functions**
- **SQL Functions (YT)**
- **String Functions (YT)**

Week 13

SQL Mastery - Joins & Subqueries

Topics:

- **Subqueries (Types & Advanced)**
- **Joins: Inner, Outer, Self, Cross, Natural**

Practice:

- SQLZoo Joins
- SQLZoo Self Join

Week 14

SQL Mastery - Constraints & Normalization

Topics:

- **SQL Constraints & Keys**
- **Normal Forms (1NF to BCNF)**
- **DDL: CREATE, DROP, ALTER**

Resources:

- MySQL Flow Control Functions
- Normalization (freeCodeCamp)
- Normalization (Microsoft Docs)

Week 15

SQL Mastery - DML & Control Commands

Topics:

- **DML: INSERT, UPDATE, DELETE**
- **TCL: COMMIT, ROLLBACK, SAVEPOINT**
- **DCL: GRANT, REVOKE**

Milestone: HackerRank SQL Certificate

Week 16

SQL Mastery - Analytical Functions

Topics:

- **Window Functions: PARTITION BY, GROUP BY**

Milestone: HackerRank Gold Badge (5 Stars)

Week 17

SQL Mastery - Advanced Concepts

Topics:

- **Indexes: Clustered vs Non-clustered, Composite vs Column Index**
- **Views: Simple vs Complex Views, Use Cases**
- **SQL Revision**

Week 18

SQL Mastery - PL/SQL

Topics:

- PL/SQL Introduction
- Procedures & Functions
- Conditional Statements & Loops
- Cursors: Types & Use
- DML & Audit Triggers
- Error Handling

Project:

- Library Management System
- Bookstore Analysis
- SQL Advanced Topics Wrap-up

Week 19

Data Engineering with AWS

Topics:

- Cloud Platform Basics
- EC2 vs In-house Servers
- IAM Roles & Users
- S3 Bucket & Access via Python
- AWS Lambda & Triggers

Week 20

Cloud Data Warehousing with Snowflake (Part 1)

Topics:

- What is Cloud Data Warehouse?
- Snowflake Features & Architecture
- Editions & Deployment Options
- File Formats: CSV, JSON, Parquet
- Internal vs External Stages
- SnowSQL CLI & COPY INTO

Week 21

Cloud Data Warehousing with Snowflake (Part 2)

Topics:

- Snowflake SQL: Joins, Aggregations, Subqueries, Window Functions, CTEs
- Temporary Tables
- Streams & Tasks
  - CDC (Change Data Capture)
  - Task Scheduling

Capstone Project:

- Data Warehouse Modeling (Star/Snowflake Schema)

Week 22

Introduction to Big Data with Hadoop

Topics:

- Big Data Concepts (Vs of Big Data)
- Hadoop Ecosystem & Architecture
- HDFS vs MapReduce
- File System Commands
- MapReduce Architecture (MV1 vs MV2)
- WordCount Problem

Week 23

Apache Spark - Modern Data Engineering Tool

Topics:

- Spark vs Hadoop
- Spark Ecosystem
- RDDs: Transformations, Actions, Lazy Evaluation
- SparkSession & SparkContext
- DAG (Directed Acyclic Graph)
- Key-Value Pair Operations
- RDD Caching

Milestone:

- Databricks Certificate: Getting Started with Databricks

Week 24

Data Lakes & Spark SQL

Topics:

- Data Lake vs Delta Lake vs Data Warehouse
- Spark SQL, DataFrames, Dataset API
- File Formats: JSON, Parquet, CSV
- Group By, Aggregation, Joins, Dashboards
- Databricks DII Platform
- Catalyst Optimizer & Tungsten Engine

Projects:

- E-Commerce Sales Analysis
- CDC Project from Database

Week 25

ETL Pipeline & Optimization

Topics:

- ETL using Spark DataFrames
- Data Cleaning & Transformation
- Optimized Queries
- Creating & Using UDFs
- Serialization & Performance Issues

Week 26

Structured Streaming & Kafka

Topics:

- Structured Streaming Overview
- Kafka, File, Socket as Data Sources
- Watermarking & Triggers
- Output Sinks
- Spark Streaming from Kafka
- Offset Management & Fault Tolerance

Week 27

Spark Performance Tuning & Debugging

Topics:

- Job/Stage Analysis with Spark UI
- Caching vs Persistence
- Tuning: Joins, Memory, Partitions
- Final ETL on Large Dataset

Capstone Project:

- Real-time Pipeline for Stock Analysis

Week 28

Kafka – Real-Time Data Processing (Part 1)

Topics:

- Introduction to Kafka
- Kafka Architecture: Brokers, Topics, Producers, Consumers
- Partitions & Replication
- Role of Zookeeper
- Kafka CLI Setup (Local / Confluent)
- Creating Topics
- Producing & Consuming Data using CLI

Week 29

Kafka – Real-Time Data Processing (Part 2)

Topics:

- Writing Data from Spark Streaming to Sinks
  - Delta Tables
  - Parquet Files
  - PostgreSQL
  - S3 Buckets
- Streaming Aggregations

Week 30

Apache Airflow – Workflow Orchestration (Intro)

Topics:

- What is Workflow Orchestration?
- Why Airflow?
- Use Cases: Data Pipelines, ML Pipelines, ELT
- Airflow vs Other Tools
- DAGs, Tasks, Operators
- Scheduler, Webserver, Worker

Week 31

Airflow – Installation & Setup

Topics:

- Airflow Installation: pip / docker-compose
- Setup Airflow UI
- Airflow Home Directory & Config Settings

Week 32

Airflow – Working with Operators

Topics:

- BashOperator, PythonOperator
- DummyOperator, EmailOperator
- BranchPythonOperator

Week 33

Airflow – Cloud & Data Tool Integrations

Topics:

- Working with Cloud Operators:
  - S3, GCS, Azure Blob
- Data Platform Integrations:
  - SnowflakeOperator
  - DatabricksSubmitRunOperator
  - SparkSubmitOperator

Week 34

Airflow – Capstone Project

Project Pipeline:

- Read Data from API
- Save to S3
- Run Spark Job on Databricks
- Write Final Output to Snowflake

Also Includes: Kafka consumer to ingest raw data

Week 35

Power BI – Getting Started

Topics:

- Introduction to Power BI
- Components: Desktop, Service, Mobile
- Power BI Workflow
- Installation & Setup
- Cardinality in Relationships
- Importing from: Excel, CSV, JSON, SQL Server, Web APIs
- Direct Query vs Import Mode
- Data Source Credentials

Week 36

Power BI – Data Modeling

Topics:

- Table Relationships: Cardinality, Cross-Filtering
- Star vs Snowflake Schema
- Applied Steps (Transform Data)
- Column Operations: Split, Merge, Fill
- Filtering & Removing Rows

Week 37

Power BI – DAX & Visualizations

Topics:

- Calculated Columns vs Measures
- Basic DAX Functions: SUM, COUNT, DISTINCTCOUNT, IF, SWITCH, CALCULATE, RELATED
- Visualizations:
    - Bar, Line, Pie, Card, KPI
    - Maps, Matrix, Tables
    - Custom Visuals from Marketplace
- Filters: Visual, Page, Report
- Slicers (Date, Category), Slicer Syncing

Projects:

- HR Attrition Dashboard
- Profit & Loss Statement
- Time-Series Analysis using DAX

JOIN OUR COMMUNITY:

For Frequent Course Updates and Information
Join our Telegram Group

JOIN TELEGRAM

For Webinar Videos and Demo Session,
Join our Youtube Channel

JOIN YOUTUBE

Want to stay updated and inspired?

JOIN INSTAGRAM

Get connected. Stay updated.

JOIN LINKEDIN

Join
100% Placement Guaranteed Programs

MORE INFO & REGISTER